

BLUEPRINT OF THE PROJECT

PREDICTING MARKET VOLATILITY USING MACRO HEADLINES

Presented by: Meet Panchal

Introduction

- On June 26th, 2015, months of debt negotiations between the Greek government and its creditors broke off abruptly. Prime Minister Tsipras called for a snap referendum regarding bailout terms.
- Within hours, the S&P had fallen significantly. Large market movements as a consequence of political and economic headlines are hardly uncommon; liquid markets are most susceptible to swings when news breaks.
- Using the VIX as a proxy for market volatility, we investigate how macroeconomic news headlines affect changes in the VIX. We predict equity market vol using tweets from major news sources, hedge funds and investment banks, and notable economists.

Buisness Problem

- Predicting market volatility using macro headlines.
- By employing three machine learning techniques for classification, we aim to predict the volatility-inducing power of a single tweet and/or news story.
- Using a training set, we will identify key words and their associated probability of increasing volatility in the markets using Naive Bayes, SVM and logistic regression. The input of our algorithm will be the word count of each bucket of tweets in a dictionary we created. We then use the three prediction methods to output a predicted increase in the VIX, which will be a binary variable. In order to create a viable trading strategy, we aim to predict moves with above 51% accuracy.

Data Overview

1. Twitter data

- The Twitter sentiment data are available in separate CSV files for each company.
- Date
- Number of negative tweets
- Number of neutral tweets
- Number of positive tweets
- Total number of tweets

2. Financial data

- The financial data are available in separate CSV files for each company.

- Also, the data for the DJIA index are contained in the CSV file financial_data_DJIA.csv.

- The columns of the CSV files are:

 - Date

 - High

 - Close

 - Open

 - Low

Type of Machine Learning project

- 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. We use three different supervised learning methods:
 - Naive bayes
 - Support vector machine (SVM)
 - Logistic regression and principal component analysis (PCA)

Performance metrics

- Using three supervised learning techniques, we have developed a methodology for predicting volatility movements, with an accuracy between 56-64%.
- By implementing a 2:1 stop-loss ratio, this can be transformed into a trading strategy, using options on the VIX