Predictive Analysis of NEPSE Using LSTM and Technical Indicators

1Bikash Kunwar, 2Paras Khati

1,2Department of Electronics and Computer Engineering, Paschimanchal Campus, Gandaki Province, Nepal

Abstract - The stock market in Nepal has gained a lot of attention in recent times and limited numbers of research are available to predict the stock market based on technical indicators. Predicting a stock market is not easy because of its non-linearity and volatility nature. It is very difficult to make a model that will accurately predict the time series data of the stock market but it can be predicted with some acceptable discrepancies.

Technical analysis is used to predict the stock market but there are some researches that support its effectiveness while other rejects this claim. There is a conflict with technical indicators work in the Nepal Stock Exchange Limited. This comparative analysis checks the effectiveness of the technical indicators to predict the stock market. Among all the other neural networks, long short-term memory is used in this research to predict the stock market.

The results of this research support the statement that technical indicators have some sort of relation between them and closing value which also shows in the correlation matrix. Fundamental and sentimental data along with technical indicators produced more effective results than those of fundamental and sentimental data.

Keywords: Stock market, technical indicators, technical analysis, Nepal Stock Exchange Limited, long short-term memory, fundamental data, sentimental data.

I. INTRODUCTION

In today's world of economies all people are searching for methods to earn money and companies are also searching for investment for their expansion. Both of these needs can be fulfilled through the stock market using various methods. The stock market is a public market where many companies are listed to get money by trading their shares at an agreed price with the people. The stock market is a volatile, non-parametric, dynamic market where the buying and selling of the shares of the company changes the price. Predicting the share price of a company is a very risky task as the stock market depends on the bank's interest rate, profit-to-equity ratio, fundamentals of a company, news sentiment, political stability, and other fundamental indicators. When the people buy the shares they can earn from either the long term or short term. Much research has been done in the field of the stock market to predict stock prices. There are various methods that can be used to predict stock prices on the basis of historical data related to a company. In this paper, we developed a program that predicts the stock prices of a particular share using previous stock prices and data treating it as training sets. The model of the program is based on the use of historical share prices, past news, and different technical indicators affecting the prices and implements Artificial Intelligence and Machine Learning techniques to speculate the stock prices. It uses the time series analysis for prediction of the stock prices considering the Nepal Stock Exchange Ltd. (NEPSE). The share prices of various companies in NEPSE depend on many factors such as earnings and profits, financial health, valuation metrics, news and events, market trends, market liquidities, etc. There are mainly two types of ways to predict the stock price, assuming the linearity and non-linearity in the data of the stock market. Since the practical data does have random non-linear attributes, we have assumed the data to be random and non-linear. Different machine learning techniques, high computational qualities, and availability of big data along with various machine learning models have made the analysis and determination of stock prices a lot easier.

There has been a lot of research undertaken on the prediction of stock prices using various deep-learning methods. Most of them have made use of multilayered neural networks, long short-term memory (LSTM), convolution neural network (CNN), gated recurrent unit (GRU), and their respective hybridization techniques in stocks. On the other hand, there has not been more prominent research to determine the stock prices of any company from NEPSE using technical indicators for market analysis.

It has been found that some of the researchers for this model say the use of these technical indicators and their efficacy in the prediction of future stock prices in the context of NEPSE is unpredictable[2]. So, our assumption is that by using these indicators doing the technical analysis, we can get adequate results to foresee the share price of a certain company. This proposed model illustrates the prediction of share prices of companies that are using NEPSE as a platform to buy or sell their stocks. It is based on the use of RNN and
LSTM along with multiple hidden layers using the financial news data and technical indicators, figuring out the score and determining the price result. The primary objective of this research is a comparative study and to find the effectiveness of technical indicators in predicting the stock market in the Nepali market, say NEPSE under an identical model [1].

II. LITERATURE REVIEW

2.1 Background

The Nepal Stock Exchange is commonly known as NEPSE. NEPSE is the only stock exchange in Nepal. NEPSE, a regulatory organization in Nepal works under the Securities Act, of 2007. Its main purpose is to regulate and enhance the marketability and liquidity of investment and trading of corporate securities in the secondary market by facilitating transactions through various market intermediaries, including brokers and market makers. All the irregularities are looked at by NEPSE, which also control and correct the irregularities by making necessary correction in the market NEPSE opened its trading floor on 13th January 1994. The government of Nepal, Nepal Rastra Bank, Nepal Industrial Development Corporation, and members are the shareholders of NEPSE [3].

All the people who are using the stock market for the aim of gaining money must be able to analyze the data and get the necessary information to predict the market trend. Predicting the market is the most important task to get benefit from the market. There are various factors to change the stock market price and analyzing all of them may be difficult but there are some statistical tools that are based on the previous data of stock called technical indicators, which help us to predict the stock market. It doesn’t depend upon the fundamentals of a company like interest rate, non-performing loans, profit, etc.

Recently after the covid, many people have started using the stock market but the reverse was different. When in the stock market and the use of technical analysis began around 5-10 years back there was a controversy that technical analysis would not work in the Nepali market. Online trading also began just 5 years back and the online charts for technical analysis have just begun in the same period. The research in this field in Nepal is limited so people have been taking either side just hypothetically theories and without any statistical data.

2.2 Related work

Ko & Chang (2021) stated that Li, Bu & Wu (2017), as well as Liu et al. (2017), described that the early research on the stock market was based on the random walk theory and the Efficient Market Hypothesis (EMH). Those past studies believe that stock market forecasting is random and no one can predict it. According to the Efficient Market Hypothesis stock market is largely driven by news like new information, rather than present and past prices, and news is unpredictable so stock market prices will follow a random pattern and cannot be predicted [4].

Gurung, (2004) takes 10-year span data from SEBON and NEPSE to study the growth trend and analyze the performance of the Nepalese securities market. He takes the number of listed and traded companies and their securities, the number of transactions, trading turnovers, paid-up value, market capitalization, and the NEPSE index as the variables. After the studies, he concluded that there is no synchronization among different securities market performance indicators, but it is true that they almost depicted an erratic trend during the observed period [5]. But Neupane, (2017) studied a stock market and concluded that the stock market prices are not random and future prices can be estimated by analyzing the historical information and claimed that the Nepalese capital market can be exploited with the help of technical analysis [6].

Vaidya, (2020), discusses Fama,(1970) work, stating that technical analysis has been a major tool for investors in analyzing and forecasting stock. Still, many academicians tend to believe that markets are informationally efficient and all available information is impounded in current prices. Later Lo and Mackinlay (1988,1999) showed that past prices may be used to forecast future returns to some degree which was taken for granted by technical analysts and rejected the Random Walk Hypothesis for US stock indexes [7].

Neely et al.(2014) compared the predictive ability of technical indicators with that of macroeconomic variables in U.S. equity risk premium and found that the technical indicators were statistically significant in both in-sample and out-of-sample predictive power. Furthermore, technical indicators better detect equity risk decline near business cycle peaks, whereas macroeconomic variables more readily pick up the rise in the equity risk premium. They get results better when the information from technical and macroeconomics is combined [8].

Karki et al. (2023) conducted research where they analyzed the data and examined the performance of technical indicators through modeling, back testing, and statistical analysis using different performance metrics. They found that the research generally supports the effectiveness of the technical analysis in NEPSE. However the RSI generates negative returns and some other indicators like the Bollinger band, simple moving average, etc. also fail and raise questions such as “Does technical analysis work on the Nepali market [9].
Joshi (2023) researched and found out that the nature of index volatility makes predicting the stock market impossible for investors to profit from stocks. He has collected the data for 6 months and used technical trading tools along with a survey where 7 people out of 62 said it does not predict the price. But, in the context of Nepal's unstable government and global crisis predicting from candlesticks shows mixed results whereas the market is in a bearish trend. So, prediction of the stock market using technical indicators in the context of Nepal is conceivable, but not certain and leads to conflicting conclusions [10].

III. METHODOLOGY

3.1 Data description and preparation

Market data plays a vital role in the analysis of the stock in the financial market. It is collected and analyzed in order to understand the patterns and insights so that it can be used for future decisions and investments. Various metrics and parameters are used to determine the risks in the market and historical analysis and technical analysis that determines the major changes occurring in the market. To perform these assessments the historical data, sentiment data, and technical indicators value are needed.

3.1.1 Fundamental data

Open price, High price, Low price, Close price, and Turnover come under the fundamental or historical data. All this is data that the stock runs into and provides the basic information about the stock. Open price is the first transaction price of the stock on that day whereas Close price is the final or last transaction price of the stock on a day. High price and low price are the highest and lowest prices or values of the stock at which the transaction happened on that day. Volume and turnover are the total number of shares and total values of securities or bonds traded in a day. All the historical data are collected from the Nepse Alpha in the Date range 1st August 2013 to 30th August 2023 as a CSV file.

<table>
<thead>
<tr>
<th>Date</th>
<th>Open</th>
<th>High</th>
<th>Low</th>
<th>Close</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>8/1/2013</td>
<td>732</td>
<td>746</td>
<td>722</td>
<td>726</td>
<td>16832</td>
</tr>
<tr>
<td>8/4/2013</td>
<td>726</td>
<td>730</td>
<td>710</td>
<td>710</td>
<td>13593</td>
</tr>
<tr>
<td>8/5/2013</td>
<td>710</td>
<td>710</td>
<td>690</td>
<td>702</td>
<td>10674</td>
</tr>
<tr>
<td>8/6/2013</td>
<td>702</td>
<td>707</td>
<td>662</td>
<td>700</td>
<td>10700</td>
</tr>
<tr>
<td>8/7/2013</td>
<td>700</td>
<td>707</td>
<td>686</td>
<td>690</td>
<td>9805</td>
</tr>
<tr>
<td>8/8/2013</td>
<td>690</td>
<td>690</td>
<td>670</td>
<td>670</td>
<td>21368</td>
</tr>
<tr>
<td>8/11/2013</td>
<td>670</td>
<td>670</td>
<td>660</td>
<td>663</td>
<td>6558</td>
</tr>
<tr>
<td>8/12/2013</td>
<td>663</td>
<td>664</td>
<td>655</td>
<td>657</td>
<td>5132</td>
</tr>
<tr>
<td>8/13/2013</td>
<td>657</td>
<td>665</td>
<td>650</td>
<td>658</td>
<td>4580</td>
</tr>
</tbody>
</table>

3.1.2 Financial news data

The financial news is very important to predict the stock market trend in present times and it will cover all the news that will affect the stock market. The financial news is collected from the website Sharesansar through web scraping using Beautiful Soup. The News data is collected for the same Date range as we have collected the fundamental data.

<table>
<thead>
<tr>
<th>Date</th>
<th>Headline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wednesday,</td>
<td>NEPSE Loses Minimal 2.82 Points to Close at 1990.59 Levels; Market Cap. at</td>
</tr>
<tr>
<td>August 30, 2023</td>
<td>Rs. 29.79 Kharba</td>
</tr>
<tr>
<td>Wednesday,</td>
<td>Implementation of the Millennium Challenge Corporation (MCC) Project</td>
</tr>
<tr>
<td>August 30, 2023</td>
<td>Begins</td>
</tr>
<tr>
<td>Wednesday,</td>
<td>See NAV of MFs Managed by Sanima Capital Limited for Shrawan</td>
</tr>
<tr>
<td>August 30, 2023</td>
<td></td>
</tr>
<tr>
<td>Wednesday,</td>
<td>11 Percent L.B.B.L. Debenture 2089 (LBBLD89) &amp; 11 Percent Mahalaxmi</td>
</tr>
<tr>
<td>August 30, 2023</td>
<td>Debenture 2089 (MLBLD89) Listed on NEPSE</td>
</tr>
<tr>
<td>Wednesday,</td>
<td>RBB Merchant Banking Published NAV Report of Two MF Schemes Until Shrawan</td>
</tr>
<tr>
<td>August 30, 2023</td>
<td></td>
</tr>
</tbody>
</table>
3.1.3 Technical Indicators

In this set of data Moving Average Convergence Divergence (MACD), Average True Range (ATR), Relative Strength Index (RSI), and Money Flow Index (MFI) are used as technical indicators. MACD measures the strength and trend or detects the price movement in the stock market. It is calculated by taking the difference between the 26-day exponential moving average and with 12-day exponential moving average [7, 11]. ATR is a market volatility indicator and is calculated as given below.

\[
True\ Range\ (TR) = \max([H - L], |H - \text{Cp}|, |L - \text{Cp}|)
\]

\[
\text{Average True Range}(ATR) = \frac{1}{n} \sum_{i} TR_i
\]

RSI, a momentum indicator that measures the change in price movements to indicate the overbought or oversold of the stock and is computed as,

\[
\text{Relative\ Strength}(RS) = \frac{\text{Average gain}}{\text{Average loss}}
\]

\[
\text{Relative\ Strength\ Index}(RSI) = 100 - \frac{100}{1 + RS}
\]

MFI is designed to detect the strength of money flowing out or in from a particular stock over a fixed period of time and also finds out the direction of short-term price movements (e.g. Bernard & Thomas, 1990). MFI is defined mathematically as below.

\[
\text{Typical price}(TP) = (\text{High} + \text{Low} + \text{Close})/3
\]

\[
\text{Raw Money Flow}(RMF) = TP \times Volume
\]

If the current TP is greater than the previous TP then RMF is said to be Positive Money Flow (PMF) and similarly, if the current TP is less than the previous one then RMF is said to be Negative Money Flow(NMF). If TP doesn't change it can't be added to both PMF and NMF. Adding all of the RMF that are said to be PMF we get the value for PMF and similarly, we get NMF by adding all the RMF values that are said to be NMF.(Pokhrel et al., 2022).

\[
\text{Money ratio}(MR) = \frac{PMF}{NMF}
\]

\[
\text{Money Flow Index}(MFI) = 100 - \frac{100}{MR}
\]

The null and blank values are handled, some are removed and some are filled using the mean from the collected data. The data after the preprocessing is shown in the table below.

<table>
<thead>
<tr>
<th>Date</th>
<th>ATR</th>
<th>MFI</th>
<th>MACD</th>
<th>RSI</th>
<th>Score</th>
<th>Open</th>
<th>High</th>
<th>Low</th>
<th>Close</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>8/1/2013</td>
<td>18.934</td>
<td>77.496</td>
<td>3.6772</td>
<td>83.870</td>
<td>0.08156</td>
<td>227</td>
<td>234</td>
<td>228</td>
<td>233</td>
<td>3874</td>
</tr>
<tr>
<td>8/4/2013</td>
<td>18.544</td>
<td>75.396</td>
<td>4.4625</td>
<td>84.848</td>
<td>0.22854</td>
<td>233</td>
<td>239</td>
<td>230</td>
<td>235</td>
<td>6768</td>
</tr>
<tr>
<td>8/5/2013</td>
<td>18.432</td>
<td>68.939</td>
<td>5.7449</td>
<td>88.095</td>
<td>0.14472</td>
<td>235</td>
<td>244</td>
<td>230</td>
<td>244</td>
<td>16970</td>
</tr>
<tr>
<td>8/6/2013</td>
<td>18.311</td>
<td>68.873</td>
<td>6.3650</td>
<td>77.5</td>
<td>0.14512</td>
<td>244</td>
<td>244</td>
<td>232</td>
<td>240</td>
<td>8059</td>
</tr>
<tr>
<td>8/7/2013</td>
<td>16.258</td>
<td>79.560</td>
<td>7.5761</td>
<td>87.234</td>
<td>0.16125</td>
<td>240</td>
<td>255</td>
<td>239</td>
<td>250</td>
<td>14766</td>
</tr>
<tr>
<td>8/8/2013</td>
<td>15.894</td>
<td>79.113</td>
<td>8.9172</td>
<td>87.5</td>
<td>0.14087</td>
<td>250</td>
<td>256</td>
<td>247</td>
<td>256</td>
<td>10429</td>
</tr>
</tbody>
</table>
3.2 LSTM (Long Short-Term Memory)

3.2.1 Background

LSTM is one of the popular variants of RNN (Recurrent Neural Network) for time series forecasting. RNN is able to extract the temporal dependencies. It can perform better than the earlier methods used in deep learning like feed-forward networks. Its cells can transform signals forward and backward, as shown below [13].

3.2.1.1 Forget Gate

The forget gate is used to remove the information in the cell state which is not useful. Two inputs \( x_t \) (input at the current time) and \( h_{t-1} \) (previous cell output) are given to the forget gate and a sigmoid activation function is used in the forget gate which gives a value in the range of 0 to 1. Value in the cell state is going to be added with the cell state value after \( ft \) multiplied with \( c_{t-1} \).

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

3.2.1.2 Input Gate

The input gate is responsible for adding useful information to the cell state using the sigmoid and tanh activation functions. First, the sigmoid function works the same as in the forget gate with the same inputs \( x_t \) and \( h_{t-1} \). Whereas the tanh function creates a vector from all the possible values of \( x_t \) and \( h_{t-1} \). Later both the output of sigmoid and tanh are multiplied to get the useful information that is going to be added with the cell state value after \( ft \) multiplied with \( c_{t-1} \).

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

\[
c_t = \tanh(W_c[h_{t-1}, x_t] + b_c)
\]

3.2.1.3 Output Gate

This portion decides what information is provided to the next layer as output by extracting useful information from the current cell state. First, a vector is generated by applying the tanh function in the cell state and then multiplied by a sigmoid activation function output, which has two inputs as in the forget gate.

\[
c_i = f_i \ast c_{t-1} + i_t \ast \hat{c_t}
\]

3.2.2 Modeling and hyperparameter tuning

In this experiment, the core of a model is composed of LSTM layers. The LSTM model has an input layer followed by the number of LSTM layers, a dropout layer, and finally a dense output layer. The number of units in the input layer is selected using hyperparameter search keeping 32 units, 256 units, and 32 units as minimum, maximum, and step respectively. Also, the number of units in the hidden LSTM layers is the same as in the input layer. In the dropout layer, the dropout rate is also finalized between 0.2 and 0.5 using a hyperparameter search having 0.1 as a step to minimize the effect of overfitting by dropping some neurons in the LSTM. Along with the dropout layer, the Early stopping function manages the overfitting and underfitting due to too many or too few epochs. In this method, we have used epoch as 100 and batch size as 30 but the function will stop the training process if there is no improvement in the model after each epoch or only negligible improvement.
The data is split into training, validation, and test data. The loss function MSE (Mean Squared Error) is calculated at the end of each epoch and the training process is stopped if there is no improvement in the model or no reduction in loss. It doesn’t mean there is no improvement further if two epochs have the same loss value. It may be better after that epoch or worsen before getting better so we have set a bit of a delay mechanism and set the value of patience as 10. So it means if the loss is nearly the same up to 10 consecutive epochs the early stopping function stops the training process as it is not improving.

Reddi (2019) described that Adam is an efficient stochastic optimization that works with a small memory requirement and requires only first-order gradients and the name Adam is derived from adaptive moment estimation. It is used to generate the adaptive learning rates for each parameter. It combines the two methods named the first popular algorithm ADAGARD (Duchi et al., 2011; McMahan & Streeter, 2010), which works well for sparse settings, and RMSProp (Tieleman & Hinton, 2012), which works well in non-stationary settings [14].

3.3 Evaluation Metrics

For the evaluation of the output from a model RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), and $r^2$ (Coefficient of determination) are used. RMSE is the square root of the mean of the squared error. It calculates the average distance between the actual and predicted values of a model. MAE is the average of the sum of the differences between the actual and predicted values without taking the direction into account. The range of RMSE and MAE is 0 to $\infty$. The coefficient of determination is a function to check how well the model is fit. Its value came in the range between 0 and 1. The higher coefficient means the higher number of points are passed through the equation. The value closer to 1 is a good fit whereas the value nearer to 0 is a bad fit. The mathematical formulas of all evaluation metrics are given below. [15]

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y}_i)^2}
\]

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |(y_i - \bar{y}_i)|
\]

\[
r^2 = 1 - \frac{\sum(y_i - \bar{y}_i)^2}{\sum(y_i - \bar{y})^2}
\]

Where $y_i$ = $i^{th}$ actual value
$\bar{y}_i$ = $i^{th}$ predicted values
$\bar{y}$ = mean of all actual values

IV. RESULT AND DISCUSSION

The model designed in this experiment was executed for the number of lag values keeping all other parameters the same. It contains input set I with technical data along with fundamental data and sentiment score while input set II with fundamental data and sentiment score only. These two sets are trained under the model and output from different input datasets based on the values of RMSE, MAE, and $R^2$ are compared. In the experiment, the number of lag values is changed in the input and so are the values of the RMSE, MAE, and $R^2$.

Table 4: Comparison of RMSE, MAE, and $R^2$ for the result

<table>
<thead>
<tr>
<th>Lag Value</th>
<th>Input set I (with technical)</th>
<th>Input set II (without technical)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>MAE</td>
</tr>
<tr>
<td>1</td>
<td>14.805</td>
<td>11.212</td>
</tr>
<tr>
<td>3</td>
<td>22.666</td>
<td>18.663</td>
</tr>
<tr>
<td>5</td>
<td>16.816</td>
<td>12.254</td>
</tr>
<tr>
<td>7</td>
<td>27.510</td>
<td>23.032</td>
</tr>
<tr>
<td>9</td>
<td>30.411</td>
<td>25.804</td>
</tr>
<tr>
<td>11</td>
<td>26.936</td>
<td>22.278</td>
</tr>
<tr>
<td>13</td>
<td>17.270</td>
<td>12.851</td>
</tr>
<tr>
<td>15</td>
<td>20.093</td>
<td>15.421</td>
</tr>
<tr>
<td>AVG</td>
<td>22.063</td>
<td>17.689</td>
</tr>
</tbody>
</table>

From the above table, it is clearly seen that the average value of the coefficient of determination ($R^2$) for input set I is greater than the input set II with a value of .932 and .844 respectively for input sets I and II. Along with the coefficient of determination ($R^2$), RMSE and MAE average values are lesser for input set I than II. Also, the best model seen in the entire table is for input set I with a lag value of 1 and with a .973 coefficient of determination.

The correlation heat map matrix shows a correlation between all the variables.
Figure 4: Correlation Matrix of dataset

The pictorial representation for the value of the predicted and actual price in the form of a bar plot is given below.

Figure 5: First 20 predicted and actual data Bar plot

The graphical representation of the error during model training is given below.

Figure 6: Learning curve for the model with lag value 1

The scatter plot for the trained model is given below.

The Kdeplot of the distribution of actual and predicted data is given as.

Figure 7: Scatter plot for the result

Figure 8: Kdeplot for the actual and predicted values of data

In the above figure, the blue line is the actual price and the yellow line is the predicted price.

V. CONCLUSION

Many people are predicting the stock markets using technical analysis to better predict the pattern and provide high accuracy. To predict a stock market overall is a difficult task and to find out the involvement of technical indicators in this prediction or say to find out if there is a contribution from technical indicators to predict the stock market is challenging. In this study we used an LSTM model with overall 10 features of which 3 are technical indicators, others are fundamental with a sentiment score. The LSTM model is trained for multiple lag values and compares the RMS, MAE, and $R^2$ values to determine the average values of RMSE, MAE, and $R^2$ as 22.063, 17.689, and .932 for Input set I whereas 32.598, 29.280 and .844 for Input set II. This illustrates that technical
indicators could be taken into account in predicting the stock market in the NEPSE platform. Also, in the correlation matrix, the correlation value of technical indicators with close is sufficiently large to provide support to the statement that technical indicators have some sort of relationship with closing values.

This experiment can further be extended to perform statistical analysis like t-test, and ANOVA in the near future, and the experiment can be performed with other companies in different sectors to implement stock forecasting with more precise outcomes. Sentiment scores can be calculated using the Nepali language and more features like inflation, interest rate, etc. can be added to increase the accuracy. We can even use various Hybrid models to predict the stock.

REFERENCES


AUTHORS BIOGRAPHY

Bikash Kunwar,
Department of Electronics and Computer Engineering, Paschimanchal Campus, Gandaki Province, Nepal.

Paras Khati,
Department of Electronics and Computer Engineering, Paschimanchal Campus, Gandaki Province, Nepal.