

Mihai ANDRONIE

# TEHNICI *DATA MINING* UTILIZATE ÎN SISTEMELE INFORMATICE ECONOMICE

Ediție revăzută și adăugită



Copyright © 2013, **Editura Pro Universitaria**

Toate drepturile asupra prezentei ediții aparțin  
**Editurii Pro Universitaria**

Nicio parte din acest volum nu poate fi copiată fără acordul scris al  
**Editurii Pro Universitaria**

**Descrierea CIP a Bibliotecii Naționale a României**

**ANDRONIE, MIHAI**

**Tehnici data mining utilizate în sistemele informatice  
economice** / Mihai Andronie. - Ed. rev. și adăug. - București : Pro  
Universitaria, 2013

Bibliogr.

ISBN 978-606-647-589-1

004.65:33(075.8)

## INTRODUCERE

Lucrarea de față tratează problema utilizării tehnicilor *data mining* în cadrul sistemelor informatice economice în vederea optimizării proceselor economice desfășurate de acestea. Se pornește de la stadiul cunoașterii în domeniul sistemelor informatice economice și al tehnicilor *data mining* pentru a se ajunge apoi la optimizarea proceselor economice folosite de către firmele cu activitate complexă (producție-comerț-servicii), având la final și un studiu de caz cu propuneri concrete.

Primul capitol al cărții tratează pe larg problematica sistemelor informatice, pornind de la stadiul cunoașterii în acest domeniu, până la sistemele informatice de management strategic. Tot în acest capitol se vorbește despre partea de management a unei organizații, precum și despre sistemele informatice utilizate pe diferite nivele de management, de la managementul operațional până la managementul tactic și management strategic.

Al doilea capitol al lucrării este dedicat tehnicilor avansate de extragere a datelor din volume mari de date, cunoscute și ca tehnici *data mining*. Se pornește mai întâi de la apariția și dezvoltarea acestor tehnici odată cu dezvoltarea altor tehnologii cum ar fi aceea a bazelor de date. Sunt prezentate în continuare principalele tipuri de date care pot fi exploatare utilizând asemenea tehnici, printre acestea numărându-se datele stocate în baze de date, datele stocate în depozite de date, fluxurile de date, datele sub formă de serii de timp și secvențe, datele cu structuri de grafuri și rețele și, nu în ultimul rând, datele spațiale, multimedia, text și web. Tot în acest capitol sunt prezentate principalele funcții *data mining* cum ar fi, de exemplu, asocierea, clasificarea, analiza excepțiilor, gruparea, regresia sau previziunile, cu diferite aplicații ale acestora în domeniul economic. La finalul capitolului este realizat un studiu comparativ referitor la diferitele instrumente *data mining* disponibile. Aici se vorbește despre diferite instrumente pentru *data mining* comerciale, produse de firme cum ar fi SPSS, Oracle, IBM sau Microsoft, fiecare având propriile caracteristici.

Cel de-al treilea capitol este o continuare firească a primelor două capitole. Aici se tratează problematica integrării tehnicilor moderne de *data mining* în cadrul sistemelor informatice economice, vorbind și despre cuplarea unui sistem *data mining* cu un sistem de gestiune a bazelor de date. De asemenea, se vorbește despre diferitele sisteme *data mining* destinate mediului de afaceri și despre caracteristicile pe care trebuie să le aibă un astfel de sistem. La sfârșitul capitolului al treilea sunt prezentate pe larg sistemele integrate de tip ERP.

Odată cu capitolul al cincilea, în lucrare se face tranziția către partea practică. La începutul acestui capitol se prezintă câteva considerente referitoare la optimizarea proceselor economice, apoi este prezentată în detaliu o metodologie formală de optimizare, anume metodologia *Six Sigma*, împreună cu baza sa teoretică și etapele acesteia. La finalul capitolului sunt prezentate și instrumentele informatice care pot fi utilizate pentru optimizarea proceselor economice, acestea utilizând și tehnici *data mining*.

În capitolul al șaselea, se propune utilizarea de tehnici *data mining* în locul tehnicilor clasice statistice care sunt utilizate în mod curent în cadrul metodologiei de optimizare *Six Sigma*. În acest capitol sunt analizate etapele metodologiei *Six Sigma* și se fac unele propuneri referitor la situațiile în care se pot aplica tehnici *data mining* în locul celor statistice. La finalul capitolului, aceste propuneri sunt verificate pe un studiu de caz referitor la o firmă care produce, distribuie și comercializează cărți cu caracter educațional.

Capitolul al șaptelea este o continuare a studiului de caz din capitolul precedent. Aici se propune completarea sistemului informatic al firmei descrise la capitolul anterior cu module pentru *data mining* care să extragă informații și cunoștințe referitoare la activitatea firmei, pe baza cărora aceasta să își poată îmbunătăți activitatea. Tot în acest capitol este realizată o prezentare a implementării acestor module, precum și a rezultatelor obținute în urma aplicării lor pe datele disponibile. La final, se formulează o serie de propuneri referitoare la măsurile pe care ar trebui să le ia firma respectivă pentru a-și optimiza activitatea.

După cum se poate observa din prezentarea pe scurt a capitolelor, prima parte a lucrării pune accentul mai mult pe fundamentarea teoretică a conceptelor utilizate, fiind o bază solidă pentru partea practică de la sfârșitul lucrării. Dacă în primele două capitole sunt tratate pe larg domenii cum ar fi cel al sistemelor informatice utilizate în domeniul economic și al analizei avansate a datelor cu ajutorul tehnicilor *data mining*, în cel de-al treilea capitol acestea sunt integrate, iar în cel de-al patrulea capitol sunt prezentate câteva domenii în care pot fi aplicate cu succes tehnicile *data mining*. Începând cu

capitolul cinci se poate astfel trata problematica optimizării proceselor economice utilizând aceste instrumente. Partea a doua a lucrării, reprezentată prin ultimele două capitole, are un pronunțat caracter aplicativ, în ultimul capitol fiind chiar implementate tehnici *data mining* pentru a fi utilizate în vederea optimizării proceselor economice desfășurate de o întreprindere.

Prin urmare, se poate afirma că lucrarea acoperă tematica utilizării tehnicilor *data mining* pe toate nivelurile, pornind de la bazele teoretice și stadiul cunoașterii în domeniu și mergând până la implementarea efectivă a unor asemenea tehnici pentru o situație reală, formulând la final propuneri de măsuri concrete care ar trebui adoptate pentru ca procesele economice desfășurate de firma prezentată în studiul de caz să poată fi optimizate.



## CAPITOLUL 1

# SISTEME INFORMATICE UTILIZATE ÎN MANAGEMENT

### 1.1. Stadiul cunoașterii în domeniul sistemelor informatice

#### 1.1.1. Rolul sistemelor informatice

Din punct de vedere lingvistic, noțiunea de sistem are semnificația de „ansamblu de elemente (principii, reguli, forțe etc.) dependente între ele și formând un tot organizat, care pune ordine într-un domeniu de gândire teoretică, reglementează clasificarea materialului într-un domeniu de științe ale naturii sau face ca o activitate practică să funcționeze potrivit scopului urmărit” [23].

Noțiunea de informatică, din punct de vedere lingvistic, este prezentată ca fiind „știința care se ocupă cu studiul prelucrării informației cu ajutorul sistemelor automate de calcul” [23].

Conceptul de „sistem informatic”, la rândul său, a primit o serie de definiții. Una dintre acestea ar fi: „Sistemul informatic este un ansamblu structurat de elemente intercorelate funcțional necesare asigurării automatizării procesului de obținere a informațiilor și pentru fundamentarea deciziilor” [46].

Se consideră că sistemul informatic este parte a sistemului informațional al unei întreprinderi. Cu alte cuvinte, putem afirma faptul că sistemul informatic se ocupă în special de gestiunea și prelucrarea datelor folosind calculatorul, pe când noțiunea de sistem informațional este ceva mai cuprinzătoare.

Din punctul de vedere al unei organizații economice, un sistem informatic „are funcția de prelucrare automată a datelor, obținute prin înregistrarea informațiilor provenite din interiorul și exteriorul unui sistem economic” [6].

*Având în vedere definițiile enunțate anterior, considerăm că un sistem informatic reprezintă un ansamblu de elemente interdependente care au rolul de a prelucra automat informațiile.*

Sistemele informatice și-au făcut loc în domeniul funcționării organizațiilor economice, ca și în multe alte domenii ale activității umane, odată ce tehnologia a permis atingerea unei puteri de calcul suficiente pentru îndeplinirea unor sarcini cum ar fi gestiunea stocurilor, a salariaților, precum și a altor resurse ale acestora. Mai târziu, odată cu dezvoltarea noilor tehnologii de comunicație, cum ar fi internetul, acestea au fost integrate în sistemele informatice ale între-prinderilor, apărând astfel noi modalități de interacțiune cu clienții, cu furnizorii etc. De asemenea, au apărut și noi posibilități și domenii de acțiune a acestora.

Datorită acestei revoluții digitale și multiplelor avantaje aduse de folosirea sistemelor informatice, acestea și-au făcut rapid intrarea pe toate nivelele de lucru în cadrul firmelor, de la simpli lucrători până la conducere. Dacă la nivelele de jos acestea au în principal rolul de a asista munca oamenilor prin sporirea productivității, la nivelul managementului instituției, sistemele informatice au fost folosite de manageri pentru a asista procesul de conducere. Au apărut, treptat, imense colecții de date cu date privind activitatea organizațiilor economice și, cu ajutorul sistemelor informatice, se puteau realiza statistici, centralizatoare care serveau managerilor instituției în munca lor. Astfel, conducerea organizației putea să aibă o viziune globală a mediului și a modului de funcționare a acesteia care permitea luarea deciziilor în funcție de contextul intern și extern, și, nu în ultimul rând, alegerea unei strategii organizaționale care să conducă organizația spre prosperitate.

Un sistem informatic prelucrează automat datele unei organizații economice stocate în baze de date, astfel încât să ofere managerilor informații suport decizional (pe baza cărora pot lua decizii în cunoștință de cauză). Folosind astfel de sisteme informatice, este posibil să se valorifice datele din trecut, înmagazinate în format electronic în baze de date, text etc. Odată ce se dispune de aceste date, apar problemele exploatării lor optime pentru extragerea de cunoștințe direct utilizabile mai departe în procesul decizional. Au apărut tehnici de analiză avansată a datelor care pot fi integrate cu sistemul informatic al organizației, numite și tehnici de *data mining*.

### ***1.1.2. Considerente privind structura sistemelor informatice***

Conform *Dicționarului de calculatoare* [24], prin structură se înțelege „concepția și compoziția” (unui sistem), exprimate de caracteristici cum ar fi cursul, ierarhia și modularitatea acestuia”.



„Structura unui sistem informatic reprezintă principiile de realizare a siste-mului informatic respectiv (model principal), stabilite pe baza elementelor fundamentale ce definesc domeniul de activitate și particularitățile sistemului eco-nomic pentru care acesta îndeplinește funcția de prelucrare automată a datelor” [6].

Sistemele informatice moderne pot avea în componența lor baze de date, depozite de date, precum și module de analiză avansată a datelor disponibile în vederea extragerii de informații și cunoștințe din acestea, numite și module *data mining*.

Unii autori consideră că un sistem informatic are o structură internă ca și cea reprezentată în Fig. 1.1, formată din trei componente principale: intrări; prelucrări; ieșiri.

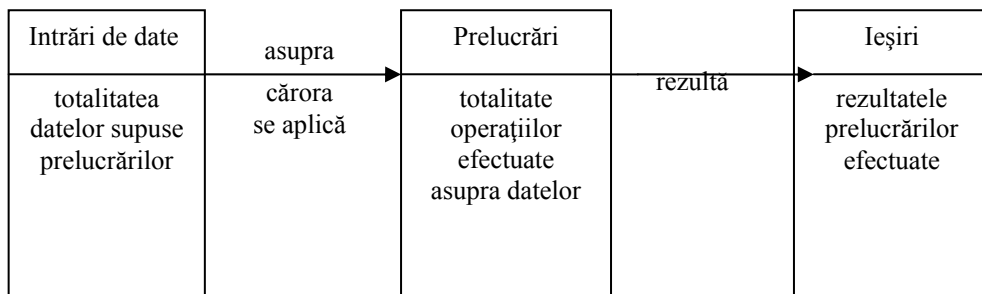


Fig. 1.1. Structura generală a unui sistem informatic. [61]

*Intrările* unui sistem informatic reprezintă datele, culese atât din interiorul, cât și din exteriorul unei organizații economice, care sunt introduse în sistem și care vor fi ulterior prelucrate spre a obține rezultate. De exemplu, pentru cazul concret al unui magazin, se pot considera intrări datele introduse zilnic în sistemul informatic (calculator), ele fiind înregistrate pentru a fi prelucrate ulterior.

*Prelucrările* efectuate de un sistem informatic reprezintă totalitatea operațiilor realizate asupra datelor provenite de la intrarea sistemului. Scopul prelucrărilor în cadrul unui sistem informatic este acela de a obține, pe baza intrărilor, rezultate care să fie apoi utilizate și valorificate de către organizația economică. Pe lângă prelucrările obișnuite, cum ar fi efectuarea unor operații matematice, în unele sisteme informatice există integrate și tehnici *data mining*, cu ajutorul cărora datele de la intrare pot fi analizate în detaliu, extrăgând cunoștințe ascunse. Astfel, prin folosirea tehnicilor

avansate de analiză a datelor, acestea pot fi valorificate și din alte puncte de vedere.

*Ieșirile* unui sistem informatic constă, de obicei, în rezultatele obținute în urma prelucrărilor, indiferent de formatul acestora (tipărit sau electronic). Aceste rezultate pot să revină în sistemul informatic ca date noi, urmând să fie prelucrate din nou, pot să fie preluate ca intrări de către un alt sistem informatic și, nu în ultimul rând, pot fi furnizate utilizatorului, fie manager, fie simplu angajat al organizației economice. În cazul rezultatelor aplicării algoritmilor *data mining*, acestea pot fi destinate managerilor organizației pentru a-i ajuta în luarea de decizii informate.

Structura unui sistem informatic prezentată mai sus este una generală, întâlnită la majoritatea sistemelor informatice. Cu toate acestea, există și diferențe între structurile diferitelor sisteme informatice, acestea depinzând în foarte mare măsură de cazul concret și de domeniul economic în care sunt folosite [6].

*Din punctul nostru de vedere, structura unui sistem informatic este compusă din elementele constitutive ale acestuia care îi permit să își îndeplinească funcțiile pentru care a fost proiectat. Aceste elemente constitutive sunt: echipamentele hardware, echipamentele de comunicație, software-ul de bază și de aplicație, datele organizate în baze de date sau depozite de date, utilizatorii, baza științifico-metodologică etc.*

Arhitectura unui sistem informatic, realizat pentru un domeniu de activitate, reprezintă un model constructiv al acestuia. [56]

Sistemele informatice executive, ca sisteme destinate asistării conducerii de vârf a unei companii, au o arhitectură specifică, în cadrul căreia pot fi identificate trei nivele: [46]

- Nivelul datelor, numit și *bottom tier*,
- Nivelul de analiză, numit și *middle tier*,
- Nivelul de prezentare, numit și *top tier*.