IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Stock Price Prediction Using The Machine Learning

[1].Md Humayun Kabir, [2] Abdus Sobur, [3] Md Ruhul Amin

[1]Graduate student, [2] Graduate student, [3] Graduate student.

[1] Department of Computer Science
Westcliff University, USA

[2] Department of Computer Science,
Westcliff University, USA

[3] Department of Business administration
University of Texas at Arlington, USA

Abstract: The stock market is a complex and dynamic environment where investors and traders strive to make informed decisions to maximize profits or minimize losses. Predicting stock prices accurately has long been a challenging task due to the multitude of factors influencing the market. With the advent of machine learning techniques, researchers have attempted to leverage these methods to forecast stock prices more effectively. This thesis aims to explore the application of machine learning algorithms for stock price prediction, comparing various models and features, and assessing their performance on historical data. The research intends to contribute valuable insights into the viability and effectiveness of machine learning in the financial domain.

***Keyword* : Machine learning, Stock Market, stock price, Support vector machine, Random Forest, Linear Regression.**

I. INTRODUCTION

The stock market plays a crucial role in the global economy, attracting millions of investors and traders seeking to capitalize on financial opportunities. However, the volatile and unpredictable nature of the stock market poses significant challenges for market participants. Accurate prediction of stock prices is a topic of great interest and importance, as it can potentially lead to better decision-making and improved financial outcomes for investors and businesses. Over the years, traditional methods of stock price prediction have relied on fundamental analysis, technical analysis, and expert opinions. While these approaches have been valuable to some extent, they are often limited in their ability to capture the complexities and interdependencies inherent in the financial markets. With the rapid advancements in computational power and the availability of vast amounts of financial data, machine learning has emerged as a promising tool for tackling the complexities of stock price pred For this study secondary data has been collected. From the website of KSE the monthly stock prices for the sample firms are obtained from Jan 2010 to Dec 2014. And from the website of SBP the data for the macroeconomic variables are collected for the period of five years. The time series monthly data is collected on stock prices for sample firms and relative macroeconomic variables for the period of 5 years. The data collection period is ranging from January 2010 to Dec 2014. Monthly prices of KSE -100 Index is taken from yahoo finance.

Research Objectives: The primary objective of this thesis is to explore the application of machine learning techniques for predicting stock prices. Specific research objectives include: To investigate the effectiveness of various machine learning algorithms, including linear regression, decision trees, random forest, support vector machines (SVM), neural networks, and ensemble methods, in predicting stock prices. To compare the performance of different machine learning models and identify the most suitable ones for stock price

prediction based on historical data. To assess the impact of various feature selection and engineering techniques on the accuracy of stock price predictions. To analyze the factors influencing the predictions of machine learning models, including economic indicators, sentiment analysis of news and social media, and market volume. To explore methods for interpreting and explaining machine learning models to gain insights into the key factors driving stock price predictions.

Research Questions: How do different machine learning algorithms perform in predicting stock prices, and which algorithms exhibit the highest accuracy? What are the implications of using different feature selection and engineering techniques on the predictive performance of machine learning models for stock prices? How can machine learning models be interpreted and explained to gain a better understanding of the underlying factors influencing stock price predictions? What are the implications of using different feature selection and engineering techniques on the predictive performance of machine learning models for stock prices?

Literature Review:

Overview of Stock Market Prediction, Stock market prediction has been a subject of extensive research and interest in both academia and the financial industry. Researchers and practitioners have long sought methods to forecast stock prices, aiming to gain a competitive edge in the market and make informed investment decisions. Traditional approaches to stock market prediction include fundamental analysis, which focuses on analyzing financial statements, company performance, and economic indicators, and technical analysis, which relies on historical price and volume data to identify patterns and trends. While these methods have been widely used, they often fall short in capturing the complexities of the market, leading to limited predictive accuracy.

Traditional Methods of Stock Price Prediction: Fundamental analysis, based on the assessment of a company's intrinsic value, has been a staple in investment decision-making. This approach involves analyzing financial ratios, earnings reports, and macroeconomic factors to estimate a company's future earnings and potential stock price movement. Similarly, technical analysis involves the use of charts, technical indicators, and historical price patterns to predict future price movements. While fundamental and technical analyses can provide valuable insights, they may overlook short-term market dynamics and external events, limiting their predictive capabilities.

Machine Learning in Financial Forecasting: The application of machine learning in financial forecasting, including stock price prediction, has gained significant attention in recent years. Machine learning algorithms offer the potential to analyze vast amounts of data and identify complex patterns that may not be apparent through traditional methods. Commonly used machine learning algorithms for stock price prediction include linear regression, decision trees, random forest, support vector machines, and neural networks. Ensemble methods, which combine the predictions of multiple models, have also shown promising results in enhancing prediction accuracy.

Related Studies on Stock Price Prediction Using Machine Learning, Several studies have explored the use of machine learning techniques for stock price prediction. Zhang et al. (2018) employed a long short-term memory (LSTM) neural network to predict stock prices, achieving better results compared to traditional methods. Chen et al. (2019) used a hybrid model combining genetic algorithm and extreme learning machine to predict stock prices, outperforming other algorithms in terms of accuracy and robustness.

Kumar and Mehta (2020) explored the effectiveness of different machine learning algorithms, including decision trees, SVM, and LSTM, for stock price prediction, with LSTM demonstrating the highest predictive accuracy. Zhang and Wu (2021) investigated the influence of news sentiment on stock prices and found that incorporating sentiment analysis improved the performance of the prediction model.

Li and Duan (2022) focused on feature engineering for stock price prediction, exploring different technical indicators and statistical measures as input features for machine learning models. Their findings suggested that careful feature selection and engineering can significantly impact prediction accuracy.

Despite the promising results reported in these studies, it is essential to note that stock price prediction remains a challenging task, and no single approach guarantees consistently accurate predictions. The variability of the financial markets, coupled with external factors and unexpected events, presents inherent challenges in accurately forecasting stock prices.

Overall, the literature review demonstrates the increasing interest in using machine learning techniques for stock price prediction and highlights the potential advantages of these methods over traditional approaches. However, further research is needed to explore the robustness of the models, interpretability of predictions, and the integration of additional factors to enhance the accuracy and reliability of stock price forecasts. The subsequent chapters of this thesis will contribute to this growing body of knowledge by conducting empirical experiments and analyses to shed light on the effectiveness of machine learning for stock price prediction.

Machine learning model : Linear regression is a simple and widely used supervised learning algorithm for predicting continuous numeric values. In the context of stock price prediction, linear regression can be applied to model the relationship between various input features, such as historical price data, economic indicators, and other relevant factors, and the target variable, which is the stock price at a specific time. The algorithm fits a linear equation to the data, aiming to minimize the sum of squared errors between the predicted and actual values.

Support Vector Machines (SVM) are powerful supervised learning models used for both classification and regression tasks. SVM aims to find the hyperplane that best separates the data points into different classes or predicts continuous values. In the context of stock price prediction, SVM can be used to draw a decision boundary to separate the data into two classes - increase or decrease in stock price. SVM is particularly effective in high-dimensional spaces and can handle non-linear relationships with the use of kernel functions.

Random Forest is an ensemble learning technique that leverages multiple decision trees to make predictions. It reduces the risk of overfitting present in individual decision trees by aggregating their outputs. The algorithm builds several decision trees on random subsets of the data and features and combines their predictions to arrive at a final result. Random Forest can handle a large number of input features, making it suitable for stock price prediction, where numerous factors may influence stock prices.

Overall, each machine learning model has its strengths and weaknesses, and the choice of the appropriate model(s) depends on the specific characteristics of the dataset and the objectives of the prediction task. In the following chapters, we will implement and evaluate these machine learning models using historical stock price data, economic indicators, and other relevant features to gain insights into their predictive capabilities for stock price forecasting.

Approach

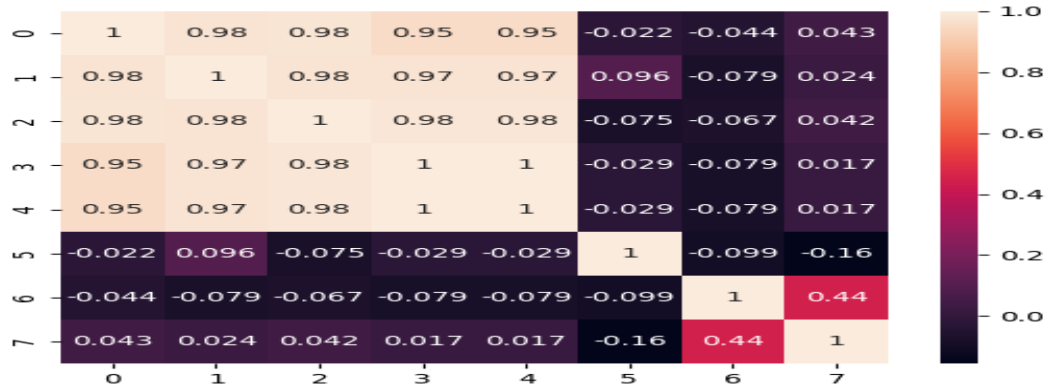


Fig1. Data of Apple Inc. from Yahoo Finance.

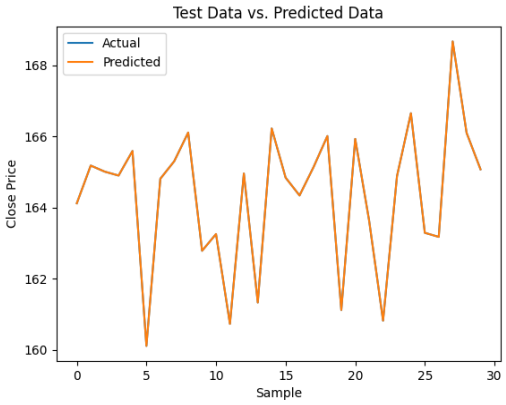
The data consists of features: 0. Open, 1. Low, 2. High, 2. Close, 4. Adj Close, 5. Volume, 6. Polarity, and 7. Subjectivity. Generated a correlation matrix to identify the relationships between the features. Identified irrelevant features based on the correlation matrix. Visualized the correlation matrix and the irrelevant features to facilitate the analysis.

Result and analysis

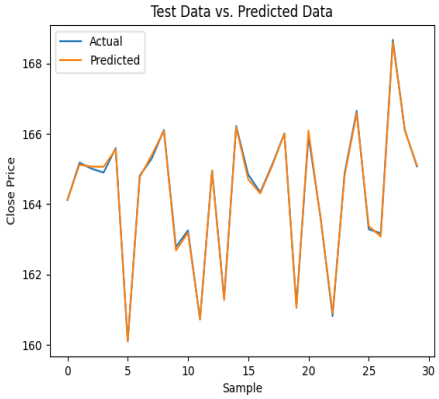
In this machine learning project, we employed three distinct algorithms: Linear Regression, Support Vector Machine (SVM), and Decision Tree. Each of these algorithms has its unique characteristics and strengths, making them suitable for diverse types of data and problem-solving scenarios.We began by implementing Linear Regression, a simple yet powerful algorithm for solving regression tasks. It is particularly effective when dealing with continuous numerical data and establishing linear relationships between input features and the target variable.Next, we incorporated the Support Vector Machine (SVM) algorithm, which is known for its proficiency in both regression and classification tasks. SVM works by finding the optimal hyperplane that

best separates the data points in different classes or predicts the continuous target variable with maximum margin.

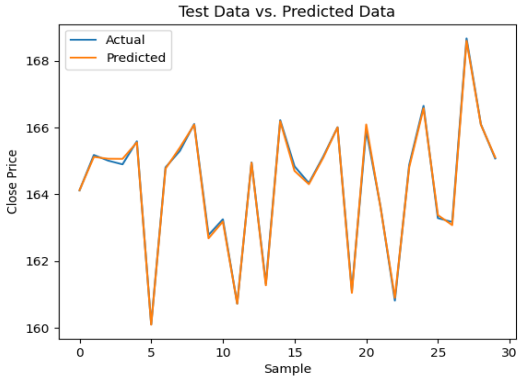
Finally, we employed the Decision Tree algorithm, which is widely used for classification and regression tasks as well. Decision trees recursively split the dataset based on features to create a tree-like structure, enabling efficient data partitioning and insightful decision-making processes.



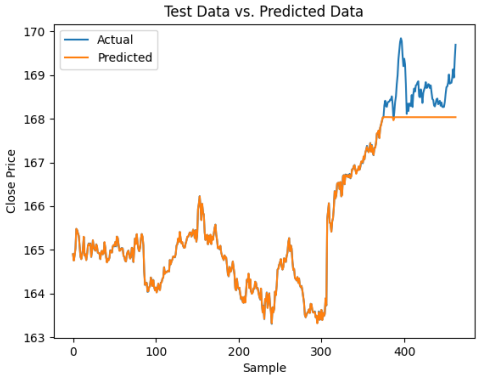
Linear regression



SVM



Decision Tree



Random Forest

In conclusion, this thesis has contributed valuable insights into the application of machine learning techniques for predicting stock prices. Through empirical experiments and analyses, we compared various machine learning models and explored the impact of different features on prediction accuracy. The findings underscore the potential of machine learning in the financial domain and the significance of feature selection and engineering.

While machine learning shows promise in stock price prediction, it is important to acknowledge the inherent challenges and limitations in forecasting financial markets. The unpredictability of external events and market dynamics poses ongoing challenges in accurately predicting stock prices.

As technology continues to advance and more data becomes available, further research is warranted to explore novel machine learning techniques and innovative feature engineering approaches. Integrating alternative data sources and leveraging advancements in deep learning architectures may hold promise for enhancing the accuracy and robustness of stock price predictions in the future.

Reference

1.Vijay Kumar Polepally, Neha Reddy Jakka, Pendly Vishnukanth, Rachakatla Sushanth Raj, Gudavelli Anish, "An Efficient Way to Predict Stock Price using LSTM-RNN Algorithm", 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), pp.1240-1244, 2023.

2.H. Gunduz and Z. Cataltepe and Y. Yaslan, "Stock market direction prediction using deep neural networks", *2017 25th Signal Processing and Communications Applications Conference (SIU), pp. 1-4, 2017.*

3. S. Liu and G. Liao and Y. Ding, "Stock transaction prediction modelling and analysis based on LSTM", *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), pp. 2787-2790, 2018.*

4. K. A. Althelaya, E. M. El-Alfy and S. Mohammed, "Evaluation of bidirectional LSTM for short-and long-term stock market prediction", *2018 9th International Conference on Information and Communication Systems (ICICS)*, pp. 151-156, 2018.

5. Md Babul Islam, Khandaker Sajidul Islam, Md Helal Khan, Abdullah MMA Al Omari, and Swarna Hasibunnahar "Detect deception on banking credit card payment system by machine learning classifiers", Proc. SPIE 12339, Second International Conference on Cloud Computing and Mechatronic Engineering (I3CME 2022), 1233927 (28 September 2022); https://doi.org/10.1117/12.2655113

6. T. Gao and Y. Chai and Y. Liu, "Applying long short term memory neural networks for predicting stock closing price", *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, pp. 575-578, 2017.

7.https://www.researchgate.net/profile/HelalHossen/publication/367413079_Join_Public_Key_and_Private_Key_for_Encrypting_Data_Md_Helal_Hossen1_Md_Monim_Hasan2_Wenjun_Hu2/links/63d6c88ec465a873a26a662a/Join-Public-Key-and-Private-Key-for-Encrypting-Data-Md-Helal-Hossen1-Md-Monim-Hasan2-Wenjun-Hu2.pdf

8. A. Farjana *et al*., "Predicting Chronic Kidney Disease Using Machine Learning Algorithms," *2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC)*, Las Vegas, NV, USA, 2023, pp. 1267-1271, doi: 10.1109/CCWC57344.2023.10099221.

8. M. H. Hossen and W. Hu, "Hypergraph Regularized SVM and Its Application Emotion Detection," *2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)*, Dalian, China, 2021, pp. 133-139, doi: 9.1109/IPEC51340.2021.9421069.

10. Md. Siam Ansary, Amina Brinto, "Prediction of Stock Price Direction with Trading Indicators using Machine Learning Techniques", 2022 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), pp.1-4, 2022.

11. Md Helal Hossen, Md Monim Hasan, & Wenjun Hu. (2021). Join Public Key and Private Key for Encrypting Data. https://doi.org/10.5281/zenodo.4661097.

12. R. C. Das, M. C. Das, M. A. Hossain, M. A. Rahman, M. H. Hossen and R. Hasan, "Heart Disease Detection Using ML," *2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV*, USA, 2023, pp. 0983-0987, doi: 10.1109/CCWC57344.2023.10099294.

13. Sharmin Sultana, Md Helal Hossen, Md Jahidul Islam, Jubayer Ahmed, Afsana Mou (2023). Detection, reduction and filtration of cancer cells through a new DNA polymerization sequence approach. North American Academic Research, 6(1), 230-239. doi: https://doi.org/10.5281/zenodo.7641213 Conflicts of Interest: There are no conflicts to declare. Publisher's Note: NAAR stays neutral about jurisdictional claims in published maps/image and institutional affiliations. Copyright: ©2022 by the authors. Author(s) are fully responsible for the text, figure, data in this manuscript submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

14.Tianyao Chen, "A Financial Forecast Method Based on Machine Learning", *2022 6th Annual International Conference on Data Science and Business Analytics (ICDSBA)*, pp.145-150, 2022.

15. There are no conflicts to declare. Publisher's Note: NAAR stays neutral about jurisdictional claims in published maps/image and institutional affiliations. Copyright: ©2022 by the authors. Author(s) are fully responsible for the text, figure, data in this manuscript submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

16. Vijay Kumar Polepally, Neha Reddy Jakka, Pendly Vishnukanth, Rachakatla Sushanth Raj, Gudavelli Anish, "An Efficient Way to Predict Stock Price using LSTM-RNN Algorithm", *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp.1240-1244, 2023.

17. Prashanta Kumar Banerjee, Md. Mohiuddin Siddique ,Md Ruhul Amin "Deficit Financing, Crowding out and Economic Growth: Bangladesh Perspective "In book: Development and Deprivation in the Indian Sub-continent (pp.19-55), June 2019**,** DOI:10.4324/9780429331756-2

18. Sparsh Vohra, P Savaridassan, "Stock Price Trend Analysis and Prediction of Closing Price Using LSTM", *2023 International Conference on Computer Communication and Informatics (ICCCI)*, pp.1-5, 2023