|  |  |
| --- | --- |
| Miskolci Egyetem  Gépészmérnöki és Informatikai Kar  Általános Informatikai Intézeti  *3515 Miskolc-Egyetemváros* |  |

**SZAKDOLGOZAT**

*Feladat címe:*

Mesterséges intelligencia alapú oktatást támogató alkalmazás tervezése

*Készítette:*

Szilvási Péter

*MSc szintű mérnökinformatika szakos*

*Alkalmazásfejlesztő szakirányos hallgató*

*Témavezető:*

Dr. Kovács László

*tanszékvezető, egyetemi tanár*

**Miskolc, 2020.**

# Bevezetés

## Téma választás ismertetése

placeholder

## Miért kell támogatni az oktatást?

placeholder

## Oktatás fontossága

placeholder

## Mesterséges intelligencia és az oktatás

placeholder

# Gépi tanulás menete

## Az adat az új üzemanyag

**Gépi tanulás**

Ha elmélyülünk a gépi tanulás definíciójában, akkor azt találjuk, hogy „*A gépi tanulás az algoritmusok és statisztikai modellek tudományos vizsgálata, amelyet a számítógépes rendszerek egy adott feladat végrehajtásához használnak kifejezett utasítások nélkül, mintákra és következtetésekre támaszkodva*”.

Ha elemezzük az előző definíciót, találunk néhány kiemelt szempontot. Hatalmas tudományos kutatás és erőfeszítés van a növekedés és fejlődés támogatására.

Algoritmusokra és statisztikai modellekre támaszkodik, amelyek azt mondják nekünk, hogy ha el akarjuk sajátítani a gépi tanulást, akkor az algoritmusok, a statisztikák és a valószínűségek átfogó megértését kell kialakítanunk. Célja egy meghatározott feladat végrehajtása, amely elmondja, hogy a gépi tanulási megoldásnak van egy bizonyos hatóköre, például előrejelzés és ajánlás.

Kifejezett utasítások nélkül, ami azt mondja nekünk, hogy a gépi tanulási megoldást nem szabad pontosan beprogramozni a dolgok megtanítására. Mintákra és következtetésekre támaszkodva, amely azt mondja nekünk, hogy gépi tanulási megoldásunk a történeti adatokból bizonyos minták megemésztésére támaszkodik a helyes működés megtanulásához.

**Megoldható problémák gépi tanulással**

Most tárgyaljuk meg a gépi tanulás néhány érdekes, valós alkalmazását a mindennapokban.

* A képfelismerés. Hogyan tudja a mobiltelefon feloldani magát azzal, hogy ha a gazdája megmutatja az arcát a kamerának, miközben nem oldja fel magát más emberek számára? Egy képfelismerési technológiával érik el ezt a funkcionalitást. Ahhoz, hogy fel tudjuk ismerni a kívánt objektumot, a képi adatok (pixeleket) egy gépi tanulási algoritmuson kell kiterjeszteni.
* A rosszindulatú programok észlelése, ahol arra vagyunk kíváncsik, hogy az antivírusok hogyan tudják hatékonyan megismerni az új vírusokat, mielőtt frissítéseket kapnának. Ismét a gépi tanulás a motor, amely ezt működteti.
* A hangfelismerés. Egy másik jól ismert példa, amely a gépi tanulást használja. Legyen szó mobiltelefonról vagy okos otthonról, a gépi tanulás lehetővé teszi a felhasználó hatékony azonosítását.

A gépi tanulás terén szerzett kompetenciák és készségek fejlesztésével fantasztikus megoldásokat tudunk kidolgozni, amelyek a valós problémákkal foglalkoznak és értéket képviselnek az emberiség számára.

Végül érdemes megvitatni, hogy a gépi tanulás miért vált manapság ilyen vonzóvá. Nem is gondolná az ember, hogy már a 60-as évek óta létezik. Azt állíthatjuk, hogy az alapstatisztika és a valószínűség valamiféle gépi tanulás. Az elmúlt években azonban volt néhány mozgatóerő.

**Adat túlcsordulás**

Először is, a problémák nagy méretűvé váltak: ***nagy dimenziószám***. A sok dimenzióval olyan adat, amelynek túl sok jellemzője van, ami szinte lehetetlenné teszi a hagyományos programozással történő elemzést. A nagyon dimenziós adatra példa a páciens egészségi állapota, ahol számos jellemző, például az immunrendszer állapota, a genetikai háttér, a táplálkozás, az operációk, a drog- és dohányfogyasztás.

Hatalmas adat túlcsordulás van, például hírcsatornák, webszolgáltatások, adatbázisok, e-mailek, felmérések adatforrásain. A vállalkozások a közelmúltban észrevették, hogy hatékony módszert kell találni ezen hatalmas adatok megértésére, amelyekből értéket kell termelniük.

A számítástechnika fejlődése, valamint a számítási és tárolási erőforrások folyamatos csökkenése, például a felhőalapú számítástechnikában, hatékonyabbá tette a gépi tanulási algoritmusok megvalósítását. Általában a gépi tanulási algoritmusok erőforrás-igényesek. A kutatások jelentősen megnövekedtek az elmúlt évtizedben. A vállalkozások elkezdték megérteni ennek értékét, ezért számos algoritmust kutattak és fejlesztettek ki. 

Tehát, figyelembe véve korábbi információkat, magabiztosan kijelenthetjük, hogy:

*„Az adatok az új üzemanyag.”*

## Miben különbözik a gépi tanulás?

Nos, egy jó kérdés merülhet fel a bennünk, hogy miért nem programozzuk be üzleti szabályainkat és tartomány (domain) ismeretünket a gépi tanulási modelljeinkbe úgy, mint a hagyományos programozásnál? Más szóval, mi a tényleges különbség a gépi tanulás és a hagyományos programozás között?

**Hagyományos programozás és gépi tanulás**

A hagyományos programozás során a teljes tartományú üzleti logikát olyan üzleti szabályok formájában rögzítjük, amelyek kifejezetten programozhatók a kódunkba. Például, ha egy adott vállalat fizetését szeretnénk kiszámítani, akkor minden forgatókönyvre világos üzleti szabályunk lesz, és pontosan tudjuk, hogy mi fog történi. Teljes mértékben megértjük az üzleti területet.

Míg a gépi tanulás során nem teljesen értjük a tartománymodellt, valamennyire tudjuk, hogyan reagál a rendszer bizonyos bemenetekre és kimenetekre. Nem igazán értjük teljesen a mögöttes bonyolult kapcsolatokat a bemenet és a kimenet között. Például egy gépi tanulási algoritmus a vásárlási viselkedés hasonlósága alapján különböző csoportokba sorolhatja az ügyfeleket, annak ellenére, hogy nem értjük teljesen az alapul szolgáló komplex összefüggést. A feladat egy gépi tanulási algoritmusra van delegálva.

A hagyományos programozásban a rendszer esztétikus. A változtatásokat kifejezetten a forráskód vagy a konfiguráció megváltoztatásával kell végrehajtani. Például, ha új szabályt szeretnénk hozzáadni a fizetési rendszerünkhöz, módosítanunk kell a forráskódot. De a gépi tanulásban a rendszer dinamikus és folyamatosan tanul, és a bemenetek és az új tanúsítványok alapján igazodik. Az ügyfelek ápolására szolgáló algoritmusunk idővel jobbá válhat, mivel több trendet tanul meg.

Egy másik különbség az, hogy a hagyományos programozásnál nincs szükségünk semmilyen történelmi adatra vagy tanulási fázisra. A logika pontosan tudja, mit kell tennie. Bérszámítási rendszerünk megfelelően működhet anélkül, hogy úgymond képzési fizetést biztosítanánk. Míg a gépi tanulás során sok történelmi adatra van szükségünk ahhoz, hogy tanulhassunk belőle. Például ügyfeleink megfelelő osztályozásához, a gépi tanulási algoritmusunknak elegendő mintát kell látnia az alapul szolgáló kapcsolatok megértéséhez. Ügyfélápoló algoritmusunknak történelmi adatokra van szüksége, hogy tanuljanak belőlük. Az embernél is hasonló a tanulás, múltbeli eseményekből vonnunk le következtetéseket, majd próbáljuk megjósolni a jövőt.

**Keretrendszerek**

A hagyományos programozásban a logika egyszerű és egyértelmű. Könnyen elolvashatjuk a forráskódot, hogy megtaláljuk az igazságot. A gépi tanulás homályos és bonyolult bizonyos matematikai algoritmusok alapján. Általában könyvtárak vagy olyan keretrendszerek, például a scikit-learn, tensorflow és a Keras segítségével elvonatkoztathatunk a gépi tanulás mögötti bonyolultságtól. A keretrendszerek absztrakciós rétegek, amelyek hozzájárulnak a gyors és egyszerű fejlesztéshez.

## A gépi tanulás folyamata

Machine learning pipeline

Adat pipeline

(Többi pipeline)

## Az adatforrás a kezdetek kezdete

Mik lehetnek az adat

Honnan szerezhető

## Saját adatforrásom ismertetése

Adat alapvetően: Ált leírás: null érték, zajok

Adathalmaz ismertetése

Átvezetés:

*Többet mond egy kép*

*mint száz bekezdés*

# **Reprezentációs módszerek**

## Miért hasznos az adatok megjelenítése?

Az adattudományban az egyik legfontosabb készség az adatok eloszlásának és összetettségének vizualizálása és megértése. Manapság azonban az online tanfolyamok nagy része elsősorban a gépi tanulásra és az algoritmus működésére összpontosít.

Ez nem azt jelenti, hogy a gépi tanulás ismerete elhanyagolandó. Nyilvánvalóan a legjobb adattudós könnyedén eligazodik a gépi tanulási algoritmusok nagy részével. Mindenesetre az adattudomány nem korlátozódik le a gépi tanulásra, inkább az egy képesség. A készség alatt azt kell érteni, hogy mennyire érted a probléma területet (domain), mennyire tudsz elmélyülni az adatokban. Elkapni a rejtett összefüggéseket és megtalálni az adatokban rejlő üzenetet. Az adat beszél magáról. A legjobb mód arra, hogy szóra bírjuk az ***adatmegjelenítés***.

**Mi az adatmegjelenítés?**

Az adatmegjelenítés az információ (adatok) felvételének és vizuális kontextusba, például térképbe vagy grafikonba helyezésének módszere. A fő cél a nagy adatkészletek vizuális grafikába történő szűrése, hogy lehetővé tegye az adatokon belüli összetett kapcsolatok könnyű megértését.

Tehát a vizualizáció választ ad azokra a kérdésekre, amelyeket nem tudunk.

**Miért fontos az adatmegjelenítés?**

Az adatmegjelenítést egyre inkább minden sikeres adatközpontú elemzési stratégia alapvető utolsó lépésének tekintik. Amikor az adattudósok egy összetett projekt közepén vannak, szükségük van egy módra az összegyűjtött adatok megértésére. Figyelemmel kell kísérni és módosítani folyamatot annak biztosítása érdekében, hogy megfelelően funkcionáljon. Az adatmegjelenítés megkönnyíti az adatcsoportok mintáinak, előítéleteinek és kiugró értékeinek észlelését is.

A célközönség vezetése arra, hogy az üzleti felismerésekre összpontosítson, hogy felfedezzék a figyelmet igénylő területeket. Korábban észrevétlen kulcsfontosságú tények feltárása az adatforrásokkal kapcsolatban, hogy a döntéshozók segítsenek adatelemzési jelentések elkészítésében. Segít az érdekelt feleknek és a csapat többi tagjának minőségi információkkal szolgálni azáltal, hogy hatalmas mennyiségű adatot könnyen érthető képpé és grafikává alakít.

**Az adatmegjelenítés előnyei**

Figyelembe véve, hogy az adatok milyen hatást gyakorolnak a vállalkozás növekedésére, íme néhány előny

* Segít felismerni a legújabb fejlesztéseket a termék fejlesztése és az üzleti nyereség növelése érdekében.
* Az adatmegjelenítések megkönnyítik a kis és nagy adatok megértését az emberi agy számára, ami jobb elemzéshez vezet.
* Segít megérteni a történetet - Az emberi agy nem képes egyszerre nagy mennyiségű számot vagy szöveget megérteni, sőt csak elképzelni. Szüksége van egy vizuális ábrázolásra, hogy értelmezze őket, és ennek következtében a nyers adatokat kézzelfogható fogalommá alakítsa.

## Ábrázolási formák

**bevezetés**: történet, gépek előtt, hogy ábrázoltunk, mit ábrázoltunk (barlangokra strigulát stb.). Az ember maximum három dimenziót képes értelmezni. Az adatelemzésnél a dimenziót úgy kell felfogni, mint a tér leírására használtat. Szélesség, hosszúság és magasság helyett adatok tulajdonsága mentén ábrázolunk. Egy csillag esetében például: hőmérséklet, tömeg és fényesség. Az adatok természetéből kifolyólag számtalan dimenzió megjelenítés lehetséges.

**különböző tulajdonságok**at ugyanazon a koordinátán való ábrázolása

**2d**

**3d**

**Hogyan tudunk háromnál több dimenziót ábrázolni?**

...

## Saját adathalmazon bemutatása

**Saját adathalmazon ábrázolása**: Milyen megjelenítést használtam, több adathalmaz bemutatása, észrevételek, következtetések stb.

**Megjelenítési forma bemutatása**: Miért ezt választottam. Mi mit jelent rajta. Hány db adat. Színek, csoportok leírása stb.

**Mit mond az ábra?** Ábráról leszűrt következtetések. Célszerű stratégiák megfogalmazása: outlier eltávolítása, dimenzió csökkentés/növelés, adat csökkentés/növelés stb.

**Az ábrázolás az érme egyik oldala**: ábrázolás + statisztika. Az ábrázolás lehet megtévesztő. Mikor lehet megtévesztő és miért? Átvezetés: Nem szabad hinni mindig a szemünkben. Az ember képes beleesni az optikai illúziók csapdáiban. Ennek érdekében nyújthat segítséget az adatok statisztikai elemzése.

# Adatelemzés lépései és adathalmazok ismertetése

placeholder