IJABER, Vol. 14, No. 5, (2016): 3059-3069

ANALYSIS OF THE JAPANESE CENTRAL BANK MONTHLY REPORTS AND NIKKEI 225 INDEX MONTHLY FOR FUTURE PREDICTION

David Ramamonjisoa* and Eichiro Moriya*

Abstract: *In this paper, we present the result of our experiment on finding the correlation between the monthly report issued by the Central Bank of Japan and the Nikkei 225 index monthly summary. We collected all monthly reports available online from December 1997 (around 30 years) and extracted several features concerning the content (term frequency, specific economics and financial terms, n-grams, …). For the Nikkei index, we put the data as monthly moving average and categorize each month into two classes (positive or negative) for moving up or down. We train a classifier such as Support Vector Machine (SVM) to assign the reports into Nikkei index class (0 for negative and 1 for positive). Then we use the classification model to predict the Nikkei index class giving a monthly report. We achieve 60% accuracy for 12 months period of test reports on Japanese language reports and above 90% on training data on English report texts. Our first conclusion is that the central bank reports analysis can give some hints to Nikkei index trends. The reports should be added to other indicators in global economy to further enhance the results.*

Keywords: *Japanese Central Bank reports, Nikkei 225 index, Prediction, Analysis.*

I. INTRODUCTION

Artificial Intelligence in finance and economy has been used since the birth of the computer. Professor Herbert Simon has developed the general problem solver in collaboration of the professor Allen Newell and attributed the Nobel memorial prize in economics during the 80s for his pioneering research into the decision-making process within economic organizations.

The financial market and world economy however are an open world system that depends on the human decisions as uncertain and unpredictable and the world situation where prices fluctuate and are unstable. This area grew exponentially during the last 4 decades. It becomes very complex system and difficult global events. We can observe the current event on the oil price bubble burst, emerging countries difficulty to maintain growth and the fight of deflation of the developed countries, the 2008 global financial crisis and the past crashes on stock market or defaulted countries.

* Faculty of Software and Information Science, Iwate Prefectural University, Japan, *E-mail: david@iwate-pu.ac.jp, g031k153@s.iwate-pu.ac.jp*

Each country has to deal with their economy recovery and stability to maintain growth, goals and policies (Ahmad & Mazlan, 2015; Oudata, Ahmadb & Yazisc, 2015). The study in this paper shows a specific situation on the Japanese financial system and economy.

The Japanese economy is now the third world economy power. The country has fighting the deflation since the stock market bubble during the 80s and 90s. This economy depended heavily on the imports and exports of goods and the corporation investments. The government and central bank policies have facilitated this economy to reach the second world economy power during the 80s and 90s despite the country size and geographical location sensitive to natural disasters such as earthquakes and volcanoes. Among those facilitations was for example the 'Plaza accord'[1] during the 80s or the collaborative interventions of the central banks during 90s, 2008 or 2011. The stock market reacted to those actions and returned to the stable prices with the effect on real world economy on increasing employment rate and GDP growth.

The investors in finance rely on the news and information available on the world wide websites. The technical analysis on the financial data based on numeric information of past performance is not enough to forecast the asset and equities future price. The fundamental analysis on the text data (e.g. news, websites, investor analysis blogs or tweets) can help to enhance the performance of the forecaster and make the decisions systematic to be learned by machine programs.

Several programs have been developed and used to predict the future price of the equities in the market based on both technical and fundamental analysis. During the 80s and 90s, expert systems were used to analyze financial risks and predict a specific stock. The knowledge inputted manually by the knowledge engineer by implementing the information with rule based systems. The knowledge base did not scale up to the prediction. Some study of the market based on the mathematical models have emerged and seemed to work. MIT professor James Simons's company Renaissance Technologies has developed a tool for managing funds in the financial market from 1994 with high returns without loss.[2]

In the recent news, the Japanese programs robot developed at the Central Bank could predict the monthly Nikkei 225 for 68% accuracy for 47 months.[3]

The text mining and machine learning are the basic tools for the fundamental analysis by selecting and extracting the patterns to be classified. The text mining concerns to extract the knowledge or valuable information from the vast amount of text data sources. The machine learning technique is used to build a model and classify the patterns (combination of many features) to a certain class of data as for example the Nikkei 225 monthly return to positive or negative. The use of the text mining in finance and mainly in the Central Bank area is surveyed in the working papers (Hendry and Madeley, 2010), (Hendry, 2012) and handbook (Bholat et al., 2015). The common techniques are the vector space models, the latent semantic analysis, topic models, and unsupervised or supervised machine learning. Before using those techniques, it is necessary to prepare the texts for analysis. This preparation is often time consuming because documents are published in different formats and should be converted into text files. Text files may need further processing to be ready for the input of the text mining programs. This text processing is referred to as the natural language processing (NLP) including the sentence segmentation, parsing, tokenization, bag-of-word representations and dimension reduction of the bag-of-words such as lemmatization, stemming, stripping out punctuation and rare words, and token normalization like a case folding. The NLP can be further extended to semantic analysis as described in (Acosta and Meade, 2015) to study every word in the Federal Open Market Committee's (FOMC) post-meeting statements by uncovering hidden truths about complex communication. Other techniques in text mining such as sentiment analysis and link-mining in financial texts are also important (Moniz and de Jong, 2014), (Baerg and Lowe, 2015). In (Moniz and de Jong, 2014), the 'Monetary Policy Committee Minutes' of the Bank of England were classified into one of the four economic aspects classes by using a multinomial Naïve Bayes model then used to predict the financial market interest rates. In (Baerg et Lowe, 2015), the n-grams topic modeling is combined with scaling methods to estimate the central bankers' preferences during the voting based on their FOMC speaker-meeting texts data.

In this paper, we present our experiment to build a model for analyzing the text data and to forecast the monthly outcome of Nikkei 225 based on the Central Bank monthly reports publicly available on the Bank of Japan website. The section 2 and 3 describe the Nikkei 225 and the monthly report issued by the Central Bank of Japan. Section 4 presents the features to build the classifier model. The section 5 deals with the experiments and results. Section 6 shows some related researches and draws conclusion and future work.

II. NIKKEI 225

The Japanese Stock Market is the world's second largest because of the government policies of liberalizing capital transactions with other countries and democratization of stock ownership. The 225 companies listed on the First Section of the Tokyo Stock Exchange are forming the Japanese Stock Market. Those 225 listed companies stocks are updated in real-time and are published by the Nihon Keizai Shinbun. They are called Nikkei 225 Average or for short Nikkei225. The Nikkei225 is an economic indicator for the country and surveyed by the government and the central bank of Japan for their monetary policy, financial stability and growth target. The Tokyo Stock Exchange has another indicator called TOPIX (Tokyo stock Price Index) representing all listed companies and the entire market performance but it is hard to follow. The Nikkei225 is the equivalent of the Dow Jones Industrial Average in

New York Stock Exchange. The chart in Figure 1 shows the yearly returns of the Nikkei225 from 1997 to 2015. We can observe the loss effect -41% caused by the financial crisis during 2008, -18% loss on March 2011 disaster and the high return +57% during the monetary easing by the central bank of Japan during 2013 until the end of 2015.

Figure 1: Annual returns of the Nikkei 225 from 1997 to 2015.

The question is now concerning the current policy from the Central Bank of Japan described in the following section. Will the negative interest rates and target 2 percent growth adopted by the central bank make the positive or negative return of the Nikkei225 for the year 2016?

III. MONTHLY REPORT OF RECENT ECONOMIC AND FINANCIAL DEVELOPMENTS (MREFD)

The Bank of Japan (central bank of Japan) releases a summary of economic and financial developments, which form the basis of the decision on the guideline for money market operations, in the Monthly Report of Recent Economic and Financial Developments (MREFD).

The Bank of Japan homepage[4] outlines the ongoing policy activities to be decided and executed by the committee members during their meeting according to their schedule. Here below is the list of their recent activities.

The Bank's Market Operations

Guideline for Money Market Operations

The Bank of Japan will conduct money market operations so that the monetary base will increase at an annual pace of about 80 trillion yen. (January 29, 2016)

Monetary Base

355,310 billion yen (February 15, 2016)

Current Account Balances at the Bank of Japan

256,460 billion yen (February 15, 2016)

Basic Loan Rate

0.3% (since December 19, 2008)

Interest Rate Applied to the Complementary Deposit Facility

minus 0.1% (applied to the Policy-Rate Balance, since February 16, 2016)

Uncollateralized Overnight Call Rate (average)

0.074% (February 15, 2016)

Bank of Japan Operations Next Monetary Policy Meeting Date

March 14 and 15, 2016

These activities can be summarized with the following two decisions as of February 2016:

- ***"Price Stability Target" of 2 Percent***
- ***"Quantitative and Qualitative Monetary Easing with a Negative Interest Rate"***

The underlying objective of financial analysis is the comparative measurement of risk and return useful for making investment, credit or regulatory decisions. The two years evaluation of the QQE (Quantitative and Qualitative monetary Easing) was reported in the Bank of Japan review in May 2015 (Monetary affairs department, 2015).

IV. FEATURE SELECTION AND EXTRACTION FROM THE TEXT DATA

We collected the published MREFD publicly available from January 2000 to December 2015 from the Bank of Japan website. The published documents are in

PDF format so they need to be processed as text before analyzing the features. Feature selection is necessary to make large problems computationally efficient—conserving computation, storage and resources for the training phase and for every future use of the classifier. Further, well-chosen features can improve classification accuracy substantially, or equivalently, reduce the amount of training data needed to obtain a desired level of performance.

The common filters applied to the text are the elimination of rare words from the whole dataset, the removing of the common words such as 'a' and 'the' as to not be discriminating for any particular class. English stop-words in any natural language processing tool contain those common words. We used those feature filters to our text data.

We converted the PDF files into text files with a plugin tool in Python scripts. For the reports in Japanese, we use a natural language processing to parse and tokenize the text to build the bag of words model and for the English reports, texts extracted from the PDF converter are directly used in the machine learning tool described in (Mayfield and Penstein Rosé, 2013). To select the feature terms, we use unigrams, bigrams of words and the term frequency. The feature extraction is depicted in Figure 1 for the English reports. For the Japanese reports, we take only the 100 terms within threshold in the TFIDF values as the representative of the each report.

The overall feature selection procedure is to score each potential feature according to a particular feature selection metric, and then take the best k features. Scoring involves counting the occurrences of a feature in training positive- and negative-class training examples separately, and then computing a function of these. The left side of Figure 2 shows the different metrics such as total hits, target hits, precision/recall/f-score, kappa, and correlation. The results are in the features table in the left in Figure 2. The correlation measure is irrelevant for our problem as the class is nominal.

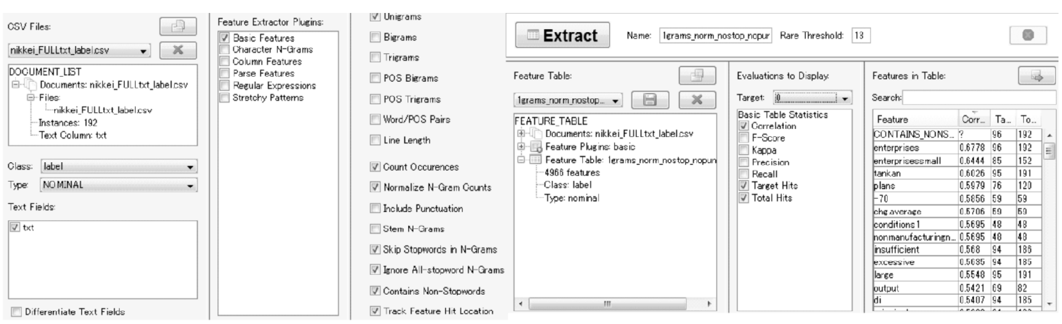

Figure 2: Basic Feature extraction based on count occurrences, skipping stopwords, unigrams and extracted features with rare threshold word counts cut equal to 13. The target hits for the class 0 (negative class for Nikkei225) are tankan, non-manufacturing data, enterprise and so on

V. MACHINE LEARNING FOR PREDICTION

A support vector machine (SVM) is conceived for binary classifier and part of the statistical learning theory by Vladimir Vapnik (Vapnik,1998). The idea is to separate two classes by calculating the maximum margin hyperplane between the training examples or instances described in (Scholkopf and Smola, 2002). SVM on text categorization has been developed and explained by Professor Thorsten Joachims (1998; 2002) in his book and research papers. In the book of professor Thorsten Joachims, a comparison of four traditional methods (Naïve Bayes, Rocchio algorithm, k-nearest neighbors, and decision trees) for learning the categorization function is presented. SVM outperforms all of them. The use of the SVM in the stock market and foreign exchange rate prediction according to economic and financial news is studied in the following research papers. Huang et al. have used SVM to forecast the movement of NIKKEI225 weekly price (Huang W., Nakamori Y., Wang S.-Y., 2005). Cao *et al.* have made a survey on SVM applied to financial research by insisting the importance of the SVM kernel and parameters (Cao, Zhan, Wu, 2009). Hagenau *et al.* have also applied SVM to german stock market prediction based on news data and obtained 65% accuracy (Hagenau M., Liebmann M., Neumann D., 2013). Therefore, we have chosen SVMs to conduct our experiments.

We train a support vector machine (SVM) classifier to learn the reports according to the Nikkei225 monthly outcome. Each outcome is categorized into 0 for the Nikkei225 minus value and 1 for positive value. The classifier is making a binary choice according the report. The building of the model is shown in Figure 3. The extracted features in the previous section are used to train the classifier. With the randomized 10 folds cross validation, we obtain a 92.71 percent of accuracy. The kappa value is 0.8542. The class 1 (Nikkei ending monthly positive) is classified 100% and the negative class 0 has 14 misclassified instances as shown in the confusion matrix. But when we reduce the training data from 2000 to 2013 (168 instances), this accuracy is decreased to 89.3%, a kappa value to 0.786 and both classes have misclassified instances after the training period.

The models comparison shown in Figure 4 gave us the best classifier with 97.92% accuracy and kappa measure to .9583 with 192 instances. This model uses a one-gram feature with word frequency cut equal to 2 and all the available options of the tool. However we selected only the top 5000 features among 13000 features extracted.

The classifier accuracy is decreased when we extended the features to contain one-gram and bigrams with a minimum cut to 1 for occurrence count and stop-words removing. The number of features in this case has doubled and was around 23000 terms. This experiment is described in Figure 4 as the competing model in the right side.

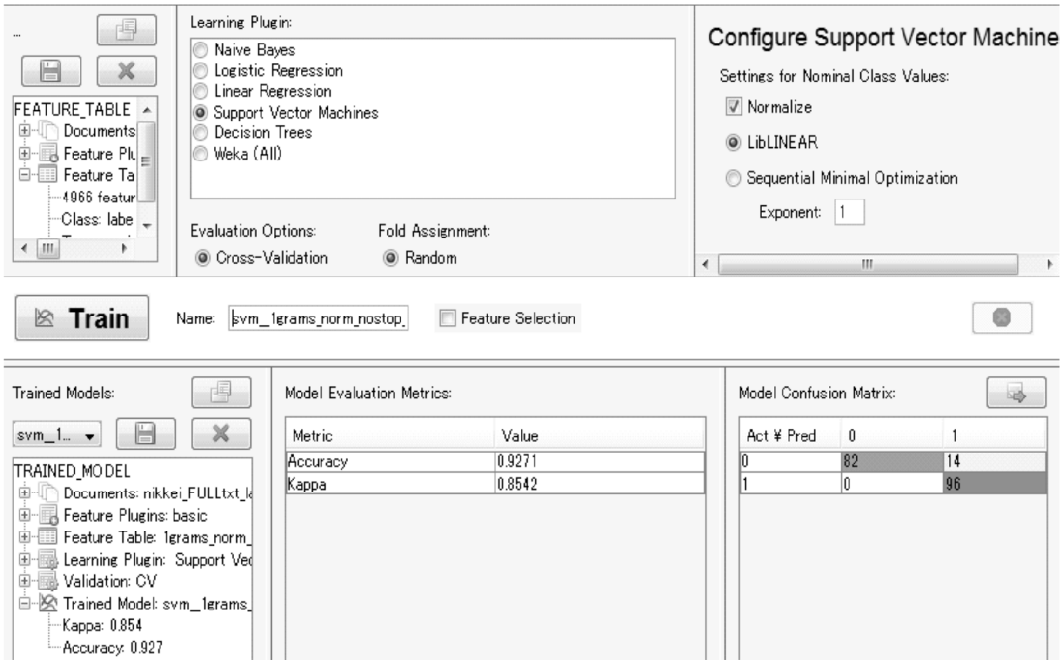

Figure 3: Selection of the classifier, building models and evaluation

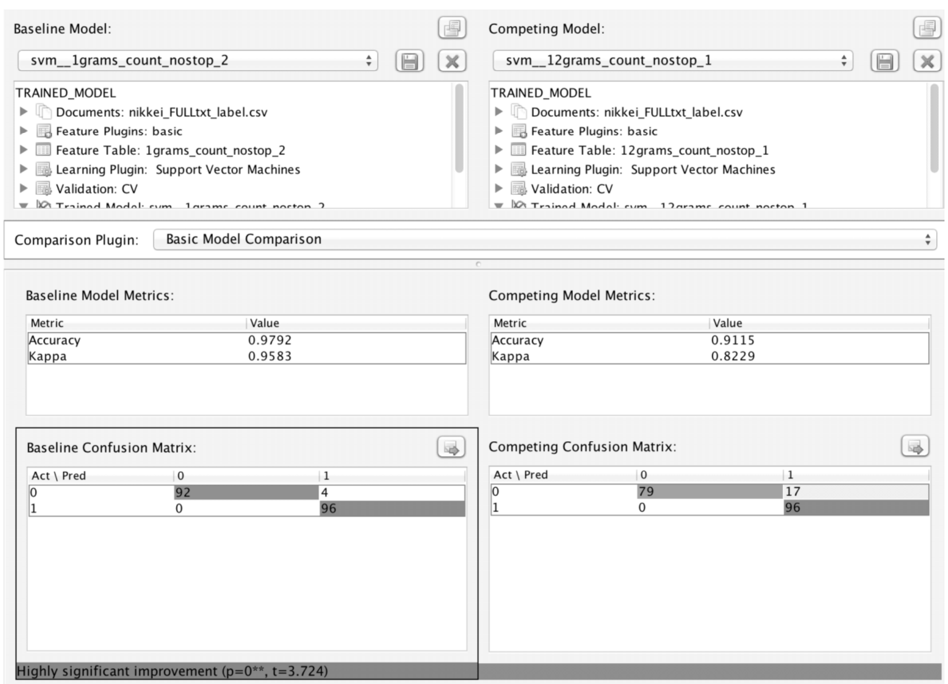

Figure 4: Comparing models by modifying features selection and classifier parameters

The SVMs with Japanese reports data had performed at best 60% accuracy that the detail is not described in this paper. The limited 100 features as input to the method is not enough in comparing to the ones used in the English text data.

VI. DISCUSSIONS

We conducted an experiment for building a classifier for categorizing the Bank of Japan monthly reports and the result of the Nikkei 225.

The results of the experiments showed that even though the reports are classified positive (class 1) as similar to the positive Nikkei monthly returns, the real returns of the Nikkei monthly are negative in all misclassified cases. This can be interpreted as the pessimistic market can be predicted for 92% at best case. It also means that the bear trends as well as bubble burst are difficult to forecast according to the data within the reports.

The promise of the Bank of Japan in 2013 to achieve a 2% growth within two years was not fulfilled whatever the causes as national or international factor (e.g. oil bubble burst in end of 2015 or world economy slow growth). This is an example of the difficulty in the long term prediction but that promise has made a good performance of the Nikkei225 during two years to double its price from 10000 in 2013 to 20000 in August 2015. The sharp decrease of the Nikkei225 at the beginning of the year 2016 has created volatilities in the financial and foreign exchange market. This situation has triggered the Bank of Japan to introduce the negative interest rate taking into effect from February 16th 2016. This policy is a surprise that the consequence for the short term is unexpected and creating more volatile asset prices.

As stated in the introduction, some forecaster algorithms based on artificial intelligence performed well and used to predict the Nikkei 225. But it is said that the system was tested only during the period of 2012 to 2015 and on mostly the positive classes (class 1)[5].

Hedge funds and institutional investors have built algorithmic trading strategies but they are very expensive that individual investors could not afford. Our research can be a start to provide some open system for everyone interested in this area.

Some research has been also investigated to study the correlation of the Central Banks monetary policy and the foreign exchange rates as described in the paper by Dominguez (Dominguez, 2006) that is also an interesting future extension.

Other researchers have used the extended SVM with other economic predictors such as GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model dealing with volatilities and ARMA (Autoregressive Moving Average) on time series data (Krishnalal, Rengarajan, Srinivasagan, 2010), (Kumar, Kumar, Prasad, 2012), (Mantri, 2013), (Leung, MacKinnon, Wang, 2014), (Seker, Mert, Al-Naami, Ozalp, Ayan, 2014).

VII. CONCLUSIONS

In this paper, we presented the analysis of the Japan Central Bank publicly available monthly reports for 19 years and the Nikkei 225 index monthly earning. We used a machine learning technique based on the Support Vector Machines to learn classifying the reports according to the Nikkei 225 index. We achieved a very good accuracy for the English reports by iteratively modifying the feature extraction criteria and selecting a part according to the Chi-Square measures for the SVM classifier. For the future, the use of several machine learning techniques combined to improve the accuracy will be investigated such as in (Patel J., Shah S., Thakkar P., Kotecha K., 2015).

Notes

1. *http://www.cfr.org/trade/plaza-accord/p19817*
2. *http://www.wsj.com/articles/SB112018150042875023*
3. *www.bloomberg.com/news/articles/2016-02-17/the-japanese-quant-who-made-a-robot-for-calling-the-stock-market*
4. *http://www.boj.or.jp/en/*
5. *http://www.valuewalk.com/2016/02/japanese-quant-bloomberg/*

References

Ahmad, N., & Mazlan, N. F. (2015). Banking fragility sector index and determinants: A comparison between local-based and foreign-based Commercial Banks in Malaysia. International Journal of Business and Administrative Studies, 1(1), 5-17

Acosta M. and Meade E. (2015). Hanging on every word: Semantic analysis of the FOMC's postmeeting statement, In FEDS Notes, USA, September 2015.

Baerg N. and Lowe W. (2015). Estimating Central Bank Preferences Combining Topic and Scaling Methods. In Munich Personal Research Archives (MPRA), paper No.61534, University Library of Munich, Germany, January 2015.

Bholat D., Hansen S., Santos P., Schonhard-Bailey C. (2015). Text mining for central banks. Cental for Central Banking Studies (CCBS) Handbook No.33. ISSN: 1756-7270 (Online).

Cao B., Zhan D., Wu X. (2009). Application of SVM in Financial Research. In proc. of 2009 International Joint Conference on Computational Sciences and Optimization, pp. 507-510, IEEE publisher, Sanya Hainan, April 24-26 2009.

Dominguez K.M.E. (2006). When do central bank interventions influence intra-daily and longer-term exchange rate movements? In Journal of International Money and Finance 25 (2006), 1051-11071.

Hagenau M., Liebmann M., Neumann D. (2013). Automated news reading: Stock price prediction based on financial news using context-capturing features. In Decision Support Systems, Volume 55, Issue 3, Pages 685-697, Elsevier Publisher, 2013.

Hendry S. (2012). "Central Bank Communication or the Media's Interpretation: What Moves Markets?," Bank of Canada Working Paper, September 2012.

Hendry S. and Madeley A. (2010). Text Mining and the Information content of Bank of Canada Communications, Working paper – Bank of Canada, November 2010.

Huang W., Nakamori Y., Wang S.-Y. (2005). Forecasting stock market movement direction with support vector machine. In Computers & Operations Research, Volume 32, Issue 10, Pages 2513-2522, Elsevier Publisher, 2005.

Joachims T. (1998). Text Categorization with Support Vector Machines: Learning with Many Relevant Features. In Proceedings of the Tenth European Conference on Machine Learning (ECML), pages 137-142, Berlin, Germany, 1998.

Joachims T. (2002). *Learning to Classify Text using Support Vector Machines*, Kluwer/Springer, 2002.

Krishnalal G., Rengarajan B.S., Srinivasagan K.G. (2010). A New Text Mining Approach Based on HMM-SVM for Web News Classification. In International Journal of Computer Applications Volume 1, No. 19, 2010.

Kumar R. B., Kumar B. S., Prasad C.S.S. (2012). Financial News Classification using SVM. In International Journal of Scientific and Research Publications, Volume 2, Issue 3, 2012.

Leung C.K.S., MacKinnon R.K., Wang Y. (2014). A Machine Learning Approach for Stock Price Prediction. In proc. of the 18th International Database Engineering & Applications Symposium IDEAS'14, pp. 274-277, ACM New York, USA, 2014.

Mayfield E. and Penstein Rosé C. (2013). Open Source Machine Learning for Text. Handbook of automated essay evaluation: Current applications and new directions.

Mantri J.K. (2013). Comparison between SVM and MLP in Predicting Stock Index Trends. In International Journal of Science and Modern Engineering (IJISME), Volume-1, Issue-9, August 2013.

Monetary affairs department, 2015. Quantitative and Qualitative Monetary Easing: Assessment of Its Effects in the Two Years since Its Introduction. Bank of Japan Review 2015-E-3, Bank of Japan, May 2015.

Moniz A. and de Jong F. (2014). Predicting the impact of central bank communications on financial market investors' interest rate expectations. In Proceedings of the European Semantic Web Conference (ESWC) 2014 Satellite Events, Greece, May 25–29, 2014. Revised Selected Papers, Lecture Notes in Computer Science 8798, Presutti V. et al. (Eds.).

Patel J., Shah S., Thakkar P., Kotecha K. (2015). Predicting stock market index using fusion of machine learning techniques. In Expert Systems with Applications, Volume 42, Issue 4, Pages 2162-2172, Elsevier Publisher, 2015.

Scholkopf B. and Smola A. (2002). *Learning with Kernels*. MIT Press, Cambridge.

Seker S.E., Mert C., Al-Naami K., Ozalp N., Ayan U. (2014). Time Series Analysis on Stock Market for Text Mining Correlation of Economy News. In International Journal of Social Sciences and Humanity Studies, Vol.6, No.1, 2014.

Vapnik V. (1998). *The Nature of Statistical Learning Theory*. Wiley, New York.

Oudata, M. S., Ahmadb, N., & Yazisc, M. (2016). The impact of macroeconomic variables and global events on banking system: Evidence from Jordanian banking system. *Journal of Administrative and Business Studies.*