Novel Approaches to Sentiment Analysis for Stock Prediction

Team 1

Agrim Rawat *2020102037*

Jewel Benny *2020102057*

KNV Karthikeya *2020102003*

Mudit Gaur *2020112008*

***Abstract*—Stock market forecasts are a suitable fit for a machine learning system due to their quantitative nature. A supervised learning model can combine technical information and qualitative sentiment from news to predict the course of stock movement. Information is encoded into real vectors of fixed length. We test numerous models to encode qualitative sentiment information into characteristics and to generate a final forecast about the direction of a particular stock given encoded news and technical aspects, which may be either up or down.**

I. Introduction

In the world of finance, stock market forecasting has long been a contentious and important topic. According to some theorists, the stock market is fundamentally unexpected because stock prices, which they contend reflect all available information, support the efficient-market hypothesis. Others have tried using fundamental research, technical analysis, and, more recently, machine learning, to forecast the market.

II. Basic Concepts

A. Support Vector Machine

Support vector machine is a supervised learning algorithm that is used to draw a decision boundary between n-dimensional data to generate classes. The algorithm takes the farthest points of the data to draw the decision boundary, which is a hyperplane. These data points are known as support vectors, hence the name.

B. Bi-directional Long Short Term Memory

Neural networks form a major portion of the deep learning algorithms and are able to mimic the structure of the human brain. It comprises of building blocks known as perceptrons which are connected to other perceptrons with associated weights and thresholds within the network. It comprises of an input layer, output layer, and hidden layers in between them. BLSTM is a type of a recurrent neural network that contains feed-back and feed-forward loops.

C. Random Forests

It is an ensemble learning method that builds multiple decision trees and outputs the class that recieves the most votes or the average of values returned by the regression operations. They are an improvement over a single decision tree to reduce the tendency of decision trees to overfit the training data.

III. Methodology

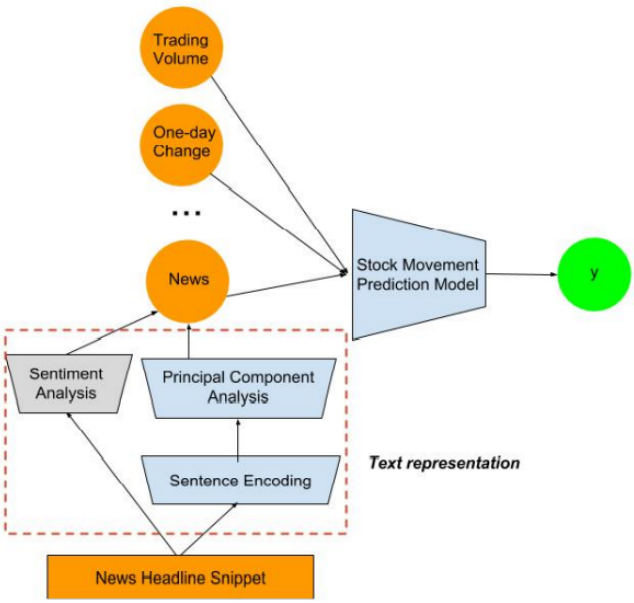

Fig. 1. Model for stock prediction

A. Data Collection

We use stock data collected from Yahoo Finance. This includes opening, closing, highest lowest price of stock and the volume of trading done for the stock over the period of 1 year. The timeframe used was 1^{st} November 2021 to 1^{st} November 2022.

B. Sentimental Analysis

As the headlines would effect the mind of a trader, this may lead to them to buy or dump the stock which would drive the price of stock up or down. Since, it is well known that many buyers are easily influenced by the actions of others and in this case the news headlines play a major role on their perception of the stock of the company.

Keeping in mind the above, to improve our predictions we are including sentimental analysis. We are using the data

headlines and snippets of news articles related to the company which is scraped from google news and finding their respective sentiment scores.

C. *Prediction Models*

We have predicted the trend of the stocks with and without sentiment analysis for three well-known companies, Meta, Apple and Amazon.

1) Without Sentiment Analysis: SVM, Random Forest, Neural Network

2) With Sentiment Analysis: BLSTM

IV. Results

We compared our outputs for the META, AAPL and AMZN tickers.

A. *Without Sentiment Analysis*

1) SVM:

- Ticker Used is META (Meta).
- Accuracy = 62.7450980392156684%
- Mean Absolute Error = 0.7450980392156863

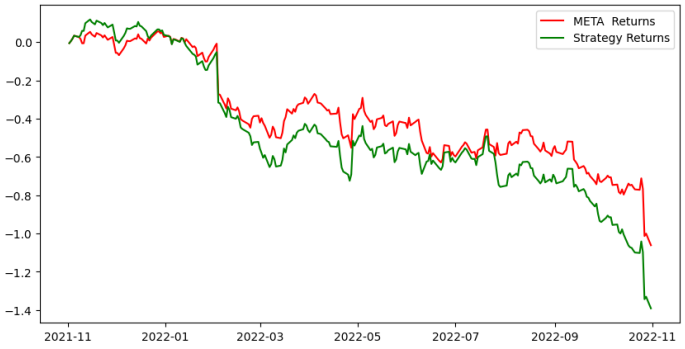

Fig. 2. Actual and Predicted returns of META using SVM

2) Random Forest:

- Ticker Used is AAPL (Apple).
- Accuracy = 99.83257035099335%
- Mean Absolute Error = 0.9844489768488489

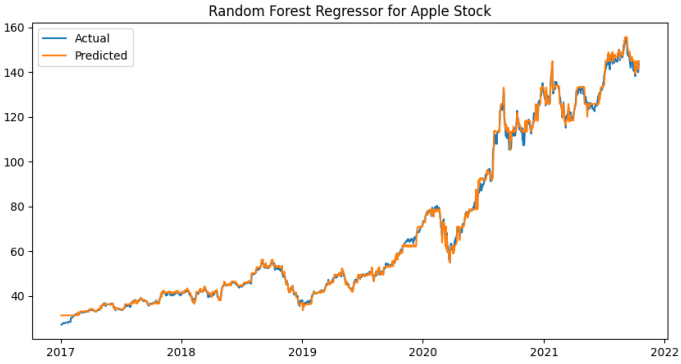

Fig. 3. Actual and Predicted returns of AAPL using Random Forest

3) Neural Network:

- Ticker Used is AMZN (Amazon).
- Accuracy (with tolerance 10) = 92.97658862876254%
- Mean Absolute Error = 5.015725789660196

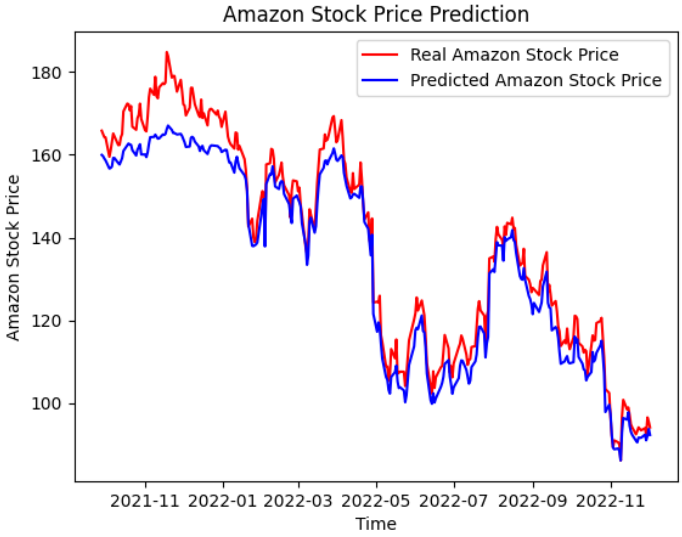

Fig. 4. Actual and Predicted returns of AMZN using Neural Network

B. *With Sentiment Analysis*

BLSTM:

- Mean Absolute Error = 7.81267497850501

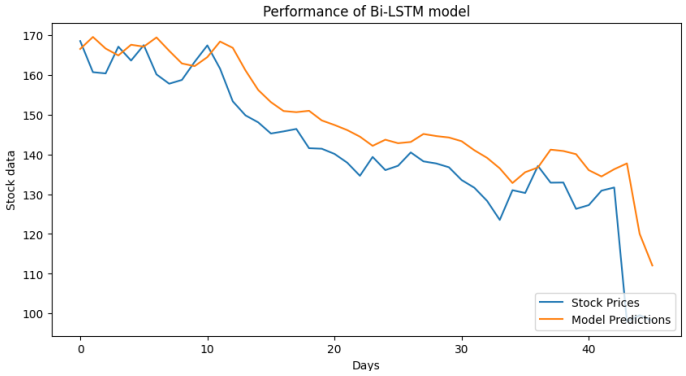

Fig. 5. Actual and Predicted returns of META using BLSTM

V. Conclusion

Model	Accuracy	Mean Absolute Error
Without Sentiment Analysis		
SVM	62.75%	0.7451
Random Forest	99.83%	0.9844
Neural Network	92.98%	5.0157
With Sentiment Analysis		
BLSTM	–	7.8127

The BLSTM model with sentiment analysis gives better prediction quality than the rest since it not only predicts on

the basis of the raw stock prices but also on the basis of the news sentiment which directly effects the buyers. The news headlines play a major role on their perception of the stock of the company.

ACKNOWLEDGMENT

We would like to thank our Professor Mr. Anoop Namboodri for giving us the opportunity to do this project. This project was a profound learning experience for all of us and we would like to thank him for allowing us to choose our topics too. We would also like to thank our TA Sanjai Kumaran for helping and guiding us whenever necessary and providing us with valuable feedback.

REFERENCES

[1] Chris Wang, Yilun Xu, Qingyang Wang, "A Novel Approach for Stock Market Prediction Using LSTM and Sentiment Analysis", 2018.