

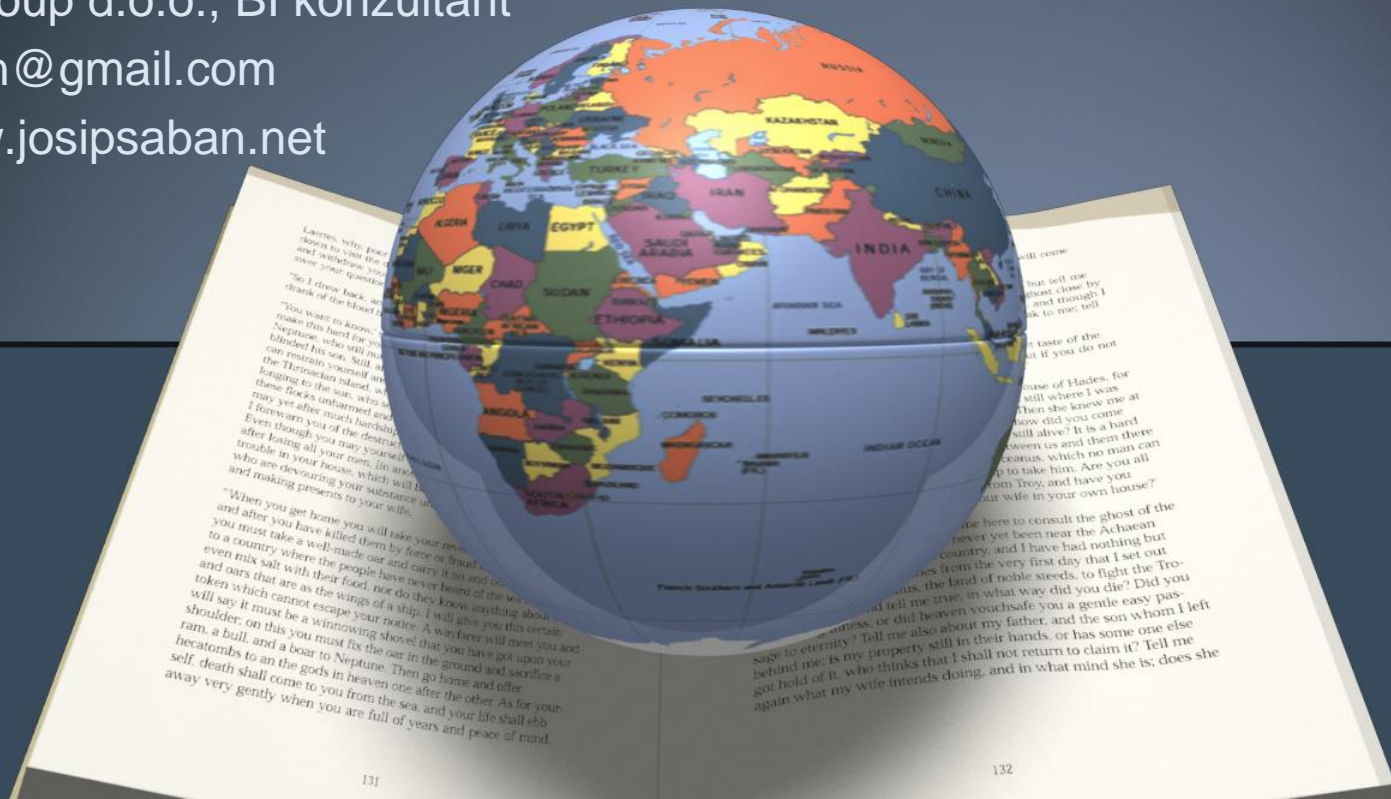
SQL Server 2005 Data Mining

Josip Šaban, dipl. ing. rač.

Alfatec Group d.o.o., BI konzultant

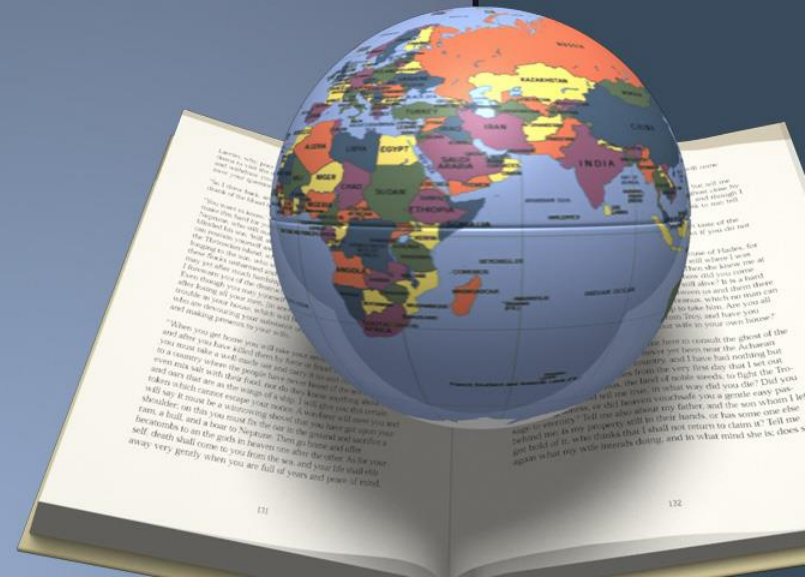
josipsaban@gmail.com

<http://www.josipsaban.net>



Preduvjeti za predavanje

- Poznavanje načina rada baza podataka
- Poznavanje matematike (statistike)
- Poznavati što je poslovna inteligencija i koja je njena uloga u IT-u
- Poznavanje koncepata poslovne inteligencije:
 - Što je skladište podataka?
 - Što je ETL proces?
 - Ostalo... 😊



Korišteni materijali

- Wiley & Sons: “SQL Server 2005 Data Mining” (programski primjeri)
- Micheline Kamber & Jiawei Han: “Data Mining: Concepts And Techniques, 2nd Edition”
- Rafal Lukawiecki – razna predavanja o rudarenju podataka sa TechNet portala (fotografije)
 - Svi navedeni materijali sa demo projektom će se nalaziti na mojim stranicama za download

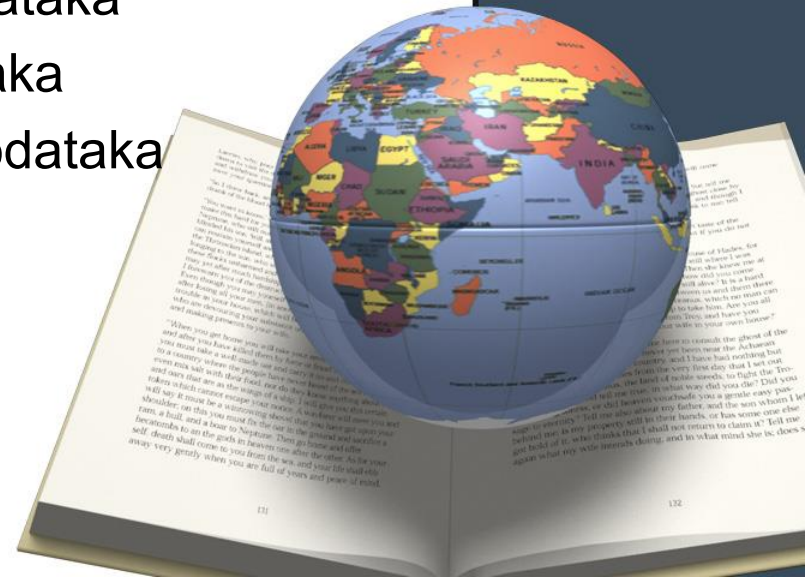




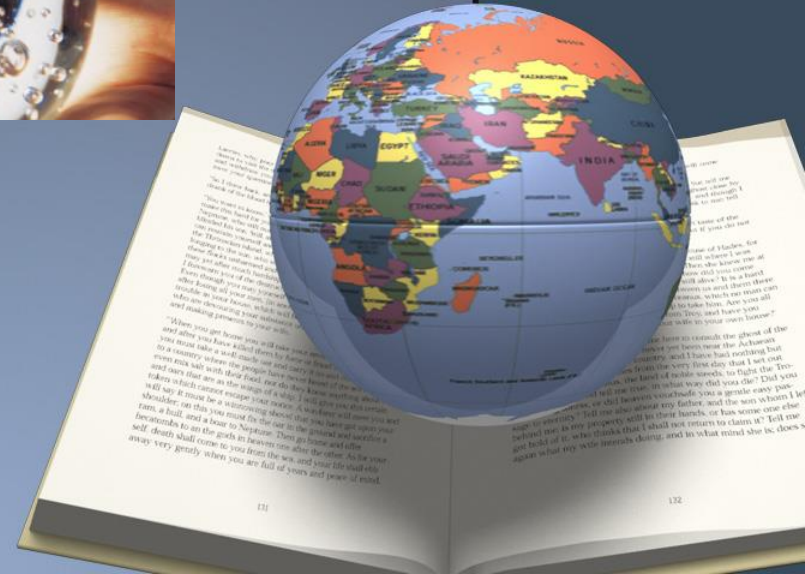
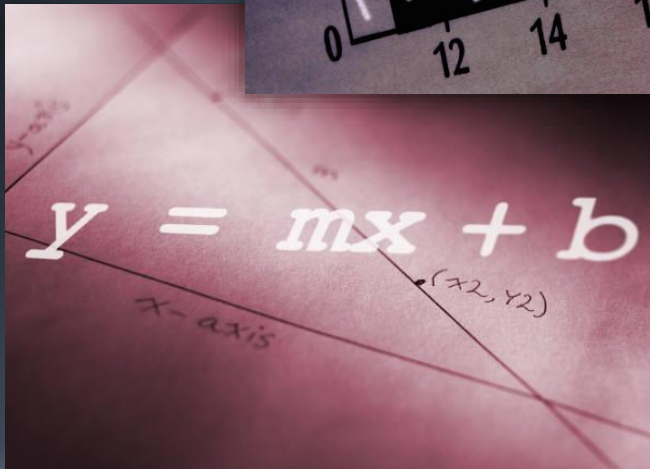
Uvod

Uvod - sadržaj

- Čarolija...ili?
- Što je rudarenje podataka?
- Koja su njegova područja primjene?
- Koje su metode rudarenja podataka te u kojim se algoritmima koriste?
- Koja su njegova područja primjene (2)?
- Životni ciklus projekta rudarenja podataka
- Tri osnovna koraka rudarenja podataka
- SQL Server podrška za rudarenje podataka
- Glavni “igrači” u rudaranju podataka
- Trenutni izazovi u razvoju rudarenja podataka



Čarolija...ili?



Što je rudarenje podataka?

- Rudarenje podataka je sastavni dio BI porodice proizvoda, zajedno sa ETL-om, OLAP-om i izvještavanjem
- Osnovna uloga rudarenja podataka je analiza podataka i pronalaženje skrivenih uzoraka korištenjem automatskih ili polu-automatskih metoda čime povećavamo vrijednost samih podataka i potičemo pretvaranje podataka (informacija) u znanje



Što je rudarenje podataka?

- Rudarenje podataka podrazumijeva primjenu algoritama (kao što su stabla odluke, klustering, asocijacija, vremenski slijedovi, ...) na skupove podataka te njihovu analizu
- Ta analiza stvara uzorke koje možemo pretraživati u potrazi za korisnim informacijama
- U ovisnosti o korištenom algoritmu ti uzorci mogu biti u obliku stabala odluke, pravila, klustera ili nizova matematičkih formula



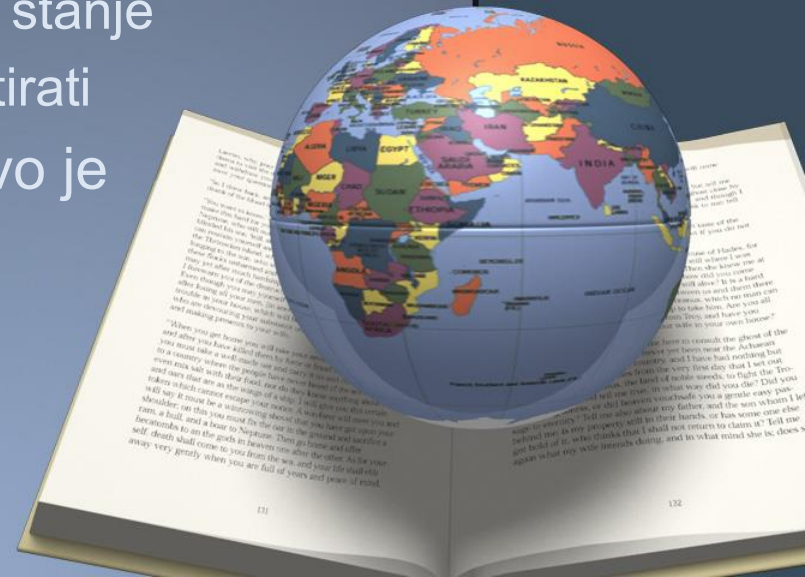
Što je rudarenje podataka?

- Informacije pronađene u uzorcima mogu biti korištene za izvještavanje, vodič za složene marketinške odluke i, najvažnije, za predviđanje
- Rudarenje podataka pruža značajnu poslovnu vrijednost ukoliko se upotrijebi na ispravan način
- Osnovne motivacija za rudarenje podataka:
 - Velika količina dostupnih podataka
 - Sve jače tržišno natjecanje
 - Spremnost tehnologije



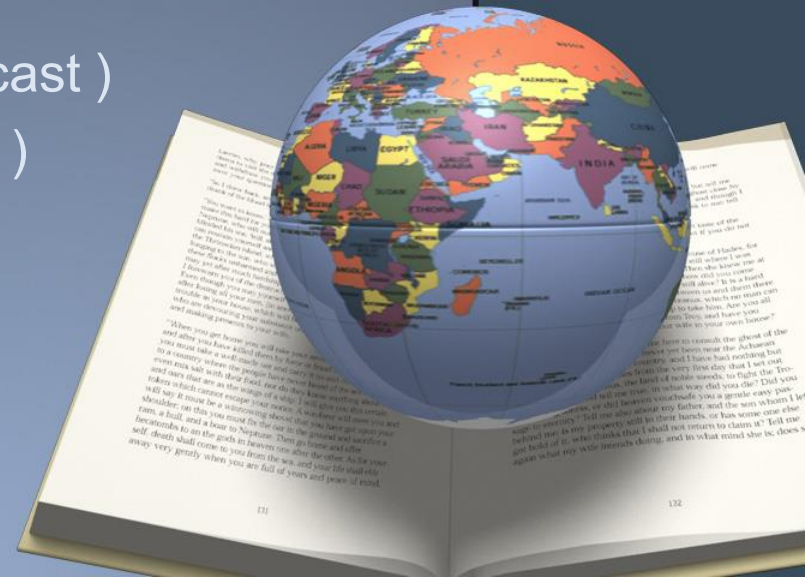
Što je...zapravo...rudarenje podataka?

- Da budemo potpuno jasni...uspjeh ili neuspjeh projekta rudarenja podataka ovisi isključivo o:
 - Kvaliteti podataka
 - Razumijevanju podataka tokom analize rezultata
- Bez ta dva čimbenika sve drugo dosad spomenuto je sekundarno jer:
 - Modeli neće predstavljati stvarno stanje
 - Krivi rezultati će se krivo interpretirati
- U konačnici alat je nebitan...no ovo je predavanje o alatu ☺
 - Za sve drugo - Kamber & Han...i puno vremena ☺



Područja primjene

- Tehnike rudarenja podataka primjenjive su na mnoga područja, od kojih su u poslovnoj okolini najčešća:
 - Analiza odljeva (Churn analysis)
 - Otkrivanje prijevara (Fraud detection)
 - Upravljanje rizicima (Risk management)
 - Segmentacija korisnika (Customer segmentation)
 - Ciljani oglasi (Targeted ads)
 - Predviđanja prodaje (Sales forecast)
 - Poprečna prodaja (Cross-selling)



Metode rudarenja podataka

- Rudarenje podataka možemo koristiti za rješavanje velikog broja poslovnih problema a ovisno o vrsti problema koristimo neku od slijedećih metoda:
 - Klasifikacija (Classification)
 - Klustering (Clustering)
 - Asocijaciju (Association)
 - Regresiju (Regression)
 - Predviđanje (Forecasting)
 - Sekvencijalnu analizu (Sequence Ana.)
 - Analizu odstupanja (Deviation Ana.)



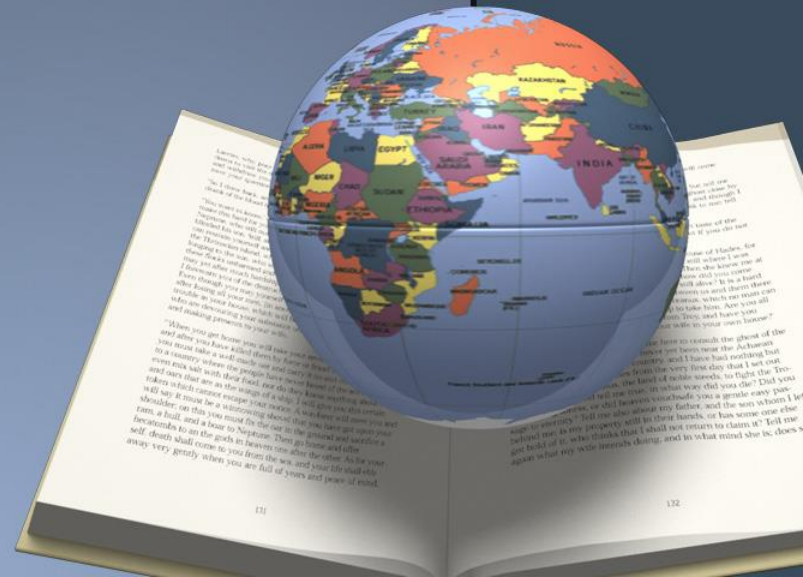
Metode rudarenja podataka

- Klasifikacija
 - Jedna od najčešćih metoda
 - Koristimo ju kod analize odljeva, upravljanja rizicima i ciljanog oglašavanja
 - To je dodijeljivanje slučajeva u kategorije u odnosu na predviđajući atribut (“predictable attribute”)
 - Svaki slučaj sastoji se on niza atributa od kojih je jedan predviđajući
 - Cilj je pronaći model koji opisuje predviđajući atribut kao funkciju ulaznih atributa

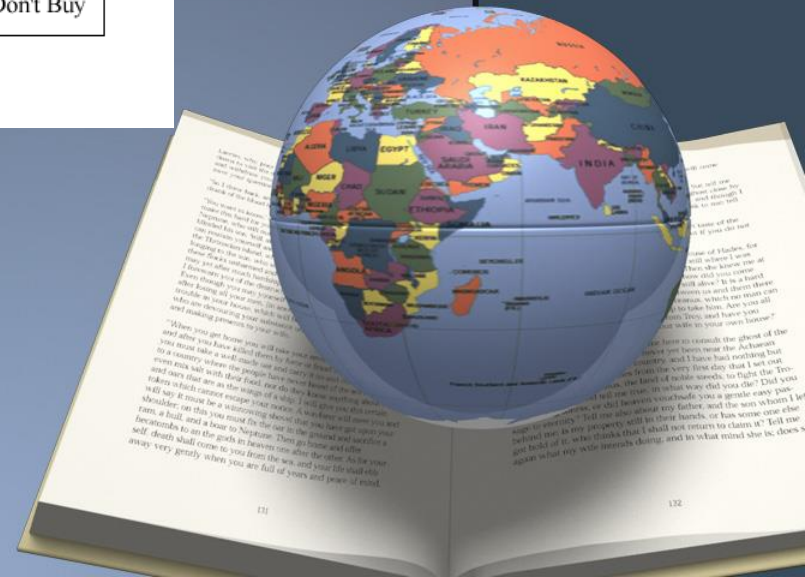
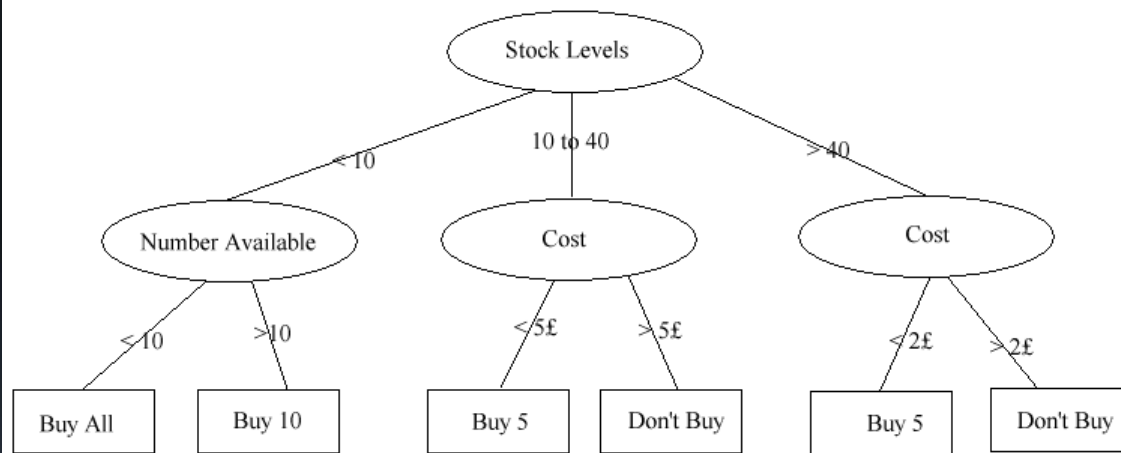


Metode rudarenja podataka

- Za treniranje klasifikacijskog modela potrebno je poznavati vrijednosti ulaznih atributa što se radi iz povijesnih podataka – metode koje zahtijevaju izvor podataka za učenje nazivaju se nadzirane (“supervised”) metode
- U klasifikacijske algoritme spadaju stabla odluke, neuronske mreže i Bayes-ov klasifikator



Metode rudarenja podataka

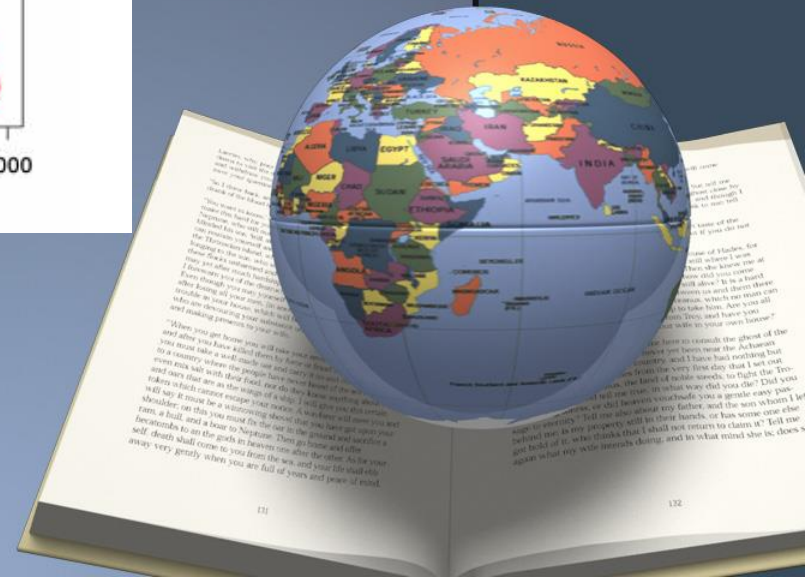
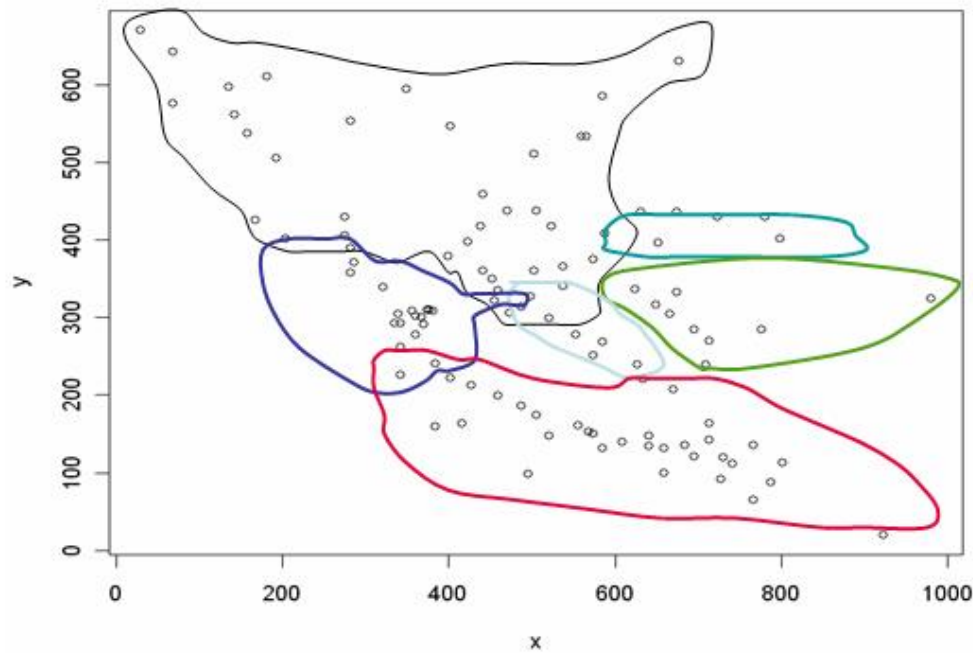


Metode rudarenja podataka

- Klustering
 - Često se naziva i segmentacija
 - Koristi se za identifikaciju prirodnih grupiranja slučajeva u odnosu na definirane attribute
 - Slučajevi koji se nalaze u istoj grupi imaju više-manje iste vrijednosti atributa
 - Klustering je nenadzirana (“unsupervised”) metoda rudarenja podataka – svi ulazni atributi (variable) se smatraju jednako važnima
 - Većina klustering metoda je iterativna
 - Iteracije se ponavljaju dok model ne konvergira, tj. dok se ne uspostave jasne granice među grupama



Metode rudarenja podataka



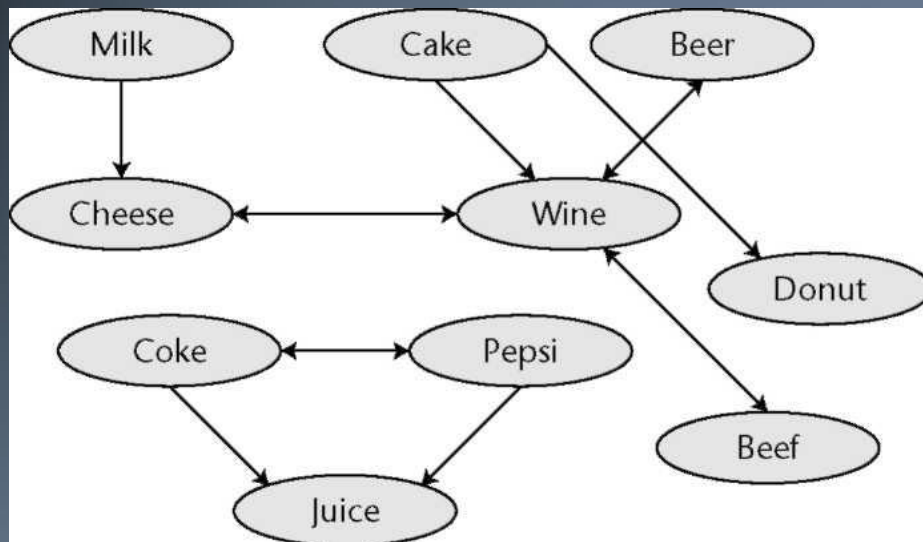
Metode rudarenja podataka

- Asocijacija
 - Često se naziva i analiza košarice (“market basket analysis”)
 - Standardni problem kod kojega se koristi ova metoda je određivanje koji se proizvodi prodaju zajedno
 - Svaki proizvod, ili generalno govoreći, svaki par atribut/vrijednost se smatraju jednom cjelinom



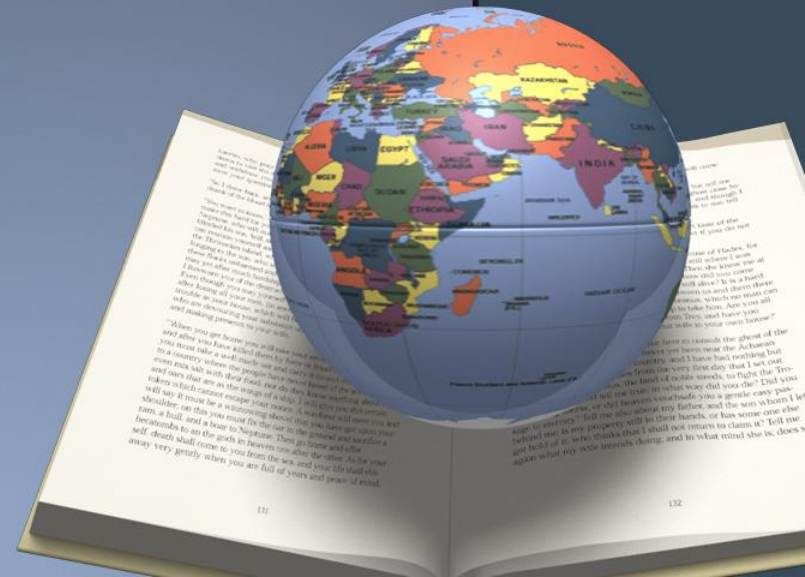
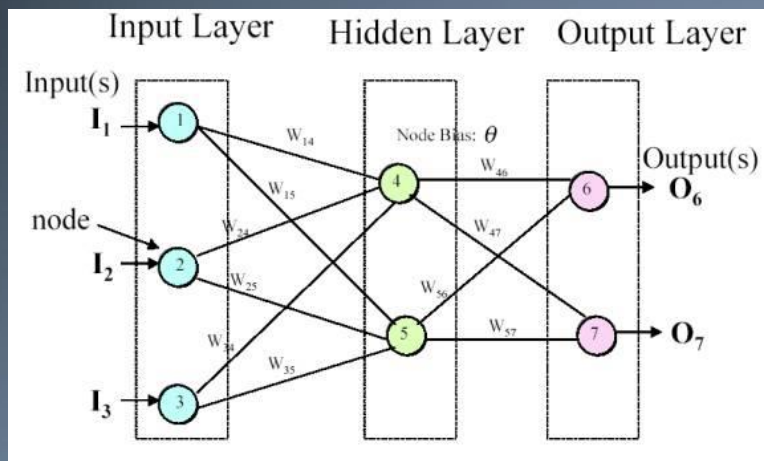
Metode rudarenja podataka

- Asocijacija ima dva osnovna cilja:
 - Pronaći parove koji “idu” zajedno
 - Pronaći asocijacijska pravila (većina asocijacijskih algoritama također pronalazi pravila te vjerojatnost njihovog ostvarenja (“confidence”) što je izraženo u postotcima)



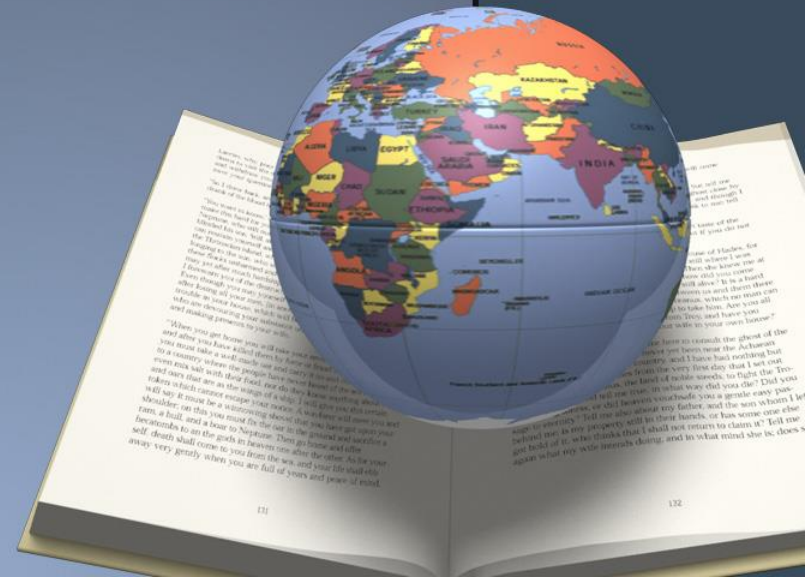
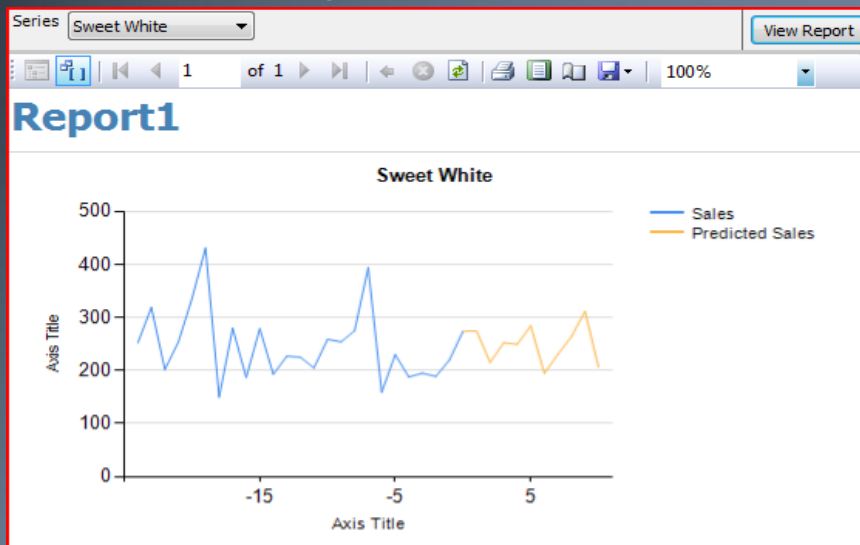
Metode rudarenja podataka

- Regresija
 - Slična kao i klasifikacija, glavna razlika je što je predvidljivi atribut slijedni broj
 - Godinama istraživana u području statistike
 - Najpoznatiji algoritmi su:
 - Linearna i logička regresija
 - Stabla regresije
 - Neuronske mreže



Metode rudarenja podataka

- Predviđanje
 - Ulaz je najčešće skup podataka koji imaju vremenski atribut (najčešće u obliku datuma stvaranja zapisa)
 - Tehnike predviđanja pomažu u rješavanju problema globalnih trendova, periodičnosti i čišćenja šuma u podacima



Metode rudarenja podataka

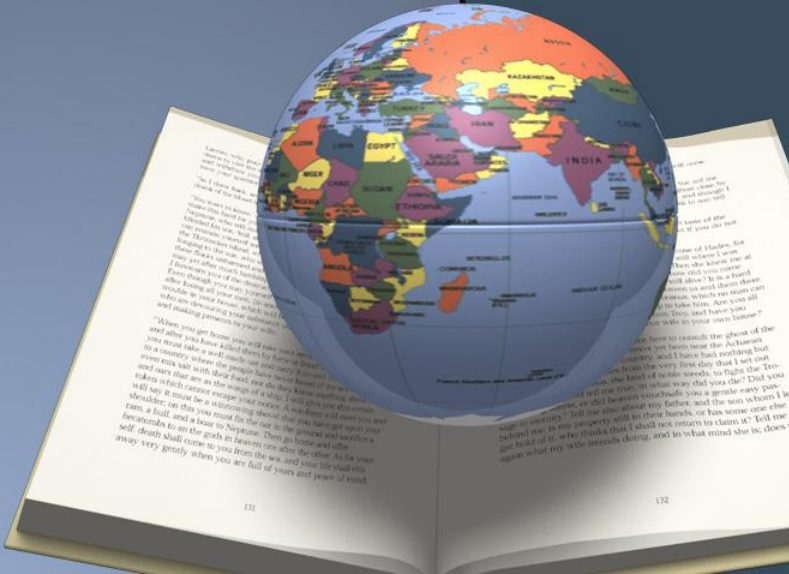
- Sekvencijalna analiza
 - Služi otkrivanju uzoraka u diskretnim sekvencama podataka (sekvenca se sastoji od niza diskretnih vrijednosti ili stanja)
 - Usporedba predviđanja i sekvencijalne analize:
 - Sekvencijalna analiza i predviđanje sadržavaju povezane podatke, ali razlika je u tome da sekvencijalna analiza radi sa diskretnim vrijednostima, dok predviđanje sadrže slijedne vrijednosti



Metode rudarenja podataka

- Usporedba sekvencijalnih i asocijacijskih modela:
 - Kod oba modela pojedini slučaj sadržava niz vrijednosti ili stanja ali sekvencijalni modeli analiziraju prijelaze između stanja, dok asocijacijski model smatra sve vrijednosti neovisnim i nezavisnim
- Primarno se upotrebljava za analizu web logova i DNK analizu
- Postoji više komercijalno dostupnih algoritama od kojih se najčešće koriste Markovljevi lanci

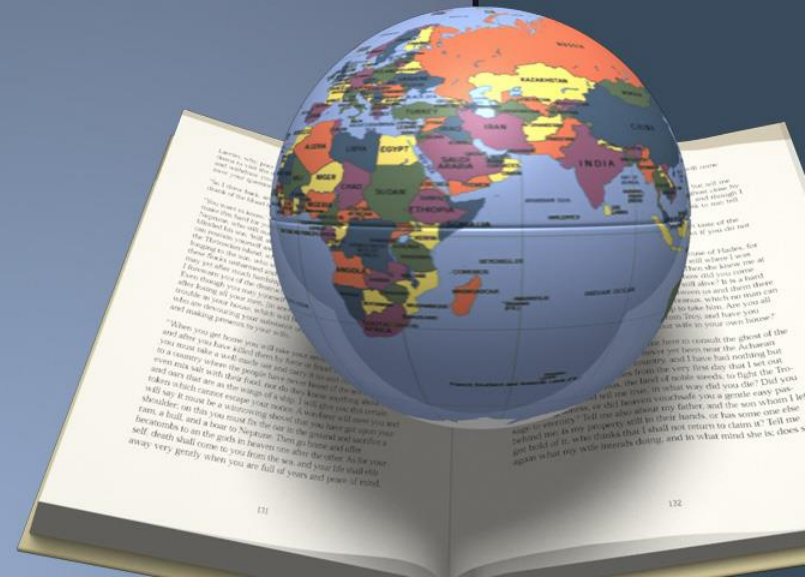




- [illegible]

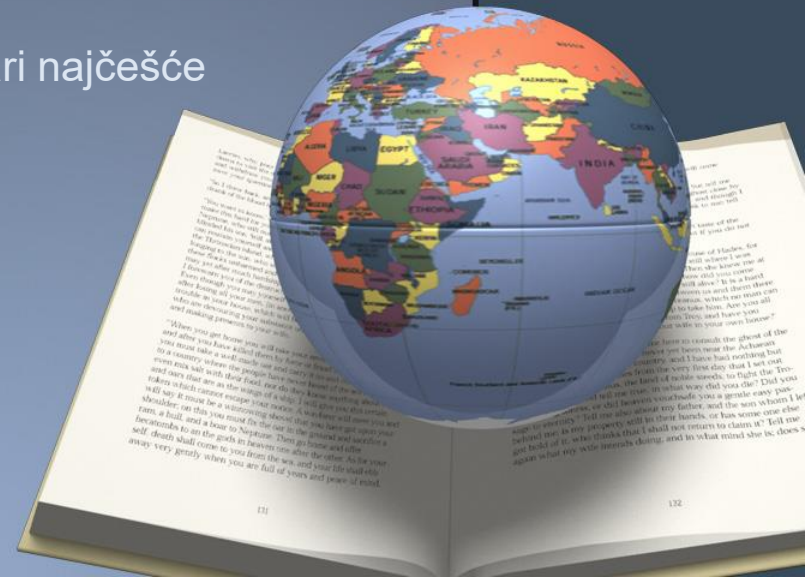
Metode rudarenja podataka

- Analiza devijacije
 - Služi pronalaženju rijetkih slučajeva koji se ponašaju puno drugačije od ostalih
 - Ponekad se naziva i “detekcija nepodobnih članova grupe” (“outlier detection”), što se odnosi na dijagnosticiranje značajnih promjena u odnosu na prijašnje stanje



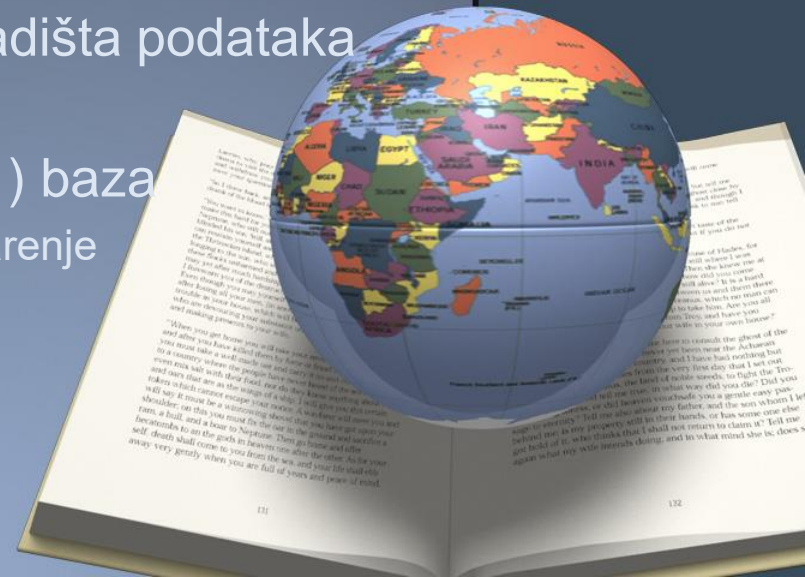
Metode rudarenja podataka

- Najčešće su primjene na području prevara kreditnih kartica (identifikacija abnormalnih slučajeva u milijunima transakcija), detekcije mrežnog napada, analizu grešaka u proizvodnji...
- Ne postoji standardna tehnika za analizu devijacije, no najčešće se koristi neka modificirana verzija stabala odluke, klasteringa ili neuronskih mreža
 - Da bi dobili značajna pravila analitičari najčešće umjetno povećavaju broj anomalija u skupu podataka za trening



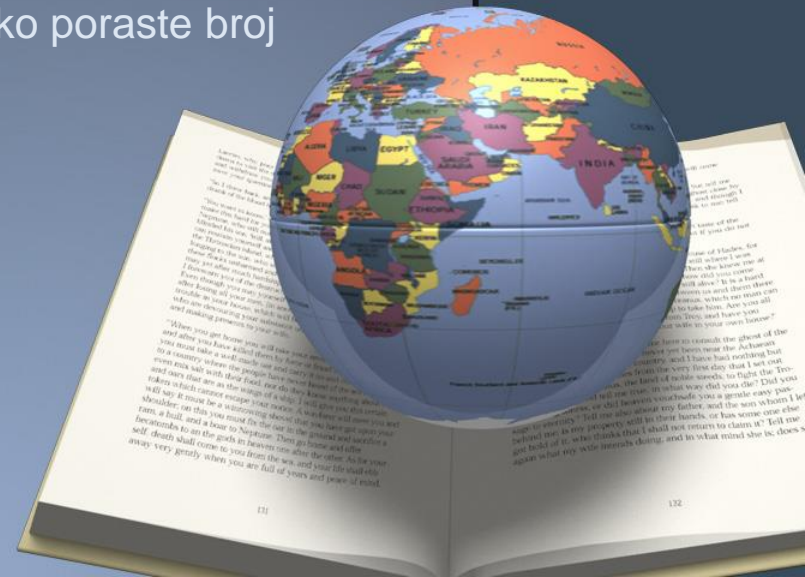
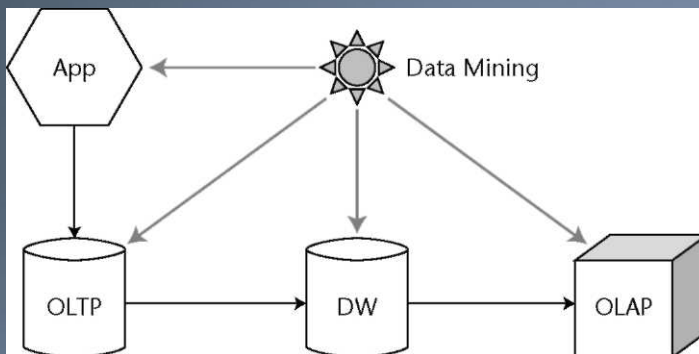
Područja primjene (2)

- Osnovna područja primjene su:
 - Statistika
 - Strojno učenje
 - Baze podataka
- Gdje se može koristiti rudarenje podataka u poslovnom svijetu:
 - Najčešće, za izvještavanje iz skladišta podataka gdje su podaci već pročišćeni
 - Direktno iz transakcijskih (OLTP) baza
 - Najčešće radimo kopiju baze za rudarenje



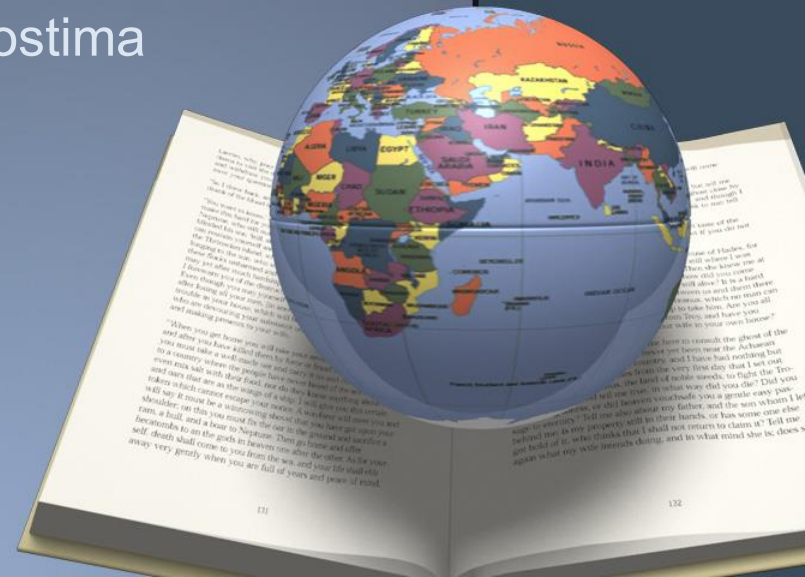
Područja primjene (2)

- Direktno povezano na poslovne aplikacije (kao predviđanje ili kao složena validacija)
 - Česta primjena je u web aplikacijama – kad korisnik stavi objekt u košaricu, izvede se predikcijski upit koji vraća listu preporučenih objekata temeljenih na asocijacijskoj analizi
- Analizu OLAP kocki
 - Radi veličine modernih kocki ponekad je teško pronaći korisne informacije u njima osobito ako poraste broj dimenzija



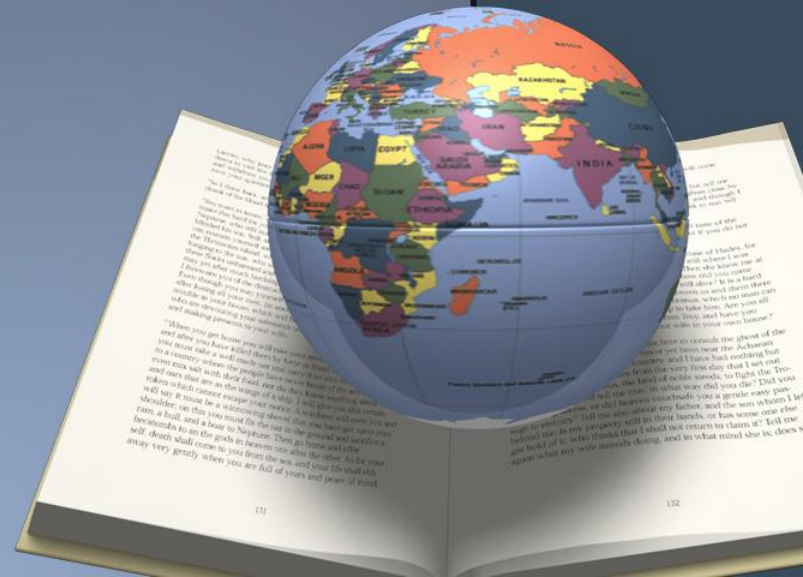
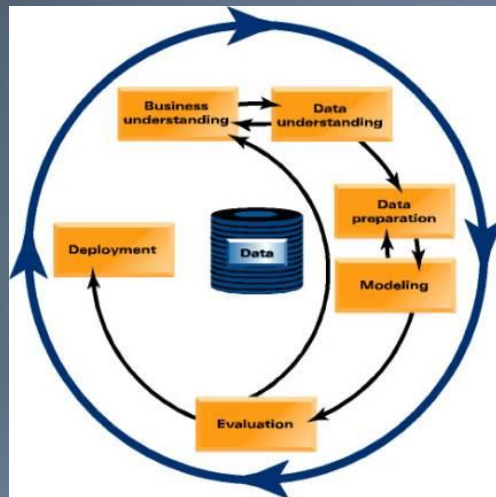
Životni ciklus DM projekta

- Skupljanje podataka
- Čišćenje podataka i transformacija
 - Transformacija tipova podataka
 - Transformacija slijednih kolona u bazi
 - Grupiranje
 - Agregacije
 - Upravljanje nepostojećim vrijednostima
 - Ukljanjanje abnormalnosti
- Izgradnja modela
 - Najčešće gradimo više modela (temeljenih na različitim algoritmima) te ih uspoređujemo



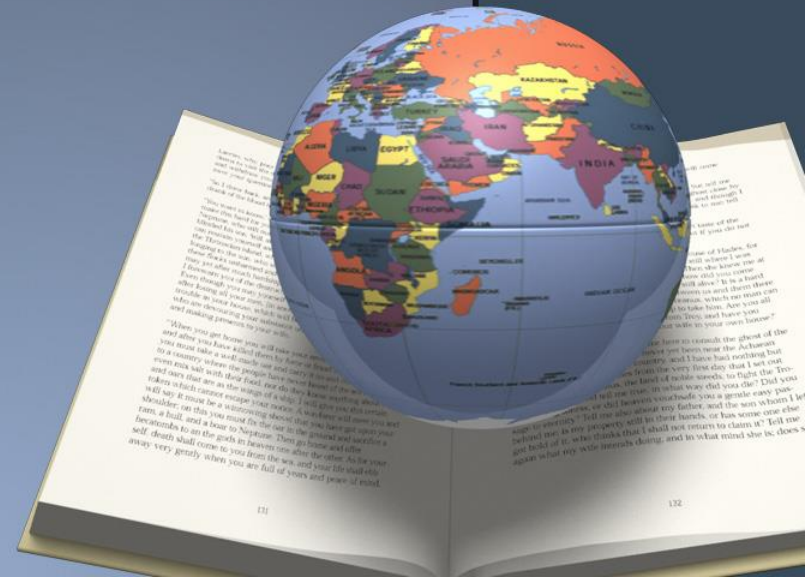
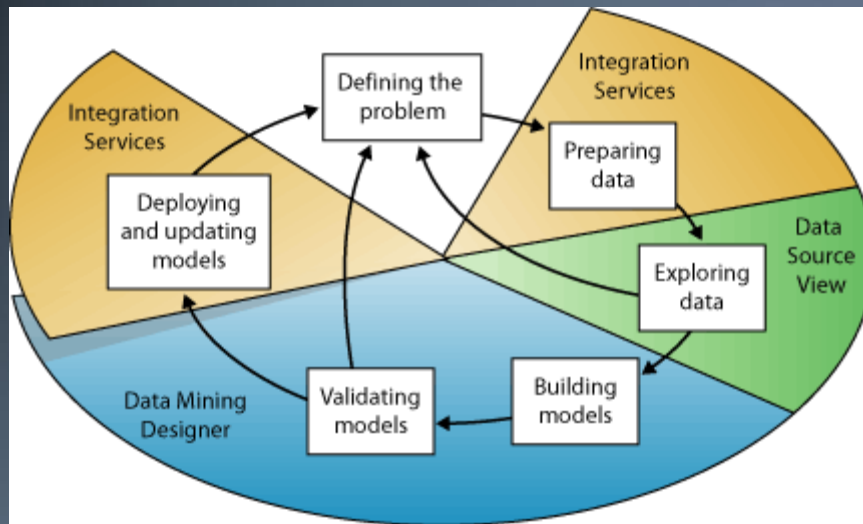
Životni ciklus DM projekta

- Analiza modela
 - Cilj je doći do korisnih informacija ili pravila
- Izvještavanje
- Predviđanje (scoring)
- Integracija sa vanjskim aplikacijama
- Upravljanje modelima



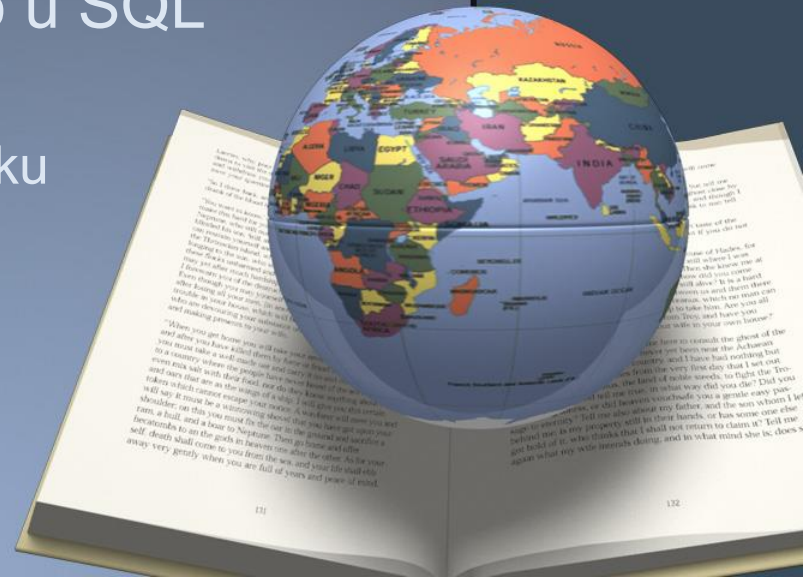
Osnovni koraci

- Stvaranje modela (Model Creation)
- Treniranje modela (Model Training)
- Predviđanje iz modela (Model Prediction)



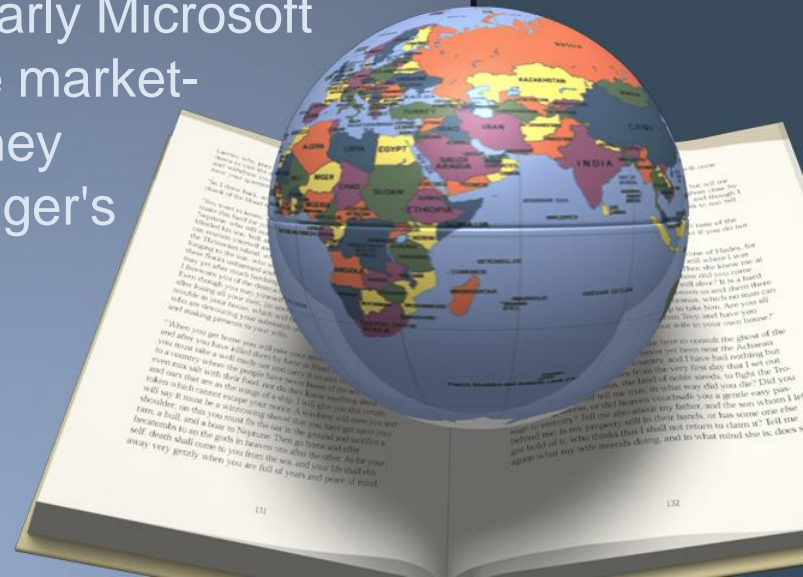
SQL Server podrška

- Osnovna podrška postojala je već u SQL Server-u 2000
 - Microsoft s time postao prvi veliki DB proizvođač koji je uključio rudarenje podataka u bazu podataka
- Značajno poboljšanje napravljeno je u SQL Server-u 2005 (i nadograđeno u SQL Server-u 2008)
 - Detalji će biti objašnjeni u nastavku
 - Za sve funkcionalnosti potrebna Enterprise licenca (Standard ne podržava text mining, ali ima potpunu data mining podršku)

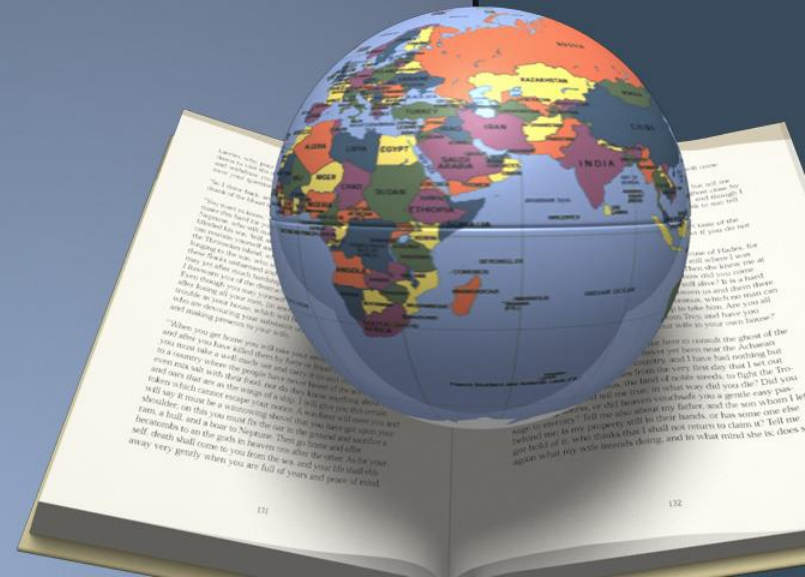


Glavni proizvođači

- SAS
- SPSS Clementine
-
 - However, CRM leaders such as Microsoft, Oracle, Siebel and SAP haven't had strong offerings, but that could change this year; once these "800-pound gorillas of CRM" – particularly Microsoft and Oracle – come out with more market-impacting, visionary strategies, they could easily take over the challenger's quadrant as soon as next year (Gartner, 2008)

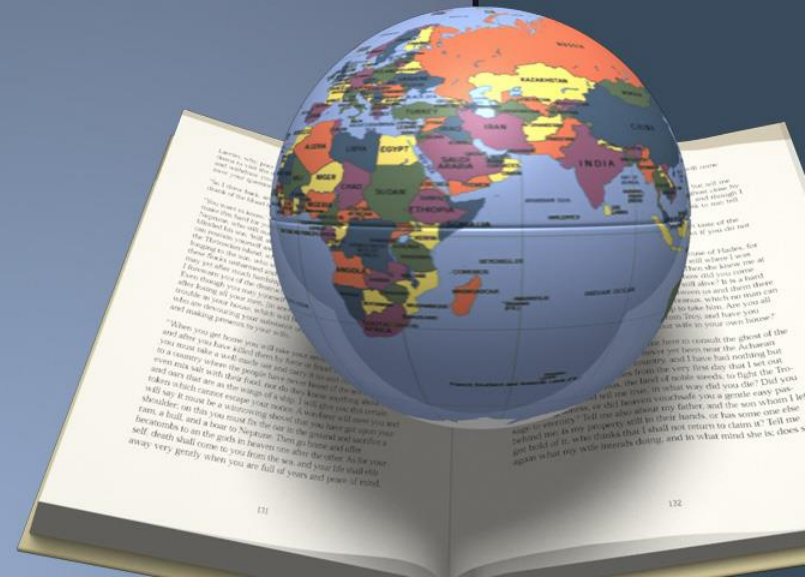


Glavni proizvođači (Gartner)



Trenutni problemi

- Zatvorenost aplikacija (nepostojanje zajedničkog API-a)
- Okrenuto analitičarima (statističarima) umjesto razvojnoj zajednici
- Nedovoljna edukacija korisnika
- Nedovoljno razvijeni algoritmi



Novi trendovi

- Ugrađivanje rudarenja podataka u aplikacije
- Pojednostavljenje i standardizacija rudarenja podataka za razvojnu zajednicu te za rješavanje specifičnih poslovnih područja
- Konsolidacija proizvoda
- PMML – standardni oblik opisnog jezika modela koji olakšava razmjenu i instalaciju
 - Pogledati <http://www.dmg.org/>



DMX – Data Mining eXtensions

- Upitni jezik za SQL Server Data Mining
- Najčešća upotreba je za integraciju modela u korisničke aplikacije
- Koristi relacijske koncepte, sintaksa je vrlo slična SQL-u

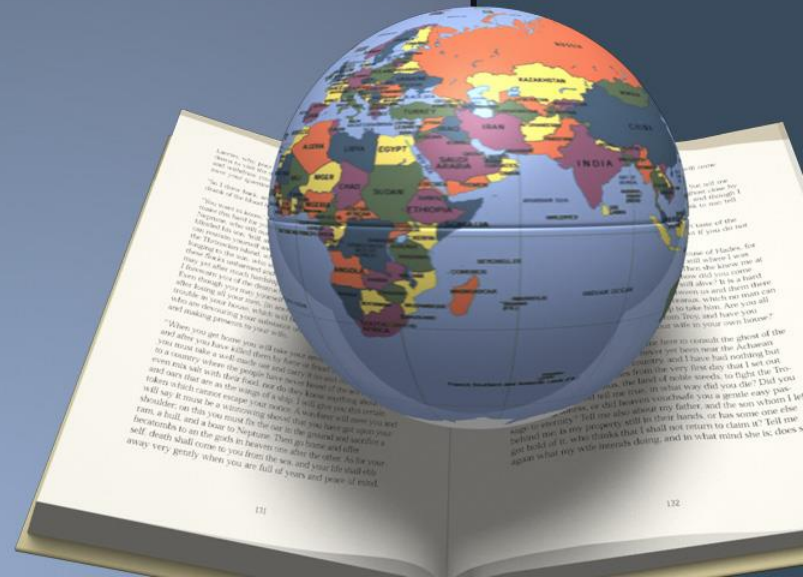
Create mining model MemberCard_Prediction

(

CustomerId long key,
Gender text discrete,
Age long continuous,
Profession text discrete,
MemberCard text discrete predict

)

Using Microsoft_Decision_Trees

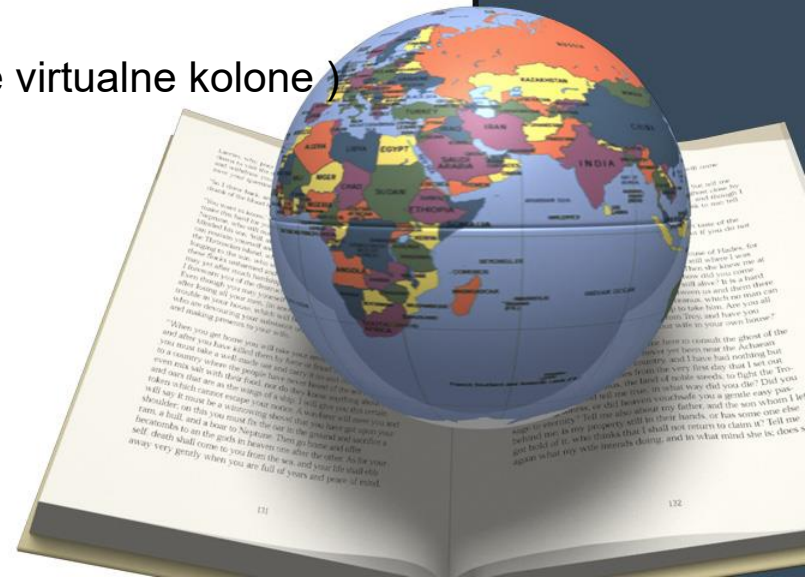




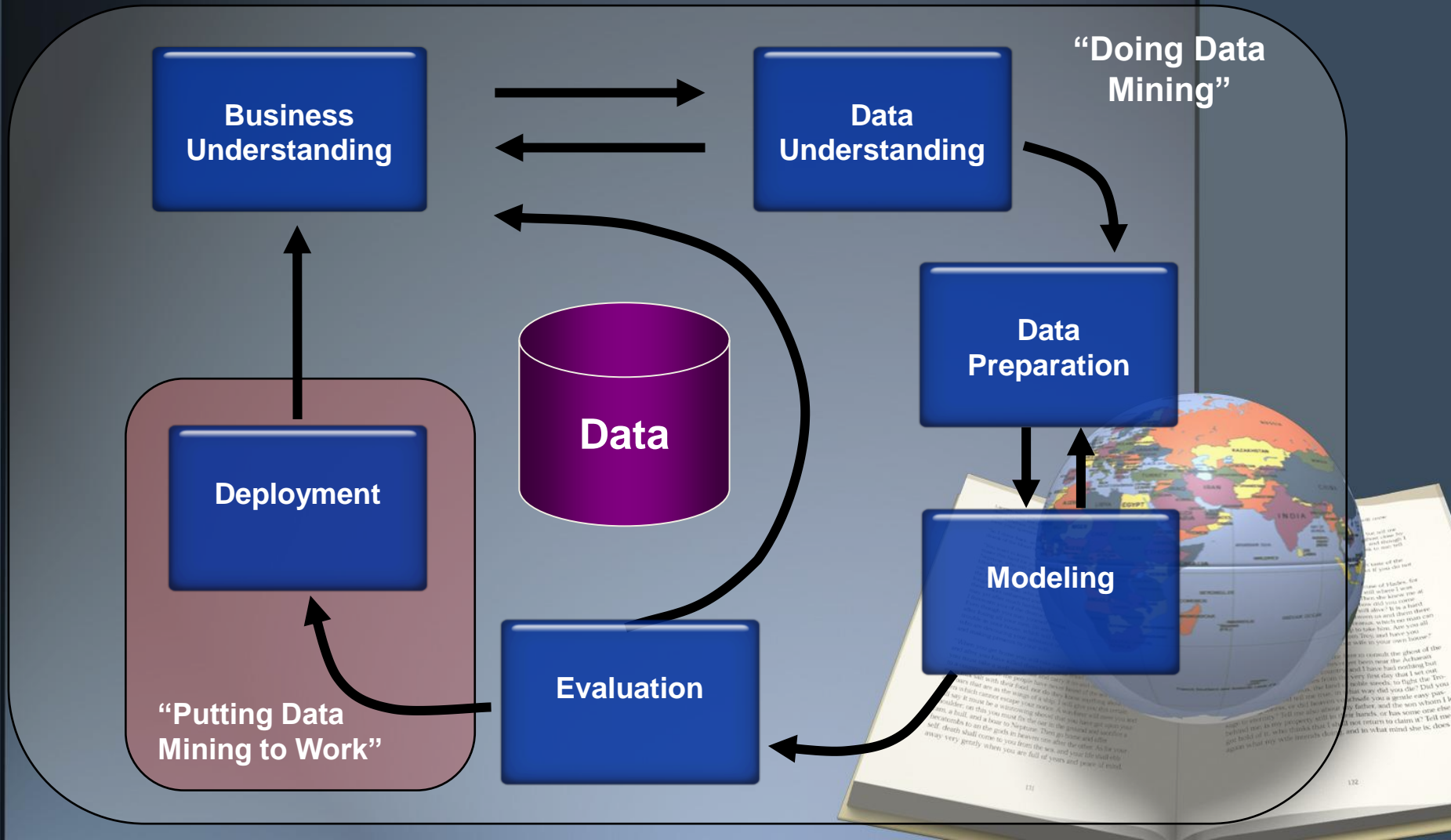
SQL Server Data Mining

SQL Server DM Demo One

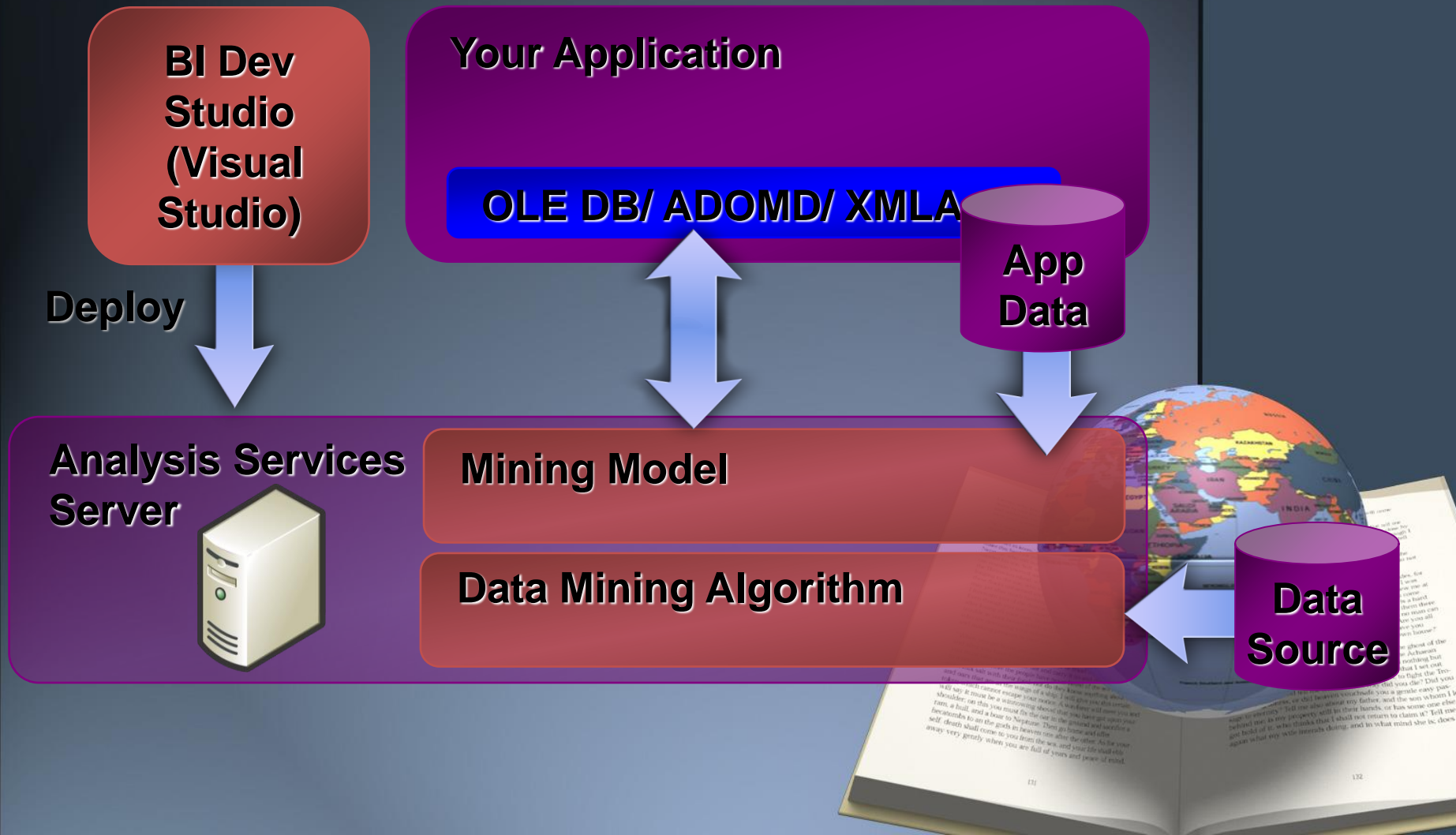
- Što zapravo radimo?
- Osnovni pojmovi
- Demo:
 - Business Intelligence Development Studio (Demo)
 - Offline i Immediate način rada
 - Stvaranje osnovnih objekata
 - Stvaranje izvora podataka (Data Source)
 - DataSource View
 - Named Queries (dodatne izračunate virtualne kolone)
 - DataSource View Diagrami



Što zapravo radimo?



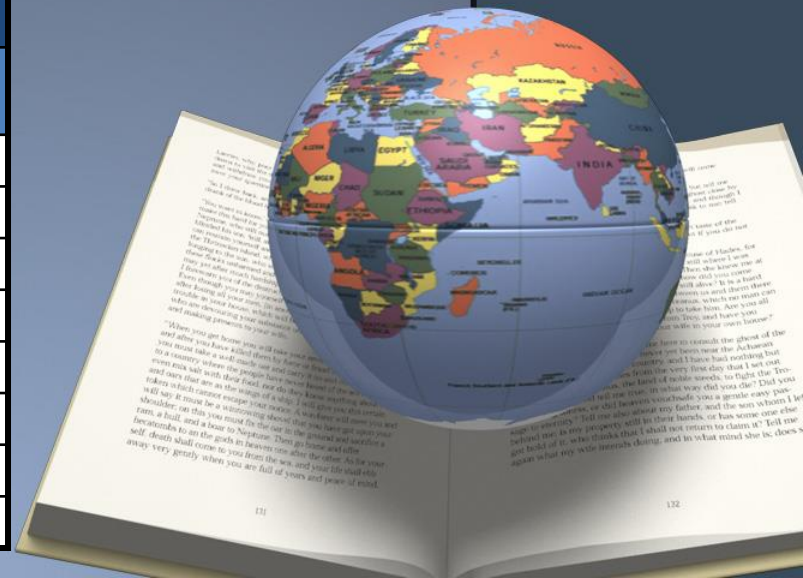
Što zapravo radimo?



Osnovni pojmovi

- Slučaj (Case)
 - Rudarenje podataka se bavi analizom slučajeva
 - Osnovna cjelina informacije
 - Sastoji se niza atributa (varijabli) koje mogu imati više stanja (npr. atribut “spol” može imati stanja “m” i “f”)
 - Slučaj može biti jednostavan ili složen – točnije, možemo imati ugniježdene slučajeve

CustID	Age	Marital Status	Car Purchases	
			Seq ID	Brand
1	35	M	1	Porch-A
			2	Bamborgini
			3	Kexus
2	20	S	1	Wagen
			2	Voovo
			3	Voovo
3	57	M	1	Voovo
			2	T-Yota



Osnovni pojmovi

- Ključ slučaja (Case key)
 - Atribut koji jednoznačno identificira svaki slučaj
 - Može biti jednostavan ili kompozitan
 - Ne sadržava podatke bitne za algoritam rudarenja podataka (samo je identifikator)
- Ključ ugniježđenog slučaja (Nested case key)
 - Ugniježđeni ključ je, s druge strane, najvažniji atribut ugniježđenog dijela
 - On nije identifikator već ključni podatak
 - Ostali atributi samo služe detaljnijem opisu ugniježđenog ključa



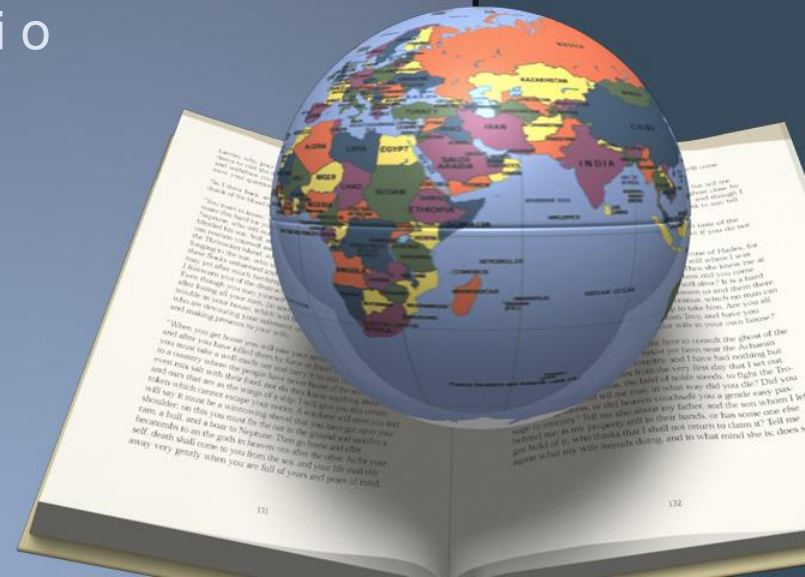
Osnovni pojmovi

- Tablica slučaja (Case table)
 - Tablica koja sadržava podatke o slučaju
- Tablica ugniježđenog slučaja (Nested case table)
 - Tablica koja sadržava podatke o ugniježđenom slučaju
 - To je često transakcijska tablica (povijest kupnje, web logovi, ...)
 - Povezuje se sa tablicom slučaja preko ključa slučaja



Osnovni pojmovi

- Skalarna kolona (Scalar column)
 - Naziva se još i varijabla ili atribut u statistici
 - U ovisnosti o upotrebi razlikujemo četiri vrste:
 - Ključ
 - Ulazna
 - Predviđajuća (cilj modela rudarenja podataka)
 - Predviđajuća i ulazna zajedno
 - Način korištenja ovih kolona ovisi o upotrijebljenom algoritmu



Osnovni pojmovi

- Tablična kolona (Table column)
 - Posebna vrsta kolone koja sadržava cijelu tablicu
 - Najčešći primjer iz komercijalnog svijeta je tablična kolona kupnji koja sadržava podatke o svim transakcijama pojedine osobe (entiteta)
- Model rudarenja podataka (Data mining model)
 - Sadržava ključeve, ulazne i predviđajuće kolone
 - Svaki model povezan je sa korištenim algoritmom nad kojim je stvoren (treniran)
 - Treniranje označava postupak stvaranja uzoraka korišnjem testnog skupa podataka
 - Nakon treninga modela on sadržava uzorke otkrivene u podacima



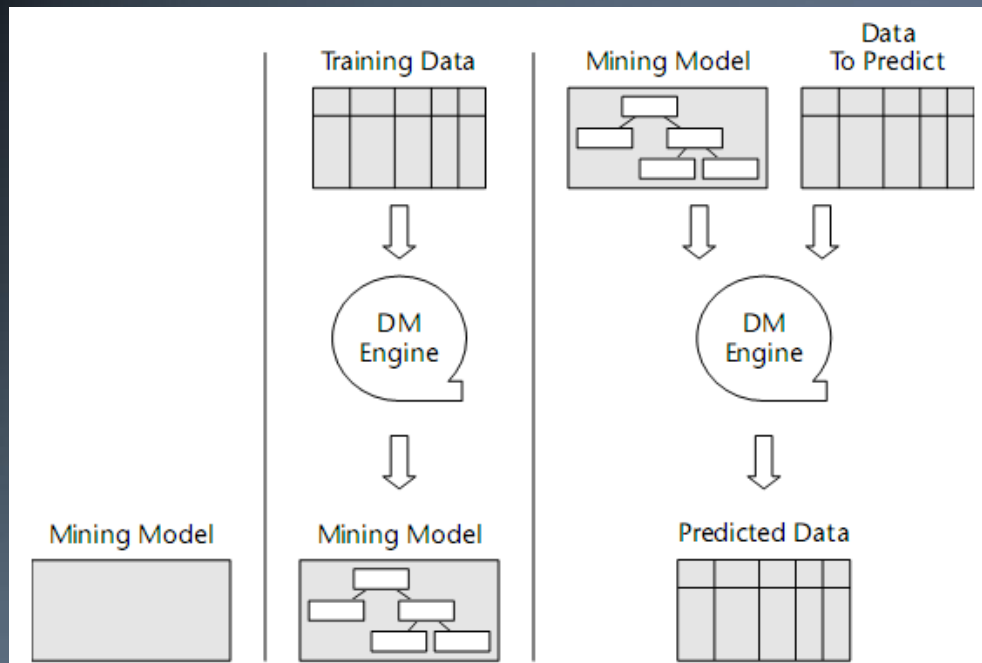
Osnovni pojmovi

- Stvaranje modela (Model creation)
 - Postupak stvaranja novog (ne-treniranog) modela – slično kao stvaranje prazne tablice u bazi
- Treniranje modela (Model training)
 - Naziva se i procesiranje modela
 - Koristi se za pozivanje algoritma rudarenja podataka i stvaranje pronađenih uzoraka iz podataka u skupu podataka za trening



Osnovni pojmovi

- Predviđanje iz modela (Model prediction)
 - Postupak u kojem primjenjujemo skup podataka za predviđanje na trenirani model



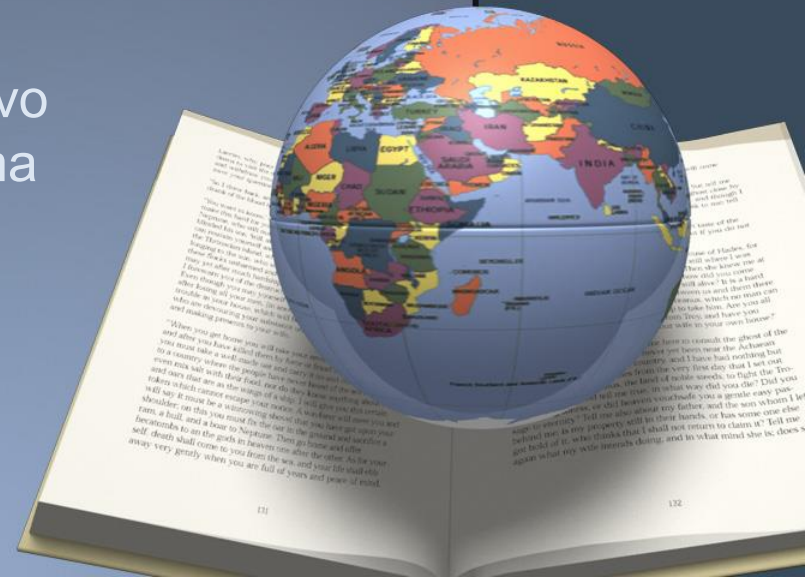
DEMO

- Osnovni alati korišteni u Microsoftovoj izvedbi rudarenja podataka
 - Analysis Services
 - BI Development Studio (Visual Studio 2005)
- Napomena o izvorima podataka:
 - Bilo što – bitni su ČISTI podaci, izvor je nebitan
 - Tekst datoteke
 - OLTP baza podataka
 - Skladište podataka
 - ODBC/OLEDb izvori
 - OLAP kocke
 - ...



DEMO

- Za potrebe ove prezentacije koristimo gotove skupove podataka iz knjige od Wiley&Sons-a
 - Priprema podataka je “skoro” uvijek jedan od najvažnijih i najtežih dijelova projekta rudarenja podataka!
 - Također je i najosjetljiviji – da rudarenje podataka ima smisla, potrebno je raditi sa pravim podacima, što dovodi u pitanje privatnost takvih podataka i njihovo korištenje u javnim prezentacijama



DEMO

- Immediate način rada
 - Direktno spojeni na server – objekti se otvaraju na serveru te se sve promjene tamo i spremaju
 - Iako je ovo intuitivno rješenje mogući su i problemi:
 - Svaka promjena modela ga postavlja u neprocesirano stanje na serveru
 - Paralelne promjene od više korisnika (dobiti ćete upozorenje ako pokušate snimiti preko već promijenjenog objekta)
 - Potrebne su dozvole administratora baze



DEMO

- Offline način rada
 - Podaci spremljeni lokalno u obliku XML datoteka
 - Deploy se pokreće ručno, stvara se skripta i pokreće na odredišnom serveru
 - Također je omogućeno korištenje softvera za verzioniranje
 - Deploy prepisuje sve promjene na odredišnom serveru
 - Radi toga što deploy briše staru i stvara novu bazu, potrebno je imati dozvole Server Administratora



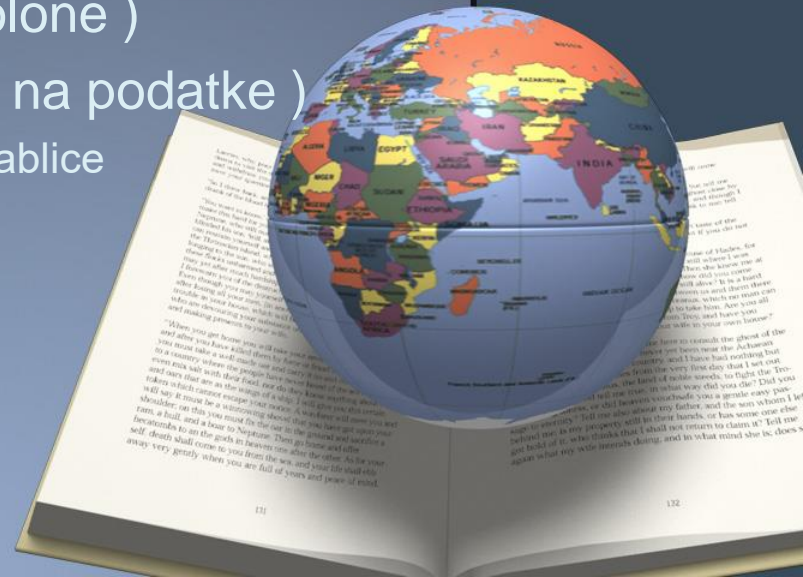
DEMO

- Dodatne mogućnosti koje treba pogledati
 - [Project] / Properties / [Project] Property Pages
 - Processing Option
 - Target server
- Data Source
 - Poveznica na izvor podataka
 - Treba paziti na:
 - Smještaj podataka (ako je bazirana na datotekama (Access/Excel) treba vidjeti da li je isto na serveru
 - Sigurnost (da li je isti korisnik kojim se autorizira u razvojnoj okolini i na produkciji)



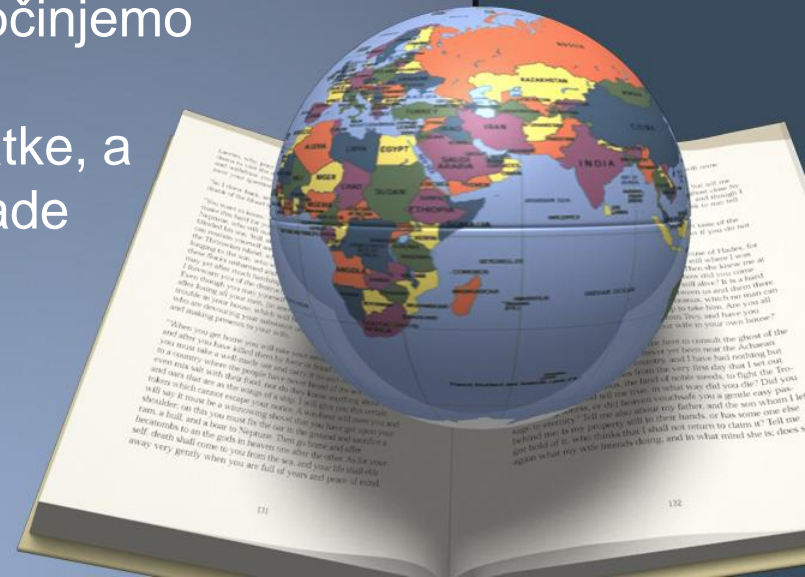
DEMO

- Data Source View
 - Abstraktni klijentski pogled na podatke
 - Dopušta izbor, organizaciju, pretragu i (donekle) manipulaciju podacima
 - Najvažnija stvar koju treba identificirati je tablica slučajeva
 - Potrebno je prikazati i tablice ugniježđenih slučajeva i sve tablice koje pružaju dodatne informacije o tablici slučajeva
 - Named Calculations (dodatne kolone)
 - Named Queries (virtualni pogled na podatke)
 - Ne prenose se ključevi iz originalne tablice
 - DSV dijagrami - dijelovi DSV-a



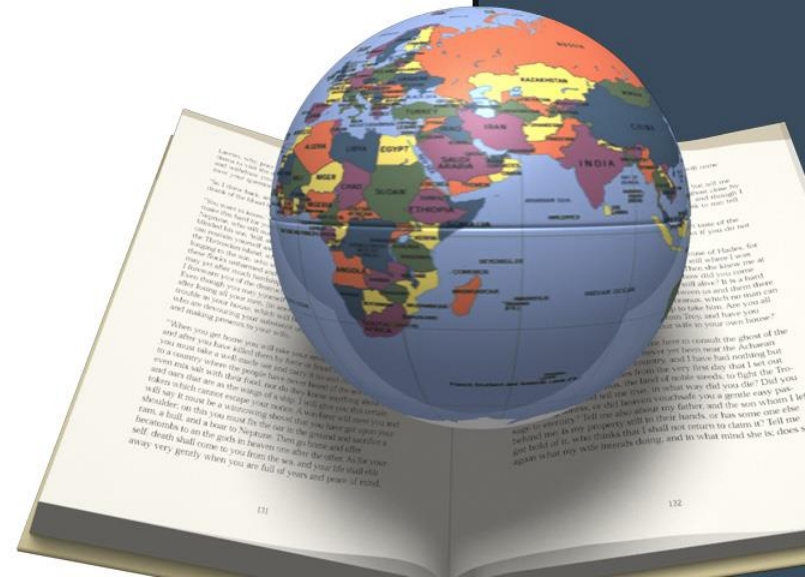
Demo

- Opcija “Exploring data”
 - Korištenjem OWC-a (Office Web Components) možemo pregledavati podatke na više načina:
 - Uzorak podataka (default 5000 zapisa)
 - Pivot table (sami ga slažemo)
 - Pivot chart (sami ga slažemo)
 - Chart
 - Ovaj dio je iznimno bitan jer tu počinjemo upoznavati svoje podatke
 - U praksi, dok ne upoznamo podatke, a i prilikom iterativnog procesa izrade modela ovdje provodimo puno vremena 😊



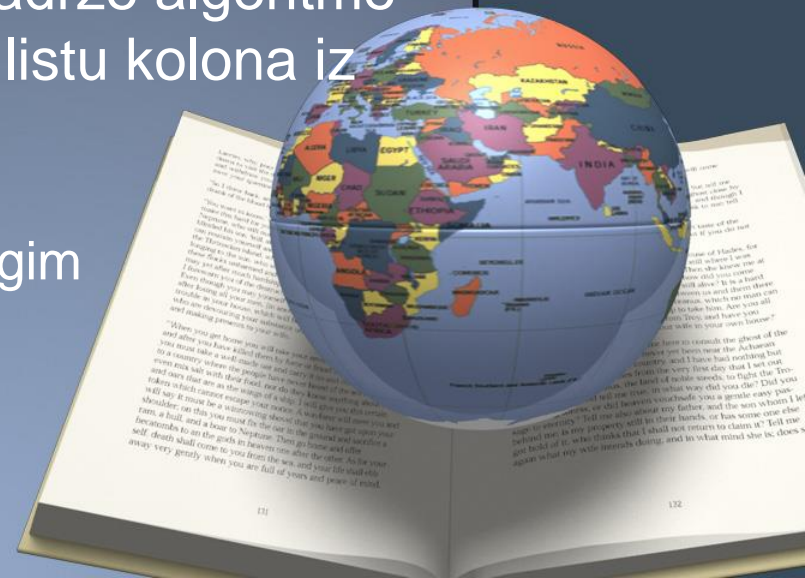
SQL Server DM Demo Two

- Struktura (“data mining structures”) i modeli (“data mining models”)
- Data Mining Wizard
- Data Mining Designer
- Procesiranje modela
- Interpretacija rezultata (Model Viewers)



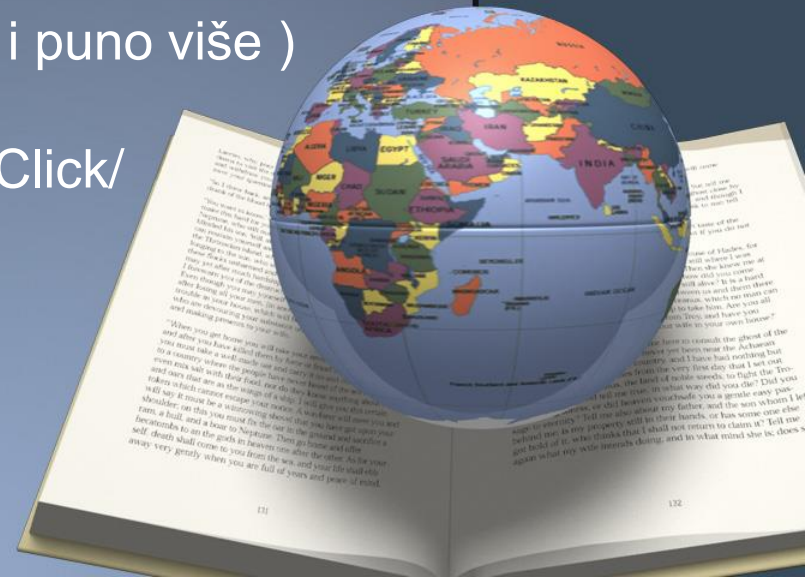
Demo

- Strukture (“data mining structures”) sadrže domenu (promatrano područje)
 - Sadrže kolone (i tipove i sadržaj), podatke o izvoru podataka, ...
 - Također sadrži listu modela koji koriste kolone iz te strukture
- Modeli (“data mining models”) sadrže algoritme sa definiranim parametrima kao i listu kolona iz strukture
 - Svaki model može koristiti drugi algoritam ili isti alogoritam sa drugim parametrima ili kolonama



Demo

- Data Mining Wizard
 - Stvara strukture i izdvaja podatke za trening te model koji primjenjuje dani algoritam nad kolonama (podacima) i definira način korištenja tih kolona (specifično za svaki algoritam)
 - Iako čarobnjak skriva neke korake od nas nužno je da smo svjesni svega što se dešava u pozadini
 - Sve stvari koje radi čarobnjak (a i puno više) moguće je raditi preko DMX-a
 - [Mining Structures]/Right Mouse Click/[New Data Mining Structure]



Demo

- Napomene:
 - Prilikom izbora koja kolona predstavlja ugniježđeni ključ moramo se uvijek sjetiti da algoritmi rudarenja podataka traže uzorke među slučajevima – to je razlog zašto je primarni ključ loš izbor za ugniježđeni ključ – jedinstven je i ne sadržava zajedničke povezeve podatke
 - Korištenje velikog broja ulaznih kolona povećava vrijeme procesiranja (a s time i zahtjeve za resurse) – stoga je potrebno poznavati podatke da bi mogli odrediti koje kolone najviše utječu na ciljanu varijablu



Demo

- Data Mining Designer se sastoji od pet dijelova:
 - Mining Structure editor
 - Omogućava promjene koje nisu moguće kroz čarobnjak
 - One uključuju postavljanje dodatnih diskretizacijskih parametara, višestruko dodavanje istih kolona u strukturu, kreiranje hijerarhija, ...
 - Mining Models
 - Služi za podešavanje parametara modela te nadzor nad svim modelima definiranim nad danom strukturom
 - Omogućava izvršavanje više različitih modela koje koriste iste ili različite alogoritme te “isprobavanje kombinacija”



Demo

– Procesiranje modela

- Process Full
- Process Default
- Unprocess
- Process Structure
- Process Clear Structure

– Mining Model Viewer

- Pruža prikaz različitih informacija i omogućava analizu podataka
- Različit je za svaki algoritam
- Mogu biti u obliku dijagrama ili tablica
- Moguća implementacija vanjski vizualizacija



Demo

– Mining Accuracy Chart

- Pomaže (ali ne apsolutno!) u određivanju točnosti dobivenog modela
- Accuracy chart - uspoređuje rezultate predviđanja modela sa poznatima podacima
- Profit chat – isto kao i accuracy chart, samo što dodatno omogućava definiranje troškova – služi za određivanje točke maksimalnog povrata sredstava
- Classification matrix – pokazuje koliko puta je predviđanje bilo točno i koji je vraćeni rezultat ako je bilo krivo

– Lift Chart

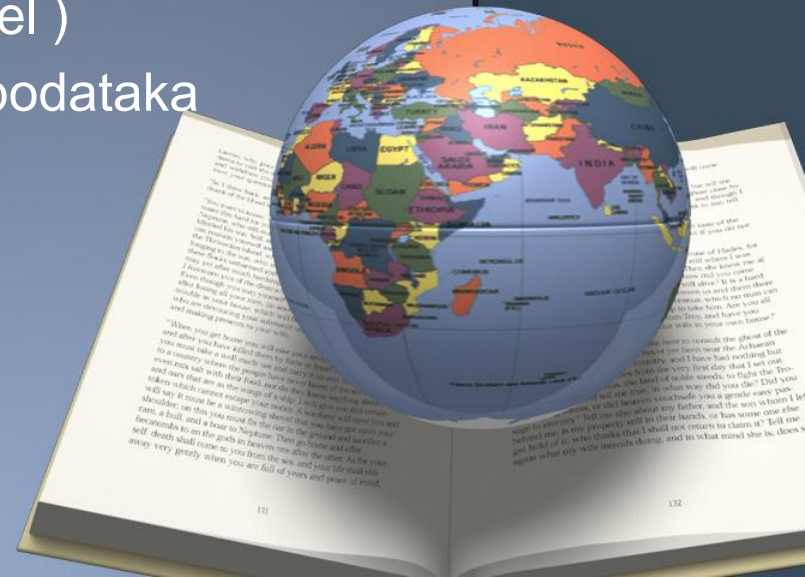
– Mining Model Prediction Builder

- Služi za izgradnju i editiranje predikcijskih upita, pregled i snimanje rezultata



Za kraj...

- Što nismo vidjeli – skoro sve ☺
 - Detaljna objašnjenja algoritama i metoda
 - Korištenje SSIS-a i rudarenja podataka
 - Programiranje (ADOMD.NET)
 - DMX na primjerima
 - Implementaciju u ASP.NET aplikaciji
 - Korištenje u front-end alatu (Excel)
 - I ono najvažnije...razumijevanje podataka
 - Ali tu vam nitko ne može pomoći ☺



Cilj predavanja

- Dati uvod u rudarenje podataka, upoznati vas sa osnovnim pojmovima i prikazati SQL Server Data Mining alat
- No, alat je samo alat, potrebno je shvatiti i naučiti koncepte a za to treba puno vremena i truda
 - Nema “jednog” načina i “kuharice”
 - Jedno od najjačih specijalizacija u IT-u danas, ali ujedno i najtežih
 - Tržište...beskonačno ☺
 - Mogu koristiti i “veliki” i “mali”
 - Sve nas čeka još puno učenja ☺





Hvala na pažnji!