

Trend Analysis of Stock Prices and Financial News

Instructor: Prof. Daniel Acuna

Team Members: Akshita Chandiramani, Ruchita Jadhav, Sanjeev Ramasamy Seenivasagamani

Problem & Objectives

- We aim to derive a quantifiable relationship between news and stock prices of a company and predict the fluctuation in the stock prices of a firm based on analysis of financial news of that firm.
- This will help investors and firms with understanding the impact of a particular news on stock prices.
- Through the textual analysis we try to convert qualitative information in the news & corporate announcements into a quantifiable measure by analyzing the positive and negative tones of information.

Data Description

The dataset consists of 11 features and has about 30000 records. It consists of data ranging from 17th April, 2017 to 17th April, 2019.

For this project we have collected data of three companies Apple, IBM and Microsoft. We have used developer API from intrinio for collecting the data. The API fetches the data through URLs for the specified companies. We input the key from API in our code to fetch the data and using data parsing techniques we were able to fetch the data. Then we combined the data for all the three companies to form one final dataset.

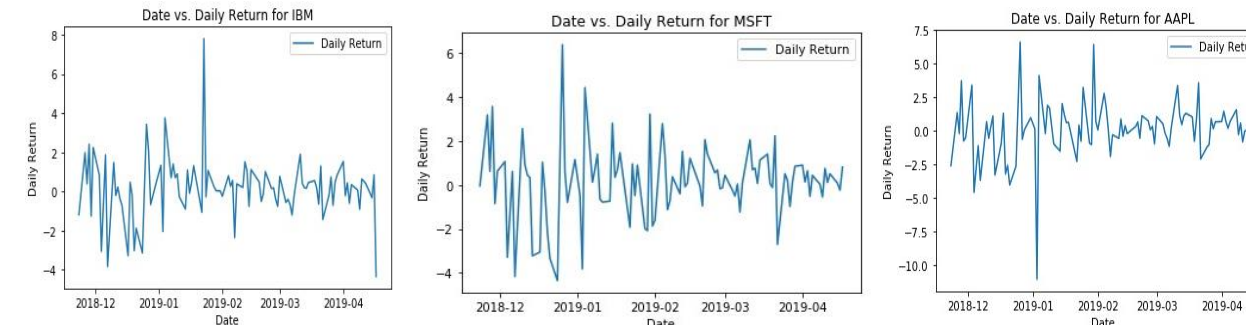
Data Collection and Pre-processing



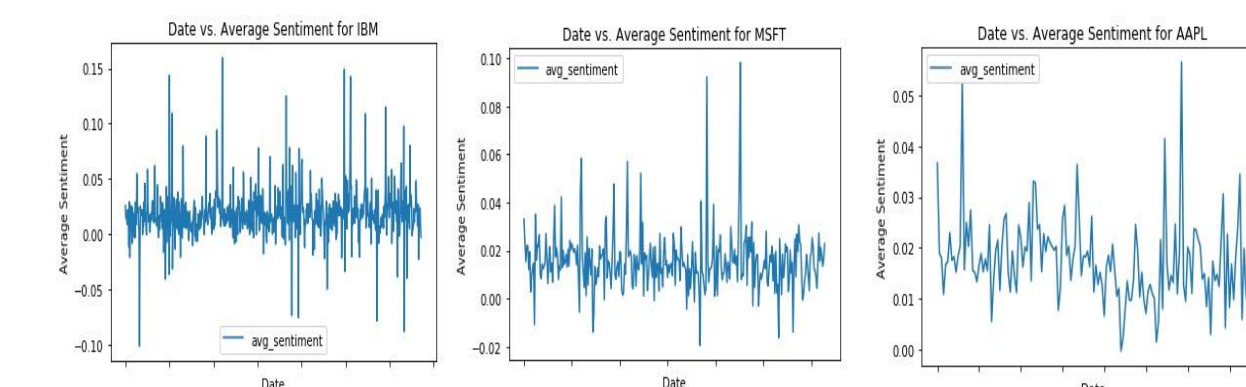
Date Vs Closing Prices across companies



Date Vs Daily Returns across companies



Sentiment Across the companies



Model Description

Model	Features	Techniques
Logistic Model using TF IDF	title tfidf, summary tfidf, stock price, date, company name	Spark pipelines, stock price difference, Regularization, Cross Validation, TFIDF conversion
Logistic Model using average Sentiment	Title sentiment, summary sentiment, stock price, date, company name	Spark pipelines, stock price difference, Regularization, Cross Validation, sentiment score
Linear regression using TFIDF	title tfidf, summary tfidf, stock price, date, company name	Spark pipelines, stock price difference, Regularization, Cross Validation, TFIDF conversion
Linear model using Average Sentiment	Title sentiment, summary sentiment, stock price, date, company name	Spark pipelines, stock price difference, Regularization, Cross Validation, sentiment score

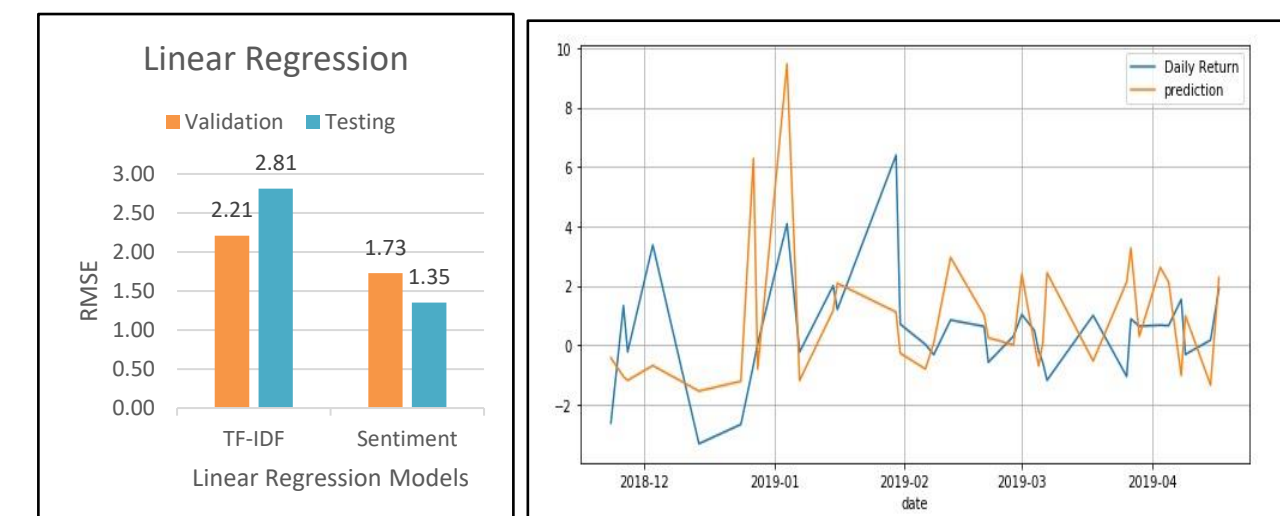
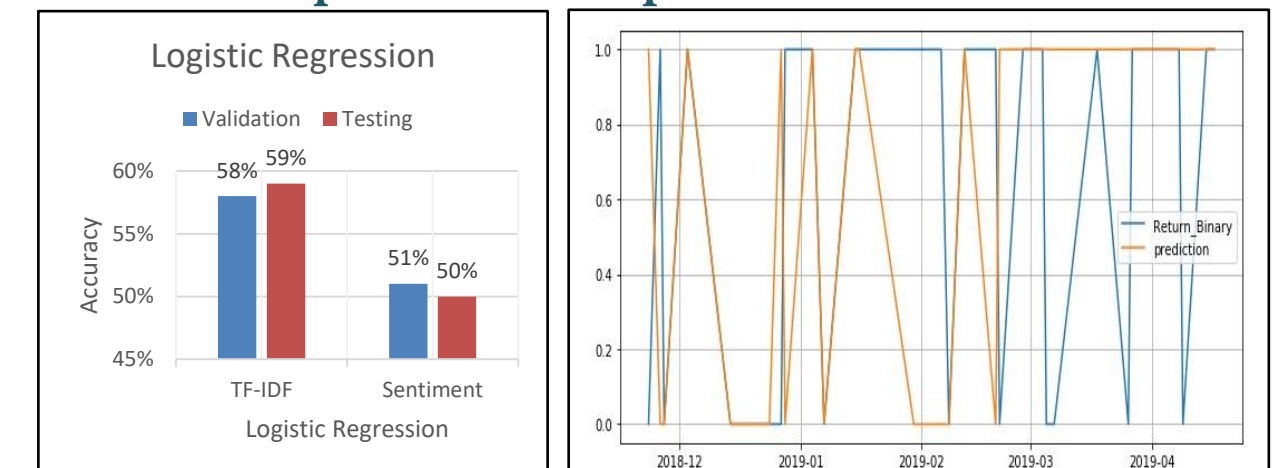
Model Comparisons measures:

Train-Validation-Test split: 60-30-10

Generalized Performance Measure:

- Logistic Regression: Accuracy
- Linear Regression: RMSE

Results –prediction performance



Results -inference

The final model chosen is logistic regression with highest accuracy of 59% and for linear regression is with TF-IDF with rmse of 2.81 since r-square value for TF-IDF is higher.

Conclusions:

Using our model the vendors/customer of application are able to predict the next day stocks of the company. We can predict the fluctuation in the stocks of company based on the title and summary of the articles. Further the model could be improvised by providing more data i.e. data collected over longer period of time and data spread over more companies. This way model would be able to predict more accurately as it learns more so that the generalization over the larger scale is easier.

Reference:

<https://www.nber.org/papers/w18725.pdf>

<https://silo.ai/relationship-between-news-and-company-stock-prices-using-ai/>

<https://www.nature.com/articles/srep03578>