Technical Analysis of Stock Market using Data Science and its tools

Shrawani Pagar
Department of Cybersecurity
Shah and Anchor Kutchhi Engineering College
Mumbai, India
shrawani.pagar15782@sakec.ac.in

Aditya Jaiswal
Department of Computer Science
Shah and Anchor Kutchhi Engineering College
Mumbai, India
aditya.jaiswal15974@sakec.ac.in

Atharva Auti
Department of Cybersecurity
Shah and Anchor Kutchhi Engineering College
Mumbai, India
atharva.auti15523@sakec.ac.in

Vansh Purohit
Department of Computer Science
Shah and Anchor Kutchhi Engineering College
Mumbai, India
vansh.purohit15866@sakec.ac.in

Vivek Mishra
Department of Cybersecurity
Shah and Anchor Kutchhi Engineering College
Mumbai, India
vivek.mishra16259@sakec.ac.in

***Abstract*— This paper presents a technical analysis approach for the stock market using data science and its associated tools. The objective is to develop a systematic framework that leverages data science techniques to analyze stock market data and make informed investment decisions. The proposed methodology involves data collection, preprocessing, feature engineering, and the application of various machine learning algorithms for predicting stock prices and identifying potential trading opportunities. The study demonstrates the effectiveness of data science in enhancing traditional technical analysis methods and provides insights into the potential benefits of integrating data-driven approaches into stock market analysis. Experimental results on historical stock data validate the proposed approach's accuracy and highlight its potential for improving investment strategies.**

Keywords—python, sql, yfinance, explosive candles, demand and supply zone, jupyter notebook, dataframe

I. Introduction

In today's fast-paced financial landscape, making informed investment decisions is of paramount importance. The stock market, with its intricate dynamics and ever-changing trends, demands sophisticated tools and methodologies to extract valuable insights from vast amounts of data. This paper presents a comprehensive analysis of the stock market using data science techniques and explores the effectiveness of various tools in aiding investors and traders in their decision-making process.

Over the past decade, data science has emerged as a powerful discipline that leverages advanced algorithms and computational tools to derive meaningful patterns and predictions from complex datasets. By harnessing the potential of data science, financial professionals can gain valuable insights into market trends, volatility, and potential investment opportunities. Technical analysis, a widely adopted approach in the financial industry, focuses on studying historical price and volume data to identify patterns and forecast future price movements.

This research aims to provide an in-depth exploration of the application of data science techniques in technical analysis of the stock market. The study investigates the performance and effectiveness of various tools and algorithms, including machine learning models, time series analysis, pattern recognition, and statistical methods. Through a rigorous analysis of historical stock data, this paper aims to determine the predictive power of these techniques and their practical implications for investors and traders.

In conclusion, this research paper offers a comprehensive examination of the technical analysis on the stock market using data science and its tools. By leveraging the power of data science, investors and traders can enhance their decision-making process and potentially achieve better returns. The findings of this study aim to contribute to the existing body of knowledge in the field of finance and provide valuable insights to practitioners and researchers alike.

II. Literature Survey

In conducting the research, the team thoroughly examined and analysed ten relevant research papers focused on the chosen topic. Each paper was scrutinized with meticulous attention to detail, ensuring a comprehensive understanding of the methodologies, findings, and implications discussed within. By studying these papers extensively, the team gained valuable insights into the application of data science techniques in technical analysis, providing a solid foundation for the current study's methodology and findings.

The paper titled “Technical Analysis Indicators in Stock Market Using Machine Learning: A Comparative Analysis” [1] investigates the application of machine learning techniques to analyze technical indicators in the stock market. The study compares the performance of various machine learning algorithms in predicting stock prices based on technical indicators. The research aims to provide insights into the effectiveness of different machine-learning approaches for stock market analysis and decision-making. The paper titled “On the Use of Technical Analysis Indicators for Stock Market Price Movement Direction Prediction” [2] explores the effectiveness of technical analysis indicators in predicting stock market price movement direction. The study analyses various technical indicators and evaluates their performance using machine learning algorithms. The research aims to provide insights into the applicability and accuracy of technical analysis indicators for predicting the direction of stock market price movements. The paper titled “Predicting stock trends through technical analysis and nearest neighbor classification” [3] explores the use of technical analysis and nearest neighbor classification for predicting stock trends. The study examines various technical indicators and applies a nearest-neighbor classification algorithm to predict future price movements. The research aims to enhance stock trading

decisions by leveraging technical analysis and classification methods for trend prediction.

The paper titled “Technical Analysis of Pattern-Based Stock Prediction Model Using Machine Learning” [4] presents a stock prediction model that combines technical analysis patterns with machine learning techniques. The study explores the performance of different machine learning algorithms in predicting stock prices based on pattern recognition. The research aims to provide insights into the effectiveness of pattern-based technical analysis combined with machine learning for stock market prediction. The paper titled “Forecast studies for financial markets using technical analysis” [5] explores the application of technical analysis in forecasting financial markets. The study investigates the performance of various technical indicators, such as moving averages and oscillators, in predicting market trends. The research analyzes historical stock data and assesses the effectiveness of technical analysis tools for short-term and long-term forecasting. The findings contribute to understanding the role of technical analysis in financial market prediction and decision-making.

The paper titled “From Technical Analysis to Text Analytics: Stock and Index Prediction with GRU” [6] presents a novel approach that integrates technical analysis with text analytics for stock and index prediction. The study utilizes Gated Recurrent Unit (GRU) neural networks to capture temporal dependencies in stock data and sentiment analysis of news articles. The results demonstrate improved prediction accuracy by combining numerical and textual information, highlighting the potential of this hybrid approach for stock market prediction. The paper titled “Ensemble of Technical Analysis and Machine Learning for Market Trend Prediction” [7] proposes an ensemble approach that combines technical analysis indicators and machine learning techniques for market trend prediction. The study integrates various technical indicators, such as moving averages and relative strength index, with machine learning algorithms, including support vector machines and random forests. The results demonstrate that the ensemble model outperforms individual approaches, highlighting the potential synergy between technical analysis and machine learning for market trend prediction. The paper titled “Integrating Fundamental and Technical Analysis of Stock Market through Multi-layer Perceptron” [8] proposes a model that combines fundamental and technical analysis for stock market prediction. The study utilizes a Multi-layer Perceptron (MLP) neural network to integrate financial ratios and technical indicators as input features. The results demonstrate the effectiveness of the integrated approach in predicting stock market trends, indicating the potential benefits of combining fundamental and technical analysis for enhanced decision-making.

The paper titled “Stock Market Analysis & Prediction” [9] investigates stock market analysis and prediction. The study explores various technical indicators and machine learning algorithms for predicting stock market trends. It analyzes historical stock data and applies machine learning models to forecast future price movements. The research aims to assist investors in making informed decisions based on accurate predictions, enhancing their understanding of stock market dynamics, and improving investment strategies. The paper titled “Discovery of Technical Analysis Patterns” [10] focuses on the automated discovery of technical analysis patterns in financial time series data. The study proposes a methodology that combines data mining techniques with technical analysis principles to identify and extract recurring patterns. The discovered patterns can provide valuable insights for traders and investors in making informed decisions. The research contributes to the development of automated tools for pattern recognition in the field of technical analysis for financial markets.

III. Methodology

A. *Python and SQL importance in Data Science*

Python has established itself as a dominant programming language in the realm of data science, primarily because of its remarkable versatility, simplicity, and an extensive array of libraries and frameworks. This powerful language provides a rich ecosystem of tools and packages that empower data scientists with efficient data manipulation, analysis, and modelling capabilities. One of the key factors driving Python's popularity in data science is its ease of use and readability. Its syntax is straightforward and intuitive, making it accessible to both novice and experienced programmers.

Furthermore, Python boasts a vast collection of libraries specifically designed for data science, such as NumPy, pandas, and sci-kit-learn. These libraries offer comprehensive functionality and powerful algorithms, enabling data scientists to perform complex operations with ease. Python's flexibility extends beyond its native capabilities, as it seamlessly integrates with SQL databases through SQL connectors. These connectors provide an interface that allows Python to connect and interact with relational databases, ensuring efficient communication with SQL servers. By leveraging SQL connectors, data scientists and analysts can retrieve and manipulate data directly from SQL databases, streamlining the process of data extraction and transformation.

The integration of Python with SQL databases through SQL connectors brings forth numerous benefits. Firstly, it empowers data scientists and analysts to effortlessly access and work with data stored in SQL databases, eliminating the need for manual data transfers or exports. This direct interaction enhances productivity and saves time by eliminating unnecessary steps in the data retrieval process. Python's vast array of libraries for data manipulation, most notably pandas, can be seamlessly combined with SQL queries to process and analyze large datasets. The flexibility of Python allows for efficient integration of SQL queries within data manipulation workflows, enabling complex data transformations and aggregations. This amalgamation of Python's data manipulation capabilities with the power of SQL provides data scientists with unparalleled flexibility in handling and analyzing large and diverse datasets.

In conclusion, Python's widespread adoption in the field of data science can be attributed to its versatility, simplicity, and the extensive collection of libraries and frameworks it offers. The seamless integration of Python with SQL databases through SQL connectors enhances data scientists' abilities to extract, manipulate, and analyze data efficiently. By leveraging Python's libraries for data manipulation in conjunction with SQL queries, data scientists can tackle complex data analysis tasks with ease, ultimately contributing to the advancement of the field of data science. Pandas is a fundamental Python library for data analysis, offering efficient and flexible data structures that simplify working with structured data. Python, known for its simplicity and wide range of libraries, is an ideal programming language for

data science tasks. It provides built-in functions for data manipulation, machine learning, and connecting to databases. Python's versatility extends to data visualization and automation of tasks. Being free, open-source, portable, and scalable, Python is accessible to all and can handle large datasets effectively. With a thriving community of developers, Python continues to be a powerful tool in the field of data analysis and beyond.

B. Preprocessing

The initial step of analysis involves gathering and preprocessing the stock dataset. To extract historical stock data from Yahoo Finance, we employ the YFinance library. This library utilizes web scraping techniques to retrieve data directly from the Yahoo Finance website. The extraction process utilizes a "Token configuration File" that contains information such as stock symbols, time, and timeframe for the desired data. By specifying the stocks and time period in the configuration file, we ensure that we gather the relevant dataset for the analysis. The extracted dataset includes important information such as open, high, low, close prices, volume, and timeframes. However, before proceeding with the analysis, we perform data cleansing and filtering to ensure the dataset's quality and integrity. This step involves removing any null values and correcting incorrectly formatted data points, ensuring that the dataset is clean and consistent.

The resulting dataset is then formatted into Pandas data frames for each stock, making it easier to manipulate and analyze the data. Stocks with insufficient data during the specified time are excluded from the analysis, as we require a sufficient data to draw meaningful insights. The preprocessed dataset encompasses the stock symbol, timeframe, candlestick data (open, high, low, close, volume), and the corresponding date range. This refined dataset serves as the input for the subsequent algorithm and analysis stages, providing us with the necessary data to perform the analysis. In summary, the process involves utilizing the YFinance library to extract the stock dataset from Yahoo Finance, specifying the desired stocks and time in a configuration file.

The extracted data then undergoes cleansing and filtering to remove null values and correct formatting errors. The resulting dataset is formatted into Pandas Data Frames, and stocks with insufficient data are excluded. This meticulous data preprocessing step ensures that we have a clean and consistent dataset for focused analysis and modelling efforts.

C. Defining the Algorithm

In order to enhance the analysis and understanding of the stock market, a series of algorithms were developed to provide valuable insights and visual representations of market trends and potential trading opportunities. The first algorithm focused on classifying different types of candles, such as explosive candles and basing candles. This classification introduced additional parameters in the Data Frame, including the freshness of the candle and its explosive score. By identifying and categorizing candles, traders gain a deeper understanding of market dynamics.

Another algorithm was specifically designed to identify demand and supply zones within the stock market. These zones serve as crucial indicators of potential buying and selling areas. By accurately pinpointing these zones, traders can make informed decisions regarding their investments. To determine the most reliable and recent zone, a Quality Factor (QF) was implemented. The QF acts as a measure of the zone's quality and fitness, allowing traders to select the zone that is most likely to yield favourable results.

Using the Quality Factor, the algorithm further calculates the net profit or loss that would be incurred if a stock is purchased within the identified zone at a specified time. This evaluation enables traders to assess the potential risk and reward associated with a particular investment decision. Armed with this information, traders can make more informed choices and minimize potential losses.

Additionally, an algorithm was developed to visualize the candlestick graph, a popular tool for analysing market trends. This algorithm utilizes packages such as mpl finance and plotly_graph_object to accurately plot the candlestick chart. Moreover, the demand and supply zones are prominently marked on the graph, providing a clear visual reference for traders and investors. This visual representation enhances the overall analysis process, allowing for quick identification of market patterns and potential trading opportunities.

In summary, these algorithmic designs significantly enhance the analysis and understanding of the stock market. By effectively identifying candle types, determining demand and supply zones, evaluating their quality, and providing visual representations of market trends, traders and investors are equipped with valuable tools to make more informed decisions. These algorithms enable a deeper level of analysis, resulting in improved trading strategies and potential for greater profitability.

D. Designing the Process

In the data analysis process, the pre-processed stock data is read in from the configuration file, encompassing essential information such as open, high, low, close, and volume for each trading day. Next, the algorithm developed previously is employed to identify the type of candle, distinguishing between explosive and basing candles. This classification assigns an explosive score and freshness value to each candle, providing insights into their characteristics. Based on the candle characteristics and the algorithm's parameters, the demand and supply zones are calculated, indicating potential areas of market interest.

To further refine the analysis, the algorithm determines the best quality factor zone, focusing on factors like zone size, age, and freshness. The zone with the highest quality factor is identified as the one most likely to yield favourable results. By considering the identified demand and supply zones, the estimated net profit or loss is calculated for a trade entered at those specific zones, aiding in decision-making. To enhance visualization, the candlestick chart is generated using the mplfinance and Plotly Graph Object libraries. The identified demand and supply zones are marked on the chart, and a legend is provided to explain these zones, facilitating a comprehensive understanding of the chart. The implementation of the program was carried out using Python within Jupyter Notebook. Several libraries, including Pandas, Numpy, Matplotlib, Plotly, and mplfinance, were utilized to support the analysis and visualization processes.

E. Calculating the Predictions Flow

The process begins with data scraping from the yfinance library. This involves applying and verifying the dates and changing the datetime format to ensure accurate data retrieval. Once the data is obtained, an analysis is performed

to determine the type of candle, categorizing it as either an explosive candle or a basing candle. To further assess the candles, factors such as freshness and quality are considered. Freshness refers to the recency of the candle, indicating how recently it occurred in the market. The quality factor is then calculated, which serves as a measure of the candle's reliability and effectiveness. If the quality factor is lower than 7, the candle is considered less fresh and may not provide reliable information. On the other hand, if the quality factor exceeds 7, the candle is deemed fresh and trustworthy for analysis.

Additionally, the concept of zones is introduced, which are specific areas in the market that offer valuable insights for trading decisions. The quality factor is applied to these zones as well. If a zone's quality factor falls below 7, it is considered less fresh and may not yield optimal results. Conversely, zones with a quality factor higher than 7 are regarded as fresh and can provide valuable trading opportunities. By incorporating these concepts of an explosive score, freshness, quality factor, and zones, the analysis becomes more robust and enables traders and investors to make informed decisions based on the reliability and recency of the data.

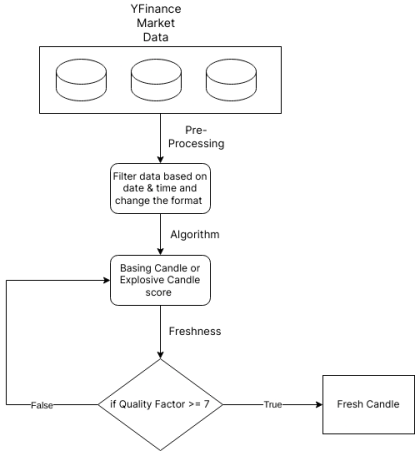

Fig. 1. Flowchart of the process

Fig. 1. provides an overview of the entire process, showcasing the sequential steps involved in collecting, preprocessing, and applying the data using the algorithm to generate fresh candles. This visual representation illustrates the systematic approach taken to ensure accurate and reliable results throughout the data analysis and candle generation process.

IV. Results

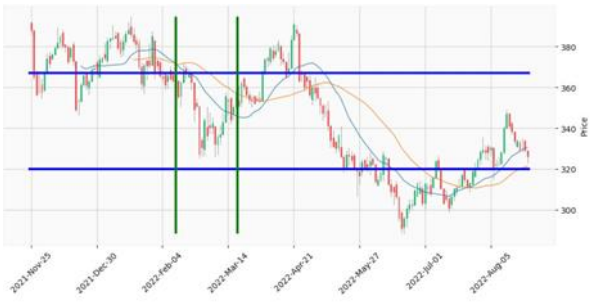

Fig. 2. Valid demand zone

Fig. 2. showcases a well-defined and valid demand zone, characterized by a noticeable upward movement in all prices. This upward trend reflects a significant buying interest within the market, indicating a favorable trading opportunity for traders and investors. The demand zone serves as a crucial indicator of a market's strength and potential profitability. It suggests that there is a substantial level of demand for the given asset, which can potentially drive prices higher in the future. Traders who recognize this demand zone can capitalize on the prevailing market sentiment by strategically entering long positions, expecting the upward momentum to continue.

By leveraging this valuable information, market participants can make informed decisions and potentially maximize their trading gains. Overall, the presence of a valid demand zone depicted in the figure provides a strong signal for traders to consider favorable trading opportunities in line with the market's buying sentiment. The analysis process involves utilizing a predefined Excel file that contains the token names corresponding to specific stocks. These tokens serve as inputs in the custom-built code for web scraping using the YFinance library. By using web scraping techniques, the code retrieves the price and data of each stock associated with the respective tokens.

Once the web scraping is performed, the obtained stock prices and data are loaded into data frames. These data frames serve as the building blocks for creating the dataset. Each data frame represents a specific stock, containing information such as open, high, low, and close prices, as well as volume and other relevant data points.

The combination of these individual data frames forms the complete dataset, which is directly sourced from Yfinance. This dataset comprises the extracted information for each stock, organized in a structured manner. It provides a comprehensive view of the historical data for the analyzed stocks, facilitating further analysis and modeling.

By leveraging the predefined Excel file to extract tokens and subsequently using web scraping techniques with Yfinance, the code ensures the seamless integration of stock prices and data into the data frames, ultimately constructing a comprehensive dataset. This approach allows for direct access to the desired data from yfinance, streamlining the analysis process and enabling efficient data-driven decision-making.

V. Conclusion

The paper highlights the significant role of data science and its tools in the field of technical analysis for stock market forecasting. The paper explores the application of various data science techniques, such as machine learning algorithms, pattern recognition, and data visualization, in analyzing stock market data and predicting market trends. The findings demonstrate the effectiveness of these tools in providing valuable insights and aiding decision-making processes in the financial domain. This research contributes to the growing body of knowledge in data-driven stock market analysis and emphasizes the importance of leveraging data science tools for informed investment strategies.

Name	Token	Start	End	timeFrame
RELIANCE	reliance.ns	8/1/2022	9/15/2022	1h
INFOSYS	infy.ns	8/1/2022	10/2/2022	1h
BPCL	bpcl.ns	9/22/2022	9/23/2022	5m
TATA	tcs.ns	9/1/2022	10/2/2022	1h
LNT	lt.ns	8/1/2022	8/15/2022	15m

Fig. 3. Tokens used for web scraping

Fig. 3. illustrates the comprehensive set of tokens employed during the web scraping process to retrieve and acquire the dataset. These tokens encompass a range of commands and parameters utilized in conjunction with the YFinance library. The image showcases the systematic use of these tokens, including the specification of stock symbols, time period, and timeframe, which are essential for retrieving the desired historical stock data from Yahoo Finance.

By visually representing the tokens utilized during the web scraping process, the image offers a clear overview of the systematic approach taken to gather the dataset. Each token serves a specific purpose, allowing for precise customization of the data extraction process. This meticulous selection of tokens ensures that the resulting dataset contains the necessary information, such as open, high, low, close prices, volume, and timeframes, for comprehensive analysis.

Overall, the image provides an insightful visual representation of the tokens employed during web scraping, showcasing the meticulous approach taken to acquire the dataset. This visual aid enhances the understanding of the web scraping process and emphasizes the importance of each token in retrieving the necessary stock data for subsequent analysis.

REFERENCES

[1] Y. K. Pardeshi and P. P. Kale, "Technical Analysis Indicators in Stock Market Using Machine Learning: A Comparative Analysis," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2021, pp. 1-6, doi: 10.1109/ICCCNT51525.2021.9580172.

[2] R. F. Oğuz, Y. Uygun, M. S. Aktaş and İ. Aykurt, "On the Use of Technical Analysis Indicators for Stock Market Price Movement Direction Prediction," 2019 27th Signal Processing and Communications Applications Conference (SIU), Sivas, Turkey, 2019, pp. 1-4, doi: 10.1109/SIU.2019.8806422.

[3] L. A. Teixeira and A. L. I. de Oliveira, "Predicting stock trends through technical analysis and nearest neighbor classification," 2009 IEEE International Conference on Systems, Man and Cybernetics, San Antonio, TX, USA, 2009, pp. 3094-3099, doi: 10.1109/ICSMC.2009.5345944.

[4] C. Dadiyala and A. Ambhaikar, "Technical Analysis of Pattern Based Stock Prediction Model Using Machine Learning," 2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2021, pp. 1-9, doi: 10.1109/ICSES52305.2021.9633961..

[5] Vinit Nagarajan, Ying Wu, Min Liu and Qing-Guo Wang, "Forecast studies for financial markets using technical analysis," 2005 International Conference on Control and Automation, Budapest, Hungary, 2005, pp. 259-264 Vol. 1, doi: 10.1109/ICCA.2005.1528128.

[6] T. . -T. Teoh et al., "From Technical Analysis to Text Analytics: Stock and Index Prediction with GRU," 2019 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM), Bangkok, Thailand, 2019, pp. 496-500, doi: 10.1109/CIS-RAM47153.2019.9095772.

[7] A. P. Ratto, S. Merello, L. Oneto, Y. Ma, L. Malandri and E. Cambria, "Ensemble of Technical Analysis and Machine Learning for Market Trend Prediction," 2018 IEEE Symposium Series on Computational Intelligence (SSCI), Bangalore, India, 2018, pp. 2090-2096, doi: 10.1109/SSCI.2018.8628795.

[8] A. Namdari and Z. S. Li, "Integrating Fundamental and Technical Analysis of Stock Market through Multi-layer Perceptron," 2018 IEEE Technology and Engineering Management Conference (TEMSCON), Evanston, IL, USA, 2018, pp. 1-6, doi: 10.1109/TEMSCON.2018.8488440.

[9] B. Shivani and S. P. G. Rao, "Stock Market Analysis & Prediction," 2021 International Conference on Forensics, Analytics, Big Data, Security (FABS), Bengaluru, India, 2021, pp. 1-5, doi: 10.1109/FABS52071.2021.9702549.

[10] U. Markowska-Kaczmar and M. Dziedzic, "Discovery of Technical Analysis Patterns," 2008 International Multiconference on Computer Science and Information Technology, Wisla, Poland, 2008, pp. 195-200, doi: 10.1109/IMCSIT.2008.4747239.