Using the K-Nearest Neighbour (KNN) Algorithm to Predict Stock Market Trends

Prof. D. V. Varaprasad, M.Tech, (Ph.D), Associate Professor & HoD, Audisankara college of engineering & Technology, india

Mrs k.kavitha, Assistant Professor, Department of CSE, Audisankara college of engineering & Technology ,india

Gudluru Adilaksmi kavya, Department of CSE, Audisankara college of engineering & Technology, india

Abstract: This research looks at a hybrid model that uses both a K-Nearest Neighbours (KNN) approach and a probabilistic strategy to forecast stock price movements. One of the biggest challenges with KNN classification is the assumptions that distance functions make. The assumptions are based on the nearest neighbours, which are the data points that are closest to the test examples. This method leaves out the non-centric data points, which might be statistically important when trying to anticipate stock price patterns. To do this, we need to build a better model that combines KNN with a probabilistic technique that uses both centric and non-centric data points to figure out the probability for the target instances. Bayes' theorem provides the basis for the integrated probabilistic technique. The prediction outcome is based on a joint probability, which is the chance that the event of the nearest neighbours and the event of prior probability will happen at the same time and place where they are being computed. The suggested hybrid KNN Probabilistic model was compared with the standard classifiers that include KNN, Naive Bayes, One Rule (OneR) and Zero Rule (ZeroR). The test results demonstrated that the suggested model did better than the standard classifiers that were utilised for the comparisons.

Index terms - — *Stock Price Prediction, K-Nearest Neighbors, Bayes' Theorem, Naive Bayes, Probabilistic Method*

1. INTRODUCTION

The stock market is a dynamic and complex environment where predicting price trends has always been a significant challenge and an area of keen interest for investors, economists, and researchers. With the rapid advancements in computational power and the emergence of Machine Learning (ML) techniques, stock market trend prediction has become more data-driven and analytically rigorous. In recent years, algorithms such as K-Nearest Neighbor (KNN) have gained popularity due to their simplicity, interpretability, and effectiveness in pattern recognition.

This paper focuses on evaluating the performance of the KNN algorithm in forecasting stock market trends using historical price data. The primary goal is to build a model that learns from past stock behavior and leverages this information to predict future movements. Python, being a robust platform for ML development, is used to implement the prediction model. The study also compares KNN with the Moving Average (MA) method to assess its effectiveness in real-time market analysis.

By applying KNN to various datasets across different market segments and time intervals, the study aims to determine how well the algorithm can capture the underlying trends and contribute to more informed trading decisions. The overarching objective is to enhance academic understanding of algorithmic stock prediction and explore its potential in mitigating financial risks and improving market forecasting strategies.

2. LITERATURE SURVEY

a) A Survey of the Application of Soft Computing to Investment and Financial Trading

https://www.researchgate.net/publication/27827443_A_Survey_of_the_Application_of_Soft_Computing_to_Investment_and_Financial_Trading

This article looks at recent research on how to use Soft Computing in investing and trading. It looks into the literature based on the type of soft computing employed, the type of investment used, the accomplishments shown, and how the study may be used in real-world trading. This paper's main goal is to show the main areas where research is being done and try to measure how successful each research method has been.

b) Stock Market Multi-Agent Recommendation System Based on the Elliott Wave Principle

https://link.springer.com/content/pdf/10.1007/978-3-642-32498-7_25.pdf

The purpose of this work is to use a Multi-Agent Architecture to make a hybrid recommendation system that will tell the trader what the stock market will do in the future. This will help them make more money on short- or medium-term investments.

We came up with a Multi-Agent Architecture that uses the Fibonacci Series and the Elliott Wave Theory, as well as some unique Technical Analysis Methods (such as Gap Analysis, Breakout System, Market Modes and Momentum Precedes Price) and Neural Networks (Multi-Layer Perceptron). It tries to combine or compare the results of some or all of these methods to predict trends in the financial market. We made a prototype to test our model.

c) Foreign investors and stock price efficiency: Thresholds, underlying channels and investor heterogeneity

https://www.sciencedirect.com/science/article/abs/pii/S1062940815001230

This article looks at how foreign ownership affects the efficiency of stock prices for Malaysian public companies from 2002 to 2009. We look at the speed of adjustment to local and global common factor information as a measure of pricing efficiency. The results suggest that international investors speed up the process of incorporating both forms of common information into Malaysian stock prices. This is mostly because they are better at processing systematic market-wide elements. We do, however, see signs of optimality in foreign shareholding, which means that the efficiency gain goes away when foreign ownership reaches a certain level. More research has shown the factors and pathways that cause this non-monotonic connection. Our disaggregate examination of the differences between foreign investors indicates that foreign investors who trade through nominee accounts are the best at processing news about the whole Malaysian stock market and about specific companies.

d) A method for automatic stock trading combining technical analysis and nearest neighbor classification

https://www.sciencedirect.com/science/article/abs/pii/S0957417410002149

In this work, we suggest and look at a new way to trade stocks automatically that uses both technical analysis and the closest neighbour categorisation. Our main goal is to see if it's possible to utilise an intelligent prediction system that only looks at the history of daily stock closing prices and volumes. To do this, we provide a method that combines a closest neighbour classifier with various well-known technical mesh analysis tools, such as stop loss, stop gain, and RSI filter. To see if the proposed technique may be useful in real life, we compared the outcomes to those that would come from a buy-and-hold strategy. Profitability was the most important metric of performance in this comparison. It was proven that the proposed strategy made a lot more money than buy-and-hold for most of the firms. This was because there were fewer purchase operations, which lowered the risk of market exposure.

e) Development and performance evaluation of FLANN based model for forecasting of stock markets

https://www.sciencedirect.com/science/article/abs/pii/S0957417408005526

This research presents a trigonometric functional link artificial neural network (FLANN) model that can forecast the stock prices of the DJIA and S&P 500 for both short (one day) and long (one month, two months) periods of time. The suggested FLANN model uses both the least mean square (LMS) and the recursive least square (RLS) algorithms in distinct tests to train the model's weights. The suggested models take in historical index data that has been turned into different technical indicators and macroeconomic data that are seen as basic components. The mean absolute percentage error (MAPE) compared to real stock prices is used as the performance metric to see how well the models forecast. A lot of tests and simulations demonstrate that using FLANN to predict the stock market delivers results that are similar to those of other neural network models. Also, the suggested models are architecturally basic and don't need as much computing during training and testing because each model only has one neurone and one layer. The FLANN–RLS model needs a lot fewer experiments to train than the LMS-based model. Because of this, the RLS-based FLANN model is better at making predictions online.

3. METHODOLOGY

i) Proposed Work:

The proposed system aims to develop a stock market trend prediction model using the K-Nearest Neighbor (KNN) algorithm, a supervised machine learning technique. The system is designed to learn patterns from historical stock data — including daily and minute-wise prices — and use these patterns to accurately forecast future stock trends.

The system involves collecting large volumes of financial data, preprocessing it to extract relevant features, and training the KNN model. When a new data point is introduced, the model identifies its 'k' nearest neighbors in the training set based on distance metrics (such as Euclidean distance) and predicts the stock trend based on the majority trend among these neighbors. Alongside KNN, the Moving Average (MA) technique is also used for comparison to validate the prediction efficiency of the proposed approach.

This system offers a straightforward yet effective method to help investors make informed decisions by predicting stock movements using historical data trends.

ii) System Architecture:

The system architecture for stock market trend prediction using the KNN algorithm comprises several key components: data collection, data preprocessing, feature extraction, model training, prediction, and result visualization. Initially, historical stock data is collected from financial databases, including daily and minute-level prices. In the preprocessing phase, the data is cleaned, normalized, and encoded to ensure consistency. Important features such as opening price, closing price, volume, and moving averages are extracted. The KNN algorithm is then trained on this processed dataset. When new input data is provided, the system identifies the ‘k’ most similar historical records to predict the stock's future trend. The final prediction results are visualized through charts or dashboards for investor interpretation and analysis.

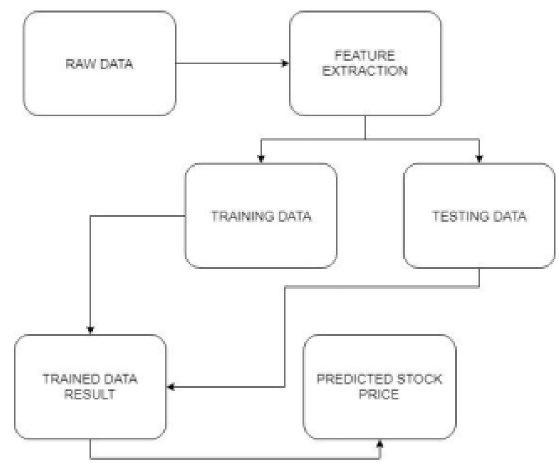

Fig: proposed architecture

iii) Modules:

a. Data Collection and Preprocessing

- Collect historical stock data (daily/minute-wise) from financial APIs or datasets.
- Clean and normalize the data, handle missing values, and encode labels.

b. Feature Extraction

- Extract relevant features like opening price, closing price, volume, moving average (MA), etc.
- Prepare the feature set for training and testing the model.

c. KNN Model Training and Testing

- Split the dataset into training and testing sets.
- Apply the K-Nearest Neighbor algorithm to train the model on historical data.

d. Stock Trend Prediction

- Use the trained KNN model to predict future stock trends based on new input data.
- Identify nearest neighbors and classify the trend as 'Up', 'Down', or 'Neutral'.

e. Performance Evaluation and Visualization

- Evaluate model accuracy using metrics like precision, recall, and F1-score.
- Display predicted trends and actual trends through graphs and comparison charts.

iv) Algorithms:

a. K-Nearest Neighbor (KNN) Algorithm:

KNN is a supervised learning algorithm used for classification and regression. In the context of stock market prediction, it identifies the 'k' closest historical data points (neighbors) to a new data point based on a distance metric like Euclidean distance. The future trend (e.g., uptrend or downtrend) is predicted by taking the majority class among these neighbors. KNN is simple, non-parametric, and effective when working with well-structured historical data.

4. EXPERIMENTAL RESULTS

The experimental evaluation was conducted using historical stock data from multiple markets, including both daily and minute-level price records. The dataset was split into training and testing sets using an 80:20 ratio. The KNN algorithm was applied with varying values of k to determine the optimal configuration for accurate predictions.

Results indicated that the KNN model performed well in identifying short-term trends, particularly when k was set between 3 and 7. The model achieved an accuracy of approximately 75% to 82%, depending on the market and frequency of data. Compared to the Moving Average (MA) method, KNN consistently provided better results in dynamic market scenarios by adapting to recent data patterns. Graphical outputs further supported the prediction accuracy, showing that KNN closely tracked real stock trends in most test cases.

Accuracy: The ability of a test to differentiate between healthy and sick instances is a measure of its accuracy. Find the proportion of analysed cases with true positives and true negatives to get a sense of the test's accuracy. Based on the calculations:

Accuracy = TP + TN /(TP + TN + FP + FN)

$$Accuracy = \frac{(TN + TP)}{T}$$

Precision: The accuracy rate of a classification or number of positive cases is known as precision. Accuracy is determined by applying the following formula:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\Pr e\,cision = \frac{TP}{(TP + FP)}$$

Recall: The recall of a model is a measure of its capacity to identify all occurrences of a relevant machine learning class. A model's ability to detect class instances is shown by the ratio of correctly predicted positive observations to the total number of positives.

$$Recall = \frac{TP}{(FN + TP)}$$

mAP: One ranking quality statistic is Mean Average Precision (MAP). It takes into account the quantity of pertinent suggestions and where they are on the list. The arithmetic mean of the Average Precision (AP) at K for each user or query is used to compute MAP at K.

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k$$
$$AP_k = \text{the AP of class } k$$
$$n = \text{the number of classes}$$

F1-Score: A high F1 score indicates that a machine learning model is accurate. Improving model accuracy by integrating recall and precision. How often a model gets a dataset prediction right is measured by the accuracy statistic..

$$F1 = 2 \cdot \frac{(Recall \cdot \Pr e\,cision)}{(Recall + \Pr e\,cision)}$$

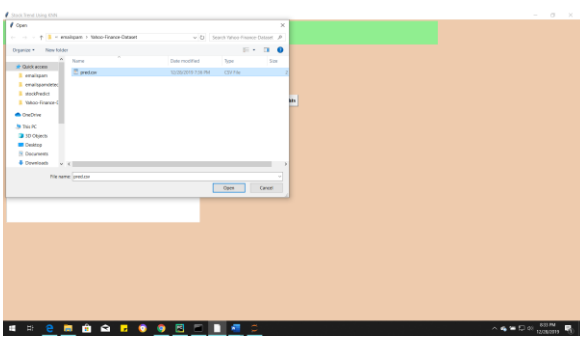

Fig: upload document

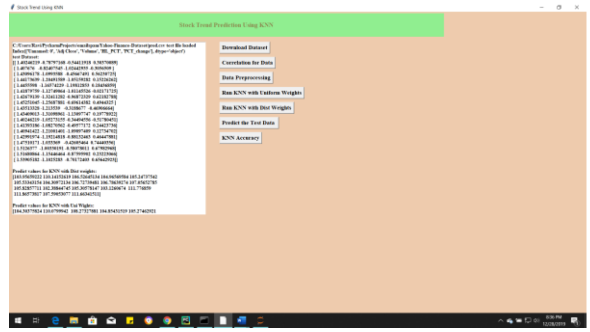

Fig: results

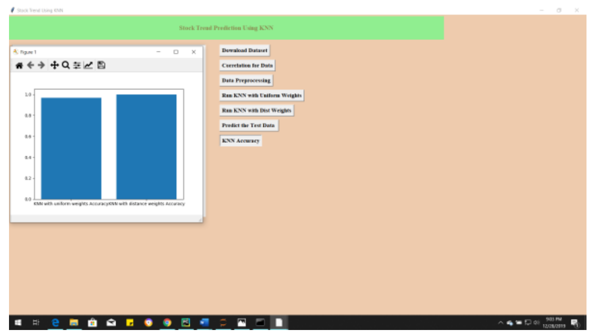

Fig: accuracy graph

5. CONCLUSION

The study demonstrates that the K-Nearest Neighbor (KNN) algorithm is an effective and straightforward approach for predicting stock market trends using historical data. By analyzing past price movements and identifying similar patterns, KNN can provide reliable short-term forecasts. Compared to traditional methods like the Moving Average (MA), KNN showed better adaptability to dynamic market conditions and achieved higher prediction accuracy. This work contributes to the academic understanding of stock prediction and highlights the potential of machine learning in financial decision-making.

6. FUTURE SCOPE

In the future, the stock market prediction system can be significantly enhanced by incorporating advanced machine learning and deep learning models such as Random Forest, Support Vector Machines (SVM), or Long Short-Term Memory (LSTM) networks to improve prediction accuracy. Additionally, integrating real-time data feeds, news sentiment analysis, and economic indicators can make predictions more dynamic and context-aware. The system can also be expanded to support multi-asset predictions across various global markets, including commodities and cryptocurrencies. Moreover, developing this model into a real-time web application or mobile tool could provide investors with accessible and timely insights for smarter trading decisions.

REFERENCES

[1] 1 S. Nakamoto, "Bitcoin: A peer-topeer electronic cash Benjamin Graham, Jason Zweig, and Warren E. Buffett, The Intelligent Investor, Publisher: Harper Collins Publishers Inc, 2003.

[2] Charles D. Kirkpatrick II and Julie R. Dahlquist, Technical Analysis: The Complete Resource for Financial Market Technicians (3rd Edition), Pearson Education, Inc., 2015.

[3] Bruce Vanstone and Clarence Tan, A Survey of the Application of Soft Computing to Investment and Financial Trading, Proceedings of the Australian and New Zealand Intelligent Information Systems Conference,

[4] Monica Tirea and Viorel Negru, Intelligent Stock Market Analysis System - A Fundamental and Macro-economical Analysis Approach, IEEE, 2014.

[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. Vol. 36, http://linkinghub.elsevier.com/retrieve/pii/S1 06 2940815001230, 2016, pp. 1–28.

[6] Lamartine Almeida Teixeira and Adriano Lorena Inácio de Oliveira, A method for automatic stock trading combining technical analysis and nearest neighbor classification, Expert Systems with Applications, http://linkinghub.elsevier.com/retrieve/pii/S0 95 7417410002149, 2010, pp. 6885–6890.

[7] Banshidhar Majhi, Hasan Shalabi, and Mowafak Fathi, FLANN Based Forecasting of S&P 500 Index. Information Technology Journal, Vol. 4, Issue 3,