# **Comparing Machine Learning Algorithms for stock prediction and stock index movement using trend deterministic data preparation techniques.**

**PROJECT SYNOPSIS**

**BACHELOR OF TECHNOLOGY**

**Computer Science and Engineering**

**SUBMITTED BY**

Group No. 4

Mohak Kulshrestha (1506410053)

Kashvi Agarwal (1506410040)

Samar Srivastava (1506410078)

Saurabh Arya (1506410079)



**Department of Computer Science and Engineering**

**Hindustan College of Science and Technology, Mathura**

**(Dr. APJ Abdul Kalam Technical University, Lucknow)**

* **PROBLEM STATEMENT:**

Improve efficiency in predicting direction of stock movements and their trends by making the prediction model “remember” the historical movement of the stocks by using algorithms like LSTM and XGBoost and comparing their results with the typically used algorithms in such scenarios like Naïve Bayes and SVM.

* **WHY IS THE TOPIC CHOSEN?**

General prediction models that are used in the previously described research papers like Naive Bayes, Support Vector Machines, Random Forest, etc., can only take in consideration the most recent previous input received to the model. In our initial research we found out that efficiency in predicting stock price movements and their trends maybe improved if the prediction model could “remember” the historical movement of the stocks.

* **OBJECTIVE AND SCOPE:**

Investment firms, hedge funds and even individuals have been using financial models to better understand market behavior and make profitable investments and trades. A wealth of information is available in the form of historical stock prices and company performance data, suitable for machine learning algorithms to process.

Can we actually predict stock movement with machine learning? Investors make educated guesses by analyzing data. They'll read the news, study the company history, industry trends and other lots of data points that go into making a prediction. The prevailing theories are that stock prices are totally random and unpredictable but that raises the question why top firms like Morgan Stanley and Citigroup hire quantitative analysts to build predictive models.

Our main objectiveto utilize Deep Learning models, Long-Short Term Memory (LSTM) Neural Network algorithm and hybrid Machine Learning Algorithm XGBoost, to predict direction of stock movement and their trends and by doing so trying to outperform the old classical algorithms that are generally used for this task.

* **Methodology:**

In our initial research we found out that efficiency in predicting stock price movements and their trends maybe improved if the prediction model could “remember” the historical movement of the stocks.

General prediction models that are used in the previously described research papers like Naive Bayes, Support Vector Machines, Random Forest, etc., can only take in consideration the most recent previous input received to the model.

Our motive is to implement the XGBoost and Long Short Term Memory networks in the prediction on stock price movements and their trends, as XGBoost uses a more regularized model formalization to control over-fitting, which gives it better performance and LSTM networks have internal contextual state cells that act as long-term or short-term memory cells.

The output of the LSTM network is modulated by the state of these cells. This is a very important property when we need the prediction of the neural network to depend on the historical context of inputs, rather than only on the very last input.

* **HARDWARE AND SOFTWARE REQUIREMENT:**
  1. Hardware Requirements(Minimum):
     1. Nvidia GeForce 920M GPU
     2. i7-6850K CPU with 40 PCIe Lanes
     3. 4 Gigabytes of DDR4-1600 RAM with 3.5GHz clock rate
  2. Software Requirements:
     1. Python 3.6
     2. Jupyter Notebook
     3. Numpy, Pandas, Matplotlib, scikit-learn, tensorflow, keras
* **What contribution would the project make?**

The project would help in predicting stock price movements and trends in a better way than all other prediction models because of the ability of LSTM networks to remember the contextual states of the cells historically.

* **Data set description:**

All the dataset will be used from quandl.com. Quandl is a platform for financial, economic, and alternative data that serves investment professionals. Quandl sources data from over 500 publishers.

We will be using stock price dataset of OHLC format of the following to train and test our prediction models:

* **STOCKS:**

1. Google (GOOGL)
2. Amazon (AMZN)
3. Apple (AAPL)

* **STOCK INDICES:**

1. S&P 500 (GSPC)
2. NASDAQ-100 (NDX)
3. DOW JONES (DJI)

* **Present Status of the Project:**

At the present state, we have successfully implemented 2 out of the 6 algorithms we have selected for the implementation of this project. These algorithms are **Random Forest algorithm** and **Naïve Bayes algorithm**. These algorithms are being used to set as a benchmark against which our new models of LSTM and XGBoost will be tested. Currently we have implemented Random Forest Classifier model using continuous data method of input and achieved an accuracy of 85.80 which is an improvement over the average accuracy of 83.59 as shown in our base research paper. We have also implemented Naïve Bayes Classifier which currently stands on an accuracy of 71.59.

* Work break down Structure (considering Project Completed by 31st March -2019)
* **Zest of Core Research papers at least 04-05 Research Papers.**

1. **Stock Trend Prediction with Technical Indicators using SVM by Xinjie Di:**

In 2014, Xinjie Di, an SCPD student from Apple Inc. submitted a paper which focused on predicting stock price trend for a company in the near future. The feature space was derived from the time series of the stock itself and was concerned with potential movement of past price. Tree algorithm was applied to feature selection and it suggests a subset of stock technical indicators are critical for predicting the stock trend. Experiment results suggested an accuracy of more than 70% on predicting 3-10 day average price trend with SVM algorithm.

1. Patel , J., et al. Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. Expert Systems with Applications (2014)

Another paper presented in Expert Systems with Applications journal under Elsevier publishing company by Jigar Patel, Sahil Shah, Priyank Thakkar, K. Kotecha was addressed the problem of predicting direction of movement of stock and stock price index for Indian Stock Markets. The paper compares four prediction models, Aritificial Neural Network(ANN), Support Vector Machine(SVM), Random Forest and Naive Bayes with two approaches for input to these models.

The first approach for input data involves computation of ten technical parameters using stock trading data(open, high, low & close prices) while the second approach focuses on representing these technical parameters as trend deterministic data. Accuracy of each of the prediction models for each of the two input approaches was evaluated. Evaluation was carried out on 10 years of historical data from 2003 to 2012 of two stocks namely Reliance Industries and Infosys Ltd.

Experiment results suggested that for the first approach of input data Random Forest outperforms other three prediction models on overall performance. Experimental results also show that the performance of all the prediction models improve when these technical parameters are represented as trend deterministic data.

* **Limitations:**

Limitations of this project lies in the fact that only technical aspects and technical indicators for predicting stock movement direction are used.

There are various other factors that can be taken into consideration while making these predictions. Some of these factors include changes in government policies, social outburst, political factors and many more. Although all these factors affect the stock prediction but incorporating all these factors in a single prediction model does not guarantee an accurate prediction.

* **References:**

**[1]** Patel , J., et al. Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. Expert Systems with Applications (2014).

**[2]** Xinjie Di, Stock Trend Prediction with Technical Indicators using SVM (2014).