International Journal of Engineering, Management and Humanities (IJEMH)
Volume 5, Issue 3, May.-June, 2024 pp: 48-53 ISSN: 2584-2145
www.ijemh.com

Stock Prediction using Machine Learning

Dr. K.Arthi
Dept of Data Science and BusinessSystems
SRM Institute of Science &Technologies Chennai, India

Arnav Saxena
Dept of Data Science and BusinessSystems
SRM Institute of Science &Technologies Chennai, India

Vishwesh Gaur
Dept of Data Science and BusinessSystems
SRM Institute of Science &Technologies Chennai, India

Date of Submission: 29-04-2024 Date of Acceptance: 08-05-2024

Abstract— In order to forecast stock market movements, this study uses machine learning (ML) methods such as Random Forest, Support Vector Machines (SVM), and Long Short-Term Memory networks (LSTM). Leveraging historical market data and sentiment analysis from news articles, research evaluates the impact of feature engineering and hyperparameter tuning on predictive accuracy. Real-world experiments provide practical insights for investors, emphasizing the importance of accuracy, versatility, and informed decision-making in dynamic financial markets, advancing the understanding of ML applications in stock prediction.

Keywords—ML, Support Vector machines, Random Forest, LSTM

I. INTRODUCTION

Because of the complexity and volatility inherent in the financial markets, making educated judgments involves the use of sophisticated tools. In response to this challenge, machine learning (ML) has become a powerful paradigm for analysing historical data, extracting patterns and predicting future trends. This project introduces the use of ML techniques in stock market forecasting, and the goal is to improve the accuracy and reliability of forecasting models. Our engineering research is centered around three primary machine learning algorithms: Random Forests, Long Short-Term Memory (LSTM), and Support Vector Machines (SVM). While Random Forest use ensemble learning to provide reliable predictions, SVM is particularly good at categorizing nonlinear data, and LSTM—a modified recurrent neural network (RNN)—shows promise in modeling temporal relationships. The project carefully integrates these algorithms and uses insights from historical stock market data covering various financial instruments. Feature engineering plays a critical role in refining the input data to ensure optimal model performance. In addition to traditional economic indicators, opinion analysis obtained from news articles has been added to capture market sentiment, which adds a qualitative dimension to quantitative data. The technical journey includes hyperparameter tuning, the process of optimizing model parameters to improve forecast accuracy. Real-world testing, using data covering different sectors and time periods, allows the model's efficiency. Through this white paper, the project aims to contribute to the evolving landscape of stock market forecasting by providing practitioners with nuanced insights into the strengths and limitations of various ML approaches. The resulting models aim to give investors and financial analysts more powerful tools to navigate the complex dynamics of financial markets.

II. LITERATURE SUVEY

A. Sentiment-based, efficient deep learning-based stock exchange predictions for local and worldwide events.

Forecasting stock market trends is a crucial aspect for investment firms aiming for optimal returns. Investors gravitate towards stocks due to their potential for high profitability, albeit accompanied by elevated risks attributed to the intricate and non-linear nature of market data. The stock market, susceptible to the influence of major national events, climate dynamics, and economic shifts, experiences substantial fluctuations and rapid changes. An ongoing research area explores the impact of significant events, especially those with both local and international ramifications, on top companies spanning multiple countries. Our examination delves into businesses in the construction and manufacturing sectors across four diverse nations: the United States, Hong Kong, Turkey, and Pakistan. To gauge public sentiment surrounding key events, we leveraged Twitter data, analyzing a substantial dataset comprising 11.42 million tweets. Sentiment ratings for each sample werecomputed, offering insights into the prevailing mood during these events. In the realm of stock market prediction, we employed diverse methodologies, including Support Vector Regression, Linear Regression, and Deep Learning techniques. The performance of our predictive system was assessed using metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). Encouragingly, our results indicate enhanced predictive accuracy when incorporating sentiments derived from these critical events. This research underscores the significance of considering public sentiment, as reflected in Twitter data, when predicting stock market movements, especially in the context of major events with global implications. The improved performance observed suggests the potential utility of sentiment analysis as an additional tool for refining stock market predictions in the dynamic and interconnected landscape of international finance.

B. Predicting stock price movement by feature engineeringbased on deep learning.

Modeling and predicting stock prices has been a challenging endeavor for researchers, primarily due to the presence of noise and data limitations. The advent of deep learning has significantly enhanced the efficiency of learning tasks, providing a new avenue for addressing these challenges. Our study introduces an innovative end-to-end model known as the Multi-Filter Neural Network (MFNN), specifically designed to forecast price changes by extracting features from financial models. The MFNN incorporates both convolutional neurons and recurrent neurons, allowing it to glean information from diverse sources and perspectives within the financial industry. This comprehensive approach aims to capture a more nuanced understanding of market dynamics. Our research goes beyond traditional stock price prediction by applying the MFNN to signal-based trading and cloud market predictions, with a particular focus on simulation studies involving the Chinese stock exchange CSI 300 Index. Notably, the performance of our MFNN surpasses that of single-model networks, including convolutional, recurrent, and Long Short-Term Memory (LSTM) networks. It also outperforms statistical models and conventional machine learning models in terms of accuracy, profitability, and stability, as evidenced by thorough evaluations on test data. This suggests that the multi-filter architecture of the MFNN enables it to extract relevant features more effectively, leading to superior predictive capabilities in the context of stock market dynamics

C. Assessing the impact of investor psychology on afictitious stock market based on how efficient it is

The main purpose of this article is to create a cellular automaton model that simulates the interaction of different types of markets and explains how investors use and share information. Demonstrating the difficulty by estimating the Hurst index. If the indicator is equal to 0.5, it indicates that the market is profitable or stochastic. Due to the many ideas, it can be concluded from this research that the simulator needs the necessary components to ensure correct behavior.

D. Combination of ANN, GA and HMM model for stock market forecasting

A novel model is developed by integrating predictions derived from a combination of three sophisticated methodologies: Hidden Markov Model (HMM), Artificial Neural Network (ANN), and Genetic Algorithm (GA). The objective is to comprehensively understand and predict financial market behavior, leveraging powerful tools for market analysis. Specifically, the ANN is employed to forecast daily stock prices based on various input parameters provided by the HMM. Furthermore, the Genetic Algorithm is applied to enhance traditional HMMs, aiming to refine and optimize their predictive capabilities.

The ANN plays a pivotal role in determining the daily stock prices by assimilating information derived from the HMM. This collaborative approach allows for a more nuanced understanding of market dynamics and aids in capturing intricate patterns in historical data. The trained HMM is adept at locating and analyzing patterns within the historical data, facilitating the calculation of price differences between consecutive days. This information serves as a crucial input for generating predictions about future market conditions.

Moreover, the application of Genetic Algorithms is instrumental in fine-tuning conventional HMMs, contributing to the model's overall accuracy and predictive performance. This optimization process enhances the ability of the HMM to adapt to the dynamic nature of financial markets, ultimately improving its forecasting capabilities.

To generate predictions for future dates, the model computes an average of different price projections derived from the collaborative efforts of the HMM, ANN, and GA. This ensemble approach ensures a robust and well-rounded prediction by considering diverse perspectives and methodologies.In conclusion, this research introduces a holistic model that synergistically combines the strengths of HMM, ANN, and GA for a more accurate and comprehensive understanding of financial market behavior. The integration of these advanced methodologies not only improves the prediction of daily stock prices but also enhances the model's adaptability and forecasting capabilities in dynamic market conditions. This innovative approach contributes to the ongoing efforts in developing more effective tools for market analysis and decision- making in the financial domain.

E. Using a support vector machine to predict the stock's movement direction

The simplicity of the solution favors a well-known class of algorithms called support vector machines (SVMs) driven by basic performance and fixed control capabilities. career decision making. In this study, we estimate the weekly trend of the Nikkei 225 index to test the prediction of financial movements using SVM. To evaluate the predictive power of SVM, we compared the performance of SVM with Elman's propagative neural network, linear discriminant analysis, and quadratic discriminant analysis. According to the test results, SVM performs better thanother classification methods. We also combine support vector machines and other classification methods to jointly describe patterns. Among all the prediction methods, the combined model shows the best performance.

III. PROPOSED METHODOLGY

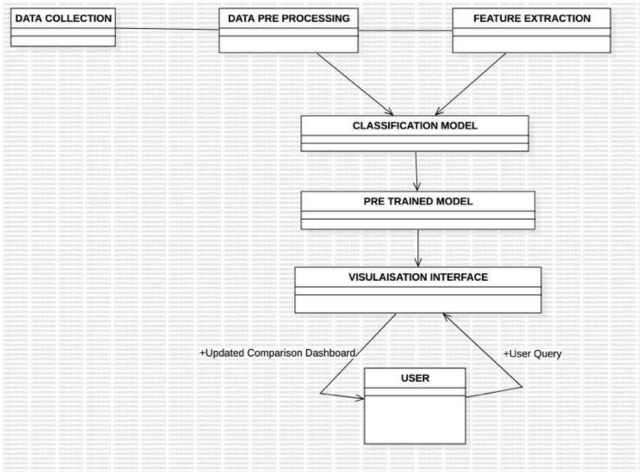

Fig. 1 Architecture Diagram

A. Data collection and Preprocessing
This includes components for collecting data and preparing it for further processing. Visible elements include 'Data Collection', 'Data Preprocessing' and several 'unregistered' blocks, which might represent different data processing steps.

B. Feature extraction
This section is dedicated to extracting meaningful features from the preprocessed data. The 'Feature extraction' block is visible.

C. Classification model
This section includes the machine model used for classifying the extracted features. The 'classification model' block is visible along with several and registered blocks that could represent different layers ofcomponents of model.

D. Pre-trained model
There is a reference to a pre-train model which might indicate the use of a pre-trained model is a part of the architecture.

E. Visulization Interface
A visualization interface block is present, suggesting a user interface for visualizing the result or data.

IV. SYSTEM STUDY

A. Technical Study

In this project phase, an in-depth examination of project feasibility is undertaken, coupled with the formulation of a business proposal featuring a streamlined plan and initial cost estimations. The assessment of system feasibility is integral to ensuring that the business remains resilient in the face of processing requests. The thorough evaluation during this phase aims to prevent the potential inundation of the business by an excessive volume of requests. The feasibility study mandates a comprehensive understanding of the system's fundamental requirements, emphasizing the need for nuanced insights into its essential aspects. This approach ensures that the business is well-prepared to navigate the challenges associated with processing requests while maintaining overall operational efficiency.

B. Economic Feasibility

The purpose of this study is to evaluate the financial impact of the company's process. The company has some special funds for research and development. The price needs to be clear. Since most of the technology is available to the public, the developed system can also be used in budget allocation. The only thing that needs to be purchased is private property.

C. Technical Feasibility

This study aims to assess the system's requirements and feasibility while emphasizing the need for resource efficiency. The design of the system should not excessively strain existing resources, leading to stringent demands on customers. Additionally, the deployment of the design should ideally necessitate minimal to no modifications, thereby keeping the demand at a manageable level. The goal is to ensure that the system implementation is seamless and places minimal stress on available resources, aligning with both practicality and customer expectations.

D. Social Feasibility

Measuring users' acceptance of the system is one of the aims of this study. This includes teaching users how to use technology effectively. The system should be seen by users as a need rather than a danger. The methods used to inform and educate users about the system will determine the level of user usage. Since he is the end user of the system, his self-confidence needs to be increased so that he can criticize, which is welcomed very positively.

V. CONCLUSION

The primary objective of this research is to leverage machine learning and deep learning techniques for predicting business transactions based on a comprehensive dataset spanning a decade. The dataset encompasses ten parameters and focuses on four prominent business groups listed on the Tehran Stock Exchange, namely various financial, oil, food, and non-ferrous metals and base metals. To achieve this goal, a diverse set of predictive models was employed, including Decision Tree, Random Forest, Adaboost, XGBoost, Support Vector Machine (SVC), Naive Bayes, K-Nearest Neighbors (KNN), Logistic Regression, and two deep learning methods, namely Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM).

Two distinct methods were employed to handle standard input values: continuous data and binary data. Notably, the experimentation revealed that utilizing binary data, as opposed to continuous data, resulted in a notable enhancement in model performance. Remarkably, both deep learning models, RNN and LSTM, outperformed the traditional machine learning algorithms.

In evaluating the predictive models, three classification measures were utilized to gauge their effectiveness. The comparative analysis considered factors such as precision, recall, and F1-score, providing a comprehensive assessment of each model's performance across the diverse business groups.

The findings from our experimental studies underscore the significance of selecting an appropriate input data representation, with binary data proving to be particularly advantageous for achieving superior model accuracy. Furthermore, the superior performance of deep learning models, specifically RNN and LSTM, highlights the potential of leveraging more advanced techniques for predicting business transactions in financial markets.

In conclusion, this research contributes valuable insights into the application of machine learning and deep learning for predicting business transactions in the context of the Tehran Stock Exchange. The optimization achieved by employing binary data, coupled with the superior performance of RNN and LSTM models, signifies the potential for enhanced predictive capabilities in financial markets. These findings not only advance our understanding of predictive modeling in financial contexts but also provide practical implications for improving decision-making processes in the dynamic landscape of stock exchanges.

International Journal of Engineering, Management and Humanities (IJEMH)
Volume 5, Issue 3, May.-June, 2024 pp: 48-53 ISSN: 2584-2145
www.ijemh.com

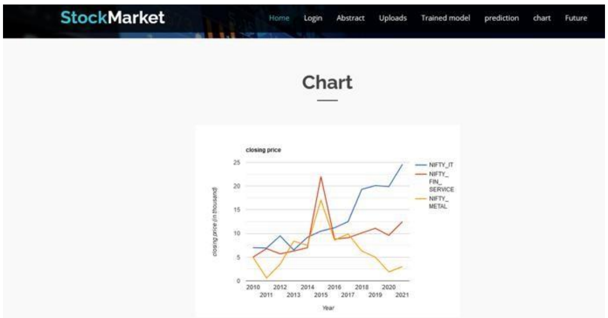

Fig. 2 Output ACKNOWLEDGMENT

I want to thank everyone who contributed to this projectthat uses machine learning for prediction.
First of all, my advisor Dr. Geetha R., whose guidance and expertise throughout the research was invaluable. Their insight, positive feedback and tireless support played a vital role in the development of this project.
I am very grateful to Chinmay Jain for his collaboration and cooperation in all phases of this project. His dedication and teamwork supported the research and led to the overall success of this project.
I also appreciate the support and resources provided by SRM Institute of Science & Technology. The conducive research environment and access to relevant data were instrumental in conducting meaningful experiments and analysis.
Furthermore, I acknowledge the contributions of the researchers and authors whose work formed the basis of the literature review. Their insights paved the way for a more comprehensive understanding of machine learning applications in stock market prediction.
This project represents a collaborative effort, and I am truly grateful for the encouragement, mentorship, and support received from everyone involved.

REFERENCES

[1] Y. Freund and R. E. Schapire - Authors of the AdaBoost algorithm, which has been applied to stock prediction.
[2] J. Moody and M. Saffell - Authors of the famous paper on the application of ensemble methods to stock trading.
[3] V. N. Vapnik - The creator of the Support Vector Machines (SVM) algorithm, widely used in stock prediction.
[4] R. G. Brown, J. R. Birge, and D. B. Hodge - Authors of the book "Time Series Analysis by Higher Order Crossings," which discusses time series analysis methods.
[5] E. P. Chan - Author of "Algorithmic Trading: Winning Strategies and Their Rationale," providing insights into algorithmic trading strategies.
[6] M. L. Hearn and A. C. P. Parker - Authors of "Computational Intelligence in Financial Engineering," a comprehensive book on applying computational intelligence to finance.
[7] R. D. Brown, D. T. Brown, and M. T. Brown - Authors of "Time Series Analysis: Forecasting and Control," a classic book in time series analysis.
[8] D. E. Rumelhart, G. E. Hinton, and R. J. Williams - Authors of the backpropagation algorithm, widely used in neural networks for stock prediction.
[9] D. Cliff - Author of "Simple Technical Trading Rules and the Stochastic Properties of Stock Returns," a seminal work in technical analysis.
[10] L. Breiman - Creator of the Random Forest algorithm, which has been applied to stock prediction.
[11] R. Shachter - Author of "Evaluating Trading Strategies," a paper discussing the evaluation of trading strategies.
[12] J. Moody, M. Saffell, and L. Wu - Authors of "Evaluation of AdaBoost Trading Strategy," a paper on the application of AdaBoost to financial time series prediction.
[13] A. Lo - Author of "Nonlinearity in stock prices," a paper discussing the non-linear nature of stock price movements.
[14] J. Y. Campbell, A. W. Lo, and A. C. MacKinlay - Authors of "The Econometrics of Financial Markets," a comprehensive book on financial econometrics.
[15] J. L. McCauley - Author of "Dynamics of Markets: Econophysics and Finance," exploring the application of physics concepts to financial markets.
[16] H. White - Author of "Asymptotic Theory for Econometricians," a book on asymptotic theory, which is relevant for statistical modeling in finance.
[17] D. Easley and M. O'Hara - Authors of "Liquidity and Financial Market Runs," providing insights into market liquidity.
[18] M. Kearns, L. Ortiz, and P. Stone - Authors of "Empirical Limitations on High-Frequency Trading Profitability," a paper discussing the limitations of high-frequency trading strategies.
[19] C. Ziemba and W. T. Ziemba - Editors of

"Worldwide Asset and Liability Modeling," a book on asset and liability management with applications in finance.
[20] S. R. Das - Author of "Market Microstructure: A Survey," providing a comprehensive overview of market microstructure, which is crucial for understanding stock price movements.