# **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)**

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

* **PROBLEM STATEMENT:**

Improve efficiency in predicting stock price movements and their trends by making the prediction model “remember” the historical movement of the stocks.

* **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:**

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.

Another paper presented in Expert Systems with Applications journal under Elsevier publishing company by Jigar Patel, Sahil Shah, Priyank Thakkar, K. Kotecha quoted as 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) 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.

* **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 Long Short Term Memory networks in the prediction on stock price movements and their trends, as 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:**
  + 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
  + Software Requirements:
    1. Python 3.6
    2. Jupyter Notebook
    3. Numpy, Pandas, Matplotlib, scikit-learn, tensorflow, keras
* 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. All Quandl's data are accessible via an API. API access is possible through packages for multiple programming languages including R, Python, Matlab.

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

* + Google (GOOGL)
  + Amazon (AMZN)
  + Apple (AAPL)
* **Present Status of the Project:**

At the present state,

* Work break down Structure ( considering Project Completed by 31st March -2019)
* Zest of Core Research papers at least 04-05 Research Papers. ( Write the summary paper wise in the synopsis)
* Limitations
* Reference

**Submit the Project Synopsis up to 26-11-2018 positively**