# Stock Market Analysis and Prediction

## Project Advisor: Prof. Weider Yu

Computer Engineering Department

Gupta, Ravi (MS Software Engineering)
Keerty, Aishwarya (MS Software Engineering)
Naik, Rushin (MS Software Engineering)
Pandya, Bhavan (MS Software Engineering)

## Introduction

Stock market prediction is the calculation of the future value of a company's stock. The stock market forecast has multiple dimensions to it which makes it a very complex problem. In our project, we gathered and refined a large dataset pertaining to the stock market and developed a recurrent neural network model to forecast the stock values and determine the difference between the predicted and actual value. The project consists of four components. Namely, the analysis component, the prediction component, the data layer and the user interface component. We created a single page responsive web application and additionally, we used Statistical arbitrage to find similar stocks based on their price movements.

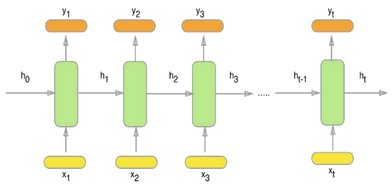

The proposed system uses an intelligent method of Machine Learning called Artificial Neural Networks (ANN). Recurrent Neural Network is a form of ANN, with dynamic temporal behavior, where the unit connections form a one-way cycle. This allows each neuron to use the input information of the previous unit.

## Methodology

### Stock Price Prediction using Recurrent Neural Net

The application uses Recurrent Neural Network (RNN) as a machine learning backbone for the prediction model. RNN has the capability to use the internal states of memory to process sequences of input. The RNN is made up of Long Short Term Memory (LSTM) units. This helps the algorithm to remember the pattern and sequences for a long period of time. Thus the algorithm was selected as it satisfies the needs for time series prediction.

The data is fed in various formats such that it helps the algorithm in analyzing various possible trends related to stock price data.




## Methodology

Three variations of the LSTM algorithm were considered and a similar grid search was performed for all of them using parameters like lookback, epochs and training size. The variations of the algorithm are as follows:

- LSTM using Lookback Window
- LSTM using Time Steps
- Stacked LSTM with Memory

The variation would include the change in number of features to the target and also the difference in how well the algorithm should remember the previous instance of the data from the training set.

The application compares the predicted value to the actual value for all the variations of the algorithm. The best hyper parameters and algorithm were selected as a part of the final prediction model.

The application allows the user to select the stocks whose value needs to be predicted. The best combination of algorithm and hyper parameters then runs on the background to determine and predict the value of the stock as desired by the user.

### Stock Price Data Analysis

The historical stock data is fetched from the Cassandra storage and then displayed to the user in the form of interactive graphs which helps the user significantly in understanding the variation in the prices of the stock of a particular company through a period of 5 years, 1 year, 1 month or 1 week.

The user has the ability to also perform the technical analysis and then use that along with the prediction model provided by the application itself to invest in the particular stock smartly.

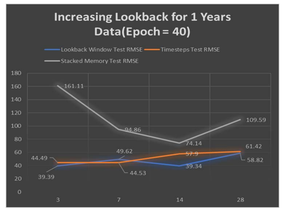
The stock data gives deep insight into the performance of the company and helps the users gauge the prospects of growth of the company to a certain extent.

The figure below gives an idea about how the applications demonstrates the stock price analysis for a given company over a period of time chosen by the user.

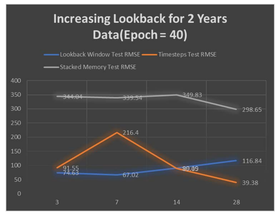

The charts can be used by users to mark support or resistance levels in the form of previous lows below the current price and previous highs above the current price respectively. The users could also couple this method with fundamental analysis to gain great results.

## Analysis and Results

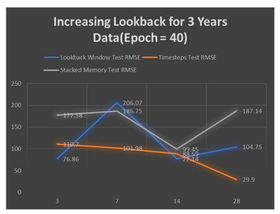
A set of algorithms were tested based changing the lookback parameter and keeping a steady number of epoch which was found to be performing best. The following figure shows a comparison of test RMSE for 3 variations of LSTM while training the data over 1 year data:




Further, increasing the size of the train set and adding one more year in consideration, the results that were obtained we varied for differently for each algorithm. The following figure shows a comparison of test RMSE for 3 variations of LSTM while training the data over 2 year data:




As a final steps towards selecting the algorithm, models were trained on 3 years stock data. The following figure shows a comparison of test RMSE for 3 variations of LSTM while training the data over 3 year data:




A further increase in the training size lead to a drastic downfall in the accuracy of the most of algorithms and results were not accurate enough to be considered. Thus, the approach with "Time Steps LSTM" proved to be the best with the lowest error rate. The next step was to increase the accuracy of the best algorithm and that task was achieved by aggressive grid search technique.

| Train Size | Lookback | Epochs | Test RMSE |
|---|---|---|---|
| 4 years | 60 | 50 | 50.84 |
| 4 years | 60 | 40 | 45.34 |
| 4 years | 28 | 40 | 39.58 |
| 4 years | 28 | 50 | 27.47 |
| 4 years | 60 | 50 | 24.19 |

The above figure shows the hyper parameters and relative test errors. The LSTM with Timesteps was run on series of combination of hyper parameters. Eventually, an increase in the epochs and loopback led to good results. Though, the train size was restricted to 4 years as the model started overfitting and had a negative impact on the accuracy.

## Summary/Conclusions

Most applications offer positive and negative classifiers for stock prices. This application uses web crawling, historical stock data analytics and a good machine learning model to predict the prices and display the analytics. The accuracy of the stock market prediction and analytics model is the prime metric that we have set as a team to determine the success of the application. The model needs to be trained better so that the benchmark in terms of throughput can be achieved.It can be concluded that the application offers foundation for technical analysis and prediction of stock prices.

## Key References

[1] Y. Yetis, H. Kaplan and M. Jamshidi, "Stock market prediction by using artificial neural network," 2014 World Automation Congress (WAC), Waikoloa, HI, 2014, pp. 718-722.

[2] Mahdi Pakdaman Naeini, H. Taremian and Homa Baradaran Hashemi, "Stock market value prediction using neural networks," 2010 International Conference on Computer Information Systems and Industrial Management Applications (CISIM), Krackow, 2010, pp. 132-136.

[3] M. Tirea and V. Negru, "Intelligent Stock Market Analysis System - A Fundamental and Macro-economical Analysis Approach," 2014 16th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, Timisoara, 2014, pp. 519-526.

[4] Shigang Liu, Min Gan and Honghua Dai, "A hybrid approach to fuzzy risk analysis in stock market," 2013 10th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Shenyang, 2013, pp. 295-298.

## Acknowledgements

The authors are deeply indebted to Professor Weider Yu and Professor Dan Harkey for their invaluable comments and assistance in the preparation of this study.

We really appreciate their insights and expertise which helped us to transform this idea to reality.