FINANCE INDIA

Vol. XXXVI No. 3, September 2022
Pages—907 - 939

The Prediction of Intraday Stock Market Movements in Developed & Emerging Markets using Sentiment and Emotions from Twitter

TALITA GREYLING*
STEPHANIE ROSSOUW**
DIMITRI H.W. STEYN***

Abstract

Paper investigates the predictability of stock market movements using text data related to stock markets extracted from the social media platform Twitter. We use high-frequency intraday data rather than daily data and analyse and compare results for both emerging and developed markets. To this end, the study uses three different Machine Learning Classification Algorithms: the Naïve Bayes, K-Nearest Neighbours and the Support Vector Machine algorithms. Several model metrics such as Precision, Recall, Specificity and the F1-Score are also used. Lastly, we use K-Fold Cross-Validation to validate our machine learning models' results and applicability to unseen data. The predictability of the market movements is estimated first by using only sentiment and then using a combination of sentiment and emotions. Our results indicate that investor sentiment and emotions derived from stock-market related tweets are significant predictors of stock market movements. This model does not only give good results in developed markets but also emerging markets.

JEL Code : C6, C8, G0, G4
Keywords : Scrips, Emotions, Sentiment Analysis, Classification, Prediction, Machine Learning, Twitter, Stock Exchange

I. Introduction

RESEARCH INTO THE field of accurate prediction of stock market movements is of interest to academics, economists, and financial analysts due to the profitability of accurately predicting the markets. Stock market movements can be explained as the up and downshift of a stock market, i.e. the deviations from its previous value. Previously, these stock market movements were predicted using rational, risk-based asset-pricing models, arguing that the prices reflect the discounted value of expected future cash

* Associate Professor, University of Johannesburg, School of Economics, D Ring 222, Auckland Park, Johannesburg, 2006, SOUTH AFRICA.
** Senoir Lecturer, Auckland University of Technology, School of Social Science & Public Policy, School of Economics, 55, Wellesley Street, East Auckland City, Auckland CBD 10010, NEW ZEALAND.
*** Junior Lecturer, The University of Johannesburg, School of Ecconomics, Auckland Park, Johannesburg, 2006, SOUTH AFRICA.

Submitted May 2021; *Accepted* March 2022

References

Abbes, H., (2015), "*Tweets Sentiment and their Impact on Stock Markets. Dissertation written with a view to obtaining a degree in Business Engineering specialized in Supply Chain Management and Business Analytics.*", Liege University, pp. 1-75.

Ageitos, E. C., (2018), "*Experiment on sentiment analysis over LSE (London Stock Exchange). Twitter data*", Trabuco de Fin de Grado Escuela de Ingeniería de Telecomunicación Grado en Ingeniería de Tecnologías de Telecomunicación. Working Paper Universida de Vigo, pp. 1-47.

Antweiler, W. and M.Z. Frank, (2004), "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards.", *The Journal of Finance*, Vol. 59, No. 3, pp. 1259-1294.

Audrino, F., F. Sigrist and D. Ballinari, (2020), "The impact of sentiment and attention measures on stock market volatility.", *International Journal of Forecasting*, Vol. 36, pp.334-357.

Baker, M. and J. Wurgler, (2006), "Investor sentiment and the cross?section of stock returns.", *The Journal of Finance*, Vol. 61, No. 4, pp. 645-1680.

Bank, E., (2019), "*Long vs. Short Stocks*", Zacks Publishing, pp. 1-3.

Black, F., (1986), "Noise", *The Journal of Finance*, Vol. 41, No. 3, pp. 528-543.

Bollen, J., H. Mao and X.J. Zeng, (2011), "Twitter mood predicts the stock market.", *Journal of Computational Science*, Vol. 2, No. 1, pp.1-8.

Bouktif, S., (2020), "Augmented Textual Features-Based Stock Market Prediction.", *IEEE Access*, Vol. 8, 4026940282.

Brown, G. W. and M.T. Cliff, (2004), "Investor sentiment and the near-term stock market.", *Journal of Empirical Finance*, Vol. 11, No. 1, pp.1-27.

Bukovina, J., (2016), "Social media big data and capital markets - An overview.", *Journal of Behavioral and Experimental Finance*, Vol. 11, (C), pp. 18-26.

Chen, R. and M. Lazer, (2011), "Analysis of Twitter Feeds for the Prediction of Stock Market Movement.", *Semantic Scholar*, pp. 1-5.

Chung, S. and S. Liu, (2011), "Predicting Stock Market Fluctuations from Twitter.", *Berkely California*, pp. 1-13.

Das, S. R. and M.Y. Chen, (2007), "Yahoo! for Amazon: Sentiment extraction from small talk on the web.", *Management Science,* Vol. 53, No. 9, pp. 1375-1388.

Da, Z., J. Engelberg and P. Gao, (2015), "The Sum of All FEARS: Investor Sentiment and Asset Prices.", *Review of Financial Studies,* Vol. 28, 1, pp. 1-32.

De Long, J.B., A. Shleifer, L.H. Summers and R.J. Waldmann, (1990), "Positive feedback investment strategies and destabilizing rational speculation.", *The Journal of Finance,* Vol. 45, No. 2, pp.379-395.

Elodie, M., (2019), "*Twitter and its relationship with returns and trading volume of European stocks*.", Working Paper, Louvain School of Management, Université Catholique de Louvain, 2019. Prom.: D'Hondt, Catherine; Desagre, Christophe, pp. 1-104.

Gentry, J., (2016), "Package 'twitter'." *The CRAN Project*, pp. 1-32.

Gholampour, V.,(2019), "Daily expectations of returns index." *Journal of Empirical Finance*, Vol.54, pp. 236-252

Hall, P., B.U. Park and R.J. Samworth, (2008), "Choice of Neighbor Order in Nearest-Neighbor Classification.", *The Annals of Statistics*, Vol. 36, No. 5, pp. 2135-2152.

Hu, M. and B. Liu, (2004), "*Mining and Summarizing Customer Reviews.*" Seattle, Washington, USA, Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Hughen, J.C. and C.G. McDonald, (2005), "Who are the noise traders?", *Journal of Financial Research*, Vol. 28, No. 2, pp. 281-298.

Jadhav, R. and M.S. Wakode, (2017), "Survey: Sentiment Analysis of Twitter Data for Stock Market Prediction.", *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 6, No. 3, pp. 507-509.

Jagdale, R. S., V.S. Shirsat, and S.N. Deshmukh, (2016) "Sentiment Analysis of Events from Twitter Using Open-Source Tool.", *International Journal of Computer Science and Mobile Computing*, Vol. 5, No. 4, pp. 475-485.

Jha, S. K., S.N. Bhattacharya and M. Bhattacharya, (2019), "Google Search Volume and Stock Market Liquidity.", *Indian Journal of Finance*, Vol. 13, No. 8, pp. 51-64.

Khan, W., U. Malik, M.A. Ghazanfar, M.A. Azam, K.H. Alyoubi and A.S. Alfakeeh, (2019), "Predicting stock market trends using machine learning algorithms via public sentiment and political situation analysis.", *Soft Computing*, pp. 1-25.

Khan, W., M.A. Ghazanfar, M.A. Azam, A. Karami, K.H. Alyoubi and A.S. Alfakeeh, (2020), "Stock market prediction using machine learning classifiers and social media news." *Journal of Ambient Intelligence and Humanized Computing*, pp.1-24.

Kolagani, S. H. D., A. Negahban and C. Witt, (2017), "Identifying trending sentiments in the 2016 us presidential election: A case study of Twitter analytics.", *Issues in Information Systems*, Vol. 18, No. 2, pp. 80-86.

Kolasani, S.V. and R. Assaf, (2020), "Predicting Stock Movement Using Sentiment Analysis of Twitter Feed with Neural Networks.", *Journal of Data Analysis and Information Processing*, Vol. 8, pp. 309-319.

Kundu, S. and A. Banerjee, (2020), "Does Twitter Activity Proxy for Idiosyncratic Information? Evidence from the Indian Stock Market.", *Indian Journal of Finance,* Vol. 15, No. 8, pp. 8-23.

Li, Q., T. Wang, P. Li, L. Liu, Q. Gong and Y. Chen, (2014), "The effect of news and public mood on stock movements.", *Information Sciences*, Vol. 278, pp. 826-840.

Maqsood, H., I. Mehmood, M. Maqsood, M. Yasir, S. Afzal, F. Aadil, M.M. Selim and K. Muhammad, (2020), "A local and global event sentiment-based efficient stock exchange forecasting using deep learning.", *International Journal of Information Management*, Vol. 50, pp. 432-451.

Maree, S. and K. Johnston, (2015), "Critical Insights into the Design of Big Data Analytics Research: How Twitter 'Moods' Predict Stock Exchange Index Movement.", *The African Journal of Information and Communication*, Vol. 15, pp. 53-67.

Mohsin, M., (2019), "10 Social Media Statistics You Need to Know in 2020.", *Oberlo Blog.*

MSCI., (2019), "Results of the MSCI 2019 global market accessibility review.", MSCI, USA

Moritz, B., (2018), "*Applications of Textual Analysis and Machine Learning in Asset Pricing.*", Dissertation an der Fakult¨at fu¨r Mathematik, Informatik und Statistik der Ludwig-Maximilians-Universita¨t Mu¨nchen, pp. 1-19.

Mohammad, S., S. Kiritchenko and X. Zhu, (2013), "*In Proceedings of the seventh international, workshop on Semantic Evaluation Exercises*", June 2013, Atlanta, USA.

Naldi, M., (2019), "*A review of sentiment computation methods with R packages.*", ArXiv 1901.08319.

Nisar, T. M. and Yeung, M., (2018), "Twitter as a tool for forecasting stock market movements: A shortwindow event study", *The Journal of Finance and Data Science*, Vol. 4, pp. 101-119.

Nti, I. K., A.F. Adekoya and B.A. Weyori, (2020), "Predicting Stock Market Price Movement Using Sentiment Analysis: Evidence From Ghana", *Applied Computer Systems*, Vol. 25, No. 1, pp. 33-42.

Oliveira, N., P. Cortez and N. Areal, (2017), "The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices", *Expert Systems with Applications*, Vol. 73, pp. 125-144.

Pagolu, V. S., K.N. Reddy Challa, G. Panda and B. Majhi, (2016), "*Sentiment Analysis of Twitter Data for Predicting Stock Market Movements*", International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)-2016.

Panday, H., V. Vijayarajan, A. Mahendran, A. Krishnamoorthy and V.B.S. Prasath, (2020), "Stock Prediction using Sentiment analysis and Long Short-Term Memory", *European Journal of Molecular & Clinical Medicine*, Vol. 7, No. 2, pp. 5060-5069.

Paramanika, R. N. and V. Singhal, (2020), "Analysis of Indian Stock Market Volatility.", *Procedia Computer Science*, Vol. 176, pp.330-338.

Qian, B. and K. Rasheed, (2007), "Stock market prediction with multiple classifiers.", *Applied Intelligence*, Vol. 26, No. 1, pp. 25-33.

Rao, T. and S. Srivastava, (2012), "*Analyzing Stock Market Movements Using Twitter Sentiment Analysis*.", Delhi, International Conference on Advances in Social Networks Analysis and Mining.

Reddy, N. N., E. Naresh and V.B.P. Kumar, (2020), "*Predicting stock price using sentimental analysis through Twitter data.*", 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT).

Rennie, J.D.M., L. Shih, J. Teevan and D.R. Karger, (2003), "*Tackling the Poor Assumptions of Naive Bayes Text Classifiers*", Washington DC: Massachusetts Institute of Technology, pp. 1-8.

Renoult, S., (2019), "Doom and Gloom in the Developed Economies: Time to Invest in Emerging Markets.", *FSG Journal*, pp.1-5.

Ruan, Y., Durresi, A. and L. Alfantoukh, (2018), "Using Twitter trust network for stock market analysis.", *Knowledge-Based Systems*, pp.1-12.

Sanjay, M., (2018), Why and how to Cross-Validate a Model? Towards Data Science, pp. 1-4.

Shen, D., L. Liu and Y. Zhang, (2018), "Quantifying the cross-sectional relationship between online sentiment and the skewness of stock returns." Physica A, Vol. 490, pp. 928-934.

Smith, T., (2019), "Market Sentiment.", *Investopedia*. Available online.

Statista., (2019), "*Leading countries based on number of Twitter users as of July 2019 (in millions).*", Statista online data retrieved from Twitter; We Are Social; Hootsuite, and DataReportal.

Tabari, N., P. Biswas, B. Praneeth, A. Seyeditabari, M. Hadzikadic and W. Zadrozny, (2018), "*Causality Analysis of Twitter Sentiments and Stock Market Returns*", Association for Computational Linguistics, Melbourne, pp. 11-19.

Valdivia, A., M.V. Luz'on and F. Herrera, (2017), "Sentiment analysis in TripAdvisor.", *IEEE Intelligent Systems,* Vol. 32, No. 4, pp. 72-77.

Wei, P., Z. Lu and J. Song, (2015), "Variable importance analysis: A comprehensive review.", *Reliability Engineering and System Safety*, Vol. 142, No. 1, pp. 399-432.

World Bank, (2019), "*GDP (current US$).*", World Bank national accounts data, and OECD National Accounts data files. The World Bank Group. Available online

Yang, S.Y., S.Y.K. Mo, and A. Liu, (2015), "Twitter financial community sentiment and its predictive relationship to stock market movement.", *Quantitative Finance*, Vol. 15, No. 10, pp.1637-1656.

You, W., Y. Guo and C. Peng, (2017), "Twitter's daily happiness sentiment and the predictability of stock returns.", *Finance Research Letters*, Vol. 23, No. C, pp.58-64.

Zehe, A., M. Becker, L. Hettinger, A. Hotho, I. Reger and F. Jannidis, (2016), "*Prediction of happy endings in German novels based on sentiment information.*", In 3rd Workshop on Interactions between Data Mining and Natural Language Processing, Riva del Garda, Italy, 2016.

Zhang, X., H. Fuehres and P.A. Gloor, (2010), "Predicting Stock Market Indicators Through Twitter I hope it is not as bad as I fear". *Procedia - Social and Behavioral Sciences*, Vol. 26, pp. 55-62.

Zhao, R., (2019), "*Quantifying the correlation and prediction of daily happiness sentiment and stock return : The Case of Singapore*", Physica A, pp. 533, 1-9.

Annexure I
Table AI A
Summary of Most Significant Studies

Paper name	Emerging market (give names)	Developed markets (names)	Sentiment	Emotion (give words)	Influencers	Viral	Daily	High frequency	Type of model	Type of machine	Control variables
Bollen, Mou and Zeng (2011)		Dow Jones Industrial Average (DJIA could be used as a US proxy)		Calm, Alert, Sure, Vital, Kind and Happy			☑		Granger Causality, Multiple Linear Regression	Fuzzy Neural Network	None but they include cross-validation by checking the effect on thanksgiving and presidential campaign day
Maree and Johnston (2015)	JSE ALSI (South Africa)			Depression, Tension, Anger, Vigor, Fatigue and Confusion			☑		Spearman Correlation, Granger Causality	Neural Network	None
Tabari, Biswas, Praneeth, Seyeditabri Hodzikadic & Zadrozny (2018)		A tweet was considered stock related if it contained at least one of the stock symbols of the first 100 most frequent stock symbols that were included in SemEval dataset form	☑		☑	☑	☑		Granger Causality	SVM, Random Forest	None
Rao and Srivastava (2012)		NASDAQ, DJIA (both from the USA) and then they included companies: Amazon Apple, Dell, eBay, etc.		☑-Using tweets got Bullishhness Message Volume and Agreement			☑		Correlation, Granger Causality, OLS and then used Expert Model Mining system to see R square and Error-values		
Zhang, Fueehers & Gloor (2010)		NASDAQ, Dow Jones and S&P 500 (ALL USA)		Hope, Happy, Fear, Worry, Nervours, Anxious, Upset, Positive, Negative	☑	☑	☑		Correlation analysis		Yes- Chicago Board Options Volalility Index (VIX) as an external benchmark of investor fear

(Contd....)

Table AIA (Continued)

Abbes (2015)	FTSE100 (U.K.)	☑	☑	☑		Causality, linear regression, Breusch-pagan, Shapiro-Wilk and Kolmogorov-Smirnov, logistic		None
You, Guo and Peng (2017)	Ten international stock markets					Granger non-causality in quintiles, Quantile regressions		None
Jadhav (2017) and Wakode (2017)	S &P 500 (USA) market	☑	☑	☑		Logistic, correlation	SVM, Random Forest Neu	None
Zhao (2019)	Singapore stock market	☑		☑		Linear quantile regression, non-liner contemporaneous correlation tests, VAR model, Granger causality	None	
Maqsood Mehmood, Maqsood, Yasir, Afzal, Adil, Selim & Muhammad (2020)	Four countries	Event sentiment			☑	Linear regression	SVM, Neural Network	None
Ruan, Durresi and Alfantoukh (2018)	Eight firms in SP500	☑ valence	☑	☑	correlationes MAE, Linear Regression,		SVR	Yes- compared treating authors equally with those that are not 'equal.'

Source: Self Compiled

Annexure II
Table AIIA
UK Evaluation Metrics

Measure	Sentiment (model 1)	Sentiment and Emotions (model 2)
	Naïve Bayes (NB)	
Accuracy	52.19	49.96
Recall	23.48	14.59
Precision	52.88	45.59
F1-Score	32.21	21.54
	K-Nearest Neighbours (KNN)	
Accuracy	55.90	52.56
Recall	56.81	57.56
Precision	56.01	52.68
F1-Score	56.41	55.00
	Support Vector Machine - Kernel (SVM-K)	
Accuracy	55.23	55.02
Recall	54.75	62.19
Precision	94.31	54.85
F1-Score	69.13	58.28

Source: Self Computed

Table A II B
Germany Evaluation Metrics

Measure	Sentiment (model 1)	Sentiment and Emotions (model 2)
	Naïve Bayes (NB)	
Accuracy	54.81	46.66
Recall	100.00	16.89
Precision	54.81	45.08
F1-Score	71.12	24.13
	K-Nearest Neighbours (KNN)	
Accuracy	54.25	53.40
Recall	77.00	69.55
Precision	55.41	55.14
F1-Score	64.39	61.48
	Support Vector Machine - Kernel (SVM-K)	
Accuracy	55.37	56.21
Recall	96.15	89.23
Precision	55.31	56.13
F1-Score	70.08	68.80

Source: Self Computed

Table A II C
Japan Evaluation Metrics

Measure	Sentiment (model 1)	Sentiment and Emotions (model 2)
	Naïve Bayes (NB)	
Accuracy	45.64	49.27
Recall	57.95	47.24
Precision	47.63	50.10
	K-Nearest Neighbours (KNN)	
Accuracy	54.12	51.09
Recall	69.86	61.52
Precision	54.29	52.00
F1-Score	61.07	56.35
	Support Vector Machine - Kernel (SVM-K)	
Accuracy	53.52	51.70
Recall	57.95	44.86
Precision	54.22	52.70
F1-Score	56.02	50.45
F1-Score	48.63	

Source: Self Computed

Table A II D
France Evaluation Metrics

Measure	Sentiment (model 1)	Sentiment and Emotions (model 2)
	Naïve Bayes (NB)	
Accuracy	55.51	52.55
Recall	100.00	86.72
Precision	55.33	54.14
F1-Score	71.57	66.55
	K-Nearest Neighbours (KNN)	
Accuracy	55.87	52.18
Recall	98.61	84.64
Precision	55.58	53.97
F1-Score	71.44	65.81
	Support Vector Machine - Kernel (SVM-K)	
Accuracy	54.40	50.34
Recall	71.44	61.03
Precision	56.05	53.20
F1-Score	62.79	56.84

Source: Self Computed

Table A II E
Spain Evaluation Metrics

Measure	Sentiment (model 1)	Sentiment and Emotions (model 2)
	Naïve Bayes (NB)	
Accuracy	55.18	52.94
Recall	89.86	84.86
Precision	55.25	54.02
F1-Score	68.31	65.91
	K-Nearest Neighbours (KNN)	
Accuracy	55.87	46.07
Recall	71.44	64.14
Precision	57.25	49.03
F1-Score	63.54	55.54
	Support Vector Machine - Kernel (SVM-K)	
Accuracy	54.43	51.81
Recall	94.86	87.00
Precision	54.63	53.29
F1-Score	69.18	65.98

Source: Self Computed