IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 6, June 2023

DOI: 10.17148/IJARCCE.2023.12645

STOCK MARKET PRICE PREDICTION

Charan pote[1], Suraj Hume[2], Tejas Deshmukh[3], Ritesh Rana[4], Yash Chahande[5], Harshal Kubde[6]

Professor, Computer technology, Priyadarshini College of Engineering, Nagpur, India[1]

Student, Computer technology, Priyadarshini College of Engineering, Nagpur, India[2-6]

Abstract: Accurate prediction of stock market values is a critical task in financial analysis, empowering investors to make informed decisions. Machine learning has emerged as a powerful approach to enhance the authenticity and effectiveness of stock market forecasting. This research paper focuses on investigating the potential of regression models and LSTM-based machine learning techniques for predicting stock values. By comparing the performance of these models in stock market valuation, we aim to uncover their strengths and limitations. Our study leverages comprehensive historical stock market data from diverse sources, which undergoes meticulous preprocessing to extract pertinent features such as price trends, trading volume, and market sentiment. Regression models such as linear regression, polynomial regression, and support vector regression are implemented and rigorously evaluated to assess their predictive capabilities in estimating stock prices accurately. Additionally, we explore the potential of LSTM-based deep learning models in capturing intricate temporal dependencies and patterns in the data.

Keywords: stock market , forecasting, price prediction ,machine learning.

I. INTRODUCTION

In this era of rapid digitization, there are numerous avenues for investment that offer the potential for profitability across multiple sources. One particularly popular option is the stock market, which has the allure of potentially generating higher returns on investment compared to alternative avenues. In India, the National Stock Exchange (NSE) and the Bombay Stock Exchange (BSE) dominate the trading landscape as the primary platforms for stock market transactions. However, due to the dynamic nature of stock prices, accurately predicting the market proves to be a complex endeavor.

Time Series Analysis

Time series analysis (TSA) is a powerful analytical method that involves the study of past trends and patterns to predict future outcomes. It finds applications in a wide range of fields, including industry and business, economics and finance, and environmental science. TSA involves analyzing sequential data points collected at regular intervals over time to identify patterns, understand underlying relationships, and make forecasts.

In business, TSA can be employed to analyze sales data, demand patterns, and market trends, enabling companies to make informed decisions about production, inventory management, and marketing strategies. In economics and finance, TSA helps economists and financial analysts assess economic indicators, stock market movements, interest rates, and exchange rates, aiding in making projections and formulating investment strategies.

The process of selecting can be following way -
1) Long term analysis
2) Medium term analysis
3) Short term analysis

Fundamental analysis

Fundamental analysis is a crucial process that involves a thorough examination of the intrinsic value of a stock or company. It aims to assess the underlying factors that can influence the future performance and profitability of an investment. By analyzing fundamental aspects such as financial statements, industry trends, competitive positioning, and management competence, analysts can make informed predictions about the potential growth or decline of a company.

Fundamental analysis provides a holistic perspective on the overall health and viability of a business. It goes beyond short-term market fluctuations and focuses on the long-term prospects of an investment. By considering factors such as revenue growth, earnings potential, cash flow, and balance sheet strength, analysts can form a comprehensive view of a company's value.

II. EXISTING WORK

The integration of machine learning algorithms has revolutionized the field of stock market forecasting, enabling more accurate predictions of stock price trends. In this innovative system, three prominent algorithms, namely LSTM, linear regression, and ARIMA, are utilized to forecast the values of stocks.

The first algorithm, LSTM, is well-suited for capturing the intricate variations and patterns in stock price trends over a specified period. LSTM leverages its ability to retain and utilize long-term dependencies in the data to make predictions. By incorporating historical stock price data, LSTM can effectively identify underlying trends and patterns, enabling it to generate reliable forecasts.

To facilitate the analysis, a large dataset comprising approximately 900,000 records of relevant stock price data is selected from trusted sources such as Yahoo Finance. The dataset is then divided into two categories: training data and testing data. This segregation allows the algorithms to learn from the historical data and assess their predictive capabilities accurately.

The LSTM algorithm is complemented by linear regression, a well-established statistical technique widely used in financial analysis. Linear regression enables the identification of linear relationships between various factors and the stock price. By fitting a linear equation to the historical data, this algorithm can provide insights into the overall trend and direction of the stock price.

III.PROPOSED WORK

In the proposed system, the system aggregates and collects the data for various parameters such as open, high, close, and low prices from a substantial and comprehensive dataset. This dataset serves as the foundation for the subsequent analysis and forecasting of stock price patterns.

The specific focus of the analysis lies in the close price value, which is regarded as a key indicator of the stock's performance. This close price data is carefully selected from the extensive dataset and utilized as the input for the algorithmic analysis.

To facilitate effective modeling and evaluation, the collected data is then divided into two distinct sets: the training set and the testing set. The allocation of data follows a standard ratio of 80:20, with 80% of the data assigned to the training set and the remaining 20% reserved for the testing set

Algorithms:
LSTM model
Linear Regression

LSTM (Long Short-Term Memory) is a specialized form of recurrent neural network (RNN) that overcomes the inherent difficulty of capturing long-term dependencies and retaining relevant context in sequential data. When confronted with tasks that require an understanding of the entire sequence's context, such as predicting the next letter in a sequence or comprehending the meaning of a sentence, conventional neural networks may struggle.

The essence of LSTM lies in its ability to incorporate memory-like properties within the neural network. This unique characteristic allows the network to selectively remember and forget specific information based on its relevance and applicability. In the case of sequential data, LSTM enables the network to remember essential features that contribute to the overall context while discarding irrelevant details.

To achieve this, LSTM introduces a specialized internal state mechanism within the RNN node structure. This internal state acts as a memory bank, allowing information to flow across multiple time steps. As each input enters the LSTM node, it receives not only the current input but also the output from the previous step and crucial state information from the LSTM's memory.

This internal state component enables the LSTM to address the challenge of long-term dependencies in sequential data. By incorporating this memory-like mechanism, the network can preserve relevant information and context across multiple time steps, facilitating more accurate predictions and improved understanding of sequential patterns.
The LSTM's ability to selectively retain or discard information based on its importance enhances its capacity to capture long-range dependencies and overcome the vanishing gradient problem that often hinders conventional RNNs. This

unique architectural design empowers the LSTM to model and comprehend intricate sequential patterns, making it highly effective in tasks that require a comprehensive understanding of the underlying context.

ARIMA

ARIMA, short for AutoRegressive Integrated Moving Average, is a sophisticated time series forecasting model that encompasses three integral components: autoregression (AR), differencing (I), and moving average (MA). Each component serves a distinct purpose in capturing and understanding the intricate patterns and dynamics inherent in time series data.

The autoregression component (AR) in ARIMA examines the interdependency between a current observation and a specific number of lagged observations. By analyzing the historical values of the time series, the AR component discerns the relationship and influence between the present observation and its preceding data points. This understanding empowers the model to generate predictions based on this learned dependency structure.

The differencing component (I) addresses the issue of non-stationarity that often characterizes time series data. Non-stationarity refers to the presence of trends, seasonality, or other irregular patterns that can hinder accurate analysis. To overcome this, differencing is applied, which involves subtracting a previous observation from the current one. This transformation effectively removes underlying patterns and temporal dependencies, ensuring the data exhibits stationarity and facilitating more reliable predictions.

The moving average component (MA) of ARIMA leverages the relationship between a given observation and the residual errors derived from a moving average model applied to previous observations. Moving average entails calculating the average of a specific number of prior observations, producing a smoothed series that reveals underlying trends and patterns. The MA component captures any remaining irregularities or fluctuations not captured by the autoregressive and differencing components, enabling the model to account for residual variations and refine its predictions.

A crucial parameter in ARIMA is denoted as 'p,' which signifies the number of lag observations considered in the model, also known as the lag order. This parameter determines the extent to which past observations influence the current prediction, providing a mechanism to adapt the model to the specific characteristics of the time series data.

By integrating the AR, I, and MA components, ARIMA offers a comprehensive and sophisticated framework for time series forecasting. Its ability to capture temporal dependencies, address non-stationarity, and account for residual fluctuations makes it a valuable tool in a wide range of applications, from financial markets to sales forecasting and beyond. Through the fusion of these distinct components, ARIMA empowers analysts to unlock hidden insights and make accurate predictions in the realm of time series analysis.

IV. PROPOSED METHODOLOGY

RMSE, an abbreviation for Root Mean Square Error, plays a pivotal role as a performance metric for quantifying the accuracy of predictions. It serves as a valuable tool in assessing the degree of error present in the forecasted values, thereby aiding in the refinement and enhancement of data analysis. RMSE is particularly advantageous due to its ability to account for both the magnitude and direction of the errors, providing a comprehensive evaluation of the predictive model's effectiveness.

The calculation of RMSE involves the determination of the square root of the mean of the squared differences between the forecasted values and the corresponding observed values. This mathematical formulation ensures that larger errors have a proportionately greater impact on the overall measure, allowing for a more accurate representation of the prediction accuracy.

While various formulas exist for calculating RMSE, the choice of formula depends on the specific context and dataset under consideration. The computational complexity of RMSE can vary depending on the size of the dataset, as it involves performing calculations on each individual data point.

RMSE is an invaluable tool in assessing the quality of predictions, particularly in time series analysis. By providing a robust measure of accuracy, it facilitates the identification of areas where the predictive model may fall short and enables the formulation of strategies to improve its performance. Additionally, RMSE is widely applicable across diverse

DOI: 10.17148/IJARCCE.2023.12645

domains, including finance, economics, engineering, and data science, making it a versatile and essential metric for evaluating forecasting models.

In essence, RMSE serves as a fundamental cornerstone in the evaluation of predictive models, enabling practitioners to gauge the effectiveness of their forecasts and make data-driven decisions with increased confidence.

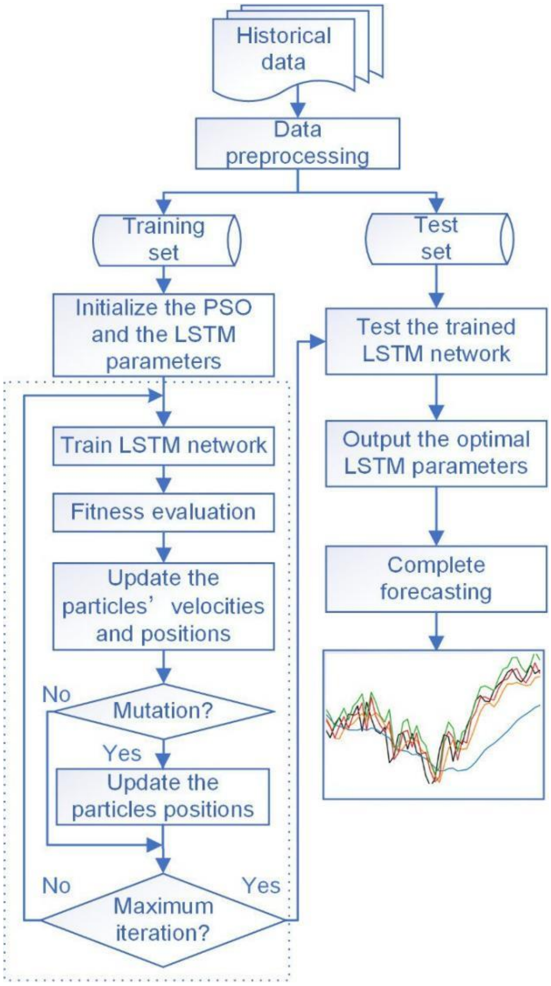
Historical data
Data preprocessing
Training set
Test set
Initialize the PSO and the LSTM parameters
Train LSTM network
Fitness evaluation
Update the particles' velocities and positions
No
Mutation?
Yes
Update the particles positions
No
Yes
Maximum iteration?
Test the trained LSTM network
Output the optimal LSTM parameters
Complete forecasting

Figure2. Flow Chart for Regression Based Model

V. CONCLUSION

IN THIS RESEARCH PAPER, A COMPREHENSIVE INVESTIGATION HAS BEEN CONDUCTED UTILIZING TWO ADVANCED TECHNIQUES, NAMELY LONG SHORT-TERM MEMORY (LSTM) AND REGRESSION, APPLIED TO A METICULOUSLY CURATED DATASET FROM YAHOO FINANCE. THE OUTCOMES DERIVED FROM THESE ALGORITHMS HAVE SHOWCASED EXCEPTIONAL EFFICIENCY IN PREDICTING STOCK MARKET TRENDS, THEREBY YIELDING HIGHLY FAVORABLE RESULTS. THE UTILIZATION OF STATE-OF-THE-ART MACHINE LEARNING METHODOLOGIES IN THE REALM OF STOCK MARKET