



WALLACHIA eHUB ROADSHOW

NOȚIUNI INTRODUCTIVE DESPRE DATA MINING

Ce este Data Mining?

- Procesul de extragere de informații netriviabile, necunoscute anterior și potențial utile din date;
- Procesul prin care informațiile și cunoștințele sunt extrase din volume mari de date folosind tehnici care sunt mai mult decât o simplă căutare în date;
- Analiza unor seturi de date (de dimensiune mare), observaționale, care are scopul de a identifica relații ascunse și de a rezuma datele în moduri noi care sunt inteligibile și utilizabile pentru proprietarul datelor

Principalele surse de date pe care se pot aplica tehnici data mining:

- Depozite de date - colecții omogene de date obținute în urma integrării mai multor surse de date, uneori diferite;
- Fluxuri de date – de exemplu analiza traficului pe internet, a datelor generate de sistemele de supraveghere în timp real, a datelor generate în urma proceselor de producție industrială și a altor tipuri de date generate de activități cu o dinamică însemnată;
- Serii de timp și secvențe - constă în secvențe de valori sau evenimente obținute în urma unor măsurători repetate de-a lungul timpului;
- Date cu structură de grafuri și rețele - permit modelarea unor structuri eterogene, cum ar fi datele de tip text, datele de tip web, rețelele sociale, și a altor date care prezintă un grad ridicat de complexitate;
- Date mai puțin convenționale - date spațiale, multimedia, text și web;

Provocări în implementarea data mining:

- Disponibilitatea datelor – existența unor volume de date suficient de mari pentru a fi exploatate cu ajutorul tehnicilor specializate data mining; în cazul în care nu există suficiente date, rezultatele aplicării unor tehnici specializate de data mining nu sunt satisfăcătoare;
- Calitatea datelor – calitatea rezultatelor obținute prin aplicarea unor tehnici avansate de analiză a



WALLACHIA eHUB ROADSHOW

datelor nu poate fi mai bună decât calitatea datelor asupra cărora acestea se aplică; Prin calitatea datelor se pot înțelege acuratețea, completitudinea, corectitudinea, consistența și actualitatea acestora.

- Complexitatea datelor – datele existente pot fi în formate diferite, unele dintre acestea dificil de procesat, lucru care îngreunează/ limitează procesul de analiză a acestora.
- Securitatea datelor – asigurarea securității datelor asupra cărora se dorește aplicarea de tehnici data mining este o problemă de maximă importanță, mai ales în contextul legislației tot mai restrictive cu privire la prelucrarea datelor cu caracter personal.

Exemple de utilizare a data mining pentru optimizarea activităților unei organizații:

- Analiza coșului de cumpărături (Market basket analysis) - este o problemă clasică de data mining care explorează tiparele de apariție împreună a unor produse în coșul de cumpărături și determină puterea legăturii dintre acestea. Această problemă mai este cunoscută și ca frequent itemset mining sau ca analiza asocierilor. Analiza coșului de cumpărături valorifică volumele mari de date rezultate din comerțul cu amănuntul și ajută comercianții să înțeleagă mai bine comportamentul clienților prin identificarea relațiilor dintre articolele cumpărate de aceștia. Astfel, comercianții pot avea grijă ca produsele care se vând împreună să fie plasate pe rafturi învecinate sau pot forma pachete de produse special concepute pentru a stimula vânzările.
- Analiza perioadelor cu vânzare mai ridicată sau mai scăzută, la nivel general sau la nivel de produs. În acest fel comercianții își pot optimiza stocurile astfel încât să nu blocheze resurse importante care ar putea fi utilizate pentru desfășurarea altor activități.
- Stabilirea unor tendințe pentru a asista managementul organizațiilor în adoptarea unor politici benefice pe termen lung – spre exemplu creșterea vânzărilor pentru anumite produse sau scăderea vânzărilor pentru altele.
- Măsurarea eficacității campaniilor de promovare prin corelarea sumelor investite pentru realizarea acestora cu rezultatele obținute. În acest fel companiile pot optimiza campaniile de promovare pe care le desfășoară pentru a le mări impactul.