Stock Analysis and Forecasting based on Airline Reviews

Milestone 3

By: Prajakta Gaydhani
Advisor: Dr. Carol Romanowski

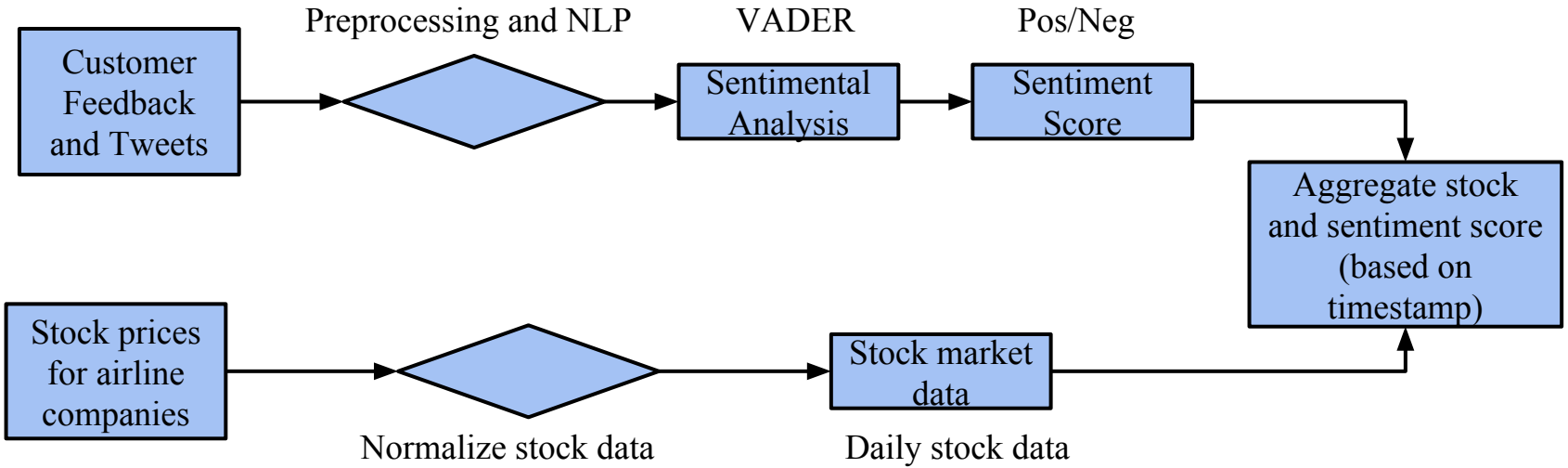
Problem Statement

Opinion Mining of different customer reviews on major Airlines in the United States and analyzing the effect of customer's sentiments on the stock market of different airline companies.

Goal of the Project

1. Correlation between customer reviews and stock price movement of different Airline Companies.
2. Accurately forecast stock prices for different Airline Companies using customer reviews and stock market data.

Review: Milestone 1 and Milestone 2



Milestone 3

1. Correlate customer's sentiments with stock market movement of different airline companies
2. Forecasting stock market prices using customer's sentiments and stock market data
3. Evaluate Results

Dataset

Date	Pos Count	Neg Count	Total Count	Open	High	Low	Close
9/27/18	10	21	31	21.41	21.42	19.10	19.30
9/28/18	17	13	30	19.30	20.53	19.20	20.52
9/29/18	20	15	35	20.44	20.58	20.10	20.21

- Removed noise, seasonality and trend from stock prices
  a. Exponential Smoothing
  b. Used statsmodels

Correlation - Stock Movement and Sentiment

- **Stock Movement (Rise/Fall):** For each day
  - If previous day's closing price > today's closing price:
    - Close_Rise_Fall = -1
  - Else if previous day's closing price < today's closing price:
    - Close_Rise_Fall = 1
  - Else:
    - Close_Rise_Fall = 0
- **Sentiment score:** For each day
  - If no_neg_tweets > no_pos_tweets:
    - Sentiment = -1
  - Else if no_neg_tweets < no_pos_tweets:
    - Sentiment = 1
  - Else:
    - Sentiment = 0

Correlation Analysis

corr() function in python to find the correlation between Sentiments and stock price movement.

American Airlines	Close_Rise_Fall Semtiment/Polarity	0.945322
United Airlines	Close_Rise_Fall Semtiment/Polarity	0.936723
Alaska Airlines	Close_Rise_Fall Semtiment/Polarity	0.899944
JetBlue Airlines	Close_Rise_Fall Semtiment/Polarity	0.908289
Delta Airlines	Close_Rise_Fall Semtiment/Polarity	0.923722

Forecasting Stock Prices using user sentiments

Models

1. Models Built
    - Machine Learning
    - Deep Learning (LSTM)
2. Evaluation Metrics
    - Mean Squared Error (MSE)
    - Root Mean Squared Error (RMSE)
    - R^2 error (for regression models)

Data Preparation

- Normalized the data before forecasting
  - Used MinMaxScaler() from scikit-learn
  - Range = (0,1)
- Denormalized the data after forecasting
- Train/Test split
  - Train Data: Jan 2010 - Dec 2016
  - Test Data: Jan 2017 - Oct 2018

Machine Learning Models

- Used scikit-learn and Python 3.6
- Tuned hyper-parameters using GridSearchCV
- Models Built:
  - Random Forest Regressor
  - AdaBoost Regressor
  - Support Vector Regression (SVR)
  - K Nearest Neighbors Regression (KNN)
  - Gradient Boosting Regressor

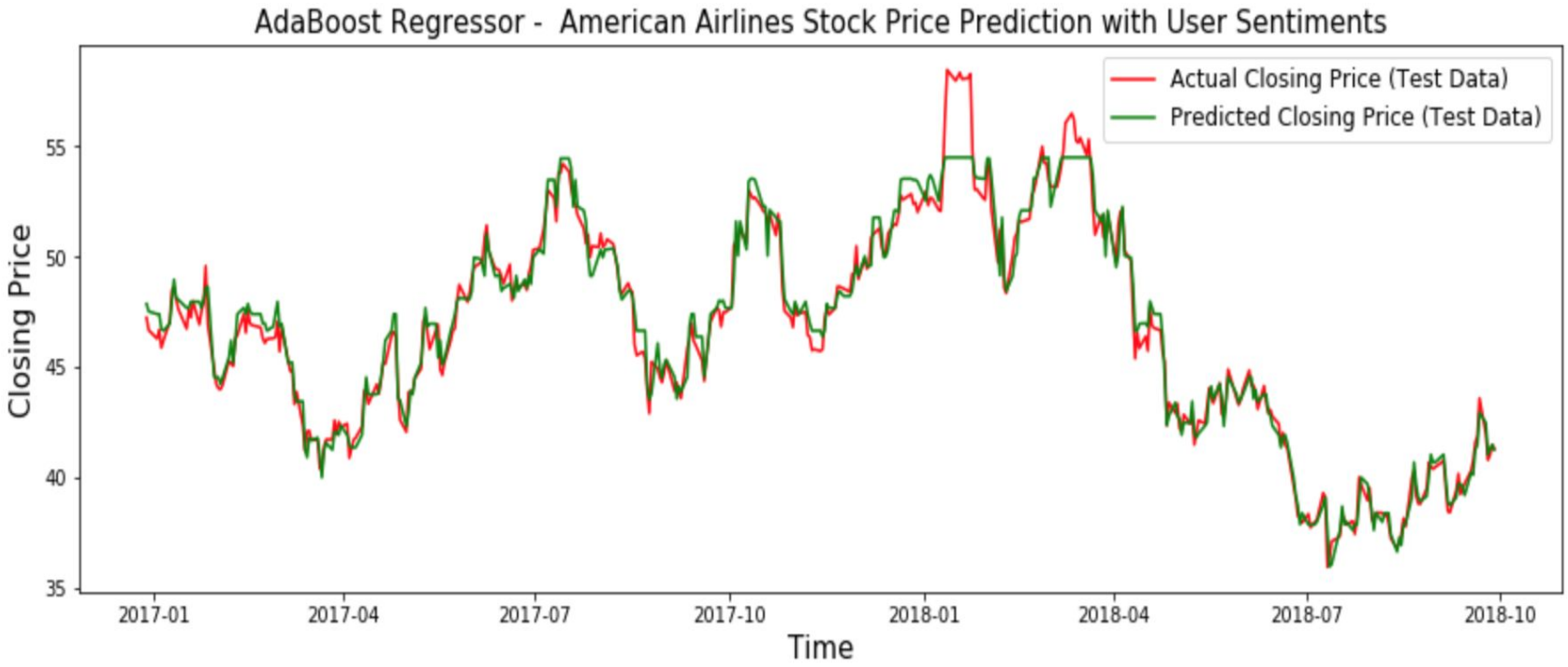
Results - American Airlines

Comparing Results - American Airlines

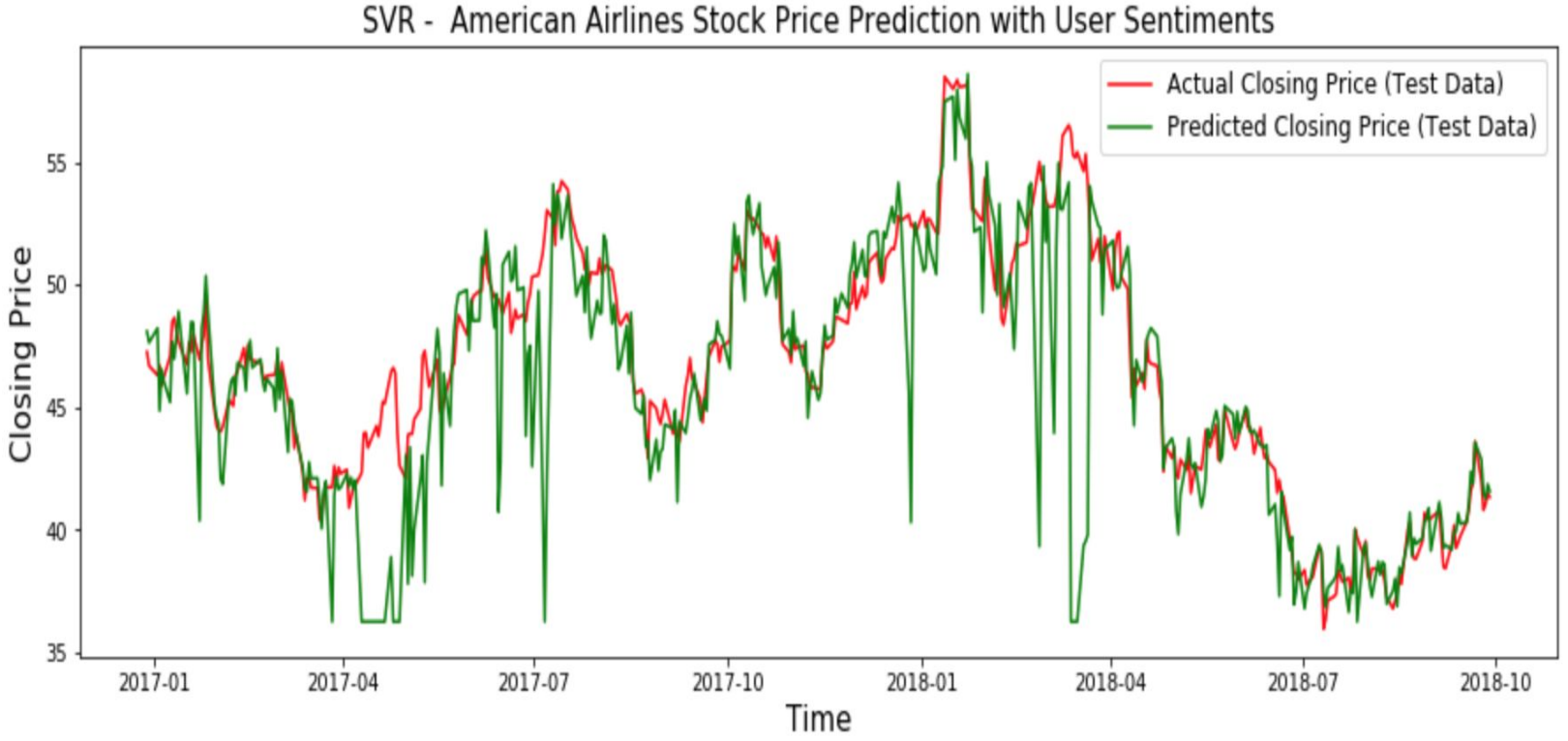
Model	MSE	RMSE	R2 error
Random Forest Regressor	0.47	0.63	0.9832
AdaBoost Regression	0.62	0.78	0.9745
Support Vector Regression(SVR)	11.82	3.43	0.5137
KNN Regressor	7.56	2.74	0.6889
Gradient Boosting Regressor	0.34	0.58	0.9856

Results - Random Forest Regressor

Results - AdaBoost Regressor



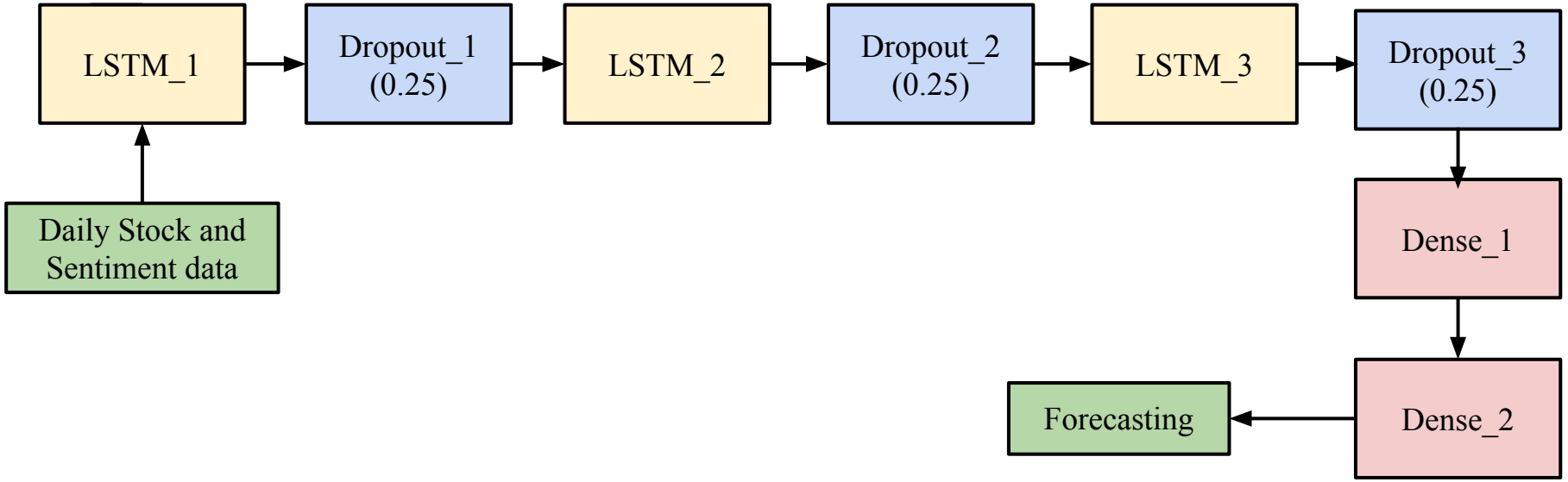
Results - Support Vector Regressor



Results - KNN Regressor

Results - Gradient Boosting Regressor

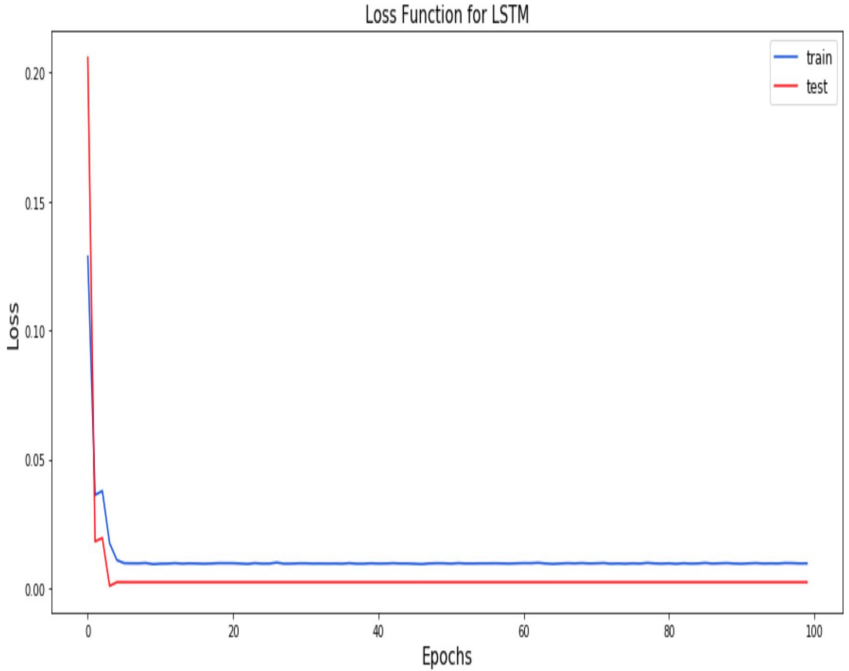
Deep Learning - Architecture (LSTM)



Hyperparameters for LSTM

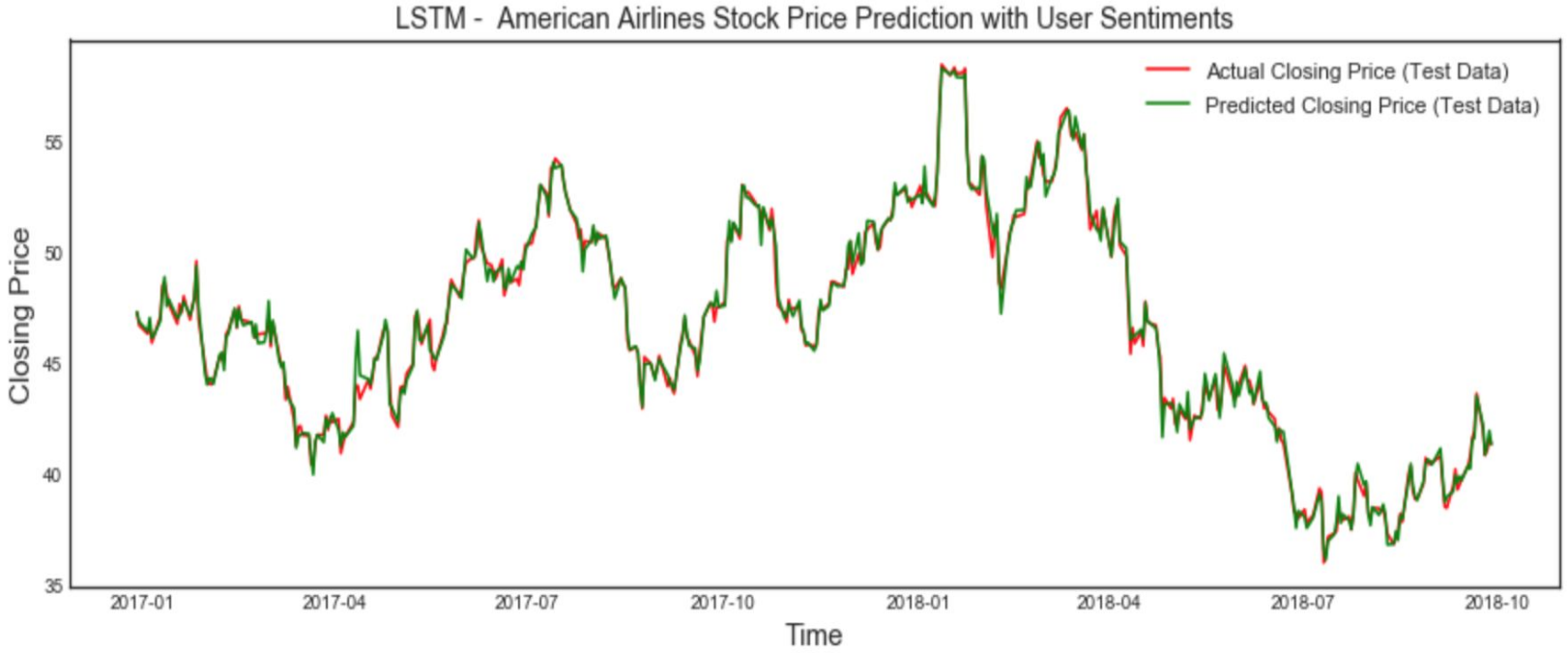
- Train/Test/Validation: 80/10/10
- Used Keras and TensorFlow
- Batch size : 64
- Epochs: 100
- Learning Rate: 0.0005
- Optimizer: Adam
- Activation Function: Relu
- Regularization: Dropout (0.25)
- Loss Function: Mean Squared Error (MSE)

Results - LSTM

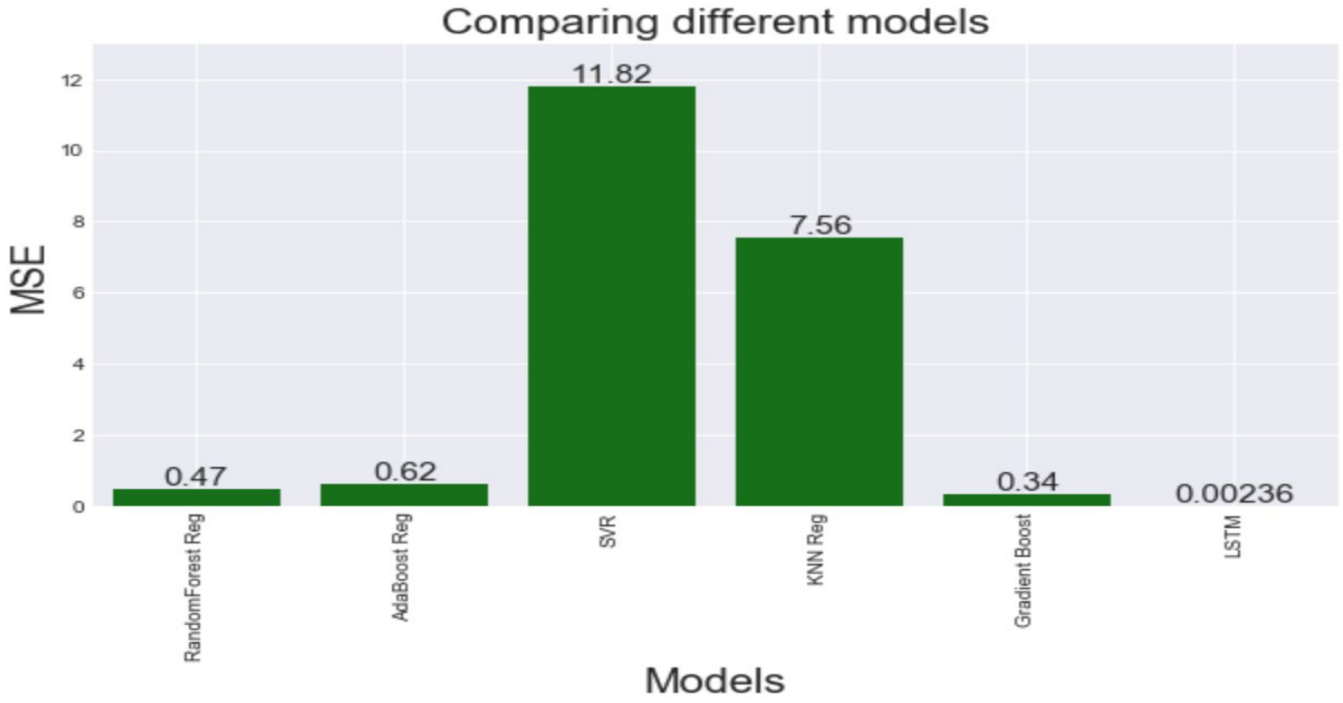


	Train	Test
MSE	0.00826	0.00236
RMSE	0.09	0.05

Results - LSTM



Comparing Results - American Airlines



Future Scope

- Add other unstructured data such as news articles about different airlines
- Find the correlation between the news articles and stock movement for different airline companies
- Try other combinations of deep learning models such as (CNN + LSTM).

Thank You!