

Universitatea Babeș-Bolyai Cluj-Napoca
Facultatea de Științe Economice și Gestiunea Afacerilor

Inteligență Economică: Data Mining aplicat datelor financiare
- Rezumatul tezei de doctorat -

Conducător științific: Prof. Dr. Ștefan Nițchi

Doctorand: Darie Moldovan

Decembrie 2011

Cuprinsul rezumatului

Cuprinsul tezei de doctorat.....	3
Cuvinte cheie.....	5
1. Introducere	5
2. Piețele financiare și tehnicile de Data Mining	8
3. Metode și unelte inteligente de calcul	10
4. Tranzacționare cantitativă	13
5. Învățarea sectoarelor pieței bursiere	20
6. Concluzii finale și posibilități de dezvoltare.....	25
Bibliografie generală	27

Cuprinsul tezei de doctorat

Abstract

Acknowledgments

1. Introduction
2. The financial markets and the Data Mining techniques
 - 2.1 Overview
 - 2.2 Financial Data Mining
 - 2.3 Conclusion
3. Computational intelligence methods and tools
 - 3.1 The Data Mining Process
 - 3.1.1 Overview
 - 3.1.2 CRISP-DM data mining process model
 - 3.2 The data mining methods
 - 3.2.1 Overview
 - 3.2.2 Statistical methods for dimension reduction
 - 3.2.3 ARIMA
 - 3.2.4 Bayesian classification
 - 3.2.5 Nonparametric learning
 - 3.3 Evolutionary methods
 - 3.3.1 Overview
 - 3.3.2 Genetic Algorithms in the Knowledge Discovery process
 - 3.4 Financial time series analysis
 - 3.4.1 Overview
 - 3.4.2 Linear time series
 - 3.4.3 Non-linear time series
 - 3.5 Conclusion
4. Quantitative trading
 - 4.1 Overview on the Algorithmic Trading
 - 4.2 Performance evaluation of the strategies
 - 4.3 The historical data simulation environment
 - 4.4 Stock market informational efficiency
 - 4.5 Risk management
 - 4.5.1 Historical Simulation
 - 4.5.2 EWMA models

- 4.5.3 GARCH models
- 4.5.4 Extreme Value Theory
- 4.6 An automatic trading strategy designed for Bucharest Stock Exchange stocks
 - 4.6.1 Motivation
 - 4.6.2 Related work
 - 4.6.3 Proposed methodology
 - 4.6.4 Experimental results
- 4.7 Conclusion
- 5 Learning the stock market sectors
 - 5.1 Overview
 - 5.2 Cluster analysis: concepts and algorithms
 - 5.2.1 Basic concepts
 - 5.2.2 Main algorithms
 - 5.3 Financial data analysis
 - 5.3.1 Clustering
 - 5.3.2 Distance measure
 - 5.3.3 Data preparation
 - 5.3.4 Tuning the GnetXP
 - 5.4. Experiments results
 - 5.4.1 Year 2000
 - 5.4.2 Year 2001
 - 5.4.3 Year 2002
 - 5.4.4 Year 2003
 - 5.4.5 Year 2004
 - 5.4.6 Year 2005
 - 5.4.7 Year 2006
 - 5.4.8 Year 2007
 - 5.5 Conclusion
- 6 Concluding remarks and future development
 - Results
 - List of Figures
 - List of Tables
 - Bibliography

Cuvinte cheie

Data Mining, Tranzacționare cantitativă, Analiza seriilor de timp, Clusterizare, Algoritm Genetic

1. Introducere

Inteligența economică crește integrarea proceselor de inovare-creare, articulând inițiativele și operațiile destinate pentru accelerarea afacerilor. Cercetările și dezbaterile în acest domeniu permit identificarea metodelor ofensive sau defensive din Inteligența Economică, promovând inovarea, optimizarea și controlul transferului de tehnologie (geografic, interdisciplinar, intercultural).

Într-o societate bazată pe cunoaștere, termenul „inteligentă” devine tot mai important la fiecare nivel al societății, de la întreprinderile mici și mijlocii până la companiile multinaționale, chiar dacă în multe sectoare termenul este încă nou și neclar pentru manageri.

Globalizarea piețelor, dezvoltarea tehnologiei informației și a Internetului au crescut nevoia de informație și au solicitat utilizarea unor unelte de procesare și analiză mai puternice. Disciplina care a apărut în intenția de a satisface aceste nevoi este Data Mining¹. A apărut datorită dezvoltării și interacțiunii a trei domenii: matematică-statistică (permițând apariția unor noi algoritmi, mai puternici), Inteligența Artificială și tehnologia bazelor de date (crescând capacitățile de stocare și viteza de acces la date). O definiție cuprinzătoare, dată de (Zanasi, 1998) afirmă ca Data Mining-ul este ansamblul tehnicilor care permit descoperirea de cunoștințe, altfel ascunse în baze mari de date. Algoritmii care fac acest lucru posibil sunt bazați

¹ Din engleza: Explorarea Datelor

pe metode matematice sofisticate și pe mijloace inteligente de utilizare a capacității de procesare.

În industria financiară, data mining-ul este utilizat de către companii cu scopul de a obține avantaje competiționale. Băncile, companiile de gestiune a activelor sau de consultanță posedă cantități mari de date, colectate din activitatea lor curentă.

Considerând aplicațiile data mining-ului asupra datelor financiare, scopul nostru s-a îndreptat către analizarea datelor de tranzacționare. Explorarea acestui tip de date presupune anumite dificultăți și surse de confuzie. Ținând cont de teoria piețelor eficiente (Fama, 1965), o tendință pe termen lung este improbabil să fie găsit, dar data mining-ul are scopul de a identifica trenduri pe termen scurt, de a le valida și de a identifica rapid momentul în care acestea nu mai sunt valabile. O altă utilizare cu impact este în managementul portofoliului și în optimizarea alegerii produselor componente.

În această teză am încercat să oferim o viziune asupra metodelor de data mining utilizate în sectorul financiar și de a le aplica pe piața bursieră românească, pe de o parte și de a realiza o clusterizare a acțiunilor componente ale indicelui american Dow Jones Composite. Având o natură interdisciplinară, principalele concepte sunt preluate din finanțe, statistică și Machine Learning².

În două dintre capitole (2 și 3) prezentăm cunoștințele teoretice de bază obținute în urma studiului acestor domenii. Începem prin a oferi o imagine asupra dezvoltărilor din ultimii ani în domeniul tranzacționării automate pe piețele financiare mondiale (Capitolul 2), expunând nevoile acestor procese în contextul globalizării piețelor și a creșterii competiției între jucători, cunoscând însă faptul că o mare parte a informațiilor referitoare la algoritmi de tranzacționare și dimensiunea reală a acestui sector rămâne parțial necunoscută datorită politicilor de confidențialitate.

Al treilea capitol prezintă metodologiile utilizate în domeniul Data Mining-ului, descriind conceptele de bază și modelul de proces. În continuare sunt descrise mai detaliat metodele, oferind imagini atât asupra celor clasice (metode statistice, ARIMA, Bayes) dar și asupra Algoritmilor Evolutivi și Genetici.

Al patrulea capitol este dedicat tranzacționării cantitative, cu scopul de a analiza condițiile de îndeplinit pentru a putea tranzacționa automat pe piața bursieră românească. Sunt descrise tendințele actuale în domeniu, arătând apoi parametrii de evaluare a strategiilor de tranzacționare și mediul de simulare pe date istorice. Întreprindem un studiu dedicat identificării existenței eficienței informaționale

² Din engleză: învățare automată.

la nivelul pieței românești. Descriem apoi o metodologie specifică pentru tranzacționare algoritmică. Am testat metodologia propusă pe date istorice de tranzacționare ale unor acțiuni de la Bursa de Valori București și am interpretat rezultatele obținute.

Combinând eficiența naturii și viteza calculatoarelor, posibilitățile sunt nelimitate – afirmă (Bauer, 1994), făcând referire la puterea computațională a Algoritmului Genetic utilizat în previziuni financiare.

Abordarea noastră din al cincilea capitol a fost să obținem un prototip care să utilizeze date financiare ca și intrări, iar prin tehnici de Data Mining să realizăm o grupare a acțiunilor analizate în funcție numai de prețul lor prin identificarea de corelații între ele. O metodologie de clusterizare propusă de Kasabov (Chan & Kasabov, 2004) (Chan, Kasabov, & Collins, 2005) care consideră evoluția în timp a valorilor genelor reglatoare a fost folosită și de noi pentru a depista interacțiunile între diverse acțiuni.

2. Piețele financiare și tehnicile de Data Mining

În ultimii zece ani colectarea datelor a devenit un fenomen normal pentru tot mai multe companii, în special date cu privire la comportamentul consumatorilor: cumpărături, locuri vizitate, tranzacții efectuate. În 1999 una dintre cele mai mari bănci de investiții americane, Goldman Sachs urmărea mai multe de un milion de serii temporale, de la informații privitoare la instrumente financiare cum sunt acțiunile sau obligațiunile la informații mai personale ale clienților lor, cum ar fi cheltuielile de vacanță. În același timp, o altă bancă, Morgan Stanley, colecta zilnic 10 Gigabytes de date (Weigend, 1999).

Este dificil de estimat astăzi care este dimensiunea datelor colectate și analizate de către cele două companii, dar ne putem face o imagine despre amploarea lor, privind rata de creștere a afacerilor în sectorul de e-banking în ultimii 10 ani: de la o valoare de 27 miliarde dolari în 2000 la 176 miliarde dolari în 2011 (US Census Bureau, 2011) (Mulpuru, 2011).

În domeniul tranzacțiilor financiare situația este și mai impresionantă. În ultima decadă piețele financiare au devenit aproape în totalitate electronice. Schimbările au fost atât de masive încât, la nivel mondial, cea mai mare parte a tranzacționării se face automat, pe baza algoritmilor de tranzacționare (Aite Group, 2010).

În opinia noastră, există o serie de factori care au făcut evoluția atât de rapidă. Dintre ei menționăm:

1. Unul dintre cei mai importanți factori este maturizarea Internetului, societatea modernă fiind familiară cu acesta și dezvoltarea rapidă a tehnologiei a dus și la dezvoltarea tehnologiilor pentru tranzacționare.

2. În același timp cu inovațiile tehnologice, investitorii au devenit mai sofisticăți, aând nevoie de executare rapidă a tranzacțiilor și instrumente de analiză puternice pentru a face față competiției.
3. Globalizarea acestui domeniu a adus ca și consecință tranzacții mai ieftine și acces la piețele mondiale pentru investitori. Fuziunile și achizițiile burselor au dterminat o dezvoltare fluentă a sistemelor de tranzacționare, direcția fiind una de compatibilizare între diferitele sisteme utilizate în lume. Piețele emergent au luat modelul piețelor dezvoltate și au creat structuri asemănătoare.
4. Perioada a fost si de maturizare a modelelor de afaceri pentru băncile de investiții care au evoluat rapid, creând o gamă largă de produse, multe dintre ele exotice, pentru a satisface nevoile investitorilor și de a alinia riscul la nivele predefinite.
5. Varietatea de instrumente a aut nevoie de unelte puternice de calcul și reacție rapidă, multe dintre ele fiind interrelaționate și acționând în cascadă. Analiza relației dintre piețe a devenit o preocupare a analiștilor financiare, cu posibilitatea de rebalansare rapidă a portofoliilor.

3. Metode și unelte inteligente de calcul

3.2 Procesul de Data Mining

Data Mining-ul este procesul prin care un proces de învățare automată este adoptat pentru a analiza și extrage date (Richard J. Roiger, 2002). O altă definiție, dată de (Edelstein, 1999) susține că data mining-ul utilizează unelte sofisticate de analiză a datelor cu scopul de a descoperi legături între date necunoscute anterior, dar valide în cadrul unor baze mari de date. O definiție simplă aleasă de (Nitchi & Avram-Nitchi, 1997) consideră data mining-ul ca fiind procesul de extragere a informațiilor predictibile, ascunse în seturi mari de date.

Data mining-ul s-a dezvoltat din trei rădăcini importante. Prima este statistica. Cele mai multe dintre metodele de data mining sunt bazate pe abordarea clasică statistică. Tehnicile dezvoltate au fost cu scopul de a îmbunătăți metodele obișnuite statistice. Conceptele de regresie, abatere medie, distribuție, varianță, clusterizare, clasificare, intervale de încredere sunt folosite în mod curent de către tehnologiile de data mining și reprezintă fundamentele algoritmilor moderni.

Cea de-a doua componentă care a influențat data mining-ul este Inteligența Artificială. Aceasta încearcă să modeleze procese specifice minții umane, fiind oarecum în opoziție cu statistica. Utilizarea Inteligenței artificiale a devenit tot mai răspândită o dată cu dezvoltarea tehnologiei. Are nevoie de capacitate mare de calcul, care a reprezentat în permanență o problemă, întrucât dimensiunile problemelor au crescut în același ritm cu creșterea capacităților de stocare și de procesare.

A treia disciplină care a ajutat la dezvoltarea sectorului, este cea a bazelor de date. În timp ce pentru un statistician data mining-ul înseamnă descoperirea modelelor, pentru cineva care lucrează în baze de date este o formă de procesare analitică (interogări care examinează seturi mari de date). Rezultatele aceleiași probleme vor fi răspunsuri la o interogare pentru persoana implicată în baze de date și parametrii unui model pentru statistician.

Data mining-ul poate fi descris ca și unificarea dezvoltărilor în statistică, Inteligență Artificială și baze de date. Succesul data mining-ului consistă în faptul că reușește să facă diferența între date și informații: transformă datele în informații.

Pentru succesul implementării data mining-ului, este nevoie de sistematizare sub forma unui model de proces. În literatură se găsesc mai multe modele dezvoltate de către producătorii de soft sau alte organizații interesate. Acestea descriu pași de urmat de către utilizator, de la culegerea datelor și până la obținerea rezultatelor. Cele mai cunoscute modele, conform lui (Edelstein, 1999) sunt 5A, dezvoltat de către SPSS, SEMMA și propus de SAS, dar cel care a atras cel mai mult atenția a fost CRISP-DM, lansat de un consorțiu format din companiile NCR, Daimler-Benz, SPSS și OHRA.

3.2 Metodele de data mining

În funcție de ținta propusă, există o gamă variată de metode de data mining, de aici nevoia de a le înțelege funcționarea, interconectivitatea și clasificarea. Două categorii mari pot fi definite: metodele orientate spre explorare și metodele orientate spre descoperire (Sayad, 2011). Metodele orientate spre explorare sunt utilizate pentru a explica trecutul și a atrage atenția asupra aspectelor importante descoperite, în timp ce metodele orientate spre descoperire încearcă să construiască modele.

În general, metodele bazate spre explorare studiază ipoteze prin metode statistice (analiza varianței, testul T, ANOVA). În principiu, Data Mining-ul încearcă să descopere cunoștințe noi prin analiza datelor, din acest motiv această categorie fiind mai puțin asociată cu conceptul, focalizarea fiind pe crearea de modele.

Metodele bazate pe descoperire au la bază două căi principale pentru atingerea obiectivelor: una bazată pe previziuni și cealaltă pe descrierea datelor. Metodele bazate pe descriere interpretează datele, studiind relațiile din cadrul acestora. Metodele de previziune se centralizează pe comportamentul datelor, creând modele iar pe baza lor realizează previziuni ale valorilor variabilelor studiate, dar și expun datele într-un mod care să fie inteligibil și util pentru utilizatorul final. Tehnicile folosite sunt de obicei bazate pe inducție, modelul învățând reguli dintr-un set de date de antrenament fiind apoi testat pe date noi, până în momentul în care este atins un nivel acceptabil de acuratețe a rezultatelor.

Dintr-o alta perspectiva de clasificare a metodelor este cea de învățare automată, care poate fi supravegheată sau nesupravegheată. Învățarea automată încearcă definirea unei funcții din studiul anumitor date cunoscute și care va oferi o anumită valoare în funcție de intrări. Acest tip de învățare poate avea la bază clasificarea sau regresia, în funcție de tipul problemei studiate. Învățarea nesupravegheată lucrează cu date neetichetate și are ca scop descoperirea structurii datelor. (Bishop, 2006)(Witten & Frank, 2005)

Metodele de Data Mining sunt influențate de anumiți parametri: datele, tipul datelor și algoritmi aplicați.

Un sistem de Data Mining trebuie integrat corespunzător pentru a putea fi utilizat în mod continuu. Cel mai important aspect este relația cu sistemul de management al bazei de date. Fie că

este un sistem online sau offline de analiza, accesul la baza de date trebuie să fie corect conceput, în special pentru domenii sensibile la timpul de reacție cum este sectorul financiar.

3.3 Metode evolutive

Creșterea nivelului de date stocate în bazele de date a condus la o nevoie crescută de analizare a lor în timp cât mai scurt, tot mai multe informații valoroase ascunzându-se în interiorul lor. Numărul de analiști fiind limitat, la fel și capacitatea lor, a apărut nevoia de procesare și analiză automată. Pe lângă metodele prezentate anterior, propune utilizarea metodelor evolutive pentru Data Mining și în particular Algoritmul Genetic.

Utilizarea algoritmilor evolutivi este reprezentată de metode stochastice de căutare, bazate pe abstractizarea evoluției biologice (reproducere, mutații, recombinări, selecții). Fiecare individ al unei populații studiate este considerat candidat pentru soluție. O funcție de evaluare calculează calitatea acestei soluții. În acest fel, folosind selecția naturală, indivizii evoluează printr-o procedură de selecție. Operatorii bazați pe genetică sunt aplicați cu probabilități predefinite în această procedură, astfel încât cu cât este mai puternică „gena” unui individ cu atât să fie mai mare probabilitatea ca părți din candidații săi să fie transmise generațiilor următoare de indivizi. Mutațiile pot da naștere unor gene care nu există în indivizii primei populații. În contrast, dacă operatorii realizează încrucișări stochastice, valorile genelor nu se vor schimba ci vor fi doar schimbate între indivizi.

Algoritmii evolutivi pot fi utilizați în rezolvarea unor probleme din diverse domenii de cercetare, principalele două dificultăți întâlnite fiind alegerea reprezentării individului (care soluție candidat să reprezinte un individ) și funcția de evaluare a indivizilor.

În problemele de descoperire a regulilor individul este reprezentat de către o regulă sau mai multe reguli candidat iar funcția de evaluare este reprezentată de măsurarea calității regulilor. Cele mai bune reguli ale unei generații vor fi selectate și operatorul genetic va transforma regulile candidat într-un set nou de reguli. Spre deosebire de alți algoritmi, algoritmii evolutivi, utilizând operatori stochastici, realizează o căutare globală a spațiului de reguli, un singur operator fiind capabil să schimbe un număr mare de seturi de reguli. Deasemenea, realizează o evaluare completă a setului de reguli candidat, nelăsând reguli candidat parțiale, dar în același timp au o putere mare de calcul, lucrând în același timp cu mai multe populații de reguli candidat.

În practică, este recomandată o combinație a algoritmilor inductivi cu cei evolutivi, cu scopul de a îmbunătăți rezultatele procesului de data mining (Freitas A.A., 2002).

Concluziile noastre, după studierea oportunității utilizării metodelor inteligente pentru analizarea datelor financiare au fost că acestea sunt recomandate de către profesioniștii în domeniu ca și unele foarte puternice pentru previziuni, în același timp existând o puternică inter-operabilitate a metodelor

între diferite domenii, datorită specificului metodelor evolutive de a simula comportamentele naturale, comune multor procese, indiferent de domeniu.

4. Tranzacționare cantitativă

4.1 Despre Tranzacționarea Algoritmă

Tranzacționarea cantitativă reprezintă procesul de tranzacționare a valorilor mobiliare în mod automat, în funcție de un algoritm, fără interacțiunea directă umană, sau potrivit lui (Chan, 2008), tranzacționarea bazată strict pe semnalele de vânzare/cumpărare ale unui algoritm.

Un raport publicat de către Aite Group (Aite Group, 2009) arată că în ultimii trei ani tranzacționarea algoritmică a devenit dominantă în piețele financiare, cu o creștere impresionantă de la an la an. În 2009 estimarea făcută a fost că aproximativ 70% din volumul zilnic de tranzacționare realizat în Statele Unite s-a realizat în acest mod automat. Expansiunea din ultimii ani a fost stimulată de către profitabilitatea ridicată pe care acești algoritmi o aduc. Conform unui raport FixProtocol (Donefer, 2008) totalul profiturilor realizate anual din tranzacționarea cantitativă automată a fost de aproximativ 20 miliarde de dolari în Statele Unite.

Algoritmii sunt dezvoltați în funcție de strategii utilizate de către specialiști, bazându-se pe datele istorice disponibile, testate și îmbunătățite. În acest fel, au fost create strategii competitive, obținându-se în mod automat soluții care pot reacționa rapid la modificarea condițiilor din piață.

Cei mai mulți dintre algoritmi sunt din categoria celor cu frecvență mare a tranzacțiilor. Diferența dintre cele două categorii, cu frecvență ridicată și scăzută a tranzacțiilor este că în primul caz se urmărește realizarea unui profit scăzut pe fiecare tranzacție, fiind însă în final consistent datorită numărului mare de tranzacții și exploatând o mare parte din mișcările pieței.

Chiar dacă se bazează pe date istorice, nu trebuie făcută confuzia cu analiza tehnică. Aceasta poate fi o parte a unei strategii dacă semnalele sale pot fi utilizate ca date de intrare utilizând un limbaj de programare. Deasemenea, date fundamentale pot fi încorporate într-o strategie, știri sau comentarii despre o anumită companie. Puterea computațională poate fi utilizată pentru a realiza comparații a mii de companii sau pentru a interpreta știri mult mai rapid decât o persoană le-ar putea citi sau înțelege.

O consecință a dezvoltării tranzacționării algoritmice este că piețele pe care activează devin mai eficiente. Algoritmii de tranzacționare tind să exploateze cât mai multe dintre ineficiențele pieței, informația nouă fiind absorbită mai repede. O altă urmare este creșterea lichidității, generată de

tranzacțiile frecvente, în beneficiul piețelor. Lichiditatea crescută în piețele unde este prezentă tranzacționarea cu frecvență ridicată determină și costuri mai scăzute pentru toți investitorii, datorită diferenței mai scăzute între cerere și ofertă și un risc mai scăzut al investițiilor, în principal al celui de contraparte.

În același timp, tranzacționarea algoritmică a dus la dezvoltarea domeniului de cercetare pentru creșterea puterii de calcul, determinând progrese tehnologice privind eficiența sistemelor de tranzacționare. În ultimii ani, timpul de execuție al tranzacțiilor a scăzut, în special datorită cerințelor venite din partea algoritmilor de tranzacționare. Competiția dintre dezvoltatorii de algoritmi, persoane fizice sau instituții are loc nu doar la nivelul pieței dar și al infrastructurii construite pentru a susține eficacitatea algoritmilor. O consecință directă a dezvoltării tehnologiei este și scăderea dramatică a timpului de menținere a unei investiții.

Tranzacționarea algoritmică este întâlnită pe piețele valutare internaționale, acțiuni și derivate, în ultimii ani fiind dezvoltate tot mai multe produse exotice destinate atât acoperirii riscului dar și pentru a produce un levier ridicat. Diseminarea știrilor, viteza și calitatea analizei datelor, a posibilității de reacție rapidă la știri a dus la creșterea transparenței piețelor.

Conform unui raport realizat în 2010 (Aite Group, 2010), în Europa tranzacțiile realizate prin intermediul algoritmilor sau a accesului direct în piață a crescut la mai mult de 50%, în timp ce în America procentajul este de peste 70. Sofisticarea investitorilor tradiționali a determinat diversificarea ofertelor. Trendul crescător al tranzacționării electronice este unul care tinde să se mențină, același raport estimând că în 2010, aproape în totalitate, acțiunile au fost tranzacționate în mod electronic în Statele Unite.

4.2 Strategie automată destinată tranzacționării acțiunilor pe Bursa de Valori București

4.2.1 Motivație

Utilizarea unor indicatori analiză tehnică în luarea deciziilor pentru investiții rămâne un subiect controversat, fiind apreciată de unii investitori, dar a fost respinsă de alții (Edwards, Magee, & Bassetti, 2007). În timp ce specialiști și cercetători din lumea academică au dezvoltat noi metode și indicatori, teste în timp real sau simulate sunt necesare pentru a le valida (Silaghi & Robu, 2005).

Predicția prețului este o problemă foarte complexă, și selectarea indicatorilor tehnici potriviți pentru o anumită acțiune este una dintre primele preocupări ale investitorilor care utilizează analiza tehnică. O dificultate este de reglarea parametrilor acestor indicatori într-un mod care semnalele lor să fie corecte într-un procent cât mai mare posibil (Bodas-Sagi, Fernández, Hidalgo, Soltero, & Risco-Martin, 2009). În timp ce comportamentul acțiunilor este diferit și suferă schimbări în timp, alegerea valorilor parametrilor devine o sarcină dificilă fără ajutorul unei metode avansate de calcul.

Metodele de data mining sunt considerate a fi o alegere inteligentă pentru selectarea indicatorilor tehnici potriviți, permițând teste pe seturi de date foarte mari (o condiție esențială, ținând cont de volumul mare de date financiare disponibile), precum și multe combinații ale valorilor parametrilor, combinând valori orare, zilnice sau săptămânale pentru teste (Bodas-Sagi, Fernández, Hidalgo, Soltero, & Risco-Martin, 2009) (Silaghi & Robu, 2005).

Obiectivul nostru este de a propune o metodologie care combină indicatori tehnici diferiți, bazată pe teste efectuate pe seturi de date colectate de pe piețele de acțiuni internaționale sau locale, precum și obținerea de semnale de tranzacționare cu o precizie îmbunătățită în comparație cu rezultatele obținute prin utilizarea utilizarea unui singur indicator, și compararea rezultatelor cu alte cercetări efectuate.

4.2.2 Metodologia propusă

Am considerat în propunerea noastră o combinație de indicatori utilizați frecvent în analiza tehnică având ca și scop demonstrarea eficacității semnalului agregat față de utilizarea singulară a indicatorilor. Cei trei indicatori sunt MACD (Moving Average Convergence-Divergence), ROC (Price Rate of Change) și STS (Oscilatorul Stochastic).

- MACD este un indicator utilizat pe scară largă și urmărește schimbările în forța, direcția, ritmul și direcția de o tendință. Se calculează luând în considerare Media mobilă exponențială (EMA), pentru două perioade diferite , comparându-le (Gerald, 1999).

Formula pentru calcularea unei EMA la un anumit punct este următoarea:

$$EMA_t = EMA_{t-1} + \alpha * (pret_t - EMA_{t-1}), \quad (4.1)$$

unde α este un factor constant de uniformizare, exprimat ca și procent sau ca număr de perioade.

În general,

$$EMA = \alpha * (p_1 + (1-\alpha)p_2 + (1-\alpha)^2 p_3 + (1-\alpha)^3 p_4 + \dots). \quad (4.2)$$

Factorul de ponderare în fiecare punct (p) este în scădere exponențială, astfel încât cu cât este mai vechi un anumit punct, cu atât va influența mai puțin rezultatul.

În continuare, formula MACD este următoarea:

$$MACD = EMA_a - EMA_b, \text{ unde } a < b. \quad (4.3)$$

Semnalele de tranzacționare sunt date când EMA pe perioadă scurtă de timp crește la o valoare mai mare decât cea pe termen ($EMA_a > EMA_b$) – semnal de cumpărare - sau când valoarea EMA pe termen scurt devine mai mică decât cea pe termen lung ($EMA_a < EMA_b$).

- Indicatorul Price Rate of Change (ROC) este un oscilator care calculează diferența dintre prețul din momentul curent și un moment timp de dinainte, cu n perioade.

Formula de calcul este următoarea:

$$ROC = Pret_t - Pret_{t-n} \quad (4.4)$$

sau valoarea relativă:

$$ROC\% = \frac{P_t - P_{t-n}}{P_{t-n}} * 100 \quad (4.5)$$

unde t este perioada curentă, iar n este numărul de perioade de timp din trecut.

Indicatorul ROC evidențiază când o anumită acțiune este supracumpărată sau supravândută, semnalele de tranzacționare fiind date atunci când apar divergențe față de evoluția curentă a prețului.

- Oscilatorul Stochastic măsoară impulsul pieței considerând un interval de tranzacționare dintr-o anumită perioadă de timp (Lane, 1984).

Pentru calcul utilizăm următoarele formule:

$$\%K = 100 \frac{C-L}{H-L} \quad (4.6)$$

unde C reprezintă prețul de închidere al acțiunii, L este minimumul din perioada analizată, iar H este maximumul.

$$\%D = 3 \text{ perioade MA}(\%K), \quad (4.7)$$

unde MA(%K) este media mobilă pentru %K.

Algoritmul intră în contact cu două entități independente și o comunicare eficientă între ele trebuie să fie asigurată, în primul rând baza de date de la care datele sunt colectate pentru analiză în timp real. Aceasta este o comunicare într-o singură direcție, de la baza de date către sistem. Un aspect

important este consistența datelor, datele care să lipsesc sau neformatate corespunzător nu sunt acceptabile, indicatorii tehnici fiind foarte sensibili la aceste aspecte.

Cealaltă entitate este piața (de exemplu, Bursa de Valori). O comunicare reciprocă este necesară în acest caz. Sistemul transmite ordine de tranzacționare către piață, iar aceasta transmite răspunsul dacă acestea au fost executate sau nu. O conexiune foarte stabilă și rapidă este necesară între cele două, viteza de execuție fiind un factor foarte important pentru succesul unui algoritm de tranzacționare, uneori, precizie trebuind să fie de milisecunde (Chaboud, Chiquoine, Hjalmarsson, & Vega, 2009).

În interiorul sistemului, algoritmul de tranzacționare este motorul principal. Toate celelalte procese transmit semnale către algoritm, care reacționează în funcție de calculele efectuate, decide dacă să transmită ordinele către piață, să închidă pozițiile deschise sau nu să reacționeze deloc, doar să aștepte modificări ale datelor (Moldova, Moca, & Nitchi, 2011).

Sistemul va efectua sarcini suplimentare, în plus față de agregarea semnalului de tranzacționare. Aceste sarcini sunt legate de gestiunea pozițiilor deschise și a riscurilor. Figura 1 prezintă integrarea algoritmului în întregul sistem.

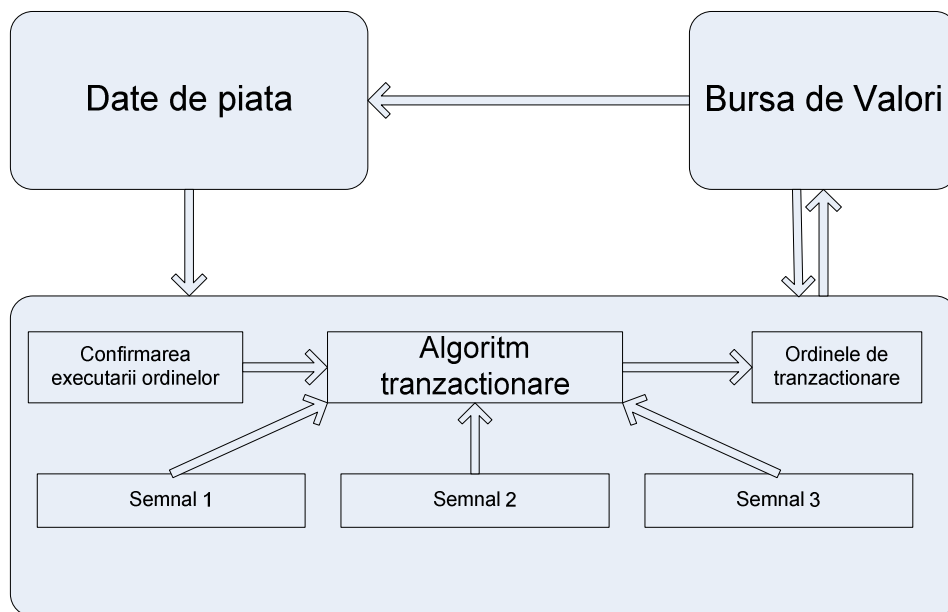


Figura 1 Integrarea componentelor

Funcționalitatea sistemului depinde nu numai de un design optim, dar și pe limbajul de programare folosit. (Russell & Yoon, 2005) consideră. NET Framework ca fiind cea mai potrivită alegere datorită flexibilității sale superioare, scalabilitate compatibilitate și interoperabilitate. Înainte de a introduce un ordin de tranzacționare, sistemul trebuie să facă unele validări cu privire la gestionarea pozițiilor deținute și de gestionare a riscurilor. O verificare repetitivă a semnalului de tranzacționare trebuie să fie efectuate (acest interval poate fi stabilit în funcție de strategia de tranzacționare, indiferent dacă acesta va fi una de tranzacționare cu frecvență ridicată sau nu). Dacă un semnal a fost emis statutul pozițiilor deschise în prezent trebuie să fie verificat. Dacă o anumită limită de expunere a fost atinsă, sistemul nu va permite ordinelor să ajungă în piață.

4.2.3 Rezultate experimentale

Pentru implementarea metodologiei am utilizat AFL –Amibroker Formula Language (Amibroker), un limbaj de programare utilizat pentru dezvoltarea de indicatori personalizați, setarea parametrilor de management a riscului și testare pe date istorice.

Pentru testarea metodologiei propuse am ales un grup de unsprezece companii listate la Bursa de Valori București, componente ale indicilor BET și BET-FI, ele numărându-se printre cele mai lichide acțiuni de pe piață; din acest motiv am considerat utilizarea lor ca fiind relevantă pentru testarea strategiei. Seriile de timp utilizate conțin date de tranzacționare istorice începând cu ianuarie 2007 și până în iulie 2011, având un număr de aproximativ 50000 de înregistrări fiecare. Datele înregistrate au o frecvență de cinci minute pe întreaga perioadă menționată mai sus.

Pentru a defini termenii de comparație a performanței metodologiei am considerat mai multe abordări. În primul rând, am testat strategia având ca și comparații strategii similare care țin cont numai de câte unul dintre indicatorii tehnici utilizați, în concordanță cu scopul declarat de a obține rezultate superioare tranzacționării pe baza semnalelor unui singur indicator.

În al doilea rând, am comparat rezultatele cu performanțele celor doi indici, BET și BET-FI. Am analizat rezultatele pe diferite intervale de timp, pe de o parte pentru fiecare an și apoi global, pentru întreaga perioadă, pentru a identifica performanțele pe diferite faze ale pieței. Strategia de comparație a presupus investirea la începutul perioadei de referință și menținerea acesteia până la finalul perioadei.

Datorită faptului că pe piața pe care a fost strategia testată este permisă intrarea numai pe poziții de cumpărare (nu este permisă vânzarea în lipsă), parametrii indicatorilor tehnici au fost modelați corespunzător pentru a identifica trendurile de creștere și a încerca să capteze cât mai mult din aceste perioade, în același timp să evite cât mai mult posibil investirea în perioadele de scădere. Din aceste motive, sistemul are anumite limitări, nefiind testat și pentru cazul în care vânzările în lipsă ar fi permise.

Suma virtuală inițială pentru tranzacții a fost definită la 100000 RON, considerând lichiditatea destul de redusă a pieței. Indicatorii tehnici utilizați au fost calculați folosind serii orare de timp, generând astfel un număr relevant de tranzacții pentru concluzia asupra performanțelor sistemului. În ceea ce privește managementul riscului, am calculat ca tranzacțiile să fie limitate la suma de 10000 RON și variația maximă permisă pentru o tranzacție la 5%.

Testele au fost realizate în două direcții, prima considerând investițiile pentru o perioadă de un an, portofoliile fiind reactualizate la suma inițială la începutul fiecărui an, neincluzând performanța anului precedent; în al doilea caz a fost luată în calcul întreaga perioadă, asumând faptul că profiturile obținute sunt reinvestite, fără a lua în calcul retrageri sau alimentări de bani. Deasemenea a fost luat în calcul un comision de 0.3% pe fiecare tranzacție.

În tabelul următor prezentăm performanțele strategiilor luate în calcul, adăugând și rata dobânzii bancare, calculată pentru fiecare an în parte și global. Valorile utilizate pentru rata dobânzii fără risc au fost cele determinate de către Banca Națională a României pentru politica monetară.

	2007	2008	2009	2010	2011	Overall
Rata dobânzii	7.42%	9.75%	9.06%	6.26%	6.25%	41.15%
BET	16.29%	-69.68%	57.2%	10.89%	-15.15%	-44.36%
BET-FI	14.95%	-83.62%	83.33%	-10.09%	-21.02%	-74.19%
MACD	19.73%	-53.88%	73.85%	-30.75	-19.16%	-45.13%
ROC	-2.74%	-8.35%	14.43%	-34.94%	-35.4%	-60.3%
Oscilatorul stochastic	18.9%	-21.88%	86%	-3.93%	-7.03%	53.13%
Metodologia propusă	13.98%	8.27%	81.81%	3.45%	14.26%	146%

Putem observa, că în ansamblu, cei doi indici utilizați ca termeni de comparație au avut performanțe negative, investiția în companiile componente ale indicilor generând pierderi semnificative. Se pot identifica însă și perioade intermediare de timp cu performanțe pozitive.

Dintre strategiile având la bază indicatori tehnici, numai cea care utilizează Oscilatorul Stochastic s-a dovedit a fi câștigătoare pe întreaga perioadă analizată.

4.3 Concluzii

Utilizarea indicatorilor analizei tehnice în luarea deciziilor de investiții în acțiuni rămâne un subiect controversat, fiind apreciat de o parte a investitorilor, dar respins de alții. În timp ce profesioniștii în

domeniu și cercetători din mediul academic dezvoltă noi metode și indicatori, acestea au nevoie de o testare intensivă pe date reale sau simulate pentru a fi validate.

Urmărind testele statistice care au fost aplicate asupra rentabilităților zilnice ale indicelui BET-FI, au fost identificate corelații liniare și neliniare, prețul acțiunilor fiind influențat de informațiile noi apărute în piață, ipoteza de mers aleatoriu fiind respinsă. Nu poate fi susținută existența unei forme slabe de eficiență informațională, nefiind în acest fel respinsă utilitatea analizei tehnice.

Combinarea a trei indicatori de analiză tehnică prin semnalele lor de tranzacționare a fost testată cu succes prin utilizarea în implementare a limbajului AFL (Amibroker). Testele au fost efectuate pe serii de timp din perioada 2007-2011, reprezentând date de tranzacționare ale unor acțiuni tranzacționate la Bursa de Valori București, obținând rezultate superioare strategiilor luate ca referință, cu toate că au fost impuse condiții prudențiale, iar controlul tranzacțiilor cu pierdere a fost corespunzător. Chiar dacă în unele cazuri strategiile cu care a concurat au înregistrat performanțe superioare, global, strategia propusă a avut performanțe mai consistente, dovedind că semnalele agregate pentru tranzacționare dau rezultate mai bune comparativ cu semnalele individuale ale indicatorilor tehnici.

Pentru optimizarea sistemului, propunem integrarea unui Algoritm Genetic sau a unei alte metode adaptive pentru modelarea parametrilor indicatorilor tehnici într-o manieră rapidă și funcțională și pentru învățarea automată de noi reguli de tranzacționare. O combinație între reguli de tranzacționare descoperite automat și regulile definite de către experți în domeniu ar putea îmbunătăți în ansamblu performanțele sistemului.

5. Învățarea sectoarelor pieței bursiere

5.1 Preambul

Comportamentul în timp al unei singure acțiuni nu poate descrie întreaga evoluție a pieței, dar studiat împreună cu cel al altor acțiuni, ponderându-le importanța, poate fi identificată direcția principală a grupului. Din acest motiv au fost dezvoltați indicii bursieri. Este însă mai ușor de previzionat evoluția unei singure acțiuni luată dintr-un grup în care majoritatea acțiunilor au un comportament similar.

Un indice bursier reprezintă o metodă care evidențiază sintetic evoluția unei secțiuni a pieței. În ultimii zeci de ani, indicii bursieri au reprezentat o preocupare intensă a managerilor de fond, crescând așteptările performanțelor lor (Burr, 2005). Creșterea companiilor de servicii financiare sau de presă financiară, indicii reprezintă principalele criterii de comparație a performanței unui portofoliu.

Există mai multe tipuri de indici, bazați pe mărime, sector de activitate, tip de management sau alte criterii considerate utile de cei care i-au construit. Abordarea noastră a fost de a obține un prototip de sistem cu suport în inteligența economică, ce utilizează ca și date de intrare informațiile de tranzacționare ale unor companii listate și prin tehnici specifice de data mining, considerând numai evoluția lor în timp, să găsim corelații noi între ele. O metodologie de clusterizare propusă de către Kasabov et al. (Chan & Kasabov, 2004), (Chan, Kasabov & Collins, 2005) pentru studiul evoluției în timp a genelor umane reglatoare, a fost adoptată pentru a fi aplicată în domeniul financiar. Pentru experimente am ales 65 de companii, componente ale indicelui american Dow Jones Average Composite (DJA). Datorită numărului relativ mic de companii în cadrul indicelui rezultatele sunt mai ușor de înțeles și verificat.

5.2 Clusterizarea datelor financiare

O clusterizare a unei serii de timp este validă dacă fluctuațiile din cadrul grupului sunt corelate, iar fluctuațiile dintre grupuri sunt decorelate, sau puțin corelate.

Statisticile desprinse din analiza clusterilor vor afișa informații despre cât sunt de apropiate comportamentele componentelor lor. Diferențierea dintre clusteri trebuie să fie vizibilă pentru a putea demonstra utilitatea clusterizării, ceea ce va conduce la posibila aplicare în domeniul financiar: prin analizarea evoluției în timp a prețului unei companii vom determina cărui grup îi aparține, ceea ce poate însemna încadrarea într-un grup diferit de sectorul în care activează și unde a fost încadrat în mod natural în cadrul indicelui.

Așa cum este discutat în (Chan, Kasabov, & Collins, 2005), posibilitatea alegerii unui algoritm pentru clusterizare, variază de la abordările clasice, cum sunt K-means, clusterizare ierarhică sau algoritm bazat pe arbori sau mai noile modele Autoregresive, B-splines și Modelele de Regresie Lineară Multiplă (MLR). Ultima variantă a fost aleasă pentru studiul traiectoriei genelor. Pentru rezolvarea problemelor de optim local a fost adoptat Algoritmul Genetic, obținând un algoritm hibrid (Figura 2).

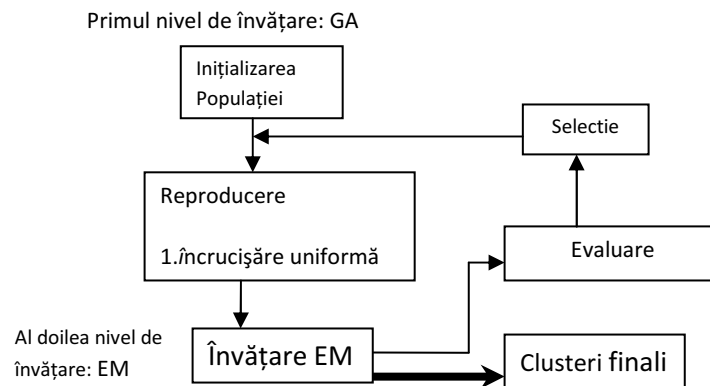


Figura 2. Algoritm de clusterizare (Chan, Kasabov, & Collins, 2005)

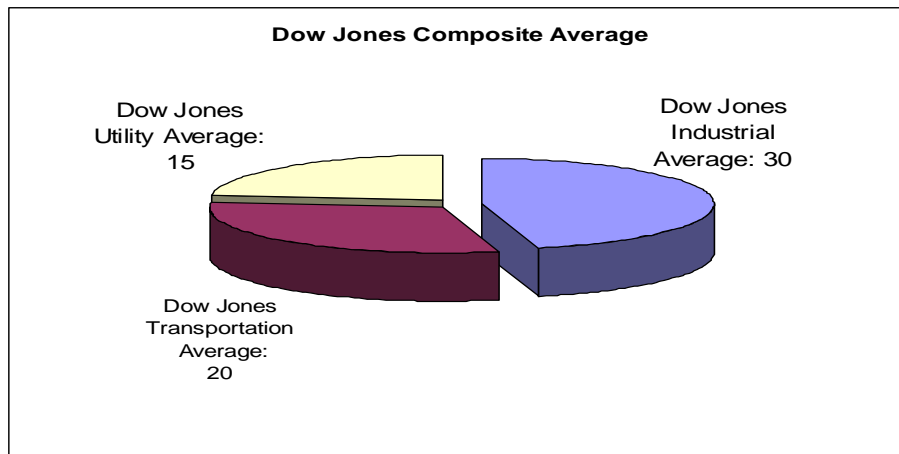


Figura 3. Structura indicelui DJA

Pentru ca datele financiare să respecte rigorile GNetXP a fost nevoie de preprocesarea lor. Pentru testele noastre am luat în considerare date zilnice ajustate ale prețului de închidere pentru acțiunile componente ale indicelui DJA pe perioada 200-2007. Indicele este divizat în trei sub-indici (Dow Jones Industrial Average – sectorul industrial, Dow Jones Transportation Average – sectorul transporturilor și Dow Jones Utility Average- sectorul companiilor din sectorul utilităților) în funcție de domeniul principal de activitate al companiilor (Figura 3).

Prețul în sine nu spune foarte mult despre valoarea unei acțiuni printre altele, fiind nevoie de o unitate unică de măsură a performanțelor pentru comparare. Am început prin a calcula rentabilitatea zilnică logaritmică pentru fiecare acțiune, obținând o imagine omogenă a evoluției lor pe parcursul perioadelor analizate.

Pentru a putea proiecta datele ca și traiectorii am realizat o scalare a lor, considerând ca punct de plecare nivelul de 100 de puncte la începutul fiecărui an și am aplicat rentabilitățile logaritmice obținute la pasul anterior. Figura 4 arată datele din anul 2000 pregătite pentru testare.

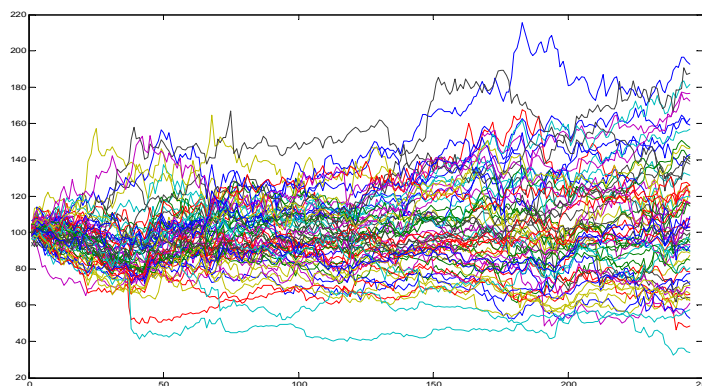


Figura 4. Datele pregătite pentru prelucrare

5.3 Rezultatele experimentale

Aplicând procedura descrisă anterior, pentru cei opt ani analizați am obținut datele pregătite pentru testare. Algoritmul a fost rulat de un număr de 30 de ori pentru a obține rezultate cu acuratețe crescută (Figura 5). Algoritmul hibrid a avut o evoluție în concordanță cu așteptările, puțini dintre clusteri fiind diferiți de la un experiment la altul.

Scopul nostru în analizarea rezultatelor s-a îndreptat în trei direcții:

- Observarea performanțelor algoritmului și asigurarea că acuratețea rezultatelor rămâne la un nivel acceptabil;
- Observarea grupării companiilor după clusterizare, în comparație cu clasificarea lor naturală, în funcție de industrie;
- Identificarea de corelații noi între companii, din punct de vedere economic (Moldovan & Silaghi, 2009).

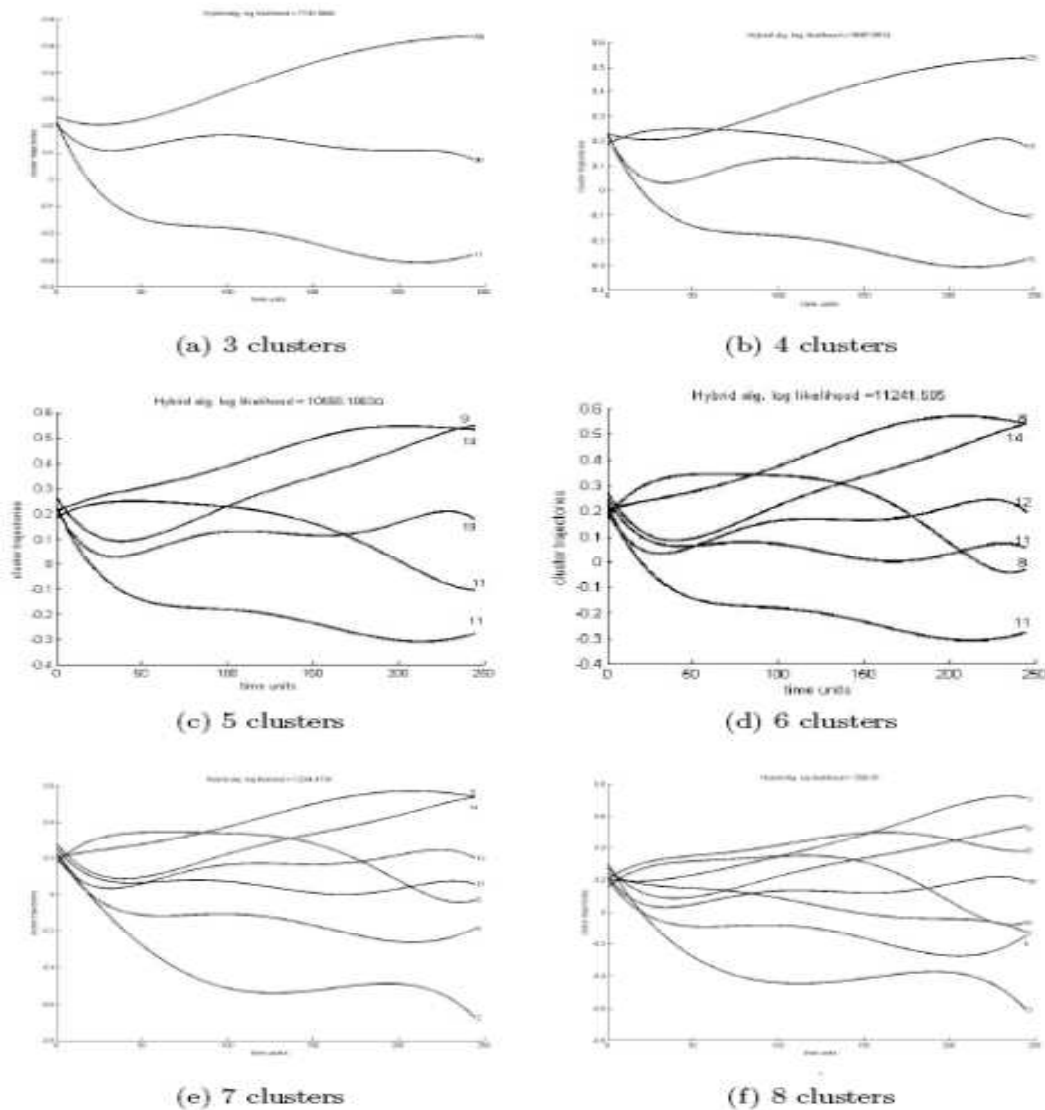


Figura 5. Clusterii obtinuti

5.4 Concluzii

Scopul cercetării a fost acela de a crea un sistem prototip în analiza datelor financiare prin utilizarea metodelor de data mining.

După preprocesarea datelor, algoritmul a fost testat și rezultatele au fost încurajatoare pentru continuarea cercetării. Analizând clusterii formați am concluzionat că algoritmul este potrivit pentru utilizarea lui în scopul clusterizării datelor financiare.

Rezultatele obținute au arătat că în multe cazuri diviziunea naturală a companiilor nu este cea mai adecvată, și că există multe corelații între evoluțiile acțiunilor care ar trebui luate în calcul pentru investiții.

Deasemenea, am descoperit că formarea clusterilor pe baza metodologiei propuse este consumatoare de resurse și că pentru o analiză mai complexă sau pentru un sistem care să ofere date în timp real este nevoie de putere computațională mult mai mare, cum ar fi un sistem de calcul paralel.

6. Concluzii finale și posibilități de dezvoltare

Teza este organizată în patru părți principale, dintre care primele două (capitolele 2 și 3) reprezintă suportul teoretic al cercetării, oferind baza de informații esențiale în dobândirea cunoștințelor necesare realizării cercetărilor practice. Cele două capitole fac atât o analiză a domeniului de afaceri cât și metodologiile utilizate în cercetare.

Scopurile noastre au fost, pe de o parte, de a crea un model de afaceri pentru tranzacționarea automată pe piața financiară românească, cercetare descrisă în capitolul 4; pe de altă parte s-a investigat în capitolul 5 posibilitatea aplicării în finanțe a unei metodologii utilizată inițial în medicină la studierea traiectoriei genelor reglatoare (Chan&Kasabov,2004)(Chan, Kasabov, & Collins, 2005).

Din cercetarea bibliografică am identificat anumite aspecte ale domeniului de Data Mining Financiar, parametrii dezvoltării sectorului financiar, evidențiind principalele direcții de interes și tendințe în sectorul metodelor computaționale inteligente aplicate în finanțe. Tehnicile de Data Mining pot fi utilizate în studiul seriilor de timp, în identificarea anumitor tipare și în detectarea anomaliilor sau în a determina scenario cu probabilitate crescută de succes sau de risc. Deasemenea, pot fi utilizate pentru analiza fluctuațiilor de preț, îmbunătățirea indicatorilor din analiza tehnică și prin utilizarea unor metode inteligente așa cum sunt rețelele neuronale artificiale sau algoritmi evolutivi să se obțină predicții cu grad ridicat de acuratețe.

După studierea oportunității utilizării metodelor inteligente pentru pentru analizarea datelor financiare am concluzionat că acestea sunt recomandate de către profesioniștii în domeniu ca și unele puternice pentru realizarea de predicții și că există o puternică interoperabilitate a metodelor între domenii.

Capitolul 4 a fost dedicat studiului tranzacționării cantitative, descriind pașii necesari în dezvoltarea unei strategii automate de tranzacționare, definind parametrii de performanță și de management a riscului.

Combinarea a trei indicatori tehnici pentru a obține semnale de tranzacționare a fost implementată cu succes. Cu toate că au fost impuse condiții prudențiale, sistemul a înregistrat performanțe superioare strategiilor comparate.

Ca și posibilități de continuare și dezvoltare a cercetării, am propus integrarea cu un algoritm evolutiv ca și soluție pentru modelarea parametrilor indicatorilor și pentru identificarea de noi reguli de tranzacționare.

Scopul cercetării prezentate în capitolul cinci a fost de a crea un prototip cu utilizare în domeniul analizei datelor financiare, utilizând metode specifice de data mining.

Modalitatea de realizare a acestuia s-a concretizat prin aplicarea în domeniul financiar a unei metodologii folosite inițial într-un domeniu diferit: studiul traiectoriei genelor reglatoare. Scopul nostru a fost acela de a studia dacă împărțirea naturală a companiilor în cadrul indicilor în funcție de domeniul lor de activitate este cea mai bună soluție, ținând cont de evoluția prețurilor acestor companii. Mai mult, au fost verificate și posibilele corelații dintre evoluțiile prețurilor companiilor respective, căutând explicații logice și economice pentru acestea.

Ținând cont de analiza clusterilor obținuți am concluzionat că algoritmul utilizat a fost potrivit pentru analiza datelor financiare și pentru realizarea clusterizării propuse.

Rezultatele obținute au arătat faptul că în multe cazuri divizarea tradițională a companiilor în funcție de domeniul de activitate în cadrul indicilor bursieri nu este cea mai potrivită, existând multe corelații între companii din domenii diferite, care trebuie luate în considerare pentru luarea deciziilor de investiții.

Analizând din punct de vedere economic clusterii obținuți, a fost confirmată și din acest unghi utilitatea algoritmului aplicat, identificând multe corelații între companii cu profiluri diferite.

Pentru îmbunătățirea rezultatelor am identificat o direcție de continuare a cercetării în sensul utilizării unui sistem de calcul paralel pentru reducerea timpului de calcul, o necesitate în cazul unor sisteme active de tranzacționare.

Bibliografie generală

Aite Group. (2009). *New World Order: The High Frequency Trading Community and Its Impact on Market Structure*. Boston.

Aite Group. (2010). *The European Equity Electronic Trading Landscape: How Deep Is Your Pool?* London.

Aldridge, I. (2009). *High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems*. Wiley.

Amibroker. (n.d.). Retrieved 2011, from <http://www.amibroker.com/>

Anand Rajaraman, J. U. (2011). *Mining of Massive Datasets*. Retrieved 2011, from <http://infolab.stanford.edu/~ullman/mmds.html>

Babu, G. M. (1994). Clustering with evolution strategies. *Pattern Recognition* , 27 (2), 3210-330.

Banzhaf, W., Nordin, P., Keller, R. E., & Francone, F. D. (1997). *Genetic Programming: An Introduction*. Morgan Kaufman.

Bauer, R. J. (1994). *Genetic Algorithms and Investment Strategies*. Wiley Finance.

Berthold, M., & Hand, D. J. (2007). *Intelligent Data Analysis*. Springer.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Bodas-Sagi, D. J., Fernández, P., Hidalgo, J. I., Soltero, F. J., & Risco-Martín, J. L. (2009). Multiobjective optimization of technical market indicators. *Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers*. (pp. 1994-2004). New York: ACM.

Bollerslev, T. E. (1988). A Capital Asset Pricing Model with Time. *Journal of Political Economy* , 96, 116-131.

Bollerslev, T. (1986). Generalised Auto-Regressive Conditional Heteroscedasticity. *Journal of Econometrics* , 307-327.

Bollerslev, T. (2008). *Glossary to ARCH (GARCH)*. School of Economics and Management, University of Aarhus.

Box, G. E. (1994). *Time Series Analysis: Forecasting and Control*. Prentice-Hall.

Burke, G. (1994). A Sharper Sharpe Ratio. *Futures* .

Burr, B. B. (2005). *Essential book of indexing*. Gale Group .

- Carlin, P. S. (1992). A Monte Carlo approach to nonnormal and nonlinear state space modeling. *Journal of the American Statistical Association* , 493-500.
- Chaboud, A., Chiquoine, B., Hjalmarsson, E., & Vega, C. (2009). *Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market*. Federal Reserve Board - International Finance Discussion Papers.
- Chan, E. P. (2008). *Quantitative Trading: How to Build Your Own Algorithmic Trading Business*. Wiley.
- Chan, Z., & Kasabov, N. (2004). Gene trajectory clustering with a hybrid genetic algorithm and expectation maximization method. *IEEE International Conference on Neural Networks* (pp. 1669 - 1674). IEEE Computer Society.
- Chan, Z., Kasabov, N., & Collins, L. (2005). A hybrid genetic algorithm and expectation maximization method for global gene trajectory clustering . *Bioinformatics and Computational Biology* 3(5) , 1227-1242.
- Chao Jin, V. (2008). An extension of MapReduce for Parallelizing Genetic Algorithms. *eScience '08. IEEE Fourth International Conference on e-Science and Grid Computing*, (pp. 214-221).
- Chen, R. T. (1993). Functional-Coefficient Autoregressive Models. *Journal of the American Statistical Association* , 298-308 .
- Chen, R. T. (1993). Nonlinear Additive ARX Models. *Journal of the American Statistical Association* , 955-967 .
- Cherkaue, K., & Shavlik, J. (1996). Growing simpler decision trees to facilitate knowledge discovery. *Proceedings of the 2nd International Conference of Knowledge Discovery & Data Mining* (pp. 315-318). AAAI Press.
- De Jong, V. (1975). Analysis of the behavior of a class of genetic adaptive systems. *Ph.D. Dissertation* . University of Michigan.
- Deza, M. M. (2009). *Encyclopedia of Distances*. Springer.
- Donefer, B. S. (2008). *Risk Management and Electronic Trading*. FIXProtocol.
- Edelstein, H. A. (1999). *Introduction to Data Mining and Knowledge Discovery*. Two Crows Corporation.
- Edwards, R. D., Magee, J., & Bassetti, W. (2007). *Technical Analysis of Stock Trends, 9th Edition (Hardcover)*. American Management Association.
- Embrechts, P. (1999). Extreme Value Theory as a Risk Management Tool. North American. *Actuarial Journal* .
- Engle, R. L. (1987). Estimating Time Varying Risk Premia in the Term. *Econometrica* , 55, 391-407.

- Engle, R. (2001). The use of ARCH/GARCH Models in Applied Econometrics. *Journal of Economic Perspectives* , 23-33.
- Fama, E. (1965). The Behaviour of Stock Market Prices. *Journal of Business* , 34-105.
- Financial Services Authority. (2009). *The turner review: A regulatory response to the global banking crisis*. Retrieved from http://www.fsa.gov.uk/pubs/other/turner_review.pdf
- Fisher, R. A. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceeding of Cambridge Philoshophical Society*, (pp. 180-190).
- Freitas, A. (2003). A survey of evolutionary algorithms for data mining and knowledge discovery. In *Advances in evolutionary computing* (pp. 819 - 845). Springer-Verlag.
- Freitas, A. A. (2002). *Data Mining and Knowledge Discovery with Evolutionary Algorithms*. Springer.
- Fuller, W. A. (1996). *Introduction to statistical time series*. Wiley Series in Probability and Statistics.
- Gavrilov, M., Anguelov, D., Indyk, P., & Matwani, R. (2000). Mining the stock market: which measure is best? *Proceedings of the sixth ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 487-496). ACM.
- Gerald, A. (1999). *Technical Analysis Power Tools for Active Investors*. Financial Times Prentice Hall.
- Glosten, L. J. (1993). On the Relation Between the Expected Value and. *Journal of Finance* , 48, 1779-1801.
- Goldeberg, D. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Professional.
- Granger, C. W. (1978). *An Introduction to Bilinear Time Series Models*. Vandenhoeur and Ruprecht.
- Hamilton, J. (1989). A New Approach to the Economic Analysis of Non-Stationary Time. *Econometrica* , 357-384.
- Haupt, R. H. (2004). *Practical Genetic Algorithms*. Wiley Interscience.
- Hoffmeister, F., & Bäck, T. (1991). Genetic self learning. *Towards a Practice on Autonomous Systems: Proceedings of the First European Conference on Artificial Life* (pp. 227-235). Paris: MIT Press.
- Holland, J. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press.
- J. Arneric, E. J. (2007). Theoretical Distributions in Risk Measuring on Stock Market. *8th WSEAS Int. Conference on Mathematics and Computers in Business and Economics*.
- Jobson, J., & Korkie, B. M. (1981). Performance Hypothesis Testing with the Sharpe and Treynor Measures. *Journal of Finance* , 889-908.

- John Y. Campbell, A. W. (1996). *The Econometrics of Financial Markets*. Princeton University Press.
- Johnson, B. (2010). *Algorithmic Trading and DMA: An introduction to direct access trading strategies*. Myeloma Press.
- Jorion, P. (2000). *Value at Risk: the new benchmark for managing financial risk*. McGraw-Hill.
- Kalyanmoy, D. (2001). *Multi-Objective Optimization Using Evolutionary Algorithms*. John Wiley & Sons.
- KDnuggets. (2007). Retrieved from KDnuggets.com:
http://www.kdnuggets.com/polls/2007/data_mining_methodology.htm
- Keating, S. (2002). A Universal Performance Measure. *Journal of Performance Measurement* .
- Kestner, L. (1996). Getting a Handle on True Performance. *Futures* .
- KIM, K. (2007). *Electronic and Algorithmic Trading Technology*. Burlington: Elsevier.
- Kovalerchuk, B., & Vityaev, E. (2000). *Data Mining in Finance: Advances in Relational and Hybrid Methods*. New York: Kluwer Academic Publishers.
- Lance, G. W. (1966). A General Theory of Classifactory Sorting Strategies. *Computer Journal* , 373-380.
- Lane, G. (1984). Lane's Stochastics. *Technical Analysis of Stocks and Commodities* (2), pp. 87-90.
- Larose, D. T. (2006). *Data Mining Methods and Models*. New Jersey: John Wiley & Sons.
- Lewis, A. S. (1991). Nonlinear Modeling of Time Series Using Multivariate Adaptive Regression Splines . *Journal of the American Statistical Association* .
- Lin, L., & Cao, L. (2008). Mining in-depth patterns in stock market. *International Journal of Intelligent Systems Technologies and Applications* .
- Longerstaey, J., & Spencer, M. (1996). *RiskMetrics—Technical Document*. New York: Morgan Guaranty Trust.
- Maggini, M., Giles, C. L., & Horne, B. G. (1997). *Financial Time Series Forecasting Using K-Nearest Neighbors Prediction*. Finance & Technology Publishers.
- Makulowich, J. (1999). Government Data Mining Systems Defy Definition. *Washington Technology* .
- Mardia, K., Kent, J., & Bibby, J. (1979). *Multivariate Analysis*. Academic Press.
- McLachlan, G. K. (1997). *The EM Algorithm and Extensions*. John Wiley and Sons.
- Mills, T. (1990). *Time series techniques for economists* . Cambridge University Press.
- Mitchell, J. (1999). *An Introduction to Genetic Algorithms* (Fifth Edition ed.). MIT Press.

- Moldovan, D. (2010). Testing the efficiency of the Romanian stock market. *International Conference on Development, Energy, Environment, Economics*, (pp. 378-381). Tenerife, Spain.
- Moldovan, D., & Silaghi, G. C. (2009). A clustering of DJA stocks – an application in finance of a method first used in gene trajectory study. *Analele Universității Oradea*, 5 (1), 1006-1011.
- Moldovan, D., & Silaghi, G. (2009). Gene Trajectory Clustering for Learning the Stock Market Sectors. *Lecture Notes in Computer Science vol. 5494*, 559-569.
- Moldovan, D., Moca, M., & Nitchi, S. (2011). A Stock Trading Algorithm Model Proposal, based on Technical Indicators Signals. *Informatica Economica*, 15 (1), 183-188.
- Mulpuru, S. (2011). *US Online Retail Forecast, 2010 To 2015*. Forrester Research.
- Mutu, S., Balogh, P., & Moldovan, D. (2011). The Efficiency of Value at Risk Models on Central and Eastern European Stock Markets. *International Journal of Mathematics and Computers in Simulation*, 5 (2), 110-117.
- Nelson, D. (1991). Conditional Heteroskedasticity in Asset Returns: A new Approach. *Econometrica*, 59 (2), 347-370.
- Nitchi, S., & Avram-Nitchi, R. (1997, Feb). Data mining, o noua era in informatica. *Byte Romania*.
- Pardo, R. (2008). *The Evaluation and Optimization of Trading Strategies*. Wiley.
- Priestley, M. B. (1980). STATE-DEPENDENT MODELS: A GENERAL APPROACH TO NON-LINEAR TIME SERIES ANALYSIS. *Journal of Time Series Analysis*.
- Rechenberg, I. (1989). Evolution Strategy: Nature's Way of Optimization. *Optimization: Methods and Applications, Possibilities and Limitations*, 106-126.
- Richard J. Roiger, M. G. (2002). *Data Mining: A Tutorial Based Primer*. Addison Wesley.
- Russel, S., & Norvig, P. (2003). *Artificial Intelligence: A Modern Approach (2nd Edition)*. New Jersey: Pearson Education.
- Russell, S., & Yoon, V. (2005). Heterogeneous Agent Development: A Multi-Agent System for Testing Stock Trading Algorithms. *AMCIS 2005 Proceedings*.
- Sayad, S. (2011). *Real Time Data Mining*. Self-Help Publishers.
- Sharpe, W. (1966). Mutual Fund Performance. *Journal of Business*, 119-138.
- Sharpe, W. (1994). The Sharpe Ratio. *The Journal of Portfolio Management*, 21 (1), 49-58.
- Shearer, C. (2000). The Crisp-DM Model: The New Blueprint for Data Mining. *Journal of Data Warehousing*, 13-22.

- Silaghi, G. C., & Robu, V. (2005). An agent strategy for automated stock market trading combining price and order book information. *2005 ICSC Congress on Computational Intelligence Methods and Applications*. IEEE Computer Society.
- Sortino, F., & van der Meer, R. (1991). Downside Risk. *Journal of Portfolio Management* .
- Subramanian, H., Ramamoorthy, S., Stone, P., & Kuipers, B. (2006). Designing Safe, Profitable Automated Stock Trading Agents Using Evolutionary Algorithms. *Proceedings of the Genetic and Evolutionary Computation Conference*.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2005). *Introduction to Data Mining*. Addison-Wesley.
- Thompson, S. (1999). Pruning boosted classifiers with a real valued genetic algorithm. *Knowledge-Based Systems* , 12 (5-6), 277-284.
- Todea, A. (2005). *Eficiența informațională a piețelor de capital. Studii empirice pe piața românească*. Cluj-Napoca: Casa Cărții de Stiință.
- Tong, H. (1983). *Threshold Models in Non-linear Time Series Analysis*. Springer-Verlag.
- Treynor, J. (1965). How to Rate Management of Investment Funds. *Harvard Business Review* .
- Tsay, R. S. (2002). *Analysis of Financial Time Series*. Wiley-Interscience.
- US Census Bureau. (2011, May 26). *E-stats*. Retrieved 2011, from US Census Bureau: <http://www.census.gov/econ/estats/index.html>
- Vlaar, P. (2000). Value at risk models for Dutch bond portfolios. *Journal of Banking & Finance* , 1131-1154.
- Ward, J. H. (1963). *Hierarchical grouping to optimize an objective function*. Am. Statist. Assoc.
- Weigend, A. S. (1999, January 20). *Data mining in finance - Course notes*. Retrieved 2011, from <http://www-psych.stanford.edu/~andreas/Teaching/DataMiningFinance/S99/Notes/01Description.html>
- Witten, I. H., & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier.
- Wold, H. (1938). *A Study in the Analysis of Stationary Time Series*. Almqvist & Wiksells.
- Young, T. (1991). Calmar Ratio: A Smoother Tool. *Futures* .
- Zanasi, A. (1998). Competitive intelligence through data mining public sources. *Competitive Intelligence Review*, 9 , 44-54.