EVEREST ENGINEERING COLLEGE

(AFFILIATED TO POKHARA UNIVERSITY)

SANEPA, LALITPUR

[Subject Code: CMP 490]

A

MAJOR PROJECT REPORT ON

"PREDICTION OF A COMPANY'S STOCK PRICE USING MACHINE LEARNING"

SUBMITTED BY

NABIN KUMAR BAMMA	**[18120047]**
RABIN KUMAR MANDEL	**[18120061]**
SANTOSH CHAPAGAIN	**[18120076]**
SUSIL KUMAR SHRESTHA	**[18120091]**

SUBMITTED TO

DEPARTMENT OF INFORMATION TECHNOLOGY ENGINEERING

EVEREST ENGINEERING COLLEGE

SANEPA, LALITPUR

SEPTEMBER, 2022

COPYRIGHT

CERTIFICATE

The undersigned certify that they have read and recommended to the Department of Information Technology Engineering for acceptance, a project report entitled "**PREDICTION OF A COMPANY'S STOCK PRICE USING MACHINE LEARNING**", submitted by **Nabin Kumar Bamma, Rabin Kumar Mandel, Santosh Chapagain and Susil Kumar Shrestha** in partial fulfillment of the requirement for the Bachelor's degree in Information Technology Engineering.

Supervisor:

....

Er. Subhadra Joshi

Lecturer

External Examiner:

....

Mr. Deepesh P. Guragain

Assistant Professor

Nepal Engineering

College

ACCEPTANCE

The project report entitled "**PREDICTION OF A COMPANY'S STOCK PRICE USING MACHINE LEARNING**", submitted by **Nabin Kumar Bamma, Rabin Kumar Mandel, Santosh Chapagain and Susil Kumar Shrestha** in partial fulfillment of the requirement for the Bachelor's degree in Information Technology Engineering has been accepted as a bonafide record of work independently carried out by the group in the department.

……………………………..

Er. Birodh Rijal

Principal

Everest Engineering College,

Sanepa, Lalitpur

……………………………..

Er. Anuj Ghimire

Head of the Department

Department of Information Technology Engineering,

Everest Engineering College,

Sanepa, Lalitpur

ACKNOWLEDGEMENT

We have taken effort in this project. However, it would not have been possible without the help and support of many individual and organization. We would like to extend our sincere thanks to all of them.

We would like to express our deepest sense of gratitude and sincere thanks to our highly respected and esteemed supervisor **Er. Subhadra Joshi** for her valuable guidance, encouragement and help for supporting the work. Her useful suggestion for this work and co-operative behaviour are sincerely acknowledged.

We are highly indebted to EEC and Pokhara University for that project & also creating friendly environment to the success of our project. We are also appreciative of the effort of HOD **Er. Anuj Ghimire** for supporting us throughout this project. We would also like to appreciate the help of various venue for providing us essential data required for our project as well as their time in guiding us so as to how our project will be more effective in users prospective.

We are also grateful to our teachers for their constant support and guidance. At the end we would like to express our sincere thanks to all the friends and others who helped us directly or indirectly during this project work.

ABSTRACT

Prediction of the stock market trend is challenging task as there is lots of uncertainty in the stock market. In this report we have mentioned various models for predicting the price of the stock. We have used three models LSTM, Support Vector Machine and random forest. The accuracy of these model is evaluated by taking AAPL data. For a good and successful investment, many investors are keen in knowing the future situation of the stock market. Good and effective prediction systems for stock market help traders, investors, and analyst by providing supportive information like the future direction of the stock market. Changes in stock prices are probably triggered by various factors, both internal and external, coming the company. Internal factors used in this study are price earnings ratio, earning per share, while external factors used are inflation rate and interest rates. The purpose of this study is to identify factors affecting stock pricing in the any stock of company, and determine which factors are most influential on stock prices company. The data used in this study were collected during the period from 2000 to 2021 of AAPL Company.

Keywords: LSTM, SVM

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATION

AI	Artificial Intelligence
ARIMA	Autoregressive Integrated Moving Average
EPS	Earnings per Share
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
PE	Price to earning
RFR	Random Forest Regressor
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
SVM	Support Vector Machine
SVR	Support Vector Regression

CHAPTER 1: INTRODUCTION

1.1 Background

A stock market is a public market where you can buy and sell shares for publicly listed companies. The stocks, also known as equities, represent ownership in the company. The stock exchange is the mediator that allows the buying and selling of shares.

Stock Price Prediction using machine learning helps you discover the future value of company stock and other financial assets traded on an exchange. The entire idea of predicting stock prices is to gain significant profits. Predicting how the stock market will perform is a hard task to do. It is a financial product characterized by high risk, high return and flexible trading, which is favoured by many investors. Investors can get abundant returns by accurately estimating stock price trends. However, the stock price is influenced by many factors they are internal and external factors involved in the prediction, such as EPS and P/E ratio are the internal and interest rate and inflation rate are the external factors and so on. All these factors combine to make share prices dynamic and volatile. This makes it very difficult to predict stock prices with high accuracy [1].

Prediction of the stock price has always been the focus and difficult research topic. Statistical and econometric models are generally used in traditional stock price prediction, but these methods cannot deal with the dynamic and complex environment of the stock market. Since 1970, with the rapid development of computer technology, researchers have begun using machine learning to predict stock prices and fluctuations, helping investors determine investment strategies to reduce risk and increase returns.

Therefore, stock price prediction is one of the most challenging problems in all kinds of prediction research. In the past decades, scholars have studied stock price prediction from many perspectives, where the improvement of prediction models and the selection of model features are the two most important directions among them. Most of the early studies used econometric models, such as autoregressive integrated moving average (ARIMA) and autoregressive conditional heteroskedastic-autoregressive integrated moving average (ARCH-ARIMA), to predict stock price. However, it is difficult for econometric models to consider the impact of other factors on stock price fluctuations and they have strong assumptions about the data, which are often difficult to meet. Therefore, machine learning has been widely used in stock price prediction in recent years and many more suitable models for stock prediction have been proposed. Many studies have shown that deep learning has superior efficiency than other models and neural network models excel regression and discriminant models. In terms of feature selection, some scholars explore the correlation between new features and stock price and some new features, including political factors, inflation rate, interest etc. Have been incorporated into the prediction model.

Previous literature extensively investigates the stock price prediction methods and many advanced prediction models are proposed. However, existing approaches on stock price prediction have two main limitations. First, although the text features are used in the existing models to better incorporate the important information in social media, they are usually mined based on traditional text mining technologies, such as the bag-of-word model. These text mining technologies cannot consider the semantic and other information in social media which are helpful to improve the performance of prediction models. Second, the feature dimensionality reduction is a basic step when balancing text features and financial features in stock price prediction. To fill the research gap discussed above, they propose a new stock price prediction method based on deep learning technology, long short-term memory (LSTM), Support Vector Machine and Random Forest Regression model. Feature extraction of text information in social media can describe the emotional tendency of investors and help to predict the stock price more accurately [2].

1.2 Motivation:

Stock market prediction aims to determine the future movement of the stock value of a financial exchange. The accurate prediction of share price movement will lead to more profit investors can make. Not all models provide the highest possible amount of the accuracy for all data. Hence the main drive to do this project is to evaluate the possible best model to predict the stock price for given data.

1.3 Aim and objectives

- To analyze, forecast and compare the three different models (SVR, LSTM, and Random Forest) based on the metrics MAE and RMSE.

1.4 Applications:

- Can be applicable to discover the future value of company stock traded on the exchange.

1.5 Project Features

- Predict stock price using daily data.
- Gives graph chart of the predicted and the original data.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

The literature review deals with the topics and the researches that would help to understand Stock Price Prediction System from the existing systems that are similar to this system. The objective of this literature review is to analyze the related work to this project and mechanisms used in previous studies.

2.2 Existing Methods

2.2.1 Stock Market Prediction Using Machine Learning

In the research work, the Yahoo finance stock data is predicted using machine learning algorithm. In this Stock to determine the future value of a stock. The technical and fundamental or the time series analysis is used by the most of the stock brokers while making the stock predictions. In this paper we propose a Machine Learning approach that will be trained from the available stocks data and gain intelligence and then uses the acquired knowledge for an accurate prediction. In this context this study uses a machine learning technique called Support Vector Machine to predict stock prices for the large and small capitalizations and in the three different markets, employing prices with both daily and up-to-the-minute frequencies [3].

2.2.2 Forecasting the Stock Market Index Using Artificial Intelligence Techniques

In this research paper [4] the forecast the future price of an asset based on the information contained in the historical prices of an asset. Furthermore, financial forecasting is a difficult task due to the intrinsic complexity of the financial system. The objective of this work was to use artificial intelligence (AI) techniques to model and predict the future price of a stock market index. Three artificial intelligence techniques, namely, neural networks, support vector machines are implemented in forecasting the future price of a stock market index based on its historical price information [5].

We used the three different algorithms namely LSTM, SVR and Random Forest algorithm for predicting stock price from past closing prices of the individual stock. The results showed that the three techniques have the ability to predict the future price of the stock with an acceptable accuracy.

2.2.3 Automated Stock Price Prediction Using Machine Learning

Next existing system that traditionally in order to predict market movement, investors used to analyses the stock prices and stock indicators in addition to the news related to these stocks. In this paper [6] we propose an automated trading system that integrates mathematical functions, machine learning, and other external factors, we aim to determine the price or the trend of a certain stock for the coming end-of-day considering the first several trading hours of the day. To achieve this goal, we trained traditional machine learning algorithms and created/trained multiple deep learning models taking into consideration the importance of the relevant news. Various experiments were conducted and predicting AAPL data.

2.2.4 Stock Price Correlation coefficient prediction with ARIMA – LSTM Hybrid Model

The research paper predicting the price correlation of two assets for future time periods is important in portfolio optimization. We apply LSTM in predicting the stock price correlation coefficient of individual stocks. The use of LSTM cells further enhances its long-term predictive properties. To encompass both linearity and nonlinearity in the model. Our work implies that it is worth considering the LSTM model to forecast the price of stock [7].

2.2.5 Technology industry on financial ratios and both factor

There are many studies related to the effect of internal factors and external factors on stock return. This study will examine both of factors [8] they examined about relationship between financial ratios and stock return in technology industry of the Stock. Five financial ratios from each category are used as sample of dependent variables; current ratio from liquidity ratio, debt to equity ratio from debt ratio, inventory turnover ratio from asset activity ratio, return on equity from profitability ratio, and price earnings ratio from market value ratio.

Thim et al [9] investigated the factors affecting the performance of many property firms listed method has been applied to represent all the variables comprising stock. Some are performance, return on equity, debt ratio, and earnings per share, and price earnings ratio, inflation, industrial production, exchange rate and interest rate.

We found from this existing previous research paper [10] explained about relationship and influence of earning per share and book value on stock prices. Ozlen & Ergun [11] investigated internal determinants of the stock price movement on sector basis. Financial ratio such as total assets turnover ratio, debt ratio, current ratio, net profit margin, price to earnings ratio and book value is the most important internal factor in explaining stock price movements for all sectors.

From all this research we could take the internal and external parameter of stock market. By interpolating monthly data could be converted into daily basis.

CHAPTER 3: METHODOLOGY

3.1 Introduction

A methodology is a development system of methods that is used to plan, structure, and control the process of developing an information system. A wide variety of published development methodologies have evolved over the years, each with its own recognized strength and weakness. Different types of system project use available methodologies that best suits a specific project based on the project's various technical development process [12].

3.2 Data Selection and Preprocessing

3.2.1 Stock Price Data

The Apple stock data are used in the experiment. The stock data contain daily stock price information such as open, high, low, close, Adj close and volume. In total, 5537 trading days are considered from 3 January 2000 to 31 December 2021.

Table 1: Sample Dataset

	A	B	C	D	E	F	G
1	Date	Open	High	Low	Close	Adj Close	Volume
2	1/3/2000	0.936384	1.004464	0.907924	0.999442	0.853355	5.36E+08
3	1/4/2000	0.966518	0.987723	0.90346	0.915179	0.781409	5.12E+08
4	1/5/2000	0.926339	0.987165	0.919643	0.928571	0.792844	7.78E+08
5	1/6/2000	0.947545	0.955357	0.848214	0.848214	0.724232	7.68E+08
6	1/7/2000	0.861607	0.901786	0.852679	0.888393	0.758538	4.61E+08
7	1/10/2000	0.910714	0.912946	0.845982	0.872768	0.745197	5.05E+08
8	1/11/2000	0.856585	0.887277	0.808036	0.828125	0.70708	4.42E+08
9	1/12/2000	0.848214	0.852679	0.772321	0.77846	0.664674	9.76E+08
10	1/13/2000	0.84361	0.881696	0.825893	0.863839	0.737573	1.03E+09
11	1/14/2000	0.892857	0.912946	0.887277	0.896763	0.765685	3.9E+08
12	1/18/2000	0.901786	0.946429	0.896763	0.928013	0.792367	4.59E+08
13	1/19/2000	0.94308	0.970982	0.922991	0.951451	0.812379	5.98E+08
14	1/20/2000	1.03125	1.084821	1.013393	1.013393	0.865268	1.83E+09
15	1/21/2000	1.020089	1.020089	0.983817	0.993862	0.848591	4.96E+08
16	1/24/2000	0.968192	1.006696	0.938616	0.948661	0.809997	4.41E+08
17	1/25/2000	0.9375	1.010045	0.914063	1.002232	0.855737	4.97E+08
18	1/26/2000	0.982143	1.019531	0.979911	0.983817	0.840014	3.67E+08
19	1/27/2000	0.97154	1.008929	0.955357	0.982143	0.838585	3.4E+08
20	1/28/2000	0.96596	0.989955	0.898438	0.907366	0.774738	4.23E+08

3.2.2 Variables

Independent variables that used in this study are Earning per Share (EPS) and Price Earnings Ratio (PER) as internal factors, meanwhile variable inflation, interest rate, are used as external variables to stock return as dependent variables.

The following are the eleven variables that are computed for designing the machine learning models.
Date: The stock value date.

Open: The price from the first transaction of a trading day is the opening price.

Close: The closing price is the last price of the trading day.

High: It is the highest price of the trading day and is typically higher than the closing or equal to the opening price.

Low: The low is the price that a stock trades for throughout a single day.

EPS: Earning Per Share is a company's net profit divided by the number of common shares. EPS indicates how much money a company makes for each share of its stock.

PE ratio: Price to Earnings Ratio is the ratio of share price of a stock to its earnings per share (EPS). In general, a high P/E suggests that investors are expecting higher earnings growth in the future compared to companies with a lower PE.

Inflation rate: It is the rate at which prices for goods and services rise. The market price of value stocks is usually directly proportional to the rate of inflation. Therefore, when the inflation rate rises, value stocks tend to perform better.

Adj Close: Amended closing price for dividends of stock value after distributing dividends.

Interest rate: When interest rates rise, stocks tend to fall in value because of lower future earnings. Higher inflation leads to higher interest rates, which do impact the stock market.

Volume: Total number of traded stocks in the market in trading day.

3.2.3 Resampling External Variables

Resampling primarily involves changing the time-frequency of the original observations. The two popular methods of resampling in time series are as follows:

- Upsampling
- Downsampling

Initially, our team collected external Parameters data from FRED Economic Data. These data in the monthly basis.

Table 2: Monthly Interest and Inflation Rate Dataset

	A	B	C	D
1	DATE	Interest_Rate	Inflation_rate	
2	1/1/2000	3.41105112	1.76	
3	2/1/2000	3.51334339	1.91	
4	3/1/2000	3.44034736	1.87	
5	4/1/2000	3.20296665	1.78	
6	5/1/2000	3.3605311	1.66	
7	6/1/2000	3.1278929	1.61	
8	7/1/2000	3.06235485	1.87	
9	8/1/2000	3.04561328	2.13	
10	9/1/2000	2.88599618	2.08	
11	10/1/2000	2.87662559	2.21	
12	11/1/2000	2.83313793	2.34	
13	12/1/2000	2.63921042	2.29	
14	1/1/2001	2.25143294	2.26	
15	2/1/2001	2.26701398	2.32	

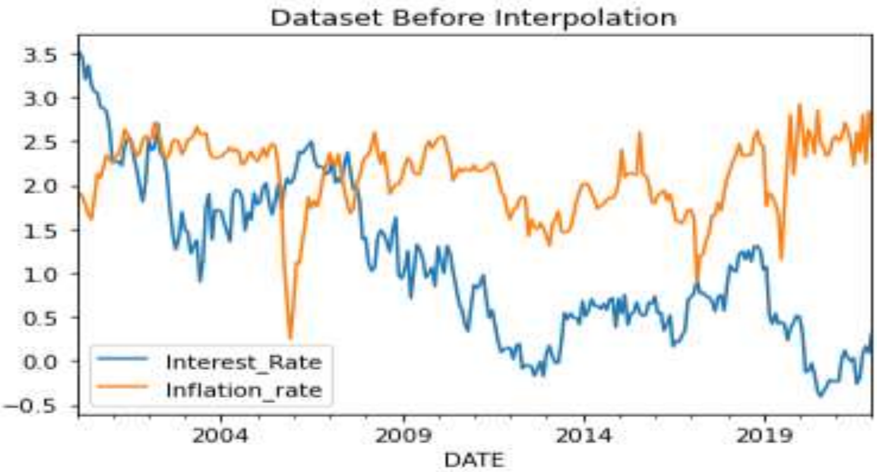

Figure 1: Plotting of the External Parameters before Interpolation

To convert the above monthly data into daily form we perform Upsampling of data. Upsampling involves increasing the time-frequency of the data, it is a data disaggregation procedure where we break down the time frequency from a higher level to a lower level. For example, breaking down the time-frequency from months to days, or days to hours or hours to seconds.

Table 3: Upsampled dataset from month to day based on mean value of the month

```
In [7]: upsampled
Out[7]:
```

DATE	Interest_Rate	Inflation_rate
2000-01-01	3.411051	1.76
2000-01-02	NaN	NaN
2000-01-03	NaN	NaN
2000-01-04	NaN	NaN
2000-01-05	NaN	NaN
...	...	...
2021-12-28	NaN	NaN
2021-12-29	NaN	NaN
2021-12-30	NaN	NaN
2021-12-31	NaN	NaN
2022-01-01	0.319495	2.53

Above figure shows a few samples of the dataset which is Upsampled from months to days, based on the mean value of the month.

The dataset has been up sampled with Nan values for the remaining days except for those days which were originally available in our dataset. Now, we can fill these Nan values using a technique called Interpolation. Pandas provide a function called DataFrame. Interpolate () for this purpose. Interpolation is a method that involves filling the nan values using one of the techniques like nearest', 'zero', 'linear', 'quadratic', 'cubic', 'spline', 'barycentric', 'polynomial'. We will choose "Spline" interpolation.

Table 4:External parameter after spline interpolation

Out[9]:

DATE	Interest_Rate	Inflation_rate
2000-01-01	3.411051	1.760000
2000-01-02	2.796146	2.488276
2000-01-03	2.795518	2.488011
2000-01-04	2.794891	2.487746
2000-01-05	2.794264	2.487482

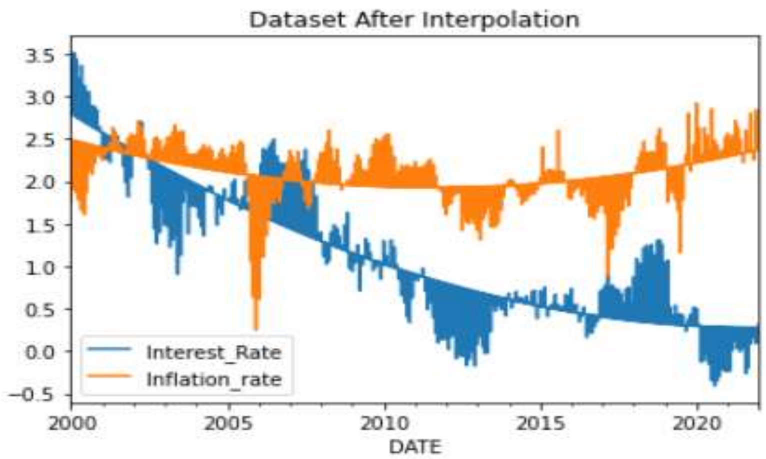

Figure 2: Interpolated Data

3.2.4 Data Merge

After interpolating the external parameter, we then merge these data with our stock datasets for further modeling. We merge the external parameter data and stock dataset with an index on ‘Date’ ranging from 01/03/2000 to 12/31/2021 with a total of 5537 rows and 11 columns.

Table 5: Final Dataset Sample

	Date	Open	High	Low	eps	pe	InRate	InfRate	Close	Adj Close	Volume
0	1/3/2000	0.936384	1.004464	0.907924	0.310000	20.770000	3.411051	1.760000	0.999442	0.853355	535796800
1	1/4/2000	0.966518	0.987723	0.903460	0.188723	21.050533	2.796146	2.488276	0.915179	0.781409	512377600
2	1/5/2000	0.926339	0.987165	0.919643	0.189153	21.045239	2.795518	2.488011	0.928571	0.792844	778321600
3	1/6/2000	0.947545	0.955357	0.848214	0.189583	21.039946	2.794891	2.487746	0.848214	0.724232	767972800
4	1/7/2000	0.861607	0.901786	0.852679	0.190012	21.034654	2.794264	2.487482	0.888393	0.758538	460734400
...	...	...	...	...	...	...	...	...	...	...	...
95	5/18/2000	0.919643	0.936942	0.898438	0.228834	20.558912	2.737512	2.463660	0.899554	0.768068	373777600
96	5/19/2000	0.886161	0.886161	0.833705	0.229258	20.553748	2.736892	2.463401	0.839286	0.716609	740667200
97	5/22/2000	0.837054	0.837054	0.767857	0.229681	20.548585	2.736272	2.463143	0.803013	0.685638	755507200
98	5/23/2000	0.808036	0.833705	0.764509	0.230105	20.543424	2.735653	2.462884	0.766183	0.654191	517585600
99	5/24/2000	0.769531	0.801339	0.741071	0.230528	20.538264	2.735033	2.462625	0.782924	0.668485	678462400

100 rows × 11 columns

3.2.5 Feature Score of the Dataset Variable

```
Feature: 0, Score: 0.00296
Feature: 1, Score: 0.00547
Feature: 2, Score: 0.00279
Feature: 3, Score: 0.00281
Feature: 4, Score: 0.53152
Feature: 5, Score: 0.41887
Feature: 6, Score: 0.02673
Feature: 7, Score: 0.00307
Feature: 8, Score: 0.00301
Feature: 9, Score: 0.00275
```

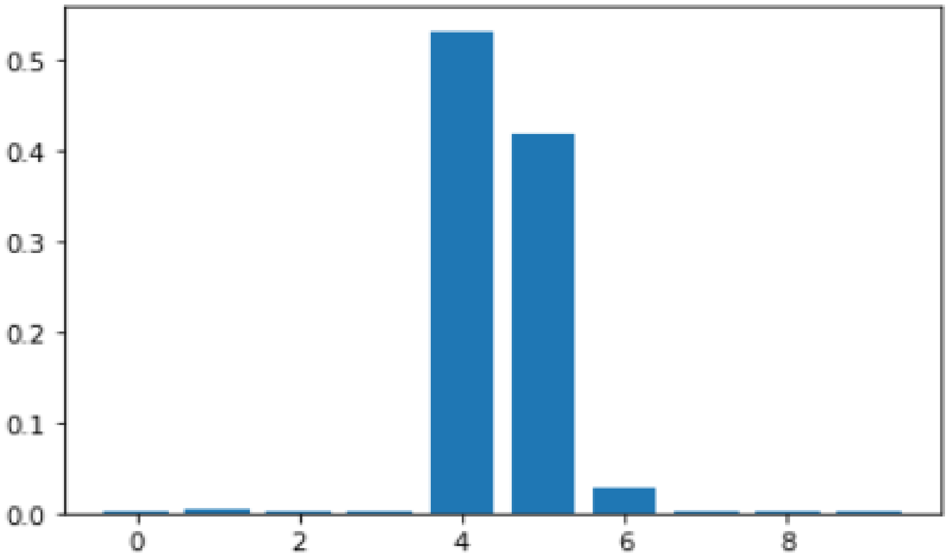

Figure 3 : Feature Selection of Variables

In this above figure 3, internal factor PE ratio and external factor interest rate has higher value. Which can gives the better prediction. Similarly other variables have very small value.so these have small effect in stock price prediction.

3.2.6 Correlation Matrix

<matplotlib.axes._subplots.AxesSubplot at 0x7f2114bf9fd0>

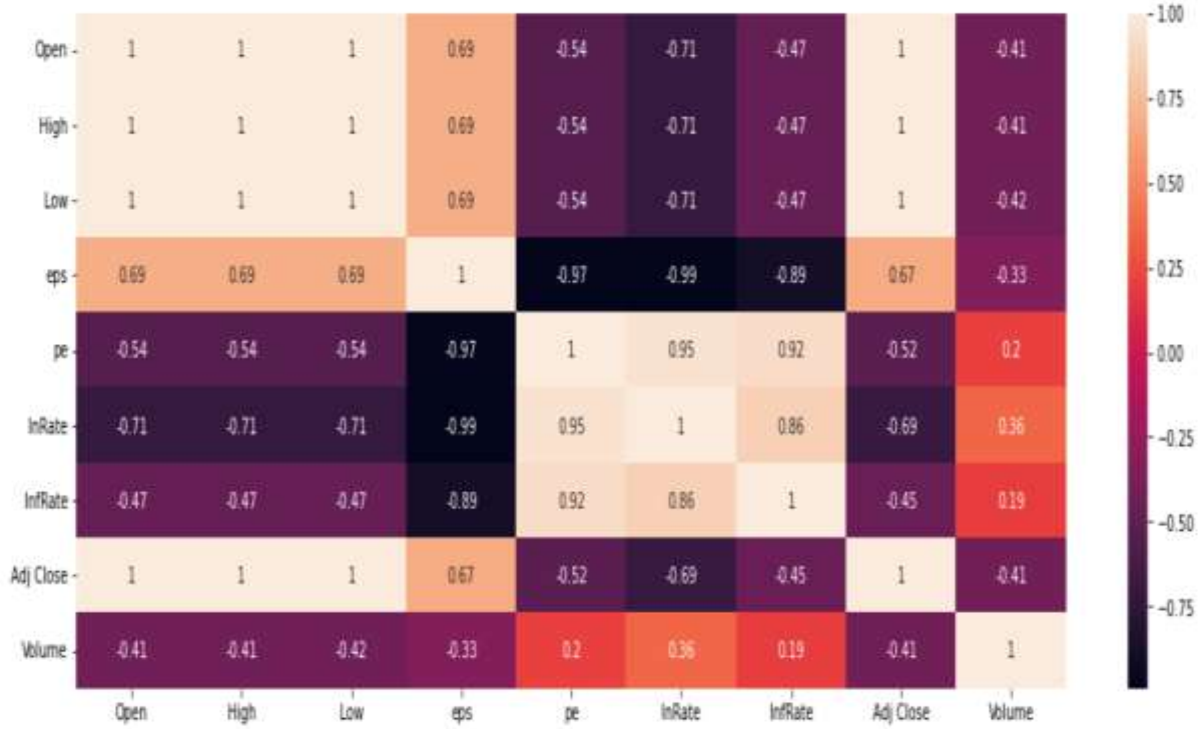

Figure 4 : Correlation Matrix

In this above figure 4 Adj Close, High, InRate and Low are correlated to each other on the basis of Close value. So all of these four variable we must kept any one and remove other remaining variables.

3.3 Proposed system analysis

3.3.1 Model Used

We used the following model in our project:

3.3.1.1 Long Short-Term Memory Model

LSTM (Long Short-Term Memory) is modified version of RNN (Recurrent Neural Network). RNN is best suitable for stock price prediction as it can analyze time-series patterns. The limitation of RNN is that it cannot save state for long-term dependencies. In RNN, values are back propagated and slope becomes small which results in vanishing gradient issue [12]. LSTM overcomes this limitation by saving states in cell state. Furthermore, there is Forget gate available in LSTM which filters whether previous state information is relevant or not. If Forget gate output is 1, cell state saves the information and if output is 0, cell state ignores the information. Input and output gates are also used in LSTM.

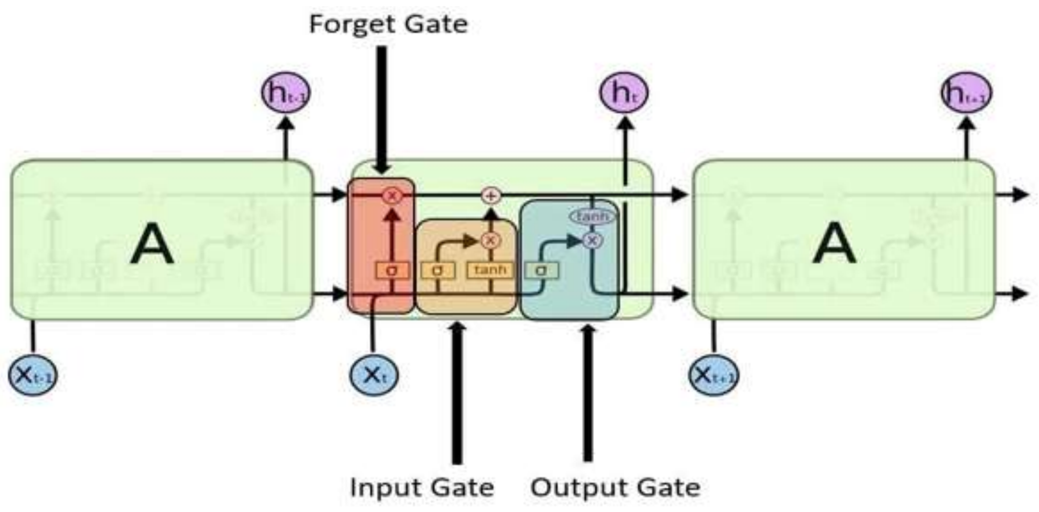

Figure 5: LSTM Architecture

It is clearly depicted in figure 2 that Input, Forget and Output gates are deployed in LSTM. Input gate filters the information from previous layers and output gate filters output that is to be sent to next layer.

In order to predict future trends, RNN's primary principle is to use the sequential observations learned from earlier phases. RNN has been modified with the Long-Short Term Memory (LSTM) model. The RNN's inability to detect long-term impacts can be overcome by this method.

The method of using components of previous sequences to predict future data is referred to as recurrent. The Long Short-Term Memory (LSTM) based on a "memory line" proved to be highly helpful in forecasting scenarios with long-time data because RNN cannot store long-time memory. An LSTM contains gates along the memory line that can be used to memorize previous stages.

We used ‘Closing Price’ from our dataset for LSTM modeling in order to predict the stock price. We then pre-process this data and transform the values in our data with the help of the fit transform function. Min-max scaler is used for scaling the data so that we can bring all the price values to a common scale. We then use 80% data for training and the rest 20% for testing and assign them to separate variables.

Finally, the Adam optimizer is used for learning the parameters, and the mean squared error is utilized as the loss function.

We used a batch size of 64 in our training procedure to speed up training and improve convergence probability. As selecting larger sizes doesn't always guarantee convergence.

In this study, we examined the stock price for the coming ten days. It is depicted in the following figure.

```
In [19]: model=Sequential()
         model.add(LSTM(50,return_sequences=True,input_shape=(100,1)))  # input_shape= (x_train.shape[1], 1)
         model.add(LSTM(50,return_sequences=True))
         model.add(LSTM(50))
         model.add(Dense(1))
         model.compile(loss='mean_squared_error',optimizer='adam')

In [20]: model.summary()

         Model: "sequential"
         _________________________________________________________________
          Layer (type)                Output Shape              Param #
         =================================================================
          lstm (LSTM)                 (None, 100, 50)           10400

          lstm_1 (LSTM)               (None, 100, 50)           20200

          lstm_2 (LSTM)               (None, 50)                20200

          dense (Dense)               (None, 1)                 51

         =================================================================
         Total params: 50,851
         Trainable params: 50,851
         Non-trainable params: 0
```

Figure 6: Summary of the LSTM model

```
In [74]: plt.plot(df3)

Out[74]: [<matplotlib.lines.Line2D at 0x1a4636b5d00>]
```

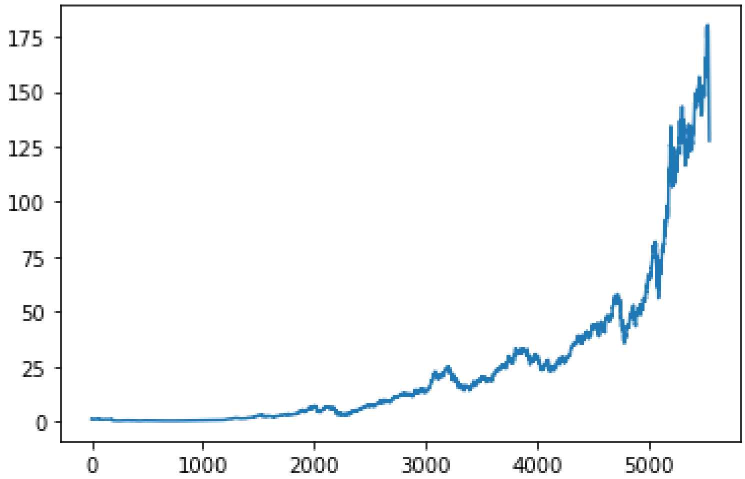

Figure 7: LSTM model for predicting the stock price for the coming ten days

```
...   Mean Squared Error: 90.12949997092743
      Mean absolute error: 81.37418995072238
      Root Mean Squared Error: 90.12949997092743
```

Figure 8: MAE and RMSE of LSTM model

3.3.1.2 Random Forest Model

The Random Forest [13] is a model based on decision trees that resolves the problem of over fitting existing in simple decision tree models by bagging on training data and variables: i.e., randomly sampling the training data with replacement and randomly selecting variables that are used to predict during each iteration. We performed a grid search of the number of estimators, maximum depth, min_sample_split, min_sample_split, bootstrap, and random state to find the best fit.

We arrange the data into X and y, and then pre-process X and y data and transform the values in our data with the help of the fit transform function. We then split 80% data for training and the rest 20% for testing and assign them to separate variables. And StandardScaler is used for scaling the data so that we can bring all feature columns values to a common scale. After that, we adjust the hyperparameters for the model to find the ones that suit the data the best. We then remodeled the data using these best parameters and calculated regression metrics. Then, we examined the stock price for the coming ten days.

```
[28]: grid_rf = {
      'n_estimators': [20, 50, 100, 500, 1000],
      'max_depth': np.arange(1, 15, 1),
      'min_samples_split': [2, 10, 9],
      'min_samples_leaf': np.arange(1, 15, 2, dtype=int),
      'bootstrap': [True, False],
      'random_state': [1, 2, 30, 42]
      }
      rscv = RandomizedSearchCV(estimator=RandomForestRegressor(), param_distributions=grid_rf, cv=10, n_jobs=-1, verbose=2, n_iter=20
      rscv_fit = rscv.fit(X_train, y_train)
      best_parameters = rscv_fit.best_params_
      print(best_parameters)
```

Apply model and Predict

```
[29]: model = RandomForestRegressor(n_estimators=100, random_state=30, min_samples_split=2, min_samples_leaf=1, max_depth=12, bootstrap
      model.fit(X_train, y_train)
      predict = model.predict(X_test)
```

Figure 9: Hyperparameter tuning of the Random Forest Regressor

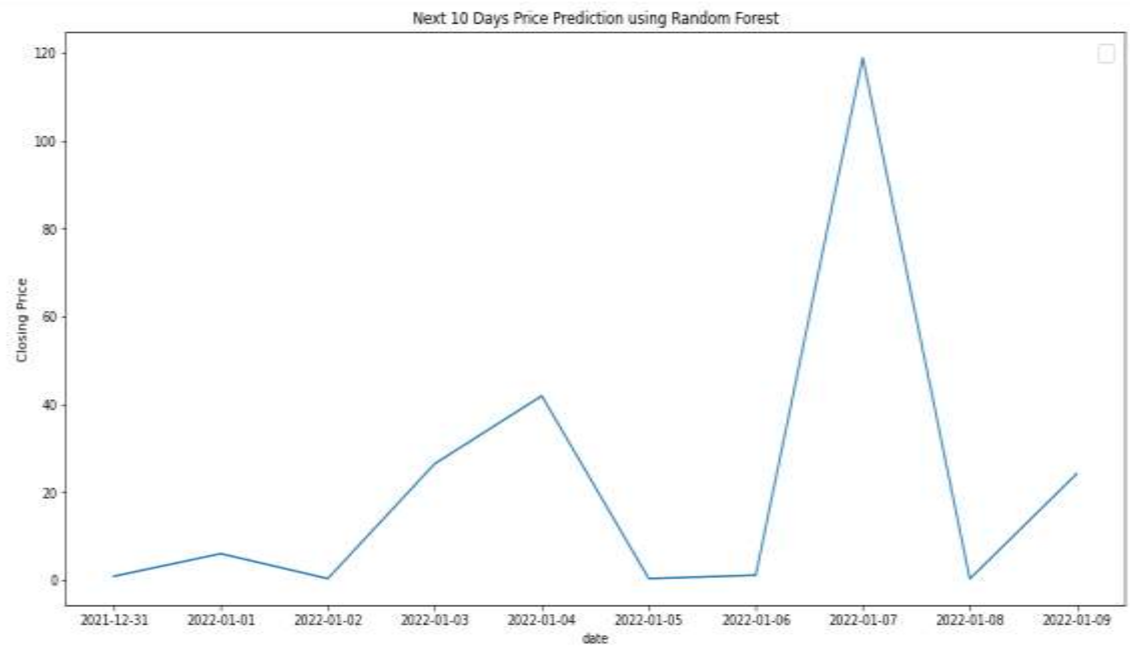

Figure 10: Random Forest Regressor model for predicting the stock price for the coming ten days

```
Mean Absolute Error: 0.03662
Mean Squared Error: 0.02138
Root Mean Squared Error: 0.14621
(R^2) Score: 0.99998
Train Score : 100.00% and Test Score : 100.00% using Random Tree Regressor.
```

Figure 11: MAE and RMSE of Random Forest Regressor

3.3.1.3 Support Vector Regression

SVR is regression technique to minimize error using most suitable hyper plane. Different kernel functions such as linear, RBF, Sigmoid and Polynomial are used and selecting kernel function is significant for regression [14]. Minimum Loss function is used to improve prediction accuracy. We performed a grid search of the number of estimators, maximum depth, min_sample_split, min_sample_split, bootstrap, and random state to find the best fit.

We arrange the data into X and y, and then pre-process X and y data and transform the values in our data with the help of the fit_transform function. We then split 80% data for training and the rest 20% for testing and assign them to separate variables. And StandardScaler is used for scaling the data so that we can bring all feature columns values to a common scale. After that, we adjust the hyperparameters for the model to find the ones that suit the data the best. We then remodeled the data the best line or decision boundary. using these best parameters and calculated regression metrics. Then, we examined the stock price for the coming ten days.

```
In [23]: from sklearn.model_selection import RandomizedSearchCV
         svm = {
             'C': [0.1, 1, 10, 100, 1000],
             'gamma': [1, 0.1, 0.01, 0.001, 0.0001],
             'kernel': ['rbf','linear', 'poly', 'sigmoid']
             }


         rscv = RandomizedSearchCV(SVR(),param_distributions=svm,cv=10,n_iter=10,verbose=True,n_jobs=-1)
         rscv_fit = rscv.fit(X_train, y_train)
         best_parameters = rscv_fit.best_params_
         print(best_parameters)

In [25]: model = SVR(kernel= 'linear', gamma= 0.0001, C=100)
         model.fit(X_train,y_train)

Out[25]: SVR(C=100, gamma=0.0001, kernel='linear')
```

Figure 12: Hyperparameter tuning of the Support Vector Regressor

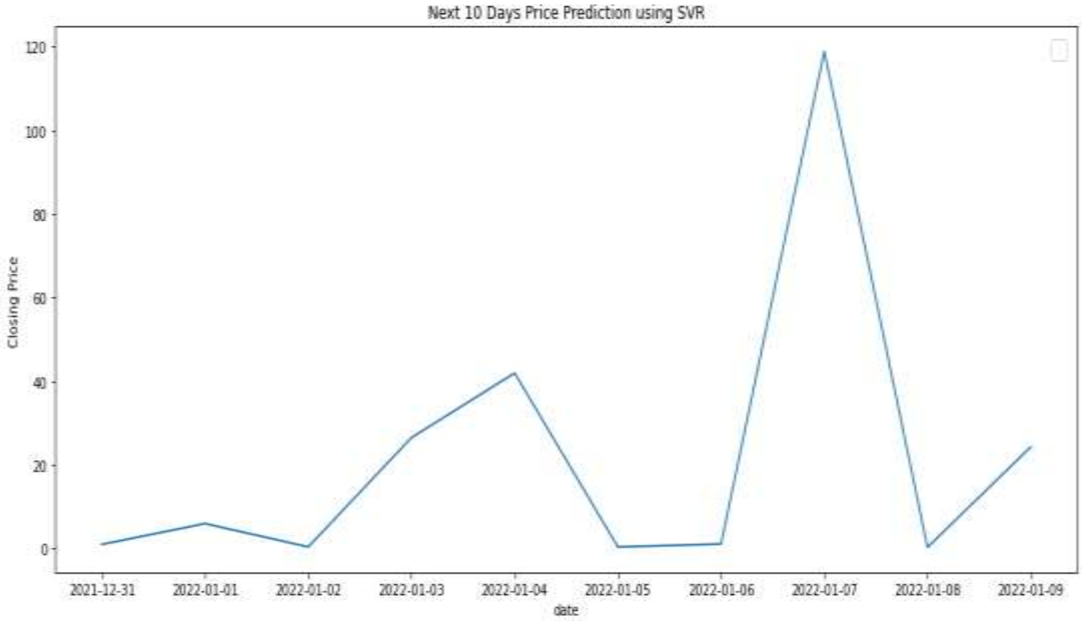

Figure 13: Support Vector Regressor model for predicting the stock price for the coming ten days

```
Mean Absolute Error: 0.04508
Mean Squared Error: 0.00271
Root Mean Squared Error: 0.05208
(R^2) Score: 1.0
Train Score : 100.00% and Test Score : 100.00% using Support Vector Regressor.
```

Figure 14: MAE and RMSE of Support Vector Regressor

3.4 System Design Diagram

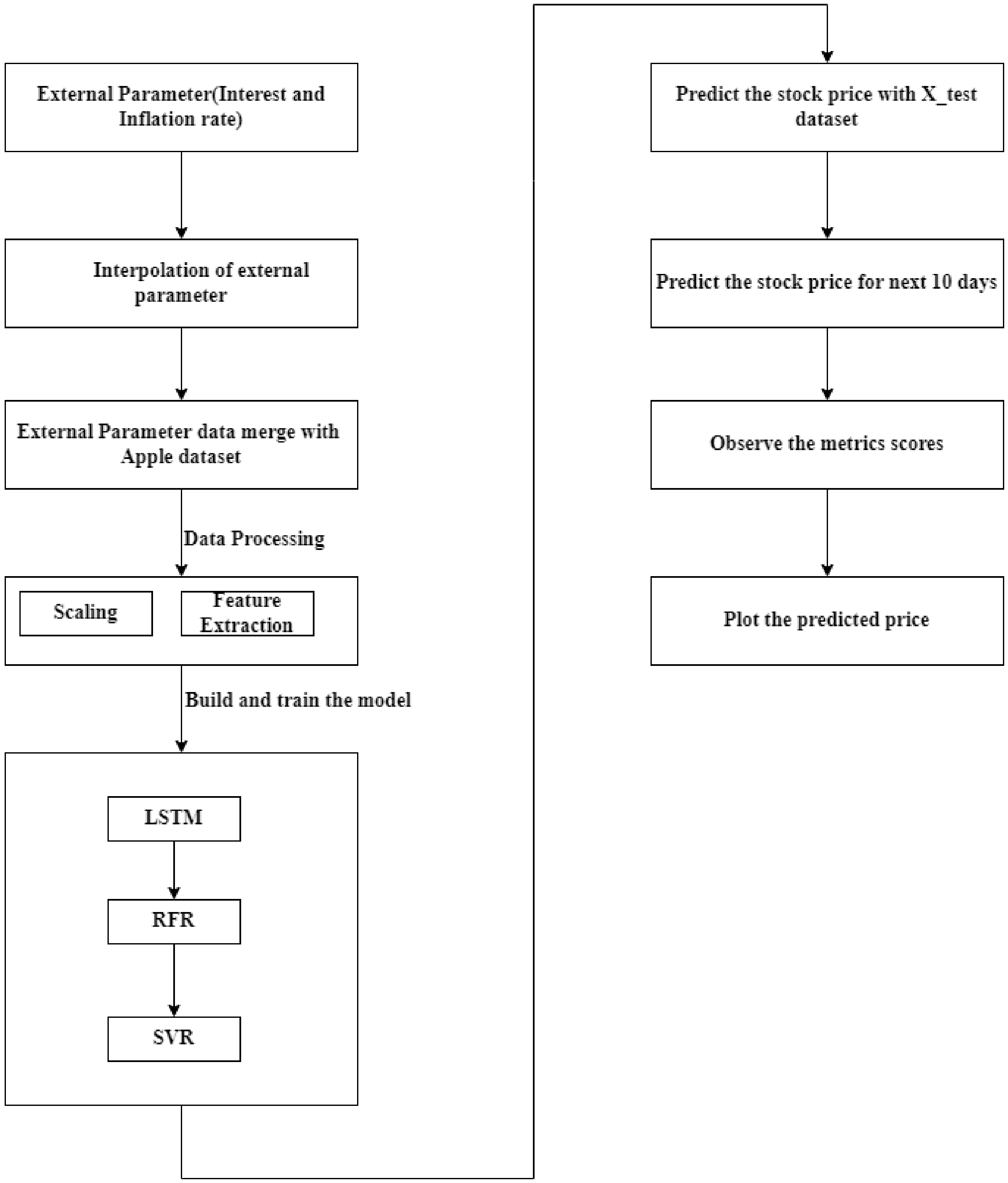

Figure 15: System Design Diagram

3.5 Implementation

3.5.1 Hardware Requirement

- Computer
- Mouse
- Keyboard
- Minimum 128 RAM Minimum 500 MB hard disk

3.5.2 Software Requirement

The software is the non-physical part of the system that uses the hardware components to successfully run the system that has been built. The system must have word processor. The system will run windows Operating System.

Operating system: Windows, Linux

Language: Python

Code Editor: VS Code, Jupyter, Google Colaboratory etc.

Spreadsheet: Excel

- **Python:**

Python is a general purpose, dynamic, high-level, and interpreted programming language. It supports Object Oriented programming approach to develop applications. It is simple and easy to learn and also provides a lot of high-level data structures.

- **Jupyter:**

Jupyter Notebook allows users to compile all aspects of a data project in one place making it easier to show the entire process of a project to your intended audience.

CHAPTER 4: RESULT CONCLUSION AND RECOMENDATION

4.1 Results and Analysis

After carefully tuning each individual model's parameters and preprocessing options, we report the results obtained from each model, as below table 1 shows.

Table 6: Compare Performance of three different predictions models

Model	Tuning parameters	MAE	RMSE
LSTM Model	2 LSTM layer, 50 nodes each, Dropout 0.2, batch size 64, epochs 100.	81.37418	90.12949
Support Vector Regressor	C=100, Gamma=0.0001, Kernel=Linear	0.04508	0.05208
Random Forest Regressor	n_estimators=100, random state=30, min_sample_split=2, min_samples_leaf=1, max_depth=12, bootstrap=True)	0.0362	0.14611

In general, the LSTM model performs worse than other models, and we view it as a baseline model. On the other hand, the Random Forest Regression model and Support Vector Regression Model gives less root mean squared error and mean absolute error accuracy than LSTM model. Experimental results showed that the best linear regression models trained is Support Vector Regressor.

4.2 Limitations

- Although there are many internal and external factors, we selected a small number of parameters to predict the stock price. As a result, the prediction is not acceptable.

4.3 Conclusion

Stock price forecasting is challenging due to noisy, dynamic, and nonlinear data in the stock market. The accurate prediction of stock prices helps investors increase profits in the financial market. Identifying the trend in the market is a challenging task. We proposed three different models (LSTM, SVR and RFR). The proposed model overcomes the instability in the prediction model. We have fine-tuned the hyper parameters manually for better accuracy, such as the model's number of layers, learning rate, neurons, and number of epochs. The stock price model is evaluated using the metrics MAE and RMSE.

Financial markets provide an excellent platform for investors and traders, who can trade from any gadget that connects to the internet. Over the last few years, people have become more attracted to stock trading. We are predicting the closing stock price of any given company, so we have used datasets belonging to Apple.

4.4 Future work

- Exploring the different internal and external parameter for possible prediction.
- Applying more models and compare the result on basis of RMSE, MSE, and Accuracy.
- Future iterations on this work should first try to improve model generalization error and reduce overfitting.

- Adding more training data, reducing the size of the feature space through feature selection or similar techniques, and tuning the model hyperparameters would help tackle overfitting.

CHAPTER 5: REFERENCE

[1] C. K.-S. R. K. M. a. Y. W. Leung, "A machine learning approach for stock price prediction.," Proceedings of the 18th International Database Engineering & Applications Symposium., 05 May 2015. [Online]. [Accessed 2 March 2015].

[2] X. J. W. a. Z. Y. Ji, " "A stock price prediction method based on deep learning technology."," *International Journal of Crowd Science* , 05 march 2021.

[3] V. K. S. Reddy, "Stock market prediction using machine learning.," *International Research Journal of Engineering and Technology (IRJET),* vol. 05, no. 10, pp. 1033-1035, 2018.

[4] L. R. Marwala, "Forecasting the stock market index using artificial intelligence techniques.," *Stock Price Prediction,* vol. i, pp. 1-166, 2010.

[5] D. V. K. a. A. M. Selvamuthu, "Indian stock market prediction using artificial neural networks on tick data," *Financial Innovation,* vol. i, no. 5, pp. 1-12, 2019.

[6] M. I. Moukalled, "Automated stock price prediction using machine learning," vol. I, no. 5, pp. 1-81, 2019.

[7] H. K. Choi, "Stock price correlation coefficient prediction with ARIMA-LSTM hybrid model.," *arXiv preprint arXiv:,* vol. v, no. 5, pp. 1-3, 2018.

[8] P. a. S. R. Petcharabul, "Technology industry on financial ratios and stock returns.," *Journal of Business and Economics,* vol. 5, no. 5, pp. 739-746, 2014.

[9] C. K. Y. V. C. a. N. Q. B. A. Thim, "Stock performance of the property sector in Malaysia.," *Journal of Modern Accounting and auditing,* vol. 8, no. 2, p. 241, 2012.

[10] W. a. A. W. Idawati, "Effect of earning per shares (EPS) and return on assets (ROA) against share price on coal mining company listed in Indonesia stock

exchange.," *Journal of Resources Development and Management,* vol. 7, pp. 79-91, 2015.

[11] U. Ergun, "Internal determinants of the stock price movements on sector basis.," *International Research Journal of Finance and Economics* , no. 92, pp. 111-117, 2012.

[12] H. D. Q. H. a. M. R. Xue, "LSTM: A hierarchical LSTM model," *LSTM for stock price prediction,* vol. iii, no. 7, pp. 1186-1194, 2018.

[13] K. O. A. A. a. B. W. Nti, "Random forest for stock price prediction," *Random forest,* vol. 7, no. 16, pp. 200-212, 2019.

[14] H. L. C. a. I. K. Yang, "Support vector machine regression for volatile stock market prediction," *international conference on intelligent data engineering and automated learning,* vol. ii, no. 6, pp. 391-396, 2002.

[15] J. S. S. T. P. &. K. K. Patel, "Predicting stock market index using fusion of machine learning techniques.," *Expert Systems with Applications,* vol. iii, no. 42, pp. 2162-2172, Mar-2015.

[16] B. W. Boehm, A spiral model of software development and enhancement, india: ieeexplore.ieee.org, 1988.