**ECHIPA D**

**Articol 1**

**Sentiment Analysis of Financial News Articles**

Autori: Robert P. Schumaker 1 , Yulei Zhang 2 and Chun-Neng Huang 2

**Obiectivul principal** reprezintăasocierea unui sistem de predicție a articolelor de știri financiare, AzFinText, cu tehnici de analiza a sentimentelor**.**

**Metodologia propusă**

Proiectarea sistemului AZFinText – componente:

* Date numerice
* Analiza textuală
* După colectarea datelor, AZFinText face predicții de preț +20 de minute pentru fiecare articol financiar de știri. **Din testele empirice anterioare**, s-a constatat că includerea reprezentării **substantivului propriu** și a **prețului acțiunilor** în momentul lansării articolului de știri, a oferit AZFinText **o performanță predictivă superioară** în comparație cu alte reprezentări textuale și diferite informații despre preț.
* Etapa **Model Building a AZFinText –** partiționează datele pentru a răspunde cel mai bine la întrebările de cercetare. În cazul obiectivității/subiectivității se folosește **OpinionFinder.** Pentru algorimtul de învățare automată s-a implementat **SVR** (**Sequential Minimal Function**)- **functia de optimizare**. Această funcție permite **predicție numerică discretă** în loc de clasificare. S-a selectat un nucleu liniar și de zece ori validare încrucișată.

**Corpusul experimentului**

**Timpul:** perioada consecutivă de 5 săptămâni (linia de bază experimentală) – 26 octombrie 2005- 28 noiembrie 2005 și încorporează 23 de zile de tranzacționare.

După niște reguli de selcție bine stabilite s-au filtrat **9 211 de articole de știri** și în mod similar s-au strâns **10 259 042 cotații de acțiuni/minut**.

Întregul corpus de articole de știri financiare sunt reprezentate de substantivele proprii în binar. Dacă o anumită caracteristică de Substantiv adecvat este prezentă în articol, acelei caracteristici i se dă 1, altfel 0 și apoi stocate în baza de date. În mod similar, fiecare articolul de știri financiare este rulat prin **OpinionFinder** pentru a identifica tonul său general, este mai mult Obiectiv sau subiectiv, și polaritatea acestuia, este articolul mai pozitiv sau negativ. In tandem, cotațiile de stoc adunate pe minut și stocate. Pentru a construi un model, am asociat mai întâi Substantivele proprii reprezentative și cotația stocului în momentul lansării articolului, pentru fiecare articol de știri financiare. Apoi, în funcție de modelul particular testat, datele sunt agregate și transmise către componenta noastră de învățare automată pentru instruire și testare. Prețul acțiunii predicțiile sunt apoi făcute pentru fiecare articol de știri financiare și transmise evaluării instrumentelor.

**Rezultatele obținute**

Din investigația cercetătorilor s-au găsit câteva rezultate interesante. Prima dintre ele a fost aceea **AZFinText** a fost cel mai în măsură să prezică articole subiective în Precizie Direcțională (59,0% până la 50,4%) și rentabilități de tranzacționare (3,30% - 2,41%), dar nu și apropiere (0,103 față de 0,0516). S-a simțit că **subiectivitatea articolelor poate fi influențată comportamentul de tranzacționare**. Al doilea rezultat notabil a fost că **AZFinText** a fost cel mai în măsură să prezică articole subiective negative în Precizie Direcțională (50,9% la 50,4%) și randamente de tranzacționare (3,04% la 2,41%), dar nu și apropiere (0,0576 versus 0,0516). Se consideră că aceste rezultate pot fi atribuite investitorilor care reacționează mai puternic la articole negative. Al treilea rezultat notabil a fost că AZFinText a găsit dovezi ale Activității contrare. AZFinText a fost mai capabil să prezică scăderi în articolele pozitive (53,5%) și creșteri în articolele negative și neutre (52,4%, respectiv 49,5%).

**Articol 2**

**SENTIMENT ANALYSIS OF FINANCIAL NEWS ARTICLES USING PERFORMANCE INDICATORS**

Autor: *Srikumar Krishnamoorthy*

Prin aceasta lucrare autorul doreste sa imbunatateaza analiza sentimentului financiar bazat pe dictionar specific domeniului, deoarece in studiile recente de cercetare, este demonstrat faptul ca, sunt rezultate superioare cand se utilizeaza un astfel de dictionar.

**Obiectivul principal** urmarit in aceasta lucrare de cercetare este de a examina utilizarea indicatorilor de performanta pentru a prezice polaritatea in textele financiare.

**Lucrarea propusa prezinta urmatoarele:**

1. O metoda de predictie a polaritatii textului financiar bazata pe indicatori de performanta financiari si nefinanciari (HSC)
2. O regula de asociere, un clasificator ierarhic pentru a prezice sentimentul

**Evaluarea experimentala** a fost realizata pe un set de date bancar, Malo2014 format din aproximativ 5000 de propozitii.

Aceasta lucrare propune un clasificator ierarhic al sentimentelor care utilizeaza indicatori de performanta pentru predictia sentimentului financiar . Abordarea propusa s-a dovedit a fi destul de eficienta, obtinand o precizie de peste 80% si o masura F de peste 68% in toate seturile de date studiate.

**Articol 3**

**Financial News Sentiment Analysis Using LexiconBased Labelling and Machine Learning-Based Algorithm**

Autori:Ling Wu, Siew Hock Ow

**Obiectivul principal** este de a analiza sentimentul de știri financiare care pot fi utilizate încontinuare pentru a analiza impactul acestor știri financiare sentimentele cu privire la mișcarea prețului bursier. Se propune utilizarea etichetării bazată pe lexic și pe algoritmi bazați pe machine learning.

**Metodologia** folosităaplică algoritmii de învățare automată supravegheați pentru a construi un clasificator, analizează datele etichetate și formează în general predicția de către caracteristici învățate. Analiza sentimentelor de știri este o componentă a sarcinii de procesare a limbajului natural (NLP) în domeniul Inteligenței Artificiale. Există diferiți algoritmi de învățare automată supravegheați, precum: Naïve Bayes, mașini vectoriale de suport (SVM), regresie logistică și arbori de decizie. Pentru implementarea părții NPL, python are un modul specific numit Natural Language Toolkit (NLTK) pentru a efectua saricinile NLP și un modul ușor de utilizat Scikit-Learn care integrează învățarea automata.

**Corpusul** este format din seturile de date utilizate atât din date primare, cât și din date secundare. Aceste seturi de date textuale ale știrilor OPEC conțin 135 de articole de știri care au fost publicate la „Comunicat de presă” a site-ului oficial al OPEC, de la data de 14 iunie 2012 până la 21 decembrie 2017. Acest set de date a fost generat de surprinzând articolele de știri publicate la rubrica „Piețe” din Wall Street Journal Site web, de la 7 iulie 2017 până la 26 iulie 2017. Aceste 50 de zile de date de știri financiare ale WSJ conțin 2.062 articole de știri financiare, care au fost înstrăinate în 37.452 de propoziții, sunt suficient pentru a efectua cercetarea. Acest set de date a fost procesat în continuare ca date de instruire pentru algoritmii de învățare automată.

**Rezultatele** aceastei cercetări arată că cei șapte clasificatori bazați pe algoritmi de învățare automată supravegheați folosiți și testați separate, prezintă în tabelul 1 raportul de clasificare a acestor clasificatori care includ rezultatele. cu cât scorul de precizie este mai mare, cu atât este mai bun performanța clasificatorului. Clasificatorul descendentului de gradient stochastic a atins un procent de 70% de precizie. Tabelul 2 prezintă scorurile sentimentelor articolelor de știri OPEC folosind algoritmul Clasificator descendent de gradient stochastic, scorurile sentimentelor variau de la -1 la 1.

**Articol 4**

**SEMANTIC BASED SENTIMENT ANALYSIS IN FINANCIAL NEWS**

Autori: Juana María Ruiz-Martínez1, Rafael Valencia-García1, Francisco García-Sánchez1

**Obiectivul principal** urmarit este de a realiza o analiza sentiment semantica asupra stirilor financiare. In cadrul acestui articol este prezentat un algoritm semantic pentru extragerea opiniilor aplicat domeniului financiar.

**Metodologia propusa** este sustinuta de metode de procesare a limbajului natural pentru a comenta stiri financiare in conformitate cu ontologia financiara, iar apoi stirile financiare comentate sunt analizate prin trecerea acestora printr-un numar de liste de gazete ce rezulta in doua seturi separate, un set cu stiri financiare pozitive si un set cu stiri financiare negative.

**Corpusul experimentului** contine 57.210 cuvinte si 900 de abstracte a stirilor financiare. Acesta a fost extras din fluxurile RSS si fiecare stire a fost etichetata manual ca fiind pozitiva sau negativa de catre 2 comentatori diferiti. Acest sistem constituie baza pentru evaluare si in continuare este descris modul in care functioneaza: daca rezultatul afisat de sistem se potriveste cu stirile etichetate manual atunci rezultatul este considerat a fi corect. In cadrul experimentului un numar total de 5 interogari au fost utilizate de sistem pentru a gasi informatii in domeniul financiar.

**Rezultatele obtinute** ale analizei sentiment sunt foarte promitatoare, cu o medie a acuratetei agregate de 87%. Important este de mentionat faptul ca in privinta rezultatelor s-a luat in considerare decizia finala a sistemului si nu procesul pe care sistemul il efectueaza pentru a produce o asemenea decizie.