

#PredictingTheDow

Daniel Mooney, Mohammed Faiz Shaikh, Apruva Tripathi

Theory

It has long been theorized that the news affects the stock market.

September 11th - when the markets opened on 9/17, NYSE went down 680 points (7.1%)

Lehman Brothers collapse led Dow closing 4.4% or 504 point down

Can we predict the stock market movement from the news?

Problem

Using /r/worldnews (Reddit) to predict whether the stock market(measured by Dow Jones Index) will go up or down

We are given Top 25 News Items of the day, along with

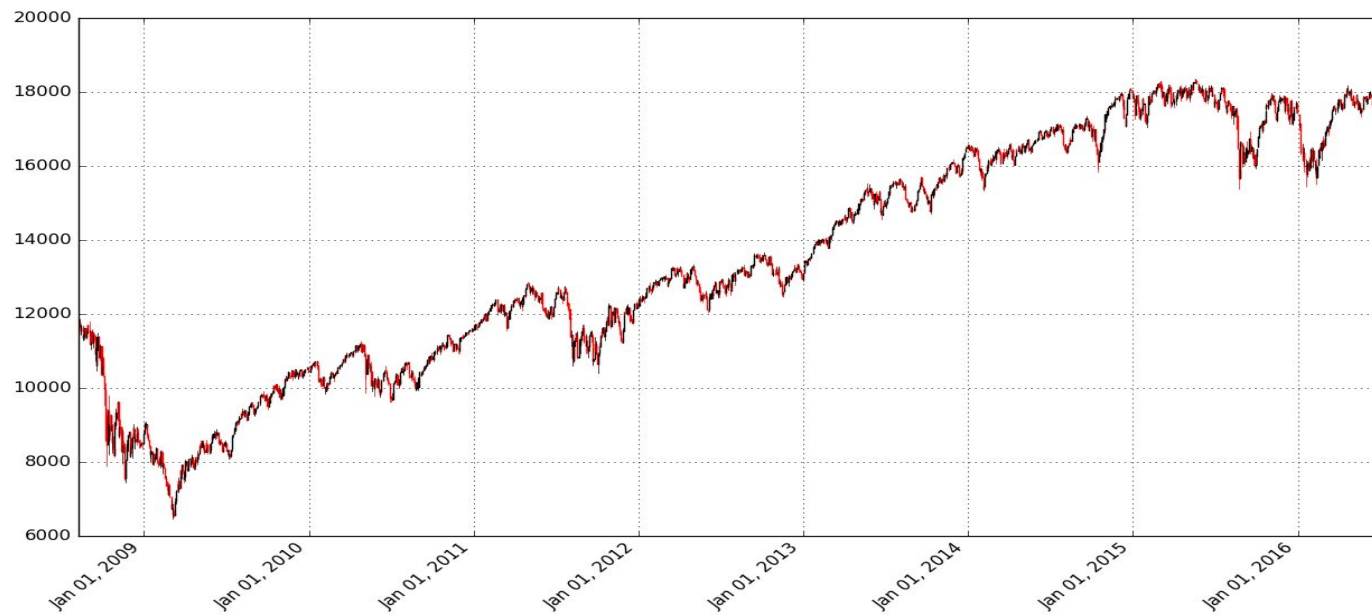
'Open', 'High', 'Low', 'Close', 'Volume', 'Adj Close'

Predictor: 'Label' (1 if Open-Close >0 else 0)

**I MADE ALL MY MONEY IN
THE STOCK MARKET**

**MOSTLY TRADED ON
THE MEOW JONES**

Exploration Time...(candlestick issues)



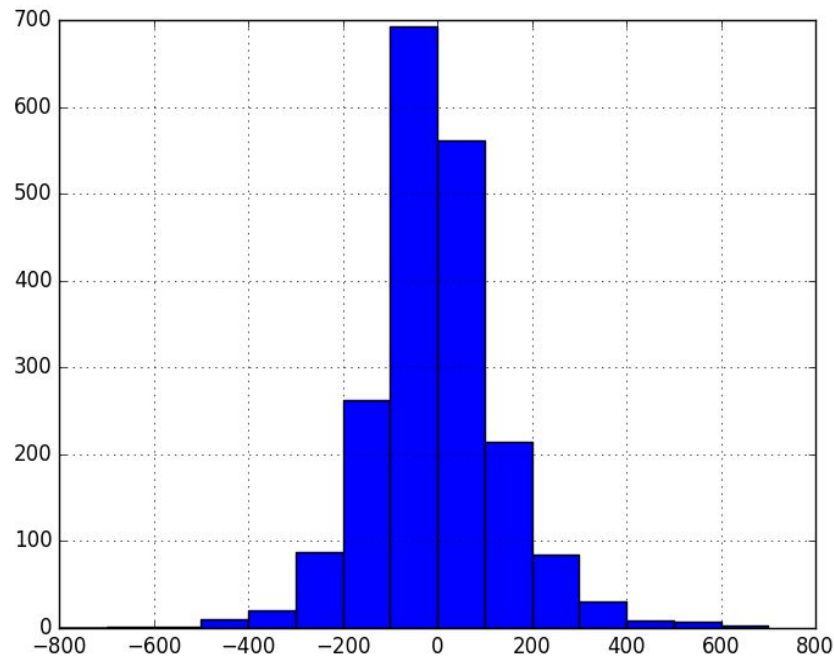
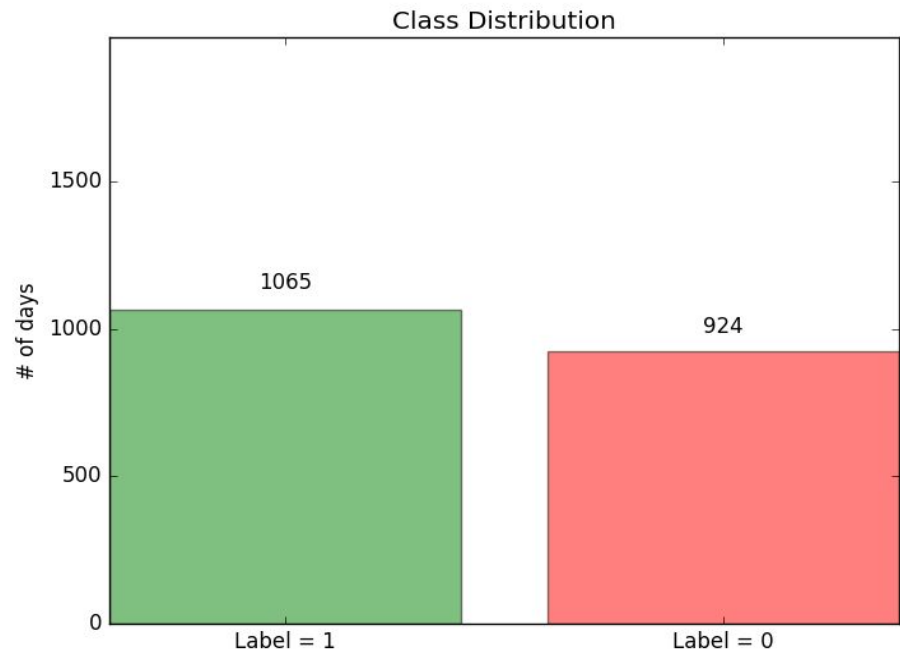
Exploration Time...(cont..)

06/08/2008 -> 07/01/2016

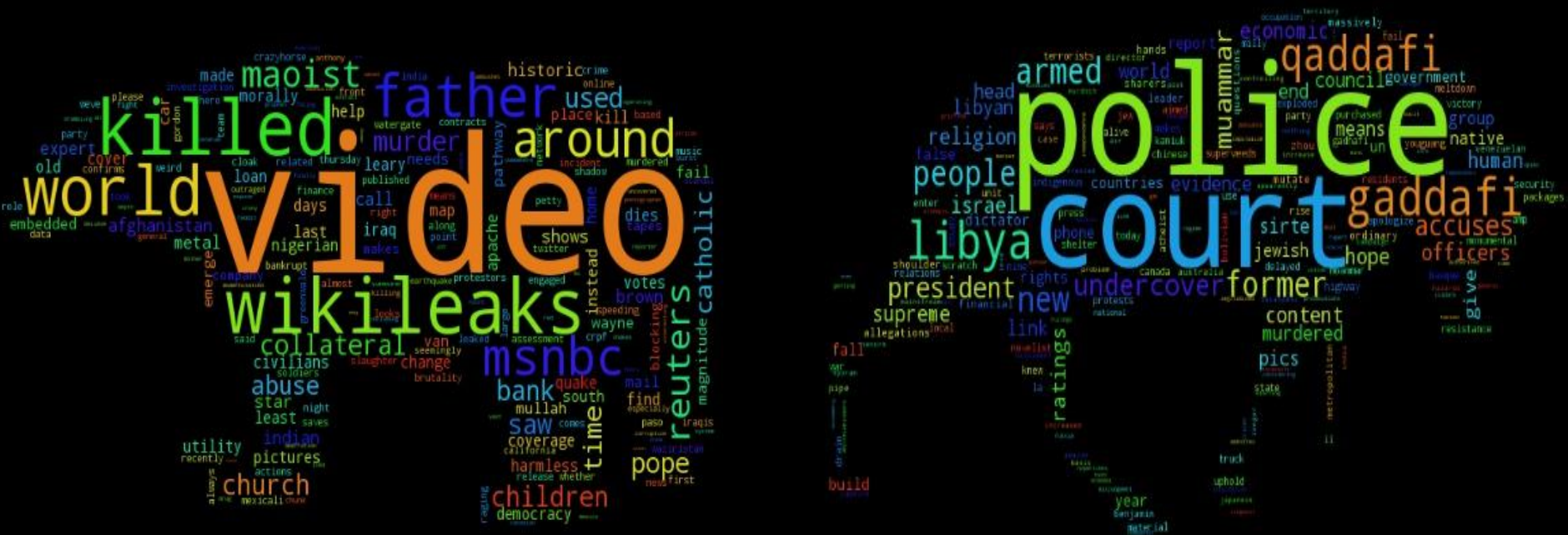
Stock Market went up 53% during that time (the goal to beat)

Can we beat it?!

Distribution



Word Clouds



We ran into problems...

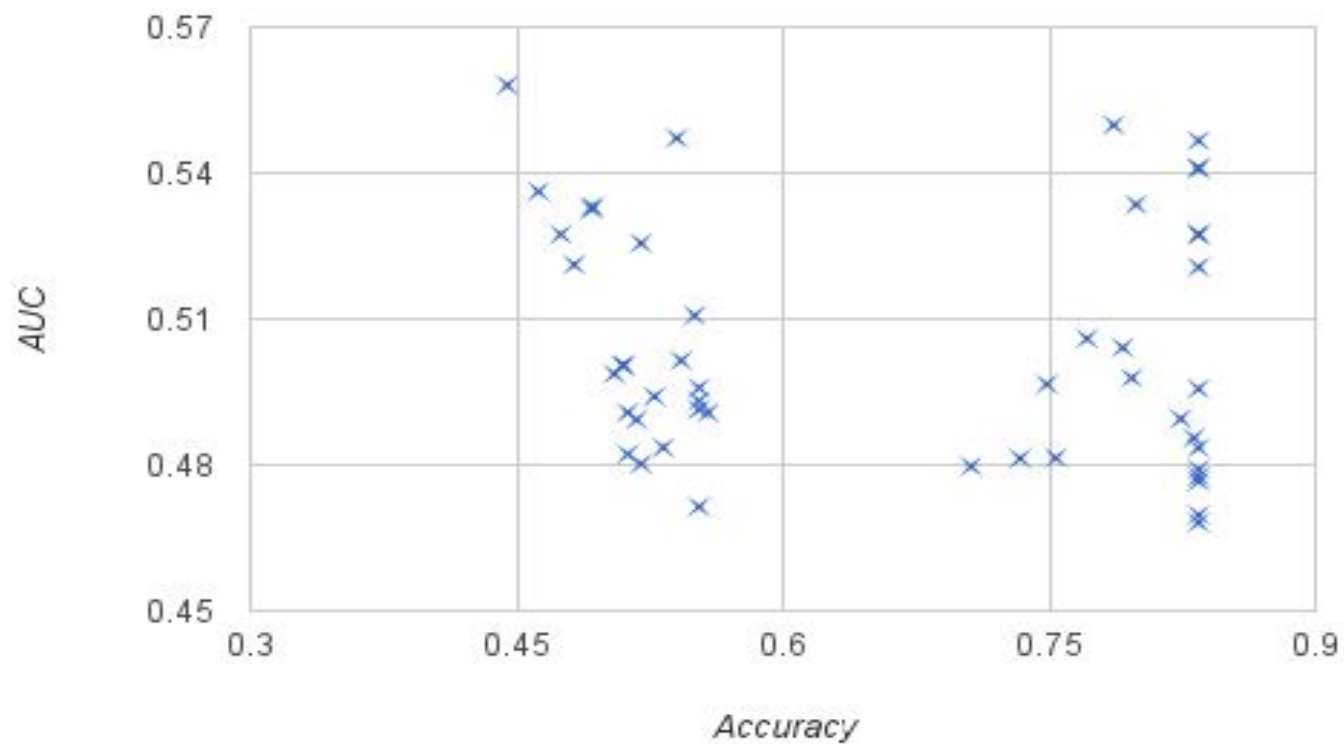
In our first or second checkpoint, we reported accuracies of 80% and better

But..we split our training and testing data incorrectly..and some of our training data found its way into our test data as well.

The data was quite messy and CountVectors and Td-idfVectors required stemword removal and stemming

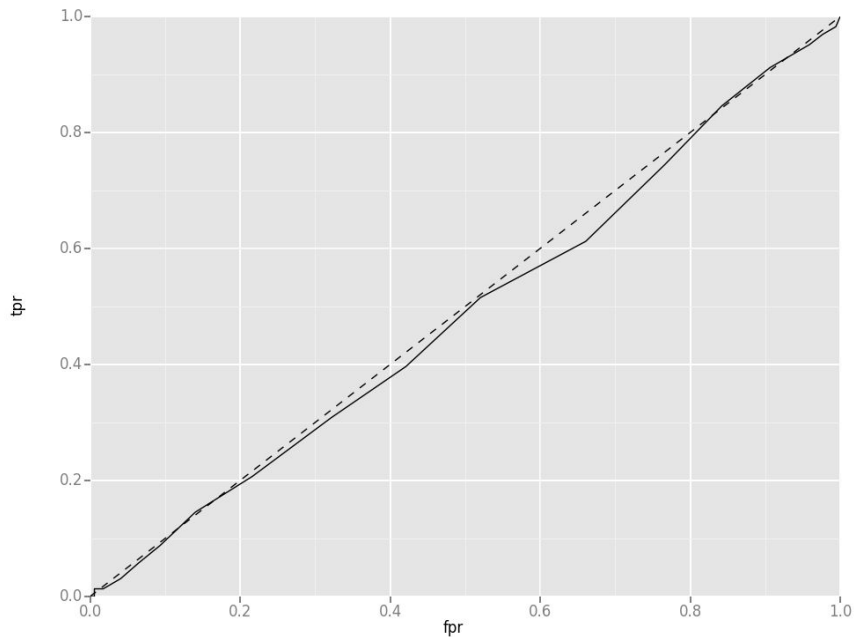
95% of words found when the stock market went down were also found in news articles when the stock market went up

AUC vs Accuracy

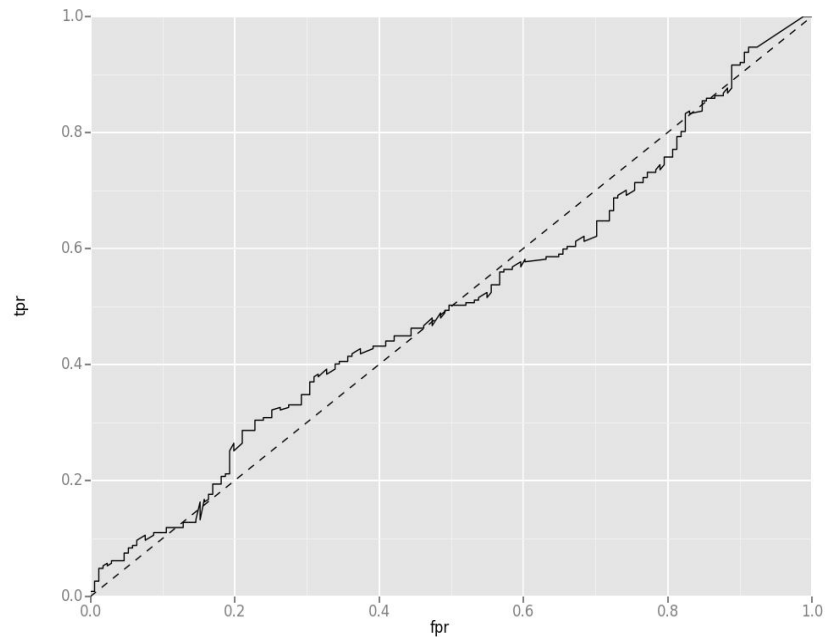


Our Best AUC Curves...

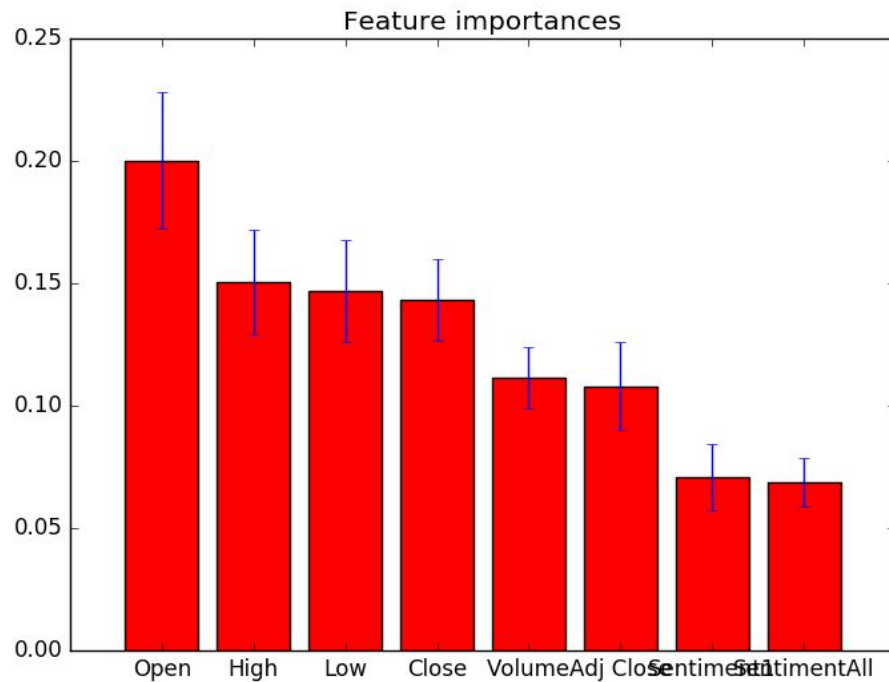
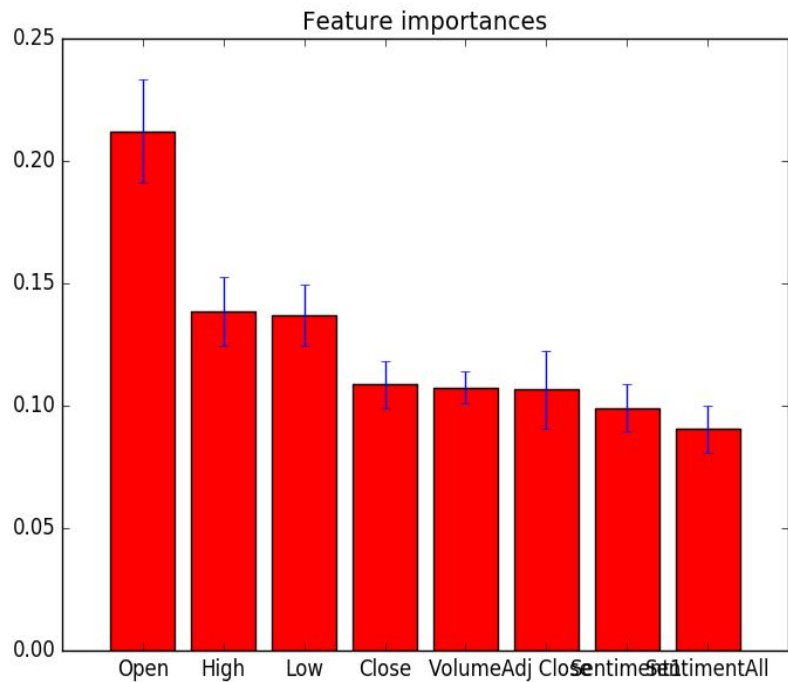
Roc Curve of KNeighborsClassifier

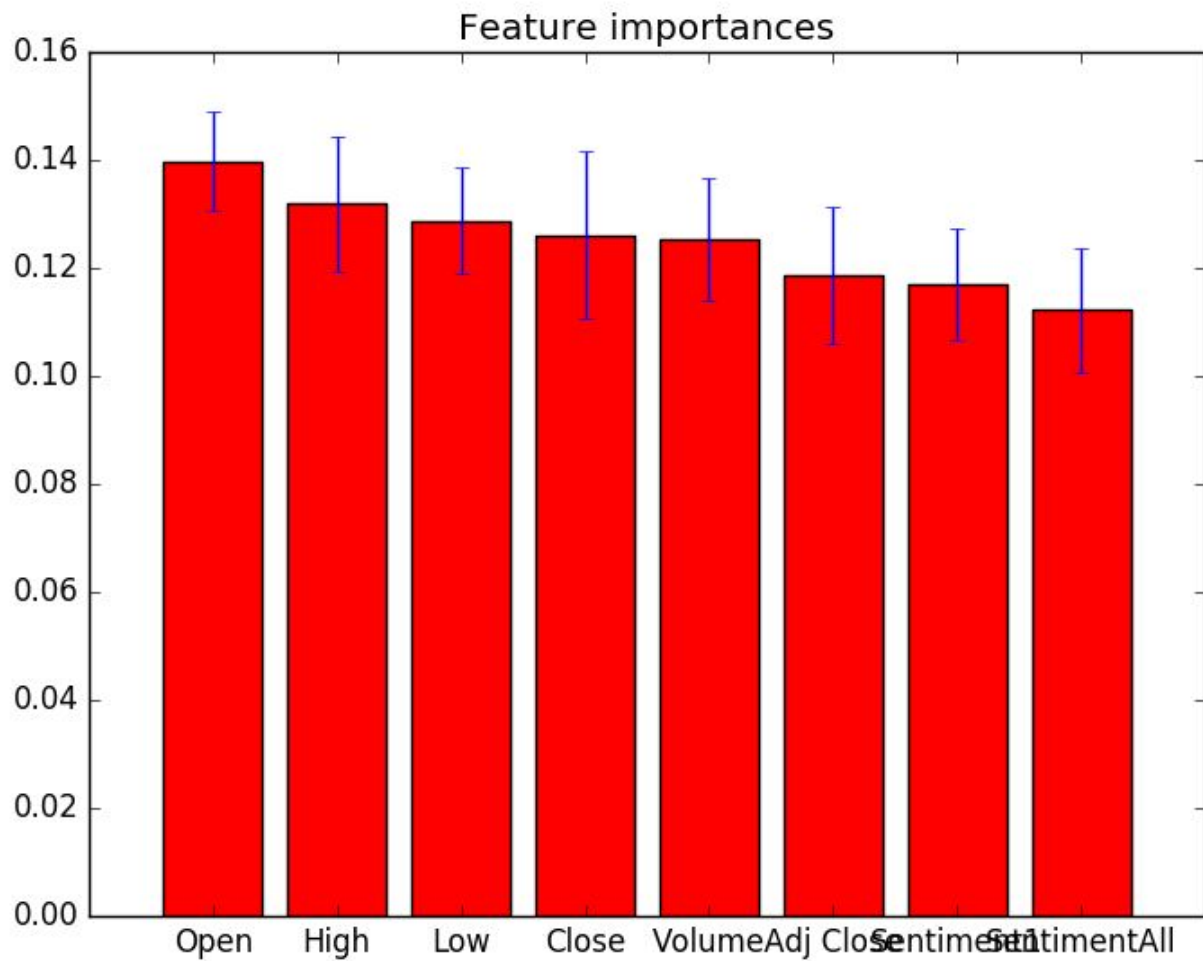


Roc Curve of SVC



Feature Selection





Conclusions

Given the data that we have, we can't accurately predict the stock market

Dataset was created for Deep Learning course, maybe a deep learning approach would be beneficial

There's a lot a foreign news that doesn't impact a the stock market. What would happen if we tried a different text source?

What if we analyzed Apple Reddit News vs Apple Stock?