Pakistan Journal of Life and Social Sciences

www.pjlss.edu.pk

https://doi.org/10.57239/PJLSS-2024-22.1.009

RESEARCH ARTICLE

Prediction of Stock Market Movement Using Long Short-Term Memory (LSTM) Artificial Neural Network: Analysis of KSE 100 Index

Atif Khan Jadoon [1*], Tariq Mahmood [2], Ambreen Sarwar [3] Maria Faiq Javaid [4], Munawar Iqbal [5]

[1,2,3,4]Department of Economics, University of the Punjab, Lahore, Pakistan

[5]College of Statistical Science (CSS), University of the Punjab, Lahore, Pakistan

ARTICLE INFO

Received: Oct 26, 2023

Accepted: Feb 14, 2024

Keywords

Artificial Neural Network (ANN)

Backpropagation algorithm

KSE 100 index

Karachi stock exchange

Long Short-Term Memory (LSTM)

Pakistan

***Corresponding Author:**

atifkhan.eco@pu.edu.pk

ABSTRACT

Predicting the movement of the stock price index has remained a challenging task for financial analysts. They are complex, given how the nation's economic, social, and political conditions affect stock market predictions. The present study is designed to build an efficient model using machine learning and taking monthly data from February 2004 to December 2020. Long- and short-term memory and backpropagation methods of an artificial neural network are utilized to predict the Karachi Stock Exchange (KSE) movement by using twenty-six economic, social, political, and administrative indicators. The model developed in this study to predict the movement of the KSE 100 index gained 99 percent accuracy. The predictions' results showed that the KSE 100 index would remain stagnant at around 40,000 points till September 2023. We compared actual and predicted values from January 2021 to July 2022 to validate our developed model. The results showed that the model developed in this study could be used to forecast stock market trends. The research predicted the KSE 100 index with 99% accuracy using LSTM and backpropagation in a neural network. The KSE 100 should stay around 40,000 points until September 2023. Despite their promise, machine learning forecasts in financial markets require continuing research, market dynamics evaluation, and liability disclosure. A thorough validation procedure and the most recent KSE 100 index prediction using LSTM and backpropagation algorithms demonstrate how machine learning may improve stock market forecasting. The study's accurate projections and consistent data through September 2023 are crucial for investors and financial experts in the Karachi Stock Exchange's changing environment.

INTRODUCTION

The stock market is a vital component of the financial sector that substantially impacts a nation's advancement (Demir, 2019). The stock market is recognized as the mechanism through which financial resources are redistributed among various economic entities. The advancement in the stock market is regarded as an indicator of the economic expansion of a nation (Arestis and Demetriades, 1997; Kolapo and Adaramola, 2012; Pan and Mishra, 2018). The stock market is an efficient platform for raising money for individuals, corporations, and states. Being highly liquid, it provides easy stock transactions among interested stakeholders. Capital and financial market development in any economy is primarily determined by the number of people engaged in stock market transactions and the stability of stock market trends (Chhajer et al., 2022). The stock market's fluctuations are a barometer for a nation's economic condition; a positive performance indicates economic expansion, while a negative performance suggests the opposite. When the stock market exhibits a bullish trend, more investors seeking profit opportunities become interested in issuing Initial Public Offerings (IPO) to generate capital. Increased investment thus leads to higher economic activity and growth (Chhajer et al., 2022; Zaffar and Hussain, 2022).

Many factors determine the performance of the stock market in any country. It is empirically tested that along with technological development and shareholder interest, several macroeconomic indicators and the administrative quality of a state do play a significant role in the stock market performance of the country (Adams and Klobodu, 2016;Asongu, 2012; Dima et al., 2018;). Most researchers in the past have found various macroeconomic indicators, including interest rate, money demand, exchange rate, gold prices, oil prices, and terms of trade, to be significant in their impact on stock market performance (Al-Ameer et al., 2018; Arjoon et al., 2012). These economic, political, administrative, and social factors directly or indirectly affect the investor's decision to invest in the country, resulting in an increase or decrease in the value of stocks. This leads to fluctuations in the value of stock market indices (Gori et al., 2023; Lehkonen and Heimonen, 2015; Santa-Clara & Valkanov, 2003). This uncertainty in the stock market imposes restrictions on the financial development of corporations and states (Asongu, 2012; Azeem et al., 2017; Demir, 2019; Ali et al., 2023). If investors are perfectly aware of changes in the value of a stock, they will gain maximum benefits by investing in stocks that are expected to increase in value. To avoid losses and gain maximum profits, investors search for techniques and tools to predict stock market movements (Ali et al., 2021; Sahu et al., 2023; Ali et al., 2023).

However, predicting the stock market is tricky because of its dynamic, stochastic, and uncertain nature (Khan et al., 2020; Parmar et al., 2018). Two primary analytical methodologies are employed in forecasting stock market trends: technical analysis and fundamental analysis. A researcher identifies the primary elements influencing the stock market in basic analysis. On the other hand, in technical analysis, a researcher can predict stock market trends by using stock prices and their trends. Predicting the stock market trend is a difficult task due to the nonlinear pattern of the stock market (Rather et al., 2015; Vaisla and Bhatt, 2010). The utilization of computational intelligence and machine learning (ML) techniques to forecast the fluctuations of the stock market has experienced a substantial rise (Kumbure et al., 2022). ML is employed to analyze and identify patterns within data. ML techniques efficiently analyze complicated data and produce accurate findings on time (Rouf et al., 2021). Artificial Neural Networks (ANN) are now the predominant approach in the literature for analyzing nonlinear pattern data (Agatonovic-Kustrin and Beresford, 2000; Rao and Reimherr, 2023). The ANN backpropagation technique is widely employed for reliable prediction of future data (Law, 2000; Siregar and Wanto, 2017). This research analyzes the data from the Pakistan stock market. It examines Pakistan's key macroeconomic and social, political, and administrative quality indicators to develop the model with the help of an Artificial Neural Network (ANN) backpropagation algorithm. Moreover, based on the developed model, the study also predicts the KSE 100 index value until September 2023.

In recent years, emerging economies have experienced significant expansion in their stock markets, presenting profitable opportunities for investment (Bosworth et al., 1999; Stiglitz, 2000). Pakistan is a developing country facing numerous challenges on the economic and political front. Pakistan's stock market performance has recorded periods of boom and bust in the recent past. Pakistan's stock market remained unstable due to the COVID-19 outbreak, the Russia-Ukrain war, and domestic political instability.

Predicting the behavior of Pakistan's stock market is crucial as it may enhance investors' trust, mitigate macroeconomic volatility, and stimulate economic growth (Ghani et al., 2022). Few studies have been conducted to predict the behavior of the stock market in Pakistan. Ali et al. (2021) predicted the direction of movement of the KSE 100 index along with the other three countries, including China, Japan, and Korea, using data from 2011 to 2020. The authors used single-layer ANN and Support Vector Machine models where opening, closing, high, and low prices were used as input layers. The study only compared the prediction accuracy of both models and concluded that ANN has better prediction accuracy. The authors suggested taking more macroeconomic indicators as input layers and using a Long-Short-Term Memory Model (LSTM) with a multilayer model.

Similarly, Jawad and Iqbal (2020) predicted Pakistan's stock market index by introducing a deep neural network model using technical indicators. The study also compared the classification accuracy with the other previously conducted studies. Usmani et al. (2018) predicted the stock market's performance. The authors used only four macroeconomic indicators, general public sentiment, and related NEWS variables. The author constructed four models based on statistical techniques (simple moving average and auto-regressive integrated moving average) and artificial neural network techniques (Single-layer perceptron and multilayer multilayer perceptron). The result from these four models merged into a single hybrid model. The study concluded that two hybrid models predicted results with less accuracy than the MultilayerMultilayer Perceptron-based sub-model. The authors also suggested using complex models like LSTM to predict stock market behavior accurately.

The present study makes the following contributions: Firstly, it has taken twenty-six economic, political, social, and administrative quality variables to develop a model to predict the movement of the KSE-100 index. No other study in the literature (to the best of the researcher's knowledge) has used such an extensive data set on various variables that directly affect stock market trends. In the literature, most studies have taken limited macroeconomic variables or political indicators to see stock market behavior. Secondly, the present study used an LSTM with three hidden layers. In the case of Pakistan, only one study has used multilayer multilayer perception with only six indicators to see stock market trends. Lastly, the model developed in this study can be utilized as a policy guide for the state of Pakistan and is beneficial for private investors. The novel use of machine learning algorithms like Long Short-Term Memory (LSTM) and backpropagation to anticipate stock markets utilizing a wide range of economic, social, political, and administrative data makes the work theoretically significant. The study advances financial market prediction modeling by predicting the KSE 100 index with 99% accuracy. This research enhances financial forecasting analytical techniques and emphasizes the need to include many factors to account for the complex impacts on stock prices. The temporal linkages of financial time series may be better understood using LSTM, a sequential data processing architecture.

MATERIAL AND METHODS

Data

The present study has used monthly data from February 2004 to December 2020. A total of twenty-six independent variables were used to predict the stock market movement of Pakistan, namely Balance of trade, Consumer financing for house building, Consumer price index (a proxy for inflation), Control of corruption, crude oil, Domestic savings, Exchange rate, External debt stocks, Foreign direct investment, Foreign exchange reserves, GDP growth rate, Gold price, Government effectiveness, household final consumption expenditure, Industrial production index, Industry value added, Labor force participation rate, Money supply, Personal remittances growth, Political stability and absence of violence or terrorism Portfolio investment Growth rate, regulatory quality, the rule of law, the treasure bill on the three-month month's treasury, and the wholesale price index. The closing price of the KSE 100 index is an output variable to measure stock market behavior. The data on the above-stated variables is taken from the World Development Indicator, State Bank of Pakistan, and Pakistan Bureau of Statistics. The artificial neural network is used to interpolate data from months to days.

Artificial Neural Network (ANN)

ANN is a biologically inspired neural network for replicating complex phenomena from the human brain into a machine. The input layer of neurons, the output layer of neurons, and the hidden layers of the neurons comprise ANN. Independent and dependent variables of the problem determine the number of neurons in the input and output layers. The hidden layers of neurons vary according to the complexity of the problem.

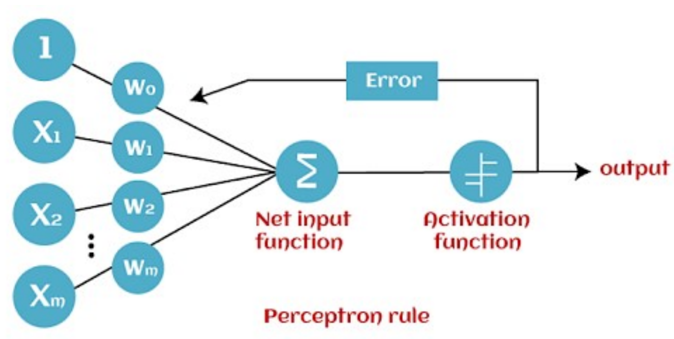

Figure 1: Working diagram of the perceptron model of ANN

Figure 1 describes the internal workings of an artificial neural network with $x_1, x_2, x_3, \ldots\ldots\ldots x_n$ inputs provided to the network. The SOMA (summing function) calculates the sum of weighted time inputs and is named a linear combiner.

$$v_k = \sum_{j=1}^{n} w_{jk} x_j$$

$$y_k = \varphi\left(v_k - \theta_k\right)$$

V_k is the weighted sum, and y_k is the final output calculated by applying suitable activation function φ.

Backpropagation algorithm

In the backpropagation ANN, the input is feed-forward from the input layer to the hidden layer and then from the hidden layer to the output layer to produce a final output of the network. Based upon the error in the final output, it sends back the error to the neurons of the output and hidden layers to remove the error in the relevant neurons to improve the overall accuracy of the results. Figure 2 represents the flow of information within AAN.

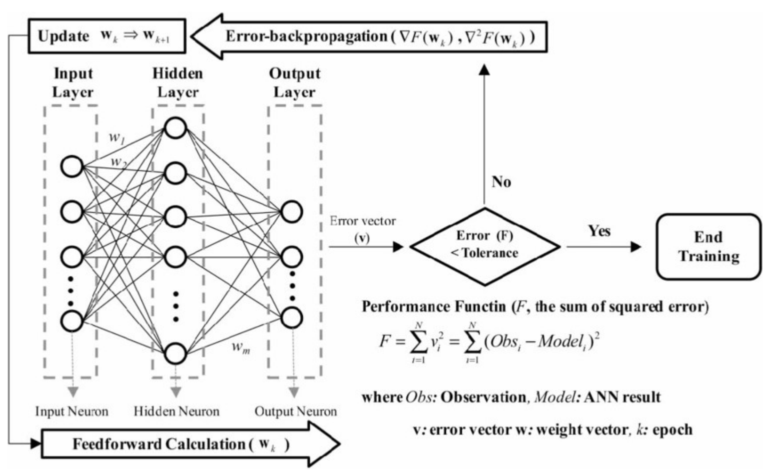

Figure 2: Schematic diagram of backpropagation artificial neural network describing the flow of information within the network (Kim and Seo, 2015)

Long Short-Term Memory (LSTM)

An ANN type called LSTM aids in processing sequential data, including time series data, text, and voice. Compared to conventional recurrent neural networks, LSTM has a more complicated structure that enables it to recall data from earlier time steps for a more extended period using a "memory cell," which can selectively forget or remember information based on input signals. Within the LSTM, several gates, such as the input, output, and forget gates, regulate the information flow and selectively store and discard the information. Due to its ability to handle long-term dependencies and avoid the vanishing gradient issue, LSTM has grown increasingly common in various research fields, including time-series predictions, voice recognition, and natural language processing. Figure 3 shows how the LSTM is structured. Figure 3 represents the structure of LSTM.

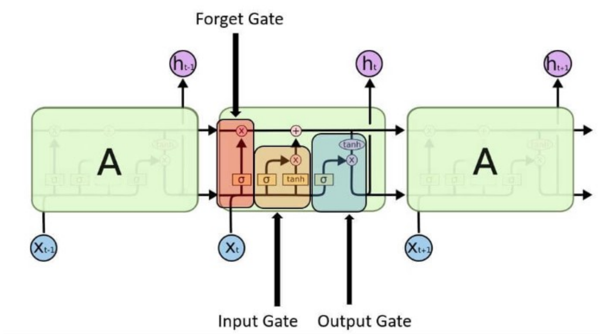

Figure 3: Schematic diagram of LSTM networks capable of learning long-term dependencies, especially in sequence prediction problems

In our research, the utilization of the ANN and LSTM techniques for data analysis is justified based on the intricate and nonlinear nature of the dataset. The complexity and nonlinear patterns within the data make neural networks, particularly LSTM, well-suited for capturing such dependencies, especially in scenarios involving time-series data with long-range dependencies. The chosen approach excels in automatically learning relevant features from the data, eliminating the need for extensive manual feature engineering, and it exhibits scalability to handle large datasets effectively. Furthermore, the adaptability of ANN + LSTM models to different data types and their capacity for generalization to unseen data positions them as versatile tools. While acknowledging the value of conventional statistical techniques, the decision to employ ANN and LSTM is substantiated by their demonstrated success in similar applications and their ability to outperform traditional methods in scenarios where intricate patterns and substantial datasets are prevalent.

METHODOLOGY

The present study used the backpropagation algorithm and LSTM to predict the KSE 100 index based on 26 macroeconomic indicators.

Table 1: Values of the parameters

Parameters	Values
Learning Rate	0.0015
Maximum Epoch	5000
Input Layer	1 With 26 Input Neurons
Hidden Layers	3 With 50 Neurons each
Output Layers	1 With One Neuron.
Activation Function for Hidden Layers	Tanh
Activation Function for the Output Layer	Linear

The backpropagation algorithm and LSTM use a supervised learning approach to forecast results. The present study has used 0 to 80% of the data for training, 80 to 90% for the test, and 90% to 100% for validation. The backpropagation algorithm's multilayer perceptron model is used to guess the validation data part, and LSTM is used to think of the unknown period. This is how it is set up in Table 1.
The output in the hidden layers is determined using the following formula with a corresponding activation function.

$$y_j = f\left(\sum_i w_{ji} x_i - \theta_j\right) = f\left(\text{ net }_j\right)$$

$n^{net} t_j = \sum_i w_{ji} x_i - \theta_j$ Where y_j is an output of the hidden layers neurons, x_i is the input, and w_{ji} are the corresponding network weights.

$$z_i = f\left(\sum_i v_{ij} y_i - \theta_i\right) = f\left(\text{ net }_j\right)$$
$$net_i = \sum_i v_{ij} y_i - \theta_i$$

Where z_j is the output of the neuron of the output layer. Error in the final output is determined as

$$E = \frac{1}{2} \sum_l \left(t_l - z_l\right) = \frac{1}{2} \sum_t \left(t_l - \mathrm{f}\left(\sum_i v_{ij} y_i - \theta_i\right)\right)^2$$
$$\mathrm{E} = \frac{1}{2} \sum \left(t_i - \mathrm{f}\left(\sum_i w_{jl} x_i - \theta_j\right) - \theta_l\right)^2$$

This error is further back-propagated to the neurons of the output layers and then to the neurons of the input and hidden layers for further modification of the corresponding weights of each layer. Error in the neurons of the input layer can be calculated as $\frac{\partial E}{\partial v_{lj}} = \sum_{k=1}^n \frac{\partial E}{\partial z_k} \cdot \frac{\partial E}{\partial v_{lj}} = \frac{\partial E}{\partial z_l} \cdot \frac{\partial z_l}{\partial v_{ij}}$ Where E consists of several z_k, but only one is associated with all the autonomous from each other.

$$\frac{\partial E_{-}}{\partial z_l} = \frac{1}{2} \sum_l \left[-2\left(t_k - z_k\right) \cdot \frac{\partial z_k}{\partial z_l}\right] = -\left(t_l - z_l\right)$$
$$\frac{\partial z_l}{\partial v_{lj}} = \frac{\partial z_l}{\partial net_l} \cdot \frac{\partial net_l}{\partial v_{lj}} = f'\left(\text{ net }_l\right) \cdot y_j$$
$$\frac{\partial E}{\partial v_{lj}} = -\left(t_l - z_l\right) \cdot f'\left(\text{ net }_l\right) \cdot y_j$$
$$\delta_l = -\left(t_l - z_l\right) \cdot f'\left(\text{ net }_l\right)$$

So, the error in the neurons of the input layer: $\frac{\partial E}{\partial v_{tj}} = -\delta_l \cdot y_j$ Errors in the neurons of the hidden layers can be calculated as $\frac{\partial E}{\partial w_{li}} = \sum_l \sum_j \frac{\partial E}{\partial z_l} \cdot \frac{\partial z_l}{\partial y_j} \cdot \frac{\partial z_i}{\partial w_i}$

$$\frac{\partial E\_1}{\partial z_l} \sum_k \left[-2\left(t_k - z_k\right) \cdot \frac{\partial z_k}{\partial z_l}\right] = -\left(t_l - z_l\right)$$
$$\frac{\partial E}{\partial w_{ll}} = \sum_l \left(t_l - z_l\right) \cdot f'\left(\text{ net }_l\right) \cdot v_{lj} \cdot f'\left(\text{ net }_j\right) \cdot x_l = -\sum_l \delta_l v_{lj} f'\left(\text{ net }_j\right) \cdot x_i$$
$$\delta_j' = f'\left(\text{ net }_j\right) \cdot \sum_l \delta_l v_{jl}$$

So the error in the neurons of the input layer: $\frac{\partial E}{\partial w_j} = -\delta_j' x_i$ The output layer is as follows:

$$\Delta v_{lj} = -\eta \frac{\partial E}{\partial v_l} = \eta \delta_l y_j$$

Where η is the learming rate. The transformation from the input layer to the hidden layer is defined as:

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ji}} = \eta' \delta_j' x_i$$
$$\delta_j' = f'(\text{ net }_j) \cdot \sum_i \delta_l v_{ij}$$

Here, $\sum_l \delta_l v_{lj}$ is showing the hidden layer's node error; δ_j' is error of the of the output node; z_l is back propagated through the weight value v_{lj} to the node y_j to become the error in the hidden layer's neurons. We have utilized the LSTM, comprising an input gate, a forget gate, and an output gate, to predict the KSE 100 index based on the macroeconomic data. The input gate modifies the memory of LSTM using input, and the sigmoid function decides whether to pass 0 or 1 data. Additionally, the tanh function gives the provided data more weight depending on the relevance on a scale from -1 to 1 .

$$i_t = \sigma(W_t \cdot [h_t - 1, x_t] + b_i)$$
$$C_t = \tanh(W_c \cdot [h_t - 1, x_t] + b_c)$$

Using the sigmoid function, Forget Gate locates the data that must be removed from the block. It generates a number between 0 (omit this) and 1 (keep this) by examining the previous state (ht-1), the content input (Xt), and each number in the cell state Ct-1.

$$f_t = \sigma(W_f \cdot [h_t - 1, x_t] + b_f)$$

Output Gate determines the output by the input and memory of the block through the following equations.

$$O_t = \sigma(W_0 . [h_t - 1, x_t] + b_0)$$
$$h_t = O_t * \tanh(C_t)$$

The flow of information in the LSTM is shown in Figure 4:

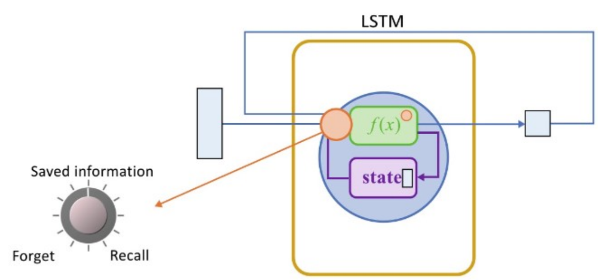

Figure 4: Flow of information in the LSTM network

RESULTS AND DISCUSSION

In this study, we predicted and forecasted the stock market trends (KSE 100 index) with the help of economic, political, social, and administrative indicators using the ANN's Backpropagation algorithm and LSTM. The model utilizes the month-end and daily closing values of the KSE 100 index.

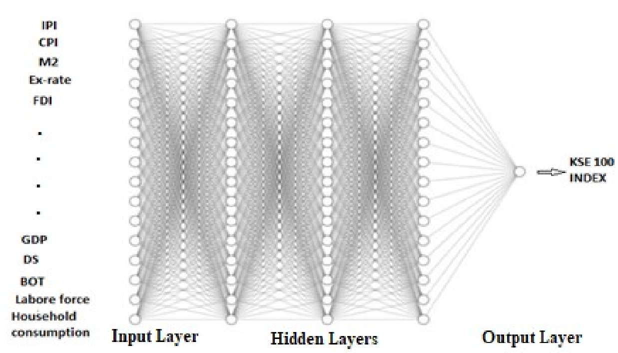

Figure 5: ANN architecture

A backpropagation of an ANN consists of an input layer, three hidden layers, and an output layer. The value of each independent variable is sent into one of the 26 neurons in the network's input layer. Each hidden layer has 50 neurons and utilizes tanh as an activation function.

In contrast, the network's output layer consists of a dependent variable (KSE 100 index). Connection weights are initialized using random distribution. The detailed structure of the backpropagation neural networks is shown in Figure 5.

In this training process, the learning rate is 0.0015, which is considered the scalar parameter used to set the rate of change. The present study has used data from February 2004 to December 2020 (203 months). Data is divided into three categories: training, validation, and testing. The training uses 80% of the data, validation 10%, and testing 10% of the data. Figure 6 shows the mean absolute error loss of the LSTM.

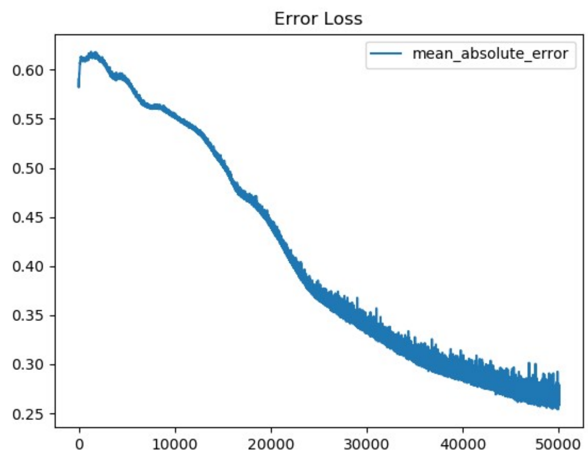

Figure 6: During training, the mean absolute error loss of the LSTM

Figure 6 shows that the mean absolute error loss of the LSTM is decreasing. In other words, a decreasing trend of mean absolute error loss shows that data accuracy increases over time. The ANN program with backpropagation algorithm developed for the present study to predict the KSE 100 index has 99% accuracy. The developed ANN model can predict the KSE 100 index values with 26 economic, social, political, and administrative variables. Since the accuracy of the ML algorithm increases with the increase in data, the results of the present study more accurately predict the KSE 100 index than the study conducted by Usmani et al.(2018).

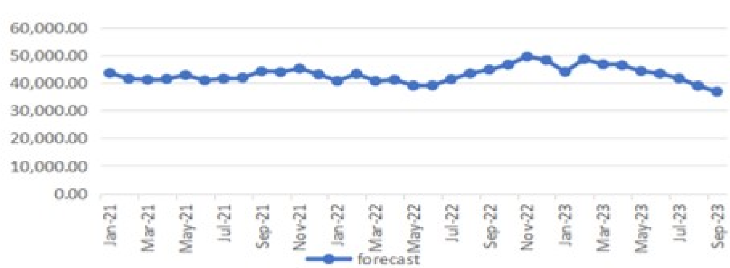

Figure 7: Forcasting of future trend of KSE 100 index using LSTM

Forecasting or predicting KSE 100 behavior for investors, financial intermediaries, and the Pakistani government was the primary goal of this study. The predicted values of the KSE 100 index using LSTM are presented in Figure 7.

In Figure 7, we predicted the value of the KSE 100 index till September 2023. Figure 7 is obtained with the help of values forecasted by LSTM based on input values of Pakistan's economic, social, political, and administrative indicators. The line graph shows that the KSE 100 index will exceed 50,000 in September 2023. This is mainly due to Pakistan's bad macroeconomic situation in the last twenty years. Moreover, Pakistan's social, administrative, and political crisis could be helping the stock market perform better. The political instability has remained one of the significant reasons for Pakistan's economic collapse. Only a few prime ministers in the history of Pakistan have completed their five-year term. Pakistan is facing multiple issues, including political crises, payment balance issues, shortage of foreign exchange, and foreign debt, that have caused investors to lose confidence in Pakistan (ProPK, 2022). Presently, Pakistan has implemented such policies, which caused the shutdown of the industrial sector of Pakistan. According to the All-Pakistan Textile Mills Association, 1600 textile mills have shut down their shops. Due to this, roughly five million people will lose their jobs as textiles are a major export of Pakistan.

Fourteen listed companies announced shutting operations from September 2022 to December 2022 due to low demand and high energy prices (Ahmed, 2023). The relatively high interest rate has also reduced investment significantly in Pakistan. In January 2023, the inflation rate increased to 27.6%, and foreign exchange reserves have been at the lowest level in the last 20 years. The expected inflation rate in Pakistan is 33% in the first half of 2023 (Bhat and Peshimam, 2023). According to Turak (2023), the energy crises, corruption, and currency devaluation brought Pakistan to collapse. These economic, political, and social issues also augment and affect the KSE 100 index. Our results through LSTM also suggest that shortly, the economy will see a slight improvement, and the value of the KSE 100 index will be less than the 40,000 mark in September 2023.

The actual and predicted values from January 2021 to July 2022 are compared to check the accuracy of the prediction made by our developed model in this study through ANN Backpropagation. The trends of actual and forecasted values of the KSE 100 index are presented in Figure 8.

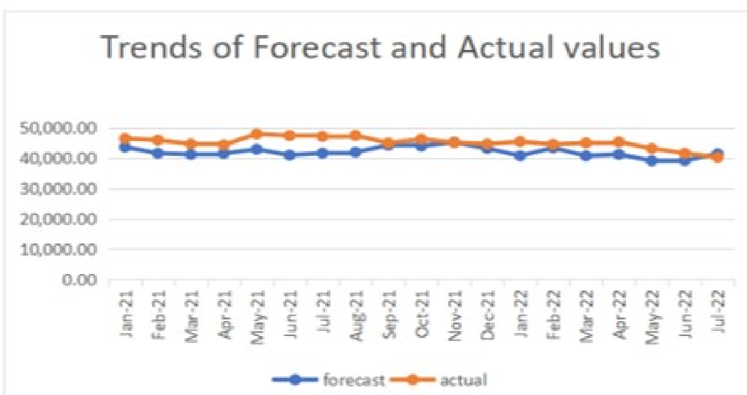

Figure 8: Line chart of predicted and actual values of the KSE 100 index according to the division of data provided to the network

The actual values of the KSE 100 index are represented with the orange line, while a blue line denotes the forecasted values of the KSE 100 index. The predicted and actual values are very close, or even in November 2021, both are the same. Figure 8 shows that the model developed through ANN Backpropagation could be used to forecast the value of the KSE 100 index. In this way, the study can significantly help investors, financial intermediaries, and even the government of Pakistan.

Theoretical contribution

In literature, three theories explain the stock market behavior: the Efficient Market Hypothesis, the Random Walk Hypothesis, and the Dow theory. According to EMH, stock prices include all the available information, and it is impossible to predict future stock price changes. Therefore, Investors can only outperform the market and make a small profit. RWM, on the other hand, states that stock prices are random, and one cannot predict future stock prices accurately because stock prices are independent of other market factors. Both these theories suggest that stock traders cannot make profits in the short run. Dow theory explains that stock prices move in trends consisting of three phases: accumulation, markup, and distribution. The strength of a trend is determined primarily by trading volume. This theory claims that current trends in stock prices can be used to predict future movements. This theory asserts that stock prices exhibit random fluctuations in the short run. However, in the long run, underlying economic and noneconomic factors can help predict future trends in the stock market. The results of this study are consistent with Dow theory as data for different economic, political, social, and administrative factors over a more extended period is used to predict future stock market values with 99 percent accuracy. These results can be helpful for traders to gain maximum benefits from investment in the stock market.

CONCLUSION

The study provides a machine learning-based solution to the problematic stock price index prediction issue. This study uses artificial neural networks' backpropagation and Long Short-Term Memory (LSTM). These methods are notable for evaluating sequential data and constructing robust prediction frameworks. The comprehensive dataset used in this study covers monthly data from February 2004 to December 2020. The model's sizeable temporal range covers many market conditions, making stock market behavior easier to explain. The model includes twenty-six economic, social, political, and administrative components. Different indicators may impact the KSE 100 index. The method predicted KSE 100 index movements with 99 percent accuracy, the study's most outstanding achievement. Accuracy shows the model's ability to recognize and respond to stock market trends. This model forecasts a KSE 100 index of 40,000 points till September 2023. It helps financial specialists and investors handle market conditions with vital information. We compared observed and expected values for January 2021 to July 2022 to assess our model's dependability and robustness. This validation method shows our model's flexibility and extrapolation, boosting its dependability in real-world market circumstances.

Implications

The analysis reveals several practical implications that affect every financial sector. Financial professionals and people may now manage stock market dynamics using a tool. Backpropagation and LSTM make machine learning dependable for investment. The model's 99 percent KSE 100 index prediction accuracy makes it a reliable risk-reward instrument. Risk management is crucial to financial decision-making, and this predictive model helps. It predicts and manages market volatility using economic, social, political, and administrative data. These estimates assist corporations and investors in managing risk and market volatility. Portfolio optimization is challenging in turbulent financial markets, but the study supports it. Portfolio managers can utilize predictions to reallocate assets to improve returns, as the KSE 100 index is predicted to stay stable until September 2023. With such market insight, portfolio managers may better allocate assets. Practical uses include market timing, a crucial trading component. Market entry and departure decisions may be made utilizing the model's findings. The expected stability of the KSE 100 index lets traders plan market-profitable deals. The study's conclusions affect strategic strategy and investment. The model may guide businesses and politicians in aligning their plans with market trends. A thorough review of economic, social, political, and administrative data helps decision-makers comprehend market dynamics.

The study makes a significant theoretical contribution to the emerging discipline of predictive modeling. The success of LSTM and backpropagation artificial neural network models emphasizes feature selection. By incorporating twenty-six indications from other domains, the model's accuracy is improved, and additional research into the potential of different indicator types in predictive model creation is encouraged. The study also sheds light on prediction model generalizability. The KSE 100 index movement is predicted with great accuracy, demonstrating generalizability. This raises questions about using machine learning models in other financial markets and scenarios besides the Karachi Stock Exchange. This study examines financial market temporal dynamics. Forecasting KSE 100 index stability within a certain period helps understand market circumstances. Financial market temporal dynamics may be studied by using this temporal viewpoint in discussions of stock market indices' stability and volatility. The study emphasizes the need for rigorous validation approaches for financial forecasting machine learning models. A complete comparison of actual and anticipated values over time evaluates the model's efficacy against empirical data. Prioritizing robust validation approaches enhances model dependability and real-world applicability.

Limitations and future direction

The study acknowledges certain limitations that require examination. First, using historical data assumes that indications and stock market changes will endure over time. Unexpected structural changes may challenge this assumption, reducing the model's accuracy. Data quality and accessibility might also restrict the model's capacity to extrapolate to new market conditions. The study's long prediction horizon, until September 2023, raises concerns about the model's long-term ability to react to changing market conditions. To keep models relevant, future research may focus on dynamic updating. Ensemble modeling may improve prediction accuracy and reduce model risk by mixing many machine learning methods. Indicators may be expanded and enhanced based on new insights or data sources to improve prediction capabilities. External inputs like financial news sentiment analysis or social media can capture instantaneous market sentiment and improve reaction to unexpected movements. Domain specialist collaboration may provide valuable insights and help fix model flaws. Strict cross-validation techniques must include out-of-sample testing to verify the model's resilience. These factors enable the improvement of stock market trend forecasting machine learning systems.

DATA AVAILABILITY STATEMENT

Monthly data from February 2004 to December 2020 is taken from the World Development Indicator, State Bank of Pakistan, and Pakistan Bureau of Statistics.

REFERENCES

Adams S, Klobodu EKM; 2016. Financial development, control of corruption and income inequality. International Review of Applied Economics, 30(6):790-808.

Agatonovic-Kustrin S, Beresford R; 2000. Basic concepts of Artificial Neural Network (ANN) modeling and its application in pharmaceutical research. Journal of pharmaceutical and biomedical analysis, 22(5):717-727.

Ahmed A, Bleak times ahead for Pakistan's manufacturing sector; 2023. https: //shorturl.at/tJVX0.

Al-Ameer M, Hammad W, Ismail A, Hamdan A; 2018. The relationship of gold price with the stock market: The case of Frankfurt Stock Exchange. International Journal of Energy Economics and Policy, 8(5):357.

Ali M, Khan DM, Aamir M, Ali A, Ahmad Z; 2021. Predicting the direction movement of financial time series using artificial neural network and support vector machine. Complexity, 2021:1-13.

Ali M, Khan DM, Alshanbari HM, El-Bagoury AAAH; 2023. Prediction of complex stock market data using an improved hybrid emd-lstm model. Applied Sciences, 13(3):1429.

Arestis P, Demetriades P; 1997. Financial development and economic growth: Assessing the evidence. The Economic Journal, 107(442):783-799.

Arjoon R, Botes M, Chesang LK, Gupta R; 2012. The long-run relationship between inflation and accurate stock prices: Empirical evidence from South Africa. Journal of Business Economics and Management, 13(4):600-613.

Asongu SA; 2012. Government quality determinants of stock market performance in African countries. Journal of African Business, 13(3):183-199.

Azeem A, Aziz B, Jadoon AK; 2017. Better tool for economic growth? Banks or secondary markets; Empirical evidence from selected ASIAN countries. Paradigms, 11(1):41.

Bhat S, Peshimam GN, Inflation in Pakistan could average 33Moody's economist; 2023. https://shorturl.at/ahJP7.

Bosworth BP, Collins SM, Reinhart CM; 1999. Capital flows to developing economies: Implications for saving and investment. Brookings papers on economic activity, 1999(1):143-180.

Chhajer P, Shah M, Kshirsagar A; 2022. The applications of artificial neural networks, support vector machines, and long-short term memory for stock market prediction. Decision Analytics Journal, 2:100015.

Demir C; 2019. Macroeconomic determinants of stock market fluctuations: The case of BIST-100. Economies, 7(1):8.

Dima B, Barna F, Nachescu ML; 2018. Does rule of law support the capital market?. Economic Research-Ekonomska Istraživanja, 31(1):461-479.

Ghani M, Guo Q, Ma F, Li T; 2022. Forecasting Pakistan stock market volatility: Evidence from economic variables and the uncertainty index. International Review of Economics & Finance, 80:1180-1189.

Gori L, Luigi Sacco P, Teti E, Triveri F; 2023. Commentary—Much ado About Something Else. Donald Trump, the US Stock Market, and the Public Interest Ethics of Social Media Communication. International Journal of Business Communication, p. 23294884231156903.

Jawad Y, Iqbal MJ.; 2020.Pakistan Stock Exchange Analysis and Forecasting using Hybrid Machine Learning Technique. In: 2020 IEEE 23rd International Multitopic Conference (INMIC) IEEE p. 1-6.

Khan W, Ghazanfar MA, Azam MA, Karami A, Alyoubi KH, Alfakeeh AS; 2020. Stock market prediction using machine learning classifiers and social media, news. Journal of Ambient Intelligence and Humanized Computing, 13:1-24.

Kim SE, Seo IW; 2015. Artificial Neural Network ensemble modeling with conjunctive data clustering for water quality prediction in rivers. Journal of Hydro-Environment Research, 9(3):325-339.

Kolapo FT, Adaramola AO; 2012. The impact of the Nigerian capital market on economic growth (1990-2010). International Journal of Developing Societies, 1(1):11-19.

Kumbure MM, Lohrmann C, Luukka P, Porras J; 2022. Machine learning techniques and data for stock market forecasting: A literature review. Expert Systems with Applications, 197:116659.

Law R; 2000. Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting. Tourism Management, 21(4):331-340.

Lehkonen H, Heimonen K; 2015. Democracy, political risks and stock market performance. Journal of International Money and Finance, 59:77-99.

Pan L, Mishra V; 2018. Stock market development and economic growth: Empirical evidence from China. Economic Modelling, 68:661-673.

Parmar I, Agarwal N, Saxena S, Arora R, Gupta S, Dhiman H, et al..; 2018.Stock market prediction using machine learning. In: 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC) IEEE p. 574-576.

ProPK, Business confidence takes shocking dip in latest OICCI survey; 2022. https://shorturl.at/hpKS5.

Rao AR, Reimherr M; 2023. Nonlinear Functional Modeling Using Neural Networks. Journal of Computational and Graphical Statistics (Just accepted), p. 1-10.

Rather AM, Agarwal A, Sastry V; 2015. Recurrent neural network and a hybrid model for prediction of stock returns. Expert Systems with Applications, 42(6):3234-3241.

Rouf N, Malik MB, Arif T, Sharma S, Singh S, Aich S, et al.; 2021. Stock market prediction using machine learning techniques: A decade survey on methodologies, recent developments, and future directions. Electronics, 10(21):2717.

Sahu SK, Mokhade A, Bokde ND; 2023. An Overview of Machine Learning, Deep Learning, and Reinforcement Learning-Based Techniques in Quantitative Finance: Recent Progress and Challenges. Applied Sciences, 13(3):1956.

Siregar SP, Wanto A; 2017. Analysis of artificial neural network accuracy using backpropagation algorithm in predicting process (forecasting). IJISTECH (International Journal of Information System and Technology), 1(1):34-42.

Stiglitz JE; 2000. Capital market liberalization, economic growth, and instability. World Development, 28(6):1075-1086.

Turak N, Blackouts, currency dives and corruption: Pakistan's economy is on the brink of collapse; 2023. https://shorturl.at/cwOP9.

Usmani M, Ebrahim M, Adil SH, Raza K.; 2018.Predicting market performance with hybrid model. In: 2018 3rd International Conference on Emerging Trends in Engineering, Sciences and Technology (ICEEST) IEEE p. 1-4.

Vaisla KS, Bhatt AK; 2010. An analysis of the performance of artificial neural network technique for stock market forecasting. International Journal on Computer Science and Engineering, 2(6):2104-2109.

Zaffar A, Hussain SA; 2022. Modeling and prediction of KSE--100 index closing based on news sentiments: An applications of machine learning model and ARMA (p, q) model. Multimedia Tools and Applications, 81(23):33311-33333.