



Stock Market Prediction Using Deep Learning by Enhancing LSTM

¹Sahunthala S, ²Sangeetha S, ³Suji G

¹Assistant Associate Professor, ²Student, ³Student

¹Information Technology,

¹Anand Institute Of Higher Technology, Chennai, India

Abstract: The volatile nature of the stock market makes it very difficult to predict future market trends and where to invest. Hence, there is a need for cross-utilization supported by ultramodern architecture. With the recent advancement of LSTM & KNN, continuous practical problems can be modeled and solved with human-level accuracy. Apart from this, in dealing with interim trading strategy, the proposed architecture is designed as a continuous training pipeline so that the stored model is up-to-date with the latest market trends by providing high accuracy in prediction. The framework outperforms basic LSTM & KNN model algorithms and maximizes portfolio returns. Experimental results show how natural language processing and statistical inference can help us pick trending stocks based on news headlines and historical data. To evaluate the performance of the proposed method, a comparison of our portfolio results was made with various LSTM & KNN model algorithms keeping the same configuration.

Index Terms – Machine Learning, Deep Learning, LSTM, KNN, Stock Market.

I. INTRODUCTION

The financial market is a composite structure that is dynamic and People can trade in virtual markets for currencies, stocks, shares, and derivatives platforms that brokers support. The stock market enables investors to purchase shares of publicly traded businesses by trading on exchanges or off-exchange markets. With relatively minimal risk compared to starting a new business or needing a high-paying job, this market has provided investors with the opportunity to make money and live affluent lives. Numerous factors have an impact on stock markets, contributing to the unpredictability and excessive volatility in the market. This is due to a variety of factors, such as market volatility. Due to the high number of variables, the stock market is known to be a complex adaptive system that is challenging to anticipate. numerous elements that affect how prices fluctuate day to day. Regression, how we accomplish this in machine learning. Here, the features are the independent factors, and the price is the dependent variable that needs to be predicted. It is clear that the features we use are not completely independent; we are aware that the closing price, the number of shares outstanding, and the return on investment are not independent. However, we are assuming this in order to make things simpler. Forecast for trading in TStock Securities item is used to estimate inventory pricing expectations for the future. The assumption that inventory would be put on the market helps both buyers and sellers of inventory. The accurate prediction future price should result in notable gains for both buyers and sellers. This might support the idea that everyone should be able to freely recognise statistics about an organisation. Which undeniably contains its price history, may currently be taken into account inside the gift value of the stock. We need to obtain the company's historical inventory cost information in order to predict the inventory value of a certain organisation. We will predict the daily open value of the inventory based on the verifiable information. Using a system mastering algorithm, a Python programme can predict the cost of the inventory. Support Calculations using Vector Regression and Linear Regression are used to predict the cost of the inventory. Depending on the company's actual inventory fee, predictions are made. A projection for the day will be made by the application after it has looked through the inventory records. In essence, the inventory trade is a conglomeration of different buyers and sellers of goods. A financial trade anticipation is the process of estimating the securities market's future performance. It is expected that the anticipation would be strong, real, and productive. The framework must function realistically and should be appropriate for real-world circumstances. Additionally, it is planned that the framework will take into account every factor that could affect the stock's performance and value. The method towards closing the stock price projection is shown in the aphase, and the unique tests conducted to ascertain the presentation of the models are demonstrated. To prepare the dataset and predict the future inventory cost, calculations involving Direct linear regression and relapse using support vectors are performed. Assignments structured along these lines are widely used to define supervised learning. It is intensely focused on one particular project, handling a wide range of calculating aids until it perfectly completes the task. This is typically the instructional type that you can almost certainly master. Stock market forecasting is generally defined as attempting to ascertain the actual value of the inventory while also giving consumers a vivid concept to understand and assume the market . The dataset's quarterly financial % is used, with a significant addition. In this approach, depending solely on a proprietary dataset may not be sufficient for the forecast and may result in final findings that are inaccurate. New investors will find data mining stock analysis handy for making stock market investments based on the different criteria that the software takes into account. Daily stock market operations include share exchanges and Sensex calculation. For trading in equities, debt instruments, and derivatives, the

exchange offers a clear and effective market. Sensex will be analyzed by our algorithms depending on firm stock value. Numerous things affect a company's stock values. The associated works are briefly discussed in section 3.

II. LITERATURE SURVEY

Jigar Patel, Shahil Shah, Priyank Thakkar, K. Kotecha et al. (2015). In this paper existing techniques are From the Indian equity markets, two indices, CNX Nifty and S&P Bombay equity Exchange (BSE) Sensex, are used. For experimental evaluation. The proposed system has two fusion stages (vector regression) and the second stage (ANN and RF). The task of predicting future values of stock market indices is focused in this paper. Experiments are carried out on ten years of historical data for two Indian stock market indices, the CNX Nifty and S&P BSE Sensex markets [1].

M. Miró-Jul, G. Fiol-Roig and A. P. Isern-Deya et al. (2010). In this paper the existing techniques are artificial intelligence to predict the stock prices. The proposed technique is data mining to be used for decision tree in stock market analysis. The proposed system gives the higher performance in the accuracy [2].

S. Kar, S. Saha, L. Khaidem, S. Basak, and S. R. Dey et al. (2019). In this paper existing system has forecasting and diffusion modelling to predict the analysis. In the proposed methodology they used the Xgboost, Random forests, Machine classification extremely accurate for medium to price direction [3].

A. F. Sheta, S. E. Ahmed and H. A. Faris et al. (2015). In this paper they use proposed system Artificial neural networks, regression Support Vector Machines to predict prices. The results were validated using number of criterion for evaluation. Future research shall focus on exploring other soft computing techniques to solve the stock market prediction problems [4].

C.-Y. Yeh, C.-W. Huang and S.-J. Lee et al. (2011). In this paper in order to forecast the stock market, vector regression has been used. However, typically, it is necessary to manually adjust the hyperparameters of the kernel functions. In proposed system Multiple-kernel learning to predict the stock prices. [5].

E. Chong, C. Han and F. C. Park et al. (2017). In this paper systematic analysis of the use of deep learning networks for stock market analysis and prediction. In the proposed system three unsupervised feature extraction methods—principal component Autoencoder, analysis, and the constrained Boltzmann machine [6].

M. Sedighi, H. Jahangirnia, M. Gharakhani and S. F. Fard et al. (2019). In this existing method stringent statistical hypotheses are essential, human interventions take part in predicting process and appropriate range. In the proposed technology they use the stock price forecasting, multi-objective optimization technical indicators ABC, ANFIS, SVM [7].

X. Pang, Y. Zhou, P. Wang, W. Lin and V. Chang et al. (2018). In this paper they used the existing method as traditional neural network algorithm. In proposed system they enhance their paper by using deep neural network with long-term memory network with long-term and short-term memory with automatic encoder (LSTM) with embedded layer and the long short-term memory neural network [8].

M. Qiu and Y. Song et al. (2016). In the proposed methodology they use the artificial neural network (ANN) model. We use evolutionary algorithms to improve the ANN model's performance in order to more accurately anticipate the future trend of the stock market index. (GA) [9].

Y. Alsubaie, K. E. Hindi and H. Alsalman et al. (2019). In this research paper existing system are technical indicators (TIs). A proposed cost-sensitive fine-tuned naïve Bayes classifier managed to achieve better overall investment performance than other classifiers [10].

B. Sezer and A. M. Ozbayoglu et al. (2018). In this paper existing system the use artificial network. In the proposed system they enhance the model by the model for automated trading. A 2-D convolutional neural network based on image processing is used by CNN-TA. properties [11].

A. Gupta and D. S. D. Sharma et al. (2014). In this research paper take a report on Stock market prediction. There are various techniques available for the stock market's forecasting value [12].

Y. Roh, G. Heo and S. E. Whang et al. (2021). We do an extensive analysis of data collecting from a data management perspective. Data collection primarily entails the gathering, labelling, and upgrading of new data or models. As part of a larger trend of combining Artificial Intelligence (AI) and Big Data, the combination of machine learning and data management for data collecting presents several chances for new research [13].

S. Lin and B. W. Kernighan et al. (1973). In this paper heuristic procedure for generating optimum and near-optimum solutions for the symmetric traveling-salesman problem. The procedure produces optimum solutions for all problems tested, "classical" problems appearing 110 cities are mentioned in the literature, along with randomly created exam problems. Runtimes increase by around n^2 [14].

S. O. Olatunji and M. S. Al-Hamdani et al. (2013). In this paper, we present an artificial neural network for Saudi stock market forecasting. The proposed predictions model's high level of accuracy makes it a potential investment advisor for Saudi stock market traders and investors. market [15].

M. Usmani, S. H. Adil, K. Raza and S. S. A. Ali et al. (2016). The Simple Moving Average (SMA) and Autoregressive Integrated Moving Average (ARIMA), two outdated statistical methods, are also utilised as input. Single Layer Perceptron (SLP) is one of the machine learning approaches used in this study to forecast the market performance of the Karachi Stock Exchange (KSE) on the day of closing. The prediction model predicts the market as Positive or Negative based on a variety of inputs. The attributes used in the model includes Oil rates, Gold & Silver rates, Interest rate, Foreign Exchange (FEX) rate, NEWS and social media feed [16].

X. Yang et al. (2011). Metaheuristic optimisation utilises metaheuristic techniques to address optimisation issues. Almost everything involves optimisation, including technical design, economics, vacation planning, and Internet routing. The most effective use of the available resources is critical because time, money, and resources are all finite [17].

Guresen E, Kayakutlu G, Daim TU et al. (2011). In the existing system neural network models. The suggested model primarily relies on historical data the Saudi Stock Exchange encompassing a considerable amount of time. Achieving a fair accuracy rate for prediction models will undoubtedly encourage more people to invest in Saudi stock market [18].

Y. Zhang et.al.(2009). In order to construct an improved bacterial chemotaxis optimisation (IBCO), the back propagation (BP) artificial neural network was presented in the paper. an efficient forecasting model for prediction of various stock indices[19].

K.J. Oh et.al.(2002) .This paper presents a piecewise nonlinear model and chaotic analysis-based stock trading model. The model's central element is made up of the following four phases: Using chaotic analysis, the first stage estimates the time-lag size in the input variables. The third phase uses backpropagation neural networks to forecast the change-point group after the second phase has detected sequential change-points in the stock market data. (BPNs)[20].

T.-S. Quah et.al.(1999) .This paper proposes that the stocks is a reflection of their profitability and the underlying company's management calibre.. Such data is reflected in technical and financial factors. As such, the ANN is used as a tool to uncover the intricate relationships between the performance of stocks and the related financial and technical variables[21].

Y. Wang et.al. (2003) This paper proposed a successful technique, a fuzzy rough set system to forecast the price of a stock at any given moment. Two agents work in our system; one assists stock dealers visually. monitor the current price of a stock and the other is a mining agent that helps stock dealers choose the best time to buy or sell stocks.We used our approach to anticipate the stronger rules of stock price, and after 180 days, we attained at least 93% accuracy, proving that it is effective.trials[22].

III. PROPOSED SYSTEM

The main design of the proposed system uses statistical analysis, sentiment analysis and the base LSTM & KNN algorithm which accurately predicts and invests with high returns and at low risks. First, models are created and news datasets are fed to pre-trained models via the news API for each stock. We also perform sentiment analysis to select the stock with highest sentiment, price prediction and statistical analysis to obtain which stock to buy, which stock to sell .

MODULES

1. COLLECTING DATA
2. PREPROCESSING
3. FEATURE EXTRACTION
4. CONTINUOUS TRAINING PIPELINE
5. SEGMENTATION
6. CLASSIFICATION

REQUIREMENTS

1. MATLAB
2. NUMPY
3. PANDAS

LSTM

extended short-term memory is an artificial(RNN) recurrent neural network architecture used in deep learning.Long short-term Memory has feedback connections in contrast to conventional feedforward neural networks., Long short-term Memory has connected to feedback. It will process single data points (e.g., images) and entire data sequences .LSTM models can store information over some time. They are remembering. capacity.This characteristic is beneficial when we deal with Time-Series or Sequential Data. When using an LSTM model, we can decide what information will be stored and discarded. We do that using the "gates."

LSTM has three gates:

1. The input gate: This gate updates the cell state with more data.
2. The knowledge is erased by the forget gate. that the model no longer requires.
3. The output gate: This gate at LSTM chooses the data shown as output.

However, some LSTM model variations, including Gated Recurrent Units (GRUs), lack an output gate. Long short-term memory networks are popularly used on time-series data for classification and making predictions. Its popularity in time-series applications is due to the possibility of several lags between significant events in a time series that are of undetermined duration.

KNN ALGORITHM

The k-nearest neighbours algorithm, often called KNN or k-NN, is a supervised learning classifier that makes predictions or classifications about the clustering of a single data point using proximity.t. Although it can be applied to classification or regression issues, it is commonly employed as an algorithm for classification because it relies on the idea that comparable points can be discovered close to one another.

EQUATION OF DISTANCE

Simply put, this is the distance in cartesian coordinates between two points in a plane or hyperplane. The length of the straight line connecting the two points under examination can also be used to represent the euclidean distance. This metric aids in the calculation of the net displacement made between an object's two states.

Left $(x,y) = \sqrt{\sum_i (x_i - y_i)^2}$, etc.

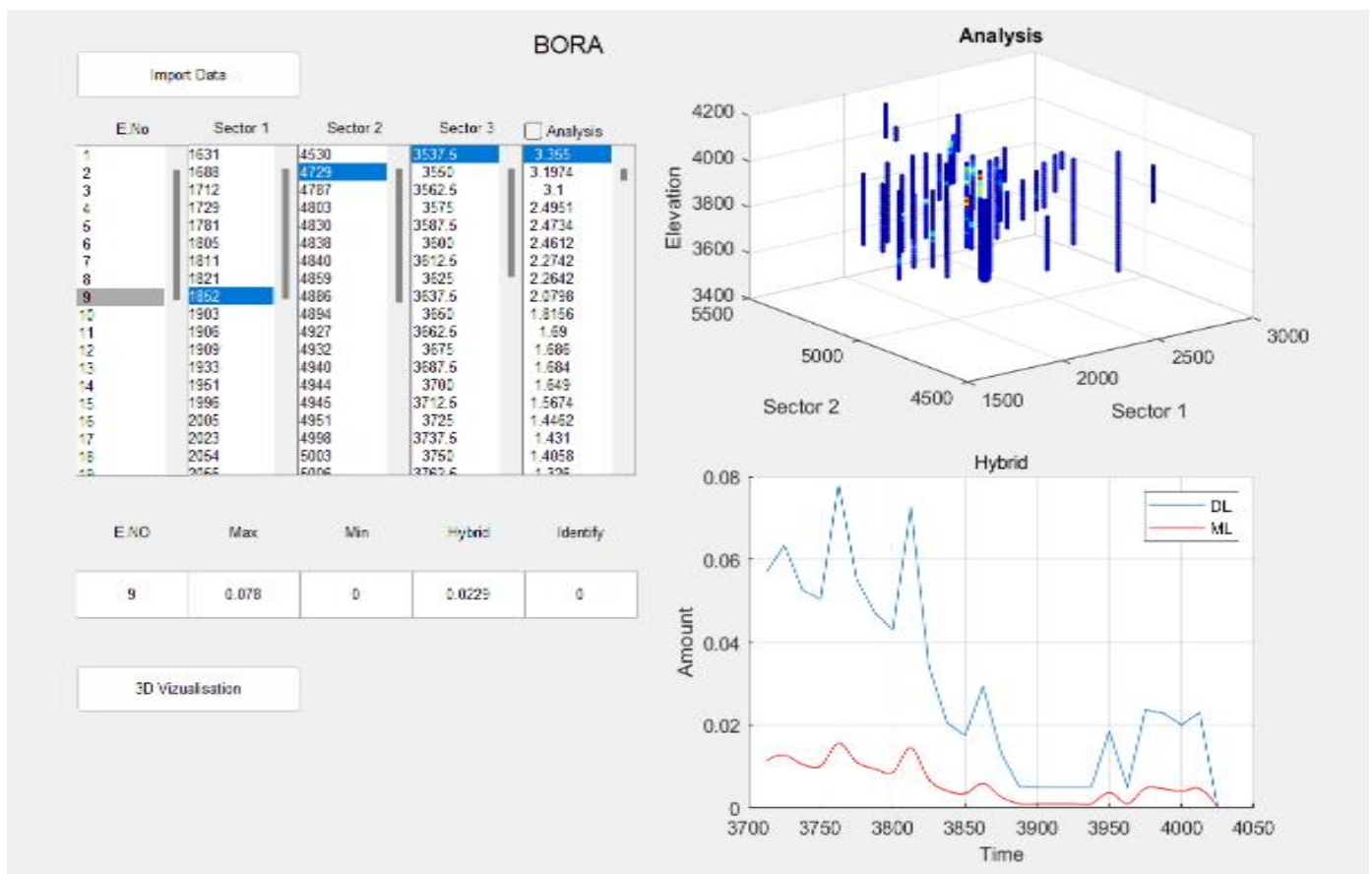
SYSTEM ARCHITECTURE



Fig 1 : System Architecture

IV. EXPERIMENT AND RESULT

- In this paper we train the dataset of nifty 50 which is the benchmark platform for the investment trade and test. Using the Matlab Gui interface.
- We collection past five years of dataset from the nifty50 and we employ deep learning and machine algorithm to train the test data and train the data.
- Finally we the predict future amount of nifty 50 values.



V. CONCLUSION

In this article, we explored LSTM and stock price using LSTM. We then visualized the future prediction. It has produced the finding that machine learning can be used to anticipate the stock market more accurately and effectively. In the future, the stock market prediction system can be further improved by utilizing a much bigger dataset.

REFERENCES

- [1]. L. A. Teixeira and A. L. I. De Oliveira, "A method for automatic stock trading combining technical analysis and nearest neighbor classification," *Expert Systems with Applications*, vol. 37, no. 10, pp. 6885–6890, 2010.
- [2]. M. Ballings, D. Van Den Poel, N. Hespeels, and R. Gryp, "Evaluating multiple classifiers for stock price direction prediction," *Expert Systems with Applications*, vol. 42, no. 20, pp. 7046–7056, 2015.
- [3]. W. Huang, Y. Nakamori, and S.-Y. Wang, "Forecasting stock market movement direction with support vector machine," *Computers & Operations Research*, vol. 32, no. 10, pp. 2513–2522, 2005.
- [4]. Y.-W. Cheung, M. D. Chinn, and A. G. Pascual, "Empirical exchange rate models of the nineties: Are any fit to survive?" *Journal of International Money and Finance*, vol. 24, no. 7, pp. 1150–1175, 2005.
- [5]. G. Armano, A. Murru, and F. Roli, "Stock market prediction by a mixture of genetic-neural experts," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 16, no. 5, pp. 501–56, 2002.
- [6]. H. Moghaddam, M. H. Moghaddam, and M. Esfandyari, "Stock market index prediction using artificial neural network," *Journal of Economics, Finance and Administrative Science*, vol. 21, no. 41, pp. 89–93, 2016.
- [7]. R. Choudhry and K. Garg, *A Hybrid Machine Learning System for Stock Market Forecasting*, vol. 39, 2008.
- [8]. Y. K. Kwon, S. S. Choi, and B. R. Moon, "Stock prediction based on financial correlation," in *Proceedings of the Conference on Genetic Evolutionary Computation*, 2005.
- [9]. L. Lacasa, V. Nicosia, and V. Latora, "Network structure of multivariate time series," *Scientific Reports*, vol. 5, 2015.

- [10]. M. Wang, Y. Chen, L. Tian, S. Jiang, Z. Tian, and R. Du, "Fluctuation behavior analysis of international crude oil and gasoline price based on complex network perspective," *Applied Energy*, vol. 175, pp. 109–127, 2016.
- [11]. B. M. Tabak, T. R. Serra, and D. O. Cajueiro, "Topological properties of stock market networks: the case of Brazil," *Physica A: Statistical Mechanics and Its Applications*, vol. 389, no. 16, pp. 3240–3249, 2010.
- [12]. L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuno, "From time series to complex networks: the visibility graph," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 13, pp. 4972–4975, 2008.
- [13]. Z. Gao, Q. Cai, Y. Yang, W. Dang, and S. Zhang, "Multiscale limited penetrable horizontal visibility graph for analyzing nonlinear timeseries," *Scientific Reports*, vol. 6, no. 1, Article ID 35622, 2016.
- [14]. E. Zhuang, M. Small, and G. Feng, "Time series analysis of the developed financial markets' integration using visibility graphs," *Physica A: Statistical Mechanics and its Applications*, vol. 410, pp. 483–495, 2014.
- [15]. R. V. Donner, Y. Zou, J. F. Donges, N. Marwan, and J. Kurths, "Recurrence networks-a novel paradigm for nonlinear time series analysis," *New Journal of Physics*, vol. 12, no. 3, Article ID 033025, 2010.
- [16]. Y. Li, H. Caö, and Y. Tan, "Novel method of identifying time series based on network graphs," *Complexity*, vol. 17, no. 1, pp. 13–34, 2011.

