OBJECTIVE

Predicting Volatility in Equity Markets using Macroeconomic News

NAVYA P

INTRODUCTION

- *On June 26, 2015, months of debt negotiation between the Greek government, headed by Prime Minister Alexis Tsipras, and its creditors, including the IMF and fellow Eurozone countries, broke off abruptly.*

- *As political and economic uncertainty grew in Europe, investors across the world were moving their funds away from risky assets that could be negatively affected by a Greek default, and towards safer assets. Days later, when the Greek situation had 'resolved', the process reversed as equities rallied.*

- *Just weeks later, crisis erupted in the Chinese equity markets, and asset values were responding to the uncertainty this provoked.*

- *The objective of this project is to how can we investigate on these macroeconomic sentiment and how it immediately impacts the volatility in liquid markets, by measuring market volatility through the VIX (volatility index of the S&P 500).*

BUSINESS PROBLEM

- *Twitter provides a plethora of market data. In this project, we will use over 200,000 tweets from various accounts to predict upward movements in the VIX.*

- *In order to create a viable trading strategy, we need to accurately predict an increase in market volatility at least 51% of the time.*

BUSINESS CONSTRAINTS

- There is a lot of research on predicting market price movements by conducting technical analysis using historical and time-series data.

- A large portion of the studies fall into developing techniques to analyze markets trends from texts such as financial market news, or social media posts.

- The Machine Learning approaches to make predictions in various financial products are Support Vector Regression, which performs linear regression in the high dimension feature space to predict stock prices, Logistic Regression and Neural Networks are also plausible in analyzing stocks trends

- All these methods are sensitive to bias and noise, and the study performances tend to improve as more noises are removed from original data sets.

DATASETS

- *Tweets are pulled from major news sources, hedge funds and investment banks, and notable economists. investigate on how **"breaking economic news"** affects the markets, and subsequently how to predict which news stories can increase volatility.*

- *We will consider a tweet **'significant'**, in that the news presented in the tweet contributes to volatility in the market which is extracted using sentiment analysis (positive tweets, negative tweets and neutral tweets).*

- *If within 30 minutes of the tweet being tweeted, the volatility of the asset increases by one-fifth of a standard deviation*

- *The Twitter API allows for the most recent 3200 tweets per account to be exported from the website.*

- *Since, our investigation is on the immediate market reaction, and therefore intraday data(at 30 min intervals) is more sufficient for our prections.*

- *This resulted in more than 200,000 tweets.*

DATASETS (CONT..)

- *Since there were 200,000 tweets, had to construct our own dictionary because the Twitter uses 'Hashtag(#)' and 'stock ticker($)' features in tweets.*

- *These tweets are cleaned by eliminating all characters that were not letters, and eliminated URLs. The dictionary used contains just over 10,000 words, which serves as our features for the algorithms.*

- *After, preprocessing the final datset contained 988 data points in which:*

Financial Data(Market Data) includes 30 companies across the globe

- *Twitter Data is extracted from 70 accounts(between 9.a.m. and 4 p.m.)*

FEATURES

- *Financial Dataset Features includes: 31-05-2013 to 18-9-2014*
  - *Date*
  - *High*
  - *Close*
  - *Open*
  - *Low*
- *Twitter Dataset Features includes: 31-05-2013 to 18-9-2014*
  - *Date*
  - *Positive Tweets*
  - *Negative Tweets*
  - *Neutral Tweets*
  - *Total Number of Tweets*

TYPE OF MACHINE LEARNING PROJECT

The model that is used in this project is Regression type of Machine Learning.

We have implemented the ML models using three different classification algorithm to find which algorithm best suits for this project.

Three different techniques as mentioned above will be implemented to predict movements :

> *1. Naive Bayes*
>
> *2. SVM and*
>
> *3. Logistic regression.*

- *Performance Matrix: Accuracy*

CONCLUSION AND FUTURE ENHANCEMENTS

- *Using three supervised model, the accuracy is between 57% - 67%*

- *Comparatively , Logistic Regression performed well than Naïve Bayes and SVM.*

- *Future enhancement will be an ongoing process, we should be extracting more tweets, eliminating tweets that are not macroeconomic headlines, or headlines unrelated.*

- *Also, by modifying and shortening the dictionary to include exclusively market related words and tokens, would lower our chance of overfitting, and see an increase in accuracy.*

- *More research is required to obtain the data from other important sources and also need to explore more on feature engineering.*