

PROFESSIONAL TRAINING REPORT

at

**Sathyabama Institute of Science and Technology
(Deemed to be University)**

Submitted in partial fulfillment of the requirements for the award
of Bachelor of Technology Degree in Information Technology

By

**GAYATHRI K S
(37120027)**



**DEPARTMENT OF INFORMATION TECHNOLOGY
SCHOOL OF COMPUTING**

**SATHYABAMA
INSTITUTE OF SCIENCE AND
TECHNOLOGY (DEEMED TO
BE UNIVERSITY)**

**Accredited with Grade "A" by NAAC
JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI - 600119
DECEMBER 2020**



SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY

(Established under Section 3 of UGC Act, 1956)
Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai -

600119 www.sathyabama.ac.in



SCHOOL OF COMPUTING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **GAYATHRI K.S. (Reg.No. 37120027)** and who carried out the project entitled “**STOCK PREDICTION**” under our supervision from July 2020 to December 2020.

-
Internal Guide

Dr. Y. BEVISH JINILA., M.E., Ph.D.,

Head of the Department

Dr. R. SUBHASHINI M.E., Ph.D.,

Submitted for Viva voce Examination held on

Internal Examiner

External Examiner

DECLARATION

I **GAYATHRI K.S. (REG NO:37120027)** hereby declare that the Project Report entitled “**STOCK PREDICTION**” done by me under the guidance of **Dr. BEVISH JINILA. Y, M.E., Ph.D.**, is submitted in partial fulfillment of the requirements for the award of Bachelor of Technology degree in Information Technology.

DATE:

PLACE:

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of management of SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. SASIKALA M.E., PhD.**, Dean, School of Computing and **Dr. R.SUBASHINI, M.E.,Ph.D.**, Head of the Department, Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. BEVISH JINILA.Y, M.E.,Ph.D.**, for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express our thanks to all Teaching and Non-teaching staff members of the Department of **INFORMATION TECHNOLOGY** who were helpful in many ways for the completion of the project.

TRAINING CERTIFICATE





RJP INFOTEK PVT. LTD.

This is to certify that

Gayathri K.S

has successfully completed

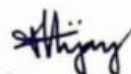
VMWARE Authorized VCA - Datacenter Virtualization Training

@ SATHYABAMA INSTITUTE OF SCIENCE & TECHNOLOGY



This certificate earned on

24/10/2019


Course Co-ordinator

This certificate can be verified at RJP INFOTEK, Ph. 2473 1102

ABSTRACT

The main objective of this project is to find the best model to predict the value of the stock market. During the process Of considering various techniques and variables that must be taken into account, I found out that techniques like random forest, support vector machine. In this project I am going to present and review a more feasible method to predict the stock movement with higher accuracy. The first thing I have taken into account is the dataset of the stock market prices from previous year. The dataset was pre-processed and tuned up for real analysis. Hence, This Project will also focus on data preprocessing of the raw dataset. Secondly, after pre- processing the data, I will review the use of support vector machine on the dataset and the outcomes it generates. The project also presents a machine-learning model to predict the longevity of stock in a competitive market. The successful prediction of the stock will be a great asset for the stock market institutions and will provide real-life solutions to the problems that stock investors face.

LIST OF FIGURES

FIGURE NO	FIGURE NAME	PAGE NO
4.1	SYSTEM ARCHITECTURE	10
4.2	USE CASE DIAGRAM	11
4.3	OUTPUT	12

LIST OF ABBREVIATIONS

ABBREVIATION	EXPANSION
SVM	SUPPORT VECTOR MACHINE
SVR	SUPPORT VECTOR REGRESSION
RAM	RANDOM ACCESS MEMORY

TABLE OF CONTENTS

CHAPTER NO	TITLE	PAGE NO
	Abstract	i
	List of Figure	ii
	List of Abbrevation	iii
1	INTRODUCTION	
	1.1 General	1
	1.2 Problem Statement	1
	1.3 Objectives	2
2	AIM & SCOPE	
	2.1 Aim	3
	2.2 Problem Definition	3
	2.3 Literature Review	4
	2.4 Proposed System	5
3	ALGORITHM & METHODS	
	3.1 Classification	6
	3.1.1 SVM Classifier	6
	3.1.2 Parameters	6
	3.2 Algorithm	7
	3.3 Modules	7
	3.3.1 Data Collection	7
	3.3.2 Pre Processing	7
	3.3.3 Training The Machine	8
4	SYSTEM REQUIREMENTS & DESIGNS	
	4.1 Requirements	9
	4.1.1 Hardware Requirements	9
	4.1.2 Software Requirements	9

	4.2 System Archeitecture	11
	4.3 Result & Discussions	12
	4.3.1 Analysis	12
	4.3.2 Result	12
5	CONCLUSION & FUTUREWORK	
	5.1 Conclusion	13
	5.2 FutureEnchacement	13
	REFERENCE	14
	APPENDIES	15

CHAPTER 1

INTRODUCTION

1.1 GENERAL

Stock Market prediction and analysis is the act of trying to determine the future value of a company stock or other financial instrument traded on an exchange. Stock market is the important part of economy of the country and plays a vital role in the growth of the industry and commerce of the country that eventually affects the economy of the country. Both investors and industry are involved in stock market and wants to know whether some stock will rise or fall over certain period of time. The stock market is the primary source for any company to raise funds for business expansions. It is based on the concept of demand and supply. If the demand for a company's stock is higher, then the company share price increases and if the demand for company's stock is low then the company share price decrease.

1.2 PROBLEM STATEMENT

Stock market is very vast and difficult to understand. It is considered too uncertain to be predictable due to huge fluctuation of the market. Stock market prediction task is interesting as well as divides researchers and academics into two groups, those who believe that we can devise mechanisms to predict the market and those who believe that the market is efficient and whenever new information comes up the market absorbs it by correcting itself, thus there is no space for prediction. Investing in a good stock but at a bad time can have disastrous result, while investing in a stock at the right time can bear profits. Financial investors of today are facing this problem of trading as they do not properly understand as to which stocks to buy or which stocks to sell in order to get optimum result. So, the purposed project will reduce the problem with suitable accuracy faced in such real time scenario.

1.3 OBJECTIVES

The project will be useful for investors to invest in stock market based on the various factors. The project target is analyses the previous stock data of companies and implement these values in machine learning algorithm to determine the value that particular stock will have in near future with suitable accuracy. These predicted and analyzed data can be observed by individual to know the financial status of companies and their comparisons. Company and industry can use it to breakdown their limitation and enhance their stock value. It can be very useful to even researchers, stock brokers, market makers, government and general people.

CHAPTER 2

AIM & SCOPE

2.1 AIM

The stock market is basically an aggregation of various buyers and sellers of stock. A stock (also known as shares more commonly) in general represents ownership claims on business by a particular individual or a group of people. To determine the future value of the stock market is known as a stock market prediction. The prediction is expected to be robust, accurate and efficient.

2.2 PROBLEM DEFINITION

Stock market prediction is basically defined as trying to determine the stock value and offer a robust idea for the people to know and predict the market and the stock prices. It is generally presented using the quarterly financial ratio using the dataset. Thus, relying on a single dataset may not be sufficient for the prediction and can give a result which is inaccurate. Hence, we are contemplating towards the study of machine learning with various datasets integration to predict the market and the stock trends. The problem with estimating the stock price will remain a problem if a better stock market prediction algorithm is not proposed. Predicting how the stock market will perform is quite difficult. The movement in the stock market is usually determined by the sentiments of thousands of investors. Stock market prediction, calls for an ability to predict the effect of recent events on the investors. These events can be political events like a statement by a political leader, a piece of news on scam etc. It can also be an international event like sharp movements in currencies and commodity etc. All these events affect the corporate earnings, which in turn affects the sentiment of investors. It is beyond the scope of almost all investors to correctly and consistently predict these hyperparameters. All these factors make stock price prediction very difficult. Once the right data is collected, it then can be used to train a machine and to generate a predictive result.

2.3 LITERATURE REVIEW

During a literature survey, we collected some of the information about Stock market prediction mechanisms currently being used. The recent studies provide a well-grounded proof that most of the predictive regression models are inefficient in out of sample predictability test. The reason for this inefficiency was parameter instability and model uncertainty. The studies also concluded the traditional strategies that promise to solve this problem. Support vector machine commonly known as SVM provides with the kernel, decision function, and sparsity of the solution. It is used to learn polynomial radial basis function and the multi-layer perceptron classifier. It is a training algorithm for classification and regression, which works on a larger dataset. There are many algorithms in the market but SVM provides with better efficiency and accuracy. The correlation analysis between SVM and stock market indicates strong interconnection between the stock prices and the market index. Financial organizations and merchants have made different exclusive models to attempt and beat the market for themselves or their customers, yet once in a while has anybody accomplished reliably higher-than-normal degrees of profitability. Nevertheless, the challenge of stock forecasting is so engaging in light of the fact that the improvement of only a couple of rate focuses can build benefit by a large number of dollars for these organizations.

2.4 PROPOSED SYSTEM

In this proposed system, we focus on predicting the stock values using machine learning algorithms like Random Forest and Support Vector Machines. We proposed the system "Stock market price prediction" we have predicted the stock market price using the random forest algorithm. In this proposed system, we were able to train the machine from the various data points from the past to make a future prediction. We took data from the previous year stocks to train the model. We majorly used two machine-learning libraries to solve the problem. The first one was numpy, which was used to clean and manipulate the data, and getting it into a form ready for analysis. The other was scikit, which was used for real analysis and prediction. The data set we used was from the previous years stock markets collected from the public database available online, 80 % of data was used to train the machine and the rest 20 % to test the data.

The basic approach of the supervised learning model is to learn the patterns and relationships in the data from the training set and then reproduce them for the test data. We used the python pandas library for data processing which combined different datasets into a data frame. The tuned up dataframe allowed us to prepare the data for feature extraction. The dataframe features were date and the closing price for a particular day. We used all these features to train the machine on random forest model and predicted the object variable, which is the price for a given day. We also quantified the accuracy by using the predictions for the test set and the actual values. The proposed system touches different areas of research including data pre-processing, random forest, and so on.

CHAPTER 3

ALGORITHMS AND METHODS

3.1 CLASSIFICATIONS

Classification is an instance of supervised learning where a set is analyzed and categorized based on a common attribute. From the values or the data are given, classification draws some conclusion from the observed value. If more than one input is given then classification will try to predict one or more outcomes for the same. A few classifiers that are used here for the stock market prediction includes the random forest classifier, SVM classifier.

3.1.1 SVM CLASSIFIER

SVM classifier is a type of discriminative classifier. The SVM uses supervised learning i.e. a labeled training data. The output are hyperplanes which categorizes the new dataset. They are supervised learning models that uses associated learning algorithm for classification and as well as regression.

3.1.2 Parameters

The tuning parameters of SVM classifier are kernel parameter, gamma parameter and regularization parameter.

> Kernels can be categorized as linear and polynomial kernels calculates the prediction line. In linear kernels prediction for a new input is calculated by the dot product between the input and the support vector.

> C parameter is known as the regularization parameter; it determines whether the accuracy of model is increases or decreases. The default value of $c=10$. Lower regularization value leads to misclassification.

> Gamma parameter measures the influence of a single training on the model. Low values signifies far from the plausible margin and high values signifies closeness from the plausible margin.

3.2 ALGORITHM

The main task of the support machine algorithm is to identify an N-dimensional space that distinguishably categorizes the data points. Here, N stands for a number of features. Between two classes of data points, there can be multiple possible hyperplanes that can be chosen. The objective of this algorithm is to find a plane that has maximum margin. Maximizing margin refers to the distance between data points of both classes. The benefit associated with maximizing the margin is that it provides reinforcement so that future data points can be more easily classified. Decision boundaries that help classify data points are called hyperplanes. Based on the position of the data points relative to the hyperplane they are attributed to different classes. The dimension of the hyperplane relies on the number of attributes, if the number of attributes is two then the hyperplane is just a line, if the number of attributes is three then the hyperplane is two dimensional.

3.3 MODULES

3.3.1 Data Collection

Data collection is a very basic module and the initial step towards the project. It generally deals with the collection of the right dataset. The dataset that is to be used in the market prediction has to be used to be filtered based on various aspects. Data collection also complements to enhance the dataset by adding more data that are external

3.3.2 Pre Processing

Data pre-processing is a part of data mining, which involves transforming raw data into a more coherent format. Raw data is usually, inconsistent or incomplete and usually contains many errors. The data pre-processing involves checking out for missing values, looking for categorical values, splitting the data-set into training and test set and finally do a feature scaling to limit the range of variables so that they can be compared on common environs.

3.3.3 TRAINING THE MACHINE

Training the machine is similar to feeding the data to the algorithm to touch up the test data. The training sets are used to tune and fit the models. The test sets are untouched, as a model should not be judged based on unseen data. The training of the model includes cross-validation where we get a well-grounded approximate performance of the model using the training data. Tuning models are meant to specifically tune the hyperparameter. Finally, we will calculate a cross-validated score, for individual sets of hyperparameters. Then, we select the best hyperparameters. The idea behind the training of the model is that we some initial values with the dataset and then optimize the parameters which we want to in the model. This is kept on repetition until we get the optimal values. Thus, we take the predictions from the trained model on the inputs from the test dataset. Hence, it is divided in the ratio of 80:20 where 80% is for the training set and the rest 20% for a testing set of the data.

CHAPTER 4

SYSTEM REQUIREMENTS AND DESIGN

4.1 REQUIREMENTS

4.1.1 HARDWARE REQUIREMENTS

- A system with minimum 2GB of RAM.
- Keyboard.
- Monitor.

4.1.2 SOFTWARE REQUIREMENTS

- Python 3.7 version installed in the system.
- using 'pip' command
 - >Import numpy
 - >Import skit learn
 - >Import quandl

4.2 SYSTEM ARCHITECTURE

The first step is the conversion of this raw data into processed data. This is done using feature extraction, since in the raw data collected there are multiple attributes but only a few of those attributes are useful for the purpose of prediction. So the first step is feature extraction, where the key attributes are extracted from the whole list of attributes available in the raw dataset. Feature extraction starts from an initial state of measured data and builds derived values or features. These features are intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps. Feature extraction is a dimensionality reduction process, where the initial set of raw variables is diminished to progressively reasonable features for ease of management, while still precisely and totally depicting the first informational collection. The feature extraction process is followed by a classification process

wherein the data that was obtained after feature extraction is split into two different and distinct segments. Classification is the issue of recognizing to which set of categories a new observation belongs. The training data set is used to train the model whereas the test data is used to predict the accuracy of the model. The splitting is done in a way that training data maintain a higher proportion than the test data. The end goal of our proposed system is to predict the price of the stock by analyzing its historical data.

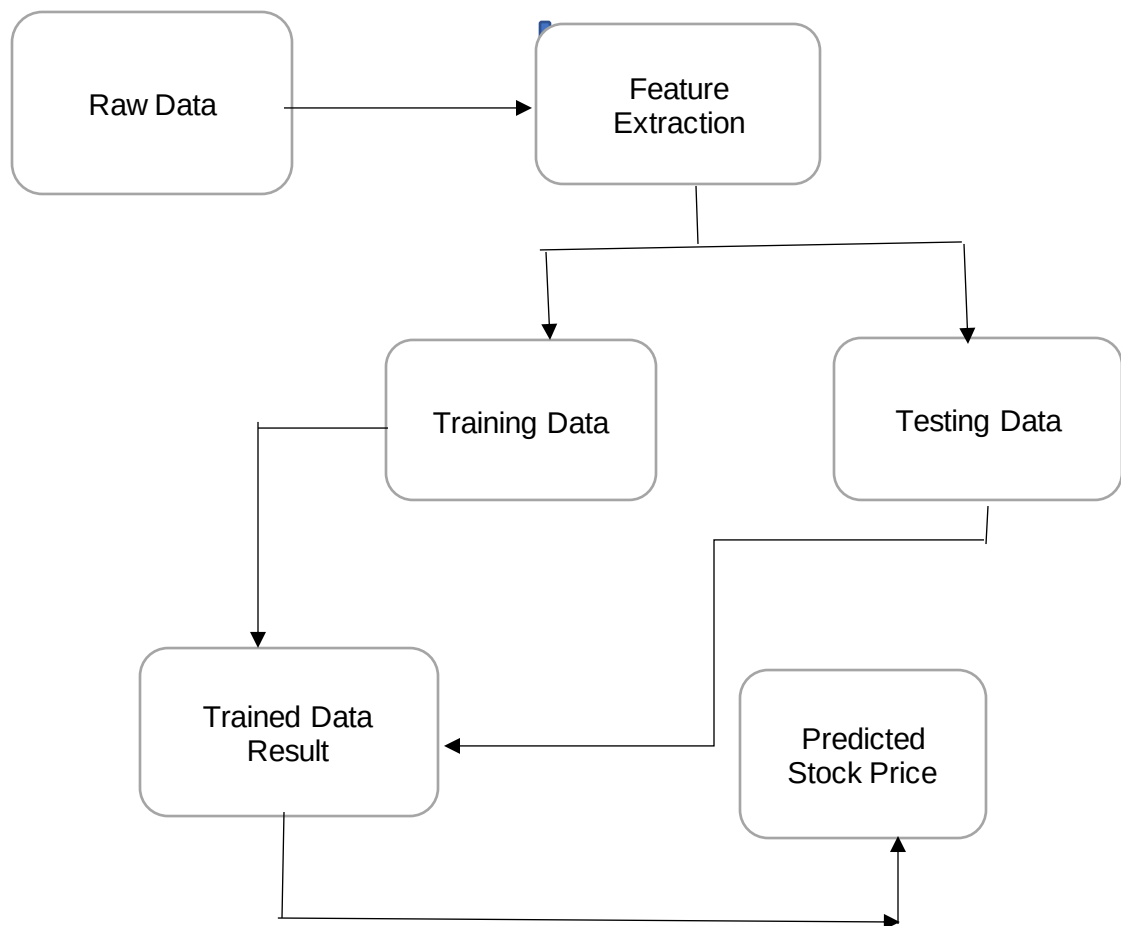


FIG NO:4.1 SYSTEM ARCHITECTURE

4.2.1 USECASE DIAGRAM

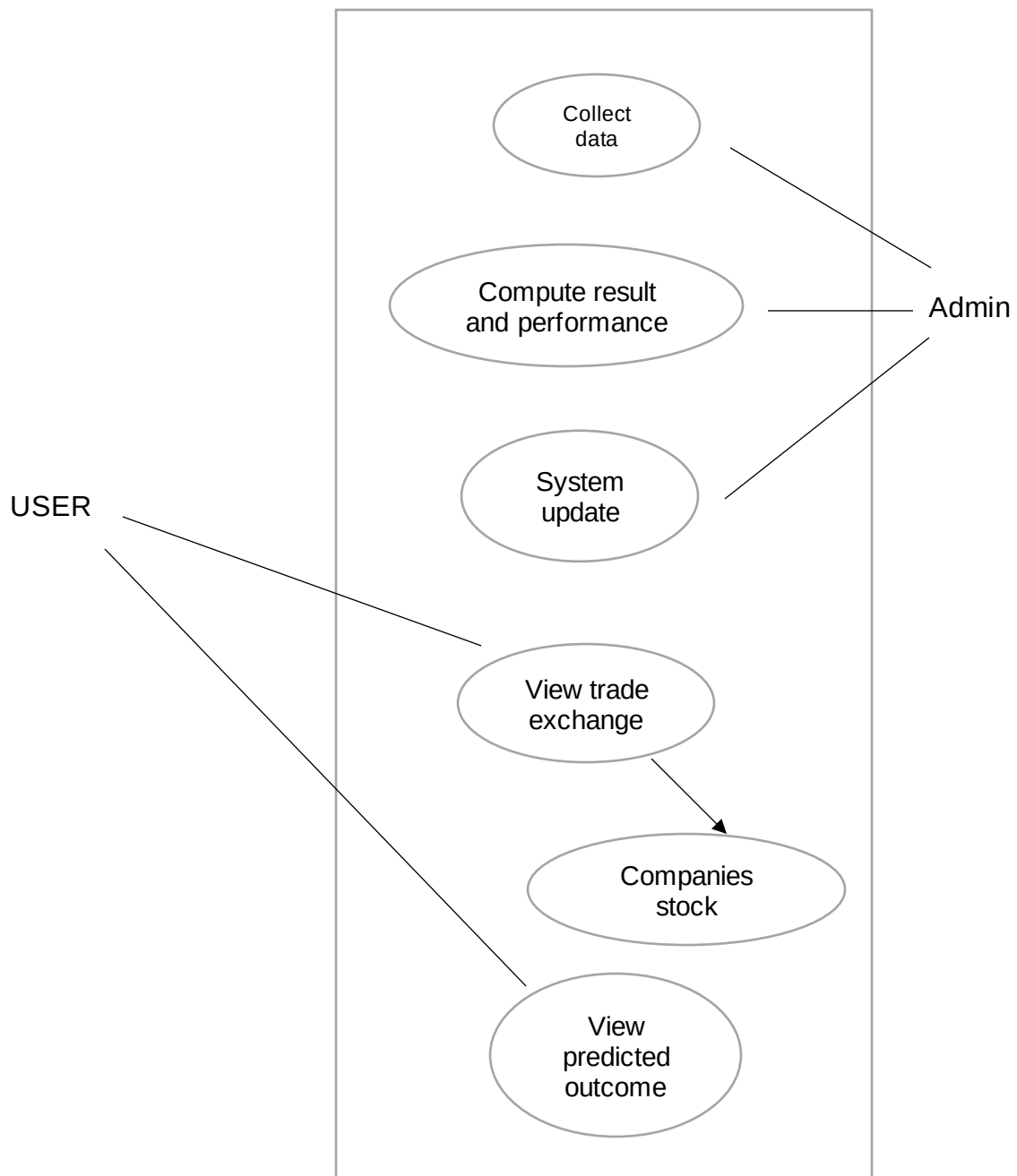


FIG NO:4.2 -USE CASE DIAGRAM

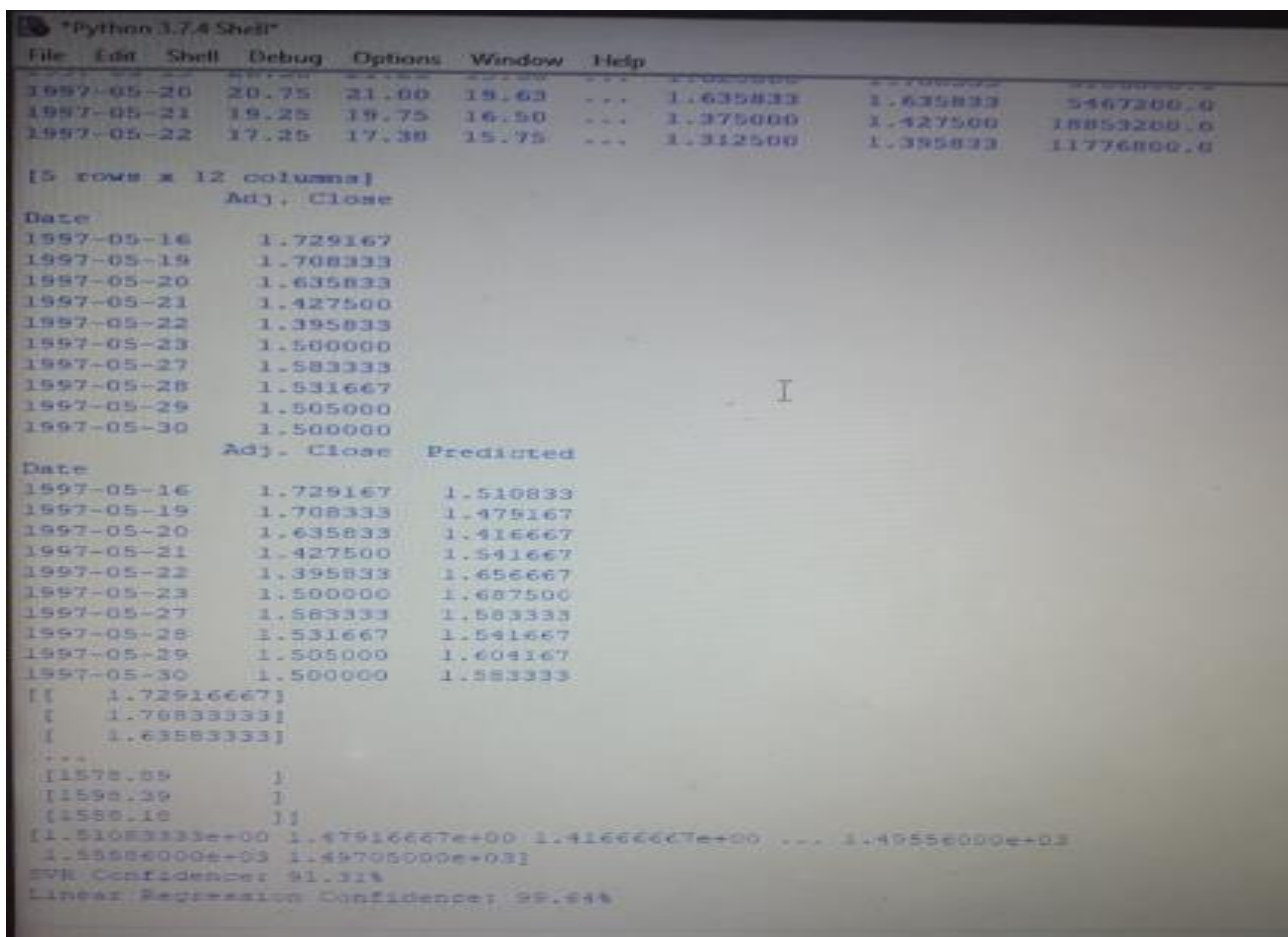
4.3 RESULT AND DISCUSSION

4.3.1 ANALYSIS

In this project, the factors that are taken into account for change in the closing price of a particular company are: General Index, Price difference, highest value, lowest value, share volume and closing price. We performed analysis on obtained data to establish relation between our output parameters and the selected factors.

4.3.2 RESULT

After collection of data the future share price is predicted using support vector machine and linear regression. The value is then compared the next day with the actual value. The results are



The screenshot shows a Python 3.7.4 Shell window with a menu bar (File, Edit, Shell, Debug, Options, Window, Help) and a table of stock data. The table has columns for Date, Adj. Close, and Predicted. The data is for the period from 1997-05-16 to 1997-05-30. Below the table, there is a list of values and a summary of the SVM and Linear Regression confidence scores.

Date	Adj. Close	Predicted
1997-05-16	1.729167	1.510833
1997-05-19	1.708333	1.475167
1997-05-20	1.635833	1.416667
1997-05-21	1.427500	1.541667
1997-05-22	1.395833	1.656667
1997-05-23	1.500000	1.687500
1997-05-27	1.583333	1.583333
1997-05-28	1.531667	1.541667
1997-05-29	1.505000	1.604167
1997-05-30	1.500000	1.583333

[[1.72916667]
[1.70833333]
[1.63583333]
...
[1578.89]
[1598.39]
[1588.18]
[1.51083333e+00 1.47516667e+00 1.41666667e+00 ... 1.40556000e+03
1.35586000e+03 1.49705000e+03]
SVM Confidence: 91.31%
Linear Regression Confidence: 99.64%

FIG NO:4.3-OUTPUT

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 CONCLUSION

The conclusion can be summarized into following aspects: SVM algorithm works on the large dataset value which collected from different global financial markets. Also SVM does not give a problem of over fitting. By measuring the accuracy of the different algorithms, we found that the most suitable algorithm for predicting the market price of a stock based on various data points from the historical data is the support vector regression(SVR) algorithm. The algorithm will be a great asset for brokers and investors for investing money in the stock market since it is trained on a huge collection of historical data and has been chosen after being tested on a sample data. The project demonstrates the machine learning model to predict the stock value with more accuracy as compared to previously implemented machine learning models.

5.2 FUTURE ENHANCEMENT

Future scope of this project will involve adding more parameters and factors like the financial ratios, multiple instances, etc. The more the parameters are taken into account more will be the accuracy. The algorithms can also be applied for analyzing the contents of public comments and thus determine patterns/relationships between the customer and the corporate employee. The use of traditional algorithms and data mining techniques can also help predict the corporations performance structure as a whole.

REFERENCES

1. Ashish Sharma, Dinesh Bhuriya, Upendra Singh. "Survey of Stock Market Prediction Using Machine Learning Approach", ICECA 2017.
2. M. Halls-Moore, "Support vector machines: A guide for beginners," September 2014. [Online]. Available: <http://www.quantstart.com/articles/Support-Vector-Machines-A-Guide-for-Beginners>
3. D. d. I. F. Rafael Rosillo, Javier Giner, "Stock market simulation using support vector machines," Journal of Forecasting, vol. 33, pp. 488–500, July 2014.
4. V. H. Shah, "Machine learning techniques for stock price prediction," pp. 1–18, 2007.
5. K. jae Kim, "Financial time series forecasting using support vector machines," Neurocomputing, vol. 55, 2003.

APPENDIES

SOURCE CODE

```
# Make necessary imports

import quandl
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.svm import SVR
from sklearn.model_selection import train_test_split

# Get Amazon stock data

amazon = quandl.get("WIKI/AMZN")

print(amazon.head())

#Get only the data for the Adjusted Close column

amazon = amazon[['Adj. Close']]

print(amazon.head(10))

#Predict for 10 days; Predicted has the data of Adj. Close shifted up by 30 rows

forecast_len=10
amazon['Predicted'] = amazon[['Adj. Close']].shift(-forecast_len)

print(amazon.head(10))

#Drop the Predicted column, turn it into a NumPy array to create dataset

x=np.array(amazon.drop(['Predicted'],1))

# Remove last 10 rows

x=x[:-forecast_len]

print(x)

# Create dependent dataset for predicted values, remove the last 30 rows

y=np.array(amazon['Predicted'])
```

```

y=y[: -forecast_len]
print(y)

#Split datasets into training and test sets (80% and 20%)
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2)

#Create SVR model and train it
svr_rbf=SVR(kernel='rbf',C=1e3,gamma=0.1)
svr_rbf.fit(x_train,y_train)

# Get score
svr_rbf_confidence=svr_rbf.score(x_test,y_test)
print(f"SVR Confidence: {round(svr_rbf_confidence*100,2)}%")

#Linear Regression
lr=LinearRegression()
lr.fit(x_train,y_train)
lr_confidence=lr.score(x_test,y_test)
print(f"Linear Regression Confidence: {round(lr_confidence*100,2)}%").

```