Investing with Machine Learning

By Peter Scagnelli
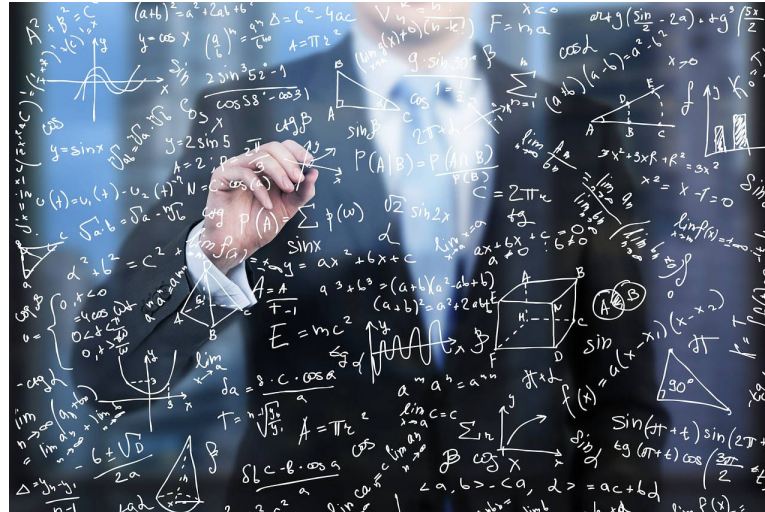
Why use Machine Learning for investing?

What are Stocks?

Quants

Data Sources

Yahoo API
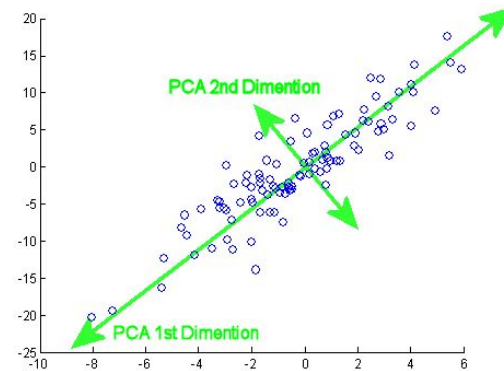

50,000 Reuters headlines

S&P company info

Machine Learning Models

Deep Learning
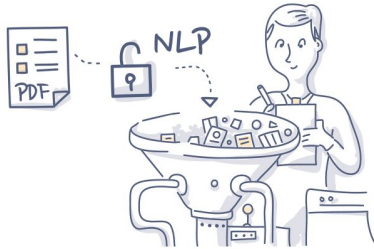
NLP

Time Series Analysis

PCA 2nd Dimention

PCA 1st Dimention

PCA

Time Series Methodology

Data

Stock Price

Model
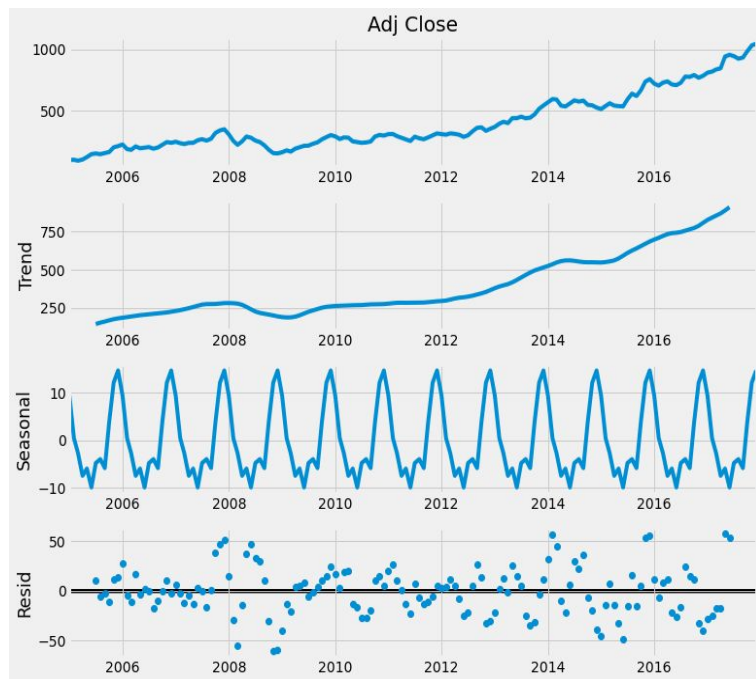
ARIMA

Output

Stock Prediction
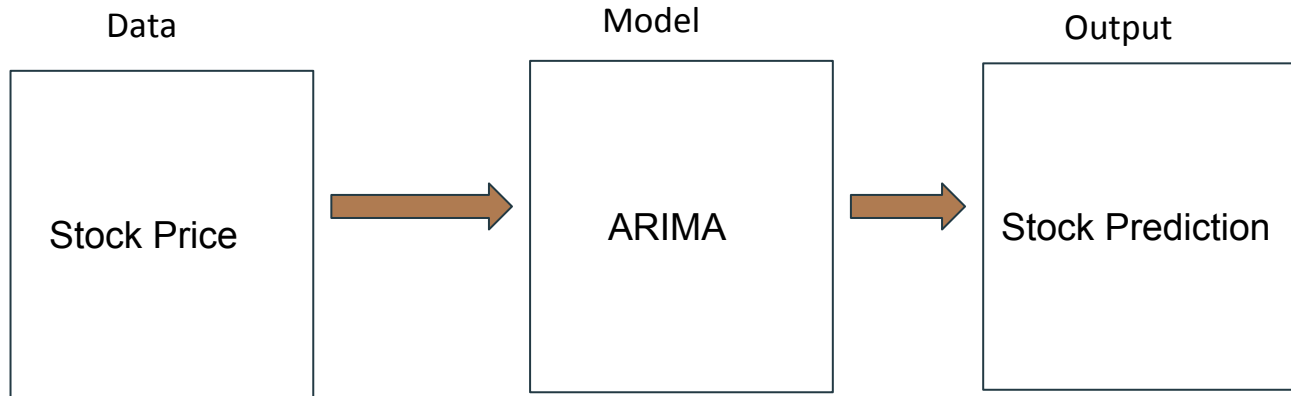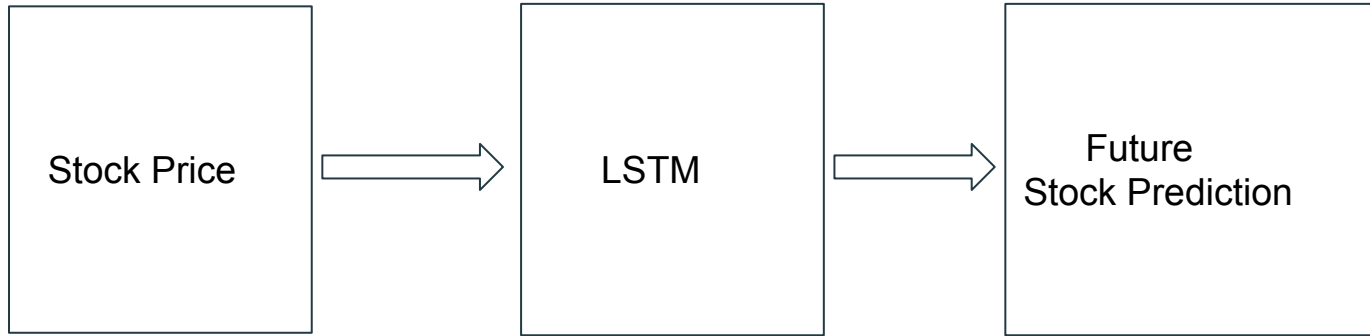
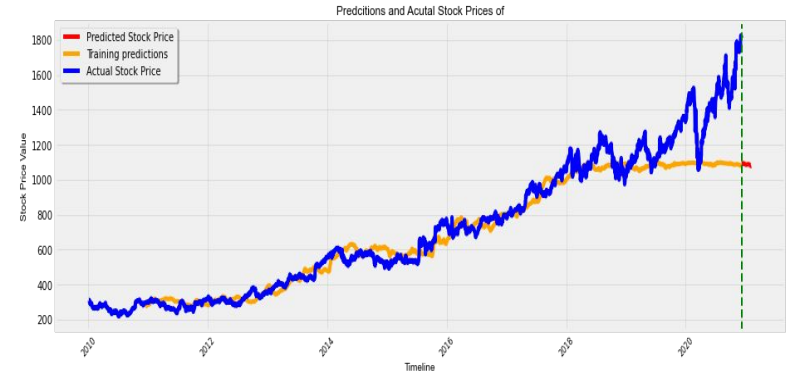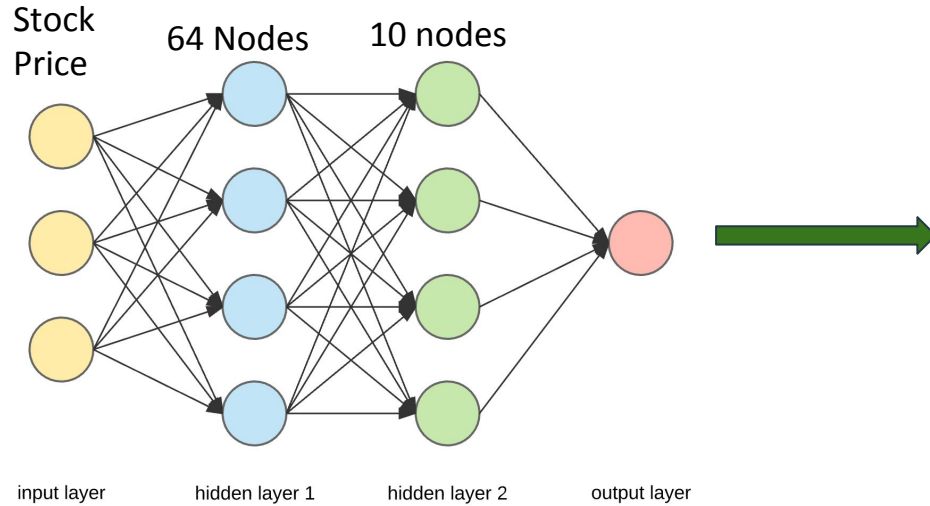# ARIMA Metrics
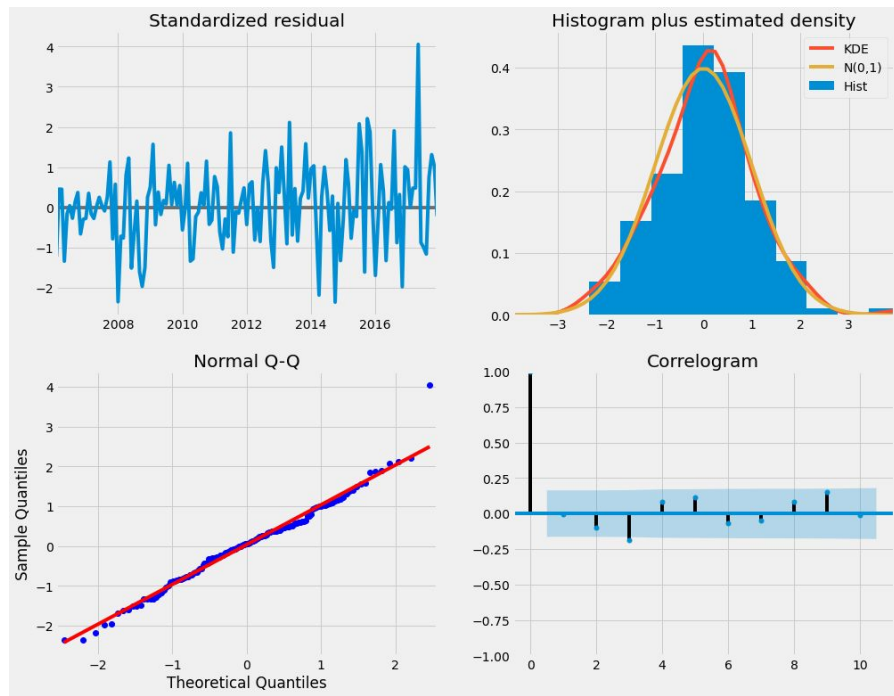
ARIMA confidence interval visualization
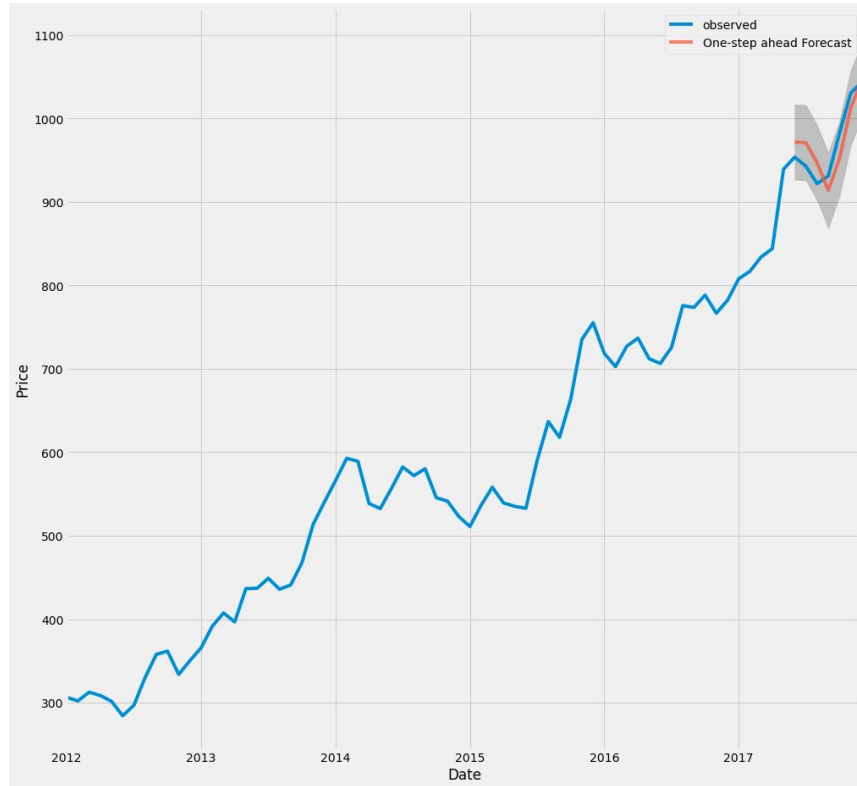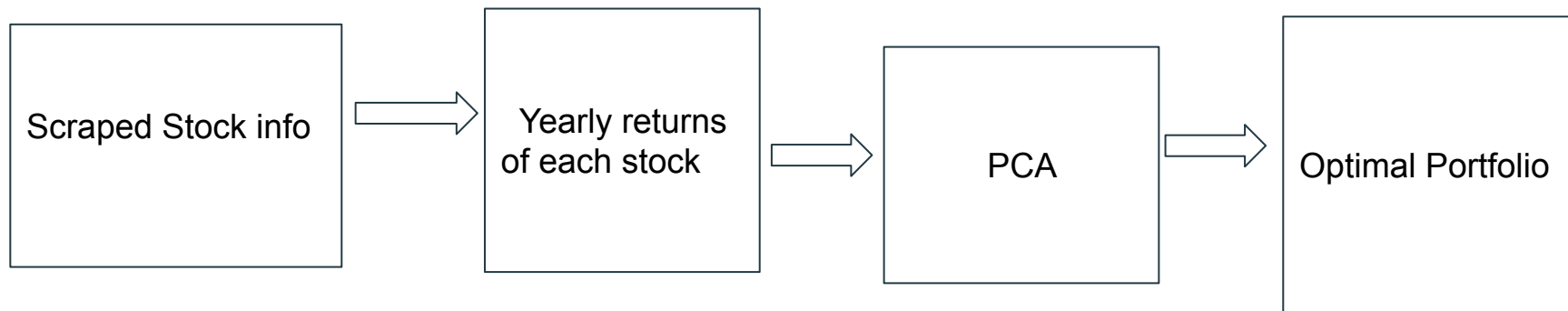

Time Series: AIC = 766.642
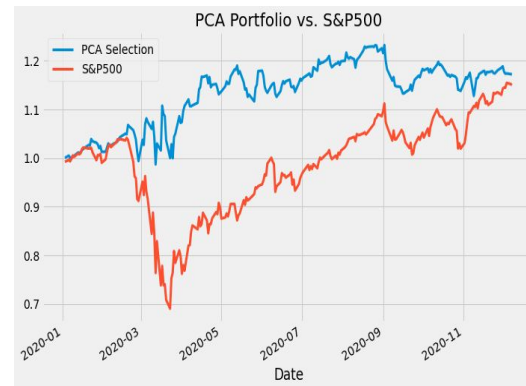
Time Series: MSE = 471.12

LSTM Methodology

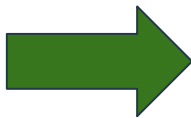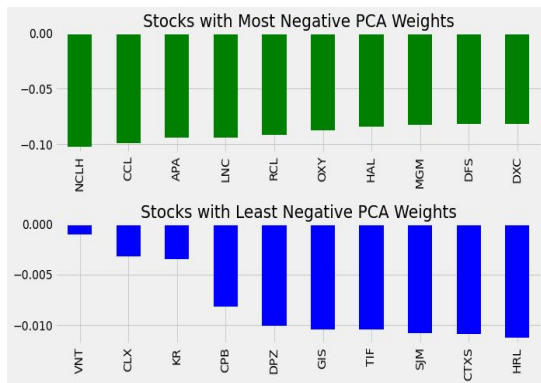# LSTM

PCA

Scraped Stock info → Yearly returns of each stock → PCA → Optimal Portfolio

Visualizations

NLP Sentiment Analysis

```
┌──────────────────┐      ┌──────────────────┐      ┌──────────────────┐
│ Stock price &    │ ──▶  │ Feature          │ ──▶  │ Sentiment        │ ──▶
│ News Dataset     │      │ Engineering      │      │ Analysis         │
└──────────────────┘      └──────────────────┘      └──────────────────┘


┌──────────────────┐      ┌──────────────────┐      ┌──────────────────┐
│ Prediction       │ ──▶  │ Regression &     │ ──▶  │ Output           │
│                  │      │ XG Boost         │      │                  │
└──────────────────┘      └──────────────────┘      └──────────────────┘
```
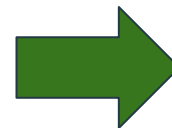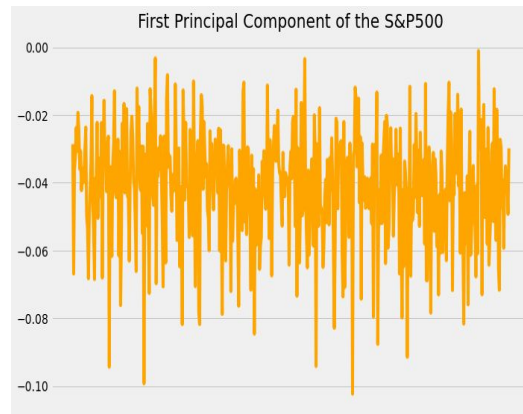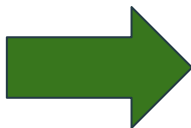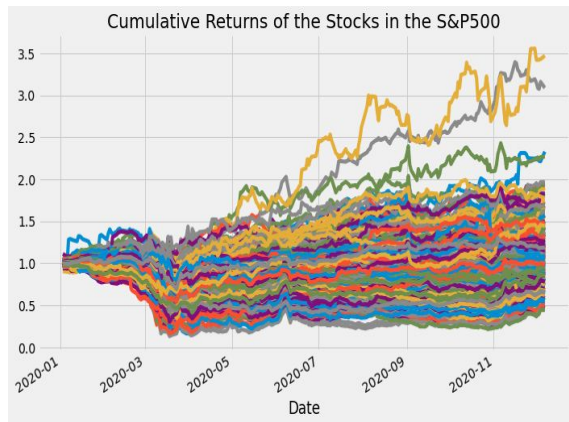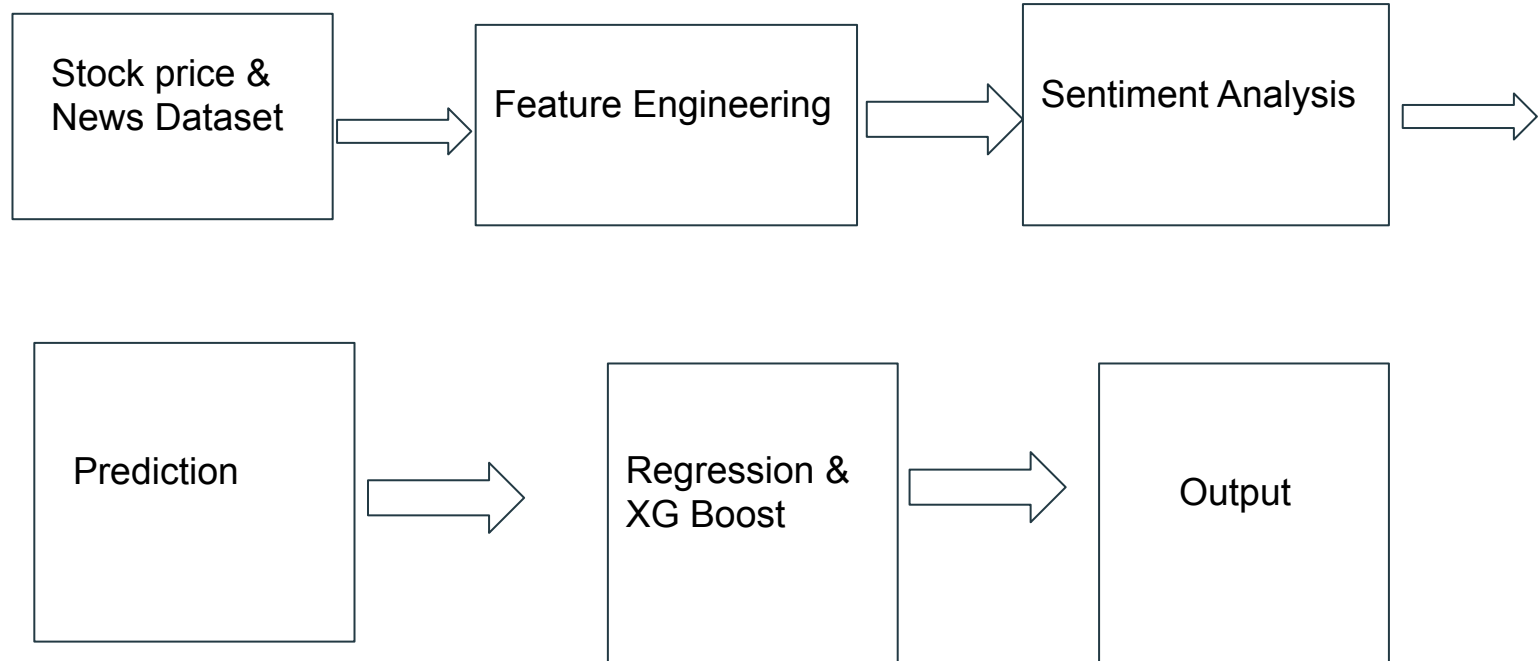
Results

Ridge Regression   Accuracy = 53.19%

XGBoost Model   Accuracy = 50.00%

Takeaways

- ML based trading works with very very good data and people who can gather extremely good insights
- These predictions were not meant to be a strategy and should not be used for any investment purposes. They do have applications in the field of Quantitative Trading, however I am not recommending that anyone use them.