REDDIT NEWS HEADLINES AND STOCK PREDICTION

DEVELOPED BY: KHANH HO

APRIL 2020

DOES REDDIT NEWS DRIVE DOW JONES INDUSTRIAL AVG CLOSE PRICES?

- Can we find a connection between the news headlines and the stock direction?
- Stocks are subject to supply, demand and patterns of human behavior
- Can we uncover a pattern?

WHY IS THIS QUESTION IMPORTANT?

- There's a lot to learn if we uncover some patterns
- There's financial incentive to find a model for all investors and traders

APPROACHES

- Natural Language Processing
- Technical Analysis

NATURAL LANGUAGE PROCESSING

- Natural Language Processing(NLP) breaks down text for machine learning algorithms and to predict outcomes
- Processed Reddit News Headlines from 2008-2014

NLP MODEL OUTPUTS

- Used four Machine Learning algorithms
- Trained and tested the model on data from 2013-2014
- We define accuracy as when the model predicted the correct market direction

Regression	Accuracy
Logistic	57%
Naïve Bayes	57%
Random Forest	55%
SVM	56%

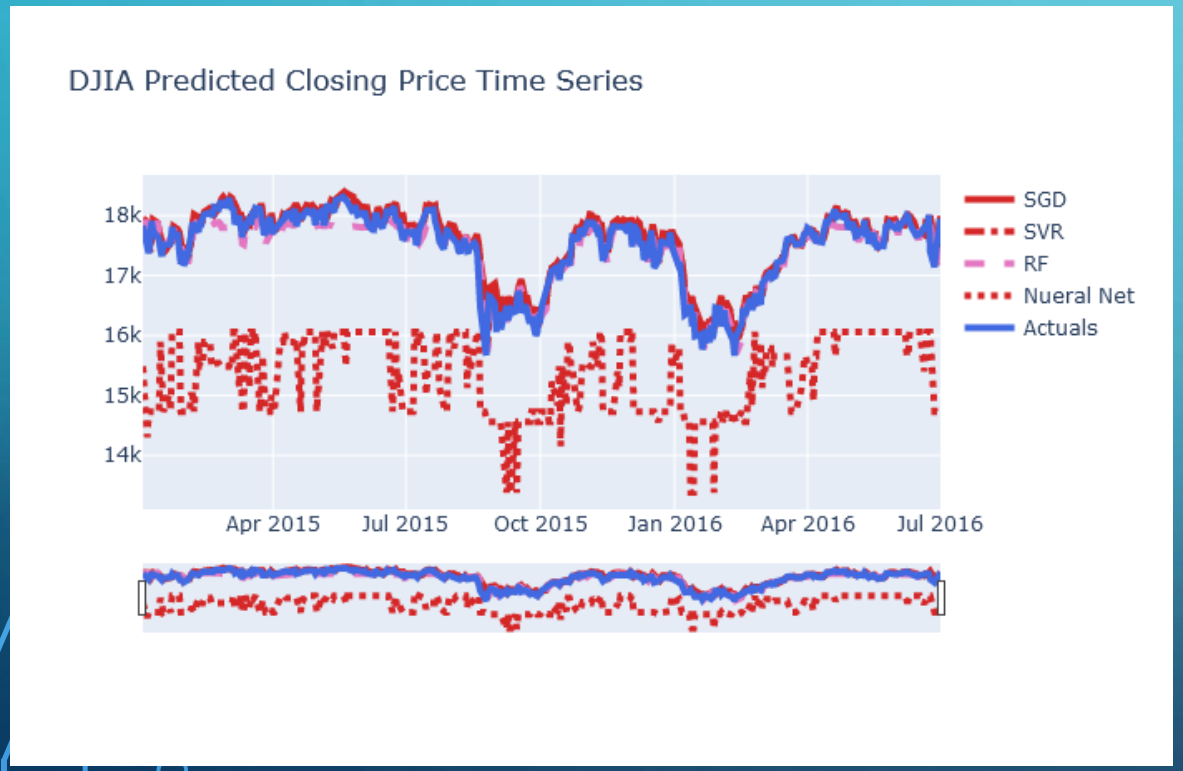
- Not the best accuracy so let's try another method!

TECHNICAL ANALYSIS

- Feature Engineering includes:
  - Previous Day's Closing Price
  - Rolling Average of Closing Prices and Volume
  - Standard Deviation of Volume and Ratio of Volume Rolling Averages
  - Return ratios and previous day return ratios

MODEL OUTPUT

- 4 regression models
- Entire forecast on the left and 5 day forecast on the right



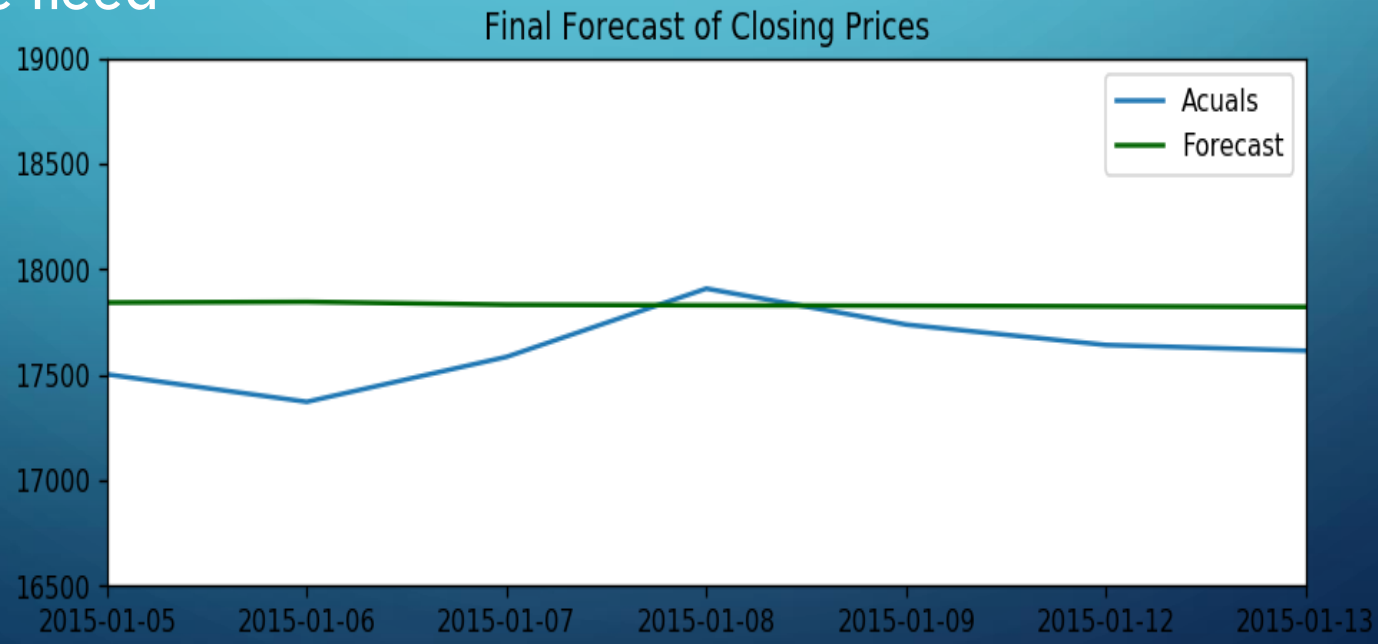
MODEL EVALUATION BY ERROR SCORING

- Best model scores are highlighted in green
- The errors are significant enough to make the models infeasible for investing

	Entire Forecast		5 Day Forecast			
Model	MSE	MAE	MAE	MAPE	ME	RMSE
SGD	41382	153	238	0.014	88	266
RF	42363	163	203	0.012	60	245
SV	31225	132	212	0.012	6	240
NN	4742582	2103	2845	0.161	-2845	2863

ARIMA FORECASTING

- ARIMA is another time series forecasting algorithm
- We predict the next 5 days of the forecast but the model doesn't have the precision we need



NEXT STEPS

- We learned what works and what didn't work
- Stock prediction is not an easy project but we can fine tune the path we take from here
- More Feature Engineering
  - This method proved to be the most accurate
  - Explore different candle sizes and various stocks
  - Fine tune the model precision to make the model more useful
- Explore Facebook Prophet package and exponential smoothing