International Journal of Engineering & Technology, 7 (2.5) (2018) 27-31

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research Paper

A scientific review of soft-computing techniques and methods for stock market prediction

Muhammad R. Islam [1]*, I.F.T. Al-Shaikhli[2], A. Abdulkadir[1]

[1]Department of Computer Science, Kulliyyah of Information and Communication Technology, International Islamic University Malaysia, 53100 Gombak, Selangor, Malaysia
[2]Department of Computer Science, Kulliyyah of Information and Communication Technology, International Islamic University Malaysia, 53100 Gombak, Selangor, Malaysia
[3]Department of Computer Science, Kulliyyah of Information and Communication Technology, International Islamic University Malaysia, 53100 Gombak, Selangor, Malaysia
**Corresponding author E-mail: rabiul.islam@live.iium.edu.my*

Abstract

Information could be power if most technological approach engaged in the era of IT. Web based text mining is one of the approach that could be analyzed in many ways through soft-computing methods and techniques. The analytical result has shown that different methods of text mining have several advantages with the gap of knowledge that is required to improve. This paper explores the performance of various methods and its impact on specific text mining field such as web based financial text analysis for stock prediction. Key research area of financial text mining is becoming one of the potential research field based on the source of online news, forums, blogs or social media.

Keywords: *Web-Text Mining, Soft-Computing Methods, Stock Market Prediction.*

1. Introduction

In the era of globalization, information technology is emerging on its usability from different sectors for further action or making decision based on World Wide Web (www). Text mining is one of the techniques of high quality informative text extraction from set of text documents. Text mining used to gain informative knowledge from text documents or web in terms of a particular operation e.g. predicting, analyzing, statistical result etc.

Financial predictive measures depend on fundamental and technical analysis quantitative market data as found on the enormous online text news source [1]. It has become the scope of information source that augmented to the mining scope through textual classification or statistical analysis in the field of computer science.

Nowadays, fundamental online data is available in newspaper, forums, social media and many sources [2]. Web-Text-Mining explored the predictability result by analytical approach of mining task. The performance of mining task among the text varies on different methods that leads to financial text to bring more clarification and direction of the stock market movement. Limited success of stock price prediction from textual data spread the existing text mining field such bag of words, linguistic textual representations, noun phrases etc [3].

All the possible combinations are widely used in preprocessing tasks for evaluating among the domains and its classification within various feature dimensions [4]. The review of this paper mainly focuses on the following. To identify the current economy methods and factors in current stock market prediction (Section 2). Web-Text-Mining techniques for stock price prediction (Section-3). The rest of this paper discusses the comparison within different soft-computing methods and interdisciplinary techniques of text mining which suggests potential further research of text mining (Section-4).

2. Methods and factors of stock market prediction

Many soft computing methods and theories are applicable to stock market predictions due to its performance on electronic media application as described below.

2.1. Fundamental and technical analysis in stock market

Stock market has been studied extensively with the predictive power of news articles to financial market. The performance of market efficiency in current economy is relatively dependent on electronic media sources such as corporate announcements, blogs, online forums and social media [5, 6]. Several theories are available based on stock market prediction. For instance, Efficient Market Hypothesis (EMH) to bring availability of information so that everyone has access to breakout further forms of market condition: weak, semi-strong and strong [7]. According to EMH, any instantaneously given news makes it impossible to outperform the market consistently [3]. In some cases, Random walk theory determined the price randomly and outperforming the market is infeasible but it has same accessible opportunity of publication information [3]. Based on these theories trading philosophies emerged on fundamental and technical analysis.

2.2. Fundamental philosophy

In this philosophy, security price will be determined through the nuts and bolts of financial numbers that are derived from the over-

all current economy condition. The structure will be figured out from inflation, joblessness, individual price to earn, return on equity and debt levels [8].

2.3. Technical analysis

Technical analysis always depends on time-series and historical data that are obvious on market timing through critical analysis and opportunities that could be found through volume movements comparing the current market prices. Stock price movements are not totally random [8]. In such cases, technical analysis generates some of the interpretation of market price movements comparing the other observation phenomenon in stock market. Certain techniques are developed based on online financial news articles for simulated traders followed by hierarchical rule which allowed traders through technical analysis that creates the time lag for receiving information to trade. Receiving information creates time lag and make a challenge for traders to tread in short period of time that can overcome through machine learning approach by different techniques and market analysis.

2.4. Market hypothesis

Efficient Market Hypothesis (EMH) and Adaptive Market Hypothesis (AMH) are both engaged informationally to market movement. However, the result of AMH is better than the result of EMH. In this case, the important time scale consistent the Hurst exponent with EMH which helps to measure the long-term time series and decrease the lag between pairs of values [9]. In a sense, available and instant information of market accurately gives to the modern investment as required time frame. Whereas AMH implies with the degree of market efficiency and behavioral based of components such as number of competitors in the market, adaptability of the market participants and magnitude of profit opportunities [10]. Basically, these hypotheses highly influence the stock market.

2.4. Stock market forecasting and analytical techniques

Stock market prediction thoroughly involved AMH and EMH with the range of behavioral economics. Real world market may perform inefficiently or efficiently due to economics fluctuation based on developed or developing countries [11]. Forecasting in short term and long term variants makes difference in predictive ability with market informative that makes more efficient overtime [12]. Other analytical techniques like sentiment and emotional based text focused on customer feedback to find the relation between current market condition and behavioral-economics [1]. This text mining technique basically focuses on positive or negative emotional stance that can bring up as the result of market prediction. Algo-Trading (AT) using sophisticated program system like decision making, dynamic planning, involves learning and reasoning to automate all or few parts of trade cycle [13]. In such cases, Evolutionary Computing (EC) techniques used like fuzzy logic, genetic programming and Genetic Algorithm (GA) [14].

3. Web-text-mining techniques for stock market prediction

Deriving required or high quality information from set of text documents is known to be text mining. Text mining could be analyzed through categorization, clustering or entity retrieval from set of text. The effective methods of text mining are described here for its performance of comparison and essential further improvement for the current problem. Address of this research problem is to find out different methods, techniques, behavioral-economics and text mining analysis that are adequately related to stock market prediction. Financial prediction is required to know the technical and fundamental analysis for trading in terms of market fluctuation.

The study of the research defines the potential topics as it is necessary to improve within comprehensive research problems. Several text mining methods followed by text extraction techniques as it is oriented within four methods [15]. i) Term Based Method (TBM) ii) Phrase Based Method (PBM), iii) Concept Based Method (CBM) iv) Pattern Taxonomy Method (PTM). Other methods of text mining analysis through discussion based on economy, web-textual information, calculation, very high dimensionality of data and large size of databases. The following discussion brings the technical analysis and problem based on current methods text mining.

3.1. Term based method (TBM)

Terms as a word in sentence, always occur on web text content or documents in where contextual phenomenon consider as “cluster”. Based on terms, two types of clustering are applicable, Frequent Term-based Clustering (FTC) and Hierarchical Frequency Term-Based Text Clustering (HFTC) that are applied on structural large set of hypertexts to determine a flat cluster [16]. The experiment result showed that the FTC is more efficient than other competitors [16]. In term based method, only polysemy and synonymy give some problem due to structural informative knowledge [15]. Such challenge especially frequent term sets efficiently discovered by using Apriori algorithm [16]. Informative web and organization intranets contain large size file, high dynamics and large diversity documents that give some challenging retrieval task which overcome through text mining methods. Term distributions are best expressed as simple distributions when text is classified into a negative or positive class as defines the result [17]. Experiment result proved that specific field of text mining from relevant web followed thoroughly FTC and HFTC.

3.2. Phrase based method (PBM)

Phrase based text mining is much discriminative and low ambiguous than TBM. However, due to low occurrence in statistical and frequency, PBM produces the low performance of diacritic text mining. In most cases, online text news contains some redundant and noisy phrase as usual other documents [18]. Specially phrase searching, information retrieval and boost query processing fairly straightforward based on grammatical features [18]. In this case, episode rules applied for phrase as the result of postprocessing. These postprocessing techniques are related to ordering, grouping and pruning those are enhanced by using knowledge based rules to make solution from single text document [18].

3.3. Concept based method (CBM)

This method relies based on natural language processing techniques. Online text involves in news portal but processing of text extraction in concept based method (CBM) is bit different than TBM and PBM. CBM able to distinguish between meaningful and non-important terms so that it can make sense what sentence mean in human understanding [19]. Large number of online text extraction has low sensitive to noise in terms of normalize about feature vectors. To overcome the problem in this model, generic argument structures able to analyze sentence but labeled verb structures may occur double in single sentence. In this condition, both verb and argument are considering as terms and these terms play important semantic roles that contribute the meaning of the sentence [19]. This concept may either be a word or phrase that depends on the semantic structure of the meaning full sentence.

3.4. Pattern taxonomy method (PTM)

In this method, pattern is the main focus in text mining that is using “is-a” relation for structural operation in sentence basis. It is also associated with related pattern mining, sequential pattern

mining, frequent item set mining and association rule mining [20]. In text mining, the association rule could be ineffective for its market analysis because having lack of information [15]. PTM may misconstruction of patterns that could lead to ineffective performance in result. Other experiment result proved that PTM give the best performance among any other methods if it overcomes misinterpretations problems of text mining.

4. Financial term based text mining technique

Financial news mining analysis through financial term based known to be financial text mining for its own perspective. Textual categorization based on online financial text perform on terms based method that tries to avoid the ambiguity problem in same document. Online financial text mining analyzed through details of financial text semantic mining that able to overcome some challenging issues like prediction [15]. The study of this research suggests to specific financial terms those are much related to financial texts that vary the meaning of financial direction and market analysis in knowledge based. Refer to Fig-1 figures out the prediction mechanism of text mining.

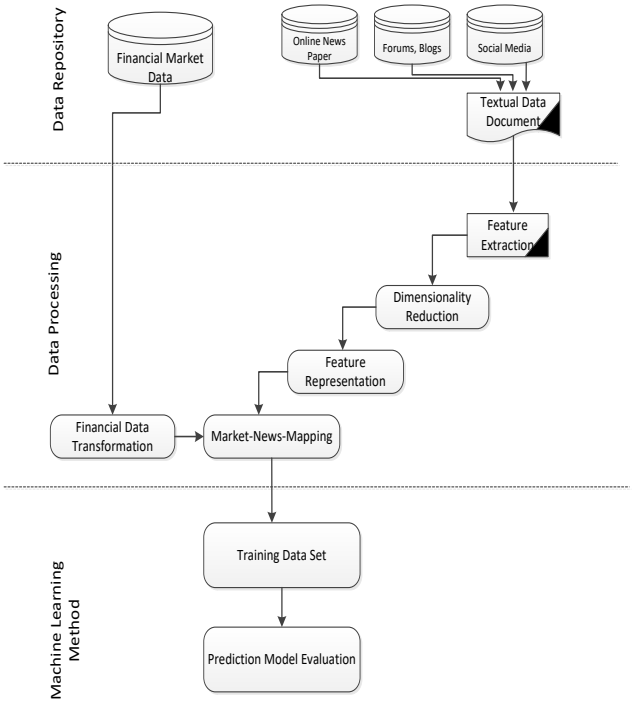

Fig 1: The Entire Path of Data Diagram and Its Components

4.1. Financial text mining classification

Online base financial data or news collected from respective companies or regular financial news sources which help to create Emotion Words Dictionary for its sentiment score of analyzing the textual data [2]. Before giving weighting score, text must be classified from online finance document model. Fig-2 showed the model of text document lemmatizing process [21]. In the first tokenization step HTML tags will be removed then replaced all the characters converting to capital letters to small letters. All the stop words removed and remaining tokens were lemmatized with the WordNet lemmatize [21].

In this process, chronological order helps to separate the textual data into two sets of classification as seen 90% training set data will be prepared for machine learning, document presentation, dimension reduction and remaining 10% were used as test set [21]. If data isn't lemmatized in chronological form, then it returns to document step. The model also shows that tokenization techniques should be repeated until required token is to be gotten back from its source. This classification is done for finance text.

Text mining with hybrid approach combining with the wrappers and filters that usually preferred methods of features selection help to process in short period of time [22, 23]. In the text mining process, filter methods consider as binomial hypothesis testing, ambiguity measure, class discriminating measures, SVM based feature selection algorithm, improved Gini index, information gain, document frequency, mutual information, term strength, odds ratio [23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34]. Text classification tokenization form, segment the text into token form known as text segmentation and in this step non-alphanumeric characters are delimited [35]. Stop-words (conjunction prepositions etc) are removed and upper case letter convert to lower case prior to text classification. Basically text classification could be operated only English language through stemming algorithm [36]. Fig-2 provides the basic structure of the document text mining.

4.2. TF-IDF method using on text mining

Method tf-idf used to find terms from the text documents defined as term frequency or invers document frequency. Term based method cover through tf-idf method as it's formulated [21]. However, the analytical terms of weighting scheme bring out individual words due to different meaning full financial terms. The requirements of applying weighting scheme for text mining prediction need weighting scheme based on knowledge [18]. Generally, in human understanding from web or any news articles written in positive sense. e.g. This sentence: "General Electric Stock Goes Negative, Sets Up for a Short Trade -- Plus Jim Cramer's Take". The words of this sentence extract word would be as result. E.g. "General", "Electric", "Stock", Goes", "Negative", "Sets-up", "Short", "trade". Only such terms are to be taken as tokenization form to be fixed for knowledge based weighting scheme [18]. In final step, the weighting scheme dictionary is analyzed by machine learning algorithm for stock market prediction

.

$$\text{tf-idf}(j) = \text{tf}(t) \times \text{idf}(j)$$

$$\text{idf} = -\log\left(\frac{\text{Number of documents}}{\text{Number of documents containing terms}}\right) \quad (1)$$

According to the 'Eq. (1)' tf-idf(j) indicates the tf-idf value of term j, tf(t) means the term frequency of term t, i.e. how many times the term occurs in single document, idf(j) is the inverse document frequency of term j. Only df(j) means document frequency of term j. Overall in document text mining, frequency term would be computed as important word of phrase [19]. Basically, three factors are considering as term. i) Term Frequency Factor (TFF), ii) Collection Frequency Factor (CFF), iii) Length Normalization Factor [37]. Among these, TFF is only given in document, CFF given to other documents total number of terms and Length Normalization Factor is dependent on first two factors since text document vary greatly in length [37].

5. Comparison of soft-computing methods and techniques for stock market prediction

Many soft computing methods and techniques are applicable for its robust comprehensive performance in the field of stock market prediction [38]. Most performing algorithm are defined here as operational techniques of text mining.

5.1. Machine learning method

Machine Learning Techniques able to get the meaningful knowledge from large set of text data that rapidly growth and available in World Wide Web (www) with the combination of Natural Language Processing (NLP) and Data Mining techniques. Text mining techniques used in various types of data like semi

supervised, unsupervised and supervised which could be enabled to be classified, presented, annotated, integrated and extracted text data from the documents via Machine Learning approach [40]. Furthermore, Hybrid machine learning system enables the Stock Market Forecasting based on Support Vector Machine (SVM) and Genetic Algorithm (GA) [41]. In this case, the performance of the result showed that GA-SVM is better than SVM Machine Learning Method which means hybrid machine learning method in data mining always performs better for stock market prediction.

5.2. Deep learning method

Deep learning method active on event-driven as short or long term for stock market prediction that mainly works with the event sequence from financial news datasets like Reuters and Bloomberg [42]. The result of this approach outperforms more than 6% from previous work. Machine learning and deep learning methods are most effective methods to large text data set based on stock market movement. Some methods and techniques are discussed in the table-1 that comes out with its advantages and disadvantages.

Table 1: Web Text Mining Related work

Author name & Paper Title	Algorithm:	Addressed Problem (Disadvantages)	Proffered Solution (Advantages)
D.K. Kirange et al. "*Sentiment Analysis of News Headlines for Stock Price Prediction*"[2].	SVM, KNN, Naïve Bayes.	Not able to classified by analysis techniques of machine learning or data mining.	Emotion classification and solve concept of stock price prediction through various classifiers.
Christian Soyland et al. "*Interday news-based prediction of stock prices and trading volume"*[19].	Supervised Automated document classification model.	This model still need to succeed in predicting based on intraday stock news.	Document in one-day trade volume experiment and classification accuracy of 78.3%.
Azadeh Nikfarjam et al. *"Text Mining Approaches for Stock Market Prediction"*[42].	Classification Algorithm: SVM & Bayesian Classification.	Unable to captured the semantic gap analysis in news content.	Use the bag of words in different method for features selection and enhance the *tf-id*f method by giving new weight.
Anurag Nagar et al. **"***Using Text and Data Mining Techniques to Stock Market Sentiment from Live News Streams"* [43].	NLP techniques.	Stock prediction from Live News Streams instead of large data.	Creating a news engine for gathering and aggregating news items for from authentic financial sources**.** Text mining solve through R and NLP tools.
Xiao Ding *"Deep Learning for Event-Driving Stock Prediction"*[41]	Deep Learning Method for Event Driving.	Relatively weaker effects of long-term events.	Able to capture the influence of news events from over history.
Ruchi Desai. *"Sdtock Market Prediction Using data Mining"* [40]	Data Mining	This method only supports to large volume of data from historic data set rather than daily data set.	Identify the hidden pattern from historic data that have predictive capability.

5.3. Further potential research work on text mining

Observed from this related work and above discussion found that online text mining is reliable for financial field based on structural words in sentence. Table-1 describes more methods, algorithms, address problem (Disadvantages) and solution (Advantages) of text mining for stock market prediction. However, there is drawback on weighting scale between stock prediction and financial text data [21]. That drawback could be removed by sentence filtering technique for noisy in text data which help build any stock strategic model [2]. So, there is need to work for financial text mining token correction to get the better result of stock market prediction with deep learning method. In the financial text mining functionality fuzzy expert system help to build any strategic model like Stock Prediction Model (SPM) or Risk Estimation Model [43].

6. Conclusion

Recently, Soft-Computing techniques are introduced for the prediction of financial stock market in order to aid brokers and make a choice for trading decisions. This research found that few models exist to perform on that and there is need for the development of more models to strengthen the risk reduction in the prediction from technique adopted. Future work should be addressed for Stock Prediction Model (SPM) to improve the performance of prediction techniques.

Acknowledgements

Authors are giving thanks to Prof. Dr. Imad Fakhri Taha Al-Shaikli, Head of Research at Kulliyyah of Information and Communication Technology for his support by the grants from science fund and endless guidance during the writing of this review paper.

References

[1] Arman Khadjeh Nassirtoussi, Saeed Aghabozorgi, The Yign Wah, David Chek Ling Ngo. 2014, "Text Mining for Market Prediction: A systematic review" Expert Systems with Applications, Vol. 41, pp. 7653-7670, Available: www.elsevier.com/locate/eswa [Accessed 8 August 2016], DOI: 10.1016/j.eswa.2014.06.009.

[2] D. K. Kirange, "Sentiment Analysis of News Headlines for Stock Sentiment Analysis of News Headlines for Stock Price Prediction," 2016, An International Journal of Advance Computer Technology. vol. 5, pp.1-6, . April 2016 pp. 1–6, DOI: 10.13140/RG.2.1.4606.3765

[3] Schumaker, Robert P.Chen, Hsinchun, February 2009, "Textual analysis of stock market prediction using breaking financial news" , ACM Transactions on Information Systems, Vol.27, Available: http://dl.acm.org/citation.cfm?id=1462204 [Accessed: 10 August 2016], DOI: 10.1145/1462198.1462204.

[4] Uysal, Alper Kursat Gunal, Serkan, 2014, "The impact of preprocessing on text classification", Information Processing and Management, Vol. 50, Available : http://dx.doi.org/10.1016/j.ipm.2013.08.006, DOI: 10.1016/j.ipm.2013.08.006.

[5] Wong, Felix Ming Fai, Liu, Zhenming, Chiang, Mung, 2014. "Stock Market Prediction from WSJ: Text Mining via Sparse Matrix Factorization", 2014 IEEE International Conference on Data Mining, 2014 pp. 430-439, DOI: 10.1109/ICDM.2014.116,

[6] Thomas, James D. Sycara, Katia, 1999, "Integrating Genetic Algorithms and Text Learning for Financial Prediction", Data Mining with Evolutionary Algorithms, 2000, pp.72-75, Available:http://citeseer.ist.psu.edu/308554.html%5Cnhttp://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.34.8655&rep=rep1&type=pdf, Accessed: [10th Octobor'16]

[7] Fama, Eugene F, "American Finance Association Efficient Capital Markets: A Review of Theory and Empirical Work", 7th Feb, 2007, Finance. 1970, Vol.25, N0.2, pp. 28-30, DOI:10.2307/2325486, MPID: 19724231

[8] Schumaker, Robert P. Chen, Hsinchun, "Textual analysis of stock market prediction using breaking financial news", February 2009,

ACM Transactions on Information Systems, 2009, Vol.27. pp.1-19. DOI: 10.1145/1462198.1462204, PMID: 20434961

[9] Reboredo, Juan C. Rivera-Castro, Miguel A. Miranda, José G V García-Rubio, Raquel,2013 "How fast do stock prices adjust to market efficiency? Evidence from a detrended fluctuation analysis", Physica A: Statistical Mechanics and its Applications, 2013, Vol.392, pp 1631-1637, DOI: 10.1016/j.physa.2012.11.038

[10] Lo, Andrew W, 2004, "The Adaptive Markets Hypothesis", The Journal of Portfolio Management, 2005, Vol. 7, pp 15-29, DOI: 10.3905/jpm.2004.442611

[11] Yu, Hao, Nartea, Gilbert V. Gan, Christopher Yao, Lee J. 2013 "Predictive ability and profitability of simple technical trading rules: Recent evidence from Southeast Asian stock markets", 2013, International Review of Economics and Finance, Vol.25, pp:356-371, DOI:10.1016/j.iref.2012.07.016

[12] Morta, Rena, Morta, Shalini, Dadios, Elmer, 2015, "Modeling Market Data and Technical Indicators with Fuzzy Logic to develop a Trading Algorithm for an Individual Stock in the Philippines", 9-12 Decemper 2015, Philippines. IEEE International Conference Humanoid, Nanotechnology, Information Technology The Institute of Electrical and Electronics Engineers Inc. (IEEE) - Philippine Section Communication and Control, Environment and Management (HNICEM), pp:9-12.

[13] Y. Hu, K. Liu, X. Zhang, L. Su, E. Ngai, and M. Liu, 2015, "Application of evolutionary computation for rule discovery in stock algorithmic trading," Applied Soft Computing Journal., vol. 36, no. C, pp. 534–551, Nov. 2015. DOI: 10.1016/j.asoc.2015.07.008.

[14] Vijay Gaikwad, P. D. Y Patil, and P. Patil, January 2014 "Text Mining Methods and Techniques," Int. J. Comput. Appl., vol. 85, no. 17, pp. 42–45, 2014.

[15] F. Beil, M. Ester, and X. Xu, 2002 "Frequent term-based text clustering," Proc. eighth ACM SIGKDD, Edmonton, Alberta, Canada,2002, KDD, Int. Conf. Knowl. Discov. data Min. KDD, pp. 436–442.

[16] S. Chakrabarti, 2003, "Mining The Web Doscovering Knowledge from Hypertext Data",University of Illinois, Urbana-Champaign. 2003. ISBN: 1558607544

[17] H. Ahonen, O. Heinonen, M. Klemettinen, and a. I. Verkamo, 1998,"Applying data mining techniques for descriptive phrase extraction in digital document collections," Proc. IEEE Int. Forum Res. Technol. Adv. Digit. Libr. -ADL'98-, pp. 2–11, Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=670374,[Accessed 12 June'16] DOI:10.1109/ADL.1998.670374

[18] S. Shehata, F. Karray, and M. Kamel, 2007, "A concept-based model for enhancing text categorization," Proc. 13th ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD, Sun Jose, California, USA, August 12-15, 2007 p. 629-637.

[19] S.-T. W. S.-T. Wu, Y. L. Y. Li, Y. X. Y. Xu, B. P. B. Pham, and P. C. P. Chen, 2004, "Automatic Pattern-Taxonomy Extraction for Web Mining," IEEE/WIC/ACM Int. Conf. Web Intell., 2004,

[20] G. a. Miller, 1995, "WordNet: a lexical database for English," Commun. ACM, 1995, vol. 38, no. 11, pp. 39–41, DOI: 10.1145/219717.219748, PMID: 17081734

[21] C. S. Oyland, 2015 "Interday news-based prediction of stock prices and trading volume," Department of Applied Mechanis, Deficon of Vehicle Engineering and Autonomous System, Ahalmers University of Technology, Sweden 2015, vol. M, 2015. ISSN: 1652-8557.

[22] Uğuz, Harun, 2011, "A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm" Knowledge-Based Systems, 2011, Vol.24, pp: 1024-1032, DOI: 10.1016/j.knosys.2011.04.014.

[23] Y. Yang, 1995, Noise reduction in a statistical approach to text categorization, in: Proceedings of the ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'95), 1995, pp. 256–263.

[24] Mladenić, Dunja, Grobelnik, Marko, 2003, "Feature selection on hierarchy of web documents", Decision Support Systems, 2003, Vol.35, pp. 45-87, DOI: 10.1016/S0167-9236(02)00097-0.

[25] Li, T. Zhang, C. Ogihara, M. 2004, "A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression" Bioinformatics, 2004, Vol. 20, pp. 2429-2437, DOI: 10.1093/bioinformatics/bth267, PMID: 15087314

[26] Liu, Huawen, Sun, Jigui, Liu, Lei, Zhang, Huijie, 2009, "Feature selection with dynamic mutual information" Pattern Recognition 2009, Vol.42, pp: 1330-1339. DOI:10.1016/j.patcog.2008.10.028, PMID: 1443.

[27] Chen, Yao Tsung, Chen, Meng Chang, 2011, "Using chi-square statistics to measure similarities for text categorization" Expert Systems with Applications 2011, Vol. 38, pp: 3085-3090, DOI: 10.1016/j.eswa.2010.08.100.

[28] Saeys, Yvan, Inza, Iñaki, Larrañaga, Pedro, 2007, "A review of feature selection techniques in Bioinformatics" Bioinformatics, 2007, Vo. 23, pp: 2507-2517, DOI:10.1093/bioinformatics/btm344, PMID: 17720704.

[29] Shang, Wenqian, Huang, Houkuan, Zhu, Haibin, Lin, Yongmin, Qu, Youli, Wang, Zhihai, "A novel feature selection algorithm for text categorization", Expert Systems with Applications 2007, Vol. 33, pp: 1-5, DOI: 10.1016/j.eswa.2006.04.001.

[30] M. Behrouzian, Said M.H, Aref Sayahi, 2014, "Feature Selection Techniques for Text Classification", International Journal of Computer Science and Network Solutions. 2014, Vol:2, pp: 90-94, Available: http://www.ijcsns.com.

[31] Wu, Ke, Lu, Bao-Liang, Uchiyama, Masao, Isahara, Hitoshi, 2007, "A probabilistic approach to feature selection for multi-class text categorization", 2007, Advances in Neural Networks–ISNN 2007, 2007, pp: 1310-1317

[32] S.S.R. Mengle, N. Goharian, 2013, "Ambiguity measure feature-selection algorithm, Journal of the American Society for Information Science and Technology", 2013, Vol. 14,pp: 90-103, 60 (5) (2009) 1037–1050. DOI: 10.1002/asi, PMID: 10.1002/asi.

[33] J. Chen, H. Huang, S. Tian, Y. Qu, 2009, "Feature selection for text classification with Naive Bayes", Expert Systems with Applications, 2009, Vol.36, pp:5432–5435. DOI:10.1016/j.eswa.2008.06.054.

[34] Shang, Wenqian Huang, Houkuan Zhu, Haibin, Lin Yongmin Qu, Youli Wang, Zhihai, 2007, "A novel feature selection algorithm for text categorization", Expert Syst. Appl. 2007, Vol. 33, pp: 1-5, DOI: http://dx.doi.org/10.1016/j.eswa.2006.04.001

[35] Uysal, Alper Kursat Gunal, Serkan, 2014 "The impact of preprocessing on text classification", Information Processing and Management, 2014, Vol.50, pp: 104-112, DOI: 10.1016/j.ipm.2013.08.006, Available: http://dx.doi.org/10.1016/j.ipm.2013.08.006

[36] Willett, Peter, 2006, "The porter Stemming Algorithm: Then and Now", Electronic Library and information Systems, 2006, Vol.40, pp: 219-223, DOI: 10.1145/1645953.1646241, PMID: 3800

[37] K.-G. Aase and P. Öztürk, 2011, "Text Mining of News Articles for Stock Price Predictions," Dep. Comput. Inf. Sci., 2011, vol. Msc., no. June, p. 82.

[38] Navale, Prof G S Dudhwala, Nishant Jadhav, Kunal Gabda, Pawan Vihangam, Brij Kishor,2016, "Prediction of Stock Market Using Data Mining and Artificial Intelligence", Vol.134, pp. 6539-6544, DOI: 10.4010/2016.1571

[39] Baharudin, BaharumLee, Lam Hong, Khan, Khairullah. February 2010, "A Review of Machine Learning Algorithms for Text-Documents Classification", Journal of Advances in Information Technology, Vol.1, pp. 4-20, DOI: 10.4304/jait.1.1.4-20

[40] Rohit Choudhry and Kumkum Garg. March 29, 2008, "A Hybrid Machine Learning System for Stock Market Forecasting", World Academy of Science, Engineering and Technology. Vol.2, pp 315-318.

[41] Ding, XiaoZhang, YueLiu, TingDuan, Junwen, 2015, "Deep Learning for Event-Driven Stock Prediction", Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence. (IJCAI), China, 2015, pp: 2327-2333

[42] M. Shrabanti and P. Anita, 2015, "New approach of text mining in r," GESJ: Computer Science and Telecommunications 2015 , vol. 1, no. 1(45), pp. 30–36, 2015.