



Széchenyi István Egyetem  
Regionális- és Gazdaságtudományi Doktori Iskola  
Prof. Dr. Rechnitzer János

# Vállalatvezetést támogató üzleti intelligencia alkalmazások sikerkritériumai

---

*Doktori értekezés*

Készítette:  
Galli Richárd

Témavezető:  
Dr. Élő Gábor

2012.

Az értekezésről készült bíráló és a védésről készült jegyzőkönyv a későbbiekben a Széchenyi István Egyetem Tudásmenedzsment Központjában elérhetőek.

## **Nyilatkozat önálló munkáról, hivatkozások átvételéről**

*Alulírott Galli Richard kijelentem, hogy ezt a doktori értekezést magam készítettem és abban csak a megadott forrásokat használtam fel. Minden olyan részt, amelyet szó szerint, vagy azonos tartalomban, de átfogalmazva más forrásból átvettem, egyértelműen a forrás megadásával megjelöltem.*

Győr, 2012. január 20.

Galli Richard

## ***Kivonat***

**Kulcsszavak:** elektronikus üzletvitel, információs rendszerek, vállalatirányítási rendszerek, vállalati információs rendszerek, vezetői információs rendszerek, döntéstámogatás, döntéstámogató rendszerek, üzleti intelligencia, analitika, valós idejű döntéstámogatás, helyszíni döntéstámogatás,

Az üzleti életet – ugyanúgy, mint mindennapi életünket is – döntések sorozata határozza meg. A legtöbb esetben döntéseken (döntések sorozatán) múlik a siker vagy a kudarca. A történelem során hosszú ideig a jó döntéshozók nagyrészt saját intuícióikra támaszkodva döntöttek – gyakran rosszul (persze inkább a sikerekről maradtak fenn legendák).

Mára azonban megváltozott a helyzet: a statisztikai módszerek kidolgozásának és az információtechnológiai megoldások fejlődésének köszönhetően lehetőségek egész sora áll a döntéshozók előtt.

Ma a vállalati döntéstámogatás legmagasabb szintjét az üzleti intelligencia képviseli. Technikák és rendelkezésre álló megoldások egész tárházát nyújtva. Bizonyítja ezt az is, hogy az üzleti intelligencia az elmúlt 15 évben egészen új piaci szektorra nőtte ki magát – sokszereplős, hatalmas pénzüsségeket mozgósító iparággá. Annak ellenére, hogy az üzleti intelligencia egy költséges beruházás, alkalmazása a legtöbb esetben jövedelmező, számos előnyt nyújt.

Disszertációmban elsőként végigtekintem azokat a döntéseméleti, matematikai és statisztikai technikákat és kérdéseket, melyek az üzleti intelligencia tudományának alapját adják, kicsit nagyobb figyelmet szentelve azoknak a statisztikai elemzési technikáknak, melyek az üzleti intelligencia (vagy az adatbányászat) során felhasználásra kerülnek.

Az informatikai kérdések megvitatására ugyancsak sor kerül – úgy a szoftverekkel kapcsolatos elvárásokat tekintve, mint az infrastruktúrával szemben támasztott követelményeket nézve. Bemutatom az üzleti intelligencia történetét, jelen állapotát és várható fejlődési irányait. Az ehhez szükséges kitekintést ugyancsak megadom.

A tézisek választása során igyekeztem olyanokat választani, melyeknek nem csupán elméleti, hanem valós, ipari (üzleti) környezetben előforduló problémákra keresik a választ.

## ***Abstract***

**Keywords:** e-business, information systems, enterprise information systems, management information system, decision support systems, business intelligence, analytics, real-time decision support, on-site decision support,

Business life – just like everybody’s everyday life – is defined by series of decisions. In most cases success or failure depends on good or bad decisions. Throughout the history decision makers had to rely on their intuitions – usually that was a bad thing, but history only kept track of the success-legends.

The situation is pretty much different today: thanks to the development of statistical methods and to the improvement of information technology, a complex range of opportunities are available.

Today the highest level of decision support is business intelligence, making available series of techniques and methods to be used in the world of business. Business intelligence went through a rapid growth and development, becoming an individual sector – despite the heavy costs of adopting business intelligence, it worth the price, in most cases adopting the technology is profitable.

In the dissertation I am going through the basis of business intelligence: decision theory, statistics and mathematics, devoting some extra notice to the techniques used in everyday business intelligence (or datamining) situations.

Disputation of the supporting information technological questions is also present, including both the software side and the infrastructural requirements as well.

I am also presenting the history of business intelligence (and therefore enterprise IT too), today’s state and the available development tracks.

As for the thesis’: I decided to take only those, which are connected to real-life business situations or questions / problems.

## ***Előszó és köszönetnyilvánítás***

Mostanra kialakult véleményem szerint minden dolgozat elkészítése valószínűleg három fázisból áll: a kezdeti hatalmas lelkesedés, amikor az ember nekiugrik a feladatnak és beleássa magát a munkába – ekkor még nem is tudja valójában, hogy mit vállal, csak a témára és az addig megszerzett tudásra támaszkodva; a második fázis feltehetően a csodálkozás és kétségbeesés része, amikor az ember rádöbben, hogy valójában mekkora feladatra vállalkozott és az egyre bővülő tudása miatt úgy érzi, hogy egyre több mindent nem tud – ekkor csökken a lelkesedés, kicsit lassul a munka üteme – majd jön a holtpon, amin átlendülni talán a legnehezebb, viszont ha megvan, akkor a harmadik fázis már ismét gyorsan megy, befejeződni látszik majd összeáll a munka. Ha mindenkinél nem is, de nálam mindenképpen így zajlott, ráadásul szerencsésnek érzem magam, hogy ilyen sok ember támogatott benne. Ettől függetlenül a disszertáció megírása nem volt könnyű feladat, rengeteg időt és fáradságot követelt. Így utólag abszolút úgy érzem: megérte.

Mindenek előtt szeretném megköszönni a lehetőséget – hogy itt dolgozhatok és doktori iskolába járhattam – a munkahelyemnek, a győri Széchenyi István Egyetemnek – és vezetőjének, **Dr. Szekeres Tamásnak** – mert lehetővé vált, hogy napi munkám mellett doktori tanulmányokkal foglalkozzam.

Szeretném megköszönni a lehetőséget a doktori iskola alapítójának és vezetőjének **Dr. Rechnitzer Jánosnak**, akinek köszönhető, hogy doktorandusz lehettem belőlem, és olyan témával foglalkozhattam tanulmányaim alatt, ami számomra is érdekfeszítő és izgalmas volt.

Szeretném továbbá megköszönni a támogatást és lehetőséget tanszékvezetőmnek, **Dr. Sziray Józsefnek**, aki egyfelől lehetővé tette, hogy munkám része is legyen a vállalatirányítási rendszerekkel és a szoftvertechnológiával, szoftverrendszerek tervezésével való foglalkozás, másfelől segítette doktori tanulmányaimat azzal, hogy részt vehettem döntéstámogatással és üzleti intelligenciával kapcsolatos egyetemi programokban.

Köszönettel tartozom még témavezetőmnek – **Dr. Élő Gábornak** – aki jó és gyakorlatias utat mutatott tanulmányaim és a disszertáció elkészítése alatt.

Köszönet jár továbbá mindazoknak a személyeknek – **ismerősöknek, kollégáknak és barátaimnak**, akiknek felsorolására itt nincs lehetőség – akik lehetővé tették és segítették a tanulmányaim és a disszertációm gyakorlati tartalommal való megtöltését, akik vállalatoknál dolgozva segítségemre voltak, és bepillantást engedtek annak működésébe.

Külön köszönet jár a **doktori iskola dolgozóinak és iskolatársaimnak**, a lelkesítésért.

Végül pedig hatalmas köszönettel tartozom **szüleimnek** is, a biztatásért és a lehetőségért. Nélkülük nemhogy befejezni nem sikerült volna, el sem kezdődött volna.

Köszönöm:

*Galli Richard*

Győr, 2012. január 20.

## Tartalomjegyzék

Ábrajegyzék.....	12
Táblázatjegyzék .....	14
1 Bevezetés .....	15
1.1 A kutatás előzményei.....	17
1.2 A kutatás tárgya és célkitűzései .....	18
1.3 A kutatás kerete, módszerei és korlátai.....	19
1.4 A disszertáció felépítése, hipotézisek .....	20
2 Vállalatvezetés és elektronikus üzletvitel .....	35
2.1 Elektronikus üzletvitel .....	36
2.2 Döntéstámogatás és informatika .....	41
2.2.1 Racionalitás és problémamegoldás.....	45
2.2.2 Döntéshozatali folyamat.....	46
2.2.3 Döntések típusai .....	50
2.3 Intelligens vállalat.....	53
2.4 Megállapítások.....	54
3 Üzleti intelligencia.....	56
3.1 Az elemzés ciklikussága .....	58
3.2 A matematika szerepe .....	59
3.2.1 Struktúra .....	60
3.2.2 Modellfejlesztés.....	61
3.2.3 Modellek fajtái.....	64
3.2.4 Eszközök.....	65
3.3 Etikai kérdések.....	66
3.4 Üzleti intelligencia alkalmazások .....	75
3.4.1 Az üzleti intelligencia bevezetése.....	75



3.4.2	Adatintegráció .....	83
3.4.3	Adatbányászat.....	86
3.5	Alkalmazási területek .....	93
4	Felhasználható matematikai és statisztikai eszközök .....	112
4.1	Adatok előkészítése .....	113
4.1.1	Módszertani kérdések .....	115
4.1.1.1	Validáció .....	116
4.1.1.2	Transzformáció.....	117
4.1.1.3	Adattömörítés .....	118
4.1.2	Technikai kérdések .....	121
4.2	Feltárás, elemzés és megjelenítés .....	122
4.2.1	Univariáns feltáró elemzés .....	123
4.2.2	Bivariáns feltáró elemzés.....	127
4.2.3	Multivariáns feltáró elemzés .....	128
4.3	Regresszió.....	129
4.3.1	Regressziós modellek felépítése.....	129
4.3.2	Prediktív változók kiválasztása .....	130
4.3.3	Validáció.....	130
4.4	Idősorok .....	131
4.4.1	Indexszámok.....	132
4.4.2	Idősor modellek .....	133
4.4.3	Komponensek elemzése .....	134
4.4.4	Exponenciális simítás .....	134
4.4.5	Autoregresszív modellek .....	135
4.4.6	Prediktív modellek.....	135
4.5	Minősítés.....	136
4.5.1	Döntési fák.....	137

4.5.2	Bayes-i módszerek.....	138
4.5.3	A k-NN algoritmus .....	140
4.5.4	Logisztikus regresszió .....	141
4.5.5	Neurális hálózatok .....	141
4.5.6	Tartó Vektor gépek.....	143
4.6	Asszociáció .....	144
4.6.1	Szintenként haladó, vízszintesen növelő technikák.....	145
4.6.2	Szintenként haladó, függőlegesen bővítő eljárás.....	147
4.6.3	Mélységben haladó, függőlegesen bővítő eljárás .....	147
4.6.4	Mélységben haladó, vízszintesen növelő technika .....	148
4.6.5	Szabályok .....	148
4.7	Klaszterezés .....	150
4.7.1	Partíció képzés.....	151
4.7.2	Hierarchikus módszerek .....	152
4.7.3	Sűrűségeen alapuló módszerek .....	154
4.7.4	Egyéb módszerek.....	155
4.7.5	Értékelés .....	155
5	Üzleti intelligencia infrastruktúra .....	157
5.1	Technológiai kérdések .....	164
5.1.1	Architektúra .....	165
5.1.2	Adattárházak.....	165
5.1.2.1	Adatpiac .....	167
5.1.3	Infrastruktúra .....	167
5.1.4	Virtualizáció .....	169
5.1.5	Gridok .....	170
5.1.6	Számítási felhők .....	172
5.1.7	Vékony kliens – Vastag kliens .....	174

5.2	Szervezeti és személyi kérdések .....	178
5.2.1	Vállalati hierarchia – BICC .....	183
6	Irodalomjegyzék .....	189
7	Melléklet .....	209

## Ábrajegyzék

1. ábra: Tézisek készítése .....	20
2. ábra: A disszertáció felépítése .....	22
3. ábra: A döntést befolyásoló tényezők.....	42
4. ábra: Döntések típusai a helyzetismeret alapján.....	48
5. ábra: Döntéshozatali folyamat .....	50
6. ábra: Vállalati döntéshozatal .....	52
7. ábra: "Tanuló" döntési modell.....	59
8. ábra: Modellek csoportosítása típus szerint.....	60
9. ábra: A modellezés folyamata .....	61
10. ábra: A modellfejlesztés folyamata .....	62
11. ábra: A problémamegoldás folyamata.....	63
12. ábra: Modellek fajtái.....	64
13. ábra: Absztrakciós piramis .....	69
14. ábra: Az üzleti intelligencia folyamata.....	70
15. ábra: Üzleti intelligencia igényei .....	71
16. ábra: Az adatintegráció folyamata.....	84
17. ábra: 32GB adat tárolása 1980-ban és ma .....	90
18. ábra: A call center kapcsolata az adattárházal .....	91
19. ábra: A mintavétel elve.....	120
20. ábra: Középértékek .....	125
21. ábra. Empírikus eloszlások.....	127
22. ábra: Viszonyszámok.....	132
23. ábra: Döntési fa felépítése .....	137
24. ábra: Bayes-i hálózat .....	139
25. ábra: Neurális hálózat .....	142
26. ábra: Hipersíkok egy tartó vektor gépben .....	144
27. ábra: Asszociációs szabályok helye a döntéshozatalban .....	145
28. ábra: Hierarchia asszociációs szabályokhoz.....	149
29. ábra: Üzleti Intelligencia rendszer felépítése.....	158
30. ábra: Új információ előállítása .....	159
31. ábra: Adattárházak, üzleti intelligencia és üzleti folyamatok kapcsolata.....	160

32. ábra: A modellek életciklusa .....	162
33. ábra: Level-of-support .....	163
34. ábra: Egy paravirtualizált rendszer felépítése.....	170
35. ábra: Vékony kliens hálózat .....	174
36. ábra: Vékony kliensek .....	176
37. ábra: SWOT analízis.....	180
38. ábra: Humán tőke menedzsment .....	182
39. ábra: Információ evolúciós modell .....	184
40. ábra: A BICC aspektusai .....	185

## Táblázatjegyzék

1. táblázat: CRM bevezetés .....	107
2. táblázat: Az analitika alkalmazásának fokozatai .....	209
3. táblázat: Adatok tulajdonságai.....	211
4. táblázat: Humán tőke optimalizáció .....	213

# 1 Bevezetés

*„Megmagyarázatlan életet élni nem érdemes” – Szókratész*

*„Megmagyarázatlan döntést nem érdemes hozni” – Thomas H. Davenport*

Már a '90-es évek óta a gazdasági folyamatokban jelentkező társadalmi-gazdasági környezetet, mint információ- és tudásközösséget említik. A termelési módszerekben és gazdasági kapcsolatokban megjelenő mélyrehatóbb változások miatt megnőtt az immateriális javak mozgatórugójának jelentősége, különösen igaz ez az információra (az helyes információ értéke olykor felbecsülhetetlen). Ennek a változásnak két fő mozgatórugója van. Az egyik a globalizáció, mint az örökké szorosabbá váló gazdasági függés az országok közt, mely lassan eléri azt az állapotot, hogy globális gazdaság néven hivatkozhatunk rá. A második az információtechnológia fejlődése, amit jól jellemez az Internet és a vezeték nélküli kommunikáció térhódítása, mivel már nem okoz gondot a nagy adatmennyiség mozgatása, valamint számos új, kifinomult kommunikációs módra ad lehetőséget. A két tényező persze erősen összefügg: az információtechnológia fejlődése nélkül nem jöhetne létre globalizáció, illetve a globalizáció által felvetett igények nélkül nem fejlődhetne ennyire rohamosan az információtechnológia – ebben a közegben pedig a lehetőségek példátlanok. Az információ és a tudás könnyű elérhetősége minden szereplő számára előnyt jelent. Így az emberek számára, akik gyorsan hozzájuthatnak a hírekhez, szolgáltatásokhoz, elektronikus úton intézhetik (pénz)ügyeiket. A vállalatok számára is kedvező, mivel az információk segítségével olyan termékeket és szolgáltatásokat fejleszthetnek, amik jobban megfelelnek a felhasználók igényeinek, megteremtve ezzel a kompetitív előnyt. Végül kedvező a helyzet a közigazgatásra nézve is, mivel e-kormányzati megoldások segítségével javíthatja szolgáltatásait az állampolgárok felé, elektronikus ügymenetet biztosítva az adminisztrációban, egészségügyben és egyéb ügyek kezelésében.

Ebben a gyorsan formálódó környezetben működő vállalkozásoknál a kapukon belül is jól érezhetőek a hatások: egyre jobban befolyásolja a vállalkozások teljesítményét az információ elérhetősége, illetve az az alapján készülő akciótervek megalkotása. A vállalkozás szó jelentését persze itt némiképp ki kell bővíteni: nem kizárólag a versenyszféra gazdasági szereplőit, hanem vállalkozó személyeket, kis-, közép- és

nagyvállalatokat, kormányügynökségeket, a pénzügyi szektor résztvevőit valamint egyéb (akár non-profit) szervezeteket is jelent.

Az adattárolási technológiák kapacitásának növekedése, árának rohamos csökkenése, valamint a szélessávú Internet és egyéb hálózati lehetőségek elérhetővé válása lehetővé tette, hogy a vállalatnál az évek során szerte felhalmozott adatmennyiség hozzáférhetővé váljon. A vállalatok mára képessé válhatnak ezen adathalmazból információt nyerni, majd azt tudássá változtatni, ami jó alapot jelenthet jövőbeli döntéseik meghozatalához, és ezáltal a kompetitív előny megteremtéséhez. Hasonló elven a mindenkori kormány is kihasználhatja ezt a lehetőséget – elemezheti a rendelkezésre álló információkat – és az állampolgárok javára fordíthatja azt, szolgáltatások formájában. Ezen célok megvalósításához persze a technológia megléte és ismerete még önmagában elégtelen, szükséges a fejlett elemzési technikák és eredményeinek értelmezésének ismerete is. Ezen felül még a döntéseket meghozó személyekre is szükség van.

Az üzleti intelligencia első definíciója tehát így hangzik: olyan matematikai modellek és elemzési technikák összessége, mely szisztematikusan segít az adatokból információ, majd tudás kinyerésére, a döntési folyamatok támogatása végett<sup>1</sup>. A megnevezésben szereplő „üzleti” jelző persze ugyancsak nem kizárólag a versenyszféra gazdasági szereplőire vonatkozik.

Az üzleti intelligencia által használt módszerek több tudományágat ölelnek fel és meglehetősen széleskörű alkalmazást tesznek lehetővé. Valamennyi technika a döntéshozatali folyamatokat vizsgálja és szervezi: az adatok összegyűjtése és tárolása a döntéshozatal könnyítése érdekében, adattárházak képzése, matematikai modellek kialakítása optimalizáció és adatbányászat céljából, valamint sokrétű statisztikai számítások alkalmazása. Az eredményeket a későbbiekben sok helyen, különböző céllal lehet felhasználni: marketing, logisztika, könyvelés, irányítás, pénzügyek, szolgáltatások, közigazgatás...

Kijelenthető, hogy az üzleti intelligencia rendszerek előnyben részesítik a tudományos és racionális alapokon nyugvó vállalkozás-szervezést. A sikerhez persze szükséges a kellő elhatározottság, és a mély folyamatismeret – a szervezet minden szintjén.

---

<sup>1</sup> Howard Dresner 1989-ben fogalmazta meg hasonlóképpen, de arról vita folyik, ki használta először a kifejezést – valószínűsíthető, hogy az első H. P. Luhn volt, 1958-ban.



Egy jó üzleti intelligencia rendszer adatfeldolgozással, matematikai modellek alkalmazásával információkat szolgáltat a döntéshozók számára. Bizonyos esetekben a számítás gyorsabb, könnyebb, kevesebb erőforrást igényel és az eredmény mindössze egy százalékos formában megmutatkozó eredmény lesz, de bonyolultabb esetekben – megfelelően kialakított elemzés és modellek esetén – optimalizáció, induktív tanulás és prediktív modellezés is szerepet játszik.

Általánosan elmondható, hogy a modell a valós rendszernek egy szelektív absztrakciója, egyszerűsítése annak érdekében, hogy a valós rendszer működése könnyebben megérthető és elemezhető legyen. A modell persze a valós rendszer csak azon elemeit tartalmazza, mely a vizsgálat szempontjából fontosnak tekinthető. A klasszikus tudományok – mint például a fizika – már nagyon régóta alkalmazzák a modellalkotást, de az „újabb” tudományok – mint például a statisztika vagy az operációkutatás – is alkalmazzák a matematikai modellek képzését olyan mesterséges rendszerek vizsgálatára, mint a vállalkozások és más összetett szervezetek.

## **1.1 A kutatás előzményei**

Jómagam informatikai szakközépiskola elvégzése után 2007-ben végeztem a Széchenyi István Egyetemen, a Gazdasági Informatika szakon. A képzés lényege az volt, hogy egyesítse a műszaki (mérnök) informatikus és a közgazdasági (gazdasági szakember) képzéseket. Jelentkezésemkor indult a szak, az első évfolyam része lettem, nem ismertem senkit, aki ilyen végzettséggel rendelkezett volna, de a tematika meggyőző volt: egyaránt tartalmazott szoftver és (informatikai) rendszertervezéssel kapcsolatos tárgyakat, valamint gazdasági ismereteket. A két szakterület ismeretei közt gyorsan épült a híd, a szak életképessége igazolásra talált a munkaerő-piaci igények tükrében, így bátran állítható, hogy érdemes a szak az oktatásra és továbbvitelre.

A diploma megszerzése után igazolódott a koncepció életképessége: az ipari / munkaerő piaci igény megvan a gazdaságinformatikusokra. Elsősorban ott lehet sikeres ilyen képzettséggel az ember, ahol gazdasági műveleteket megvalósító, összetett informatikai rendszereket kell tervezni / fejleszteni / karbantartani / felügyelni / üzemeltetni.

Ez a halmaz pedig egybeesik az ERP-rendszerek és az üzleti intelligencia megoldások alkalmazásának helyével. Jobban mondva egybeesett kezdetben, mert mára kiszélesedett

az a termék- és megoldáspaletta, ami ezeket a célokat szolgálja – már a kisebb vállalkozásoknak is elérhető.

Bár a témaválasztás már korábban megtörtént (a témaválasztás akkori okai érdeklődési köröm és addigi szakmai ismereteim), az igazi nagy lendület akkor jött, amikor az Egyetem is bekerült az egyik üzleti intelligencia megoldásokat szállító nemzetközi vállalat által hirdetett Egyetemi Programba.

A program során megismerkedtem a technológia adta lehetőségekkel: azokkal amik már bizonyítottan jók és a vállalatok előszeretettel használják, valamint azokkal a lehetőségekkel, amik adottak ugyan, de még elterjedésre várnak.

## **1.2 A kutatás tárgya és célkitűzései**

A kutatás tárgyát olyan kérdések megválaszolása képezi, amelyek az üzleti intelligencia témakörében aktuálisan felmerülnek úgy Magyarországon, mint világszinten. A kutatásba beletartozik egyfelől a technológia létrejöttét lehetővé tevő matematikai és statisztikai háttér megismerése, másfelől az üzleti intelligenciához (vitatottan, de sokak véleménye szerint szorosan) kapcsolódó támogatás is. Különösen nagy figyelmet kell fordítani az olyan kérdéseknek, amelyek azt címezik, hogy miként lehetséges a technológia használata olcsóbban, esetleg nagyobb sikerarány mellett.

Feltételezésem szerint lehetséges lehet az egyidejű költségcsökkentés és hatékonyságnövelés a vállalatoknál. E kérdés megvitatásához persze itt a vállalatok informatikai lehetőségeit (és azok szervezési oldalát) és hatásait kell nagyító alatt vizsgálni.

Az tényként kezelhető, hogy a szoftverpiacot évtizedek óta nagyon rossz (rendszer) bevezetési sikerarány jellemzi – különösen igaz ez az olyan összetett rendszerekre, mint amilyenek a nagyvállalatoknál használt vállalatirányítási és döntéstámogató rendszerek. Feltételezésem szerint javíthatónak kell lennie az amúgy katasztrofális aránynak<sup>2</sup> (különösen akkor, ha a költségvonzatokat is figyelembe vesszük).

Az is tény, hogy ezek az informatikai rendszerek rendkívül költségesek. Ennek oka az, hogy minden összetevőjük rendkívül drága, úgy a felhasználandó hardver és

---

<sup>2</sup> A legismertebb évről-évre megújuló mérés és hozzá tartozó tanulmány a Standish Group „CHAOS” jelentése.

szoftverelemek, mint a szükséges személyzet tudása. Sok vállalatot (különösen a KV szektorban) már az állítólagos árak hírei elriasztanak egy ekkora átállási / bevezetési projekt elkezdésétől. Ugyan az informatikai beruházásokban rendkívül nagy pénz van, ennek ellenére feltételezésem szerint mégsem szükséges minden esetben a minden tekintetben drága beruházást választani, megoldható lehet a bevezetés némiképp költségkímélőbb módon is.

Feltételezésem szerint ebben a meredeken felfelé ívelő iparágban (az üzleti intelligencia rendszerek területe), mely már-már kényszerítve van a sikerre, és nagyrészt a vállalatokhoz beáramló irdatlan mennyiségű információra támaszkodik – és az elérhető sikerek egyre kisebbek lesznek – lehet még különbséget tenni: a nem számszerűsíthető és minőségi információk felhasználásával. Nehéz kérdés, de reményeim szerint megválaszolható.

Remélem, hogy munkám segít rávilágítani néhány, a témával kapcsolatos gyakori problémák megoldására, és bízom benne, hogy a versenyszférában is használható tudást mutat fel, hozzájárulva a működő és a jövőbeni üzleti intelligenciával kapcsolatos rendszerek működtetéséhez és tervezéséhez.

Céлом persze a kutatás további folytatása, lehetőség szerint szeretném majd a közeljövőben inkább gyakorlatiasabb és úttörőbb síkra terelni annak menetét.

### **1.3 A kutatás kerete, módszerei és korlátai**

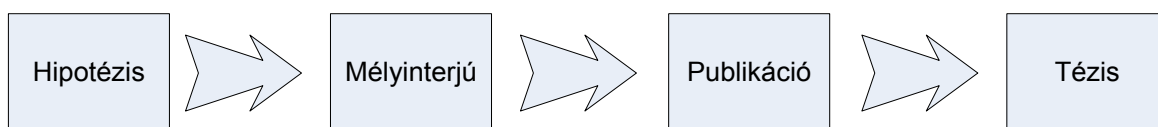
Mivel a választott téma sokrétű – de szakterületem miatt kifejezetten kedvemre való – több területet kell felölelni a bejárásához. A kutatási terület két legfőbb kontextusa a statisztika (és ez által a matematika is), valamint az informatika. Ezek mellett a téma erős etikai és közgazdasági kérdéseket is felvet.

A felhasználható matematikai és statisztikai technikák ismertetésekor igyekeztem a gyakorlatban alkalmazott módszereket ismertetni, saját, illetve vállalati tapasztalatok alapján. Ehhez természetesen szükséges volt a téma irodalmának áttekintése is.

Az informatikai kérdések megvitatásához kollégáim mellett a témát gyakorló szakemberek segítségét is igénybe vettem. A téma informatikai oldalának magyar nyelvű irodalma gyakorlatilag néhány könyvre szűkül, míg a külföldi (elsősorban Egyesült Államokbeli) irodalma meglehetősen szerteágazó – ami persze érthető, hiszen ott van igazi múltja a technológiának.

A gyakorlati problémák és tapasztalatok megvitatásához fel kellett keresnem több céget is, ahonnan a mélyinterjúk nagy része is származik. Ennek kapcsán lehetőségem nyílt eszmecserére informatikai, munkapiaci és tudás-igény tapasztalatokat illetően.

Mivel a disszertáció lényegét a tézisek (azok cáfolata vagy igazolása) teszi ki, ezért azokat hipotézisekből kiindulva alkottam meg [1. ábra]. A hipotézisek felállítása után mélyinterjúkat készítettem több vállalkozás / vállalat munkatársaival, akik jórészt informatikus beosztásban dolgoznak. A mélyinterjúk kiértékelése és eredményeik felmutatása cikkekben, tanulmányokban és előadásokban történt, majd ezen publikációk alapján dolgoztam ki a téziseket.



*forrás: saját munka*

*1. ábra: Tézisek készítése*

A kutatás sikeréhez jelentős mennyiségű kvantitatív adatgyűjtést kellett végezni, ennek jó része az üzleti intelligencia területén vezető pozíciót betöltő cégektől, illetve független elemzőcégektől származik, kiküszöbölve az esetleges marketingcélú túlzásokat.

Helyzetem szerencsésen alakult, mivel nagy mennyiségű kapcsolódó adattal rendelkeztem a már említett részben általam oktatott tárgyak kapcsán. Az adatgyűjtés további részét személyesen végeztem el a mélyinterjúk alatt.

Ezen kívül levelezést folytattam a terület több gyakorló szakemberével, valamint a technológiát alkalmazó cégek több szakemberével.

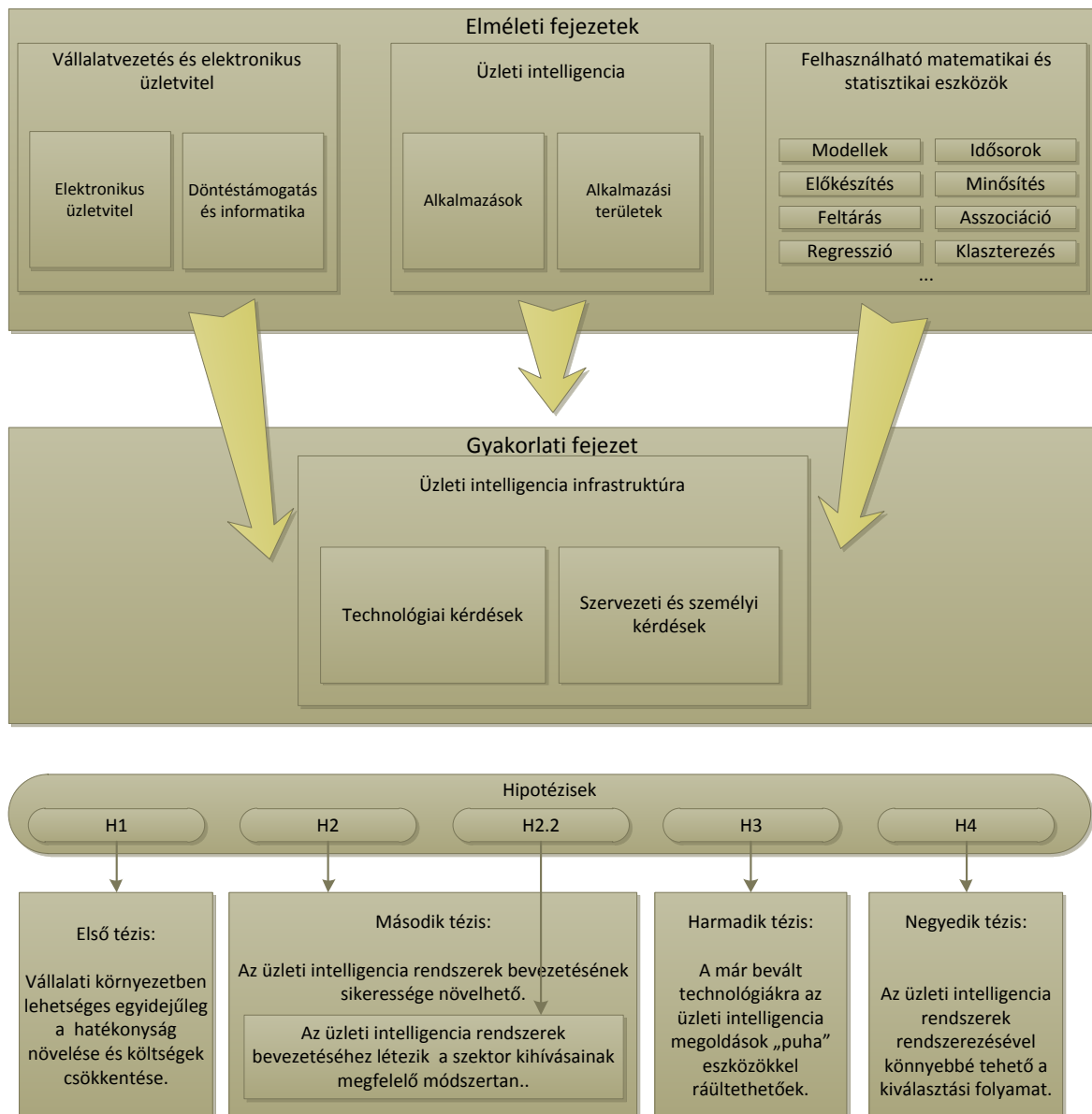
A kutatás során erős korlátnak bizonyult a felkeresett cégek témával kapcsolatos stratégiájának titkossága – persze a körülmények közötti viselkedés érhető ebben a helyzetben, a vállalatok számára létkérdéssé válhat a versenytársak terveinek ismerete (saját érdekeiket is védeniük kell).

## **1.4 A disszertáció felépítése, hipotézisek**

Céлом volt olyan munkát készíteni, mely felépítését tekintve logikus, rávezeti az olvasót a problémákra, majd kellő mélységben kifejtve azt, ismerteti a megoldásokat [2. ábra].

A szakirodalmi áttekintés rész adja a kutatás alapját a dolgozatban – megléte hasznos, rávezeti az olvasót a témára, érezhetővé válik a technológia és a módszertanok fejlődése, megismerhető belőle, hogy milyen úton jutott el oda az üzleti élettel a világ, ahol most van, és mennyire jó lehetőségek is adóttak jelen pillanatban. A matematikai és statisztikai rész segít éreztetni a megoldandó feladatok összetettségét és bonyolultságát, ismeretével jobban értékelhetővé válik a rendelkezésre álló információtechnológia, mely oly erősen meghatározza korunkat.

A tartalom tekintetében igyekeztem az érintett témákat kellő mélységben, ám nem mély informatikai tekintetben ismertetni, inkább a gazdasági összefüggéseket, folyamatokat hangsúlyozom. Így véleményem szerint jobban illik egy gazdaságtudományi doktori iskola profiljába. A kimaradó, mélyebb informatikai részeket igyekszem inkább az oktatás során, publikációkban és tantárgyfejlesztés hasznosítani.



2. ábra: A disszertáció felépítése

A külalakkal kapcsolatban az is célom volt, hogy olyan emberek számára is érthető és valamelyest élvezhető olvasmányt készítek, akik annyira nem járatosak az informatika, információrendszerek és vállalatirányítás témakörében, így a szövegben vagy elmélkedésben felmerülő szakszavak és rövidítések definícióit a lábjegyzetekben ismertetem.

A tézisek kidolgozása során célom volt a felmerülő kérdés (és kapcsolódó probléma) hátterének, okainak ismertetése, majd a kutatás során kialakult megoldás ismertetése. A következtetéseket külön kiemeltem.

## Hipotézisek

A kutatásom során megfogalmazott hipotéziseimet a következő sorrendben prezentálom:

### **H1. Vállalati környezetben lehetséges egyidejűleg a hatékonyság növelése és a költségek csökkentése.**

Az első tézis szerint létezik olyan információtechnológiai megoldás (vagy megoldások), mellyel a két – gyakorlatban sokszor egymással ellentétes – folyamat párhuzamosítható; melynek alkalmazásával a vállalat ily módon kettős haszonra tehet szert. Amennyiben létezik ilyen eszköz (vagy technika), lehetséges-e eredmények elérése tisztán az információtechnológia használatának megváltoztatásával? Ha létezik ilyen, milyen egyéb tényezők segíthetik hatásának erősítését? A tézis kidolgozása során vizsgáltam, hogy melyek azok az eszközök, melyek a vállalatvezetés számára elérhetőek, illetve információtechnológiai vonatkozásúak. Tételezzük fel, hogy igenis lehetséges a hatékonyságnövelés és a költségcsökkentés egyidejűen. Ezt persze rengeteg tényező befolyásolja. Elsőként, szervezeti oldalról befolyásoló tényező a szervezeti hierarchia (azaz, hogy a cég hogyan van felépítve) állapota, az üzleti intelligencia rendszer állapota, illetve az, hogy milyen mértékben használnak a vállalatnál informatikai megoldásokat, valamint azok technológiát tekintve milyenek. A tétel (és cikk) mint eszmefuttatás íródott, annak érdekében, hogy igazoljuk (vagy cáfoljuk) ezt a feltevést. Vizsgálatra kerül, melyek azok a lehetőségek, melyek biztosan egy költségcsökkentésre készülő vállalat elé vetődnek, melyeknek mi az előnye-hátránya, illetve várható hatása. Persze több dolgot kell megvizsgálni: meg kell határozni, hogy a kérdéses folyamatok (költségcsökkentés, hatékonyságnövelés) milyen eszközei érhetőek el, különös tekintettel az informatikai megoldásokra, illetve milyen metrikával mérhető a hatékonyság növekedése, végül pedig feltárni, hogy a kérdéshez hogyan kapcsolódik az üzleti intelligencia alkalmazása (szervezeti, szoftver, hardver és menedzsment terén). A vizsgált megoldások közé tartozik a távmunka (*WESSELÉNYI 1999*) (nagyreszt információtechnológiai értelemben (*BREINERNÉ 2004*)), az üzleti intelligencia alkalmazása (*HOWSON 2007*), a virtualizáció (*SABHERVAL ÉS BECERRA-FERNANDEZ 2010*), a hierarchia átalakítása (*ING 2007 ÉS MILLER,*

*BRAUTIGHAM ÉS GERLACH 2006)* és a kiszervezés (*MILLER, BRAUTIGHAM ÉS GERLACH 2006*).

A feltételezés gyakorlati vizsgálatához mélyinterjúk elvégzésére volt szükség. Az interjúkat olyan vállalatokhoz volt célszerű szervezni, amik jelentős informatikai háttérrel rendelkeznek, azaz a napi feladatok elvégzésében több informatikai rendszer vesz részt, sok számítógéppel, jelentős a napi keletkező adatmennyiség, valamint a nagy infrastruktúra miatt megvan a hozzá szükséges személyzet.

A mélyinterjúkat a vállalatok informatikai szakembereivel és gazdasági (belső gazdálkodásért felelős) szakembereivel folytattam. Több mélyinterjú rávilágított arra, hogy a két – látszólag egymásnak ellentétes – folyamat kivitelezése nem csak lehetséges egyidejűleg, hanem bizonyos esetekben a költségek csökkenése a hatékonyságnövelő stratégia velejárója. Az egyik megnevezhető ilyen stratégiai lépés a virtualizáció használata. (Bizonyos feltételek mellett persze, amennyiben például nem lehetséges számítási felhők igénybe vétele.) A költségek csökkenése alapvetően két irányból érkezik: egyik az infrastruktúra oldalról, mivel magas kihasználtság mellett kevesebb hardverelem szükséges, másfelől személyzeti oldalról kevesebb ember (de különleges képesítéssel) képes ellátni ugyanazt az infrastruktúra-menedzsment feladatot.

A tézis igazolásához kapcsolódó elméleti és gyakorlati eredményeket bemutató cikksorozat első része 2008-ban a Széchenyi István Egyetemen rendezett Kautz Konferencia keretein belül *A hitelválság hatása az informatikai stratégiákra* címmel hangzott el; majd ugyan ebben az évben a székesfehérvári Versenyképesség – Változó Menedzsment / Marketing Tudományos Konferencián *Nagyvállalatok; Az informatikai lépéstartás és költségmegtakarítás lehetőségei* címmel folytatódott előadás és cikk formájában, majd a következő rész 2009-ben komáromi Világméretű pénzügyi és hitelpolitikai válság és Magyarország konferencián *A válság és az információtechnológia kölcsönhatása* című előadásban és cikkben jelent meg; míg a két végső rész a móri IV. és V. KHEOPS Tudományos Konferencián hangoztak el és jelentek meg *Az IT infrastruktúra fejlesztése és a vállalati költségcsökkentés*, valamint egy évre rá *Méreték és technológiák – hatékonyság a főszerepben* címmel. A cikksorozat utolsó része és a tézis következtetése az Alma Mater újságban jelent meg 2011-ben *Paralel hatékonyságnövelés és költségcsökkentés a stratégia és hierarchia tekintetében* címmel.



A vizsgálatok eredménye miatt tehát **a tézis igazoltnak tekinthető**, ám azt a kikötést mindenképp meg kell ejteni, hogy ez csak a jelen korra, a néhány éves időintervallumra igaz – ahogy a technológia fejlődik és ahogy a gazdaság változik, idővel új megoldások és lehetőségek láthatnak napvilágot.

## **H2. Az üzleti intelligencia rendszerek bevezetésének sikerességi aránya növelhető.**

A második tézis szerint az üzleti intelligencia rendszerek bevezetésének sikerességi aránya növelhető. A hipotézis ellenőrzéséhez be fogom mutatni, hogy milyen általánosan alkalmazható szoftverfejlesztési és szoftverrendszer bevezetési módszerek állnak rendelkezésre – különös tekintettel az irányítási kérdésekre. A bevezetési arány meglehetősen rossz minden területen, nem kivétel ez alól az üzleti intelligencia sem, lennie kell azonban módnak arra, hogy ezt a területet ebből a szempontból ne általánosítva, hanem specializált eszközökkel kezeljük. Az üzleti célú szoftverek fejlesztése teszi ki a piacon lévő szoftverek legnagyobb részét (persze vannak ezek között nem kizárólag üzleti célokra használható szoftverek). Fejlesztési feladatként fogható fel sok bevezetési / tesztelési feladat is, mivel egy nagyon hasonló folyamatot kell végigcsinálnia a teamnek – a követelmények felmérése, a helyzet és a problémák elemzése, a tervezés, a kivitelezés és a tesztelés majd az üzemeltetés / karbantartás nagyon hasonló folyamat mindkét esetben – a különbség leginkább úgy fogalmazható meg, hogy egy ilyen (üzleti) rendszer bevezetésénél már sokkal nagyobb mennyiségű a kész részek száma, mint egy „zöldmezős” esetben. Általánosan elmondható emellett, hogy a szoftverek fejlesztése összetett probléma – több tudományterületet kell egyszerre ismerni, azok eredményeit alkalmazni. Általában a projektek átfutási ideje és költségvetése előre szabott, jól kiszámított. Mindezek ellenére a statisztika lehangoló képet fest az üzleti intelligencia bevezetési projektekről. Felmérések szerint a projektek kevesebb mint harmada tekinthető sikeresnek (*WILSON 1998*), majdnem negyedük törölve lesz – a fent maradó 50% körüli rész pedig a kihívásokkal küzdő projekteket jelöli, ez nagyjából azt jelenti, hogy a termék elkészül ugyan és átadásra kerül, de komoly késéssel, komoly költség-túllépéssel, vagy ezek elkerülése végett komoly csonkításokkal (*PERKS 2003*). Sikeresnek tehát akkor tekinthető bármilyen projekt, ha idő és költségkereten belül, az eredeti felhasználói / üzleti igények alapján készül el (*WHITTAKER 1999*). A sikertelenségnek persze több oka is lehet:

ezek a rendszerek már olyan nagyméretűek, és olyan magas komplexitással bírnak, hogy elkészítésükhöz meglehetősen nagy csapatra van szükség. A nagy csapat (nagy team) önmagában nagyobb kockázatot hordoz, mivel több ember között kell a kommunikációt megoldani, több dokumentum keletkezik, ezek pedig sok hibalehetőséget jelentenek. Tehát a kihívás kettős: szervezési és informatikai.

## **H2.1. Az üzleti intelligencia rendszerek bevezetéséhez létezik megfelelő, a szektor kihívásainak megfelelő módszertan.**

A második hipotézis (altézise) egyfajta igazolása lehet annak a meglátásnak az igazolása, mi szerint egy hatékony módszertani keretrendszer alkalmazása növelheti az üzleti intelligencia rendszerek bevezetési hatékonyságát. A hipotézis ellenőrzéséhez be fogom mutatni a legmegfelelőbb specializált módszertant (*HOWSON 2007 ÉS MOSS, ATRE 2003*), majd gyakorlati eset során igazolom annak helyességét, illetve amennyiben szükséges, módosításokat javaslok. Az üzleti intelligencia alapú döntéstámogatás meglehetősen drága. Az összeegyeztethetetlen adatokat ki kell szűrni, majd mindent egyesíteni az on-line tranzakció feldolgozó rendszerekből (OLTP), a batch feldolgozó rendszerekből és más külső adatforrásokból. A sikerhez a legtöbb esetben új technológiák használatára is szükség van, rengeteg új feladatot kell ellátni, új szerepeket és felelősségeket kell betölteni, ráadásul ezeket a megoldásokat kellő sebességgel kell adoptálni, mialatt az általános minőséget is elfogadható szinten kell tartani (adatminőség, folyamatminőség). Hihetetlennek tűnik, de az üzleti intelligencia bevezetés projektek mintegy 60%-a sikertelenséggel vagy félbehagyottsággal végződik (*STEPANEK 2005*). A nagy arány okai elsődlegesen az elégtelen tervezés, a kimaradt logikai lépések, a határidők betartatlansága, a követelmények nem (teljeskörű) teljesülése vagy az általánosan gyenge minőségű átadott rendszer (*EWUSI-MENSAH 2003*). Ezek az indokok nagyon jól mutatják, hogy végeredményképpen itt is szoftverprojektekről van szó – nagyon hasonló arány, és hasonló indokok vezetnek az egyéb üzleti célú szoftver bevezetések sikertelenségéhez. A megfigyelt jelenségek nagyrészt megegyeznek a George Stepanek (*STEPANEK 2005*), Frederick P. Brooks Jr. (*BROOKS 1995*) és Kweku Ewusi-Mensah (*EWUSI-MENSAH 2003*) által megfigyelttel. Már megfogalmazott, de kiemelendő jelenség a fejlesztők és kivitelezők technológiai függése, azaz az általuk támogatott, preferált és a projekt kezdetéig sokrétűen (lehet, hogy addig sem minden esetben helyesen) alkalmazott

informatikai megoldás problémára-igazítása és azzal a technológiával történő megoldása (a külföldi szakirodalomban Silver-Bullet néven hivatkozzák a jelenséget (*BROOKS 1995*), míg Magyarországon inkább a „ha valakinek van egy jó kalapácsa, akkor mindent szögnek néz” mondással utalnak rá). Persze üzleti intelligencia szoftvereknél a jelenség némiképp árnyaltabb, de megtalálható. Több esetben elhangzott, hogy anti-pattern megoldások is fejlesztésre és bevezetésre kerültek, ennek következtében a folyamatok szervezése és az információrendszerek fejlesztése területén. Az jelenség egy speciális, megfigyelt változata az, amikor a fejlesztő / kivitelező cég ragaszkodik saját termék fejlesztéséhez, ahelyett, hogy a céloknak tökéletesen megfelelő, már piacon lévő és testre szabható termék megvásárlása és alkalmazása mellett döntene. Ez a cselekmény jelentős anyagi és időbeli hátrányt okoz a megrendelőnek – mindez gyakran a funkcionalitás rovására megy. Megfigyelhető volt több esetben a nem megfelelő projektvezetési koncepció alkalmazása – mely tapasztalatlanságból adódó személyi probléma – amibe bele tartozott a nem megfelelő fejlesztési módszertan kiválasztása és követése, így aztán a későbbiekben sokkal nehezebb volt a változó igényekhez való igazodás, mely jelentős időhátralékot okozott a projektben. A jelenség oka több esetben az volt, hogy a szakterülethez értő, de projektvezetésben és az informatika technológiai vonatkozásaiban nem jártas személy (vagy személyek) vezették a projektet. A vezetés nem megfelelő mértékű elhivatottsága megjelent ugyan, mint probléma, de nem számít jelentősnek a felmérés során. A vizsgálat azt mutatja, hogy ez a probléma gyorsan és jól kezelhetővé vált az üzleti intelligencia rendszerek esetében. Ennek az oka az, hogy legtöbb esetben az ösztönzés a fejlesztésre a felső vezetésen át érkezik, így sok esetben ők már eleve meggyőződtek annak szükségességéről és a várható haszon jellegéről és mértékéről. Az elemzésből kiderül továbbá, hogy a fejlesztést kivitelező team-ek összetétele kielégítő, legtöbb esetben a coaching<sup>3</sup> jól alkalmazott, a tervek megvalósításához a csapatok tagjai szükséges technikai és technológiai ismeretekkel rendelkeznek (*KIS 2010*). Ugyanakkor ezzel ellentételesen jellemzőek a projektetek szervezésének nehézségei. Ennek oka kettős: részben a már említett személyi döntések miatt, részben pedig az alkalmazott módszertan (vagy projektmenetrend) miatt. Főként ott jellemző az utóbbi, ahol nem tettek szert megfelelő tapasztalatra a fejlesztési projekt vezetői, mivel a vizsgált fejlesztés az egyik kezdő projektjük volt. Ugyanez igaz egy magasabb szinten is: az olyan cégek, akik új területre

---

<sup>3</sup> Coaching: olyan, vállalatoknál alkalmazott személyzeti (és menedzsment) technika, mely célja a tudás-átadás révén a teljesítmény növelése, a kihívások felfedezése, a munka közbeni tanulás és az öröm felfedezése a munkában.

próbálnak bejutni, és olyan fejlesztést vállalnak, amiben előtte nem szereztek gyakorlatot – nagy valószínűséggel problémák merülnek fel (tipikusan ilyen a vállalatirányítási rendszerek, eszközök vagy üzleti intelligencia rendszerek területe). Kirívóan nagy figyelmet kaptak a projektek igazolásával (azaz, hogy érdemes-e elindítani egyáltalán a projektet) kapcsolatos problémák (*EWUSI-MENSAH 2003*). Sok esetben hiányzik, vagy nem áll rendelkezésre elegendő információ az igazolás egészéhez, annak ellenére, hogy ez egy fontos döntési pont lehet. Ugyanígy hiány merült fel sok esetben a kockázatok elemzésével kapcsolatban – mely egy sor további hiányosságra utal például a BCP<sup>4</sup>, DRP<sup>5</sup> területeken. A kockázatok elemzése a vizsgálat szerint sok esetben elmarad, és a fennmaradó néhány esetben is csak kezdetleges módon van elvégezve. A kockázatelemzés eredménye persze a cégen belül több helyen felhasználható lenne a későbbiekben. A projektek kapcsán a helyzetértékelés is komoly hiányosságokat mutat, pedig üzleti intelligencia rendszerek fejlesztése és bevezetése kapcsán különösen nagy figyelmet érdemelne.

A hipotézis vizsgálatához megint mélyinterjúk készítéséhez kellett folyamodni. Az interjú eredményességéhez olyan vállalatoknál dolgozó szoftvermérnökök véleményére kellett rákérdezni, akik már vettek részt cégüknél üzleti intelligencia rendszerek bevezetésében (vagy olyan tanácsadók, akik kifejezetten erre specializálódtak) és így kellő rálátással lehetnek a felmerülő problémákra.

Az elméleti kutatást alátámasztották a mélyinterjúk; megállapítható, hogy ahhoz, hogy egy üzleti intelligencia rendszer bevezetés hatékony legyen, a következő felkészülésre van szükség: Először biztosítani kell, hogy rendelkezésünkre álljanak megfelelő metrikák és mérési pontok – kettős céllal: egyfelől meg kell határozni, hogy mely metrikákon szeretnénk javítani a rendszer bevezetésével, másfelől a bevezetés sikerességének kiértékelése ezek alapján történik. Másodszor igazolni kell, hogy a beruházás / erőfeszítés megtérül, meg kell határozni az üzleti lehetőségeket és az üzleti hajtóerőt. Harmadszor pedig pontos tervet kell készíteni a kivitelezéshez, megfelelő párhuzamosításokkal, pontos felelősség-kiosztással.

---

<sup>4</sup> BCP: Business Continuity Planning, azaz üzletmenet-folytonosság tervezés. Olyan rendszerszeméletet igénylő tervezési eljárás, melynek célja felkészülni az üzletmenetet megzavarni képes hatásokra, kivédeni azokat, vagy csökkenteni hatásukat.

<sup>5</sup> DRP: Disaster Recovery Planning, azaz katasztrófa bekövetkezés utáni helyreállítás tervezés. Olyan tervezési és megvalósítási technikák összességét jelenti, melyek célja valamilyen ügymenetet megzavaró esemény bekövetkezte utáni helyreállítás rendezett és lehető leggyorsabb módon történő kivitelezése.

A rendelkezésre álló elméleti anyagokból leszűrt lehetőségek és a mélyinterjúk eredményei először 2010-ben a Szatmárnémetiben megrendezett XX. SZÁMOKT – Számítástechnika és oktatás Tudományos Konferencián kerültek bemutatásra a *Software technology: Specialized development / launch method for BI applications* című előadásban és kapcsolódó cikkben. (A vizsgálatához csak felületesen kapcsolódó, de a szakemberek képzésének és a széles látókörök kialakítását hangsúlyozza a XIX. SZÁMOKT – Számítástechnika az oktatásban konferencián, Marosvásárhelyen elhangzott *SAS Egyetemi Program belülről – üzleti intelligencia oktatási tapasztalatok* című előadás és a kapcsolódó cikk.) A speciális rendszer-bevezetési módszer két cikkben, *Felvetések az üzleti intelligencia rendszerek bevezetéséhez*; valamint *Módszerek az üzleti intelligencia rendszerek sikeres bevezetéséhez* kerül ismertetésre. A vizsgálat kulcslépéseit a *Vállalatvezetést támogató üzleti intelligencia* című tankönyv tartalmazza.

A vizsgálat bizonyította tehát, hogy a feltüntetett BI ütemterv megoldást nyújt a tipikusan ilyen rendszerek esetén jelentkező problémákra. (Ellenben vetett fel egy újat is, ami miatt inkább egy személyes módosítás utáni változat javasolható.)

A fenti okfejtés miatt **a második tézis részben igazolt**, ám véleményem szerint további megerősítés szükséges, ám nem mélyinterjú, hanem célzott, de szélesebb szakembercsoportnak szóló kérdőív segítségével (kérdéses persze, hogy lehetséges-e kellő kitöltés összegyűjtése). Továbbá attól még, hogy egy módszer sikerre visz egy ilyen projektet, nem zárható ki, hogy létezik más, ennél sokkal hatékonyabb megoldás.

### **H3. A már sikerrel alkalmazott technológiákra az üzleti intelligencia rendszerek „puha” eszközökkel telepíthetőek.**

Általános probléma (a gazdasági válság miatt kiemelt figyelmet kapott) az üzleti intelligencia rendszerek bevezetése során azok forrásrendszereivel való kapcsolatuk, és az ehhez kapcsolódó hardver-beszerzések magas ára. Azonban szükséges-e a sok esetben rendkívüli hardverteljesítmény megvásárlása, vagy egyszerűbb megoldások is léteznek, és az üzleti intelligencia rendszerek „puha” eszközökkel rátelepíthetőek (ráültethetőek) a már meglévő rendszerekre? A igazoláshoz sorra fogom venni az elérhető üzleti intelligencia rendszereket, a hozzájuk ajánlott hardverkonfigurációt, illetve annak lehetőségét, hogy a

ráültetés kivitelezhető-e. A tézis alapfeltevése az, hogy lehetséges a legújabb, legmodernebb üzleti intelligencia megoldások adaptálása egy olyan környezetbe, ahol már meglévő infrastruktúra üzemel – hardvert és szoftvert beleértve. A feltételezés azért különösen aktuális, mivel számtalan tanulmány lát napvilágot olyan témákkal, hogy hogyan célszerű, hogyan költséghatékony egy vállalati informatikai rendszer felépítése, komoly üzleti intelligencia megoldások alkalmazása mellett. Ugyanakkor a tanulmányok nagy része figyelmen kívül hagyja azokat a tényezőket, melyek igencsak jellemzőek a vállalati rétegre. Az első tényező, mi szerint a vállalat már régóta üzemel, és az üzleti intelligencia bevezetése alatt is üzemelnie kell – általában egyáltalán nem, vagy csak rendkívül korlátozott mértékű leállás engedhető meg. A második tényező az, hogy a vállalati rendszerek meglehetősen heterogének hardver és szoftver megoldások terén, azaz sokféle hardver dolgozik együtt, és sok fajta szoftver van használatban (*HODGE 2004*). Ennek oka általában a vállalatok evolúciós fejlődésében található, ugyanis sok cégnél megtartják a régi rendszereket – vagy mert muszáj, annak célja miatt, vagy mert nincs pénz illetve erőforrás annak kiváltására. Az üzleti intelligencia bevezetési projektek tervezésekor (*BUSSLER 2007*) ezeknek a tényezőknek külön figyelmet kell szentelni (*BURSTEIN 2008*) – mely a szoftvermérnököket gyakran válaszüti elé állítja. A jóval költségesebb megoldás az, ha a bevezetés egyben bizonyos szintű átszervezést és modernizálást is jelent, ekkor az alkalmazott megoldások átgondolása is szerepet kap. A költségkímélőbb megoldás az, ha megvizsgáljuk, hogy a már meglévő infrastruktúra képes-e ellátni a neki szánt többletfeladatot (*COSTA – SUBIRANA – VALOR – VALOR ÉS MENKEN – BLOKDIJK 2008*), és ha igen, akkor ráültethető az üzleti intelligencia megoldás. Ám ha nem, akkor meg kell vizsgálni, milyen lehetőségek vannak világszerte, melyeket segítségül lehetne hívni. Ahhoz persze, hogy a tervező csapat számolni tudjon az infrastruktúrával szembeni elvárásokkal, már a tervezés korai szakaszában ismerniük kell azokat a követelményeket, melyeknek a készülő üzleti intelligencia rendszer meg kell feleljen.

A tézis igazolásához (vagy elvetéséhez) elsőként össze kellett gyűjteni az összes olyan, a világban előforduló szoftverrendszert (és szoftvert), mely üzleti intelligencia megoldást (vagy alkalmazható részmegoldást) nyújt, majd egyesével meg kellett vizsgálni képességeiket. A vizsgálat 54 üzleti intelligencia funkcionális rendszerre terjedt ki. A vizsgált rendszerek nagy része kereskedelmi forgalomban kapható, de a valóságnak megfelelően tartalmaz a vizsgálat ingyenes, illetve nyílt forráskódú termékeket. A vizsgálat

egyfelől bebizonyította, hogy ebben a szektorban is folyamatban van a klasszikus kliens-szerver architektúra elhagyása és webes architektúrára áttérés. Azok a rendszerek amelyek vagy a klasszikus kliens-szerver architektúrára, vagy desktop környezetre épülnek, (kliensoldalt tekintve) leginkább az elterjedt keretrendszereket alkalmazzák, amit a JAVA technológia vagy a .NET keretrendszer nyújt. A megállapítás pozitív oldala az, hogy mivel mindkét keretrendszer elérhető az elterjedt operációs rendszereken, így ezen az oldalon nem jelentkezhethet probléma a használatot illetően. A meglévő szervereken futó operációs rendszerek tekintetében (amennyiben új hardver illetve operációs rendszer is kerül beszerzésre az üzleti intelligencia rendszer bevezetésekor, az nagyobb árkategóriába sorolja ugyan a beszerzést, ám nem is jelentkezik a megkötés a rendszer architektúrájának tervezése során) a termékek rugalmassága más képet fest. Ebben a tekintetben könnyedén osztható két részre a „mezőny”, a csoportosításnak jó alap az, hogy szerveroldalon több operációs rendszer támogatott-e. A nagyobb, összetettebb rendszerek (főként azok, melyek tekintélyes több évtizedes múltra tekintenek vissza), evolúciójuknak köszönhetően általában támogatják az iparban elterjedt szerver operációs rendszereket. A másik csoportba a nem összetett, funkcionalitásukat tekintve szegényesebb szoftverrendszerek tartoznak, melyek csak egy-egy részfeladat megoldására alkalmasak, és nem rendelkeznek ekkora rugalmassággal sem telepíthetőség tekintetében. A vizsgálat eredményeként kijelenthető, hogy a vizsgált rendszerek 44,4%-a képes valamilyen gyártó-specifikus adatforrás hasznosítására – ez nem előnyös tulajdonság összekapcsolhatóság szempontból, de 72% képes XML adatforrás hasznosítására és a vizsgált megoldások 54%-a képes ODBC adatforrások hasznosítására (a vizsgált megoldások 25,9%-a íródott JAVA nyelven, ezek mindegyike képes JDBC adatforrások használatára).

A kutatás és vizsgálatok eredményeit – így a tézis igazolását is – A 2011 januárjában a Híradástechnika újságban megjelent *Már alkalmazott technológiák és BI megoldások összehangolása* címmel megjelent cikk tartalmazza. (Felületesen kapcsolódik a már említett Alma Mater cikk, *Paralel hatékonyságnövelés és költségcsökkentés a stratégia és hierarchia tekintetében*; valamint technológiai kérdésekben a *Vállalatvezetést támogató üzleti intelligencia* című tankönyv is.)

A ismertetett vizsgálat eredményei miatt **a tézis igazoltnak tekinthető.**

#### **H4. Az üzleti intelligencia rendszerek rendszerezésével könnyebbé tehető a kiválasztási folyamat.**

Az előző tézis gondolatmenetét folytatva – és a munkahelyi vita ajánlását követve – felvetődött a kérdés, hogy lehetséges volna-e az üzleti intelligencia rendszerek kategorizálása annak érdekében, hogy egy felvetőtő bevezetési igény esetén gyorsan le lehessen szűkíteni a szóba jövő eszközök körét. A rendszerezést persze nem infrastrukturális és kompatibilitási, hanem funkcionális oldalról lenne célszerű elvégezni.

A taxonómia egy hierarchikus osztályozási vagy rendezési elvet jelent, mely történhet tetszőleges számú dimenzió mentén. A taxonómiával kapcsolatos fontos elvárás az, hogy a leírt tér minden eleme elhelyezhető legyen benne, azaz jelen esetben: minden üzleti intelligencia eszköz (meglévő és jövőben fejlesztett) elhelyezhető legyen benne.

Amennyiben pedig minden megoldás elhelyezhető egy ilyen taxonómiában, akkor az felhasználható új rendszerek bevezetése előtt az alternatívák vizsgálatához, vagyis: készíthető segítségével egy ajánlás, mely néhány kezdeti információ megadásával jelentősen szűkíteni képes a szóba jövő megoldások körét.

Mivel az üzleti intelligencia rendszerek az elmúlt és jelen évtized egyik legfelkapottabb rendszerei (szoftverei), ezért még a rendszerezésükkel kapcsolatos publikációk száma nagyon nagyra nőtt. A szám persze jelentősen csökken, ha eltekintünk a marketing-célú írásoktól. Kiemelkedik azonban a Colin White taxonómia, mely az egyik legtöbbet hivatkozott és mai napig tartós besorolható segítségével minden rendszer.

A fent leírt rendszerezés nagyon jó volt, ám a készítése óta eltelt 8 év alatt számos új üzleti intelligencia rendszer és alkalmazás látott napvilágot, a verseny azóta is erősen élénkül ebben a piaci szegmensben és az ezt követő piaci átrendeződés miatt a csoportok határai már nem egyértelműek, valamint nagyon sok olyan összetett termék került piacra, mely a definiált tér nagy részét lefedi. Ám az egyszerűsítésre okot adó tény az, hogy a trendeknek (és a felhasználható technológiák árának csökkenésének) hála több dimenzió egymással ok-okozati kapcsolatba került, így egy egyszerűbb, gyorsabban használható taxonómia megalkotásához tehát célszerű a vizsgálati dimenziók csökkentése.

Az eredeti hat dimenziót (időtáv, latencia, finomság, adattárak, szabályok és típus szerint) az elemzés során sikerült háromra csökkenteni. A legfontosabb kérdés egy rendszer funkcionalitását illetően ebben a kérdéskörben az, hogy milyen időtávra szóló döntések



meghozatalát képes segíteni, azaz stratégiai, taktikai vagy operatív célokat szolgál-e – így az első dimenzió az időtáv (ez persze összefügg az elvárt funkcionalitással, az alkalmazható adattárakkal és szabályokkal is). A második a latencia, azaz, hogy a rendszer valós idejű, közel valós idejű, vagy történeti adatok használatára képes (itt is összefüggés állapítható meg egyrészt a funkcionalitás miatt kis mértékben az időtávval, valamint az adattárakkal). A harmadik dimenzió újdonság: ez tartalmazza a termék kereskedelmi megjelenését, vagyis azt, hogy egy személyre szabott egyszeri, in-house fejlesztés eredményéről, vagy kereskedelmi forgalomban kapható termékről (COTS – Commercial Off-The-Self, vagyis „polcra levehető”) beszélünk – utóbbi három kategóriára bontható: Kereskedelmi termék, mely mindössze csatlakoztatható, kereskedelmi termék, mely mély beállításokkal rendelkezik, kereskedelmi termék, mely fejleszthető. A dimenzió persze összefügg a funkcionalitással, a humán-erőforrás igényvel és a megoldás által nyújtott lehetőségekkel.

Az elemzés egyik mellékes eredménye, a megoldások átlagos ára (TCO számítással kapható) és a taxonómiában elfoglalt hely közti összefüggés, így kijelenthető, hogy egy üzleti intelligencia megoldás annál költségesebb, minél hosszabb távú döntések támogatására képes, minél közelebb képes kerülni a valós idejűséghez és minél nagyobb fejlesztési lehetőség jár vele (természetesen az in-house megoldások a legköltségesebbek).

A 2011 januárjában a Híradástechnika újságban megjelent *Már alkalmazott technológiák és BI megoldások összehangolása* címmel megjelent cikk folytatása a 2012 első negyedévi Acta Technica Jaurinesis tartalmazza, *Taxonómia a már alkalmazott technológiák és új üzleti intelligencia megoldások összehangolásához*. (Felületesen kapcsolódik (a mélyinterjú elvégzése hozzájárult a tézis végkifejletéhez) a már említett Alma Mater cikk, *Paralel hatékonyságnövelés és költségcsökkentés a stratégia és hierarchia tekintetében*; valamint technológiai kérdésekben a *Vállalatvezetést támogató üzleti intelligencia* című tankönyv is.)

A vizsgálatban részt vevő 54 üzleti intelligencia alkalmazás mindegyike besorolható az így előálló taxonómiába. Interjúk alkalmával kérdésként hangzott el, hogy hasznos lenne-e a teljes vizsgálat interaktív publikálása, megkönnyítené-e egy esetleges beszerzési / bevezetési eljárás megindítását annak korai szakaszában? A válaszok ~70%-a egyértelműen igen volt, azzal a kikötéssel persze, hogy mivel a három feltétel alapján történő szűrés még így is jelentős mennyiségű (eltérő és azonos funkcionalitású) terméket takar, további

elemzésre, megvalósíthatósági tanulmányokra (itt játszik szerepet a cikk második részében boncolgatott kompatibilitási kérdéskör) és azok alapján bevezetési tervekre van szükség, hogy kiderüljön, az adott vállalat céljait hogyan képesek a termékekkel megvalósítani. Ennek ellenére hasznos lenne a teljes vizsgálat online publikálása és mindenki által böngészhetővé tétele, segítségével gyorsan következtethetők a lehetséges alternatívák, a szükséges előfeltételek és erőforrások.

A ismertetett vizsgálat és interjúk eredményei miatt **a tézis igazoltnak tekinthető.**

## 2 Vállalatvezetés és elektronikus üzletvitel

A vállalatirányítás menedzsment. Ám a menedzsment története indokoltan jóval régebbre nyúlik vissza: már az ókorból maradtak fenn olyan emlékek, melyek létrehozása rendkívüli szervezőkészségről tanúskodik, elegendő csupán a piramisépítésekre gondolni, ott is több százezer ember napi munkáját kellett koordinálni a cél elérése érdekében. A korai kezdetek ellenére a menedzsment története négy szakaszra bontható.

Az első szakaszt a tudományosság előtti időszaknak szokták nevezni. A korszak lényege, hogy nem igazán volt tudományos alapú kiválasztás sem a szervezői, sem a végrehajtói oldalon: nagyrészt származási alapon válhatott valaki vezetővé. Idővel persze a jól bevált módszereket rögzítették, ami az iparosodás elengedhetetlen kelléke volt, kialakulhatott a tervezés és a kiválasztás gyakorlata is. A szakasz egészen 1880-ig tartott, lezárulásához nagy mértékben hozzájárult Eli Whitney, aki már 1800-ban minőségellenőrzést vezetett be, Charles Babbage munkássága (akit a számítástechnika egyik korai úttörőjeként is emlegetnek), valamint az első üzleti iskola megalapítása.

Az 1880-tól 1927-ig tartó időszakot a tudományos vezetéselmélet időszakának nevezik, ennek legfontosabb motívuma az ipari forradalom, a gépek és a mechanika térhódítása. Szervezési oldalon elsősorban a munkaszervezésre és a bérezésre kellett hangsúlyt fektetni – a Frederick Winslow Taylor által lefektetett alapokat ezeken a területeken a mai napig használják – ekkor alakult ki a munkaelemzés (mozdulatvizsgálat), teljesítmény arányos bérezés és a normák számításának technikája (*TAYLOR 1911*).

A következő időszakot (1927–1950) az emberi kapcsolatok időszakának nevezik. Az előző szakaszban gyakran felmerült az ember-ellenesség vádja, mi szerint nem etikus az embereket szervezési szempontból így (sokak szerint a normák léte rideg, emberellenes környezetet teremt) kezelni. Az előző kor hibáiból tanulva George Elton Mayo (üzempszichológus) igyekezett a lehető legemberibb környezetet kialakítani a munkához, bízott benne, hogy a jó környezet és a gyakori pihenők jó hatással lesznek a teljesítményre. Vizsgálata eredményeként felállította az ipari szociálpszichológia, menedzsment és az emberi kapcsolatok alapjait (*MAYO 1945 ÉS 1993*).

Az utolsó szakasz 1950-től a mai napig tart, ez a szintézis időszaka. Az időszak elejét meghatározza Abraham Maslow szükségletpiramisa (*MASLOW 1954*), és az egyén fontosságának hangsúlyozása (Herzberg ösztönzéselmélete (*HERZBERG 1954*)). Többek

szerint a korra leginkább a már korábban kitalált módszerek a rendelkezésre álló technológia által történő finomítása jellemző.

Mivel sokan sokféleképpen próbálták az információmenedzsment korszakait felírni, ezért talán nehéznek tűnhet annak meghatározása, hogy mikor voltak a nagy fordulópontok. Ennek elsődleges oka az, hogy a korszakképzés alapjára a felhasznált technológiát próbálják ráerőltetni. Jobb megközelítés lehet azonban az, ha a felhasznált technológia célját elemezzük. Így könnyebben felállítható a következő korszakképzés (MAY, 2009):

- 1985-1995: az információtechnológia alkalmazásának célja a hatékonyságnövelés és a költségcsökkentés;
- 1995-2000: a cél az innováció (mely sokak szerint gyakran káoszba csap át);
- 2000-2005: az információtechnológia alkalmazásának célja újra a hatékonyságnövelés és a költségcsökkentés;
- 2006-2017 (várhatóan): az alkalmazás célja együttesen a hatékonyságnövelés, a növekedés és az innováció.

## 2.1 Elektronikus üzletvitel

Az ERP<sup>6</sup> rendszerek feltehetően az MRP<sup>7</sup> rendszerek következő generációját jelentették. Üzleti szemszögből nézve az ERP hatóköre kibővült, mivel eleinte csak a termelő folyamatok koordinálása volt a cél, míg később már vállalati szintű támogatás. Technológiai szemszögből nézve ez az őrendszer megvalósításoktól a rugalmas, jól testre szabható megoldások irányába mozdulást jelent (SYSOPTIMA.COM 2009).

Az 1960-as években lehetett először vállalati informatikai rendszerekről beszélni, elsőként a készletgazdálkodás és a vezetés területére tudott beférközni az akkor új technológia. Ez akkoriban nagyjából a megfelelő készletmennyiségek figyelését jelentette az raktáraknak. Már ekkor a végrehajtandó feladatok közé tartozott a készletezéssel kapcsolatos követelmények azonosítása, a célértékek beállítása, a feltöltési technikák és lehetőségek koordinálása, az árucikkek fogyásának figyelése, a készletegyensúly fenntartása és kezdetleges jelentéskészítés is. Az egyik első, de mindenképpen a leghíresebb, sikeres megoldás J.I. Case és az IBM együttműködésének eredményeként született. A termék /

---

<sup>6</sup> Enterprise Resource Planning, eredetileg készletgazdálkodást automatizáló informatikai rendszer, később azonban a hatókör kibővülése miatt vállalatirányítási rendszert jelent.

<sup>7</sup> Manufacturing Requirements Planning, azaz a gyártási feladatok végzéséhez szükséges anyag és eszközgazdálkodást segítő / automatizáló informatikai rendszer.

rendszer neve MRP (*LUNN ÉS NEFF 1992*) volt (innen a későbbi kategória megnevezés), és alapfunktionalitását a nyersanyagok beszerzésének tervezése és ütemezése jelentette (*WHYBARK 1975*).

Az első üzleti intelligencia definíció még 1958-ban hangzott el egy IBM kutatótól (*CERUZZI 2003*), miszerint „a tények közti összefüggések olyan feltárása, mely a kívánt cél eléréséhez hozzásegíthet”, persze ekkor a technológia nem tartott ott, hogy komolyabban lehessen vele foglalkozni (*IBM 1973*).

A '60-as évek közepén jelentek meg a PICS<sup>8</sup> rendszerek, melyek főleg termeléssel kapcsolatos vállalatoknál jelentek meg; a cél a termelési folyamatok lehető legnagyobb automatizálása volt: főként készletezés, rendelések és rutin termelésirányítás téren.

Egy évtizeddel később, az 1970-es években jelentek meg szélesebb körben az MRP rendszerek (*ORLICKY 1975*), melyek fő célja a termelési folyamatok ütemezése volt. Segítségükkel a műveletek és a nyersanyag beszerzések a késztermék szükségletre alapozva voltak számíthatóak, lehetőség volt a termelési rendszer struktúrájának módosítására, valamint a készletszintek és a szükséges raktárméret kiszámolására. A kor sajátossága volt a mainframe-es működés, így az üzemeltetéshez nagyon nagyszámú támogató személyzetre volt szükség. Összességében elmondható, hogy ezek a megoldások nagyok, nehezen kezelhetőek és nagyon drágák voltak. 1972-ben a németországi Mannheimben megalapította az SAP céget (*HENDERSON 2009*). A cél olyan integrált üzleti megoldás létrehozása volt, mely a későbbiekben szabványként szolgálhat. A piac gyorsan bővült, több cég fedezte fel a szegmensben rejlő lehetőségeket, így 1975-ben megalapították a Lawson Software nevű céget (*GRANT 2006*). Az ő céljuk nem a testre szabott megoldások fejlesztése és piacra dobása volt, hanem inkább általánosan használható termékeket szerettek volna árulni. Ettől függetlenül, 1976-ban az MRP válik a termelő ipar irányítási alapfogalmává. 1977-ben alakul meg a JD Edwards és az Oracle Corporation (*ENSOR, NORGAARD, GORMAN, KOLK 2004*). 1978-ban már Baan néven informatikai és pénzügyi, adminisztrációs tanácsadással foglalkozó céget alapítanak (*GARG ÉS VENKITAKRISHNAN 2004*). 1979-ben kerül forgalomba (az Oracle-től) az első kereskedelmi SQL relációs adatbázis kezelő rendszer – a mai napig ennek alapjain nyugszik a legtöbb ilyen szoftver.

---

<sup>8</sup> Production Information Control System, mely az IBM egyik terméke volt, de tudásával nevet adott a termékcsoporthoz. Hivatkozható PCS (Process Control System – Folyamatirányító rendszer) néven is, így inkább érezhető gyűjtőfogalomnak.

Az 1980-as évekre nem az új technológiák bevezetése, mindinkább a már meglévő rendszerek bővítése volt a jellemző (*HARRIS 1983*). Az ekkor tervezett és kivitelezett rendszereket MRP-II-nek nevezik (*WRIGHT 1981*), utalva a fejlődésre. Ekkor vált igazán elérhetővé a termelő folyamatok koordináció, a tervezés, az anyagbeszerzés, a készletnyilvántartás és a termékelosztás támogatásával (*TOOMEY 1996*). Ekkor teljesül az igény az elosztási és üzletszintű aktivitások támogatására. 1987-ben megalakul a PeopleSoft, egy évvel később tőlük kerül piacra az első HR irányú szoftver (*ERPANDMORE.COM 2008*). Ennek oka az, hogy gyorsan nőtt a vállalatok digitális adathalmaza; egyre több üzleti egység volt képes digitálisan adatokat tárolni.

Az 1990-es években pörgött még inkább fel az információtechnológia piaca. Ennek oka főként a személyi számítógépek megjelenése és gyors terjedése volt. Lehetővé vált a számítógépek lejtatása a vállalat minden pontjára, minden irodájába. Ekkor kerültek bevezetésre az első ERP rendszerek (*HIGGINS, LE ROY ÉS TIERNEY 1996*), ekkor következett be a ma „.com buborék” néven elhíresült gazdasági hullámvölgy<sup>9</sup> is (*MILLER, VANDOME ÉS MCBREWSTER 2010*). Ezek a vállalati rendszerek már nagyrészt mind modulárisak voltak (természetesen a modularizáció, mint technika már korábban megjelent, de ekkorra vált bevett szokássá), fő céljuk a belső folyamatok valamint a szervezeti egységek közötti üzleti folyamatok optimalizálása, kezdve a termék tervezésétől és az alkatrészek beszerzésétől, a készletgazdálkodáson és a terjesztésen át, a teljesítésig és a megrendelésekig. A későbbi változatok már egyre gyakrabban tartalmaztak marketing, pénzügyi, könyvelő, HR és egyéb modulokat. Ekkor a Baan cég már indirekt módon 35 országban van jelen, Azt ERP megnevezésre azért is volt szükség, hogy nevében is utaljanak a kibővült szolgáltatási körre, utalva az elérhető tervezést támogató, pénzügyi, HR és projekt menedzsment modulokra. 1991-re már a PeopleSoft is jelen van minden földrészen, 1999-re pedig a JD Edward 100 országban van jelen, minden második HR megoldás a PeopleSoft-tól származik (*BROMWICH 1999*) és a vezető vállalati szintű szoftver szállító az SAP, 50 országban van jelen. az évtized elején vált elérhetővé a grafikus megjelenésű web, amit hamarosan mint marketing eszközt kezdett használni az üzleti élet – persze ekkor még az oldalak nagy része csak kapcsolati információkat tartalmazott, mint telefonszámok, vagy e-mail címek. Az elektronikus (kis)kereskedelem

---

<sup>9</sup> Szakértők szerint ez persze nem a technológia önmagáért való túlpörgése volt, hanem ez volta az információtechnológia igazi útkeresése: meg kellett várni, míg kialakulnak a széles felhasználói tömegek és meg kellett várni, míg rendelkezésre áll olyan technológia, mely alkalmas a felhő alapú szolgáltatások nyújtására.

1994-ben kezdődött, ekkor vált elérhetővé az interneten keresztül történő hitelkártyás fizetés (az SSL titkosítás megjelenésének (*STAPKO ÉS STAPKO 2008*) köszönhetően), ráadásul a rohamos terjedésnek köszönhetően milliókhoz juthatott el a reklám és termékek megvásárlásának lehetősége. Aztán a szervertechnológia fejlődése lehetővé tette, hogy nagy weboldalak lehessen építeni, tisztán interneten működő web-áruházaknak, nagy adatbázis-háttérrel. Ekkor jelent meg a világon az eBay (*COLLIER 2009*) és az Amazon is. A nagy áttörést a hatékony és kevesebb plusz feladatot jelentő internetes kereskedelembe a dinamikus oldalak megjelenése jelentette, így már nem kellett minden terméket egyesével felvinni, hanem akár több százezer termék is feltölthető volt adatbázisból – kimutatható az összefüggés a termékszám növekedése és a forgalom, valamint az eladások növekedése között (*OPENSOURCEERP.GURU.COM 2009*).

A '80-as és '90-es évek voltak az igazán jó időszakok az ERP szoftverrendszereket készítő cégeknek (*WIGHT 1995*), a fellendülő piac éves bevétele millió dollárokról milliárdokra nőtt néhány év leforgása alatt. A '90-es évek elején látott napvilágot a második megfogalmazás, ami az üzleti intelligenciát mint gyűjtőfogalmat használja, miszerint „olyan koncepciók és módszerek összessége, mely a tényeken alapuló döntéstámogatással segíti az üzletmenetet”. Ebben az évtizedben – a vállalati rendszerek terjedésének köszönhetően – számos informatikai rendszer került bevezetése, saját adatbázissal, mind-mind adott funkciót kiszolgálva. Ezek az adatbázisok „szigetalkalmazások” lettek, mivel információikhoz semmi más alkalmazás nem férhetett hozzá. Ezek a szigetek az alkalmazások körének bővülésével pedig gyorsan szaporodtak. Az okozott problémát még a később bekövetkező egyesülések és felvásárlások sem oldották meg, mivel a felvásárolt termékek nagyon hasonló feladatot láttak el, azokat hiába integrálták.

Ahogy egyre több cég jelezte igényét integrált megoldások iránt, úgy reagált rá a piac; megjelentek az olyan megoldások, melyek képesek voltak a sok szigetalkalmazás információtartalmát egyesíteni, majd elemezni azokat. Ehhez meg kellett alkotni az adattárházak modelljét.

Persze nem volt sokáig érezhető a dübörgő piac –a kezdeti lelkesedés után alább maradt, a tökéletlenségek láthatóvá váltak, és sok cég kezdte úgy nézni az ERP-koncepciót, mint egy rendkívül drága megoldást, ami túl komplikált, és aminek működtetése és karbantartása rengeteg erőforrást emészt fel. A részleges kiábrándulás egyik oka a bevezetésre szánt költségkeret és a határidők majdnem minden esetben történő túllépése is volt. Így aztán sok

cég kezdett kételkedni az integrált koncepció létjogosultságában – rájátszott erre számos nagy botrányt kavaráó rosszul sikerült implementáció is, ami különösen a kisebb cégek számára rontotta a helyzetet (ADAM ÉS SAMMON 2004).

Az ezredforduló után némi csökkenés volt érezhető az ERP-rendszerek piacán, ezt a csökkenést akkor a 9/11-es terrortámadásokkal hozták összefüggésbe, de valószínűsíthető, hogy azon események elmaradása esetén is tapasztalható let volna a visszaesés. Az ok az lehet, hogy kellően sok cég kezdett üzleti intelligencia rendszerek árúsításába és ez a piaci szegmens telítetté vált, valamint a cégek kezdtek rájönni, hogy ez a technológia sem mindenre orvosság, csak jól használva néhány addig ismeretlen lehetőség kiaknázására képes. 2002-től vált követelménnyé ezen rendszerekkel kapcsolatban az internetes hozzáférhetőség, beszállítói és megrendelői oldalról egyaránt. Igazi technológiai áttörés talán 2004-ben történt, amikor végre szabványosították a SOA<sup>10</sup> szoftverarchitektúrát (KRAFZIG, BANKE ÉS SLAMA 2005), jelentősen könnyítve a különböző rendszerek közti kommunikációt (ERL 2008).

A SOA architektúra erősen támaszkodik a SOAP, XML és WSDL szabványokra. A SOAP<sup>11</sup> képezi a legtöbb webszervíz alapú rendszer kommunikációjának alapját (persze vannak más, szabványos megközelítések, de ez a legelterjedtebb), használata során más protokollokra is támaszkodik, mint a HTTP<sup>12</sup>, vagy az RPC<sup>13</sup>. SOAP alapú webszervízek esetén az üzenetek XML-alapúak, ebben a formában adják vissza a kért adatokat. Az XML internetre tervezett strukturált adatok megosztására készített jelölő nyelv. A WSDL<sup>14</sup> egy XML formátumú dokumentum, mely a webszolgáltatás leírását tartalmazza.

A 2003-2005 közti időszak jelentős átrendeződést tartogatott a vállalati informatika piacán: Az Oracle felvásárolta a PeopleSoft-ot, a JD Edwards-ot, a Seibel-t és az E-Business Suite-ot. Közben a Microsoft is meglátta a lehetőséget a vállalatirányítási informatika szektorban, így saját fejlesztés helyett inkább felvásárolta a Navision, az Axapta, a Great Plains és a Solomon cégek termékeit. Az Infor felvásárolt a Baan és Mapics mellett még több kisebb céget, míg a Sage megvette a Best Software-t (BI.HU 2005).

---

<sup>10</sup> Service-Oriented Architecture, vagyis szolgáltatásorientált architektúra. Használatának célja egy biztonságos üzleti infrastruktúra kialakítása, erősen támaszkodik az újrafelhasznált komponensekre – webszervízként valósul meg.

<sup>11</sup> SOAP: Simple Object Access Protocol, vagyis egyszerű objektum hozzáférést biztosító protokoll.

<sup>12</sup> HyperText Transport Protocol, alkalmazási protokoll, mely főként weben való információterjesztésre szolgál.

<sup>13</sup> Remote Procedure Call, vagyis külső procedúra hívás.

<sup>14</sup> Web Services Description Language



Az egyesülések áttértek az üzleti intelligencia szektorra is, mivel a szoftvercégek felismerték, hogy a teljes körű megoldások szállításához ilyen jellegű szolgáltatás nyújtása elengedhetetlen. Ennek jegyében az SAP felvásárolta a Business Objects-et (*BI.HU 2007*).

A konszolidációt követően azonban újabb változások vannak készülóban a piacon, megjelentek az első nyílt forráskódú és megfelelő funkcionalitással rendelkező ERP-rendszerek. Ezen nyílt megoldások célcsoportja elsősorban a kis- és középvállalkozási (KKV) szektor, azon belül is azok a cégek, melyek SaaS<sup>15</sup> megoldást választanának, mivel így interneten nyújtható a teljes funkcionalitás (*BI.HU 2010*).

## 2.2 Döntéstámogatás és informatika

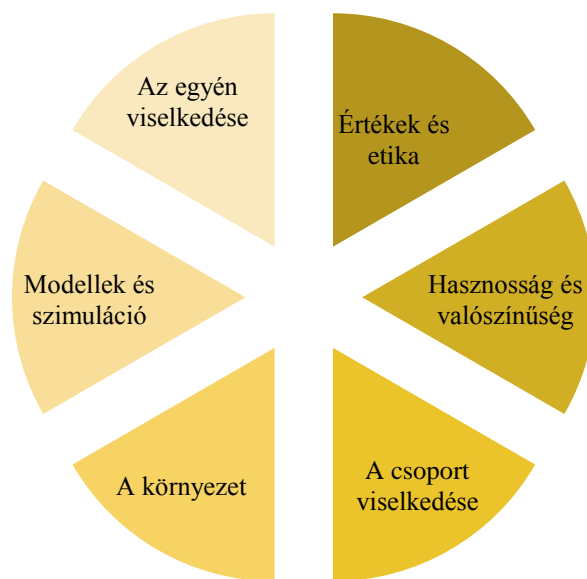
A döntés olyan lehetőségek közötti választás, mely előre rögzített kritériumok szerint történik, valamilyen előre kigondolt stratégia alapján. Döntési helyzet persze akkor áll elő, ha mérlegre kerül a valóság, illetve az elvárások; döntés akkor szükséges, ha a kettő között a bíráló személy (vagy rendszer) számára látható eltérés van. Azért, hogy ne kizárólag személyes döntések szülessenek, korán érkezett igény (elsőként az ismétlődő) döntések automatizálására – így a döntéshozó (aki ismeretei, készségei és tapasztalatai alapján mérlegeli a lehetőségeket azok várható kimenetelei alapján) bizonyos esetekben kiváltható; a döntés (mely nem kizárólag szakmai, hanem (területfüggően) társadalmi folyamat) automatizálható (*KŐ-LOVRICS 1997*).

Alapvetően háromfajta döntési eljárást lehet megkülönböztetni: az empirikust vagy intuitívat, mely egyéni tapasztalatokra alapoz, az algoritmus alapján történő döntéshozatalt, ami számszerűsíthető problémák esetén használható és jól támogatható, illetve a heurisztikus eljárást (*SKIENA 2008*), ami logikai következtetés alapján dolgozik, modelleket alkalmaz, és a tapasztalatok alapján fejlődik – már ez is jól támogatható (*FANK, HIPEL ÉS KILGOUR 1993*).

A döntéstudomány azonban számos segédtudománnyal bír, melyek mind hasznos adalékot nyújtanak a döntések meghozatala során.

---

<sup>15</sup> Software as a Service: Szoftvert szolgáltatásként, mely egyben jelent értékesítési filozófiát és szoftverfejlesztési irányt; Erősen támaszkodva a web és a számítási felhők nyújtotta lehetőségekre.



*forrás: saját munka*

*3. ábra: A döntést befolyásoló tényezők*

A hasznosság egy döntési alternatíva relatív kívánatosságának mérésére szolgáló mennyiség. A döntéshozó személyéhez kötött, ezért relatív és szubjektív. A valószínűség azt vizsgálja, hogy a döntéshozó által kiválasztott alternatíva alkalmazása esetén mekkora az esély az elképzelt következményállapotba való eljutásra (*SIMON 1957*). Minden döntés valamilyen közegben, környezetben születik meg. Ez a környezet meghatározza a döntéseket. A matematika a döntéelmélet és módszertan segédtudománya. Azokat a modelleket, technikákat teszi elérhetővé, melyeket elsősorban az elemzés fázisában használunk (létezik persze sok olyan felhasználás, amikor később is szükség van rá, ilyen az optimalizáció, illetve a modellek összehasonlítása is). Aranyszabály, hogy mindig a problémához kell módszert, modellt választani, és nem fordítva. Akárhány szereplője van a döntéshozatali folyamatnak, a főszereplő mindig az egyéni döntéshozó marad (felelőségek és szerepkörök figyelembevételével). Döntéseinket befolyásolják [3. ábra], sőt meghatározzák azok az értékek, amelyeket követünk. Az etika olyan kérdéseket boncolgat (*ZSOLNAI 2000*), hogy meddig terjednek a döntési kompetenciák, hol vannak a döntések határai, ki tehető felelősség bizonyos döntésekért (*ENYEDI 2004*).

Az irányított döntéshozatal története ugyan az őskorban kezdődik, de a téma szempontjából lényeges vállalati döntéshozatal megszületése is az 1830-as évekre tehető,

ugyanis 1839-ben kezdték meg a borítékok tömeggyártását, majd 1842-ben kidolgozzák az akkori legnagyobb koordinálandó üzleti vállalkozás felépítését (ezek akkoriban egyértelműen a vasúttársaságok voltak). Ugyanebben az évben alapítják meg az első gazdasági főiskolát (Eastman Commercial College). Hosszú ideig ezek után a húzóágazat az irányításban (és emiatt a döntéshozatalban is) a vasút volt, aztán idővel persze valamelyest felzárkóztak a távirati irodák és a kórházak. Valamivel később a kereskedőházak megjelenése (*MICHIGAN 1887*) és a tőzsde beindulása is gyorsítani kezdte a folyamatot (*SCOTT ÉS GRAS 2009*). A fejlődés (részben a technológiának, részben a gyakorlatok terjedésének köszönhetően) dinamikusan megy végbe.

Az irányítás és döntéshozatal következő fejezete a második világháború után kezdődik, ekkor indul meg az információalapú fejlődés, mely Nyugat-Európára, az Egyesült Államokra, Japánra és Kanadára volt jellemző – mely az információs társadalom kialakulásához vezetett. Már a '60-as években komoly tervekkel és kérdésekkel rendelkeztek a tudósok a számítógépek automatizálásra való használatával, illetve a mesterséges intelligencia kutatás néhány gyökere is az akkori döntésautomatizálásban található (*SIMON 1982*). Ekkor kezdtek komolyabban foglalkozni a döntéstámogatás technológiai (radikális – akik szerint hamarosan a gépek felválthatják az embereket – és konzervatív – akik szerint a gépek csak céljukat tudják majd elvégezni – eszmék feszültek egymásnak), gazdasági (ugyancsak radikális – akik szerint az automatizálás és a születő mesterséges intelligencia látszata majd hatalmas munkanélküliséget és szolgáltatási áradatot hoz – és konzervatív – szerintük mindez csak az ipari forradalom egy új vívmánya – gondolkodásúak vitája kerekedett) és filozófiai (itt a radikálisok már azt hangoztatták, hogy miként lehet egy gép egyenlő egy emberrel, míg a konzervatívok inkább a határ örökre meg fog maradni) dimenzióival. Az információk beszerezhetősége hozzásegítette a vezetőket a jól strukturált problémák feltárásához, valamint a technikai emberek a „heurisztikus”<sup>16</sup> programozás” lehetősége a problémák megoldásához. A technológia és a társadalmi helyzet lehetővé tette tehát a főként információk feldolgozására és szétosztására épülő szolgáltatások létrejöttét (*EDWARD.WARD ÉS BYTHEWAY 1999*), melyek azóta is meghatározzák ezen területek arculatát (*BERINGER 2004*). Ahogy a technológia fejlődött és a matematikai eszközök használhatóvá váltak, bekapcsolódhatott a döntéshozatal automatizálásába a statisztika, az operációkutatás, (az elektrotechnika és) a programozás.

---

<sup>16</sup> A heurisztikusság itt elsősorban mintakeresésre és mintafelismerésre utal.

A döntéstámogató (informatikai) rendszerek definiálása szakirodalom szerte nehéz feladatnak minősül, mivel gyakorlatilag minden olyan megoldás nevezhető döntéstámogató rendszernek, amely segít bizonyos döntések meghozatalában. Az ilyen mértékű általánosítás persze nem szerencsés, mivel egy egyszerűbb eszközt (táblázat vagy egyszerű weblap) nem is igazán lehet összehasonlítani a '90-as években fellángoló döntéstámogató rendszerekkel, és főként nem a mai legmodernebb megoldásokkal (*BERKELEY, WIDMEYER, BREZILLON ÉS RAJKOVIC 1998*). Helyesebb definíció tehát az, ha a döntéstámogató rendszereket komplex problémák megoldására irányuló, adatbázis- és modell-alapú informatikai rendszernek nevezzük (*BAIER ÉS DECKER 2005*).

Az informatikai döntéstámogatás hőskorában – a '80-as években – ezek a rendszerek főleg nagyon jól specifikált problémákkal kapcsolatos döntésekben voltak képesek segítséget nyújtani, nagyrészt célalkalmazásokként (*POWER 2002*). A legelterjedtebb döntéstámogatási informatikai szolgáltatás a pénzügyi szektorban a vállalati kimutatások elemzése, valamint a kereskedelemben a beszállító-elemzés (persze a mai lehetőségekhez mérten csak kezdetleges formában) volt.

Azóta, ahogy a szektor a technológia fejlődésének köszönhetően bővült, igen szerteágazó megoldások jelentek meg – főként elméletben: szöveg-, adatbázis-, táblázat-, egyenlet-, szabály alapú, illetve kommunikáció-, adat-, dokumentum-, tudás-, web-, és modellvezérelt döntéstámogató megoldások. Idővel persze megszűnt a megkülönböztetés, és az összetettség a sokszorosára növekedhetett – elengedhetetlen volt ugyanakkor a relációs adatbázisok, az OLAP és az adattárház technológiák és az adatbányászat megjelenése (*LINSTEDT, GRAZIANO ÉS HULTGREN 2008*). Így, ezekkel már a mai döntéstámogató rendszerek (gyakorlatilag nincs is határ az üzleti intelligencia és a döntéstámogatás közt) valóban sokcélúak, és valóban sokrétű problémák megoldására alkalmasak (*HAMMERGEN ÉS SIMON 2009*).

Ehhez hasonlóan, a határok ugyanígy elmosódnak a korábban még jól kategorizálható rendszerek között: mára általánosan elfogadott, hogy egy vállalati integrált információrendszert ERP jelzéssel illetnek, mivel nem sorolható be egyértelműen sem döntéstámogató (DSS, GDSS), sem vezetői információs, sem csoportmunka-támogató, vagy munkafolyamat támogató rendszerek közé; helyette adott inkább a mára mindenre ráhúzható ERP megnevezés (holott ez sem igazán találó), és a régen még külön külön rendszerként elérhető funkciók már modulokként érhetők el. Így a legtöbb

vállalatirányítási rendszer (mely kíván versenyben maradni, és emiatt folyamatos fejlesztés alatt van) rendelkezik a döntések automatizálását segítő modullal, vezetői információs modullal, valamilyen üzleti intelligencia és elemzői szolgáltatással, valamint munkafolyamat- és csoportmunka-támogatással.

Az informatikai döntéstámogatáshoz sok esetben jó eszköz a szimuláció használata. A szimuláció nem más, mint rendszerekhez kapcsolódó modellek vizsgálata, kiértékelés és vizsgálat céljából (*JÁVOR 1979*). Megjegyzendő persze, hogy a szakirodalom nagyon változatosan definiálja a szimuláció jelentését, illetve azt, hogy mi tartozik a fogalmi körbe.

### **2.2.1 Racionalitás és problémamegoldás**

Az adott rendszerben a döntéshozó akkor hoz racionális döntést, ha megfelel az axiómákban megfogalmazott követelményeknek. Minden döntési problémában két fontos paramétert vesznek figyelembe: mindegyik lehetséges eredmény szubjektív értékét (hasznosságát), és a szóban forgó eredmény bekövetkezésének valószínűségét. A leghíresebb axiómarendszer a Neumann-Morgenstern-féle (*LEONARD 2010*):

- Az összehasonlíthatóság (komparabilitás) axiómája azt fejezi ki, hogy két dolog (a döntésekben két alternatíva) esetében a döntéshozó kész és tud dönteni: az egyiket preferálja a másikhoz képest vagy fordítva (*PATAKI 2008*).
- A tranzitivitás axiómája feltételezi, hogy egy „racionális döntéshozó” preferenciái kielégítik a tranzitivitás követelményét. (Ha  $A > B$  és  $B > C$  akkor  $A > C$  mintájára a tranzitivitás axiómája azt mondja ki, hogy ha A preferált B-vel szemben és B preferált C-vel szemben, akkor ebből az A preferált C-hez képest preferencia relációnak kell következnie (*PATAKI 2008*). Edwards kimutatta, hogy aki nem elégíti ki a tranzitivitás axiómát a gyakorlatban kizsákmányolható. Ez az axióma valamennyi normatív axiómarendszerben szerepel, sőt egyes leíró döntéselméleti modellek is megtartják.
- A dominancia axiómája kimondja, ha egy S1 cselekvési változat olyan eredménnyel jár, amelyik legalább annyira preferált a döntéshozó számára, mint egy másik S2 cselekvési változat eredménye, továbbá legalább egy lehetséges tényállapot esetében S1 preferáltabb eredménnyel jár, akkor a dominancia axiómája

értelmében (*PATAKI 2008*) a racionális döntéshozónak nem szabad S2 változatot preferálni S1 változattal szemben, vagyis az S1 változat domináns.

- A függetlenség axiómája azt fejezi ki, hogy az eredmények hasznosságának és valószínűségének egymástól függetlennek kell lenniük. A racionális döntéshozónak nem szabad tehát sem pesszimistának, sem optimistának lennie, hanem „realistának” kell maradnia (*PATAKI 2008*).

A jól automatizálható probléma tehát olyan, melynek lehetséges kimenetelei egymással összehasonlíthatóak, közöttük hasznossági reláció állítható fel, a kimenetek egymástól függetlenek és kiválasztható domináns kimenetel. Jól automatizálható problémákra (azok természetétől és a rendelkezésre álló adatok milyenségétől függően) többféle elemzési technika alapján lehet ajánlható kimenetelt választani. Valószínű azonban, hogy sok esetben a különböző elemzési technikák más-más hasznosságú kimenettel rendelkező döntést javasolnak, ám éppen ezért célszerű már több elemzési eszközt és azok lehetséges kimeneteleit összehasonlítani, majd közülük (a többféle képpen elérhető kedvező megoldások közül) az optimálisat kiválasztani.

### 2.2.2 Döntéshozatali folyamat

Minden döntési helyzet alapvetően hét fázisra osztható <sup>17</sup>. Ezek: a döntési helyzet felismerése, helyzetfelismerés, helyzetelemzés, célkitűzés és a döntési kritériumok meghatározása, a döntés, majd annak végrehajtása és ellenőrzése, értékelése (*S. NAGY 2007*). Természetesen egy ilyen döntési folyamat erősen ideális jellegű, megvalósítása nem minden esetben lehetséges. A legnagyobb problémát általában a helyzetelemzés során végzendő probléma feltárás és meghatározás jelenti – ezen kívül sok esetben ugyancsak nehéz lehet a választási lehetőségek összegyűjtése, azok kimeneteleinek meghatározása.

Az első fázis a döntési helyzet <sup>18</sup> felismerése [5. ábra]. Több szakirodalom szerint ez a folyamat legnehezebb pontja. A felismeréshez több módon lehet eljutni: az első, legrosszabb lehetőség a kényszerítő nyilvánvalóság, amikor már döntési kényszer áll fenn (ezt célszerű elkerülni, eleve nem szabadna hagyni, hogy a helyzet idáig fajuljon); a másik mód ha létezik valamiféle figyelmeztető rendszer, míg a harmadik a kutatás–tényfeltárás útján előálló helyzetfelismerés. A technológiai lehetőségek adottak ahhoz, hogy a

---

<sup>17</sup> Egyes szakirodalmak fázisokat összevonva öt vagy három fázisról beszélnek, ám a folyamat lényege ugyanez, az említett folyamatlépések mindegyikben megtalálhatóak.

<sup>18</sup> Általános döntési helyzet akkor áll fenn, ha van mód több lehetőség közül választani.

figyelmeztető rendszerek végezzék a dolgukat, automatizáltan. Ugyancsak nagy szerep jut a technológiának a problémakutatás során: itt ugyan jelentős emberi beavatkozásra van szükség, de a technológia használata nélkül időben lehetetlen lenne a kutatás kivitelezése (*BORDÁNE 1986*) (Gondoljuk csak el, hiába állna rendelkezésre a hatalmas adatmennyiség, valamint az elemzéshez szükséges statisztikai módszer, a technológia használata nélkül nem lennének képesek időben rájönni a problémát keltő összefüggésre – a legegyszerűbb példa lehet áruházi termék gyenge eladási oka: vásárlói kosár elemzéssel felkutatható, hogy nem jól van elhelyezve, vagy a helyettesítő termék túlságosan előnyös áron kapható, vagy nem fogynak kellően a segédtermékek sem).

A második fázis a helyzetfelmérés, itt lényegében az információk összegyűjtése, majd csoportosítása zajlik. Itt meg kell jegyezni, hogy az automatizált döntések információigénye valamelyest kevesebb (persze a szabályrendszert egyszer ki kellett alakítani, és ahhoz jelentősen nagyobb mennyiségű információra volt szükség). A helyzetfelmérés során vizsgálni kell a környezetet (a feltételek feltárása), a rendelkezésre álló erőforrásokat (korlátozások ismerete), a rendszer struktúráját és felépítését (lehetőségek), valamint a vezetést.

A helyzetfelmérés után következik a helyzet elemzése. Itt az információké és a matematikáé a főszerep: az információkból tudást kell képezni. Ezt az információk csoportosításával, elemzésével lehet elvégezni. Meg kell határozni a döntés során alkalmazható mutatókat vagy metrikákat is; meg kell határozni az alkalmazható módszereket és algoritmusokat. Az első három fázisra együttesen a döntési helyzet azonosításaként szoktak hivatkozni.

A negyedik fázis a célkitűzés és a döntési kritériumok meghatározása. Fontos a döntési kritériumok és azok prioritásainak jó meghatározása, mivel a későbbiekben ezek alapján lesznek összehasonlítva a cselekvési változatok, amiből aztán az ajánlható sorrend előáll. A nem megfelelő prioritizálás eredményeként nem optimális döntés születhet (*PÁLINKÓ ÉS SZABÓ 2006*).

Ezután következhet a cselekvési alternatívák kidolgozása és értékelése, összehasonlítása. Az alternatíva-választó algoritmusok közül az egyik legegyszerűbb a Kesselring-algoritmus (Az algoritmus lényege, hogy különböző szempontok alapján súlyozással hasonlít össze kategóriákat, termékeket. Az értékelés végén a legelőnyösebb a legmagasabb pontszámot elért változat.) (*RAFFAI 2003*), Combinex (csak pontszámítás és

súlyozás), de használják a JOKER (geometriai módszer, az ideálshoz leginkább hasonlító alternatíva kiválasztására) módszert (MIKOLICS 2000), az AMxM (az átlagosan legjobb alternatíva kiválasztása, fa struktúra képzésével az információk átlag számítása) módszert és a WINGDSS (Az értékelési szempontok alapján fa-struktúra készül, melyet súlyokkal kell ellátni, és a súlyozókat is értékeli; részfák súlyai alapján lehet dönteni.) eljárást (MIKOLICS 1997) is. Léteznek más választást segítő technikák is, mint a költség-haszon elemzés, vagy a határhaszon elemzés. Az alternatívák kiértékelése és összehasonlítása információtechnológiával jól támogatott folyamat, több szoftvereszköz áll rendelkezésre. Az alternatívák kidolgozásába beletartozik azok végrehajtásának javasolható menetének kidolgozása is. Az első öt fázist közösen döntés előkészítésnek nevezik.

Az alternatívák értékelése után a döntést a vezetőknek kell meghoznia. Célszerű a döntést a döntés előkészítés után azonnal meghozni. A vezetői döntéshozatalnak persze több részfolyamata van: az akaratképzés (ide tartozik a döntési helyzet felismerése, a döntés előkészítése, a variációk kidolgozása és értékelése, majd a döntés), az akarat érvényesítés (a döntés megvalósításának lépéseinek kidolgozása, feltételek biztosítása, az utasítás kiadása, majd annak megvalósítása) és az ellenőrzés (ez már részben a visszacsatolás, részben monitorozó tevékenység). A vezetői döntések kétszintűek: az első szint az, ha a vezető saját cselekvései céljára hoz döntést, a második szint, ha mások cselekvéseire hoz döntést. A helyzetismeret alapján a döntés lehet biztos (amikor a helyzet jól ismert, ismertek az ok-okozati összefüggések), bizonytalan (ha a helyzet nem ismert), illetve kockázatos (a helyzet nem teljes egészében ismert, de a körülmények bekövetkezésének valószínűsége ismert). Célszerű a biztos döntések meghozatalára törekedni [4. ábra], ez a célja az információtechnológiai (BEROGGI 2004) támogatásnak (amennyiben ez nem lehetséges, úgy a kockázatos döntések támogatása, majd a bizonytalan döntések támogatása a cél).



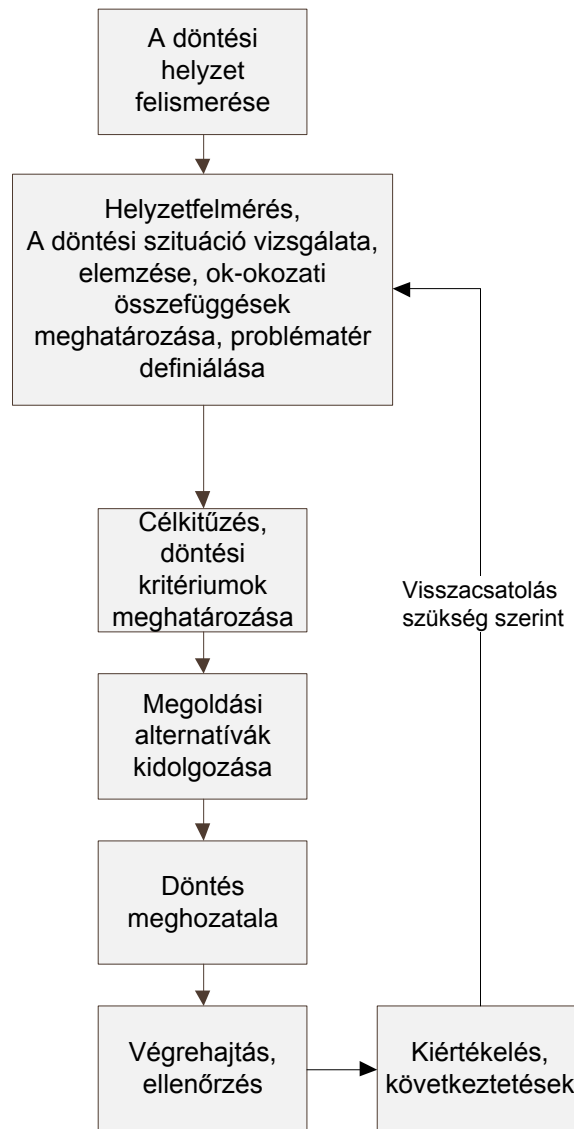
forrás: saját munka

4. ábra: Döntések típusai a helyzetismeret alapján



Az utolsó fázis a végrehajtás és ellenőrzés. A végrehajtást a vezetőnek kell megszerveznie. Célszerű, ha a szervezet figyelmet fordít az utólagos kiértékelésre is, így lehet tanulni az elkövetett hibákból, ezért célszerű visszacsatolást alkalmazni.

A visszacsatolás során kell hangsúlyt adni azoknak a behatásoknak, melyek a választott megoldás valóságba (gyakorlatba) való átültetés során jelentkeznek. Elképzelhető, hogy olyan szempontok is kihatással vannak a végkifejletre, amikről eddig nem volt tudomás. A visszacsatolás adja meg a lehetőséget ezen szempontok következő alkalommal való figyelembe vételére. Mivel az esetleges hiba elkövetésével új ismeretre teszünk szert, ezért tudásunk bővül, amit a következő modell vagy választási lehetőség megalkotása során felhasználhatunk; ez pedig hozzájárul a modellek egyre finomabbá válásához, valamint az elemzés körforgásához is.



forrás: saját munka

5. ábra: Döntéshozatali folyamat

### 2.2.3 Döntések típusai

A döntéseket sokféleképpen lehet csoportosítani. A feladat jellegéből adódóan lehet programozott döntés (a rutinszerű döntési helyzetekre utal, az ismétlődő probléma esetén többször alkalmazható – nagyon jól támogatható, automatizálható) vagy nem programozott döntés (új, egyszeri vagy ritka helyzetekre – szintén jól támogatható elemzéssel és adatbányászattal, de nem automatizálható) (ANDICS 1977). A döntéshozó hatásköre alapján a döntés lehet kötelesség vagy felelősség. Időtartam szerint megkülönböztetünk rövid vagy középtávú döntéseket. Jelentőség alapján négy szintet lehet megkülönböztetni: a döntés lehet alapvető vagy üzletpolitikai, stratégiai, taktikai illetve operatív. Szükségesség szerint

lehet szükséges egy döntés vagy szükségtelen. Egy döntés lehet végleges, vagy további döntéseket igénylő. Illetve egy döntés lehet egyszemélyes vagy csoportos döntés.

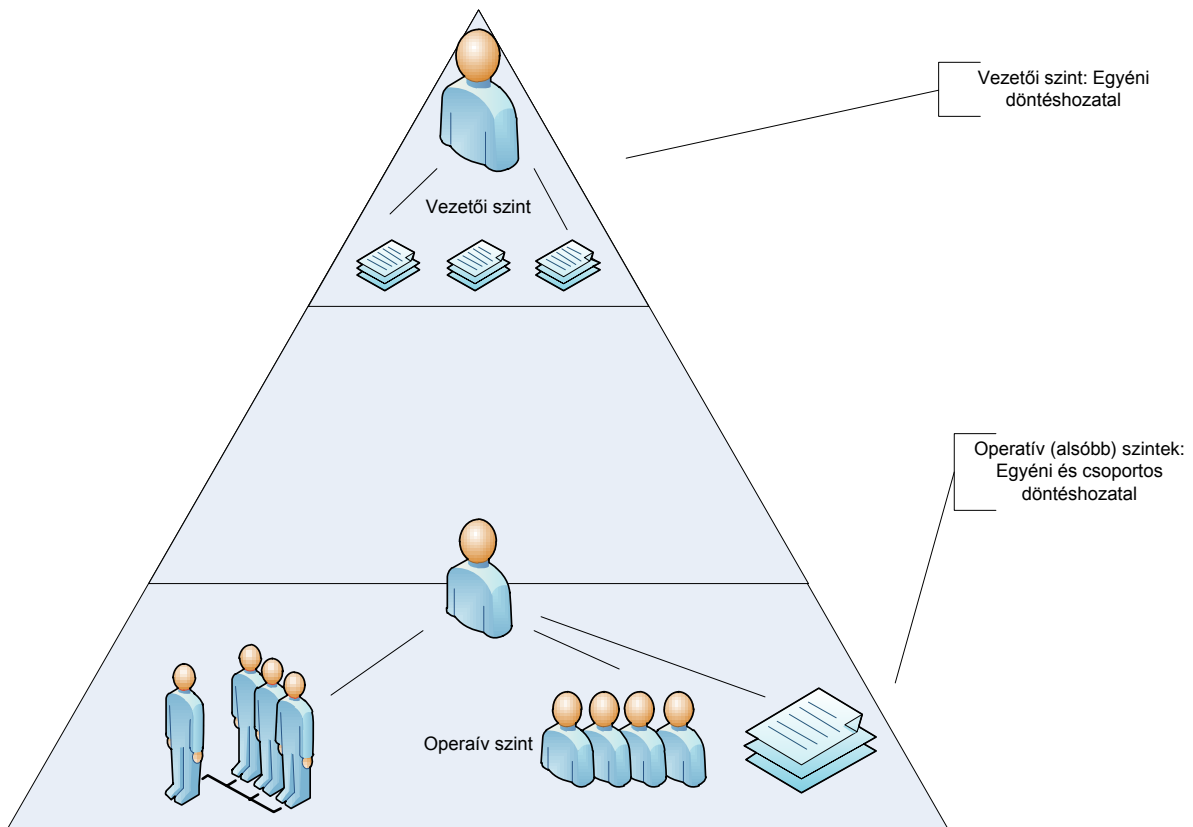
A szervezeti döntéshozatalnak ugyancsak több modellje ismert. Ezek abban térnek el egymástól, hogy más előfeltevésekkel élnek a döntéshozókra, valamint a döntéshozók közötti szervezeti kapcsolatokra vonatkozóan (*ALLISON 1969*).

A racionális egységes cselekvő megközelítés egyszemélyes döntéshozatalt feltételez, ahol a döntéshozó az optimális megoldás elérésére törekszik. Ez egy normatív modell, az elemzést állítja a középpontba. Azt feltételezi, hogy minden szükséges információ rendelkezésére áll, vagy beszerezhető. Általában optimális, a vállalkozás profitját maximalizáló döntést lehet hozni, ezért ezt a felfogást szokták optimalizáló közelítésmódnak is nevezni (*ZOLTAYNÉ ÉS SZÁNTÓ 2006*).

A szervezeti modellben több döntéshozó is szerepet játszik a döntéshozatalban, akik egy közös célért tevékenykednek, de azért vannak saját céljaik is. Olyan kognitív korlátokkal rendelkeznek, amelyek miatt korlátozottan racionális döntéshozóként nyilvánulnak meg. E modell azt feltételezi, hogy nem lehet minden szükséges információt beszerezni. A problémák általában túl komplexek, magas a bizonytalansági szintjük. A szervezet céljainak a kielégítő megoldások is megfelelnek. A kielégítő közelítésmóddal ésszerű ráfordítás mellett általában jó eredményt lehet elérni (*ZOLTAYNÉ 1997*).

Napjainkban a vállalat és az emberek vezetésében az egyéni és a csoportos döntések meghozatalára egyaránt szükség van [*forrás: saját munka*

6. ábra]. Míg az operatív szintű vezetés esetén az egyéni döntések dominálnak, de a beosztottak is bevonásra kerülnek a döntéshozatalban, addig minél feljebb megyünk a vállalati hierarchiában, egyre nő a csoportos döntéshozatal szükségessége (*DURÁNYIK 2007*).



forrás: saját munka

6. ábra: Vállalati döntéshozatal

A döntéshozatalban érdekelt felek részt vesznek a csoportos döntéshozatalban (eljárási formula, mód), mert csak a részvételük nyomán előálló megegyezés alapján lehet megalkotni a vállalat reális stratégiáját. A csoportos teljesítmény gyakran jobb, mint az átlagos csoporttagé (PINTÉR 2004). A csoporttagokat (csapattagokat) fel lehet használni a problémák innovatív és kreatív megoldásainak kifejlesztésére. Ezekre különböző technikák szolgálnak: brainstorming, okokat vagy megoldásokat kereső módszerek illetve gondolati képekkel segített csoportmunka (SZVITACS 2009).

A döntéstámogatás információtechnológiai alapokra helyezésével megszüntethető (persze hozzáfűzendő, hogy rengeteg múlik a döntés alapját képező szoftveres szabályrendszer megalkotásán) a megérzésen / intuíción alapuló döntéshozatal, mely ugyan számos esetben sikeresnek bizonyult, ám a látszati magas sikerarány betudható annak, hogy csak a legendásan jó döntések, és azok sikeres következményei anekdotává.

## 2.3 Intelligens vállalat

Az intelligens vállalat koncepció először 1992-ben jelent meg (*QUINN 1992*). A kifejezés olyan menedzsment megközelítést takar, mely az üzleti teljesítmény növelése érdekében alkalmazza újfajta szemléletben szolgáltatási paradigmákat és technológiai megoldásokat alkalmaz. A megközelítés kiindulópontja az, hogy a vállalatnál felhalmozott értelem és tudás lesz minden szolgáltatás létrehozásának és nyújtásának alapja (*SCHILLER 1999*). (Az alkalmazott technikák gyűjtőneve más egyes irodalmakban tudásmenedzsment (*BARTENSTEIN 2003*).)

A versenyszektorban (elsősorban a gyártási / előállítási szektorra igaz) a szolgáltatások költsége általában 65-75% között mozog, vagyis ennyit tesz ki átlagosan a hozzáadott érték a végtermék árában – az intelligens vállalat filozófia ezeket a költséghelyeket igyekszik inkább haszonközponttá alakítani (*THANNHUBER 2005*). Egyik kedvelt eszköze az outsourcing<sup>19</sup> (kihelyezés vagy kiszervezés), mely eleinte kifejezetten az informatikai részlegekre vonatkozott, itt kezdődött, és az informatikai részlegektől terjedt szét ennek gyakorlata (*SCHNIEDERJANS ÉS SCHNIEDERJANS 2007*).

Ha a kiszervezés révén a szolgáltatások aránya megnő, akkor az makroszinten is előnyként jelentkezik, mivel a keletkező szolgáltatások átformálják a gazdaságot (*BUTLER 2001*) – igaz ez kicsiben és nagyban egyaránt. Ahhoz persze, hogy a paradigma széles körben alkalmazható legyen, sok vállalatnál nézetváltásra van szükség: meg kell érteni és el kell fogadni, hogy a szolgáltatások is rendelkezhetnek magas értékkel, valamint hogy ugyanannyira függenek a technológiától és pénzáramlástól, mint például a termelőipar, el kell fogadni, hogy van annyi potenciál a szolgáltató iparágban, hogy kutatásokat finanszírozzon, és hasznosítsa az ott megszerzett tudást, végül pedig, hogy a szolgáltató iparág is életképes hosszútávon.

Felmerülnek persze etikai és biztonsági kérdések is, mi szerint valóban szabad-e mindent kiszervezni, szolgáltatásként igénybe venni, vagy adni. Általánosan elterjedt például, hogy a pénzintézetek (bankok, biztosítók) nem veszik szolgáltatásként igénybe az információtechnológiát, inkább megoldják „házon belül”. Ennek oka egyfelől a működési szabályzatban és a vonatkozó törvényekben keresendő (érzékeny adatokat nem is szabad külső félre bízni), másfelől bizalmatlanság és versenyszempontok képezik alapját.

---

<sup>19</sup> Outsourcing: Az outsourcing (kiszervezés) alatt azt a folyamatot értjük, amiben egy vállalat az addig szokásosan maga által végzett valamelyik tevékenységét átadja, egy, az adott tevékenységre szakosodott vállalatnak és a jövőben azt a bizonyos tevékenységet – díjazás ellenében – ezen másik vállalat látja el.

Gazdasági kérdések is jelentkeznek, és döntések előtt komoly szerepet játszanak – ugyanis a kiszervezés nem mindenesetben a legjobb pénzügyi megoldás, köszönhetően a szolgáltatások, illetve azon nyújtásához szükséges kellékek árának ingadozása miatt. Mérlegelendő tehát egyfelől a kiszervezni kívánt tevékenységhez szükséges technológia értéke (ha kevésbé fontos, akkor inkább ajánlható a kiszervezés), másfelől annak teljesítménye, elérhetősége (gyenge teljesítmény esetén inkább javasolható a kiszervezése, míg kielégítő teljesítmény esetén inkább a bent tartás).

A megközelítés alkalmazásával elérhető, hogy a vállalat értékét az ott dolgozók által felhalmozott ötletesség (*GUPTA ÉS SHARMA 2004*), innováció és tudás adja, ne elsősorban a birtokolt tulajdon értéke (*INFORMATIONWEEK.COM*).

## 2.4 Megállapítások

Véleményem szerint mára a nagyvállalatok és a gazdaság működése olyan mértékben vált bonyolulttá, hogy annak átlátása elengedhetetlenné teszi az elektronikus üzletvitel adta előnyök kihasználását.

Az elektronikus ügymenettel és a folyamat-automatizálással jelentős energia és költség takarítható meg.

A döntések meghozatala során pedig a bonyolult gazdasági-piaci összefüggések és a számos tényező hatásai miatt egyre inkább elengedhetetlen az információtechnológia használata. Ez persze nem jelenti azt, hogy a sikerhez száz százalékig elengedhetetlen, de a sikerre való esélyt megsokszorozza.

Az infúziós-diffúziós mértékkel (*RAFFAI 2006*) kapcsolatban megállapítható, az információtechnológiát magas szinten (beleértve az üzleti intelligencia által nyújtott lehetőségek kiaknázását) használó vállalatra vonatkozó infúzió mértéke<sup>20</sup>, mely a magas összefonódás és az erősen az információkra történő támaszkodás miatt az eddigieknél várhatóan magasabb értéket fog felvenni – érthető, hiszen az üzleti intelligencia rendszerek bevezetésének célja éppen az adatokból kinyerhető tudás extra profittá alakítása. Ezzel párhuzamosan, a diffúzió<sup>21</sup> mértéke ugyanakkor (a várakozásokkal némiképp ellentétesen) csak közepes szintet fog elérni, mivel az üzleti intelligenciához használatos adatokat a vállalat minden részén gyűjtik, (jó esetben nagyon sok területen kovácsolnak belőle

---

<sup>20</sup> Az infúzió mértéke határozza meg a vállalat függését az informatikai rendszereitől.

<sup>21</sup> A diffúzió mértéke az informatikai irányítás decentralizáltsága alapján határozható meg.

előnyöket is), de az adatok információvá, majd tudássá alakítása központosítottan célszerű, hogy történjen (vonatkozik ez az informatikai infrastruktúrára és a személyzetre egyaránt).

Megállapítható az igény arra, hogy a vállalatok az elektronikus üzletviteli szolgáltatásokat inkább szolgáltatásként vegyék igénybe. Ez több tényező együttes eredménye: egyfelől a szélessávú és megbízható internet-szolgáltatás már kellően elterjedt (a szolgáltatást igénybe vevő szempontjából), a hardverárak csökkenése és a technológia fejlődése lehetővé teszi az ilyen szolgáltatások erőforrás-hatékony<sup>22</sup> nyújtását (*GALLI 2008, GALLI 2009, GALLI 2010 ÉS GALLI 2011*) (a szolgáltatás nyújtója szempontjából), valamint kialakult gazdasági trend, mi szerint az ilyen szolgáltatások díját célszerű használatarányosan számítani (a szolgáltatást igénybe vevő és a szolgáltatást nyújtó fél közös érdeke).

---

<sup>22</sup> Erős szerepet játszik a virtualizációs technikák, a hálózati megoldások, valamint az informatikai menedzsment technikák fejlődése.

### 3 Üzleti intelligencia

Az üzleti intelligencia szó először 1989-ben került használatba. Legfőbb oka, hogy a szoftvercégek reklámszakemberei jobban eladhatónak érezhették ezt a kifejezést, az addig használatos döntéstámogató rendszer helyett. A két kifejezés között gyakorlati határvonal nincs. Minimális különbségként talán az alkalmazhatóságot lehet felhozni: kellően strukturált-e<sup>23</sup> a megoldandó probléma? Régebben annak kellett lennie, ha döntéstámogató rendszert akart használni a vállalat, mára ez már nem feltétlen követelmény: üzleti intelligencia rendszereknél több út is járható (*SABHERWAL BECERRA-FERNANDEZ 2010*). Kezdetben kizárólag a jól strukturált problémák voltak automatizálhatóak, mivel azok ismétlődőek és kidolgozott módszer van megoldásukra (algoritmus). A rosszul strukturált problémák ugyanakkor nem rendelkeznek bejáratott megoldó módszerrel – azonban statisztikai (és más informatikai eszközök bevonásával) úton, a kockázatok csökkentése mellett megoldhatóvá válnak.

Az IDC<sup>24</sup> kutatásai alapján az üzleti intelligenciának, mint iparágnak az átlagos forgási ideje körülbelül 15 év (*IDC HUNGARY 2008*). Az első ilyen ciklus 1975 és 1990 közt volt, melyet főként korai jelentések készítése a termelésben jellemzett. A technológiai oldalt nézve ez még a mainframe időszak, némi korai statisztikai szoftverrel ellátva. A második – „modern kor” néven emlegetett – ciklus 1990 és 2005 közt lehetett, melyet technológiai oldalról a kliens-szerver architektúra jellemez OLAP támogatással, (szoftverek tekintetében elérhetővé váltak célzottan BI felhasználásra szánt eszközök) összetett, de könnyen használható lekérdező, jelentés készítő szoftverekkel, és a végkifejletet erről a kliens-szerver technológiáról a web-bázisú megoldásokra való áttérés jellemzi, így szélesedett azon eszközkészletek száma és elérhetősége, melyek BI célokat szolgálhatnak (*VESSET 2009*).

2005 lehetett a fordulópont ebben az iparágban (és ha a 15 éves trend megmarad, akkor egészen 2020-ig így folytatódhat), mivel itt fedezték fel újra ennek a technológiának a jelentőségét, jól mérhető ez a szektorba fektetők számának növekedésével. Így mára a fókusz inkább az üzleti intelligencia határainak bővítésén van, vállalaton kívül és belül

---

<sup>23</sup> Akkor nevezhető jól strukturáltnak egy probléma vagy döntési helyzet, ha megismerhető minden lehetőség, azok értékei, és egymáshoz viszonyított preferenciái.

<sup>24</sup> Az IDC (International Data Corporation) egy 1964-ben alapított piaci elemző cég, mely elsősorban információtechnológiai és telekommunikációs megoldásokra specializálódott. Világszerte több mint 110 országban (így Magyarországon is) van jelen. A vállalat az IDG (International Data Group) leányvállalata, mely főként médiában és technológiai kutatásokban tevékenykedik



egyaránt, valamint a jelentéskészítéstől folytatódóan egyre inkább a döntéstámogatás irányába mozdításán. Ezek alapján a várható két fejlődő és terjedő megoldás a következő: egyfelől az üzleti intelligenciát a tömegek felé kell mozdítani (hogy ne csak kevesek kiváltsága legyen a használata – ide tartozik az árak csökkenése, a felhasználás egyszerűsítése), másfelől ezt döntés-központú módon kell megtenni.

A klasszikus BI-piac<sup>25</sup> főleg az elemzőkre koncentrált, ami csak nagyon kis részét jelentette bármilyen vállalatnak. Ezeknek az eszközöknek komoly jelentés-fejlesztő, OLAP és fejlett elemző támogatásuk van – ami nem éppen megfelelő a döntéshozók számára. Az üzleti stratégia elsődleges célja, hogy a megfelelő információt, a megfelelő embereknek a megfelelő időben kell eljuttatni – az IDC elemzése szerint csak a cégek 15%-a képes erre. Ezt a problémát elsősorban úgy lehetne megoldani, hogy a rendelkezésre álló eszközöket sokkal „felhasználóbarátabbá” kell tenni az adatok vizualizációjával, az interaktivitás fokozásával, csoportmunka támogatással, valamint a szövegbányászati eszközök keresőjének egyszerűsítésével. Az átmenet ideális végcélja az, hogy mindezek a szolgáltatások on-line formában, egy portálon gyűjtve legyenek elérhetőek.

Egy másik (érdekesebb) megközelítés az, hogy a lehető legtöbb BI funkciót illeszteni a már működő szoftverekbe. (Ennek a megoldásnak megvan az a veszélye, hogy Software bloat<sup>26</sup> keletkezik.) A megoldást intelligens folyamatautomatizálás néven emlegetik (IPA – Intelligent Process Automation), de felfogható úgy is, mint az IPA-megoldások az eredeti BI-funkciók és a BPA (Business Process Automation – Üzleti folyamatautomatizálás) megoldások közelítése (*MOHAPATRA 2009*). Főként ott lehet ez a hibrid megoldás sikeres, ahol nagyszámú, ismétlődő döntés meghozására van szükség, melyeket az előre lefektetett üzleti folyamatok határoznak meg. A mai napig kereskedelmi forgalomban nem érhető el IPA-szoftver, a vállalatok maguk, szükségszerűen kombinálják a fenti megoldásokat ilyenné – ez persze hatalmas energiabefektetéssel jár, és ez a többletmunka némiképp rontja a hatékonyságot.

A fejlesztések másik célja hogy az információszoftvertől (jelentések készítése) elmozduljon az üzleti intelligencia, az együttműködésre képes, döntés-központúság felé (DCBI – Decision-centric Business Intelligence) (*FENG 2009*). A piacon megtalálható legtöbb megoldás a főként OLAP alapokon működő, történeti adatelemző és

---

<sup>25</sup> A BI-piac kifejezés alatt a szakirodalom (és főként szakmai cikkek), mint az üzleti intelligenciát támogató és segítő, vagy kapcsolódó szolgáltatást nyújtó szoftverrendszerek piacát érti.

<sup>26</sup> Software bloat vagy bloatware: felesleges funkciókkal teletűzdelt szoftver.

jelentéskészítő eszköz. Ez persze önmagában kevés a döntéstámogatóhoz. Amire szükség volt, és meg kellett hogy jelenjen a problémát orvosolandó, az a prediktív elemzés, valamint a valós-idejű üzleti monitoring. Az utolsó generációs BI-eszközökben már megtalálhatóak ezek a megoldások (legalábbis a prediktív elemzés mindenképp), így lehetőség nyílik a döntéseket megelőzően minták keresésére, modellek kialakítására, valamint a döntési kimenetek szimulálására.

Az üzleti intelligencia<sup>27</sup> (BI – Business Intelligence) az információtechnológiai piac azon jól körülhatárolható területét jelenti, mely szoftvereket lekérdezésekre, jelentéskészítésre, OLAP megoldásokra, adatbányászatra illetve statisztikai célkora nyújt. Felfogható úgy is, hogy az üzleti intelligencia az elemző piac egy szűkebb rétege, mely magába foglalja az adattárházak és BI-eszközöket, valamint egyéb, kapcsolódó alkalmazásokat – ilyenek lehetnek a CRM-megoldások, az ellátási lánc menedzsmenttámogató eszközök, a pénzügyi tervezés valamint a munkaerő-elemzés; mindezt azon cél érdekében, hogy adatokat tudjuk követni, modellezni és végső soron döntések meghozatalához segítségként felhozni (GALLIER, LEIDNER ÉS DOROTHY 2004).

### 3.1 Az elemzés ciklikussága

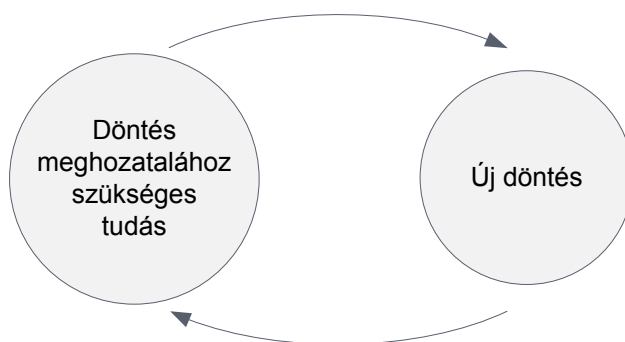
Az elemzés körforgása nem jelent mást, mint a folyamatosan finomodó döntési modellek fejlődését [7. ábra]. Vannak olyan rutinszerű döntések, amelyeket egyszerű szabályrendszer alapján könnyű automatizálni (SHTUB 1999) (jó példa erre a készletnyilvántartás során történő készletfigyelés – ha a készlet bizonyos mennyiség alá csökken, rendelni kell). Más a helyzet azonban az olyan döntésekkel, amelyek egyrészt több tényezőtől függenek, másrészt nem egyszerű és/vagy kapcsolat van a döntési feltételek közt. Az ilyen összetettebb problémák esetén fennálló döntési helyzetek megoldására jó megoldás a folyamatosan finomodó („tanuló”) modellek alkalmazása (TSYPKIN 1973).

Ilyen esetben a lehető legjobb lefedés érdekében az új eredmények kidolgozása során figyelembe kell venni az addigi eredményeket (és azok valóságát) is. Aztán a következő

---

<sup>27</sup> A magyar megnevezés – üzleti intelligencia – az angol business intelligence szóból került fordításra. Megemlíteném ugyanakkor, hogy a fordításban sajnos eltérés van a hazai és az angol szavak jelentése között: eredetileg az elnevezés arra utalt, hogy a folyamat során elemzés és megfigyelés segítségével lehet majd új információkat szerezni; a magyar elnevezésből ez már nem érezhető; hibásan inkább készségek meglétére utal.

döntés meghozatala során az előzmények kiértékelése alatt ez az eredmény is (ugyan csekély mértékben csak) befolyásoló tényező lesz.



forrás: saját munka

7. ábra: "Tanuló" döntési modell

Jó példa erre az adatbányászat során bevett azon technika, mely szerint minden elemzést több féle képpen célszerű elvégezni, majd azok eredményeit megversenyeztetni, a legjobbat kiválasztani és ezt a döntést feljegyezni. (Bevett technika tanító-, ellenőrző-, és teszhalmazokra bontani a vizsgált sokaságot, majd az összes kiválasztott elemző technikát mindegyiken ugyanúgy elvégezni, aztán a kimeneteket összehasonlítani, és így könnyebb eldönteni, melyik technika alkalmas leginkább az adott helyzetben az elemzés (pl.: minősítés) elvégzésére.)

### 3.2 A matematika szerepe

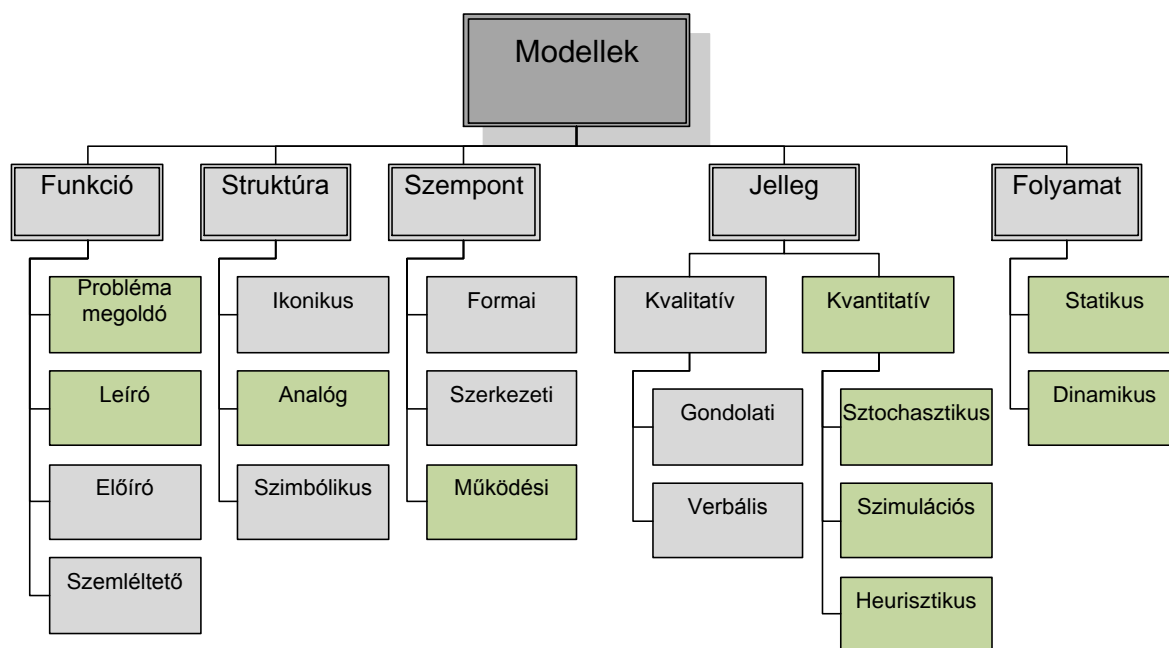
Mivel a vállalatirányítás során hozott döntések nagy többségének komoly anyagi vonzata van, ezért különösen nagy figyelmet kell fordítani a modellek precizitásának, mert a modellből származó eredménye lesz a későbbiekben a döntés alapja (PACZOLAY 1973).

A modell szónak sokfajta definíciója adott, ám a téma szempontjából leginkább releváns megfogalmazást (meglepően nem a Vállalatgazdasági, hanem) a Biológiai lexikon adja: „Rendszerint bonyolult, részleteiben nem ismert fizikai, kémiai, biológiai stb. rendszerek működésének magyarázatára készített sematikus elképzelés, amelyből új összefüggésekre lehet következtetni, vagy amely alkalmas arra, hogy a rendszer jelenségei matematikailag leírhatóak legyenek. A modell a valódi rendszereknek többnyire csak főbb tulajdonságait tükrözi, egyszerűsített formában. A jól szerkesztett modell mégis alkalmas a jelenségek

vizsgálatára, következtetések vonhatók le belőle a modellezett rendszer viselkedését illetően.” (STAUB 1975)

### 3.2.1 Struktúra

A modellek csoportosítása többféleképpen lehetséges. Az egyik megközelítési mód, ha a modellezett folyamat (gazdasági) az alap. Ekkor a modelleket típus szerint csoportosítani lehet azok funkciója, struktúrája, szempontja, jellege és folyamatai alapján [8. ábra].



forrás: Kerékgyártó, Mundruczó és Sugár 2001 alapján saját munka

8. ábra: Modellek csoportosítása típus szerint

Funkcióját tekintve egy modell lehet problémamegoldó, leíró, előíró vagy szemléltető. Az üzleti folyamatok matematikai modelljei általában leíró természetűek (de minősíthetők probléma megoldónak is alkalmazásuk során történő bővülésük miatt).

Struktúra szerint lehet ikonikus, analóg vagy szimbolikus – általában analóg modelleket célszerű készíteni.

Szempont alapján megkülönböztethető formai, szerkezeti illetve működési modell. Ebből a szempontból működési modellek a relevánsak.

Jelleg alapján kvalitatív (gondolati vagy verbális) és kvantitatív (sztokasztikus, szimulációs és heurisztikus) modellek megkülönböztethetők. Általában kvantitatív

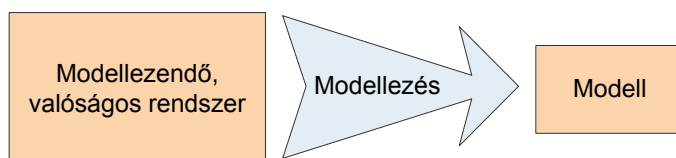
modelleket célszerű használni, még akkor is, ha a modell tartalma valamilyen kvalitatív információ.

Folyamataikat tekintve a modellek lehetnek statikusak vagy dinamikusak. Míg a statikus modellek állandóak, a dinamikus modellek képesek a változásra, finomodásra illetve a tanulásra, ennek ellenére az egyszerűbb rutinszerűbb feladatokhoz tökéletesen megfelelőek a statikus modellek is.

### 3.2.2 Modellfejlesztés

A modell általában létező tárgyak, folyamatok, rendszerek célszerűen egyszerűsített megfelelője: olyan fizikai vagy csupán leírt térbeli konstrukció, amely sematikus formában, vagy egyezményesen elfogadott jelölésekkel mintáz le, illetve ábrázol bizonyos tárgyi területet, folyamatot vagy rendszert. Minél bonyolultabb a rendszer, annál nagyobb szükség van modellezésre (BODLAKI 2004).

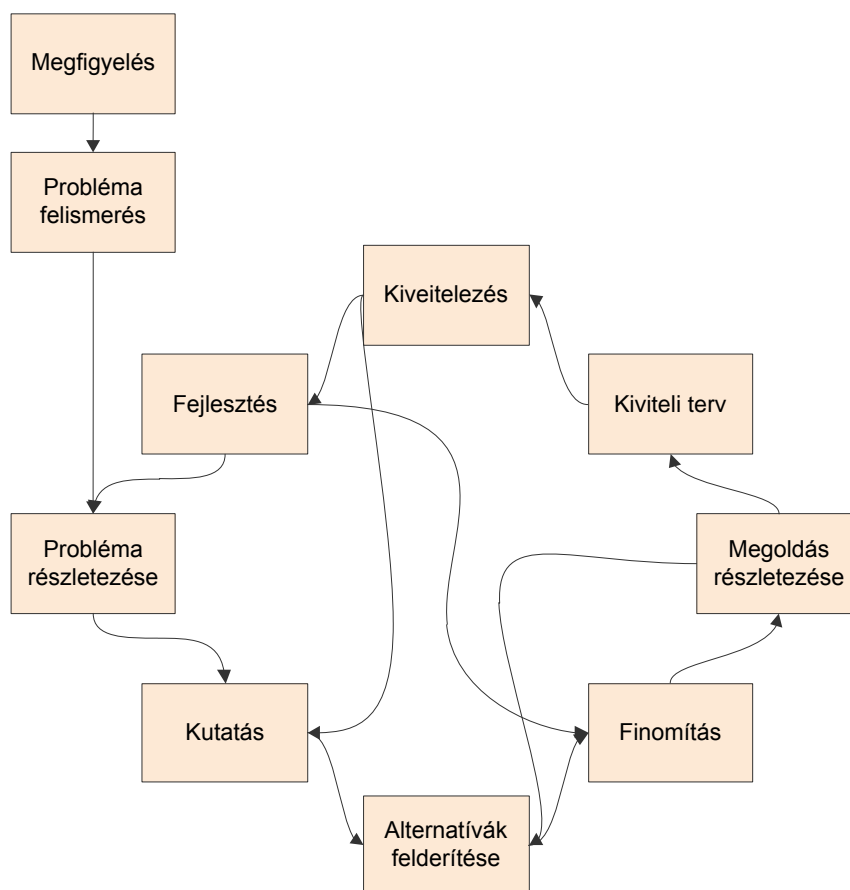
A fenti definíció alapján a modellezés a modellek létrehozásának folyamata. A modell minden esetben hasonló lesz a modellezetthez (örökli annak releváns tulajdonságait), vagyis közöttük hasonlósági reláció lesz jelen [9. ábra]. Megállapítható továbbá, hogy minden modell információt adó rendszer lesz (STAHLY, KISCHKA 2001).



forrás: saját munka

9. ábra: A modellezés folyamata

A modellezés kiindulópontja minden esetben egy megfigyelés [10. ábra]. Legtöbb esetben a megfigyeléssel egy időben, vagy némileg utána felmerül az igény, mi szerint a megfigyelés valamilyen matematikai módon jellemezhető (FODOR ÉS ROUBENS 1994). A probléma felismerés jelenti annak a kérdésnek a megfogalmazását, amit majd a modell kifejlesztése és alkalmazása után annak segítségével megválaszolhatunk.

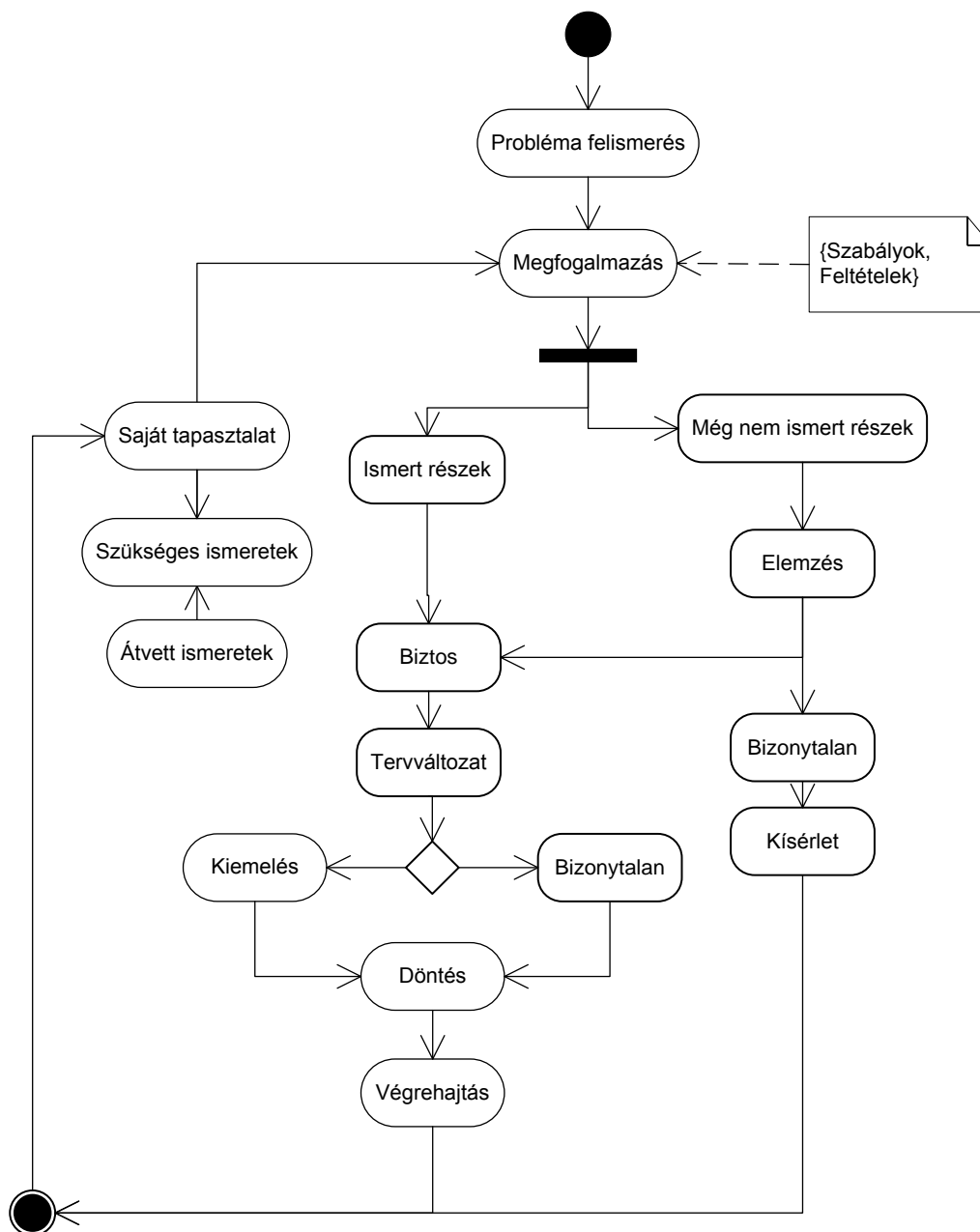


forrás: Fodor és Roubens 1994 alapján saját rajz

10. ábra: A modellfejlesztés folyamata

Ha a probléma felismerése megtörtént, a folyamat innentől kezdve ciklust alkot. A probléma részletezése során kell megfogalmazni az elvárt eredményeket. A kutatási fázis során kell megvizsgálni, hogy a probléma megoldása miként lehet megoldható, ahhoz milyen feltételek szükségesek, illetve a részletezés során megfogalmazottakra miféle megoldás adható. Az alternatívák felderítése abban az esetben lényeges, ha a problémára több egymástól eltérő megoldás adható. A finomítás során kell kiválasztani, hogy adott szempontok alapján melyik alternatíva a legelőnyösebb. A választás után részletezni kell a megoldást (szakterületi szabályok szerint – új információ felderítése esetén sok esetben adott a lehetőség az alternatívákhoz történő visszacsatolásra), majd kiviteli tervet kell készíteni (ugyancsak szakterületenként eltérő). A kivitelezés szakasza a végrehajtás. Amennyiben itt váratlan korlátok merülnek fel, ugyancsak van lehetőség a visszalépésre (általában a kutatási fázishoz, mert ott adott a lehetőség a korlátok legyőzésére). A ciklus utolsó lépése a fejlesztés, a modell kidolgozása, alkalmazása, a tapasztalatok levonása. A

leggyakoribb innen a finomításhoz való visszanyúlás, és a gyakorlati tapasztalatok elméletbe implementálása (GACHET 2004).



forrás: saját munka

11. ábra: A problémamegoldás folyamata

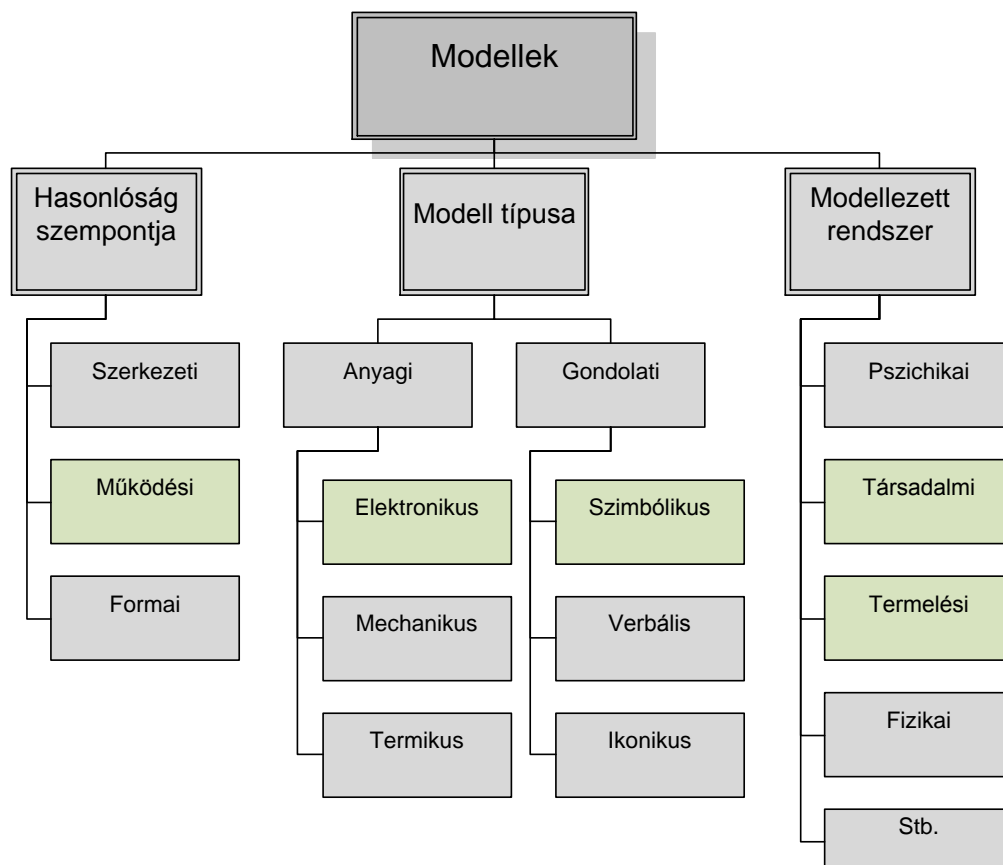
A modellfejlesztéshez hasonló folyamat a problémamegoldás elméleti folyamata [11. ábra]. A problémamegoldás is – hasonlóan – felismeréssel kezdődik. A pontos megfogalmazáshoz szükségesek a környezetből kapott szabályok és korlátozó feltételek is. Ezzel szokott a legtöbb esetben párosulni a tapasztalati úton megszerzett tudás (korábbi, hasonló problémák megoldásából), valamint egyéb, átvett ismeretek. Amennyiben a

probléma megoldása kizárólag ismert részeket tartalmaz, akkor készíthető megoldási terv (amennyiben vannak nem ismert részek, akkor kutatásra, vagy kísérletezésre van szükség). Végül a terv finomításával, az elindítási döntés meghozatalával és a terv végrehajtásával a probléma megoldódik.

### 3.2.3 Modellek fajtái

Modellezési technika alapján (amivel történik a modell létrehozása) háromféleképpen történhet a csoportosítás: lehet a hasonlóság szempontja vagy a modell típusa vagy a modellezett rendszer alapján [12. ábra].

A témának megfelelő matematikai modell legfontosabb tulajdonsága, mely a gazdasági folyamatokat fogja reprezentálni (amire a későbbiekben döntéseket lehet alapozni) az, ha a hasonlóság szempontja szerint a modell és a modellezett rendszer közt működési hasonlóság van – ez elengedhetetlen a döntések során.



forrás: forrás: Kerégyártó, Mundruczó és Sugár 2001 alapján saját munka

12. ábra: Modellek fajtái



Mivel azonban vállalati információrendszerekről van szó, ezért a felhasználás során kétfajta modellt kell megkülönböztetni: az egyik a már említett matematikai modell (mely megköveteli a működési hasonlóságot), amely a vállalat külső folyamataival foglalkozik, a másik a vállalat információrendszer modellje, mely a belső folyamatokat veszi alapul (ebben ugyancsak fontos a működési hasonlóság).

A matematikai modell fejlesztéséért közgazdászok, matematikusok és statisztikusok felelősek, míg az információrendszer modelljét elsősorban gazdasági informatikus szakemberek készítik (BOSTROM, FRENCH ÉS GOTTLIEB 2008).

*„Azt a módot, ahogy a valós rendszer működését feltárjuk, elemeiről, jellemzőiről, szerkezetéről, környezetével való és elemei közötti kapcsolatáról ismereteket szerzünk, majd ezeket az ismereteket felhasználva egy, a rendszer céljait megvalósító, a számítógépek által is kezelhető rendszert képezzünk, információrendszer-modellezésnek nevezzük” (RAFFAI 2003)*

### **3.2.4 Eszközök**

A döntési modellek megalkotása során a siker érdekében elengedhetetlen felsőbb matematika alkalmazása.

Mivel a döntések során a tényezők változónak tekinthetők, és a kimeneti érték (mely a döntés közvetlen alapja lesz) ezek függvényében változik, jól használható matematikai terület a függvény-analízis és a lineáris algebra.

Mivel sok olyan döntési helyzet is elő fog fordulni, ahol nem lehet egyértelmű döntést hozni, fontossá válik a döntéstámogatás során a valószínűségszámítás – nem hiába a döntéstámogatás egyik ága a (LEVIN, ZAHAVI ÉS OLITSKY 1995) valószínűségvezérelt döntéstámogatás (probability-driven decision support).

A diszkrét matematika használata főként gráfelmélettel kapcsolatos döntési hierarchiák során hasznos (Gráf-alapú modellezés) (MAYR, SCHMIDT ÉS TINHOFFER 1995).

Ugyanakkor mindezen részek mellett legnagyobb súllyal a matematikai statisztika kap szerepet, mivel a nagy mennyiségű (és egyre csak növekvő mennyiségű) adathalmaz feldolgozásához, összefüggések és trendek megállapításához ez elengedhetetlen.

Az üzleti intelligencia (és így az adatbányászat is) számos esetben szembesül optimalizáció révén megoldható problémával. Az optimalizálás a lehető legjobb állapot vagy akció kiválasztására alkalmas eljárás, előre meghatározott feltételrendszer alapján (*KOMLÓSI 2001*).

Az optimalizálás a problémamegoldás során több helyen merülhet fel: Gyakori elemhalmaz keresése esetén (pl. APROIRI algoritmusok), osztályozás során (pl. regresszió, neurális hálózatok, fák, tartó vektor gépek), klaszterezés és elemkiválasztás során (pl. genetikai algoritmusok). Az optimalizálási eljárások alapvetően három csoportra oszthatóak: gradiens alapú módszerek (ezen belül megkülönböztethetők direkt és indirekt módszerek), véletlent használó módszerek (eszközei a tabukeresés, a szimulált lehűlés és az evolúciós algoritmusok – mely tovább bontható evolúciós stratégiákra (a vállalatirányítás egyik eszköze) és genetikai (soros illetve párhuzamos) algoritmusokra) és kimerítő keresésen alapuló módszerek (ennek eszköze a dinamikus programozás) (*MINIEKA 1978*).

A genetikai algoritmusok azon adaptív, lágy számítási módszerek gyűjtőneve, melyek a darwini evolúció elvére épülnek és főként keresési és optimalizálási feladatok megoldására használhatóak (*ÁLMOS, GYŐRI, HORVÁTH ÉS VÁRKONYINÉ 2002*). Olyan módszer, mely az élőlények öröklődési tulajdonságait figyelembe véve (előnyös tulajdonságok öröklődnek), egyszerre többpontos keresést végez és ennek megfelelően több közel-optimalis megoldást tud nyújtani (még akkor is, ha bizonyos keresési ágak zsákutcának minősülnek).

Szimulációs eljárásokkal számos megalkotott modell minősége ellenőrizhető. A szimulációt főként tömegkiszolgálási és forgalmi problémák modellezése során használják – ennek oka, hogy ezekben a modellekben olyan sok változó és ismeretlen (hatás és összefüggés) szerepel, hogy matematikai leírása emberileg áttekinthetetlen lenne, illetve időben nem készülne el (a szimuláció fekete-dobozos esete). Ettől függetlenül persze használható a szimuláció fehér-dobozos esetekben is, amikor minden részlet ismert; ekkor a cél a modell viselkedésének megismerése (*JÁVOR 1979*).

### **3.3 Etikai kérdések**

Bár a jogi szabályozás meglehetősen kiterjedt, és igyekszik jól szabályozni az adatok felhasználásának módját és feltételeit, az internet adta adatterjedéssel nem igazán képes lépést tartani.

Minden adatbányászati alkalmazás (független attól, hogy mik a célok) az adatok összegyűjtésével kezdődik. Ma a technikai lehetőségeknek köszönhetően mindenki rengeteg adatot oszt meg magáról mindennapi élete során: a vásárlások során kártyával fizet, esetleg törzsvásárlói kártyával rendelkezik, az üzletekben kamerák vannak, melyek figyelhetik a vásárlók cselekedeteit, viselkedésüket, telefonhívások esetén ismerhető azok célja és időtartama, marketing jellegű hívások esetén a válaszok mellett, a beszéd (hangsúly, hanglejtés, idők) elemzésével további viselkedési adatok szedhetőek össze, és akkor még nem esett szó az internetes tevékenységekről, mint az on-line vásárlási vagy böngészési szokások ismerhetősége, a levelek, fórumok beszélgetések tartalmának vizsgálhatósága. Ezek olyan információk melyek kikerülését nagyon nehéz is lenne megakadályozni. Az adatvédelmi törvény úgy fogalmaz, hogy előre ismertetni kell a célt, amire majd az adatokat fel szeretné használni a vállalat. Sok esetben ennek letiltása sem lehetséges, valamint rengeteg esetben nem is tudható előre a cél sem (mivel adatbányászat során összefüggéseket is kereshet a cég). Ez a rengeteg adat mind összegyűjthető a személyekről – vannak kifejezetten erre szakosodott cégek is.

Az elemzési folyamatban előbb-utóbb sorra kerül valamilyen osztályozási művelet. Etikailag kifogásolható, az emberek adatok alapján, de ismeretség nélkül a nagy számok törvényei alapján történő minősítése. A döntésautomatizálással kapcsolatban ez etikai kérdéseket vethet fel – szabad-e az embereket adathalmazként kezelni (a döntési fák használata némiképp jobb és kevésbé diszkriminatív). A probléma különösen akkor érezhetőbb, ha kevés attribútum áll rendelkezésre (a jó adatbányászati lehetőségek sok attribútum jelenléte esetén használhatóak ki). Ide kapcsolódik a diszkrimináció kérdése is, valamint a kényes adatok kezelése. Mérlegelendő, hogy fel szabad-e használni vásárlói halmazok kialakítására a vallási, nemi, etnikai hovatartozásra vagy egészségi állapotra utaló adatokat (és még szóba sem került az, hogy vannak olyan adatok is, melyekből nagy biztonsággal lehet következtetni ezekre). Figyelni kell a negatív és pozitív diszkriminációra egyaránt.

Napjainkban, a technológiai fejlődésnek és a technológia egyre inkább elérhetővé válásának (árak csökkenése, tömeggyártás és új gyártási technikák használata) köszönhetően a technológiát és az üzleti életet elválasztó határvonal egyre inkább nemhogy vékonyodik, hanem homályosabb lesz. Köszönhető ez a számos sikertörténetnek és azon vezetőknek, akik észreveszik és elismerik, hogy mekkora jelentősége van az információnak – legyen szó külső vagy belső információról. Ezen felismerés és egy hosszú

evolúciós folyamat eredménye az üzleti intelligencia megszületése lett: ez magába foglalja azon eszközöket, módszereket, folyamatokat, technológiát és gyakorlatokat, melyet segítségével a nyers adatok felhasználható tudássá válhatnak (*LOSHIN 2003*).

Az új szakterület megszületéséhez ötvözni kellett az üzleti élet és az információtechnológia legjavát: egyfelől meg kell érteni, hogy milyen információk kellenek a vállalat működéséhez, azokat honnan lehet beszerezni, milyen módon, hogyan lehet segítségükkel a munkát gördülékenyebbé és hatékonyabbá tenni, ehhez érteni kell az üzleti folyamatok működését és felépítését; másfelől ismerni kell mindezen kérdések megválaszolására alkalmas technológiákat, technikai megoldásokat, azok felépítését és működését. Mindennek tetejébe – ebből a szempontból – nincs két egyforma vállalat, minden eset és probléma valamelyest különböző (ez persze nem jelenti, hogy ne lennének értelmezhetőek tervezési minták).

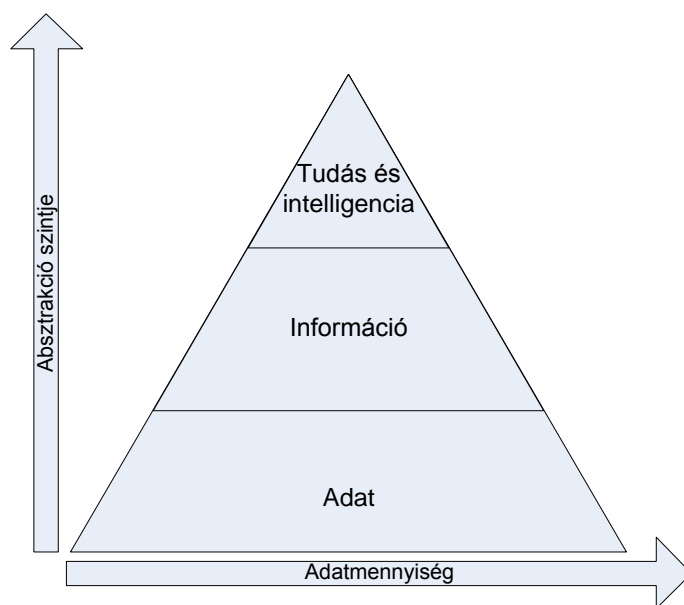
Annak ellenére, hogy vállalatok tömegei költenek hatalmas összegeket üzleti intelligencia keretrendszerek<sup>28</sup> és alkalmazások építésére, fejlesztésére és indítására, a befektetett összeg nagy része az infrastruktúra fejlesztésére megy, míg sokkal kevesebb jut a vállalati adatok menedzselésére és kiaknázására (*RABIN ÉS JACKOWSKI 1988*). Ennek egyik oka, hogy az üzleti intelligencia fogalma nagyon lazán van csak meghatározva, így rengeteg vállalatvezető úgy értelmezi a fogalmat, ahogyan éppen neki az utolsó szoftver vagy hardvergyártó/terjesztő mondta. Annak a ténynek, hogy az üzleti intelligencia bevezetések nagy százaléka sikertelen, ez az egyik oka. A sikertelenség megnyilvánulhat a projekt törlésében, a költségkeret vagy a határidő túllépésében vagy a követelmények nem teljeskörű teljesítésében. A jó meglátás persze az, ha a vállalat elkezd adataira úgy tekinteni, mint egy értékes erőforrásra / eszközre (*LUFTMAN ÉS BULLEN 2004*), és felismerik, hogy a megfelelő gyűjtés, aggregáció és elemzés olyan információkat fedhetnek fel, amik akár jelentősen is javíthatják a vállalat teljesítményét.

Éppen ezért, mindenkinek, aki a vállalaton belül a BI kezdeményezések hatósugarába kerül, fontos értenie, hogy igenis olyan információkat kell gyűjteni, ami később a vállalat működését jobbá teheti (*HOSNI ÉS KHALIL 2004*).

---

<sup>28</sup> A keretrendszer nem más, mint egy olyan “építőelem”-gyűjtemény, melynek elemeit később (alkalmazásfejlesztés során) felhasználva gyorsabban végezhető el a feladat, egységesebb eredményt kapva. Önmagában persze egyetlen keretrendszer sem használható, de adott típusú feladatok véghezvitelét jelentősen képes megkönnyíteni., ezért javasolt használata.

Az információk hatékony gyűjtése, feldolgozása és értékelése nem egyszerű feladat. Mint ahogyan az absztrakciós piramis is rámutat, a rendelkezésre álló adatok mennyiségétől függ a megszerezhető információk és megszerezhető tudás mennyisége, de minden adag így elérhetővé tett tudáshoz rengeteg adatra van szükség.



forrás: saját munka

13. ábra: Absztrakciós piramis

Az absztrakciós piramis [13. ábra] értelmezéséhez (KOVÁCS 2004) tisztázni kell néhány alapvető fogalmat:

Az adat lehet bármilyen mérési eredmény, észlelés, ismerv, hír vagy közlemény (RAFFAI 2003).

Az információ az adatok értelmezéséből adódik. Megfogalmazható úgy is, hogy az adatok gyűjtésének és oly módú rendezésének eredménye, mely az adatok közti kapcsolatokra fényt derít, azokat tartalommal és értelmezéssel ellátva.

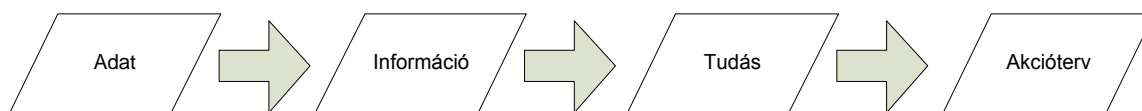
A tudás<sup>29</sup> valamilyen tárgykör ismeretét jelenti, beleértve az információkat, tényeket és leírásokat. A tudás megszerzése persze történhet oktatás vagy tapasztalat által – jelen esetben a tapasztalat által megszerezhető tudásra kell koncentrálni. A tudás megléte annyit tesz, hogy a tárgykört elméleti és gyakorlati oldalról egyaránt értjük. Tudás továbbá az

---

<sup>29</sup> Mivel a tudás számos területen használt és értelmezett szó, ezért jelentését az értelmezési környezethez kötetten érdemes megadni – itt a tudás információtechnológiai szempontból kerül értelmezésre.

információk megértésén és azokban minták felismerésén alapuló elv, mely mély belelátást biztosít az információkba.

A folyamat végeredményeként előálló tudás [14. ábra] hasznosításának módja az, ha akcióttervet készít a vállalat, majd meg is valósítja azt.



*forrás: saját munka*

*14. ábra: Az üzleti intelligencia folyamata*

Az a folyamatrész, mely során az adatokból információkat képezünk, megfogalmazható úgy is, hogy meg kell határozni, melyik adatot, milyen összefüggésben kell gyűjteni, illetve kezelni. A környezetből kiemelve (vagy környezet nélkül) az adatok kevésbé értékesek (ilyen lehet egy vásárlói névhalmoz), ám megfelelő környezetben gyűjtve, azokat értelmezéssel ellátva sokkal nagyobb értéket nyernek (vásárlói adatok összegyűjtve: nevek, hozzájuk kapcsolódó vásárlások, egyéb értékek) (DAVENPORT ÉS PRUSAK 2000).

Elmondható, hogy a BI-folyamatok közül főként az adatok kezeléséről és megjelenítéséről van szó, mely fizikai megjelenésüket tekintve főként hardverplatformokat, relációs adatbázisokat és azokat kezelő szoftvereket jelentenek – funkciójuk tekintetében lekérdezések futtatására és jelentések<sup>30</sup> készítésre van szükség. Ehhez a fázishoz az emberigény elsősorban adatkezelésben, adatintegrációban és koordinációban jártas szakemberekben merül ki.

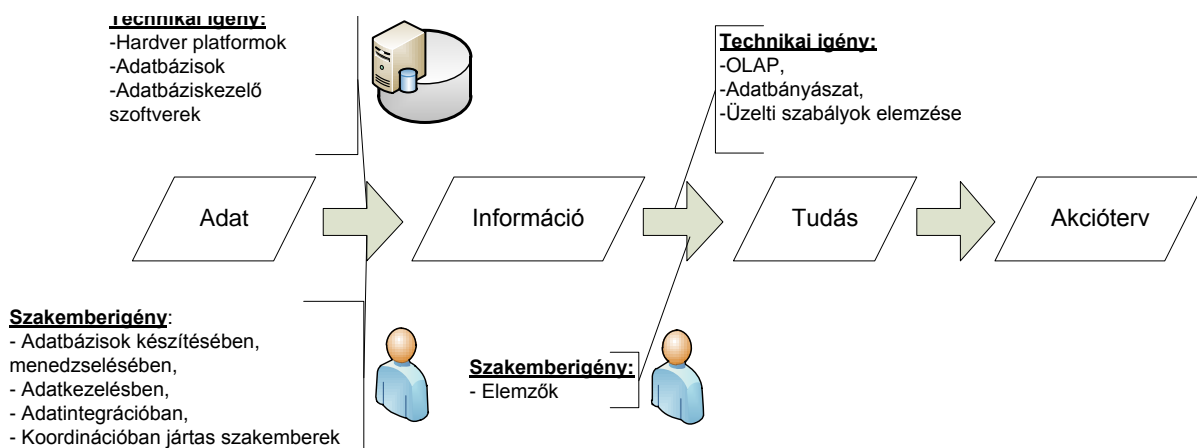
Ha a vállalat odáig eljutott, hogy rengeteg információval rendelkezik, akkor azokat többféle módon kell elemezni, hogy abból tudás nyerhető legyen. A nyert tudás felhasználható arra, hogy akcióttervek készüljenek belőle, majd azokat megvalósítva megoldást nyújtsanak valamilyen problémára.

Itt elmondható, hogy a BI-folyamatok közül itt főként az elemzést támogató komponensekre van szükség [ forrás: *saját munka*

15. ábra], mint amilyenek az adattárházak, az analitikai feldolgozás (OLAP), az adatbányászat, az üzleti szabályok elemzése, az adatok körvonalazása és minősítése. Itt a

<sup>30</sup> A jelentés itt az angol report szóból fordított riport szinonímája; szakemberek szerint előnyösebb a jelentés szó használata, mivel így az anyanyelvi szaknyelv kialakítható és mélyíthető.

szakemberigény elsősorban olyan szakemberekben nyilvánul meg, akik ezeket az informatikai eszközöket hatékonyan tudják kezelni, hogy kell a megfelelő eredményeket kinyerni – e nélkül semmit sem érnek a drága szoftverek.



forrás: saját munka

15. ábra: Üzleti intelligencia igényei

Valószínűleg az utolsó fázis a legfontosabb lépés, itt keletkezik a valódi, kézzel fogható és alátámasztható tartalom (tudásból akcióterv), valamint ez az amire a vezetés is vágyik. Az egész üzleti intelligenciára alapozott vállalati stratégia kulcsfontja a reagálás képessége. Ezen keresztül lesz megállapítható az információtechnológiára fordított beruházás megtérülése is. Az akciók segítségével növelhető lesz az üzleti hatékonyság, az eladások, javítható a vásárlók megcélzása, csökkenthetőek a szolgáltatások költségei, felismerhetőek a csalások és összességében a költségek csökkentésével nőhet a profit.

A PriceWaterhouseCoopers elemzése (*PRICEWATERHOUSECOOPERS 2001*) szerint azok a vállalatok, akik az adataikat stratégiai erőforrásként kezelik és hajlandóak pénzt fektetni az adataik minőségébe, már most érezhető előnyre tesznek szert reputáció és termelékenység területén. A kijelentést továbbgondolva megállapítható, hogy az adat (és a folyamat egyik terméke az információ is) igenis stratégiai erőforrásként kezelendő. A hagyományos felfogás szerint az adat nem más mint az automatizálendő üzletmenet hajtóanyaga, arra utalva, hogy a vállalatnál már informatizálva van a munka. Az előre tekintőbb felfogás ugyanakkor az, hogy az üzletmenet javításához szükséges stratégiai tudás a felhalmozott adathalmazban van, onnan kell kinyerni. Ebből viszont ismételt újabb következtetés vonható le: a vállalatoknak magukat mindinkább információs üzletként kell, hogy lássák, a

tradicionális vertikális megközelítés helyett. A folyamat persze hosszadalmas, melyben az adat stratégiai erőforrássá változik.

A legtöbb erőforrásra (mint például emberi tudás, tárgyi eszközök vagy akár informatikai erőforrások is) igaz, hogy a használat növekedésével tapasztalható valamilyen értékcsökkenés (az erőforrás értékében). Ez a tulajdonság az információra nézve azonban nem igaz, annak értéke nem csökken megtekintéseinek számával. Ám, ha mindenki tudja, hogy melyik információ hol és milyen módon érhető el a vállalaton belül, akkor annak értéke hirtelen növekedni kezd. Az adatok is, ha tárolva vannak, menedzselve, de azok sosem lesznek felhasználva, akkor nem is lehet hozzáadott értékről beszélni, inkább terhessé válnak.

Az nyilvánvaló, hogy pontos (jó minőségű) információra támaszkodva jobb és több profitot termelő döntéseket lehet hozni. Persze nem csak azt kell elvárni egy információtól (ennek kapcsán az információforrás megbízhatósága is fontos), hogy az jó minőségű legyen, hanem rendelkezni kell metrikákkal, annak minőségének mérésére. Ha léteznek ilyen metrikák, és rendszeres mérések az adatok minőségét illetően, akkor a döntéshozó már tud döntései során kockázattal is számolni.

Az érték persze további növekedést mutat azzal (*LAWRENCE 1999*), ha a több forrásból származó információkat kombinálni lehet (kereskedelemben: például az eladási oldal és a beszerzési oldal információ alapján az áruk mozgásának üteme számítható). Ez a kombinálás alapján további információkhoz jutás az egész BI folyamat egyik kulcsa. Elmondható tehát hogy ha az információk kombinálásával olyan új tudás (ismeret) áll elő, amely anélkül nem került volna napvilágra, akkor sikerrel volt növelhető a kompetitív előny.

Szintén ellentétesen a többi erőforrással, az információ mennyiségének növelésével nem feltétlenül növekszik annak értéke (*COSTA, SUBIRANA, VALOR-SABATIER ÉS VALOR 2004*). Némelyest alátámasztja ezt az a tény is, hogy ha az évente keletkező hihetetlen információmennyiséget kellene integrálni a már meglévő halmazzal; ez hatalmas feladatot jelentene. A másik pedig: ha egy személynek egyre több és több információt adunk egyszerre, a mennyiséggel fordítottan arányosan csökken a megjegyzett adag.

Az adatok integrálásának komplexitása meredeken növekszik az információforrások számának emelkedésével. A komplexitás növekedése persze nem csak a számszerű



növekedésből adódik, hanem belejátszik az információforrások diverzitása, a szerzett információ természete és értelmezhetősége is. Ennek részben az az oka, hogy a különböző forrásból érkező adatok csak nagyon ritkán használhatóak ugyanazon adatmodellen, reprezentációjuk is eltér. Minden új adathalmazt, amit hozzáadunk, minden egyes már a repozitóriumban meglévő halmazzal integrálni kell, ez pedig sok megoldandó problémát okoz.

Másfelől persze az emberek felfogása olyan, hogy minél több az információ, annál jobb. De azért meghúzható (még ha homályosan is) az a határvonal, amely elválasztja a már károsan sok információt a megfelelő mennyiségűtől. A megfelelő mennyiség úgy definiálható: az a mennyiség, mely kielégíti a vállalkozás követelményeit, feltéve, hogy jól és időben integrálható.

Egy másik mennyiségi kérdés az, hogy az azonos forrásból származó nagy adatmennyiség-e a jobb, vagy a sok forrásból származó, azonos mennyiségű – ez a felhasználás módjától függ. Így például trendszámításnál, illetve minta felfedezéshez a nagy mennyiségű, azonos forrásból származó történeti adatok a megfelelőek, míg piaci részesedés megállapítása esetén a több információforrás job eredményre vezet.

Az információ értékének mérése persze nem könnyű feladat, de használhatóak a más erőforrások mérésére szolgáló hagyományos modellek (*GELINAS ÉS DULL 2007*):

- Történeti költség: a módszer lényege, hogy összesíti azokat a költségeket, melyek az adatok összegyűjtése és az információ előállítása vagy az információ beszerzése során merültek fel (vagy mennyibe kerülne pótolni azt).
- Piaci érték: a számítás során azt kell figyelembe venni, hogy más mennyit lenne hajlandó fizetni a már meglévő információért. Főként az adatszolgáltatók használják ezt a modellt.
- Használati érték: a számítás lényege, hogy azt az értéket kell figyelembe venni, ami az információ alapján (kialakított akcióterv alapján hozott döntés alapján) előállt.

Mint minden erőforrás, az információ csak akkor hozhat hasznot, ha felhasználásra kerül. Minden BI-folyamat az adatokat információvá majd tudássá változtatja, ám ez a tudás is csak akkor hozhat hasznot, ha az alapján megfelelő cselekmények zajlanak le. Ez persze nem technológiai kérdés, mindinkább szervezeti: a tudás megléte egy dolog, ám a cselekményt vagy akciót, amit a tudás birtoklása kivált meg is kell tervezni, szervezni,

ehhez pedig jó szervezeti felépítésre, gyors reagálásra, és olyan emberekre van szükség, akik képesek ezt végigvinni. Ezért, minden BI-programnak része kell legyen valamilyen megtérülés számítás.

Ahhoz hogy ez jól mérhető legyen, elemezni kell a költségeket, a bevételek alakulását, különösen azokat, melyek közvetlenül a cselekményhez köthetőek (*KAPLAN ÉS COOPER 1998*). Ehhez szükséges:

- A fix költségek már eleve a BI infrastruktúra részei legyenek (hardverek és szoftverek beszerzése, adatbázis kialakítás);
- A változó költségeket tevékenységekhez kell kötni (különleges mérőberendezések);
- Ezen tevékenységek folyamatos költségeit jegyezni kell;
- A cselekvés eredményeként létrejövő hasznosság értékének megadása;
- Más területek azon költségei, amelyek adott aktivitás lebonyolításához szükségesek;
- A cselekménytől elvárt értékmodell kialakítása;
- Annak valószínűségét kell számolni, hogy mekkora eséllyel alkalmazhatóak ezen cselekmények úgy, hogy az elvárt értékmodell teljesüljön;
- Megtérülési időket kell számolni a jövedelmezőség méréséhez.

Jó példa a fentiekre a következő cél: olyan CRM-rendszer és adattárház kialakítása, melynek célja az összes információ összegyűjtése minden vásárlóról, azért, hogy kitolhassa a vállalat a vásárlói életciklust. Az adattárház megépítés több költséget jelent: hardver és szoftvervásárlás, az adatbázis megtervezése és kivitelezése, néhány szoftvereszköz beszerzése, valamint mindezek integrálása a vállalatba. A következő költségcsoport az adattárház modellek kialakítása kapcsán van: fel kell tárni az adatforrásokat, meg kell oldani az adattárházba való áttöltés módját, valamint a rendszer karbantartása is rendszeres költségekkel jár. A beruházástól elvárt haszon az, hogy növekedjen a vásárlói életciklus a bevezetéstől számított harmadik évtől. Figyelni kell a költségek alakulását, valamint azt, hogy mikor jön el az a pont, amikor a beruházás és az üzemeltetés költsége kevesebb lesz, mint a vásárlói életciklus növekedése miatt előálló haszon. Ha ez nem következik be, át kell vizsgálni, hogy milyen más érték vonható ki a rendszerből. Ha ez sem elegendő, akkor komoly számítási és tervezési hibák voltak jelen a projekt korai szakaszában.

### 3.4 Üzleti intelligencia alkalmazások

A komplex üzleti intelligencia rendszerek mára lehetőségek egész sorát nyújtják (nagyreszt lefedve a vállalat minden ide kapcsolódó folyamatát), mely mind-mind a vállalati teljesítmény növeléséhez járul hozzá. Ezen eszközök feladatok egész sorát képesek támogatni, legfőképpen olyan céllal, hogy az üzleti / gazdasági folyamatok során előálló döntési helyzetekben az optimális döntés legyen meghozható. Ezek közé az elvégezhető feladatok közé tartozik tehát az adatgyűjtés és tárolás, az interaktív jelentéskészítés, az (esetleg online) analitika, az adatintegráció, az eseménykezelés és az együttműködés a vállalat többi informatikai rendszerével (*VERCELLIS 2009*).

Ezeknek az eszközöknek megvan az a tulajdonsága, hogy képesek hatalmas mennyiségű adat tárolására – a későbbi elemzés céljából. Az itt tárolt (adattárházban) archív és kurrens adatok pedig a vállalat számára igény szerint folyamatosan elérhetőek. Ezzel lehetővé válik minden művelet információszükségletének kielégítése.

Az üzleti intelligencia rendszerekben összegyűjtött adatok igény szerint elemezhetőek és adhatóak ki, segítségükkel jól alátámasztható minden céges kezdeményezés. Az elemzések eredményeinek átvizsgálásával adott a lehetőség a vállalati teljesítmény és okainak vizsgálatára. De az elemzések egyértelműen legfontosabb felhasználási módja a döntések megalapozása kell legyen (*KUDYBA ÉS HOPTROFF 2001*).

Járulékosan ezen eszközök használatával csökkenthető az idő, energia és egyéb erőforrás tervezési ráfordítása. Köszönhető ez az adatok mindenkori rendelkezésre állásának és az azokból különböző elemzési technikákkal készülő jelentések elérhetőségének – így a döntések meghozataláért mindenkor felelős személyek gyorsan és hatékonyan lehetnek képesek következtetések levonására és megfelelő megoldások formálására. Így a vezetők és menedzserek több időt tudnak fordítani azok megvalósítására és ellenőrzésére, biztosítva, hogy a tervek kivitelezése megfelelően alakuljon.

#### 3.4.1 Az üzleti intelligencia bevezetése

Sok vállalat ott kezdi az analitika bevezetését, ahonnan a pénz jön – a vásárlóknál, jobban mondva a vállalati ügyfélkapcsolat menedzsmenttel. A CRM-rendszerekben szereplő ügyfelek aztán szegmentálhatóak, elemezhető a vásárlók viselkedése, megjósolható a vásárlói jövőbeni cselekedetei, ezek alapján a hirdetések és termékek jobban címezhetővé válnak. Biztosíthatóvá válik, hogy a marketingtevékenység az eddig legsikeresebb

kampányokra irányul a legsikeresebb csatornákon. A termékek árazhatósága is változik, mivel kideríthető, hogy egy termékért mennyi az a maximális ár, amit még hajlandóak megfizetni a vásárlók. Segítségével a vásárlómegtartás is hatékonyabban címezhető, időben és térben egyaránt (*MOSS ÉS ATRE 2003*).

A második lépés általában a logisztika szokott lenni, azon belül is az ellátási lánc menedzsment. Itt főként a készletezés és a szállítási utak optimalizálhatóak jól. A cél annak biztosítása, hogy a megfelelő termékek a megfelelő helyen illetve a megfelelő mennyiségben és minőségben legyenek jelen.

A munkaügy – HR – is egyre inkább használja az analitikát. Ezen a területen a két fő direktíva a megfelelő célra a megfelelő emberek alkalmazása, illetve a dolgozói állomány megtartása. Ez a részleg ebből a szempontból meglehetősen ösztönösen működik a mai napig sok cégnél.

Nem meglepő módon a számvitel és pénzügy területén is rendkívül hatékonyan alkalmazható az analitika – köszönhetően a terület numerikus vonatkozásainak. Az elsődleges célok közé tartozik a kockázatok figyelése és csökkentése (ez főként a hitel- és biztosítóintézeteknél fontos) valamint az ok-okozati összefüggések vizsgálata és a trendek felfedezése.

Bár egyértelműen kijelenthető, hogy a legtöbb szervezet számára jövedelmező az analitikus megközelítés a döntéshozatalban, mégis vannak olyan helyzetek, amikor felesleges ennek a látásmódnak az erőltetése.

Az egyik tipikusan ilyen helyzet, ha nincs elegendő idő az adatok összegyűjtésére (az elemzésre és kiértékelésre már nem szükséges számottevő idő a valósidejű döntéshozatali lehetőségek miatt), vagy ha a döntésnek még az adatok beérkezése előtt kell megszületnie (*BIERE 2003*).

Abban az esetben ha még nem történt precedens értékű helyzet, ugyancsak nem ajánlható, mivel a még be nem következett eseményekről nagyon nehéz adatot találni. Ebben az esetben, ha mégis szeretné a szervezet igénybe venni a technológiát, célszerű egy kicsi, de véletlenszerű mintán lefuttatni a felmerülő ötleteket – ettől függetlenül vannak olyan helyzetek, amikor még a tesztek készítése sem lehetséges. Mindenesetre a következő hasonló helyzet előfordulásakor már lesz vizsgálható adat, addig pedig a tapasztalatokra kell támaszkodni.

Előállhat olyan helyzet is, amikor ugyan létezik történeti adathalmaz, de az félrevezető lehet. Tipikusan ilyen helyzetekkel találkozhatnak a brókerek nap mint nap – fel kell ismerni, hogy melyek azok a helyzetek, amikor a statisztika félrevezető lehet. Az ilyen szituációt nevezik „fekete hattyú”-nak. A megoldás persze az lehet, hogy olyan mintákat kell keresni az adatokban, ami egy ilyen szituáció jelenlétét jelzik.

Sok esetben a tapasztalat ütheti a statisztikát – vagy legalábbis feleslegessé teszi az adatok gyűjtését és az algoritmusok használatát. Jó példa a becslő szakma erre: egy jó becslő tapasztalatból tudja, hogy mennyit ér az adott tárgy, számára felesleges lenne adatgyűjtéssel foglalkozni, valamint drága szoftverbe és hardverbe fektetni.

Vannak továbbá olyan helyzetek is, amikor egyszerűen a változók nem mérhetőek – elsősorban személyek kvalitatív vizsgálatánál. Itt sok múlik a személyeken, amik nem mérhetőek. Az analitika alkalmazása persze itt is egy jó kezdőpontot adhat, esetleg segíthet szűkíteni a kört, de mindig szükséges lesz emberi beavatkozás.

Mindezen okok mellett nagyon sok vállalkozásnál tapasztalható, hogy az adatok gyűjtése és elemzése mindinkább a már meghozott döntések alátámasztására szolgál, mintsem a döntéseik megalapozására. Természetesen nem várható el, hogy minden vállalkozás azonnal áttérjen egy analitikai döntéshozatali ügymenetre, de az adatok gyűjtése és elemzése (akár utólag is) jó kezdetnek minősül (*BLOKDIJK 2008*).

Kétségtelen, hogy az egyre gyorsuló világ egyre inkább megköveteli a jól alátámasztott döntések meghozatalát – a világ ebbe az irányba halad. Ám ahogyan az analitika nélküli döntéshozatalba is csúszhatnak logikai vagy folyamatbeli hibák, itt is szembesülhet velük bárki (*THIERAUF 2001*).

Mivel minden kvantitatív elemzés következtetéseken alapszik, fontos figyelni arra, hogy miután a mintára felállított modell tökéletesen működik, működjön a teljes sokaságra is, azaz a minta jól tükrözze a valóságot. A legtöbb felállított modellben a pontosság csak az összegyűjtött adatok tartományában értelmezhető. A modell tartományán kívül eső faktorok megítélése komoly problémákat idézhet elő.

Ezek alapján a döntési hibákat két csoportba oszthatjuk: a logikában rejlő és a folyamatokban rejlő hibákra. A logikai hibák a következők lehetnek:

- Nem a megfelelő kérdések kerülnek felszínre.

- Rossz következtetések kerülnek levonásra és elmarad a következtetések tesztelése.
- A statisztika oly módon történő használata, hogy már egy meghozott döntés alátámasztására keresnek adatokat.
- Az adatokat félreértelmezik, vagy nem kerülnek az alternatívák kiértékelésre.

A folyamatokkal kapcsolatban a következő hibák kerülhetnek elő:

- Gondatlanságból elkövetett hibák:
  - Félreértelmezett adatok vagy
  - Hibás modell
- A döntések meghozatala előtt elmarad az elemzés és ezért nem valósul meg a megfelelő belátás.
- Nem kerülnek kellő komolysággal az alternatívák áttanulmányozásra.
- A döntéshozatali feltételek nem eléggé kiterjedtek.
- Túl későn kerülnek az adatok összegyűjtésre.
- Késői döntések meghozatala, az elégtelennek vélt adatok miatt.

A bevezetés technikájának elnevezése a görög delta (az analitika deltá-ja) betűből adódóan a változásokra utal – valójában nem három, hanem öt faktor játszik szerepet abban, hogy a vállalatnál sikerre vihető legyen az üzleti folyamatok analitikai alapokra helyezése. A delta betűszó az angol kezdőbetűkből áll össze (**D** – data, **E** – enterprise, **L** – leadership, **T** – targets, **A** – analysts). Ez az öt tényező:

- Hozzáférhető, jó minőségű adatok,
- Vállalati elkötelezettség,
- A vezetés analitikai alapokra helyezése,
- Lefektetett stratégiai célok,
- Elemzők megléte.

A sorrend persze nem a fontossági sorrendet reprezentálja. Egyformán fontos minden tényező. Minden elemzés megkezdése előtt szükséges azonban a megfelelő adatok megléte: pontosnak kell lennie az adatforrásnak és megfelelő formátumúnak. Ha több forrásból származik (mint vállalatok esetében a legtöbb esetben) akkor integrálnak és konzisztensnek is kell lennie (*DAVENPORT, HARRIS ÉS MORISON 2010*).

Az adatok hozzáférhetőségét illetően sokszor gondolható, hogy célszerű kicsiben kezdeni, eleinte csak egy üzleti funkciót támogatni, majd lépésről lépésre bevenni az összest. Általában ez az út nem járható, mivel:

- egyfelől minden olyan analitikai alkalmazás, mely ténylegesen javít a teljesítményen és a versenyképességen, adatait a vállalat több részéről szedi össze, érintve így azon területeket;
- ha az alkalmazások funkcionalitásában van átfedés vagy kapcsolódás, akkor nem érdemes lokálisan kezelni az erőforrásokat:
  - adatokat,
  - elemzőket és
  - technológiát;
- valamint kimutatható, hogy komoly vállalati perspektíva nélkül valószínűleg lesz sok kis kezdeményezés, mely analitikát érint, de egy igazán nagy sem.

A célkitűzés és a vezetés elhatározottsága elsősorban az analitikai projekt bevezetésével és megtérülésével kapcsolatosan fontos – ennek megtervezése a vezetés feladata. Az elemzők feladatai a következők:

- modelleket kell építeniük, hogy a célok elérhetőek legyenek,
- valamint bevezetik az analitikát a cég különböző területeire.

Azonban a sikerhez célszerű a DELTA minden fokán egyszerre előremozdulni a vállalatnak [2. táblázat: Az analitika alkalmazásának fokozatai]. Az analitika DELTA-ja<sup>31</sup> öt fokozatú.

Azt mondják, hogy a cég az analitikai DELTA 1. fokán áll, ha jelen állapotában a vállalatnál egy vagy több feltétel megléte hiányzik ahhoz, hogy komolyan vehető analitikai projekt kezdődhessen: vagy az adatok hiányoznak, vagy az elemzői képességek vagy a vezetői hozzáállás nem megfelelő, ezért analitikai megoldásokra alkalmatlan. A problémák forrása többek közt lehet az adatdefiníciók hiánya, vagy egyenesen az adatok hiánya, ám a problémák leggyakoribb forrása a rosszul, vagy egyáltalán nem integrált rendszerekből adódik. Ha ez az eset áll fenn, a továbblépéshez össze kell gyűjteni a döntéshozatal

---

<sup>31</sup> Az analitika DELTA-jával kapcsolatban a SASCOM magazin több számában jelentek meg értékes cikkek, számos szerző közreműködésével.

szempontjából releváns adatokat, létre kell hozni a master / referencia adatok jegyzékét, illetve létre kell hozni az adatkereskedelmi réteget.

Az analitikai DELTA második fokán álló vállalatoknál lokalizált analitika alkalmazható, mivel vannak ugyan analitikát alkalmazó megoldások a vállalaton belül, ám ezek nem központilag koordináltak, és emiatt nem is rendelhetőek alá a stratégiai céloknak. A probléma legtöbbször, hogy a megfelelő adat hiányzik a döntéshozatalhoz. A továbblépéshez szükséges vállalati szinten meghatározni az analitikai célokat és azok adatszükségletét, valamint különböző üzleti doménekhez tartozó adattárházakat kell létrehozni, illetve a vezetői rétegnek támogatni kell az adatok kapcsolását.

Egy vállalat akkor áll a harmadik fokozat szintjén, ha a törekvések megvannak, mivel van centralizált kezdeményezés az analitika alkalmazását illetően, ám nincs meg minden feltétel, és ennek köszönhetően a haladás lassú. A problémákat főként az okozza, hogy az adathalmaz nincs teljesen integrálva, és/vagy nem szabványosított, valamint nem minden tekintetben elérhető. A továbblépéshez elengedhetetlen a vállalati szintű adattárház létrehozása (külső adatok kombinálásával), vezetői szinten az EDW tervek támogatása, valamint a vállalat minden szintjén figyelni az új lehetséges adatforrásokat.

Az az analitikát alkalmazó vállalat vagy a DELTA negyedik szintjén, amelyik már rendelkezik a szükséges humán és technológiai erőforrásokkal, vannak működő és pozitív eredményeket produkáló analitikai alkalmazások, de a stratégiának nincsenek jól alárendelve. Ilyen esetben problémaforrás lehet, hogy néhány folyamat automatizálva van ugyan és az üzleti logika a rendszerbe van építve, ám ez a jelenség még nem terjed ki a vállalat egészére. Elvégzendő feladatként marad tehát az újabb egyedi adatforrások feltárása, az adatkezelési szabályzat létrehozása, vállalatmérettől függően BICC kialakítása, valamint az analitika további lehetőségeinek felkutatása.

Egy vállalat akkor érdemli ki az analitikai úton versengő vállalat címet, ha már rutinszerűen használja az analitika nyújtotta lehetőségeket, annak hatása az egész vállalaton belül érezhető, a vezetőség elkötelezett – ez látható a vállalaton kívül és belül egyaránt. Ekkor mondható, hogy a vállalat az analitikai DELTA ötödik fokán áll.

Az elemzéshez használható adattal szemben a következő elvárások (ALLEN ÉS TERRY 2005) vannak [3. táblázat: Adatok tulajdonságai]: strukturálnak, egyedülállónak és integráltnak



kell lennie, valamint jó minőségű kell legyen. Szempont ezeken kívül még az adat hozzáférhetősége, biztonságossága és kezelhetősége.

Strukturáltság tekintetében nehézséget okoz annak megállapítása, hogy milyen természetű adatról van szó. Az adatokat strukturáltság szempontjából három féle képpen lehet tárolni. Képezhetőek belőle kockák, tömbök vagy nemnumerikus leírások. A struktúrától függ nagyrészt hogy a későbbiekben milyen elemzés alapja lehet. A tranzakciós rendszerekben (és a forgalomban lévők túlnyomó többsége ilyen) az adatok táblákba rendezve találhatóak (ez a forma a működéshez ideális, azonban elemzéshez nem) – és mivel ez nem tartalmaz történelmi adatokat általában, ezért az adattárházban kockákat képeznek belőle. A képzett adatkockákon már számos új művelet értelmezhető, ám inkább feltárássra és nyomozásra valók, mintsem elemzésre – ennek legfőbb oka a limitált változó mennyiség. A tömbök felépítése előre meghatározott, rengeteg változót tartalmazhat, elemzési célokra ez éppen megfelelő ugyan, de aki nem jártas a technológiában, annak rettenetesen bonyolult lehet. A nemnumerikus adatok közé tartoznak a hang és videofelvételek, az internetes vagy belső adatbázisból származó szöveges és médiatartalmak – elmondható, hogy ezen kategória elemzése a legnehezebb, ám egyben a leginkább fejlődő irányzat is egyben.

Egyedülállóság tekintetében arról kell beszélni, hogy hogyan lehetne hasznosítani a szóban forgó adatot, főleg akkor, ha ez eddig egyáltalán nem történt meg. A kiinduló probléma az, hogy ha a szektor minden résztvevője ugyanazokkal az adatokkal rendelkezik, akkor mindenki ugyanolyan elemzési módszerrel, nagyrészt hasonló eredményeket kap. Ezért fontos, hogy legyenek egyedülálló adatok – ezek nyilván azok amelyek a vállalatnál kívülről nem publikusak. Az egyedülállóság azonban máshogy is biztosítható: olyan adatokat kell felkutatni, és összefüggéseket felrajzolni, melyekre más addig még nem gondolt.

Integrálás során adatokkal kapcsolatban a legnagyobb feladat annak meghatározása, hogy az hogyan konszolidálható több helyről. A fő cél itt az adatok összegyűjtése és integrálása. Ha a vállalatnál olyan ERP rendszer üzemel, mely a vállalat sok vagy összes részlegét felöleli, nagyban megkönnyíti az adatok szűrését és egyesítését – ám kevés helyzetben van így<sup>32</sup>. A felsorakozó adatok ráadásul sokféle képpen értelmezhetőek és magyarázhatóak – sokban függ az érdekeltségtől és az előadó személyétől – és számuk is igen magas lehet. Éppen ezért fontos a döntéshozatal szempontjából releváns adatokat összegyűjteni; ezt

---

<sup>32</sup> MDM : A Master Data Management: olyan gyűjtőfogalom, mely azokat a folyamatokat és technológiákat tartalmazza, melyekkel adatobjektumok kezelhetőek vállalat szerte. Egyik legfontosabb eszköze az adatintegráció.

nevezzük master vagy referencia adatoknak. Tökéletes adatintegráció persze még a legelőrehaladottabb vállalatoknál sem létezik, de fontos a törekvés és a folytonos munka – az alkalmankénti integrációs erőfeszítések kárba veszhetnek.

Adatminőség tekintetében annak megbízhatóságát kell megítélni. Fontos az adatok jó minősége – de nem elsődleges, itt (elemzési célokra) nem kell az a pontosság, mint üzleti jelentések készítésére vagy tranzakciós ügyviteli rendszerekhez. A hiányzó adat nem probléma ugyanígy, mivel statisztikai módon pótolható. A félrevezető adatok viszont problémát okoznak – mivel az adatok nagy része tranzakciós rendszerekből érkezik majd, fontos az esetleges félrevezetésre okot adó események kiszűrése. Itt segítséget nyújthat az ok-okozati összefüggések vizsgálata, segítségével a problémát a gyökerénél lehet megszüntetni.

Minden adatra meg kell határozni, hogy az hogyan válik hozzáférhetővé. Ahhoz hogy az adatok elemzésre hozzáférhetőek legyenek, célszerű azokat elválasztani a tranzakciós (forrás) rendszereiktől. Erre a célra tökéletesek az adattárházak (akár szervezeti egységenként is), vagy az egyetlen célt szolgáló adatpiacok; de a legjobb megoldás a centralizált EDW létrehozása.

Az adatok biztonsága ebben a tekintetben máshogy értelmezhető, itt azt kell kialakítani, hogy azok hogyan biztosíthatóak – nem összekeverendő a klasszikus adatbiztonság fogalmával. Itt, a folyamat során ki kell alakítani azon szabályok körét, melyek az adatok beszerzését, tárolását, felhasználását szabályozzák. A feladat igen összetett, mivel több ország törvényeit és jogszabályait is figyelembe kell venni, azzal az elsődleges direktívával, hogy kár ne keletkezessen (ebből a szempontból hasonlít biztonságkritikus rendszerek tervezéséhez).

Kezelhetőség tekintetében azt kell meghatározni, hogy az adatok hogyan szervezhetőek. Az hatékony adatkezelés-szervezéshez be kell vezetni néhány szerepkört a vállalatnál. Ezek a következők: BI architektúra tervező, adat adminisztrátor, adatbányász, adatminőség elemző, ETL fejlesztő, meta adat adminisztrátor (ÉLŐ, SZABÓ ÉS GALLI 2011).

### 3.4.2 Adatintegráció

Az adatintegráció nem más, mint az elemzés számára fontos adatbázisok tartalmának és egyéb adattartalmak egyesítése (*BODON 2010*). Ha az adatintegráció az egész vállalatra kiterjed, akkor a célterület adattárház (illetve lokális értelemben lehet adatpiac is).

Egy adatintegrációs project számos problémát vet fel: inkonzisztens és duplikált adatok létezése esetén meg kell határozni, hogy melyik adat lesz a megbízható, az elsődleges.

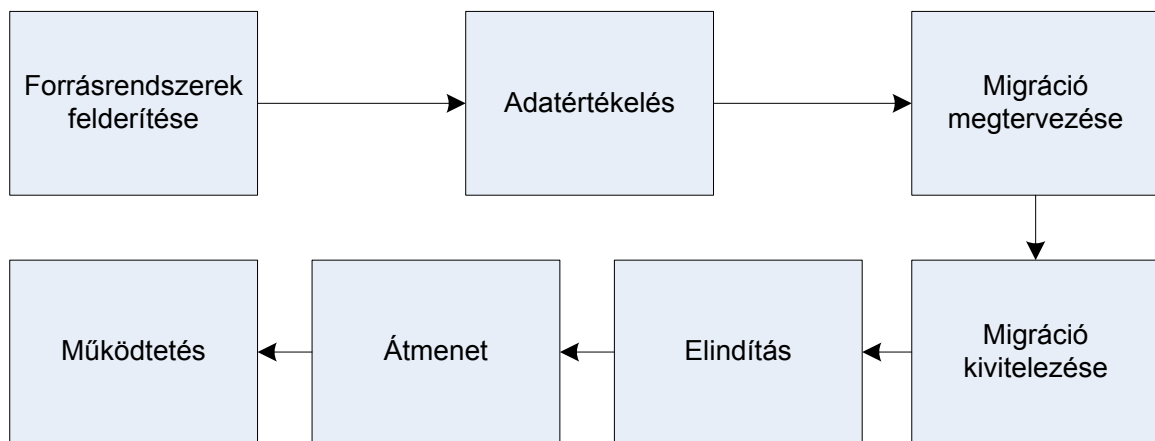
Természetesen a vállalati infrastruktúra heterogenitása (különböző beszerzési lehetőségek miatt, valamint az informatikai stratégia hiánya miatt kialakult) növeli az adatintegráció nehézségét (ugyanígy növeli a nehézségét, ha a céges hierarchia változik, összevonások miatt adatforrások összevonására is sor kerül).

Sajnos, ha projektek sokaságát nézzük, az adatmigrációs próbálkozásokról általánosan elmondható, hogy nem jól sikerülnek. A Bloor Research kutatása szerint a próbálkozások mintegy 60%-a sikertelen (*HOWARD 2007*). Sok adatintegrációs kezdeményezés azzal válik sikertelenné, hogy nem tudja megoldani az adatminőség problémáját, valamint csak nehezen képes követni a változásokat.

A legtöbb adatintegrációs folyamat része az adatmigráció. Sajnos a legtöbb szervezet esetében nem egyszeri, hanem folyamatos tevékenységről van szó.

A migráció ebben az esetben persze nem a népvándorlásra utal, mindinkább az adatok / adatbázisok / szoftverek mozgatására. Leggyakrabban a migráció egy régebbi adatbázisból az adatok kinyerését, majd (átalakítva) egy új adatbázisba való betöltését jelenti (*BARKAWAY 2009*). Mára számos BI termékcsomag rendelkezik adatintegrációt segítő szolgáltatással; előredefiniált folyamatrészekkel.

Létezik azonban egy járható, tapasztalati úton kialakított módszer, amit követve sikerre vihető a projekt.



forrás: saját munka

16. ábra: Az adatintegráció folyamata

Az adatmigrációs projekt első fázisa az adatforrások felkutatása [16. ábra] és azonosítása. Az azonosítás legelfogadottabb módja az, ha az adatokat csoportosítjuk a vásárlói, termék, folyamat, cím illetve egyéb hovatartozás alapján – mindezt persze befolyásolja a tervezett modell is (KHAN 2005).

Könnyen megeshet, hogy a forrásrendszer több ezer mezőt is tartalmazhat, de a gyakorlat azt mutatja, hogy ezek közt rengeteg lesz a duplikált, illetve lesz sok olyan is, ami nem lesz hasznosítható az új, célrendszerben. Ebben az állapotban egyelőre azt kell megjegyezni, hogy melyik adat az ami fontos, hol található, esetlegesen hol található meg redundánsan; fel kell jegyezni a szükségtelen adatokat is.

Ennek megfelelően azonban ha a kiinduló rendszer nem tartalmaz minden adatot, amit a célmodell igényelne, akkor ott hézag van. Ebben az esetben a hézag eltüntetéséért konszolidációra lehet szükség: ekkor több helyről kell adatokat összevonni, és olyan adatot képezni, ami megfelel a célmodellnek. Az adatok kategóriákba sorolása is itt végezhető el.

Több adatforrás használata persze több validációs pontot is jelent, ezért még inkább megbízható adatokat kapunk.

A következő logikai lépés a forrásadatok minőségének értékelése kell legyen, ugyanis ha az új rendszer hibásan működik adat-inkonzisztencia, redundancia vagy hézagok miatt, akkor értelmét veszti a migrációs próbálkozás (OLSON 2003). Erre szolgál az adatok jellemzésvizsgálata (data profiling).

A jellemzésvizsgálat nem más, mint minden adatbázis, minden táblájának, minden oszlopának tartalmának minőségi vizsgálata. Ezzel kiszűrhetőek a hibás/defektes adatok. Ez a vizsgálat hosszadalmas ugyan, de a folyamat fontos része, mivel biztosítható vele az adatok megegyezősége a célrendszer számára.

A vizsgálatba beletartozik a rekordok és azok metaadatainak<sup>33</sup> (*SMIRAGLIA 2005*) vizsgálata. Annak ellenére, hogy ezzel a lépéssel biztosítható, hogy a projekt nem fog több időt igénybe venni a tervezettnél, nagyon sok kezdeményezés adatvizsgálat nélkül indul.

A jellemzésvizsgálat segítségével meghatározható (*ZENG ÉS QIN 2008*), hogy:

- Az adatok megfelelnek-e az üzleti céloknak,
- Milyen legyen az integrációs stratégia, mivel a normálistól eltérő adat-előfordulásokat megismerjük
- Végrehajtható-e a későbbiekben automatizáltan az integrációs folyamat.

A fázis végeredményeként a team megismeri az adatok minőségét és mennyiségét a forrásrendszerben, a felmerült problémákat. Kialakulnak a korrigálási szabályok. Kialakítható egy nagyon absztrakt, felső szintű nézet az integrációs folyamat tervét illetően (*INMON, O'NEIL ÉS FRYMAN 2008*).

A migráció tervezése fázisban a technikai architektúra és a migrációs folyamat kialakítása a cél. Ehhez még szigorúan hozzá kell venni a tesztelési eljárások megtervezését, és azt hogy hogyan lesznek végrehajtva a célrendszeren (*HODGE 2004*). El kell dönteni, hogy lesz-e a régi és új rendszernek párhuzamos futása, vagy nulla-időkieséses migráció lesz, vagy egyszerűen ha az új rendszer üzemképes az átállás után, a régi rendszer dekommissiózására<sup>34</sup> kerül-e sor (*HAY 2006*).

Fontos, hogy a terveket írásban rögzítsük. A leírásba bele kell foglalni a követelmények mellett az ütemrendet, illetve a már eldöntött technikai részleteket.

A negyedik fázis a migráció részletes tervezése és kivitelezése. Vigyázni kell azonban, általában csak egyszer hajtható végre (sok esetben azért csak egyszer hajtható verge, mert a hatalmas adattömeg mozgatásának tervezése és mozgatása sok időt vesz igénybe), ezzel magyarázható a migrációs projektek sikertelensége / időtúllépése. Jó lépés ha a már

---

<sup>33</sup> A metaadat nem más, mint „adat az adatokról”, vagyis olyan információcsomag, mely definiálja egy szervezet / rendszer működéséhez szükséges adatokat, azok kezelésének módjával együtt.

<sup>34</sup> Dekommissiózás: terv szerinti használatból való kivonás, itt leállítás és a rendszerből való eltávolítás is.

kategorizált adatok közül csak az egyikkel kezdjük. Nagy projektek esetén fejleszthetőek a kategóriák párhuzamosan is. Célszerű először egyetlen kategórián/komponensen belül egy kisebb adathalmazt alapul venni, erre kidolgozni a transzformációs műveletsorozatot, kipróbálni és ezzel biztosítani, hogy működik, majd az adatmennyiség növelésével tovább tesztelni, és végül az összes így elkészült programot összevonni egyetlen alkalmazásba.

A fázis eredményeként előáll egy testreszabott, tesztelt és rugalmas migrációs alkalmazás.

Ezután következik az elindítás fázis. Mindenre kiterjedő tesztelés után elindítható a migrációs folyamat. Erre az időre szokás szerint leállítják a forrásrendszereket, hogy ne keletkezzen új adat. Az üzletmenet fennakadását elkerülendő ezt általában hétvégén vagy ünnepnapon végzik. Azokban az esetekben, amikor a rendszernek napi 24 órában, heti 7 napban, az év 365 napján kell rendelkezésre állnia<sup>35</sup>, nulla-időkieséses migrációra lesz szükség – ekkor a kezdeti adatok áttöltése után szükséges az áttöltés ideje alatt keletkezett adatok szinkronizálása.

Az átmenet a következő fázis, itt vándorol a valódi munka és terhelés az új rendszerre, és ha indokolt, a régi rendszer leállítására is itt kerül sor. Meg kell győződni, hogy az átállás sikeres volt, ez elsősorban a naplófájlokból deríthető ki. Ha sikeres volt, a terhelést át kell helyezni az új rendszerre (az adatok forrásait is ide kell irányítani).

Az utolsó fázis a működtetés. A tervezési folyamatban leírtak szerint kell eljárni, a régi rendszert leszerelni. Arra azonban figyelni kell, hogy a régi rendszer nem minden eleme lesz valószínűleg leállítva, így az adatok minőségének kérdése továbbra is figyelemmel követendő marad.

Ha a migrációs projekt sikerrel zárult, várhatóan lesz következő. Ehhez azonban már rendelkezésre állhat számos anyag a mostani kategorizálási / tervezési fázisokból.

### **3.4.3 Adatbányászat**

Az adattárolás árának általános csökkenése és a szélessávú hálózati hozzáférés általánossá válásával vállalatok egész sora előtt nyílt meg a nagymennyiségű adatfelhalmozás lehetősége. Persze az adathalmaz a legtöbbször heterogén származását, tartalmát és megjelenítését tekintve. Ennek oka, hogy a vállalat különböző pontjain keletkezett

---

<sup>35</sup> Rendelkezésre állás: Az informatikai elem vagy szolgáltatás egy adott időpontban vagy intervallumban a normál működésre kész állapotát jelenti. Ez a jellemző adott intervallumra vonatkoztatva a rendelkezésre állás tényleges és előírt értékének hányadosával jellemezhető.

adatokról van szó. Néhány gyakori példa: lehet kereskedelmi, pénzügyi és adminisztrációs tranzakciós bejegyzések, böngésző előzmények, e-mailek, egyéb szövegek illetve hipertextek, vagy teszteredmények. Ezen adatok hozzáférhetősége érdekes lehetőségeket nyit: ezen adathalmaz felhasználásával információk, majd abból tudás nyerhető, amit később a döntések során fel lehet használni, javítva a vállalat esélyeit a sikerre (VERCELLIS 2009).

A mai kor egyik legkorszerűbb elemzési módszerének az adatbányászat minősül. Ennek létrejöttéhez persze nagyban hozzájárult a gépi tanulás jelentős fejlődése (így azt ugyancsak nevezhető kapcsolódó tudománynak).

Az adatbányászat is olyan döntéstámogatást szolgáló folyamat, mely érvényes, hasznos, előzőleg még nem ismert információt tár fel (ABONYI 2006). Fontos jellemzője, hogy nagy mennyiségű adat szükséges hozzá – minnél több adat áll rendelkezésre, annál könnyebb az adatbányászat kivitelezése.

Az adatbányászat, mint folyamat lehet megerősítő vagy felfedező, ezen belül pedig predikciós vagy leíró (ABONYI 2006). Predikciós feladatoknál a cél egy még nem ismert változó jövőbeni értékének valószínűségi jóslása, leíró feladatoknál pedig már meglévő adatokból összefüggések feltárása.

Mivel az adatbányászat célja hasznos összefüggések, vagy új ismeretek feltárása, valamilyen nagy adathalmazban, sokszor azonosítják a gépi tanulással<sup>36</sup>. Az adatbányászat azonban sok esetben több, mint a gépi tanulás, mivel míg a gépi tanulás a legtöbb esetben mintafelismerést és trandfelismerést jelent, addig az adatbányászat sokkal szélesebb skálán mozgó technikákat használ (amik ugyancsak nem az adatbányászat sajátjai): pl. osztályozás, klaszterezés.

Az adatbányászat mára széles körben elismert, kutatott és nagyszámú irodalommal rendelkező tudomány; ennek megfelelően a gyakorlatba is számos új technika kerül átültetésre – az adatbányászatot segítő rendszerek verzióról verzióra új algoritmusokkal és lehetőségekkel bővülnek, ezért mindegyik ismertetése nehéz lenne, ám minden technika visszavezethető valamilyen (klasszikus adatbányászati) alapl műveletre. Ezek az (statisztikai) alapl műveletek: csoportosítás, osztályozás, gyakori elemhalmazok keresése, vagy asszociáció, regresszió és idősoelemzés.

---

<sup>36</sup> A külföldi szakirodalom nagyon szélesen, *machine learning* elnevezéssel tárgyalja.

A csoportosítás<sup>37</sup> célja az adathalmazban egymáshoz hasonló elemeket összegyűjteni. Ezek a csoportok persze lehetnek egymást kizáróak, vagy átfedőek, illetve képezhetőek hierarchiák, al-csoportokkal – vagyis, a csoportokon belül nagyon hasonló elemeket kell összegyűjteni, a csoportok között pedig cél a nagy eltérés felfedése. Az osztályozás célja ismert változók alapján az elemek kategóriákba sorolása. A művelet bemeneteit az elemek tulajdonságai adják, a kimenetet pedig a generált kategória értéke. Az osztályozás során, a modell generálása alatt kell meghatározni azt a függvényt, mely ezt a hozzárendelést megvalósítja (meglévő adatok alapján), majd az előrejelzés során új adatokra is meghatározható a kategória. A gyakori elemhalazok kinyerésénél (vagy asszociációs szabályok felállítása során) a cél az elemek közötti összefüggések feltárása. A regresszió célja ugyancsak a jövőre vonatkozó numerikus érték (kategorikus változó) megállapítása, a jelen adatai alapján – összefüggések feltárására van szükség, a kulcs a függvény megadása. Az idősor-elemzés időrendi adatok elemzését jelenti, összefüggések, minták feltárásával.

Léteznek klasszikus, jól kitaposott úton járó adatbányászati alkalmazások. A leginkább ismert és leginkább lendületet kapott alkalmazási terület a vásárlók megtartása és kezelése (CRM, később részletesebben), a második jó eséllyel a csalás felderítés, utána a keresztértékesítés, majd a minőség-ellenőrzés (ABONYI 2006).

Míg CRM alkalmazás esetén az adatbányászat műveletei közül a célszerű az adat szegmentáláshoz (valamilyen csoportosítási technika alkalmazását jelenti), majd a prediktív modellezéshez (regressziós és osztályozási technikákkal az eddig ismeretlen értékek jóslása) nyúlni, addig csalás felderítés során az extrém értékeket észlelése (csoportosítási és regressziós technikák kombinálásával) és a szegmentáció az ajánlott. Keresztértékesítés során egyértelműen kapcsolat analízisre (asszociációs szabályok felállítása és csoportosítási technikák alkalmazása) van szükség, esetleg prediktív modellezéssel összekötve. Minőség ellenőrzési feladatok esetén ugyancsak az extremitások észlelése és a kapcsolat-analízis, majd a prediktív modellezés a járható út. Természetesen nem minden adatbányászati probléma megoldása ilyen egyszerű, sok esetben több irányban célszerű elindulni és a végén összehadonlítani, melyik technika hozott jobb eredményt, aztán persze feljegyezni, visszacsatolni (FAJSZI, CSER ÉS FEHÉR 2010).

---

<sup>37</sup> A csoportosítási művelet különböző megnevezései: szegmentálás, szegmentáció, klaszterezés, klaszterképzés.



Szövegbányászatnak az a folyamatot nevezik, amellyel felfedezhetőek és kivonhatóak értelmezhető minták és összefüggések szöveggyűjteményekből. Egyszerűbben fogalmazva a szövegbányászat nem más, mint adatbányászat természetes nyelvfeldolgozással egybekötve (*FELDMAN ÉS SANGER 2007*).

**Szövegbányászat = adatbányászat + természetes nyelvfeldolgozás**

Mára már az üzleti információk több csatornáról gyűjthetőek össze, és ennek köszönhetően sokféle formában érkeznek. Ezt a sokszínű információáradatot kell valamilyen az üzleti intelligencia által használható formára hozni (szervezetten, címkézve – strukturált adathalmazzá), ugyanúgy, mint a friss adatokat, amiket először manuálisan kell feldolgozni, majd címkézni (ez a strukturálatlan adat) (*BERRY, BERRY ÉS KOGAN 2010*).

Az eltérő formában megjelenő adatok feldolgozásának problémája persze nem várható, hogy a közeljövőben megoldódik. Valójában persze minél több vállalat érez rá a folyamatos innováció és gyorsaság nyújtotta előnyökre, a növekedési stratégiák készítése iránti kereslet erősödni fog – így az örökké növekvő méretű adathalmaz egyre jobb bázist adhat a tény-alapú döntések meghozatalában (*CRISSEY 2007*).

A tradicionális adathalmaz és a szöveg természetű adatok összefésülése persze új lehetőséget ad a vállalatok számára, hogy továbblépjenek a jelenleg elterjedt cél-orientált keresési metodológiákon (*WEISS 2005*). Integrált szövegbányászati megoldásokkal a gyorsabban cselekvő vállalatok képesek lehetnek elektronikus úton felfedni a piaci hézagokat, és még azelőtt kihasználni a kínálkozó lehetőségeket, mielőtt a vetélytársaik felfigyelnének azokra.



forrás: Internet<sup>38</sup>, saját módosításokkal

17. ábra: 32GB adat tárolása 1980-ban és ma

Az adattárolási hatékonyság növekedésének, a háttértár- és memóriaárak csökkenésének (MANSOUR 2009) köszönhetően [17. ábra] sok nagyvállalat hatalmas mennyiségű adatot volt képes (repozitóriumként) felhalmozni. Ez egy jó lépés, ám még a feldolgozási sebesség hihetetlennek tűnő növekedésének és az elemzési technikák fejlődésének ellenére is ezek a nagy adathalmazok érintetlenek (elemzési és kihasználtság szempontjából) és feldolgozatlanok. Van köztük ami a vásárlók véleményét, interakcióit, óhajait tartalmazza – ezen információk kinyerése megoldható.

A vásárlói megelégedettségre utaló adatok szavak formájában van jelen a halmazban, akár többféle nyelven, akár többfajta mondat szerkezetben. A szövegbányászat elérhetővé válása előtt ilyen információk kinyeréséhez, értelmezéséhez és valamiféle trendek felfedezéséhez legalább egy csapatnyi képzett nyelvészre lett volna szükség – ezért az ilyen árnyalatnyi információk használható adatokká alakítása hatalmas feladat volt.

Mára a globális vállalatok (nagy termelő vállalatok) nagy kiterjedésű ügyfélkapcsolati menedzsment csapattal gyakorlatilag folyamatos szövegalapú információáradattal állnak szemben, e-mailekből, vásárlói véleményekből, garancia-ügyintézésből, mérnöki és eladói visszajelzésekből. Ez éves szinten több-százezer rekordot jelenthet. Nem sokkal ezig lehetetlen volt egy ilyen információhalmaz feldolgozása, mivel arra lett volna szükség, hogy ezeket a szabad szövegeket valaki strukturált mezőkbe beírva értékes (akkor már feldolgozható) információvá alakítsa, melyeket az adatbányászat és a statisztikai elemzés segítségével fel lehetett volna döntések meghozatalához használni. Mostanra, a szövegbányászat elérhetősége miatt mindez könnyebben feldolgozhatóvá válik. A nagy

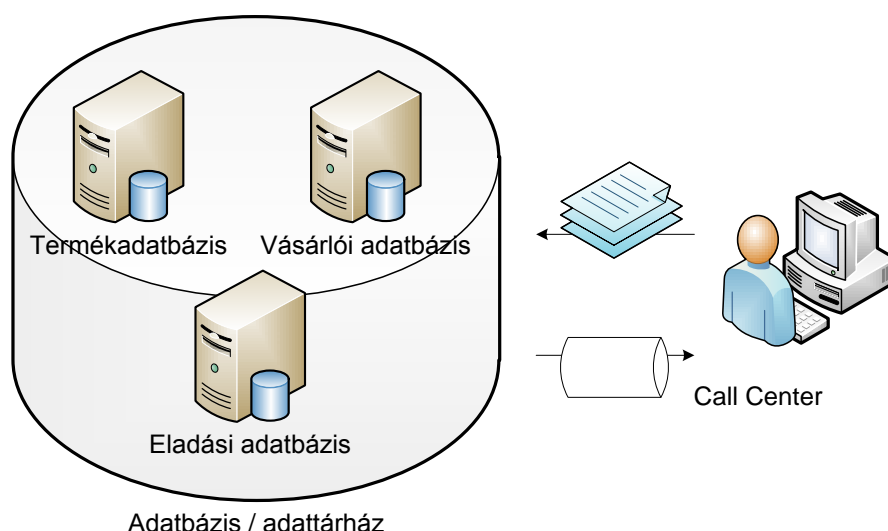
<sup>38</sup> A Shareables weboldal <http://thenextweb.com/shareables/2010/03/21/20gb-1980-32gb-2010-pic/> cikke alapján.

adathalmazt szövegbányászati eszközök segítségével elemzik, felfedve a megbúvó témákat és koncepciókat. Ezek után csoportosíthatóak (témánként való klaszterezés) illetve kategóriákba rendezhetőek. A szöveg-természetű adatok és a strukturált adatok integrálásával lehetőség nyílik a prediktív modellezés gazdagabbá tételére, valamint az üzleti döntések mélyreható ügyfél-információkra való alapozására – így megszüntethető a tapasztalt munkatársak megérzéseire való alapozás (*SHRIVASTAVA, SHRIVASTAVA ÉS SAHAMI 2009*).

A szövegbányászat elérkezése előtt a termékosztályozás unalmas, fárasztó, és rengeteg időt felemésztő feladat volt, ráadásul, mivel végig kellett menni az információkon egyesével, gyakran a statisztikusok sem értették a valódi lényegét, mivel az gyakran termék-specifikus volt.

A szövegbányászati megoldások elérhetősége segít abban, hogy mindez automatizálható legyen, így a kategorizálás is – nincs szükség a termékismeretre, a szükséges ráfordítandó idő a töredékére csökkenhet (*BERRY 2004*).

A szöveges információnak a meglévő adatbányászati módszerekhez adásával [18. ábra] a cégek belekeverhetik azokat a viselkedési indikátorokat a kutatásukba, amik a *miért?* típusú kérdésekre adhatnak választ – enélkül csak a múltbéli adatokhoz lehetne hasonlítani a jelenlegieket, esetlegesen a *mit?* típusú kérdéseket megválaszolva (*KISS ÉS RÁCZ 2003*).



forrás: saját munka

18. ábra: A call center kapcsolata az adattárházzal

A szövegbányászat emellett az akadémiai és a non-profit szektorban is bizonyít: használható kórházakban a kezelések és hatásaik összefüggéseinek vizsgálatára, költségcsökkentési helyek kereshetőek vele, stb. (Akadémiai körökben a szövegbányászat egyik elterjedt felhasználása a kapcsolódó cikkek keresése: segítségével egy nagy halmazban elhelyezkedő cikkekről / írásokról megállapítható, hogy milyen témában íródtak, illetve összetettebb elemzés esetén megállapítható az állásfoglalás is.)

Szövegbányászat használatával némely esetekben előre lehet vetíteni a várható kiadásokat: ha megnövekszenek adott termékkel kapcsolatban a panaszok, várhatóvá válnak a termék javításából fellépő költségek, így azokra időben fel lehet készülni, esetleg megelőző lépéseket tenni (*FRANKE, NAKHAEIZADEH ÉS RENZ 2003*).

A szövegbányászati eszközök persze bemenetként szolgálhatnak a látványos és interaktív ábrákat felvonultató elsősorban menedzserek számára, gyors döntések meghozatalára szánt dashboarding alkalmazásoknak.

A következő várható lépés a vállalatok részéről az lehet, hogy együtt szeretnék elemezni a strukturált illetve a nem strukturált adatokat.

Szövegbányászathoz hasonló, de attól némiképp eltérő, nagyon érdekes és fellendülőben lévő alkalmazási terület a multimédia-adatbázisok bányászata (érdemes megemlíteni, hogy míg a klasszikus / leírás<sup>39</sup> alapú megközelítés a multimédia anyagok metaadatainak bányászatára tör, addig új / tartalom<sup>40</sup> alapú megközelítések lehetővé teszik magának a multimédia anyag tartalmának bányászatát), valamint a web-bányászat (azon belül is megkülönböztethető weblinkstruktúra bányászat<sup>41</sup>, webtartalom bányászat<sup>42</sup> és webhasználat bányászat<sup>43</sup> – gyorsan fejlődő technológia, ám rengeteg a megoldandó probléma). Az új generációs okostelefonok elterjedésével új lökést kapott a

---

<sup>39</sup> A leírás alapú multimédia bányászat nagy alapos munkát, és jelentős emberi beavatkozást igényel, elvégre minden elemre meg kell adni annak címkéit. Hosszadalmas és lassú, automatizálva ellenben nem ad megbízható eredményeket.

<sup>40</sup> Tartalom alapú megközelítés során az adatbázisban nem maga az anyag kerül tárolásra, hanem annak generált leírói: általános leírók (fájl és webinformációk), jellegzetességeleírók (leggyakoribb szín vektor, leggyakoribb irányultság vektor) és elrendezésleírók (szírelrendezés leíró vektor és peremelrendezés leíró vektor), majd ezen adatokon végezhetőek elemzések.

<sup>41</sup> A webstruktúra bányászata az adott témához kapcsolódó autentikus és csomóponti weboldalak azonosításával foglalkozik.

<sup>42</sup> A webtartalom bányászat célja webes dokumentumok / weboldalak automatikus osztályozása.

<sup>43</sup> A webhasználat bányászata a webes hozzáférések metaadatainak elemzését jelöli. Itt nem a tartalom hangsúlyos, hanem a webszerver által erre leadott és tárolt kérések metaadatai.

helymeghatározási <sup>44</sup> technológia (a kérés hatására vezettek be pl. SQL Serverben térinformációkat tároló adattípust), mely nagy mennyiségű térinformáció kinyerését eredményezte, melyekből téradatbázisok építhetők (*FÜSTÖS – MESZÉNA – SIMONNÉ 1997*), majd azokon elemzések végezhetőek (*HAN – KAMBER 2004*).

### 3.5 Alkalmazási területek

Jelenleg a világ bármely ipari szegmensét vizsgáljuk, mindenhol változást látni. Kérdéses persze, hogy mi a sorrend, ahogy az információtechnológia és az analitika átformálja a világot. Feltehetőleg a bankrendszer volt az első, ami átalakult. A folyamatot Myron Scholes és Fischer Black indították el az 1970-es évek közepén a Wall Streeten – elgondolásuk azóta a Black-Scholes modell néven ismert. Módszerükkel a derivatívák értékének meghatározása egyszerűbb, mivel így rendelkezésre áll egy konceptuális keretrendszer, mellyel értékelhetővé válik a pénzügyi termékek áradata. Őket megelőzően csak 1956-ban merült fel, hogy számítógépekkel helyettesítsék a hitelbírákat. Mára persze olyan mértékben integrálódott az elemzés technikája a pénzügyi szektorban, hogy anélkül már működésképtelen lenne – az ügyfélmegtartástól kezdve a csalási kísérletek azonosításáig mindenre használható (*MAY 2009*).

A második szegmens, ahol már most is jelentősen érezteti hatását az üzleti analitika / intelligencia jelenléte, az egészségügy. Az egészségügyi adatok eddig csak nagyon szűk réteg számára voltak elérhetőek, ám az Internet megjelenésével megváltozott a helyzet: az adatok bárhol is hozzáférhetőek, valamint léteznek online analitikai megoldások is. Egészségügyi kutatás során szövegbányászat segítségével kereshetőek tudományos anyagok, de komoly sikerek érhetőek el mint a felismeréssel az orvosi képalkotásban (*MEYER-BASE 2004*) és gyakori halmazok kinyerésével vagy már statisztikai eljárásokkal gyógyszerek fejlesztése során (*LÜ ÉS FANG 2003*).

A kiskereskedelem valószínűleg ebből a szempontból nagy változások előtt áll. (Az előző nagy változás a II. Világháborút követően a professzionális logisztika alkalmazása volt.) Mára nem alkalmazható egységesen az a séma, miszerint a vásárlók majd mindent megvesznek, amit lehet; mára sokkal inkább ezt úgy kell értelmezni: a vásárlók mindent megvesznek, amit szeretnének, éppen ezért kell kideríteni, hogy mi is a vágyuk. Itt lép a képbe az üzleti intelligencia. A legnagyobb eredmények az élelmiszerboltokban várhatóak,

---

<sup>44</sup> Főként a világon mindenhol elérhető GPS – Global Positioning System használata, de alakul az európai GALILEO-Program is (2011 végén már kísérleti műholddal).

mivel itt minden nap sok ember fordul meg, ezért sok adat keletkezik, abból pedig könnyebb mintákat létrehozni és az egyéb hatásokat kikutatni.

A kiskereskedelemben az üzleti intelligencia használata az eladási minták felfedezésével, valamint az ügyfeleknek szóló személyre szabott ajánlatok készítésével tud segíteni. Eladási minták felfedezéséhez szükséges az elmúlt időszak összes vásárlói kosarának ismerete. Ebből kinyerhetők azok a termékcsoportok, melyek gyakran szerepelnek együtt értékesítésben. Ezt az információt többféle képpen lehet hasznosítani: ha az árukészlet csoportosítása mellett dönt a vezetés, akkor célszerű ezeket a termékeket egymás mellé helyezni, vagy célszerű még távozás előtt valamilyen módon ajánlani a termékeket a vásárlóknak. Természetesen más elgondolások is szóba jöhetnek, ezek már részben marketing kérdések. Személyre szabott értékesítéshez már nem elegendő a vásárlói kosarak ismerete, hanem ezeket személyekhez is kell tudni kötni. Innen aztán mintafelismeréssel megállapítható, hogy ki, milyen termékeket, milyen gyakorisággal vásárol, ez az információ pedig felhasználható személyes akciók vagy személyes ajánlatok készítéséhez. Ez a vásárlói analitika területe.

A vertikális piacok közül persze feltehetőleg a logisztika volt az első, amely teljes mértékben ki tudta aknázni az üzleti analitika lehetőségeit. Segítségével csökkenthetők voltak az átfutási idők, az üzemanyag és energiafelhasználás, növelhető volt a kihasználtság és a rendszerek teljesítőképessége. Nem titok, hogy a hadászatban is hasonló célok miatt vezették be az üzleti analitika alkalmazását.

A nagykereskedelemben és termelésben is jó eredmények érhetőek el a rendelési életciklusok és a készletek elemzésével – sok hasonlóság fedezhető fel a vásárlói analitikával. A kis- és nagykereskedelemben egyaránt előnyös a készletek elemzése, mivel változásaiban egyfelől minták fedezhetőek fel, másfelől külső események behatásai is megállapíthatóak, abból a célból, hogy azokra legközelebb felkészülhessen a vállalat.

A technológia jól használható államigazgatásban is, már több országban használják az adózási elemzésre (adófizetési és segélyezési csalások felderítésére), valamint bűnügyek felderítésére.

Az államigazgatáshoz sokban hasonlít a pénzügyi szektorban történő üzleti intelligencia felhasználás. A három fő cél a hitel- és pénzügyi kockázat elbírálás / elbírálás, a csalások felderítése valamint az eszközök kezelése.

Kiemelendő még a telekommunikációs iparági felhasználás. Nagy előny, hogy itt a vásárlók / ügyfelek eleve azonosítva vannak, valamint minden tevékenységüknek digitális formában nyoma van. A csalás felderítés mellett itt fontos szerepet kap a vásárlói viselkedés elemzés, valamint egyszerűvé válik a reklámok / akciók eredményességének vizsgálata.

Szektorról független felhasználási mód a fentieken kívül még a jövedelmezőség vizsgálat (nyereség és veszteség elemzés, valamint ok-okozati összefüggések keresése, aztán jövedelmezőség kialakítás), a teljesítmény elemzés és az értéklánc elemzés.

Érdekes megfigyelni, hogy hányféleképpen lehet az adatokat felhasználni, és hányféle jelentéssel válik értelmezhetővé az adathalmaz – felhasználástól függően. A felhasználást kétfelé lehet bontani: operatív és stratégiai felhasználás. Jelen pillanatban az operatív felhasználás van többségben, ennek célja az üzletmenet fenntartása, míg a stratégia felhasználás célja az üzletmenet jobbá tétele. Tisztán látható, hogy mindkét felhasználási cél nagyon fontos, hiszen operatív felhasználás nélkül a vállalkozás nem működhetne. Azt azonban, hogy mennyire értékes az információ, majd a stratégiai célok megvalósítására felhasználója tudja megmondani. Ezek alapján a stratégiai felhasználásnak több területe lehet, ezek között van olyan ami erősen iparág-függő, de nagy részük inkább univerzálisan használható, ezek: vásárlói, értékeítési csatorna, ellátási lánc analitika, viselkedés vizsgálat (*BURSTEIN ÉS HOLSAPPLE 2008*).

Vásárlói analitika: A fogalom erősen összefonódik a CRM-rendszerekkel, valamint azok magabiztos marketingkampányaival. Valójában persze minden sikeres CRM-kezdemenyezés / megoldás / rendszer mögött néhány olyan analitikai módszer áll, amely segít a vállalkozással megértetni a vásárlók szokásait, igényeiket és viselkedésüket. Ezen elemzések eredményeit, ha jól értelmezik és hasznosítják, akkor annak pozitív hatásai egyértelműek – a vásárlói elégedettség nőni fog (*LOSHIN 2003*). Ilyen egyszerűbb analitikai módszerek a vásárlók kategorizálása, a célzott marketing, a személyre szabás és az együttműködők szűrése.

- Vásárlók kategorizálása: A klasszikus marketing általában nagy tömegeket céloz, azt remélve, hogy a lehető legtöbb emberhez eljut az üzenet. Mára azonban a felfogás megváltozott, a célközönségről kiderült, hogy senki sem egyforma, és különösen nehéz jó szegmenseket kialakítani. Ezért, a vásárlók elemzésével

folyamatosan finomodhatnak azok a profilok, melyek demográfia, pszichológiai tényezők, viselkedés alapján jellemeznék embereket.

- Célzott marketing: sokszor nevezik a vásárlók elemzésének melléktermékének, ám lényege, hogy olyan nagyon szűk vásárlói csoportokat céloznak meg a profiljuk alapján olyan ajánlatokkal, melyek feltehetően felkeltik érdeklődésüket.
- Személyre-szabás: Annyit tesz: olyan bemutatkozás a vállalat részéről, mely a vásárlói profil alapján készült. Segíti a terjedését, hogy rengeteg esetben már az interneten való megjelenés a cég első kapcsolata a potenciális vásárlóval. Ha a kedvében jár azzal, hogy egy neki feltehetően tetsző, személyre szabott felületet biztosít, máris jobb helyzetben van. Elsősorban grafikai elemeket, másodsorban a kialakított profil alapján megjelenő tartalmat jelent.
- Együttműködő szűrés: A vásárlói csoportok preferenciáinak egyezésein alapuló ajánlattétel. A legtöbbet használt, és legjobban bevált online értékesítési adalék. Nagy előnye a kapcsolódó termékek felfedezésénél, illetve azok eladásánál érezhető.

Értékesítési csatorna elemzése: Szakértők körében a vélemények megoszlanak, hogy érdemes-e a kategóriát külön kezelni, sokak szerint az üzleti termelékenység elemzés témakörébe tartozik ez is.

- Marketing: Mind a marketing program finomhangolása, mind annak hatékonyságának mérése az értékesítési csatorna elemzésével érhető el. A vásárlók kategorizálásán alapuló stratégia kidolgozása tipikusan iteratív feladat. A hatékonyság az elvárások és a valós eladási adatok összehasonlításával számolható. A számított arányok persze a későbbiekben felhasználhatóak, mint adatforrások, a következő stratégia elkészítéséhez.
- Értékesítési út és teljesítmény-elemzés: Az értékesítést végző csapat adatainak eredményeikkel való összevetése. Segítségével azonosíthatóak azok a változók, melyek befolyásolják a teljesítményt, kideríthető, hogy min kell változtatni, az értékesítési életciklus csökkentéséért.

Az ellátási lánc elemzés célja a vállalat ellátási csatornáinak jellemzése és teljesítményének mérése. Beletartozik a szállítók és gyártók elemzése, a belső készletnyilvántartás elemzése, illetve a termékek vásárlóhoz jutásának menetének elemzése is. Aspektusai a következők:



- Beszállító és eladó menedzsment<sup>45</sup>: sok vállalkozás számára az i nehézséget jelent, hogy meghatározzák, hány szállítóval rendelkeznek – termékeket és szolgáltatásokat illetően. A technika használatával monitorozhatóvá válik a szállítók teljesítménye és megbízhatósága, mérhetővé válik a szolgáltatások és termékek minősége, így a szállítókkal a kapcsolat optimalizálható költségek, beszerzési képességek illetve kockázat alapján (*BASCHAB ÉS PIOT 2007*).
- Szállítás: Minden vállalat többféleképpen juttathatja el a termékeit a vásárlókhoz – minden változat külön költségmodellel rendelkezik. (Levegőben szállítani egészen biztosan drágább, mint hajóval, de gyorsabb is.) Használatával a vállalatok minimalizálhatják a szállítási költségeket, a megfelelő szállítási mód kiválasztásával, valamint biztosítható hogy a termékek időben a megfelelő helyen legyenek (persze ez a szállítókon is múlik). Sokban összefügg az erőforrás tervezéssel / nyilvántartással (*REYNOLDS 2001*).
- Készletnyilvántartás: Főként olyan termékek értékesítésénél van nagy szerepe, amik gyorsan vesztenek árukból, és ennek köszönhetően piaci kockázatot jelentenek. Egyik végtet sem jó: nem elfogadható a kizárólag megrendelésre történő szállítás, valamint a felhalmozás sem. Azonban ha rendelkezésre állnak az értékesítési, termelési és az ellátási láncból származó információk, kiszámíthatóvá válnak a készletezéssel szemben támasztott követelmények. Lehetséges továbbá a felmerülő kockázatok számszerűsítése és minősítése is (*ANANDARAJAN, ANANDARAJAN ÉS SRINIVASAN 2003*).
- Elosztás elemzés: Elsősorban a kiskereskedelemben használható annak megállapítására, hogy melyik kiskereskedelmi egységben mekkora kereslet várható, illetve oda milyen útvonalon, honnét érdemes a javakat szállítani. A szállítási és raktározási költségek minimalizálhatóak.

A viselkedésvizsgálat során a többi alkalmazási területtel ellentétben nincs nagymennyiségű számszerűsített adat, nincsenek adatkockák, illetve nem értelmezhetőek az azokon végezhető műveletek sem. A módszer lényege a jelentős eseményeket megelőző viselkedések vizsgálata, majd ebből következtetések levonása, annak érdekében, hogy a későbbiekben előre jelezhető legyen adott esemény adott viselkedési minta alapján (*LOSHIN 2003*).

---

<sup>45</sup> A szaknyelvben már megragadt a *vendor* angol szó magyar környezetben történő használata.

Ügyfelek kezelése: A CRM mint fogalom valamikor az 1990-es évek közepén alakult ki, és az évtized vége felé kezdett erősen hirdetni lenni (a terület azonban csak a 2000-es évek után kapott nagyobb figyelmet (*SARNEVITZ 2009*), és a teljes potenciált is csak az utóbbi években kezdi elérni).

A CRM (Customer Relations Management – Ügyfélkapcsolati Menedzsment) egy olyan üzleti megközelítés, mely integrálja az embereket, a folyamatokat és az alkalmazható technológiát, így maximalizálva a vásárlókkal a kapcsolatot. Az utóbbi néhány évben egyre inkább az internetre támaszkodva végzi mindezt, mivel így jól koordinálhatóvá válnak a vásárló-központú funkciók (*DYCHÉ 2002*).

Ahogy az internet és a mobil technológiák használat egyre inkább a valós időbe vezetik a CRM-et, úgy képes egy vállalat a nap 24 órájában éberrel működő idegrendszerré (*GOLDENBERG 2002*) változni, képessé válik célzott egyének értesítésére, amennyiben változás áll be a vásárlói szükségletekben, a versenyképességi elemzésekben, a készletekben, a nyersanyagok rendelkezésre állásában, vagy a jövedelmezőségben (*PEEL 2002*). Ezekhez a lehetőségekhez még ha hozzáadjuk az internet és a szociális hálókat nyújtotta lehetőségeket, és rögtön megkapjuk célunkat: a tökéletes környezetet a valós idejű vállalatához (Real-Time Enterprise - RTE).

Arra, hogy miért is érdemes egy vállalat számára automatizálni a sales, marketing és CRM rendszereiket, három fő indok létezik:

- A legfőbb ok közvetlen értékesítés (direct sales) magas költsége, mely – ha hinni lehet a trendnek – tovább nő. Egy jó CRM megoldással az értékesítés termelékenységét növelni lehet, kordában tartva, vagy kedvező esetben csökkentve az értékesítés költségeit.
- A második ok a globális verseny kiéleződése. Mára eljutott a világ (sebessége és összekapcsoltsága) oda, hogy a nagyobb üzleti döntések mind a globális piacra hatnak. Ahhoz hát, hogy egy helyi vagy külföldi piacon versenyképes lehessen egy vállalat, a rendkívül hatékony piaci intelligencia már elengedhetetlen. Egy jó CRM megoldás segít a vállalatoknak abban, hogy hatékonyabban figyeljék és kövessék a piac lehetőségeit.
- A harmadik azonosítható ok egyértelműen az információigény. A sales, a marketing és a vásárlói támogatás tipikusan információ-érzékeny részek. A siker két kulcstényezője a következő: egyfelől az, hogy sikerül-e egy hatékony marketing

mix stratégiát megvalósítani, illetve az, hogy jól értelmezzük-e (címezzük-e) a versenytársaink marketing stratégiáját. Egy jó CRM megoldás segítségével összegyűjthető, szűrhető és kiválasztható a releváns információ a piacról, különösen figyelve a vásárlókra.

A CRM alkalmazásából származó kulcshaszon tényezők: Már most köztudott, hogy a CRM megoldások alkalmazásának haszna az, hogy a vállalatok képesek segítségével elsősorban a sales, marketing illetve az ügyfélkapcsolati rendszereiket automatizálni. Így a legjellemzőbb haszontényezők a következők lehetnek:

Jobb minőségű információink lehetnek a sales és marketing részlegek számára: A vásárlók adatai, háttérük, szükségleteik és versenyképességi helyzetük tipikusan olyan információk, melyek egy CRM megvalósításból adódnak.

A javuló termelékenység a piac szegmens sikeres azonosításának és megcélzásának, a rossz adatok eliminálásának, a készletek azonnali hozzáférhetőségének illetve a rendelések akárholnan feladhatóságának eredménye, mivel mind-mind hozzájárul az értékesítési ciklus rövidüléséhez (*LOMBARDO 2003*).

A vásárlói gondoskodás alaposabbá és teljesebbé válása annak köszönhető, hogy az értékesítésre fordított adminisztráció mennyisége lényegesen csökken, így marad erőforrás a vásárlói szükségletek gyors lereagálására, valamint a meglévő illetve potenciális problémák felderítésére illetve megoldására.

A legtöbb szervezet számára az egyik legfontosabb céllá vált a vásárlói megítélés (és az abból következő lojalitás) javítása. Ez a terület is tipikusan jól támogatható CRM megoldásokkal. Vizsgálatok kimutatták, hogy a vásárlói hűség növekedésével együtt nő a vásárlón keletkező haszon (*BLIGH ÉS TURK 2004*).

Vizsgálatok kimutatták továbbá, hogy minimum 10 százalékos bevétel növekedés várható a CRM rendszer bevezetése után, mivel segítségével a sales-es képviselők képesek lesznek jobban megcélozni az ügyfeleket, növelve a hatékonyságot, időkihasználás és sikeresség tekintetében egyaránt. Másik várható eredmény az első három évben egy 5%-os adminisztrációs költségcsökkenés, mivel a használt megoldások segítségével már nem lesz többé szükséges a céltatlan, nagy mennyiségű reklámanyag előállítás és kiküldése. További 5%-al nőhet a sikeres üzletkötések száma, mivel az alkalmazott megoldás segítségével az eleve kudarcra ítélt üzleti megkeresések kiszűrhetőek, több időt hagyva a

jól célzott üzletek kialakítására. Mivel a kereskedők a rendszer segítségével jól elkülöníthető csoportokra bontott vásárlókkal dolgozhatnak (és célozhatják őket), ezért alakítható úgy a csoport, hogy a vásárlók nagyobb hangsúlyt fektessenek a vásárlásra, jobban reagáljanak a kedvezményekre, elegendővé válhat a kedvezmények csökkentése és ritkítása. További 5%-os javulás várható el realisan a vásárlói megítélésben, ami annak köszönhető, hogy a vásárlók gyorsabban hozzájuthatnak a kívánt, számukra releváns információkhoz, ezért elégedettebbé is válnak.

Minden CRM megoldás tervezésénél persze felmerül a kérdés (elsősorban a vezetés részéről), hogy hogyan lesz majd mérhető a CRM rendszer bevezetésének, vagyis a beruházásnak (idő, pénz, erőforrás, személyzet) a haszna? Mára sok metrika áll rendelkezésre úgy a direktben mérhető, mint a nehezen számszerűsíthető oldalon egyaránt (*FINNEGAN ÉS WILLCOCKS 2007*).

A könnyedén számszerűsíthető eredmények közé tartozik például az az időmennyiség amit a már meglévő vásárlóval kell tölteni (ez lehet telefonhívás / nap vagy eltöltött óra / nap, a lényeg, hogy a már meglévő vásárlóval történő interakció mennyiségét kell mérni), vagy a potenciális új vásárlók száma (jó mérés a meglévő ügyfelekre jutó potenciális ügyfelek számának mérése, naponta, hetente, havonta vagy negyedévente), vagy a menedzserek által szerződéskötéssel töltött idő, esetleg a vezetők és beosztottak által közösen a vásárlói problémák megismerésére és megoldására szánt idő. Itt jön képbe a coaching, mint módszer hasznossága. Ugyancsak jól mérhető a vásárlók kiszolgálásának hatékonysága (minden probléma ciklusidejének mérése, illetve a téves információkból adódó hibák számának mérése), valamint az ügyfélkövetés pontossága (az ügyfél megkeresése és az azt követő információcsere közt eltelt idő). A két legjobban mérhető valószínűleg a sikeres üzletek kötésének aránya, illetve a bevétel növekedés. A számszerűsíthetően mérhető eredmények közé tartozik még az üzleti eredmény. Mérhető az is, hogy egy CRM rendszer segítségével mennyivel gyakrabban lesz a vállalat neve az emberek (meglévő és potenciális ügyfelek) előtt (*TSIPTSIS ÉS CHORIANOPOULOS 2010*). A vásárlói elégedettség szintén jól mérhető, ehhez kérdőívekre van szükség. Javulhat a cégen belüli kommunikáció is, mivel ahogy egyre többen foglalkoznak a cég alkalmazottai közül az ügyfelekkel és lehetőségekkel, úgy nő az igény a belső szoros kommunikációra is – jó mérési alap lehet a cég kirendeltségei közötti kommunikáció idejének vizsgálata. Javulni fog a sikeres üzletek aránya, valamint csökkenni fog várhatóan a sikeres üzletkötésre fordított idő.

A nehezebben számszerűsíthető hasznok az úgynevezett lágy kritériumok, melyek között megemlítendő az a tény, hogy a CRM rendszer egyik hatása az lehet, hogy általánosan simábbá válik a cégen belüli ügymenet, mivel csökken a felesleges adminisztrációval (betanulással) való időtöltés, így a felszabaduló idő nem információkeresésre, hanem inkább az információk felhasználására és munkavégzésre használható fel. Nőhet az alkalmazottak motiváltsága és megelégedése (bár ez nagyon nehezen mérhető), ehhez vagy a CRM rendszer felhasználóitól kapott visszajelzéseket használhatjuk, vagy mérhetjük az alkalmazottak cserélődésének mértékét (*LEVITAN 2008*). A rendszer használatával jobban képzett szakembereink lehetnek, sales, marketing és ügyfélszolgálati részeken. Egy jól kialakított rendszer nagyon jó képzési lehetőségeket nyújt a vállalat termékeiről és szolgáltatásairól. Mérhetjük a szükséges, releváns adatok előteremtéséhez szükséges időt – csökkenni fog. Egy jó megoldás használatával nőni fog a mobil hozzáférések száma. Ez fontos, mert bár egyénenként eltérő a mobil technológiák és eszközök elsajátításának ideje, lesznek olyanok, akik munkájához elengedhetetlen a mobil hozzáférés, és az ő elégedettségük is ugyanannyira fontos kell hogy legyen a cég számára. A rendszer használatából (és az előzőekből) adódik, hogy a rendelkezésre álló információink sokkal naprakészebbek lehetnek, valamint elérhetőségük szintén javul. Azt, hogy mennyire segíti ez az ügyfeleket, már szubjektív, ám mérhető, például úgy, hogy mennyi idő telt el az információigény és az információ megtalálása közt. Hasonlóképpen mérhető és javítható a reakció a vásárlói igényekre (*REYNOLDS 2002*). Összességében az ügyfelekkel való kapcsolattartás javulása mindenképp a cégről kialakult kép javulását vonja maga után: az ügyfelekkel való kapcsolattartás kezelése (és kézben tartása) játszhatja a legfontosabb szerepet a jó megítélés/kép kialakításában. Ennek mérése nehézkes, ám ha a meglévő és leendő vásárlók véleményét kikérjük a cég marketing tevékenységéről, az megadhatja a kellő adatokat, ahonnan már viszonyítani tudjuk a változásokat. Ha javul a cégről kialakult kép, az egyúttal azt is jelenti, hogy a vásárlók kis idő után jól el tudják különíteni a céget a versenyszektortól, amiben dolgozik. A jó mérési alapot e tekintetben a vásárlói hűség mérése, valamint annak mérése, hogy mennyire bírják a céget megkülönböztetni a versenytársaktól.

Vizsgálatok kimutatták, hogy a CRM rendszereknek köszönhető automatizálás (bizonyos területeken persze, mint a folyamatkövetés, időbeosztás, valamint profilkészítés) átlagosan egy órát takarít meg minden ember napi munkaidejéből. Ez úgy is felfogható, mintha egy évben 26 nappal többet dolgoznának. Ez a kiemelkedő időmegtakarítás elsősorban a sales

részlegesen jelentkezhethet, ám hasonló (napi 20-30 perces) időmegtakarítás érhető el más részlegeken is, mint az általános menedzsment vagy a vásárlói szervizcenter. Az időmegtakarítás persze nem a kizárólagos eredmény. Az ajánlatok készítése, mely eddig fáradságos munkaként nehezedett a marketing és sales osztályokra, egy rendszer implementálásával jóval inkább felhasználóbarát lesz, mivel az összes szükséges információt (szövegeзések, embercsoportok, a címzés adatai) mind-mind a rendszerben tárolva vannak. A leglátványosabb várható haszon, viszont a 14%-os növekedés a munka termelékenységében (elsősorban az értékesítési részlegen).

A kihívások: A világ legnagyobb szoftverbevételeit a CRM szoftverlicencek jelentik. 2006-ban ez a bevételi érték 31 milliárd dollárt tett ki. Míg a CRM "iparág" 2004-ben 5%-os növekedést hozott, addig az általa támogatott iparágaknál 10% volt mérhető. Egyes becslések szerint a gazdasági hullámvölgy elmúlását követően ez a tendencia még évekig megmaradhat. A növekedésnek köszönhetően a CRM iparág növekedni kezdett, már kínálnak megoldásokat szolgáltatásként (*SaaS*), illetve a mobil CRM megoldások is terjedőben vannak.

Egy CRM kezdeményezés sikeres kivitelezésének leggyakrabban a következő tények állják útját: az elkötelezettség hiánya, a sales- marketing- és ügyfélszolgálati stratégia hiánya, a belső politika, a megfelelő képzés hiánya, a tudásbázis meglétének a hiánya, a felhasználók idegenkedése (*GENTLE 2002*).

A stratégia hiányában a cég nem lesz képes priorizálni azokat a területeket, melyek a későbbiekben nyertesként kell, hogy kijöjjenek a rendszer bevezetéséből. A cég elkötelezettségének hiánya a leggyakoribb ok a kudarcra. Egészen a legfelsőbb szinteken is komoly támogatással kell rendelkeznie a projektnek. Ahhoz, hogy a projekt sikeres lehessen, mindenkinek meg kell osztania az információit a cégen belül. A megfelelő képzés megléte szintén létfontosságú a projekt alatt. Ez egészen olyan extremitásokhoz is vezethet, hogy az informatikai hardverkötség 3x-15x költségét jelentheti a képzés ára; ezért kell a felsővezetők támogatása is. A tudásbázis hiányát vagy annak belső megteremtésével vagy importálásával lehet megtenni. A felhasználók idegenkedése pedig már az informatikai rendszerek kezdetek óta probléma, ám megfelelő információáramlással és információ-megosztással a probléma kiküszöbölhető, mivel ha mindenkivel ismertetik azokat az előnyöket, melyekhez a vállalat és az ő személye is jut, csökken az ellenállás. Ennek a (tájékoztatási) kezdeményezésnek fentről kell indulnia a vállalati hierarchiában, és

mindenkire ki kell terjednie, aki (bármilyen módon) érintett az informatikai rendszerek használatában.

Az időbeosztás kezelése: A CRM szoftverek világában naptárkezelés szempontjából két etalon említhető: az egyik a Microsoft Outlook, a másik a Lotus Notes. Mindkettő egyben e-mail kliens is, és meglehetősen jó naptár funkcióval rendelkezik – bizonyítja ezt az a tény is, hogy számos rivális szoftvertermék funkcionalitása és kinézete hasonlóná vált ezekéhez. Az etalonhoz igazodás odáig fokozódott, hogy mára minden jelentősebb CRM megoldástól elvárható a kétirányú Outlook integráció. Ez annyit tesz, hogy az Outlook naptárába beírt időpontok megjelennek a CRM rendszerben, illetve a CRM rendszerben keletkezett időpontok bekerülnek az Outlook naptárba. Ehhez még természetesnek mondhatóan hozzá adódik a csoportok kezelése, a csoportos foglalkozások kezelése, illetve a tennivalók listáinak kezelése.

Értékesítés management: Ide tartozik elsősorban a kapcsolattartási adatok kezelése (illetve a megtörtént események kezelése), valamint az ajánlatkészítéshez használatos funkciók (dokumentum generátorok, illetve szolgáltatás/termékválasztó modulok). Itt használható és integrálható funkció a pipeline analízis, az előrejelzés illetve a trend/ciklus keresés, az értékesítési metrikák, valamint az adatkockákból történő jelentések készítése.

Telemarketing / Telesales: Ha egy CRM rendszer része egy telemarketing támogató modul, akkor ott elvárhatóan megvalósítandó funkció a hívási listák összeállítása és kezelése, a hívás lebonyolítása (akár *Voice-over-IP* megvalósításban), a híváskövetés.

Vásárlói kapcsolattartás: A vásárlóközönséget kiszolgáló funkciók, mint az esemény hozzárendelés / feljebbvitel / követés / jelentéskészítés (*level-of-support*) és a probléma menedzsment/megoldás, a garanciális ügyek, a szerződések kezelése tartozik ide. Fejlettebb, a web nyújtotta lehetőségeket jobban kiaknázó megoldások esetén már a vásárlók/ügyfelek maguk is hozzá tudnak férni az ügyfélközpont bizonyos részeihez, gyorsabbá, egyszerűbbé és olcsóbbá téve így az egyszerűbb ügyek intézését.

E-business: Minden CRM rendszernek, mely tartalmaz lehetőséget e-kereskedelem megvalósításra, célja hogy az áruk / szolgáltatások weben keresztül történő cseréjét gyorsabbá tegye. Utóbbi években jelentősen fellendültek a B2B és B2C megvalósítások. Mai napig ennek ellenére kevés az olyan CRM alkalmazás, melyben minden e-business funkció megtalálható. Ami viszont szokássá válik, az a megvalósítás, amikor a CRM

rendszerben megtalálható egy külső web alapú e-commerce szoftver front-end interfésze, melynek segítségével jól konfigurálható, kezelhető és ellenőrizhetőek a web-áruház, illetve a vásárlások, statisztikák (cross-selling, up-selling).

Ellátási lánc menedzsment: Az e-business támogatottsághoz hasonlóan ritka sajnos az ellátási lánc menedzsmentjének belső támogatása, ám a legtöbb CRM-megoldás jól integrálódik az ellátási lánc menedzsmentet támogató szoftverekkel – így a CRM-et használó cégek mindig a legújabb megoldásokat használhatják, beleértve ebbe a B2B lehetőségek kiaknázását is.

Sokrétű hozzáférés támogatása: Talán az ügyfelek részéről a legszembetűnőbb változás az lehet, hogy így már többféleképpen is képesek lesznek a céggel kapcsolatba kerülni: levelezés (hagyományos és remélhetőleg egyre nagyobb részben e-mail), telefon (és üzenethagyás), fax és webes lehetőségek (fórumok, szociális hálók és weboldalak). Emiatt fontos követelmény egy CRM-rendszerrel szemben, hogy az ügyfelek akármilyen módon is érkezzenek, a cégnek ugyanazzal az arculatával/megjelenésével találkozzanak.

Csoportmunka / Együttműködés: A kereskedelmi részek, a vásárlói kapcsolattartók és a vásárlók egyaránt szeretnék mindig naprakész (percre-kész) információkat kapni a rendelésekkel kapcsolatban, de a különböző osztályok menedzserei is szeretnék elemzéseket és előrejelzéseket kapni. Ezen igények magas szintű adatmegosztást és együttműködést követelnek meg a CRM rendszertől, ezért a CRM megoldások szolgáltatói vagy eleve integrálnak ERP megoldásokat a termékeikbe, vagy jó lehetőséget biztosítanak a más ERP-rendszerekkel való integrációra. Gyakran a két termék integrációja csak egy harmadik szoftver termék bevonásával lehetséges (third party tools). Az adatok megfelelő és időben pontos szinkronizációja / az adatok megosztása így elsődleges fontossággal bír (ez a szinkronizáció lehet egyfelől mobil eszközök között, egyre terjedőbben vezeték nélküli protokollokon keresztül, vagy másfelől lehet a cég adatbázis/alkalmazásszerverei között). Mivel még az adatok szinkronizációjára nem létezik szabvány, ezért a gyártók vagy saját szinkronizációs eljárásukat használják, vagy harmadik cégre bízzák a funkció ellátását végző szoftver kivitelezését. Ebből következően a szinkronizáció terhelhetősége és skálázhatósága gyakran kérdéses marad a vásárló számára – szerencsére már vannak szinkronizációt tesztelő létesítmények (a HP Compaq és a Microsoft is rendelkezik ilyen megoldással), akik tanúsítványok kiállításával igazolják a gyártó megoldásának terhelhetőségét.



Web bázisú CRM: A CRM piac mai dinamizmusában a kliens-szerver megoldások egyre inkább teret vesztenek a web-bázisú e-vásárló kezdeményezésekkel szemben.

A kliens-szerver megoldások nagyban támogatják a vásárló központú és belső sales vagy marketing vagy vásárlói ügyfélkapcsolati célú CRM rendszereket, míg az újabb web-bázisú e-vásárló rendszerben a vásárlók közvetlenül hozzáférhetnek a számukra lényeges, vállalat specifikus információkhoz (*PEEL 2002*).

A legvalószínűbb jövőkép az elkövetkezendő évekre a két megoldás párhuzamos fejlődése, illetve kombinálódása, mivel a kliens-szerver megoldások már komoly vásárló (céges) táborral rendelkeznek, akik már profitáltak a technológia létezéséből, ám ezen megoldások szolgáltatói lépésről lépésre folyamatosan adoptálják a web alapú e-vásárló megközelítést. Az előző változatok továbbfejlesztése miatt a szoftverek nagy része nem lesz Web-alapú, de mégis mindenképpen web-kompatibilisek lesznek, hogy kiszolgálhassák a nagyszámú felhasználót. Ennek ellenére vannak olyan szolgáltatók akik teljesen újraírták megoldásukat a tökéletes webes használhatóság érdekében. A többi szolgáltató ügyfelei pedig lépésenként (verzióként) haladhatnak a webes e-vásárlói megközelítés felé. Ennek az átállásnak a lassúsága tartja életben a kissé idejétmúlt kliens-szerver látásmódot.

A web-alapú e-vásárló megközelítésű CRM szolgáltatók így már saját vásárlói körüket építhetik. Az általuk kínált megoldások rendkívül impresszívek: egyszerre kínálják a vásárlói önkiszolgálás és az analízisek lehetőségét, a tartalomkezelés funkcionalitásával együtt. (Ilyen megoldásokat kínálnak a következő szolgáltatók: Salesforce.com, NetSuite, BroadVision, KANA, Infor CRM Epiphany, Vignette.)

CRM-bevezetés: Hála a technológiának és a bevált módszerek jó alkalmazásának mára a CRM kezdeményezések sikerszázaléka valamivel 70% felett van. Ezt az adatot látszik megerősíteni az InfoTrends és a CRMGuru.com elemzőcégek által készített tanulmány is – így nyilvánvalóvá vált, hogy egy jól implementált CRM megoldás a cég számára sokrétű és mérhető eredményekkel kecsegtet.

Egy CRM üzleti terv – beleértve az értékteremtést is – a legjobb módja annak, hogy sikerre vigye bármely cég a CRM projektjét. Egy ilyen tervtől a következő dolgok várhatóak el: Először is meg kell határozni azokat a metrikákat, melyek az automatizálandó funkciókat mérik. Másodsorban meg kell határozni ezen metrikák célértékeit (számítva, alátámasztva, hogy reális elvárások szülessenek). Harmadsorban pedig rendszeres (havi, negyedéves

vagy fél éves) jelentéseket kell készíteni minden ilyen célértékről a menedzsment felé (DYCHÉ 2002). Ez a háromlépéses terv egyszerűnek hat ugyan, ám mégis meglepő, hogy milyen kevés CRM bevezetés részét képezik a mérések vagy akár a beruházás megtérülésének (ROI) mérése.

***CRM: Megfelelő emberek, megfelelő folyamatokkal és megfelelő technológiával támogatva.***

Minden CRM-megvalósítás az alábbi három összetevő zökkenőmentes integrálásán múlik: az emberek, a folyamatok illetve a technológia. Azért ez a három tényező, mivel ezzel a hárommal érintkeznek az ügyfelek a leggyakrabban. Ezen tényezőket kell úgy összehozni, hogy az ügyfeleknek/vásárlóknak a segítségére tudjanak lenni, függetlenül attól, hogy mikor, hol vagy hogyan kommunikálnak a vállalattal. És bár a három tényező külön-külön is nagy kihívást jelent, integrálásuk sokkal nagyobb energia befektetését igényli, ha a cél a valós idejű CRM rendszer megvalósítása.

Az „emberek” (a vállalat munkatársai) rész jelentheti a legnagyobb kihívást, tekintve hogy mennyire érzékenyek is tudnak lenni az emberek a változásokra. Minden CRM megvalósítás, mely valamilyen folyamat automatizálását jelenti, szükségessé teszi a dolgozók napi munkamenetének változását. Ameddig pedig nem értik a munkatársak a változás célját/céljait, addig az ellenállásuk (érthetően) nőni fog, különösen abban az esetben, ha a megfelelő képzés vagy képzettség is hiányzik. Így a munkatársak támogatásának és a bevezetési projekttel szemben támasztott bizalomnak az elnyerése kulcsfontosságú.

A „folyamatok” résznek azért kell központi témának lennie, mivel egy rosszul célzott automatizálás lehet hogy hibás folyamatokat fog felgyorsítani. Míg rengeteg cégnek az ügyfeleket jól bevonó üzleti folyamatai vannak (ilyen lehet például a termék/szolgáltatás kiválasztása, fizetése vagy használata), fontos, hogy ezeket a folyamatokat felülvizsgálják, finomítsák a CRM bevezetés alkalmával. Éppen ezen okból lehet veszélyes egy olyan CRM megoldás megvásárlása és bevezetése, mely előre definiált folyamatokkal rendelkezik, azokat automatizálja és támogatja, ráerőltetve azokat a vásárló cégre. Az ilyen megoldások megvásárlása általában akkor olcsóbb, de később további költségek merülhetnek fel (BERGERON 2002).

A „technológia” komponens túlzott hangsúlyossága érthető ugyan (a technológiai megoldások szüntelenül újuló és sokasodó mivolta miatt), de gyakran a teljes projekt kárára van. A technológiával kapcsolatban így két kihívással kell mindenkinek szembesülnie. Az egyik, hogy meg kell birkózni a megoldások szállítóival, kiválasztani a megfelelőt, a másik, hogy a technológiai trendek tetején kell maradni. Innen nézve pozitív, hogy az Internet mennyire forradalmasította a vásárlói kapcsolattartás mikéntjét, ezért mint megoldás, elengedhetetlen a használata.

A fenti három elem megfelelő kezelésével és integrálásával a CRM bevezetési projekt minden fázisában biztosítható a megfelelő arány.

*1. táblázat: CRM bevezetés*

Megvalósítandó, kulcsfontosságú CRM lépések	Hangsúlyos részei
<b>Üzleti követelmények meghatározása</b>	Munkatársak és nagyrészt folyamatok
<b>A bevezetési projekt csapatának menedzsmentjének összeállítása</b>	Nagyrészt munkatársak és folyamatok
<b>A régi és a szükséges rendszerek integrációja</b>	Technológia
<b>A választott CRM megoldás tesztre szabása</b>	Munkatársak, folyamatok és technológia
<b>Pilot CRM rendszer</b>	Munkatársak és technológia
<b>CRM rendszer kiterjesztése</b>	Munkatársak és technológia
<b>CRM rendszer támogatásának kialakítása</b>	Nagyrészt munkatársak és folyamatok
<b>CRM rendszer tovább növelése</b>	Munkatársak, folyamatok és technológia

A fenti táblázatból [1. táblázat: CRM bevezetés] kiderül, hogy melyek azok a kulcsfontosságú lépések, melyeket végig kell járni egy CRM rendszer bevezetése alkalmával, és hogy mely komponensekre kell különös figyelmet fordítani adott lépések során.

Elsőként, az üzleti követelmények meghatározása során egy megfelelően strukturált folyamatot kell végigjárni ahhoz, hogy azonosítsuk és priorizáljuk, hogy mik a felhasználók igényei. Ez főként emberekkel való munkát jelent, meg kell értetni mindenkivel, hogy igenis érdemes átgondolni, hogy melyek a jelenlegi és melyek lehetnek a jövőbeni igényei és elvárásai egy ilyen rendszerrel szemben, illetve milyen elvárásai vannak a CRM megoldással szemben, valamint a megoldás hogyan fog hatni azokra. Mivel a CRM rendszer bevezetési projekt nagyban hasonlít egy szoftverfejlesztési projektre (mivel lényegében, leegyszerűsítve és informatikai szemmel nézve arról van szó), ezért az

információ rendszerek fejlesztése során alkalmazható igényfelmérési és követelmény-összegejtési technikák itt is jól alkalmazhatóak.

Hasonlóan tekintve, az emberekkel való munka fogja a központi szerepet játszani a bevezetési projekt csapatának menedzsmentjének összeállítása során. El kell dönteni, és szigorúan meg kell határozni a felelőségeket, hogy ki melyik szakaszért, részfeladatért vagy területért felel. A team-ek összeállítása a felelőségek meghatározásával és a vezetők kijelölésével kezdődik. következő lépés lehet a team tagjainak kiválasztása. Ezen fázisban még a technológia nagyon kis szerepet játszik.

Ellenben a régi és szükséges rendszerek integrációja során a technológiáé a főszerep. A vállalati alkalmazás architektúra (Enterprise Application Architecture – EAA) kiválasztása, az alkalmazandó keretrendszerek kiválasztása, a közbenső réteget jelentő eszközök kiválasztása mind különösen fontos lesz ebben a fázisban. A CRM projektekre jellemző, hogy mindenki azt gondolja majd, hogy az ő rendszerét (vagy részrendszerét, adatbázisát) célszerű először integrálni, ám mivel ebben a fázisban a technológia a főszereplő, először ebből a szempontból kell meghatározni, hogy mely rendszereket érdemes integrálni, és azokat milyen sorrendben.

Később, a választott CRM megoldás testre szabása során már fontos szerepet játszanak a munkatársak, az üzleti folyamatok és a technológia egyaránt. Az emberek fontosak, mert véleményezniük kell, hogy a testre szabás mennyire elégíti ki szükségleteiket, illetve, hogy az üzletmenet mennyire és hogyan befolyásolja a felhasználóbarátságot. A folyamatok fontosak, mivel ki kell alakítani a meghatározott új ügymenetet – a technológia nyújtotta lehetőségeket kihasználva. A technológia másik fontos része a rendszer felületeinek kialakítása, módosítása vagy törlése. Ki kell alakítani, hogy hogyan célszerű a felületek közti navigáció.

Valós idejű CRM: Persze mára a CRM – mint iparág – megértése rendkívül bonyolulttá vált. Ez köszönhető annak az évi 40%-os növekedésnek amit a szektor mutatott 1995-től egészen a 2000-es évek elejéig, illetve annak, hogy emiatt a gyors növekedés miatt egyre több és több szereplő szállt be erre a piacra, aminek köszönhetően bővültek a lehetőségek, úgy a beszállítói oldalt tekintve, mint a lehetőségek (megoldások) kínálatát. Ahhoz, hogy a cégek kitűnhessenek, újabb és egyedibb megoldásokat kellett fejleszteniük, és ezért gyakran nehéz (gyakorlatilag nem tanácsos) megválaszolni elemzés nélkül a kérdés, miszerint „Mi a legjobb CRM megoldás?”. Egy ilyen – folyamatosan új piaci szereplőkkel,

az új termékekkel, új modulokkal és komplex technológiai alternatívákkal színesedő – szektort egyre időigényesebb átlátni (*GOLDENBERG 2008*).

A gazdasági válság okozta növekedési lassulás ellenére még mindig a CRM az információtechnológiai piac leggyorsabban fejlődő része. A recesszió előtt, 2002-ben például 30 milliárd dolláros bevételt könyvelhetett el a CRM iparág, beleértve ebbe szolgáltatásaik és szoftvereik értékesítését. 2002-ben ugyan a növekedés mértéke a szektorra nézve évi 10%-ra csökkent (magyarázható ez az akkori gazdasági hullámvölgygel), de aztán a 2003-2007 közti időszakban ismételten magas, 15-20%-os éves növekedés volt jellemző.

Szakemberek szerint jelen pillanatban két hatalmas kihívással kell megküzdenie a CRM szektornak: egyfelől túl kell élni a gazdasági válság okozta pénzhiányt, másfelől meg kell birkózniuk az új megközelítésű termék architektúrára való átállással.

Alkalmazott megközelítések: Érezhető némi – technológiai felépítésből és architektúra váltásból adódó – feszültség a CRM piacon. A két alkalmazott technológiai megközelítés a régebbi, kliens-szerver kapcsolaton alapuló, melyre inkább az alkalmazottak felé fordultság a jellemző (cél az alkalmazottak és a cég belső ügymenetének legteljesebb kiszolgálása), illetve az új, web-alapú – gyakran ”e-vásárló” címkével ellátott – megközelítés, mely sokkal inkább az ügyfél információval és szolgáltatással történő ellátására fókuszál. Természetesen a web-alapú fejlesztőeszközök terjedése és fejlődése (az új technológiák térhódítása) mind-mind inkább az web-alapú CRM rendszerek terjedésének kedveznek.

A CRM Kliens-szerver megközelítése: A kliens-szerver kiépítés jelen pillanatban hatalmas ügyfélkörrel rendelkezik (az ügyfelek persze ebben az esetben a CRM szoftvert/szolgáltatást megvásárló cégek, akik saját forgalmukat szeretnék növelni a megoldás alkalmazásával), akik már az iparág ébredése óta élvezik az ötlet előnyeit – az ügyfelek így ezeket a megoldásokat jól bevált CRM szolgáltatójuktól veszik, bízva a folytonos sikerben. Ennek ellenére a legtöbb CRM szolgáltató folyamatosan áll át az új web-alapú CRM megoldások fejlesztésére és szolgáltatására. Ez az átállás persze folyamatos, viszonylag nem gyors folyamat, mivel rendszereik alapjai kifejezetten nem webes közegbe lettek tervezve, de folyamatos a fejlődés.

A valós idejű CRM nem más, mint az a trend, minek lényege, hogy összekapcsolja a vállalatok belső rendszereit és folyamatait hálózatokon és interneten keresztül, kihasználva az IT és a mobil eszközök lehetőségeit, előmozdítva így a valós idejű vállalat létrejöttét.

A módszer alkalmazásával a vállalat idegrendszer szerű hálózattá változtatható, így azonnal reagálni lehet a vásárlói igények, a versenyképességi analízisek eredményeinek, a készletek, a nyersanyagok rendelkezésre állásának vagy a jövedelmezőségi jellemzők változására. Manapság a mindig vásárlásra kész és mindig online vásárló minden eddiginél jobban támaszkodik a mobileszközök és az internet nyújtotta lehetőségekre, ezért fontos szem előtt tartani, hogy bár az alkalmazott technológiai megoldások vannak az előtérben, de a főszerepet még mindig az emberek, a folyamatok és a technológia játssza egy CRM projektben.

CRM trendek: Az alábbi trendek figyelhetők meg jelenleg a CRM piacon (illetve jelölhetők meg mögöttes indokként):

A legtöbb vállalat, aki áldoz egy nagyszabású CRM projekt végig vitelére, tudja, hogy az emberek/folyamatok/technológia megfelelő bevonási aránya megközelítőleg 50% / 30% / 20%. Ezen vállalatok sikerének az oka az, hogy egyfelől rendezik a vásárlóközpontú folyamataikat, másfelől ráveszik üzleti partnereiket (belső munkatársra, üzleti partnerek, illetve ügyfelek/vásárlók), hogy használják a megoldásaikat, illetve harmadrészt ezeket a folyamatokat beveszik és támogatják CRM megoldásukba.

Minden CRM kezdeményezést indító vállalatnak képesnek kell lennie egy CRM terv elkészítésére. Ebben ki kell emelni a teremtendő értékeket, a növelendő termelékenységet, az alacsonyabb elvárt költségeket, a javuló munkaközösség morált, a szélesebb látókörű ügyfeleket, a megnövelendő vásárlói elégedettséget, lojalitást és visszatérést. A felsővezetők általában megértik, hogy ebben nem a technológia kiaknázása a legfontosabb, mindinkább a strukturálás illetve a mérhetőség, ezért követelik általában jól mérési pontokat tartalmazó tervek elkészítését. Az alapszintű metrikák (*baseline metrics* – a kiinduló állapot) és a célmetrikák (ahová meg kell érkeznie a cégnek a projekt sikerével), illetve a mérések elkészítésével igazolható a CRM értelme és értéke.

A technológia fejlődésével és egyre precízebb kiaknázásával a költségek csökkennek és felszínre kerülnek a korai CRM megvalósítások gyengeségei és azok orvoslásának lehetőségei. Jó példa erre a Microsoft .NET platform terjedése és használata (különösen

úgy, hogy a cég belépett a CRM megoldások szállítóinak piacára), segítségével jól és könnyedén integrálhatóak a belső (régebbi ERP megoldások) és külső (például ellátási-lánc menedzsment rendszerek vagy külső információs) rendszerek. Várható, hogy erre a megoldásra válaszul több CRM szolgáltató új modulok fejlesztésével válaszol, mely jó, hiszen a verseny ezáltal csak erősödik (*BUSINESS INTELLIGENCE LIMITED 2008*). Az üzleti analitika fejlődése (és az egyre jobb és kiforrottabb támogató eszközök) is egyre jobb CRM megvalósításokhoz vezet, a mobil eszközök terjedése és a vezetékek nélküli kommunikáció fejlődése lehetővé teszi a költséghatékony hozzáférést, bármikor, nagyon sok helyről, ráadásul egyre inkább készüléktől függetlenül. Ugyanígy jó alapot nyújt az e-learning rendszerek fejlődése, melynek segítségével megvalósítható a költséghatékony, életen át tartó tanulás (Life Long Learning). Továbbmenve, az e-vásárló önkiszolgálást támogató alkalmazások pedig csökkentik az online vásárlások növekvő számához társuló költségeket.

Az ISM megfigyelései szerint 2002 óta a CRM bevezetés költsége 50%-al csökkent. Ez nagyrészt a piaci verseny kiéleződésének köszönhető, valamint az ebből adódó ténynek, mi szerint a szolgáltatók inkább fejlesztik saját CRM megoldásukat, mintsem hogy más megoldáshoz hasonló hibába essenek. Ugyancsak javult a külső CRM bevezetők sikeres bevezetéseinek aránya, valamint egyre több projekt fejeződik be sikerrel, az előre megadott költségkereten belül. A trendekből sejthető, hogy a költségek a közeljövőben tovább fognak csökkenni, ahogy a nyílt szabványokon alapuló felhasználóbarát fejlesztőeszközök rendszeressé válnak.

A siker további sikert szül – tartja a mondás; ezt kihasználva egyre gyakoribb a sikeres CRM megvalósítások publikálása a médiában/online. A CRMGuru.com kiadványai szerint a CRM megvalósítások 67%-a végződik a cég részéről átlagos vagy átlagon felüli eredménnyel – mivel az egykoron megszokott hibák ma már előre jelezhetőek és kiküszöbölhetőek.

A következő néhány évben tovább folytatódhat a CRM szolgáltatók konszolidációja, a vertikális piaci ajánlatok bővülésével, valamint a CRM piac éves 10-15%-os bővülése. Ennek köszönhetően továbbra is bővülő körből válogathatnak a cégek szolgáltatókat és lehetőségeket tekintve, ennek köszönhetően pedig tovább fog nőni a sikeres CRM bevezetések száma, valamint a valós idejű CRM jelentősége.

## 4 Felhasználható matematikai és statisztikai eszközök

Ha már vállalati rendszerekről, és azok modellezéséről van szó, célszerű meghatározni: mit nevezünk rendszernek.

Bár a rendszer definíciójáról a mai napig tudományos vita zajlik (mint ahogy arról is, hogy kell-e definiálni), mégis az általánosan elfogadott meghatározás szerint a rendszer olyan meghatározott struktúra szerinti elemek együttese, melyek egymással kapcsolatban és kölcsönhatásban állnak, egymással és a struktúrával összhangban (*BODLAKI 2009*).

Minden üzleti intelligencia rendszer használata magában foglal jelentős mennyiségű adatkezelést, mivel a cél az, hogy a rendelkezésre álló adathalmazból valamiféle új tudásra tegyen szert a vállalat, ami valamilyen mérhető módon pozitív hatást fejt ki a későbbiekben.

Üzleti intelligencia rendszerek esetén ez az adatkezelés folyamatos – talán ez a nehezen meghatározható különbség az üzleti intelligencia rendszerek és az adatbányászati rendszerek között, hogy az ezek a rendszerek folyamatosan működnek, folyamatos adatkezelést végeznek. Az adatkezelésbe beletartozik az adatok előkészítése (hogy alkalmasak legyenek bármiféle elemzés elvégzésére), a feltárás, a kezdeti elemzés és a kezdeti megjelenítés (meg kell találni bizonyos alap összefüggéseket, esetleg ábrázolni kell ezeket), majd mélyebb elemzések elvégzését. Ezek lehetnek regressziós műveletek (több változó közti függvényszerű összefüggés megállapítása), idősorelemzés (idővel való összefüggések megállapítása), minősítés (az előfordulások különböző minták alapján előre definiált osztályokba sorolás), asszociáció (gyakori elemhalmazok és a köztük fennálló minták keresése) illetve csoportosítás (a vizsgált objektumok / előfordulások jól elkülönülő, előre nem ismert csoportokba / kalszterekbe sorolása). Az adatokkal való foglalkozás itt látszólag véget ér, de az üzleti intelligencia egyik lényege a folyamatos felhasználás, illetve az információ majd tudás előállítása – tehát a folyamat folytonos.

Gyakorlatilag a fenti lépések különböző statisztikai műveletek elvégzését, algoritmusok futtatását jelentik az adathalmazon. Míg matematikából inkább a függvényanalízis módszerei kerülnek használatra, addig statisztikából a teljes paletta használt, ám ezek áttekintése roppant nehéz feladat, egyrészt a sorétűség miatt, másrészt a folyamatos algoritmus-kutatás miatt.



## 4.1 Adatok előkészítése

Mivel szinte minden elemzési feladat elvégzése megkövetel valamilyen bemeneti állapotot, ezért sok esetben (gyakorlatilag mindig alkalmazásra kerül, bizonyos mértékben) szükség van az adatok előkészítésére (vagy előfeldolgozására).

Az adatok előkészítése magában foglalja az adatok egy helyre gyűjtését (adatintegráció, adattárházi kezeléssel), rendezését, az adatok megismerését – így megtudható, milyen hibákra kell számítani, azok javíthatóak, vagy kiküszöbölhetőek – aztán az adatok átalakítását (megfelelő formára hozás), illetve a feldolgozandó adatmennyiség csökkentését (vagy számosságában, vagy dimenzióiban).

Míg az adatintegráció képes megoldani olyan problémákat, mint a több helyen másként szereplő adatok problémája, és egy jól kivitelezett adatintegrációs projekt tökéletes alapot képes biztosítani az üzleti intelligencia rendszernek (azon belül többek között az adatbányászati lépéseknek), addig az adatok elemzésével nem foglalkozik, így a látszólag jó adatok is több hibát tartalmazhatnak – ezért van szükség azok megismerésére. (Attól függően persze, hogy milyen adatintegrációs eszközt használ a vállalat, annak képességei eltérőek – van köztük olyan, mely képes pl. adattisztításra, de van amelyik erre nem koncentrálna.)

A helyes előfeldolgozás számos problémára megoldást nyújthat (jobban mondva helyesen elvégezve ezek a problémák nem fognak előjönni), mint a hiányos adatok problémája, inkonzisztencia és inkompatibilitási problémák valamint a redundanciából és zajosságból adódó problémák. Ezt a műveletet szokta a szakirodalom adattisztításnak is nevezni – melynek kimenete így a tiszta adat (más-más adattípusokat persze máshogy kell tisztítani, illetve amennyiben fontos, megoldható az adattípusok közötti váltás is).

Az inkonzisztencia belső ellentmondást jelent az adatokban. Az egyik kapcsolódó jelenség az integritási szabályok hiánya – azaz, hogy nincs megkövetelt formátum számos esetben; a másik – szintén integráció során gyakran – felmerülő gond a mezők elnevezéséből adódik: ugyanarra az adattartalomra két helyen kétféle elnevezést használnak<sup>46</sup>, vagy más adattartalomra ugyanolyan<sup>47</sup> elnevezést. Nem probléma, de minden esetben figyelni kell az esetleges váltószámok használatára (egyik helyen milliméterben, másik helyen méterben van megadva ugyanaz az érték; egyik helyen USD, másik helyen EUR az árazási alap) –

---

<sup>46</sup> Más néven homonímia hiba.

<sup>47</sup> Más néven szinoníma hiba.

ennek speciális esete, amikor értékek keveredhetnek összegértékekkel. Minden esetben találkozni hiányzó értékekkel: ezeket lehet figyelmen kívül hagyni (ami csökkenti a rendelkezésre álló adatmennyiséget), lehet átlaggal, vagy kategóriaátlaggal helyettesíteni, vagy valamilyen konstans<sup>48</sup> értékkel, de elterjedt a lehetséges értékek közüli sorsolás, vagy a valószínűségi értékadás. A legtöbb adatintegrációs szolgáltatást nyújtó szoftver képes az inkonzisztencia-hibák kiküszöbölésére (be- és kimeneti formátumok kezelésével).

Inkompatibilitás vagy összeférhetetlenség főleg akkor alakul ki, amikor egymással összefüggő, vagy egymásból számított adatok kerülnek elemzésre – tény, hogy egymásból számítható adatokat egy adatbázisban tárolni normálformát sért, ám jelentős idő takarítható meg ilyen értékek tárolásával bizonyos esetekben. Természetesen ekkor külön feladat az értékek frissítésének ellenőrzése.

Redundancia kétféle képpen jelentkezhet: mező szinten (ez már az integráció során előjön), illetve attribútum szinten (egyik attribútum meghatározza a másikat – korrelációanalízis megadja a választ).

Az adatok zajossága nem más, mint a mért értékekre rakódott véletlen hibák összessége. Zaj az adatokban többféle képpen érződhet: folyamatos (a valóstól való) eltérések, vagy kiugró értékek, illetve a kettő kombinációja (*FÜSTÖS ÉS KOVÁCS 1989*). Zaj minden esetben az adatok beszerzésekor (számlálási, vagy szenzor eredmények), vagy továbbításakor (analóg hálózat) keletkezhet. A zaj meghatározható szórással (illetve a zaj mértéke eloszlásfüggvényekkel becsülhető), vagy nagyrészt eltüntethető adatsimítással (de nem teljesen). Adatsimításhoz is több technika áll rendelkezésre: kosarazás (értékehelyettesítés mozgó átlaggal vagy mediánnal), klaszterezéssel (és a szélső értékek elhagyásával), kivételek vizsgálatával vagy regresszióval.

A döntéstámogatási projektekre is igaz az a szomorú statisztikai adat, mi szerint az ezen projektek kevesebb mint fele sikeres csak (*FRESE ÉS SAUTER 2003*). Ennek több oka lehet:

- adatbányászati szakember-hiány
- célterület-specialista szakember-hiány
- szervezési-együttműködési problémák
- nem megfelelő adatok használata: az adatokkal kapcsolatban több elvárás is van:

---

<sup>48</sup> Konstans lehet az adatbázis kezelő rendszer *null* értéke is, ám nem szerencés – ezt az értéket is egyes adatbázis kezelők másként tárolják.

- annál jobb, minél nagyobb mennyiségű adat áll rendelkezésre. Ha kellően nagy mennyiségű adat áll rendelkezésre, job eséllyel zárhatóak ki az eseti összefüggések és a véletlenek. A nagy mennyiségű adat feldolgozása azonban drágább: időben és infrastruktúra oldalról nézve mindenképpen.
- az attribútumok számának növekedése ugyancsak jó az elemzés szempontjából, mivel így több lehetőség adott összefüggések felderítésére. Ha csak kevés attribútum van jelen, nem szükségesek a fejlett elemzés eszközei, mivel hagyományos módon is feltárhatóak az összefüggések.
- az adatoknak tisztáknak<sup>49</sup> kell lenniük, de mivel ez sok esetben elérhetetlen, elegendő az is, ha ismerjük a zaj / torzítás mértékét
- az adatoknak torzítatlannak kell lenniük. Figyelni kell a kiválasztás menetének torzításmentességére.
- metrikák és mérési pontok megléte: fontos szempont, hogy kimutatható legyen a projekt és a befektetések megtérülése.

Kimondható tehát, hogy amennyiben a fent megnevezett (és a korábban említett) problémaforrásokat sikerül időben kiiktatni, a sikerre jelentősen nő az esély.

#### 4.1.1 Módszertani kérdések

A módszertani kérdések tisztázásához először tisztázni kell, mi is a különbség a módszer és a módszertan közt. Az ide leginkább illő meghatározás Raffai Máriától: *„A módszer bizonyos feladatok elvégzéséhez szükséges, meghatározott feltételek közt érvényes szisztematikus végrehajtási mód és ennek előírása. A módszertan pedig a módszerek egységes rendszere, amely meghatározott filozófiai álláspontra helyezkedve specifikálja az eszközöket és a technológiát is.”*

A döntéstámogatás során felhasznált módszerek statisztikai műveletek és leírásaik, a módszertan a későbbiekben egy bevált, ajánlható rendszer lehet, megadva az eszközöket és az ajánlható technológiát.

A döntéstámogatás matematikája nagyrészt statisztikai műveletek sokaságát jelenti. Egyfelől validálni kell az adatokat, hogy egyértelmű legyen, valóban jók az adatok a

---

<sup>49</sup> Az adattisztítás az adatok elemzésre való felhasználását megelőző lépés. Ennek során kell a benne rejlő hiábkat eltávolítani, ilyenek általában a hiányzó értékek, beviteli hibák, extrém értékek és lehetetlen értékek.

vizsgálathoz, elvégezhető az adathalmazon a számítás, és végeredményként meghozható a döntés.

Mivel az adatok nagyon sokfélék lehetnek, azokat át kell alakítani valamilyen jobban használható formára (ezt a transzformációt nem szabad összekeverni az ETL-folyamatok<sup>50</sup> közben zajló transzformációval – bár mindkettő adatokkal foglalkozik, itt azok előkészítése zajlik, míg ETL során már az adatbázisba töltésük).

Végül valamilyen módon tömöríteni kell az adatokat, hogy az egész számítást ne a teljes adathalmazon kelljen elvégezni. Így csökken a számítási kapacitás igény, valamint idő és erőforrás takarítható meg. Idővel, ahogy a rendelkezésre álló számítási kapacitás tovább növekedni fog, kérdéses lesz az adatok tömörítésének szükségessége (legalábbis a végső elemzés esetére).

#### **4.1.1.1 Validáció**

A validáción jelen esetben<sup>51</sup> az adatok megfelelőségének eldöntését kell érteni. A felhasználásra kerülő adatok validálásának három vetülete lehet. Az első, hogy meg kell birkózni a hiányzó adatokkal, a második a befolyásolt (vagy zajos) adatok szűrése, a harmadik pedig annak eldöntése, hogy valóban az adott adatokra van szükség a döntés meghozatalára, illetve elméletileg lehetséges-e azokból az adatokból meghozni a döntést (*SAS PUBLISHING 2009*).

Mivel a döntéstámogató / üzleti intelligencia rendszerekben általában nagymennyiségű, felhalmozott és ugyancsak nagy mennyiségű, új adat áll rendelkezésre, melynek egy része automatikusan lett rögzítve, másik része manuálisan került rögzítésre, sajnos elkerülhetetlen, hogy a rendszerbe hiányzó adatok is kerüljenek. Ennek oka lehet emberi figyelmetlenség, mulasztás, vagy valamely információtechnológiai rendszer hibája, de az is feltételezhető, hogy nem állt rendelkezésre adat. Ebbe a problémakörbe sorolhatóak az olyan esetek is, amikor duplikált vagy érvénytelen adatok szerepelnek. Mára az összes rendelkezésre álló, valóban nagyvállalati környezetben használható üzleti intelligencia rendszerben meghatározható olyan szabályrendszer, mely segítségével gyorsan, nagy

---

<sup>50</sup> Az ETL-folyamat az angol Extract-Transform-Load szavakból képződött, jelentése: Forrásadatbázisból adatok kinyerése, azok megfelelő formára hozása, átalakítása, majd a céladatbázisba / adattárházba való betöltése.

<sup>51</sup> Informatikában a validáció általában olyan kiértékelést jelent, mely az aktuális folyamat eredményeét vizsgálja azzal a céllal, hogy az megfelel-e a követelményeknek (*SZIRAY 2009*).

menyiségű adat „javítható”, felmentve a statisztikusokat a manuális szűrés alól (*FERNÁNDEZ 2003*).

A befolyásolt (vagy zajos) adatok problémája különösen a tanuló rendszereknél jelenthet problémát. A ok lehet mérési hiba, emberi tényező, szakértői hiba a tanító példákat illetően, vagy a hiányzó adatok nem megfelelő pótlása (*REFAAT 2007*). A tanító adathalmazban levő ilyen hibák komoly következményekkel járhatnak:

- lehet, hogy az előrejelzések annyira rosszak lesznek, hogy már használhatatlanná válik (az ellenőrző adathalmaz használatával persze kiderül a hiba megléte), illetve
- túlságosan bonyolult döntési fa állhat elő, ami már emberek számára átláthatatlan (ekkor a döntési fa visszanyesése lehet jó megoldás, de törekedni kell egyszerűbb döntési fák képzésére).

Azt pedig, hogy a rendelkezésre álló adatokból meg lehet-e hozni egy döntést, inkább alapvető átgondolás, a vállalaton belül és a vállalat környezetében zajló folyamatok beható ismerete és az összefüggések átlátása szükséges. Ezek tekintetében az adatok validációja lehetséges.

#### **4.1.1.2 Transzformáció**

Az adatok használatra való felkészítésére (is transzformáció, azonban nem keverendő össze az ETL folyamatok során végzett transzformációval) főként a bevezetés során, az első használat előtt, nagy mennyiségű adatok esetén van szükség (persze később is, folyamatosan, de addigra már megvan a kialakult gyakorlat, a rendszer működik).

Itt transzformáció alatt elsősorban olyan statisztikai műveleteket értünk, melyek alkalmassá teszik az adathalmazt számítógépes feldolgozásra és értelmezhető statisztikák készítésére. Ilyen műveletek a standardizálás és a jellemző-kiválasztás (feature extraction) (*LIU ÉS MOTODA 1998, GUYON 2006*).

A standardizálás célja, a heterogén adathalmaz (mely statisztikai szempontból a sokaságot jelenti) főátlagainak és intenzitási viszonyszámainak elemzésével a részátlagok és összetétel (súlyok) különbözőségéből adódó hatásokat különválassza és mérje. Ezt a funkciót a legtöbb statisztikai szoftver támogatja.

A jellemző-kiválasztás az adatok dimenziócsökkentésének egyik módja (*GORBAN, KÈGL ÉS WUNSCH 2007*). Hagyományosan a képi információk feldolgozásánál használták, de már

minta felismerés során is régebb óta alkalmazzák. Különösen ott ajánlott a használata, ahol rengeteg adat áll rendelkezésre, de a mennyiség annyira redundáns, hogy csak mérsékelt információtartalommal rendelkezik. Ilyenkor a feldolgozhatóság és kezelhetőség érdekében célszerű néhány főbb jellemzőt kiválasztani, hogy ettől a ponttól kezdve ez reprezentálja az adathalmazt. Amennyiben a kiválasztás során a megfelelő jellemzők maradnak a halmazban, az adatmennyiség információtartalma (közel) változatlan marad, de a feldolgozhatósága jelentősen javul. Ahhoz, hogy a kiválasztás során megfelelő eredmények legyen elérhetőek, a szakterületen jártas személy részvétele szükséges – így az elérendő célhoz kötődő jellemzők nagyobb valószínűséggel választhatóak ki. Léteznek persze általános algoritmusok is a kiválasztásra (főkomponens elemzés, fél-végleges beágyazás, multifaktor és nemlineáris dimenzió redukció, független komponens elemzés, isomap, kernel PCA, rejtett jelentés elemzés, stb.), de a szakértői segítség bizonyíthatóan hasznosabb (PERNER 2009). A kiválasztás szakértői folyamata, és általános eljárásai mára a jelentősebb adat-előkészítő és adatbányászati szoftverben támogatottak.

A transzformálás / transzormáció során szokott még szerepet kapni a simítás, az általánosítás (bizonyos értékcsoportok konkrét értékekkel való helyettesítése), normalizálás (adatok levetítése az egyik értelmezési tartományról a másikra (FÜSTÖS KOVÁCS MESZÉNA ÉS SIMONNE 2004); többféle normalizálás használható: lineáris, logaritmikus; de a széthúzás is bizonyos értelemben normalizálásnak minősül), de lehetséges új attribútumok készítése is (FÜSTÖS ÉS KOVÁCS 1989).

#### **4.1.1.3 Adattömörítés**

Sokrétű üzleti intelligencia-elemzésekhez általánosan elmondható, hogy a nagyobb rendelkezésre álló adathalmaz előnyösebb – persze az egyes elemzésekhez nincs szükség minden adatra. Ha az adatmennyiség a dimenziók számának csökkentése után is túl nagy a kezeléshez és az elemzések ésszerű időn belül történő elvégzéséhez, akkor valamilyen adattömörítési módszerre (BRYMAN ÉS HARDY 2009) van szükség (a tömörítés itt nem keverendő össze a fájlműveletek során használt tömörítéssel, a közös vonás csak a méretcsökkenés).

Az adattömörítésre több lehetőség van: a mintavétel, a kiválasztás, a főkomponens elemzés, illetve a diszkretizálás.

A mintavétel célja egy olyan [19. ábra] előállítás, mely méretben ugyan kisebb az eredetinél, ám jellemzőit tekintve megegyezik azzal – legyen reprezentatív, legyen véletlen

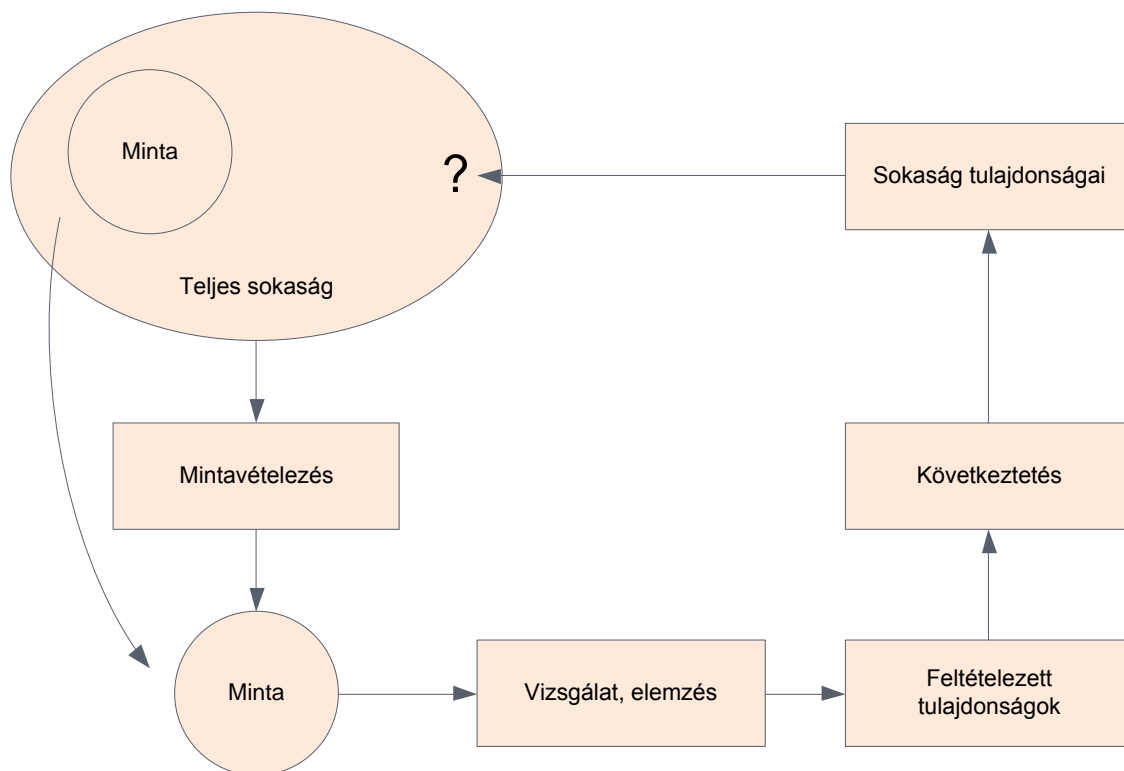
és az elemzések értékelhető elvégzéséhez kellő méretű. Itt kell megjegyezni, hogy a számítási kapacitás rohamos növekedésének, illetve az adattárolás árának drasztikus csökkenése miatt egyre inkább adott a lehetőség az összes rendelkezésre álló adat tárolására, illetve egyre kevésbé lesz szükséges a mintavételezés használata – jelenleg is számos olyan (számítógéppel támogatott) elemzés van, ahol már nem alkalmaznak mintavételt, hanem a teljes adathalmazon megy végbe az elemzés – ilyen esetekben persze általában nem szükséges a gyors válaszidő (*HSIAO 2003*).

A minta a legtöbb esetben tartalmaz hibákat, ennek oka főleg a tömörítés, a közelítés és a becslés lehet (*CHAMBERS ÉS SKINNER 2003*). Mivel azonban a mintavételi hiba a minta nagyságától függ, és a méretével fordítottan arányos, ezért idővel (→számítási kapacitás növekedése) ez a tényező már nem fog gondot okozni.

A mintavétel módja a kiválasztás. Alapvetően kétfajta kiválasztási mód adott: véletlen, nem véletlen kiválasztás. Véletlen mintavétel esetén véletlenszerű a választás, minden elem (objektum vagy rekord) azonos valószínűséggel kerül a mintába, egyszerű véletlen (EV) minta esetén visszatevés nélkül (ha a minta elemei nem függenek egymástól, a minta növelésével nő az esély az elemek mintába kerülésére). Fontos jellemző a kiválasztási arány. Alkalmazható persze visszatevéses kiválasztás (a mintavételek számának növelését kűszöböli ki, de vannak hátrányai) is.

Létrehozható rétegzett (heterogén sokaság esetén homogén rétegek képzése, átfedésmentesen, rétegenként véletlen kiválasztással), csoportos (előre adott tulajdonságok szerint) és többlépcsős minta (több tulajdonság esetén) is.

A nem véletlen kiválasztásos minták közé tartoznak a szisztematikus kiválasztás, a kvótás kiválasztás (ha a minta összetétele előre adott), a koncentrált kiválasztás (a leginkább jellemző elemek kerülnek a mintába), a hólabda kiválasztás és az önkényes (vagy szubjektív) kiválasztás (*DAVENPORT ÉS PRUSAK 2001*).



forrás: Forrás Zoltán: Tudásmenedzsment című műve alapján saját munka

19. ábra: A mintavétel elve

A mintavételezést több szabvány is szabályozza. A jelentősebb<sup>52</sup>, ide tartozók:

- ISO 2859 (a nemzetközi minősítési rendszerből, mely megadja az attribútumok szerinti mintavételezés lehetséges módjait)
- MIL STD 105 D (katonai minősítési rendszer, mely ugyancsak az attribútumok szerinti elemzéssel foglalkozik)
- Léteztek kifejezetten magyar szabványok is (pl.: MSZ 247-1 és MSZ 213), de ezeket a nemzetközi szabványokkal való átfedés miatt már visszavonták.

A főkomponens elemzés részben dimenziócsökkentő eljárás ugyan, mivel célja azon jellemzők kiválasztása, melyek a legnagyobb magyarázóerővel bírnak, így azok kiválaszthatóak és a dimenziók csökkenésével jelentősen csökkenthető a kezelendő adatmennyiség.

A diszkretizálás ugyan elsősorban inkább információelméleti kérdés lehet, ám fontos vetületei vannak adatfeldolgozás terén, mivel a nagyvállalati körben számos olyan adatforrás van,

<sup>52</sup> A teljes mintavételezést befolyásoló szabványgyűjtemény elérhető a <http://www.variation.com/techlib/standard.html> weboldalon



mely automatikus adatgyűjtőkön keresztül érkezik be. Már az üzleti intelligencia rendszer tervezésekor ki kell térni olyan kérdésekre, hogy milyen statisztikai folyamatokat kell a későbbiekben támogatni, és azoknak milyen adatigénye van (*GISBERT 2005*). Ennek kapcsán meg kell határozni, hogy milyen időközönként szükséges az adatok gyűjtése ezekről az automatikus adatszolgáltató érzékelőkről (a folyamatos kiolvasás a számítógépek és minden digitális eszközök korlátai miatt nem lehetséges, a túl sűrű kiolvasás felesleges adatokat és nagy adathalmazt eredményez, míg a túl ritka ronthatja az elemzések hatásfokát). A diszkretizálás persze értelmezhető más típusú adatokra is, de a cél minden esetben a megfelelő mennyiség előállítása.

A vizsgálat és elemzés során a tisztított, szűrt, mintából szeretnénk következtetéseket levonni, majd azokat a teljes sokaságra kivetíteni. Az elosztott adatfeldolgozás üzleti informatika rendszerekben történő elterjedése persze lehetővé tette, hogy ne kelljen mintavételezést végezni, mivel megoldható már a teljes sokaságon az elemzés elvégzése. Ennek ellenére célszerű még mindig mintán dolgozni, az elemzést először mintán elvégezni, esetleg itt van lehetőség a különböző elemzési módszerek gyors összehasonlítására is (az elemzési módszer egy jó mintán ugyanúgy viselkedik, mint a teljes sokaságon), majd a preferrált módszert kiválasztva, a folyamatot kialakítva annak forrásául nem a mintát, hanem a teljes sokaságot felhasználni – az már tervezési döntés, hogy erre szükség van-e, de bizonyos esetekben az eredmény pontosságát, megbízhatóságát jelentősen növeli a teljes sokaságon végzett elemzés.

#### **4.1.2 Technikai kérdések**

A technológiai haladásnak köszönhetően (a számítási kapacitás rohamos fejlődése) a statisztikai számítások ideje nagymértékben rövidült, ezzel párhuzamosan a hardverárak csökkentek, valamint számos új elemzési technika látott napvilágot, így egyre több cég számára vált (válík) elérhetővé az üzleti intelligencia alkalmazása.

Technikai kérdésként felmerül ugyanakkor, hogy a vállalat milyen üzleti intelligencia megoldást válasszon. A kiválasztás során számos szempont vehető figyelembe:

- a vállalati folyamatok komplexitása (milyen bejövő adatokra, milyen folyamatok vonatkoznak, a folyamatokban hol és milyen szereplők vesznek részt),
- más rendszerekkel való együttműködési igény (a vállalatnál milyen olyan más információs rendszerek működnek, melyekkel együtt kell működni, adatokat kell

cserélni, vagy csatlakozni kell), ide tartozik főként a vállalatirányítási rendszerekhez való csatlakozás igénye,

- a kimeneti igények (vagyis mik a rendszer elvárt eredményei, azoknak hol, milyen módon és rendszerességgel kell megjelenni, mely folyamatok alapján),
- gazdasági számítások,
- az elvégzendő feladat összetettsége (statisztikai és elemző képességek iránti igények).

Megvalósíthatósági tanulmányban kell elemezni a szóba jövő alternatívákat (*SOMMERVILLE 2007*). Azt, hogy melyik szempont milyen súllyal kerül számba véve, cégfüggő. Általánosan azonban elmondható, hogy az egyik legfontosabb szempont (a mindenképpen magas szoftverárak és rendkívüli bevezetési költségek miatt) a már meglévő szoftverrendszerekhez való legteljesebb igazodás. A szoftvercégek konszolidációjának köszönhetően több olyan szoftvercég is van már, melyek egyidejűen árulnak vállalatirányítási és üzleti intelligencia rendszereket – ilyen esetben ugyancsak bevett a teljes technológiai stack egyetlen beszállítótól történő beszerzése (*BI.HU 2009*).

## 4.2 Feltárás, elemzés és megjelenítés

Amennyiben az adatok elemzése közben számít a minőség, célszerű első lépésnek a feltárást választani. Ez különösen igaz az adatbányászatra, amikor a meglévő adatokban kell addig ismeretlen összefüggéseket keresni (*NICHOLSON 1974*). A feltáráshoz persze az adatoknak már eleve rendezett formában kell lenniük, hogy aztán elemezhető és végeredményként jellemezhető legyen. A jellemzés / leírás történhet statisztikai számításokkal vagy vizuális megjelenítéssel – a lényeg, hogy a céloknak megfeleljen (*ELLIOTT ÉS MARSH 2009*).

Első ránézésre a feltárás folyamata sokban egyezik magának az adatbányászatnak a folyamatával, ám közelebbről nézve más a helyzet: statisztikai szemszögből nézve a feltárás leíró technikákat alkalmaz, míg az adatbányászat a leíró módszerek mellett valószínűségeen alapuló módszereket is használhat. A két eljárás célja között is különbség van: a feltárás célja az adatok struktúrájának és a bennük rejlő összefüggések leírása, hogy a modellezés során később használható legyen, míg az adatbányászat célja a döntési szabályok kialakítása, a kialakított modellre és összefüggésekre alapozva. Az alkalmazott technikák között is vannak különbségek: a feltárás során több, az adathalmaz különböző

aspektusát vizsgáló feltáró technika áll rendelkezésre, míg adatbányászat során a különböző technikák / módszerek kiértékelése zajlik, abból a célból, hogy melyik segíthet jobb döntési szabály létrehozásában.

Adatelemzést persze kétféle képpen lehet végezni. Megkülönböztethetőek felátró<sup>53</sup> típusú és igazoló<sup>54</sup> elemzési eljárások. A feltáró elemzések célja (korlátozott értelmű) magyarázatok, összefüggések és hipotézisek felállítása, melyből sokrétű technikákkal modellek alkothatóak, és később igazoló elemzéssel ellenőrizhető a helyessége (*FÜSTÖS ÉS KOVÁCS 1989*).

Eldöntendő még az elemzés korai szakaszában, hogy az adathalmaz, ami rendelkezésre áll, mennyire biztos. Az adatok biztosságával kapcsolatban két jellemző áll rendelkezésre: a megbízhatóság és az érvényesség. A megbízhatóság (illetve bizonyos aspektusban az érvényesség is) a szórás segítségével és korreláció számítás segítségével számítható.

A feltárás során lehetőség van (célszerű is) az adatok megjelenítésére. Attól függően, hogy milyen adataink vannak, rengeteg ábrázolási technika közül lehet választani (az adatvizualizációnak komoly szakirodalma van). Két attribútum közti összefüggést általában jól lehet leírni pont diagrammal, vagy vonal diagrammal, amennyiben egy attribútum alakulásának szemléltetése a cél, akkor az oszlop és sávdigramok (gyakorisági értékek mutatására a hisztogram alkalmas – mely az oszlopdiagramból származhat) állnak rendelkezésre, míg százalékos értékek megmutatására a kördiagram és leszármazottai alkalmasak. Több változó együtte megjelenítésére már inkább célszerű pontdiagram mátrixot vagy három dimenziós diagramokat rajzolni, vagy akár a páruzámos koordináta megjelenítést alkalmazni.

Természetesen látványos és könnyedén áttekinthető dolog a mélyrehatóbb elemzés elkészítése után az ábrázolás megismétlése – amennyiben értelmes. Az így kapott ábrákat lehetséges akár vezetői információs rendszerek (vagy üzleti intelligencia rendszerek ilyen moduljával) dashboard moduljával összefogni.

#### **4.2.1 Univariáns feltáró elemzés**

Az előzetes elemzés során a változók külön-külön történő vizsgálata fontos szempont, mivel a későbbi több változós elemzésekhez és modellkészítéshez fontos információkat

---

<sup>53</sup> A feltáró elemzés másik neve exploratív elemzés.

<sup>54</sup> Az igazoló elemzés a szakirodalomban sok helyen konfirmatív elemzésként fordul elő.

szolgáltathat. A főbb rendelkezésre álló eszközök a grafikonok és az összeg indexek. A felhasználható grafikonok típusai az adat típusától függenek (legáltalánosabban oszlopdiagramokat vagy hisztogramokat <sup>55</sup> (kategória-frekvencia összefüggés) és tortadiagramokat használnak).

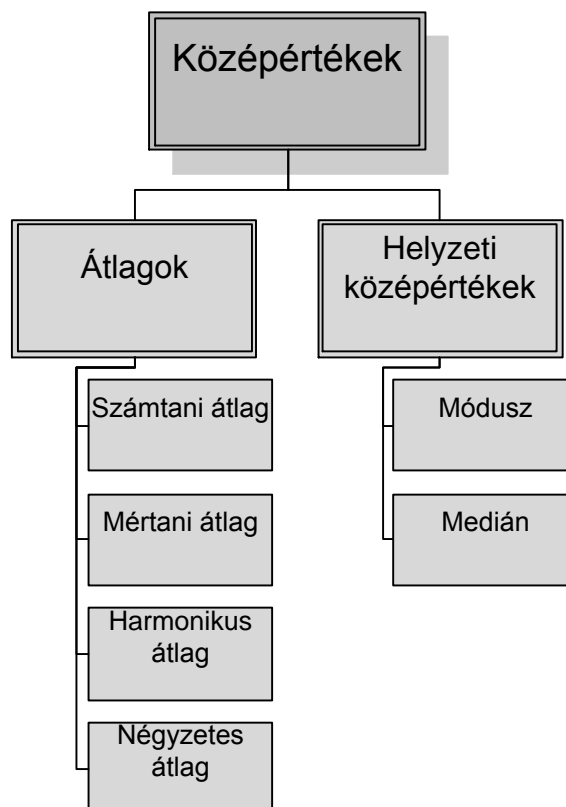
A folyamatos kvantitatív változók frekvencia eloszlásának számításához először a változókat diszkrétizálni kell (ROYSE, THYER ÉS PADGETT 2009). Először az intervallumok szélességét kell meghatározni: célszerű vagy azonos szélességeket választani, vagy változó szélességeket, de azonos frekvenciával. Megjegyzendő persze, hogy itt információveszteség keletkezhet, mivel feltételezzük, hogy a változó egyenletesen oszlik el minden kategóriában. Szintén megjegyzendő, hogy az ismételt klasszifikáció érdekes mintákat fedhet fel. az intervallumokba helyezett folyamatos változók grafikai megjelenítésére a hisztogram a legalkalmasabb (az intervallumok egymás mellett helyezkednek el, a magasságot pedig a frekvenciasűrűség határozza meg). Folytonos ismérv alapján vonaldiagrammal gyakorisági poligon, illetve gyakorisági görbe képezhető.

A statisztikai indexek számítása [20. ábra] akkor lehet hasznos, ha a megfigyelések fontos aspektusainak összegzését kell elvégezni. Az egydimenziós statisztikai indexek lehetnek az elhelyezkedésre, változékonyságra, a heterogenitásra, a koncentrációra, az aszimmetriára illetve a lapultságra vonatkozóak.

Az elhelyezkedés értékelésének leggyakoribb eszköze az átlag számítása – csak kvantitatív változókra alkalmazhatóan. Mivel a számításban minden érték részt vesz, így az extrém kiugró értékek eltéríthetik az eredményt (pénzügyi területen pedig gyakoriak a kiugró értékek). Ezért javasolható inkább valamilyen robosztusabb számítás használata. Jobb megoldás tehát a súlyozott számtani átlag használata – főleg akkor, ha a frekvencia eloszlás már rendelkezésre áll.

---

<sup>55</sup> A hisztogram koordináta rendszerben elhelyezett olyan oszlopdiagram, melynek oszlopai hézag nélkül illeszkednek egymáshoz; az x tengelyen a mennyiségi ismérv osztályközeit, kategóriáit, az y tengelyen pedig a gyakoriságokat, illetve relatív gyakoriságot kell feltüntetni.



*forrás: Kerékgyártó, Mundruczó és Sugár 2001 alapján saját munka*

*20. ábra: Középértékek*

Az elhelyezkedés mérésének második eszköze a módusz számítása, ez azt adja meg a legnagyobb frekvenciával előforduló értéket. Előnye, hogy alkalmazható diszkrét vagy minőségi értékek használata esetén is. A folytonos ismerv módusza az az érték, ami a gyakorisági görbe maximumán található. A harmadik eszköz a medián használata. Használatához az adatoknak eleve rendezettnek kell lenniük. A medián értékét a sorban a középső elem értéke adja. Kvantitatív változók mellett ordinális minőségi változókra is alkalmazható, valamint sokkal kevésbé befolyásolják értékét a kiugró értékek.

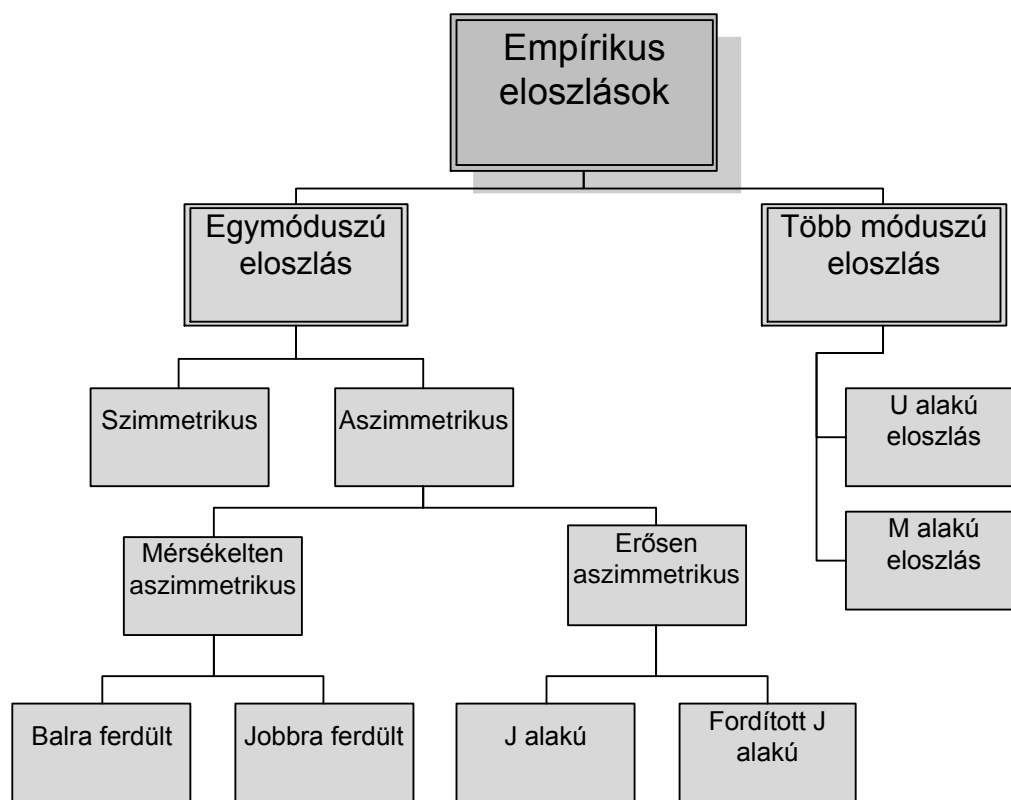
A kvantilis értékek a számszerűen megfigyelt adatok nagyság szerinti rendezését szolgálják, a sokaság mennyiségi ismerv szerinti eloszlásának, az ismervértékek elhelyezkedésének tömör leírását adják. Két részre osztás esetén a medián, négy részre osztás esetén a kvartilisek, öt rész esetén a kvintilisek, tíz résznél decilisek, míg száz rész esetén percentilisek nyerhetők.

Általában érdekes a szóródás vizsgálata. Az értékek különbözősége, változékonyság a szóródás. Értéke mindig a legkisebb és legnagyobb érték által meghatározott terjedelmen belül helyezkedik el. Mivel a tartomány nagyon érzékeny az extrém értékekre, ezért jobb

megoldás az interkvartilis terjedelem használata. Az interkvartilis terjedelem a felső és alsó kvartilis különbsége. A szóródás leggyakrabban alkalmazott mérőszáma a szórás (sokféleség), mely az ismértértékek számtani átlagtól való eltérését használja fel. Előnye, hogy valamennyi értéket figyelembe veszi és algebrailag könnyen kezelhető. Hátránya, hogy ugyancsak érzékeny a kiugró értékekre.

A koncentráció sokban kötődik a sokféleséghez. Mondható úgy is, hogy a frekvencia eloszlás maximálisan koncentrált, ha annak nulla a szórása, illetve minimálisan koncentrált, ha sokfélesége maximális. Számításának célja, hogy kiderüljön, hogy az értékösszeg mennyire koncentrálódik a sokaság bizonyos értékeire. <<Az értékösszeg kevés számú egységre való összpontosulását koncentrációnak nevezzük.>> Megkülönböztetünk abszolút és relatív koncentrációt. Az abszolút koncentráció abban az esetben fordul elő, ha a teljes értékösszeg kevés számú egységhez tartozik. Relatív koncentráció akkor fordul elő, ha az értékösszeg egyenetlenül oszlik el a sokaság egységei között. A relatív koncentráció mérésére a szóródás számításán kívül a koncentrációs tábla, a kvantilis eloszlás és a Lorenz-görbe ad segítséget (*LAMBERT 2001*).

A statisztikai megfigyelés eredményeként kapott gyakorisági sorok, illetve ábrák igen nagy változatosságot mutatnak, ám nagy részüknél felfedezhető valamiféle hasonlóság, mely alapján néhány egyszerűbb eloszlástípusba sorolhatóak. Az (empirikus) eloszlásokat a gyakorisági görbéjük alapján lehet megkülönböztetni (persze gyakorisági sor vagy hisztogram alapján is lehet következtetni).



forrás: Kerékgyártó, Mundruczó és Sugár 2001 alapján saját munka

21. ábra. Empírikus eloszlások

Az egy móduszos eloszlásoknak [21. ábra] egy sűrűsödési helyük van (maximum érték), míg a több móduszos eloszlásoknak két vagy több tömörülési helye van (lokális maximummal), ezek a bimodális és multimodális eloszlások. A szimmetrikus eloszlásoknál a számtani átlag, módusz és medián értéke megegyezik.

#### 4.2.2 Bivariáns feltáró elemzés

A mennyiségi ismérvek közötti sztochasztikus összefüggés a korreláció. A korrelációs kapcsolatok statisztikai módszerekkel végzett elemzésével nagymértékben gazdagíthatóak a jelenségekről és összefüggéseikről alkotott ismeretek. Itt azonban a statisztika kizárólagos ismerete és használata nem elegendő, nagyon fontos a célterület szakmai ismerete, hogy az ok-okozati viszonyok megfelelően értelmezve legyenek feltárhatóak (HARTWIG ÉS DEARING 1979). Az időbeni változásokat tanulmányozva a különböző jelenségek gyakran úgy járnak be hasonló utakat, hogy közöttük valódi összefüggés lenne. A kapcsolat elemzésének eszközei a regresszió számítás és a korrelációs számítás. A regresszió számítás az összefüggésekben lévő tendenciát vizsgálja, és a kapcsolat

természetét valamilyen függvénnyel írja le, az ok-okozati összefüggések ily módon számszerűsíthető válnak. A korrelációs számítás a vizsgált változók közti kapcsolat erősségét, intenzitását vizsgálja. A két számítás kiegészíti egymást, elvégzésük együtt ajánlott.

A változók közti összefüggések felderítésében nagy szerepet kap a grafikus ábrázolás. Kétváltozós vizsgálatok esetén a pontdiagram a megfelelő eszköz (scatterplot – az X tengely a magyarázó változó, Y tengely a függő változó értéke). A diagram pontjai által felvett alakzatból lehet következtetni a regresszió típusára. A lineáris regresszió függvényének meghatározása során arra kell törekedni, hogy a vizsgált értékpárokra leginkább illeszkedő egyenes legyen rajzolható (a függvény alapján). Ez általában a legkisebb négyzetek módszerével érhető el (*ENDERS 2010*).

A korreláció számítás során a függetlenségi kritériumból kiinduló a kapcsolat szorosságát jellemző kovariancia mérőszám (mely függetlenség esetén nulla értéket vesz fel) és lineáris korrelációs együttható számítható.

A nemlineáris modellek jelentős részénél a változók transzformálásával visszavezethető lineáris modellre – nagyban megkönnyítve a munkát.

#### **4.2.3 Multivariáns feltáró elemzés**

Mivel egy jelenség változása általában több másik változó értékétől függ, ezért célszerű a magyarázó modellbe további változók felvétele. Erre ad lehetőséget a többváltozós regresszió számítás. Így a bevezetett parciális regressziós együtthatók megmutatják, hogy adott tényező változó változása esetén mekkora változás várható az eredményben, a többi tényező hatásának figyelembe vétele nélkül (*YATES 1987*).

Hasonlóképpen, a parciális korrelációs együtthatók mutatják meg, hogy milyen szoros a kapcsolat az adott magyarázó változó és a célváltozó közt (ilyenkor a többi változó okozta hatásokat mindkét értékből ki kell zárni). A többszörös korrelációs együttható a tényezőváltozók és az eredményváltozó korrelációjának szorosságát jellemzi. (A determinációs együtthatók a ezen együtthatók négyzetét jelentik.) A determinációs együttható rámutat, hogy az egyes magyarázóváltozók az eredményváltozók variációjának hány százalékát magyarázzák meg (*HARTWIG ÉS DEARING 1979*).



A többváltozós lineáris regressziós modell széles körben használt, hatékony elemzési eszköz az olyan esetekben, amikor valamelyik jelenségnek több más jelenségtől való egyidejű függésének vizsgálata a cél.

### 4.3 Regresszió

A regresszió két változó közti kapcsolatot vizsgál. Míg a regresszió számítás az összefüggésekben lévő tendenciát vizsgálja, és a felfedezett kapcsolatot valamilyen függvénnyel írja le, az ok-okozati összefüggést számszerűsítve, addig a korrelációs számítás a kapcsolat erősségét jellemzi. A legelterjedtebb regressziós technika a lineáris regresszió (ABONYI 2006). A pontdiagramra legjobban illeszkedő egyenes meghatározásához a legkisebb négyzetek módszere használt (SEBER ÉS WILD 2003).

Gazdasági helyzetekben azonban csak ritkán lehet használni a lineáris modelleket, mindinkább a nemlineáris modellek alkalmazására van szükség (pl. a Cobb-Douglas termelési függvény hatványkitevős). A nemlineáris modellek egy részét azonban vissza lehet vezetni lineárisra.

Nemlineáris modellt alkot például a paramétereiben lineáris regresszió, az interpoláció, a használtak neuális hálózatok (és bázisfüggvények) is.

Az interpoláció használata ott hasznos, ahol csak kevés számú mérés áll rendelkezésre, és valamilyen közbülső érték meghatározása a cél. Közbenső értékek számíthatóak interpolációval, ellenben extrapoláció, vagyis előreutató becslést nem lehet vele végezni. Az interpoláció három lépésből áll: részhalmaz választás (a keresett érték szomszédos pontjainak azonosítása), polinomiális illesztés (regresszió számítása a kiválasztott értékekre) és értékelés (célponti érték kiszámítása).

#### 4.3.1 Regressziós modellek felépítése

Az elemzés és az összefüggést leíró függvény alakjának meghatározása után lehet eldönteni, milyen regressziós modell is a legmegfelelőbb a célnak. Ehhez manuális számítás esetén hasznos a pontdiagram (scatterplot), amin a pontok vonulási irányának eldöntése a cél. Ha egyenessel jól jellemezhető, akkor lineáris regressziós modell szükséges, de alkalmazható hatványkitevős, exponenciális, logisztikus vagy többváltozós modell is. Többváltozós modell esetén az vizsgálható, hogy az egyes változók milyen mértékben képesek megváltoztatni a kimeneti értéket (HAND ÉS MANNILA, SMYTH 2001).

A modellek felépítése során azonban fontos kritérium az, hogy legalább egy nagyságrenddel több adat álljon rendelkezésre, mint ahány (magyarázó)változó szerepel. Ez a kritérium az elektronikusan nagy mennyiségben rendelkezésre álló adatoknak hála már nem jelent problémát. A másik probléma a túl sok rendelkezésre álló változó miatt szokott fennállni: bár a regresszió számítás nem rendelkezik nagy számítási igényvel, mégis célszerű a változók számának csökkentése. Itt kell megemlíteni a minőségi ismérvek kezelésének lehetőségét (de megnöveli a változók számát).

#### **4.3.2 Prediktív változók kiválasztása**

A regresszió számítás célja az előrejelzés. Amennyibe a már meglévő elemek változói közt felírható valamilyen kapcsolat, ami függvénnyel jellemezhető, lehetségessé válik az elkövetkező értékék becslése (*KHOSROWPOUR 1997*).

Mivel vállalati rendszerekben előfordul, hogy ezres nagyságrendben vannak változók, és adatbányász szemszögből is több százas lehet a nagyságrend, szinte minden esetben csökkenteni kell a prediktív változók számát (*LEIGH 2004*). Erre több megoldás lehetséges. Az első, ha van szakterületi specialista, akkor ő a tudása alapján képes lesz kiszűrni a felesleges (predikcióra alkalmatlan) változók jelentős részét, de előfordulhat, hogy még így is túl sok változó marad. A második ajánlható lehetőség a keresztkorreláció alkalmazása: a két adatsor közti összefüggés megtalálható – ha ilyen nincs (vagy nagyon gyenge), a változó kizárható az elemzésből. Ahhoz persze, hogy minden szükségtelen változó kizárható legyen, minden elemre el kell végezni a keresztkorrelációt, ami nagy adatmennyiség esetén óriási számítási igényt jelent – fel kell rá készülni.

#### **4.3.3 Validáció**

Mivel a cél a predikció, és az azon alapuló döntések meghozatala, ezért fontos kritérium az eredmények megfelelőségének vizsgálata. A validáció itt, ebben az értelemben azt jelent, hogy meg kell vizsgálni: a megfelelő eredményeket adta-e az eljárás (*SZIRAY, BENYÓ, GÓTH, MAJZIK, PATARICZA 2000*).

A bevett gyakorlat szerint az adathalmazt fel kell darabolni, ki kell alakítani egy olyan validáló-halmazt, melyben tárolt adatok nem vesznek részt a „tanításban”, azaz nem kerülnek számításba a modell kialakítása során. Amikor a modell elkészült, előrejelzést készítenek az alapján, majd az addig érintetlen adatokat újakként adják meg, és

összehasonlítják a kapott eredményeket. Azt persze, hogy mekkora fedés esetén minősül jónak a modell, egyéni döntés szabja meg.

A felosztás speciális esete a *bootstrap* tesztelési technika, melynek során annyi elem kerül kiválasztásra véletlenszerűen, ismétlődésekkel, ahány megfigyelés adott. Matematikailag bizonyítható, hogy így ~63.2% lesz a tanítóhalmaz<sup>56</sup>, a fentmaradó 36.8% lesz a teszhalmaz.

Fejlettebb technika a kereszt-validálás (vagy n-szeres kereszt-validálás), melynek lényege, hogy felosztásra kerül az adathalmaz, majd első körben egy véletlenszerűen kiválasztott halmaz lesz a teszhalmaz, az összes többi a tanító halmaz. A pontosság számítása után a következő halmaz lesz a teszhalmaz, a többi tanítóhalmaz. Annyiszor kell elvégezni a tesztet, ahány halmazra történt a felosztás. A végső pontosságot pedig a pontosságok számtani átlaga fogja adni. Bevett szokás szerint 10 részre szokták osztani az adathalmazt. Amennyiben a részhalmazokban egyenletesen van elosztva (szándékosan) minden osztályba tartozó elem, akkor rétegezett kereszt-validálásról van szó. Másik speciális eset a leave-one-out validálási technika, mely esetén annyi részhalmazt kell használni, ahány elem / megfigyelés van. Az előny a pontosság, a hátrány az extrém mértékben megnövekedett futási idő.

## 4.4 Idősorok

Minden gazdasági szervezet számára nagy előnyt nyújthat a jövő ismerete. A termékek iránti kereslet változása vagy a foglalkoztatottság, munkanélküliség, nyersanyag rendelkezésre állás átváltozásai ismeretében hatékonyabb és gazdaságosabb stratégia alakítható ki. Ehhez azonban az időbeli változások vizsgálata szükséges (*HAMILTON 1994*).

„Az *idősorok elemzése segít megérteni és megmagyarázni a múltat és előre jelezni, tervezni a jövőt.*” (*KERÉKGYÁRTÓ, MUNDRUCZÓ ÉS SUGÁR 2001*) Az idősorok elemzése történhet kvantitatív és kvalitatív szempontok szerint (*WEI 1990*). A kvantitatív elemzéshez számszerűsített, múltba visszamenő adatok szükségesek (*BOX, JENKINS ÉS REINSEL 2008*), melyeket elemezve (regresszió segítségével akár más hatásokat is bevonva) előre jelezhető a változók értékének jövőbeni alakulása. Persze nem minden esetben áll rendelkezésre nagy mennyiségű számszerűsített adat (idősor formájában), ekkor ugyancsak szakterületi specialisták vagy szakértők bevonásával készíthető kvalitatív elemzés.

---

<sup>56</sup> Ezért is szokta az angol nyelvű szakirodalom 0.632 *bootstrap*-nek nevezni.

#### 4.4.1 Indexszámok

Az idősorok elemzésének egyszerűbb módszerei az indexszámok, viszonyszámok és átlagok számítása. Ide tartozik a grafikus megjelenítés (általában pont vagy vonal- illetve oszlopdiagram).

A viszonyszámok [22. ábra] mindig két egymással összefüggő statisztikai érték hányadosai. Három fő típusa ismert:



*forrás: Kerékgyártó, Mundruczó és Sugár 2001 alapján saját munka*

22. ábra: Viszonyszámok

- A megosztási viszonyszám egy részadat (időszakos vagy kategória) az egészhez való viszonyát fejezi ki,
- Az intenzitási viszonyszám két egymással összefüggő adat hányadosa (pl.: fogyasztási értékek, színvonalmutatók),
- A dinamikus viszonyszámok összehasonlított időszakok hányadosai (ha több adatra kell alkalmazni, akkor a bázis- és láncviszonyszámok használata kell).

Az indexszámok olyan összetartozó mennyiségek, melyeket közvetlenül összesíteni ugyan nem lehet, de azok térbeli vagy időbeli összehasonlítása értelmes és magyarázó.

Az átlagok számítása itt kétféle céllal történhet:

- Első esetben egy átlagos szinttel jellemezhető az egész idősor (mozgó sokaság esetén számtani átlag, álló sokaság esetén kronologikus átlag használata a bevett),
- Második esetben az idősoron belül, az egyes időszakok átlagértékeivel jellemezhető a változás menete (átlagos abszolút változás a szomszédos időszakok különbségeinek számtani átlaga, illetve átlagos relatív változás az egymást követő

adatok hányadosainak mértani átlaga) (*KERÉKGYÁRTÓ, MUNDRUCZÓ ÉS SUGÁR 2001*).

Megjegyzendő azonban, hogy ezek az egyszerűbb elemzési technikák nem alkalmasak arra, hogy az idősorok jellegzetességeit megragadják. Előnyük a gyors (folyamatos) számíthatóság. Átlagok számítása akkor javallott inkább, ha van ugyan ingadozás az értékek között, de azok szóródása egy konstans körül történik.

A grafikus megjelenítés elsősorban a vizualizációt segíti, a számításokban nem szerepel – viszont a dashboard-ok (műszerfalak) kialakításánál és használatánál fontos szerepet tölt be, hiszen ott a cél a gyors, áttekinthető jellegű grafikák (esetleg interaktív) megjelenítése.

#### **4.4.2 Idősor modellek**

Az idősor elemzésben alapvetően kétfajta elemzési technikát lehet megkülönböztetni (*KERÉKGYÁRTÓ, MUNDRUCZÓ ÉS SUGÁR 2001*):

- Determinisztikus idősor elemzés (Az elemzés alapfeltevése szerint a vizsgált értékek várhatóan előre jelezhető értékeket fognak felvenni. Itt az elemzés célja ennek a pályának a meghatározása. Ehhez szét kell választani a komponenseket, és hasznos, ha hosszú távra elérhető adatok állnak rendelkezésre. Az eljárás során a zajt ki kell szűrni), illetve a
- Sztochasztikus idősor elemzés (ahol a cél mindössze a rövidtávú törvényszerűségek megismerése, viszont itt felhasználásra kerül a zaj is).

Azonban a modelleket technikájuk alapján is lehet csoportosítani (*KERÉKGYÁRTÓ, MUNDRUCZÓ ÉS SUGÁR 2001*):

- Dekompozíciós idősor modellek (megkülönböztethető additív illetve multiplikatív modell, aszerint, hogy a hosszú távú irányzathoz hozzáadódnak vagy azt sokszorozzák a rövidebb távú, szezonális ingadozások és hosszabb távú ciklusok),
- Simító eljárások (melyek a sztochasztikus és determinisztikus elemzések keverékei, viszont folyamatos felülvizsgálatra szorulnak),
- ARMA modellek (melyek kifejezetten sztochasztikus eljárások, kiemelt szerepet kapnak benne a véletlenek, viszont iterativitásuk miatt ugyancsak jelentős számítási igényeket támasztanak).

### 4.4.3 Komponensek elemzése

Az idősor felfogható úgy is, mint egy érték (adat) múltbéli hatások összegének változásainak sora (*KERÉKGYÁRTÓ, MUNDRUCZÓ ÉS SUGÁR 2001*). Minden idősor szétbontható az alábbi komponensekre (ezek összessége adja az értékeket):

- A trend (a leghosszabb távú alakulás – néhány évet is jelenthet – az idősor értékeinek legfőbb mozgási iránya), melynek detektálása azért fontos, mert így valamelyest simább idősor az eredmény, azaz lehetőség van a többi komponens felderítésére is. A két használt számítási módszer: a mozgó átlagok módszere (ekkor az idősor dinamikus átlaga kerül számításra, amihez előny a magas tagszám megléte), illetve az analitikus trendszámítás (ekkor a megbúvó trendet valamilyen függvénnyel lehet jellemezni – a függvény lehet lineáris, exponenciális, másod illetve p-ed fokú polinom, hiperbola vagy logisztikus).
- A ciklus (a trend fölötti vagy alatti tortósabb mozgás, általában periodikus, de nehezen felfedhető, szabálytalan, mivel hosszú távú adatokra van hozzá szükség). A ciklikusság vizsgálata a trendszámítás alapján is két féleképpen lehetséges: az egyik a mozgóátlagokból kiinduló, a másik az analitikus trend illesztéséből.
- A szezonális ingadozás (szabályos – tehát jól előre jelezhető – rendszeres időközönként jelentkező változás, melynek iránya mindig azonos, a periódus legfeljebb egy év lehet), arra ad választ, hogy a periódus egyes szakaszaiban milyen mértékben és irányban téríti el az értékeket az alapirányzattól. Megkülönböztetett állandó (ugyancsak lehet additív vagy multiplikatív) és változó szezonális ingadozás.
- Véletlen ingadozás (szabálytalan, előre nem jelezhető hatások, melyek jellemzője, hogy annál jelentősebbek, minél rövidebb időre szól a megfigyelés).

### 4.4.4 Exponenciális simítás

Az előrejelzés egyik egyszerűbb és gyakran használt technikája az exponenciális simítás. A technika használatának előfeltétele, hogy nem remélhető az adatokban (idősorban) sem trend, sem szezonális ingadozás.

A számítás során a következő periódus értékét az aktuális periódus mért értéke súlyozva és a korábban előre jelzett érték súlyozott szorzata adja, így egy adott időponthoz tartozó értékben benne foglaltatnak a korábbi értékek is, időben visszafelé haladva egyre kisebb súllyal. A súly értékét vállalati körben 5% és 30% közé kell választani, de a kérdésben a

téma szakértői is megosztottak (*SIPOS 1985, BACSKAY 1984, MAKRIDAKIS – WHEELWRIGHT 1978, GILCHRIST 1978*).

A legjobb eredmények eléréséhez sokszori számításra van szükség, a súlyozást folyamatosan finomítva – ettől függetlenül azonban a számítási igény nem jelentős (*BÁLINT 2002*).

#### **4.4.5 Autoregresszív modellek**

Az autoregresszív modellek olyan folyamatokat próbálnak szimulálni, melyekben nagy szerepet kapnak a véletlen behatások. Tipikusan ilyen behatások a társadalmi és természeti jelenségek. Alkalmazásuk akkor javallott, ha rövidtávon összefüggő folyamatok modellezése a cél (nem alkalmasak hosszú távú folyamatok leírására).

Megvalósítást tekintve többfajta autoregresszív modell (AR) létezik (*LAW ÉS KELTON 2000*), melyek más-más területen használhatóak (főbb különbség a korrelációjukban):

- p-ed rendű autoregresszív folyamat (AR),
- Diszkrét autoregresszív folyamat (DAR – mely alkalmas stacionárius v.v sorozat előállítására),
- Vektor autoregresszív folyamat, VAR,
- Mozgóátlagos autoregresszív folyamatok (ARMA),
- Integrált mozgóátlagos autoregresszív folyamat (ARIMA / ARFIMA – főként nem stacionárius folyamatokhoz),
- Valamint univerzálisnak szánt ARTA (AutoRegressive To Anything).

Általánosan elmondható, hogy az autoregresszív modellek nagy számítási kapacitát igényelnek, viszont

#### **4.4.6 Prediktív modellek**

A prediktív modellezés egy létező vagy hipotetikus rendszer jövőbeli működését hivatott előre jelezni.

A prediktív modellek ugyancsak „tanulási” módon készülnek, a tanító (adat)halmazon, majd ugyanúgy validálni kell azokat. a validálás során mutatott eredményekből lehet következtetni a kialakított modell hatékonyságára, precizitására.

Mára ez az irányzat vált az adatbányászat egyik vezető funkciójává, gyakorlatilag az üzleti intelligencia rendszerek minden moduljában megtalálható bizonyos mértékben: általánosan használt marketing, telefonközpont, vásárlói megelégedettség mérése során, de terjedően van a web-bányászat és szövegbányászat ágakban is (SARMA 2007).

## 4.5 Minősítés

A minősítés vagy klasszifikáció<sup>57</sup> egy diszkrét kategória-skálán mért változó előrejelzése. Az eredmény előállításához számos magyarázó (bemeneti) változó kombinációját kell vizsgálni, majd azok értékei alapján besorolni a kimenetet<sup>58</sup> (KERÉKGYÁRTÓ, MUNDRUCZÓ ÉS SUGÁR 2001). A minősítés is használható előrejelzésre.

Minden minősítési probléma a következőképpen fejthető ki: megadható egy tanítóhalmaz, melyben össze vannak rendelve a bemenetkehez tartozó elvárt minősítések. A feladat minden esetben egy olyan szabály létrehozása, mely minden beérkező elemet képes minősíteni. Célszerű a minősítést becslési feladatként felfogni és a feltételes valószínűséget becsülni arra nézve, hogy mekkora a valószínűsége annak, hogy a beérkező elem adott minősítési osztályba tartozzon.

A másik felmerülő probléma mindig annak eldöntése, hogy bináris vagy multiklassz minősítést kell alkalmazni. Bináris minősítés esetén csak két kategória szerepel, míg multiklassz esetén több. Mivel azonban a bináris minősítésre nagyon sokfajta módszer lett kifejlesztve, célszerű minden több kategóriát igénylő minősítési műveletet visszavezetni több bináris minősítésre.

Minden osztályozási feladat alapvetően három lépésre vezethető vissza: modellgenerálás (először meg kell határozni az osztályokat, amikbe szeretnénk a megfigyeléseket sorolni; emellett fel kell darabolni a rendelkezésre álló adathalmazt tanító mintákra és ellenőrző mintákra; modell kiválasztása és tanítás, majd a szabályok kinyerése), ellenőrzés (a megkapott szabályrendszer ellenőrzése az ellenőrző mintával, majd annak vizsgálata, hogy annak elemei a megfelelő osztályokba lettek-e sorolva; ez iteratív folyamat, folyamatosan finomodva egy behangolt modellt eredményezve) és felhasználás (az új beérkező elemek osztályozása).

---

<sup>57</sup> A minősítés, a klasszifikáció és az osztályozás mind a szakirodalomban megjelenő szinonímák erre a műveletre.

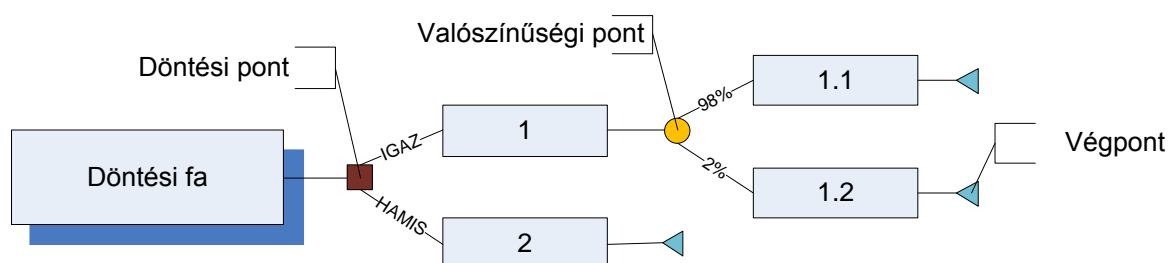
<sup>58</sup> Gyakran klaszterezéssel keverik: a fő különbség az, hogy osztályozás esetén ismertek a csoportok, míg klaszterezés során a csoportok kialakítása is cél a besorolás mellett.



A minősítéshez nagyon sokféle modell áll rendelkezésre, ám a legelterjedtebbek a döntési fák, a Bayes-i módszerek, a tartó vektor gépek<sup>59</sup> és a neurális hálók. Azt persze, hogy melyik algoritmus lesz a célnak legmegfelelőbb, sok tényezőn múlik: szakterület, bemeneti adatok száma, minősége és mivolta.

#### 4.5.1 Döntési fák

A döntési fák [23. ábra] olyan döntéstámogató eszközök, melyek gráfok formájában reprezentálják a döntési lehetőségeket, azok következményeit, esetleg beleértve valószínűséget, költséget (FREIEDMAN 1996). Gyakran használt eszköz az operációkutatásban, valamint a döntések elemzése során, hiszen segít a stratégia kialakításában, valamint a feltételes valószínűségek meghatározásánál.



forrás: vonatkozó Wikipédia cikk példája alapján saját munka

23. ábra: Döntési fa felépítése

A fák három építőeleme a következő: döntési csomópontok, valószínűségi csomópontok, illetve végpontok (ROKACH ÉS MAIMON 2008). A gyöktől indulva minden döntési pont egy kérdést reprezentál, és a kiinduló ágak a döntésre adott lehetséges értékeket (attribútum értékek). A fa levelei lesznek a ténylegesen javasolható döntések.

Használatának előnye, hogy:

- könnyű megérteni és értelmezni,
- könnyen kinyerhetőek belőle ha-akkor típusú osztályozási szabályok (minden döntési pont különböző értékei közti ÉS kapcsolat segítségével),
- ha szakértők készítik, akkor az avatatlanok számára is gyors belátás nyílna egy problémába,

<sup>59</sup> A szakirodalomban sok helyen használják a szupport vektor gép elnevezést is (angolul Support Vector Machine – SVM)

- előrejelzés esetén jól láthatóvá teszi a végső döntéshez szükséges döntéssorozatot
- fehér-dobozos modellek alkalmazására is használható, mivel az eredmények könnyen ellenőrizhetők,
- jól kombinálható más döntési technikákkal.

Gyakorlatban persze az egyszerűség kedvéért bináris döntési fák kerülnek használatra (minden csomópont két gyermeket tartalmaz), ám bármilyen nem bináris döntési fa átalakítható bináris fává – a döntési pontok számának növelésével.

Döntési fák két lépcsőben készíthetők, az első lépés a fa elkészítése, a második lépés a már kész fa tisztítása.

A döntési fa megépítése elején még minden elem a fa gyökerénél található. Minden döntési pont beiktatása a fába felfogható egy-egy particionáli lépésnek. Cél olyan döntési pont beiktatása, mely a következő szinten növeli az elemek homogenitását. A döntési pontok beiktatása rekurzív folyamat, és akkor van vége, ha már nem lehet a fentmaradó elemeket particionálni, vagy nincs több változó, vagy elfogytak az adatpontok, vagy egyszerűen elérte a fa az előre meghatározott méretet.

A fa tisztításának célja a fa egyszerűsítése azáltal, hogy törli azokat az ágakat, melyek a valószínű túltanulás<sup>60</sup> miatt keletkeztek. Cél egy egyszerűbb, de ugyanúgy értékes információt tartalmazó fa kialakítása. A tisztításnak kétféle megközelítése (*BRESHOW ÉS AHA 1997*) van: az előmetszés (a fa felépítésének korai megállítása, mely történhet automatikusan, szignifikanciázt alapján, vagy manuálisan – ekkor a csomópontdöntési pontból levéllé alakul, címkéje a legvalószínűbb érték lesz ajánlhatóan) és az utómetszés (a fa teljes felépítése után manuálisan egyes részfákat levelekké alakítunk). A faépítő algoritmusok többsége<sup>61</sup> az utómetszést preferálja.

#### 4.5.2 Bayes-i módszerek

A Bayes-i módszerek esetén két témát kell ismertetni: a naiv Bayes minősítő eljárást, valamint a Bayes hálózatokat.

---

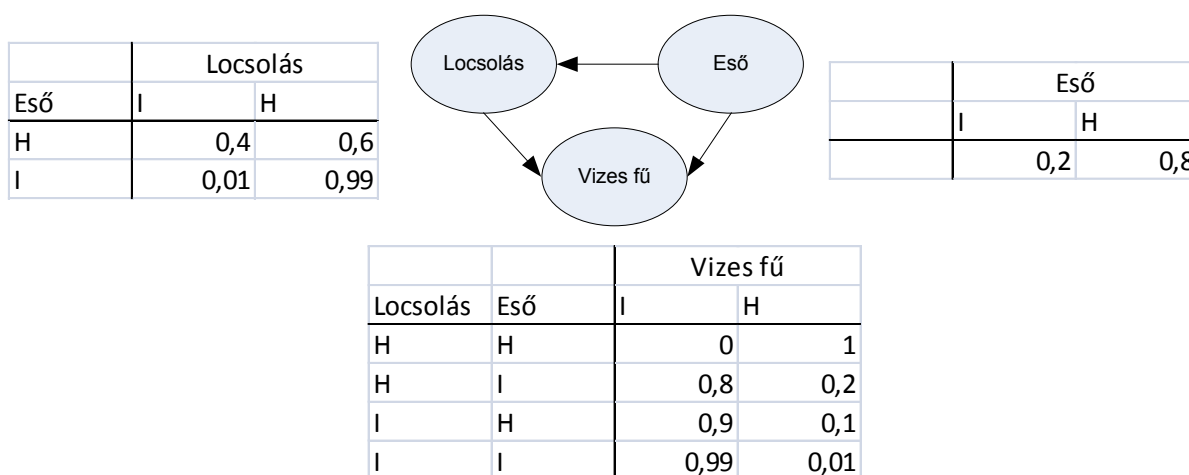
<sup>60</sup> A túltanulás / túltanítás (*overfitting*) a modell túlságos illeszkedését jelenti a tanító halmazon. A hátránya, hogy így megtanulja a tanító halmazban lévő zajt is, mely kizárólag ott érvényes, és ez később rossz teljesítményt okoz.

<sup>61</sup> Ilyen algoritmusok például a CART vagy az ID3.

A Bayes-i minősítő egy olyan valószínűségi minősítő, mely Bayes-tételt<sup>62</sup> alkalmazza, erős függetlenségi feltevással (*POURRET, NAIM ÉS MARCOT 2008*). Segítségével megbecsülhető egy adott minta egy osztályba tartozásának valószínűségét. Gyakorlatból megállapítható, hogy döntési fákhhoz hasonló pontosság érhető el használatával.

A valószínűségi modell precizitásától függően a naiv Bayes minősítők felügyelt tanulás mellett nagyon pontos szabályok kialakítására képesek. Használatának nagy előnye, hogy viszonylag kis tanítóhalmazzal is működik.

A Bayes-i minősítő használata során meg kell fogalmazni egy sor felétételt (mely attribútumokat kell figyelni, milyen értékekkel), majd a feltételhez tartozó kimeneti osztályt. A működés során a bemenet egy elem (attribútumaival) lesz, a kimenet pedig egy valószínűségi érték lesz, mely megmutatja, hogy az adott elem mekkora valószínűséggel esik az adott osztályba.



forrás: vonatkozó Wikipédia cikk példája alapján újrarajzolva

24. ábra: Bayes-i hálózat

A Bayes-i hálózat [24. ábra] olyan valószínűségi grafikai modell, mely több véletlen változót és azok feltételes valószínűségét képes megjeleníteni irányított ciklusmentes gráfon (DAG – Directed Acyclic Graph). A csomópontok minden esetben a véletlen változókat jelölik, ezek lehetnek mennyiségek, megbúvó változók, ismeretlen paraméterek vagy hipotézisek. Az élek feltételes függést jelölnek. A nem összekötött csomópontok egymástól független változókat jelölnek. Minden csomópontozhoz hozzárendelésre kerül egy

<sup>62</sup> A Bayes-tétel egy feltételes valószínűség és annak fordítottja közt állít fel kapcsolatot; azt adja meg, hogy ha adott egy hipotézis és egy esemény, akkor az esemény megfigyelése miként befolyásolja a hipotézis helyességébe vetett hitet – azaz az új információk fényében mindig újraértékeli az eseményrendszer egyes tagjainak bekövetkezési valószínűségét.

valószínűségi függvény, melynek bemenete(i) a csomópont szülőjétől származnak, és a kapott eredmény a csomópont által szimbolizált változó valószínűsége (*JENSEN ÉS NIELSEN 2007*).

Bayes-i hálózatkora azért volt szükség, mivel a Bayes-i minősítők alapfeltétele (attribútumok függetlensége) a valóságban gyakran nem teljesül. A háló szerkezete mellett meg kell adni a feltételes valószínűségi táblázatokat is, minden csomópontra.

A Bayes-i hálózatokra több sikeres, megvalósított tanuló algoritmus létezik. Ha a hálózat képes jelsorozat feldolgozására, akkor dinamikus Bayes-i hálózatnak nevezik. A befolyási diagramok (influence diagram) generalizált Bayes-i hálózatok (*MITTAL ÉS KASSIM 2007*).

### 4.5.3 A k-NN algoritmus

A k-legközelebbi szomszéd technika<sup>63</sup> megvalósítása. A technika feltételezi, hogy létezik egy mintahalmaz, melyen az osztálybasorolás már megtörtént. Az új, beérkező elemeket abba a csoportba / osztályba sorolja, melyhez a legkisebb a Manhattan távolsága. Térbeli reprezentációja úgy képzelhető el, hogy a tanító halmaz elemei egy n-dimenziós térben helyezkednek el, ahol n az attribútumok száma.

Mivel általában több ismerv (attribútum) alapján történik az elemzés, ezért normalizálni kell azok értékeit. Ez persze felveti azt a problémát, hogy így minden attribútum egyformán fontossá válik – mely kijelentés valószínűleg nem tükrözi a valóságot.

Az elnevezésben a k betű azt jelenti, hogy hány szomszéd közelségét veszi figyelembe az algoritmus. Alapbeállításként sok szoftverben egyetlen, legközelebbi szomszéd figyelembe vétele jelenik meg. Ekkor egyszerű a grafikus ábrázolás, mivel az aktuális állapot jól reprezentálható egy egymással illeszkedő poligonokkal (Voronoi cellák, illetve Voronoi ábrázolás)-

Problémát szoktak jelenteni a hiányzó értékek, erre gyakorlatban a legrosszabb eset feltételezése a megoldás, ez persze sok esetben hibás kimenetet eredményez.

Abban az esetben, ha nem egyetlen szomszéd figyelése a beállítás, további probléma szokott lenni az, hogy az algoritmus az összes szomszédot ugyanolyan fontosnak tekinti. Erre megoldás az, ha a szomszédok a távolságukkal ellentétesen arányos súlyokat kapnak –

---

<sup>63</sup> Az eredeti elnevezése k-Nearest Neighbor, innen a rövidítés.

ez növeli a későbbi hatékonyságot is – ekkor az elnevezés távolság-súlyozott k-NN osztályozás.

A legnagyobb probléma ugyanakkor az, hogy a teljes mintát a memóriában kell tartani ahhoz, hogy hatékonyan futtatható legyen; a nagy memóriaigény miatt kevés elemző eszközben van implementálva.

#### **4.5.4 Logisztikus regresszió**

A logisztikus regresszió a minősítés egy klasszikus módszere, alkalmazása széles körben elterjedt. Nevezik logit-modellnek is.

A minősítés során akkor lehet logisztikus regresszióról beszélni, ha az eredmény jellegű változó bináris, azaz csak két értéket vehet fel (dichotom logisztikus regresszió) (*KLEINBAUM ÉS KLEIN 2010*).

Az eljárás lényege, hogy a bemeneti adatok alapján egy olyan logisztikus görbét alakít ki, mely szétválasztja a kimeneti változók értékeit – a görbe alatti rész lesz az egyik kimeneti érték, a görbe feletti rész lesz a másik.

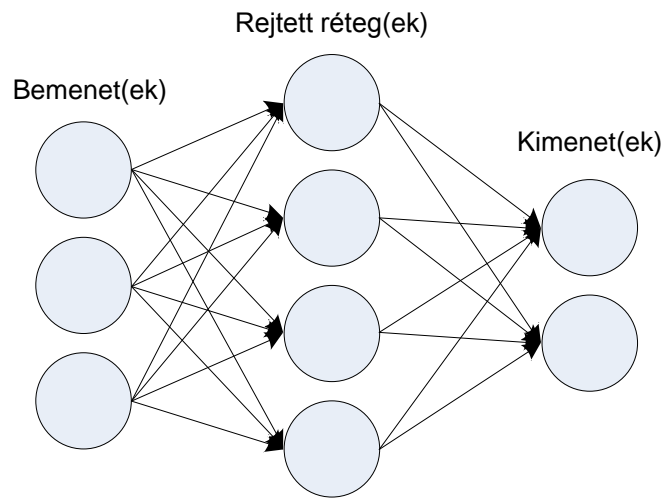
A folyamat eredményessége azonban erősen összefügg a bemeneti adatok mennyiségével (a minta mérete), a multikollinearitással és a kívülálló értékekkel. A modell (működését tekintve) teljesen ekvivalens az egyszerű perceptronnal (mely egy egyszerűbb neurális hálózat – lineáris minősítésre alkalmas) (*HOSMER ÉS LEMESHOW 2000*).

#### **4.5.5 Neurális hálózatok**

A neurális hálózatok [25. ábra] (a pontos megnevezés mesterséges neurális hálózatok) a biológiai neurális hálózatok felépítését és működését utánzó matematikai vagy számítási modellek. Egy neurális hálózat mesterséges neuronok összekapcsolt hálózatából áll, így dolgoz fel információkat. Ezek a hálózatok általában adaptív módon működnek, ennek köszönhetően a belső hierarchia a tanulási fázisban érkező adatok alapján alakul ki (*GOURNEY 1997*).

Minden neurális hálózat több rétegből áll. A legegyszerűbb felépítés három rétegű. Az első rétegben helyet foglaló input neuronok az összeköttetések révén az adataikat a második rétegbe továbbítják, majd onnan azok a harmadik rétegbe, a kimeneti neuronokhoz. A kapcsolatok (szinapszisok) súlyokat tartalmaznak, így az adatok átmenet közben is

módosulhatnak (*HASSOUN 1995*). A hálózat felépítését tekintve lehet bonyolultabb, tartalmazhat több neuront, több réteget, akár több rétegben be- és kimeneti neuront is.



*forrás: saját munka*

25. ábra: Neurális hálózat

A hálózatoknak két nézete képzelhető el: az egyik a funkcionális nézet, amit főként optimalizálás során használnak, a másik a valószínűségi nézet, ami elsősorban a grafikai modellezés során használt (*GALUSHKIN 2007*). Az egyszerűbb hálózatok szigorúan nem ciklikusak és irányítottak, de bonyolultabb esetben készíthető ciklikus, vagy akár rekurzív hálózat is.

A neurális hálózatok alkalmazási sikerességének kulcsa tanuló képességük (*MICHIE ÉS SPIEGELHALTER 1994*). A bennük definiált költségfüggvényt úgy állítják be, hogy semmi más megoldás ne legyen elfogadható, csak az optimális, mivel annak lesz legkevesebb a költsége. A költségfüggvény definiálása tehát az egyik kritikus pont. Így aztán a tanuló algoritmus az összes lehetséges megoldást végigpásztázza, hogy meglegelje a lehető legkisebb költséggel járó (*ANDERSON 1995*).

Az algoritmus három egyszerű tanulási paradigmával jellemezhető (*DAVALO ÉS NAIM 1991*):

- Felügyelt tanulás (ahol adatpárok adottak és a cél a bemenet  $\rightarrow$  kimenet függvény kialakítása, így használják a minta felismerés és regresszió során),

- Felügyelet nélküli tanulás (ahol a modell egy része adott, valamint a költségfüggvény, és a változók értékeit keresi, így használható eloszlások vizsgálatára, klaszterezésre, valamint adatok szűrésére), illetve
- Megerősítő tanulás (amikor a beérkező inputokra adott válaszokat figyeli az algoritmus, próbálva szabályrendszert felállítani – vezérlési feladatok esetén használt).

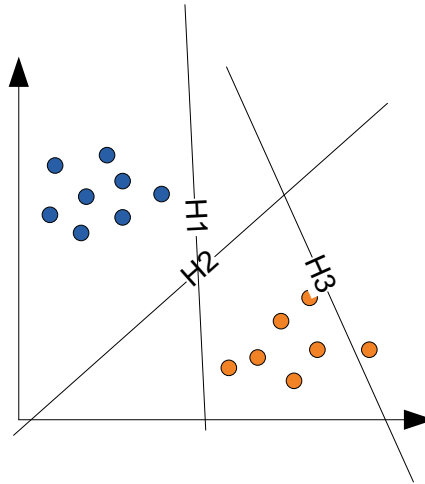
A neurális hálózatok alkalmazásának legnagyobb szerepe abban rejlik, hogy képesek függvények előállítására a már meglévő megfigyelések alapján (*TAKEFUJI 1992*). Ebben sok esetben a genetikus algoritmusokra való támaszkodás segít – használhatóak a háló súlyainak meghatározásánál, a neuronok közötti kapcsolat kialakításánál vagy megszüntetésénél (*ÁLMOS – GYŐRI – HORVÁTH – VÁRKONYINÉ 2002*).

A döntéstámogató felhasználásuk miatt nem lineáris statisztikai modellező eszköznek minősülnek, használatuk két esetben javasolt: ha a bemeneti és kimeneti változók közti összefüggés rendkívül bonyolult, vagy ha mintafelismerésre van szükség (*SKAPURA 1996 KARAYIANNIS ÉS VENETSANOPOULOS 1993*).

#### **4.5.6 Tartó Vektor gépek**

A tartó vektor gép, valamint a szupport vektor gép (SVM) elnevezés egyaránt elterjedt – mindkettő ugyanarra a Vapnik által 1979–1992 között kidolgozott felügyelt tanuló eljárásra vonatkozik, melyet főként statisztikában, minősítésre és regresszió analízisre használnak. Mára számos finomítás mellett új tudományággá nőtt (*WANG 2005*).

A módszer lényege, hogy a tanulási fázisban az algoritmusnak a bemeneti értékek mellett meg kell adni, hogy azok melyik kategóriába esnek (*JOACHIMS 2002*). A tanulás közben kialakuló minősítő algoritmus grafikus alapokon dolgozik, a pontdiagramon elhelyezkedő bementeket igyekszik a lehető legjobb egyenessel (hipersíkkal vagy hipersíkokkal) szétválasztani [26. ábra], oly módon, hogy az új bemenet, melynek kategóriája eldöntésre vár az egyenes alatt vagy felett elhelyezkedve meghatározza a kategóriát. További cél olyan egyenes definiálása (*SUYKENS, GESTEL ÉS BRABANTEL 2002*), mely a lehető legtávolabb helyezkedik el az adatokat reprezentáló pontoktól.



forrás: vonatkozó Wikipédia cikk alapján saját rajz

26. ábra: Hipersíkok egy tartó vektor gépben

Megkülönböztetett lineáris, illetve nem lineáris SVM, valamint több bináris (két kategóriát megkülönböztetni képes) gép kombinálásával multiklassz vektor gépek is létrehozhatóak.

## 4.6 Asszociáció

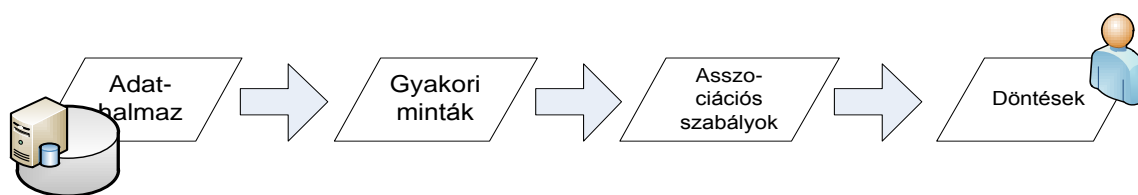
Az asszociáció a korrelációval ellentétben nem mennyiségi összefüggéseket, hanem minőségi és / vagy területi ismérvek közötti kapcsolatot elemez. Az üzleti intelligenciában a minőségi ismérvek, és kapcsolatok tanulmányozása során kap kiemelt figyelmet – tipikusan az asszociáció van ott hasznosítva, ahol valamilyen termékkapcsolás, terméksorrend kialakítás, kereszteladás vagy áruajánlat kerül bevezetésre (HAN ÉS KAMBER 2001).

Az asszociációs szabályrendszer létrehozásának célja minden esetben először a gyakori minták<sup>64</sup> felderítése. Gyakori minták az adathalmazban [27. ábra] sokszor előforduló kapcsolatokat jelölik (KOH ÉS ROUNTREE 2009).

A minták alapján lehet létrehozni asszociációs szabályokat, amiket megvizsgálva

<sup>64</sup> A gyakori mintákat a szakirodalomban szokás gyakori elemhalmazként ismertetni.





*forrás: saját munka*

27. ábra: Asszociációs szabályok helye a döntéshozatalban

Az itt felhasználható algoritmusok (keresési technikák) alapvetően négy csoportra oszthatóak, vagyis megkülönböztethetőek

- keresés szerint szintenként haladó, ám gyakoriság számítás szerint vízszintesen növelő;
- keresés szerint szintenként haladó, ám gyakoriság számítás szerint függőlegesen növelő;
- keresés szerint mélységben haladó, de gyakoriság számítás szerint vízszintesen növelő, valamint
- keresés szerint mélységben haladó, de gyakoriság számítás szerint függőlegesen növelő algoritmusok.

Gyakori elemhalmazok kinyerése esetén felmerülő probléma, hogy olyan elemhalmazokat kapunk, melyek nem járulnak hozzá új információ kinyeréséhez, redundancia vagy tartalmazás miatt. Így aztán megkülönböztethetőek maximális vagy nem bővíthető gyakori elemhalmazok (olyan halmazok, melyekhez elemeket adva már nem lesznek gyakoriak) illetve zárt gyakori elemhalmazok (olyan halmazok, melyekhez elemet adva változik a gyakoriság értéke).

A nem bővíthető elemhalmazok általában sok elemet tartalmazó halmazok – értékesítésnél hasznos a meglétük, de nem feltétlenül tartalmazzák a rész-elemhalmazok gyakorisági értékeit, ilyen adatok kinyeréséhez ismételt futtatás szükséges.

#### 4.6.1 Szintenként haladó, vízszintesen növelő technikák

Az APRIORI algoritmust 1993-ban hozták létre, leváltandó az addig használt AIS-t. Jól alkalmazható gyakori minták keresésére – így asszociációs szabályok kinyerésére is (YE 2003).

Az algoritmus iteratív, szélességi bejárást végez. Minden ciklusban újabb szintet vizsgál, egyre több elemmel. Minden gyakori minta jelölt lesz. A vizsgált mintát jelölteként fogja fel, hamis jelöltté nyilvánítva azokat, ha túl ritkák (elhanyagolt mintának, amik be sem kerülnek – akkor jó ha ilyen nincs). Az algoritmus először jelölteket állít (az előző szint gyakori halmazainak összeillesztésével, azzal a feltétellel, hogy a gyakorisági küszöb feletti elemek megegyeznek), majd azok gyakoriságát ellenőrzi (végiglépked minden adatbázis rekordon és a ténylegesen megjelenő jelöltek gyakoriságát növeli, majd a végeredményt visszaadja: a küszöb feletti értékeket produkáló halmazok). Az algoritmus annyiszor nézi át a teljes adatbázist, amekkora a halmazok maximális mérete.

Alkalmazásának több előnye van: rendkívül egyszerű, gyors és kicsi a memóriaigény (bizonyos megvalósításokban) – így nem támaszt nagy hardverkövetelményeket. Minden megvalósítása feltételezi az anti-monoton<sup>65</sup> elvet és ebből következően az Apriori-elvet<sup>66</sup> - ennek köszönhető a csökkentett memóriaigény és a gyors feldolgozás (illetve innen kapta az elnevezését is).

Az eredeti APRIORI algoritmus mellett hasonló elven működnek az AprioriTID, az Apriori-Hibrid, a DIC és a DHP algoritmusok.

Az AprioriTID<sup>67</sup> algoritmus ott képes jobb teljesítményt elérni elődjénél, ahol nagyon nagyszámú tranzakció és nagyon nagyszámú elem fordul elő (például egy hipermarketben, ahol rengeteg vásárlás történik és sok termékből állhatnak össze a vásárlói kosarak). Ugyanúgy két lépésben dolgozik, az első lépés változatlan, ám a második lépéshez a forrásadatbázist kivonatotja; abból több lépésben tranzakciókat kizár (kimaradnak azok a tranzakciók, melyek eleve csak egy elemet tartalmaznak, melyek eleve nem tekinthetők gyakorinak, mert kevés alkalommal fordulnak elő), így futása felgyorsul. A végeredményt a kapott különböző hosszúságú elemhalmazok uniója adja.

Az Apriori-Hibrid módszert az IBM vállalat a két előd különböző szituációkban elért teljesítmény-különbsége (kísérleteik bizonyították, hogy a hagyományos APRIORI algoritmus gyorsabb kevesebb elemet tartalmazó elemhalmazok esetén, míg egy bizonyos határ felett az AprioriTID gyorsabb) miatt alkotta meg. A kettő algoritmus közti váltás

---

<sup>65</sup> Az anti-monoton elv alapja az a kijelentés, mi szerint ha egy elemhalmaz nem gyakori, akkor azt bármilyen elemmel bővítve sem kapható gyakori elemhalmaz.

<sup>66</sup> Az apriori-elv az anti-monoton elv megfordítása: amennyiben található egy gyakori elemhalmaz, akkor annak minden részhalmaza is gyakorinak minősül.

<sup>67</sup> AprioriTID: Az Apriori elvet megvalósító, adatbázis kivonatólással működő változat; nevét az elsődleges kulcsok használatáról kapta (Transaction IDentifier - tranzakcióazonosító).

időpontját memória mérethez kötik; ha már a kivonatolás miatt elfér a memóriában a teljes adatbázis, célszerű AprioriTID-re váltani (az eredményt nem befolyásolja, csak a végrehajtás idejét).

A DIC<sup>68</sup> algoritmust azért alkották meg, hogy ne kelljen annyiszor végignézni az adatbázist – ez jobb végrehajtási időt eredményez. Működése annyiban tér el az APRIORI algoritmustól, hogy az adott körben nem kizárólag a körhöz tartozó elemhalmazokat vizsgálja, hanem bemeneti paraméterétől függően folyamatosan bekapcsolja a vizsgálatba az egyre nagyobb hosszúságúakat. Olyan esetekben hatékony, ahol a tranzakciók száma magas.

A DHP<sup>69</sup> eljárás megszületésének két oka van: a memóriaigény csökkentése és az a tény, hogy minden tranzakcióhalmazon a legtöbb gyakori elemhalmaz a kételeműek közül kerül ki, így a futás itt a leglassabb. Hash-fa készítésével jelentősen csökkenti a kételemű jelöltek halmazát, így memóriát spórol. Az előzőeknél hatékonyabb, ám helytelen paraméterezés eredményeként gyakran állapíthat meg olyan elemhalmazt, mely valójában nem az.

#### **4.6.2 Szintenként haladó, függőlegesen bővítő eljárás**

Szintenként haladó, ám függőleges bővítést valósít meg a partícionáló algoritmus. Működése során két alkalommal halad végig a teljes adatbázison: első körben az összes lehetséges gyakori elemhalmaz halmazát készíti el, majd második körben a kapott elemhalmazok gyakoriságát vizsgálja. Gyakorlatilag két fázissal szokták megvalósítani, az első fázis partíciókat képez (szétbontja az adatbázist), a második fázisban Apriori algoritmus segítségével gyakoriságot használ. Előnye a jó párhuzamosíthatóság (ebből adódóan gyorsan képes futni pl. elosztott rendszereken), ellenben hátránya, hogy bizonyos valójában gyakori minták kimaradhatnak a megállapított gyakori elemhalmazok közül.

#### **4.6.3 Mélységben haladó, függőlegesen bővítő eljárás**

Mélységben haladó, ám függőlegesen növelési technikával dolgozik az ECLAT<sup>70</sup> módszer. Az eljárás használatához szükség van az adatbázis vertikális sémájának<sup>71</sup> létrehozására,

---

<sup>68</sup> DIC: Dynamic Itemset Counting, azaz dinamikus elemhalmaz számlálás.

<sup>69</sup> DHP: Direct Hashing and Pruning, vagyis közvetlen hash-elés és tisztítás.

<sup>70</sup> ECLAT: Equivalence CLAss Transformation: egyezőség szerint / ekvivalencia alapú osztály transzformáció.

mely némi időt igényel. Az algoritmus rekurzív, minden körben célszerű rendezést végezni az állított jelölteken, így gyorsítható a futása (a másik teljesítmény-javító módszer a tranzakciólisták helyett azok különbségi értékeinek tárolása). Minden körben az előző kör eredményei alapján vizsgálatra kerül, minden két halmaz metszete. A legnagyobb továbbjutnak a következő körre.

#### 4.6.4 Mélységben haladó, vízszintesen növelő technika

Mélységben haladó, ám vízszintes növelési technikát valósít meg az FP-growth<sup>72</sup> algoritmus. Működése során nem szükséges minden elemre az előfordulások számítása, e helyett inkább fát épít, minden elem egy láncolt listát tartalmaz, mely az összes olyan tranzakciót tartalmazza, melyben előfordul.

#### 4.6.5 Szabályok

Az asszociációs szabály jelentése üzleti szférára értelmezve a következő: ha egy fogyasztói kosárban előfordulnak bizonyos elemek, akkor nagy valószínűséggel még egy adott másik elem is elő fog fordulni. Azt, hogy ez a valószínűség mekkora, a konfidencia adja meg – célszerű csak a valóban nagy valószínűségű szabályokra alapozni. A valószínűség mellett azonban a támogatottság (szupport) is fontos, mivel a szupport mérőszám magas értéke biztosítja, hogy a szabály valóban sok esetben előfordul. Így egy elemhalmaz akkor mindősül gyakorinak, ha legalább egy előre meghatározott százaléknál magasabb arányban fordul elő (*SOMAN, DIWAKAR ÉS AJAY 2006*). Az asszociációs szabályok persze nem kizárólag a vásárlói korás elemzés során használhatóak; bankügyletek elemzése, valamint orvosi mintakeresés során is általánosan használják őket.

Asszociációs szabályok között megkülönböztetünk kétértékű<sup>73</sup> és mennyiségi<sup>74</sup> szabályokat; ezen kívül lehet szó egy- vagy többszintű szabályokról. A mennyiségi asszociációs szabályok kialakítása még ma is kutatás tárgyát képezik – a terület szerencsére gyorsan fejlődik. A kihívás az, hogy a több értéket felvenni képes ismérvek asszociációban történő alkalmazására azokat célszerű szétbontani, annyi ismérvet képezve,

---

<sup>71</sup> Olyan esetekben hozható létre a vertikális séma, amikor egy tranzakcióazonosítóhoz vannak eredetileg rendelve elemek (pl.: vásárlás azonosítóhoz megvett termékek) – ez a horizontális séma. A vertikális séma létrehozásakor az elemekhez kerülnek hozzárendelésre a tranzakcióazonosítók (pl.: termékeket melyik vásárlói kosarakban vásárolták meg).

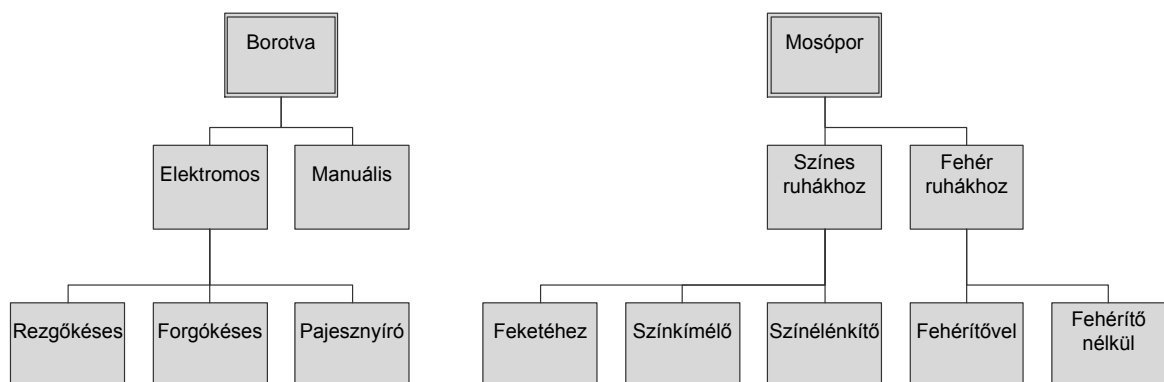
<sup>72</sup> FP-growth: Frequent Pattern growth, vagyis gyakori minta növelés.

<sup>73</sup> Kétértékű asszociációs szabály, vagy más néven boole-féle: csak két értéket vehet fel.

<sup>74</sup> A mennyiségi asszociációs szabályokat szokták még numerikus vagy kvantitatív szabályoknak is nevezni.

ahány lehetséges (vagy ha túl sok lenne, akkor lehet partíciókat képezni az ismérvből) érték van – jelentősen megnövelve egy adatbázisban a mezők számát – majd az így nyert kétértékű elemekre készíteni elemzést; két probléma jelentkezhet: az éles határok problémája, illetve az, hogy minden partíció minden eleme azonos súllyal esik számításba. Mindkét problémára megoldást nyújt a fuzzy intervallum elemek használata (mint ahogy nagyon sok más elemzési módszer is használja a fuzzy eszközöket).

Építhetők fák is hierarchikus asszociációs szabályok segítségével. Ekkor a gyakori elemhalmazok kinyerése közben célszerű már a fa elkészítése az ősök segítségével – ehhez szükséges az elemek taxonómiájának ismerete.



*forrás: Kerékgyártó, Mundruczó és Sugár 2001 alapján saját munka*

28. ábra: Hierarchia asszociációs szabályokhoz

Hierarchikus asszociációs szabályok [**Hiba! A hivatkozási forrás nem található.**] bsztraktabb termékkategóriák közti összefüggések keresésére alkalmasak, ahol nem mutatható ki asszociációs kapcsolat konkrét termékek közt (LIU 2007).

Jelenleg az asszociációs szabályokat főleg taktikai és stratégiai eszközként használják, de térnyerése a döntéstámogatásban is folyamatos.

A leglátványosabb eredmények asszociáció segítségével egyértelműen termékek eladása kapcsán érhetők el:

- árukapcsolással (vagy ha a vevő megvásárolja A terméket, valószínű, hogy B termékből is vesz, ezt kell reklámozni vagy együtt akciózni – ezt elegendő viszonylag ritkán felülvizsgálni, de első alkalmakkor várható a nagyobb profitnövekedés),

- az áruház átrendezésével (a gyakori termékpárokat egymás mellé tenni – itt nagy lift érhető el, mivel csak ritkán kell elvégezni, ám akkor logisztikai feladattal is párosul), illetve
- áruajánlattal (a többi vásárló ezen termékek mellé még az alábbiakat vette, valamint szabályok alapján történő ajánlás – ez folyamatos feladat, de mivel a szabályokat elegendő ritkábban frissíteni, nem igényel nagy számítási kapacitást).

Általános érvényes tehát itt is az a szabály, hogy első alkalommal lehet látványos eredményt elérni, aztán a további körökben a hozamok egyre kisebbek lesznek.

A szabályok kialakítása persze felvet néhány problémát is, melyek kiküszöböléséhez elengedhetetlen szakterületi specialista jelenléte:

- Elképzelhető, hogy túl sok szabály fog keletkezni – manuálisan válogatni nem lehetséges. Ezen lehet segíteni a küszöbszámok növelésével, de az egyensúly megtalálása nem egyszerű.
- Nem minden asszociációs szabály fog valódi oksági viszonyt tartalmazni, nem elég tehát csak a valószínűség vizsgálata. Értelmezni kell a függetlenségi változót is, valamint próbákat végezni.
- A legtöbb szabály nem lesz „érdekes”<sup>75</sup>, mert azok csak már meglévő szabályokat specializálnak. Ilyenkor az általánosabb szabálynak kell érvényesülni.

## 4.7 Klaszterezés

A klaszterképzés lényege, hogy a rendelkezésre álló adatok alapján a sokaságot (egyedeket, rekordokat) csoportokra lehessen bontani – valamilyen hasonló tulajdonság vagy tulajdonságcsoport alapján (*XU ÉS WUNSCH 2008*). A lényeg, hogy az egy csoportba / klaszterbe / partícióba kerülő elemek sokkal inkább hasonlóak egymáshoz, mint azokhoz amelyekkel nem kerültek egy csoportba. A klaszterezés előnye, hogy – felügyelettel – bármilyen adathalmazra alkalmazható. A klaszterezés egyik nehézsége, hogy sok esetben a bonyolult klaszterezési szabályok miatt nehéz nevet adni vagy megfogalmazni a klaszter lényegét.

---

<sup>75</sup> Az érdekesség mérésére különböző mutatók állnak rendelkezésre, pl: Lift-mutató, RI mutató (kovariancia analízis) és  $\phi$  együttható (korrelációs analízis) számítása – ám a feladattól függően lehet más érdekességi mutatókat is használni (Koszinusz mutató, Természetes tényező, Hozzáadott érték mutató, Jaccard mutató, Kloggen mutató), vagy akár egyénit definiálni.

Gyakori a klaszterezés és az osztályozás összevetése: a technika egymáshoz valóban nagyon hasonló, célját és eredményeit tekintve, azonban megjegyzendő, hogy míg a klaszterezés nem felügyelt, addig az osztályozás az (*BANKS 2004*).

A klaszterezés eleinte szigorúan statisztikai feladat volt, ám az adatbányászat megjelenésével ez lett az egyik első alkalmazott ott technika.

Informatikai szempontból a klaszterezés jól automatizálható feladat, ezt bizonyítja, hogy szinte az összes adatbányász és statisztikai szoftver képes rá. A probléma azonban ott jelentkezik, hogy a művelet jelentős számítási feladattal jár, ami nagy mértékben növekedhet a dimenziók (tulajdonságok) számának növelésével (*ARABIE, HUBERT ÉS SOETE 1996*). Ahhoz, hogy ezek a számítások belátható idő alatt, vagy gazdaságosan befejeződhessenek, célszerű a dimenziók számának csökkentése.

Több algoritmus létezik, melyekkel klaszterek képezhetők. Ezek működésük szerint a következőképpen csoportosíthatóak: partícióképző módszerek, hierarchikus módszerek, valamint sűrűsége alapuló módszerek (*PEDRICZ 2005*).

#### **4.7.1 Partíció képzés**

A partícionáló klaszterező algoritmusok folyamatosan finomuló klaszterek kialakításán alapulnak, az iterációk során egyre inkább a klaszterek végállapota áll elő – a kezdeti klaszter reprezentációk lehetnek véletlenek is, vagy előállhatnak valamilyen heurisztikus technika eredményeként (*MIRKIN 1996*).

Mindkét általában alkalmazott algoritmus a térben elhelyezkedő pontok (az adatok reprezentálása scatterplot ábrán) egymáshoz viszonyított helyzetét vizsgálja. Míg a K-KÖZÉP<sup>76</sup> algoritmus a kialakuló csomópontok közepéből indulva alakítja ki a klasztereket (kategrikus értékek esetén nem is használható, valamint a megfelelő eredményhez többszöri futás szükséges, így némiképp komolyabb hardverkövetelményeket vagy időszükségletet támasztva), addig a K-MEDOID csak a pontok közti távolságokat vizsgálja. Mindkét függvény a távolságnégyzetek összegének minimalizálásával dolgozik, a fenti okok miatt gyakorlatban inkább az utóbbit használják.

A K-KÖZÉP eljárás első megvalósítása a FORGY (*FORGY 1965*) algoritmus, mely véletlenszerű elemkiválasztással és középponthez rendeléssel működik – számos

---

<sup>76</sup> A szakirodalom K-ÁTLAG módszernek is hívja.

változatban létezik, többféle megvalósítással. A másik eljárás az ISODATA (BALL ÉS HALL 1965) algoritmus, mely a klaszterek szétbontását szórásküszöb átlépéséhez köti. A K-KÖZÉP módszer fuzzy megvalósítása az EM, mely alkalmas folytonos és kategorikus adatok kezelésére is.

A K-MEDOID eljárásnak több megvalósítása létezik, ezek a (korai) PAM<sup>77</sup>, CLARA<sup>78</sup> és egy altípusa a CLARANS<sup>79</sup>. A PAM (KAUFMANN ÉS ROUSSEUW 1990) algoritmus minden elemre kiszámítja a minőségi mutató értékét (ha az lenne a medoid), majd a legkedvezőbbet kiválasztja. Ugyanezt teszi a CLARA (KAUFMANN ÉS ROUSSEUW 1990) algoritmus is, mindössze annyi kikötéssel, hogy kezdetben reprezentatív mintát vesz (ezt több mintahalmazzal is megteszi). A CLARANS (NG ÉS HAN 1994) folyamatosan változtatja a mintahalmazokat, majd gráfot épít a medoidok alapján.

#### 4.7.2 Hierarchikus módszerek

*„A hierarchikus eljárások onnan kapták a nevüket, hogy az elemeket egy hierarchikus adatszerkezetbe (fába, dendogramba, taxonómiába) rendezik el. Az adatpontok a fa leveleiben találhatók. A fa minden belső pontja egy klaszternek felel meg, amely azokat a pontokat tartalmazza, melyek a fában alatta találhatók. A gyökér az összes adatpontot tartalmazza.” (LUKÁCS ÉS FOGARAS 2006)*

A hierarchia felépítésének szempontjából meg lehet különböztetni felhalmozó<sup>80</sup> és lebontó<sup>81</sup> eljárásokat. A felhalmozó (vagy bottom-up) eljárások elején minden elem külön klaszter, majd a fában felfelé haladva van esély az egyesülésre. Másik esetben, a lebontó (vagy más néven top-down) eljárások esetén eleinte minden elem ugyanahhoz a klaszterhez tartozik, majd a fában lefelé haladva van esély azok szétválására.

A hierarchikus módszerek előnye, hogy bármilyen adattípusra képesek működni, és programozásuk is könnyű. Hátrányuk, hogy a fában magasabban bekövetkező hibák nem javulnak / nem javíthatóak, és nehéz eldönteni, hogy meddig épüljön fel a fa.

---

<sup>77</sup> PAM: Partitioning Around Medoids azaz partícióképzés medoidok mentén.

<sup>78</sup> CLARA: Clustering LARge Applications – nagy adathalmazok klaszterezésére.

<sup>79</sup> CLARANS: Clustering Large Applications based upon RANdomized Search, vagyis az elődje, véletlen kereséssel.

<sup>80</sup> A felhalmozó hierarchikus módszereket a szakirodalom egy része egyesítő hierarchikus módszerek néven kezeli.

<sup>81</sup> Az előbbihez hasonlóan a lebontó eljárásokat a szakirodalom több helyen felosztó eljárásokként említi.



Minden ilyen rendszerű algoritmus kritikus pontja a klaszterek távolságát mérő függvény megválasztása, ezen a bemeneten múlik a végeredményként előálló klaszterek halmaza.

A felhalmozó módszerek közül alapvetően három algoritmus emelkedik ki. Ezek a CURE<sup>82</sup>, ROCK<sup>83</sup> és CHAMELEON<sup>84</sup>. A CURE (GUHA, RASTOGI ÉS SHIM 1998) kizárólag numerikus adatok kezelésére alkalmas, ám véletlen kezdeti mintavételezésének köszönhetően nagy adatbázisokon is jó teljesítménnyel működik. A véletlen mintahalmazon is tovább partíciónál, majd reprezentatív elemeket határoz meg, melyek távolságai később az összevonás alapját adják – nem kizárólag gömb alakú csoportok felderítésére alkalmas. A ROCK (GUHA, RASTOGI ÉS SHIM 1999) algoritmus kifejezetten kategorikus értékek alapján történő klaszterképzésre szolgál (SARKER, ABBASS ÉS NEWTON 2002). Futása során gráfot épít, melynek ágai a még különböző csoportba tartozó, de már közös szomszédokkal rendelkező egyedekből keletkeznek. A CHAMELEON (KARYPIS, HAN ÉS KUMAR 1999) a másik kettő ismeretében ékszült, azok hátrányait kivédeni – nem kizárólag az egyes klaszterek közötti távolságokat figyeli, hanem a klaszteren belüli távolságokat is figyelembe veszi. A csoportokat adatpontok alapján dinamikusan építi fel, így bármilyen alakú csoportok felismerésére alkalmas.

A lebontó eljárások lényege (az előzővel ellentétesen), hogy kezdetben minden elem / megfigyelés ugyanabba a csoportba tartozik, majd különböző kritériumok alapján a csoportok szétvágásra kerülnek. A vágási feltétel meghatározása során vannak eljárások, mely egy- és vannak eljárások, melyek több attribútumot vesznek figyelembe. A leggyakrabban használt (ABONYI 2006) egyváltozós hierarchikus felosztó módszer az asszociáció analízis (WILLIAMS ÉS LAMBERT 1959), melynek lényege, hogy minden vágás előtt megkeresi azt a tulajdonságot, mely a legerősebb asszociációt (khi-négyzet statisztikával) mutatja a többivel. Csak bináris adatmátrixon működik, így általában megelőző lépésként adattranszformáció szükséges. (Létzeik még információtartalom alapú vágás is.) A többváltozós felosztó hierarchikus módszerek (mint amilyen a gyakori MacNaughton-Smith eljárás is (MACNAUGHTON-SMITH, WILLIAMS, DALE ÉS MOCKETT 1964)) két lépésből állnak: objektumkiválasztás és távolságszámítás, majd az objektumhoz mért kisebb távolságú objektumok mozgatása az új csoportba. Használatának előnye, hogy

---

<sup>82</sup> CURE: Clustering Using REpresentatives, vagyis klaszterezés jelöltek segítségével.

<sup>83</sup> ROCK:

<sup>84</sup> CHAMELEON:

lényegesen kisebb számítási igénnyel rendelkezik (ABONYI, 2006), mint bármely egyesítő algoritmus.

### 4.7.3 Sűrűségeen alapuló módszerek

A sűrűség alapú módszerek lényege, hogy a már vizualizált adatokat (általában pontokként, több dimenzióban) tekintve még emberi szemmel is könnyű meghúzni a klaszterek határait, el lehet dönteni, hogy melyek azok az előfordulások, melyek közelebb vannak egymáshoz, azaz egy klaszterbe célszerű kerülniük, és melyek azok melyek kívül esnek ezeken a határokon. A sűrűségeen alapuló algoritmusok legjobb példái a DBSCAN és az OPTICS

A DBSCAN (Density-Based Spatial Clustering of Applications with Noise – azaz zajos halmazok sűrűség-alapú térbeli klaszterezése) algoritmust 1996-ban hozták létre Martin Ester, Hans-Peter Kriegel, Jörg Sandler és Xiaowei Xu, azóta ez az egyik legelterjedtebb klaszterező folyamat. Az algoritmus két kezdőváltozó megadása után végiglépdel a pontthalmazon és minden pontra megvizsgálja, kellően közel van-e az aktuálishoz, lehet-e vele egy klaszterben (TERANO, LIU ÉS CHEN 2000). Előnyei, hogy tetszőleges alakú klasztereket tud előállítani (tehát nem kizárólag kör vagy gömb alakúakat), jól kezeli a zajt, valamint nem kell előre ismerni, hogy hány klaszterből is fog állni az adathalmaz. Hátránya azonban, hogy rosszul kezeli a változó sűrűségű halmazokat, és a kezdeti változók rossz megadása esetén helytelen eredményeket szolgáltat. Az alterek klaszterezésére vonatkozóan több algoritmus alapja a DBSCAN (SZCZUKA, KRYSZKIEWICZ, RAMANNA ÉS JENSEN 2010), ilyenek a PreDeCon és a SUBCLU.

Az OPTICS (Ordering Points To Identify the Clustering Structure – azaz Pontsorrend felállítása a klaszterképző struktúra megállapításához) algoritmus a DBSCAN-nel hasonló elgondolás, de inkább hierarchikus módszernek tekinthető – létrehozói Mihael Ankerst, Markus M. Breunig, Hans-Peter Kriegel és Jörg Sandler. Hasonlóan az előzőhöz, itt is két változó megadásával indul a folyamat, de jóval kevésbé érzékeny azok rossz értékére. A folyamat célja, hogy kiküszöbölje a változó sűrűségből adódó detektálási hibákat (HAN ÉS KAMBER 2001). Itt minden pontra számítható egy a környezetet jellemző sűrűségi érték, így a folyamat befejeztével többféle klaszterezés is rendelkezésre áll. Ezen kívül azonban számos hasonló tudású algoritmus áll rendelkezésre: OPTICS-OF, DeLi-Clu, HiSC, HiCO, DiSH – melyek mind az OPTICS kiterjesztései.

#### 4.7.4 Egyéb módszerek

Léteznek persze más csoportosító eljárások is, ezen a területen is használják a Fuzzy módszereket és a neurális hálókat.

Az itt alkalmazott fuzzy módszerek lényege, hogy míg minden más módszer esetén a végeredményben egyértelműen eldönthető, hogy melyik elem melyik csoportba tartozik (és emiatt kemény csoportosító módszereknek is hívják őket), addig itt egy elem egyszerre több halmaz része is lehet (és emiatt lágy csoportosító módszernek is nevezik). Fuzzy csoportosítás esetén a cél a tagsági függvény kialakítása, mely megmondja, hogy egy elem mennyire tartozik a halmazba (értéke 0, ha egyáltalán nem és 1, ha teljesen – ha csak ez a két érték fordulna elő, akkor nem Fuzzy megoldásról lenne szó). A módszer konkrét megvalósítása az FCM (*BEZDEK 1981*) algoritmus, melynek kidolgozásához Dunn és Bezdek járult hozzá.

A neuronhálók klaszterezéshez való felhasználását Hertz (*HERTZ, KROGH ÉS PALMER 1991*) javasolta először, majd felhasználásukat Kohonen (*KOHONEN 1997*) tökéletesítette, megalkotva az önszerveződő hálózatok elméletét.

Léteznek rács és modell alapú megközelítések is: Az egyik legegyszerűbb, modell alapú növekményes algoritmus (mely nem sorolható be a fenti kategóriákba) a COBWEB (*FISHER 1987*), mely egy osztályozási fát épít, minden csomópont valószínűségi értékekkel van ellátva, a gyökér a teljes adathalmaz és folyamatosan bővül.

#### 4.7.5 Értékelés

Bár viszonylag sok lehetőség adott algoritmusválasztás tekintetében – ráadásul elméletileg lehetséges új algoritmusok definiálása is, célszerű azokat különböző szempontok szerint értékelni (*LUKÁCS ÉS FOGARAS 2006*).

Az értékeléshez (és kiválasztáshoz) a következő szempontokat érdemes megfontolni:

- Adattípusok, melyekkel működni képes: fontos, hogy képes legyen az adathalmazban előforduló minden adattípus kezelésére. A felhasználható adattípusok lehetnek numerikusak, binárisak vagy kategorikusak (minőségi adatok).

- Robosztusság: A zajokra való érzékenység alacsony legyen (itt a robusztusságnak nem kell kiterjednie hibátűrésre, mivel az adatok előkészítése során már azokat a hibáktól és hiányosságoktól mentesíteni kell).
- Skálázhatóság: Az algoritmus kiválasztásánál figyelembe kell venni, hogy idővel mekkorára nőhet az adatmennyiség, illetve a dimenziók száma miként változhat. Így fel lehet készülni a teljesítmény-problémák elkerülésére.
- Klaszter-geometria preferencia: Célszerű olyan algoritmust választani, mely nem rendelkezik geometria preferenciákkal.
- Sorrend-függetlenség: Az algoritmus kimenete (a klaszterek) nem függhet az adatok beolvasási sorrendjétől.
- Valamint a Kleinberg-tétel<sup>85</sup> szerint az alábbi három tulajdonság egyikének megléte ugyancsak ajánlott (egymást kizáró tulajdonságok):
  - Skála-invariancia: Ha minden elempár távolsága ugyanannyiszorosára nő, az eredménynek ugyancsak változatlanul kell maradnia.
  - Gazdaság: Ha előre adottak a klaszterek, léteznie kell olyan távolságmátrixnak, hogy az előre adott eredményt szolgáltatassa.
  - Konzisztencia: Ugyanazt az eredményt kell akkor is szolgáltatnia, ha az azonos klasztereken belül lévő elempárok távolsága nem nő, illetve a különállóknak nem csökken.

---

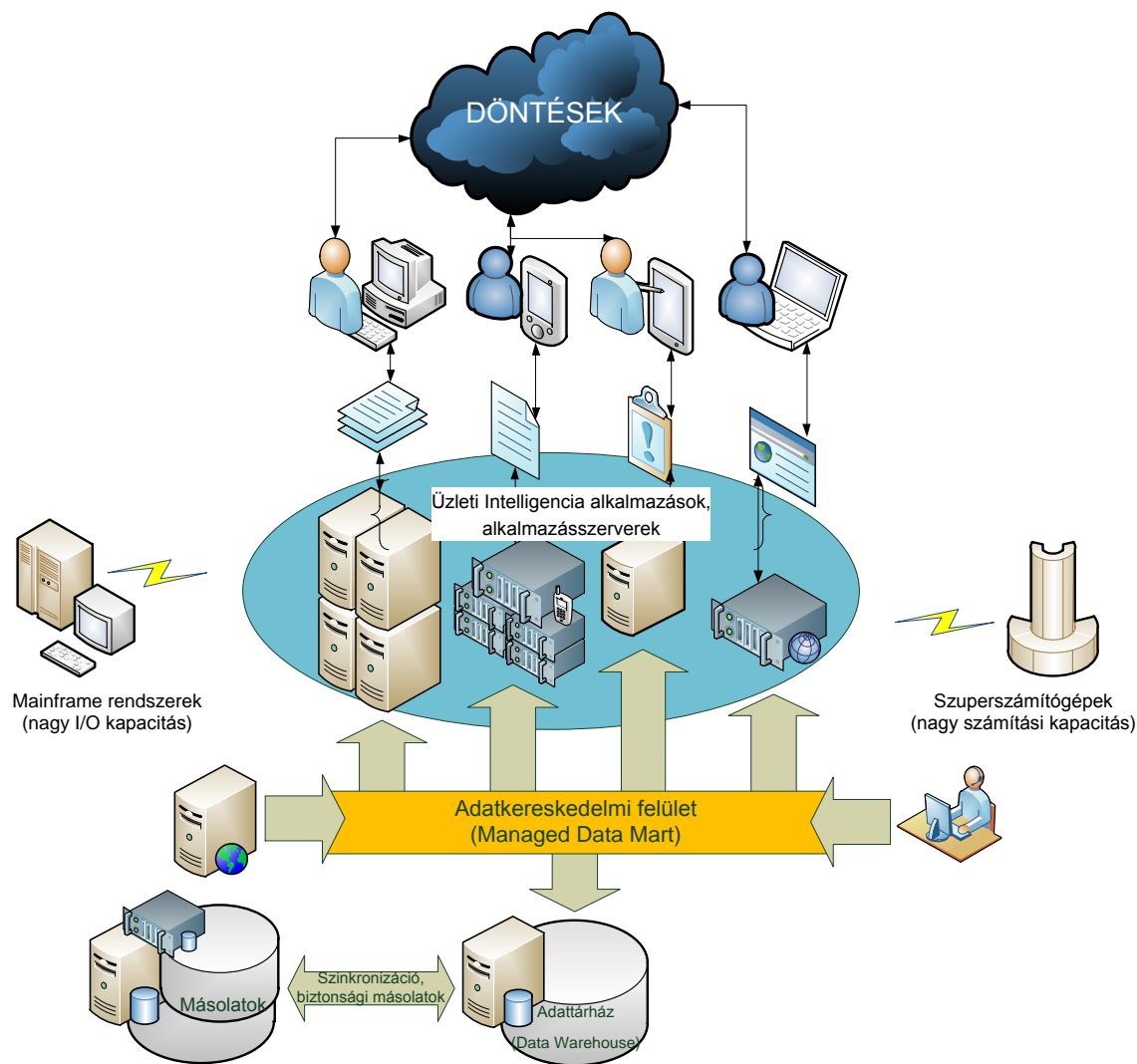
<sup>85</sup> Kleinberg-tétel: Nem létezik olyan klaszterező eljárás, ami egyszerre rendelkezik invariancia, gazdaság és konzisztencia tulajdonságokkal.

## 5 Üzleti intelligencia infrastruktúra

Mára hagyományosnak mondhatóan a BI feladata az, hogy olyan metrikát teremtsen a vállalat számára, mely megvalósulás esetén növeli a teljesítményt, és segít a jobb pozicionálásban. Az üzleti intelligencia legáltalánosabban döntéstámogatási céllal terjedt el, mivel a megfelelő adatok szolgáltatásával kedvezően befolyásolhatóak a stratégiai, taktikai és működési célok (*ALLAN 2007*).

A BI tradicionális felhasználása abban rejlik, hogy a döntésekért felelős személyekkel segít megértetni azt, hogy pontosan hogyan is működik a szervezet, és milyen hatások várhatóak bizonyos helyzetekben – így kideríthető az is, hogy mit kell tenni a jobb eredmények eléréséért. Ám ekkor a folyamatok rögzítettek. Felmerülhet a kérdés, hogy lehetséges-e magukat a folyamatokat optimalizálni valamilyen BI eszköz segítségével? Igen, lehetséges – ám ehhez tovább kell lépni, fejlettebb megoldásokat kell igénybe venni. A három megemlíthető és alkalmazandó megoldás: a visszacsatolási hurkok képzése a vállalaton belül, a valós idejű üzleti intelligencia (*BUSSLER 2007*), illetve az üzleti aktivitás figyelése.

A klasszikus BI infrastruktúra felépítése [29. ábra]: A tipikus üzleti intelligenciát ellátó infrastruktúra tartalmaz operatív elemeket, a hozzájuk tartozó adatbázisokat, illetve azok potenciális instanciáit, beleértve ebbe az adattárházakat és adattárakat [31. ábra]. Az adattárházak felett általában egy úgynevezett adatkereskedelmi (adatcsere) felület található, mely a különböző analitikai alkalmazások speciális adatigényének kielégítésére szolgál. Az ábrán látható modellben az elemzők (emberek) a jelentéskészítésből és analitikai alkalmazásokból származó adatokhoz férhetnek hozzá – többféle interfészen keresztül (*HOWSON 2007*).



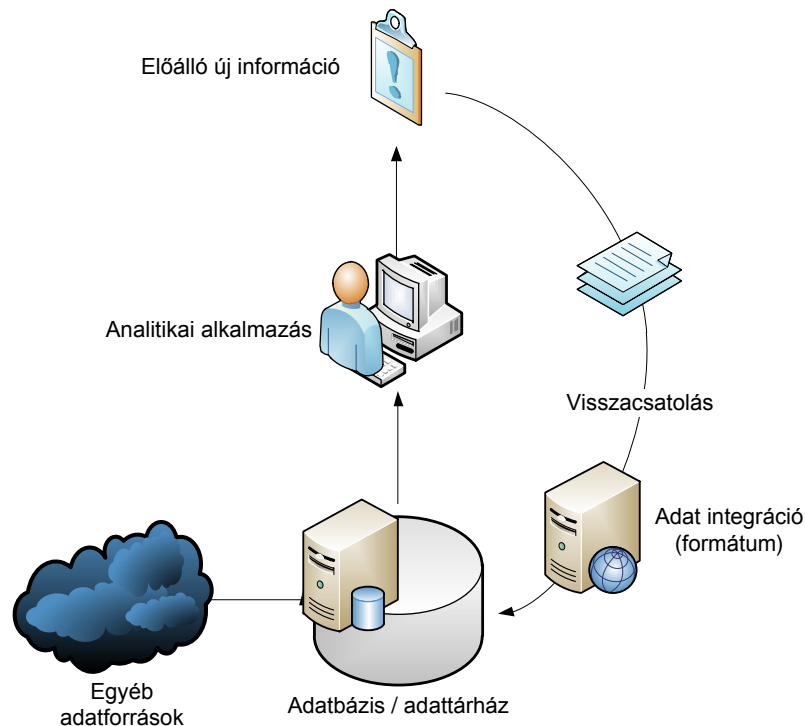
forrás: saját munka

29. ábra: Üzleti Intelligencia rendszer felépítése

Így az elemző felelőssége az, hogy jól kommunikáljon minden új információt, mely a műveletek során előáll. Ilyen művelet lehet a prediktív modellezés, az előrejelzés, az optimalizáció mellett sok egyéb más is. Az előálló információ lehet a későbbi döntések alapja, elsődleges fogyasztója a vezetés (THIERAUF 2001).

Ahhoz, hogy hatékonyan működjön az üzleti intelligencia rendszerünk, célszerű visszacsatolást létrehozni. Az ilyen hurkok létrehozásával [30. ábra] szorosabbá válhat a kapcsolat a BI-rendszer, illetve az üzleti folyamatokat megvalósító szoftverek (és üzemeltetőik, felhasználói) közt. A hatékonyság növelésével persze további profitemelés érhető el. Egy ilyen modellben azok az eredmények, melyek az analitikát megvalósító szoftverben előállnak, nem kizárólag az elemzőhöz (és rajta keresztül a döntéshez) jutnak,

hanem vissza az operatív rendszerbe is (PAREEK 2007). Így az előrejelzésből, költségmodellezésből, optimalizációból származó eredmények már adat szinten (lent) integrálhatóak, valamint visszajuttathatóak bemenetként az alkalmazások adatbázisaiba, vagy az adattárházba.



forrás: saját munka

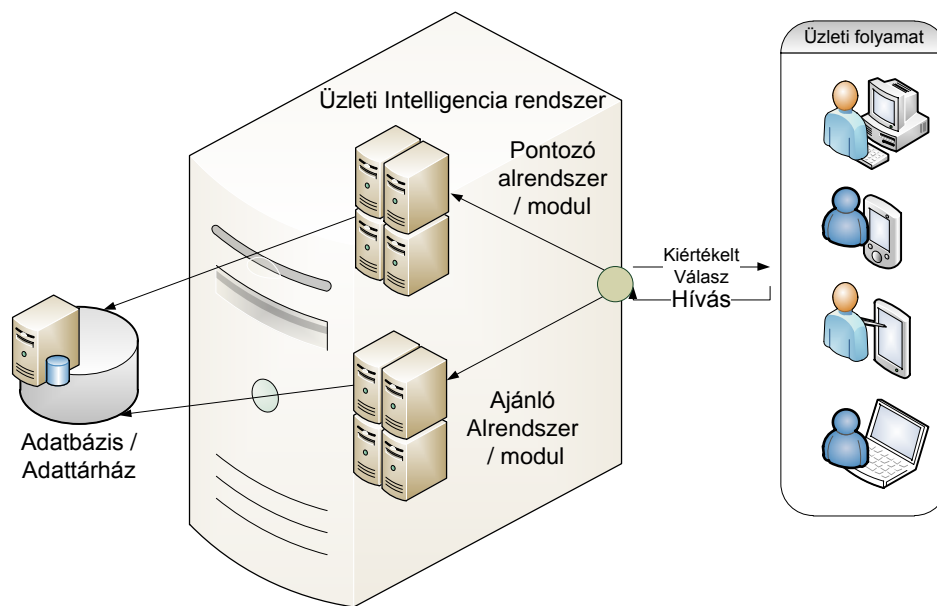
### 30. ábra: Új információ előállítása

Ez a technika ott a leghatásosabb, ahol léteznek ciklikus folyamatok, mint például havi, heti vagy napi beszerzés és tervezés. Az információ feldolgozói ebben az esetben majdnem azonnal érezhetik a pozitív hatást, mivel az alkalmazás minden esetben figyelembe fogja venni az újabb és újabb eredményeket (RASMUSSEN, GOLDY ÉS SOLLI 2002).

Egy ilyen visszacsatolásos modell létrehozásához az elemzések eredményeit egy soros feldolgozóba vezetve ellenőrzik az adatok formátumának helyességét (vagy az adatok formátumát kell az operatív alkalmazás számára fogadhatóra alakítani, vagy a fogadó alkalmazást kell képessé tenni az adatok eredeti formában való fogadására). Célszerű a tervezési és tesztelési szakaszban adatellenőr szakemberek bevonása.

A második fontos, megfontolandó technika a valósídejű, helyszínen meghozható döntéseket lehetővé tevő megoldás. A valósídejű döntéshozatallal főként a kapcsolódó

termékek eladásánál, csalásfelderítésnél lehet javulást elérni. Ilyen esetekben az üzleti folyamatnak, vagy alkalmazásnak hozzá kell tudnia férni az üzleti intelligencia infrastruktúrájához. Ez a hozzáférés azt a reakciót váltja ki, hogy a BI rendszer ajánló vagy pontozó alrendszere kiértékeli a szituációt (a lehető leggyorsabban), és válaszol a kérésre. A válasz lehet igen/nem jellegű vagy a következő ajánlat kereszteladáshoz (BUSSLER 2007).



forrás: saját munka

31. ábra: Adattárházak, üzleti intelligencia és üzleti folyamatok kapcsolata

Ezekben az esetekben előzetes pontozási technika is elterjedt, de az előre rögzített modellek nem kellő gyorsasággal a vásárló által okozott helyzetre, mivel ezek a pontszámok nem sokkal csak a következő ciklus előtt kerülnek számításra.

Ahhoz persze, hogy valós idejű üzleti intelligencia rendszert építhessünk, ami történeti adatokra épül, még mindig szükség van a klasszikus BI lépéseire. A különbség persze szembevető teljesítményben és sebességben egyaránt: most már megoldható, hogy az operatív alkalmazás meghívja a folyamatosan finomodó modell eredményét; ahelyett hogy egy hetekkel ezelőtt frissített fájlból kellene kiolvasni az eredményeket (CASTENALLOS, DAYAL ÉS SELLIS 2009).

Az említett megoldás elsőként a bankrendszerben terjedt el, mivel minden hitelkártya tranzakció előtt, néhány milliszekundum alatt kiértékelhető az ügyfél hitelképessége (ez elsősorban az USA-beli szokások miatt ott terjedt el, de mindenhol használható – az

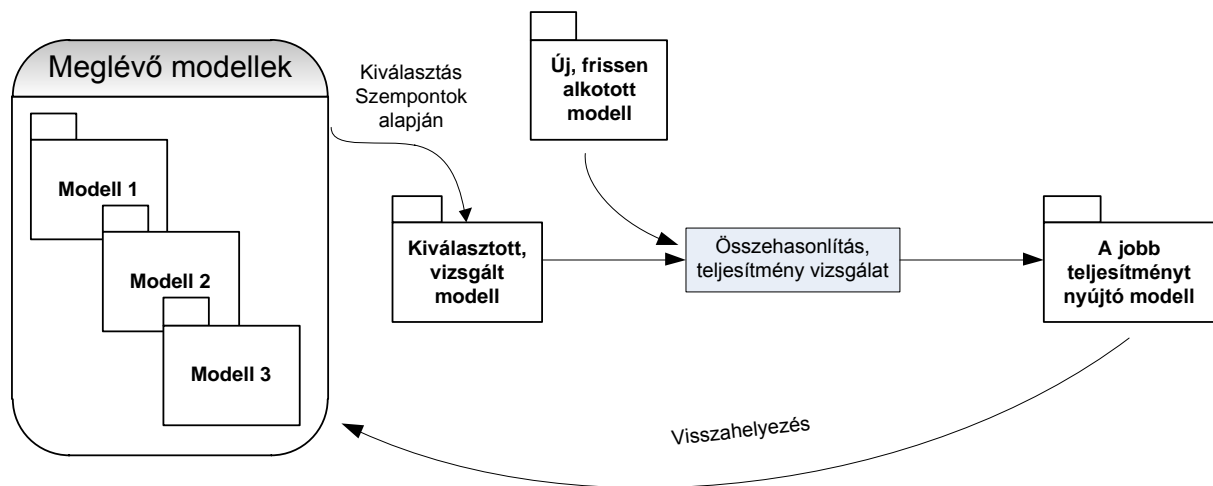


eredmény itt igen/nem/továbbküldendő értéket vehet fel). A megoldás úgy adoptálható jól a csalások felderítésére, ha az adatok először egy adatbányászati alkalmazásba kerülnek, még az adattárházba írás előtt, egyenesen az operatív alkalmazásból, így rögtön összehasonlítható a már meglévő mintákkal (*CASTENALLOS, DAYAL ÉS MILLER 2010*).

Ez, a valós idejű adatelemzés megköveteli, hogy az alkalmazások / rendszerek jól tudjanak kommunikálni egymással – ezt könnyebb nyitott infrastruktúrával elérni. Két megoldás létezik: az egyik szerint meg kell valósítani az operatív alkalmazás és a BI rendszert összeköttetését. Ez némi fejlesztést igényel, ugyanis ismerni kell az operatív alkalmazás, valamint a BI rendszer felépítését, működését és a kívánt kommunikáció menetét. A másik megoldás persze az, ha a felállított modellt az operatív adatbázison belül futtatjuk. Ez utóbbi megoldás inkább lehetővé teszi, hogy a lehető legtöbb adatot gyűjtsük össze és használjuk fel a döntéshozatalhoz, ráadásul mindez gyorsabban hajtható végre, mivel minden adat ugyanott található (*SABHERWAL ÉS BECERRA-FERNANDEZ 2010*).

Minden szervezetnek szükséges érdeke, hogy számon tartsa és figyelje a meglévő modelleket (amelyek alapján a döntéshozatal zajlik), valamint fontos, hogy rendszeres időközönként megvizsgálja a modell teljesítményét – elképzelhető, hogy a világ előbb-utóbb elmozdul, és akkor a modell már nem fogja azt kellő hűséggel reprezentálni. Amint van észrevehető változás, célszerű egy másik – kihívó – modellt létrehozni és megvizsgálni, hogy az új vagy a régi modell teljesítménye jobb. Amennyiben az új modell teljesítménye jobb, a csere indokolt. A modellek életciklusát [*forrás: saját munka*

32. ábra] célszerű figyelemmel követni, a modelleket tárolni, és rendszeresen megvizsgálni – ilyen feladatokra számos szoftvereszköz áll rendelkezésre (*ALONSO, DADAM ÉS ROSEMAN 2007*).



forrás: saját munka

32. ábra: A modellek életciklusa

A harmadik figyelemre méltó megoldás a figyelmeztetések (és ezzel egyidejűen ajánlások) küldése a folyamatok javítása érdekében.

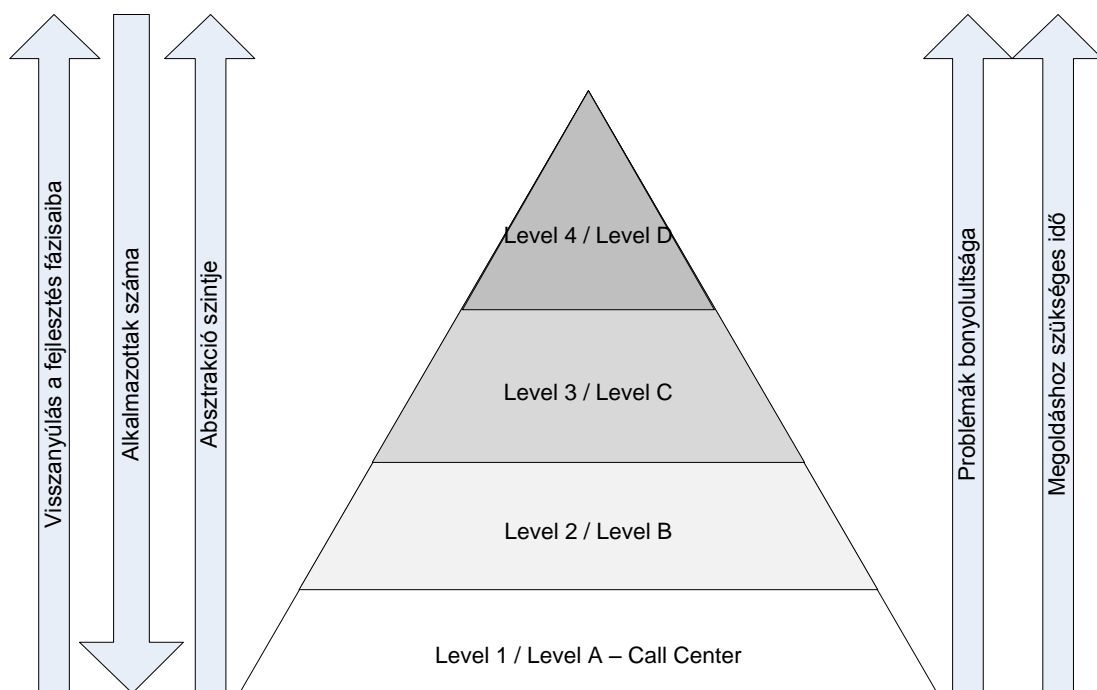
Azt az összeállítást, ha az üzleti intelligencia infrastruktúra felügyeli az üzleti folyamatokban a hibák keresését, és javítását, BAM megoldásnak (Business Activity Monitoring – Üzletmenet felügyelet) nevezzük (KRISHNA 2006). Ilyen megoldás alkalmazása esetén, ha a rendszer a normálistól eltérő üzletmenetet észlel, vagy üzenetet (vagy másfajta kommunikációt) küld egy kijelölt döntéshozó személynek; vagy maga az üzleti intelligencia rendszer felügyeli az operatív alkalmazás és a felhasználó/aktuális döntéshozó kommunikációját (HAVEY 2005).

A második esetben a BI infrastruktúra valós időben felügyeli az eseményeket, és elemzi azokat a megadott szabályok szerint, normálistól eltérő ügymenetet keresve. Ilyen eltérés lehet a kiemelkedően nagymennyiségű tranzakciótípus (egy adott fajtából) vagy az egyik ügymeneti fázis hosszúra nyúlása.

Ahhoz hogy egy jól működő BAM megoldás kivitelezhető legyen, szükség van két további komponensre a rendszerben: az egyik az üzenetküldő komponens a másik a szabályokra vonatkozó repozitórium (GREENBERG 2004).

Az üzenetküldő komponens feladat, hogy szükség esetén értesítse a kijelölt személyt a beavatkozás szükségességéről. Meg kell tudnia határozni, hogy ki az illetékes személy, ez függhet a személy beosztásától, munkaidejétől, szabadságaitól, a probléma típusától és helyétől. Itt is érvényesül a level-of-support rendszer (TECHTARGET.COM 2002).

A level-of-support [33. ábra] vagy más néven a támogatás szintje arra utal, hogy a technikai segítségnyújtás több szintre bontható. A legalsó szinten általában call center szolgál, ahol alapvető segítségnyújtás zajlik csak, előre megírt forgatókönyvek alapján, kérdés válasz rendszerben. Ahogy a felmerülő problémák bonyolultabbak lesznek, úgy ha az adott szinten arra nem adható megoldás, akkor azt az eggyel feljebb álló szintnek kell továbbítani. Így a legfelső szintre kevés probléma jut el, viszont azok meglehetősen súlyosak. A hierarchiában feljebb mozdulással nő a lehetőség a fejlesztés korábbi (tervezési) fázisaiba visszanyúlásra, csökken az alkalmazottak / beavatottak száma, az absztrakciós szint és az adott probléma megoldásához szükséges átlagos növekszik.



forrás: saját munka

33. ábra: Level-of-support

A döntéshozó persze használhat akármilyen segítséget a döntés meghozatalához, választhat dashboard alkalmazást, vagy statisztikai alapokon nyugvó analitikai alkalmazást is. Megoldható, hogy a BAM ne csak értesítsen a probléma létéről, hanem – amennyiben a folyamat befolyásolható szoftverek által – cselekedjen is. Ilyen cselekvés lehet közlekedési dugó okozta hosszúra nyúló szállítási idő esetén a további szállítmányok útvonalának áttervezése.

A másik az értesítés költséghatékony módja: használjon e-mailt, belső telefonhálózatot, mobil vagy külső telefonhálózatot, SMS-t vagy ezek valamilyen akár időben eltérő kombinációját.

A már meglévő BI rendszer mellé újabb szolgáltatásokat / rendszereket fejleszteni / vásárolni és telepíteni nem egyszerű feladat, mivel minden esetben figyelni kell a rendszerek megfelelő integráltságára. Az összekapcsoltság növelésével hibák szűrhetőek ki, növelhető az automatizáltság és további összefüggések, hatások fedhetőek fel. Amire azonban mindenképpen érdemes odafigyelni, az a következők:

- Klasszikus üzleti intelligencia műveletek, mint OLAP kockák szerkesztése, vagy ETL folyamatok kezelése. (ETL: Extract-Transform-Load, vagyis az adatintegráció három lépése: az adatok kigyűjtése, azok használható formára hozása és az adatbázisba töltése.)
- A modellek életciklusának felügyelete, vizsgálata (különösen ahogy a modellek száma növekszik).
- BI szolgáltatások, melyek idővel egyre fontosabbá lesznek.
- BAM stratégiák létrehozása, a szabályrendszer kezelése valamint a rendszerek összekapcsoltságának vizsgálata.

Ezért megállapítható, hogy a megfelelő nem csak implementáció esetén egy üzleti intelligencia rendszer jól integrálható abba a folyamatba is, mely a vállalatok napi működését szabályozza, hanem képes javítani is azokon.

## 5.1 Technológiai kérdések

Általánosan elmondható, hogy az üzleti intelligencia rendszerek több szempont miatt is rendkívül bonyolultnak számítanak:

- Egyfelől a felhasznált matematika / statisztika miatt,
- Másrészt a magas integráltság és
- Harmadrészt a magas technológiai színvonal miatt.

Ez a három tényező a következő képpen jelenik meg: az adatbányászat, mint felhasznált és jól körülhatárolható eszköz rendkívül bonyolult algoritmusokat képes hasznosítani, melyekben felsőbb matematikai és statisztikai eszközök egyaránt megtalálhatóak, szinte minden esetben a mélyebb üzleti intelligencia része. Ilyen elemzések elvégzéséhez (jobban

mondva ahhoz, hogy ténylegesen jó eredményt adjanak) sok adatra van szükség, ezeket tárolni és mozgatni kell – a vállalat méretével arányosan nő a keletkező adatmennyiség (különösen akkor, ha már rutinosan használják az üzleti intelligenciát). Ahhoz persze, hogy sok adatot bonyolult módszerekkel tudjunk elemezni, sok erőforrásra van szükség – akkorára, hogy egyetlen szerver sok esetben nem elegendő; célszerű tehát valami nagyobb erőforrást nyújtó lehetőséget keresni, valami olyat, ami megkönnyíti a szükséges hardver és szoftverelemek kezelését és karbantartását is. A nagy számítási teljesítmény létrehozásával és kiosztásával kapcsolódik tehát a virtualizáció, a gridek építése és a számítási felhők használata. Az adatok tárolásában és rendszerezésében pedig az adattárház filozófia segít.

### **5.1.1 Architektúra**

Az üzleti intelligencia architektúra (BI Architecture) olyan gyűjtőfogalom, melybe olyan a vállalat szerte használt rendszerek, alkalmazások, folyamatok tartoznak, melyek kifinomult analitikát valósítanak meg, adatoké folyamatok, tartalmak áramoltatásához azokhoz a célszemélyekhez, akiknek szükségük van erre. Ezek a személyek lehetnek statisztikusok, menedzserek, folyamatirányítók, információkkal dolgozók – számukra fontos, hogy időben, pontos és megbízható információhoz jussanak a változó komplexitású döntéseik meghozatalához. Az információnak persze nagyon sok csatornája lehet: hagyományos jelentések, ad-hoc elemző eszközök, dashboard-ok, spreadsheet-ek, e-mailek vagy automatikus figyelmeztetések (*PAGE 2009*).

### **5.1.2 Adattárházak**

Az EDW (Enterprise Data Warehouse) magyar megnevezése vállalati adattárház. Ebben az információkötegben megtalálható minden adat: történeti és kurrens egyaránt (*DYCHÉ 2000*). Bár a felhalmozott adatmennyiség gyakran túl sok lehet egyes elemzők számára, meglétének célja mégis az, hogy egy helyen lehessen elérni minden a céget érintő információt.

Azaz, az adattárház gyűjti össze a vállalat minden részlegének összes adatát (*HAMMERGREN 2009*). A központi kezelésnek több előnye van: egyszerűbbé válik az adatkezelés (a centralizált hozzáférés-jogosultság kiosztás javít a biztonságon, a centralizált tárolás pedig olcsóbbá és könnyebbé teszi az archiválást és a biztonsági

másolatok készítését és kezelését), és jelentősen egyszerűsödik az adatok beszerzésének költsége (idő takarít meg).

Az adattárház létrehozása hatalmas projekt. Össze kell szedni, hogy az egész vállalatnál milyen adatok vannak, milyen természetűek, hol és milyen formában keletkeznek, és később hol lehetnek felhasználva (*DYCHÉ 2000*). A tény pozitív oldala az, hogy számos módszertan (*MOSS ÉS ATRE 2003*) és ajánlás foglalkozik adattárházak építésével. Ettől függetlenül persze lépcsőzetesen célszerű végrehajtani – valamint itt is igaz az általános mérnöki igazság, mi szerint az alapos és átgondolt tervezés nagyban képes hozzájárulni a termék (itt adattárház) elkészültéhez és magas minőségéhez, jó használhatóságához (*JARKE 2003, KHAN 2005, GALLI 2010*).

Ha egy adattárház már üzemben van, nagyon jó lehetőséget ad a vállalatnak arra, hogy a meglévő adathalmazt elemezve, segítse a döntési folyamatot. Az elemzések során gyakran adatbányászati eszközök kerülnek használatra, melyek eredményei riportokban jelennek meg. Ezek a riportok aztán főként taktikai és stratégiai döntések meghozatalában segítenek. Felmerül a kérdés persze, hogy operatív döntések során lehet-e használni a technológiát, de megállapítható egyrészt, hogy rendkívüli erőforrásigénye miatt csak nehezen és költségesen valósítható meg, másrészt az ilyen elemzések számítás- és adatigényük miatt nagyon hosszúra nyúlnak – így egy operatív döntés esetén nem lehetséges a legtöbb esetben az eredmények kivárása. (Ez az elemzés mára olyan munkakörök megszületését tette lehetővé, mint a riportkészítő elemző <sup>86</sup> vagy adatbányászati elemző.)

Nagy (és főként régóta üzemelő) adattárházak esetén a legnagyobb problémát a megfelelő teljesítmény nyújtása adja (*KHAN 2005*). Nem ritka, hogy az adattárházban található adatbázisok összmérete több 10 terabájtos értéket ér el. Egy ekkora adatmennyiséget persze célszerű eloszlatni, hogy egy része háttértáron, egy (aktuálisabb rész) a memóriában tárolható – ekkor a friss adatok kezelése és lekérdezése gyors lesz ugyan, de a régebbi adatok elérése lassúvá válik, és ez különösen akkor nyújtja nagyon hosszúra az adatok elemzését, ha az sok adatot érint (*WESTERMAN 2001*).

Az olyan elemzések gyorsíthatóak fel, melyek kifejezetten a vállalat egy területéről származó adatokkal foglalkozik – ebben nyújt segítséget az adatpiacok alkalmazása.

---

<sup>86</sup> Mint sok más üzleti-információtechnológiai munkakörnél, itt sem található jó magyar megfelelő; jelen esetben a *reporting analyst* kifejezésre.

(Elengedhetetlen persze az adatpiacok és az adattárház közötti megfelelő sávszélességű hálózati kapcsolat.)

#### **5.1.2.1 Adatpiac**

Az adatpiac (datamart) a szervezeti szintű / méretű adattárház. Gyakran az IT-től függetlenül hozzák őket létre. Ebben az esetben az integráltság hiánya komoly korlátokat állít az elvégezhető elemzések elé, de mégis számos olyan esetről tudni, amikor ez a megoldás oldotta meg az EDW mérete okozta fennakadást (*WESTERMAN 2001*).

Az adatpiac alkalmazásának ugyanakkor számos előnye van, melyek közösen a jobb teljesítmény elérésére vezethetők vissza. Adatpiacok létrehozásának főleg olyan infrastruktúra esetén van értelme, ami egyfelől nagy és nagyon adatintenzív vállalatnál van kiépítve – mely gyakran hatalmas adattárház elkészítését eredményezi. Ebben az esetben a központi adattárház részben tehermentesíthető (vagy terhlése csökkenthető) adatpiac létrehozásával. Az adatpiacot célszerű nem a központi adattárházzal egy (térbeli) helyen elhelyezni, hanem a vállalat kihelyezett részlegénél. Olyan adatokkal kell feltölteni, melyeket ott sűrűn használnak, így adhatja elemzések alapját (nyilván rövidtávú döntésekhez), vagy jelentések bemenetét. Minden esetben komoly figyelmet kell szentelni az adatpiac adatainak az adattárházba töltésére. A leggyakrabban alkalmazott forgatókönyv szerint ez az áttöltés / szinkronizálás egy alacsonyabb terhelésű időszakban célszerű – így elosztva a terhelési hullámokat. Figyelembe kell venni, hogy ebben az esetben persze a legújabb / legaktuálisabb adatok az adatpiacban találhatóak (és ez jó mert az operatív döntéseket ezek alapján is meghozhatják) és egészen a szinkronizálás befejeztéig az adattárházban lévő adatok (vagy változatok) régebbiek lesznek. Nem okoz gondot a technológia számára az sem, ha időközben más forrásból adatok kerülnek az adattárházba, vagy az ott lévők módosulnak, mivel minden módosítás időbélyeggel van ellátva, így a szinkronizálás során az időrend megmarad. Bevett gyakorlat bizonyos műveletek előtt / után a szinkronizálás kikényszerítése (*JARKE 2003*).

#### **5.1.3 Infrastruktúra**

Mivel az idő és a technikai lehetőségek adottak ahhoz, hogy (idővel) a vállalati adatbázisok elérjék a petabájtos nagyságrendet, ennek az adatmennyiségnek a feldolgozására fel kell készülni. A ma kapható alsó- és középkategóriás számítógépek / szerverek nem igazán

alkalmasak erre. Azonban az 64-bites szerverek (illetve rendszerek) sokkal gyorsabban képesek feldolgozni (*THE SERVER SIDE.NET 2005*).

A legtöbb klasszikus vállalat azonban több nehézséggel is küzd: az egyik, hogy nem tudják, hogyan érdemes feldolgozni és mire használni a hatalmas mennyiségű adatot; a másik, hogy informatikai részlegeik sok esetben nehezen birkóznak meg még a támogatási feladataikkal is. Az, hogy még a jól kialakított tranzakciós rendszerrel rendelkező vállalatok is küzdenek az egyszerűbb feladatokkal, mint amilyen az adattisztítás az integráció során. Vagyis: hiába a rohamos technológiai fejlődés, legtöbbször a probléma inkább az üzleti folyamatokkal van, mint az adatok kezelése, vizsgálata és az eredmények alkalmazása (*JARKE 2003*).

Fontosnak mondható, hogy minden komoly analitikai törekvésekkel rendelkező vállalat rendelkezzen egy erős információtechnológiai csoporttal. Így például a megfelelő adatok kiválasztásával és a megfelelő analitika az üzleti folyamatokba ágyazásával az IT szervezeti egység segít a vállalat kompetitív előnyének kiépítésében és megtartásában (*DAVENPORT ÉS HARRIS 2007*).

Ám fontos figyelni arra, hogy ez a feladat nem delegálható egy-az-egyben az IT részleghez: szoros együttműködés szükséges az üzleti illetve az informatikai szakemberek közt a követelmények vizsgálatához. Ez a szervezeti elrendezés sok jól működő vállalatnál megtalálható, azonban érdemes felállítani néhány szabályt, amivel biztosítható, hogy az informatikai befektetések a vállalat érdekeit szolgálják:

- Az ütköző információforrásokkal kapcsolatos kockázatokat ki kell zárni.
  - A hazard olyan helyzet, amely valóságos vagy lehetséges veszélyt jelent az emberek vagy a környezet számára.
  - Kockázat minden olyan lehetséges helyzet, ami a projekt sikeres kivitelezését akadályozhatja. A kockázatok egy hazardhoz kapcsolódó események és azok következményei, valószínűségi alapon. Azt fejezi ki, hogy milyen valószínűségű veszélyes következményei lehetnek egy hazardnak (*SZIRAY 2008*).
- Az alkalmazásokat integrálni kell, mivel az analitika mindenképpen adat-intenzív technológia: a vállalat minden szegmenséből fog adatokat gyűjteni.
- Az analitikát mindenképp az üzleti stratégia részeként kell kezelni.



Az adatok beszerzése, a technológia és a folyamatok jó kiválasztása mind az IT Architekt névvel emlegetett tervezői szerepkör felelősségei közé tartoznak. Ez általában vezető pozíció a vállalatban belül, megköveteli a CIO-val a napi szoros munkát (*MILLER, BRAUTIGAM ÉS GERLACH 2006*). Kettejüknek kell meghatározni, hogy milyen infrastruktúra (szoftverek, hardverek és hálózatok) és hogyan fogják az üzleti folyamatokat adatokkal és technológiával ellátni. Ez a feladat kezdő vállalatoknak könnyebb, mivel ilyen esetben nulláról lehet felépíteni mindezt, nem kell a már meglévő korlátokhoz igazodni. Általában a már meglévő informatikai rendszerek tökéletesen ellátják ugyan azokat a feladatokat, amikre eredendően kitalálva / kiépítve lettek, de további, más célú felhasználásuk nagyon nehézkes lehet.

#### **5.1.4 Virtualizáció**

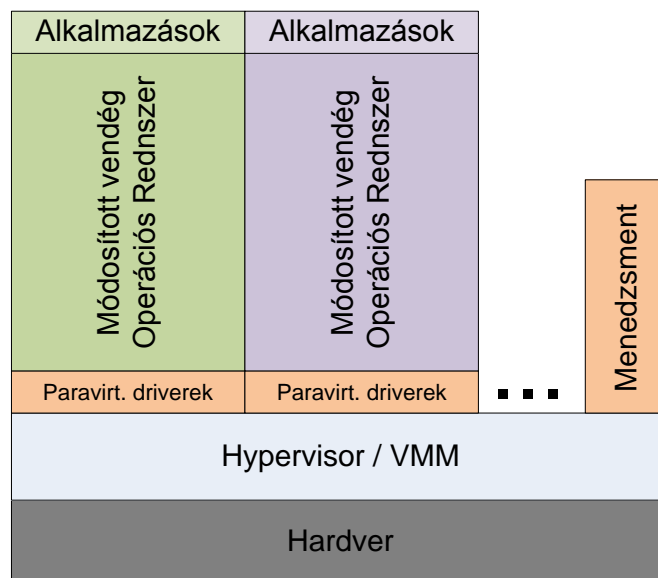
Az informatikában a virtualizáció egy széleskörűen használt fogalom, mely elsősorban a számítógépek erőforrásainak absztrakcióját jelenti (*RUEST ÉS RUEST 2009*). Négy főbb típusát különböztetjük meg. Ezek: platform virtualizáció, mely a fenti módon elválasztja az operációs rendszert az alatta elhelyezkedő platformtól, erőforrás virtualizáció, mely egy adott rendszer erőforrás virtualizációját jelent, mint például háttértárak vagy hálózati erőforrások, alkalmazás virtualizáció, mely alkalmazások kiajánlását teszi lehetővé idegen hardverre és szoftverre valamint megkülönböztetünk még desktop virtualizációt, mely egy távoli számítógép asztalának manipulálását jelenti (*GOLDEN 2007*).

A virtuális számítógép tehát egy szimulált számítógépet jelent. „A számítógépek általában fizikailag létező dolgok: elektronikai elemekből (integrált áramkörökből) felépített központi egység értelmezi és hajtja végre a programokat. A virtuális számítógép fizikailag nem létezik: a felépítése csupán egy szimuláció, egy olyan számítógépes program ami egy létező fizikai számítógépet, vagy egy fizikailag nem felépített számítógép működését szimulálja.” [34. ábra] Így valójában egy "teljes számítógép egy másik számítógépen belül" (*GOLDEN 2007*).

Napjainkban egyre több vállalat és intézmény dönt a virtualizációs megoldások használata mellett, amelytől nem csak jelentős költségmegtakarítást, de nagyobb biztonságot, az üzemeltetés egyszerűsítését, hatékonyabb helykihasználást és a leállási idő csökkentését is várják (*MENKEN ÉS BLOKDIJK 2008*).

Jelenleg a világban mintegy 6 millió darab (!) szerverfeladatot ellátó számítógép működik. Felhasználási területük (és a rajtuk futtatott operációs rendszerek és szoftverek valamint az általuk nyújtott szolgáltatások) ugyan nagyon szerteágazóak, de egy valami közös bennük: a kihasználatlanság. Ugyanis ezen számítógépek átlagos terhelése kb. 15%. Természetesen vannak időszakok, amikor szükség van a teljes teljesítményükre, de ezek az időpontok általában nem egyszerre következnek be (*BUYTAERT, DITTNER ÉS RULE 2007*). Ezen számítógépek üzemeltetése rengeteg pénzbe kerül. És az üzemeltetési költségek kb. 70%-át teszi ki az energiaszámla.

Ezen probléma enyhítésére nyújt jó megoldást a virtualizáció. Jelenleg a világ szervereinek mintegy 13%-a lát el valamilyen virtualizációs feladatot. Nemzetközi elemzések szerint a virtualizáció lesz a legmeghatározóbb trend az IT-infrastruktúra üzemeltetésének terén. Különösen igaz ez a nagyvállalati IT-infrastruktúra üzemeltetésre, mivel ott a költséghatékonyság nagyobb prioritást kap. Ez a trend minden valószínűség szerint egészen 2013-ig döntő lesz.



forrás: saját munka

34. ábra: Egy paravirtualizált rendszer felépítése

### 5.1.5 Griddek

A grid számítási hálók lényegében nem másról szólnak, mint sok számítógépes erőforrás összekapcsolásáról (így számítási teljesítmény és tárhely), mely így egy sokkal erősebb, virtuális gépet eredményez. A technológia használatával a különálló alkalmazások sokkal

gyorsabban futhatnak, valamint több alkalmazás közt eloszthatóvá válik a keletkezett nagy teljesítmény, mindez jól skálázhatóvá válik, ráadásul nincs szükség az infrastruktúra ismeretére az alkalmazások futtatása során (*RABB ÉS DONINGER 2007*).

Ez a technológia az Internet felemelkedéséhez hasonlóan gyorsan tört előre és hódított az ipari szektorban. Vállalatok nagy és kis hálózatokkal egyaránt adoptálták, mivel így csökkenthetőek voltak a futásidők, illetve kihasználhatóvá váltak az erőforrások (*JOSEPH ÉS FELLESTEIN 2004*). Használatot tekintve a pénzügyi és biztosítási szektor vezet, majd a telekommunikációs és gyártói szektor követi őket.

A grid számítási hálók lényegében párhuzamosított és elosztott feldolgozást használnak, hogy dinamikusan elosszák és összegezzék a számítógépes erőforrásokat. Általában ezek a számítási tömbök számítógépek hálózatából épülnek fel. Ez a hálózat állhat egy részlegen néhány gépből, de a mérete terjedhet több száz vagy több ezer gépig, akár geográfiailag több földrészre / városba elosztva is. A hálóba küldött alkalmazások először egy köztes réteg szoftverrel – middleware – kerülnek kapcsolatba, mely a számítási igény alapján eldönti, hogy melyik gép kezdi el annak futtatását (*BERMAN, FOX ÉS HEY 2003*). A döntés futás időben történik, figyelembe véve a rendelkezésre álló erőforrásokat is.

A grid technológia használatával jelentős teljesítménynövekedés érhető el, mivel az alkalmazások számítási igényét „feldarabolja” majd külön, párhuzamosan végzi el a kisebb műveleteket. Azokat a szoftvereket érdemes ilyen környezetben futtatni, melyeknek normál esetben órákra, vagy akár hetekre van szükségük, vagy bizonyos esetekben egyáltalán nem futtathatóak – ha egy ilyen hosszú futási idővel rendelkező alkalmazás ugyanazt az alapvető feladatot hajtja végre futása közben sok alkalommal (*FOSTER ÉS KESSELMAN 2004*), vagy kisebb részekben (kisebb adatszerkezetekkel) dolgozik tovább akkor tapasztalható különösen nagy javulás a futási idő tekintetében.

Emellé társul az a lehetőség is, mi szerint így hatékonyabban lehet kihasználni a már meglévő erőforrásokat, valamint a terheléelosztásnak, illetve a prioritizációnak köszönhetően a konkurens folyamatok is hatékonyan futtathatóvá válnak (*PLASZCZAK ÉS WELLNER 2006*).

### 5.1.6 Számítási felhők

A számítási felhő nem más, mint egy erőforrás-bérlési szolgáltatás mögött elhelyezkedő filozófia, lényege, hogy a szolgáltató hatalmas szerverfarmot épít – ez képezi a számítási felhő hardverét – majd ennek egy szerződésben meghatározott részét a szolgáltatást igénybe vevő részére bocsátja (általában havidíj ellenében). Attól függően tehát, hogy csak (virtuális) infrastruktúrát, vagy platformot is a felhasználó rendelkezésére bocsát, beszélhetünk IaaS vagy PaaS megoldásokról. Jellemzőbb (és a szolgáltatás díja általában magasabb is) az a forgatókönyv, hogy a felhasználó összeállíthatja, hogy milyen platform-elemekre van szüksége az így bérelt virtuális infrastruktúrán – így rendelkezhet webserverekkel, adatbázisszerverekkel, vagy akár egyszerűen csak „nyers” operációs rendszerrel – nyilvánvalóan ezek is paraméterként számítanak a díjazás során. A felhasználó előnye, hogy nem kell foglalkoznia a hardver- és szoftverkarbantartással, valamint a biztonsági másolatok készítésével / kezelésével; ugyancsak nem terheli a folyamatos energiaellátás biztosítása, vagy a rendszereinek fizikai védelme – a felhasználót terhelő összeg csak a havidíjat jelöli. Ezen felül, ha a felhasználó igényli, tetszőlegesen változtatható a rendelkezésre álló erőforrás: az elérhető processzorok száma és teljesítménye, az elérhető memória és háttértár mérete, a hálózati kapcsolat sebessége és beállításai. A rendkívüli skálázhatóság jól igazodik a szolgáltatások iránti változó igényhez: amennyiben egy felhőben futó szolgáltatás iránt kicsit az érdeklődés (például szezonális okok miatt), akkor az alatta fekvő erőforrások csökkenthetőek, jelentősen olcsóbbá téve a szolgáltatás fenntartását; ellenkező esetben a fellendülés alatt lehetséges a szolgáltatást nyújtó virtuális szerverek erőforrásainak lépcsőzetes növelése (ami a havidíj emelkedésével jár). Attól függően, hogy mi a felhőben futó szolgáltatás, változhat annak igénybevételi módja: ha webes szolgáltatásról van szó, akkor azt végső soron az azt igénybe vevők webböngészővel érhetik el, ám némiképp érdekesebb az az eset, amikor a felhőben elhelyezett virtuális gépekhez kell csatlakozni – ez lehetséges lehet távoli számítógéppel vagy akár vékonykliens segítségével is (BERTOLINI 2005).

A számítási felhő<sup>87</sup> elnevezés ugyanúgy a 2010-es évek felkapott fogalmai közé tartozik: erősen kapcsolódik a fejezetben tárgyalt virtualizációs és grid technológiákhoz. A számítási felhők hardverének alapját szerverfarmok adják, melyekben az ott elhelyezett szerverek gridbe kötve végzik a kiszolgálást. Az ilyen farmok több fronton is előnyt

---

<sup>87</sup> A számítási felhő angol elnevezése *computing cloud*, de sokkal gyakrabban haszált az ennek igénybe vételére utaló *cloud computing* megfogalmazás.

élveznek; egyrészt elhelyezésket tekintve megoldható olyan helyre történő telepítésük, ahol olcsó az elektromos áram (a szerverek darabja képes 2-3kW áramfelvételre, valamint a farm klimatizálása is komoly tényező – amennyiben szükséges), megfelelő a klíma (ezért népszerűek az északi országok farmépítésre, mivel nem kell klimatizálni, a külső levegő hőmérséklete egész évben elegendő a befogadó épületek hűtésére – innen Izland népszerűsége). Több esetben a keletkezett felesleges hőt vízmelegítésre és / vagy épületek (környező lakóépületek) fűtésére használják. A telepítési hely kiválasztása során még gyenge szempontként figyelembe célszerű venni a közvetlen üzemeltetés lehetőségét, mivel a közvetlen hardverkarbantartás ekkora gépszámok mellett rendszeres emberi beavatkozást igényel. Az előny másik forrása az üzemeltetés térbeli elkülönülése – azaz a szoftveres üzemeltetés gyakran más országban található (ide kapcsolódik az indiai call centerek / support centerek népszerűsége).

Számítási felhők használata esetén két lehetőség tárulkozik minden vállalat elé: egyfelől bérbe lehet venni erőforrásokat interneten keresztül, ezek nyilvánosak, rendkívül jól skálázhatóak, és több tanulmány szerint egyre inkább gazdaságos használatuk egyre kisebb és kisebb vállalatméretek esetén is (ilyen szolgáltatást nyújt is többek között az Amazon, a Google és a Microsoft is). A másik lehetőség a saját számítási felhő (másnéven private cloud) létrehozása. Kétségtelen, hogy a saját infrastruktúra felépítése és üzemeltetése nagyságrendekkel tud drágább lenni, és megtérülésére is csak nagyon hosszú (akár évtizedes) távon lehet számítani. A privát felhők építésére ugyanakkor több hajtóerő is létezik: bizalmas adatokat sok ország törvényei tiltják, hogy harmadik félre legyenek bízva – és ez kizárja a szolgáltatásként igénybe vehető felhők használatát. Ezen ok miatt építenek általában pénzügyi és biztosítók saját felhőket; ezek akkora méretben is képesek elkészülni, hogy skálázhatóságuk gyakorlatilag megfelel a szolgáltatásként igénybe vehető társaikkal ( - nem ritka a több ezer szervert tartalmazó farmok építése).

A számítási felhők az évezred elején még nagyon drága megoldásnak számítottak (bizonyos méretig olcsóbbnak számított a fizikai hardver beszerzése), de mára teljesen jó alternatívának számít ez a telepítési modell (pénzügyileg és technológiai oldalról egyaránt). Bizonyos esetekben a hardveren futó szolgáltatások felhőbe történő átültetése nem igényli a szolgáltatást nyújtó szoftver megváltoztatását.

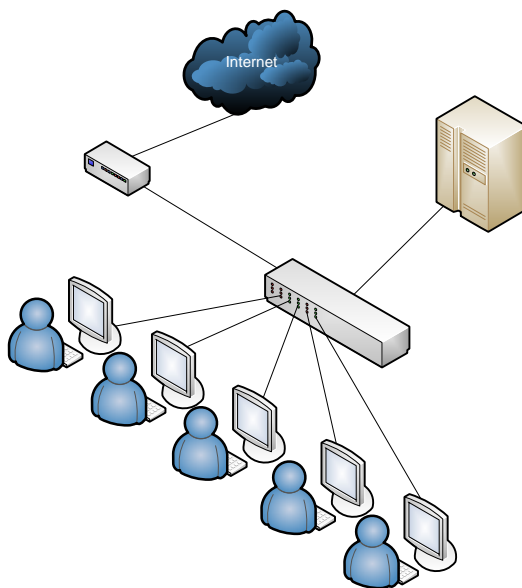
Létezik persze az a vállalati méret és szektor, ahol kifejezetten megéri a privát felhők építése – ez persze a multinacionális vállalatok legfelső hányada; tipikusan olyan vállalatok

lehetnek, melyek vagy valamilyen pénzügyi szolgáltatást nyújtanak, vagy nagyon erősen támaszkodnak az információtechnológiai megoldásaikra (gyártó és termelő vállalatok).

### 5.1.7 Vékony kliens – Vastag kliens

A vékony kliens technológia olyan kliens-szerver architektúra, mely elsősorban a központi szerver számítási kapacitására alapoz [35. ábra], és főként a felhasználó és a szerver közti információcserére fókuszál. Összehasonlításképpen a vastag kliens megoldásnál a kliens annyi számítási feladatot végez amennyit csak tud, így minimalizálva a szerverek felé forgalmazandó adatok mennyiségét (AGRAWAL, ANSARI ÉS TIWARI 2005).

A vékony kliens gépek általában csak web browsereket, vagy távoli asztali kapcsolódásra alkalmas szoftvert futtatnak, ezért jelentős terhelés jut a szerverekre. (Már forgalomban vannak olyan vékony kliensnek mondott gépek is, melyek saját operációs rendszert futtatnak, de ezeket inkább hibrid kliensnek nevezzük.)



*forrás: saját munka*

35. ábra: Vékony kliens hálózat

Ezek a vékony kliensek is PC-k [36. ábra], csak hiányzik belőlük több építőelem. Elterjedésük főként a nagysebességű és megbízható hálózatoknak köszönhető. A gyakorlatban egy Gb/s sebességű hálózat gyorsabb a PCI buszoknál, és sok merevlemeznél is. Éppen ezért helyezhetők át ezek az erőforrások máshová.

Ilyen architektúráknál a vékony kliensre a felhasználóval való kapcsolattartást szolgáló szoftver kerül (esetleg még a nagyon sűrűn használt programok egy része is) valamilyen hálózati operációs rendszerrel (ez betölthető háttértárról helyben vagy a indításkor a szerverről is). Mivel a terhelés a vékony kliensen kicsi, ezért lehetnek kicsik és alacsony fogyasztásúak, így bekerülésük és üzemeltetésük is jóval kedvezőbb. Így minden terhelés a néhány szerverre jut (programok futtatása, adatok tárolása, szolgáltatások) – mindez könnyebb menedzselhetőséget, és nagyobb biztonságot eredményez. Mivel minden adat egy helyre tárolódik, a biztonsági másolatok készítése / visszaállítás is sokkal gyorsabb és egyszerűbb művelet (*SURHONE, TIMPLEDON ÉS MARSEKEN 2009*).

A terhelés összehasonlításához egy átlagos PC körülbelül 5, egy nagyobb teljesítményű PC vagy szerver már akár 100, míg egy „high-end” szerver akár 700 vékony klienst is kiszolgálhat.

Használatukkal lényegesen egyszerűbbé válik a lemezképek (boot image) kezelése, mivel általában elegendő egyetlen kép többfajta igény kielégítéséhez, ráadásul mindez központilag kezelhetővé válik.

Mivel a klienseket majdnem teljesen a szerverről lehet konfigurálni, csökkennek az IT adminisztráció költségei, és csökken a hardver meghibásodásának lehetősége, mivel kevesebb hardverelem van használva (háttértárak híján bizonyos kártékony programoktól már eleve védve van a rendszer).

A rendszer beállítható úgy, hogy használat után semmi adat- vagy alkalmazástöredék nem marad a klienseken. Ez szintén növeli a biztonságot, és csökkenti az adatlopás esélyét.

Mivel minden művelet a szerveren folyik, ezért a kliensoldalon történő komoly hardverhiba vagy baleset esetén sem kell adatvesztéssel számolni.



forrás: Internet:HP.com, ChipPC.com

### 36. ábra: Vékony kliensek

Mivel a vékonykliensekben sem merevlemez, sem nagyméretű memóriái, sem erős processzor nincs, ezért önmagukban is olcsóbbak, valamint ezen alkatrészek hiánya miatt sokkal lassabban avulnak el. Az egyetlen olyan paraméterük, melyet nem a szerveroldal határoz meg, az a képernyő felbontás (ha a modell képernyővel egybeépített változat), de szintén nagyon lassan avul el. Míg a vastag klienseket átlagosan 3 évente cserélik (vagy hardverhibák kiküszöbölése miatt, vagy erőforrásaik elégtelensége miatt) addig a vékonykliensek még egy 10 éves periódus végén is ugyanolyan jól tudnak teljesíteni. Így még a szerverkövetelmények figyelembevételével is sokkal alacsonyabb költségvetés társul kivitelezésükhöz és üzemeltetésükhöz. Használatuk jobb CPU kihasználtságot is eredményez, mivel míg a vastag kliens megoldásoknál a processzor az ideje nagy részében tétlen, és annyi példányban kell annyi munkaállomás memóriájába tölteni a használt alkalmazást, ahányan használják, addig vékonykliensek esetén elég ezt a szerveren egyszer megtenni (BERTOLINI 2005).

Ez persze jelentős energia megtakarítással jár, nem csak azért, mert a vékony kliensek szervereikkel együtt is kevesebb áramot fogyasztanak, hanem mert hűtési igényük is sokkal kevesebb, ezért nem kell a légkondicionáló berendezéseket sem fejleszteni / üzemeltetni.



Ha mégis hardverhiba árnyékolja a képet (vékonykliens hibája), elegendő azt cserélni, ami olcsóbb, és a felhasználónak elegendő a munkát pár percre megszakítani, nem kell adatvesztéstől tartani.

Vannak speciális vékony kliensek, melyek képesek a legbarátságtalanabb környezetben is működni, mivel sem mozgó sem hűtendő alkatrészeket nem tartalmaznak (így kevesebb zajt termelnek).

Megemlítendő, hogy így csökkenhet a lopások gyakorisága is, mivel kisebb értékű tárgyakról van szó.

Elsőre ellentmondásnak tűnhet, hogy használatukhoz kevesebb hálózati erőforrás szükséges, de vastag kliensek esetén ha egy 10 MB méretű fájlt nyitunk meg, akkor azt átkerül a felhasználó gépére, majd mentéskor mindez vissza, így 20MB-nyi adatforgalom keletkezett, ami nyomtatás esetén duplázódhat is. Vékony kliensek használatával csak a billentyűzet, egér és képernyő változásai vándorolnak a hálózaton.

Szükség esetén sokkal könnyebb fejleszteni a rendszeren. Egyszerűbb és olcsóbb (néha leállással sem jár) a szerverhez erőforrásokat adni, legyen az tápegység, processzor, memória vagy háttértár. Ilyen esetekben nem is mindig keletkezik szemét. Kevesebb hardver-szemét pedig kevesebb kidobott nehéz fémeket jelent, ami kisebb környezeti terhelést okoz.

Általában a szerveroldali 70%-os terhelés mondható ideálisnak. Ekkor még egy nagyobb terhelési hullám lekezelhető (több felhasználó egyidejű nagy erőforrást igénylő munkája miatt). Ha ennél nagyobb az átlagos terhelés, akkor célszerű szerver-bővítéseken gondolkozni.

Bár nagyvállalati (sok munkaállomás és hálózat) környezetben nagyon sok érv szól a vékonykliens technológia alkalmazása mellett, illendő megemlíteni a vastag kliensek előnyeit is. Ide tartozik, hogy ebben csökkennek a szerver követelmények (elsősorban a memóriakövetelmények); jobb a multimédiás teljesítményük, különösen nagy sávszélességet igénylő műveleteknél (például audio/video szerkesztés); rugalmasabbak; sokkal nagyobb periféria ellátottság tartozik hozzájuk; és rosszminőségű/lassú hálózat esetén is jól használhatóak.

## 5.2 Szervezeti és személyi kérdések

Vezetői körökben axiómaként elfogadott, hogy a vállalat értékébe jelentősen számítanak az alkalmazottak (ezért a humán kapita optimalizáció minden vezető fontos célja). A munkaerő a szervezeti megújulás és innováció mozgató rugója. Kicsit egyszerűbben szólva a siker a munkaerő hatékonyságán múlik, azon, hogy mennyire támogatják és mennyire érzik sajátjuknak a vállalat misszióját, céljait és tervet. A cégek szignifikáns előnyre tehetnek szert a munkaerő és létszám optimalizálásával, különösen akkor, ha hosszú távra tervezve integrálják az üzleti stratégiába is (*GOREN 2008*).

Ezt a feltevést alátámasztandó, egy kutatás szerint (*MCKINSEY 2007*) a cégvezetők több mint 70%-a úgy gondolja, hogy „az innováció a növekedéshez vezető három legfontosabb prioritás közt van”. A kutatásból kiderül továbbá, hogy a vezetők szerint a legnagyobb kihívás megtalálni és megfelelő pozícióba „nevelni” azokat az embereket akik az innováció kulcsfigurái lehetnek. Kicsit távolabbról nézve az innovációnak, mint folyamatnak két fontos alappillére van a vállalatoknál: az egyik a vállalati kultúra, a másik az ott dolgozó emberek. Megállapítható, hogy az innováció, az agilitás és a kompetitív előnyök biztosítása mind feltételei a hatékony munkaerőnek és munkavégzésnek.

Míg egyre több cégvezető ismeri fel az üzleti intelligenciában és a fejlett analitikában rejlő lehetőségeket a teljesítmény növelése érdekében (a vásárlók, a költségek és a jövedelmezőség elemzésével mérhető a vásárlói életciklus, érték és jövedelmezőség termékekre nézve is, akár szektorokra bontva is), vagy az ellátási lánc optimalizálására, azonban meglepően kevesen ismerték eddig fel, hogy a munkaerő analizálásával egy hatékonyabb humán kapita stratégia alakítható ki, mely képes folyamatosan igazodni a felmerülő üzleti igényekhez [38. ábra].

A legtöbb cég jelenleg nem figyeli a kritikus dolgozóit, nem figyel arra, hogy valószínűsíthetően ki fogja elhagyni a céget és milyen okból – emiatt nem is alakítható ki hatékony munkaerő megtartó stratégia. Az Accenture tanácsadó cég által végzett felmérésből (*ACCENTURE 2006*) megtudható, hogy a megkérdezett vállalatok 40%-nál nincs semmilyen formális mód vagy lehetőség arra, hogy mérjék a HR és képzéseik hatását; másik 39%-nál van ugyan mérési pont, de csak néhány kezdeményezésre; a megkérdezettek majdnem fele gondolta úgy, hogy az üzleti-célú (bevételek és jövedelmezőség) metrikák alkalmazása erre a célra.

Ezekre alapozva sejthető, hogy a munkaerő analitika (Workforce Analytics) a hiányzó láncszem az üzleti stratégiákban. A legtöbb cél a célokat anélkül fokozza, hogy utánanézne, ki az a csapatában, aki éppen nyugdíj előtt áll, vagy éppen olyan kritikus időszak van az életében, ami elszakíthatja a cégtől. Így a készített stratégia nagyobb kockázatot hordoz magában. A célok eléréséhez a fókusz ebben az esetben célszerű kicsit a cégen felül helyezni, hogy biztosítsuk, hogy a megfelelő munkaerő a megfelelő helyen van, és a jövőben is ott lesz. A belső lehetőségek felderítésével – a munkaerő analitika bevezetésével és a vásárlói intelligenciához kapcsolva – kiegyensúlyozottabbá tehető az üzleti stratégia.

A mai egyre laposodó és szétterülő gazdasági környezetünkben a nagyvállalatok jó része globális, az egész világon jelen van, a hét minden napján és a nap minden órájában. Ennek egyik hozadéka a kiszervezés általánossá válása volt, méretben sokféleképpen, taktikai funkciótól kezdve akár teljes üzleti egységekig.

A szellemi tulajdon és az innováció az, ami fenntarthatja a kompetitív előnyöket, ezért lehet ez a munkaerő-megtartás kulcsa. A feladat pedig összetett: a gyereket vállalók egy része rövid időre kiesik a cégtől, a frissen végzettek pedig teljesen más prioritásokkal vágnak bele a munkába, ugyanakkor figyelni kell az idősödő szakemberek megtartására; ezért fontos nem csak a munkaerő megtartása, hanem a megszerzése is – így vihető kellő szakértelem, mélység és érték a vállalatokhoz.

Ebben mindenki számára megvan a megfelelő szerep:

- A menedzsereknek fel kell deríteniük a szükséges tudásbeli réseket, valamint biztosítaniuk kell az utat a feltörekvő, leendő vezetőknek;
- A pénzügyi szakembereknek meg kell határozni egy pozíció megüresedésének árát, az okozott túlóra mennyiségét, az esetleges kiszervezés lehetőségét, a toborzás menetét, a kulcsemberek pótlását, és ezek alapján kell kialakítani a stratégiát;
- A személyügyi szakembereknek (HR) fel kell deríteniük a változó munkaerőben rejlő trendeket és erre alapozva kell kialakítani a személyügyi stratégiát – mindezt a többi vezetővel együtt működve, hogy kellő időben lehessen változásokat kezdeményezni és munkaerőt felvenni.

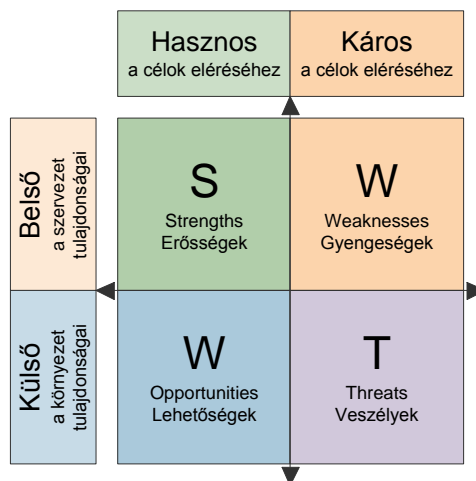
A munkaerő analitika segítségével kellő belátás nyerhető a kockázatokba és versenykérdésekbe, valamint mérhetővé válik a folyamatos fejlődés – hozzájárulva

mindezzel a cég céljaihoz. Így, kellően előretekintve a következő hozadéai lehetnek a technológia alkalmazásának:

- Javulhat a vállalat teljesítménye, az innováció, az agilitás és a kompetitív előnyök.
- Egyidejűen csökkenhetnek a költségek, és javulhatnak az eredmények.
- Nőhet a termelékenység és javulhat a csapatszellem.

Azok a vállalatok, akik alkalmazzák az új lehetőséget, kompetitív előnyökre tehetnek szert – persze csak akkor ha felfedik és megértik a munkaerejükben rejlő erősségeket, gyengeségeket, kockázatokat és veszélyeket.

A SWOT analízis olyan stratégiai tervezőeszköz, mely segít értékelni és felfedni azon tényezőket, melyek egy meghatározott cél elérése érdekében meghozandó döntés során felmerülhetnek. A SWOT analízis [37. ábra] tartalmazza az egyén belső és külső környezetének felmérését, így támogatva őt abban, hogy a legfontosabb témakörökre összpontosítson. Az elemzés célja a meghatározó belső és külső tényezők/adottságok (SWOT faktorok) azonosítása. Az elemzés során beszerzett információk két fő kategóriába: „belső” és „külső” tényezőkbe sorolhatók.



forrás: Kerekes – Kindler 1997 alapján saját rajz

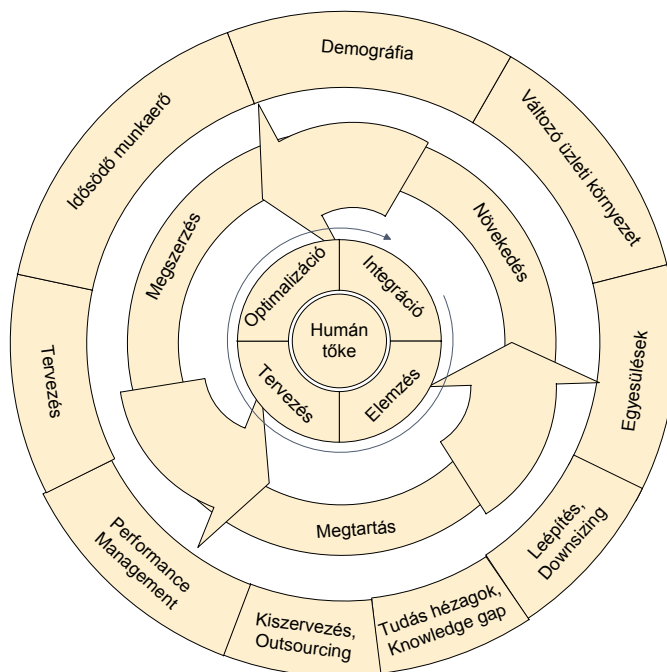
37. ábra: SWOT analízis

Bár az elemzés nem hívja fel a figyelmet arra, hogy melyek a stratégiaileg fontos kérdések, illetve nem kínál választ arra, hogyan dolgozzuk ki a stratégiánkat, viszont abban segít, hogy észrevegyük, a hatékony stratégia alapjául az erősségek szolgálnak, illetve hogyan használjuk ki a rendelkezésre álló lehetőségeket, hogyan küzdjük le, vagy minimalizáljuk a

gyengeségeket és a veszélyeket. Az előrejelzés és modellezés segítségével a munkaerő hatékonyabban felsorakoztatható az üzleti célok mögött; felfedhetővé válnak az előjövő képességbeli hiányok, így azokra fel lehet készülni. A folyamat során megérthetővé válik a szükséges munkaerő összetétele és mennyisége az üzletmenet fenntartásához illetve fokozásához.

A human tőke optimalizáció folyamata öt szakaszra bontható. Az első szakasz célja a munkaerő magas szintű célokhoz igazítása. Ennek során előre kell jelezni, hogy milyen képességekre lesz szükség az üzleti stratégia kivitelezéséhez, be kell gyűjteni a lehető legtöbb információt a jövőbeni döntések meghozatalához, a célokhoz igazodva kell kialakítani a munkaerő összetételét és meg kell állapítani, hogy mik a hiányzó képességek (talent gaps). Az üzleti intelligencia itt főként idősor elemzéssel és prediktív analitikával tud segíteni. A második szakasz célja, hogy meg legyenek címezve a munkaerővel kapcsolatos elvárások a teljes tehetség-ciklus kapcsán. Három vetületet kell figyelembe venni: a munkaerő megszerzése, a munkaerő növekedése (képzése), illetve a munkaerő megtartása. A megszerzéssel el kell érni, hogy a megfelelő alkalmazott, megfelelő képességekkel, a megfelelő időben és költségekkel, a megfelelő pozícióba kerüljön. A növekedési rész célja a már meglévő munkaerő képzése, hogy továbbfejlődhessen, és a jövőben kimagasló eredményt érhessen el. A megtartással kapcsolatban ugyanakkor figyelmet kell fordítani és fel kell készülni a változó demográfiai állapotokra és trendekre. Információtechnológiai támogatás itt trendszámítással adható. A harmadik szakasz eredményeként meg kell határozni a kockázatokat és mérsékelni kell azokat. Lépései a következők lehetnek: A múlt segítségével (de a jövőbe tekintve) meg kell határozni a trendeket, és okokat a felmondást, hiányzást és egyéb okokat illetően, meg kell határozni, a szervezeti változások hogyan érinthetik a munkaerőt, meg kell határozni, a jövőben hol várható üresedés, és hol lehet szükség új vezetőre, valamint meg kell érteni a munkaerőben rejlő keresleti és kínálati mintákat, ezt figyelembe véve kell az ellátásról gondoskodni és a stratégiát kialakítani. Az üzleti intelligencia itt trendszámítással, ok-okozat analízissel, valamint minta-analízissel tud segíteni. A negyedik szakasz célja felkészülni az üzletmenet változásaira (egyesülés, leépítés). Szükség lehet forgatókönyvek készítésére a különböző helyzetekhez és helyszínekhez, valamint döntéseket kell hozni, hogy miként lehetne a jól teljesítőket megtartani és a rosszul teljesítőket vagy hiányzókat kiszűrni. Az IT itt a scenario-készítésben és a döntéstámogatásban tud segítséget nyújtani. A végső – ötödik – szakasz célja az operatív és pénzügyi stratégiák szinkronizálása. Ennek során minden

alkalmazottra meg kell határozni a háttér információkat: hiányzások okai, túlórák, képzési költségek, létszámok, fizetések és egyéb juttatások, valamint igazolható álláspontot kell kialakítani azt illetően, hogy ezek a költségek hogyan hatnak a vállalatra. Itt az analitika és optimalizáció hívható segítségül.



forrás: hivatkozott cikk alapján saját munka

### 38. ábra: Humán tőke menedzsment

Látható, hogy a vállalati human tőke optimalizálásnak négy kulcstényezője van. Ezek: adatintegráció, elemzés, tervezés, optimalizáció (joggal fedezhetőek fel benne a mérnöki folyamatra utaló jelek).

- Adatintegráció, de nem csak a vállalaton belül, hanem harmadik féltől származó adatokkal és külső teljesítmény-értékeléssel dúsítva,
- SWOT analízis készítése részlegi és vállalati szinten a jelenlegi állapot és az elérhető célok megismerése céljából,
- Az üzleti célok és piaci trendek alapján stratégiai terv készítése, human capital scorecard és stratégiatérkép segítségével,
- A létrehozott forgatókönyvek segítségével a meglévő munkaerő optimalizálása, valamilyen alacsony kockázattal rendelkező környezetben a stratégia kipróbálása és validálása, majd éles alkalmazása.

Belátható tehát, hogy a vállalati növekedés szempontjából ugyanolyan fontos a jó munkaerő megtartása, mint a jó ügyfelek vagy vásárlók megtartása. Míg hagyományosan a humán kapita kérdése tipikusan HR feladatnak gondolt, belátható az is, hogy igenis célszerű az analitika bevezetése erre a területre is, és hogy hozzájárul az üzleti célok eléréséhez hosszú és rövidtávon egyaránt. Meg kell állapítani, hogy hogyan lehet olyan cégek kultúrát létrehozni, amely önmagában vonzó tényező a friss munkaerő számára és előrebocsátja a munkaerő javulását, fejlődését.

### **5.2.1 Vállalati hierarchia – BICC**

A BetterManagement.com közvélemény kutatására alapozva kijelenthető, hogy még a rengeteg energia befektetés ellenére is a cégek jelentős része még nem elégedett az információk döntésekhez való felhasználásának módjával, mivel a felmérés szerint a megkérdezettek 60%-a soha, ritkán vagy csak néha kapja meg a megfelelő információkat a döntéseik megalapozásához. És itt van a probléma gyökere: a megkapott információ nem az (vagy nem olyan), mint ami szükséges volna. Ennek egyik oka a nem egybefüggő BI rendszerek kialakítása, mely mindinkább információ raktárakat (*Information silo*) hoz létre, valamint a menedzsment is nehezebbé válik, mivel a több részleget nehezebb összefogni, koordinálni – persze szervezési és menedzselési szempontokból nem egyszerű egy olyan egységesített BI rendszer létrehozása sem, mely a vállalat majdnem minden területét érinti (*ING 2007*).

A sikerhez a vezetőségnek két dolgot kell figyelembe venni: az egyik, hogy az üzleti intelligencia mára kulcstényezővé vált – mint ahogyan az információ gyűjtése és felhasználása is – a másik annak felderítése, hogy milyen külső és belső információforrásaink lehetnek, amik a tény alapú döntési rendszerhez hozzájárulhatnak. Ezek után két tényezők múlik a siker: az összegyűjtött információ minőségén és az az azokat értelmező és felhasználó embereken [4. táblázat: Humán tőke optimalizáció]. Tapasztalatokból megállapítható, hogy minden cég számára üzleti intelligenciát négy helyre a legfontosabb bevezetni ezek: ügyfelekkel kapcsolatosan, pénzügyekkel kapcsolatban, az operatív munkával kapcsolatban és a kockázatok kapcsán (*MILLER, BRAUTIGAM ÉS GERLACH 2006*).

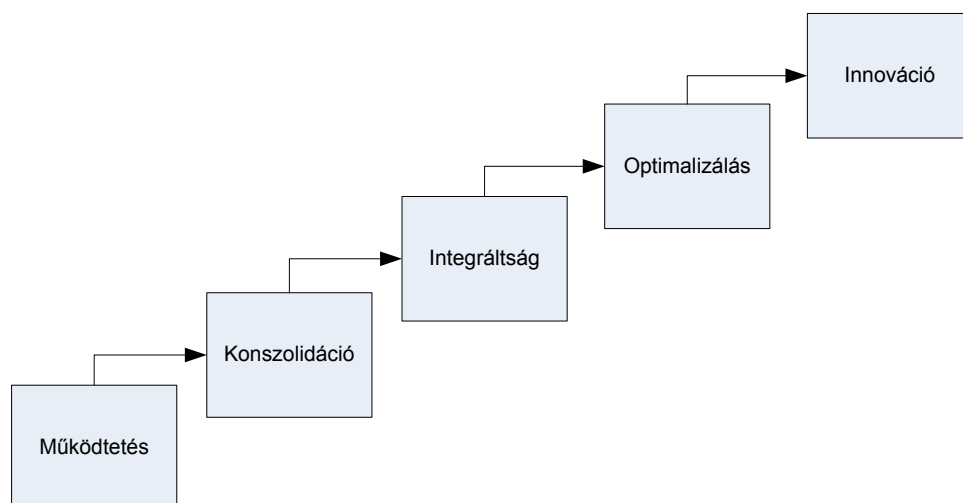
Ahhoz, hogy a vállalat e négy fontos szegmensébe bevezethető legyen az üzleti intelligencia adott vetülete, fontos, hogy legyen egy személy, aki rendelkezik a megfelelő

jogosultságokkal. Itt merülhet először fel a BICC (Business Intelligence Competency Center – Üzleti Intelligencia Kompetencia Központ) részleg alkalmazása, melynek feladata az üzleti intelligencia kidolgozása. Ez többről szól persze, mint egyszerűen a technológia bevezetése. Ugyanúgy fontosak az emberek, a folyamatok, az infrastruktúra és a kultúra.

Egy ilyen egység létrehozásához először azt kell meghatározni, hogy milyen fajta információra van szükségünk. Össze kell gyűjteni, hogy hol (melyik szinten), milyen üzleti döntéseket kell meghozni napi, heti, havi, negyedéves vagy egyéb rendszerességgel. Azt is meg kell határozni, hogy a majd létrejövő információkat hogyan és kinek kell eljuttatni – ezt nagyon sok minden befolyásolhatja, a vállalat típusától és méretétől, a felépítésén keresztül a az információkezelésig (Howson 2007).

Figyelembe kell azonban venni, hogy a cég intelligencia érettségi szempontból a skála melyik fokán áll. A skála az Információ Evolúciós Modell, öt foka pedig a következő [39. ábra]:

- Működtetés: BI rendszer működik.
- Konszolidáció: Az információk osztály/részleg szinten vannak tárolva.
- Integráltság: a vállalati szabványok rendelkezésre állnak és használatban vannak.
- Optimalizálás: az információ rendelkezésre áll a mérések elvégzésére és javításokra.
- Innováció: Az innováció folyamatos a cégnél, a vállalati kultúra része.



forrás: saját munka

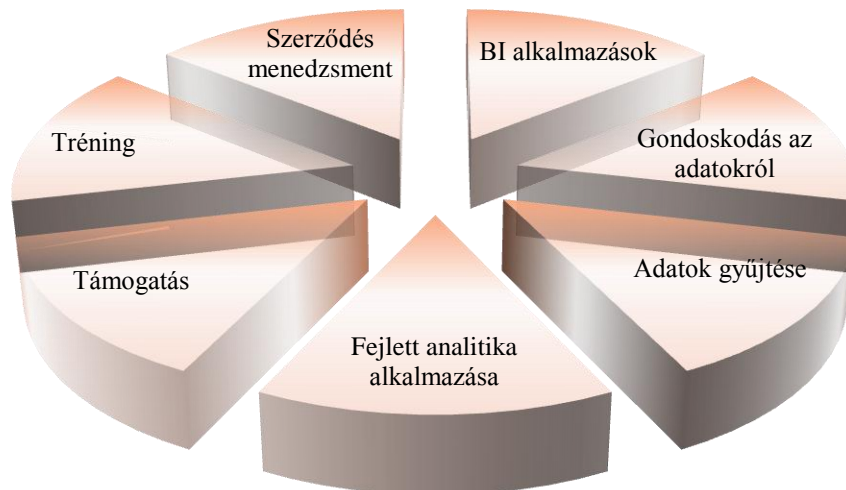
39. ábra: Információ evolúciós modell



Mindezen szinteknek persze van emberi, folyamati, infrastruktúra és kulturális oldala. A BICC létrehozása persze nem függ attól, hogy hol áll a vállalat ezen a skálán, de meghatározhatja annak szerkezetét (DAVIS, MILLER ÉS RUSSEL 2006).

A kezdetleges BICC tartalmazhat csak támogató funkciókat, de több tanácsadó cég szerint ajánlható, hogy az alábbi aspektusokat is vegyük figyelembe [40. ábra]:

- BI programok: melyek felelnek meg legjobban a stratégiának és melyeket célszerű alkalmazni.
- Gondoskodás az adatokról: Az adatok minőségéért, irányításáért és a szabályzatokért a BICC felel.
- Adatok gyűjtése: adat integráció, ETL és tárolás.
- Fejlett analitika alkalmazása
- Támogatás: a vállalat felhasználói felé
- Tréning: változás menedzsment, projekt menedzsment és folyamatintegráció
- Partnerszerződés menedzsment.



forrás: saját munka

40. ábra: A BICC aspektusai

Fontos, hogy a BICC beépülhessen a vállalati hierarchiába. A lehető legjobb, ha közvetlenül egy vezető alá tartozik a részleg – az kevésbé fontos, hogy CEO<sup>88</sup>, CIO<sup>89</sup> vagy CFO<sup>90</sup> az illető. Lényeges szempont, hogy itt találkozhassanak a technikai és üzleti szempontok. A részleg összetételét illetően fontos, hogy egyaránt jelen legyenek benne az üzleti és az informatikai emberek.

Tipikus betöltendő gazdasági/üzleti munkakörök az üzleti elemző, az iparági specialista, projekt menedzserek. Tipikus technológiai szerepkörök a technikai vezető, az adatbányászok, az adattárház tervezőmérnök, a szoftverfejlesztő és a konzulens. ezeken kívül persze szükség van olyan kommunikációs szakemberekre is, akik a változásokat, amin a cég keresztülmegy, megfelelően kommunikálják a vállalat dolgozói, partnerei és ügyfelei felé. Sok menedzser számára ez a részleg jelentheti a kezdeti munkakört, valamint idővel ugródeszkeként szolgálhat egyéb vezetői pozíciók felé. Ennek oka az elsősorban, hogy itt összefutnak a folyamatok, valamint kellő rálátás alakul ki az emberekben a vállalat egészére nézve (*MILLER, BRÄUTIGAM, GERLACH ÉS DAVIES 2005*). A BICC teljes mérete kedvező esetben egy tucat fő körül alakul, ám hatalmas vállalatok esetén elérheti a száz főt is (*SCHEPS 2008*).

A BICC (vezető) menedzser feladata a BI rendszer céljainak egyeztetése a vállalati stratégiával, valamint a kellő vezetői támogatás biztosítása.

Az adattárház tervező mérnök (Warehouse Architect) feladata a kivitelezésre kerülő megoldás megtervezése, beleértve ebbe a teljesítményi megfontolásokat, az adatsémák kidolgozását, a tárolás módjának kidolgozását, a repozitórium megtervezését, és az objektumok azonosítását.

Az üzleti elemzőnek (Business analyst) kell értenie részletesen a vállalat folyamatait és szabályait, ismeri azt, hogy a vállalat miként használja az adatokat, beleértve ebbe az adatok manipulálásának módját és szabályait.

A technikai konzulens: (Technical Consultant) biztosítja a kiválasztott BI rendszer megfelelő implementációját, ő fogalmazza meg a technikai, biztonsági, összekapcsolhatósággal kapcsolatos követelményeket.

---

<sup>88</sup> CEO (Chief Executive Officer): A legfőbb tisztségviselő, gyakran a vezérigazgató.

<sup>89</sup> CIO (Chief Information Officer): Az informatikai vezető.

<sup>90</sup> CFO (Chief Financial Officer): A pénzügyi vezető.

Projekt menedzser felügyeli a team napi feladatainak ütemezését, jelentéseket készít a támogatók felé, ő felügyeli az erőforrások elosztását a projektben. Ilyen erőforrás lehet az ember, a pénz, a szoftver, a helység vagy egyéb más felszerelés.

Az analitikai konzulens: (Analytical Consultant) feladata az alkalmazott modellek és jelentések ismerete, valamint a felmerülő üzleti problémákra analitikai megoldások képzése. Segítenie kell a megfelelő BI megoldás kidolgozásában, valamint a vállalatban belüli felhasználóknak a rendszerrel kapcsolatos feladatok nehézségeiben.

Az adatkezelő: (Chief Data Steward) kezeli az adatokkal kapcsolatos problémákat az egész vállalat területén. Feladata azon rutinok kialakítása, mely az adatok keletkezésével, gyűjtésével, terjesztésével, kezelésével, manipulálásával és megőrzésével kapcsolatos. Mivel az adatok cégen belüli származásáért valakit felelőssé kell tenni, célszerű kerülni az adattulaj (Owner) kifejezést, és ajánlott az adatkezelő használata – ez kevesebb feszültséget eredményez. Főként menedzseri feladatkör, itt nem szükséges a technológia mélyreható ismerete (*SABHERWAL ÉS BECERRA-FERNANDEZ 2010*). Általában az alábbi felelősségeket jeleni:

- Üzleti definíciók és szabványok: A rendelkezésre álló információk folyamatos értelmezése és azok integrálási lehetőségeinek kidolgozása.
- Információminőség: a pontosságot, konzisztenciát, időbeliséget, validitást és teljességet illetően.
- Információ védelem: a biztonsági és adatvédelmi szabályok betartatása.
- Információ életciklus: a feladat az adatok keletkezésétől / felfedezésétől azok gyűjtésén és felhasználásán át egészen azok használaton kívül helyezéséig és végső tárolásáig tart.

Ez a funkció (kellően nagy szervezeti méret esetén) több szinten is értelmezhető:

- Stratégiai szint: az információs stratégia és tervek 3-5 évre előre kidolgozása; a korábbi tervek felülvizsgálata.
- Operatív szint: Információ menedzsment terv kidolgozása a stratégiának alárendelve.
- Taktikai szint: Információs folyamatok fejlesztése, szállítása és karbantartása a vállalati információ menedzsment támogatása érdekében.

Továbbá felelős döntéshozó(k) (Executive Decision Makers): azon személy (vagy személyek csoportja), aki feladata eldönteni, hogy melyik adat fontos a döntési célok érdekében. Tagjai általában a felső vezetés köréből kerülnek ki annak ellenére, hogy ott már hagyományosan nem téma az adatok eredetéről, felügyeletéről vagy a stratégiához való viszonyáról beszélni, ám mivel nem lehet minden adat megfelelő, az ő feladata dönteni. Másodrészt az ő feladatuk az adatokkal kapcsolatos beruházások elfogadása és megtérüléseik felügyelete.

Az információs képviselő szervezet (Analytical Data Advocate) vagy más néven információ menedzserek (IM Group). Elsősorban olyan vállalatoknál van szerepe, ahol nem BICC jellegű szervezeti egység található, hanem annál némiképp tágabb spektrumot felölelő tevékenység (tehát nem szigorúan informatikai jellegű feladatok és megközelítés). Így feladatai valamelyest redundánsak lehetnek: stratégiafejlesztés, modellképzés és döntéshozatal. Bővebb értelemben ide tartozhat az adattárházak, adatbányászati és BI eszközök felügyelete, de a látókör tovább szélesedhet, egészen az IT infrastruktúra felügyeletéig.

Értelemszerűen a BICC és az IM csoport megléte kizárja egymást, ám nem lehet eléggé hangsúlyozni, mennyire fontos is az IT és az üzleti döntések egymáshoz közel tartása.

## 6 Irodalomjegyzék

1. Accenture: The high-performance workforce study 2006: the lack of business connections. Accenture, 2006.
2. Adam, Frédéric – Sammon, David: The enterprise resource planning decade: lessons learned and issues for the future. Idea Group Inc (IGI), 2004.
3. Agrawal, Neeraj – Ansari, Nasimuddin – Tiwari, Shekhar: Practical Handbook Of Thin-Client Implementation. New Age International, 2005.
4. Allan, Russell: You can't do that with BI- or can you? SASCOM, 2007
5. Allen, Sharon – Terry Evan: Beginning relational data modeling. Apress, 2005.
6. Allison, Graham T.: Conceptual Models and the Cuban Missile Crisis. American Political Science Review, 1969.
7. Álmos Attila – Győri Sándor – Horváth Gábor – Várkonyiné Kóczy Annamária: Genetikus algoritmusok. Typotex, 2002.
8. Alonso, Gustavo – Dadam, Peter – Rosemann, Michael: Business process management. Springer, 2007.
9. Anandarajan, Murugan – Anandarajan, Asokan – Srinivasan, Cadambi A.: Business intelligence techniques: a perspective from accounting and finance. Springer, 2003.
10. Anderson, James A.: An Introduction to Neural Networks. MIT Press, 1995.
11. Andics Jenő: A technikai haladás társadalmi problémái a gazdasági szervezetekben. Akadémiai Kiadó, 1977.
12. Anke Meyer-Base: Pattern recognition for medical imaging. Academic Press, 2004.
13. Arabie, Phipps – Hubert, Lawrence J. – Soete Geert de: Clustering and classification. World Scientific, 1996.
14. Armstrong, Michael – Baron, Angela: Performance management: the new realities. Institute of Personnel and Development, 1998.
15. Bacsokay Zoltán: Ökonómiai elemzési módszerek a mezőgazdaságban. Mezőgazdasági Kiadó, 1984.

16. Baier, Daniel – Decker, Reinhold – Schmidt-Thieme, Lars – Gaul Wolfgang: Data analysis and decision support. Springer, 2005.
17. Bálint János: Gazdasági előrejelzés idősor elemzéssel. Kertgazdaság Online. [www.kertgazdasag.hu](http://www.kertgazdasag.hu). Budapest. 2002.
18. Ball, G.H. – Hall, D.J.: Isodata, a novel method of data analysis and classification. Stanford, 1965.
19. Bánéri István Dr. – Gyurcsókné Dr. Lajtha Ildikó – Dr. Kiss János – Szucsákné Dr. Ónódi Ágnes: Kereskedelemszervezés és technológia. Budapest, KVIF, 1999
20. Banks, David L.: Classification, clustering, and data mining applications. Springer, 2004.
21. Barkaway, David: How to migrate your data in 7 steps. SASCOM, 2009/1.
22. Bartenstein, Oskar: Web knowledge management and decision support. Springer, 2003.
23. Baschab, John – Piot, Jon: The executive's guide to information technology. John Wiley and Sons, 2007.
24. Beniger, James R.: Az irányítás forradalma - az információs társadalom technológiai és gazdasági forrásai. Gondolat–Infonia 2004.
25. Bergeron, Bryan P.: Essentials of CRM: a guide to customer relationship management. John Wiley and Sons, 2002.
26. Berkeley, Dina – Widmeyer, George – Brezillon, Patrick – Rajkovic, Vladislav: Context Sensitive Decision Support Systems. Chapman&Hall, 1998.
27. Berman, Fran – Fox, Geoffrey – Hey, Anthony J. G.: Grid computing: making the global infrastructure a reality. John Wiley and Sons, 2003.
28. Beroggi, Giampiero E. G.: Designing and evaluating e-management decision tools. Springer, 2004.
29. Berry, Michael J. – Berry, Michael W. – Kogan, Jacob: Text Mining: Applications and Theory. John Wiley and Sons, 2010.
30. Berry, Michael W.: Survey of text mining: clustering, classification, and retrieval. Springer, 2004.
31. Bertolini, Phil: Thinking Thin. CIO, 2005/1
32. Bezdek, J.C.: Pattern recognition with with fuzzy objective function algorithms. Plenum, 1981.

33. Biere, Mike: Business intelligence for the enterprise. Prentice Hall PTR, 2003.
34. Bligh, Philip – Turk, Douglas: CRM unplugged: releasing CRM's strategic value. John Wiley and Sons, 2004.
35. Blokdijk, Gerard: Business Intelligence 100 Success Secrets - 100 Most Asked Questions. Lulu.com, 2008.
36. Bodlaki Tamás: Bevezetés az operációs rendszerekbe. PTE, 2004.
37. Bodlaki Tamás: Bevezetés az operációs rendszerekbe. PTE, 2009.
38. Bolen, Alison: Predicting the future of business intelligence. SASCOM, 2007.
39. Bordáné Rabóczki Mária: Nyereségtervezés és –elemzés – vállalatvezetés. Közgazdasági és Jogi Könyvkiadó, 1986.
40. Bostrom, Ann – French, Steven P. – Gottlieb, Sara J.: Risk Assessment, Modeling and decision support. Springer, 2008.
41. Box, George E. P. – Jenkins, Gwilym M. – Reinsel, Gregory C.: Time series analysis: forecasting and control. John Wiley, 2008.
42. Breinerné Varga Ildikó: A távmunka humánpolitikája. Emberi erőforrásémenedzsment. 2004/6.
43. Breshow, L.A. – Aha, D.W.: Simplifying decision trees: A survey. Naval Research Laboratory, 1997.
44. Bromwich, Adam T.: PeopleSoft HRMS reporting. Prentice Hall PTR, 1999.
45. Brooks, Frederick P. Jr.: No Silver Bullet-Essence and Accidents of Software Engineering. Nick Heap: Information technology and society: a reader. SAGE, 1995.
46. Bryman, Alan – Hardy, Melissa A.: Handbook of Data Analysis. SAGE Publications Ltd, 2009.
47. Burstein, Frada – Holsapple, Clyde W.: Handbook on decision support systems. Springer, 2008.
48. Business Intelligence Limited: Evaluating CRM Trends: An Assessment of Key CRM Strategies, Implementation, Technologies, Management and Measurement. Business Intelligence Limited, 2008.
49. Bussler, Christoph: Business intelligence for the real-time enterprises. Springer, 2007.

50. Bussler, Christoph: Business intelligence for the real-time enterprises. Springer, 2007.
51. Butler Group: The intelligent enterprise: competitive use of intelligence in business : in-depth report.competitive use of intelligence in business. Butler Group, 2001.
52. Buytaert, Kris – Dittner, Rogier – Rule, David: The best damn server virtualization book period. Syngress, 2007.
53. Castenallos, Malu – Dayal, Umesh – Miller, Renee: Enabling Real-Time Business Intelligence. Springer, 2010.
54. Castenallos, Malu – Dayal, Umesh – Sellis Timos: Business Intelligence for the Real-Time Enterprise. Springer, 2009.
55. Ceruzzi, Paul E.: A history of modern computing. MIT Press, 2003.
56. Chambers, Ray L. – Skinner, C. J.: Analysis of survey data. John Wiley and Sons, 2003.
57. Collier, Marsha: EBay for Dummies. For Dummies 2009.
58. Collins, Keith – Goodnight, Jim – Hagström, Mikael –Davis, Jim: The Future of business intelligence. SASCOM 2007/1.
59. Collins, Keith– Goodnight, Jim Dr. – Hagström, Mikael– Davis, Jim: The Future of business intelligence. SASCOM, 2009/2.
60. Costa, Joan E. Ricart i –Subirana, Brian –Valor-Sabatier, Josep –Valor, Josep: Sources of information value: strategic framing and the transformation of the information industries. Palgrave Macmillan, 2004.
61. Crissey, Mary: Text mining reveals hidden value and opportunities. SASCOM, 2007/1.
62. Davalo, Éric – Naim, Patrick: Neural networks. Macmillan Education, 1991.
63. Davenport, Thomas – Prusak, Laurence: Working knowledge: how organizations manage what they know. Harvard Business Press, 2000.
64. Davenport, Thomas H. – Harris, Jeanne G. – Morison, Robert: Analytics at Work: Smarter Decisions, Better Results. Harvard Business Press, 2010.
65. Davenport, Thomas H. – Prusak, Laurence: Working Knowledge. Kossuth, 2001.
66. Davenport, Thomas H. –Harris, Jeanne G.: Competing on analytics: the new science of winning. Harvard Business Press, 2007.



67. Davis, Jim – Miller, Gloria J. – Russell, Allan: Information revolution: using the information evolution model to grow your business. John Wiley and Sons, 2006.
68. Dr. Abonyi János: Adatbányászat, a hatékonyság eszköze. Computerbooks, 2006.
69. Dr. Bodon Ferenc: Adatbányászati algoritmusok. BME, 2010.
70. Dr. Bodon Ferenc: Adatbányászati algoritmusok. Budapest, 2010.
71. Dr. Élő Gábor – Dr. Szabó József – Galli Richárd: Vállalatvezetést támogató üzleti intelligencia. Győr, 2011.
72. Dr. Élő Gábor – Galli Richard Péter: Vállalatvezetést támogató üzleti intelligencia. Győr, 2012.
73. Dr. Kiss Ferenc – Rácz Csaba: Szervezeti kihívások – informatikai megoldások. BME GTK, 2003.
74. Dr. Kiss Imre: Az üzleti informatika elmélete gyakorlatban. Információs Társadalomért Alapítvány, 2007.
75. Dr. Pap István: Bevezetés a menedzsmentbe- GAMF, 1994.
76. Dr. Papp Ottó: Projektmenedzsment a gyakorlatban. LSI, 2002.
77. Dr. Raffai Mária: Az információ: szerep, hatás, menedzsment. Palatia, 2005.
78. Durányik Nóra: A teljesítmény-elvű vezetés és a public relations értéke. Budapesti Gazdasági Főiskola, 2007.
79. Dyché, Jill: e-Data: turning data into information with data warehousing. Addison-Wesley, 2000.
80. Dyché, Jill: The CRM handbook: a business guide to customer relationship management. Addison-Wesley, 2002.
81. Edward, Chris – Ward, John – Bytheway, Andy: Az információs rendszerek alapjai. Budapest, Panem, 1999
82. Elliott, Jane – Marsh, Catherine: Exploring data: an introduction to data analysis for social scientists. Polity, 2009.
83. Enders, Craig K.: Applied Missing Data Analysis. Guilford Press, 2010.
84. Ensor, Dave – Norgaard, Mogens – Gorman, Tim – Kolk, Anj: Oracle insights: tales of the Oak Table. Apress, 2004.

85. Enyedi Miklósné dr.: Az etika szerepe a döntéshozatalban. Budapesti Műszaki Főiskola, 2004.
86. Erl, Thomas: SOA: principles of service design. Prentice Hall, 2008.
87. Ewusi-Mensah, Kweku: Software development failures: anatomy of abandoned projects. MIT Press, 2003.
88. Fajszí Bulcsú – Cser László – Fehér Tamás: Üzleti haszon az adatok mélyén. Az adatbányászat mindennapjai. Alinea Kiadó, 2010.
89. Fang, Liping – Hipel, Keith W. – Kilgour, D. Marc: Interactive decision making. Wiley, 1993.
90. Feldman, Ronen – Sanger, James: The text mining handbook: advanced approaches in analyzing unstructured data. Cambridge University Press, 2007.
91. Feng, Xue: The Road to Decision-Centric Business Intelligence. International Conference on Business Intelligence and Financial Engineering, 2009.
92. Fernández, George: Data mining using SAS applications. Chapman & Hall/CRC, 2003
93. Finnegan, David Jesse – Willcocks, Leslie P.: Implementing CRM: from technology to knowledge. John Wiley and Sons, 2007.
94. Fisher, D.: Knowledge acquisition via incremental conceptual clustering. Machine Learning, 1987.
95. Fodor János – Roubens, Marc: Fuzzy Preference Modeling and multicriteria decision support. Kluwer, 1994.
96. Forgy, E.: Cluster analysis of multivariate data: Efficiency versus interpretability of classification. Biometrics, 1965.
97. Foster, Ian – Kesselman, Carl: The grid: blueprint for a new computing infrastructure. Morgan Kaufmann, 2004.
98. Franke, Jürgen – Nakhaeizadeh, Gholamreza – Renz, Ingrid: Text mining: theoretical aspects and applications. Physica-Verlag, 2003.
99. Friedman, Ken: The Decision Tree. Heart Pub., 1996.
100. Füstös László – Kovács Erzsébet – Meszéna György – Sinonné Mosolygó Nóra: Alakfelismerés (Sokváltozós statisztikai módszerek). UNK, 2004.
101. Füstös László – Kovács Erzsébet: A számítógépes adatelemzés statisztikai módszerei. Tankönyvkiadó, 1989.

102. Füstös László – Meszéna György – Simonné Mosolygó Nóra: Térstatisztika: három- és többdimenziós adattömbök elemzési módszerei. Aula Kiadó, 1997.
103. Gachet, Alexandre: Building Model Driven Decision Support Systems with Dicoless. VDF, 2004.
104. Galli Richárd Péter (2010) – Méretek és technológiák - hatékonyság a főszerepben. V. KHEOPS Konferencia, Mór, 2010.
105. Galli Richard Péter: A válság és az információtechnológia kölcsönhatása, Világméretű pénzügyi és hitelpolitikai válság és Magyarország Tudományos Konferencia, Komárom, 2009.
106. Galli Richard Péter: Az IT infrastruktúra fejlesztése és a vállalati költségcsökkentés, IV. Kheops Tudományos Konferencia, Mór, 2009.
107. Galli Richard Péter: Már alkalmazott technológiák és BI megoldások összehangolása. Híradástechnika, 2011.
108. Galli Richard Péter: Nagyvállalatok; Az informatikai lépéstartás és a költségmegtakarítás lehetőségei, Versenyképesség - Változó menedzsment / Marketing Tudományos Konferencia, Székesfehérvár, 2008.
109. Galli Richard Péter: Paralel hatékonyságnövelés és költségcsökkentés a stratégia és a hierarchia tekintetében. Alma Mater, 2011.
110. Galli Richard Péter: Possibilities for data management and online business at small and midsized businesses. FIKUSZ 2011.
111. Galli Richard Péter: SAS Egyetemi program belülről - üzleti intelligencia oktatási tapasztalatok., XIX. SZÁMOKT - Számítástechnika az oktatásban Tudományos Konferencia, Románia, Marosvásárhely, 2009.
112. Galli Richard Péter: Software technology: Specialized development / launch method for BI applications, XX. SZÁMOKT 2010- Számítástechnika és oktatás Konferencia, Románia, Szatmárnémeti, 2010.
113. Galli Richard Péter: Taxonomy for choosing BI systems into an already running infrastructure. Acta Technica Jaurinesis, 2012.
114. Galli Richard Péter: Using quality information in analysis – with experiment. ITS2011, 2011.
115. Galli Richard: A hitelválság hatása az informatikai stratégiákra. SZE, 2008
116. Galliers, Robert – Leidner, Dorothy E.: Strategic information management: challenges and strategies in managing information systems. Elsevier, 2004.

117. Galushkin, Aleksandr Ivanovich: Neural networks theory. Springer, 2007.
118. Garg, Vinod Kumar – Venkitakrishnan, N. K.: Enterprise resource planning: concepts and practice. PHI Learning Pvt. Ltd., 2004.
119. Gelinas, Ulric J. – Dull, Richard B.: Accounting Information Systems. Cengage Learning, 2007.
120. Gentle, Michael: The CRM project management handbook: building realistic expectations and managing risk. Kogan Page Publishers, 2002.
121. Georg Simmel: Soziologie. Duncker & Humblot, 1908.
122. Gilchrist, Guy: Superkernel: fun's a poppin'. Xerox Education Publications, 1978.
123. Gilleland, Jeff– Heath, Karen: Using your data to know your customers. SASCOM, 2009/3.
124. Gisbert, Stoyan: MATLAB. Typotex Kft, 2005.
125. Golden, Bernard: Virtualization for Dummies. For Dummies, 2007.
126. Goldenberg, Barton J.: CRM Automation. Prentice Hall PTR, 2002.
127. Goldenberg, Barton J.: CRM in Real Time: Empowering Customer Relationships. Information Today, Inc., 2008.
128. Gorban, Alexander N. – Kègl Balázs – Wunsch, Donald C.: Principal manifolds for data visualization and dimension reduction. Springer, 2007.
129. Goren, Becca: Success – from the inside out. SASCOM, 2008/2.
130. Grant, Tina: International Directory of Company Histories, St. James Press, 2006. 76. kötet.
131. Greenberg, Paul: CRM at the speed of light: essential customer strategies for the 21st century. McGraw-Hill Professional, 2004.
132. Guha, S. – Rastogi, R, - Shim K.:Cure: an efficient clustering algorithm for large databases. Seattle, 1998.
133. Guha, S. – Rastogi, R, - Shim K.:Rock: a robust clustering algorithm for categorical attributes. Sydney, 1999.
134. Gupta, Jatinder N. D. –Sharma, Sushil Kumar: Intelligent enterprises of the 21st century. Idea Group Inc (IGI), 2004
135. Gurney, Kevin: An introduction to neural networks. UCL Press, 1997.

136. Guyon, Isabelle: Feature extraction: foundations and applications. Springer, 2006.
137. Hamilton, James Douglas: Time series analysis. Princeton University Press, 1994.
138. Hammergren, Thomas C. – Simon, Alan R.: Data Warehousing for Dummies. For Dummies, 2009.
139. Han, Jiawei – Kamber, Micheline: Adatbányászat koncepciók és technikák. Panem, 2004.
140. Han, Jiawei – Kamber, Micheline: Data mining: concepts and techniques. Morgan Kaufmann, 2001.
141. Hand, D. J. – Mannila, Heikki – Smyth, Padhraic: Principles of data mining. MIT Press, 2001.
142. Harris, Roy D.: MRP: a materials requirements planning model. Dept. of Management, College of Business Administration and Graduate School of Business, University of Texas at Austin, 1983
143. Hartwig, Frederick – Dearing, Brian E.: Exploratory data analysis. Sage Publications, 1979.
144. Hassoun, Mohamad H.: Fundamentals of artificial neural networks. MIT Press, 1995.
145. Havey, Michael: Essential business process modeling. O'Reilly Media, Inc., 2005.
146. Hay, David C.: Data model patterns: a metadata map. Morgan Kaufmann, 2006.
147. Henderson, Harry: Encyclopedia of computer science and technology. Infobase Publishing, 2009.
148. Hertz, J. – Krogh, A. – Palmer, R.: Introduction to the theory of neural computation. Santa Fe, 1991.
149. Herzberg, Frederick: One more time: how do you motivate employees? Harvard Business Review vol. 46/1, 1968.
150. Heteyi József: Vállalatirányítási rendszerek Magyarországon 2. ComputerBooks, Budapest, 2000.
151. Heteyi József: Vállalatirányítási rendszerek Magyarországon. ComputerBooks, Budapest, 1999.
152. Higgins, Paul – Le Roy, Patrick – Tierney, Liam: Manufacturing planning and control: beyond MRP II. Springer, 1996.
153. Hodge, Gail M.: Understanding metadata. NISO Press, 2004.

154. Hosmer, David W. – Lemeshow, Stanley: Applied logistic regression. John Wiley and Sons, 2000.
155. Hosni, Yasser A. – Khalil, Tarek M.: Management of technology: opportunities and challenges for developed and developing regions of the world : selected papers from the Eleventh International Conference on Management of Technology. Emerald Group Publishing, 2004.
156. Howard, Philip: A Data Migration. Bloor research, 2007
157. Howson, Cindi: Successful Business Intelligence: Secrets to Making BI a Killer App. McGraw-Hill Professional, 2007.
158. Hsiao, Cheng: Analysis of panel data. Cambridge University Press, 2003.
159. Illangasekare, Tissa H. – Mahutova, Katarina – Barich, John J.: Decision Support for Natural Disasters and Intentional Threats to Water Security. Springer, 2009.
160. Ing, Steven: A strategic approach to intelligence. SASCOM. 2007/3.
161. Inmon, William H. – O’Neil, Bonnie K. – Fryman, Lowell: Business metadata: capturing enterprise knowledge. Morgan Kaufmann, 2008.
162. International Business Machines Corporation: Highlights of IBM history. International Business Machines Corp., 1973
163. Jarke, Matthias: Fundamentals of data warehouses. Springer, 2003.
164. Jávor András – Benkő Tiborné: Diszkrét rendszerek szimulációja. Műszaki Könyvkiadó, 1979.
165. Jensen, Finn V. – Nielsen, Thomas Dyhre: Bayesian networks and decision graphs. Springer, 2007.
166. Joachims, Thorsten: Learning to classify text using support vector machines. Springer, 2002.
167. Joseph, Joshy – Fellenstein, Craig: Grid computing. Prentice Hall PTR, 2004.
168. Kaplan, Robert S. – Cooper, Robin: Cost & effect: using integrated cost systems to drive profitability and performance. Harvard Business Press, 1998.
169. Karayiannis, Nicolaos B. – Venetsanopoulos, Anastasios N.: Artificial neural networks: learning algorithms, performance evaluation, and applications. Springer, 1993.
170. Karypis, G: - Han E.H. – Kumar V.:Chameleon: a hierarchical clustering algorithm using dynamic modeling. Computer, 1999.

171. Kaufmann, L. – Rousseeuw, P.: Finding groups in data: an introduction to cluster analysis. Wiley, 1990.
172. Kerekes Sándor – Kindler József: Vállalati környezet-menedzsment. Aula, 1997.
173. Kerékgyártó Györgyné – Mundruczó György – Sugár András: Statisztikai módszerek és alkalmazásuk a gazdasági, üzleti elemzésekben. Aula, 2001
174. Khan, Arshad: SAP and BW Data Warehousing: How to Plan and Implement: iUniverse, 2005.
175. Khosrowpour, Mehdi: Managing information technology resources and applications in the world economy. Idea Group Inc (IGI), 1997.
176. Kis Endre: Szolgáltatások és folyamatok életciklusa. Computerworld, 2010 április 13.
177. Kleinbaum, David G. – Klein, Mitchel: Logistic Regression: A Self-Learning Text. Springer, 2010.
178. Koh, Yun Sing – Rountree, Nathan: Rare Association Rule Mining and Knowledge Discovery: Technologies for Infrequent and Critical Event Detection. IGI Global snippet, 2009.
179. Kohonen, T.: Self-organizing maps. Springer, 1997.
180. Komlósi Sándor: Az optimalizáláselmélet alapjai. Dialóg Campus Kiadó, Budapest, 2001.
181. Kovács László Dr.: Adatbázisok tervezésének és kezelésének módszertana, Computerbooks kiadó , 2004.
182. Kő Andrea – Lovrics László: Döntéstámogató rendszerek. Széchenyi István Főiskola, 1997.
183. Krafzig, Dirk – Banke, Karl – Slama, Dirk: Enterprise SOA: service-oriented architecture best practices. Prentice Hall PTR, 2005.
184. Krishna, S. Jaya: Business activity monitoring: an introduction. ICFAI University Press, 2006.
185. Kudyba, Stephan – Hoptroff, Richard: Data mining and business intelligence: a guide to productivity. Idea Group Inc (IGI), 2001.
186. Lambert, Peter J.: The distribution and redistribution of income. Manchester University Press, 2001.
187. Law, Averill M. – Kelton, David W.: Simulation modeling and analysis. McGraw-Hill, 2000

188. Lawrence, David B.: The economic value of information. Springer, 1999.
189. Leigh, James R.: Control Theory. IET, 2004.
190. Leonard, Robert: Von Neumann, Morgenstern, and the Creation of Game Theory: From Chess to Social Science, 1900-1960. Cambridge University Press, 2010.
191. Levin, Nissan – Zahavi, Jakob – Olitsky, Morris: AMOS — A probability-driven, customer-oriented decision support system for target marketing of solo mailings. European Journal of Operational Research, 87/3, 1995
192. Levitan, Jeff: Transform customer data into profit. SASCOM, 2008/2.
193. Linstedt, Daniel – Graziano, Kent – Hultgren, Hans: The New Business Supermodel: The Business of Data Vault Modeling. Lulu.com, 2008
194. Liu, Bing: Web data mining: exploring hyperlinks, contents, and usage data. Springer, 2007.
195. Liu, Huan – Motoda, Hiroshi: Feature extraction, construction and selection: a data mining perspective. Springer, 1998.
196. Lombardo, Russ: Crm For The Common Man: The Essential Guide To Designing And Planning A Successful Crm Strategy For Your Business. Russ Lombardo, 2003.
197. Loshin, David: Business Intelligence: The savvy Manager's Guide. Morgan Kaufmann, 2003.
198. Luftman, Jerry N. – Bullen, Christine V.: Managing the information technology resource: leadership in the information age. Prentice Hall, 2004.
199. Lukács András – Fogaras Dániel: Klaszterezés. SZTAKI, 2006.
200. Lunn, Terry – Neff, Susan A.: MRP: integrating material requirements planning and modern business. Business One Irwin, 1992.
201. MacNaughton-Smith, P. – Williams, W.T. – Dale, M.B. – Mockett L.G.: Dissimilarity analysis. Nature, 1964.
202. Makridakis, Spyros G. – Wheelwright, Steven C.: Forecasting: methods and applications. Wiley, 1978
203. Maslow, Abraham: Hierarchy of needs, 1954.
204. May, Thorton: Advanced information management. SASCOM, 2009/3.
205. May, Thorton: The end of forecasting? SASCOM, 2008/2.



206. May, Thorton: The end of forecasting? SASCOM, 2008/2.
207. May, Thorton: The social side of analytics. SASCOM, 2007/3.
208. May, Thorton: Top 8 things transformed by analytics in 2009. SASCOM,
209. Mayo, George Elton: The Human Problems of an Industrial Civilization: Early Sociology of Management and Organizations. Routledge, 1993.
210. Mayo, George Elton: The Social Problems of an Industrial Civilization. Routeledge, 1945.
211. Mayr, Ernst W. – Schmidt, Günther – Tinhofer, Gottfried: Graph-Theoretic Concepts in Computer Science. Springer, 1995.
212. McKinsey Survey, How Companies approach innovation: A McKinsey Global Survey [McKinsey&Company, 2007].
213. Menken, Ivanka – Blokdijk, Gerard: Virtualization - the Complete Cornerstone Guide to Virtualization Best Practices: Concepts, Terms, and Techniques for Successfully Planning, Implementing and Managing Enterprise IT Virtualization Technology. Lulu.com, 2008.
214. Michie, Donald – Spiegelhalter, D. J. – Taylor, C. C.: Machine learning, neural and statistical classification. Ellis Horwood, 1994.
215. Michigan University: History of the New York stock exchange. Michigan, 1887.
216. Mikolics Mihály: Döntéselőkészítési esettanulmányok. GATE, 2000.
217. Mikolics Mihály: Rangsorolási módszerek és döntési technikák. GATE, 1997.
218. Miller, Frederic P. – Vandome, Agnes F. – McBrewster, John: Dot-com Bubble: Economic Bubble, Nasdaq Composite, Stock Market, Quaternary Sector of the Economy, Boom and Bust, World Wide Web, Mosaic (web Browser), Internet, Dot-com Company. Alphascript Publishing, 2010.
219. Miller, Gloria J. – Brautigam, Dagmar – Gerlach, Stefanie Virginia: Business intelligence competency centers: a team approach to maximizing competitive advantage. John Wiley and Sons, 2006.
220. Miller, Gloria J. – Bräutigam, Dagmar – Gerlach, Stefanie Virginia – Davies John: The business intelligence competency center: a SAS approach. SAS Institute, 2005.
221. Minieka, Edward: Optimization algorithms for networks and graphs. Marcel Dekker, 1978.

222. Mirkin, Boris Grigor'evich: Mathematical classification and clustering. Springer, 1996.
223. Mittal, Ankush – Kassim, Ashraf: Bayesian network technologies: applications and graphical models. Idea Group Inc (IGI), 2007.
224. Mohapatra, Sanjay: Business Process Automation. PHI Learning Pvt. Ltd. 2009.
225. Moss, Larissa Terpeluk – Atre Shaku: Business intelligence roadmap: the complete project lifecycle for decision-support applications. Addison-Wesley, 2003.
226. Ng, R. – Han, J.: Efficient and effective clustering methods for spatial data mining. Chicago, 1994.
227. Nicholson, Wesley Lathrop: Exploring data analysis: the computer revolution in statistics. University of California Press, 1974.
228. Olson, Jack E.: Data quality: the accuracy dimension. Morgan Kaufmann, 2003.
229. Orlicky, Joseph: Material requirements planning: the new way of life in production and inventory management. McGraw-Hill, 1975.
230. Paczolay Gyula: Tudományok és rendszerek, 1973, Akadémiai Kiadó, Budapest
231. Page, Jon: Building a business intelligence architecture fit for the 21st century. Melrose Books, 2009.
232. Pálinkó Éva – Szabó Márta: Vállalati pénzügyek. Typotex Kft, 2006.
233. Papp Ilona – Szabó Zsolt Roland: Stratégiai menedzsment. Universitas, 2008.
234. Papp Ottó: A hálós programozási módszerek gyakorlati alkalmazása. Közgazdasági és Jogi Könyvkiadó, 1969.
235. Pareek, Deepak: Business intelligence for telecommunications. Auerbach Publications, 2007.
236. Pataki Béla Dr.: Változásmenedzsment. BME, 2008.
237. Pedricz, Witold: Knowledge-based clustering: from data to information granules. John Wiley and Sons, 2005.
238. Peel, Jeffrey: CRM: redefining customer relationship management. Digital Press, 2002.
239. Perks, Mike: Best practices for software development projects. IBM, 2003.

240. Perner, Petra: Machine Learning and Data Mining in Pattern Recognition. Springer, 2009.
241. Pintér Szilvia: A döntés fogalma, és folyamata. A döntések csoportosítása, döntés, hatáskör, felelősség összefüggése. KözigKonsult Bt, 2004.
242. Plaszcak, Pawel – Wellner, Richard: Grid computing: the savvy manager's guide. Elsevier, 2006.
243. Pourret, Olivier – Naim, Patrick – Marcot, Bruce: Bayesian networks: a practical guide to applications. John Wiley and Sons, 2008.
244. Power, Daniel J.: Decision support systems: concepts and resources for managers. Greenwood Publishing Group, 2002.
245. PriceWaterhouseCoopers: Global Data Management Survey, 2001
246. Quinn, James Brian: Intelligent Enterprise: A Knowledge and Service Based Paradigm for Industry, Free Press, 1992.
247. Rabb, Merry –Doninger, Cheryl: Grid computing basics. SASCOM, 2007/1.
248. Rabin, Jack – Jackowski, Edward M.: Handbook of information resource management. M. Dekker, 1988.
249. Raffai Mária: Információrendszerek fejlesztése és menedzselése. Novadat, 2003.
250. Rasmussen, Nils – Goldy, Paul S. – Solli, Per O.: Financial business intelligence: trends, technology, software selection, and implementation. John Wiley and Sons, 2002.
251. Refaat, Mamdouh: Data preparation for data mining using SAS. Morgan Kaufmann, 2007.
252. Reynolds, Janice: A practical guide to CRM: building more profitable customer relationships. Focal Press, 2002.
253. Reynolds, Janice: Logistics and fulfillment for e-business: a practical guide to mastering back office functions for online commerce. Focal Press, 2001.
254. Robert Frese – Dr. Vicki Sauter: Project success and failure: what is success, what is failure, and how can you improve your odds for success? University of Missouri, 2003.
255. Rokach, Lior – Maimon, Oded Z.: Data mining with decision trees: theory and applications. World Scientific, 2008.
256. Royse, David – Thyer, Bruce A. – Padgett, Deborah K.: Program Evaluation: An Introduction. Cengage Learning, 2009.

257. Ruest, Danielle – Ruest, Nelson: Virtualization: a beginner's guide. McGraw Hill Professional, 2009.
258. S. Nagy Katalin: Szociológia közgazdászoknak. Typotex Kft, 2007.
259. Sabherwal, Rajiv –Becerra-Fernandez, Irma: Business Intelligence (Prectices, Technologies and Management). John Wiley and Sons, 2010.
260. Sarker, Ruhul A. – Abbass, Hussein A. – Newton, Charles Sinclair: Heuristics and optimization for knowledge discovery. Idea Group Inc (IGI), 2002.
261. Sarma, Kattamuri S.: Predictive Modeling with SAS Enterprise Miner: Practical Solutions for Business Applications. SAS Publishing, 2007.
262. Sarnevitz, Alexi: When knowing the customer isnt enough. SASCOM, 2009/1.
263. SAS Publishing: SAS Data Integration Studio 4.21: User's Guide. SAS Publishing, 2009
264. Scheps, Swain: Business Intelligence for Dummies. For Dummies, 2008.
265. Schiller, Stefan: The intelligent enterprise: from a competence management perspective. BAS, 1999.
266. Schniederjans, Marc J. – Schniederjans, Ashlyn M.: Outsourcing management information systems. Idea, 2007.
267. Scott, Norman – Gras Brien: An Introduction to Economic History. Bibliobazaar, 2009.
268. Seber, George Arthur Frederick – Wild, Christopher John: Nonlinear regression. Wiley-IEEE, 2003.
269. Shrivastava, Ashok Narain – Shrivastava, Ashok – Sahami, Mehran: Text mining: classification, clustering, and applications. CRC Press, 2009.
270. Shtub, Avraham: Enterprise resource planning (ERP): the dynamics of operations management. Springer, 1999.
271. Simon, A. Herbert: A vezetői döntés új tudománya. Statisztikai Kiadó Vállalat, 1982.
272. Simon, Herbert: Általános Problémamegoldó Gép. 1957.
273. Sipos Ida: A környezetvédelem ökonómiai kérdései a mezőgazdaságban. Mezőgazdasági Kiadó, 1985.
274. Skapura, David M.: Building neural networks. Addison-Wesley, 1996.

275. Skiena, Steven S.: The algorithm design manual. Springer, 2008.
276. Smiraglia, Richard P.: Metadata: a cataloger's primer. Routledge, 2005.
277. Soman, K.P. – Diwakar, Shyam – Ajay V.: Insight Into Data Mining: Theory and Practice. PHI Learning Pvt. Ltd., 2006.
278. Sommerville, Ian: Szoftver-rendszerek fejlesztése, Software Engineering, (Seventh Edition), PANEM Könyvkiadó, Budapest, 2007.
279. Stahly, Paul – Kischka, Peter: Models, methods and Decision Support for Management. Physica, 2001.
280. Stapko, Timothy – Stapko, Timothy John: Practical embedded security: building secure resource-constrained systems. Newnes, 2008.
281. Staub, F. Brunó: Biológiai Lexikon. Akadémiai Kiadó, Budapest, 1975.
282. Stepanek, George– Software Project Secrets: why software projects fail. Apress, 2005.
283. Surhone, Lambert M. – Timplendon, Miriam T. – Marseken, Susan F.: Thin Client: Computer Program, Information Processing, Fat Client, Persistent Data Structure, Grid Computing, Graphical User Interface. Betascript Publishers, 2009.
284. Suykens, Johan A. K. – Gestel, Tony Van – Brabanter Jos De: Least squares support vector machines. World Scientific, 2002.
285. Szczuka, Marcin – Kryszkiewicz, Marzena – Ramanna, Sheela – Jensen, Richard – Hu, Qinghua: Rough Sets and Current Trends in Computing. Springer, 2010.
286. Szeifert Ferenc – Chován Tibor – Nagy Lajos – Almásy Gedeon: Rendszermodellek – Rendszeranalízis. Veszprém, 2000.
287. Szép Jenő – Forgó Ferenc: Bevezetés a játékelméletbe. Közgazdasági és Jog i Könyvkiadó, 1974.
288. Sziray József – Benyó Balázs – Góth Júlia – Majzik István – Pataricza András: Szoftver rendszerek minőségbiztosítása és verifikálása, SZIF-BME-VE jegyzet, Budapest, 2000.
289. Sziray József Dr.: Szoftver rendszerek minőségbiztosítása. NOVADAT, 2008.
290. Sziray József Dr.: Szoftvertechnológia. NOVADAT, 2008.
291. Szücs Ervin: Rendszer és modell I. Nemzeti Tankönyvkiadó, 1994.

292. Szvitacs István dr.: Minőségmenedzsment. PMMF 2009.
293. Takefuji, Yoshiyasu: Neural network parallel computing. Springer, 1992.
294. Taylor, Winslow Frederick: Time and motion – the principles of scientific motion, 1911
295. Terano, Takao – Liu, Huan – Chen, Arbee L. P.: Knowledge discovery and data mining. Springer, 2000.
296. Thannhuber, Markus J.: The intelligent enterprise: theoretical concepts and practical implications. Springer, 2005.
297. Thierauf, Robert J.: Effective business intelligence systems. Greenwood Publishing Group, 2001.
298. Toomey, John W.: MRP II: planning for manufacturing excellence. Springer, 1996.
299. Tsiptsis, Konstantinos – Chorianopoulos, Antonios: Data Mining Techniques in CRM: Inside Customer Segmentation. John Wiley and Sons, 2010.
300. Tsypkin, Iakov Zalmanovich: Foundations of the theory of learning systems. Academic Press, 1973.
301. Vercellis, Carlo: Business Intelligence: Data Mining and Optimization for Decision Making. John Wiley and Sons, 2009.
302. Vesset, Dan: Emerging trends. SASCOM, 2009.
303. Wang, Lipo: Support vector machines: theory and applications. Springer, 2005.
304. Wei, William W. S.: Time series analysis: univariate and multivariate methods. Addison-Wesley Pub., 1990.
305. Weiss, Sholom M.: Text mining: predictive methods for analyzing unstructured information. Springer, 2005.
306. Wesselényi Andrea: A távmunka térhódítása a világban, Európában és lehetőségei Magyarországon. Teleházak és távmunka Magyarországon. IGE, 1999.
307. Westerman, Paul: Data warehousing: using the Wal-Mart model. Morgan Kaufmann, 2001.
308. Whittaker, Brenda: What went wrong? Unsuccessful information technology projects. KPMG, 1999.

309. Whybark, D. Clay: MRP: a profitable concept for distribution. Institute for Research in the Behavioral, Economic, and Management Sciences, Krannert Graduate School of Industrial Administration, Purdue University, 1975
310. Wight, Olivier: The Executive's Guide to Successful MRP II. John Wiley and Sons, 1995.
311. Williams, W.T. – Lambert, J.M.: Multivariate methods in plant ecology; association analysis in plant communities. Journal of ecology, 1959.
312. Wilson, Sheila: Failed IT Projects. University of Maryland Bowie State University, 1998. május 16.
313. Wright, Oliver W.: MRP II: unlocking America's productivity potential. O. Wight Ltd. Publications, 1981.
314. Xu, Rui – Wunsch, Donald C.: Clustering. Wiley, 2008.
315. Yates, Allen: Multivariate exploratory data analysis: a perspective on exploratory factor analysis. SUNY Press, 1987.
316. Ye, Nong: The handbook of data mining. Routledge, 2003.
317. Ying Lü – Ji-Qian Fang: Advanced medical statistics. World Scientific, 2003.
318. Zeng, Marcia Lei – Qin, Jian: Metadata. Neal-Schuman Publishers, 2008.
319. Zoltayné Paprika Zita – Szántó Richárd: Pillanatfelvétel a menedzsment képességekről és döntéshozatali megközelítésmódokról az EU csatlakozáskor. Versenyképesség Kutatások Műhelytanulmány sorozat, 2006.
320. Zoltayné Paprika Zita: A stratégiai döntéshozatal módszerei és a versenyképesség. Budapesti Közgazdaságtudományi Egyetem, 1997.
321. Zsolnai László: A döntéshozatal etikája. Kossuth Kiadó, Budapest, 2000.

### ***Internetes Források***

322. bi.hu: BI Portál az üzleti intelligenciáról. <http://www.bi.hu>
323. ErpAndMore.com: ERP And More! <http://www.erpandmore.com>
324. IDCHungary.hu: IDC Hungary Kft. <http://www.idchungary.hu>

- 325. [intelligent-enterprise.informationweek.com](http://intelligent-enterprise.informationweek.com): Intelligent Enterprise – better insight for business decisions
- 326. Mario Mansour: How much did a hard drive cost in 1982? [Http://www.mariomansour.org/how-much-did-a-hard-drive-cost-in-1982/](http://www.mariomansour.org/how-much-did-a-hard-drive-cost-in-1982/) 2009. október 6.
- 327. [opensourceerp.guru](http://opensourceerp.guru): Open Source ERP Guru. <http://www.opensourceerp.guru>
- 328. SysOptima.com: E-business Insight – ERP, CRM and Supply Chain Management. <http://www.sysoptima.com>
- 329. Techtarget.com: Techtarget, where serious technology buyers decide. <http://www.techtarget.com>
- 330. Theserverside.net: .NET community discussing. <http://www.theserverside.net/>



## 7 Melléklet

### Táblázatok

2. táblázat: Az analitika alkalmazásának fokozatai

Fokozat / Fázis	Cég állapota	Jelen állapot	Problémák	Teendők; Mi kell a továbblépéshez?
1. fokozat:	Analitikai megoldásokra alkalmatlan	A vállalatnál egy vagy több feltétel megléte hiányzik ahhoz, hogy komolyan vehető analitikai projekt kezdődhessen: vagy az adatok hiányoznak, vagy az elemzői képességek vagy a vezetői hozzáállás nem megfelelő.	<ul style="list-style-type: none"> <li>• Adatdefiníciók hiányoznak</li> <li>• Adatok hiányoznak</li> <li>• Rosszul vagy egyáltalán nem integrált rendszerek</li> </ul>	<ul style="list-style-type: none"> <li>• Össze kell gyűjteni a döntéshozatal szempontjából releváns adatokat;</li> <li>• Létre kell hozni a master / referencia adatok jegyzékét, illetve;</li> <li>• Létre kell hozni az adatkereskedelmi réteget.</li> </ul>
2. fokozat:	Lokalizált analitika alkalmazható	Vannak ugyan analitikát alkalmazó megoldások a vállalatban, ám ezek nem központi módon koordináltak, és emiatt nem is rendelkeznek a stratégiai célokkal.	<ul style="list-style-type: none"> <li>• A megfelelő adat hiányzik a döntéshozatalhoz</li> </ul>	<ul style="list-style-type: none"> <li>• Vállalati szinten kell meghatározni az analitikai célokat és azok adatszükségletét;</li> <li>• A különböző üzleti doménekhez tartozó adattárházak létrehozása;</li> <li>• Támogatni kell az adatok kapcsolását.</li> </ul>
3. fokozat:	Meglévő törekvés	Van ugyan centralizált kezdeményezés az analitika alkalmazását	<ul style="list-style-type: none"> <li>• Az adathalmaz nincs teljesen integrálva,</li> <li>• Nem</li> </ul>	<ul style="list-style-type: none"> <li>• Vállalati szintű adattárház létrehozása, külső adatok kombinálásával;</li> <li>• EDW tervek támogatása;</li> </ul>

		illetően, ám nincs meg minden feltétel, és ennek köszönhetően a haladás lassú.	szabványosít ott és •Nem minden tekintetben elérhető	•Figyelni kell az új lehetséges adatforrásokat
4. fokozat:	Analitikát alkalmazó vállalat	A vállalat már rendelkezik a szükséges humán és technológiai erőforrásokkal, vannak működő és pozitív eredményeket produkáló analitikai alkalmazások, de a stratégiának nincsenek jól alárendelve.	•Néhány folyamat automatizálv a, az üzleti logika a rendszerbe építve	•Hirdetni kell az analitika további lehetőségeit, •Újabb egyedi adatforrásokat kell feltárni •Szigorú adatkezelést kell létrehozni, •Ajánlott a BICC létrehozása.
5. fokozat:	Analitikai úton versenyző vállalat	A vállalat - rutinszerűen használja az analitika nyújtotta lehetőségeket, annak hatása az egész vállalaton belül érezhető, a vezetőség elkötelezett – ez látható a vállalaton kívül és belül egyenként.	-	-

### 3. táblázat: Adatok tulajdonságai

Tulajdonság	Nehézség	Kihívások és célok
strukturáltság	Milyen természetű az adat?	<p>Az adatokat strukturáltság szempontjából három féle képpen lehet tárolni. Képezhetőek belőle kockák, tömbök vagy nemnumerikus leírások. A struktúrától függ nagyrészt hogy a későbbiekben milyen elemzés alapja lehet.</p> <p>A tranzakciós rendszerekben (és a forgalomban lévők túlnyomó többsége ilyen) az adatok táblákba rendezve találhatók (ez a forma a működéshez ideális, azonban elemzéshez nem) – és mivel ez nem tartalmaz történelmi adatokat általában, ezért az adattárházban kockákat képeznek belőle. A képzett adatkockákon már számos új művelet értelmezhető, ám inkább feltárássra és nyomozásra valók, mintsem elemzésre – ennek legfőbb oka a limitált változó mennyiség.</p> <p>A tömbök felépítése előre meghatározott, rengeteg változót tartalmazhat, elemzési célokra ez éppen megfelelő ugyan, de aki nem jártas a technológiában, annak rettenetesen bonyolult lehet.</p> <p>A nemnumerikus adatok közé tartoznak a hang és videofelvételek, az internetes vagy belső adatbázisból származó szöveges és médiatartalmak – elmondható, hogy ezen kategória elemzése a legnehezebb, ám egyben a leginkább fejlődő irányzat is egyben.</p>
egyedülállóság	Hogyan lehetne hasznosítani, főleg akkor ha ez eddig egyáltalán nem történt meg?	<p>A kiinduló probléma az, hogy ha a szektor minden résztvevője ugyanazokkal az adatokkal rendelkezik, akkor mindenki ugyanolyan elemzési módszerrel, nagyrészt hasonló eredményeket kap. Ezért fontos, hogy legyenek egyedülálló adatok – ezek nyilván azok amelyek a vállalatnál kívülre nem publikusak.</p> <p>Az egyedülállóság azonban máshogy is biztosítható: olyan adatokat kell felkutatni, és összefüggéseket felrajzolni, melyekre más addig még nem gondolt.</p>
integráltság	Hogyan konszolidálható több helyről?	<p>A fő cél itt az adatok összegyűjtése és integrálása. Ha a vállalatnál olyan ERP rendszer üzemel, mely a vállalat sok vagy összes részlegét felöleli, nagyban megkönnyíti az adatok szűrését és egyesítését – ám kevés helyzetben van így.</p> <p>A felsorakozó adatok ráadásul sokféle képpen értelmezhetőek és magyarázhatóak – sokban függ az érdekeltségtől és az előadó személyétől – és számuk is igen magas lehet. Éppen ezért fontos a döntéshozatal szempontjából releváns adatokat</p>

<b>minőség</b>		összegyűjteni; ezt nevezzük master vagy referencia adatoknak. Tökéletes adatintegráció persze még a legelőrehaladottabb vállalatoknál sem létezik, de fontos a törekvés és a folytonos munka – az alkalmankénti integrációs erőfeszítések kárba veszhetnek.
	Mennyire megbízható?	Fontos az adatok jó minősége – de nem elsődleges, itt (elemzési célokra) nem kell az a pontosság, mint üzleti jelentések készítésére vagy tranzakciós ügyviteli rendszerekhez. A hiányzó adat nem probléma ugyanígy, mivel statisztikai módon pótolható. A félrevezető adatok viszont problémát okoznak – mivel az adatok nagy része tranzakciós rendszerekből érkezik majd, fontos az esetleges félrevezetésre okot adó események kiszűrése. Itt segítséget nyújthat az ok-okozati összefüggések vizsgálata, segítségével a problémát a gyökerénél lehet megszüntetni.
<b>hozzáférhetőség</b>	Hogyan férhető hozzá?	Ahhoz hogy az adatok elemzésre hozzáférhetőek legyenek, célszerű azokat elválasztani a tranzakciós (forrás) rendszereiktől. Erre a célra tökéletesek az adattárházak (akár szervezeti egységenként is), vagy az egyetlen célt szolgáló adatpiacok; de a legjobb megoldás a centralizált EDW létrehozása.
<b>biztonság</b>	Hogyan biztosítható?	Ki kell alakítani azon szabályok körét, melyek az adatok beszerzését, tárolását, felhasználását szabályozzák. A feladat igen összetett, mivel több ország törvényeit és jogszabályait is figyelembe kell venni, azzal az elsődleges direktívával, hogy kár ne keletkezhessen.
<b>Kezelhetőség</b>	Hogyan szervezhető?	Az hatékony adatkezelés-szervezéshez be kell vezetni néhány szerepkört a vállalatnál. Ezek a következők:

4. táblázat: Humán tőke optimalizáció

Sor-szám	Cél	Lépések	IT-Támogatás
1.	A munkaerő a magas szintű célokhoz igazítása	<ul style="list-style-type: none"> <li>Előre kell jelezni, hogy milyen képességekre lesz szükség az üzleti stratégia kivitelezéséhez,</li> <li>Be kell gyűjteni a lehető legtöbb információt a jövőbeni döntések meghozatalához,</li> <li>A célokhoz igazodva kell kialakítani a munkaerő összetételét,</li> <li>Meg kell állapítani, hogy mik a hiányzó képességek (talent gaps).</li> </ul>	Idősor-elemzés Prediktív analitika
2.	Meg kell címezni a munkaerővel kapcsolatos elvárásokat a teljes tehetség-ciklus kapcsán	<ul style="list-style-type: none"> <li>Megszerzés: a megfelelő alkalmazott, megfelelő képességekkel, a megfelelő időben és költségekkel, a megfelelő pozícióban legyen,</li> <li>Növekedés: Képezni kell a már meglévő munkaerőt, hogy továbbfejlődhessen, és a jövőben kimagasló eredményt érhessen el,</li> <li>Megtartás: fel kell készülni a változó demográfiai állapotokra és trendekre.</li> </ul>	Trendszámítás
3.	Meg kell határozni a kockázatokat és mérsékelni kell azokat	<ul style="list-style-type: none"> <li>A múlt segítségével (de a jövőbe tekintve) meg kell határozni a trendeket, és okokat a felmondást, hiányzást és egyéb okokat illetően,</li> <li>Meg kell határozni, a szervezeti változások hogyan érinthetik a munkaerőt,</li> <li>Meg kell határozni, a jövőben hol várható üresedés, és hol lehet szükség új vezetőre,</li> <li>Meg kell érteni a munkaerőben rejlő keresleti és kínálati mintákat, ezt figyelembe véve kell az ellátásról gondoskodni és a stratégiát kialakítani.</li> </ul>	Trendszámítás, Ok-okozat analízis Minta-analízis
4.	Fel kell készülni az üzletmenet változásaira (egyesülés, leépítés...)	<ul style="list-style-type: none"> <li>Forgatókönyvek készítése a különböző helyzetekre és helyszínekre,</li> <li>Döntéseket kell hozni, hogy miként lehetne a jól teljesítőket megtartani és a rosszul teljesítőket vagy</li> </ul>	Scenario-készítés Döntéstámogatás

5.	Szinkronizálni kell az operatív és pénzügyi stratégiákat	<p>hiányzókat kiszűrni.</p> <ul style="list-style-type: none"> <li>• Minden alkalmazottra meg kell határozni a háttér információkat: hiányzások okai, túlórák, képzési költségek, létszámok, fizetések és egyéb juttatások,</li> <li>• Igazolható álláspontot kell kialakítani azt illetően, hogy ezek a költségek hogyan hatnak a vállalatra</li> </ul>	Analitika, Optimalizáció
----	--	--	--------------------------

