Stock Market Investment Suggestion using Deep learning

1. Tanmay Changade, computer dept., G.H.Raisoni Institute of Engineering and Technology, Pune
2. Atharva Dhabu, computer dept., G.H.Raisoni Institute of Engineering and Technology, Pune
3. Nikhil Hingane, computer dept., G.H.Raisoni Institute of Engineering and Technology, Pune
4. Parth Kawale, computer dept., G.H.Raisoni Institute of Engineering and Technology, Pune
5. Prof.Sakharam Kolpe, G.H.Raisoni Institute of Engineering and Technology, Pune

***Abstract* – The task of forecasting the outcome of a stock market has traditionally posed a challenge for professionals in the fields of statistics and finance. The primary justification for this projection is the strategy of investing in stocks that are anticipated to appreciate and simultaneously divesting those that are anticipated to depreciate. When attempting to forecast subsequent years' prices for shares and the trajectory of the stock market, typically one of two methodologies is employed. The initial type of examination is referred to as "vital assessment," which involves utilizing a company's methodology and fundamental data such as market standing, expenses, and annualized growth rates to make inferences. The latter pertains to a profound examination that involves scrutinizing historical share prices alongside estimates to forecast the future of the company. This study examines historical graphs and developments to predict the future movements in the market. The present article offers a thorough exposition of the actualization of our strategy for foreseeing the stock market's movements. The methodology employs K-nearest-neighbor clustering, in tandem with Long Short-Term Memory and Decision Making, to operationalize the investment recommendation. The efficacy of the strategy has been successfully confirmed via a sequence of experiments that yielded satisfactory results.**

Keywords— *Stock market prediction, K Nearest Neighbors, Linear Regression, Deep Belief Network and Decision making.*

I. INTRODUCTION

Making investments in stocks presents a favorable opportunity to capitalize on market volatility and yield financial gain. The capacity to anticipate fluctuations in stock prices is a crucial aspect of engaging in stock market transactions. Engaging in financial trading entails a certain degree of risk as

the potential for profit or loss is contingent upon the fluctuation of stock prices. The typical investor anticipates a higher return on investment from their stock portfolio while assuming a lower level of risk. The people who depend solely on their own personal assessments and speculation to forecast and manipulate stock prices significantly elevate the level of risk and diminish potential rewards. Consequently, there exists a need for a robust and pragmatic framework capable of effectively predicting stock market trends.

During the early stages, economists possessed the ability to accurately predict the trajectory of the stock market. The advancement of pedagogical techniques has enabled data researchers to address forecasting challenges. In a further development, researchers in the field of software development have initiated the utilization of machine learning methodologies to enhance the efficiency of techniques for estimation. The utilization of deep learning techniques represents a progressive advancement towards the development of models for forecasting with enhanced accuracy. Forecasting the financial markets poses a formidable challenge, and analysts frequently encounter obstacles in their endeavors. The correlation between investor psychology and market dynamics, along with the instability that plagues financial markets, pose several challenges, such as complexity and non-linear behavior

The trajectory of the stock market is evidently impacted by a multitude of indeterminate factors, encompassing individuals' perceptions of business organizations and the geopolitical climate of countries. Thus, it is possible to predict the trajectory of the stock market and indices through efficient data preprocessing techniques applied to share prices, coupled with the utilization of suitable algorithms. The integration of deep-learning along with machine learning methodologies into share market forecasting models may offer advantages to financial analysts and the shareholders. The objective of such techniques is to independently recognize and identify patterns within vast quantities of data. Algorithms possess the ability to autonomously develop understanding, thereby enabling them to effectively predict future price fluctuations. This skill can be leveraged to refine strategies for trading.

C. Bousoo-Calzon et al. identified two Homological invariants, namely the 0th alongside 1st constant Betti numerals, to classify professionals into two categories according to their behaviors within the decision-making procedure of the Prediction Market [1]. The differentiation between logical and irrational agents in an AI-driven market model is more discernible. Ultimately, the authors collaborated with experts in the insurance sector to carry out experimental investigations utilizing four authentic SPMs. By leveraging these market places, individuals may discover a correlation between the proficiency of various participants and the dependability of their forecasts. This highlights the notion that professionals can be classified based on their conduct as opposed to their personal traits. Furthermore, the authors analyze the market by examining PM's acquisition of specialized information and shedding light on the observation that prices may serve as noisier signals compared to portfolios of assets.

Kabbani et al. [2] proposed a Deep Reinforcement Learning approach to identify the optimal trading approach for assets, incorporating indicators of volatility and sentiment information. The findings indicate that an ongoing process space is more advantageous than an isolated one in addressing the trading problem. Moreover, the inclusion of both technical variables and emotional assessments of recent headlines in the current state illustration has significantly enhanced the agent's effectiveness. The assets distribution challenge was examined with regards to the potential application of an Actor-Critic approach (TD3). The method's performance was evaluated using the averaged Sharpe ratio, resulting in a value of 2.68, which is considered favorable by investors. In a subsequent investigation, the approach could potentially derive advantages from augmented computational capabilities, thereby facilitating the modeling of supplementary circumstances and a more comprehensive evaluation of the outcomes.

X. Li et al. have conducted studies on stock forecasting algorithms that rely on news as a driving factor, yet have been taught on stocks with limited financial information. The authors of this study introduce a technique known as emotional transferable learning, which involves the application of knowledge acquired from analyzing enterprises with a high volume of news coverage to those with a comparatively lower volume of news coverage (target). As a component of the transfer process, three discrete transfer principles are formulated and implemented to ascertain the stocks that will be used as the basis. The study indicates an elevated relationship among the previous price time frame of the original stock and the target stock's assets, both of which belong to the exact same field. The original supplies, getting among the most news-rich in the industry in question, exhibits the highest prediction accuracy in the validation information set. [3] In order to enhance the dependability of the original choice of stocks, an overwhelming voting technique is employed to amalgamate all three requirements. The economic headline of every company is projected onto a shared feature space that is constructed based on sentiment measurements. Forecasts regarding stocks are generated through algorithms that have undergone training utilizing the data present in the feature space of the model.

Section 2 of the research paper analyzes prior studies. Section 3 provides a comprehensive analysis of the approach, whereas section 4 concentrates on the assessment of the outcome. Ultimately, the fifth section serves as the concluding segment of this article and provides suggestions for potential avenues of inquiry in subsequent studies.

II. LITERATURE SURVEY

S. Bouktif et al. [4] have presented a methodology that involves the extraction of various text-based features to enhance the conveying of emotions. Ultimately, the approach involves aggregating each model in order to optimize the performance of base stock orientation classifiers. This is achieved through the implementation of various methods for selecting features, which enable the contextual choice of suitable feature sets for varying scenarios. The findings of the experimental study indicated that the stock market failed to adhere to a standard distribution. This was evidenced by the differing degrees of efficacy demonstrated by distinct machine learning computations and feature selection techniques when applied to distinct stocks [4]. The study's authors deduce that the sentiments and opinions of the general public, as conveyed through Twitter, have an impact on the stock valuation of a corporation. In order to make precise predictions regarding the fluctuations of the stock market as a whole, the authors necessitate more sophisticated approaches to sentiment analysis.

The introduction of a novel technique by M. Wen et al. involves the utilization of high-order structures, also known as motifs, for the purpose of reestablishing time series. This approach has been proposed as a means of predicting developments in economic time series. The utilization of convolutional neural networks is employed to acquire knowledge of the fundamental patterns present in the reconstructed sequence. This knowledge can subsequently be utilized to predict forthcoming fluctuations [5]. In contrast to previous studies utilizing sequential models such as recurrent neural networks, the suggested methodology exhibits superior performance in computational efficiency. The recommended technique's authority over real time series of financial companies datasets has been demonstrated by empirical results, which also confirm its effectiveness in acquiring pattern knowledge regarding stock shares. The technique that has been developed offers a novel perspective on the prediction of prices and provides insight into the identification of macro-level sequences in the financial time series.

Alsubaie et al. conducted a comprehensive investigation into the selection of a limited number of technical indicators (TIs) for the purpose of predicting the direction of stock prices. The act of selecting over 30 TIs was found to have a negative correlation with accuracy of predictions, an increase in incorrect categorization cost, and a decrease in investment return, as per a study conducted [6]. Regarding the particulars, the data indicate that the highest levels of accuracy were attained with a minimum of 10 TIs, the least expensive scores were achieved alongside 5 TIs, as well as the optimal investment achievement was observed with either 5 or 10 TIs. The study utilized an investing simulator to showcase that the classification methods with the greatest precision scores or the smallest expense scores did not yield the most favorable investment outcome. The execution of cost-sensitive adjusting of the NB classifier (CSFTNB) was determined to offer the optimal balance among possible earnings and possible harm. The CSFTNB classifier, as proposed, enhances the conventional NB classifier through the incorporation of a fine-tuning stage that facilitates the reduction of classification expense by adjusting the value of probabilities.

Chen et al. put forward a pattern forecasting model (TPM) that operates in two phases with regard to equity markets. During the initial phase of data preparing, the authors utilize the PLR approach and CNN to segregate the dual characteristics that signify the historical data's long-term pattern from the fundamental market knowledge that pertains to the short-term. The authors propose a new framework for time series simulation that integrates a short-term characteristic encoder and a long-term feature decoder. The researchers incorporated the attention process throughout the encoder and decoder to effectively select and incorporate the most important elements of features throughout all temporal points. Finally, according to reference [7], the TPM has the ability to accurately predict both the direction and duration of trends. The findings of the experiments demonstrate that the utilization of our proposed TPM results in a reduction of RMSE by 13.74% and 17.63% when contrasted with conventional approaches.

S. M. Idrees et al present an introduction to time series assessment and projections, with a focus on the Indian economy. In light of the significant depreciation of the Indian rupee, precise prediction of stock market trends has become crucial in safeguarding investor funds. The objective of this study is to develop a proficient ARIMA model for the purpose of forecasting the swings of the Indian stock market. The present study utilized publicly available time series information compared to the Indian stock market [8]. Upon comparing the anticipated and factual time series, it was observed that the Nifty and Sensex exhibited a mean proportion discrepancy of approximately 5%. Various tests can be employed to verify the precision of the projected time series.

In their study, S. Kim et al. utilized a sample of 55 stocks from 11 industries and examined six separate instances of recessions in the US market. Their findings indicate that the time-varying ETE, as determined by the 3M as well as 6M proceeding windows, possesses substantial explaining power with respect to market behavior. Temporarily, we observe an increase in industries associated with crisis management and the substantial influx of data across the market. Ultimately, across all evaluated scenarios involving the LR, MLP, RF, XGB, and LSTM models, the incorporation of ETE network indications as supplementary features has been shown to improve the accuracy of stock price forecasting. The period of prediction and precision thereof exhibit enhancement with the passage of time and a reduction in the lag of prediction, as noted in reference [9]. In the fourth instance, the authors employ a novel outcomes measurement, namely adjusted accuracy, which is centered on the concept of risk-adjusted reward. Based on this metric, the authors conclude that the MLP and LSTM are the most effective machine learning algorithms for predicting the trajectory of stock prices. The sectors in which ETE network variables can be of utmost utility have been disclosed.

In their publication, D. Lien Minh et al presented a novel model known as the Two-stream Gated Recurrent Unit (TGRU), that exhibits superior performance compared to its predecessors, namely the GRU and LSTM representations. The proposed model's ability to absorb information in both forward and backward instructions enhances its capacity to acquire pertinent data, a crucial aspect in addressing text processing concerns [10]. The Stock2Vec embedding is constructed utilizing a financial dataset alongside the Harvard IV-4 sentiment a lexicon In comparison to conventional embedding strategies like Glove and Word2Vec, the Stock2Vec embedding exhibited superior performance in evaluations due to its incorporation of the emotional value of individual words. The incorporation of financial metrics is significant as they constitute a fundamental component of a dependable financial evaluation. The model's efficacy is demonstrated by its capacity to precisely forecast the S&P 500 measure and the patterns of specific corporations. The author's technique enabled investors to accurately measure their returns with greater precision. This approach demonstrated resilience in the face of fluctuations in markets and was flexible enough to account for the risk inherent in the current market through simulation.

K. Huang et al. [11] propose a neural network framework utilizing multilayer graph attention to predict market trends. The authors have developed a novel model based on a graph-based neural attention network to address the limitations of current forecasting stock methods. This model integrates various types of data, including financial market data, information, and operations between businesses, through a specialized feature extraction component. ML-GAT employs a range of attention mechanisms at different levels of complexity to acquire knowledge about the characteristics of nodes which are advantageous for forecasting duties. This is achieved by specifically screening diverse types of data to create a consolidated graph, which aids in the learning of node feature representations. The utilization of graph-based learning is anticipated to enhance the precision of predictions. In order to evaluate the effectiveness of the suggested approach, the author compares ML-GAT with established benchmark algorithms that have been developed using publicly-accessible datasets.

Kim et al. [12] aim to improve the efficacy of stock price forecasting through the incorporation of entropy-driven variables through machine learning methodologies. Considerable scholarly effort has been devoted to amalgamating network-driven indicators, yet comparatively little has been undertaken to explore the potential of entropy in the context of stock price prediction. Previous studies on TE have primarily focused on clarifying the Granger-causal relationship among different monetary variables. Insufficient research has already been conducted on the analysis of technological efficiency (TE) and economic-technological efficiency (ETE) at the intra-sector as well as dynamic-interval levels. Previous studies have primarily concentrated on inter-market along with static-interval evaluation. The authors have utilized a moving-window methodology to examine 55 corporations spanning 11 businesses, with a particular focus on the financial sector of the United States, which is the largest individual market in the worldwide banking system. The shifting window approach enables the monitoring of fluid and time-varying phenomena across a spectrum of crisis alongside non-crisis periods through the utilization of interval assessment.

III PROPOSED WORK

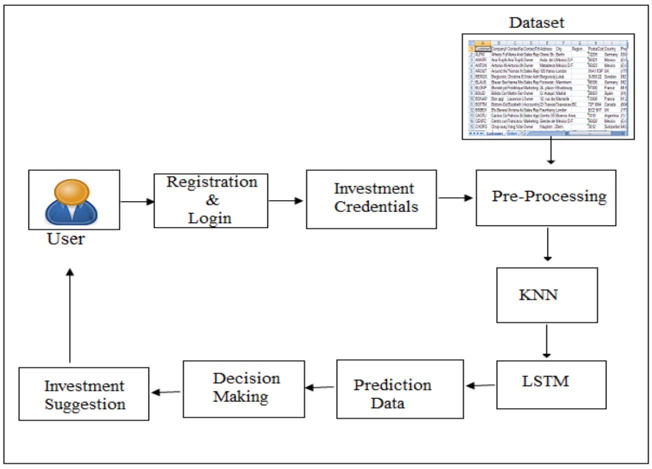

Figure 1: System overview

The research study outlines the methodology for the investment recommendation utilizing a Deep Learning model which is shown in the figure 1 above. The following steps provide a comprehensive description of the proposed approach.

Step 1: Data Gathering – The proposed approach for providing investment recommendations initiates with an interactive online platform that is purposefully designed to facilitate the submission of a dataset comprising stock-related information by the administrator. The web-based software was created through the utilization of Java programming language, utilizing a JSP page and subsequently deploying it onto a Glassfish server situated on the local host.

The administrator employs the webpage to register for the system by means of the sign-up feature. Upon selecting this option, a webpage containing diverse attributes of the administrator is exhibited in a structured format. To complete this form, one must input pertinent details such as a username, mobile number, email address, password, and other required information. Upon successful validation and storage of the aforementioned details in the database, the administrator is able to utilize the sign-in functionality to gain entry into the system. This system affords the administrator the capability to input datasets, which will be further elaborated upon in the subsequent step of the approach.

Step 2: Dataset Reading and Preprocessing – The file picker has been initiated and is presently capable of accommodating the dataset integration into the system. The aforementioned procedure entails making use of the input dataset procured via the Yahoo Finance service. The Yahoo Finance infrastructure enables the expeditious retrieval of stock data. The dataset contains current data on selected stocks that are utilized for the purpose of making predictions. The dataset is acquired for a specified duration and presented in the shape of a comma-separated values document, which is then provided to the system for further examination.

The utilization of JXL libraries by the Java program enables the access and interpretation of the dataset in the workbook format. The dataset's details is acquired and afterwards transformed towards a list format, that is eventually processed through preprocessing techniques. For the particular objective under consideration, a subset of characteristics has been chosen from the dataset for the intent of extraction, whereas the rest of the characteristics have been eliminated. The previously mentioned attributes encompass the stock market, factor, objective, duration, investmonitor, and initial and concluding values.

After the provision of the dataset to the system and its appropriate preprocessing and transformation through an operational format for the source code, the user can utilize the interactive web page to enter user attributes. The first step for the user involves completing the registration process via the specially created sign-up a page on the web page. The individual employs the web form to enter personal information, including their username, date of birth, mobile number, email address, and password. This data is then verified and saved in the database. The user is able to use the login information to gain access to the system through the webpage. After the successful verification of the login credentials, the system authorizes the user to enter their user attributes to facilitate forecasting.

The above-mentioned features and the associated numerical representations are selected and transmitted to the subsequent phase in order to classify them according to the user-provided data inputs from the previous step through the application of an interactive web page. Subsequently, the ensuing procedures are executed utilizing the Python programming language.

Step 3: K-nearest Neighbor Clustering – The current phase of the approach entails the acquisition of user feedback and preprocessed dataset obtained from the prior stages. The data provided above is being employed to estimate the spatial distance between the user's input and each row in the preprocessed data set. The calculation of distance is determined through the utilization of equation 1, as demonstrated.

$$ED = \sqrt{\sum (ATi - ATj)^2} \quad \_\_\_\_\_\_\_\_ (1)$$

Where,
ED=Euclidian Distance
A_{Ti}=Attribute at index i
A_{Tj}= Attribute at index j

Following the computation and addition of the distance from every row to its equivalent row, the resultant list is organized in ascending order utilizing the bubble sort algorithm. The present instance involves a value of k that has been established as 2, thereby resulting in the configuration of two distinct clusters. Subsequently, the clusters are forwarded to the subsequent stage to generate suggestions utilizing the LSTM model. Algorithm 1 presents a concise summary of the complete procedure of K-Nearest Neighbors (KNN).

ALGORITHM 1: KNN Classified Cluster Formation

//Input : Sorted Distance List SD_L,
//Output:Cluster List K_{CL}
1: ***Start***
2: $I_L = \emptyset$ [Inner Layer] $O_L = \emptyset$ [Outer Layer], $K_{CL}=\emptyset$
3: MIN= 0 , MAX=SDL_{SIZE-1}
4: K= (MAX-MIN) /2
5: K=MIN+K
6: ***for*** i=0 to Size of SD_L
7: R = $SD_{L\,[i]}$
8: ***if***(i<=K), ***then***
9: $I_{L} = I_{L+R}$
10: ***else***
11: $O_L = O_{L+R}$
12: ***end for***
13: $K_{CL[0]} = I_L$
14: $K_{CL[1]} = O_L$
15: return K_{CL}
16: ***Stop***

Step 4: Long Short Term Memory – The clusters obtained from the previous step are effectively taken as an input along with the dataset that has been previously preprocessed. These elements are utilized for the purpose of attaining the prediction through the use of the long short term memory. The initial step in this approach is the splitting of the dataset into the training and testing datasets. The dataset consists of attributes such as date, open, high, low, close and adj close. This dataset is divided in a 67:33 division into two parts, one (67) is the training part that is used to train the LSTM model and the testing part (33) which is used to test the trained model.

Once the split is done, the min max scaler and the transform data is performed. The calculation is achieved by the minmaxscaler on both the training and testing splits to achieve the transformation of the data. The transformation generates effective data that can be provided as an input to the LSTM model for further processing.

The transformed data before being applied to the LSTM model the data needs to be reshaped. The reshaping accurately converts the data from a 2-dimensional data to a 3-dimensional data. The training split is first reshaped into a 3-dimensional data following which the testing dataset is reshaped into a 3-dimensional shape. The data now can be directly utilized into the LSTM model.

The LSTM model is initiated in a sequential manner by adding the layers sequentially. The first layer in the model utilizes 50 units with the return sequences set to true and the input shape is the shape of the modified 3-dimensional input training data generated previously. This layer is followed by the dropout layer set at 0.2. The next layer initializes 50 units and the return sequences set to true along with a dropout layer set at 0.2. Another layer with 50 units and the return sequences enabled with another dropout layer set at 20%. The next layer for LSTM is set with 50 units accompanied with a dropout layer with similar parameters. The model is then culminated in a dense layer with one unit as the parameter. Once the model is achieved as a .h5 file it is being used perform the prediction on the user data, which results probability scores in a list that needs to be classified through decision making in the next step of the approach.

Step 5: Decision Making – This step involves the ingestion of a sorted list that comprises the probability scores attained from the LSTM operation. The probability scores generated by the user may not be directly employed as they fail to accurately reflect the desired outcome. Thus, it is imperative to appropriately categorize these probability scores in order to derive pertinent forecasts for the stock market. The method of decision-making utilizes precise if-then rules to categorize the probability scores and generate appropriate investment suggestions. The webpage graphical user interface presents the user with categorized suggestions.

[1] IV RESULTS AND DISCUSSIONS

The current investigation unveils a proposed approach for attaining investment suggestions through the implementation of machine learning methodologies. The methodology was devised employing the Java programming language. The development machine consists of a standard configuration, which includes an Intel Core i5 processor, eight gigabytes of RAM with random access, and one terabyte of solid-state drive storage. The methodology has been developed to employ a dataset of movements in stocks as input and implement algorithms grounded in machine learning for creating forecasts.

Performance Evaluation based on Precision and Recall

The assessment of a component's precision and recall measurements is a successful means of measuring its performance reliability within a given framework. The precision of an element is a determining factor in its relative accuracy, which comprises a wide range of reliability.

The present methodology ascertained the precision metric by computing the ratio of correct investment recommendations acquired in relation to the total number of conducted trials. The recall criteria function as a complementary measure to the precision metric, aiding in the determination of the overall dependability of the LSTM component.

The recall can be computed using this approach, which involves dividing the number of correct investment suggestions by the overall number of incorrect investment suggestions. The subsequent equations offer a numerical elaboration of the previously mentioned notion.

Precision and Recall can be depicted as below:

- ✓ X = The number of accurate investment suggestions
- ✓ Y= The number of inaccurate investment suggestions
- ✓ Z = The number of accurate investment suggestions not done

So, precision can be defined as

Precision = (X / (X+ Y)) *100
Recall = (X / (X+ Z)) *100

Table 1 presents the precise experimental outcomes obtained using the previously mentioned equation. The utilization of statistical elements is being used to provide a graphical depiction, as demonstrated in Figure 2.

No. of Trials	Accurate Investment Suggestions (X)	Inaccurate Investment Suggestions (Y)	Accurate Investment Suggesstions not done (Z)	Precision	Recall
1	1	0	0	100	100
4	3	0	1	100	75
7	6	1	0	85.71429	100
10	8	2	0	80	100
13	10	1	2	90.90909	83.33333
16	13	2	1	86.66667	92.85714
19	17	1	1	94.44444	94.44444
22	17	3	2	85	89.47368
25	20	3	2	86.95652	90.90909
28	24	2	2	92.30769	92.30769

Table 1: Precision and Recall Measurement Table

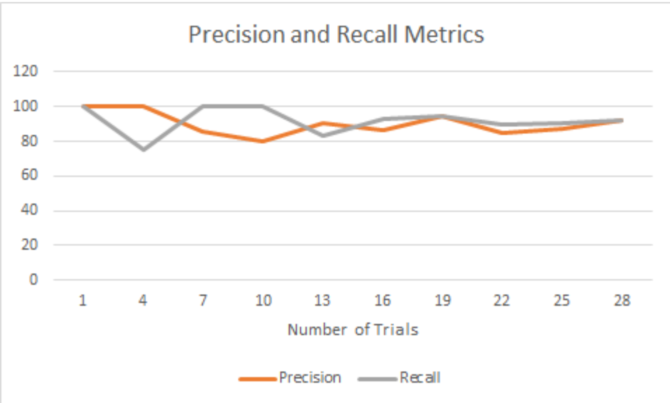

Figure 2: Comparison of Precision and Recall

1

The graph presented depicts the effectiveness of the LSTM model and its associated suggestions, as determined by input data, across various trial counts. The approach's exceptional precision and recall scores of 90.19 percent and 91.83 percent, respectively, serve as a demonstration of its remarkable accuracy. The numerical values are of significant importance in the initial implementation of the stated process and have produced an acceptable result for the investment suggestions.

V CONCLUSION AND FUTURE SCOPE

The present research article delineates the proposed methodology for attaining accurate and precise investment recommendations. The forecasts were generated through the application of machine learning techniques. The procedure initiates with the administrator's authentication and authorization for using the system, followed by the submission of the dataset that pertains to the stock market patterns of Yahoo Finance as the input data. The dataset underwent preprocessing procedures to remove extraneous components and inadequate data points. Following this, the preprocessed dataset is provided for further evaluation. Following this, the user moves forward to gain entry into the system through the use of the designated login information that were provided over the registration procedure. After the user has been successfully authenticated, they begin submitting their user characteristics to enable prediction. Subsequently, the aforementioned attributes are merged with the dataset provided by the administrator and analyzed using the K-Nearest Neighbors method. Preceding any examination, it is imperative to group the information into clusters. The K nearest Neighbors technique will cluster the preprocessed dataset and user attribute information in a subsequent manner. The clusters derived from the previous stage serve as an input for the Long Short Term Memory (LSTM) model, which generates probability scores that necessitate proficient categorization for the purpose of achieving precise recommendations. The Decision Making approach entails the process of classifying the possible outcomes, which are then proffered to the consumer as an investment suggestion. The methodology has been evaluated for its ability to forecast inaccuracies using precision and recall metrics, and the outcomes are deemed satisfactory.

In order to improve the future strategy of the research, it is advisable to enhance the approach by enabling real-time operation on a platform that is hosted in the cloud, which would additionally enhance the user experience.

REFERENCES

[1] C. Bousoño-Calzon, H. Molina-Bulla, J. J. Escudero-Garzas and F. J. Herrera-Galvez, "Expert Selection in Prediction Markets With Homological Invariants," in IEEE Access, vol. 6, pp. 32226-32239, 2018, DOI: 10.1109/ACCESS.2018.2846878.

[2] T. Kabbani and E. Duman, "Deep Reinforcement Learning Approach for Trading Automation in the Stock Market," in IEEE Access, vol. 10, pp. 93564-93574, 2022, DOI: 10.1109/ACCESS.2022.3203697.

[3] X. Li, H. Xie, R. Y. K. Lau, T. -L. Wong and F. -L. Wang, "Stock Prediction via Sentimental Transfer Learning," in IEEE Access, vol. 6, pp. 73110-73118, 2018, DOI: 10.1109/ACCESS.2018.2881689.

[4] S. Bouktif, A. Fiaz, and M. Awad, "Augmented Textual Features-Based Stock Market Prediction," in IEEE Access, vol. 8, pp. 40269-40282, 2020, DOI: 10.1109/ACCESS.2020.2976725.

[5] M. Wen, P. Li, L. Zhang, and Y. Chen, "Stock Market Trend Prediction Using High-Order Information of Time Series," in IEEE Access, vol. 7, pp. 28299-28308, 2019, DOI: 10.1109/ACCESS.2019.2901842.

[6] Y. Alsubaie, K. E. Hindi and H. Alsalman, "Cost-Sensitive Prediction of Stock Price Direction: Selection of Technical Indicators," in IEEE Access, vol. 7, pp. 146876-146892, 2019, DOI: 10.1109/ACCESS.2019.2945907.

[7] Y. Chen, W. Lin, and J. Z. Wang, "A Dual-Attention-Based Stock Price Trend Prediction Model With Dual Features," in IEEE Access, vol. 7, pp. 148047-148058, 2019, DOI: 10.1109/ACCESS.2019.2946223.

[8] S. M. Idrees, M. A. Alam, and P. Agarwal, "A Prediction Approach for Stock Market Volatility Based on Time Series Data," in IEEE Access, vol. 7, pp. 17287-17298, 2019, DOI: 10.1109/ACCESS.2019.2895252.

[9] S. Kim, S. Ku, W. Chang, and J. W. Song, "Predicting the Direction of US Stock Prices Using Effective Transfer Entropy and Machine Learning Techniques," in IEEE Access, vol. 8, pp. 111660-111682, 2020, DOI: 10.1109/ACCESS.2020.3002174.

[10] D. Lien Minh, A. Sadeghi-Niaraki, H. D. Huy, K. Min, and H. Moon, "Deep Learning Approach for Short-Term Stock Trends Prediction Based on Two-Stream Gated Recurrent Unit Network," in IEEE Access, vol. 6, pp. 55392-55404, 2018, DOI: 10.1109/ACCESS.2018.2868970.

[11] K. Huang, X. Li, F. Liu, X. Yang, and W. Yu, "ML-GAT: A Multilevel Graph Attention Model for Stock Prediction," in IEEE Access, vol. 10, pp. 86408-86422, 2022, DOI: 10.1109/ACCESS.2022.3199008.

[12] S. Kim, S. Ku, W. Chang, and J. W. Song, "Predicting the Direction of US Stock Prices Using Effective Transfer Entropy and Machine Learning Techniques," in IEEE Access, vol. 8, pp. 111660-111682, 2020, DOI: 10.1109/ACCESS.2020.3002174.

*****