STOCK MARKET CLASSIFICATION MODEL USING SENTIMENT ANALYSIS BASED ON HYBRID NAÏVE BAYES CLASSIFIERS

GHAITH ABDULSATTAR A. JABBAR ALKUBAISI

DOCTOR OF PHILOSOPHY
UNIVERSITI UTARA MALAYSIA
2019

Permission to Use

In presenting this thesis in fulfilment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the Universiti Library may make it freely available for inspection. I further agree that permission for the copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to:

Dean of Awang Had Salleh Graduate School of Arts and Sciences
UUM College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok

Abstrak

Analisis sentimen telah menjadi satu kaedah lazim untuk mengklasifikasi tingkah laku pasaran saham. Malahan, analisis sentimen kian menjadi penting pada dekad ini terutamanya dengan ketersediaan data daripada media sosial seperti Twitter. Walau bagaimanapun, ketepatan model klasifikasi pasaran saham masih rendah, dan ini secara negatifnya memberi kesan kepada petunjuk pasaran saham. Tambahan pula, terdapat pelbagai faktor yang memberi kesan langsung kepada ketepatan model klasifikasi yang tidak diambil kira dalam kajian terdahulu. Salah satu faktornya adalah pengecualian ciri spatial-temporal. Faktor lain yang penting adalah teknik pelabelan automatik yang menjurus kepada ketepatan klasifikasi yang rendah disebabkan oleh ketiadaan leksikon khusus. Kesesuaian pengklasifikasi terhadap ciri data dan domain juga adalah faktor lain yang memberi kesan kepada ketepatan klasifikasi. Dalam kajian ini, model klasifikasi pasaran saham berdasarkan analisis sentimen telah dibangunkan. Model ini direka bentuk untuk meningkatkan ketepatan klasifikasi dengan penggabungan ciri tweet timestamp dan lokasi, teknik pelabelan pakar domain pasaran saham dan pembangunan pengklafikasi Naïve Bayes hibrid untuk mengklasifikasi sentimen pasaran saham. Metodologi kajian ini terdiri daripada enam fasa. Fasa pertama adalah pengumpulan data, dan fasa kedua merupakan fasa penting yang melibatkan pelabelan dimana polariti data ditentukan sebagai nilai negatif, positif atau neutral. Fasa ketiga melibatkan pra-pemprosesan data yang mana hanya ciri berkaitan sahaja diambil kira. Fasa keempat adalah klasifikasi dimana corak pasaran saham yang sesuai dikenal pasti melalui penghibridan pengklasifikasi Naïve Bayes. Fasa kelima adalah penilaian dan prestasi, dan fasa terakhir iaitu pengecaman tingkah laku pasaran saham. Model ini menghasilkan dapatan yang signifikan dalam mengklasifikasi tingkah laku pasaran saham dengan ketepatan melebihi 89%. Model ini bermanfaat kepada pelabur dan penyelidik. Bagi pelabur, ia membolehkan mereka merumus pelan berdasarkan ketepatan petunjuk di mana ia mengurangkan risiko dalam pembuatan keputusan. Dari segi penyelidik, ia menarik perhatian terhadap kepentingan kejuruteraan ciri, teknik pelabelan, dan penghibridan pengklasifikasi dalam meningkatkan ketepatan klasifikasi.

Kata Kunci: Klasifikasi pasaran saham, Pengklasifikasi Naive Bayes hibrid, Analisis sentimen, Pelabelan pakar, ciri *spatial-temporal*.

Abstract

Sentiment analysis has become one of the most common method to classify stock market behaviour. Moreover, sentiment analysis has gained a lot of importance in the last decade especially due to the availability of data from social media such as Twitter. However, the accuracy of stock market classification models is still low, and this has negatively affected the stock market indicators. Furthermore, there are many factors that have a direct effect on the classification models' accuracies which were not addressed by previous research. One of the factors is the exclusion of spatial-temporal features. Another important factor is the automatic labelling technique which leads to low classification accuracy due to the absence of specific lexicon. The appropriateness of the classifiers to the data features and domain is also another factor, which affect the classification accuracy. In this research, a model for stock market classification based on sentiment analysis is constructed. It is designed to enhance the classification accuracy by the incorporation of tweet timestamp and location features, stock market domain expert labelling technique and the construction of a hybrid Naïve Bayes classifiers to classify the stock market sentiments. The methodology for this research consists of six phases. The first phase is data collection, and the second phase represents the most important phase, which is labelling, in which polarity of data is specified as negative, positive or neutral values. The third phase involves data pre-processing, which is conducted to get only relevant features. The fourth phase is classification in which suitable patterns of the stock market are identified by hybridizing different Naïve Bayes classifiers. The fifth phase is performance and evaluation, and the final phase is recognition for the stock market behaviour. The model produced a significant result in classifying stock market behaviour with accuracy more than 89%. The model is beneficial for investors and researchers. For investors, it enables them to formulate their plans based on accurate indicators whereby it reduces the risk in decision making. For researchers, it draws their attention to the importance of feature engineering, labelling technique, and the classifiers hybridization in enhancing the classification accuracy.

Keywords: Stock market classification, Hybrid Naive Bayes classifiers, Sentiment analysis, Expert labelling, Spatial-temporal features.

Acknowledgment

All thanks to almighty Allah

O Lord, to You is praise as befits the Glory of Your Face and the greatness of Your Might.

يَا رَبِّ لَكَ الْحَمْدُ كَمَا يَنْبَغِي لِجَ لَالِ وَجْهِكَ وَلِعَظِيمِ سُلْطَانِكَ.

First and foremost, I am heartily thankful to my supervisors, Associate Prof. Dr. Siti Sakira Kamaruddin and Associate Prof. Dr. Husniza Husni for their appreciated guidance and continuous support from the initial to the final level in this research. The honest supervision and real encouragement that they gave truly help the headway of this research. It is an honor for me to have both of you as my supervisors. May Allah reward you well (جزاكم الله خير).

My genuine and heartfelt thanks to my dear parents and my big brother, My father Prof. Dr. Abdulsattar Al-kubaisi, my mother Siham Al-zubaidi, and my big brother Laith Al-kubaisi. My heartfelt thanks and appreciation are also extended to my parents in law, my father in law Noori Alani and my mother in law Alya Alani. Thank you for your support, continual prayers, and patience for our parting. Nothing in this world is equal to your boundless giving and support.

I wish to express my gratitude and thanks to the academic and supporting staff in AHSGS and SOC, especially to the dean of AHSGS Prof. Dr. Ku Ruhana Ku-Mahamud, Associate Prof. Dr. Yuhanis Yusof, Dr. Farzana Kabir Ahmad, Dr. Juhaida Abu Bakar, and Dr. Nor Hazlyna Harun.

Special thanks are due to my dear friends, Ghanim Shamas, Qais Alrubaiei, and Muthana Alani. Dear Ghanim and dear Qais, I will never forget our happy moments in UUM, thanks for everything. Dear Muthana, I will never forget your support.

Last but not least, to the person who made my life beautiful, my wife, thanks a lot for your love and support during my difficult times. My dear son, Ayham, my dear daughter, Elaf, thanks for your patience and continual prayers for your daddy.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ADX	Abu Dhabi Securities Exchange
API	Application Programming Interface
ASA	AlSafi Arabia
B2B	Business to Business
BN	Bayesian Network
BNB	Bernoulli Naive Bayes
CSV	Comma Separated Values
DFM	Dubai Financial Market
DM	Dubai Mall
DMM	Dubai Marina Mall
EM	Expectation Maximization
EMH	Efficient Market Hypothesis
FN	Falls Negative
FP	Falls Positive
GCC	Gulf Cooperation Council
GNB	Gaussian Naïve Bayes
HNBCs	Hybrid Naïve Bayes Classifiers

HTTP	Hypertext Transfer Protocol
JSON	JavaScript Object Notation
KNNs	K-Nearest Neighbours
LR	Linear regression
ME	Maximum Entropy
ML	Machine Learning
MNB	Multinomial Naïve Bayes
MS	Microsoft
MSA	Modern Standard Arabic
NB	Naïve Bayes
NBC	Naïve Bayes Classifier
NBCs	Naïve Bayes Classifiers
NNs	Nerul Networks
OAUTH	Open Authorization
REST	Representational State Transfer
RF	Random Forest
R-Reqs	Research Requirements
S&P	Standard and Poor

SA	Saudi Arabia
SSNB	Semi-Supervised Naïve Bayes
SVM	Support vector machine
TF-IDF	Term Frequency-Inverse Document Frequency
TN	True Negative
TP	True Positive
UAE	United Arab Emirates
URL	Uniform Resource Locator
UTF	Unicode Transformation Format

CHAPTER ONE
INTRODUCTION

This chapter presents an overview of stock market investment and stock market classification models, to introduce the research. It explains the problem statement and proposed solutions, discusses the research questions, and introduces the purpose of the study by presenting the research objectives, the motivation for the study, the research scope, the research significance, and finally the thesis organization.

1.1 Overview

Investors and business people need to decide on an effective approach to improve the outputs of their investments and to avoid massive financial losses, mainly on investment in the stock market (Nassirtoussi, Aghabozorgi, Wah, & Ngo, 2014; Ren, Wu, & Liu, 2018). The stock market is important because a company's stock prices play a pertinent role in all economic sectors (Baker, Stein, & Wurgler, 2002; Pan & Mishra, 2018). The global increment of the stock exchanges has raised the need for an in-depth decision-making tool using a stock market classification model (Bartov, Faurel, & Mohanram, 2017; Ruan, Durresi, & Alfantoukh, 2018).

Accurate classification of the data sources in the stock market domain is necessary for investors to make suitable decisions, such as selling or buying stocks (Guresen, Kayakutlu, & Daim, 2011; Hsu, Lessmann, Sung, Ma, & Johnson, 2016; Zhong & Enke, 2017). These kinds of investments need a pattern (Smedt & Daelemans, 2012; Fortuny, Smedt, Martens, & Daelemans, 2014) to assist decision makers in the stock market reach the right decision with minimal risk (Fortuny et al., 2014; Nguyen, Shirai, & Velcin, 2015). To determine a suitable pattern, trends must be followed by

The contents of the thesis is for internal user only

REFERENCES

Abbasi, A., Chen, H., & Salem, A. (2008). Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Transactions on Information Systems (TOIS), 26*(3), 12.

Abdelwahab, O., Bahgat, M., Lowrance, C. J., & Elmaghraby, A. (2015, 7-10 Dec). Effect of training set size on SVM and Naive Bayes for Twitter sentiment analysis. *Proceedings of the IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*.

Abdulqader, K. S. (2015). GCC's Economic Cooperation and Integration: Achievements and Hurdles. *Aljazeera Centre for Studies*.

Agrawal, R., Imielinski, T., & Swami, A. (1993). Database mining: A performance perspective. *IEEE Transactions on Knowledge and Data Engineering, 5*(6), 914-925.

Aizawa, A. (2003). An information-theoretic perspective of tf–idf measures. *Information Processing & Management, 39*(1), 45-65.

Alagić, D., & Šnajder, J. (2015). Experiments on Active Learning for Croatian Word Sense Disambiguation. *Proceedings of the 5th Workshop on Balto-Slavic Natural Language Processing*, (pp. 49-58).

Ali, R., Lee, S., & Chung, T. C. (2017). Accurate multi-criteria decision making methodology for recommending machine learning algorithm. *Expert Systems with Applications, 71*, 257-278.

Alpaydin, E. (2009). *Introduction to machine learning*: MIT press.

AL-Rubaiee, H., Qiu, R., Alomar, K., & Li, D. (2018). Techniques for Improving the Labelling Process of Sentiment Analysis in the Saudi Stock Market. *International Journal of Advanced Computer Science and Applications, 9*(3), 34-43.

Andreevskaia, A., & Bergler, S. (2008). When Specialists and Generalists Work Together: Overcoming Domain Dependence in Sentiment Tagging. *Proceedings of the ACL-08: HLT,* 290-298.

Anjaria, M., & Guddeti, R. M. R. (2014). Influence factor based opinion mining of Twitter data using supervised learning. *Proceedings of the 6th IEEE International Conference on Communication Systems and Networks (COMSNETS)*, (pp. 1-8).

Aramaki, E., Maskawa, S., & Morita, M. (2011). Twitter catches the flu: detecting influenza epidemics using Twitter. *Proceedings of the conference on empirical methods in natural language processing*, (pp. 1568-1576). Association for Computational Linguistics.

Araque, O., Corcuera-Platas, I., Sanchez-Rada, J. F., & Iglesias, C. A. (2017). Enhancing deep learning sentiment analysis with ensemble techniques in social applications. *Expert Systems with Applications, 77*, 236-246.

Araújo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. *Trends in ecology & evolution, 22*(1), 42-47.

Arceneaux, N., & Schmitz Weiss, A. (2010). Seems stupid until you try it: Press coverage of Twitter, 2006-9. *New Media & Society, 12*(8), 1262-1279.

Argamon, S., Bloom, K., Esuli, A., & Sebastiani, F. (2007). Automatically determining attitude type and force for sentiment analysis. *Proceedings of the Language and Technology Conference*, (pp. 218-231). Springer, Berlin, Heidelberg.

Arvanitis, K., & Bassiliades, N. (2017). Real-Time Investors' Sentiment Analysis from Newspaper Articles. In *Advances in Combining Intelligent Methods*, (pp. 1-23): Springer, Cham.

Atefeh, F., & Khreich, W. (2015). A survey of techniques for event detection in twitter. *Computational Intelligence, 31*(1), 132-164.

Auria, L., & Moro, R. A. (2008). Support vector machines (SVM) as a technique for solvency analysis.

Baker, M., Stein, J. C., & Wurgler, J. (2003). When does the market matter? Stock prices and the investment of equity-dependent firms. *The Quarterly Journal of Economics, 118(3)*, 969-1005.

Barbosa, L., & Feng, J. (2010). Robust sentiment detection on twitter from biased and noisy data. *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, (pp. 36-44). Association for Computational Linguistics.

Bartov, E., Faurel, L., & Mohanram, P. (2017). Can Twitter help predict firm-level earnings and stock returns?. *The Accounting Review*, *93(3),* 25-57.

Baştanlar, Y., & Özuysal, M. (2014). Introduction to machine learning. In *miRNomics: MicroRNA Biology and Computational Analysis,* (pp. 105-128). Humana Press, Totowa, NJ.

Beel, J., Breitinger, C., & Langer, S. (2017). Evaluating the CC-IDF citation-weighting scheme: how effectively can 'Inverse Document Frequency'(IDF) be applied to references. *Proceedings of the 12th iConference*.

Bermingham, M. L., Pong-Wong, R., Spiliopoulou, A., Hayward, C., Rudan, I., Campbell, H., . . . Navarro, P. (2015). Application of high-dimensional feature selection: evaluation for genomic prediction in man. *Scientific reports, 5*, 10312.

Berthon, P. R., Pitt, L. F., Plangger, K., & Shapiro, D. (2012). Marketing meets Web 2.0, social media, and creative consumers: Implications for international marketing strategy. *Business horizons, 55*(3), 261-271.

Bhardwaj, B. K., & Pal, S. (2012). Data Mining: A prediction for performance improvement using classification. *arXiv preprint arXiv:1201.3418*.

Bhattacharjee, S., Das, A., Bhattacharya, U., Parui, S. K., & Roy, S. (2015). Sentiment analysis using cosine similarity measure. *Proceedings of the 2nd IEEE International Conference on the Recent Trends in Information Systems (ReTIS)*, (pp. 27-32).

Bhattu, N., & Somayajulu, D. (2012). Semi-supervised Learning of Naive Bayes Classifier with feature constraints. *Proceedings of the 24th International Conference on Computational Linguistics,* (pp. 65-78).

Bijalwan, V., Kumar, V., Kumari, P., & Pascual, J. (2014). KNN based machine learning approach for text and document mining. *International Journal of Database Theory and Application, 7*(1), 61-70.

Bing, L., Chan, K. C., & Ou, C. (2014). Public sentiment analysis in Twitter data for prediction of a company's stock price movements. *Proceedings of the 11th IEEE International Conference on e-Business Engineering (ICEBE)*, (pp. 232-239).

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*: " O'Reilly Media, Inc.".

Blankespoor, E., Miller, B. P., & White, H. D. (2014). Initial evidence on the market impact of the XBRL mandate. *Review of Accounting Studies, 19*(4), 1468-1503.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2*(1), 1-8.

Bonde, G., & Khaled, R. (2012). Extracting the best features for predicting stock prices using machine learning. *Proceedings on the International Conference on Artificial Intelligence (ICAI),* (p. 1).

Borio, C., Gambacorta, L., & Hofmann, B. (2017). The influence of monetary policy on bank profitability. *International Finance, 20*(1), 48-63.

Bratu, C. V., Muresan, T., & Potolea, R. (2008). Improving classification accuracy through feature selection. *Proceedings of the 4th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP),* (pp. 25-32).

Brazdil, P., Carrier, C. G., Soares, C., & Vilalta, R. (2008). *Metalearning: Applications to data mining*. Springer Science & Business Media.

Cakra, Y. E., & Trisedya, B. D. (2015). Stock price prediction using linear regression based on sentiment analysis. *Proceedings of the IEEE International Conference on Advanced Computer Science and Information Systems (ICACSIS),* (pp. 147-154).

Canuto, S., Gonçalves, M. A., & Benevenuto, F. (2016). Exploiting new sentiment-based meta-level features for effective sentiment analysis. *Proceedings of the ninth ACM international conference on web search and data mining,* (pp. 53-62).

Cao, G. (2017). The Impacts of Information on Stock PricesAssessed by Social Media Sentiment. *Proceedings of the IEEE International Conference on the Internet of Things (iThings),* (pp. 1-8).

Castillo, O., Melin, P., & Pedrycz, W. (2007). *Hybrid Intelligent Systems: Analysis and Design*, (Vol. 208). Springer.

Catal, C., & Nangir, M. (2017). A sentiment classification model based on multiple classifiers. *Applied Soft Computing, 50*, 135-141.

Chakraborty, P., Pria, U. S., Rony, M. R. A. H., & Majumdar, M. A. (2017). Predicting stock movement using sentiment analysis of Twitter feed. *Proceedings of the 6th IEEE International Conference on the Informatics, Electronics and Vision (ICIEV)*, (pp. 1-6).

Chandra, S., Khan, L., & Muhaya, F. B. (2011). Estimating Twitter user location using social interactions--a content based approach. *Proceedings of the 3rd IEEE International Conference on Social Computing (SocialCom)*, (pp. 838-843).

Chandrasekar, P., & Qian, K. (2016). The Impact of Data Preprocessing on the Performance of a Naive Bayes Classifier. *Proceedings of the 40th IEEE Annual Computer Software and Applications Conference (COMPSAC),* (Vol. 2), (pp. 618-619).

Chanthinok, K., Ussahawanitichakit, P., & Jhundra-indra, P. (2015). Social media marketing strategy and marketing outcomes: A conceptual framework. *Proceedings of the Allied Academies International Conference: Academy of Marketing Studies*, (Vol. 20, No. 2), (p. 35). Jordan Whitney Enterprises, Inc.

Chaovalit, P., & Zhou, L. (2005). Movie review mining: A comparison between supervised and unsupervised classification approaches. *Proceedings of the 38th IEEE Annual Hawaii International Conference on the System Sciences*, (pp. 112c-112c).

Chen, R., Chu, T., Liu, K., Liu, J., & Chen, Y. (2015). Inferring Human Activity in Mobile Devices by Computing Multiple Contexts. *Sensors, 15*(9), 21219-21238.

Chen, S.-H., & Chen, M.-C. (2013). Addressing the advantages of using ensemble probabilistic models in estimation of distribution algorithms for scheduling problems. *International Journal of Production Economics, 141*(1), 24-33.

Chojaczyk, A., Teixeira, A., Neves, L. C., Cardoso, J., & Soares, C. G. (2015). Review and application of artificial neural networks models in reliability analysis of steel structures. *Structural Safety, 52*, 78-89.

Choudhery, D., & Leung, C. K. (2017). Social media mining: prediction of box office revenue. *Proceedings of the 21st ACM International Database Engineering & Applications Symposium*, (pp. 20-29).

Conneau, A., Kiela, D., Schwenk, H., Barrault, L., & Bordes, A. (2017). Supervised learning of universal sentence representations from natural language inference data. *arXiv preprint arXiv:1705.02364*.

Culnan, M. J., McHugh, P. J, & Zubillaga, J. I. (2010). How large U.S. companies can use Twitter and other social media to gain business value. *MIS Quarterly Executive*, *9*(4).

De Fortuny, E. J., De Smedt, T., Martens, D., & Daelemans, W. (2014). Evaluating and understanding text-based stock price prediction models. *Information Processing & Management, 50*(2), 426-441.

Deans, P. C. (2011). The impact of social media on C-level roles. *MIS Quarterly Executive, 10*(4), 187-200.

Dubai Financial Markets (2014). Dubai Financial Markets. Retrieved 7-Nov, 2016, from http://www.dfm.ae/products/ivestor.

Di Nunzio, G. M., & Sordoni, A. (2012). How well do we know Bernoulli? *IIR, 835*, 38-44.

Di Nunzio, G. M., & Sordoni, A. (2012). A visual tool for bayesian data analysis: the impact of smoothing on naive bayes text classifiers. *Proceedings of the 35th ACM SIGIR international conference on research and development in information retrieval*, (pp. 1002-1002).

Domingo, P., & Pazzani, M. (1996). Beyond independence: conditions for the optimality of the simple bayesian classier. *Proceedings of the 13th International Conference on Machine Learning*, (pp. 105-112).

Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning, 29*(2-3), 103-130.

Donmez, P., Carbonell, J. G., & Schneider, J. (2009). Efficiently learning the accuracy of labeling sources for selective sampling. *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, (pp. 259-268).

Dougherty, G. (2012). *Pattern recognition and classification: an introduction*: Springer Science & Business Media.

Duda, R. O., Hart, P. E., & Stork, D. G. (2012). *Pattern classification*. John Wiley & Sons.

Duggan, M., Ellison, N. B., Lampe, C., Lenhart, A., & Madden, M. (2015). Social media update 2014. *Pew Research Center, 9*.

Elyazji, J. (2015). Temporary migration of Syrian investments, *Alaraby*. Retrieved from https://www.alaraby.co.uk/supplementeconomy/2015/10/14.

Eyheramendy, S., Lewis, D. D., & Madigan, D. (2003). On the naive bayes model for text categorization.

Fernández-Avilés, G., Montero, J.-M., & Orlov, A. G. (2012). Spatial modeling of stock market comovements. *Finance Research Letters, 9*(4), 202-212.

Fiarni, C., Maharani, H., & Pratama, R. (2016). Sentiment analysis system for Indonesia online retail shop review using hierarchy Naive Bayes technique. *Proceedings of the 4th IEEE International Conference on Information and Communication Technology (ICoICT)*, (pp. 1-6).

Forman, G. (2003). An extensive empirical study of feature selection metrics for text classification. *Journal of Machine Learning Research, 3*(Mar), 1289-1305.

Frank, E., & Bouckaert, R. R. (2006). Naive bayes for text classification with unbalanced classes. In *European Conference on Principles of Data Mining and Knowledge Discovery*, (pp. 503-510). Springer, Berlin, Heidelberg.

Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning, 29*(2-3), 131-163.

Friedrichs, F., & Igel, C. (2005). Evolutionary tuning of multiple SVM parameters. *Neurocomputing, 64*, 107-117.

Gantz, J., & Reinsel, D. (2011). Extracting value from chaos. *IDC iview, 1142*, 1-12.

Gay, R. D. (2016). Effect of macroeconomic variables on stock market returns for four emerging economies: Brazil, Russia, India, and China. *The International Business & Economics Research Journal (Online), 15*(3), 119.

Gentry, J., Gentry, M. J., RSQLite, S., & Artistic, R. L. (2016). Package 'twitteR'. *R package version, 1*(9).

Geser, H. (2011). Has Tweeting become Inevitable? Twitter's strategic role in the World of Digital Communication. *Sociology In Switzerland: Towards Cybersociety and Vireal Social Relations.*

Ghorpade, T., & Ragha, L. (2012). Featured based sentiment classification for hotel reviews using NLP and Bayesian classification. In *the IEEE International Conference on Communication, Information & Computing Technology (ICCICT)*, (pp. 1-5).

Giatsoglou, M., Vozalis, M. G., Diamantaras, K., Vakali, A., Sarigiannidis, G., & Chatzisavvas, K. C. (2017). Sentiment analysis leveraging emotions and word embeddings. *Expert Systems with Applications, 69*, 214-224.

Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford, 1*, 12.

Gong, Z., & Yu, T. (2010). Chinese Web Text Classification System Model Based on Naive Bayes. In *the IEEE International Conference on E-Product E-Service and E-Entertainment (ICEEE)*, (pp. 1-4).

Gregg, D. G. (2017). Why use SVM?. Retrieved 24-Feb, 2018, from http://blog.yhat.com/posts/why-support-vector-machine.html.

GulfNews. (2015, November 7). GCC News. Retrieved 7-Nov, 2016, from http://gulfnews.com/news/uae.

Guresen, E., Kayakutlu, G., & Daim, T. U. (2011). Using artificial neural network models in stock market index prediction. *Expert Systems with Applications, 38*(8), 10389-10397.

Hamed, A.-R., Qiu, R., & Li, D. (2015). Analysis of the relationship between Saudi twitter posts and the Saudi stock market. In *the IEEE Seventh International Conference on Intelligent Computing and Information Systems (ICICIS)*, (pp. 660-665).

Hamed, A.-R., Qiu, R., & Li, D. (2016). The importance of neutral class in sentiment analysis of Arabic tweets. *Int. J. Comput. Sci. Inform. Technol, 8*, 17-31.

Han, W., Nan-feng, X., & Zhao, L. (2011). An enhanced EM method of semi-supervised classification based on Naive Bayesian. In *the Eighth IEEE International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, (Vol. 2), (pp. 987-991).

Hardeniya, N., Perkins, J., Chopra, D., Joshi, N., & Mathur, I. (2016). *Natural Language Processing: Python and NLTK*. Packt Publishing Ltd.

Har-Peled, S., Roth, D., & Zimak, D. (2002). Constraint classification: A new approach to multiclass classification. *Advances in neural information processing systems, 15*.

Harrington, P. (2012). *Machine learning in action*, (Vol. 5): Manning Greenwich, CT.

Hasbullah, S. S., Maynard, D., Chik, R. Z. W., Mohd, F., & Noor, M. (2016). Automated Content Analysis: A Sentiment Analysis on Malaysian Government Social Media. *Proceedings of the 10th ACM International Conference on Ubiquitous Information Management and Communication*, (p. 30).

He, Y., & Zhou, D. (2011). Self-training from labeled features for sentiment analysis. *Information Processing & Management, 47*(4), 606-616.

Hellerstein, J. L., Jayram, T., & Rish, I. (2000). *Recognizing end-user transactions in performance management*. IBM Thomas J. Watson Research Division.

Hiemstra, D. (2000). A probabilistic justification for using tf× idf term weighting in information retrieval. *International Journal on Digital Libraries, 3*(2), 131-139.

Hong, L., Ahmed, A., Gurumurthy, S., Smola, A. J., & Tsioutsiouliklis, K. (2012). Discovering geographical topics in the twitter stream. *Proceedings of the 21st ACM international conference on World Wide Web*, (pp. 769-778).

Hsu, M.-W., Lessmann, S., Sung, M.-C., Ma, T., & Johnson, J. E. (2016). Bridging the divide in financial market forecasting: machine learners vs. financial economists. *Expert Systems with Applications, 61*, 215-234.

Hu, J. (2017). *Sentiment Analysis on Social Media Platforms.* (Bachelors Research Project), The University of Arizona.

Huang, Y., & Li, L. (2011). Naive Bayes classification algorithm based on small sample set. In *the IEEE International Conference on Cloud Computing and Intelligence Systems*, (pp. 34-39).

Hung, C., & Chen, J.-H. (2009). A selective ensemble based on expected probabilities for bankruptcy prediction. *Expert Systems with Applications, 36*(3), 5297-5303.

Hussien, W., Tashtoush, Y. M., Al-Ayyoub, M., & Al-Kabi, M. N. (2016). Are emoticons good enough to train emotion classifiers of Arabic tweets?. *CSIT. IEEE*, 1-6.

Iacomin, R. (2016). Feature optimization on stock market predictor. In *the IEEE International Conference on Development and Application Systems (DAS)*, (pp. 243-247).

Ifrim, G., Shi, B., & Brigadir, I. (2014). Event detection in twitter using aggressive filtering and hierarchical tweet clustering. In *the Second Workshop on Social News on the Web (SNOW)*, (pp. 33-40). Seoul, Korea.

Imandoust, S. B., & Bolandraftar, M. (2013). Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background. *International Journal of Engineering Research and Applications, 3*(5), 605-610.

Islam, M. J., Wu, Q. J., Ahmadi, M., & Sid-Ahmed, M. A. (2007). Investigating the performance of naive-bayes classifiers and k-nearest neighbor classifiers. In *the IEEE International Conference on Convergence Information Technology (ICCIT 2007)*, (pp. 1541-1546).

Jain, A., & Mandowara, J. (2016). Text classification by combining text classifiers to improve the efficiency of classification. *International Journal of Computer Application (2250-1797), 6*(2).

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112), (p. 18). New York. Springer.

Janssen, M., van der Voort, H., & Wahyudi, A. (2017). Factors influencing big data decision-making quality. *Journal of Business Research, 70*, 338-345.

Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we twitter: understanding microblogging usage and communities. *Proceedings of the 9th WebKDD and 1st SNA-KDD workshop on Web mining and social network analysis*.

Jeon, S., Hong, B., & Chang, V. (2017). Pattern graph tracking-based stock price prediction using big data. *Future Generation Computer Systems*, *80*, 171-187

Jha, N., & Mahmoud, A. (2017). Mining user requirements from application store reviews using frame semantics. In *the International Working Conference on Requirements Engineering: Foundation for Software Quality,* (pp. 273-287). Springer, Cham.

Jiang, L., Wang, D., Cai, Z., & Yan, X. (2007). Survey of improving naive Bayes for classification. In *the International Conference on Advanced Data Mining and Applications,* (pp. 134-145). Springer, Berlin, Heidelberg.

Kafai, Y. B., Fields, D. A., & Burke, W. Q. (2010). Entering the clubhouse: Case studies of young programmers joining the online Scratch communities. *Journal of Organizational and End User Computing (JOEUC), 22*(2), 21-35.

Kaji, N., & Kitsuregawa, M. (2006). Automatic construction of polarity-tagged corpus from HTML documents. *Proceedings of the COLING/ACL on Main conference poster sessions*, (pp. 452-459). Association for Computational Linguistics.

Kanayama, H., & Nasukawa, T. (2006). Fully automatic lexicon expansion for domain-oriented sentiment analysis. *Proceedings of the 2006 conference on empirical methods in natural language processing*, (pp. 355-363). Association for Computational Linguistics.

Kazienko, P., Lughofer, E., & Trawiński, B. (2013). Hybrid and ensemble methods in machine learning J. UCS special issue. *J Univers Comput Sci, 19*(4), 457-461.

Kesavaraj, G., & Sukumaran, S. (2013). A study on classification techniques in data mining. In *the Fourth IEEE International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, (pp. 1-7).

Khaidem, L., Saha, S., & Dey, S. R. (2016). Predicting the direction of stock market prices using random forest. *arXiv preprint arXiv:1605.00003*.

Khan, K., Baharudin, B., Khan, A., & Ullah, A. (2014). Mining opinion components from unstructured reviews: A review. *Journal of King Saud University-Computer and Information Sciences, 26*(3), 258-275.

Kharde, V., & Sonawane, S. (2016). Sentiment Analysis of Twitter Data: A Survey of Techniques. *arXiv preprint arXiv:1601.06971*.

Khreisat, L. (2009). A machine learning approach for Arabic text classification using N-gram frequency statistics. *Journal of Informetrics, 3*(1), 72-77.

Kim, S.-M., & Hovy, E. (2004). Determining the sentiment of opinions. *Proceedings of the 20th international conference on Computational Linguistics*, (p. 1367). Association for Computational Linguistics.

Kim, Y., Jeong, S. R., & Ghani, I. (2014). Text opinion mining to analyze news for stock market prediction. *Int. J. Adv. Soft Comput. Its Appl, 6*(1).

Kohavi, R. (1995). *Wrappers for performance enhancement and oblivious decision graphs.* Carnegie-Mellon, Univ Pittsburgh Pa, Dept Of Computer Science.

Kokalj-Filipovic, S., Greco, M., Poor, H., Stantchev, G., & Xiao, L. (2018). Introduction to the Issue on Machine Learning for Cognition in Radio Communications and Radar. *IEEE Journal of Selected Topics in Signal Processing, 12*(1), 3-5.

Kolchyna, O., Souza, T. T., Treleaven, P., & Aste, T. (2015). Twitter sentiment analysis. *arXiv preprint arXiv:1507.00955*.

Kordonis, J., Symeonidis, S., & Arampatzis, A. (2016). Stock price forecasting via sentiment analysis on Twitter. *Proceedings of the 20th ACM Pan-Hellenic Conference on Informatics*, (p. 36).

Korjus, K., Hebart, M. N., & Vicente, R. (2016). An efficient data partitioning to improve classification performance while keeping parameters interpretable. *PloS one, 11*(8), e0161788.

Kouloumpis, E., Wilson, T., & Moore, J. D. (2011). Twitter sentiment analysis: The good the bad and the omg! *ICWSM, 11*(538-541), 164.

Kshirsagar, M., Thomson, S., Schneider, N., Carbonell, J., Smith, N. A., & Dyer, C. (2015). Frame-semantic role labeling with heterogeneous annotations. *people, 3*, A0.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*, (Vol. 26). New York. Springer.

Kumar, S., Morstatter, F., & Liu, H. (2014). *Twitter data analytics,* (pp. 1041-4347). *New York*. Springer.

Kumar, V., Choi, J. B., & Greene, M. (2017). Synergistic effects of social media and traditional marketing on brand sales: capturing the time-varying effects. *Journal of the Academy of Marketing Science, 45*(2), 268-288.

Kuncheva, L. I., Bezdek, J. C., & Duin, R. P. (2001). Decision templates for multiple classifier fusion: an experimental comparison. *Pattern Recognition, 34*(2), 299-314.

Last, M., Tassa, T., Zhmudyak, A., & Shmueli, E. (2014). Improving accuracy of classification models induced from anonymized datasets. *Information Sciences, 256*, 138-161.

Lee, J., & Kim, D.-W. (2017). SCLS: Multi-label feature selection based on scalable criterion for large label set. *Pattern Recognition, 66,* 342-352.

Lee, L. F., Hutton, A. P., & Shu, S. (2015). The role of social media in the capital market: evidence from consumer product recalls. *Journal of Accounting Research, 53*(2), 367-404.

Li, R., Lei, K. H., Khadiwala, R., & Chang, K. C.-C. (2012). Tedas: A twitter-based event detection and analysis system. In *the IEEE 28th International Conference on Data Engineering*, (pp. 1273-1276).

Li, X., Xie, H., Wang, R., Cai, Y., Cao, J., Wang, F., . . . Deng, X. (2016). Empirical analysis: stock market prediction via extreme learning machine. *Neural Computing and Applications, 27*(1), 67-78.

Lin, J., & Ryaboy, D. (2013). Scaling big data mining infrastructure: the twitter experience. *ACM SIGKDD Explorations Newsletter, 14*(2), 6-19.

Liu, B. (2010). Sentiment Analysis and Subjectivity. *Handbook of natural language processing, 2*, 627-666.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies, 5*(1), 1-167.

Liu, B., Blasch, E., Chen, Y., Shen, D., & Chen, G. (2013). Scalable sentiment classification for big data analysis using Naive Bayes Classifier. In *the IEEE International Conference on Big Data*, (pp. 99-104).

Livingstone, D. J., Manallack, D. T., & Tetko, I. V. (1997). Data modelling with neural networks: advantages and limitations. *Journal of computer-aided molecular design, 11*(2), 135-142.

Louridas, P., & Ebert, C. (2013). Embedded analytics and statistics for big data. *IEEE software, 30*(6), 33-39.

Ludvigson, S. C., & Steindel, C. (1998). *How important is the stock market effect on consumption?*. (Vol. 9821). Federal Reserve Bank of New York. New York.

Ludwig, S., De Ruyter, K., Friedman, M., Brüggen, E. C., Wetzels, M., & Pfann, G. (2013). More than words: The influence of affective content and linguistic style matches in online reviews on conversion rates. *Journal of Marketing, 77*(1), 87-103.

Lughofer, E. (2011). *Evolving fuzzy systems-methodologies, advanced concepts and applications*, (Vol. 53). Springer.

Lughofer, E., Eitzinger, C., & Guardiola, C. (2012). Online quality control with flexible evolving fuzzy systems. In *Learning in Non-Stationary Environments*, (pp. 375-406). Springer.

Luo, X., Dong, L., Dou, Y., Zhang, N., Ren, J., Li, Y., . . . Yao, S. (2017). Analysis on spatial-temporal features of taxis' emissions from big data informed travel patterns: a case of Shanghai, China. *Journal of Cleaner Production, 142*, 926-935.

Mahajan Shubhrata, D., Deshmukh Kaveri, V., Thite Pranit, R., Samel Bhavana, Y., & Chate, P. (2016). Stock Market Prediction and Analysis Using Naïve Bayes. *International Journal on Recent and Innovation Trends in Computing and Communication, 4*(11), 121-124.

Majhi, R., Panda, G., Sahoo, G., Dash, P. K., & Das, D. P. (2007). Stock market prediction of S&P 500 and DJIA using bacterial foraging optimization technique. In *the IEEE congress on evolutionary computation*, (pp. 2569-2575).

Makice, K. (2009). *Twitter API: Up and running: Learn how to build applications with the Twitter API*: " O'Reilly Media, Inc.".

Makrehchi, M., Shah, S., & Liao, W. (2013). Stock prediction using event-based sentiment analysis. In *the IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, Volume 01, (pp. 337-342).

Manek, A. S., Shenoy, P. D., Mohan, M. C., & Venugopal, K. (2017). Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier. *World wide web, 20*(2), 135-154.

Marsland, S. (2015). *Machine learning: An algorithmic perspective* (Second ed.): CRC press.

McCallum, A., & Nigam, K. (1998). A comparison of event models for naive bayes text classification. *In the AAAI-98 workshop on learning for text categorization*, (Vol. 752, No. 1), (pp. 41-48).

Meesad, P., & Li, J. (2014). Stock trend prediction relying on text mining and sentiment analysis with tweets. In *the 4th IEEE World Congress on Information and Communication Technologies (WICT)*, (pp. 257-262).

Melin, P., Castillo, O., Ramírez, E. G., & Pedrycz, W. (2007). *Analysis and design of intelligent systems using soft computing techniques*, (Vol. 41). Springer Science & Business Media.

Melville, P., Gryc, W., & Lawrence, R. D. (2009). Sentiment analysis of blogs by combining lexical knowledge with text classification. *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, (pp. 1275-1284).

Menezes, S. L., & Varkey, G. (2013). Prediction of Missing Items Using Naive Bayes Classifier and Graph Based Prediction. In *the Third IEEE International Conference on Advances in Computing and Communications (ICACC),* (pp. 39-45).

Michaelidou, N., Siamagka, N. T., & Christodoulides, G. (2011). Usage, barriers and measurement of social media marketing: An exploratory investigation of small and medium B2B brands. *Industrial marketing management, 40*(7), 1153-1159.

Michie, D., Spiegelhalter, D. J., & Taylor, C. C. (1994). Machine learning, neural and statistical classification. *Neural and Statistical Classification, 13*.

Miranda, F., & Abreu, C. (2016). *Handbook of Research on Computational Simulation and Modeling in Engineering*. Engineering Science Reference.

Mitchell.T.M. (1997). Machine Learning. *McGraw-Hill*, 1, 27.

Moraes, R., Valiati, J. F., & Neto, W. P. G. (2013). Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications, 40*(2), 621-633.

Mund, S. (2015). *Microsoft azure machine learning*: Packt Publishing Ltd.

Naaman, M., Boase, J., & Lai, C.-H. (2010). Is it really about me?: message content in social awareness streams. *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, (pp. 189-192).

Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: an introduction. *Journal of the American Medical Informatics Association, 18*(5), 544-551.

Nagwani, N. K., & Verma, S. (2014). A Comparative Study of Bug Classification Algorithms. *International Journal of Software Engineering and Knowledge Engineering, 24*(01), 111-138.

Narayanan, V., Arora, I., & Bhatia, A. (2013). Fast and accurate sentiment classification using an enhanced Naive Bayes model. *In the International Conference on Intelligent Data Engineering and Automated Learning*, (pp. 194-201). Springer, Berlin, Heidelberg.

Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., & Ngo, D. C. L. (2014). Text mining for market prediction: A systematic review. *Expert Systems with Applications, 41*(16), 7653-7670.

Navale, G., Dudhwala, N., Jadhav, K., Gabda, P., & Vihangam, B. K. (2016). Prediction of Stock Market Using Data Mining and Artificial Intelligence. *International Journal of Engineering Science, 6539*.

Navigli, R. (2009). Word sense disambiguation: A survey. *ACM Computing Surveys (CSUR), 41*(2), 10.

Netti, K., & Radhika, Y. (2015). A novel method for minimizing loss of accuracy in Naive Bayes classifier. In *the IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, (pp. 1-4).

Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications, 42*(24), 9603-9611.

Nichols, J., Mahmud, J., & Drews, C. (2012). Summarizing sporting events using Twitter. *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*, (pp. 189-198).

Novalić, A. (2013). Introduction to Tweepy. Retrieved 3-May, 2017, from https://www.pythoncentral.io/introduction-to-tweepy-twitter-for-python/.

Oliveira, N., Cortez, P., & Areal, N. (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, 73*, 125-144.

Padmaja, S., & Fatima, S. S. (2013). Opinion Mining and Sentiment Analysis-An Assessment of Peoples' Belief: A Survey. *International Journal of Ad Hoc, Sensor & Ubiquitous Computing, 4*(1), 21.

Pan, L., & Mishra, V. (2018). Stock market development and economic growth: Empirical evidence from China. *Economic Modelling, 68*, 661-673.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval, 2*(1-2), 1-135.

Perrin, A. (2015). Social media usage. *Pew Research Center*, 52-68.

Podgorelec, V., Kokol, P., Stiglic, B., & Rozman, I. (2002). Decision trees: an overview and their use in medicine. *Journal of medical systems, 26*(5), 445-463.

Provost, F. J., Fawcett, T., & Kohavi, R. (1998). The case against accuracy estimation for comparing induction algorithms. In *the ICML*, (Vol. 98), (pp. 445-453).

Prusa, J., Khoshgoftaar, T. M., & Dittman, D. J. (2015). Using ensemble learners to improve classifier performance on tweet sentiment data. In *the IEEE International Conference on Information Reuse and Integration (IRI)*, (pp. 252-257).

Prusa, J. D., Khoshgoftaar, T. M., & Dittman, D. J. (2015). Impact of Feature Selection Techniques for Tweet Sentiment Classification. In *the Twenty-Eighth International FLAIRS Conference*.

Prusa, J. D., Khoshgoftaar, T. M., & Napolitano, A. (2015). Using feature selection in combination with ensemble learning techniques to improve tweet sentiment classification performance. In *the 27th IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*, (pp. 186-193).

Prusa, J. D., Khoshgoftaar, T. M., & Seliya, N. (2016). Enhancing Ensemble Learners with Data Sampling on High-Dimensional Imbalanced Tweet Sentiment Data. In *the twenty-ninth international the FLAIRS Conference*.

Qasem, M., Thulasiram, R., & Thulasiram, P. (2015). Twitter sentiment classification using machine learning techniques for stock markets. In *the IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, (pp. 834-840).

Raschka, S. (2014). Naive bayes and text classification i-introduction and theory. *arXiv preprint arXiv:1410.5329*.

Raschka, S., & Mirjalili, V. (2017). *Python Machine Learning*: Packt Publishing Ltd.

Ren, R., Wu, D. D., & Liu, T. (2018). Forecasting Stock Market Movement Direction Using Sentiment Analysis and Support Vector Machine. *IEEE Systems Journal*.

Rennie, J. D., Shih, L., Teevan, J., & Karger, D. R. (2003). Tackling the poor assumptions of naive bayes text classifiers. *Proceedings of the 20th international conference on machine learning (ICML-03)*, (pp. 616-623).

Roesslein, J. (2009). Tweepy Documentation. Retrieved 6-May, 2017, from http://tweepy.readthedocs.io/en/v3.5.0/getting_started.html#introduction.

Rosenthal, S., Farra, N., & Nakov, P. (2017). SemEval-2017 task 4: Sentiment analysis in Twitter. *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, (pp. 502-518).

Rout, J. K., Choo, K.-K. R., Dash, A. K., Bakshi, S., Jena, S. K., & Williams, K. L. (2018). A model for sentiment and emotion analysis of unstructured social media text. *Electronic Commerce Research, 18*(1), 181-199.

Ruan, Y., Durresi, A., & Alfantoukh, L. (2018). Using Twitter trust network for stock market analysis. *Knowledge-Based Systems, 145*, 207-218.

Ruiz, E. J., Hristidis, V., Castillo, C., Gionis, A., & Jaimes, A. (2012). Correlating financial time series with micro-blogging activity. *Proceedings of the fifth ACM international conference on Web search and data mining*, (pp. 513-522).

Russell, M. A. (2013). *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*: " O'Reilly Media, Inc.".

Ruth, M., & Hannon, B. (2012). Positive Feedback in the Economy. In *Modeling Dynamic Economic Systems* (pp. 71-76). Springer, New York, NY.

Santos, M. A. (2015). *Integrated Ownership and Control in the GCC Corporate Sector*: International Monetary Fund.

Sarkar, B. K., & Sana, S. S. (2009). A hybrid approach to design efficient learning classifiers. *Computers & Mathematics with Applications, 58*(1), 65-73.

Sathyadevan, S., Sarath, P. R., Athira, U., & Anjana, V. (2014). Improved document classification through enhanced Naive Bayes algorithm. In *the IEEE International Conference on Data Science & Engineering (ICDSE)*, (pp. 100-104).

Schmid, H. (2013). Probabilistic part-ofispeech tagging using decision trees. In *the New methods in language processing*, (p. 154).

Schumaker, R. P., Zhang, Y., Huang, C.-N., & Chen, H. (2012). Evaluating sentiment in financial news articles. *Decision Support Systems, 53*(3), 458-464.

Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., & Lichtendahl Jr, K. C. (2017). *Data mining for business analytics: concepts, techniques, and applications in R*: John Wiley & Sons.

Shoeb, M., & Ahmed, J. (2017). Sentiment Analysis and Classification of Tweets Using Data Mining. *work, 4*(12).

Skuza, M., & Romanowski, A. (2015). Sentiment analysis of Twitter data within big data distributed environment for stock prediction. *In the IEEE Federated Conference on Computer Science and Information Systems (FedCSIS)*, (pp. 1349-1354).

Smedt, T. D., & Daelemans, W. (2012). Pattern for python. *Journal of Machine Learning Research, 13*(Jun), 2063-2067.

Smith, A. (2010). Government online: The internet gives citizens new paths to government services and information. *Pew Internet & American Life Project*.

Sohangir, S., Wang, D., Pomeranets, A., & Khoshgoftaar, T. M. (2018). Big Data: Deep Learning for financial sentiment analysis. *Journal of Big Data, 5*(1), 3.

Song, Z., & Xia, J. C. (2016). Spatial and Temporal Sentiment Analysis of Twitter data. *European Handbook of Crowdsourced Geographic Information*, 205.

Srivastava, A., Han, E.-H., Kumar, V., & Singh, V. (1999). Parallel formulations of decision-tree classification algorithms. *High Performance Data Mining*, (pp. 237-261). Springer.

Sul, H., Dennis, A. R., & Yuan, L. I. (2014). Trading on Twitter*:* The financial information content of emotion in social media*.* In *the 47th IEEE Hawaii International Conference on System Sciences*, (pp. 806-815).

Sulthana, A. R., Jaithunbi, A., & Ramesh, L. S. (2018). Sentiment analysis in twitter data using data analytic techniques for predictive modelling*.* In *the Journal of Physics*: *Conference Series*, (Vol. 1000, No. 1), (p. 012130). IOP Publishing.

Tan, P.-N., Steinbach, M., & Kumar, V. (2006). Classification: basic concepts, decision trees, and model evaluation. *Introduction to data mining, 1*, 145-205.

Tan, S., & Zhang, J. (2008). An empirical study of sentiment analysis for chinese documents. *Expert Systems with Applications, 34*(4), 2622-2629.

Tan, T. Z., Quek, C., & Ng, G. S. (2005). Brain-inspired genetic complementary learning for stock market prediction*.* In *the IEEE Congress on Evolutionary Computation*, (Vol. 3), (pp. 2653-2660).

Tang, B., He, H., Baggenstoss, P. M., & Kay, S. (2016). A Bayesian classification approach using class-specific features for text categorization. *IEEE Transactions on Knowledge and Data Engineering, 28*(6), 1602-1606.

Tang, H., Tan, S., & Cheng, X. (2009). A survey on sentiment detection of reviews. *Expert Systems with Applications, 36*(7), 10760-10773.

Tanwani, A. K., Afridi, J., Shafiq, M. Z., & Farooq, M. (2009). Guidelines to select machine learning scheme for classification of biomedical datasets. In *the European Conference on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics*, (pp. 128-139). Springer, Berlin, Heidelberg.

Tanwani, A. K., & Farooq, M. (2010). Classification potential vs. classification accuracy: a comprehensive study of evolutionary algorithms with biomedical datasets. In *Learning Classifier Systems*, (pp. 127-144): Springer.

Thompson, C. (2008). Brave new world of digital intimacy. *The New York Times, 7*.

Ticknor, J. L. (2013). A Bayesian regularized artificial neural network for stock market forecasting. *Expert Systems with Applications, 40*(14), 5501-5506.

Ting, S., Ip, W., & Tsang, A. H. (2011). Is Naive Bayes a good classifier for document classification. *International Journal of Software Engineering and Its Applications, 5*(3), 37-46.

Tripathy, A., Agrawal, A., & Rath, S. K. (2016). Classification of sentiment reviews using n-gram machine learning approach. *Expert Systems with Applications, 57*, 117-126.

Trupthi, M., Pabboju, S., & Narasimha, G. (2017). Sentiment analysis on twitter using streaming API. In *the 7th IEEE International Advance Computing Conference (IACC)*, (pp. 915-919).

Tugores, A., & Colet, P. (2014). Mining online social networks with Python to study urban mobility. *arXiv preprint arXiv:1404.6966*.

Turney, P. D., & Littman, M. L. (2002). Unsupervised learning of semantic orientation from a hundred-billion-word corpus. *arXiv preprint cs/0212012*.

Uusitalo, L. (2007). Advantages and challenges of Bayesian networks in environmental modelling. *Ecological modelling, 203*(3-4), 312-318.

Vinodhini, G., & Chandrasekaran, R. (2012). Sentiment analysis and opinion mining: a survey. *International Journal, 2*(6).

Vu, T.-T., Chang, S., Ha, Q. T., & Collier, N. (2012). An experiment in integrating sentiment features for tech stock prediction in twitter. *Proceedings of the workshop on information extraction and entity analytics on social media data*, (pp. 23-38).

Wang, H., Wu, J., Zhang, P., & Zhang, C. (2016). Temporal Feature Selection on Networked Time Series. *arXiv preprint arXiv:1612.06856*.

Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., & Gong, Y. (2010). Locality-constrained linear coding for image classification. In *the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (pp. 3360-3367).

Weise, K., & Woger, W. (1994). Comparison of two measurement results using the Bayesian theory of measurement uncertainty. *Measurement Science and Technology, 5*(8), 879.

Welpe, I., & Sprenger, T. (2010). Tweets and Trades: The Information Content of Stock Microblogs. *SSRN eLibrary*.

Whitelaw, C., Garg, N., & Argamon, S. (2005). Using appraisal groups for sentiment analysis. *Proceedings of the 14th ACM international conference on Information and knowledge management*, (pp. 625-631).

Wijeratne, S., Sheth, A., Bhatt, S., Balasuriya, L., Al-Olimat, H. S., Gaur, M., . . . Thirunarayan, K. (2017). Feature Engineering for Twitter-based Applications. *Feature Engineering for Machine Learning and Data Analytics*, 35.

Williams, J., & Dagli, C. (2017). Twitter language identification of similar languages and dialects without ground truth. *Proceedings of the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial),* (pp. 73-83).

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Xia, R., Zong, C., & Li, S. (2011). Ensemble of feature sets and classification algorithms for sentiment classification. *Information Sciences, 181*(6), 1138-1152.

Xu, F., & Keelj, V. (2014). Collective Sentiment Mining of Microblogs in 24-Hour Stock Price Movement Prediction. In *the 16th IEEE Conference on Business Informatics*, (Vol. 2), (pp. 60-67).

Yang, A., Zhang, J., Pan, L., & Xiang, Y. (2015). Enhanced Twitter Sentiment Analysis by Using Feature Selection and Combination. In *the International Symposium on Security and Privacy in Social Networks and Big Data (SocialSec),* (pp. 52-57).

Yengi, Y. K., Karayel, M., & Omurca, S. İ. (2015). An Alternative Method for Sentiment Classification with Expectation Maximization and Priority Aging. *International Journal of Scientific Research in Information Systems and Engineering (IJSRISE), 1*(2), 91-96.

Yong, S., TANG, Y.-r., CUI, L.-x., & Wen, L. (2018). A text mining based study of investor sentiment and its influence on stock returns. *Economic Computation & Economic Cybernetics Studies & Research, 52*(1).

Yousef, M., Saçar Demirci, M. D., Khalifa, W., & Allmer, J. (2016). Feature Selection Has a Large Impact on One-Class Classification Accuracy for MicroRNAs in Plants. *Advances in bioinformatics, 2016*.

Yousif, A. (2014). Increase the migration of Iraqi capital (Press release). Retrieved from http://www.aljazeera.net/news/ebusiness/2014/9/16.

Yu, L., Zhou, R., Tang, L., & Chen, R. (2018). A DBN-based resampling SVM ensemble learning paradigm for credit classification with imbalanced data. *Applied Soft Computing, 69*, 192-202.

Yu, L.-C., Wu, J.-L., Chang, P.-C., & Chu, H.-S. (2013). Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news. *Knowledge-Based Systems, 41*, 89-97.

Zakka's, K. (2016). A Complete Guide to K-Nearest-Neighbors with Applications in Python and R. Retrieved 24-Feb, 2018, from https://kevinzakka.github.io/2016/07/13/k-nearest-neighbor/

Zhang, L. (2013). *Sentiment analysis on Twitter with stock price and significant keyword correlation* (Doctoral dissertation). Available from The University of Texas at Austin, Department of Computer Science.

Zhang, W., Yoshida, T., & Tang, X. (2011). A comparative study of TF* IDF, LSI and multi-words for text classification. *Expert Systems with Applications, 38*(3), 2758-2765.

Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting stock market indicators through twitter "I hope it is not as bad as I fear". *Procedia-Social and Behavioral Sciences, 26*, 55-62.

Zhang, Z., Ye, Q., Zhang, Z., & Li, Y. (2011). Sentiment classification of Internet restaurant reviews written in Cantonese. *Expert Systems with Applications, 38*(6), 7674-7682.

Zhong, X., & Enke, D. (2017). Forecasting daily stock market return using dimensionality reduction. *Expert Systems with Applications, 67*, 126-139.

Zhou, L., Zhang, P., & Zimmerman, H. (2011). Call for papers for a series of special issues: Social commerce. *Electronic Commerce Research and Applications*.

Zhou, Z., Wen, C., & Yang, C. (2015). Fault detection using random projections and k-nearest neighbor rule for semiconductor manufacturing processes. *IEEE Transactions on Semiconductor Manufacturing, 28*(1), 70-79.

Zhou, Z.-H. (2012). *Ensemble methods: foundations and algorithms*: Chapman and Hall/CRC.

Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing, 74*(2), 133-148.