

# E-COMMERCE

## Curs 9



- 1. Data mining - definire**
- 2. Data mining și personalizarea**
- 3. Web Usage Mining**
- 4. Tipuri de tehnici data mining pentru e-commerce**
- 5. Tipuri de reguli data mining pentru e-commerce**
- 6. Studiu de caz: ebay**



## Definiție generală

---

Data mining = proces de:

- analiză a unor cantități mari de date
- extragere a informațiilor relevante din aceste date folosind metode matematice și statistice

Termenul este utilizat de obicei de către:

- organizațiile ce se ocupă cu prelucrarea informațiilor despre companii
- analiștii financiari

Este folosit din ce în ce mai mult și în domeniul științific pentru extragerea informațiilor din seturi mari de date, generate de exemplu de experimente moderne.



# Data mining din perspectiva e-commerce

= un set de tehnici analitice care caută după anumite șabloane într-o bază de date sau în data warehouse, încercând să definească un model de comportament al clientului.

Astfel datele unui site web pot fi “săpate” astfel încât să se poată dezvolta profile ale vizitatorilor și ale clienților.



Profilul unui utilizator = set de reguli ce descrie un comportament tipic al unui client sau grup de clienți pe site. Profilele ajută la stabilirea unor șabloane specifice care pot fi identificate.



### *Exemplu:*

- orice tranzacție financiară ce are loc este procesată de o aplicație *data mining* pentru detectarea fraudelor.
- Companiile de telefonie mobilă monitorizează îndeaproape utilizarea telefoanelor mobile pentru a detecta șabloane tipice telefoanelor furate sau a apelurilor telefonice neobișnuite.



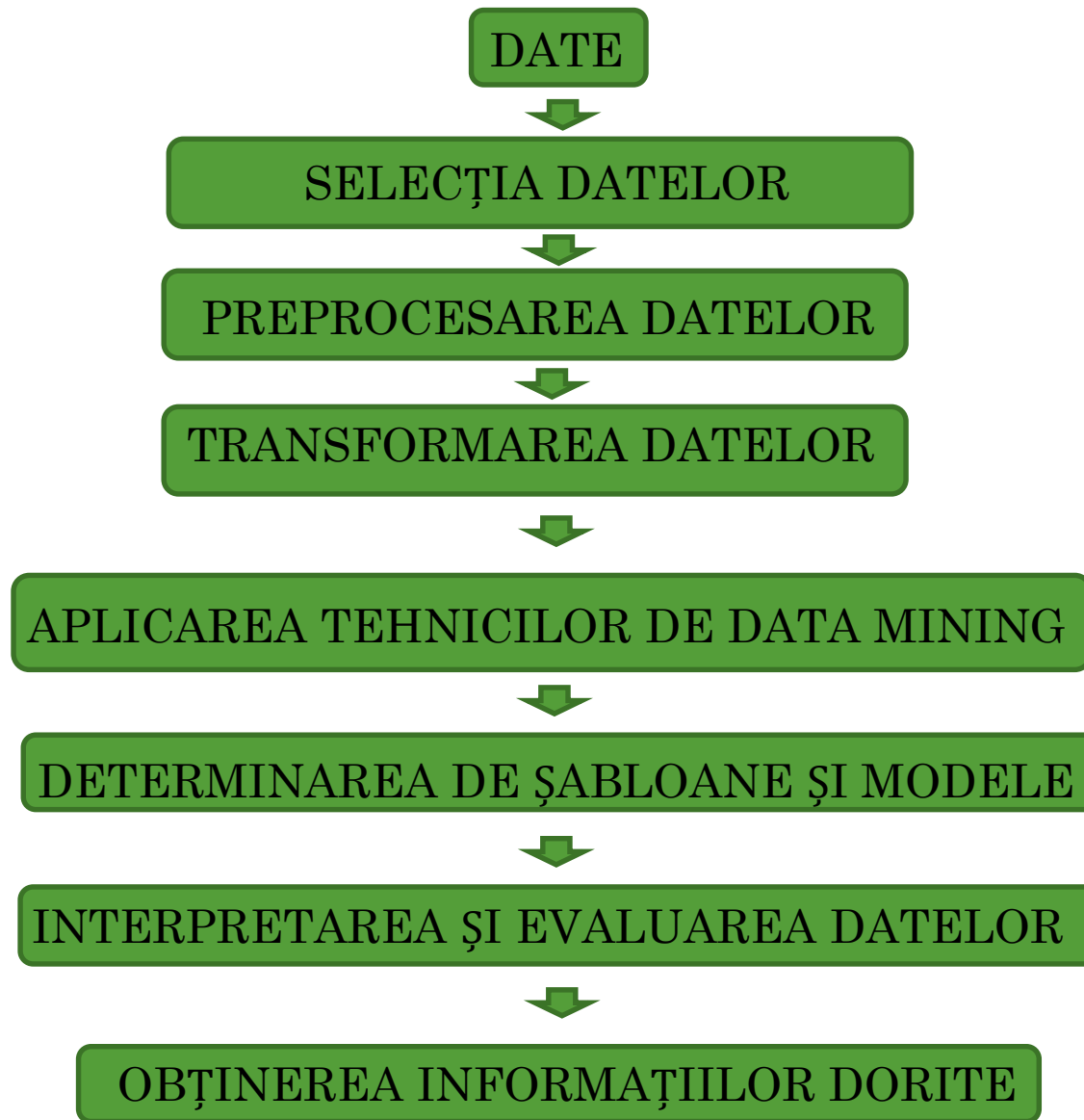
Companiile financiare și cele de telefonie mobilă folosesc tehnici de data mining pentru a dezvolta profilul unei fraude. O tranzacție este oprită dacă este detectat un comportament asociat unui astfel de profil.

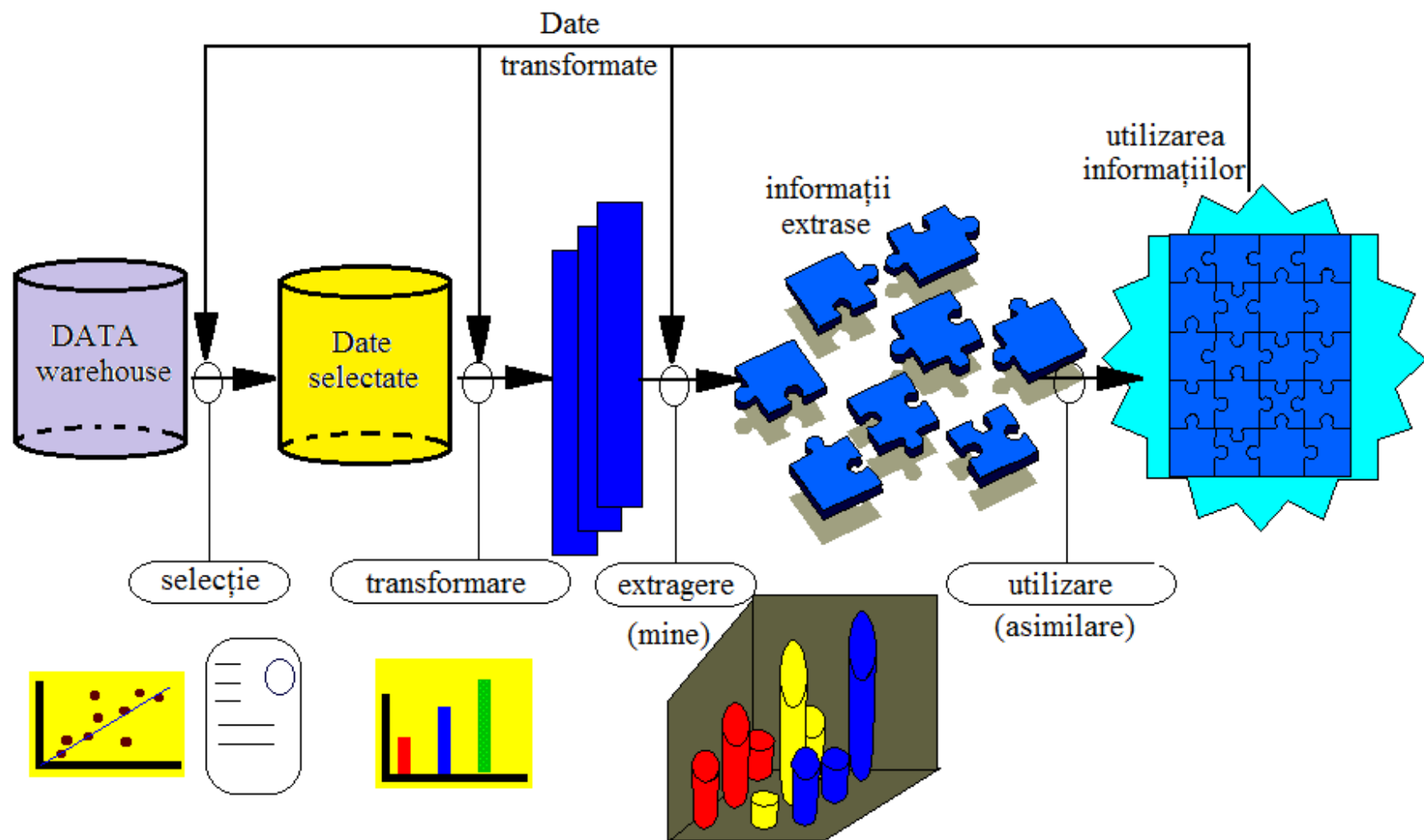
Tehnicile de data mining:

- ✓ extrag datele tranzactionale
- ✓ stocheaza si administreaza datele într-un sistem multidimensional de baze de date
- ✓ analizeaza datele
- ✓ prezinta datele sub formă de tabele si grafice



# ETAPE DATA MINING



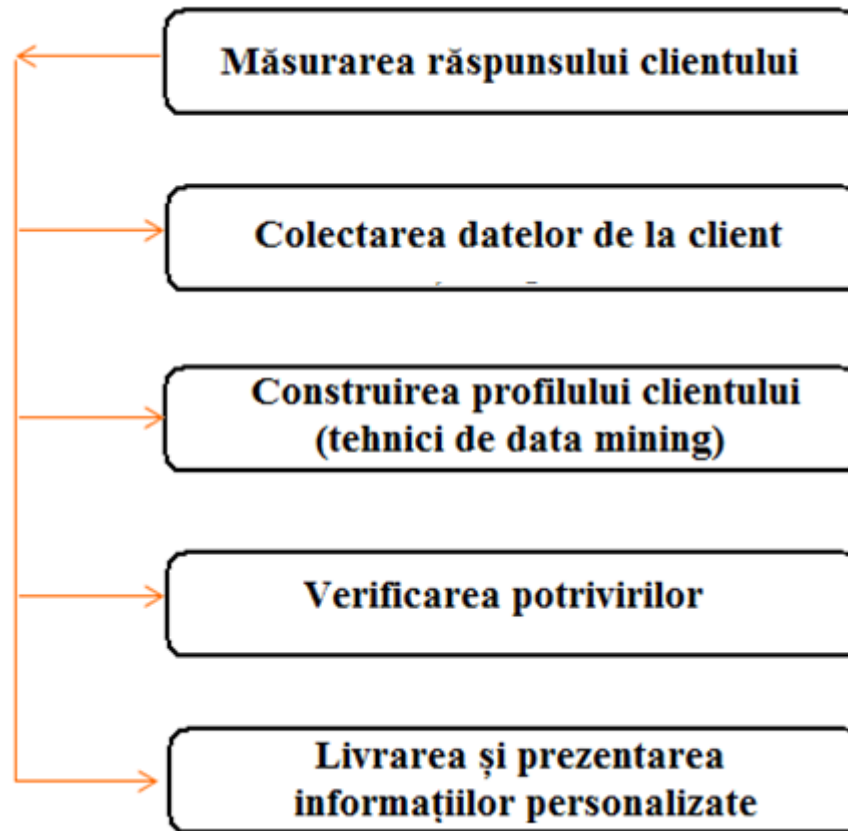


- ❑ Datele de intrare se pot găsi într-o varietate de formate: fișiere text, baze de date relaționale, date semistructurate (de exemplu XML, HTML), imagini, filme etc.
- ❑ Datele se selectează din întreaga colecție de surse.
- ❑ Preprocesarea și transformarea datelor sunt etape ce pot fi realizate chiar și în 60% din timpul total al unui proces de extragere a informațiilor relevante.
- ❑ Există numeroase tehnici de data mining și, de multe ori, sunt testate mai multe metode.
- ❑ La final, informațiile rezultate sunt post-procesate, eliminându-se astfel rezultatele invalide sau neinteresante. Forma de prezentare poate fi sub formă de reguli sau integrate în alte sisteme (de exemplu un sistem pentru detectarea fraudelor).





## Data mining și personalizarea



Personalizarea conținutului și a datelor de marketing se bazează pe tehnici de data mining care produc un profil real al comportamentului utilizatorului.



**Web Usage Mining** este partea cea mai relevantă din punct de vedere al marketing-ului, deoarece explorează căile de navigare și comportamente din timpul vizitei în cadrul unui site web al unei companii.

*Web usage minig* este descris ca aplicarea de tehnici data mining pe logurile de acces web pentru optimizarea sitului web în beneficiul utilizatorilor.

- Metodele de extragere a regulilor de asociere sunt folositoare pentru obținerea corelațiilor dintre diferitele pagini vizitate în timpul unei sesiuni.
- Regulile de asociere secvențiale sau modele de serii de timp pot fi folosite pentru a analiza datele utilizate dintr-un site web luând în calcul o dinamică temporală a utilizării site-ului.
- Web usage mining se bazează în principal pe analiza secvenței de pagini vizitate în cadrul unei sesiuni date, numită *analiza clicurilor web*.
- Informații despre comportamentul de cumpărare al vizitatorilor se pot obține în cadrul site-urilor web de comerț electronic analizând clicurile web.



**Click-stream** = o secvență de pagini web vizualizate de un utilizator, paginile fiind afișate câte una pe rând la un moment dat. În momentul în care un vizitator accesează un site web, serverul reține toate acțiunile făcute de vizitator într-un fișier jurnal(log file).



O sesiune (sesiune a server-ului) reprezintă totalitatea paginilor unui site web din cadrul unei sesiuni de utilizator.



Fiecare click de mouse corespunde unei cereri de pagina web, succesiunea click-urilor corespunde astfel succesiunii link-urilor. Analiza fluxului click-urilor succesive poate fi folosită pentru a înțelege cea mai probabilă cale de navigare într-un site web cu **scopul previziunii online a paginilor pe care un vizitator le va accesa** dată fiind succesiunea link-urilor (calea) pe care a urmat-o până atunci.



## Tipuri de analiză a click-streamurilor:

1. Analiza traficului:
  - ➡ Implică analiza drumului pe care utilizatorul îl urmează prin site bazându-se pe informațiile adunate la nivel de server folosind date din analiza click-urilor.
  - ➡ Înregistrează numărul încărcărilor paginilor în timpul navigării.
2. Analiza e-commerce
  - ➡ Folosește date din click-stream pentru a determina eficiența de marketing a unui site prin cuantificarea comportamentului utilizatorului în timp ce vizitează site-ul.
  - ➡ Înregistrează traducerea navigării vizitatorului pe site în tranzacții de vânzare.
  - ➡ Este indicator al gradului de comoditate al utilizatorului în folosirea interfeței, coșului de cumpărături, plății etc.



## Tipuri de tehnici data mining pentru e-commerce

---

### ❑ Tehnici bazate pe o întrebare specificată (query-driven data mining)

*Ex:* Care este relația dintre un moment al zilei și cumpararea unor anumite produse? ...se explorează data warehouse și se realizează o tabelă ordonată cu primele 10 produse vândute în fiecare oră a zilei. Astfel vânzătorul poate să schimbe conținutul site-ului pentru a stimula vânzările, reliefând anumite produse în anumite momente de timp sau plasând anumite produse în pagina principală care se schimbă în funcție de momentul zilei.

### ❑ Tehnici bazate pe un model (model-driven data mining)

Implică utilizarea unui model care analizează variabilele de interes și astfel se iau anumite decizii.

*Ex:* Vânzătorul dorește să reducă inventarul, eliminând produsele care se dovedesc a fi neprofitabile. Astfel se construiește *un model financiar* ce calculează gradul de profitabilitate al fiecărui produs de pe site..... Astfel se ia o decizie!



## ❑ Tehnici bazate pe reguli (rule-based data mining)

Examinează datele demografice individuale sau ale grupurilor de utilizatori ai site-ului și astfel sunt trasate niște reguli de comportament ale vizitatorilor.

Astfel, în funcție de produsele cumpărate și de valoarea lor, este formată o tabelă a profilelor vizitatorilor, segmentându-se piața în grupuri bine definite.

*Ex:* Utilizatorii care au cheltuit aproximativ 50 Euro în urma unei vizite pe site și au vizualizat articole pentru călătorie, vor avea afișate și oferte de excursii.

Regulile sunt specificate de managerii de marketing

## ❑ Filtre colaborative (collaborative filtering)

Abordează diferit analizarea comportamentului utilizatorului. Nu impune reguli, ci clasifică vizitatorii în foarte multe grupuri caracterizate de un interes comun. Produsele sunt recomandate în funcție de cumpărăturile care s-au făcut recent în grup.

*... dezvoltate de MIT Media Lab*



## Tipuri de reguli data mining pentru e-commerce

---

- Reguli de asociere (relații și dependențe între câmpurile datelor) pentru analiza coșului de cumpărături .
- Reguli de clasificare (maparea elementelor în una sau mai multe categorii predefinite) pentru a găsi potențialii clienți pentru un anumit produs sau clasificarea comportamentului vizitatorilor site-ului .
- Reguli caracteristice (descoperirea specificului unui anumit element) pentru a aborda vânzările încrucișate.
- Reguli secvențiale (modelarea stărilor și șabloanelor unui proces) pentru a detecta căile tipice pe care cumpărătorii tind să le aleagă.
- Clustering (găsirea grupurilor de entități similare) pentru a găsi tipologii asemănătoare de vizitatori, care duc la interesul pentru același produs.



ebay <—> data mining



Procesează petabytes de date îmbunătățind constant 5 arii importante în e-commerce :

1. Căutarea de produse
2. Recomandarea de produse
3. Detectarea produselor similare
4. Detectarea fraudelor
5. Business Intelligence







## 1. Căutarea de produse = “sortarea” produselor în funcție de popularitate. (product ranking)

Rank-ul unui produs crește în funcție de:

- numărul de vizualizări
- numărul de articole vândute

## 2. Recomandarea de produse

Se creează anumite pattern-uri de utilizatori.

Se recomandă cele mai populare produse vizualizate de utilizatorii din aceeași categorie cu tine.





### 3. Produse similare.

Se analizează:

- Proprietățile produsului
- Gama de prețuri
- User-ii din aceeași categorie cu utilizatorul subiect ce au fost atrași

Se returnează produsele similare.





## 4. Detectia fraudelor.

- Vanzatorii plaseaza produsele lor in categorii mai populare, ceea ce le face produsele mai usor de gasit.

*Ebay* analizeaza produsul (folosind data mining), detecteaza aceste fraude si plaseaza produsele respective in categoriile corespunzatoare.

- Cumparatorii folosesc carti de credit furate.

Se analizeaza :

- Profilul si activitatea cumparatorului
- Gama de preturi in care se incadreaza cumparaturile obisnuite.
- Categoriile de produse de care era interesat

Se detecteaza anomalile si se monitorizeaza tranzactiile.





## 5. Business intelligence

- Inventar inteligent. Fiind o gama extraordinar de variată, este foarte greu sa inventariezi și să împarți pe categorii produsele fără un sistem inteligent de clasificare automata.
- Stocuri modificate în timp real.
- *Ebay* investește continuu în data mining
- Raportare
- Procesarea evenimentelor complexe
- Analiza predictivă
- Managementul performanței afacerii
- Tehnici de machine learning și text mining

