Sentiment Analysis with a Focus on Stock Market Trend Prediction

Student Name: Aaron Meade
Student ID:
Course Code: LM051
Supervisor: Annette McElligott

Outline

What is Sentiment Analysis?

Foundation for my project

A Review of Sentiment Analysis

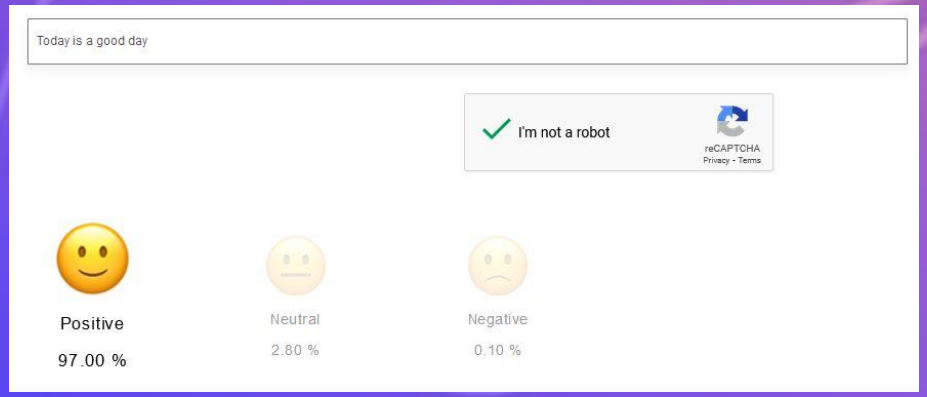
What I am proposing to do

How I am going to do that

What Is Sentiment Analysis?

- “Sentiment Analysis (SA) or Opinion Mining (OM) is the computational study of people’s opinions, attitudes, and emotions toward an entity. The entity can represent individuals, events, or topics.” (Medhat W et al, 2014)

- Customer feedback, survey responses, product reviews, social media monitoring and reputation management.



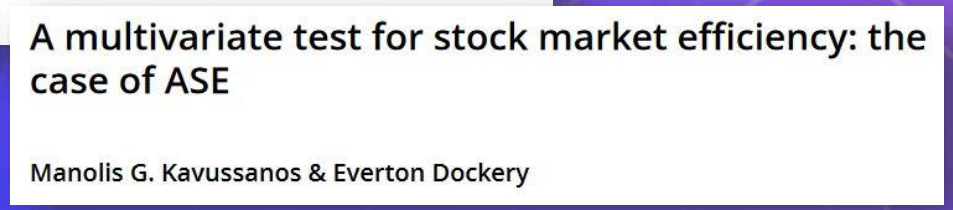
(https://komprehend.io/sentiment-analysis)

Foundation for my project

- Random walk algorithm
- The efficient market hypothesis states that when new information comes into the market, it is immediately reflected in stock price.
- social media = access to public emotion
- In the case of my project applying sentiment analysis to social media data should give me the ability to predict stock price movement to a certain extent

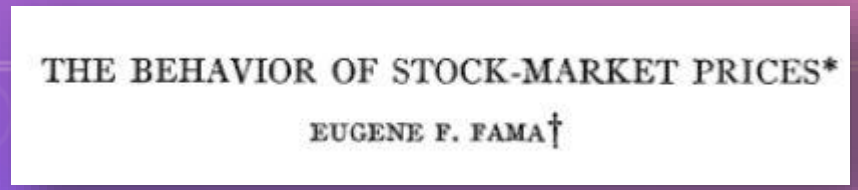

Efficiency and inefficiency in thinly traded stock markets: Kuwait and Saudi Arabia

Kirt C. Butler, S.J. Malaikah


A multivariate test for stock market efficiency: the case of ASE

Manolis G. Kavussanos & Everton Dockery

(Studies that show that stock market prices do not follow a random walk)


THE BEHAVIOR OF STOCK-MARKET PRICES*

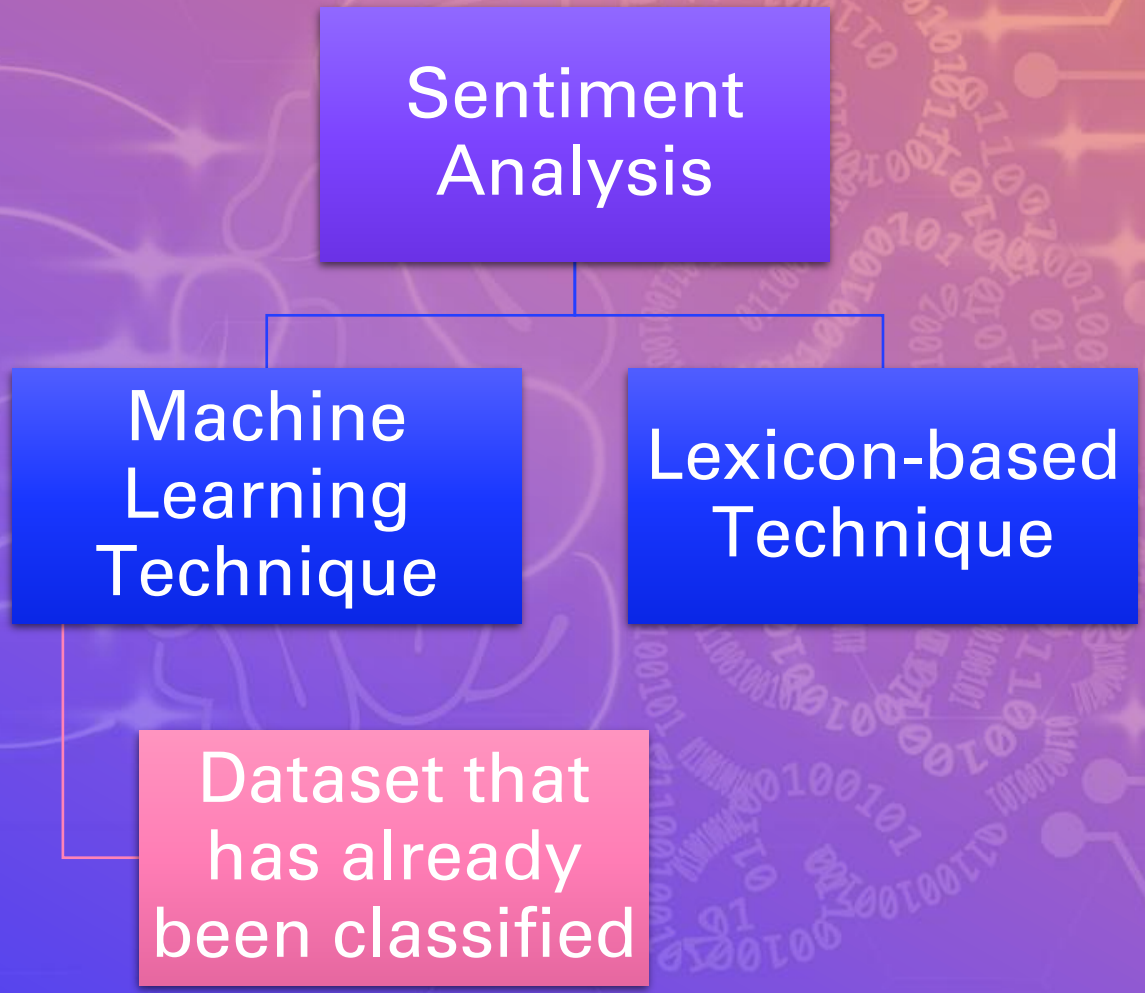
EUGENE F. FAMA†

(Efficient Market Hypothesis)

A Review Of Sentiment Analysis

- Sentiment Analysis Techniques
  1. Machine Learning Technique
  2. Lexicon Based Technique
- Data Collection
- Smart User Approach

Sentiment Analysis Techniques

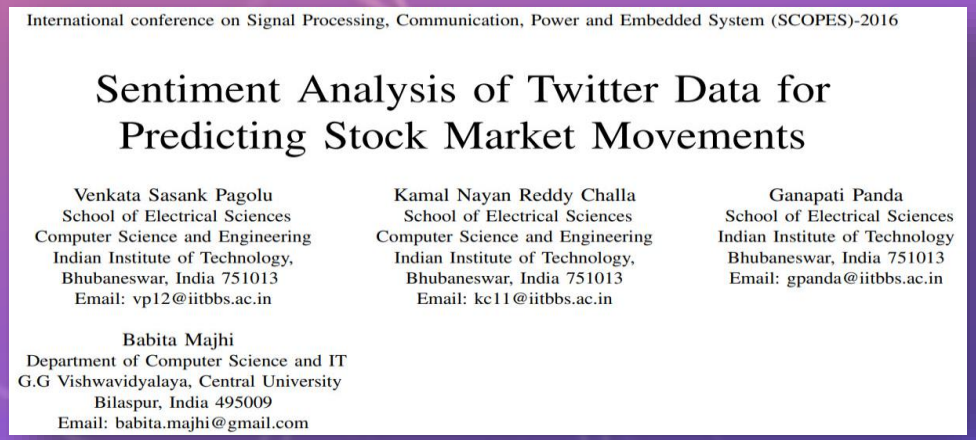


Machine Learning Technique

- Sentiment analysis is very much field specific.
- There are some open-source sentiment analysis tools available.
- They are trained with a different corpus.

Machine Learning Technique

- Developed their own sentiment analyser.
- They asked people to annotate each post / entry in a subset of their dataset.
- Features extracted using Word2vec representation.
- Trained using random forest algorithm.
- Accuracy = 70.2%
- Human accuracy = between 70% and 79% (brnrd.me)



International conference on Signal Processing, Communication, Power and Embedded System (SCOPES)-2016

Sentiment Analysis of Twitter Data for Predicting Stock Market Movements

Venkata Sasank Pagolu
School of Electrical Sciences
Computer Science and Engineering
Indian Institute of Technology,
Bhubaneswar, India 751013
Email: vp12@iitbbs.ac.in

Kamal Nayan Reddy Challa
School of Electrical Sciences
Computer Science and Engineering
Indian Institute of Technology,
Bhubaneswar, India 751013
Email: kc11@iitbbs.ac.in

Ganapati Panda
School of Electrical Sciences
Indian Institute of Technology
Bhubaneswar, India 751013
Email: gpanda@iitbbs.ac.in

Babita Majhi
Department of Computer Science and IT
G.G Vishwavidyalaya, Central University
Bilaspur, India 495009
Email: babita.majhi@gmail.com

Data Collection → Human annotation → Features extracted: Word2Vec → Trained model: Random Forrest Algorithm

Machine Learning Technique

- StockTwits
- Bullish = Positive , Bearish = Negative
- Feature extraction TF-IDF
- Logistic regression.
- Accuracy between 75% and 85%

Sentiment Analysis for Stock Price Prediction

Rubi Gupta, Min Chen
Computing and Software Systems, School of STEM
University of Washington Bothell
Bothell, WA, USA
{rubig, minchen2}@uw.edu

Stock	Accuracy
Apple	75.7
Amazon	76.8
General Electric	76.5
Microsoft	84.8
Target	74.6

(Sample of results from Gupta, R et al., 2020)

Lexicon based Technique

Twitter mood predicts the stock market

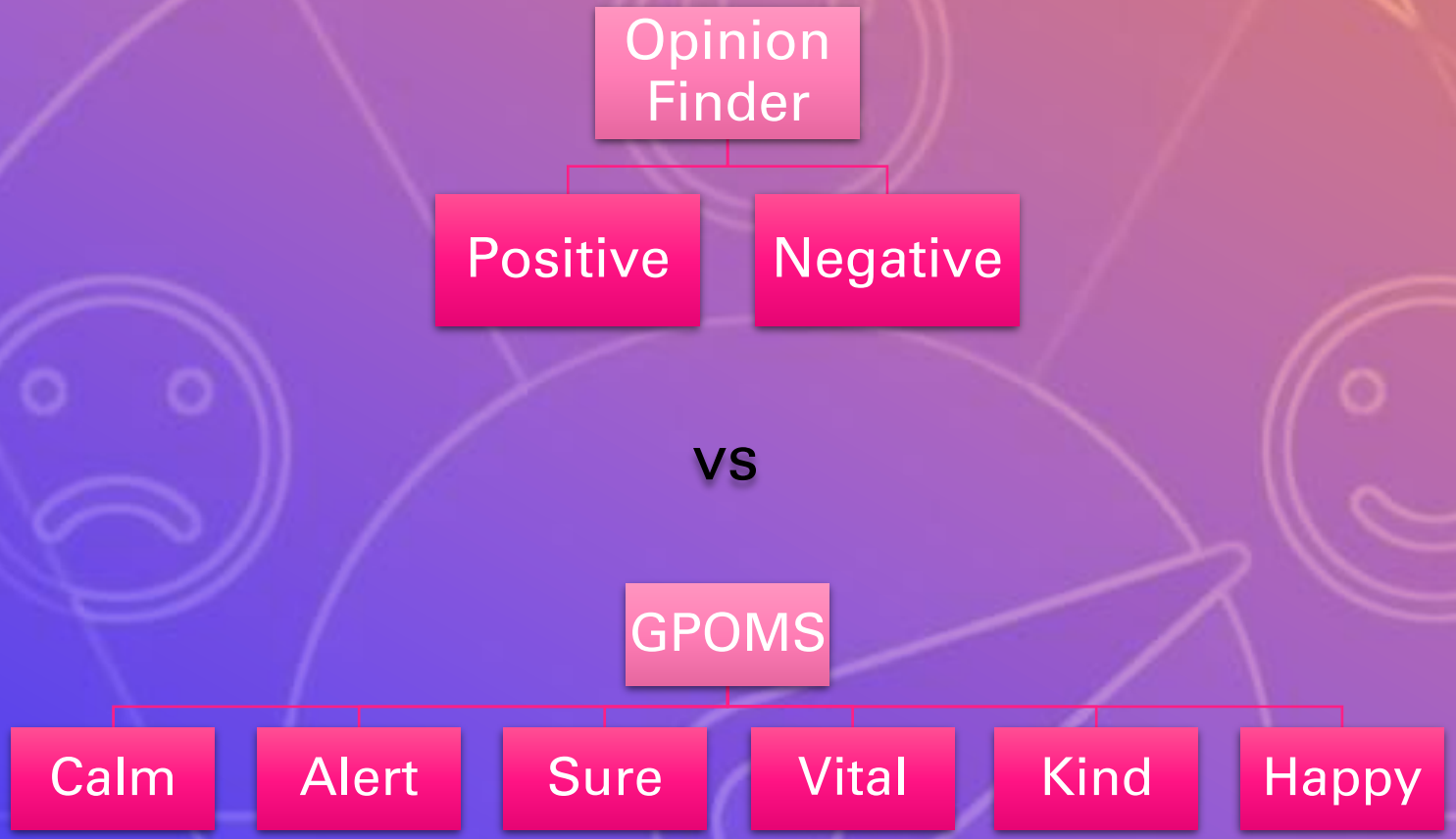
Johan Bollen[a,*,1], Huina Mao[a,1], Xiaojun Zeng[b]

[a] School of Informatics and Computing, Indiana University, 919 E. 10th Street, Bloomington, IN 47408, United States
[b] School of Computer Science, University of Manchester, Kilburn Building, Oxford Road, Manchester M13 9PL, United Kingdom

- OpinionFinder (OF) which is a publicly available software package for sentiment analysis.
- Each post they determined whether it contained any number of positive or negative terms.

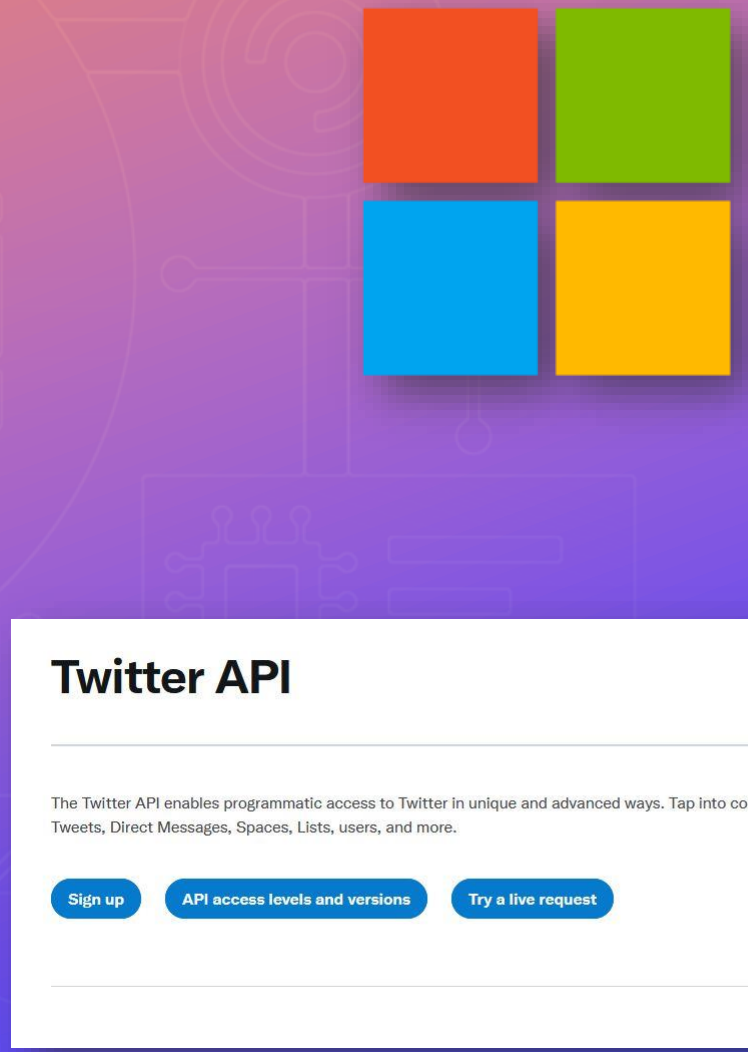
(Bollen et al., 2011. Opinion Finder used to track publics mood from October 2008 to December 2008.)

Lexicon based Technique



Data Collection

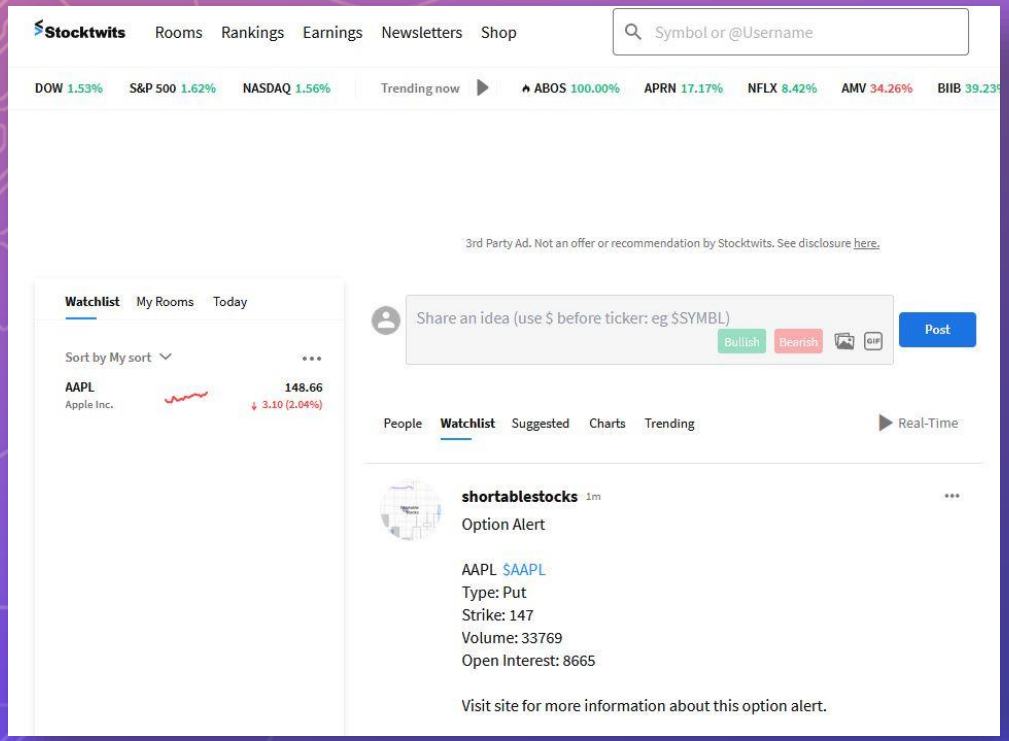
- (Pagolu V.S et al, 2016) collected 250,000 tweets on Microsoft
- Filtered using keywords like $MSFT, Microsoft and Windows etc
- Tweets extracted must represent the exact emotions of the public about Microsoft.
- Preprocessing: 1. Tokenization, 2. Stop word removal, and 3. Special character removal.
- “Twitter API has daily usage limits and only allows us to search for tweets in the past 7 days” (K Hu et al).



(https://developer.twitter.com/en/docs/twitter-api)

Data Collection

- (Coelho et al, 2019) analyzed posts from StockTwits.
- Partner level access to StockTwits provided them with a years' worth of data.
- Gupta, R et al., 2020 pre-processing: Case-normalization and no tokenization



(https://stocktwits.com/)

$AMZN still saying this ends the week down. Holiday sell off probable. Too much opportunity for a poorly timed tweet

amzn still saying ends week down holiday sell probable much opportunity poorly timed tweet

(Example of pre-processing output from Gupta, R et al., 2020)

Smart Users and Stock Selection

- (K Hu et al) proposed assigning more weight to tweets from verified users with many followers and retweets.

- (Coelho et al, 2019) used smart user approach
- Smart user is someone who's sentiment for a stock in the past has been correct at least 80% of the time.
- Narrow dataset
- Only popular/ common stocks can be used.

- (Chang, J et al, 2020) founded combining data from users "with different levels of expertise" improved their accuracy.

Stock	Users	Smart Users
GOOG	556	6
NFLX	3319	36
AMZN	2579	36
AAPL	2579	36
TSLA	3627	23

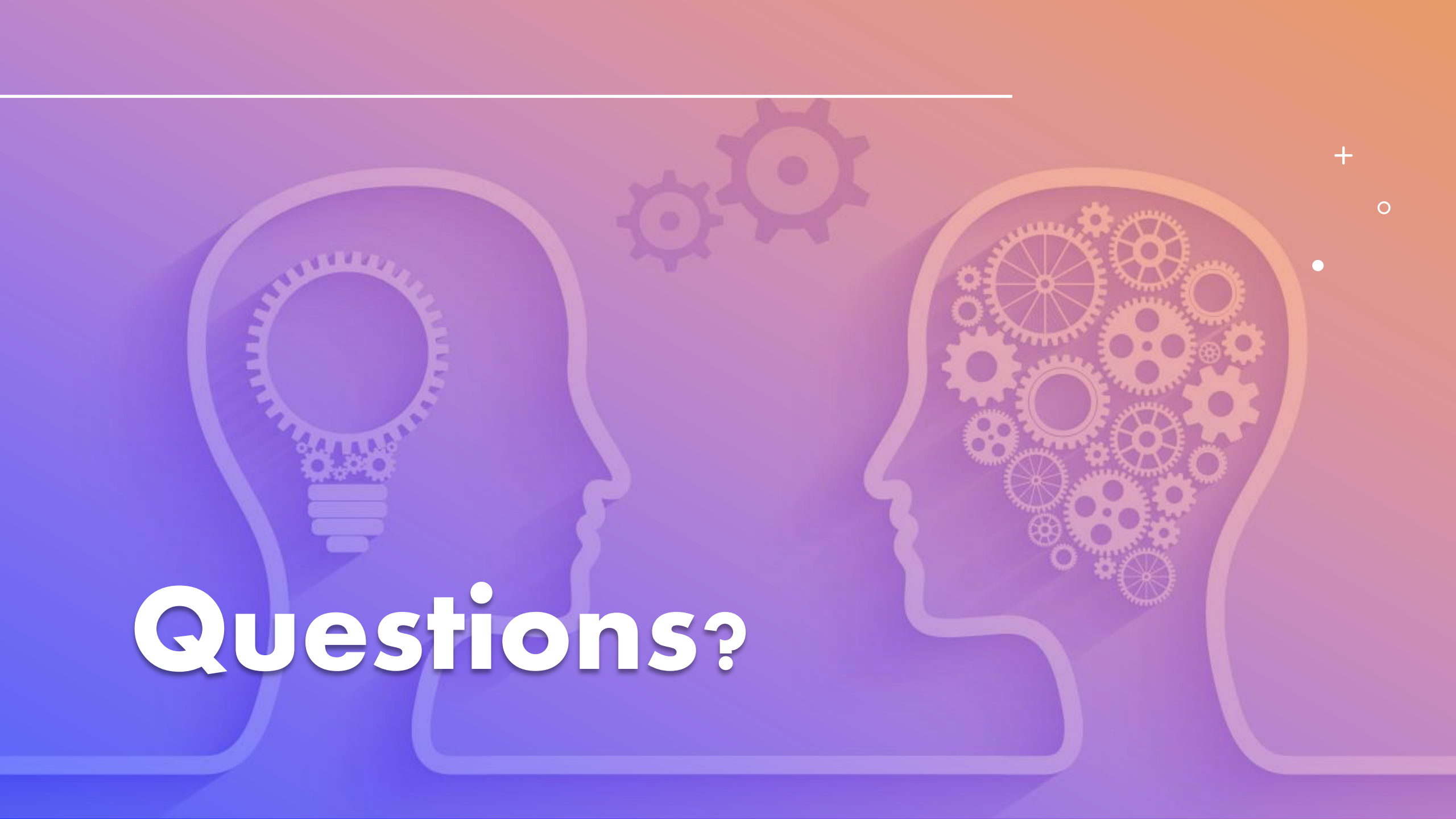
(Coelho et al., 2019)

What am I Proposing to do?

- For my project I hope to train sentiment analysis models based on a dataset comprised of social media posts. I then want to use one model on a live stream of social media posts to determine the sentiment of users over a period of time and investigate if this has any correlation to the price movement of a selected stock(s)

How Am I going to do that?

- Machine Learning Technique
- Word2Vec and TF-IDF
- Random Forest Algorithm and Logistic Regression
- Stock Market Tweets Data-NLP-2021 and / or StockTwits
- Tokenization, Stop word removal, Case normalization and Regex matching of special characters
- TweePy
- Smart User Approach


Questions?

References

- Bollen, J., Mao, H. and Zeng, X., 2011. Twitter mood predicts the stock market. *Journal of computational science*, *2*(1), pp.1-8.
- Butler, K.C. and Malaikah, S.J., 1992. Efficiency and inefficiency in thinly traded stock markets: Kuwait and Saudi Arabia. *Journal of Banking & Finance*, *16*(1), pp.197-210.
- Brnrd.me. 2022. *brnrd.me*. [online] Available at: <http://brnrd.me/social-sentiment-sentiment-analysis/> [Accessed 12 October 2022].
- Chang, J., Ding, Y. and Tu, W., 2020, November. FollowAKOInvestor: Using Machine Learning to Hear Voices from All Kinds of Investors. In *2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI)* (pp. 875-882). IEEE.
- Coelho, J., D'almeida, D., Coyne, S., Gilkerson, N., Mills, K. and Madiraju, P., 2019, July. Social media and forecasting stock price change. In *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)* (Vol. 2, pp. 195-200). IEEE.
- Coyne, S., Madiraju, P. and Coelho, J., 2017, November. Forecasting stock prices using social media analysis. In *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)* (pp. 1031-1038). IEEE.
- Fama, E.F., 1965. The behavior of stock-market prices. *The journal of Business*, *38*(1), pp.34-105.

References

- Gupta, R. and Chen, M., 2020, August. Sentiment analysis for stock price prediction. In *2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)* (pp. 213-218). IEEE.
- https://www.cfainstitute.org/en/research/cfa-digest/2003/11/the-efficient-market-hypothesis-and-its-critics-digest-summary
- Hu, K., Grimberg, D. and Durdyev, E., Twitter Sentiment Analysis for Predicting Stock Price Movements.
- Kavussanos, M.G. and Dockery, E., 2001. A multivariate test for stock market efficiency: the case of ASE. *Applied Financial Economics*, *11*(5), pp.573-579.
- Medhat, W., Hassan, A. and Korashy, H., 2014. Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, *5*(4), pp.1093-1113.
- Mittal, A. and Goel, A., 2012. Stock prediction using twitter sentiment analysis. *Standford University, CS229 (2011 http://cs229. stanford. edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis. pdf)*, *15*, p.2352.
- Pagolu, V.S., Reddy, K.N., Panda, G. and Majhi, B., 2016, October. Sentiment analysis of Twitter data for predicting stock market movements. In *2016 international conference on signal processing, communication, power and embedded system (SCOPES)* (pp. 1345-1350). IEEE.

- Touparis, F., 2018. Predicting Stocks Movement using Social Media Analytics

Images

- Getthematic.com. 2022. [online] Available at: <https://getthematic.com/assets/img/sentiment-analysis/aspect-based-sentiment-analysis.png> [Accessed 12 October 2022].
- GitHub. 2022. *Microsoft*. [online] Available at: <https://github.com/microsoft> [Accessed 12 October 2022].
- Krasamo.com. 2022. [online] Available at: <https://www.krasamo.com/wp-content/uploads/0730Building-ML-Models.jpg> [Accessed 12 October 2022].
- M.foolcdn.com. 2022. [online] Available at: <https://m.foolcdn.com/media/dubs/original_images/Intro_slide_-_digital_stock_chart_going_up_-_source_getty.jpg> [Accessed 12 October 2022].
- Martech.org. 2022. [online] Available at: <https://martech.org/wp-content/uploads/2018/04/AI-model.png> [Accessed 12 October 2022].
- Memind.eu. 2022. [online] Available at: <https://www.memind.eu/wp-content/uploads/2022/01/sentiment-analysis-emotion-detection.jpg> [Accessed 12 October 2022].
- Mindfullyurs.com. 2022. [online] Available at: <https://www.mindfullyurs.com/assets/images/approach.jpg> [Accessed 12 October 2022].