Stock Market Price Trend Prediction in Indian Stock Exchange

B.Tech CSE - IT, 8th Semester, 2023-24

Mentor

Prof. Sudipta Dutta

Assistant Professor
Dept. of CSE & IT
Bengal Institute of Technology, Kolkata

Group Members

Name	MAKAUT Roll No.	Stream
Subham De	12100120092	CSE
Shahobaz Ahmed	12100120063	CSE
Arnab Chatterjee	12100120060	CSE
Subhasish Das	12100220015	IT

BENGAL INSTITUTE OF TECHNOLOGY

No. 1, Basanti Highway, Brahmapur Government Colony, Bagdoba,
Kolkata, 700150

Acknowledgement

This project would not been a successful without the sincere cooperation and guidance of our mentor Prof. Sudipta Dutta, Assistant Professor, Dept of CSE & IT, Bengal Institute of Technology, Kolkata, who has provided us with useful resources and motivation through the various phases of this project which has prove to be crucial for the completion of this documentation.

Signature of Group Members :

Name :
MAKAUT Roll No. :

Name :
MAKAUT Roll No. :

Name :
MAKAUT Roll No. :

Name :
MAKAUT Roll No. :

Signature of Mentor

Prof. Sudipta Dutta
Assistant Professor
Dept. of CSE & IT
Bengal Institute of Technology

Signature of HOD

Dr. Shanta Phani
Head of the Department
Dept. of CSE & IT
Bengal Institute of Technology

TABLE OF CONTENTS

ABSTRACT

In the past decades, there is an increasing interest in predicting markets among economists, policymakers, academics and market makers. The objective of the proposed work is to study and improve the supervised learning algorithms to predict the stock price. Stock Market Analysis of stocks using data mining will be useful for new investors to invest in stock market based on the various factors considered by the software. Stock market includes daily activities like Sensex calculation, exchange of shares. The exchange provides an efficient and transparent market for trading in equity, debt instruments and derivatives. Our aim is to create software that analyses previous stock data of certain companies, with help of certain parameters like delivery percentage, open & close value, market capitalization, total turnover, candlestick chart pattern, price to book ratio etc. which affect stock value. We are going to implement these values in data mining algorithms and we will be able to decide which algorithm gives the best result. This will also help us to determine the values that particular stock will have in near future. We will determine the best stocks that has grown in the stipulated time period. We will collect the data of the stock market from NSE and based on the data we will determine the list of stocks that has given the best value and we will also predict which stock going to be best based on the parameters in near future.

CHAPTER 1
INTRODUCTION

1.1 OVERVIEW

In recent times stock market predictions is gaining more attention, maybe due to the fact that if the trend of the market is successfully predicted the investors may be better guided. The profits gained by investing and trading in the stock market greatly depends on the predictability. If there is a system that can consistently predict the direction of the dynamic stock market will enable the users of the system to make informed decisions. More over the predicted trends of the market will help the regulators of the market in taking corrective measures.

1.2 AIM AND OBJECTIVE

The aim of the project is to examine a number stocks to predict future stock returns based on past returns and the pattern of the candlestick chart to construct a portfolio of multiple stocks in order to diversify the risk. We do this by sorting and analyzing the stocks for a given period of time to give the stock price forecasting by interpreting the seemingly chaotic market data.

1.3 STOCK MARKET [1]

A stock market, equity market or share market is the aggregation of buyers and sellers (a loose network of economic transactions, not a physical facility or discrete entity) of stocks (also called shares), which represent ownership claims on businesses; these may include securities listed on a public stock exchange as well as those only traded privately. Examples of the latter include shares of private companies which are sold to investors through equity crowd funding platforms. Stock exchanges list shares of common equity as well as other security types, e.g. corporate bonds and convertible bonds. Stock price prediction is one of the most widely studied problem, attracting researchers from many fields. The volatile nature of the stock market makes it really difficult to apply simple time-series or regression techniques. Financial institutions and active traders have created various proprietary models to beat the market for 2 themselves or their clients, but rarely did anyone achieve consistently higher than the average returns on investment. The challenge of stock market price forecasting is so appealing because an improvement of just a few points of percentage can increase the profit by millions of dollars. This paper discusses the application of Support Vector Machines and Linear Regression in detail along with the pros and cons of the given methods. The paper introduces the parameters and variables which can be used to recognize the patterns in stock prices which can be helpful in future stock prediction.

1.4 MOTIVATION

Stock price prediction is a classic and important problem. With a successful model for stock prediction, we can gain insight about market behavior over time, spotting trends that would otherwise not have been noticed. With the increasingly computational power of the computer, we have tried to find the possible stock market analysing pattern. Thus, our motivation is to design a public service incorporating historical data and users predictions to make a stronger model that will benefit everyone.

CHAPTER 2

LITERATURE SURVEY

1. **Survey of stock market prediction using machine learning approach**
   Authors: Ashish Sharma ; Dinesh Bhuriya ; Upendra Singh
   2017 International conference of Electronics, Communication and Aerospace Technology [2]

   Stock market is basically nonlinear in nature and the research on stock market is one of the most important issues in recent years. People invest in stock market based on some prediction. For predict, the stock market prices people search such methods and tools which will increase their profits, while minimize their risks. Prediction plays a very important role in stock market business which is very complicated and challenging process. Employing traditional methods like fundamental and technical analysis may not ensure the reliability of the prediction. To make predictions regression analysis is used mostly. In this paper we survey of well-known efficient regression approach to predict the stock market price from stock market data based. In future the results of multiple regression approach could be improved using more number of variables.

2. **NSE Stock Market Prediction Using Deep-Learning Models**
   Authors : Hiransha M ; Gopalakrishnan E.A ; Vijay Krishna Menon ; Soman K.P
   Centre for Computational Engineering and Networking, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore-641112, India. [3]

   In this work they used four DL architecturesfor the stock price prediction of NSE and NYSE ,which are two different leading stock markets in the world. Here we trained four networks MLP, RNN, LSTM and CNN with the stock price of TATA MOTORS from NSE. The models obtained were used for predicting the stock price of MARUTI, HCL and AXIS BANK from NSE stock market and also for predicting the stock price of BANK OF AMERICA (BAC) and CHESAPEAK ENERGY (CHK) from NYSE. From the result obtained, it is clear that the models are capable of identifying the patterns existing in both the stock markets. This shows that there exist an underlying dynamics, common to both the stock markets. Linear models like ARIMA is a univariate time series prediction and hence they are not capable of identifying underlying dynamics within various time series. From the result, we can conclude that DL models are outperforming ARIMA model. In the proposed work , CNN has performed better than other three networks as it is capable of capturing the abrupt changes in the system since a particular window is used for predicting the next instant.

3. **Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications**
   Authors : Gaurang Sonkavde; Deepak Sudhakar Dharrao; Anupkumar M. Bongale; Sarika Doreswamy [4]

   In this review, several conventional, machine learning, and deep learning techniques that are employed in stock market forecasting are investigated. This review describes various machine learning techniques, deep learning techniques, and time series forecasting techniques. This article presents recent applications of machine learning and deep learning models, and an ensemble model is also tested on the TAINIWALCHM and AGROPHOS stock datasets. Despite the existence of several popular methods for stock price forecasting, even today, there is no universal solution to accurately predict the stock price or trend of the market.

CHAPTER 3
PROPOSED METHODOLOGY

1. **Data Collection :-** The initial step in our methodology involves collecting historical stock market data. This data will be sourced from the NSE [10] database which includes the following features –
    - Opening & Closing Value
    - High & Low Price
    - Total Turnover
    - Delivery Percentage
    - Total Count
    - Date
    - Series
    - Symbol

2. **Data Processing :-** Data preprocessing is critical for ensuring the quality and reliability of the predictions. The steps involved are –
    - **Data Cleaning** : Removing any null values or outliers that could skew the analysis.
    - **Feature Engineering** : Creating new features from the existing data, such as moving averages, relative strength index (RSI), and other technical indicators.
    - **Normalization** : Scaling the features to ensure they have a standard range, which helps in improving the performance of machine learning algorithms.
    - **Data Splitting** : Dividing the dataset into training, validation, and test sets to evaluate the performance of the models.

3. **Exception Handling :-** In Python, the try-except block is used to handle exceptions (errors) that might occur during the execution of a block of code. The try block contains the code that may potentially cause an exception. The except block contains the code that will run if an exception occurs. The purpose of this exception handling block is to ensure that if an error occurs during data processing for a particular date, the program :
    - Catches the error without crashing.
    - Provides a meaningful error message for debugging purposes.
    - Returns a None value to indicate that the data processing was unsuccessful for that specific date.

4. **Libraries & Functions :-** Several python libraries have been used in this project like numpy and pandas where By using pandas for the majority of the operations, the function takes advantage of its robust and efficient data manipulation capabilities, while NumPy provides optimized numerical operations when needed. To fetch the market data nselib and to fetch the date formatting datetime libraries are also used.

- **Numpy :** ‘numPy’ is highly efficient for numerical operations and array manipulations. Initially, np.argsort was used to sort indices, which is a common use case for NumPy due to its optimized performance with numerical data.

- **Pandas :** ‘pandas’ is a powerful library for data manipulation and analysis, providing high-level data structures and tools designed to make data analysis fast and easy.

- **Nselib :** ‘nselib’ appears to be a custom or third-party library in Python, specifically designed for interacting with the National Stock Exchange (NSE) of India. Libraries like nselib typically provide functions to fetch and process stock market data, such as daily market summaries, historical prices, and other financial metrics.
- **Datetime :** The ‘datetime’ library in Python is a module within the standard library that provides classes for manipulating dates and times. It allows for the creation, manipulation, and formatting of date and time objects, making it a versatile tool for a wide range of date and time-related tasks.

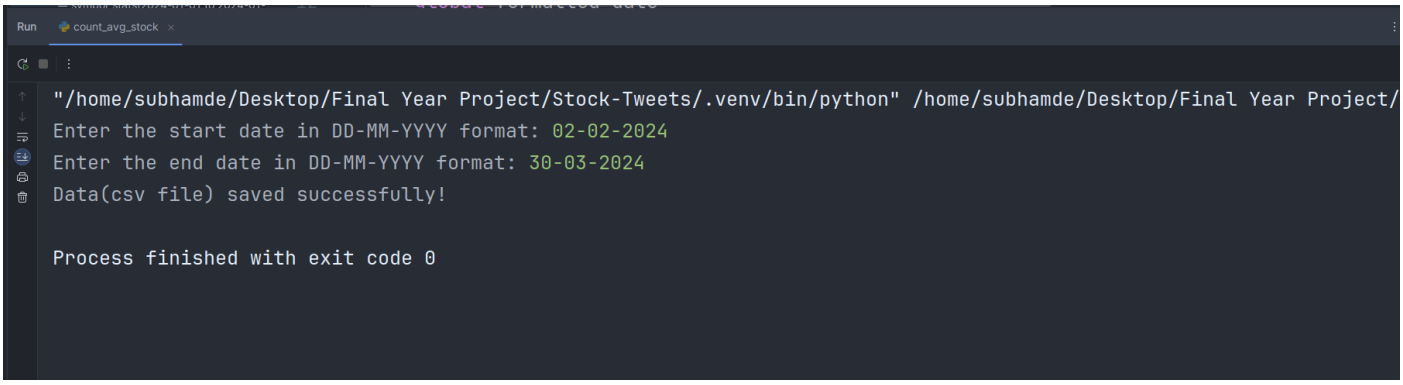
Explanation of the functions :

- **“is_weekend(date_input)” :** To check if a given date falls on a weekend (Saturday or Sunday).
- **“fetch_and_sort_data(date_input)” :** To fetch, filter, sort, and return market data for a specified date.
  - **Formatting Date :** Converts the date_input into a formatted string using strftime.
  - **Fetching Data :** Calls capital_market.bhav_copy_with_delivery() to get market data for the specified date.
  - **Data Processing :**
    - Converts the 'DELIV_PER' column to numeric using pd.to_numeric.
    - Applies filtering conditions to the DataFrame to include only equity series with turnover greater than 100 and delivery percentage greater than 85%, excluding certain types of symbols.
    - Sorts the filtered DataFrame by delivery percentage in descending order and selects the top 15 rows.
  - **Returns :** Returns a DataFrame containing the symbols and their delivery percentages, or None if there's an error.
- **“process_data_range(start_date, end_date)” :** To process and aggregate market data over a range of dates, and save the results to a CSV file.
  - **Date Conversion :** Converts the start_date and end_date strings into datetime objects.
  - **Date Validation :** Checks that start_date is before end_date; raises a ValueError if not.
  - **Data Processing :**
    - Iterates over each date in the specified range, skipping weekends.
    - Calls fetch_and_sort_data() for each valid date to fetch and process market data.
    - Aggregates data into dictionaries (count_dict and sum_dict) to count occurrences and sum delivery percentages for each symbol.
    - Creates a DataFrame (result_df) to store the total count and average delivery percentage for each symbol.
    - Filters out symbols with a total count less than or equal to 5 and sorts the remaining symbols by average delivery percentage in descending order.
  - **CSV Output :** Saves the resulting DataFrame to a CSV file named based on the date range.

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

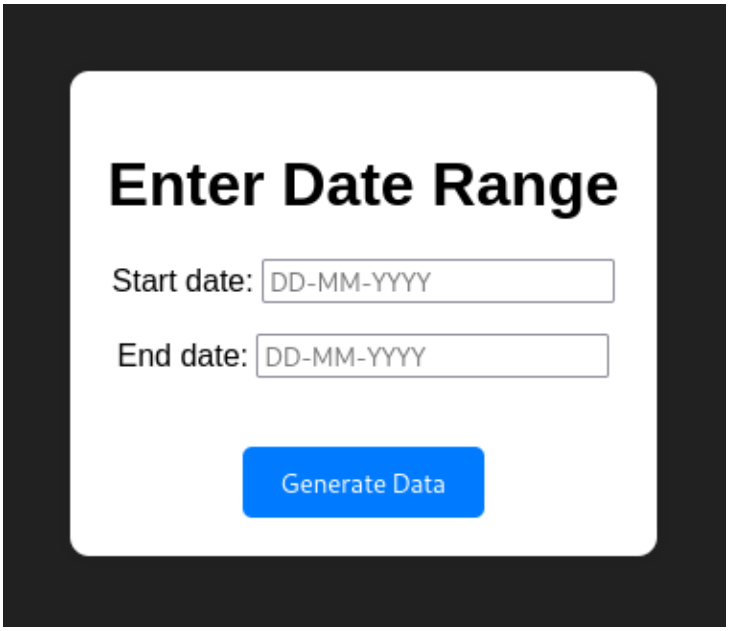
Screenshots of the Output in the terminal :-

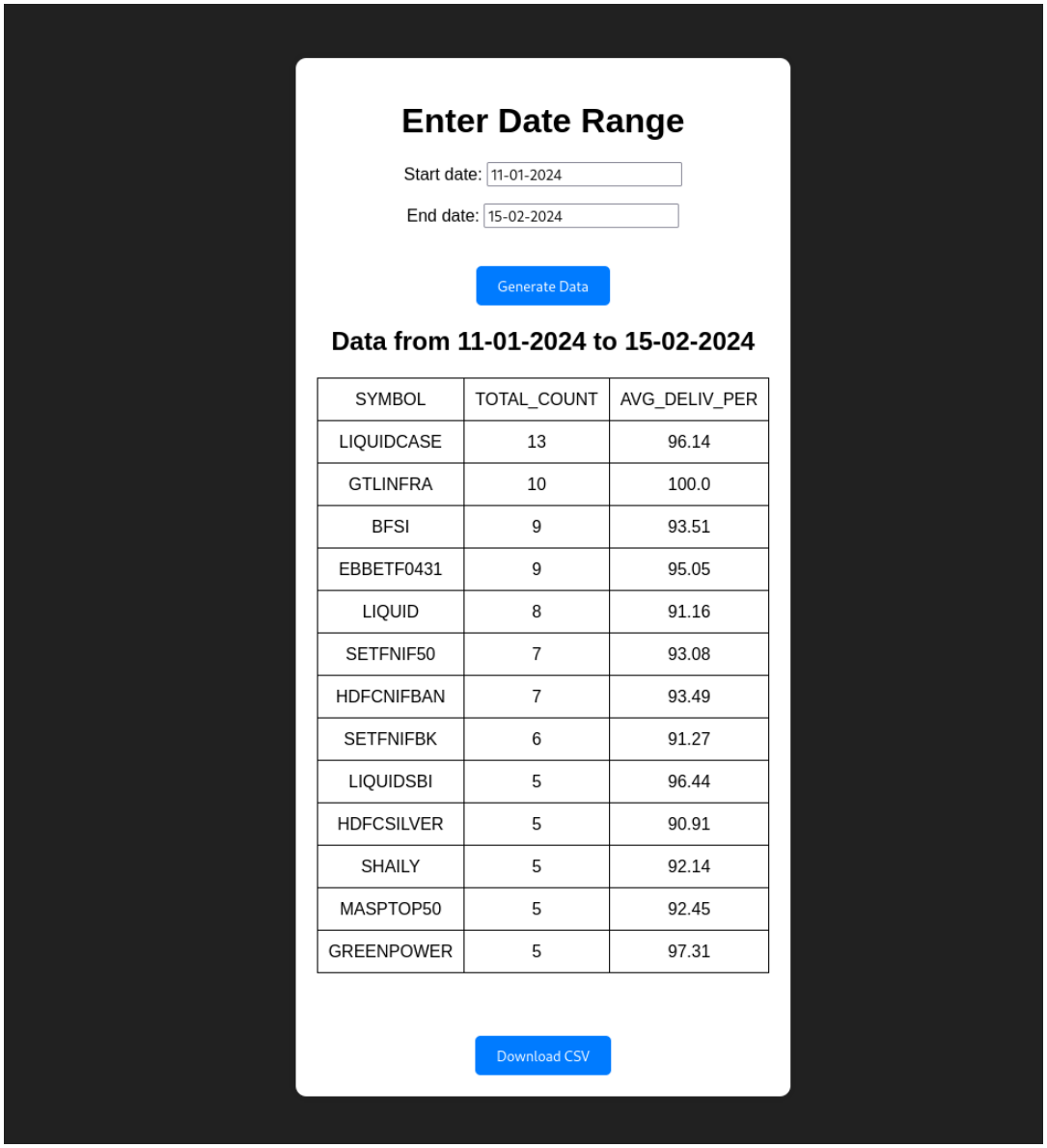


symbol_stats(2024-02-02 to 2024-03-30).csv | count_avg_stock.py

	SYMBOL	TOTAL_COUNT	AVG_DELIV_PER
2	LIQUIDCASE	32	95.3325
3	SETFNIF50	25	93.53200000000001
4	SETFNIFBK	20	94.21899999999997
5	GOLDETFADD	11	98.5009090909091
6	SMALLCAP	9	93.10333333333332
7	LIQUID	9	90.44222222222223
8	ABSLLIQUID	7	99.05428571428571
9	PGHL	7	92.88285714285715
10	HDFCNIFBAN	7	96.95428571428572
11	ROUTE	6	89.86333333333334
12	TIMKEN	6	90.24166666666667
13	IREDA	6	96.89666666666666
14	ENDURANCE	6	87.62333333333332
15	ICICIB22	6	88.78166666666668
16	GREENPOWER	5	97.31199999999998
17	MATRIMONY	5	90.248
18	GRINDWELL	5	91.596
19	KAJARIACER	5	88.456
20	GOCOLORS	5	87.852
21	APTUS	5	91.12
22	PAYTM	5	99.282
23	HDFCSML250	5	89.542
24	GALAXYSURF	5	90.06199999999998
25	CARBORUNIV	5	89.13000000000001
26	BANKETFADD	5	96.86600000000001
27			

Screenshots of the output in the web application :-





SYMBOL	TOTAL_COUNT	AVG_DELIV_PER
LIQUIDCASE	13	96.14
GTLINFRA	10	100.0
BFSI	9	93.51
EBBETF0431	9	95.05
LIQUID	8	91.16
SETFNIF50	7	93.08
HDFCNIFBAN	7	93.49
SETFNIFBK	6	91.27
LIQUIDSBI	5	96.44
HDFCSILVER	5	90.91
SHAILY	5	92.14
MASPTOP50	5	92.45
GREENPOWER	5	97.31

CHAPTER 5

CONCLUSION & FUTURE WORK

In this project, we developed a robust methodology to predict stock market prices by analyzing the delivery percentages of various stock symbols over a specified date range. The implemented Python script efficiently processes stock data, handling tasks such as date validation, data fetching, and calculation of average delivery percentages for each stock symbol.

Key Achievements

1. **Data Validation and Processing** : The script successfully converts user-provided date strings into datetime objects and ensures the start date is not after the end date. This validation step is crucial for accurate data processing.
2. **Data Fetching and Aggregation** : By iterating over each date in the specified range and skipping weekends, the script fetches stock data and aggregates delivery percentages for each symbol. This approach ensures comprehensive coverage of trading days.
3. **Statistical Analysis** : The calculation of average delivery percentages for each symbol, combined with filtering out symbols with fewer than five occurrences, ensures that the analysis focuses on stocks with sufficient trading activity.
4. **Output and Reporting** : The results are saved to a CSV file, making it easy to access and analyze the data further. The file is named based on the specified date range, facilitating organized record-keeping.

Insights and Implications

The project's methodology highlights the importance of delivery percentages as an indicator of stock performance. By focusing on symbols with high average delivery percentages, investors can identify stocks with consistent interest and potentially higher stability. This approach can complement other stock analysis techniques, offering a more comprehensive view of market trends.

Future Work [5]

1. **Incorporating Additional Indicators** : Future iterations of this project could incorporate other stock market indicators, such as volume, price trends, and sentiment analysis from news and social media, to enhance the predictive accuracy.
2. **Real-Time Data Processing** : Enhancing the script to process real-time data would provide more timely insights, benefiting short-term traders and allowing for quicker decision-making.
3. **Machine Learning Integration** : Integrating machine learning models could further refine predictions by identifying complex patterns and correlations in the data that are not immediately apparent through basic statistical analysis.

CHAPTER 6

REFERENCES

1. STOCK MARKET PREDICTION by Sanjith & Romal Fernando ; Sathyabama Institute of Science and Technology, Jeppiaar Nagar

2. Survey of stock market prediction using machine learning approach by Ashish Sharma ; Dinesh Bhuriya ; Upendra Singh ; 2017 International conference of Electronics, Communication and Aerospace Technology

3. NSE Stock Market Prediction Using Deep-Learning Models by Hiransha M ; Gopalakrishnan E.A ; Vijay Krishna Menon ; Soman K.P ; Centre for Computational Engineering and Networking, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore-641112, India.

4. Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications by Gaurang Sonkavde; Deepak Sudhakar Dharrao; Anupkumar M. Bongale; Sarika Doreswamy

5. Coffee Can Investing: The Low-Risk Road to Stupendous Wealth by Pranab Uniyal, Rakshit Ranjan, and Saurabh Mukherjea

6. Basic Python in Finance : How to Implement Financial Trading Strategies and Analysis using Python by Bob Mather

7. https://www.codecademy.com/learn/paths/finance-python

8. https://www.geeksforgeeks.org/python-numpy

9. https://www.w3schools.com/python/pandas/default.asp

10. https://www.nseindia.com/