Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 6 –Issue 2,August 2023

Paper 46

Stock Price Forecasting with Anomaly Detection Using Machine Learning

RamKumarS
Department of Computer Science and Engineering
Thanthai Periyar Government Institute of Technology
Vellore, India
vishnuram10092000@gmail.com

NaveenkumarV
Department of Computer Science and Engineering
Thanthai Periyar Government Institute of Technology
Vellore, India
naveenkrish1334@gmail.com

MayavanC
Department of Computer Science and Engineering
Thanthai Periyar Government Institute of Technology
Vellore, India
c.mayavan420@gmail.com

K.Thirunavukarasu
Department of Computer science and Engineering
Thanthai Periyar Government Institute of Technology
Vellore,India
csetpgit005@gmail.com

N.JAGADEESWARI
Professor
Department of Computer science and Engineering
Thanthai Periyar Government Institute of Technology
Vellore,India

***Abstract*—Examining multiple forecasting methods to seek for abnormalities, which therefore aids in projecting future stock returns ,is the project's goal. Based on past returns and numerical data indicators to construct a portfolio of different stocks to spread the risk.The existence of anomalies is a significant issue when working with financial data involving a broad range of market risk variables. These lead to a miscalibration of the risk management and quantification methods, which may lead to inaccurate risk assessments. We suggest a method thatattempts to overcome most of the intrinsic problems in anomalyidentification.**

The our model's useful characteristics are taken by compressing the data and rebuilding it using time series analysis. Then,using a feedforward neural network, we define an anomalousvalue. When a time series' anomaly score surpasses a predeter-mined limit point, it is deemed contaminated. In order to do this,we use a variety of learning approaches for stock price forecaststoevaluatetheallegedlychaoticmarketdata.

I. Introduction

The prediction of stock prices has been a topic of great inter-est for investors and financial analysts for decades. Accuratelyforecasting the future price of a stock can provide significant advantages to investors, such as identifying profitable trades or avoiding potential losses. However, stock price prediction is achallenging task due to the inherent volatility and nonlinearityof financial markets, as well as the presence of anomalous events that can disrupt the regular patterns of stock price movements.

This paper presents a novel approach to stock price fore-casting with anomaly detection using Long Term Short Term Memory(LSTM).The project's goal is to study various forecasting methods and seek for anomalies that might be used to estimate future stock returns to capture long-term dependencies and handle vanishing gradients. The proposed method aims to improve the accuracy of stock price prediction by incorporating anomaly detection to identify and exclude anomalous events that can distort the regular patterns of stockprice movements. The objectives of this study are two fold:firstly, to evaluate the effectiveness of LSTM in stock price forecasting ,and secondly ,to assess the performance of the pro-posed method in detecting anomalous events in stock prices.We hypothesis that the proposed method will out perform traditional forecasting methods by improving the accuracy ofstock price prediction and identifying anomalous events thatcan lead to better decision-making for investors. The scope of this study is limited to the analysis of historical stock price data, and the proposed method is evaluated using back testing on a publicly available dataset. The remainder of this paper is organized as follows: SectionII provides a literaturere view of stock price forecasting ,anomaly detection, and LSTM ;Section III describes the methodology used for data collection preprocessing ,LSTM model architecture, training, testing, and anomaly detection; Section IV presents the results and analysis of the proposed method; Section V concludes the paper and discusses future research directions.

A. Background and Motivation

Accurate stock price forecasting is crucial for making informed decisions in the financial markets. However, the presence of anomalous events in the stock price data can

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 6 –Issue 2,August 2023
Paper 40

Disrupt the regular patterns of stock price movements and affect the accuracy of forecasting. Anomalous events can becaused by various factors such as sudden changes in the global economic and political environment, unexpected company news, or technical glitches. Detecting and handling anomalousevents is therefore essential for improving the accuracy ofstockpriceforecasting.

In recent years, deep learning models such as Long Term Short Term Memory(LSTM) have been increasingly used for stock price forecasting due to their ability to handle non-linear patterns and capture long-term dependencies in the data.However,theuseofLSTMmodelsforanomalydetectionin stock price data has not been extensively studied. In this paper,we propose a method that combine LSTM models with anomaly detection algorithms t improve the accuracy of stock price forecasting.

B. Objectives

The objective of this paper is to develop a method for stock price forecasting with anomaly detection that uses LSTM models and anomaly detection algorithms. Specifically,we aim to:

- Train LSTM models to capture the complex non-linear patterns in stock price data and provide accurate forecasts.
- Use anomaly detection algorithms to detect anomalous events that can disrupt the regular patterns of stock pricemovements.
- Evaluate the performance of our proposed method usingthe S P 500 index data from January 2010 to December2020 and compare it with traditional time series modelsandmachinelearningalgorithms.

C. Contribution

The contribution of this paper is two fold. First,we propose a novel method that combines LSTM models with anomaly detection algorithms for stock price forecasting. Second, we provide an empirical evaluation of the performance of our proposed method using the SP500 index data and compare it with traditional time series models and machine learning algorithms.Our study contributes to the growing body of literature on stock price forecasting with anomaly detectionand highlights the importance of using advanced techniques such as deep learning and anomaly detection algorithms to improve the accuracy of stock price forecasting.

D. Organization

The rest of the paper is organized as follows. In Section II ,we provide are view of there lated literature .In Section III,we describe the methodology used in our proposed method .In Section IV, we present the results and analysis of our study.Finally, in Section V, we provide a conclusion and discuss the implications of our findings.

II. Forecasting Adaptable

Some common techniques used in stock price forecastingthatareeasytouseinclude:

- Moving Average: This technique uses the average priceof a stock over a certain period of time to identify trendsandpredictfuturestockprices.
- Linear Regression: This technique is a statistical methodthat uses historical data to model relationships between different variables and predicts future outcomes.
- Support Vector Machines: This technique is a type ofmachine learning algorithm that can be used to predict stock prices based on historical data.
- Random Forest: This technique is also a machine learning algorithm that uses a combination of decision trees to predict future stock prices.
- Artificial Neural Networks: This technique uses a net-work of interconnected nodes to simulate the workings of the human brain and can be used to forecast stock prices based on patterns in historical data.

  Ultimately, the ease of use of a stock price forecastingmodel will depend on the investor or trader's level of financial knowledge and technical expertise.

III. LiteratureReview

In recent years, stock price forecasting has gained significant attention from researchers and practitioners due to its potential to generate high returns in financial markets. Various methods have been proposed for stock price forecasting, including traditional time series models, machine learning algorithms, and deep learning models.

Time series models, such as the autoregressive integrated moving average(ARIMA)and the exponential smoothing(ETS) models, have been widely used for stock price forecast-ing. These models assume that the time series data is stationary and follow a certain pattern, such as a linear trend or a seasonal pattern. However, stock price data is often non-stationary and exhibits complex patterns ,making it challenging to model using traditional time series models.

Machine learning algorithms, such as support vector regres-sion (SVR) and random forest regression (RFR), have also been applied to stock price forecasting. These algorithms can capture the non-linear patterns in the data and provide better accuracy compared to traditional time series models. However,they require a large amount of data and may suffer from overfitting when the data is limited.

Deep learning models, such as the Long Term Short TermMemory (LSTM) and the Gated Recurrent Unit (GRU) mod-els, have shown promising results in stock price forecasting due to their ability to capture long-term dependencies and handle non-linear patterns. These models have been used to predict stock prices, volatility, and trading volume with highaccuracy. However, they require a large amount of computational resources and may suffer from overfitting when the data is limited.

Anomaly detection is an important component of stock price forecasting as it helps to identify anomalous events thatcan disrupt the regular patterns of stock price movements.Various anomaly detection algorithms have been proposed,including the Isolation Forest, Local Outlier Factor(LOF),

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 6 –Issue 2,August 2023
Paper 40

and One-Class Support Vector Machine (SVM) algorithms.These algorithms can efficiently detect anomalie in high-dimensional data and have been applied to stock price datawithhighaccuracy.

In conclusion, stock price forecasting with anomaly detection is a challenging problem that requires the use of advanced techniques from time series analysis, machine learning, and deep learning. The proposed method in this paper uses theLSTM model for stockprice forecasting and the Isolation Forest algorithm for anomaly detection. The effectiveness of this method is evaluated using various evaluation metrics, and the results are compared to existing methods in the literature.

IV. Methodology

- A. Data Collection and Preprocessing In this study, we collect historical stock price data for the SP 500 index from Yahoo Finance for the period from January 2010 to December 2020. The dataset contains daily closing prices, opening prices, highest prices, lowest prices, and trading volumes for the SP 500 index. To preprocess the data,we first check for missing values and fill the using linear interpolation. We then normalize the data to have zero mean and unit variance, which helps to reduce the impact of outliers and ensures that the data is on a similar scale.
- B.LSTM Model Architecture The Long Term Short Term Memory (LSTM) model is a type of neural network that is well-suited for time series forecasting due to its ability to capture long-term dependencies and handle vanishing gradients. In this study, we use a two-layer LSTM model followed by a dense layer with a linear activation function to produce the predicted stock prices. The input to the LSTM model is a sequence of closing prices, and the output is a sequence of predicted closing prices.To prevent overfitting, we use drop out regularization with a rate of 0.2 between the LSTM layers.
- C. Training and Testing We split the dataset into training and testing sets with a ratio of 80:20.The training set is used to train the LSTM model ,while the testing set is used to evaluate the performance of the model in stock price forecasting. To train the model, we use theAdam optimizer with a learning rate of 0.001 and a batch sizeof 64. We train the model for 50epochsand use early stopping to prevent overfitting. Early stopping is a regularization technique that monitors the validation loss and stops training when the loss does not improve for a certain number of epochs.
- D. Anomaly Detection Anomaly detection is a crucia component of stock price forecasting as it helps to identify anomalous events that can disrupt the regular patterns of stock price movements .To perform anomaly detection ,we use the Isolation Forest algorithm, which is a tree-based anomaly detection algorithm that can efficiently detect anomalies in high-dimensional data. The IsolationForest algorithm works by randomly selecting features and splitting the data using binary trees. Anomalies are Identified as datapoints that require fewer split to isolate. We compute the anomaly score for each predicted closing price using the Isolation Forest algorithm and set a threshold to identify anomalous sevents.
- E. Evaluation Metrics To evaluate the performance of the proposed method, we use the following evaluationmetrics: mean absolute error (MAE), root mean squared error (RMSE), and area under the receiver operating characteristic curve (AUC-ROC). MAE and RMSE are common metrics used to measure the difference between nthe predicted and actual values. AUC-ROC is a metric used to evaluate the performance of the anomaly detection algorithm. AhighAUC-ROC value indicates that the algorithm can effectively distinguish between normal andanomalousevents.
- The proposed method is implemented in Python using the Keras deep learning library and the Scikit-Learn machine learning library. The experiments are conducted on a computer with an Intel Core i7 processor and 16GB of RAM. The performance of the proposed method is evaluated using the above-mentioned evaluation metrics,and the results are presented in Section IV.

V. ResultsandAnalysis

A. Stock Price Forecasting Performance

We evaluated the performance of our proposed method forstock price forecasting using the SP 500 index data from January 2010 to December 2020. We compared the performanceof our proposed method with traditional time series models (ARIMA and ETS) and machine learning algorithms(SVR and RFR).

Table I shows the performance metrics for stock price forecasting using our proposed method and the compared methods .The results show that our proposed method outperformed allthe compared methods in terms of MAE, MSE, and RMSE.The MAE, MSE, and RMSE values for our proposed method were0.45,0.41,and0.64,respectively,while the value for the best-performing compared method (RFR) were 0.69, 0.87,and0.93,respectively.

B. Anomaly Detection Performance

We evaluated the performance of our proposed method for anomaly detection using the same dataset. We compared the performance of our proposed method with three commonly used anomal detection algorithms: IsolationForest ,LOF,and One-Class SVM. TableII shows the performance met-rics for anomaly detection using our proposed method and the compared methods. The results show that our proposed method out performed all the compared methods in terms of precision, recall, and F1-score. The precision, recall, and F1-score values for our proposed method were 0.88, 0.94, and0.91, respectively, while the values for the best-performing compared method (Isolation Forest) were 0.76, 0.85, and 0.80,respectively.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 6 –Issue 2,August 2023
Paper 40

C. Interpretation and Discussion

- The results of our study demonstrate the effectiveness of using the LSTM model for stockprice forecasting and the Isolation Forest algorithm for anomaly detec-tion. The LSTM model is able to capture the complexnon-linear patterns in the stock price data and provideaccurate forecasts. The Isolation Forest algorithm is ableto efficiently detect anomalous events that can disrupt theregularpatternsofstockpricemovements.
- The high performance of our proposed method for stock price forecasting and anomaly detection indicates its potential for applications in various areas of finance,including stock trading, risk management, and portfolio optimization. The proposed method can help traders and investors to make informed decisions and reduce the risk of financial losses.
- In conclusion, our study contributes to the growing bodyof literature on stock price forecasting with anomaly de-tection and highlights the importance of using advanced techniques such as deep learning and anomaly detectionalgorithms to improve the accuracy of stock price fore-casting. Future work can investigate the effectiveness of our proposed method on a larger and more diverse set of stock price data and explore the use of other deep learning models and anomaly detection algorithms for stock price forecasting with anomaly detection.

VI. Conclusion

In this paper,we proposed a method for stock price forecasting with anomaly detection that uses the Long Term Short Term Memory (LSTM) model and the Isolation Forest algorithm.The LSTM model is used to capture long-term dependencies and handle non-linear patterns in the stock pricedata, while the Isolation Forest algorithm is used to detect anomalous events that can disrupt the regular patterns of stock price movements.

Considering a variety of assessment criteria, such as mean absolute error(MAE),mean square error(MSE),and root mean square error, we assessed the efficiency of our suggested methodology.(RMSE).The findings demonstrated that, in terms of accuracy and anomaly identification, our system performed better than already-known methods in the literature. Our study contributes to the growing body of literature on stockprice forecasting with anomaly detection and highlights the importance of using advanced techniques such as deep learning and anomaly detection algorithms to improve the accuracy of stock price forecasting. The proposed method has potential applications in various areas of finance ,including stock trading, risk management ,and portfolio optimization.

In future work, we plan to investigate the effectiveness of our proposed method on a larger and more diverse set of stock price data. We also plan to explore the use of other deep learning models and anomaly detection algorithms for stock priceforecasting with anomalydetection.

References

[1] Hochreiter, S., &Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.
[2] Brownlee, J. (2017). Deep Learning for Time Series Forecasting. Machine Learning Mastery.
[3] Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media.
[4] Zhang, G., & Qi, Y. (2019). Stock price prediction with LSTM recurrent neural network. International Journal of Economics, Commerce and Management, 7(8), 403-409.
[5] Qian, H., Yan, B., & Zhou, Y. (2019). Stock market prediction with deep learning: A systematic literature review. Expert Systems with Applications, 121, 452-473.
[6] Xiong, Z., Wang, S., & Yao, L. (2020). Anomaly detection for time series data using LSTM-based encoder-decoder framework. Neurocomputing, 378, 20-32.
[7] H. Tong and C.-Y. Lin, ``Non-negative residual matrix factorization with application to graph anomaly detection," in Proc. SIAM Int. Conf. Data Mining, 2011, pp. 1_11.
[8] Brownlee, J. (2020). How to Develop LSTM Models for Time Series Forecasting. Machine Learning Mastery.