

**UNIVERSITATEA DIN CRAIOVA  
FACULTATEA DE ECONOMIE ȘI ADMINISTRAREA AFACERILOR**

**REZUMATUL TEZEI  
DE DOCTORAT**

**INTEGRAREA TEHNICILOR DATA  
MINING ÎNTR-O PLATFORMĂ DE  
COMERȚ ELECTRONIC**

**Conducător Științific,  
Prof. univ. dr. Vasile GEORGESCU**

**Doctorand,  
Daniel HUNYADI**

**- Craiova, 2010 -**

## Cuprins

Introducere .....	5
1. Arhitectura unui site de comerț electronic .....	10
1.1. Dezvoltarea comerțului electronic .....	10
1.2. Modele de afaceri în comerțul electronic .....	12
1.2.1. Modelul Business-to-business .....	13
1.2.2. Modelul Business-to-consumer .....	14
1.2.3. Modelul Consumer-to-consumer .....	15
1.3. Planificarea unui site de comerț electronic .....	15
1.3.1. Factori care nu pot fi planificați .....	15
1.3.2. Etapele planificării site-ului .....	18
1.4. Etapele creării unui site de comerț electronic .....	22
1.4.1. Analiza/planificarea sistemelor .....	23
1.4.2. Proiectarea sistemelor de comerț electronic – Platforme hardware și software .....	24
1.4.3. Implementarea sistemului .....	26
1.4.4. Testarea aplicațiilor .....	29
1.4.5. Implementarea și promovarea site-ului Web .....	31
1.5. Aspecte de securitate în comerțul electronic .....	32
1.5.1. Cerințe privind securitatea tranzacțiilor în comerțul electronic .....	32
1.5.2. Vulnerabilități și atacuri în comerțul electronic .....	33
1.5.3. Soluții tehnologice pentru securizarea site-urilor Web de comerț electronic .....	36
1.6. Avantaje și dezavantaje ale comerțului electronic .....	47
2. Analiza coșului pieței din perspectivă de marketing .....	51
2.1. Planificarea activității de marketing online .....	51
2.1.1. Convențiile de planificare a marketingului online .....	52
2.1.2. Strategiile marketingului electronic .....	54
2.1.3. Importanța orientării spre client .....	56
2.1.4. Identificarea clientelei online .....	59
2.2. Cercetarea pieței online .....	61

2.2.1. Cercetarea pieței și Internetul .....	63
2.2.2. Avantaje și dezavantaje ale cercetării de marketing online .....	66
2.2.3. Culegerea datelor .....	68
2.2.4. Cercetarea primară .....	70
2.2.5. Motoare de căutare și liste de adrese .....	73
2.2.6. Metode de sondare online .....	74
2.2.7. Cercetarea concurenței .....	80
2.2.8. Cercetarea pieței prin intermediul site-ului propriu .....	81
2.2.9. Cercetarea automată .....	82
2.2.10. Elemente privind prelevarea eșantioanelor .....	83
3. Managementul cunoștințelor .....	85
3.1. Concepte, modele și tehnici Data Mining .....	85
3.1.1. Analiza Factorială .....	91
3.1.2. Analiza Componentelor Principale .....	92
3.1.3. Analiza Clasificării .....	94
3.1.4. Regresia Liniară Multiplă .....	95
3.1.5. Regresia logistică .....	97
3.1.6. Analiza Discriminant .....	98
3.2. Analiza coșului pieței folosind tehnici Data Mining .....	100
3.2.1. Căutarea asocierilor .....	101
3.2.2. Descrierea formală a problemei .....	104
4. Algoritmi pentru determinarea mulțimilor frecvente de articole și a regulilor de asociere .....	110
4.1. Algoritmul Apriori .....	111
4.1.1. Explorarea mulțimilor de articole .....	114
4.1.2. Explorarea regulilor de asociere .....	115
4.1.3. Structuri de date folosite .....	120
4.1.4. Optimizări ale algoritmului Apriori .....	122
4.2. Algoritmul Eclat .....	129
4.3. Algoritmul FP-Growth .....	133
4.4. Algoritmul RElim .....	137
5. Dezvoltarea unei aplicații de comerț electronic, cu suport data mining	

integrat .....	142
5.1. Integrarea Rapid Miner într-o aplicație web de comerț electronic .....	142
5.1.1. Operatorul Read AML .....	145
5.1.2. Operatorul IdToRegular .....	146
5.1.3. Operatorul Example2AttributePivoting (Pivot) .....	147
5.1.4. Operatorul Numerical To Polynomial .....	147
5.1.5. Operatorul Attribute Subset Preprocessing .....	148
5.1.6. Operatorul FP-Growth .....	149
5.1.7. Operatorul Create Association Rules .....	151
5.2. Proiectarea și descrierea bazei de date .....	151
5.3. Prezentarea aplicației .....	154
5.3.1. RMEC Client .....	156
5.3.2. RMEC Server .....	158
5.4. Prezentarea rezultatelor obținute .....	159
Concluzii .....	163
Anexa 1 .....	167
Anexa 2 .....	170
Anexa 3 .....	172
Lista abrevierilor .....	175
Lista figurilor .....	176
Lista tabelor și a listingurilor .....	178
Bibliografie .....	179
Listă Lucrări .....	185

Integrarea și utilizarea tehnicilor data mining în economie deschide noi posibilități în ceea ce privește manipularea informațiilor în timp real între toți factorii de decizie din întreprinderi, în scopul realizării unor noi avantaje atât la nivelul firmei, cât și la nivel global prin obținerea unui grad superior de eficiență economică a tuturor factorilor de producție implicați în procesul productiv.

Programarea la nivelul întreprinderii a evoluat foarte mult în ultimii ani. În mod deosebit, aceasta evoluție se bazează pe existența *Web*-ului. Acesta permite întreprinderilor să își automatizeze și integreze afacerile.

Organizațiile și-au dat seama de necesitatea de abordare a unor noi sisteme, capabile să exploateze beneficiile derivate dintr-o colaborare mai strânsă cu partenerii și furnizorii, care pot adăuga valoare prin *know-how*-ul și experiența lor. Sistemele trebuie să fie redefinite și extinse pentru a cuprinde *Internetul*, sistemele *CRM* și noile modele de *ecommerce*, *B2B* și *B2C*.

În multe privințe, *Internetul* este cel mai eficient instrument de marketing care a existat vreodată. El dă posibilitatea firmelor să interacționeze cu piața la o dimensiune nemaiîntâlnită anterior. Firmele se pot antrena într-un tip de relație bilaterală inimaginabilă în „era producție”. *Internetul* poate să ajute la eliminarea presupunerilor în detectarea dorințelor consumatorului. Dar el nu poate să indice propriile obiective.

Așa cum economia de astăzi este un amestec de vechi și nou, majoritatea companiilor de renume sunt un hibrid între vechi și nou. Deși competențele de marketing stabilite sunt încă în vigoare, noi competențe și cunoștințe adaugă un potențial îmbunătățit dezvoltării. În același mod, consumatorii pot fi priviți acum foarte bine ca și hibrizi. Ei profită de avantajele cumpărăturilor online dar încă mai vizitează magazinele pentru a avea o interacțiune umană și o „experiență a cumpărăturilor”.

Tranzacționarea electronică a mărfurilor și serviciilor constituie o extensie a comerțului actual. Se obține astfel o eficiență sporită, în sensul reducerii costurilor și, în ceea ce privește efectul marketingului, îmbunătățește relația beneficiar-client. Comerțul electronic facilitează și cooperarea între firme: reduce costurile de marketing și de livrare, susține strategia de marketing a firmei și oferă acces pe noi piețe. În viitorul apropiat, comerțul electronic va avea un impact puternic asupra competitivității firmelor. Mai mult, comerțul electronic nu este restricționat de frontierele statelor, dar depinde de existența sau inexistența rețelelor de calculatoare și a infrastructurii bancare care să permită efectuarea plăților prin

Internet. Comerțul electronic permite chiar și celor mai mici furnizori, indiferent de proveniența lor geografică, să fie omniprezenți și să facă afaceri în întreaga lume. El influențează pozitiv nu numai activitatea întreprinderilor mici și mijlocii, dar vine și în sprijinul clienților, oferindu-le nenumărate opțiuni.

Subiectele tratate în această teză sunt de maximă actualitate pentru domeniul studiat. Având în vedere eforturile privind transpunerea în realitate a societății informaționale, au fost studiate, actualizate și implementate o serie de tehnologii cu impact în maniera de abordare a unor afaceri desfășurate de consorții de firme pe o piață globală.

Data mining, cunoscută ca și *Descoperirea Cunoștințelor în baze de date* (Knowledge Discovery in Databases - KDD), este un instrument modern și puternic al Tehnologiei Informației și Comunicațiilor, instrument ce poate fi folosit pentru extragerea unor informații utile dar încă necunoscute. Acest instrument automatizează procesul de descoperire a unor relații și combinații în datele brute, iar rezultatele găsite ar putea fi încadrate într-un sistem automat de suport a deciziei.

Metodele data mining provin din calculul statistic clasic, din administrarea bazelor de date și din inteligența artificială. Ele nu înlocuiesc metodele tradiționale ale statisticii, ci sunt considerate a fi extinderi ale tehnicilor grafice și statistice. Rezultatele metodelor data mining vor trebui supuse în mod sistematic unei supravegheri umane deoarece aplicațiilor software le lipsește intuiția umană pentru a face deosebirea între ceea ce este relevant și ceea ce nu este.

Data mining-ul implică aplicarea tehnicilor de transformare a datelor în informații. Totodată data mining reprezintă interfața critică dintre cunoștințele sintetice sau tiparele generate de către mașini și cunoștințele semantice cerute de oameni pentru raționamentele cu privire la lumea reală.

Cea mai mare parte a companiilor au acumulat de-a lungul timpului cantități mari de date de la clienții lor. Cu aplicații de comerț electronic care au crescut repede, companiile au avut o creștere semnificativă de date în câteva luni, nu în câțiva ani. Data mining are rolul de a descoperi tendințe, modele, corelații, anomalii în aceste baze de date, care ne pot ajuta să luăm decizii corecte pe viitor.

Data mining se ocupă cu prelucrarea datelor complexe de dimensiuni mari. Sunt necesare instrumente robuste pentru a se recupera semnale slabe. Aceste instrumente necesită algoritmi extremi de eficienți care să realizeze prelucrările dorite. Descoperirea regulilor de asociere este unul dintre cele mai populare și mai de succes instrumente din data mining.

Sunt disponibili algoritmi eficienți, care combină concepte și intuiții din probabilități și combinatorică.

Pentru unele firme, comerț electronic înseamnă orice tranzacție financiară care utilizează tehnologia informației. Pentru altele, noțiunea de comerț electronic acoperă circuitul complet de vânzări – inclusiv marketingul și vânzarea propriu-zisă. Mulți oameni consideră comerțul electronic ca fiind orice tranzacție comercială condusă electronic pentru cumpărarea unor produse. Dar, comerțul electronic are, în sens larg, un impact mult mai profund asupra evoluției afacerilor și cuprinde nu numai noile achiziții comerciale ci și totalitatea activităților care susțin obiectivele de marketing ale unei firme și care pot include, spre exemplu, publicitate, vânzări, plăți, activități post-vânzare, servicii către clienți, etc.

Inițiativele de comerț electronic pot genera scăderi ale costurilor, creșteri ale veniturilor și eficiență operațională pentru companiile care urmăresc să dobândească un avantaj în mediul economic competitiv din zilele noastre. El presupune utilizarea Internetului, a comunicațiilor digitale și a aplicațiilor software în cadrul proceselor de vânzare/cumpărare, fiind o componentă a procesului de e-business.

Prezenta teză de doctorat, prin subiectul abordat, dorește să se înscrie la nivelul preocupărilor existente pe plan internațional în domeniul tehnologiilor informatice și a integrării acestora în aplicațiile economice.

În primul capitol al tezei este prezentat conceptul de comerț electronic și sunt introduse câteva idei generale despre ce trebuie făcut atunci când se dorește aducerea afacerii proprii în piața virtuală. Este analizat conceptul de comerț electronic din patru puncte de vedere: al comunicațiilor, al proceselor de afaceri, al serviciilor și al accesibilității în timp real. Sunt prezentate modelele de afaceri întâlnite în comerțul electronic, grupate în funcție de natura tranzacției (Business-to-Business, Business-to-Consumer, Consumer-to-Consumer, Business-to-Guvern) și tehnologiile folosite (Peer-to-Peer, Mobile Commerce).

În continuare sunt prezentate aspecte de care trebuie ținut cont la planificarea unui site. Sunt prezentați factorii care nu pot fi planificați, factori ce țin de comportamentul utilizatorului, browser-ul folosit, linkurile către paginile externe existente în site, etc. Sunt prezentate etapele planificării site-ului: stabilirea audienței, stabilirea scopului, definirea obiectivelor, colectarea informațiilor, stabilirea specificațiilor și stabilirea modului de prezentare. La crearea unui site de comerț electronic trebuie parcurse următoarele etape:

analiza/planificarea sistemelor, proiectarea sistemelor, construirea sistemelor, testarea, implementarea și promovarea.

Asigurarea securității comerțului în Internet reprezintă cea mai importantă provocare ce stă în fața specialiștilor securității informatice. Pentru a se asigura protecția tranzacțiilor comerciale care se pot desfășura prin Internet, trebuie avute în vedere următoarele cerințe fundamentale de securitate: confidențialitatea, autentificarea originii tranzacției, integritatea datelor, împiedicarea nerecunoașterii tranzacției de către expeditor și aplicarea selectivă a unor servicii. În continuare sunt prezentate vulnerabilități și atacuri în comerțul electronic și soluții tehnologice pentru securizarea site-urilor de comerț electronic.

În ceea ce privește avantajele utilizării comerțului electronic, acestea pot fi analizate din trei puncte de vedere: al companiei, al consumatorului și al societății.

Al doilea capitol al tezei se ocupă de analiza coșului pieței folosind instrumente de marketing. Pentru început sunt prezentate elemente cheie de planificare a marketingului pentru Internet. Acestea includ: analiza situației, legătura dintre afacerea electronică și strategia marketingului electronic, obiectivele planului, un plan de implementare, bugetul și un plan de evaluare a reușitei. Sunt prezentate elemente legate de convențiile de planificare ale marketingului online, strategiile marketingului electronic, importanța orientării spre client și identificarea clientelei online.

Cercetarea este fundamentul pe care trebuie să se clădească întreaga activitate de marketing online. Dacă organizarea unei cercetări complexe eșuează, va fi aproape imposibilă orientarea efectivă spre publicul țintă. Internetul este, înainte de toate, o sursă de informații. De aceea reprezintă, sub multe aspecte, instrumentul perfect de cercetare a pieței. El oferă posibilitatea căutării informațiilor în toată lumea. Mai oferă și o bună cunoaștere a activităților concurenței și o estimare a părerilor cumpărătorilor țintă. Mai mult, datorită tehnologiei motoarelor de căutare, informațiile relevante pot fi căutate foarte repede. Aceste informații vor crește valoarea eforturilor de marketing electronic.

Principalii factori determinanți ai activității de cercetare sunt: constrângerile bugetare, nivelul de detaliere al informațiilor, disponibilitatea informațiilor, dinamica necesarului de informații și accesul la „populația” relevantă. În continuare sunt prezentate avantaje și dezavantaje ale cercetării de marketing online, metode de culegere a datelor, metode de utilizare a motoarelor de căutare și a listelor de adrese, metode de sondare online, etc. În final



sunt prezentate metode de cercetare a concurenței, metode de cercetare a pieței prin intermediul site-ului propriu și elemente privind prelevarea eșantioanelor.

În capitolul trei s-a introdus conceptul data mining și sunt prezentate principalele două tipuri de modelări din cadrul data mining-ului: modelarea supervizată și modelarea nesupervizată. Sunt prezentate principalele metode nesupervizate (analiza componentelor principale, analiza factorială și analiza clasificării) iar apoi se continuă cu metodele de învățare supervizată (regresia liniară multiplă, regresia logistică, rețelele neurale și analiza funcțiilor discriminant).

Este realizată o analiza a coșului pieței cu ajutorul regulilor de asociere. Este realizată o descriere formală a problemei și sunt prezentate tehnici data mining de analiză a coșului pieței și de căutare a regulilor de asocierilor. În finalul capitolului sunt prezentați câțiva algoritmi reprezentativi pentru generarea regulilor de asociere folosind tehnici data mining.

În capitolul patru s-a realizat o descriere a principalilor algoritmi data mining folosiți pentru analiza tranzacțiilor și examinarea comportamentului clienților în funcție de produsele cumpărate. Este descris principiul de funcționare și succesiunea operațiilor pentru algoritmi Apriori, Eclat, FP-Growth și RElim. S-a realizat o prezentare a regulilor de asociere obținute pe o colecție de date de test și sunt prezentate structurile de date folosite și posibile optimizări ale acestor algoritmi.

În ultimul capitol s-a realizat o descriere a aplicației de comerț electronic realizate și a modului în care s-au integrat tehnicile data mining în această aplicație. RapidMiner este un mediu dezvoltat pentru procese data mining. Conține o colecție de operatori modulari ce permit proiectarea de prelucrări complexe pentru un număr mare de probleme data mining. Operatorii nu trebuie să facă față formatului actual sau diferitelor reprezentări ale datelor, nucleul RapidMiner ocupându-se de transformările necesare. Manipularea datelor se realizează transparent, putând fi preluate mai multe tipuri de surse de date. Datorită conceptului modular al operatorilor, de cele mai multe ori doar un operator trebuie înlocuit pentru a se putea evalua performanțele sale, în timp ce restul procesului rămâne neschimbat. Aceasta este o caracteristică importantă atât pentru cercetare științifică cât și pentru optimizarea aplicațiilor din lumea reală.

A fost implementată o aplicație web de comerț electronic intitulată Rapid Miner E-Commerce. Realizarea designului aplicației Web a pornit de la ideea integrării funcționale a aplicației de comerț electronic cu un procesor data mining puternic și flexibil precum RapidMiner.

Aplicația este construită pe o arhitectură client-server. Clientul reprezintă un magazin virtual în cadrul căruia orice utilizator poate realiza tranzacții, iar serverul este consola de administrare a magazinului virtual, care pe lângă servicii de administrare oferă și servicii de prelucrare a datelor prin intermediul procesului RapidMiner. Comunicarea clientului cu serverul se realizează prin intermediul serviciilor web. Există două servicii expuse de către server: serviciul de securitate și serviciul de prelucrare logică. Fiind dezvoltată după o arhitectură orientată către serviciu, aplicația este divizată în servicii și module specializate în executarea și rezolvarea anumitor sarcini. Această arhitectură oferă aplicație o flexibilitate ridicată, ceea ce o face ușor de modificat, de distribuit și de îmbunătățit.