Thesis Proposal Ba Economie & Bedrijfseconomie

|  |
| --- |
| General Information |
| Student Name **Menno van Dijk** ANR **7 2 4 1 7 1** Phone# **0683900014** email **m.a.c.vandijk@tilburguniversity.edu**  Supervisor **prof. dr. E.C.M. van der Heijden** Date **03/03/2017** |
| Proposal information |
| (Working) Title of the thesis Twitter Sentiment on Trump and its Effect on Financial Markets |
| Problem Identification / Motivation for projectPrior research in combining social media analytics with the stock market has examined how Twitter can help predict firm-level earnings and stock returns (see, e.g., Bollen, Mao, & Zeng, 2011; Bartov, Faurel, & Mohanram, 2016). Additionally, Kearney and Liu (2014) provide a very concrete overview of all the research that has been done on textual analysis in finance. While the latter focusses more on earnings press releases, 10-Ks and a variety of other corporate news-outlets, the main focus in my paper will be on Twitter data.Expanding on the literature, I will research what the effect of Twitter sentiment is when these tweets are filtered into a single category, namely the 45th U.S. President, Donald J. Trump. Since these presidential elections were very recent, little to none research has been done on the effect of the Tweets concerning Trump on stock market prices. Furthermore, when taking into account that Trump himself is an active user of the social media platform, averaging around 18.4 tweets per day, the effect of his tweets on general sentiment (more on how to measure this later) might affect stock market pricing.In a sense, the tweets sent on Twitter will be used as a proxy to measure sentiment regarding Trump. Hopefully the paper will find a statistically significant relationship between tweets sent and stock market prices. |
| Research QuestionDoes sentiment surrounding president Trump on Twitter affect stock market pricing? |
| Proposed methodologyI will use an empirical model to analyze the data, it is a necessity in order to see the correlation between tweets and stock pricing. |
| Methodology : Empirical researchIn order to get the Twitter data, the paper will make use of the Twitter Streaming API in order to load in the tweets real-time. A program has been written in Python in order to tap in to this API, and to get all of the tweets in live. The main benefit from gathering the data in this way instead of doing it via third-party providers is that it remains a very flexible process, if any adjustments need to be made it can be done without a lot of effort. Stock market data will be loaded in directly using the Google Finance API, it is already coded into the existing program.In order to extract sentiment from a tweet I will use a similar classification system as used in Narayanan, Arora and Bhatia (2013). This classifier is provided for free by the TextBlob module in Python. In addition to this module, I will also use a Naïve Bayes Classifier that I have trained myself on a subset of tweets. The two different classification systems will run alongside each other and the data will hence stem from both of these classifiers. I will digress further on the classifiers used in the paper, since they are not quite the same and might therefore yield different results.Sentiment will eventually be extracted from a tweet, and if it meets the threshold of being negative/positive enough with sufficient certainty, it will be labelled as such. The threshold of labelling a tweet positive or negative will depend on a measure called polarity, this is a score ranging from [-1, 1], where a sentence with a polarity of 1 would be a purely positive sentence. The threshold of determining whether a sentence will be classified as positive or negative will be algorithmically determined, in order to induce the highest accuracy of the classifiers at hand.Eventually, when all of the tweets and stock prices for a given trading day are gathered, they will be printed onto an Excel sheet for further analysis. Multiple other variables that determine stock market price fluctuations can then be added in order to perform various OLS regressions. |
| Methodology: Theory modelsNot applicable |
| Literature studyNot applicable. |
| Relevant literatureI mentioned all of the most important literature already in section 1.1, when I do start writing my thesis however the list will most likely become a lot bigger. The papers mentioned already are most applicable and have steered me in the direction I am heading in right now, hence their mention. |
| Reference list (use APA system of referencing)  * Bartov, E., Faurel, L., Mohanram, P. S. (2016). Can Twitter Help Predict Firm-Level Earnings and Stock Returns?. *Rotman School of Management Working Paper No. 2631421.* * Bollen, J., Mao, H., Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2(1), 1-8.* * Kearney, C., Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis, 33, 171-185.* * Narayanan, V., Arora, I., Bhatia, A. (2013). Fast and Accurate Sentiment Classification Using an Enhanced Naive Bayes Model. *Intelligent Data Engineering and Automated Learning IDEAL*, *8206, 194-201.* |
| **Note or remark**  For any questions regarding any part of the proposal, I’ll be more than willing to answer them! If something is unclear or written down incorrectly I would love to hear about it.  Thank you again for the meeting we had earlier, I appreciated it greatly and it was of great help. |