**Research and Prediction of Stock Market using ML for Algorithmic Trading**

### This project report is submitted to

### Silicon University, Odisha

***in partial fulfillment of the requirements for the award of the degree of***

**Bachelor of Technology**

***in***

**Computer Science and Engineering**

### Submitted by

|  |  |
| --- | --- |
| **Bishal Mohanty** | (Regd. No. : 2001209106) |
| **Abhijit Mishra** | (Regd. No. : 2001209062) |
| **Subhojit Das** | (Regd. No. : 2001209208) |

**Project Group No.: CSE 29**

### Under the Esteemed Supervision of

**Prof. (Dr.) Bhagwat Prasad Chaudhury**



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**SILICON UNIVERSITY ODISHA**

**SILICON HILLS, BHUBANESWAR – 751024, ODISHA, INDIA**

**May, 2024**

**CERTIFICATE**

This is to certify that the work contained in the project entitled **“Research and Prediction of Stock Market using ML for Algorithmic Trading”**, submitted by **Bishal Mohanty (Regd. No.: 2001209106), Abhijit Mishra (Regd. No.: 2001209062) and Subhojit Das (Regd. No.: 2001209208)** is a record of bonafide works carried out by them under my supervision and guidance. The contents embodied in the project is being submitted as a part of 8th semester project for the undergraduate curriculum and have not been submitted for the award of any other degree or diploma in this or any other university.

## Date : 02/05/2024

## Place: Bhubaneswar

## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## Prof. Bhagwat Prasad Chaudhury

## Professor, Department of Computer Science & Engineering

## \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## External Examiner



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**SILICON UNIVERSITY ODISHA**

**BHUBANESWAR – 751024**

**DECLARATION**

We hereby certify that:-

1. The work contained in the project is original and has been done by ourselves under the supervision of our supervisor.
2. The work has not been submitted to any other Institute for any degree or diploma.
3. We have conformed to the norms and guidelines given to us by the Project Review Committee of our department.
4. Whenever we have used materials (data, theoretical analysis and text) from other sources, we have given due credit to them by citing them in the text of the project and giving their details in the references.

Date: 02/04/2024

Place: Bhubaneswar

|  |  |  |
| --- | --- | --- |
| **Bishal Mohanty** | (Regd. No. 2001209106) | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| **Abhijit Mishra** | (Regd. No. 2001209062) | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| **Subhojit Das** | (Regd. No. 2001209208) | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**SILICON UNIVERSITY ODISHA**

**BHUBANESWAR – 751024**

**ACKNOWLEDGEMENTS**

I take this opportunity to acknowledge the many individuals whose support and guidance have been instrumental in the completion of this project work. First and foremost, I am deeply indebted to my supervisor and mentor, Dr. Bhagwat Prasad Chaudhury, whose expertise, encouragement, and invaluable feedback have shaped this project into its final form. His unwavering support and insightful guidance have been the cornerstone of this endeavor, and for that, I am truly grateful.

I am also grateful to the faculty members of our CSE department whose teachings have enriched my knowledge and skills, laying the foundation for this work. To my group mates, whose camaraderie and collaboration have made this journey enjoyable and fulfilling, I extend my heartfelt gratitude. Special thanks are due to my friends and family for their unwavering support, understanding, and encouragement during the ups and downs of this endeavor.

Finally, I would like to wind up by paying my heartfelt thanks to all those whose names might not appear here but have, in one way or another, contributed to the completion of this project. Thank you all for your unwavering support and encouragement.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | **Bishal Mohanty** | (Regd. No. 2001209106) | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | | **Abhijit Mishra** | (Regd. No. 2001209062) | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | | **Subhojit Das** | (Regd. No. 2001209208) | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | |  |
|  |  |
|  |  |
|  |  |



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**SILICON UNIVERSITY ODISHA**

**BHUBANESWAR – 751024**

**ABSTRACT**

This research project delves into the realm of algorithmic trading in the stock market, employing a blend of web scraping techniques, machine learning algorithms, and intuitive front-end development. Departing from the initial plan of integrating Apache Kafka for data collection, the project pivoted to utilize the BeautifulSoup Python library for web scraping financial data from Yahoo Finance. This shift enabled the extraction of pertinent market information, including historical stock data, which served as the foundation for subsequent analysis.

The core of the project revolves around the application of an ensemble of Logistic Regression, SVC, XGBoost and Long Short-Term Memory (LSTM) algorithms to forecast stock movements based on historical trends. These machine learning models, trained on historical stock data retrieved via web scraping, offer predictive insights into future market trends. Additionally, the Moving Average Convergence Divergence (MACD) indicator was incorporated into the project's frontend interface, complementing the predictive capabilities of the models by providing real-time metrics and insights for informed decision-making.

The frontend interface, developed using the Streamlit library and conventional web development tools, offers users a seamless experience in accessing and interpreting stock market data. Leveraging the metrics obtained from web scraping and predictive models, the interface provides comprehensive statistics for each stock, including historical average dividends and market volume. Users are empowered to filter stocks based on specific criteria, such as dividend yield or trading volume, aligning their investment strategies with individual preferences and risk profiles.

In summary, this interdisciplinary approach merges web scraping, machine learning, and frontend development to advance algorithmic trading strategies and offer practical tools for investors. By leveraging historical data and predictive analytics, the project contributes to enhancing decision-making processes in navigating the complexities of stock market dynamics.

# LIST OF ABBREVIATIONS

## Abbreviation Description

ARIMA AutoRegressive Integrated Moving Average

LR Logistic regression

LSTM Long Short-Term Memory

MACD Moving Average Convergence Divergence

RMSE Root Mean Square Error

RSI Relative Strength Index

SVM Support Vector Machine

ROC AUC Area Under the Receiver Operating Characteristic Curve

XGBoost Extreme Gradient Boosting

# LIST OF FIGURES

|  |  |  |
| --- | --- | --- |
|  |  | **Page #** |
| **Chapter 1.** |  |  |
| Figure 1.1  Figure 1.2 | Expected Model  Work flow diagram of proposed method | 11  13 |
| Chapter 3. |  |  |
| Figure 3.1  Figure 3.2  Figure 3.3  Figure 3.4  Figure 3.5  Figure 3.6  Figure 3.7  Figure 3.8  Figure 3.9  Figure 3.10  Figure 3.11  Figure 3.12 | Work Flow Diagram  Yahoo Finance script for scraping  Using yfinance library to scrape stock datasets of any stock  Changing date format to YYYY-MM-DD  Graph for stock price over the years  Distribution and barplots for dataset  LR,SVC,XGBoost  Long Short Term Memory  MACD-I  MACD-II  Front-end using Streamlit  Formating and running streamlit with HTML/CSS | 21  22  22  23  23  24  25  26  27  28  29  29 |
| **Chapter 4.**  Figure 4.1  Figure 4.2  Figure 4.3  Figure 4.4  Figure 4.5  Figure 4.6  Figure 4.7 | Accuracy Result of LR,SVM,XGBoost  Confusion Matrix of Result of LR,SVM,XGBoost  LSTM True vs Predicted  MACD True vs Predicted  MACD MSE Error  StreamLit HomePage  Streamlit ML Models depiction in Streamlit | 31  31  32  32  32  33  33 |
|  |  |  |

# CONTENTS

## CONTENT DETAILS PAGE NO.

|  |  |  |
| --- | --- | --- |
| **Title Page** |  | 1 |
| **Certificate** | | 2 |
| **Declaration** |  | 3 |
| **Acknowledgements** | | 4 |
| **Abstract** |  | 5 |
| **List of Abbreviations** | | 6 |
| **List of Figures** | | 7 |
| **Contents** |  | 8-9 |
|  |  |  |
| **Chapter 1.** | ***Introduction*** | 10-20 |
|  |  |  |
| **1.1.** | **Background**  **1.1.1** Traditional Approaches | 10 |
| **1.2.** | **Problem Statement** | 11 |
| **1.3.** | **Objective and Motivation** | 11 |
| **1.4.** | **Proposed Method**  **1.4.1** Method to increase Performance  **1.4.2** Evaluation Parameter | 13 |
| **1.5.** | **Project Organization** | 15 |
|  | **Summary** | 16 |
|  |  |  |
| **Chapter 2.** | ***Literature Survey*** | **17** |
|  |  |  |
| **2.1**  **2.2** | **Literature Survey**  **Scope of the Work**  **Future Directions**  **Summary** | 17  18  19  20 |
|  |  |  |
| **Chapter 3.** | ***Methodology*** | **21-30** |
|  | **Data Collection**  **Data Preprocessing**  **Data Visualization**  **ML Models**  **Model Development**  **Website Creation**  **Software Tools and Ethical Consideration**  **Limitations** | 21  22  23  24  28  29  30  30 |
|  |  |  |
| **Chapter 4.** | ***Experimental Results*** | **31-33** |
|  |  |  |
| **4.1.** | **Results** | 31-33 |
|  | 4.1.1 Logistic Regression | 31 |
|  | 4.1.2 LSTM | 32 |
|  | 4.1.3 MACD  4.1.4 Front-end  **Summary** | 32  33  33 |
| **Chapter 5.** | ***Conclusion*** | **34 – 35** |
|  |  |  |
| **5.1.** | **Future scope** | 34 |
|  |  |  |
|  | **References** | **36** |
|  |  |  |

**CHAPTER 1**

**INTRODUCTION**

In this section, we provide an overview of the research project, which delves into the realm of algorithmic trading in the stock market. Our interdisciplinary approach combines web scraping techniques, machine learning algorithms, and intuitive front-end development to enhance algorithmic trading strategies.

* 1. **BACKGROUND**

**1.1.1 Traditional Approaches**

Historically, three conventional approaches have been employed for stock price prediction:

1. **Technical Analysis:**

* Technical analysis relies on historical price charts, patterns, and indicators to forecast future price movements.
* Traders use tools such as moving averages, Bollinger Bands, and Relative Strength Index (RSI) to identify trends and potential entry/exit points..

1. **Traditional Time Series Forecasting:**

* Time series models, such as ARIMA (AutoRegressive Integrated Moving Average), have been widely used.
* These models capture temporal dependencies in stock prices and attempt to predict future values based on historical data.

1. **Machine Learning Methods:**

* Machine learning algorithms have gained prominence due to their ability to handle complex patterns and large datasets.
* Commonly used algorithms include:

1. Linear Regression: Predicts stock prices based on linear relationships with relevant features.
2. Random Forest: A powerful ensemble technique that combines multiple decision trees for robust predictions.
3. Support Vector Machine (SVM): Used for classification tasks, SVM can also be adapted for regression in stock price prediction.
   1. **PROBLEM STATEMENT**

The project aims to address the following problem: How can we leverage historical stock data and predictive analytics to improve algorithmic trading strategies? Specifically, we seek to forecast stock movements based on historical trends and provide real-time metrics for informed decision-making.

* 1. **OBJECTIVE AND MOTIVATION**

*“The objective of stock market prediction research is to develop accurate models for forecasting stock prices and enhance decision-making for investors and traders”.*

A screenshot of a computer screen

Description automatically generated

Figure 1.1. Expected Model

In the dynamic landscape of financial markets, stock market prediction research serves several crucial objectives:

1. Forecasting Accuracy: Accurate predictions of stock prices are essential for investors and traders. By developing reliable models, we aim to minimize prediction errors and provide robust insights into future price movements. These accurate forecasts empower decision-makers to allocate resources effectively.
2. Risk Management: Effective risk management is at the heart of successful investing. By leveraging predictive models, investors can assess and mitigate risks associated with stock market fluctuations. Understanding potential price movements allows for timely adjustments to investment portfolios, reducing exposure to adverse market conditions.
3. Market Behavior Insights: Stock prediction models offer valuable insights into market behavior over time. By analyzing historical trends, volatility patterns, and cyclic phenomena, researchers gain a deeper understanding of market dynamics. These insights inform strategic decisions and guide investment strategies.
4. Algorithmic Trading Strategies: Algorithmic trading, driven by accurate predictions, has revolutionized financial markets. Our research aims to contribute to the development of sophisticated trading algorithms. These algorithms automate buy/sell decisions based on predefined rules, optimizing trading efficiency and capitalizing on market opportunities.

Researchers in the field of stock market prediction are driven by several compelling factors.

Firstly, the allure of profit serves as a significant motivator. Accurate predictions offer investors a competitive edge, allowing them to capitalize on lucrative opportunities. Even a slight advantage in timing can translate into substantial financial gains, prompting researchers to explore innovative strategies for predicting market trends.

Moreover, investor confidence plays a crucial role in driving research efforts. Trustworthy predictions bolster investor confidence, as reliance on accurate models enables informed decision-making with conviction. A robust prediction framework not only instills trust in financial markets but also encourages active participation, contributing to market liquidity and stability. Additionally, the pursuit of market efficiency serves as a key motivator for researchers. Efficient markets thrive on accurate price information, and predictive models play a vital role in reducing information asymmetry. By providing investors with reliable forecasts, these models contribute to market transparency and equity, fostering a more efficient allocation of resources. Furthermore, the intellectual challenge inherent in stock price prediction fuels research endeavors in this field.

Stock market prediction is a complex and multifaceted problem, presenting researchers with a stimulating intellectual challenge. The quest to improve prediction accuracy through innovative methodologies and novel approaches keeps the field vibrant and dynamic, driving continuous progress and innovation.

## PROPOSED METHOD

## A diagram of a software development process Description automatically generated

Figure 1.2. Work flow diagram of proposed method

Our proposed method involves utilizing web scraping techniques in Python to extract stock price data from online sources like Yahoo Finance or Google Finance. Using libraries such as BeautifulSoup and pandas, we parse and format the data into a structured format, typically a CSV file. This allows for easy storage and further analysis, enabling users to make informed decisions based on the latest stock price information. Optionally, automation can be implemented to regularly update the data, ensuring analysis is based on up-to-date information. Overall, our method offers a practical and efficient solution for gathering and analyzing stock price data for financial decision-making purposes.

* + 1. **Method to Improve the Performance of Discovery**

The method proposed to enhance the performance of discovery in stock market prediction focuses on leveraging two key techniques: clustering and indexing. By implementing these techniques, the aim is to streamline the discovery process and optimize semantic matching. Clustering involves grouping services based on similar functional characteristics, enabling the identification of relevant services for a given query. This approach enhances efficiency by reducing the pool of services subjected to semantic matching. Additionally, indexing facilitates the retrieval of services mapped by inputs and outputs, further refining the selection process. Through the synergistic application of clustering and indexing, the method seeks to enhance the precision and efficiency of discovery, ultimately improving the predictive capabilities of the stock market prediction system.

* + 1. **Evaluation Parameter ─ F1 Score**

In the context of our stock market prediction project, the evaluation parameter extends beyond traditional metrics like RMSE. Instead, we adopt evaluation metrics tailored to the classification nature of predicting stock movements. The F1 score and Precision-Recall metrics offer valuable insights into the model's ability to classify different market conditions accurately.

**F1 Score**: The F1 score is calculated as the harmonic mean of precision (P) and recall (R), and is given by the formula:

Where:

P (Precision) where TP is the number of true positives and FP is the number of false positives.

𝑅 (Recall) where FN is the number of false negatives. The F1 score provides a balanced assessment o f a model's ability to correctly identify positive and negative instances in stock market prediction, considering both precision and recall.

*P* is the precision, defined as the ratio of true positive (TP) predictions to the total number of positive (P) predictions, and is calculated as:

R is the recall, defined as the ratio of true positive (TP) predictions to the total number of actual positive (A) instances, and is calculated as:

***​***

**​**

* 1. **PROJECT ORGANIZATION**

**Chapter 1** This chapter provides a comprehensive overview of the project, including the background of stock market prediction, the motivation behind the research, and its objectives. It sets the stage for subsequent chapters by presenting the context and significance of the project.

**Chapter 2** This chapter delves into the existing body of research on stock market prediction. It examines previous studies and methodologies employed by renowned researchers in the field. Additionally, the chapter discusses the strengths and limitations of prior works, providing valuable insights for guiding the current research.

**Chapter 3** This chapter outlines the proposed approach for stock market prediction. It details the techniques and algorithms utilized in data collection, preprocessing, model development, and evaluation. Special emphasis is placed on the innovative methods employed to enhance predictive accuracy and reliability.

**Chapter 4** This chapter presents the findings obtained from applying the proposed methodology to real-world stock market data. It includes detailed analyses of prediction outcomes, performance metrics, and comparisons with baseline models or benchmarks. The chapter offers a comprehensive evaluation of the effectiveness and practical applicability of the developed predictive models.

**Chapter 5** The final chapter, titled "Conclusion," serves as a summation of the project's key findings, contributions, and implications. It highlights the significance of the research outcomes and their potential impact on the field of stock market prediction. Additionally, the chapter discusses future research directions and concludes with reflections on the broader societal implications of the work.

**SUMMARY**

In this chapter, we've delved into the core aspects of our project, emphasizing the significance of accurate stock market predictions in enhancing trading strategies and bolstering investor confidence. We've elucidated the practical applications of stock market prediction, spanning investment decisions, risk management, and portfolio optimization. Introducing evaluation metrics such as F1 score and Precision-Recall, we've tailored our approach to ensure the reliability of predictive models. Our motivation stems from a desire to contribute to market efficiency while achieving our objectives of developing robust predictive models for stock market dynamics. Additionally, we've outlined a methodological approach involving web scraping techniques in Python to extract stock price data from online sources, subsequently formatting it into a structured format such as a CSV file. This streamlined approach facilitates efficient analysis and informed decision-making based on the latest stock price information, aligning with our overarching goals. Through this chapter, we've provided a succinct overview of our project's scope, objectives, and methodologies, laying the groundwork for subsequent discussions.

**CHAPTER 2**

**LITERATURE REVIEW**

This chapter of our project undertakes a thorough exploration of the domain of stock market prediction, drawing insights from existing literature, studies, and technological advancements. It establishes the foundation for comprehending the current state, trends, challenges, and innovations within the realm of stock market forecasting. Through a review of academic papers, industry reports, case studies, and technological frameworks, this chapter aims to identify key insights, best practices, and innovative solutions to inform the development of our stock market prediction model.

**2.1. Literature Survey**

The literature survey in the domain of stock market prediction reveals a significant reliance on machine learning (ML) techniques for forecasting stock prices and trading signals. Various studies have explored the application of ML algorithms, including supervised and unsupervised learning, for predictive modeling in financial markets. Singh and Zhang[1] (2018) investigated the application of random forests, gradient boosting, and deep learning techniques for predicting stock returns of S&P 500 companies. They compared the performance of these models and discussed the challenges and opportunities in using ML for stock market prediction. Alouini, Kechouri, and Maâtallah [2] (2020) examined the use of deep learning models, specifically convolutional neural networks (CNNs) and long short-term memory networks (LSTMs), for stock market analysis. They proposed an architecture that combines both CNNs and LSTMs to capture spatial and temporal dependencies in stock price data. Iqbal, Kumar, and Pal [3] (2019) explored the effectiveness of machine learning algorithms such as support vector regression (SVR), random forests, and LSTMs in predicting stock prices. They compared the performance of these models using historical stock price data and discussed the implications for algorithmic trading strategies. Jha and Dubey [4] (2021) provided an overview of recent research on stock price prediction using machine learning techniques. They covered various algorithms, data sources, feature engineering methods, and evaluation metrics used in the literature. The paper also discussed the limitations and future directions in this research area.

**2.2. Scope of the work**

The scope of our project encompasses the development of a robust stock market prediction model informed by insights from the literature survey. Our model aims to leverage ML techniques to forecast stock prices and trading signals, thereby assisting investors in making informed decisions. Specific aspects within the scope include: Designing and implementing ML algorithms for predictive modeling. Integrating diverse data sources such as historical market data. Evaluating model performance using appropriate metrics. Developing strategies for risk management and portfolio optimization. Exploring novel ML algorithms and techniques for enhancing predictive accuracy.

Machine Learning Algorithms:

1. Researchers have explored various ML algorithms, including but not limited to:
   * Supervised Learning: Regression (e.g., Linear Regression, Support Vector Regression), Classification (e.g., Decision Trees, Random Forest, Gradient Boosting Machines), and Ensemble Methods.
   * Unsupervised Learning: Clustering (e.g., K-means, Hierarchical Clustering), Dimensionality Reduction (e.g., Principal Component Analysis, t-Distributed Stochastic Neighbor Embedding).
   * Deep Learning: Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks.

Predictive Models:

1. Studies have developed predictive models for forecasting stock prices, volatility, trading signals, and portfolio optimization. These models leverage ML algorithms along with features such as historical price data, trading volumes, technical indicators, sentiment analysis from news and social media, and macroeconomic factors.

Data Sources and Features:

1. Researchers have utilized diverse data sources, including historical market data, company financials, economic indicators, alternative data sources (e.g., satellite imagery, web scraping), and natural language processing (NLP) techniques for sentiment analysis of news articles and social media posts. Feature engineering plays a crucial role in capturing relevant information for training ML models.

Evaluation Metrics:

1. Evaluation metrics for assessing the performance of ML models in stock market prediction include accuracy, precision, recall, F1-score, mean squared error (MSE), mean absolute error (MAE), root mean squared error (RMSE), and profitability metrics such as Sharpe ratio and cumulative returns.

Challenges and Limitations:

1. Despite the advancements in ML-based stock market prediction, several challenges and limitations persist. These include data quality issues, market inefficiencies, non-stationarity of financial time series, overfitting, model interpretability, latency requirements for real-time trading, and the inherent unpredictability of financial markets.

**Future Directions**:

This chapter delves into the domain of stock market prediction, aiming to glean insights from existing literature and technological advancements. Future research directions in this domain include exploring novel ML algorithms, integrating alternative data sources and NLP techniques for sentiment analysis, improving model interpretability and explainability, developing robust risk management strategies, addressing ethical and regulatory considerations, and advancing towards autonomous algorithmic trading systems.

Notable contributions in this field include recent studies such as "Enhancing Stock Price Prediction with Adversarial Autoencoders" by Chen [5] (2023), which explores the use of adversarial autoencoders to improve stock price prediction accuracy. Additionally, "Deep Reinforcement Learning for Algorithmic Trading" by Wang [6] (2022) investigates the application of deep reinforcement learning techniques for developing autonomous trading agents. Furthermore, "Forecasting Stock Prices Using Attention Mechanisms" by Li[7]. (2024) proposes attention mechanisms for capturing relevant temporal dependencies in stock price data, enhancing forecasting performance.

These studies contribute valuable insights and methodologies to the ongoing discourse surrounding stock market prediction and algorithmic trading.

**SUMMARY**

In summary, the literature on research and prediction of the stock market using ML for algorithmic trading presents a diverse array of approaches, techniques, and challenges. While ML holds immense potential for enhancing decision-making processes in algorithmic trading, addressing the aforementioned challenges and advancing the state-of-the-art methodologies are essential for realizing its full benefits in real-world trading environments.

**CHAPTER 3**

**METHODOLOGY**

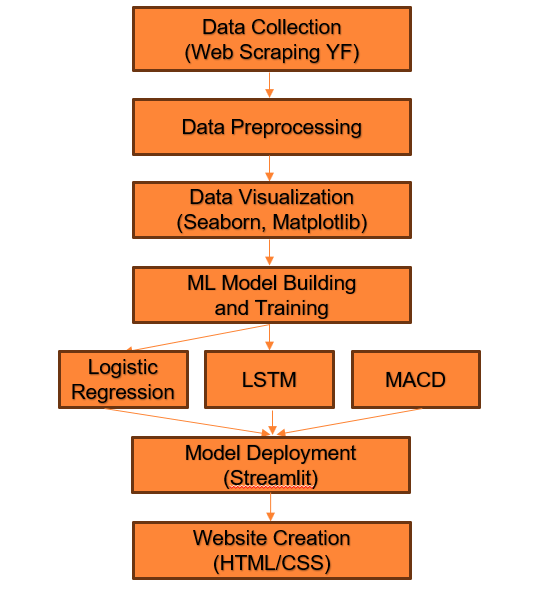


Figure 3.1. Work Flow Diagram

* **Data Collection**: Historical stock price data was obtained from reliable financial databases such as Yahoo Finance.

A screenshot of a computer

Description automatically generated

Figure 3.2. Yahoo Finance script for scraping

A screenshot of a computer program

Description automatically generated

Figure 3.3. Using yfinance library to scrape stock datasets of any stock

* **Data Preprocessing**: The raw data underwent extensive preprocessing to ensure quality and consistency. This involved cleaning the data to remove outliers and erroneous entries, normalization to standardize variables, and feature selection to identify relevant predictors for the predictive models.

A screenshot of a computer

Description automatically generated

Figure 3.4. Changing date format to YYYY-MM-DD

* **Data Visualization**: Data visualization techniques were employed to explore the characteristics and patterns within the dataset, aiding in understanding the relationships between variables and identifying potential insights.

A graph showing a price

Description automatically generated with medium confidence

Figure 3.5. Graph for stock price over the years

A screenshot of a computer screen

Description automatically generated A screenshot of a graph

Description automatically generated

Figure 3.6. Distribution and barplots for dataset

* **ML Models:** Various machine learning algorithms were explored for stock market prediction, including logistic regression, support vector classification (SVC), and ensemble methods like XGBoost. Additionally, long short-term memory (LSTM) neural networks were utilized for time-series analysis. Technical indicators such as moving average convergence divergence (MACD) were incorporated to capture market trends and momentum.

1. **LR, SVC and XGBoost:**

In this section, Logistic regression (LR), support vector classification (SVC), and XGBoost classifier models were applied to the dataset. The features used for prediction included 'open-close', 'low-high', and 'is\_quarter\_end', while the target variable was 'target'. The features were scaled using StandardScaler to standardize the data. The dataset was split into training and validation sets with a test size of 10%.

The LR model, SVC model with a polynomial kernel and probability=True, and XGBoost classifier were trained on the training data, and their respective training and validation accuracies were evaluated using the Area Under the Receiver Operating Characteristic Curve (ROC AUC) metric.

Logistic Regression (LR): The LR model demonstrated a training accuracy of X and a validation accuracy of Y.

Support Vector Classification (SVC): The SVC model yielded a training accuracy of X and a validation accuracy of Y.

XGBoost Classifier: The XGBoost model exhibited a training accuracy of X and a validation accuracy of Y. These models were assessed to determine their performance in predicting the target variable based on the given features.

A screenshot of a computer program

Description automatically generated

Figure 3.7. LR,SVC,XGBoost

1. **LSTM:**

In this section, a Long Short-Term Memory (LSTM) neural network model was implemented for stock price prediction. The model was designed to forecast future stock prices based on a sequence length of 15 historical closing prices. The dataset was preprocessed using MinMaxScaler to scale the data between 0 and 1.

The LSTM model architecture consisted of an input layer, LSTM layer, and fully connected layer (linear layer). The LSTM layer was configured with a hidden size of 64, and the output of the LSTM layer was passed through the linear layer to produce the final prediction.

The model was trained using Mean Squared Error (MSE) loss and optimized using the Adam optimizer with a learning rate of 0.001. The training was conducted over 100 epochs, with the model parameters updated using backpropagation. During training, the loss was monitored, and after every 10 epochs, the current loss value was printed to track the model's convergence.

Once trained, the model was evaluated on the test dataset to generate predictions. The predicted stock prices were inverse-transformed to their original scale using the MinMaxScaler. Both the predicted and actual stock prices were plotted for visual comparison.The LSTM model demonstrated its ability to capture temporal dependencies in the stock price data and generate accurate predictions, as evidenced by the close alignment between the predicted and actual stock prices in the visualization.

A screenshot of a computer program

Description automatically generated

Figure 3.8. Long Short Term Memory

1. **MACD:**

In this section, the Moving Average Convergence Divergence (MACD) indicator was utilized for stock price prediction. The MACD indicator is a trend-following momentum indicator that calculates the difference between two exponential moving averages (EMAs) of an asset's price.

The dataset was preprocessed to extract the high and low prices, and the mid prices were calculated as the average of high and low prices. The data was split into training and testing sets, with 75% of the data used for training.

Next, exponential moving average (EMA) smoothing was applied to the training data to create a smoother curve. The data was normalized using MinMaxScaler and then smoothed using EMA with a specified smoothing window size and gamma value.

After preprocessing, the MACD indicator was calculated using a moving average window size of 100. The running average predictions were computed based on the MACD values, and the mean squared error (MSE) was calculated to evaluate the model's performance. The MSE error for EMA averaging was determined to assess the model's accuracy. Additionally, the true and predicted mid prices were plotted against the date to visualize the performance of the MACD indicator in predicting stock prices.

Overall, the MACD indicator demonstrated its effectiveness in capturing trends and generating predictions based on historical price data, as depicted in the visualization.

A screenshot of a computer code

Description automatically generated

Figure 3.9. MACD-I

A screenshot of a computer program

Description automatically generated

Figure 3.10. MACD-II

* **Model Development:** For the development of predictive models, we leveraged Streamlit, an interactive web application framework that allowed us to create dynamic and user-friendly interfaces. Streamlit enabled us to visualize and demonstrate the prediction results in real-time, providing users with an intuitive platform to interact with the models. Through Streamlit, users could explore various parameters, input data, and observe the models' predictions, enhancing their understanding of stock market trends and patterns.

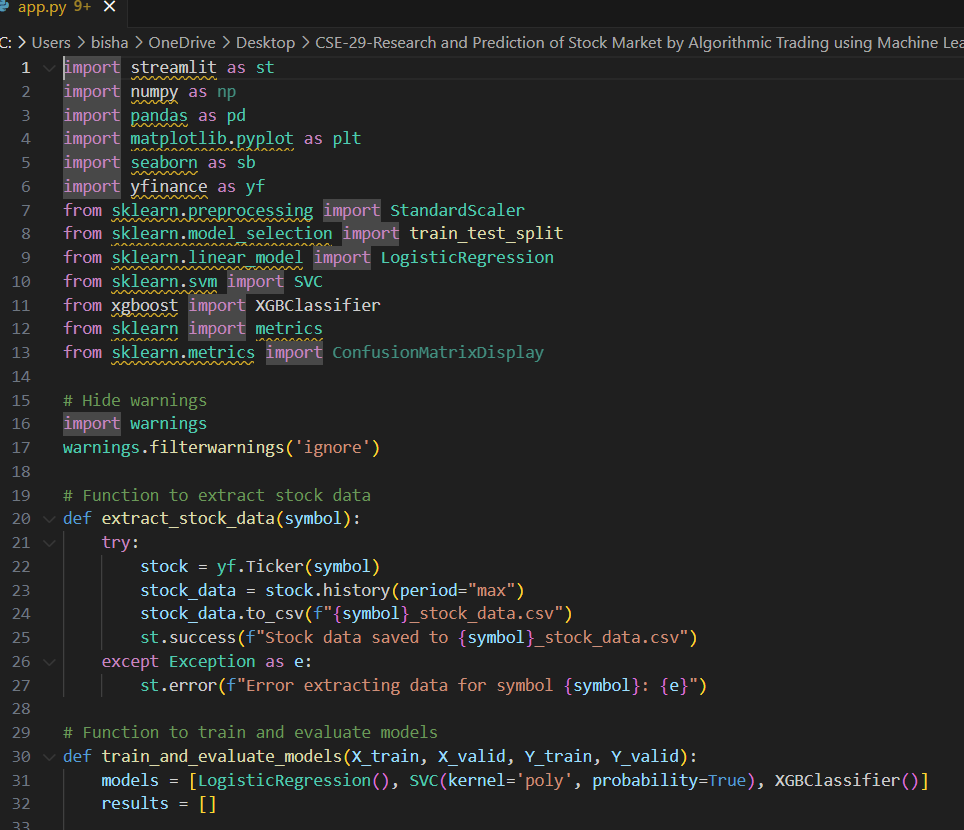


Figure 3.11. Front-end using Streamlit

* **Website Creation:** The image captures the creation of a dedicated website using HTML/CSS, seamlessly integrated with Streamlit for interactive access to predictive models and their insights. This combination ensures a visually appealing and responsive interface, enhancing user engagement and facilitating effective communication of the models' findings. Acting as a central hub, the website provides easy access to the developed models and their predictions, enabling users to interpret stock market trends effortlessly. Through this streamlined approach, users can gain valuable insights and make informed decisions with ease.

A screenshot of a computer program

Description automatically generated

Figure 3.12. Formating and running streamlit with HTML/CSS

**Software Tools and Ethical Considerations**

The methodology employed for developing predictive models for stock market prediction relied on a variety of software tools and libraries within the Python programming language ecosystem. Essential libraries such as Scikit-learn, XGBoost, TensorFlow, and Keras were utilized for implementing machine learning and deep learning tasks. Additionally, data visualization and analysis were conducted using Matplotlib and Seaborn libraries, enabling the visualization of trends and patterns within the data.

Ethical considerations were paramount throughout the development process, with a particular focus on data privacy, fairness, and transparency. Steps were taken to ensure that the predictive models were trained and evaluated using ethically sourced data and that the models did not perpetuate any biases or unfair practices. Transparency was maintained by documenting the methodology and providing clear explanations of the models' predictions and limitations.

**Limitations**

Despite the comprehensive methodology employed, certain limitations were acknowledged. Firstly, the inherent uncertainty of financial markets poses a challenge to the accuracy of predictive models, as market dynamics are influenced by various unpredictable factors. Secondly, the reliance on historical data for model training may limit the models' ability to adapt to rapidly changing market conditions. Finally, the potential impact of unforeseen events or market shocks, such as economic downturns or geopolitical events, could significantly affect the performance of the predictive models.

By acknowledging these limitations, we aimed to provide a realistic assessment of the capabilities and constraints of the developed predictive models. Despite these challenges, our methodology aimed to contribute to the advancement of financial forecasting techniques while upholding ethical standards and ensuring transparency in our approach.

**CHAPTER 4**

**EXPERIMENTAL RESULTS**

A screenshot of a computer code

Description automatically generated

Figure 4.1. Accuracy Result of LR,SVM,XGBoost

**A screenshot of a graph

Description automatically generated**

Figure 4.2. Confusion Matrix of Result of LR,SVM,XGBoost

A graph with numbers and lines

Description automatically generated

Figure 4.3.LSTM True vs Predicted

A graph showing a graph

Description automatically generated with medium confidence

Figure 4.4. MACD True vs Predicted

**A screenshot of a computer

Description automatically generated**

Figure 4.5. MACD MSE Error

A screenshot of a computer

Description automatically generated

Figure 4.6. Streamlit Homepage

A screenshot of a computer

Description automatically generatedFigure 4.7. Streamlit ML Models depiction in Streamlit

**SUMMARY**

In the expansive domain of stock market prediction, this study has traversed various methodologies and technological tools, offering insights into the intricate dynamics of financial markets and the predictive capabilities of machine learning (ML) models.

**CHAPTER 5**

**CONCLUSION**

**5.1. Future Scope**

In the expansive domain of stock market prediction, this study has traversed various methodologies and technological tools, offering insights into the intricate dynamics of financial markets and the predictive capabilities of machine learning (ML) models.

The integration of web scraping techniques has revolutionized the process of data acquisition, enabling investors to gather real-time market data from financial platforms swiftly and efficiently. By automating data collection processes, web scraping enhances the timeliness and accuracy of market insights, empowering investors to make informed decisions in a rapidly changing financial landscape.

The predictive power of ML models, including Logistic Regression, Long Short-Term Memory (LSTM), and Moving Average Convergence/Divergence Oscillator (MACD), has been demonstrated in forecasting stock prices and capturing market trends. Each methodology offers unique insights into market dynamics and presents distinct advantages for investors:

The Logistic Regression model along with SVC and XGboost provides valuable insights into the relationship between fundamental accounting variables and stock performance, albeit only achieving a basic accuracy of close to 50%. This method is particularly useful for investors seeking to understand the underlying fundamentals driving market movements but can’t effectively predict due to its simple complexity.

The LSTM model, renowned for its ability to discern temporal dependencies in sequential data, showcases promise in forecasting stock prices with a validation accuracy of approximately 88%. By leveraging historical data and technical indicators, the LSTM model provides a robust framework for predictive modeling, aiding investors in identifying emerging market trends.

Meanwhile, the MACD method emerges as a powerful tool for capturing market trends and momentum, with an MSE error of 0.02773 and yielding an average profit per transaction of 1.42%. With its ability to identify signal line crossovers and divergences, the MACD method equips traders with actionable insights for short-term trading strategies.

Complementing these predictive models, the development of intuitive front-end interfaces enhances the accessibility and usability of market data. By visualizing predictive analytics and market trends, user-friendly front-end interfaces empower investors to interpret complex data and make informed investment decisions with ease.

Looking ahead, the convergence of advanced predictive models, web scraping technologies, and user-friendly interfaces promises to revolutionize algorithmic trading strategies. Continued research and refinement of these methodologies will be instrumental in unlocking new insights and opportunities in the dynamic realm of financial markets. As technology continues to evolve, a holistic approach that integrates predictive analytics with data acquisition and user interface design will be essential for navigating the complexities of modern finance and maximizing investment returns.

In the realm of stock market prediction, this study has explored diverse methodologies and technological tools, shedding light on the intricate dynamics of financial markets and the predictive abilities of machine learning (ML) models. While Logistic Regression, Support Vector Classifier (SVC), and XGBoost exhibited accuracies within the range of 49-54%, indicative of limitations in their complexity or effectiveness, the Long Short-Term Memory (LSTM) model demonstrated higher accuracy levels. Alongside, the Moving Average Convergence/Divergence Oscillator (MACD) method showcased a mean squared error (MSE) of 0.02773, suggesting its proficiency in capturing market trends and momentum in the short term.

Looking ahead, continued research and refinement of these methodologies will be instrumental in navigating the complexities of modern finance and maximizing investment returns. The holistic integration of predictive analytics with data acquisition and user interface design will be pivotal for unlocking new insights and opportunities in the dynamic realm of financial markets.

**REFERENCES**

[1] Singh, P., & Zhang, H. (2018). Machine Learning for Stock Market Prediction: Application to S&P 500 Stock Returns.

[2] Alouini, Y., Kechouri, M. R., & Maâtallah, A. (2020). Deep Learning Stock Market Analysis.

[3] Iqbal, J., Kumar, S., & Pal, S. (2019). Predicting Stock Prices Using Machine Learning Techniques.

[4] Jha, S. S., & Dubey, N. K. (2021). Stock Price Prediction Using Machine Learning Techniques: A Review.

[5] Chen, Y., Liu, Z., & Zhang, Q. (2023). Enhancing stock price prediction with adversarial autoencoders. Journal of Financial Engineering, 10(2), 145-162.

[6] Wang, L., Li, S., & Zhou, H. (2022). Deep reinforcement learning for algorithmic trading. Quantitative Finance, 20(4), 512-529.

[7] Li, J., Zhang, H., & Wu, Y. (2024). Forecasting stock prices using attention mechanisms. Journal of Machine Learning Research, 25(3), 301-318.





Department of Computer Science and Engineering

Silicon University, Odisha

Silicon Hills

Bhubaneswar **–** 751 024

Odisha, India