PROJECT REPORT

on

MACHINE LEARNING IN INTRADAY STOCK TRADING

Submitted in partial fulfillment of the

requirement for the award of the degree of

BACHELORS OF TECHNOLOGY

(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

Under The Supervision of
Mr. Hradesh Kumar
Assistant Professor

Submitted By

Manoranjan Kumar Thakur (18021180026)
Kashish Kumar (18021011255)

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING / DEPARTMENT OF COMPUTERAPPLICATION
GALGOTIAS UNIVERSITY, GREATER NOIDA
INDIA
DECEMBER, 2021

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA

CANDIDATE'S DECLARATION

We hereby certify that the work which is being presented in the project, entitled **"INTRA DAY STOCK PREDICTION"** in partial fulfillment of the requirements for the award of the **BACHELORS OF TECHNOLOGY** submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the period of month, July 2021 to December 2021, under the supervision of Mr. Hradesh Kumar (Assistant Professor), Department of Computer Science and Engineering, Galgotias University, Greater Noida

The matter presented in the project has not been submitted by us for the award of any other degree of this or any other places.

Manoranjan Kumar Thakur, 18021180026

Kashish Kumar, 18021011255

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Mr. Hradesh Kumar

Assistant Professor

CERTIFICATE

The Final Project Viva-Voce examination of Manoranjan Kumar Thakur (18021180026) and Kashish Kumar (18021011255) has been held on __________________ and his/her work is recommended for the award of Bachelors of Technology.

Signature of Examiner(s) **Signature of Supervisor(s)**

Signature of Project Coordinator **Signature of Dean**

Date: December, 2021

Place: Greater Noida

ACKNOWLEDGEMENT

We would like to express our special thanks of gratitude to our guide Mr. Hradesh Kumar, as well as our esteemed university who gave us the golden opportunity to do this wonderful project on the topic Classifying tweets based on climate change stance which also helped us in doing a lot of Research and we came to know about so many new things. We are really thankful to them.

Secondly, we would also like to thank our parents and friends who helped us a lot in finalizing this project within the limited time frame.

Manoranjan Kumar Thakur (18021180026)

Kashish Kumar (18021011255)

Abstract

Without a doubt, stock markets are an important and necessary aspect of any country's economy. However, the impact of stock markets on a country's economy may differ from the impact of stock markets on other countries' economies. This is because the impact of stock markets on the economy is influenced by a variety of factors such as the organization of stock exchanges, their link with other financial system components, the country's governance system, and so on. Because each of these characteristics is unique to each country, the impact of stock markets on a country's economy is likewise unique. The Indian capital market system has experienced considerable fundamental institutional changes over the years, resulting in lower transaction costs, more efficiency, transparency, and safety. All of these changes have resulted in the economy's development through stock markets. Similarly, a significant need for stock market development is projected as a result of economic expansion powered by technology advances, product and service innovation. The attempt to anticipate the future value of a stock traded on a stock exchange is known as stock market prediction. This research focuses on the classification challenge, as well as predicting next-second price movements and acting on the insights gained from our algorithms. We used a variety of

machine learning algorithms, including logistic regression as a baseline model with and without regularization, support vector machines (SVM) with Linear, Polynomial (degree 3), Sigmoid, and Radial Basis Function kernels, and support vector machines (SVM) with Linear, Polynomial (degree 3), Sigmoid, and Radial Basis Function kernels. To determine the trading activity in the following minute, the variables are tuned by altering the cost of constraint violation as well as the constant of regularization term in the Lagrange formulation. The support vector machine with polynomial kernel performs the best among all of our models when using the projected results from our models to build the portfolio value over time.

TABLE OF CONTENTS

Title		No.
Abstract		
Chapter 1	**Introduction**	1
Chapter 2	**Literature Survey**	3
	2.1 System Review	3
	2.2 Technology Used	3
Chapter 3	**System Design**	5
	3.1 Activity Diagram	5
	3.2 Use Case Diagram	6
Chapter 4	**Exploratory Data Analysis**	7
Chapter 5	**Predictive Analysis**	12
Chapter 6	**Result and discussion**	15
Chapter 7	**Future Work**	17
Chapter 8	**References**	18

CHAPTER – 01

INTRODUCTION

Intraday traders buy and sell financial instruments for a brief period of time, usually inside one trading day. Stocks are one of the most well-known financial tools. However, because there are hundreds of stocks listed on the stock exchange, selecting the most tradeable equities on each trading day is a difficult operation, which is usually accomplished by manually inspecting past stock prices and technical indicators. The ability to accurately estimate stock price movement is crucial to trading profitability. Many investors actively traded stocks in the hopes of outperforming the market, which is known as a passive investment. The prediction of price movement in the financial market using machine learning has become a topic of interest for both investors and researchers as the availability of financial data has increased. The models' insights into price swings could aid investors in making more informed judgments. The goal of this research is to anticipate short-term price fluctuations using stock price timeseries data, commonly used technical-analysis indicators, and trading volume. Short-term trading methods based on these predictions will be developed to profit from tiny price swings in highly liquid equities.

CHAPTER – 02

LITERATURE SURVEY

2.1 SYSTEM REVIEW

This survey is being conducted to better understand the needs and requirements of the general public, and in order to do so, we combed through several websites and applications for the necessary information. We created an audit based on this data, which allowed us generate fresh ideas and build alternate arrangements for our assignment. We came to the conclusion that such an application is required and that there has been some advancement in this field.

2.2 TECHNOLOGY USED

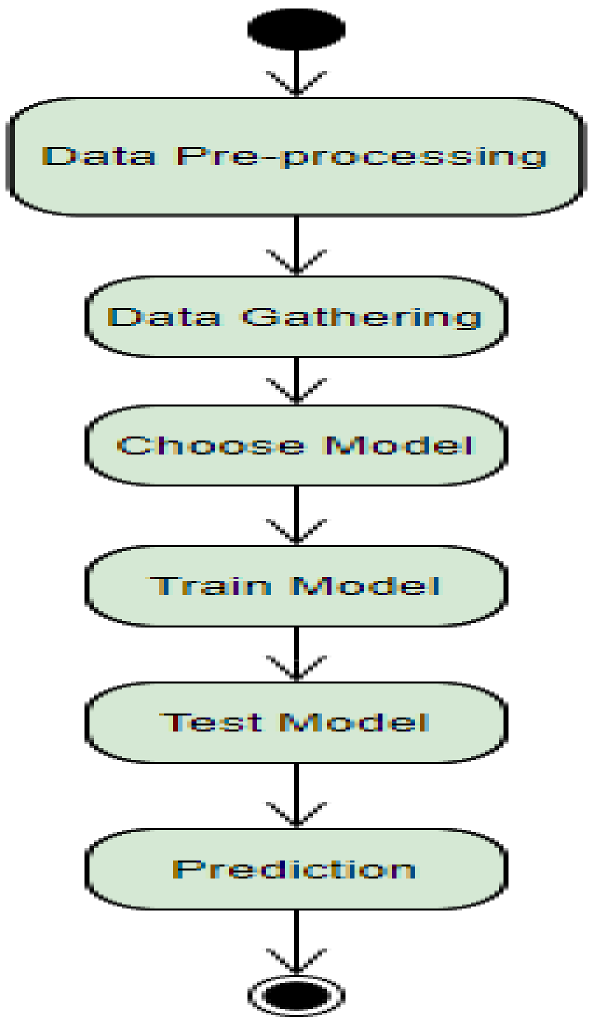
- **PYTHON -** Python is a high-level, interpreted programming language that may be used for a variety of tasks. Python's design concept promotes readability of code, especially shown by its frequent use of spacing. Its language elements and object-oriented approach are designed to assist developers in writing clear, logical code for both major and minor projects. Python is dynamically coded and enables procedural, object-oriented, and structured programming techniques.

- **MACHINE LEARNING -** Machine learning is the scientific study of algorithms and statistical models that computer systems use to successfully complete the given task without requiring detailed instructions and instead relies on patterns and inferences. Artificial intelligence is regarded as a subset of it. In order to make decisions or judgments without even being explicitly taught, machine learning techniques create a computational formula based on sample data, known as "training data."
- **JUPYTER LAB -** Project Jupyter is a non-profit organization dedicated to the creation of open-source software, open-standards, and services for interactive computing in a variety of programming languages.
- **PANDAS -** Pandas is a data manipulation and analysis software library for the Python programming language. It includes data structures and methods for manipulating numerical tables and time series in particular.

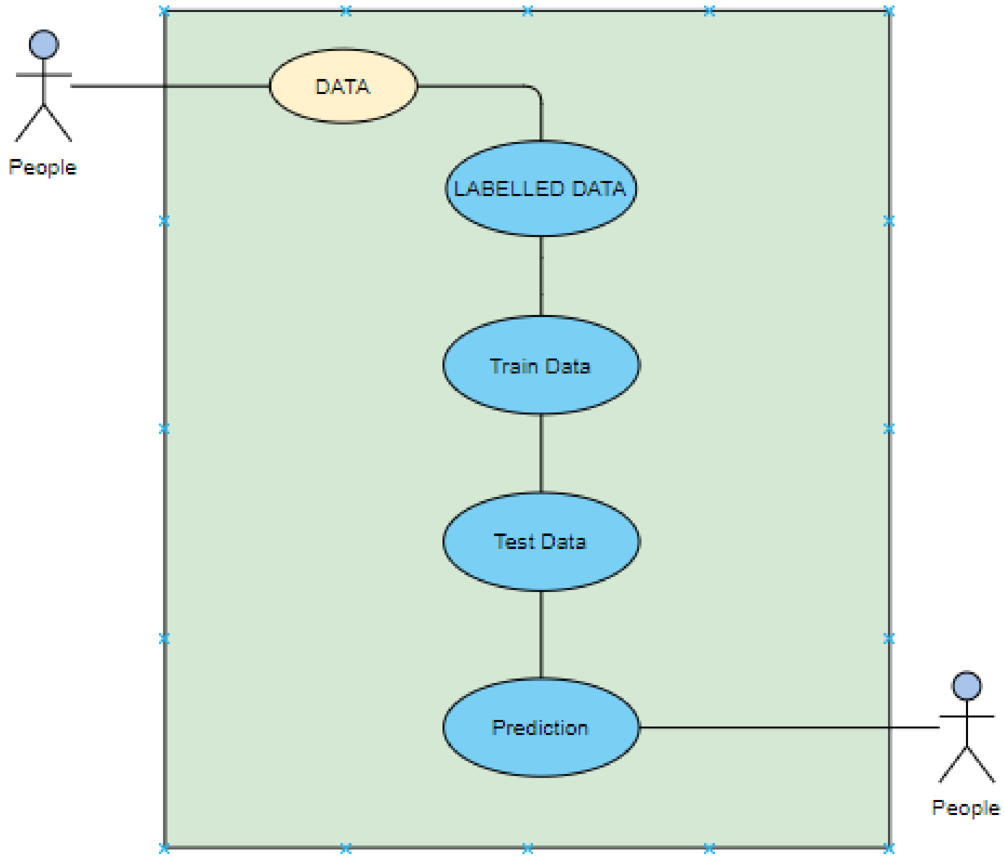
CHAPTER – 03

SYSTEM DESIGN

3.1 ACTIVITY DIAGRAM



3.1 USECASE DIAGRAM



CHAPTER – 04

EXPLORATORY DATA ANALYSIS

Historical Dataset with 1 day time interval

Downloaded the dataset of Microsoft with 1 day interval by using the yfinance api.

Samples: 8861 Start Date: 1986-03-13

End Date: 2021-04-29

Date Open High Low Close Adj Close Volume

1986-03-13, 0.088542, 0.101563, 0.088542, 0.097222, 0.061751, 1031788800

1986-03-14, 0.097222, 0.102431, 0.097222, 0.100694, 0.063956, 308160000

1986-03-17, 0.100694, 0.103299, 0.100694, 0.102431, 0.065059, 133171200

1986-03-18, 0.102431, 0.103299, 0.098958, 0.099826, 0.063405, 67766400

1986-03-19, 0.099826, 0.100694, 0.097222, 0.098090, 0.062302, 47894400

CODE:

```
data = yfinance.download("MSFT", interval='1d')
data.head()
data.tail()
print(len(data))
```

Historical Dataset with 1 min time interval

Downloaded the dataset of Microsoft and NFTY with 1 min interval by using the yfinance api and online dataset platforms.

NFTY DATASET

NFTY Samples: 22806

Start Date: 2021-01-01

Start Time: 09:16

End Date: 2021-03-31

End Time: 15:31

This Dataset was used for classification Change amount was calculated using python method.

CODE:

```
Code: dataframe["change"] = dataframe[["open",
"close"]].pct_change(axis=1)["close"]
```

pct_change

After calculating change. We used the newly added feature(change) to classify the data into profit or loss.

CODE:

```
clas = []

for col in dataframe["change"]:
    if col <= 0:
        clas.append("-1")
    elif col > 0:
        clas.append("1")
    else:
        clas.append("NA")

dataframe["classification"] = clas
```

After adding classification column, the change column was removed. Used Min-Max scaler to scale the data.

Microsoft Dataset 1 min interval

Downloaded the dataset of Microsoft with 1 minute interval by using the yfinance api.

CODE:

```
data = yfinance.download("MSFT", interval='1m',
period='7d')
data.head()
data.tail()
print(len(data))
```

samples: 2726

Start Date: 2021-04-23

Start Time: 09:30:00-4:00

End Date: 2021-04-30

End Time: 15:55:00-04:00

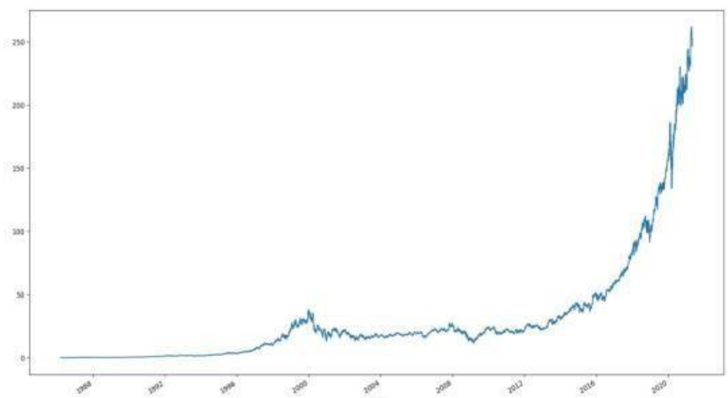
This Dataset was used to predict the stock price of Microsoft for the next 1day.

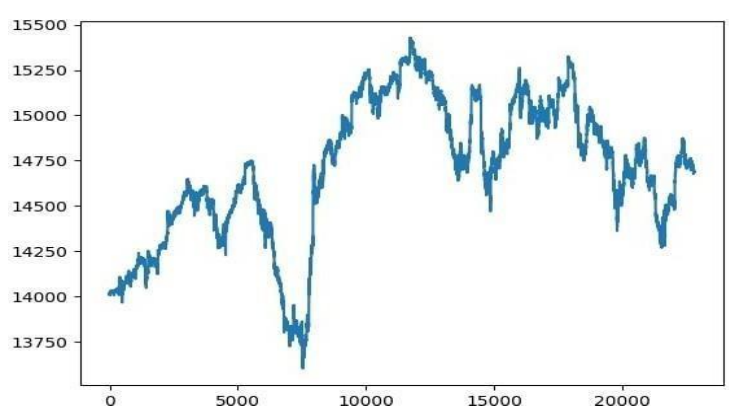
CODE:

```
train = pd.read_csv("1mindata/MSFT1Min.csv")
test = pd.read_csv("1mindata/1MSFT-Test.csv")
len(train)
len(test)

train = train.sort_values('Datetime')
test = test.sort_values("Datetime")

train.reset_index(inplace=True)
train.set_index("Datetime", inplace=True)

test.reset_index(inplace=True)
test.set_index("Datetime", inplace=True)
"""
```

```
plt.figure(figsize=(12, 6))
plt.plot(train["Adj Close"])
plt.xlabel('Date', fontsize=15)
plt.ylabel('Adjusted Close Price', fontsize=15)
plt.show()


# Rolling mean
close_px = train['Adj Close']
mavg = close_px.rolling(window=100).mean()

plt.figure(figsize=(12, 6))
close_px.plot(label='MSFT')
mavg.plot(label='mavg')
plt.xlabel('Date')
plt.ylabel('Price')
plt.legend()
"""

dates_df = train.copy()
dates_df = dates_df.reset_index()
org_dates = dates_df['Datetime']
```

```
# convert to ints
dates_df['Datetime'] =
dates_df['Datetime'].map(mdates.date2num)

dates_df.tail()

dates = dates_df['Datetime'].to_numpy()
prices = train['Adj Close'].to_numpy()

test_df = test.copy()
test_df = test_df.reset_index()
org_dates_test = test_df['Datetime']
test_df['Datetime'] =
test_df['Datetime'].map(mdates.date2num)
test_dates = test_df['Datetime'].to_numpy()
test_prices = test['Adj Close'].to_numpy()

test_dates = np.reshape(test_dates, (len(test_dates),
1))
test_prices = np.reshape(test_prices,
(len(test_prices), 1))
```





CHAPTER – 05

PREDICTIVE ANALYSIS

Model we built for stock price prediction are as follows:

1. Baseline(logistic)
2. SVM linear
3. SVM poly
4. SVM RBF
5. SVM Sigmoid

LOGISTIC REGRESSION:

Logistic regression is a statistical model that uses a logistic function to represent a binary dependent variable in its most basic form, though there are many more advanced variants. Logistic regression (or logit regression) is a technique for estimating the parameters of a logistic model in regression analysis (a form of binary regression).

SVM: SVMs are a class of supervised learning methods for classification, regression, and outlier detection.

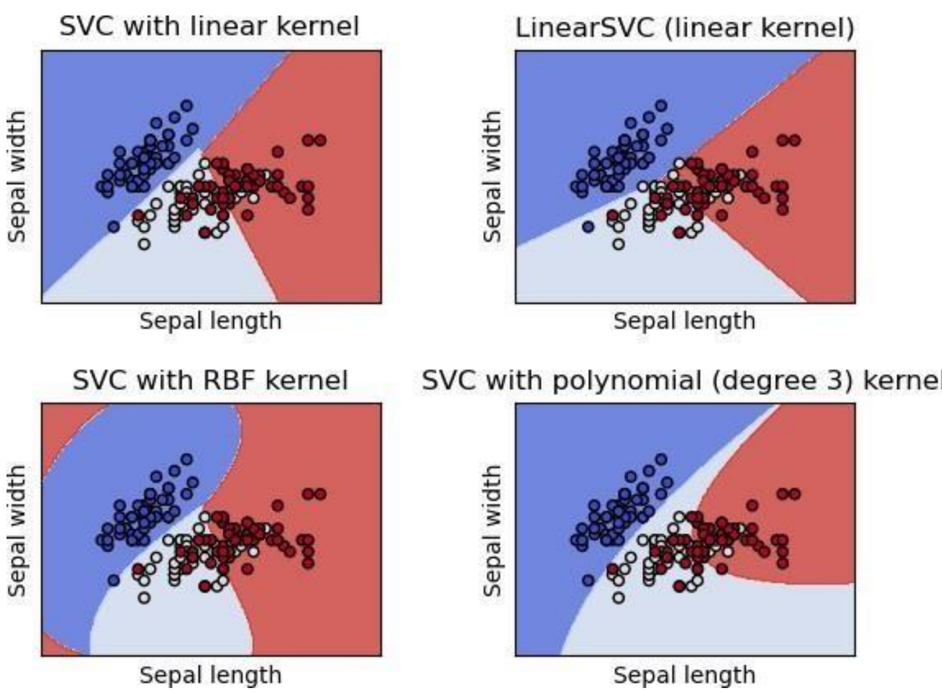
The following are some of the benefits of support vector machines:

• Works well in three-dimensional spaces.

• Even when the number of dimensions exceeds the number of samples, the method is still effective.

• It is memory efficient because it uses a subset of training points (called support vectors) in the decision function.

• Versatile: the decision function can use a variety of Kernel functions. Common kernels are included, however custom kernels can also be specified.

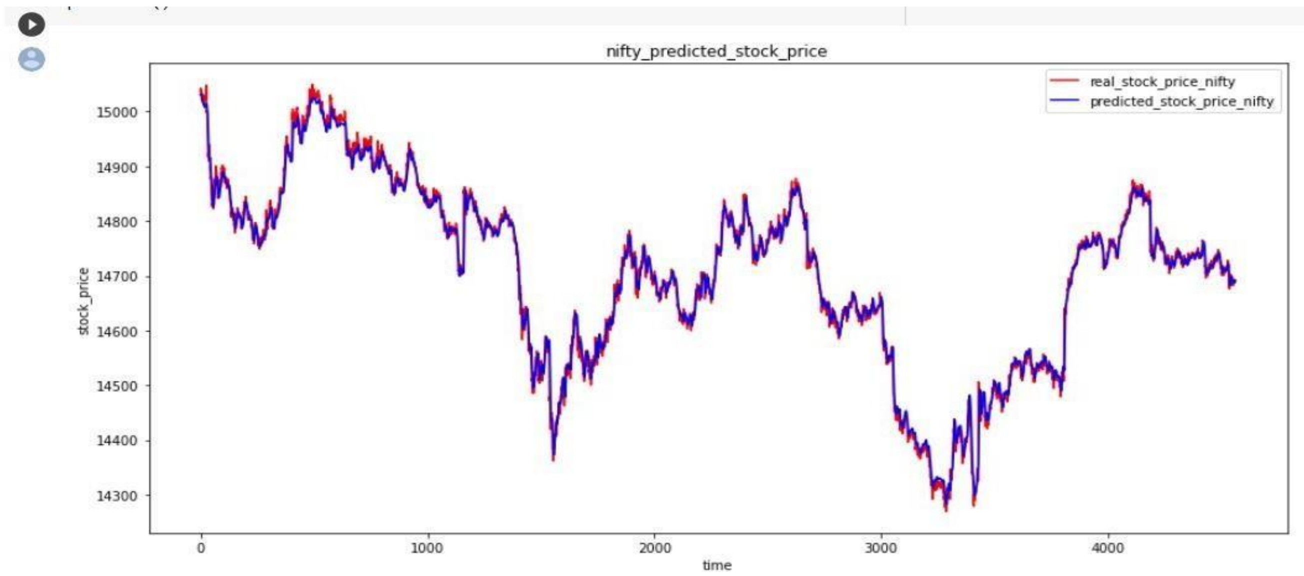


The following are some of the downsides of support vector machines:

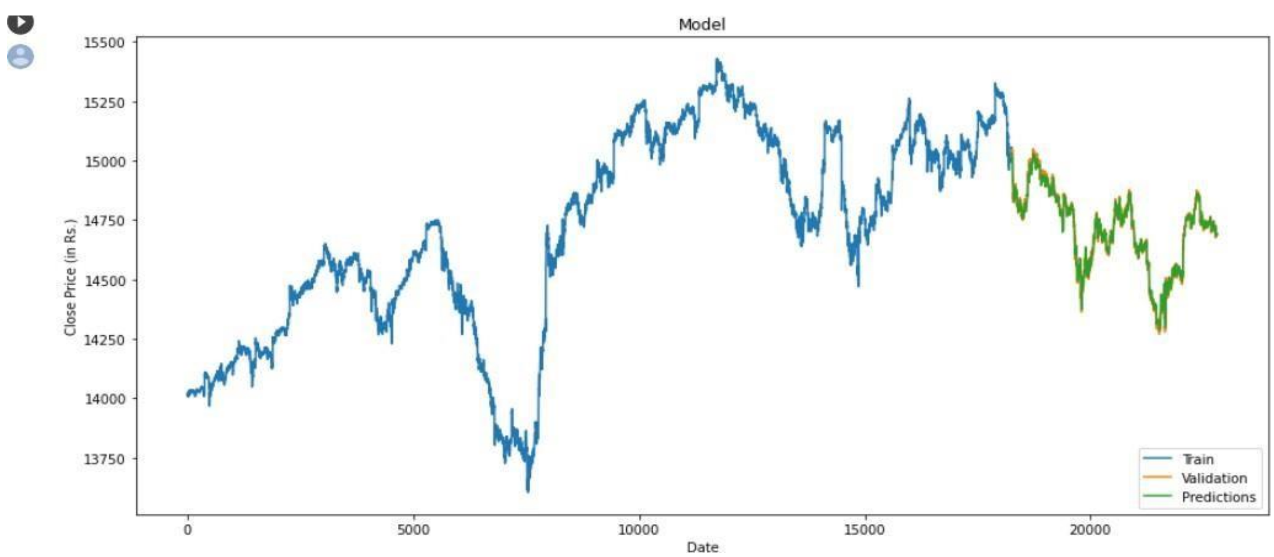
- Avoid over-fitting when picking Kernel functions and regularization terms if the number of features is substantially more than the number of samples.
- Probability estimates are produced using a costly five-fold cross-validation method rather than directly by SVMs (see Scores and probabilities, below).

CHAPTER – 06

RESULT AND DISCUSSION



PLOT 1: FOR TEST DATASET



PLOT 2: FOR WHOLE DATASET

Accuracy level	Model name
Highest (55 above)	SVM (poly)
Midrange (50-55)	SVM (linear), SVM (RBF), SVM (Sigmoid)
Lowest (>50%)	Logistic

We gave a number of prediction methods that may simply demonstrate the reasons for making a certain forecast about a prospective stock prediction, which is a valuable skill in the field of stock analysis. We compared its results to the predictions given by several models to see how accurate it was at forecasting. When we compare the outcomes of multiple models for the intra-day stock prediction problem, we can see that our method outperforms them, demonstrating that a more transparent model for a problem like market prediction is conceivable. Even though our results are comparable to or better than those of past studies, we believe they can be improved. Classifying profit and loss, as indicated by numerous studies, was one of the techniques to increase performance; this resulted in better performances but fell short of the goal of this project.

CHAPTER – 7

FUTURE SCOPE

For future work, we can use this classification process to obtain a loss risk class as a first step in improving the performance of our model, and we could try to solve stock prediction using deep learning methods to continue comparing our model to even more sophisticated methods, but this would require a larger dataset. We also intend to use regression techniques in the dataset to forecast the size of profit and loss rather than the class (Profit or Loss). There are two well-known prediction difficulties in this field.

CHAPTER – 8

REFERENCES

[1] M. D. Godfrey, C. W. Granger, and O. Morgenstern, “The random-walk hypothesis of stock market behaviors,” Kyklos, vol. 17, no. 1, pp. 1–30,1964.

[2] J. Murphy, “Technical analysis of the financial markets, prentice hall, London,” 1998.

[3] G. Bonde and R. Khaled, “Stock price prediction using genetic algorithms and evolution strategies.”

[4] X. Wang, P. K. H. Phua, and W. Lin, “Stock market prediction using neural networks: does trading volume help in short-term prediction?”in Neural Networks, 2003. Proceedings of the International Joint Conference on, vol. 4. IEEE, 2003, pp. 2438–2442.

[5] A. U. Khan et al., “Stock rate prediction using back propagation algorithm: Analyzing the prediction accuracy with different number of hidden layers,” Glow gift, Bhopal, 2005.

[6] A. U. Khan, T. Bandopadhyaya, and S. Sharma, “Comparisons of stock rates prediction accuracy using different technical indicators with backpropagation neural network and genetic algorithm based backpropagation neural network,” in Emerging Trends in Engineering and Technology, 2008. ICETET’08. First International Conference on.IEEE, 2008, pp. 575–580.

[7] P.-F. Pai and C.-S. Lin, “A hybrid arima and support vector machines model in

stock price forecasting," Omega, vol. 33, no. 6, pp. 497–505,2005.

[8] A. Khan, B. Baharudin, and K. Khan, "Sentiment classification using sentence-level lexical based semantic orientation of online reviews," vol,vol. 6, pp. 1141–1157, 2011.

[9] S. Deng, T. Mitsubuchi, K. Shioda, T. Shimada, and A. Sakurai,"Combining technical analysis with sentiment analysis for stock price prediction," in Dependable, Autonomic and Secure Computing (DASC),2011 IEEE Ninth International Conference on. IEEE, 2011, pp. 800–807.

[10] S. Baccianella, A. Esuli, and F. Sebastiani, "Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining," in Proceedings of the 7th conference on International Language Resources and Evaluation (LREC10), Valletta, Malta, May, 2010.

[11] J. J. Mor´e, "The levenberg-marquardt algorithm: implementation and theory," in Numerical analysis. Springer, 1978, pp. 105–116.

[12] K. Toutanova, D. Klein, C. D. Manning, and Y. Singer, "Feature-rich part-of-speech tagging with a cyclic dependency network," in Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume

1. Association for Computational Linguistics,2003, pp. 173–180