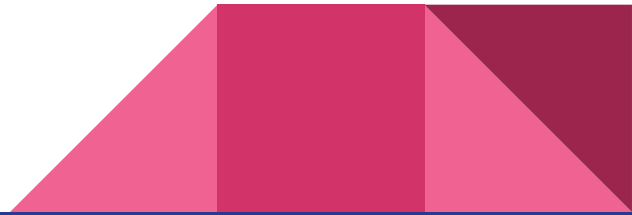
Stock Price Prediction

Final Presentation

Ethan Benville, Carson Earnest, Alex Reyna,
Jim Sherman, Conor Walsh

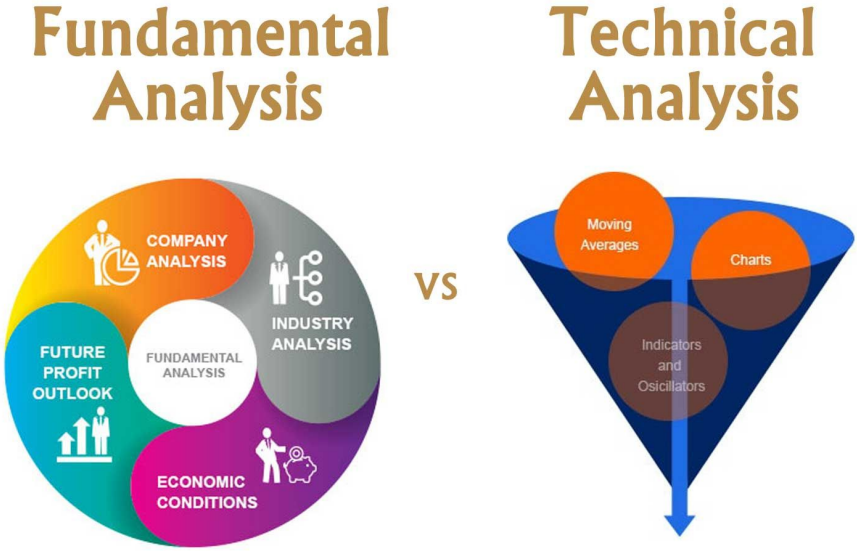
Intro/Background

- Analyzing stock prices for companies on S&P 500 index
  - Evaluated using variety of metrics like earnings or revenue
- Metrics used to assess financial performance affected by various factors
  - New business developments (ex. product launches, mergers)
  - Supply chain demand changes
  - Industry competitor sales growth
- As stock prices which take in these variables are extremely volatile, machine learning algorithms can be used to help predict future changes in stock prices based on historical data



Problem Definition

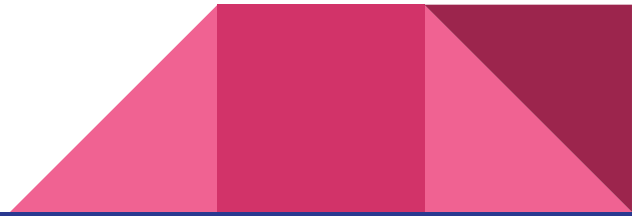
- Overall goal: Utilize ML algorithms to accurately predict future stock prices
- Focusing on technical analysis because:
  - Historical fundamental analysis hard to find
  - Fundamental analysis can be subjective
  - Technical analysis involves finding mathematical patterns in price data
- Expect algorithm to be better suited to predicting price for shorter time periods



Data Collection

- Yahoo Finance API to pull historical daily stock pricing data
  - Period: September 2010 until the end of 2020
  - Tickers: VOO, MSFT, AAPL, AMZN, GOOGL, NVDA, JNJ, DIS, BIIB, SYY, PPG, GIS, YUM, KHC, WHR, PKG, PTC, HD, MRO

- Pulled current financial metrics metrics of the same companies
  - Trading volume, 52 week high, 52 week low, price to book, forward PE, trailing PE, shares outstanding, market capitalization



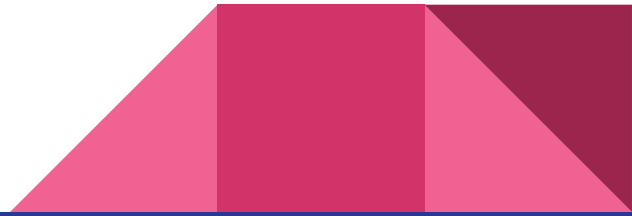
Data Preprocessing

- Created 5-day, 20-day and 260-day forward returns for each company's stock prices
  - Each time division corresponds to one week, one month and one year of trading days
  - Calculated using formula price of day i+[x] - price of day i)/(price of day i) where x in {4,19,259}
- Also created 10, 50, and 200-day moving averages for each company's stock prices
  - Calculated by pulling the average price over the specified period (10, 50 and 200 day window)

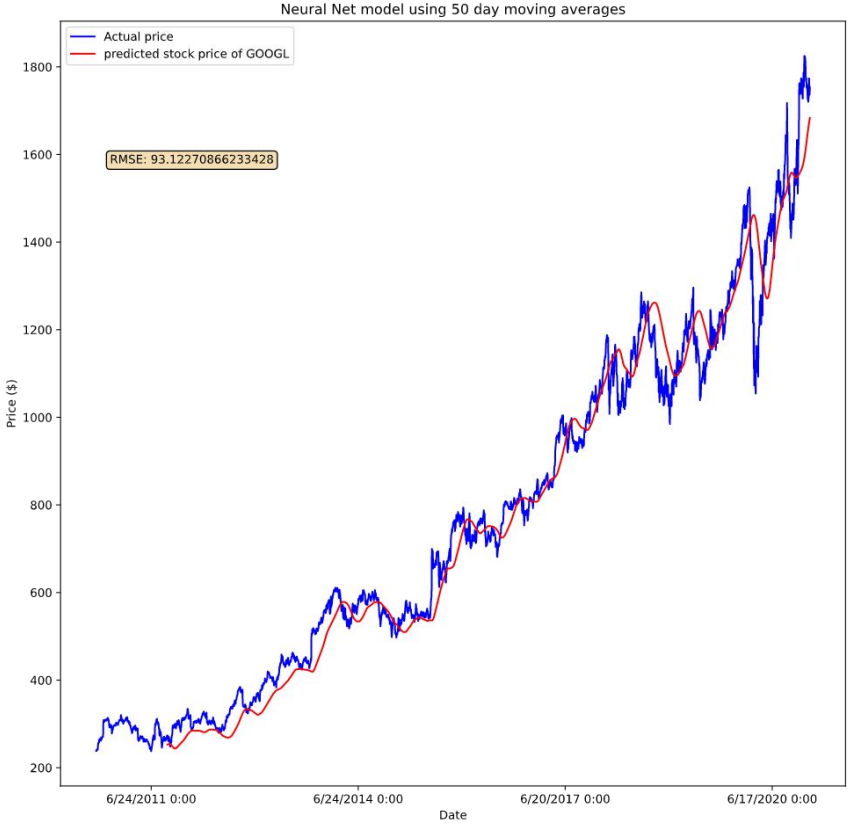
HD_10day_MA	HD_50day_MA	HD_5day_mvg_return
28.646	28.4372	0.029665314
28.723	28.4306	0.037955466
28.824	28.4174	0.070603675
28.896	28.4052	0.060057697
28.902	28.3902	0.022291343
28.9	28.3724	0.010483252
28.876	28.3566	-0.035443038
28.793	28.34	-0.053376365
28.755	28.3304	-0.036944583
28.742	28.326	-0.036224741
28.716	28.3218	-0.025172754
28.708	28.315	0.001491795
28.731	28.3092	0.024814531
28.754	28.3038	0.020880503
28.714	28.2866	0.016048144
28.65	28.2552	-0.024733269
28.572	28.2196	-0.064160881
28.458	28.1742	-0.068432154
28.197	28.1114	-0.047300526

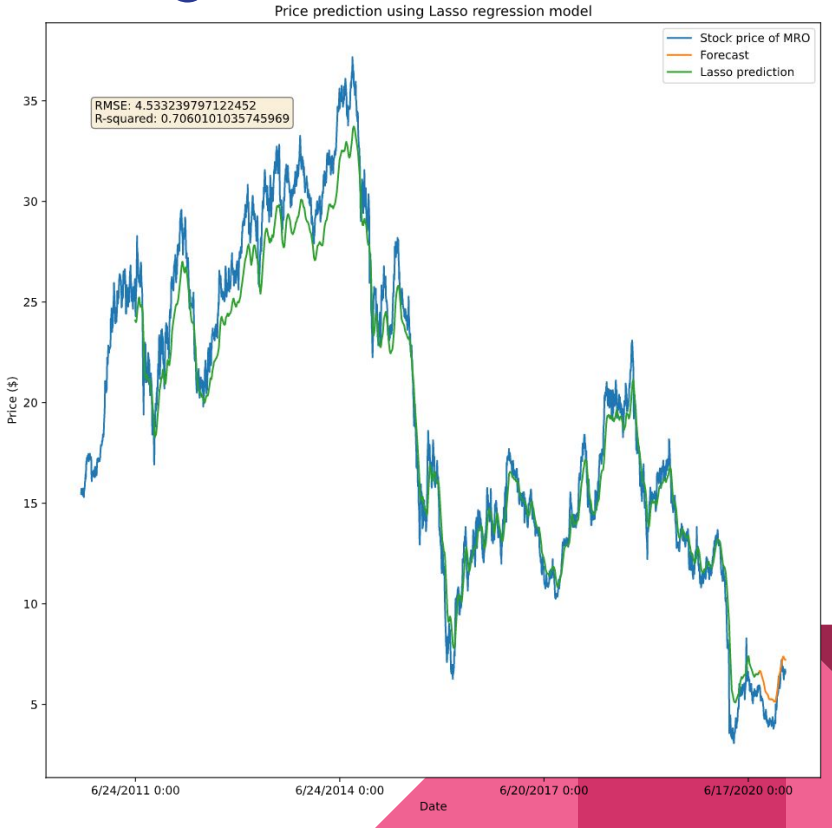
Methods

- Models
  - Ridge Regression
  - LASSO
  - Neural Network
- Tools
  - SKLearn
  - Keras
  - Tensorflow
- Error Measures
  - RMSE
- Training and Test Data
  - Stock Moving Average Data



Results: Stock Price Modeling





Results: Stock Price Modeling of GOOGL

RMSE and R-squared values of different models used to predict GOOGL stock price

Model	**RMSE value**	**R-squared value**
Ridge Regression	69.010904217	0.970958556
Lasso Regression	82.2146226	0.951645504
LSTM Neural Net using 10 day moving price averages	91.60613555	NA
LSTM Neural Net using 50 day moving price averages	93.1227086	NA
LSTM Neural Net using 200 day moving price averages	150.24497814	NA

Conclusion: Ridge Regression works best on predicting prices that follow a linear pattern

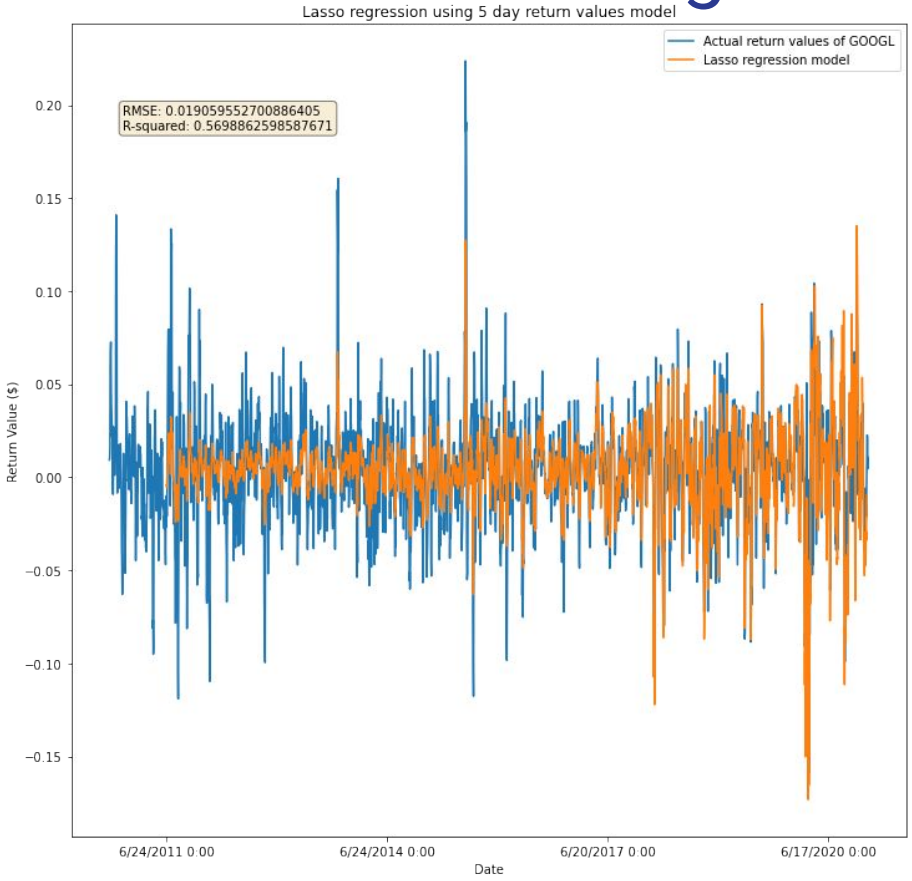
Results: Stock Price Modeling of MRO

RMSE and R-squared values of different models used to predict MRO stock price

Model	**RMSE value**	**R-squared value**
Ridge Regression	4.518773588	0.68083489
Lasso Regression	4.5332398	0.7060101
LSTM Neural Net using 10 day moving price averages	0.827341418	NA
LSTM Neural Net using 50 day moving price averages	0.261573764	NA
LSTM Neural Net using 200 day moving price averages	1.96069617	NA

Conclusion: Neural net performs best at predicting prices that fluctuate

Results: Predicting Return Values



Results: Predicting Return Values of GOOGL

RMSE and R-squared values of different moving return predictions of GOOGL stock using ridge regression and lasso regression

Model/what is being predicted	**RMSE value**	**R-squared value**
ridge/5 day return	0.0202955	0.59544785
lasso/5 day return	0.01905955	0.56988626
ridge/20 day return	0.03929537	0.5829511
lasso/20 day return	0.044526050	0.541465807
ridge/260 day return	0.12431293	0.4715617
lasso/260 day return	0.121992451	0.43286538

Results: Predicting Return Values of MRO

RMSE and R-squared values of different moving return predictions of MRO stock using ridge regression and lasso regression

Model/what is being predicted	**RMSE value**	**R-squared value**
ridge/5 day return	0.03985127	0.5938015
lasso/5 day return	0.030613120	0.617875485
ridge/20 day return	0.09171964	0.6481888
lasso/20 day return	0.09060504	0.5740656
ridge/260 day return	0.253353404	0.49152697
lasso/260 day return	0.256229780	0.55424029

Conclusion: Ridge regression and Lasso regression work the same!