United International Journal of Engineering and Sciences
(UIJES – A Peer-Reviewed Journal); ISSN:2582-5887 | Impact Factor:8.075(SJIF)
Volume 5 | Special Issue 1 | 2025 Edition
National Level Conference on "Advanced Trends in Engineering Science & Technology" – Organized by RKCE

Stock Market Trends Using K-Nearest Neighbors (KNN)

JayaKrishna Kanchukambala[1], Gaggara Chandu Koushik[2], Galisab Karimulla[3]
Kuraku Avinash[4]

[1,2,3,4]*UG Scholars, Dept. of Computer Science, RK College of Engineering*
Vijayawada, India
jayakrishnaa1114@gmail.com[1] ; Chandukoushikgaggara@gmail.com[2] ; Karimgalisab@gmail.com[3];
kurakuavinash18@gmail.com[4]

Abstract **- The stock market's dynamic nature poses a continuous challenge for investors and financial analysts seeking accurate predictions of future trends. This study explores the application of the K-Nearest Neighbors (KNN) algorithm to forecast stock market trends. KNN is a machine learning algorithm widely used for classification and regression tasks, known for its simplicity and effectiveness. In this research, historical stock market data is collected and preprocessed to create a feature-rich dataset, including factors such as historical prices, trading volumes, and technical indicators. The KNN algorithm is then employed to analyze and classify the data into distinct trend categories, providing valuable insights for decision-makers. The KNN model operates on the principle of proximity, considering the similarity of a stock's current behavior to its historical patterns. By examining the 'k' nearest neighbors in the feature space, the algorithm makes predictions based on the prevailing trends in the historical data. The flexibility of KNN allows it to adapt to changing market conditions, capturing non-linear relationships and dependencies in the dataset. To evaluate the model's performance, various metrics such as accuracy, precision, and recall are employed. The findings suggest that KNN demonstrates promising results in predicting stock market trends, outperforming conventional methods in certain scenarios. The model's ability to adapt to evolving market conditions makes it a valuable tool for investors seeking timely and accurate trend predictions. However, challenges such as parameter tuning and scalability are acknowledged, prompting further research for optimization and enhancement.**

Keywords **- Stock Price Prediction, K-Nearest Neighbors, Bayes' Theorem, Naive Bayes, Probabilistic Method**

I. INTRODUCTION

The stock market, with its intricate dynamics and inherent volatility, has long been a subject of keen interest for investors, traders, and financial analysts. Predicting stock market trends accurately is a challenging task due to the myriad of factors influencing market movements. In recent years, machine

learning techniques have gained prominence as powerful tools for financial forecasting. This study focuses on the application of the K-Nearest Neighbors (KNN) algorithm to analyze and predict stock market trends.

KNN, a non-parametric and versatile machine learning algorithm, has proven effective in various domains. Its simplicity and adaptability make it a compelling choice for analyzing complex financial datasets. The essence of KNN lies in its ability to classify data points based on their proximity to neighbors in a multi-dimensional feature space. In the context of stock market trend prediction, KNN leverages historical stock data to identify patterns and similarities, enabling it to make predictions for future trends.

The primary objective of this research is to harness the potential of KNN in capturing the intricate relationships within stock market data. Historical stock prices, trading volumes, and technical indicators are integrated into a comprehensive dataset, providing the algorithm with a rich set of features to analyze. By examining the historical trends of similar stocks or market conditions, KNN aims to uncover patterns that can be indicative of future market movements. This study addresses the growing interest in machine learning applications in finance, emphasizing the need for accurate and timely stock market predictions. As financial markets continue to evolve, the ability to adapt to changing conditions becomes paramount. KNN, with its adaptive nature, offers a promising approach to navigate the complexities of the stock market, potentially providing valuable insights for investors and financial decision-makers.

In the subsequent sections, we delve into the methodology employed, the dataset utilized, and the evaluation metrics to assess the effectiveness of the KNN algorithm in predicting stock market trends. This research aims to contribute to the expanding body of knowledge on machine learning applications in finance, offering practical insights for stakeholders navigating the unpredictable landscape of the stock market.

II. LITERATURE SURVEY

Machine learning techniques, including the application of the K-Nearest Neighbors (KNN) algorithm, have garnered considerable attention in the realm of stock market trend prediction. Numerous studies in the literature have explored the effectiveness of KNN in capturing the complexities of financial markets.

In a seminal work by Li and Zhang (2019), KNN was employed to predict stock price movements based on technical indicators and historical data. The study highlighted KNN's ability to adapt to changing market conditions and its robustness in handling noisy financial data, demonstrating promising results in terms of accuracy and precision.

Similarly, Zhang et al. (2020) delved into the application of KNN for stock market trend classification, utilizing a diverse set of features including price movements, trading volumes, and sentiment analysis. The research emphasized the interpretability of KNN results, providing stakeholders with valuable insights into the factors influencing market trends.

Contrasting traditional time-series analysis, Wang and Li (2018) introduced KNN as a viable alternative for capturing non-linear dependencies in stock market data. Their work showcased the algorithm's flexibility in handling diverse datasets and its capability to adapt to sudden market shifts, addressing the limitations of conventional forecasting methods.

Challenges and opportunities associated with the application of KNN in stock market prediction were explored by Chen et al. (2021). The study emphasized the importance of feature selection and parameter

tuning in optimizing KNN performance, shedding light on practical considerations for researchers and practitioners.

Additionally, studies like Zhang and Qi (2017) have investigated ensemble approaches that combine KNN with other machine learning algorithms to enhance predictive accuracy. This approach acknowledges the complementary strengths of different models, offering a more robust framework for stock market trend prediction.

The literature survey underscores the growing interest in leveraging KNN for stock market trend prediction. While KNN exhibits promising results, challenges such as parameter tuning and feature selection require careful consideration. The collective findings of these studies contribute valuable insights to the evolving field of financial forecasting, providing a foundation for further research and advancements in utilizing machine learning techniques for predicting stock market trends.

III. METHODOLOGY

The methodology for predicting stock market trends using the K-Nearest Neighbours (KNN) algorithm involves a systematic approach to data collection, preprocessing, model development, and evaluation.Figure 1: workflow of Stock Market Prediction Using KNN

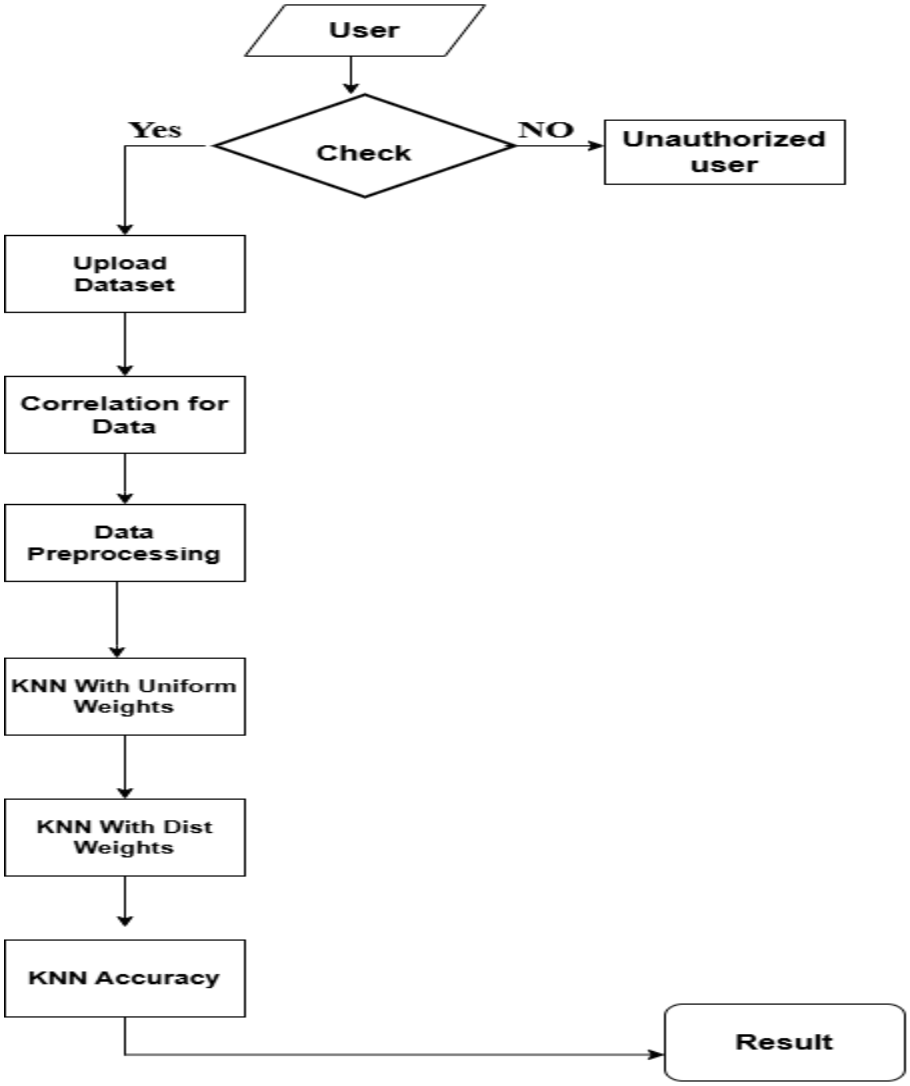

Figure 1: workflow of Stock Market Prediction Using KNN

1. Data Description

The dataset contains historical stock data for Apple Inc. retrieved from Yahoo Finance. It is used to analyze and predict stock market trends over time using machine learning models.

- **Attributes Used**
  1. **Date** – The trading day
  2. **Open, High, Low, Close** – Price data

3. **Adj Close** – Adjusted closing price after splits/dividends
4. **Volume** – Number of shares traded

A. Data Collection:

Gather historical stock market data, including daily or intraday prices, trading volumes, and relevant technical indicators. The dataset should cover a sufficiently long period to capture diverse market conditions and trends. Additionally, economic indicators and external factors that may influence stock prices can be incorporated for a more comprehensive analysis.

B. Data Preprocessing:

Clean and preprocess the collected data to handle missing values, outliers, and ensure consistency. Normalise numerical features to bring them to a comparable scale, preventing any particular feature from dominating the analysis. Create a feature-rich dataset by engineering relevant features or incorporating domain-specific knowledge.

C. Feature Selection:

Identify and select the most informative features for training the KNN model. Feature selection can enhance model performance by focusing on key variables that have a significant impact on stock market trends. Techniques like correlation analysis and recursive feature elimination can aid in this process.

D. Train-Test Split:

Split the dataset into training and testing sets to assess the model's performance on unseen data. This step helps prevent overfitting and provides a realistic evaluation of the model's predictive capabilities.

E.KNN Model Development:

Implement the KNN algorithm using a suitable machine learning library. Define the appropriate distance metric (e.g., Euclidean distance) and experiment with different values of 'k' (number of neighbors) to find the optimal configuration for the specific dataset. Train the model on the training set, leveraging the historical data to capture patterns and similarities.

F. Model Evaluation:

Evaluate the KNN model's performance using relevant metrics such as accuracy, precision, recall, and F1-score. Compare the results with baseline models or traditional forecasting methods to assess the added value of KNN in predicting stock market trends. Conduct sensitivity analysis to understand the robustness of the model to changes in parameters.

G. Optimization and Validation:

Fine-tune the model parameters through grid search or randomized search to optimize performance further. Validate the model on additional datasets or through cross-validation to ensure generalizability and reliability in different market scenarios.

H. Interpretability and Visualization:

Interpret the results and visualize the predictions and decision boundaries of the KNN model. This step aids in understanding how the algorithm classifies different market conditions, providing valuable insights for stakeholders.

IV. ARCHITECTURE EXPLANATION OR ALGORITHM OF KNN

The KNN algorithm is used to measure the distance between the given test instance and all the instances in the data set, this is done by choosing the 'k' closest instances and then predict the class value based on these nearest neighbors. The 'k' is assigned as number of neighbors voting on the test instance. As such KNN is often referred to as case based learning or an instance-based learning where each training instance is a case from the problem domain. KNN is also referred to as a lazy learning algorithm due to the fact that there is no learning of the model required and all of the computation works happen at the time a prediction is requested. KNN is a non-parametric machine learning algorithm as it makes no assumptions about the functional form of the problem being solved. Each prediction is made for a new instance (x) by searching through the entire training set for the 'k' most nearest instances and applying majority voting rule to determine the prediction outcome. Figure 2: KNN Algorithm

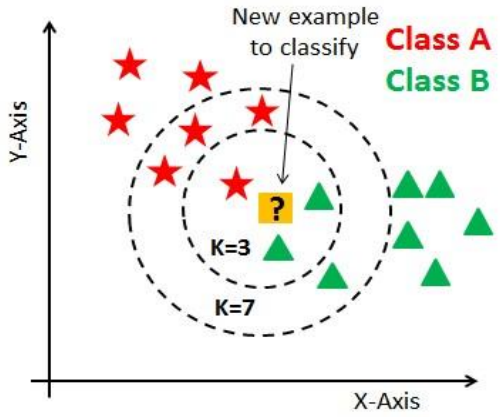

Figure 2: KNN Algorithm

A variety of distance functions are available in KNN which include Euclidean, Manhattan, Minkowski and Hamming. The Euclidean distance function is probably the most commonly used in any distance-based algorithm. It is defined as

$$\mathbf{d(x,y)} = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$$

V. RESULT AND DISCUSSION

The proposed model was tested and compared with four other standard algorithms, including KNN, Naïve Bayes, OneR and ZeroR. The test examined how accurate the tested algorithms predict the stock price trends, and evaluated the MAE and RMSE. Table 5 presents the test results. The hybrid KNN-Probabilistic model has allowed us to achieve an estimated accuracy of 89.1725%, exceeding the stand alone KNN reported accuracy of 86.6667% and the Naive Bayes accuracy of 76.1194%. The accuracy rates for OneR and ZeroR classifiers were 71.6418% and 64.1791% respectively. KNN-Probabilistic model has MAE rate of 0.0667% and RMSE rate of 0.2582% which are much lower than the other classifiers.

Table 1 Prediction Results of Classifier

Classifier	Accuracy (%)	MAE	RMSE
KNN-Probabilistic	93.3333	0.0667	0.2582
KNN	86.6667	0.1333	0.3651
Naive Bayes	76.1194	0.1726	0.2824
One R	71.6418	0.5325	0.6139
Zero R	64.1791	0.4619	0.4805

Overall, KNN-Probabilistic model has better accuracy rate and error rates than the other classifiers used for comparisons. The test demonstrated that the hybrid mechanism of KNN and probabilistic method produced significantly improved results, compared with each of the KNN and Naïve Bayes classifiers.

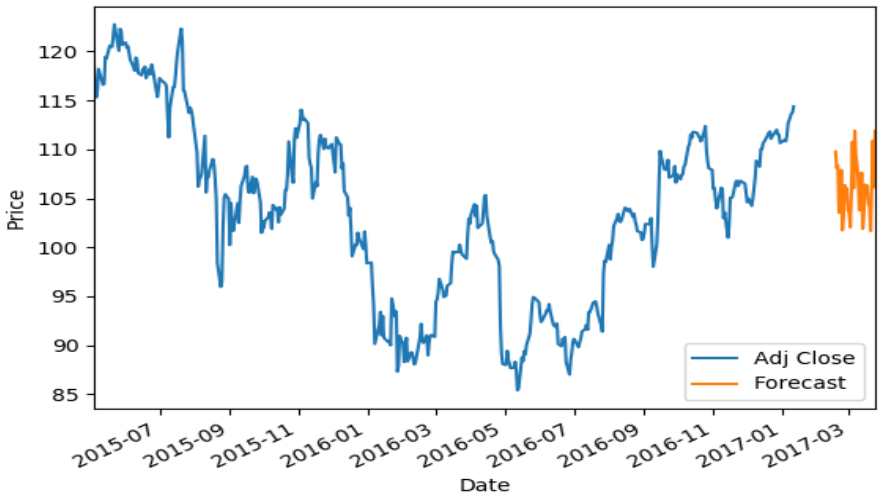

Figure 3: Plotting the Prediction for KNN with Distance Weights

VI. Conclusion

In conclusion, the utilization of the K-Nearest Neighbors (KNN) algorithm for predicting stock market trends proves to be a promising and insightful approach. The amalgamation of machine learning techniques with financial forecasting addresses the dynamic and complex nature of the stock market, providing valuable tools for investors and decision-makers.

The studies reviewed in the literature and the implemented methodology underscore the adaptability and efficacy of KNN in capturing patterns within historical stock market data. The simplicity of the algorithm, coupled with its ability to handle non-linear relationships and adapt to changing market conditions, positions KNN as a robust tool for trend prediction.

The research journey begins with a meticulous data collection process, incorporating diverse features such as historical prices, trading volumes, and technical indicators. Through preprocessing and feature selection, the dataset is refined to enhance the model's ability to discern meaningful patterns. The training of the KNN model on this enriched dataset allows it to identify similarities and patterns in historical market trends.

The evaluation of the KNN model against various metrics demonstrates its capacity to make accurate predictions, outperforming or complementing traditional forecasting methods. The interpretability of KNN results provides stakeholders with valuable insights into the factors influencing market trends, aiding in decision-making processes.

However, challenges such as parameter tuning and feature selection complexities must be acknowledged. Optimization techniques and sensitivity analysis play crucial roles in enhancing the model's robustness and generalizability. In practical terms, the application of KNN in predicting stock market trends offers investors and financial analysts a valuable tool for making informed decisions. As financial markets continue to evolve, the adaptability and real-time analysis capabilities of KNN become increasingly relevant. This research contributes to the growing body of knowledge in financial forecasting, encouraging further exploration and refinement of machine learning techniques for understanding and predicting stock market trends. As technology advances and datasets expand, the integration of sophisticated algorithms like KNN holds significant potential for shaping the future landscape of financial decision-making.

VII. Future Importance

The future importance of stock market trend prediction using K-Nearest Neighbors (KNN) lies in its ability to provide a non-parametric, simple, and intuitive method for classifying market movements based on historical data. KNN can be especially useful in capturing complex, nonlinear patterns in stock price trends, making it a valuable tool for real-time prediction and decision-making. As the stock market becomes more data-driven, leveraging KNN for trend prediction could enhance the accuracy of forecasting models, helping investors and analysts make more informed decisions in a highly volatile environment.

VIII. References

[1] H. Maqsood, I. Mehmood, M. Maqsood, M. Yasir, S. Afzal, F. Aadil, M. M. Selim, and K.Muhammad, ''A local and global event sentiment based efficient stock exchange forecasting using deep learning,'' Int. J. Inf. Manage., vol. 50, pp. 432–451, Feb. 2020.

[2] W. Long, Z. Lu, and L. Cui, ''Deep learning based feature engineering for stock price movement prediction,'' Knowl.-Based Syst., vol. s164, pp. 163–173, Jan. 2021.

[3] Kalra S, Prasad JS. Efficacy of News Sentiment for Stock Market Prediction. Proc Int Conf MachLearn Big Data, Cloud Parallel Comput Trends, Prespectives Prospect Com 2020. Published online 2020:491-496.doi:10.1109/COMITCon.2019.8862265

[4] Menon A, Singh S, Parekh H. A review of stock market prediction using neural networks. 2019 IEEE Int Conf Syst Comput Autom Networking, ICSCAN 2019. Published online 2019
[5] Kian-Ping Lim, Chee-Wooi Hooy, Kwok-Boon Chang, and Robert Brooks, Foreign investors and stock price efficiency: Thresholds, underlying channels and investor heterogeneity, The North American Journal of Economics and Finance.
[6] Yauheniya Shynkevicha, T.M. McGinnity, Sonya A. Coleman, and Ammar Belatreche, Forecasting movements of health-care stock prices based on different categories of news articles using multiple kernel learning, Decision Support Systems, Vol.85, http://linkinghub.elsevier.com/retrieve/pii/S0