Project Proposal

Stock price prediction via improved Machine Learning Techniques

Group I

Content

1. Introduction

In today's whole world is facing a financial crisis due to the covid-19 pandemic situation lots and lots Of investors are bankrupt due to unpredictable stock prices therefore investors lose their confidence in the investment. Financial markets have a significant impact on the economies in which they operate. Often Business organizations in modern economies are highly dependent on Funds generated by these financial markets. Therefore, Analyze the behavior and performance of financial markets have become an important area of research. These analyses include Predict the price of securities (for example stocks, bonds, etc.),

Exchange rates, market indicators, securities benefits, and trade volumes, inventory classification, etc. [1] in this research present a method to increase the investor's confidence by predicting the stock prices using machine learning.

Rising stock markets can indicate favorable economic conditions for companies, which successively translate into higher profits. On the opposite hand, the declining stock exchange indicates an economic downturn. Within the future, these trends are likely to point out the economy and stocks tandem. Within the case of stock markets, particularly, the task is gaining importance due to dynamic changes in market behavior and immeasurable economic benefits. By predicting stock indices, risk managers and professionals can see if their portfolio will decline within the future and should want to sell before it loses value. Therefore, research to predict future trends in financial indices is critical and necessary for investors that have an interest within the stock markets.

However, the behavior of stock markets depends on many factors like political situations, economic, Natural factors, and lots of others. The stock markets are dynamic and exhibit wide variation, and therefore the prediction of the stock exchange may be a highly challenging task because of its highly nonlinear nature and sophisticated dimensionality. Statistic forecasting is that the basic study to investigate data processes over ages. This is often a series of statistical observations recorded over statistic. It may be wont to realize past behavior of the series and supported past behavior it can forecast future behavior of the series. The target of sales forecasting is to assist the organization to work out the strain of products and improve its strategy for the long run [2].

This project purpose is to form a stock closing price prediction model to Sri Lankan Stock market which is called Colombo Stock Exchange (CSE). The resulting model is meant to be used as a call support tool or as an autonomous tool that predicts the long run value of the stock prices by analyzing the previous stock price data and ASPI (All Share Price Index).

Mainly this project proposes a solution for the unpredictable stock price fluctuation using Machine learning techniques such as LSTM Recurrent Neural Network. The outcomes of this research are a model that can predict stock closing prices and increase the investor confidence about the investment, then investors tend to invest in stocks that will be a favorable impact on the Sri Lankan economy as well.

The primary objective of this project is to develop a model to predict the stock closing prices shares using machine learning techniques therefore there can be several benefits as follows.

- Building confidence in the investment of local and foreign investors.

- To attract local and foreign investors to the stock market.
- Contributing to the country's economy through the stock market.
- Reduce stock and company analyzing time.

2. Problem Statement

Most of the investors in the Colombo stock exchange as well as other stock markets have difficulties regarding their investment whether the particular share price in a certain company will increase or decrease in the future. So investors lose their confidence regarding their investment. In today's whole world is facing a financial crisis due to the covid-19 pandemic situation lots and lots of investors are bankrupt due to unpredictable stock prices therefore investors lose their confidence in the investment. The significant impact on economies has been seen by the financial market. Often Business organizations in the current economy are highly dependent on the stock market. Therefore, examining the way of acting and performance of the stock exchange is an essential area of research. This analysis mainly includes the prediction of the price of securities for example stocks, bonds, etc. Exchange rate, securities benefit, market indicators, trade volumes, inventory classification, in this research present a method to increase the investor's confidence by predicting the stock closing prices using machine learning techniques.

3. Challenges

One of the main challenges associated with stock price prediction is the unpredictable nature of stock markets. The stock prices are affected by various internal and external factors such as economic indicators, news events, and market sentiment, making it difficult to accurately predict their future values. Another challenge is the large amount of data that needs to be analyzed, as stock prices are constantly changing and historical data must be carefully considered. Additionally, many machine learning algorithms struggle with modeling non-linear relationships, which are often present in stock market data. To overcome these challenges, it is important to carefully preprocess the data, select the most appropriate machine learning techniques, and evaluate the results effectively.

4. Initial Literature review

In [1], Adil MOGHARa et al developed a predictive RNN with Prediction of future stock prices using LSTMs, with varying number of epochs employed for training the data sets, such as 12,25,50, and 100 epochs. NYSE, GOOGL, and NKE companies from the New York Stock Exchange for stock price prediction. Historical data gathered from August 19, 2004 to December 19, 2019.from GOOGLE and NKE. This paper proposes RNN-supported LSTM for building then future forecast open prices for both NKE and GOOGL shares. During this research, Future work proposed to determine the optimal data collection length and increase

the training epochs number to improve the accuracy of stock predictions. The challenges of the area include dealing with non-stationary and highly volatile data, and accurately modeling the complex relationships between various factors that influence stock prices. In this research strengths were ability to handle sequential data and ability to capture long-term dependencies in the data. Its sensitivity to the choice of parameters and initial conditions and large amounts of data to train were the main limitations of this research.

In [2], Kesavan et al. developed an algorithm to predict the stock prices of the Indian market, where fluctuations are often driven by multiple factors such as political events, news, and public safety incidents. The study aimed to create a model that merges sentiment analysis of news sources (Twitter) with historical stock market data using LSTM. The polarity of sentiments was included to enhance prediction accuracy, enabling investors to make more informed investment decisions. In this research preprocessing is used to remove irrelevant data. The use of NLP aims to understand investor psychology and its impact on stock prices by introducing polarity scores for social media and news data. The next step involves incorporating the historical time series data.They finally used deep learning techniques like RNN to predict the historical pattern of the data. The challenges of the area include the high volatility of stock market prices, the vast amount of data available, and the difficulty of accurately predicting stock market trends. The challenges of the area include the high volatility of stock market prices, the vast amount of data available, and the difficulty of accurately predicting stock market trends. In this research main strengths were able to handle large amounts of data and is scalable and high accuracy rate of stock market prediction. In this research weaknesses were that method is only applicable to certain types of stock markets and may not be generalizable to other markets and method is only applicable to certain time periods and may not be generalizable to other time periods. There were some limitations in this research which is only applicable to certain types of stock markets and time periods. And method is dependent on the availability and quality of social media data.

A paper by A.J.P. Samarawickrama wt al.(referenced as [5]) developed a model to predict Colombo Stock Exchange share prices in selected sectors using RNN and identified limitations of the algorithms to measure it's accuracy. The architectural models chosen for this research were Feedforward, Long Short-Term Memory(LSTM), gated recurrent unit (GRU), and Simple Recurrent Neural Network (SRNN), and these models were deemed suitable for the purpose of this study.. Feed Forward Neural Network architecture also was used for the comparison of feedforward neural networks, also called multiple layer perceptron. Low, High, and Closing prices were used as the input variable for predicting future prices. This research analyzed three companies from different sectors, namely Commercial Bank from the banking and finance sector, Royal Ceramics Limited from the Manufacturing sector, and Jhon Keels Holding from the Diversified Holding sector. Data from 2002-01-01 to 2013-06-30 was collected and used to build a module. The study found that the LSTM model performed the best among all the RNN models and had the lowest error compared to other feedforward networks, but in some scenarios the feedforward network performed better. The GRU network

showed the highest error compared to other neural network methods. According to historical studies, recurrent neural networks, especially LSTM, produce the highest accuracy output compared with other models. MLP models have given the best result for this research because the previous two days have been used as an input variable. When considering the number of past days RNN module is given the best accuracy. There are some limitations also can be seen in this research like sometimes LSTM models is fails to store information for a long period of time and capacity of LSTM models have no limits.

Kang Zhang conducted research on stock market forecasting using a Generative Adversarial Network (GAN) in [6]. An LSTM RNN was used as the generator and GAN as the discriminator to predict future closing stock prices. The study showed promising results using the LSTM as the generator and a Multi-layer Perceptron as the discriminator. The model was trained using past data to forecast the closing price for the day. The research was based on actual stock prices from the Standard and Poor's 500 in Shanghai and focused mainly on IBM from the NYSE and PAICC from Ping An Insurance Company in China. MSFT from the NASDAQ, which is a national association of securities dealers automated quotation. Finally, in this research, compute the RMSE of the FIVE datasets for the evaluation criteria purpose, and MAE and HR are also calculated in that way. Typically, stock market prediction is made using classical methods such as Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM), and Support Vector Regression (SVR). This research also used these methods for comparison purposes. Results showed that the Generative Adversarial Network (GAN) model performed best with the lowest error, as measured by MAE, RMSE, and MAPE.

In [3] Jinghua Zhao et al conducted a stock market prediction model with recurrent neural networks. In share, the market has a long period of previous data so stock prices are closely related to each other which means the current transaction is normally based on previous transactions. RNN has better time series forecasting power than other frequently used models. This paper proposed different models like RNN, LSTM, and GRU respectively. This research introduces some attention mechanisms based on A-LSTM, AT-GRU, and AT-RNN. Using attention mechanism prediction accuracy can be increased. According to this research six models were implemented which are according to LSTM, RNN, and GRU. This research showed results as GRU-M and LSTM-M is performed well than RNN-M.GRU-M was given the highest accuracy than LSTM-M. Attention mode is given high accuracy. The relationship of this paper to others in the field of stock market prediction is that it adds to the existing body of literature by proposing a new prediction model using a recurrent neural network. The authors aim to address some of the challenges in the field, such as the non-linearity and volatility of the stock market, by using a recurrent neural network that can capture the dynamic relationships between different factors influencing stock prices. There are some weaknesses can be seen in this research also stock price trends are influenced by many factors and can be difficult to predict, the validity of the results

can also be dependent on the quality and representativeness of the data used to train the mode and Recurrent neural networks can be computationally intensive and time-consuming to train, especially for large datasets.

In [4] Priyank et al, developed a model for stock price prediction using BERT and GAN. Stock price prediction technical analysis is the most important thing. Technical analysis is done with the help of technical indicators because of the technical analysis investors can decide to invest or not for the share prices. One of the most important aspects is sentiment analysis which means the emotion of the investors that shows the willingness the investment. This research proposes state-of-the-art methods to forecast stock prices. Firstly, the sentiment analysis part of the news and headlines for Apple Inc was done with the help of BERT. BERT is a pre-trained transformer model developed by google Natural Language Processing (NLP). Currently, stock sentiments play a vital role in the stock market share prices Fin-Bert is used to analyze the sentiments of the Apple Inc stock price. In this research for the comparison purpose comparison Root Mean Square Error (RMSE). In this research proposed model has captured the trends and patterns with the historical data with the sentiments. In this research, the 5-day prediction proposed model has the highest accuracy but in the 30 days, prediction LSTM has the highest prediction. This research has limitations due to the varying impact of sentiment on different stock market sectors, such as banking and finance, the appeal sector, and manufacturing sector. A sector-wise analysis of sentiment and historical share prices is necessary for improved model accuracy. The research is also limited by the availability and quality of social media data, and the high cost and computational requirements of the BERT model used.

In [7] et al have implemented the model using machine learning techniques, like K-nearest-neighbour, Support Vector Regression, Random Forest Regression, linear Regression, for analyzing time-series data to predict stock price. In this research was developed a forecasting model by stacking multiple methods to find the best future forecast of the stock price. In this work was proposed an efficient distribution method which used the same training set with the parallel usage of the averaging methodology which combines Support Vector Regression, Linear regression model, KNN regression model, so four techniques were used of the machine learning algorithm. K – Nearest Neighbor, Support Vector Regression, Linear regression, and Random Forest regression. As data set time series data were taken by a variable over time such as daily sales revenue, monthly overheads, weekly orders, yearly income, daily stock prices and tabulated or plotted as chronologically ordered numbers or data points. This research also was given different accuracy for different approaches. so, there were some limitations in this research too slow for practical problems, sentiment analysis is not considered, low accuracy level and there can be other relevant factors that would be extremely affected for share prices. Therefore, to take the highest accuracy Recurrent neural network LSTM and deep learning are used for stock market prediction.
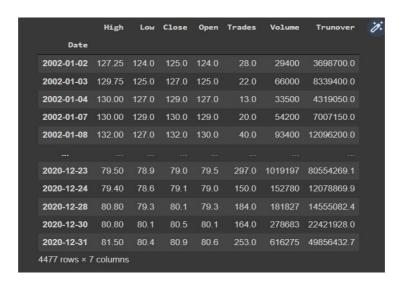
There are some limitations can be seen in the all the above research one thing is they don't consider some other factors which effected to the share prices like sentiments in our project the main scope is to predict based stock prices based on historical data as well as using sentiment data line sweeter sentiment data.

5. Proposed Model

The project is to develop a machine learning model to increase the investor's confidence by predicting the stock closing prices using machine learning techniques and identifying the most suitable machine-learning algorithm to predict the future stock and use sentiments to get predicted price accurately and increase investor confidence based on their investment.

6. Dataset description

Dataset was collected for this research from the Colombo stock exchange (CSE). In this research different data was used as a historical dataset which includes closing, opening, high, low, trades, volume, and turnover for each company. Other than that, there are a lot of data like All Share Price Index (ASPI), Sector indices, foreign holdings, foreign buying and dividends payments…. etc. For building the stock market closing price prediction model Sri Lankan stock market data was collected in Commercial bank (COMB.N) from 2002 to 2020 from the Colombo stock exchange data library as mentioned in the below image. Other than that throw-out this project we hope to get tweets to get more accuracy based on prediction.

Date	High	Low	Close	Open	Trades	Volume	Trunover
2002-01-02	127.25	124.0	125.0	124.0	28.0	29400	3698700.0
2002-01-03	129.75	125.0	127.0	125.0	22.0	66000	8339400.0
2002-01-04	130.00	127.0	129.0	127.0	13.0	33500	4319050.0
2002-01-07	130.00	129.0	130.0	129.0	20.0	54200	7007150.0
2002-01-08	132.00	127.0	132.0	130.0	40.0	93400	12096200.0
...	...	...	...	...	...	...	...
2020-12-23	79.50	78.9	79.0	79.5	297.0	1019197	80554269.1
2020-12-24	79.40	78.6	79.1	79.0	150.0	152780	12078869.9
2020-12-28	80.80	79.3	80.1	79.3	184.0	181827	14555082.4
2020-12-30	80.80	80.1	80.5	80.1	164.0	278683	22421928.0
2020-12-31	81.50	80.4	80.9	80.6	253.0	616275	49856432.7

4477 rows × 7 columns

7. Evaluation plan

Model evaluation can be done using R Square/Adjusted R Square, Mean Square Error (MSE)/Root Mean Square Error (RMSE) or Mean Absolute Error (MAE).

References

[1]   K. J. Kesavan M, "Stock Market Prediction with Historical Time Series Data and Sentimental Analysis of Social Media Data," 2020.

[2]   T. F. A. Samarawickrama, "A Recurrent Neural Network Approach in Predicting Daily Stock Prices, University of Sri Jayewardenepura, 2018.," 2018.

[3]   M. H. M. H. Adil MOGHARa, "Stock Market Prediction Using LSTM Recurrent Neural Network," 2020.

[4]   Z. D. Zhao J, "Prediction model for stock price trend based on recurrent neural network," Journal of Ambient Intelligence and Humanized Computing, 2021.

[5]   V. B. Priyank Sonkiya∗, "Stock price prediction using BERT and GAN," 2020.

[6]   G. KangZhang, "Stock Market Prediction Based on Generative Adversarial Network," 2019.

[7]   R. A. Md.Tanvir Rahman, "Forecasting Stock Market Price Using Multiple," 2018.

Member's contribution to the deliverable

Dilhan M.W.S. 239155F : Introduction, proposed solution

Samarasinghe T. A. A. 239173H : Evaluation plan

Wickramasinghe J.A.D.L. 239187E : Problem definition , Challenges

Wijesena C.M.C.H. 239188H : Dataset description