Frontiers in Economics and Management Volume 2 Issue 7, 2021
ISSN: 2692-7608 DOI: 10.6981/FEM.202107_2(7).0039

Leveraging Financial News Analysis to Predict Stock Price Movement

Mingchen Xu[1,*], Yuxuan Wang[2], Yiran Huang[3], Mulei Xu[4]

[1]School of Journalism and Communication, Shanghai International Studies University, Shanghai 201600, China;

[2]Department of Mathematics & Statistics, Mount Holyoke College, South Hadley, MA 01075, United States;

[3]Department of Mathematics, Nankai University, Tianjin 300000, China;

[4]Obridge Academy, Chongqing 400020, China.

*Corresponding author. Email: 0171127040@shisu.edu.cn

Abstract

Stock market predictions have been prominent among scholars. Sentiment analysis applying financial news articles for predicting stock has also grown in popularity over the past few decades. Our research indicates the impact of sentiment analysis of financial news on forecasting asset prices. In this research, sentiment is culled from financial news of companies over the past three years. Subsequently the asset prices can be predicted using stock market data, alongside sentiment analysis data with several machine learning algorithms. The results show a more convincing level of precision in the aggregation of both stock market data and sentiment analysis data.

Keywords

Stock Market Predictions; Sentiment Analysis; Natural Language Processing; Machine Learning.

1. Introduction

1.1. Stock Market Predictions

Stock market prediction has always been the subject of emphasis in the financial world. Due to the growing interest in investing in stock markets and the objective of maximizing returns on investments in asset markets and minimizing risk, researches of stock predictions become highly prominent for the recent decades [1-5].

Early in the 1960s, economists had diverse perceptions about stock market predictions. Efficient Markets Hypothesis [6] and Random Walk Theory [7] both pointed out that stock market prediction is complicated and almost impossible. Nevertheless, other researchers have proved them obsolete by performing various forms of predictions [3, 5, 8-10].

Two main schools of thought in predicting stock markets include fundamental and technical analysis. Fundamental analysis is based on unstructured data, for instance, financial status of a company, climatic circumstances such as natural disasters [2,3,11]. Technical analysts attempt to predict asset prices through the use of historical and time series data [4,11,12]. It was used by Mills in 1997 in a study on the predictable power of several simple technical trading rules in London.

Furthermore, with the development of machine learning and Natural Language Processing (NLP), increasing number of machine learning algorithms are used for predicting the stock markets, such as Latent Dirichlet Allocation (LDA) based topic extraction mechanism, Support

ISSN: 2692-7608 DOI: 10.6981/FEM.202107_2(7).0039

Vector Machine (SVM), Neural Networks, Random Forest, Regression Trees and so on [13-17]. Among these, neural networks and SVM were discovered to be especially popular for predicting the stock market [4, 14, 17-19]. Examples such as Yao and his team's paper that presented a study of artificial neural nets for use in stock index forecasting in 1999 [4], the 2017 study of Nelson et al on the use of LSTM networks for predicting future pattern in stock prices on the basis of price history [19] and the study of Shen et al. (2012) using the temporary correlation among world stock markets, as well as various financial products for predicting the next-day stock trend with the aid of SVM [16].

1.2. Financial News Sentiment Analysis

Text sentiment analysis is extensively applied in a range of business applications and has been an effective tool used in financial research [1, 20-22]. Due to the advancement of communication technologies, large new data sources on our information consumption are produced from our interactions on the Internet, and decisions to purchase or sell stocks are influenced through various sources of information in the environment of trading [23-25].

Financial news is among the most essential information sources influencing the investors' decisions. Researches have evaluated the impact of sentiments extracted from financial news on stock predictions. For example, the study of Kim et al., (2014) revealed that news tends to be used for both upward and downward movements in stock prices [26]. Over the past few decades, there has been more research on the use of financial news to predict asset prices. [1,17,27,28]. In 2005, Yoo et al., showed in their work that integrating event information from news with prediction model played an extremely important roles for more precise prediction [25]. Tetlock (2007) reveals the impact of high media pessimism in daily content from a popular Wall Street Journal column on financial market returns [29]. Garcia (2013) assesses financial market sentiment from New York Times financial columns [30].

Meanwhile, a plethora of NLP applications and the availability of high-performance computing systems add to the opportunities to extract effective information from news articles [26-28]. In 2007, Zhai et al. proposed a system that integrated the information from both related news releases and technical indicators for enhancing the predictability of the daily stock price trends. The performance indicated that this system is capable of achieving higher accuracy and return than a single source system [5]. Mahajan et al. (2008) introduced a text mining system that analyzed news from the Indian stock market [31]. Another study of 9,211 news articles and 10,259,042 stock quotes over a five-week period was conducted in 2009 by Schumaker and Chen. They introduced the AZF in text system and reached a directional precision of 57.1% [3].

Text sentiment analysis is extensively applied in a range of business applications and has become an effective tool in economic and financial research [32]. Tetlock (2007) indicates the impact of high media pessimism in daily content from a popular Wall Street Journal column on financial market returns [29]. Bollen and Mao (2011) explored whether measurements of collective mood states derived from large-scale Twitter feeds share correlation with the value of the Dow Jones Industrial Average (DJIA) over time [33]. Garcia (2013) uses the New York Times financial column to evaluate the financial market sentiment [30]. Shapiro and Wilson (2019) assessed the objective function of the central bank on the basis of sentiment expressed in internal discussions of the Federal Open Market Committee [34]. Subsequently, Shapiro (2020) demonstrated state-of-the-art text sentiment analysis tools during the development of a new time-series measure of economic sentiment derived from economic and financial newspaper articles from January 1980 to April 2015 [32].

Certain scholars have already assessed the impact of sentiments extracted from financial news on the prediction of stock price. In Kim's research in 2014, the researchers showed that news' sentiment tends to be used in predicting stock price fluctuations, whether up or down [26]. Nguyen develops a model in 2015 to predict price movements of stock using sentiment on social

ISSN: 2692-7608 DOI: 10.6981/FEM.202107_2(7).0039

media. From the result it is shown that incorporating sentiment information from social media tends to assist in the improvement of stock forecasts [20]. Xu & Cohen (2018) presented a model for predicting stock price movement from tweets and historical stock prices with a new topic model TSLDA to capture the feature [22].

Our research will assess if sentiment extracted from a leading news agency can be utilized as a proxy for short-term asset price movement and consequently used in an automated trading strategy.

2. Data

This work uses financial news to predict the price movement of the asset. The work uses a range of data from Kaggle.com consisting of over 100,000 financial news articles as well as stock market data based on stocks featured in Yahoo Finance articles, including daily highest and least price. These stock prices and computed AR, BR and MA5 (Moving Average Five) are also interpolated. All of those data are used for predicting asset prices.

All the articles in the dataset are originally found on Reuters.com, prominent for breaking news concerning market and business. The articles run between March 2017 and May 2020 with 3,538 stocks listed. Each stock comprised of 30 articles on the average. Table 1 provides more information on the distribution of these articles by warehouse.

Table 1. Shows more details about the distribution of these articles for each stock.

Features	Num of Stock
Count	3538
Mean	30.77473
Std	96.51833
Min	1
Max	2340
25%	5
50%	10
75%	20

Three stocks with the most articles are presented in Table2. The first three include Boeing Co, Apple Inc., Facebook Inc., insinuating that the articles contained in this dataset are inclusive of companies' stocks from different aspects.

Table 2. Presents statistics for three stocks that have the most frequent appearance in Reuters' full datasets. For each stock, it belongs to a company in a specific area. It equally shows the number of news articles whereby the stock was mentioned.

Company	Stock	Num of News Articles
Boeing	BA	2340
Apple	AAPL	1879
Facebook	FB	1368

In the same vein, through the organization of the news articles, the distribution for each month is depicted in Figure 1 below. Among the 100,000+ articles, there are significantly more articles between June 2017 and June 2018. No considerable social changes have occurred during this period of time, thus the assumption is that the dataset contains less news articles for this time period.

The main components of each article were also presented in this dataset. The headers of these news articles describe shorter information about the stock markets, and the entire article has more detailed information on stock sentiment, each with its own column. The dataset likewise

ISSN: 2692-7608 DOI: 10.6981/FEM.202107_2(7).0039

contains the VADER polarity score (compound score) of the news headers processed through the use of NLTK VADER Sentiment Intensity Analyzer. Overall, the dataset includes the URL of the articles, the headers (both raw and processed header), the publication date and time, the full text of the news articles, as well as the VADER polarity scores of the news headers. Descriptions about each of these variables are presented in Table 3.

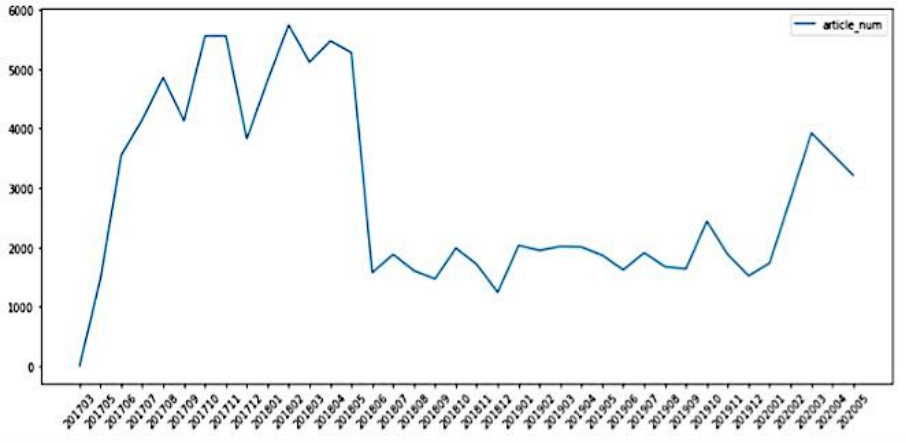

Figure 1. Depicts the distribution of total numbers of news articles over the time period spanning from March 2017 to May 2020.

Table 3. Presents the description of the columns of the dataset.

Columns of the Dataset	Descriptions
Header (raw)	The original headline of the news article, which captures the major idea of the article
URL	The address of the article on Reuters.com
Stock	The ticker of the stock this article is mentioned
Article's Publish Date	The exact date and time of the publication to Reuters.com of this article
Full Article Content	The whole original news article
Header (Processed)	The headline of the news article without tags (such as 'BRIEF')
Negative Sentiment	VADER (composite value) polarity scores in titles of news is a metric that computes the sum of all ratings in the lexicon, which are already normalized between -1 (most extreme negative) and +1 (most extreme positive). Positive sentiment: compound score >= 0.05; Neutral sentiment: -0.05 <= compound score < 0.05; Negative sentiment: compound score <= -0.05
Neutral Sentiment	
Positive Sentiment	

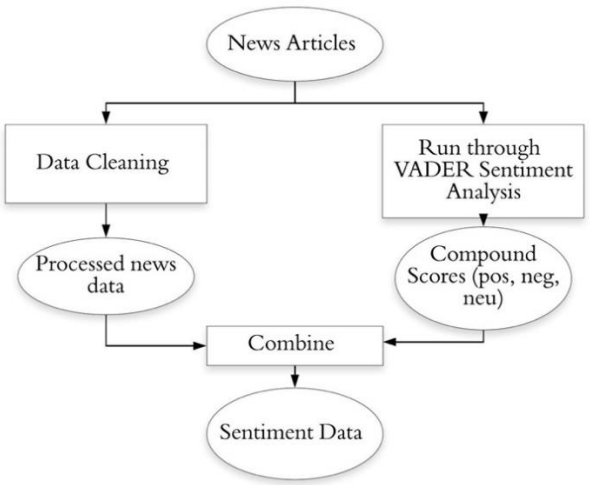

Figure 2. Depicts the process of the sentiment analysis system of this work.

ISSN: 2692-7608 DOI: 10.6981/FEM.202107_2(7).0039

3. Methodology

3.1. Sentiment Recognition

The general outline of our sentiment analysis system is revealed in figure 2. The first approach, this job, conducts a pre-process of all the data and marks every object that identifies the company (s). Financial news from American companies including Boeing, Apple and Facebook are chosen over period of approximately 3 years and stored the data in a database. The next step comprises of natural language processing to extract sentiment from unstructured news articles.

For the purpose of extracting the sentiment for each news article, VADER is applied in this work, which is a ready-made Python machine learning library for natural language processing, specifically suitable for reading the emotions expressed in media.

For sentiment computation, VADER pays special attention to the recognition of capital letters, but equally recognizes slang, exclamation marks and the most common emojis. The range of emotional scores is from extreme negative (-1) to extreme positive (+1), and neutral is displayed as being close to 0. The composite score is a standardized, weighted composite score that is calculated from the sum of the score of valence for each word in the lexicon and adjusted in accordance with the rules. News data that are missing is substituted with the average value of the sentiment function. Through this approach, the data can be integrated with stock data. The sentiment features described above are presented in Table 4 and Figure 3.

Table 4. News Sentiment Statistics. This table presents the percentile statistics of the useful metrics for sentiment measures in the data set provided by Reuters.com.

Percentile	Negative	Neutral	Positive	Compound
0%	0	0.726	0	-0.993
25%	0.026	0.881	0.052	0.036
75%	0.041	0.894	0.065	0.268
100%	0.051	0.909	0.072	0.656

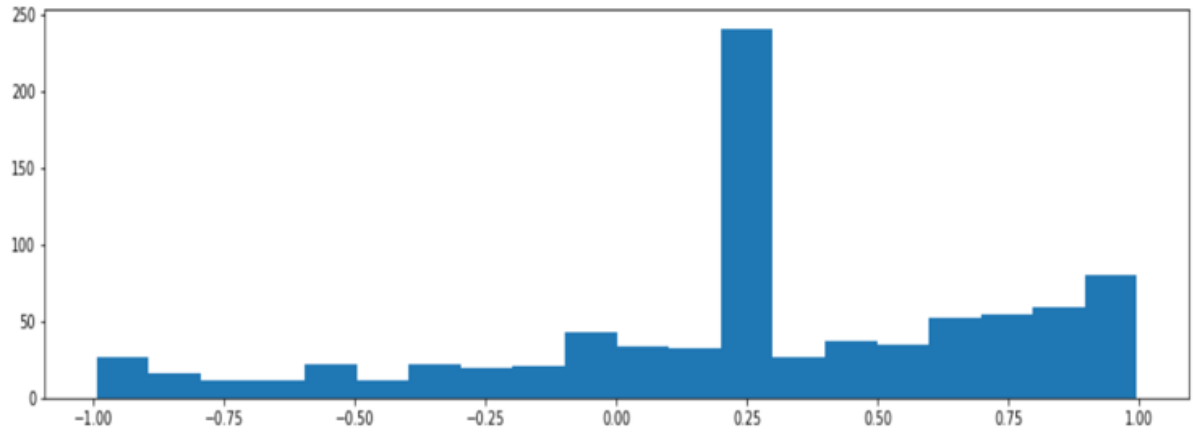

Figure 3. Presents how the 'Compound' features are distributed.

Figure 4 depicts stock data gathering and pre-processing. The historical stock price data obtained from Yahoo Finance is preprocessed for data interpolation.

To ensure compatibility with daily news, inventory data is interpolated over weekends. Subsequently the relevant indices MA5 (Moving Average 5), AR and BR are computed and prepared for the prediction.

ISSN: 2692-7608 DOI: 10.6981/FEM.202107_2(7).0039

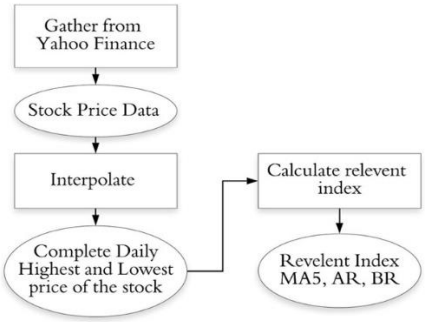

Figure 4. Presents data cleaning and pre-processing of the stock market data.

3.2. Forecasting

3.2.1. Machine Learning Models

The prediction for our research is the majority vote of the subsequent machine learning models: Logistics Regression, Gaussian Naïve Bayes, as well as Random Forest Regression.

Logistic Regression denotes a generalized linear regression model that fits a logistic function [35-38]. Unlike simple linear regressions, the generation of a linear equation is limited between 0 and 1, which makes classification with logistic regression better than with linear regressions [38]. Logistic regression is adopted by several scholars. It's not only a classification model, but also provides probabilities as outputs. Nevertheless, Logistic Regression tends to suffer from total separation. Assuming a feature exists that would perfectly separate the two classes, the logistic regression model can no longer be trained [35, 36, 38]. In this research, logistics regression is applied only once as it is a categorical algorithm.

The Naive Bayes classifier is in accordance with the Bayes' Theorem. It computes the class probabilities for each feature independently, which corresponds to a strong assumption of independence of the features [38-40]. Thus, it is crucial to perform a Principal Component Analysis (PCA) prior to applying Naive Bayes. PCA is a multivariate technology that assesses a table of data with significant proportion of inter-correlated quantitative dependent variables [41, 42]. It extracts essential information from a dataset and expresses using a set of new orthogonal variables known as principal components. Naive Bayes classifier works more suitably than other models when assumption of independent features holds true. However, this assumption is likewise the main limitation. In actual forecasts, it is complicated to obtain functions that are completely independent of each other [38, 43]. In this study, the Naive Gaussian Bayesian model was chosen to predict the daily closing price of the stock.

Random Forest denotes an ensemble learning method that constructs a collection of decision trees and subsequently aggregates the predictions of each tree for determining the final prediction [44]. Decision trees, simply defined according to its name, a tree branch of the random forest model that decides or classifies for the final prediction. Each individual tree will split a class prediction and the class with the highest vote will become the final prediction of the model [14, 45]. Overall the Random Forest model denotes "public wisdom". Random forest is applied not only to solve classification problems, but also to solve regression problems [15, 44]. It has been consistent and stable among scholars for a few decades. Nevertheless, it equally requires more computational powers due to its complexity and completeness [14, 45]. In this research, Random Forest Regression is applied for predicting short-term asset prices.

3.2.2. Daily Closing Price Predictions

Figure 5 depicts the overall outline of our prediction system. For more convincing clarification whether the sentiment features of the financial news shares correlation with the asset price, this work first completes the fitting and forecasting of time-series data on stock price data

ISSN: 2692-7608 DOI: 10.6981/FEM.202107_2(7).0039

without utilizing the sentiment data. The subsequent step consists of separate sentiment predictions.

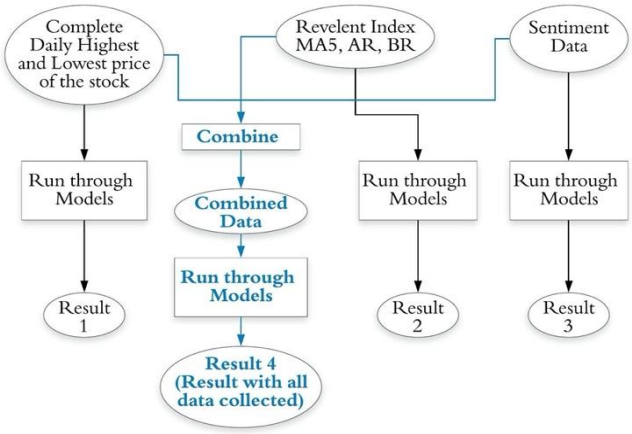

Figure 5. Shows the overall outline of our prediction system.

The third step combines both market data analysis and sentiment forecasting, and subsequently draws comparisons for those forecasts. Hopefully a higher model performance will be displayed in the aggregated one.

(1) Prediction from the Stock Market Data Analysis

This work makes predictions on the daily closing price based on the daily highest and lowest price adopting the prominent Logistic Regression, Gaussian Naive Bayes and Random Forest Regression. MA5, AR and BR index are equally applied for predictions. AR and BR index is directly computed from the daily opening, highest, lowest, and closing prices of the stock in the last 26 days without the usage of sentiment data.

(2) Prediction from the Sentiment Data

To assess the impact of financial news analysis on the asset price forecast, use the positive, negative, and composite values in the VADER sentiment computation results as attributes for machine learning models trained to determine the prices of asset. This work evaluates the sentiments extracted from Reuters real time press-release filtered from the large set of stock news between May 18, 2017 and May 28, 2020. The negative, positive and compound scores along with the daily opening price have been applied for predicting the Daily Closing Price.

(3) Combination of the Market Data Analysis and the Sentiment Prediction

To aggregate both the separate market data analysis and financial news sentiment analytics for testifying whether the aggregated one displays more suitable model performance, this work used the positives, negatives and compounds included in the results of the VADER sentiment calculation, as well as the daily open, high and low price included in the stock analysis and Index for predicting the Closing Price.

4. Results

This work has identified various prediction methods as the data and algorithm applied in the prediction procedure. The well-known Naïve Bayes and Random Forest are utilized for stock market analysis, sentiment predictions, as well as the aggregated predictions. The fitting degree of logistic regression rate is considerably lower compare to the other two models because it is a logistic function and cannot predict continuous values. Work computed for each root Mean Square Error, a measure for determining the performance of a particular algorithm. The work has assessed the performance of prediction with Random Forest and Naïve Bayes as presented in tables 5, 6 and 7.

ISSN: 2692-7608 DOI: 10.6981/FEM.202107_2(7).0039

Table 5. Presents the results of MSE computations for Boeing Co. Datasets.

Predictions	Data	Algorithm	MSE
Stock Market Data Analysis	Daily Highest and Lowest Price	Random Forest	19145.94771
		Naïve Bayes	9438.19608
	BRAR Index	Random Forest	41193.95890
		Naïve Bayes	33394.19863
	MA5	Naïve Bayes	41674.45395
Sentiment	Sentiment Scores	Random Forest	4635.09468
Aggregated	All Above	Random Forest	542.07362

Table 6. Presents the results of MSE calculations for Facebook Inc. Datasets.

Predictions	Data	Algorithm	MSE
Stock Market Data Analysis	Daily Highest and Lowest Price	Random Forest	10965.11585
		Naïve Bayes	4025.43293
	BRAR Index	Random Forest	72575.43949
		Naïve Bayes	59318.23567
	MA5	Naïve Bayes	29860.76687
Sentiment	Sentiment Scores	Random Forest	7722.91975
Aggregated	All Above	Random Forest	638.61783

Table 7. Presents the results of MSE calculation for Apple Inc.

Predictions	Data	Algorithm	MSE
Stock Market Data Analysis	Daily Highest and Lowest Price	Random Forest	180.22892
		Naïve Bayes	2659.40964
	BRAR Index	Random Forest	50919.59748
		Naïve Bayes	55876.25876
	MA5	Naïve Bayes	12600.42424
Sentiment	Sentiment Scores	Random Forest	2005.70122
Aggregated	All Above	Random Forest	233.08805

For stock market analysis, the precision level of prediction from Daily Highest and Lowest Price is unsatisfactory. The deficiencies of this prediction are as follows. The fitting method requires daily highest and lowest price which cannot be determined until the end of one day, thus the data can only reflect the recent situation. Moreover, it's usually difficult to obtain the latest data.

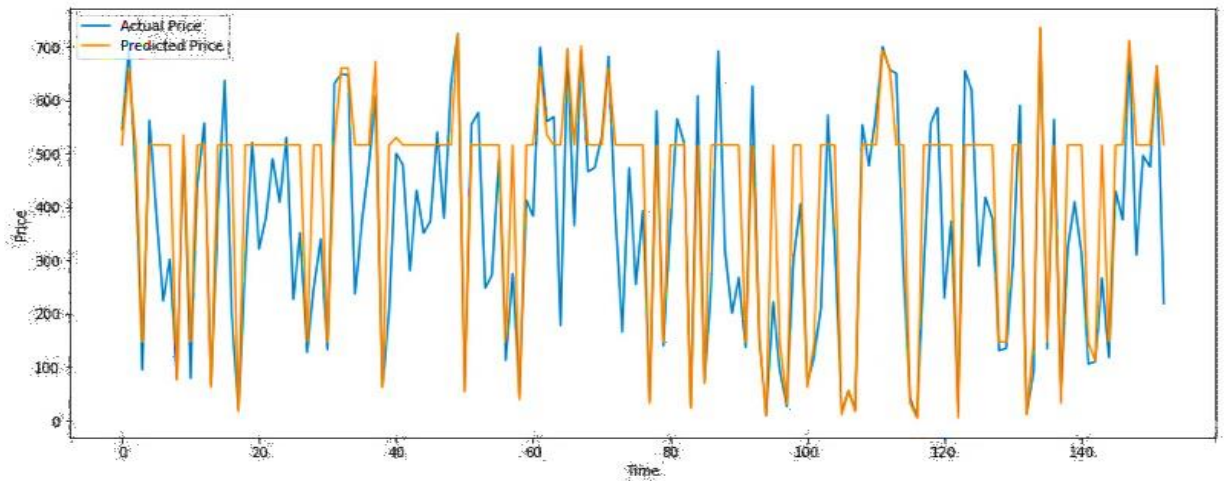

Figure 6. Shows the curve fitting using Daily Highest & Lowest Price for Boeing Co. Daily Closing Price predictions. This curve fitting is predicted through the application of Random Forest Regression.

The closing price is fitted by using indicators including MA5, AR and BR.

ISSN: 2692-7608 DOI: 10.6981/FEM.202107_2(7).0039

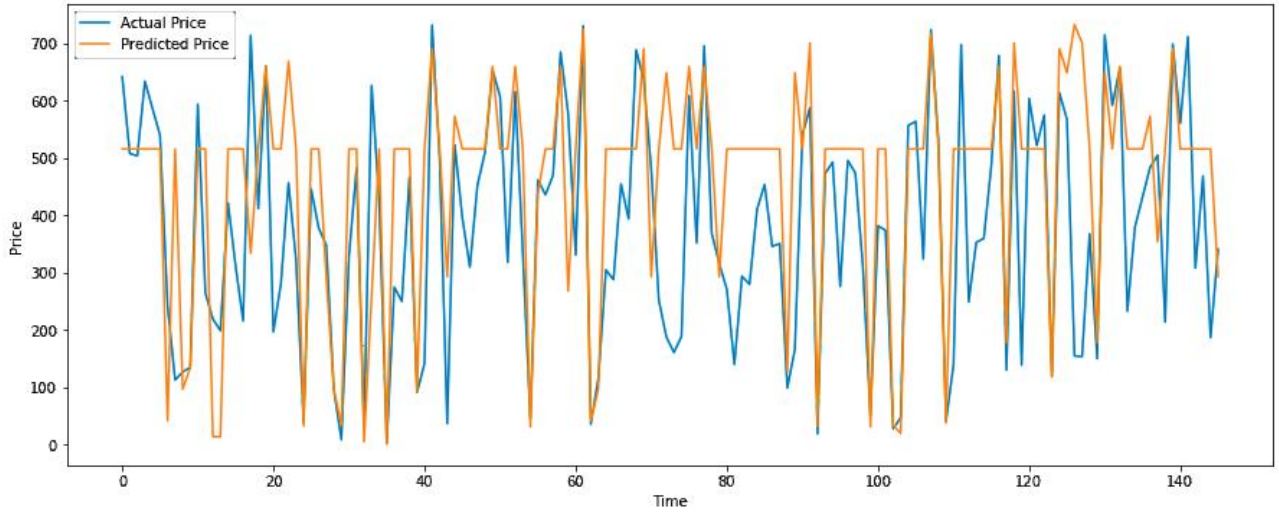

Figure 7. Shows the curve fitting using BRAR Index for Boeing Co. Daily Closing Price prediction. This curve fitting is predicted through the application of the Random Forest Regression.

The fitting degree appears low mainly due to the inefficiency of the index, implying this form of index is insufficient to indicate the rapid changes of the stock market. Furthermore, a small error may cause a significantly different outcome in the prediction process and therefore destroy accuracy.

As shown in the performance of the model, sentiment extracted from news articles is highly correlated with traders' readiness for purchasing or selling, and thus sentiment affects the price of the asset. This work also showed that the degree of adjustment is more suitable when the deferred time is set to 0, 7 and 14 days. It can be surmised that sentiment extracted from financial news tend to be applied as a proxy for short-term asset price movement. First, sentiment expressed in news may cause immediate action of the traders in an extremely brief period, for instance, in a single day. Second, sentiment from news media influences the market movement in the subsequent weeks. Traders can consider a week to be a normal period of time to consider and not change their decisions. The sentiment prediction is dependable to certain extent, and the demand for stock data in this prediction is lower, which may enhance the time value of the prediction results.

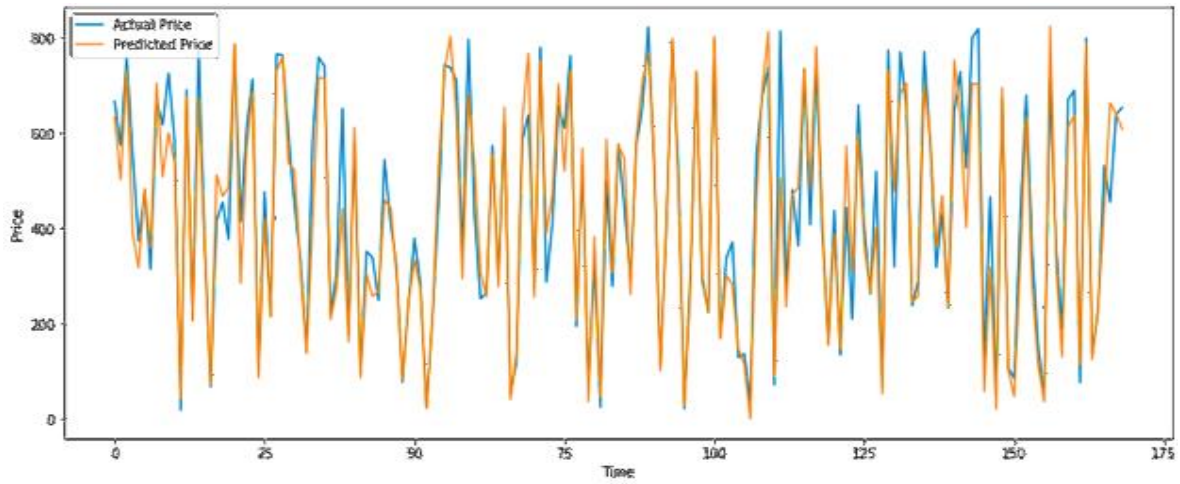

Figure 8. Depicts the curve fitting using Sentiment Scores for Boeing Co. Daily Closing Price prediction. This curve fitting is predicted through the application of the Random Forest Regression.

ISSN: 2692-7608 DOI: 10.6981/FEM.202107_2(7).0039

The aggregate method, which combines both market data analysis and sentiment prediction, displays more suitable performance of all the three models. It provides supportive evidence that stock market data and sentiment data separately are insufficient to illustrate the complexity of the stock market.

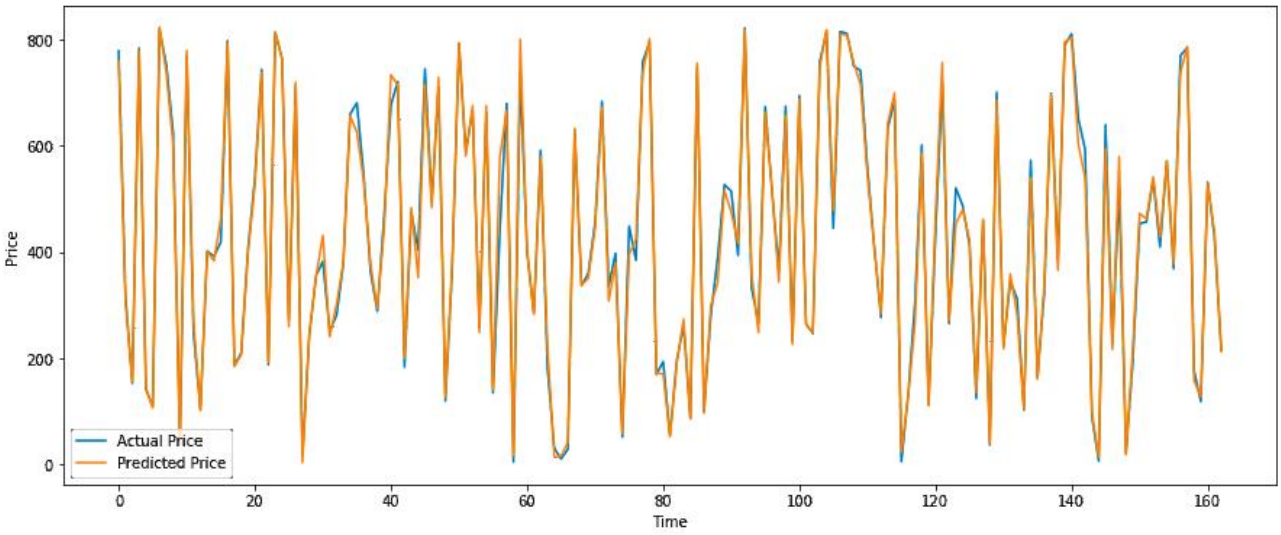

Figure 9. Shows the curve fitting using Sentiment Scores along with the Daily Highest & Lowest Price and BRAR Index for Boeing Co. Daily Closing Price prediction. This curve fitting is predicted through the application of the Random Forest Regression.

5. Conclusion

This paper shows that a feasible approach of predicting short term asset price movement through the use of extracted sentiments from financial news is put forward. First, the financial news articles relevant for the study are already obtained from Kaggle.com. Then they have been subjected to analysis to extract sentiment around companies. To test efficient market hypothesis and clarify the reliability of sentiment data in prediction, the work draws comparison among predictions from stock market analysis, sentiment analysis, as well as a hybrid of both. By relating various indicators to stock prices and using MSE to measure the level of precision of stock price forecasts, the work shows that sentiment analysis increases the performance of the model and that the level of precision of the method forecast, which uses all the information gathered, has the highest satisfactory performance. It can be surmised that financial news sentiment proves to be effective for short-term stock price forecasting.

The predictive model can subsequently be developed through the incorporation of more complex proxies and analysis techniques of machine learning or data mining. The future scope of this research tends to be varied. Feelings can be divided into diverse categories and the difference in predictive skills tends to be displayed. Furthermore, the sentiment around one business venture tends to affect another business venture over time. Utilizing the proposed method, several financial modeling and other modeling will be applied in automated trading strategies.

References

[1] Nikfarjam, A., Emadzadeh, E., & Muthaiyah, S. (2010, February). Text mining approaches for stock market prediction. In 2010 the 2nd international conference on Computer and automation engineering (ICCAE) (Vol. 4, pp. 256-260). IEEE.

ISSN: 2692-7608 DOI: 10.6981/FEM.202107_2(7).0039

[2] Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2019). A systematic review of fundamental and technical analysis of stock market predictions. Artificial Intelligence Review, 1-51.

[3] Schumaker, R. P., & Chen, H. (2009). Textual analysis of stock market prediction using breaking financial news: The AZFin text system. ACM Transactions on Information Systems (TOIS), 27(2), 1-19.

[4] Yao, J., Tan, C. L., & Poh, H. L. (1999). Neural networks for technical analysis: a study on KLCI. International journal of theoretical and applied finance, 2(02), 221-241.

[5] Zhai, Y., Hsu, A., & Halgamuge, S. K. (2007, June). Combining news and technical indicators in daily stock price trends prediction. In International symposium on neural networks (pp. 1087-1096). Springer, Berlin, Heidelberg.

[6] Fama, E. F. (1965). The behavior of stock-market prices. The journal of Business, 38(1), 34-105.

[7] Malkiel, B. G. 1973. A Random Walk Down Wall Street. W.W. Norton, New York.

[8] Fung, G. P. C., Yu, J. X., & Lam, W. (2002, May). News sensitive stock trend prediction. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 481-493). Springer, Berlin, Heidelberg.

[9] Lam, M. (2004). Neural network techniques for financial performance prediction: integrating fundamental and technical analysis. Decision support systems, 37(4), 567-581.

[10] Soni, A., van Eck, N. J., & Kaymak, U. (2007, April). Prediction of stock price movements based on concept map information. In 2007 IEEE Symposium on Computational Intelligence in Multi-Criteria Decision-Making (pp. 205-211). IEEE.

[11] Petrusheva, N., & Jordanoski, I. (2016). Comparative analysis between the fundamental and technical analysis of stocks. Journal of Process Management. New Technologies, 4(2), 26-31.

[12] Mills, T. C. (1997). Technical analysis and the London Stock Exchange: Testing trading rules using the FT30. International Journal of Finance & Economics, 2(4), 319-331.

[13] Khan, Z. H., Alin, T. S., & Hussain, M. A. (2011). Price prediction of share market using artificial neural network (ANN). International Journal of Computer Applications, 22(2), 42-47.

[14] Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. J. O. G. R. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geology Reviews, 71, 804-818.

[15] Segal, M. R. (2004). Machine learning benchmarks and random forest regression.

[16] Shen, S., Jiang, H., & Zhang, T. (2012). Stock market forecasting using machine learning algorithms. Department of Electrical Engineering, Stanford University, Stanford, CA, 1-5.

[17] Shynkevich, Y., McGinnity, T. M., Coleman, S., & Belatreche, A. (2015, July). Stock price prediction based on stock-specific and sub-industry-specific news articles. In 2015 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.

[18] Hegazy, O., Soliman, O. S., & Salam, M. A. (2014). A machine learning model for stock market prediction. arXiv preprint arXiv:1402.7351.

[19] Nelson, D. M., Pereira, A. C., & de Oliveira, R. A. (2017, May). Stock market's price movement prediction with LSTM neural networks. In 2017 International joint conference on neural networks (IJCNN) (pp. 1419-1426). IEEE.

[20] Nguyen, T. H., & Shirai, K. (2015, July). Topic modeling based sentiment analysis on social media for stock market prediction. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (pp. 1354-1364).

[21] Serban, I. V., González, D. S., & Wu, X. (2014). Prediction of changes in the stock market using twitter and sentiment analysis.

[22] Xu, Y., & Cohen, S. B. (2018, July). Stock movement prediction from tweets and historical prices. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 1970-1979).

[23] Alanyali, M., Moat, H. S., & Preis, T. (2013). Quantifying the relationship between financial news and the stock market. Scientific reports, 3, 3578.

ISSN: 2692-7608 DOI: 10.6981/FEM.202107_2(7).0039

[24] Ingle, V., & Deshmukh, S. (2016, August). Hidden Markov model implementation for prediction of stock prices with TF-IDF features. In Proceedings of the International Conference on Advances in Information Communication Technology & Computing (pp. 1-6).

[25] Yoo, P. D., Kim, M. H., & Jan, T. (2005, November). Machine learning techniques and use of event information for stock market prediction: A survey and evaluation. In International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06) (Vol. 2, pp. 835-841). IEEE.

[26] Kim, Y., Jeong, S. R., & Ghani, I. (2014). Text opinion mining to analyze news for stock market prediction. Int. J. Advance. Soft Comput. Appl, 6(1), 2074-8523.

[27] Peng, Y., & Jiang, H. (2015). Leverage financial news to predict stock price movements using word embeddings and deep neural networks. arXiv preprint arXiv:1506.07220.

[28] Van Bunningen, A. H., Nijholt, A., Poel, M., & Van Otterlo, M. (2004). Augmented trading: From news articles to stock price predictions using syntactic analysis.

[29] Tetlock P. (2007). Giving content to investor sentiment: The role of media in the stock market. The Journal of Finance, 62(3), 1139-1168.

[30] Garcia, D. (2013). Sentiment during recessions. The Journal of Finance, 68(3), 1267-1300.

[31] Mahajan, A., Dey, L., & Haque, S.K. (2008). Mining Financial News for Major Events and Their Impacts on the Market. 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 1, 423-426.

[32] Shapiro, A. H., Sudhof, M., & Wilson, D. (2020, March). Measuring news sentiment. Federal Reserve Bank of San Francisco.

[33] Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of computational science, 2(1), 1-8.

[34] Shapiro, A. H., & Wilson, D. (2019, February). Taking the Fed at its Word: Direct Estimation of Central Bank Objectives using Text Analytics. Federal Reserve Bank of San Francisco.

[35] Hosmer, D. W., Hosmer, T., Le Cessie, S., & Lemeshow, S. (1997). A comparison of goodness-of-fit tests for the logistic regression model. Statistics in medicine, 16(9), 965-980.

[36] Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: a methodology review. Journal of biomedical informatics, 35(5-6), 352-359.

[37] Anzai, Y. (2012). Pattern recognition and machine learning. Elsevier.

[38] Molnar, C. (2019). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable (2019). URL https://christophm. github. io/interpretable-ml-book.

[39] Lewis, D. D. (1998, April). Naive (Bayes) at forty: The independence assumption in information retrieval. In European conference on machine learning (pp. 4-15). Springer, Berlin, Heidelberg.

[40] Zhang, H. (2004). The Optimality of Naive Bayes,". In Proc. Seventeenth Int. Florida Artif. Intell. Res. Soc. Conf. FLAIRS 2004 (Vol. 1, No. 2, pp. 1-6).

[41] Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. Chemometrics and intelligent laboratory systems, 2(1-3), 37-52.

[42] Abdi, H., & Williams, L. J. (2010). Principal component analysis. Wiley interdisciplinary reviews: computational statistics, 2(4), 433-459.

[43] Keogh, E. (2006). Naive bayes classifier. Accessed: Nov, 5, 2017.

[44] Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

[45] Peters, J., De Baets, B., Verhoest, N. E., Samson, R., Degroeve, S., De Becker, P., & Huybrechts, W. (2007). Random forests as a tool for ecohydrological distribution modelling. ecological modelling, 207(2-4), 304-318.