

# Predicting the Behavior of Financial Markets from News Headlines



Kate Mohajery

GA Data Immersive Program Capstone

Feb 10 2017

# Overview

- ❖ Introduction
- ❖ Datasets
- ❖ Data processing
- ❖ Models
- ❖ Results



# Introduction

- ❖ Behavioral Economics believe in importance of emotions in human decision making process and its influence on market
- ❖ Online methods replacing traditional surveys for getting public's sentiment and mood state
- ❖ Financial market prediction based on online sentiment tracking

**In this project, market trends were predicted using news headlines**



# Datasets

- ❖ News Headlines
  - New York Times business news headlines (Jan. 2016 - Jan. 2017)
  - Reddit hot business news headlines (Aug. 2008 -Aug. 2016) from kaggle
- ❖ Market Indicators
  - Dow Jones Industrial Average (DJIA)
  - Volatility index of S&P 500 by CBOE (VIX)
  - Gold Price <http://www.gold.org/>
- ❖ Dictionary
  - List of negative words (Harvard Psychosociological Dictionary )
  - List of negative words developed for financial texts developed by Loughran and McDonald at University of Notre Dame



# Data Processing

Calculating negative news sentiment score (NNSS)

- Count the number of words in each headline that is listed as negative in the dictionary
- Divide by total number of words in the headline
- Add this ratio for all headlines in a day
- Divide by total number of news headlines in that day

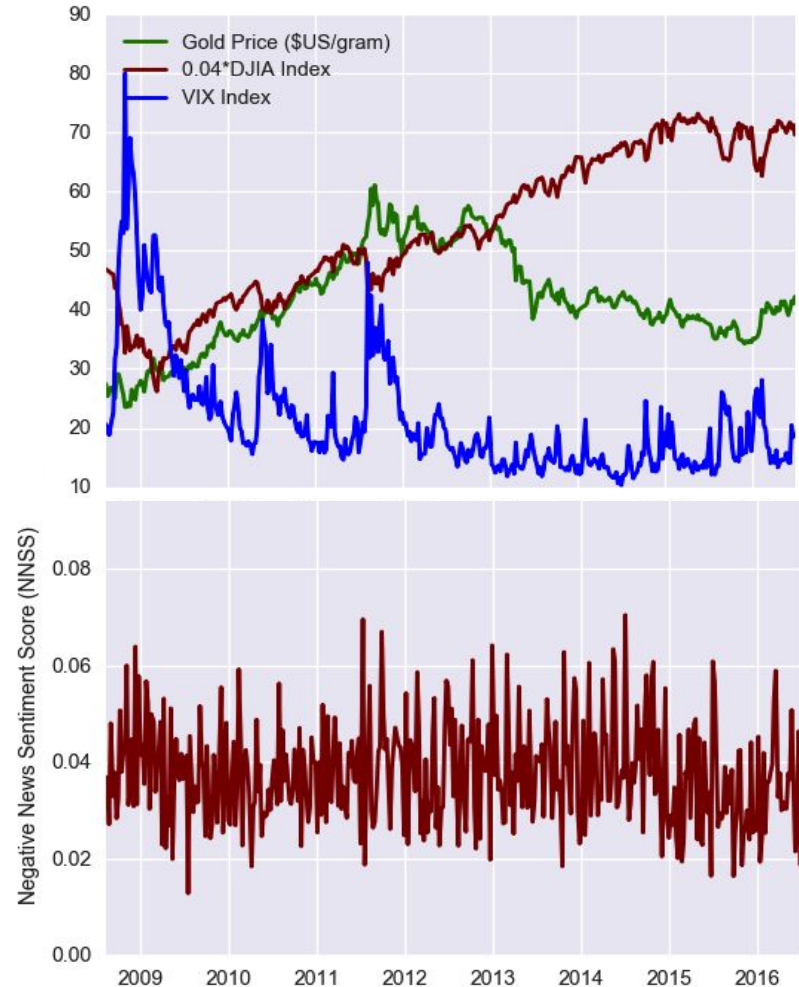


[illegible]

A 3D graphic where the word 'MARKET' is spelled horizontally and 'FUND' is spelled vertically, intersecting at the letter 'N'. The letters are represented by white blocks with black outlines, except for the 'M' in 'MARKET' and the 'T' in 'FUND', which are red blocks. The blocks are arranged on a white background with soft shadows.

# Exploring Datasets

- ❖ DJIA and VIX have inverse relation
- ❖ Negative News Sentiment Score is very noisy- no visible patterns



# Modeling

- ❖ Granger causality of market indicators and NNSS
- ❖ NNSS2 is 2-day moving average of negative news sentiment score
- ❖ VIX index has the highest correlation with NNSS
- ❖ Linear Regression: non-satisfactory results

Correlation matrix between market indices and NNSS

	NNSS	NNSS2 *	DJ Index	VIX Index	Gold price
NNSS	1.000	0.724	-0.055	0.082	0.026
NNSS2	0.724	1.000	-0.077	0.110	0.037
DJ Index	-0.055	-0.077	1.000	-0.682	0.183
VIX Index	<b>0.082</b>	<b>0.110</b>	-0.682	1.000	-0.408
Gold price	0.026	0.037	0.183	-0.408	1.000





# Classifying Market as Positive or Negative

- ❖ Closing price of DJIA higher than opening value  $\Rightarrow$  Class 1
- ❖ Closing price of DJIA lower than opening value  $\Rightarrow$  Class 0
- ❖ Using bag of words and TF IDF method on news headlines
- ❖ Models: random forest, support vector machine, gradient boosting and logistic regression



# Results

- ❖ Random forest and Support Vector Machine most promising results
- ❖ Employing n-grams of 2 and 3 improved the model
- ❖ Using previous day headlines Increased accuracy to 0.69

## Summary of scores for models

Model	AUC Score	Accuracy score
Random Forest	0.63	0.64
SVM	0.63	0.65

\* Baseline Accuracy = 0.53



# Conclusion

- ❖ Using news headlines improved the prediction of market trends
- ❖ Low correlation between negative news sentiment score and market indicators and no overlap between most predictive words and negative lists suggest that the negative word dictionary used was not appropriate



# Recommendations

- ❖ Use of more market related news websites such as bloomberg, forbes or Financial Times
- ❖ Better weighting method or word lists required for calculating negative news sentiment score
- ❖ Weak signal from news headlines can be amplified if used along other online sentiments such as search engine data ( such as Google Insights for Search ) or social media data (twitter data)



## Most important words

