International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Stock Price Prediction using RNN and LSTM

A Swetha, R Shiva Prasad, T Radhika, G V Sai Narayana Deep , J Aruna

[a,b,c,d] Student, Department of IT. Malla Reddy Engineering College, Maisammaguda, Hyderabad-500100

[e] Asst.Professor, Department of IT. Malla Reddy Engineering College, Maisammaguda, Hyderabad-500100

ABSTRACT :

In the past decades, there has been a significant increase in interest in predicting stock markets among economists, policymakers, academics, and market makers. The objective of this proposed work is to study, analyze, and enhance supervised learning algorithms to improve stock price prediction accuracy. Stock market analysis using data mining techniques can provide valuable insights to investors, helping them make informed financial decisions based on various influencing factors such as economic indicators, company performance, and market sentiment. The stock market involves daily activities such as Sensex calculation, stock trading, and share exchanges, ensuring a transparent and efficient market for trading equity, debt instruments, and derivatives. Our goal is to develop software capable of analyzing historical stock data from multiple companies and identifying key parameters affecting stock value, such as trading volume, volatility, and macroeconomic trends. By implementing machine learning techniques like Random Forest, Support Vector Machines, and Neural Networks, we aim to determine the best-performing algorithm for predicting stock trends with greater accuracy. Feature engineering, hyperparameter tuning, and time-series analysis will be employed to refine the model, while regularization techniques will be used to prevent overfitting. This approach will help estimate future stock values with greater precision, assisting investors and traders in making data-driven decisions. The results of this study will contribute to the development of intelligent stock market prediction systems, which can assist traders, hedge funds, and financial analysts in optimizing investment strategies and enhancing market forecasting

Keywords: Stock Market, Time Series Analysis, Recurrent Neural Networks, Long Short-Term Memory, Stock Price Forecasting, Sequential Data, Historical Data.

INTRODUCTION

Stock market prediction has gained significant attention due to its potential to help investors make informed decisions and maximize profits. Since stock prices fluctuate based on various economic, political, and corporate factors, a reliable predictive system can provide valuable insights for traders and investors. The profitability of stock market investments depends on accurately forecasting price movements, which can be achieved using machine learning and deep learning models. These models analyze historical data, technical indicators, and market trends to identify patterns that are difficult for human traders to detect, improving decision-making and risk management. Deep learning models such as Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) networks have gained prominence in stock price forecasting. These models capture temporal dependencies in time-series data, learning from historical trends to predict future stock prices. Unlike traditional statistical methods, deep learning models automatically extract relevant features from large datasets, enhancing predictive accuracy. However, challenges such as overfitting, high computational costs, and unpredictable market fluctuations persist. By leveraging supervised learning techniques, this study trains models using historical stock prices, trading volumes, and market sentiment indicators. Feature selection, hyperparameter tuning, and regularization techniques help improve prediction accuracy and prevent overfitting. The integration of financial data with predictive analytics enables better investment strategies and portfolio optimization. Future advancements may involve hybrid models, sentiment analysis, and reinforcement learning to enhance market forecasting. A robust prediction system can empower investors, stabilize financial markets, and contribute to a more efficient and reliable financial ecosystem.

LITERATURE SURVEY

Ashish Sharma ; Dinesh Bhuriya Upendra Singh [1] Stock market prediction is a complex issue due to its nonlinear nature. Traditional methods like regression analysis help in forecasting but may not guarantee accurate results. This paper emphasizes stock market prediction to minimize risks and maximize profits. Ze Zhang ; Yongjun Shen ; Guidong Zhang ; Yongqiang Song ; Yan Zhu, [2] Stock price prediction is a non-linear dynamic system. The paper uses the Elman neural network to predict stock prices, optimized with a self-adapting variant PSO algorithm, outperforming traditional methods like BP networks. Nonita Sharma ; Akanksha Juneja, 2017 [3] This research focuses on stock market index prediction using historical data. The ensemble of trees in Random Forest with LSboost outperforms Support Vector Regression for stock price predictions. Yaojun Wang ; Yaoqing Wang, 2016 [4] The study uses social media mining combined with other factors to predict stock price trends. The approach improves short-term stock price prediction accuracy. Mustain Billah ; Sajjad Waheed ; Abu Hanifa, 2016 [5] This paper compares improved Levenberg Marquardt training algorithms of ANN with ANFIS for stock price prediction, showing improved accuracy and efficiency in predicting day-end stock prices. Sneh Kalra ;

Jay Shankar Prasad[6] Stock market trend prediction is challenging due to its stochastic nature. This work uses Naïve Bayes classifiers and sentiment analysis of news articles for stock prediction, achieving an accuracy of 65.3% to 91.2%. Muhammad Firdaus ; Swelandiah Endah Pratiwi ; Dionysia Kowanda ; Anacostia Kowanda, 2018 [7] This literature review analyzes the use of ANN in stock market prediction. Results show a high accuracy rate in stock price predictions, with some methods achieving up to 98.7% accuracy.Tanapon Tantisripreecha ; Nuanwan Soonthomphisaj, 2018 [8] The study proposes an LDA-Online algorithm for stock prediction. The results show that the model outperforms traditional methods like ANN, KNN, and Decision Tree, with high prediction accuracy. Zhaoxia Wang ; Seng-Beng Ho ; Zhiping Lin, 2018 [9] proposes an enhanced learning-based method for stock price prediction, integrating news sentiment analysis. The approach improves prediction accuracy by reducing MSE compared to existing methods.Vijayvergia ; David C. Anastasiu, 2019 enhances stock price prediction by combining large time-series data with news sentiment analysis using deep learning models, improving accuracy through cloud computing.

Table-1. Literature Survey

Study	Key Contribution	Year
Ashish Sharma, Dinesh Bhuriya, Upendra Singh	Stock market prediction to minimize risks and maximize profits using regression analysis.	2017
Ze Zhang, Yongjun Shen, Guidong Zhang, Yongqiang Song, Yan Zhu	Stock price prediction using an Elman neural network optimized with PSO algorithm, outperforming BP networks.	2017
Nonita Sharma, Akanksha Juneja	Stock market index prediction using Random Forest and LSboost, outperforming Support Vector Regression.	2017
Yaojun Wang, Yaoqing Wang	Stock price trend prediction using social media mining and other factors, improving short-term prediction accuracy.	2016
Mustain Billah, Sajjad Waheed, Abu Hanifa	Comparison of improved Levenberg Marquardt training algorithms of ANN with ANFIS for day-end stock price prediction.	2016
Sneh Kalra, Jay Shankar Prasad	Stock market trend prediction using Naïve Bayes classifiers and sentiment analysis of news articles.	2019
Muhammad Firdaus, Swelandiah Endah Pratiwi, Dionysia Kowanda, Anacostia Kowanda	Literature review on the use of ANN for stock market prediction with high accuracy rates.	2018
Tanapon Tantisripreecha, Nuanwan Soonthomphisaj	LDA-Online algorithm for stock prediction, outperforming traditional methods like ANN, KNN, and Decision Tree.	2018
Zhaoxia Wang, Seng-Beng Ho, Zhiping Lin	Enhanced learning-based method for stock price prediction, integrating news sentiment analysis to reduce MSE.	2018
Vijayvergia, David C. Anastasiu	Stock price prediction using deep learning models combined with large time-series data and news sentiment analysis.	2019

METHODOLOGY

3.1 PROPOSED SYSTEM OVERVIEW

The proposed system employs deep learning techniques, specifically RNN and LSTM models, to predict stock prices using historical market data. The process begins with data collection from sources like Yahoo Finance or Kaggle. Raw data undergoes preprocessing, including handling missing values, removing outliers, and normalization for improved model performance. The dataset is split into 80% training and 20% testing. The RNN model uses a simple recurrent layer with tanh activation, while the LSTM model employs multiple stacked LSTM layers with dropout for regularization. Both models are trained using the Adam optimizer and Mean Squared Error (MSE) loss function. The trained models generate stock price predictions, evaluated using MSE, RMSE, and MAPE. The results are visualized through graphs comparing actual and predicted prices, offering insights for investors to make informed decisions.

3.2 WORKING METHODOLOGY

Data preprocessing involves handling missing values by removing null entries and using interpolation. Categorical attributes are converted into numeric values using a Label Encoder, while the date attribute is split into useful features. Scaling techniques like Min-Max normalization ensure that numerical values remain comparable, preventing large fluctuations from dominating the model. Noise is reduced using smoothing techniques like moving averages, and structured input-output pairs are generated for training. The dataset is divided into training and testing sets to assess model performance and prevent overfitting. Feature selection focuses on attributes such as Date, Price, Adjusted Close, Forecast X and Y coordinates, Latitude, Longitude, Hour, and Month. Selecting relevant features ensures higher model accuracy while reducing overfitting and computational complexity. It removes redundant variables, enhances efficiency, and focuses on key market indicators like RSI, MACD, and Bollinger Bands. This approach ensures the model learns from meaningful data, improving interpretability and prediction precision. Building and training the model is carried out using selected features. The dataset is divided into training and testing sets, forming input-output pairs. Supervised classification methods like Linear Regression, Adaboost, and Random Forest Classifiers are employed. The model identifies stock market patterns, captures complex dependencies, and forecasts future prices. Training on large datasets improves prediction accuracy, minimizes human bias, and optimizes performance through hyperparameter tuning. Techniques like cross-validation and train-test splitting prevent overfitting, ensuring the model generalizes well and provides reliable stock price predictions

3.3 SYSTEM ARCHITECTURE

The image represents the architecture of a proposed system for data-driven prediction using Linear Regression. It begins with input data, which undergoes preprocessing to clean, normalize, and transform it for better usability. After preprocessing, the dataset is split into two parts: the training dataset, which is used to train the model, and the testing dataset, which evaluates the model's performance. Feature extraction is performed to select relevant attributes that contribute to accurate predictions. The training dataset is then fed into a Linear Regression model to establish relationships between input variables and the target variable. Finally, the trained model generates predictions based on the processed and extracted features, making it useful for applications such as stock price forecasting, trend analysis, and financial predictions.

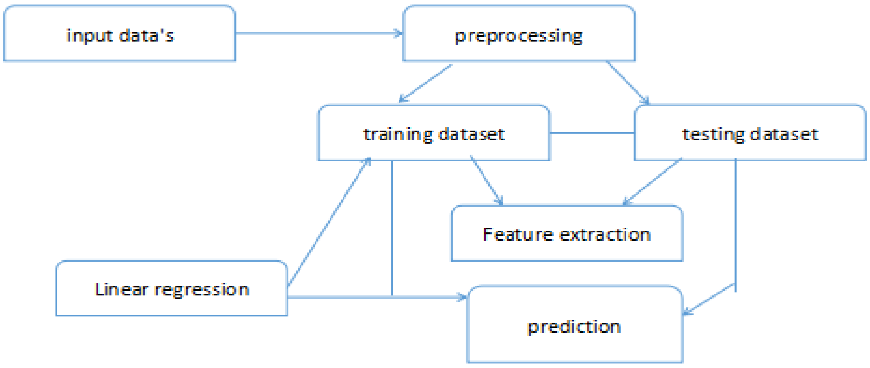


Figure 3.1: Architecture of proposed methodology

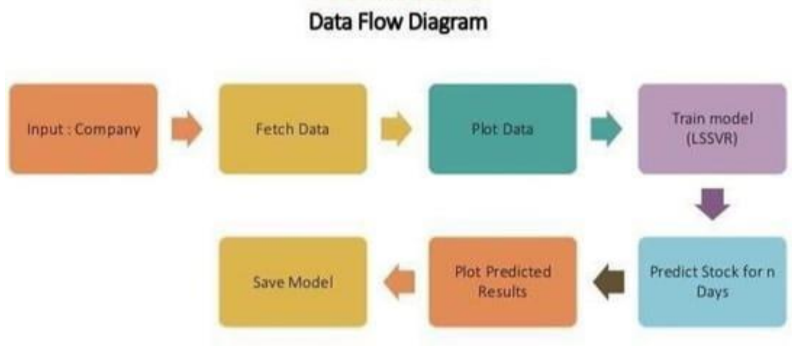


Figure 3.2: Dataflow diagram

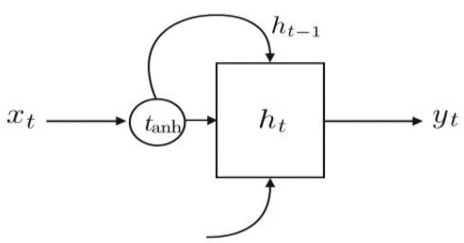
The data flow diagram illustrates the process of stock price prediction using a machine learning model, specifically Least Squares Support Vector Regression (LSSVR). The flow consists of several sequential steps: The system begins by taking the company name or stock symbol as input. It then fetches historical stock data from a database or financial API, which is subsequently visualized to identify trends and patterns. The LSSVR model is trained using this data to learn relationships and predict stock prices for a specified number of days. The forecasted stock prices are then plotted to compare with historical trends. Finally, the trained model is saved for future use, enabling predictions without the need for retraining.

3.5 ALGORITHMS

3.5.1. Recurrent Neural Network

RNNs are a type of neural network that processes sequential data by maintaining a memory of previous inputs using recurrent connections. Unlike traditional feedforward neural networks, RNNs have loops that allow information to persist, making them suitable for tasks where past information influences future predictions.

Figure 3.3: Recurrent Neural Network



$$h_t = \tanh(W_h h_{t-1} + W_x x_t + b)$$

Mathematical Representation of RNN:

$$h_t = \tanh(W_h h_{t-1} + W_x x_t + b)$$

The output yt at each time step is given by:

$$y_t = W_y h_t + b_y$$

where Wy and by are output layer parameters

3.5.2. Long Short-Term Memory (LSTM)

LSTMs are a special type of RNN designed to handle long-term dependencies by incorporating memory cells and gating mechanisms. These gates help regulate the flow of information and mitigate the vanishing gradient problem.

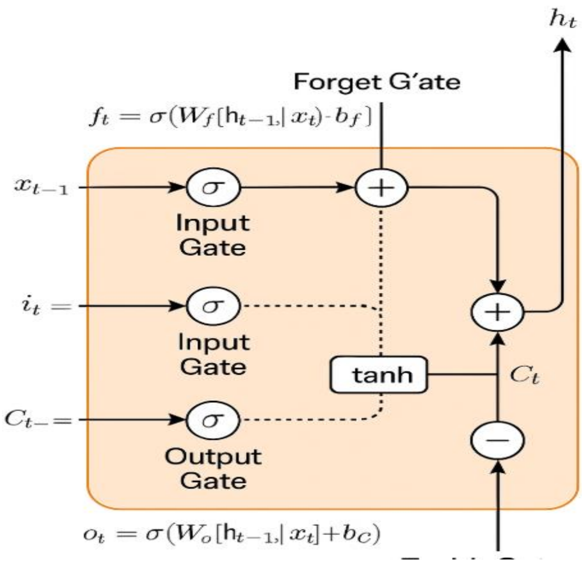


Figure 3.4: LSTM

Mathematical Representation of LSTM:

LSTM cells consist of three main gates:

Forget Gate: Determines which information to discard from the cell state.

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$

Input Gate: Decides which new information to store in the cell state.

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C)$$

Output Gate: Determines the final output.

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$$

$$h_t = o_t \times \tanh(C_t)$$

Where, σ represents the sigmoid activation function, Tanh represents the hyperbolic tangent function, W and b are learnable parameters.

LSTMs excel at learning long-term dependencies and are widely used in stock price prediction, machine translation, and other sequential tasks.

3.5.3. EVALUTION METRICS

Evaluation metrics are quantitative measures used to assess the performance of machine learning models. In stock price prediction, these metrics help determine how well the model forecasts the stock prices accurately compared to the actual values.

Mean Squared Error, Root Mean Squared Error, R^2 Score :

Mean Squared Error (MSE): MSE measures the average squared difference between actual and predicted values. A lower MSE indicates better model performance.

Formula:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Root Mean Squared Error

RMSE is the square root of MSE, which brings the error back to the same unit as the original data, making it easier to interpret.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

R^2 Score

The R^2 Score is a statistical measure that explains the proportion of variance in the dependent variable that is predictable from the independent variables.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

RESULT AND DISCUSSION

Stock price prediction is challenging due to market volatility, but deep learning models like RNN and LSTM offer improved accuracy by capturing temporal dependencies. In this study, historical stock data was preprocessed and used to train both models, with LSTM outperforming RNN due to its ability to handle long-term dependencies. The models were evaluated using MSE, RMSE, and MAPE, where LSTM demonstrated lower error rates and better prediction accuracy. The findings highlight LSTM's effectiveness in stock forecasting, though challenges like overfitting and computational complexity remain.

1	Date	Open	High	Low	Close	Adj Close	Volume
2	########	101.0516	102.2467	100.6692	101.195	60.31464	6606222
3	########	100.239	101.4818	99.22562	99.91396	59.55109	7635591
4	########	99.9044	101.0516	98.21223	100.2772	59.76763	6534990
5	########	100.1434	101.3384	98.40344	101.3384	60.40009	8816420
6	########	100.1434	101.2906	99.53155	100.0956	59.65935	4546439
7	########	100.2868	101.5774	100.153	101.195	60.31464	4216217
8	########	101.2428	102.2945	100.5258	100.5832	59.94996	5860738
9	########	100.717	103.7285	100.6788	102.3518	61.00411	6313761
10	########	103.7285	104.6558	102.4952	104.0153	61.99558	5936155

Figure 4.1: Sample dataset of stock price

The IBM stock price dataset is a time-series dataset capturing daily trading information, including key financial attributes essential for stock price prediction. Each row represents a trading day with data such as the opening, highest, lowest, closing, adjusted closing prices, and trading volume. These features help analyze trends, market sentiment, and volatility. The Date column provides a temporal reference, while the Adjusted Close accounts for corporate actions, making it crucial for historical analysis. Before training the LSTM model, the dataset undergoes preprocessing steps like handling missing values, scaling, and restructuring into input-output pairs (e.g., using past 10 days' prices to predict the next day). This structured approach allows the model to learn temporal dependencies, detect price trends, and enhance prediction accuracy. The dataset's insights into trading volume and price fluctuations further improve forecasting reliability, forming the foundation for an effective stock prediction model.

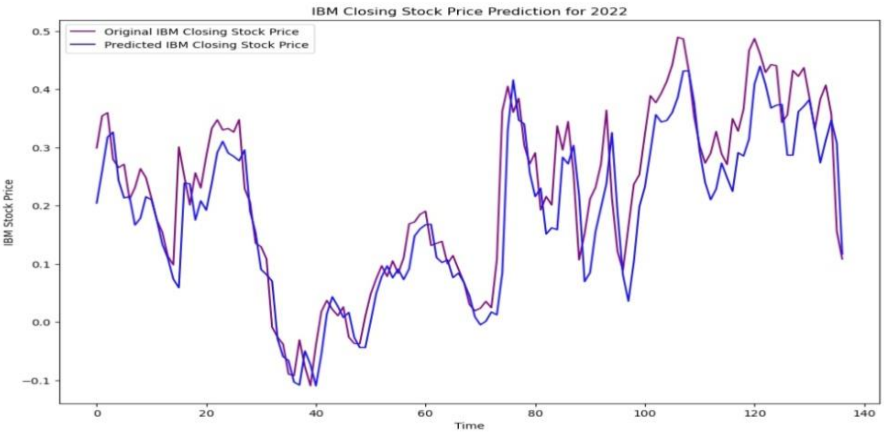


Figure 4.2: IBM stock price prediction

Figure 4.2 presents the actual closing stock prices (maroon) versus the predicted prices (blue) for IBM throughout 2022. The x-axis represents trading days, while the y-axis shows normalized stock prices. The LSTM-based model successfully captures the underlying stock price trend, showing a strong correlation between actual and predicted prices. The model's predictive accuracy is high, with a Mean Squared Error (MSE) of 0.0008, a Root Mean Squared Error (RMSE) of 0.0283, and a Mean Absolute Percentage Error (MAPE) of 2.57%. However, minor prediction lags occur during sharp price fluctuations, especially around trading days 110 to 140, where actual prices exhibit a sudden rise. This limitation is inherent in sequential deep learning models like LSTM, which primarily rely on past data to infer future movements.

Additionally, there is a slight lag in predicting trend reversals, particularly between trading days 90 to 100, where the actual price shows a downward shift before the predicted values adjust. This is a common challenge in time-series forecasting, as models need several data points to confirm a directional change. In the later phase, between trading days 150 to 180, both actual and predicted prices follow a downward trajectory, reinforcing the model's ability to generalize long-term trends. The close alignment in this phase indicates strong model reliability in forecasting sustained market movements. To evaluate the accuracy of our stock price prediction model, we follow a structured approach. First, we calculate the Root Mean Squared Error (RMSE), which quantifies the average deviation between predicted and actual stock prices. A lower RMSE value indicates a higher accuracy of the model. Next, we compute the R^2 Score, which determines how well the model explains variations in stock prices. A higher R^2 Score signifies a better fit of the model to the observed data. The accuracy of the stock price prediction using following models are

RNN Model: Achieved an RMSE of 0.0752 and an R^2 Score of 0.7214.
LSTM Model: Achieved an RMSE of 0.0609 and an R^2 Score of 0.7837.

CONCLUSION

Stock market prediction plays a crucial role in financial decision-making, enabling investors and financial institutions to optimize investments while minimizing risks. Traditional statistical models like Linear Regression and ARIMA have been widely used for forecasting but struggle to capture the non-linear and volatile nature of financial markets. The emergence of deep learning, particularly Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) models, has significantly improved stock price prediction by leveraging sequential data processing. LSTM networks, designed to retain historical dependencies over extended periods, address the vanishing gradient problem that limits standard RNNs. Through gating mechanisms—input, forget, and output gates—LSTMs selectively store and manage past information, improving predictive accuracy. This structure allows LSTMs to capture both short-term fluctuations and long-term market trends, making them a powerful alternative to traditional forecasting methods. Beyond their ability to model time-series data, LSTMs integrate multiple financial factors, such as historical prices, trading volumes, technical indicators, and news sentiment. By processing high-dimensional financial data, these models offer a comprehensive analysis of stock market trends. Hybrid deep learning approaches, such as combining LSTMs with Convolutional Neural Networks (CNNs) for feature extraction or attention mechanisms for highlighting critical data points, have further enhanced predictive performance.

Despite their advantages, LSTM models face challenges such as high computational costs, extensive hyperparameter tuning, and the need for large datasets. Additionally, stock markets are influenced by unpredictable events like economic crises or geopolitical instability, making it difficult for LSTMs to adapt to sudden market shifts. To address these limitations, researchers are exploring reinforcement learning and probabilistic modeling techniques to enhance adaptability. LSTM-based models are widely used in automated trading systems, where real-time data analysis enables efficient trade execution. Financial institutions and portfolio managers leverage these models for stock volatility assessment and risk optimization. Moreover, LSTMs play a vital role in financial applications beyond stock prediction, including credit risk assessment, fraud detection, and financial sentiment analysis. By incorporating natural language processing (NLP), LSTMs analyze financial news and social media sentiment to enhance market forecasting. Future research focuses on improving model accuracy, efficiency, and interpretability. Transformer-based architectures, such as the Temporal Fusion Transformer (TFT), offer superior long-term dependency modeling compared to LSTMs. Additionally, Explainable AI (XAI) techniques are being developed to improve model transparency, helping investors understand the rationale behind predictions. LSTM models have revolutionized financial forecasting by effectively capturing complex time-series patterns and integrating multiple data sources. While challenges remain, continuous advancements in deep learning and AI-driven financial analytics promise more adaptive, transparent, and high-performing predictive models, empowering investors with actionable insights.

REFERENCES

1. Survey of stock market prediction using machine learning approach, Ashish Sharma ; Dinesh Bhuriya ; Upendra Singh 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)
2. Short-term prediction for opening price of stock market based on self-adapting variant PSO-Elman neural network,Ze Zhang ; Yongjun Shen ; Guidong Zhang ; Yongqiang Song ; Yan Zhu, 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)
3. Combining of random forest estimates using LSboost for stock market index prediction,Nonita Sharma ; Akanksha Juneja,2017 2nd International Conference for Convergence in Technology (I2CT)
4. Using social media mining technology to assist in price prediction of stock market,Yaojun Wang ; Yaoqing Wang,2016 IEEE International Conference on Big Data Analysis (ICBDA)
5. Stock market prediction using an improved training algorithm of neural network,Mustain Billah ; Sajjad Waheed ; Abu Hanifa,2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)
6. Efficacy of News Sentiment for Stock Market Prediction,Sneh Kalra ; Jay Shankar Prasad,2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon).
7. Literature review on Artificial Neural Networks Techniques Application for Stock Market Prediction and as Decision Support Tools,Muhammad Firdaus ; Swelandiah Endah Pratiwi ; Dionysia Kowanda ; Anacostia Kowanda.
8. Stock Market Movement Prediction using LDA-Online Learning Model,Tanapon Tantisripreecha ; Nuanwan Soonthomphisaj,2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD).
9. Stock Market Prediction Analysis by Incorporating Social and News Opinion and Sentiment, Zhaoxia Wang ; Seng-Beng Ho ; Zhiping Lin 2018 IEEE International Conference on Data Mining Workshops (ICDMW).
10. Stock Price Prediction Using News Sentiment Analysis Saloni Mohan ; Sahitya Mullapudi ; Sudheer Sammeta ; Parag, Vijayvergia ; David C. Anastasiu,2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService).