Google Trends and Stock Market

Team Members

Akshay Patil (apatil17)
Onur Baydar(obaydar)
Calvin Moore (cmoor206)

Abstract

Historical trading data is widely used to predict the movement in the stock market. With the popularity of social networking and Internet search tools, information collection ways have been widely classified. Considering the amount of data exchanged on various websites, In this project, we tried to classify the stock returns of the Dow Jones Index by utilizing the Google trend Index for different sets of keywords. We accomplished this task by using a basic 2 hidden layer sequential neural network model, LSTM model, and Convolutional neural network model that correlates changes in the Google trend index with changes in the movement of the Dow Jones index. The important challenge that we faced was a shortage of data as google places an upper limit on the amount of data that can be downloaded. The models were implemented on two different datasets and the results were compared. While the scope of our study was limited to the Dow Jones Index, it can be extended to other stocks and indexes.

Introduction

Buying the rumor, and selling the news is a phenomenon that happens in most markets particularly financial markets. Sometimes, traders use this idea as a trading strategy that can bring some opportunities in the market. Even though most of the transactions are made by robots these days, there is still human behavior and reactions intact for most of the decisions. With the popularity of searching for Internet tools and sharing information on social media, the noise in the market for particular stocks has been changing. The advent of the google trend website has made it possible for everyone to monitor the internet habits of other humans. With this newfound ability to access information, we decided to classify the movements of stock markets or some individual assets by data from Google Trends.

Related Work

Hassan et al. stated that predicting stock markets using complex calculations does not assist the user. He proposed a forecasting technique combining the hidden Markov Model and Fuzzy Concept to predict stock market rates.

Yang et al. discovered the relationship between Internet search trends and suicide deaths. The conclusion gained from this discover, was the fact that suicide related search terms were correlated to a suicide death. Thus, keyword driven search results from Internet are essential knowledge to reduce suicide attempt rates and deaths.

Frijters et al. studied about the relationship between macroeconomic conditions and an indicator of problem drinking data from Google searches. The results showed that macroeconomic conditions are associated with health in some ways. The real time data provided by Google searches are crucial information for policy makers.

Smith looked into forecasting foreign currency exchange rates by using three Google search keywords and time-series models. The results gathered demonstrated that the information from Google searches is an important tool in forecasting the market for foreign currency.

Takeda and Wakao conducted a study about the relation between the Google search intensity of stock trading volume and their relation to stock prices. It was found that the positive relationship between Google search intensity and trading volume, is greater than Google search intensity and stock prices.

Araz et al. used Google Flu Trends data to forecast influenza like illness. During research it was found that there was a strong positive relation between Google Flu Trends data and influenza illness revealed.

Bollen et al. famously demonstrated that constructing a statistical model using Twitter mood as an indicator, could generate a directional forecast for movements in the Dow Jones Industrial Average index (DJIA) at an extremely high accuracy level.

Rubin and Rubin (2010) discovered from tracking the editing frequency on Wikipedia articles of 30 Dow Jones Industrial firms, that a higher editing frequency corresponded to smaller analyst forecast errors and lower forecast dispersion. They declared that editing frequency could serve as a proxy for investor interest.

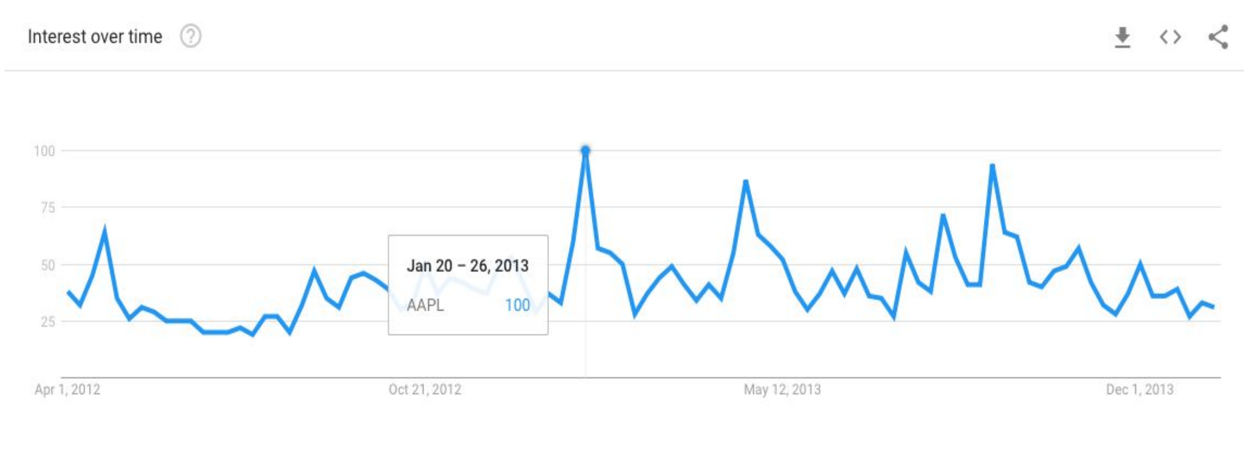
Carneiro et al. proposed Google search is by far the most popular search engine on the Web. Several researchers have used Google Trends as a tool in their research in recent years to show correlation between search volume increase of epidemics and diseases and their spreading.

The pioneering paper of Preis, Reith, and Stanley (2013) provided evidence that there is a statistically relationship between weekly Google search volumes of S&P 500 companies and weekly transaction volumes of corresponding stocks. Also, present stock prices are found to affect online search volumes of the respective companies of which prices are being changed in the following weeks .

As you can see many researchers have already started researching the relationship between Google search volume and their relation to other topics. From what is already mentioned it became easier for our group to separate this information into pros and cons; in relation to our approach. Authors such as Hassan et al. became quite useful to learn about and see what kind of approaches they took to the similar problem we were working on. Other authors that were developing research using Google search volume and terms were useful in the sense of developing how search terms can affect a given topic. The only con to using them, was the fact that they were not the exact same topic we were improving on.

Motivation

From recent research papers, it was evident that it is still difficult to predict stock prices. After doing the literature survey research papers suggested there is a correlation between GTI and the stock market. To confirm our notion we had a look at the ‘Apple’ (AAPL) stock prices and also the Google Trend Index for Apple.



The first plot is Apple’s stock price from August 2012 to November 2014. The second plot is about the Google Trend Index for the keyword ‘AAPL’ for the same time frame. From the plot it is evident that when the GTI increases the stock prices did decrease as compared to previous week. So it confirmed our notion and we attempted to incorporate the data from google trends to predict if the stock market particularly Dow Jones Index will go up or go down.

Google Trend Index

Google Trends is a search trends website provided by the largest search engine Google. It shows how frequently a given search term is entered into Google’s search engine relative to the site’s total search volume over a given period of time. These are the features of GTI:

- Google Trends provides a sample of search requests made to Google in a specified time frame.
- Google normalizes the search data to make comparisons between terms.
- The numbers are scaled on a range of 0 to 100 based on a term proportion to all searches on all topics.
- A value of 100 means that it was the peak popularity among those terms for that particular time frame.
- For instance, a value of 50 would mean that term is half as popular.
- For the same timestep, GTI will differ if the overall time frame is different.

Method:

Approach / Hypothesis / Difference:

The project should be a novel idea or different from the existing approach was the main criteria. Considering this criteria we had first decided to compare various machine learning techniques for classifying and calculating the Loan default probability. But since comparison couldn’t be the main task of our project we decided with a complete new Idea. After reading the various topics and research papers we decided to do something related to Google Trends Index. Before 2017 there was Google Correlate but Google had shut it down as nobody was using it. As our group was interested in Finance we decided to read about the research based on google trend index and relevant topics. Various papers have tried to understand the correlation between the GTI and stock prices or stock volumes. Some also had used Google search volume for influenza prediction. The research paper by Preis, Reith, and Stanley (2013) had suggested that there is statistical relation between stock market and the google search volume.

So to come up with a novel and different idea we decided to make use of predicting the movement of the stock market using certain keywords. We wanted to study given the values of google trend index will the market be classified into giving positive returns or negative returns. Various researchers have tried to predict the stock prices or volume of stock traded using the google search volumes but no one has tried to classify it based on weekly returns.

Preis, Reith, and Stanley (2013) tried finding relation between GTI for each keyword and the stock market. We decided to use GTI all the set of keywords proposed by them to classify the market index. We also came up with an idea to include a different dataset. We downloaded the GTI for all the stocks in the Dow Jones index and prepared our second dataset. To add to it we considered a deep learning framework to classify the returns. We implemented three different deep learning neural net models on our both data sets to classify the weekly returns of Dow Jones Market Index and compared them.

Dataset

The first important part of the project was data collection. For the project, we made use of two datasets. To be able to see how search engine query data can reveal the trend change or performance of the entire financial security or index, we analyzed the relationship between movement market indices and search volume of keywords. The first dataset has 98 search terms. Google Trend Index of 98 terms is collected from *Google Trends* for the U.S. market. These terms are derived from the work of Preis, Moat, and Stanley (2013) who use search volumes of these terms to evaluate trading decisions. Fig 1 lists these keywords. The dataset was downloaded using the publicly available by trends API. As Google places an upper limit on how much the data can be downloaded, 5 years of weekly data was downloaded. So the total number of observations was 266 for 98 keywords.

98 Search Terms Related to Financial Markets (Preis, Moat, and Stanley 2013)

debt	society	water	trader
color	leverage	rich	rare earths
stocks	loss	risk	tourism
restaurant	cash	gold	politics
portfolio	office	success	energy
inflation	fine	oil	consume
housing	stock market	war	consumption
dow jones	banking	economy	freedom
revenue	crisis	chance	dividend
economics	happy	short sell	world
credit	car	lifestyle	conflict
markets	nasdaq	greed	kitchen
return	gains	food	forex
unemployment	finance	financial markets	home
money	sell	movie	crash
religion	invest	nyse	transaction
cancer	fed	ore	garden
growth	house	opportunity	fond
investment	metals	health	train
hedge	travel	short selling	labor
marriage	returns	earnings	fun
bonds	gain	arts	environment
derivatives	default	culture	ring
headlines	present	bubble	

Fig. 1

The second dataset consists of the tickers of 30 stocks from Dow Jones that we have obtained from Yahoo Finance. The Google trend Index for these keywords was also downloaded using the pytrends library. The list of these keywords is displayed in Fig 2.

Tickers	Names
AXP	American Express Co
AMGN	Amgen Inc
AAPL	Apple Inc
BA	Boeing Co
CAT	Caterpillar Inc
CSCO	Cisco Systems Inc
CVX	Chevron Corp
GS	Goldman Sachs Group Inc
HD	Home Depot Inc
HON	Honeywell International Inc
IBM	International Business Machines Corp
INTC	Intel Corp
JNJ	Johnson & Johnson
KO	Coca-Cola Co
JPM	JPMorgan Chase & Co
MCD	McDonald's Corp
MMM	3M Co
MRK	Merck & Co Inc
MSFT	Microsoft Corp
NKE	Nike Inc
PG	Procter & Gamble Co
TRV	Travelers Companies Inc
UNH	UnitedHealth Group Inc
CRM	Salesforce.Com Inc
VZ	Verizon Communications Inc
V	Visa Inc
WBA	Walgreens Boots Alliance Inc
WMT	Walmart Inc
DIS	Walt Disney Co
DOW	

Fig. 2

The plot of GTI for various keywords is shown in following figures:

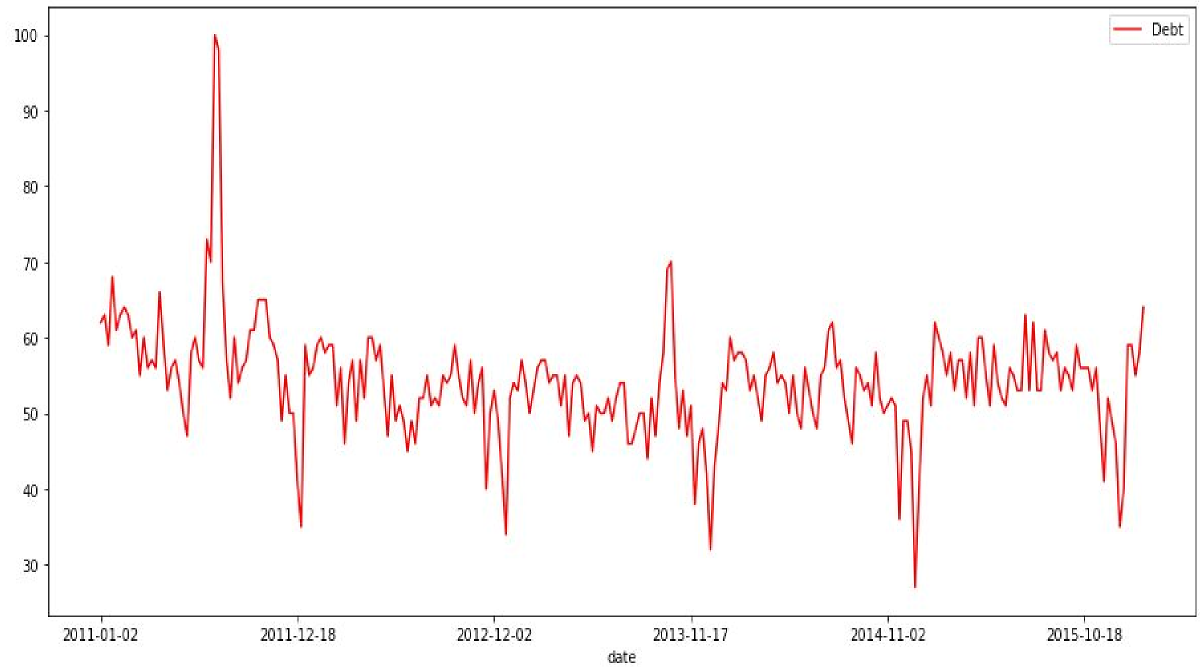


Fig.2

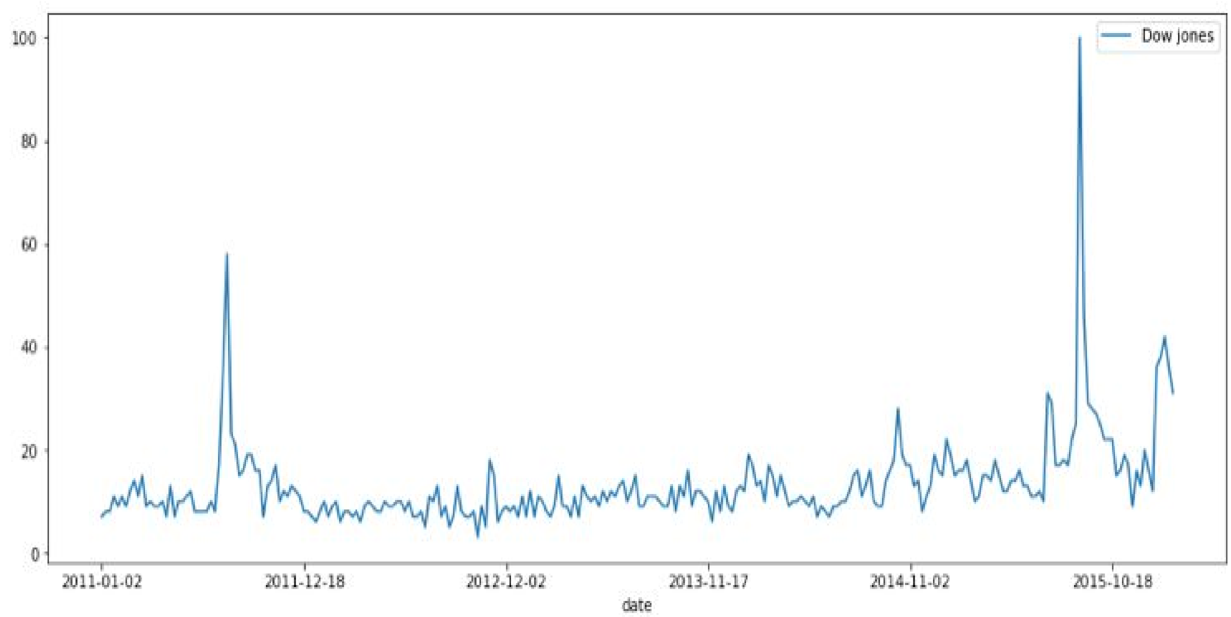


Fig. 4

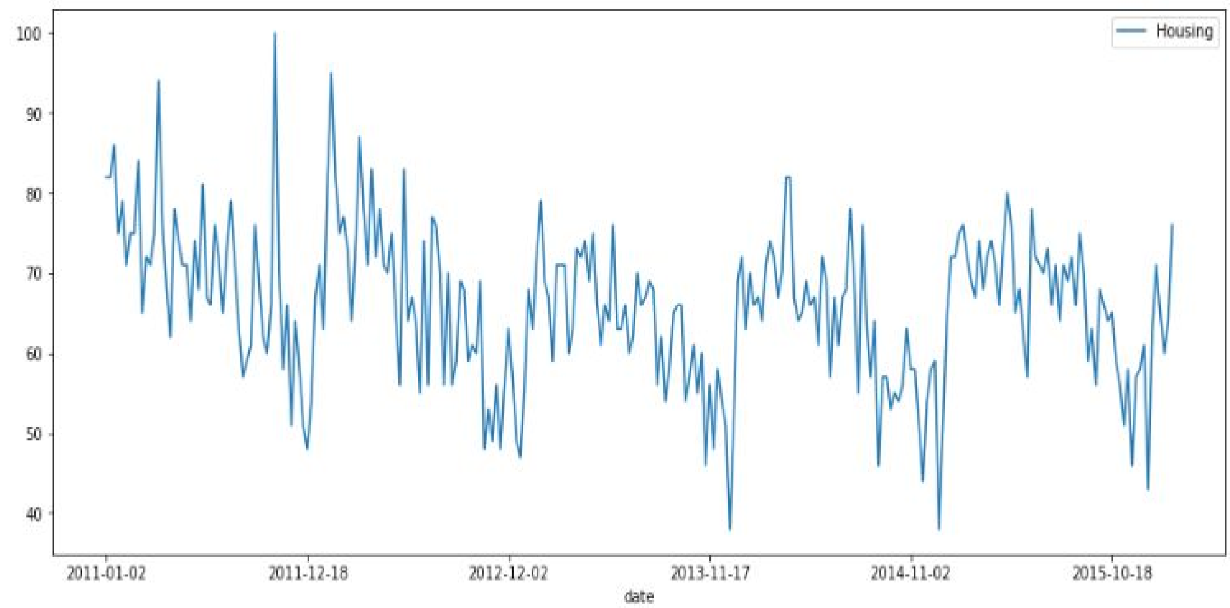


Fig. 5

Following plot helps us to compare the GTI for three different keywords. It can be seen it is easy to compare and analyze them after standardizing the GTI.

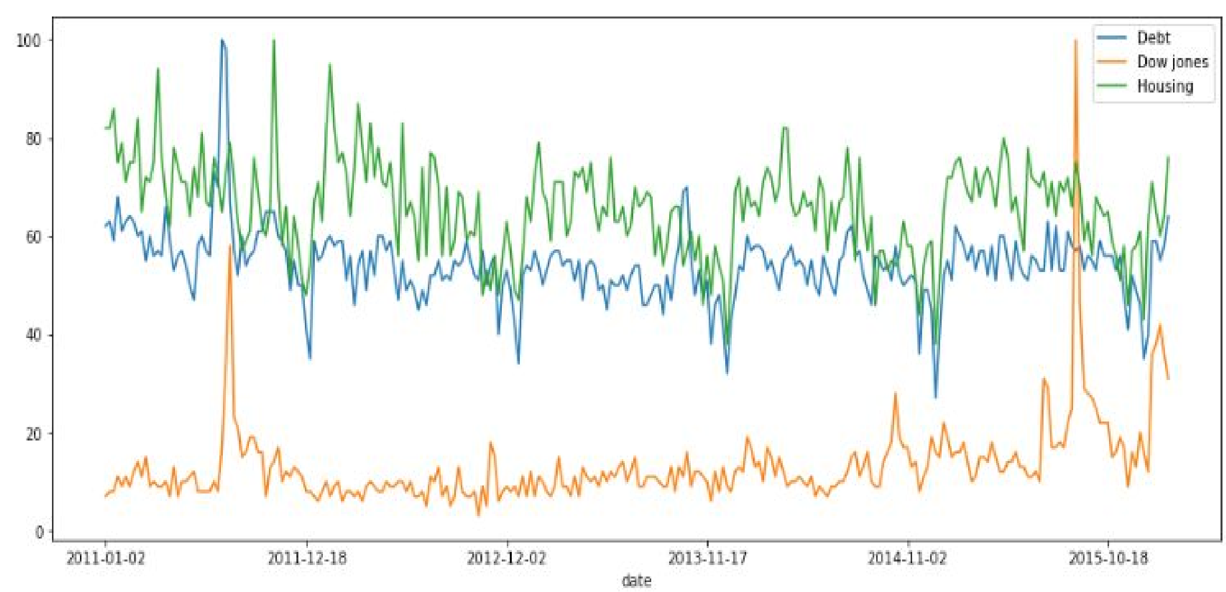


Fig. 6 Before Standardizing

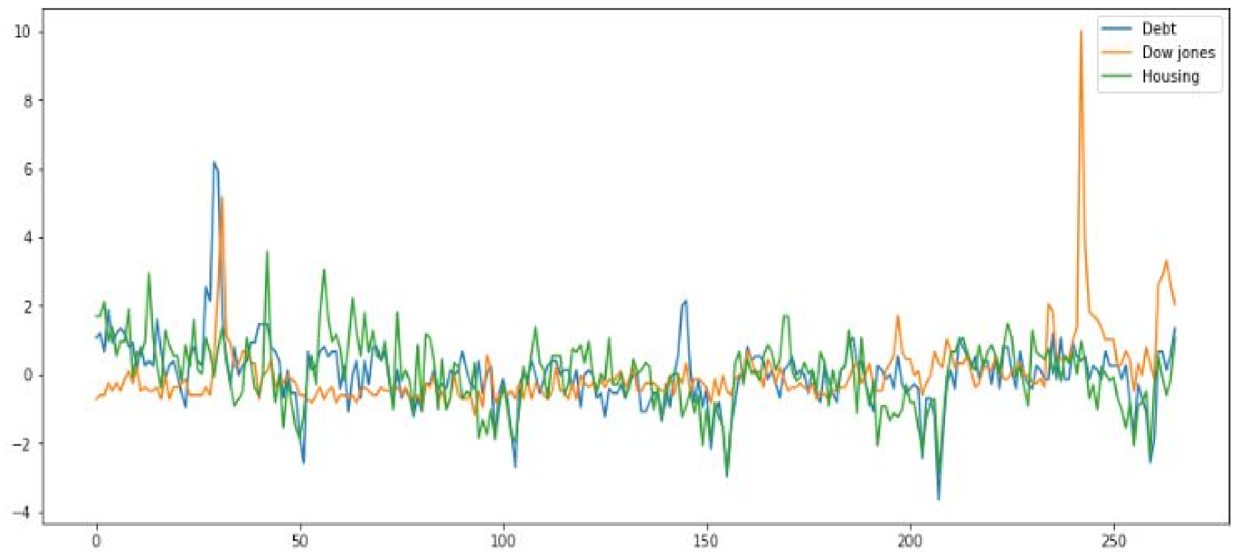


Fig. 7 After Standardizing.

Once the Google Trend Index for all the keywords was downloaded. The historical data for Dow Jones Index was downloaded for almost the same time frame from the yahoo finance website. The GTI for all keywords was downloaded from a time frame for 2011/01/02 to 2016/01/31. The Dow Jones weekly historical data ranged from 2011/01/03 to 2016/02/02.

Data Pre-Processing

Once all the data that is needed was ready we needed to create the labels for our analysis. We intended to classify the stock market movements on the basis of weekly returns. By using the adjusted closing price we calculated the weekly returns. Weekly returns were calculated by taking the ratio of the current week's adjusted closing price to the previous week's adjusted closing price. We took the log of those weekly returns and the negative log-returns were labeled as '-1' and positive returns were labeled as '1' and stored in a direction column. The Google trend index was normalized to a scale between 0 to 1 so it can be ready for our Deep Learning Model.

Experiment

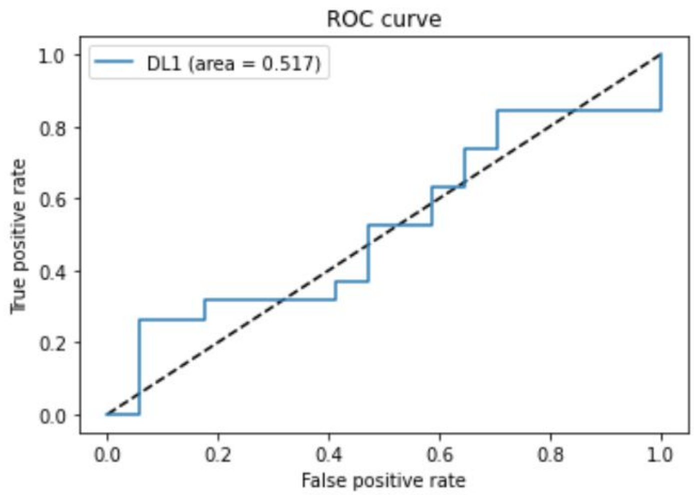
Dataset :

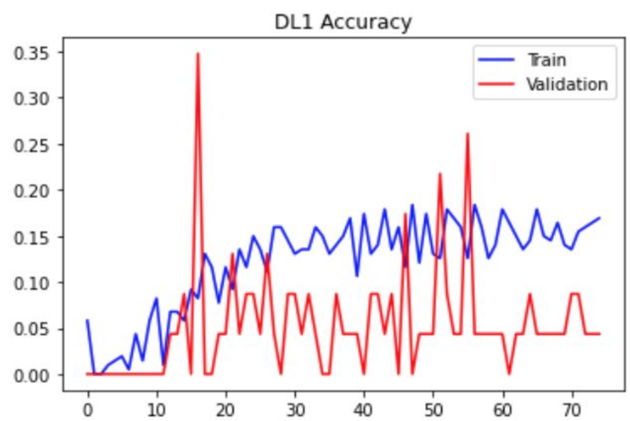
When the data was pre-processed we were faced with the most important challenge. The challenge was to partition the data into training and testing sets. The sample size of data is 266 weekly observations. Due to the small sample size, it posed an interesting challenge. As we were making use of deep learning models it was known that the deep learning models require more data to train. So the challenge was to partition the data in such a way that the Deep learning model can be trained on and also be left with enough data to make predictions. After some deliberation and lots of trial and error, we decided that we will be using 230 observations for training the models and the remaining 36 observations as test data.

Once the data was partitioned we implemented three different neural network models on both the dataset. To summarize again, the first dataset consists of 98 keywords adopted from the Preis research paper. These keywords were selected by them on the basis that they were the trending words and were used in news during the time frame 2011 and 2016. Let's call it ‘dataset 1’. The second dataset consists of GTI for 29 stocks which make up the Dow Jones Index. Let’s call it ‘dataset 2’.

Model 1

The first model was a neural network with 1 hidden layer with 128 neurons and the activation function used is ‘ReLu’. The activation function for the output layer is ‘sigmoid’ as we are making the binary classification. The optimizer used was ‘Adams’ and the loss function used is ‘binary cross-entropy’. The batch size was kept 1 due to the low sample size and the number of epochs was 75. The model with the best validation accuracy was saved using callback and was later used to classify. Since our dataset was not imbalanced (‘-1’ - count 155 and ‘1’ count 111) we made use of AUC as our metric. The AUC score for this model was 0.517.

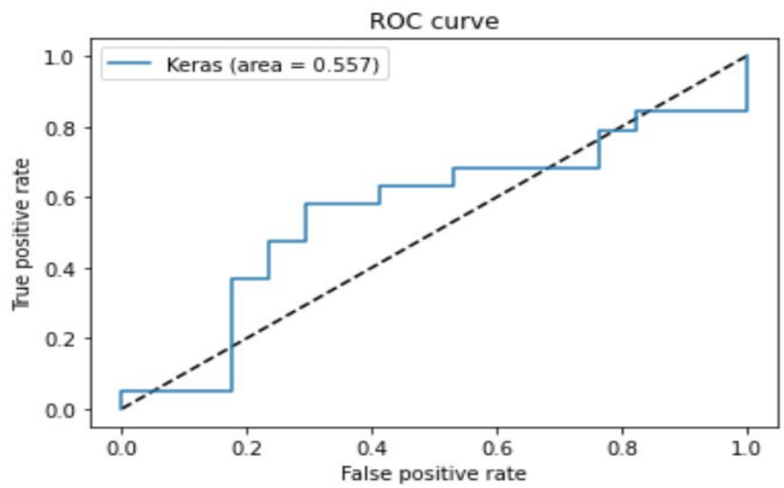

ROC curve
DL1 (area = 0.517)
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

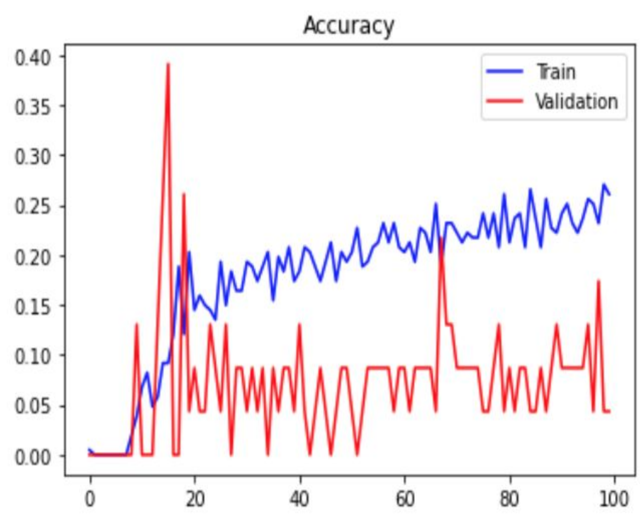

DL1 Accuracy
Train
Validation
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0
10
20
30
40
50
60
70

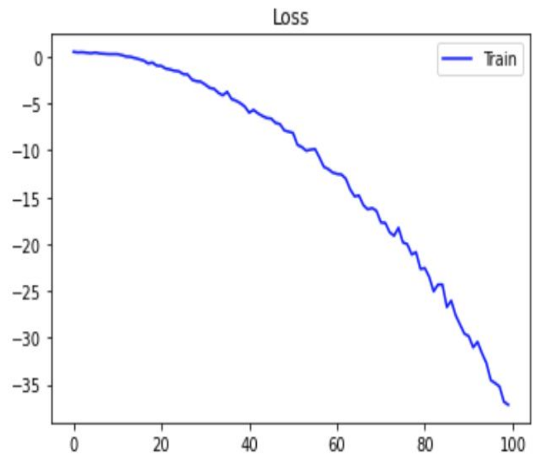

DL1 Loss
Train
0
−5
−10
−15
−20
−25
0
10
20
30
40
50
60
70

Model 2:

The second model was an LSTM neural network with 1 hidden layer with 128 neurons and the dropout was 0.2. The activation function for the output layer is 'sigmoid' as we are making the binary classification. The optimizer used was 'Adams' and the loss function used is 'binary cross-entropy'. The batch size was kept 1 due to the low sample size and the number of epochs was 100. The model with the best validation accuracy was saved using callback and was later used to classify. Since our dataset was not imbalanced ('-1' - count 155 and '1' count 111) we made use of AUC as our metric. The AUC score for this model was 0.557.

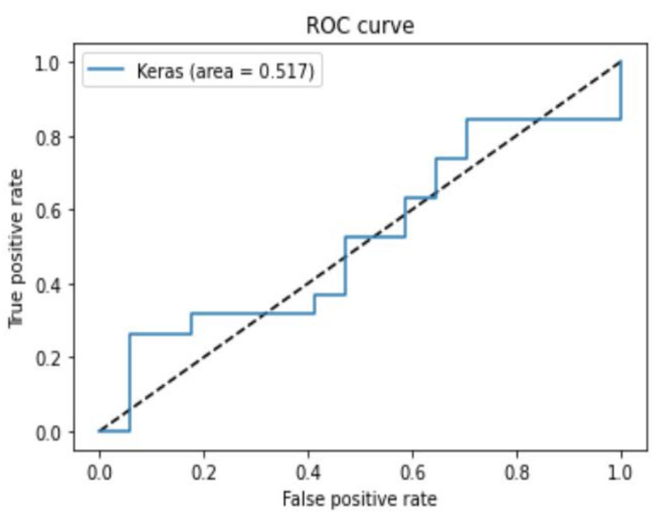


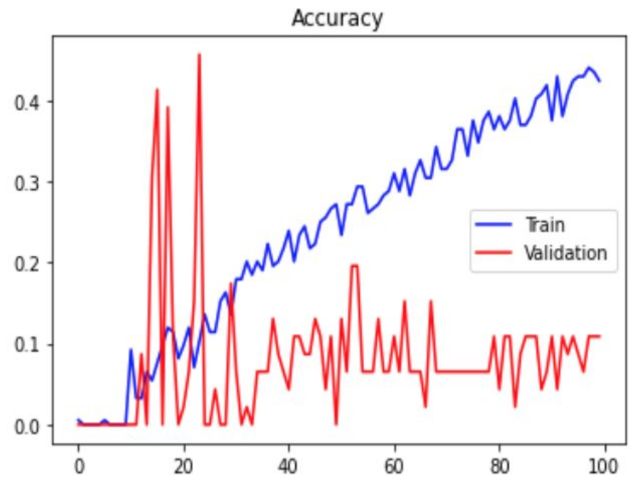


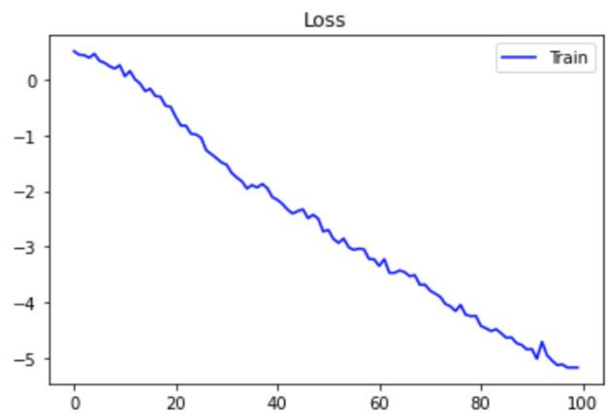


Model 3:

The third model was a Convolutional 1d neural network with 1 hidden layer with 128 neurons and the activation function was 'ReLu'. The activation function for the output layer is 'sigmoid' as we are making the binary classification. The optimizer used was 'Adams' and the loss function used is 'binary cross-entropy'. The batch size was kept 1 due to the low sample size and the number of epochs was 100. The model with the best validation accuracy was saved using callback and was later used to classify. Since our dataset was not imbalanced ('-1' - count 155 and '1' count 111) we made use of AUC as our metric. The AUC score for this model was 0.51.



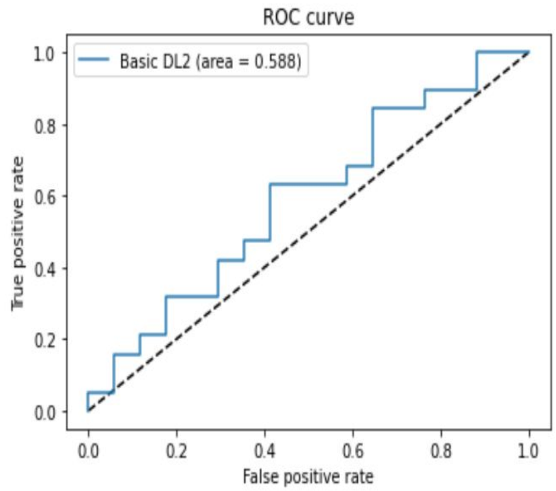


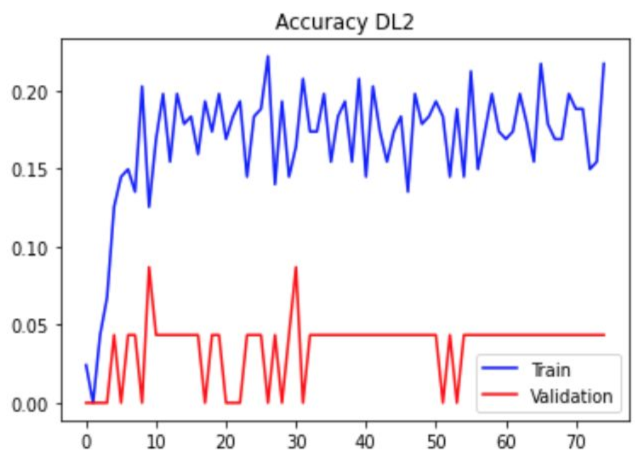


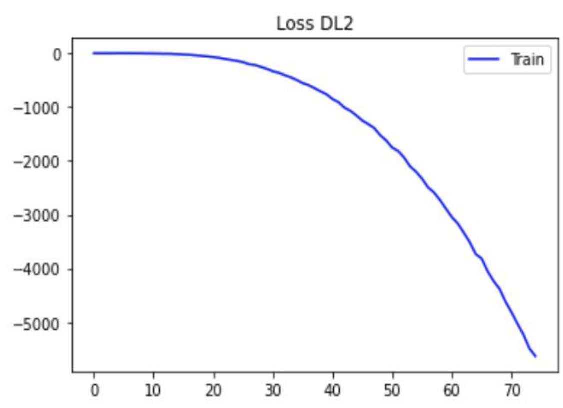
Dataset 2:

Model 1

The first model was a neural network with 2 hidden layers with 128 neurons and the activation function used is 'ReLu'. The activation function for the output layer is 'sigmoid' as we are making the binary classification. The optimizer used was 'Adams' and the loss function used is 'binary cross-entropy'. The batch size was kept 1 due to the low sample size and the number of epochs was 75. The model with the best validation accuracy was saved using callback and was later used to classify. Since our dataset was not imbalanced ('-1' - count 155 and '1' count 111) we made use of AUC as our metric. The AUC score for this model was 0.588.

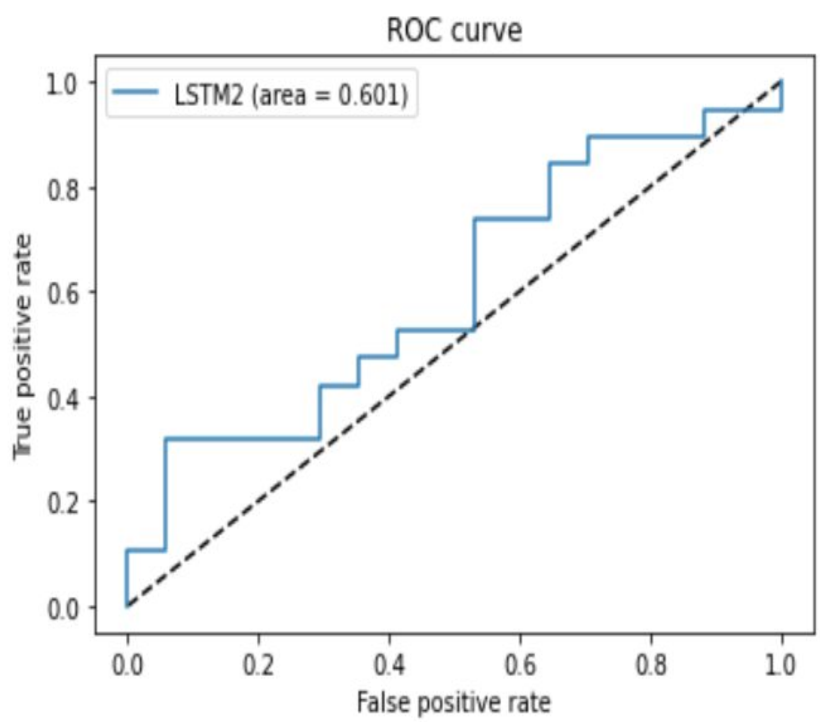

ROC curve
Basic DL2 (area = 0.588)
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

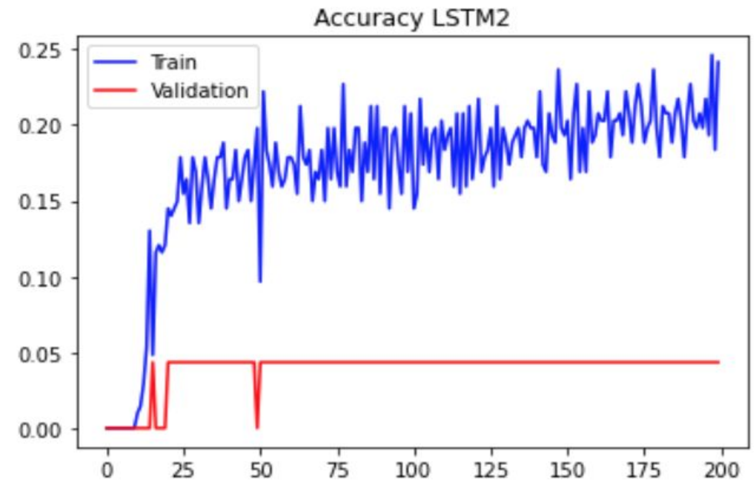

Accuracy DL2
Train
Validation
0.00
0.05
0.10
0.15
0.20
0
10
20
30
40
50
60
70

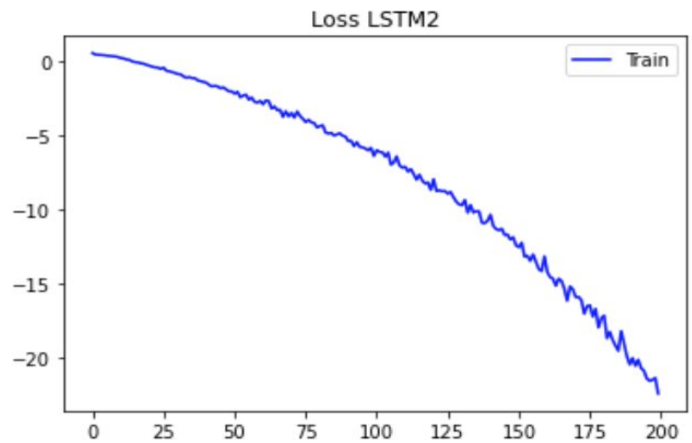

Loss DL2
Train
0
−1000
−2000
−3000
−4000
−5000
0
10
20
30
40
50
60
70

Model 2:

The second model was a LSTM neural network with 1 hidden layer with 128 neurons and the dropout was 0.2. The activation function for the output layer is ‘sigmoid’ as we are making the binary classification. The optimizer used was ‘Adams’ and the loss function used is ‘binary cross-entropy’. The batch size was kept 1 due to the low sample size and the number of epochs was 200. The model with the best validation accuracy was saved using callback and was later used to classify. Since our dataset was not imbalanced (‘-1’ - count 155 and ‘1’ count 111) we made use of AUC as our metric. The AUC score for this model was 0.601.

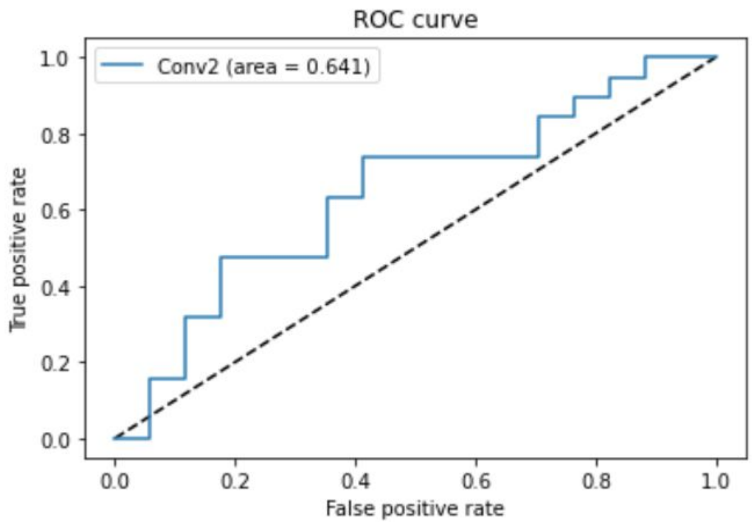


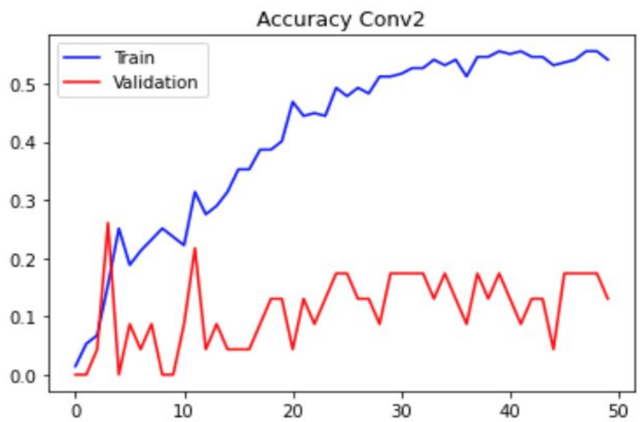


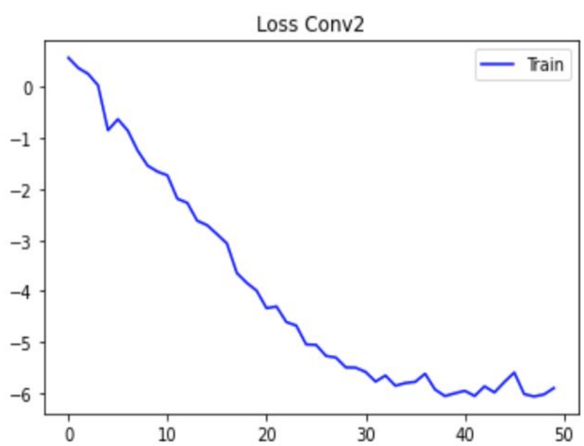


Model 3:

The third model was a Convolutional 1d neural network with 2 hidden layers having 256 and 512 neurons and the activation function was 'ReLu'. The activation function for the output layer is 'sigmoid' as we are making the binary classification. The optimizer used was 'Adams' and the loss function used is 'binary cross-entropy'. The batch size was kept 1 due to the low sample size and the number of epochs was 50. The model with the best validation accuracy was saved using callback and was later used to classify. Since our dataset was not imbalanced ('-1' - count 155 and '1' count 111) we made use of AUC as our metric. The AUC score for this model was 0.64.


ROC curve
Conv2 (area = 0.641)
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0


Accuracy Conv2
Train
Validation
0.0
0.1
0.2
0.3
0.4
0.5
0
10
20
30
40
50


Loss Conv2
Train
0
-1
-2
-3
-4
-5
-6
0
10
20
30
40
50

Experimentation and Analysis:

Our first initial finding was to classify the data for daily returns but since google places an upper limit only 180 observations could be downloaded for daily data. The accuracy of the model was very bad. So we decided to make use of weekly data. All the above models were selected after lots of trial and error. Also, the epochs, batch size, the number of layers were changed to avoid underfitting as well as overfitting. The second dataset works better in classifying the market. The AUC for all the models is pretty low but the second dataset having fewer variables works well. The best model was the convolutional neural network model with an AUC of 0.64 amongst all the models. It works well than the dumb classifier.

Challenges and reflection:

Our original plan was to compare the predictive accuracy of loan default for different Machine Learning models. Based on the feedback, we planned to pursue a new project topic. We were also asked to give a detailed approach and difference that we have provided earlier. For the new task, it was a challenging and valuable experience that being able to understand the use of google trends, scraping websites. Understanding the different models as well as their limitations to find the model which is suitable for classification was also very challenging during our project. In addition to that, we implemented models on the daily Google Trend Index but the accuracy was low. The model can be improved more by using the following methods:

- Making use of google trends, scraping websites like Twitter if needed.
- Understanding of different models, their limitations, and selecting the model suitable one for the classification will be a difficult problem.
- The risk of underfitting is our main challenge so provided with enough data we train the model for better accuracy.

Conclusion :

We have described a computational procedure for relating the google trend Index with respect to classifying the market based on weekly returns. If the model is improved it can be used as a technical indicator for carrying out trades, If google removes the upper limit, the model can perform better, as it will have more observations to be trained on.

References & Citation

- M. R. Hassan, "A combination of hidden Markov model and fuzzy model for stock market forecasting," *Neurocomputing*, vol. 72, no. 16–18, pp. 3439–3446, 2009.
- Bollen J, Mao H, Zeng X (2011) Twitter mood predicts the stock market. Comput Sci 2:1–8
- Chan WS (2003) Stock price reaction to news and no-news: drift and reversal after headlines. J Financ Econ 70(2):223–260
- G. P. Smith, "Google internet search activity and volatility prediction in the market for foreign currency," *Finance Research Letters*, vol. 9, no. 2, pp. 103–110, 2012.
- F. Takeda and T. Wakao, "Google search intensity and its relationship with returns and trading volume of Japanese stocks," *Pacific-Basin Finance Journal*, vol. 27, pp. 1–18, 2014.
- M. Araz, D. Bentley, and R. L. Muelleman, "Using Google Flu Trends data in forecasting influenza-like-illness related ED visits in Omaha, Nebraska," *The American Journal of Emergency Medicine*, vol. 32, no. 9, pp. 1016–1023, 2014.
- A. C. Yang, S. J. Tsai, N. E. Huang, and C. K. Peng, "Association of Internet search trends with suicide death in Taipei City, Taiwan, 2004–2009," *Journal of Affective Disorders*, vol. 132, no. 1-2, pp. 179–184, 2011.
- Hisano R, Sornette D, Mizuno T, Ohnishi T (2013) High quality topic extraction from business news explains abnormal financial market volatility. PLoS ONE 8(6):e64846
- Ingle V, Deshmukh S (2016) Live new streams extraction for visualization of stock market trends. In: Lecture notes in electrical engineering, vol 395

- Rubin A, Rubin E (2010) Informed investors and the internet. J Bus Financial Acc 37(7/8):841–865
- Carneiro and Mylonakis, 2009 H.A. Carneiro, E. Mylonakis. Google trends: A web-based tool for real-time surveillance of disease outbreaks. Clinical Infectious Diseases, 49 (2009), pp. 1557-1564
- Ginsberg et al., (2008) Detecting influenza epidemics using search engine query data. Nature, 457 (2008), pp. 1012-1014
- Preis, T., H. S. Moat, and H. E. Stanley. 2013. Quantifying trading behavior in financial markets using google trends. *Scientific Reports* 3: 1684.

Team Contributions

Task (Updated Plan)	Akshay Patil	Onur Baydar	Calvin Moore
Proposal	✔✔	✔✔	✔✔
Literature review	✔✔	✔✔	✔✔
Dataset 1using pytrends	✔✔	✔	
Classification using dataset	✔✔		
Mid Progress report	✔✔	✔✔	✔✔
Feature selection and Extraction from google trend	✔✔		
Dataset 2 using pytrends	✔✔	✔	
Model selection and Parameter tuning	✔✔		
Model training	✔✔		
Model evaluation and comparison	✔✔		
Final Report	✔✔	✔	✔