

6. Rudarenje podataka (Data Mining)

Zekić- Sušac

1

Što ćete naučiti u ovom poglavlju?

- Što je rudarenje (ili kopanje) podataka (eng. Data mining)?
- Koje su metode data mining-a
- Kako koristiti kvantitativne sofisticirane metode koje se koriste za traženje skrivenih zakonitosti u podacima tvrtke
- Koji su alati za datamining

Zekić- Sušac

2

Rudarenje podataka (Data mining)



Poznato otkriće rudarenja podataka za P&G:
Pelene kupuju očevi petkom poslijepodne, i uz njih obično kupe pivo za gledanje utakmice preko vikenda!

Korist od ovog otkrića: ↓

Reklame za pelene treba slati očevima a ne majkama!
Zalihe treba planirati tako da petkom bude najviše robe na policama.

Zekić- Sušac

3

Što je rudarenje podataka?

- Cilj rudarenja podataka je identificirati vrijedne nove, potencijalno korisne veze i uzorce u postojećim podacima (Jackson, 2002)

Rudarenje podataka uključuje primjenu tehnika:

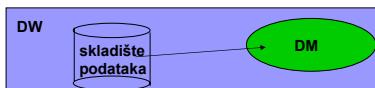
- numeričke analize
- traženje uzorka i strojno učenje,
- stabla odlučivanja, **neuronske mreže**, genetički algoritmi i dr.

Zekić- Sušac

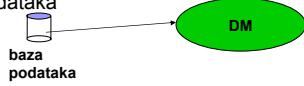
4

Data Mining i Data Warehousing

a) ako DW postoji, DM može biti dio DW



b) ako DW ne postoji, DM se može koristiti na raspoloživoj bazi podataka



Zekić- Sušac

5

Data Mining i OLAP

- Razlika: OLAP je tehnologija za sumiranje i agregaciju podataka, a DM je metodologija za traženje uzorka i veza među tim podacima (Jackson, 2002)
- Integracija OLAP-a i DM: OLAM (On-line Analytical Mining)

Zekić- Sušac

6

Što možemo napraviti data mining-om?

- segmentirati tržište
- otkriti profil tipičnog klijenta određene vrste proizvoda
- sklonosti klijenta kupnji i motivatore za cross selling
- napraviti model potrošačke košarice (market basket)
- detektirati potencijalne prijevare na prodajnim mjestima
- predvidjeti trendove pojave, npr. kretanja profita
- otkriti sličnosti među tržišnim kategorijama, i dr.

Zekić- Sušac

7

Primjeri i upotrebe data mining-a

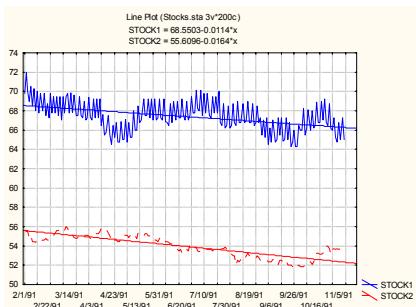
- Ako smo vlasnik ili manager u lancu trgovina i želimo saznati postoji li pravilnost u trendovima prodaje u nekoj od trgovina početkom tjedna ili krajem tjedna i za što
 - ili želimo saznati da li će se povećati prodaja nekog artikla X ako se prodaje u paru s artiklom Y
- Ako smo manager u banci i želimo saznati profil klijenata koji će vjerojatno otici u minus na tekućem računu ili profil onih koji neće vraćati kredit
- Ako smo vlasnik proizvodne tvrtke i želimo znati koje su karakteristike proizvoda koje imaju produž na tržištu,
- i dr.

Zekić- Sušac

8

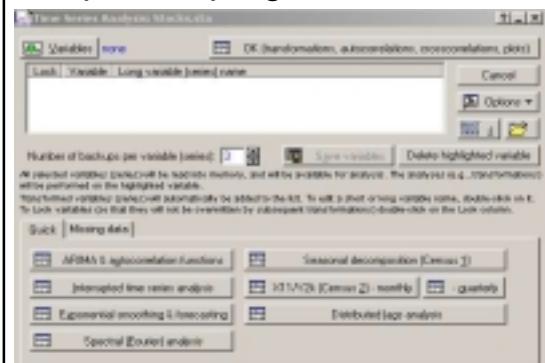
Prognostičke metode

Primjer 1. Predviđanje cijene dionica



9

Napredne prognostičke metode



10

Veza DM i poslovnih ciljeva



Tvrte često koriste reklamu putem pošte direktno na adrese potrošača.

Stopa reakcije na ovaku reklamu u SAD je oko 2%.

DM može pomoći u smanjivanju ovog uzaludnog napora i troška **predviđanjem reakcije potrošača, i time slanjem reklama samo onim skupinama za koje postoji velika vjerojatnost da će kupiti proizvod.**

Zekić- Sušac

11

Primjeri primjene DM

- Trgovina - Potrošačka košarica – za identificiranje artikala koji se prodaju zajedno, za određivanje načina oglašavanja artikala, za predviđanje efektivnosti promocije i reklame
- Banke i osiguravajuće kuće – za identificiranje "dobrih" i "loših" klijenata
- Medicinske ustanove – za predviđanje uspješnosti operacija, medicinskih testova, ili lijekova
- Proizvodne tvrtke – za procjenu vjerojatnosti kvarova proizvoda
- i slično.

Zekić- Sušac

12

Slučaj primjene DM

- Tvrta Petrol napravila je dvije analize:
 - analizu veleprodaje i identificirala kupce za koje je vjerojatno da će biti slabi platise u budućnosti
 - analizu računa kupaca kako bi otkrila u kojim danima u tjednu treba imati veće zalihe goriva
- Chase Manhattan bank – koristi DM za određivanje kojim klijentima dodijeliti kredit

Zekić- Sušac

13

Koraci u primjeni DM

Koraci prema CRISP-DM metodologiji:

- razumijevanje poslovnih ciljeva (što želimo postići)
- razumijevanje podataka (prikupiti, opisati i verificirati podatke)
- priprema podataka (izabrati, očistiti, integrirati i formatirati podatke)
- modeliranje (izbor tehnike, testiranje, izgradnja modela, ocjena modela)
- evaluacija (ocjena rezultata, pregled napravljenog, plan za sljedeći korak)
- primjena plana, izrada konačnog izvještaja, revizija projekta

Zekić- Sušac

14

Tehnike koje koristi DM

- Neuronske mreže (Neural Networks)
- Zaključivanje na temelju slučajeva (Case-Based Reasoning – CBR)
- Genetički algoritmi (Genetic Algorithms)
- Stabla odlučivanja (Decision Trees)
- Asocijacijska pravila (Association Rules)
- Statističke metode: deskriptivne i vizualizacijske tehnike, klaster analize, korelacijske analize, diskriminantnu analizu, faktorsku analizu, regresijsku analizu, logističku regresiju i dr.)

Zekić- Sušac

15

DM se može koristiti za:

- Sumiranje podataka
- Segmentaciju
- Klasifikaciju
- Predviđanje
- Asocijaciju (prepoznavanje uzorka)

Zekić- Sušac

16

Koju tehniku DM koristiti?

TEHNIKA	Sumiranje podataka	Segmentaciju	Klasifikacija	Predviđanje	Asocijacija
Deskriptivna stat. i vizualizacija	+	+			+
Korelacijska analiza					+
Klasteri		+			
Diskriminantna analiza			+		
Regresijska analiza				+	+
Neuronske mreže		+	+	+	+
CBR					+
Stabla odlučivanja			+	+	
Pravila asocijacije					+

Alati za DM

- Statistički softverski paketi (npr. SAS, Statistica, SPSS, i dr.)
- Matematički softverski paketi (npr. MathLab, Matematica)
- Alati uključeni u skladištenje podataka (OLAP) ili sustav za upravljanje bazom podataka (npr. Microsoft SQL Server Business Intelligence – uključuje i Enterprise Miner)
- Specijalizirani alati za općenite ili poslovne primjene (npr. DataMiner, IntelliMiner, i sl.)

Zekić- Sušac

18

SQL Server Data Mining



Zekić- Sušac

19

Budućnost upotrebe DM u poslovanju

- DM metode se integriraju u alate za upravljanje bazama podataka, te tako softverski postaju raspoložive brojnim tvrtkama
- Uklapljene u web alate postaju dostupne svim uposlenima
- Problem:
- Još uvijek zahtijevaju visoku razinu stručnog znanja o metodama, pa se razvijaju alati koji će omogućiti korištenje menadžerima i bez tog specijaliziranog znanja stručnog znanja
- Radi se na edukaciji korisnika

Zekić- Sušac

20

Literatura:

- Čerić, V., Varga, M., Birola, H.: Informacijska tehnologija u poslovanju, Element, Zagreb, 2006.
- Bidgoli, H., Intelligent Management Support System, Quorum, London, 1998.
- Jackson, J., Data Mining: A Conceptual Overview, Communications of the AIS, Volume 8, 2002, pp. 267-296.
- Klepac, G., Mršić, L., Poslovna inteligencija kroz poslovne slučajeve, Lider Press, Tim Press, Zagreb, 2006.
- Panian, Ž., Klepac, G., Poslovna inteligencija, Masmedia, Zagreb, 2003.

Zekić- Sušac

21