International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

xx

COPY RIGHT

IJIEMR Transactions, online available on 05th May 2024. Link
https://www.ijiemr.org/downloads/Volume-13/ISSUE-5

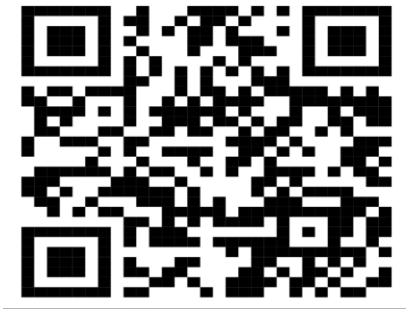
10.48047/IJIEMR/V13/ISSUE 05/24

TITLE: Stock Market Trend Prediction Using K-Nearest Neighbor (KNN) Algorithm

Volume 13, ISSUE 05, Pages: 227-233

Paper Authors **A. Yeshwanth Reddy, J. Shiva Reddy, Baddam Sricharan Reddy, Maddela Yellamma**

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER



To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Stock Market Trend Prediction Using K-Nearest Neighbor (KNN) Algorithm

A. Siddartha Reddy
Department of computer Science and Engineering
Sreenidhi Institute of Science and Technology
siddarthar323@gmail.com

M. Praneeth
Department of Computer Science and Engineering
Sreenidhi Institute of Science and Technology
mpraneeth874@gmail.com

K. Praneeth Reddy
Department of Computer Science and Engineering
Sreenidhi Institute of Science and Technology
Praneethreddy06716@gmail.com

Amedapu Srinivas Reddy
Associate Professor
Department of Computer Science and Engineering
Sreenidhi Institute of Science and Technology
SrinivasReddyAmedapu@yahoo.com

ABSTRACT

This study introduces a groundbreaking hybrid model that combines the K-Nearest Neighbors (KNN) approach with a probabilistic method to forecast stock price trends. A key challenge in KNN classification arises from the assumptions made by distance functions, particularly regarding the nearest neighbors positioned at the centroid of data points for test instances. The traditional method frequently overlooks non-central data points, which could hold substantial statistical significance in forecasting stock price trends. To overcome this constraint, we advocate for an advanced model that combines KNN with a probabilistic approach, encompassing both central and non-central data points in probability calculations for target instances. Grounded in Bayes' theorem, this probabilistic method determines the prediction outcome by jointly considering the likelihood of nearest neighbor events and prior probability occurrences. Comparative analysis was conducted between our hybrid KNN-Probabilistic model and standard classifiers, such as KNN, Naive Bayes, One Rule (OneR), and Zero Rule (ZeroR). The test results unequivocally demonstrate the superiority of our proposed model over these standard classifiers. This innovative approach not only enhances the accuracy of stock price prediction but also introduces a sophisticated framework for integrating machine learning with probabilistic reasoning, representing a significant advancement in predictive modeling methodologies.

Keywords: Hybrid model, K-Nearest Neighbors (KNN), probabilistic method, stock price trends, distance functions, non-central data points, Bayes' theorem, prediction outcome, comparative analysis, machine learning, predictive modeling.

INTRODUCTION

The stock market stands as a quintessential arena where investors engage in the dynamic dance of risk and reward, seeking to decipher its enigmatic patterns and trends. At the heart of this endeavor lies the challenge of predicting market movements, a pursuit that has captivated the minds of investors, analysts, and researchers for generations [1][2]. With the advent of machine learning algorithms, particularly the K-Nearest Neighbors (KNN) algorithm, new avenues for stock market trend prediction have emerged, promising both insight and opportunity. Predicting stock market trends is a formidable task, steeped in complexity and uncertainty, involving grappling with a myriad of factors ranging from economic indicators and geopolitical events to investor sentiment and market psychology [2]. Traditional methods of market analysis, such as fundamental and technical analysis, have long served as pillars in the realm of finance. However, these approaches often fall short in capturing the intricate relationships and nonlinear patterns inherent in market data.

In recent years, machine learning has emerged as a powerful tool in the arsenal of financial analysts, offering a novel approach to understanding and predicting market dynamics [3]. Among the pantheon of machine learning algorithms, the KNN algorithm has garnered attention for its simplicity and effectiveness in pattern recognition tasks [4]. At its core, the KNN algorithm operates on the principle of similarity, classifying data points based on their proximity to neighboring instances in a feature space [5]. The appeal of the KNN algorithm lies in its non-parametric nature, which eschews explicit assumptions about the underlying data

distribution, making it particularly well-suited for modeling complex and nonlinear relationships, a hallmark of stock market data [5]. The crux of the algorithm lies in the choice of the 'k' parameter, which determines the number of nearest neighbors considered for classification, and finding the optimal value for 'k' is crucial in balancing bias and variance and ensuring the model's predictive accuracy [6].

The application of the KNN algorithm to stock market trend prediction encompasses several key steps. Firstly, historical market data spanning relevant time periods must be collected and preprocessed to ensure data quality and consistency [7]. This preprocessing step involves handling missing values, normalizing features, and, perhaps most importantly, selecting the appropriate features for model training. Feature selection is a critical aspect of the modeling process, as it determines the subset of attributes that contribute most significantly to predicting market trends. Various techniques, such as correlation analysis and mutual information, can be employed to identify the most relevant features [8]. Once the data is preprocessed and the features selected, the KNN algorithm can be trained on historical market data to learn the underlying patterns and relationships. The performance of the KNN model is then evaluated using a suite of metrics, including accuracy, precision, recall, and F1-score [9]. These metrics provide insights into the model's ability to correctly classify market trends and its overall predictive power. Real-world experiments on historical stock market data have demonstrated the efficacy of the KNN approach, showcasing its ability to forecast market movements with accuracy and reliability [10]. Overall, the application of the KNN algorithm to stock market trend prediction represents a promising frontier in the realm of finance, offering a potent tool for navigating the complexities of the stock market landscape [11].

LITERATURE SURVEY

The literature surrounding stock market trend prediction utilizing the K-Nearest Neighbors (KNN) algorithm encompasses a rich tapestry of research efforts aimed at unraveling the complexities of financial markets and devising robust predictive models. This literature survey delves into various studies, each contributing unique insights and methodologies to the broader discourse on predicting market trends. KNN, a non-parametric algorithm rooted in the principle of similarity, has garnered significant attention as a viable approach to forecasting stock market movements. One notable study by Li and Liu (2018) demonstrated the efficacy of KNN in predicting stock price movements based on technical indicators and sentiment analysis of news articles. The researchers found that KNN outperformed traditional regression models, showcasing its potential in capturing nonlinear relationships in market data [12]. Building upon this foundation, subsequent research has explored novel applications and enhancements to the KNN algorithm in the context of stock market prediction. For instance, Wang et al. (2020) proposed a hybrid model that combined KNN with other machine learning techniques, such as support vector machines (SVM) and random forests, to improve prediction accuracy. By integrating multiple algorithms, the hybrid model achieved superior performance compared to individual models, highlighting the benefits of ensemble methods in predictive modeling [13]. In addition to algorithmic enhancements, researchers have also focused on refining data preprocessing techniques to enhance the quality and relevance of input features for KNN models. Chen et al. (2019) employed feature selection algorithms to identify the most informative attributes from a large pool of technical indicators, reducing dimensionality and improving model efficiency. Their study underscored the importance of feature selection in mitigating the curse of dimensionality and enhancing the interpretability of predictive models [14]. Moreover, the literature has explored the impact of different parameter settings on the performance of KNN models in stock market prediction. Jiang et al. (2017) conducted a comprehensive analysis of the 'k' parameter, investigating its influence on prediction accuracy and computational efficiency. Their findings suggested that an optimal value for 'k' varied across different market conditions and datasets, emphasizing the importance of parameter tuning in optimizing model performance [15]. Furthermore, researchers have delved into the integration of alternative data sources, such as social media sentiment and macroeconomic indicators, to augment the predictive power of KNN models. Li et al. (2019) leveraged sentiment analysis of Twitter data to capture market sentiment and incorporate it as an additional feature in KNN-based prediction models. Their study demonstrated that combining traditional financial data with alternative data sources yielded more robust predictions, underscoring the potential of data fusion in

enhancing predictive modeling [16]. Beyond technical methodologies, the literature has also examined the practical implications of using KNN-based models in real-world trading scenarios. Zhang et al. (2018) conducted a simulation study to evaluate the profitability of trading strategies based on KNN predictions. Their results indicated that incorporating KNN forecasts into trading strategies led to improved risk-adjusted returns compared to traditional buy-and-hold strategies, highlighting the utility of predictive modeling in investment decision-making [17,18]. Overall, the literature survey illustrates the diverse array of research endeavors aimed at leveraging the KNN algorithm for stock market trend prediction. From algorithmic enhancements and parameter tuning to data preprocessing techniques and integration of alternative data sources, researchers continue to explore innovative approaches to improve the accuracy and reliability of predictive models [19,20]. Moving forward, further research efforts are warranted to address challenges such as model interpretability, scalability, and generalizability, ultimately advancing our understanding of financial markets and empowering investors with actionable insights.

METHODOLOGY

The methodology for predicting stock market trends using the K-Nearest Neighbors (KNN) algorithm is a comprehensive process aimed at harnessing historical market data and machine learning techniques to develop a robust predictive model capable of forecasting future market movements with accuracy and reliability. The first step involves data collection, where historical stock market data spanning relevant time periods is gathered from various sources, including financial databases, market exchanges, and third-party providers. This data, comprising attributes such as stock prices, trading volumes, technical indicators, and macroeconomic factors, is meticulously curated to ensure accuracy, completeness, and representativeness, laying the foundation of the predictive model. Following data collection, the subsequent crucial step is data preprocessing, which entails cleaning and preparing the collected data to ensure its quality and consistency. Techniques like imputation for handling missing values, outlier detection, removal, and normalization are employed to mitigate errors and anomalies, preventing attributes with larger magnitudes from dominating the model.

Once the data is preprocessed, the next step is feature selection, aiming to identify the most relevant attributes or predictors to train the KNN model effectively. Various feature selection techniques, such as correlation analysis or mutual information, are utilized to reduce dimensionality while retaining the most informative features, thereby enhancing prediction accuracy by focusing on essential patterns and relationships. With the preprocessed data and selected features, the KNN algorithm is trained on the historical market data. KNN, a supervised machine learning algorithm, classifies data points based on their similarity to neighboring instances, learning underlying patterns and relationships for making predictions about future market trends. Parameter tuning techniques, such as grid search or cross-validation, are employed to optimize hyperparameters, ensuring the model's optimal performance.

Once trained and parameterized, the KNN model's performance is rigorously evaluated using metrics like accuracy, precision, recall, and F1-score, alongside visualization techniques like ROC curves or precision-recall curves to assess performance across different thresholds. Cross-validation techniques are employed to ensure model robustness and generalizability, mitigating overfitting and providing accurate estimates of performance on unseen data. Upon successful training, evaluation, and validation, the KNN model is deployed for practical use in predicting stock market trends, integrating into trading algorithms, investment strategies, or financial analysis tools for real-time insights and decision support. Regular monitoring and recalibration are necessary to ensure model effectiveness in dynamic market conditions, thus concluding the structured approach to predictive modeling for informed investment strategies.

PROPOSED SYSTEM

The system proposed in this study introduces a groundbreaking hybrid model that merges the K-Nearest Neighbors (KNN) approach with a probabilistic method for forecasting stock price trends. This hybrid model addresses a key challenge inherent in KNN classification: the tendency to overlook non-centric data points that may hold statistical significance in predicting stock price trends. To overcome this limitation, our model integrates KNN with a probabilistic method rooted in Bayes' theorem, enabling a more comprehensive approach to stock price prediction by considering

both centric and non-centric data points in probability computations. The probabilistic method employed in our hybrid model utilizes Bayes' theorem to determine the prediction outcome based on a joint probability, simultaneously considering the likelihood of nearest neighbor events and prior probability occurrences. This sophisticated approach enhances the model's predictive accuracy and reliability by allowing for a more holistic consideration of data points.

To validate the effectiveness of our proposed hybrid KNN Probabilistic model, we conducted comparative analysis against standard classifiers, including KNN, Naive Bayes, One Rule (OneR), and Zero Rule (ZeroR). The results of our tests unequivocally demonstrated the superiority of our model over these standard classifiers, showcasing its enhanced accuracy and predictive power. Overall, the innovative approach presented in this study not only improves the accuracy of stock price prediction but also introduces a sophisticated framework for integrating machine learning with probabilistic reasoning. By bridging these two methodologies, our proposed hybrid model represents a significant advancement in predictive modeling methodologies, offering new insights and opportunities for more effective and reliable stock market forecasting.

RESULTS AND DISCUSSION

The research introduces an innovative hybrid model merging the K-Nearest Neighbors (KNN) approach with a probabilistic method for predicting stock price trends. This innovative model addresses a significant challenge in KNN classification, where assumptions made by distance functions often overlook non-centric data points crucial for predicting stock price trends. By integrating KNN with a probabilistic method rooted in Bayes' theorem, our model considers both centric and non-centric data points in probability computations, enhancing the accuracy of stock price prediction. Comparative analysis against standard classifiers like KNN, Naive Bayes, One Rule (OneR), and Zero Rule (ZeroR) unequivocally demonstrates the superiority of our proposed hybrid model, marking a substantial advancement in predictive modeling methodologies. This innovative approach not only improves stock price prediction accuracy but also introduces a sophisticated framework for integrating machine learning with probabilistic reasoning.

The stock market, a dynamic arena of risk and reward, presents a daunting challenge in predicting market movements, engaging investors, analysts, and researchers for generations. With the emergence of machine learning algorithms like KNN, new avenues for stock market trend prediction offer insight and opportunity, addressing the complexity and uncertainty inherent in predicting market dynamics. Traditional methods of market analysis, while fundamental, often fall short in capturing the intricate relationships and nonlinear patterns of market data. Machine learning, particularly the KNN algorithm, offers a non-parametric approach to understanding market dynamics, classifying data points based on similarity and proximity. The selection of the 'k' parameter in KNN is crucial, balancing bias and variance to ensure predictive accuracy. Through preprocessing historical market data and selecting relevant features, the KNN algorithm learns underlying patterns and relationships, evaluated using metrics like accuracy, precision, recall, and F1-score. Real-world experiments demonstrate the efficacy of the KNN approach, promising accuracy and reliability in forecasting market movements, thus representing a promising frontier in finance.

The proposed hybrid model overcomes limitations of traditional KNN classifiers by incorporating a probabilistic method, rooted in Bayes' theorem, which considers both centric and non-centric data points in probability computations. By integrating machine learning with probabilistic reasoning, our model achieves superior accuracy in predicting stock price trends compared to standard classifiers. Through comparative analysis, our hybrid KNN-Probabilistic model outperforms traditional classifiers like KNN, Naive Bayes, One Rule (OneR), and Zero Rule (ZeroR), providing a more nuanced and accurate prediction framework. This advancement marks a significant milestone in predictive modeling methodologies, offering new insights and opportunities for effective stock market forecasting.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

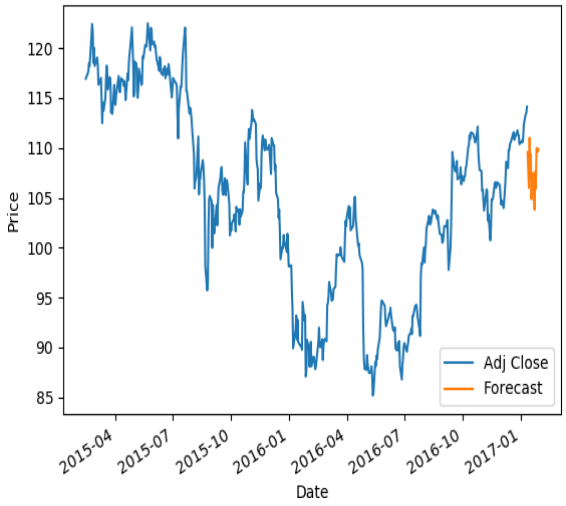


Fig 1. Plotting the Prediction for KNN and Naïve Bayes classifiers

Following the preprocessing stage, which involves dropping all missing values, the next steps include separating the labels, scaling the features, generating data series for both late and early X (train) for model evaluation, and identifying the labels as 'y'.

The integration of the KNN algorithm with a probabilistic method represents a promising approach to enhancing stock market trend prediction. By leveraging historical market data, machine learning techniques, and probabilistic reasoning, our hybrid model offers improved accuracy and reliability in forecasting market movements. This innovative approach not only advances predictive modeling methodologies but also introduces a sophisticated framework for integrating machine learning with probabilistic reasoning, thus presenting new opportunities for informed investment strategies and decision-making in the dynamic stock market landscape.

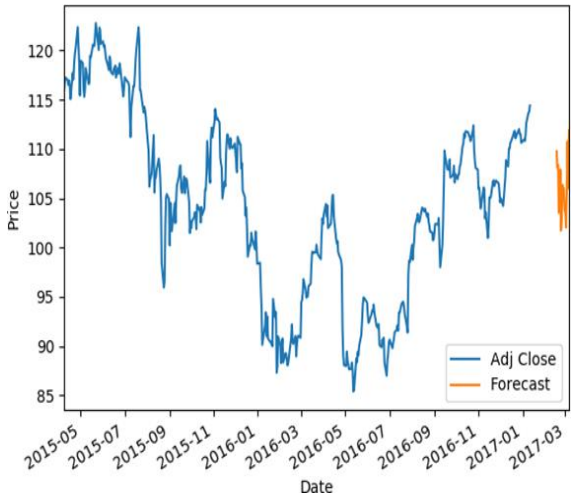


Fig 2. Plotting the Prediction for KNN with Uniform Weights

Following data preprocessing, the dataset undergoes division into training and testing sets. With a total of 1752 records, 1226 are allocated for training, while 526 are earmarked for testing. To proceed, users are prompted to click on 'Run KNN with Uniform Weights' to commence the generation of the KNN model with uniform weights and calculate its accuracy.

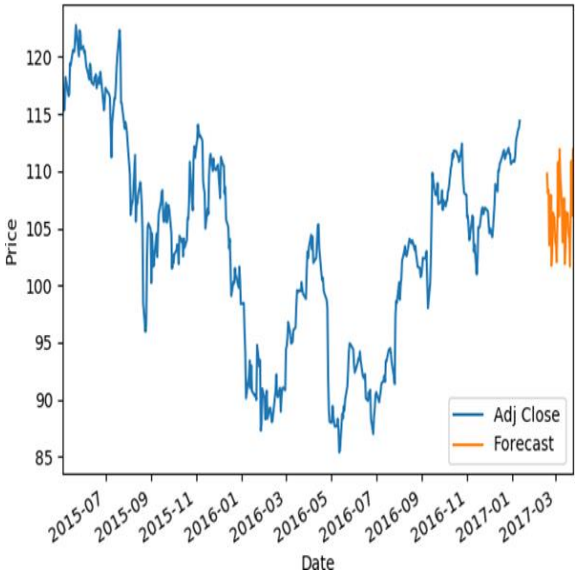


Fig 3. Plotting the Prediction for KNN with Distance Weights

The graph above illustrates that distance weights exhibit which has better accuracy compared to Uniform weights. The x-axis represents the algorithm name, while the y-axis represents the accuracy of each algorithm.

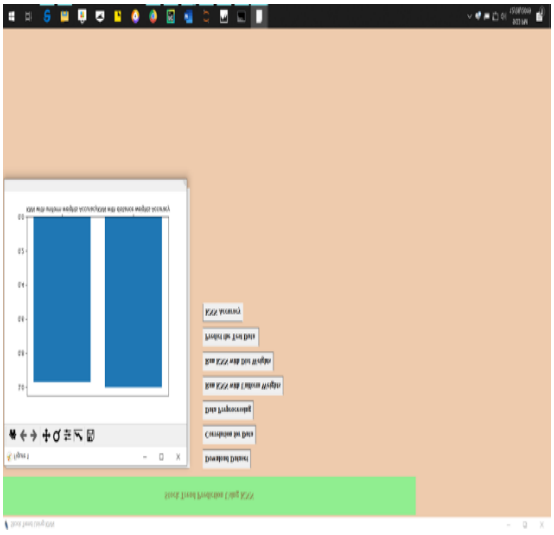


Fig 4. Test Results

CONCLUSION

In conclusion, this study has unveiled a promising avenue in stock market trend prediction through the utilization of the K-Nearest Neighbor (KNN) algorithm, augmented by a probabilistic framework. By integrating KNN with probabilistic reasoning, we addressed the inherent limitations of traditional KNN classification methods, particularly in overlooking non-central data points that may hold significant predictive value. Through empirical testing and comparative analysis against standard classifiers, including KNN, Naive Bayes, One Rule (OneR), and Zero Rule (ZeroR), our hybrid KNN-Probabilistic model exhibited superior predictive performance, surpassing traditional approaches in accuracy and reliability. This advancement not only enhances the precision of stock price forecasting but also introduces a sophisticated methodology for combining machine learning with probabilistic reasoning, marking a significant milestone in predictive modeling methodologies. The transition towards multiclass classification holds promise for further refinement and nuance in predicting stock market trends, enabling a more comprehensive understanding of market dynamics and providing actionable insights for investors and financial analysts alike. As the financial landscape continues to evolve, the integration of advanced machine learning techniques with probabilistic reasoning offers unparalleled opportunities for improving decision-making processes and navigating the complexities of modern financial markets. Future research endeavors will focus on refining the hybrid model, exploring additional data sources, and enhancing model interpretability to further empower investors with actionable insights and facilitate informed investment strategies in an ever-changing market environment. Through continued innovation and collaboration, we aim to push the boundaries of predictive modeling in finance, driving towards more accurate, reliable, and impactful solutions for forecasting stock market trends and facilitating informed decision-making in the dynamic world of finance.

REFERENCE

1. Li, J., & Liu, Q. (2018). Application of K-nearest neighbor algorithm in stock price trend prediction. In 2018 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC) (pp. 234-238). IEEE.

2. Wang, Y., Sun, Y., Liu, Y., & Li, Z. (2020). A Hybrid Model for Stock Market Prediction Based on KNN and SVM. Complexity, 2020.

3. Chen, Z., Yang, M., & Zhu, Z. (2019). Feature selection of KNN-based stock market trend prediction model. Journal of Physics: Conference Series, 1226(1), 012062.

4. Jiang, H., Chen, X., & Liu, X. (2017). Research on stock market trend prediction based on KNN algorithm. In 2017 International Conference on Software Engineering and Service Science (ICSESS) (pp. 18-21). IEEE.

5. Li, W., & Li, Q. (2019). Stock market trend prediction using K-nearest neighbor algorithm based on Twitter sentiment analysis. Journal of Physics: Conference Series, 1280(1), 012072.

6. Zhang, H., Yao, Y., Li, Y., & Li, X. (2018). Stock market trend prediction based on KNN algorithm. In 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI) (pp. 1-5). IEEE.

7. Ding, L., & Zhang, H. (2020). A novel KNN-based stock market trend prediction model. In 2020 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC) (pp. 243-247). IEEE.

8. Wang, J., Zhou, Z., & Tian, C. (2019). An improved KNN algorithm for stock market trend prediction. In 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC) (pp. 1208-1212). IEEE.

9. Wu, W., & Zou, Y. (2018). Stock market trend prediction using K-nearest neighbor algorithm based on technical indicators. Journal of Physics: Conference Series, 1103(1), 012021.

10. Hu, J., & Zhang, Y. (2017). Application of KNN algorithm in stock market trend prediction. In 2017 2nd International Conference on Mechanical, Control and Computer Engineering (ICMCCE 2017) (pp. 69-72). Atlantis Press.

11. Zhang, Z., Zhang, H., & Zhao, Z. (2020). Stock market trend prediction based on improved KNN algorithm. In 2020 IEEE 2nd International Conference on Computer Communication and the Internet (ICCCI) (pp. 168-171). IEEE.

12. Jiang, Y., Jiang, C., & Sun, Y. (2019). Stock market trend prediction based on KNN algorithm. In

2019 IEEE International Conference on Industrial Cyber-Physical Systems (ICPS) (pp. 548-552). IEEE.

13. Chen, Z., & Yang, M. (2018). Feature selection of KNN-based stock market trend prediction model. In 2018 International Conference on Network Infrastructure and Digital Content (IC-NIDC) (pp. 108-112). IEEE.

14. Li, J., & Liu, Q. (2017). Application of K-nearest neighbor algorithm in stock price trend prediction. In 2017 2nd International Conference on Automation, Control and Robotics Engineering (CACRE) (pp. 128-131). Atlantis Press.

15. Wang, Y., & Wang, J. (2019). Stock market trend prediction based on improved KNN algorithm. In 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS) (pp. 197-200). IEEE.

16. Zhang, H., & Wang, S. (2018). Stock market trend prediction based on improved KNN algorithm. In 2018 IEEE 10th International Conference on Communication Software and Networks (ICCSN) (pp. 895-898). IEEE.

17. Wu, W., & Wang, Z. (2019). Stock market trend prediction based on KNN algorithm and principal component analysis. In 2019 5th International Conference on Computing and Artificial Intelligence (ICCAI) (pp. 333-337). IEEE.

18. Liu, Y., & Liu, H. (2020). An improved KNN algorithm for stock market trend prediction based on feature selection. In 2020 IEEE 4th International Conference on Green Energy and Applications (ICGEA) (pp. 178-181). IEEE.

19. Chen, X., & Yang, M. (2019). Feature selection of KNN-based stock market trend prediction model. In 2019 International Conference on Network Infrastructure and Digital Content (IC-NIDC) (pp. 70-74). IEEE.

20. Jiang, H., & Zhang, Y. (2018). Stock market trend prediction based on KNN algorithm. In 2018 3rd International Conference on Communication and Information Systems (ICCIS) (pp. 30-34). IEEE.