TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

THESIS No: 073MSICE603

STOCK PRICE PREDICTION USING DEEP NEURAL NETWORK

BY

BASANT RAJ PHULARA

A THESIS REPORT
SUBMITTED TO THE DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN INFORMATION AND COMMUNICATION ENGINEERING

DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING

August 2021

STOCK PRICE PREDICTION USING DEEP NEURAL NETWORK

BY

BASANT RAJ PHULARA
073/MSICE/603

Thesis Supervisor
Dr. Basanta Joshi

A thesis submitted in partial fulfillment of the requirements for the degree of Masters of Science in Information and Communication Engineering

Department of Electronics and Computer Engineering
Institute of Engineering, Pulchowk Campus
Tribhuvan University
Lalitpur, Nepal

August 2021

COPYRIGHT©

DECLARATION

The dissertation is submitted for the fulfillment of degree in **Master of Science in Information and Communication Engineering (MSICE)** to Department of Electronics and Computer Engineering, Pulchowk Campus, Institute of Engineering under the title **“STOCK PRICE PREDICTION USING DEEP NEURAL NETWORK”**. I swear this work has not been previously submitted and is completely my own work. This work has not been previously submitted to university to fulfill any academic award.

Department of Electronics and Computer Engineering, Pulchowk Campus has authorized to freely distribute this thesis to other institutions or any individuals for further scholarly research work.

Basant Raj Phulara
073MSICE603
Date: August, 2021

RECOMMENDATION

The undersigned certify that they have read and recommended to the Department of Electronics and Computer Engineering for the acceptance, a thesis entitled **"STOCK PRICE PREDICTION USING DEEP NEURAL NETWORK"**, submitted by **Basant Raj Phulara** in partial fulfillment of the requirement for the award of the degree of **"Master of Science in Information and Communication Engineering"**.

....

Supervisor: Asst. Prof. Basanta Joshi, PhD
Department of Electronics and Computer Engineering
Institute of Engineering, Tribhuvan University

....

External Examiner: Adesh Khadka
IT Director,
Ministry of Education, Science and Technology

....

Committee Chairperson: Asst. Prof. Basanta Joshi, PhD
Program Coordinator, Master of Science in Information and Communication Engineering
Department of Electronics and Computer Engineering
Institute of Engineering, Tribhuvan University

Date: August, 2021

DEPARTMENTAL ACCEPTANCE

The thesis entitled **"STOCK PRICE PREDICTION USING DEEP NEURAL NETWORK"**, submitted by **Basant Raj Phulara** in partial fulfillment of the requirement for the award of the degree of **"Master of Science in Information and Communication Engineering"** has been accepted as a bonafide record of work independently carried out by him in the department.

....

Prof. Dr. Ram Krishna Maharjan
Head of the Department
Department of Electronics and Computer Engineering,
Pulchowk Campus,
Institute of Engineering,
Tribhuvan University,
Nepal.

ACKNOWLEDGMENT

I heartily express my sincere gratitude to my supervisor **Dr.Basanta Joshi**, Assistant Professor of Pulchowk Campus, for his encouragement, support, and continuous guidance throughout my thesis work.

I would like to express gratitude to all respected teachers for their guidance, encouragement, and support for the thesis proposal development. I would also like to acknowledge the authors of various papers and developers of the programming libraries especially python that has referenced and will utilize in my project.

I would like to thank Mr. Bharat Bhatta, Mr. Baikuntha Acharya, and Mr. Bishal Singh for supporting me in this work.

I would like to thank our Department of Electronics and Computer Engineering for providing an opportunity of learning and implementing the knowledge in the form of a thesis.

I am also thankful to all the friends, relative and family members whose direct and indirect support is very valuable in this research work.

Basant Raj Phulara

073/MSICE/603

ABSTRACT

Forecasting the financial market is one of the practical problems in the economic field. The noisy and the volatility are the two characteristic that hinders the timely prediction of the stock future price. In order to further resolve the drawbacks of the existing models in dealing with non-stationary and non-linear characteristics of high frequency financial time series data, this research work proposes the Wavelet transform based data preprocessing and developing the LSTM-attention model including the human sentiment for stock price prediction. The financial time series is smoothened by Wavelet transform, LSTM and attention mechanism is used to extract and train its features. Also the impact of human sentiment is investigated by adding the sentiment polarity score to historical dataset. The results of the proposed model are compared with the other two models, including LSTM and GRU on four different stocks ADBL, NIB, NABIL and SCB datasets. The Performance of the different models is evaluated based on RMSE, MAE, coefficient of determination R^2, and MDA. The results from the experiments on all the stock datasets shows that RMSE, and MAE is less than 2.5 and 2.2 respectively and R^2 is greater than 0.94 and MDA greater than 0.79. The results show that the proposed model along with the addition of human sentiment outperforms other similar models.

Key Words: Sentiment analysis, stock price prediction, Long short term memory (LSTM), Gated Recurrent Unit (GRU), Attention mechanism, Deep learning

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ANN	Artificial Neural Network
BoW	Bag of Words
CNN	Convolutional Neural Network
DMA	Mean Directional Accuracy
DNN	Deep Neural Network
EMD	Empirical Model Decomposition
EMH	Efficient Market Hypothesis
IPO	Initial Public Offering
MAE	Mean Absolute Error
MDA	Mean Directional Accuracy
MSE	Mean Square Error
LSTM	Long Short-Term Memory
NLP	Natural Language Processing
R^2	Coefficient of Determination
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SAEs	Stack Auto Encoders
SNR	Signal to Noise Ratio
SVM	Support Vector Machine
SVR	Support Vector Regression
TFIDF	Term Frequency Inverse Document Frequency
VADER	Valence Aware Dictionary and Sentiment Reasoner
WT	Wavelet Transform

CHAPTER 1

INTRODUCTION

1.1 Background and Motivation

Stock market is a place where there is buy and sell of the share of publicly listed companies. Initially company raises their fund by offering initial share which is called Initial Public Offering (IPO). The price of the stock market will go up or fall down depends on the market demand and supply. The more the investor willing to buy the stock price will go high and vice-versa. Although one can judge the number of sell and buy, it is very difficult to find out what factors contribute these transactions. This may be due to different reasons like market behavior, inflation, trends, and also the investor sentiment and other news.

The investor aims to make profit doing investment in the stock market by carefully forecasting the future trends. In order to make the fruitful decisions, one must go through the technical analysis that include company charts, stock market indices, and user posts , comments in the forums.

Therefore, to forecast the future trends automatically many Artificial Intelligence (AI) methods have been devised. The first research was done in 1991; by using machine learning models a comparative study was done [1]. Since then much more researches have been done in this field for precisely forecasting the financial time series.

There are many statistical models like autoregressive integrated moving average (ARIMA) to forecast the share market using the historical financial data [17]. But due to the non-linear behavior of the time series data these statistical models are not very efficient.

It is not possible to interpret the random nature of the stock market manually and is really boring and tedious. But with the advent of the artificial intelligence, big data and increased computational capabilities one can automatically forecast the future stock price and the forecasting become more feasible. By the advent of deep learning, a branch of machine learning, it is possible to make an appropriate forecasting models, and there are a lot of research has been done that will precisely forecast the future prices.

1.2 Problem Statement

The stock market is the place where the share of public companies are bought and sold. In the past, very few people invest their money and make a profit. But nowadays, it is the hotspot to invest money and everyone is willing to invest and make benefits. One may bear the loss if s/he invests money haphazardly because the stock market is governed by different risk factors. Forecasting future stock price is matter of economic research but it is so challenging the timely prediction of the market due to non-stationary, volatile nature.

Previously, it is assumed that the trend in the stock market depends on historical financial data. But it is studied that along with financial time series data many other factors determine the trend of the future stock price. There are more factors such as technical indicators of stock trading, the macroeconomics variables, and more important the human sentiments. The news headlines, social media tweets, company news, and other sentiment factors largely affect the stock market prediction scheme.

In the past time, the research on financial time series forecasting has been based on the machine learning models, such as Artificial Neural Network (ANN) and Support Vector Regression (SVR), to forecast the stock price and gain high popularity. Forecasting the financial time series has been the active research area from the long time. Only the trader’s market experience and intuition is insufficient therefore an efficient, scientific and effective method is needed to direct stock trading movement. Therefore not only considering the financial historical data, but other factors such as human sentiment must be considered to effectively predict stock future prices. The main concern of this work is to precisely predict future stock price considering data denoising using wavelet transform taking the daily trading data, fundamental technical indicators, and of course the human sentiment by making a model for forecasting future stock prices deep neural network due to its good self-learning capabilities .

1.3 Objectives

There are a lot of researches and studies have been done for forecasting financial time series. This work will focus on building the stock price prediction model by using a deep neural network that considers historical financial data, technical indicators, and financial news and posts as a large volume of the training dataset to yield less prediction error.

The main objective of this research work can be illustrated as:

a) To perform the data preprocessing using wavelet transform and develops the LSTM-attention mechanism for stock price prediction.

b) To evaluate the prediction performance by adding sentiment polarity score and comparing other deep learning models.

1.4 Thesis Contribution

With this approach, the work will be concerned on verifying that the use of attention mechanism in deep learning architectures, wavelet transform as denoising technique, and including human sentiment can lead to a significant improvement on the results of the previous works based on statistical methods.

1.5 Scope of Work

Some of the fields where the carried out work can be fruitful are:

a) Share market prediction: Share market is considered as one of the trending business these days and everybody want to involve in it. As we know there is a market risk in the share market so one should pay more attention to the future trends of the stock market by predicting in advance. The main aim of this work is to efficiently predict the future stock prices so that the investors can take benefits.

b) Business and marketing: Text mining and sentiment analysis used in this work are very helpful for the business to cope with the customer's advice. The positive or negative comments for the products help to improve the quality of stuff and overall growth of the business.

CHAPTER 2
LITERATURE REVIEW

2.1 Related Work

Forecasting the financial time series is the job of obtaining the future prices of the stock listed on the stock exchange [2].

Lin et al. [3] proposed a stock market system based on SVM, and the research based on the Taiwan datasets that choose a suitable attributes, controlling over-fitting, and judging stock indicators. They resulted in better accuracy than the traditional forecast system. The main problem while using the SVM is that the number of input attributes, especially when the features ranges in hundreds, model takes huge amount of computation and memory to train.

Wei Bao, et al. [4] presented a stock market prediction method where the stacked auto encoders, LSTM, and wavelet transform based data denoising scheme are collectively used. The wavelet transform was used for reducing the noise from the data and then for the generation of the deep high-level feature, SAEs are used and finally the long-short term memory (LSTM) predict the next day closing price. This model claimed that they outperform other similar models with respect to both predictive accuracy and profitability performance.

Jiayu Qiu, et al. [5], did a research based on wavelet transform. They used LSTM and an attention mechanism to find and train the features and establish the forecasting model.

Jin et al. [6] added the human sentiment in the model analysis. They used empirical model decomposition (EMD), along with attention based LSTM for predicting the future stock price.

Yahya, et al. [9] introduced the forecasting model in the Indonesian Stock Market considering the importance of the human sentiment form different forums and social media websites. They mainly used naïve Bayes, support vector machines, and random forest algorithms to classify tweets about companies and compared the results of the different algorithms.

Fisher did the financial price prediction task on S&P500 data from December 1992 until October 2015. They used LSTM as the recurrent neural network and found that LSTM model outperformed other traditional machine learning approach [10].

Liu introduced the end to end attention mechanism based forecasting model which include the effects of event counts, short term, medium term and long term influence as well as the movement of stock prices[11]

Li et al.[14] used both stock indicators and investor sentiment in CSI300 index to predict the stock prices. They used LSTM model with four layers and 30 nodes. The naïve Bayes classifier was used to get the investor sentiments. The model accuracy is greater than the support vector machines (SVR).

Saud and Shakya [15] did a research on the NEPSE by comparing the performance of financial time series forecasting from three different deep learning models: Vanilla RNN, LSTM and GRU. Their research work found that the GRU outperformed the other models. They did not consider the sentiment of the investors and news.

Shahi et al. [16] used GRU and LSTM model to forecasting the stock price of NEPSE as an deep learning models. Their study was mainly concerned on the comparison of the GRU and LSTM model with and without including the human sentiments. This research found that adding the news sentiment enhanced the prediction accuracy of the model. The drawback of this model is that they didn't consider the data denoising as the stock data are highly volatile and noisy and the second one, they used VADER for finding the sentiment polarity, instead some other machine learning approach can be used so that the sentiment can be more polished.

2.2 Limitations on existing models

Most of the research work was done by taking only the historical price data, and there is not much effort on considering the investor sentiment which can be the crucial features for accurately predicting the future stock prices.

News related to politics, news about the positive and negative aspects of companies, media and shareholders’ post on social media has a great influence in the rise and fall of the future stock price. Therefore extracting news about the company and related posts yield better results in predicting the stock prices. Textual analysis based forecasting [12], sentiment polarity analysis using a cohesion-based approach [13] are some of the major work.

So considering both the historical stock data and textual information related to indices help in the correct prediction of future stock prices.

CHAPTER 3
METHODOLOGY

As we realize the stock price relies upon on lots of factors such as historical time series records, essential technical indicators, macroeconomics variables, and off course direction the sentiment of the buyers. Therefore in this research, we can perform both technical analysis and sentiment analysis on the stocks and information approximately the agencies and tweets/conversations made via the traders. Technical evaluation is executed by treating the stock traits as one dimensional time series and attempts to forecast the future stock charges by means of watching the prices of past historic records and the essential technical signs. In sentiment analysis, we can use the agency's news and tweets of the investors and other relevance sentiment data and study their influences on inventory expenses by using calculating every day sentiment polarity. Eventually, the polarity acquired from the sentiment analysis is mixed with the other characteristic used to carry out the technical evaluation to see the effect of the human sentiment on the motion of future stock price.

3.1 Technical Analysis

Forecasting the financial market is one of the practical problems in the economic field. The noisy and the volatility are the two characteristic that hinders the timely prediction of the stock future price. For numerical analysis of stock data a neural network based stock market prediction model is considered. This model has three parts namely: wavelet transform to denoise historical stock data, LSTM, due to its unique memory structure avoids long-term dependence and helps to predict financial time series, and finally attention mechanism to efficiently extract specific information.

The workflow for the technical analysis of the stock price prediction model is shown in figure 3.1. This numerical analysis is to build a recurrent neural network (RNN) which are very efficient at learning and predicting time series data.

Dataset preparation is one of the major tasks in any machine learning algorithm. After the dataset preparation, data cleaning is carried out, and after that features are extracted. The extracted features are then passed into the neural network for model building.

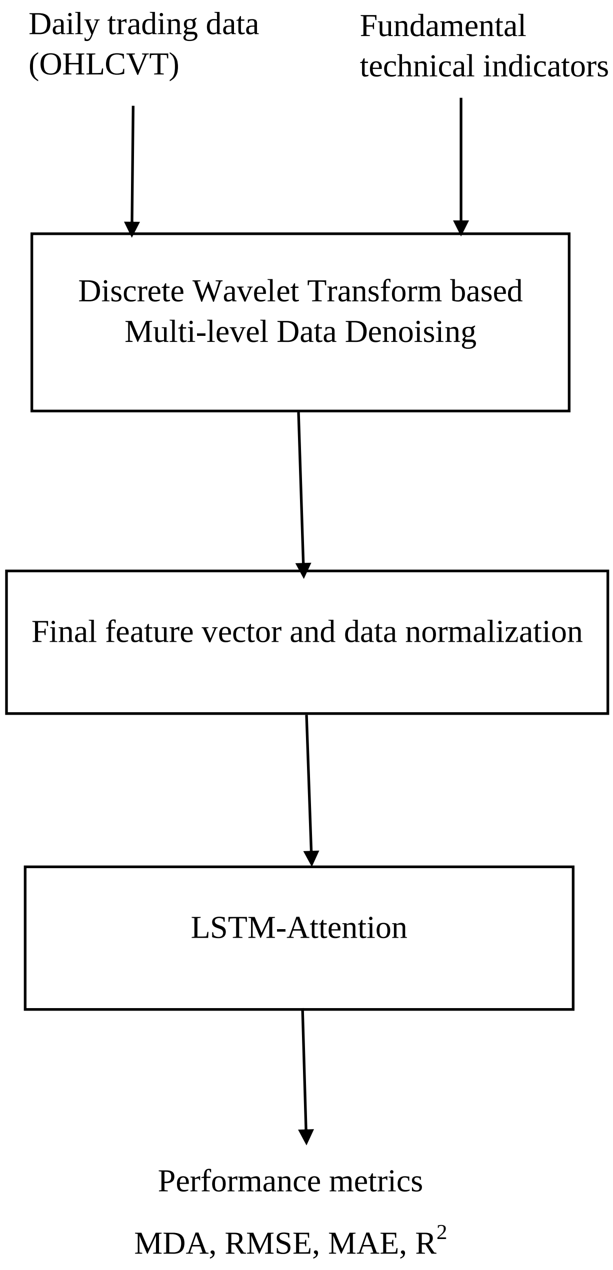

Figure 3.1: Attention-based LSTM model flow diagram for numerical analysis

3.1.1 Data Preparation

The historical stock prices of public companies listed in the Nepal Stock Exchange (NEPSE) are taken as the experimental data. There are six key features in the basic transaction dataset that includes daily open price, daily high price, daily low price, daily close price, and daily trading volume, and trading turnover (OHLCVT). From these basic daily trading data, fundamental technical indicators are calculated as additional input features and then added to the dataset. Along with these basic daily transaction data, seven more technical indicators are added in the dataset. The technical indicators used in this research work is shown in table 3.1

Table 3.1 Some of the technical indicators

Full Name/Symbol	**Description**
Daily Return	The percentage change in daily closing price
High-Low	Difference of daily high price to daily low price
Close-Open	The difference of daily closing price to daily opening price
Weighted Moving Average (WMA)	The moving average where the most recent prices are given greater weight and prior price are given less weight
Price Rate of Change(ROC)	Measure the percent change between the most recent price and price in the past and is used to identify the trends
Triple Exponential Moving Average (TEMA)	Smooths price fluctations and making it easier to identify the trends.
On-Balance Volume(OBV)	Momentum indicator that uses volume flow to predict the change in stock price.

3.1.2 Preprocessing and Normalization

Data Preprocessing Using Discrete Wavelet Transform

Forecasting financial time series is one of the most challenging problems amongst time series prediction because of its noisy and non-stationary and volatile traits. Wavelet transform is the transform that provides the time-frequency representations. Two major concepts behind the wavelet transform are scaling (shrinking and expanding of wavelet) and shifting. Some common mother wavelets are Morlet, Daubechies, Coiflets, symlets, etc. Wavelet transforms can be classified into three classes: continuous wavelet transforms (CWT), discrete Wavelet transforms (DWT) and multi-resolution-based discrete wavelet transform. The level 4 decomposition diagram of DWT is illustrated in figure 3.2. The cA_n are the approximation coefficients and cD_n are the detail coefficients. The large amplitude coefficients are produced by the signal and small coefficients represent the noise. The wavelet threshold method can be divided into three steps:

1) Find the wavelet coefficients by choosing the suitable wavelet basis and decomposition scale,
2) The estimated value of the wavelet coefficient can be obtained by selecting the proper threshold and thresholding function,
3) Using inverse Wavelet transform the signal is reconstructed based on estimated values.

The major issue in the wavelet based data denoising is to find the optimal threshold. Special attention is given to choose the threshold, otherwise the noise level still exist if threshold is too small, and the signal will be discarded if the threshold is too high.

In this research work, the universal thresholding method is used because of its simplicity and effectiveness. The formula for the universal threshold is expressed as follows:

$$\lambda = \sigma\sqrt{2\ln(N)},$$

Where σ is the variance of the noise signal and N is the length of the time-series data.

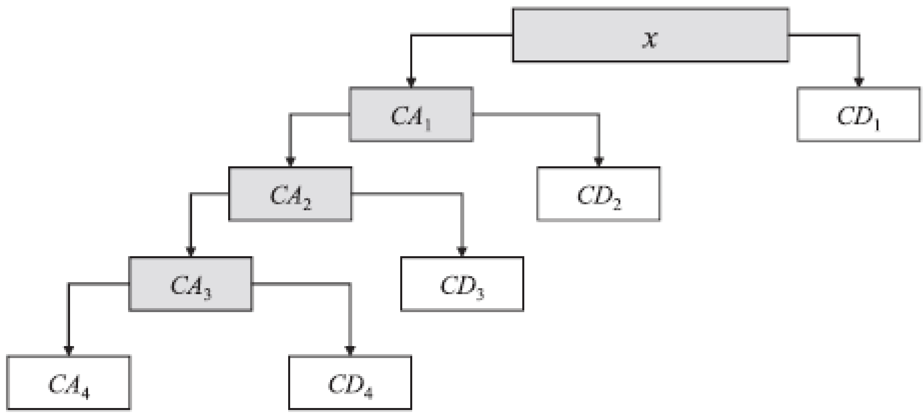

Figure 3.2: Four level decomposition diagram of DWT

The overall wavelet transform-based denoising the 1-D signal is illustrated in figure 3.3. Firstly the noisy signal is provided to the discrete wavelet transform that decomposes the signal into approximation coefficients and detail coefficients, this process is called signal analysis. Then, the coefficients are passed to the thresholding where the noise level is estimated and the threshold is set. Finally, the inverse wavelet transform is used to reconstruct the noise-free signal.

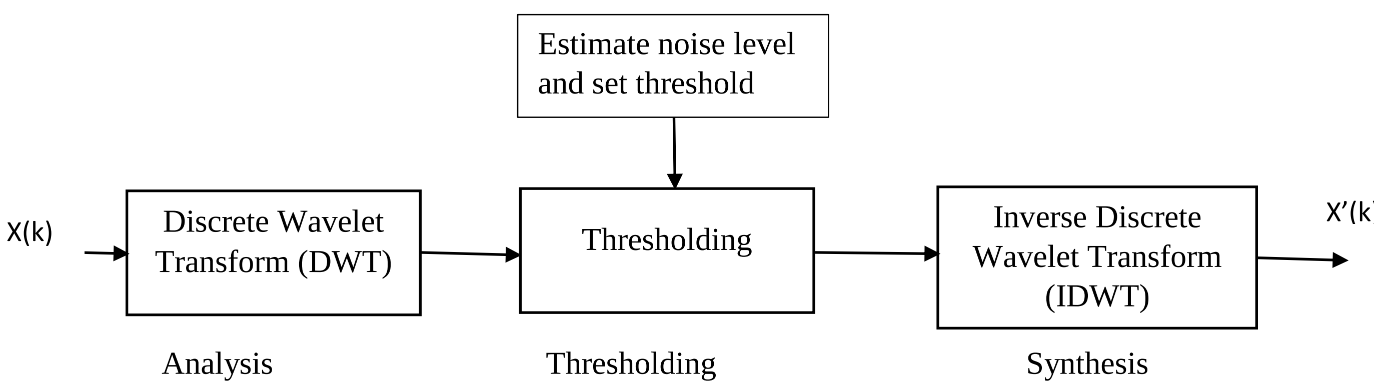

Figure 3.3: Wavelet transform-based denoising

The wavelet transform perform both time domain and frequency domain analysis and therefore is considered more useful for extremely irregular financial sequences.

In this work, multi-resolution based Wavelet transform is used to extract the noise free signal. Denoising the financial time series data using Denoising the financial time series data using

Wavelet Transform mainly consists of wavelet decomposition, threshold processing, and reconstruction of signals. The RMSE and SNR are normally used to evaluate the effect of the wavelet transform for denoising. The higher the SNR and lower the RMSE, the better the denoising effect of the wavelet transform. The signal to noise ratio is the ratio of signal power to noise power, expressed in decibel (dB). Higher the value of SNR and lower the value of RMSE, the more accurate is the denoising. The signal-to-noise ratio (SNR) and root mean square error can be obtained by using the following formula.

$$SNR = 10\log[\frac{\sum_{j=1}^{N} x_j^{\,2}}{\sum_{j=1}^{N}(x_j - x_j)^2}]$$

$$RMSE = \sqrt{\sum_{j=1}^{N}\frac{(x_j - x_j)^2}{N}}\ ,$$

Where x_j and x_j are the denoised and original signal respectively, and N is the length of the signal.

Data Normalization

Normalization is a technique that is often used as part of data preparation for machine learning. The main of normalization is to transform features to be on a same scale. This help to easily converse the performance and training stability of the model. The need of normalization is when the features have different ranges. There are several ways for normalization; three most popular techniques are described below.

Min-max normalization

It is also known as rescaling and can be calculated as

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

i) Mean Normalization

This method uses the mean of observations in the transformation process and can be calculated as

$$x^{'} = \frac{x - average(x)}{\max(x) - \min(x)}$$

ii) Z-score Normalization

Z-score Normalization is also called standardization and uses Z-score or standard score. This method is generally used in machine learning algorithms such as Support Vector Machine (SVM) and logistic regression. It can be calculated using the following formula,

$$z = \frac{x - \mu}{\sigma}$$

In this research work, generally the Volume and turnover attribute have a very high range or there is high deviation in these features. Min-Max normlization method is used to range all the features in the range of 0 to 1.

3.1.3 Model Establishment and Training

Gated Recurrent Unit (GRU)

In 2014 Chao et.all introduced the type of RNN for the purpose of solving the vanishing gradient problem, called Gated Recurrent Unit (GRU). This simplified version of LSTM uses two gates which are update and reset gate. The main advantages of using GRU are that it has fewer parameters and trains faster or need fewer data to generalize.

The update gate and the reset gate results the intermediate values z_t and r_t respectively and the output h_t is stored in the final memory of GRU [18].The objective of the update date is to decide how much input x_t and the previous output h_{t-1} to be passed to the next cell, which is controlled by the weight w_z. The use of the reset gate is to determine how much past information to forgot. The following are the main equations representing the functionality of the GRU.

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$

$$r_t = \sigma(W_z \cdot [h_{t-1}, x_t])$$

$$\widetilde{h_t} = \tanh(W[r_t * h_{t-1}, x_t])$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * h_t$$

The structure of GRU is shown in figure 3.4.

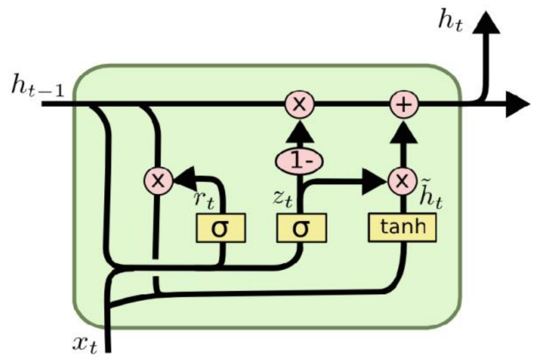

Figure 3.4: Internal structure of GRU

Long-Short Term Memory (LSTM)

Long-Short Term Memory which is common forms of RNN, meant to avoid vanishing gradient problems and used in processing and predicting time series. In comparison to GRU, LSTM has three gates input, forgotten, and output gates. Each memory cell has three sigmoid layers and one *tanh* layer. Figure 3.5 shows the structure of LSTM cells.

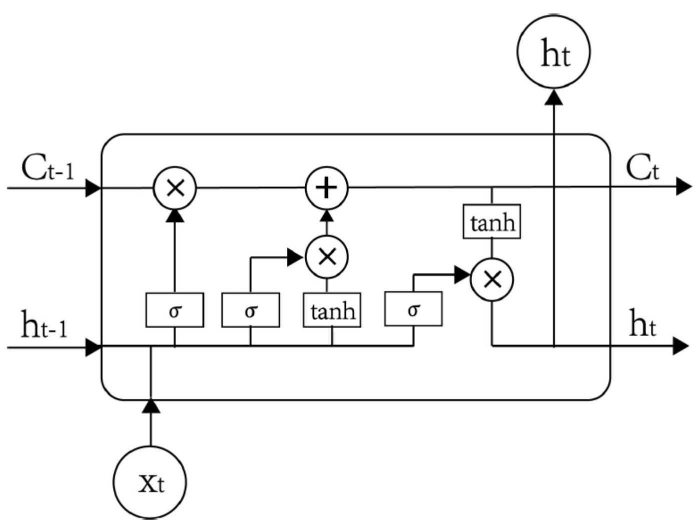

Figure 3.5: Structure of long-short term memory

The forgotten gate of the LSTM unit is used to determine which cell state information is omitted from the model. The memory cell accepts the previous moment h_{t-1} and the current input information x_t and combines them in a long vector $[h_{t-1}, x_t]$ through σ transformation to become

$$F_t = \sigma \left(W_f . [h_{t-1}, x_t] \right) + b_f$$

The input gate is used to find out how much of the current inputs x_t is reserved into the cell C_t, which avoids insignificant content from entering the memory cells

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

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

$$C_t = f_t * C_{t-1} + i_t * C_t$$

The output gate is used to control how much of the current cell state is omitted. Firstly,the sigmoid layer determine the output information, and then the cell state is processed by *tanh* and finally, multiplied by the output of the sigmoid layer to obtain the final output portion,

$$O_t = \sigma\left(W_\sigma . \left[h_{t-1}, x_t\right] + b_0\right)$$

The final output value of the cell is defined as:

$$h_t = O_t * tanh\left(C_t\right)$$

Attention Mechanism

Normally, the attention mechanism is very useful in speech recognition, machine translation and part of speech tagging. Recently the attention mechanism gain very high popularity for the time series data as well. Attention mechanism can be self attention, global attention or local attention. The hard attention concern on one element in the input selecting in formationn based on mximum or random sampling that need more training for good results. Whereas soft attention mechanism enables more efficient use of informaion by assigning weights to all information. The soft attention mechanism can be formulated as,

$$e_t = \tanh(w_a[x_1, x_2, ..., x_T] + b)$$

$$a_t = \frac{\exp(e_t)}{\sum_{k=1}^{T} \exp(e_k)}$$

The attention mechanism has mainly two steps: the first one is to calculate the attention distribution and second one is the calculation of the weighted average of the input informaion according to attention distribution. Firstly, the input information, $x=[x_1,x_2, \ldots, x_T]$ is first passed to attenion scoring function S, and then passed the result to the softmax layer to obtain the attention weights $[\alpha 1, \alpha 2, \ldots, \alpha N]$. Finally the attention weight vector is weighted and averaged with the input information to obtain the final result. The attention mechanism is shown in figure 3.6.

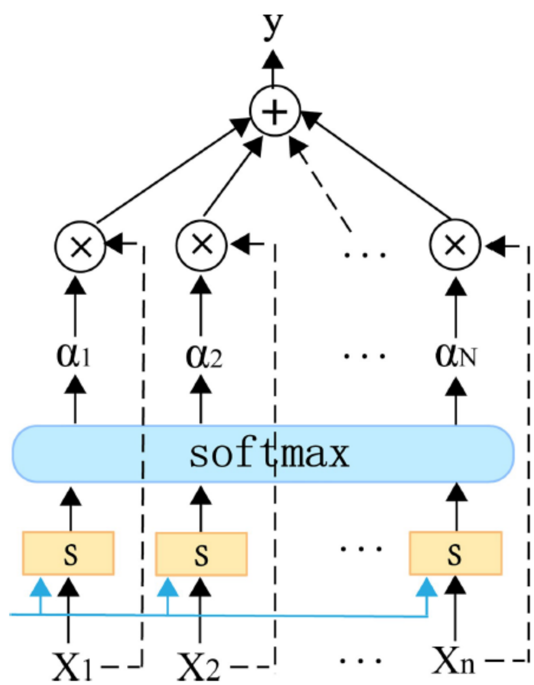

Figure 3.6: The basic structure of the attention model

3.1.4 Evaluation of Experimental Results

The performance of the model will be assessed by using MAE, RMSE, mean directional accuracy (MDA) and R^2. The smaller the value of MAE, RMSE the closer the predicted value to the actual value. The closer the coefficient of determination R^2 to 1 better the fit of the model.

3.2 Sentiment Analysis

In this research work, sentiment analysis is used to find out the sentiment polarity of the investors posts, news and other articles to calculate the daily sentiment polarity so that it can become one of the prominent feature for stock price forecasting.

The very first step to any experiment is to collect the data and preprocessed them. The data for sentiment analysis are collected from the various stock related websites, and forums. The collected data have a lot of tags, stop words and regular expressions that do not give any sentiment. The data are collected in the daily basis and treated as combined news

In order to calculate the daily polarity sentiment, in this research work Valence Aware Dictionary and Sentiment Reasoner (VADER) is used which is a lexicon and sentiment intensity analyzer specially designed to obtain sentiments expressed in social media. The figure 3.7 shows the block diagram of the sentiment analysis from which the sentiment polarity is obtained.

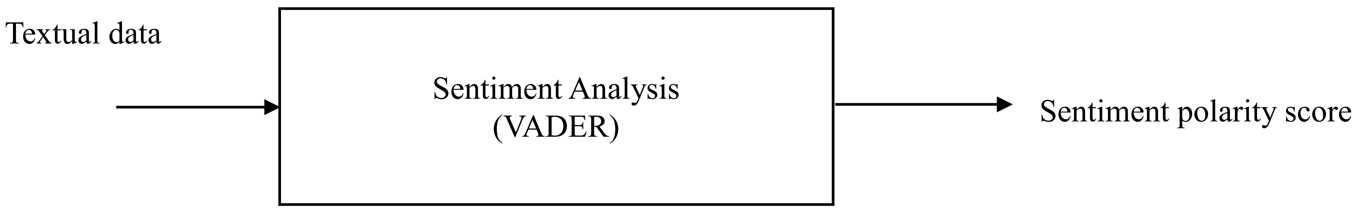

Figure 3.7: Obtaining the sentiment polarity using VADER

VADER actually provide the sentiment polarity score from -1 to 1. The range from 0 to 1 is considered as a positive, the range from -1 to 0 is used as negative sentiment and 0 is considered as neutral. In this research work, VADER is used to obtain the sentiment polarity for each day and the daily sentiment score is used as an additional feature to our dataset.

After obtaining the dataset for the sentiment analysis the model is fit and the performance will be evaluated.

3.3 Merged Model

The sentiment analysis is used to find the user opinion and their tendency towards the specific stock. The daily sentiment score is the new feature for our model. This feature or attribute isadded with the daily trading data and fundamental technical indicators to obtain the final feature vectors.

Now, the final feature vector with daily sentiment polarity is passed to our LSTM-attention model to forecast the future stock prices. The merged model is shown in figure 3.8.

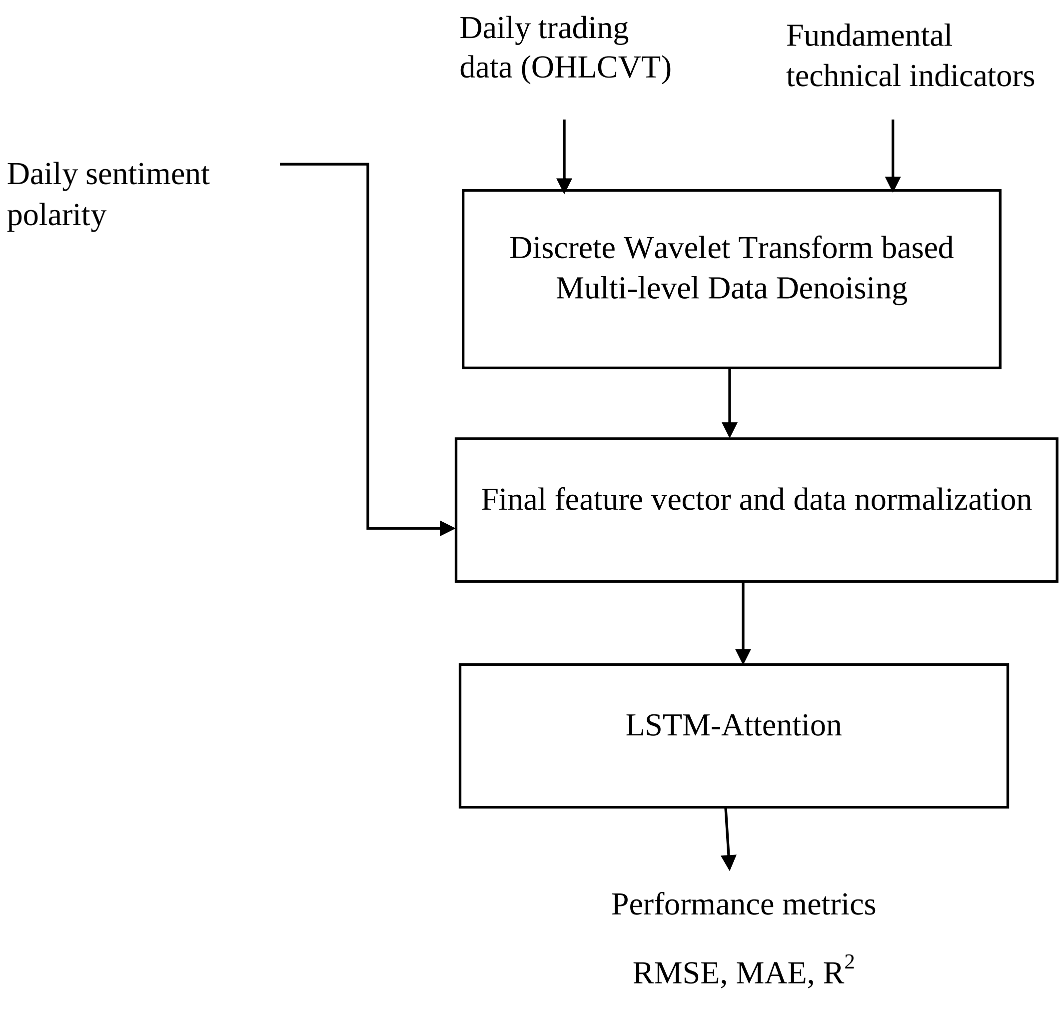

Figure 3.8: Attention-based LSTM workflow for merged model

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Train/test Split and Experimental Setup

The thesis work is carried out on Google Collab.

The stock price data are collected from the websites sharesansar.com. The data are collected in between 2011-03-20 to 2021-01-13, with the total number of trading days 2222 which are our number of samples. The news headlines and text related to stock market are also collected for the same date and obtained the daily polarity score as an additional attribute.

For the training and testing the model, the data is split into train and test as 80% of the total data is used for training the model and rest 20% is used to test the model and 20% of the training data are used for model validation. Figure 4.1 shows the stock data for ADBL closing price as index.

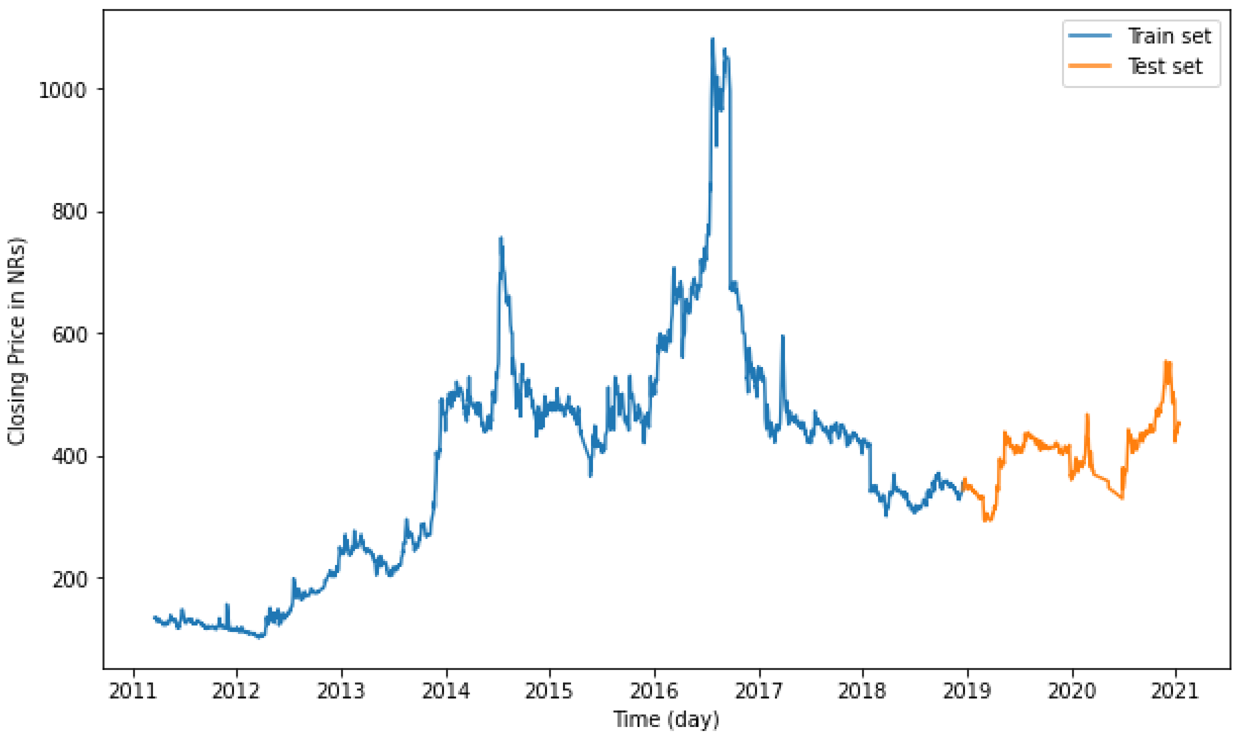

Figure 4.1: Train/test split and closing feature of ADBL

The dataset contain the OHLCV and the polarity sentiment as the features. There are 199 negative, 273 neutral and 1791 positive sentiment data. This shows that the huge volume of data has the daily positive sentiment. The Pearson Correlation matrix among the different features is shown in figure 4.2. From the figure, we can say that there are significant correlation between sentiment score and the closing price.

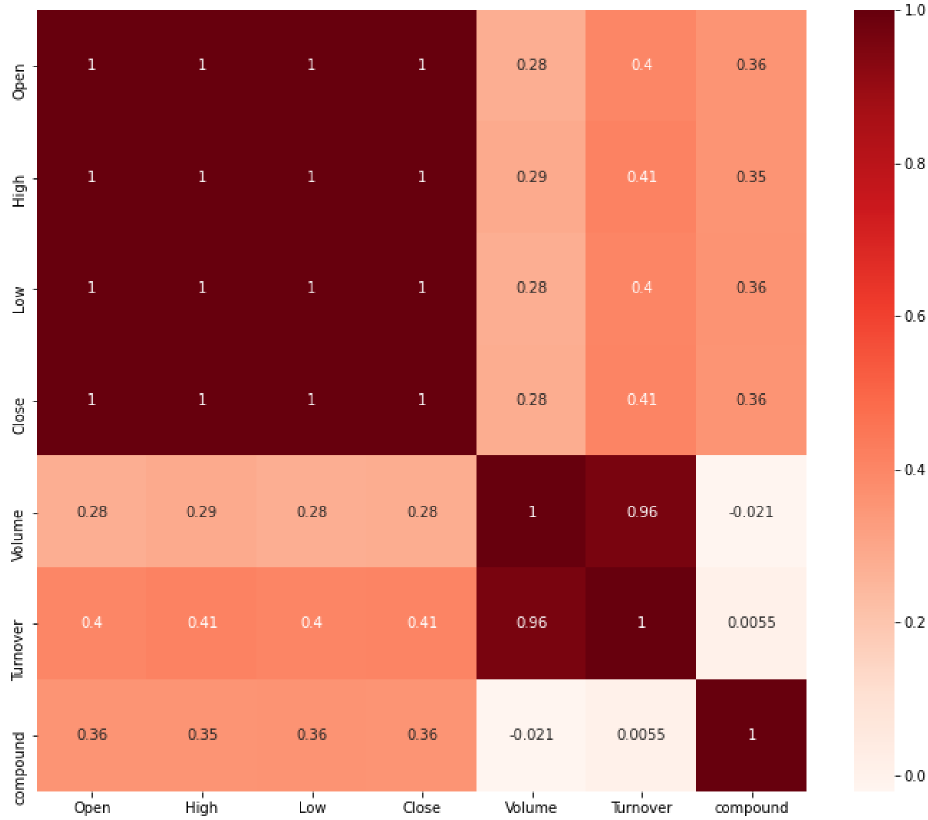

Figure 4.2: Pearson correlation matrix among different features

Hyperparameter Tuning

Hyperparameter tuning is the process of finding the best suite combination of the hyperparameters for the models. The tuned optimization parameters are number of units, batch size, learning rate and dropout.

a) Number of units: From the optimization technique the number of the units of each LSTM and GRU layer is set to 120.
b) Batch size: The batch size is set to be 32 for tuning the model
c) Learning rate: The learning rate for an Adam optimizer is set to 0.01
d) Dropout layer: The model performs well in training and poor in testing or validation phase. This is the very serious problem in deep learning models since they need much

more data for training. Dropout mechanism is the simple regularization technique to avoid overfitting for neural networks. It randomly drops the cells of the recurrent neural network. The typical value of the dropout is generally 0.2.

The training of the model may undergo over-fit and under-fit. In order to prevent from the situations due to too many and too few epochs, early stopping criteria is generally employed that allows us to specify large number of training epochs and stop training once the parameter of the model has stopped improving on the validation set.

The complete specifications of parameters for training the model is shown in table 4.1.

Table 4.1: Specification of parameters for training

Parameters	Values
Number of nodes in input layer	Number of input features *look back period
Number of epochs	100 with early stopping criteria of patience of 10 epoch
Batch size	32
Hidden layer	1 LSTM/GRU layer with 120 units
Dropout layer	0.2 dropout rate
Look back period	10
Output layer	1

4.2 Evaluation Metrics

The mean square error (MSE), root mean squared error (RMSE), mean absolute error (MAE), coefficient of determination (R^2) and mean directional accuracy (MDA) are used to analyze the performance of the technical analysis and merged model. Mean directional accuracy is generally used to judge the ability of the model to predict the direction of change than the magnitude of the forecasting error. It also compares the predict direction to actual direction. Formula for the root mean square error, coefficient of determination and mean absolute error and directional accuracy is shown below.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - y_i)^2}$$

$$R^2 = \frac{\frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{y}_i)^2}{\frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{y}_i)^2}$$

$$MAE = \frac{1}{N}\sum_{j=1}^{N} x_j - x_j$$

$$MDA = \frac{1}{N}\sum_{j=1}^{N} 1_{sign(X_j - X_{j-1})}$$

Where x_j and x_j are actual and predicted data and N is the total number of data

4.3 Experimental Results

4.3.1 Comparison of different company's closing price

Figure 4.3 shows the closing price of four different companies for the same time frame. All the four companies have the same trend except in the year 2014 and 2015.

Figure 4.3: Comparison of the closing price for four commercial bank

4.3.2. Discrete Wavelet Transform based financial time series data denoising

The multi-level discrete wavelet transform is used to denoise the financial time series data. Figure 4.4 shows the plot of raw and denoised data of Nepal Investment Bank Limited (NIBL).

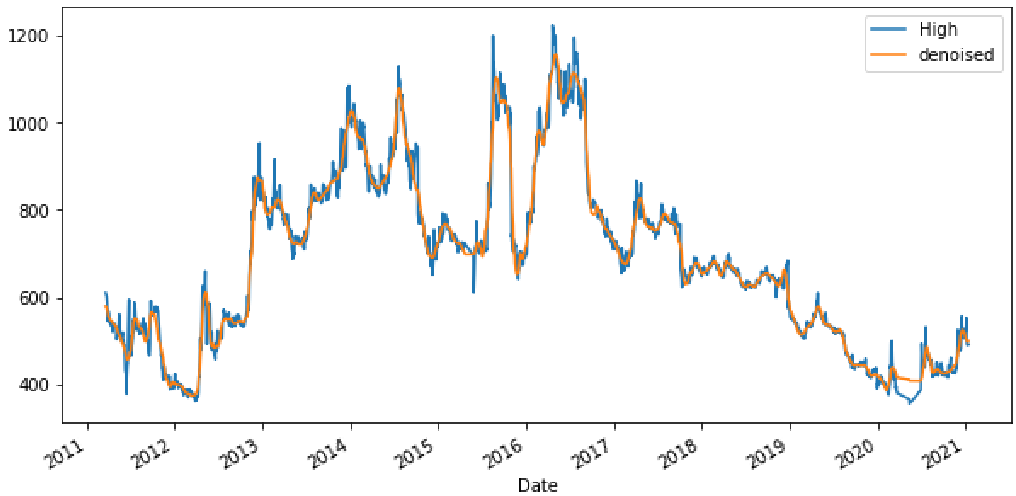

Figure 4.4: Plot of raw and denoised data of Nepal Investment Bank Limited

The performance comparison of different wavelets for four levels of decomposition is shown in table 4.2.

Table 4.2: Performance comparison of different wavelet for four levels of decomposition

Wavelets/performance measure	**Haar**	**dB N (dB3)**	**Symlets (Sym3)**	**Coiflets (Coif3)**
Signal to noise ratio(SNR)	28.0692	**29.7618**	29.761	29.3578
Root mean square error(RMSE)	28.2817	**23.274**	23.2742	24.3822

From the above table, it is concluded that dB3 outperform the other wavelets for four-level of decomposition in both the measure (SNR and RMSE). The signal to noise ratio must be higher and the root mean squared error must be lower for efficient performance. Therefore, the historical data are denoised using this db3.

4.3.3. Experiment of GRU, LSTM and LSTM-AT on historical raw data (OHLCVT) and technical indicators

Using the historical data along with the technical indicators, the performance of different models is tabulated in Table 4.3. From the table it is observed that the attention mechanism employed

LSTM perform better than the other two neural networks. And may be due to small size of the data GRU perform better than LSTM. The comparison is made for the same number of lookback value.

Table 4.3 Performance comparison of different model on ADBL dataset

Performance Metrics/Models	LSTM	GRU	LSTM-AT
RMSE	15.334	13.23	11.734
MAE	12.4817	9.61	7.604
R-squared	0.894	0.91	0.937
MDA	0.577	0.67	0.743

The predicted closing price with the actual closing price for all the models is shown in figure 4.5.

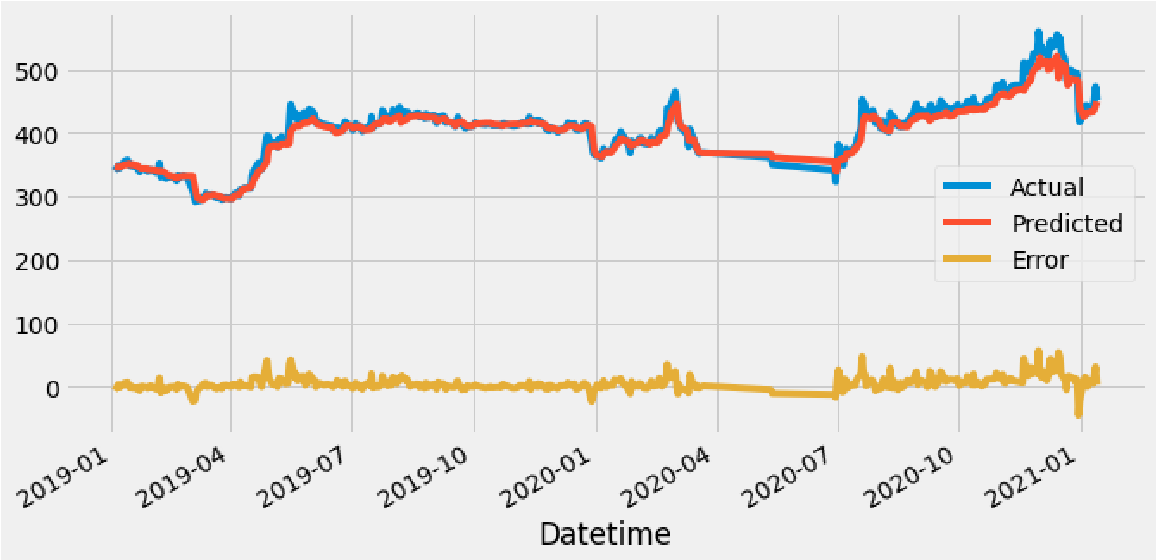

Figure 4.5 (a): Actual versus predicted closing price for the GRU model

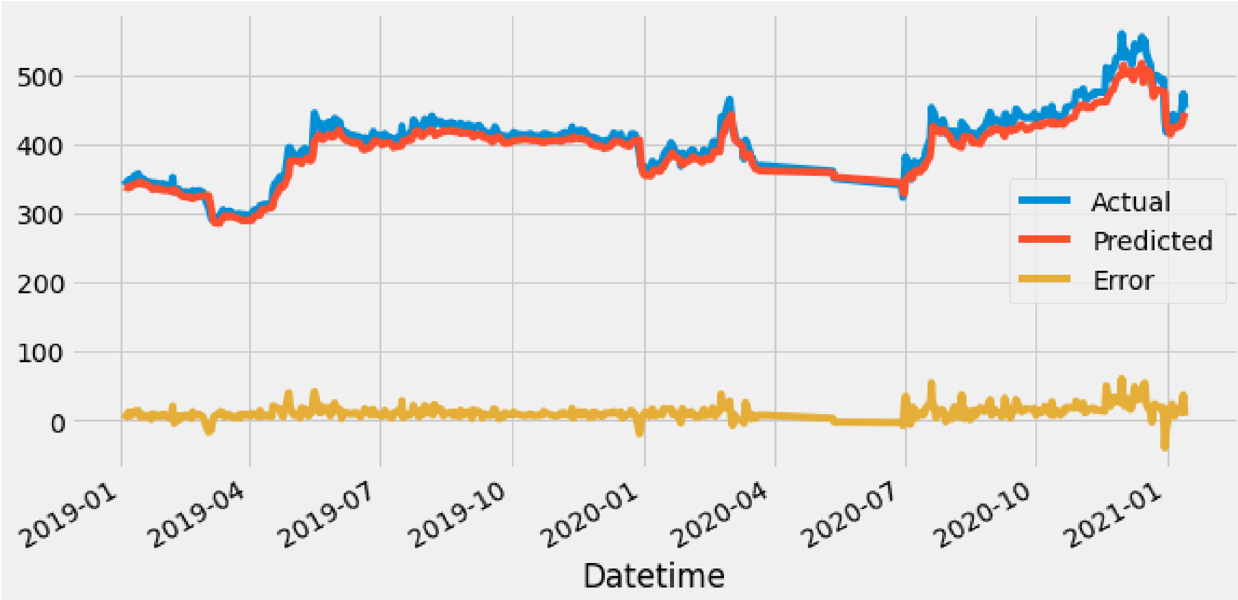

Figure 4.5 (b): Actual versus predicted closing price for the LSTM model

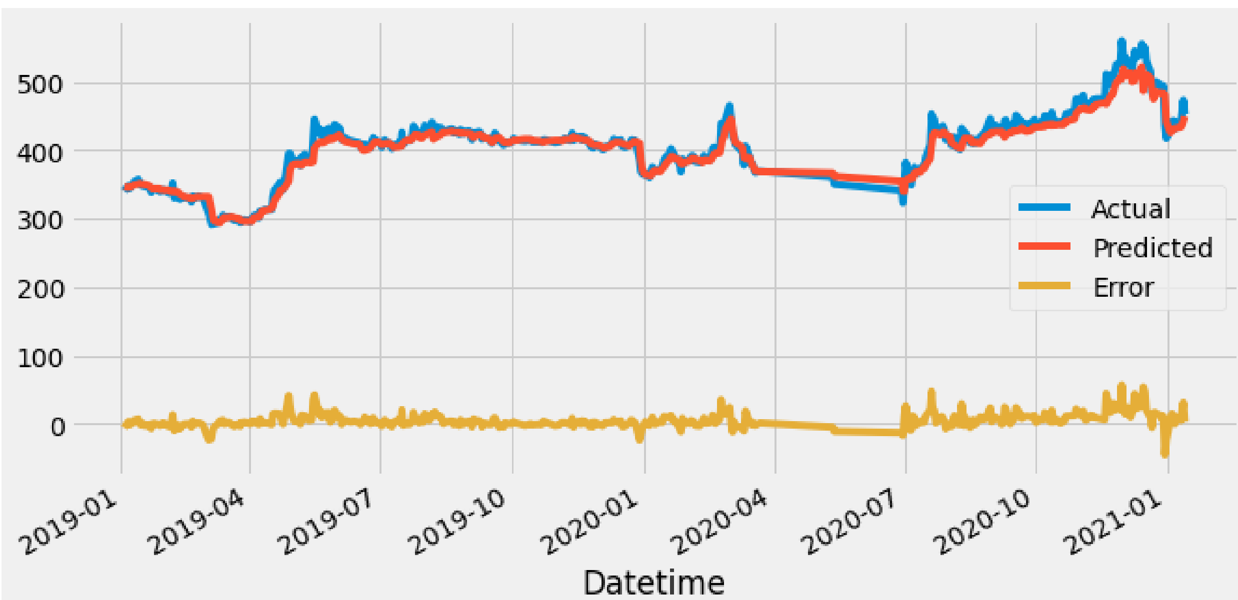

Figure 4.5 (c): Actual versus predicted closing price for the LSTM-AT model

4.3.4 Experiment of GRU, LSTM and LSTM-AT for sentiment features on ADBL dataset

The OHLCVT historical features combined with the sentiment score makes the total dataset which is fed on all the three models and are compared with the performance metrics RMSE, MAE, r-squared and MDA which is shown in Table 4.4.

Table 4.4 Performance comparison of different model for sentiment features on ADBL dataset

Performance Metrics/Models	LSTM	GRU	LSTM-AT
RMSE	9.239	8.236	7.357
MAE	5.9767	5.9757	5.12
R-squared	0.965	0.965	0.96
MDA	0.7042	0.7043	0.75

From above we come to a conclusion that including human sentiment makes the prediction better than just taking the historical data.

Figure 4.6 shows the plot of the actual and predicted and the error for all the three models.

Figure 4.6 (a): comparison plot for LSTM

Figure 4.6 (b): comparison plot for GRU

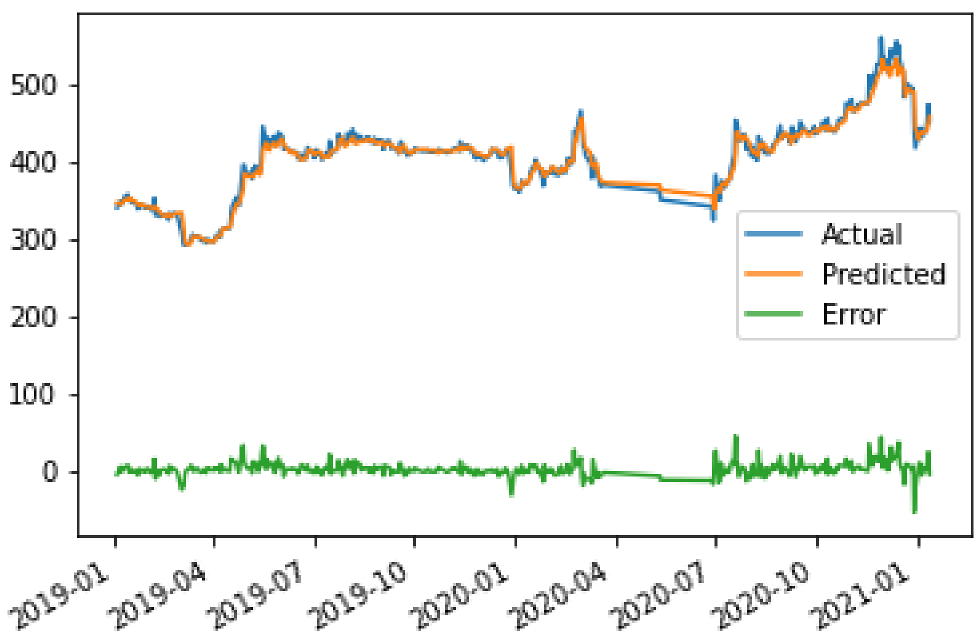

Figure 4.6 (c): comparison plot for LSTM-AT

4.3.5 Experiment on GRU, LSTM and LSTM-AT model with the OHLCVT Denoised Data

For the denoised data the performance comparison for GRU, LSTM and LSTM-AT with the lookback period 10 is shown in table 4.5. It is also noticed that the use of Wavelet Transform as a denosing helps to better predict the future values and reduce the prediction error. In comparison with the the raw data, the denoised data perform very well in all the three model and comparing the models LSTM-AT outperforms others.

Table 4.5 Comparison of GRU, LSTM and LSTM-AT with denoised features

Performance Metrics/Models	LSTM	GRU	LSTM-AT
RMSE	5.783	7.935	3.75
MAE	4.255	5.37	3.54
R-squared	0.986	0.97	0.986
MDA	0.771	0.632	0.78

Figure 4.7 shows the comparison plot for the actual, predicted and the error for all the three models.

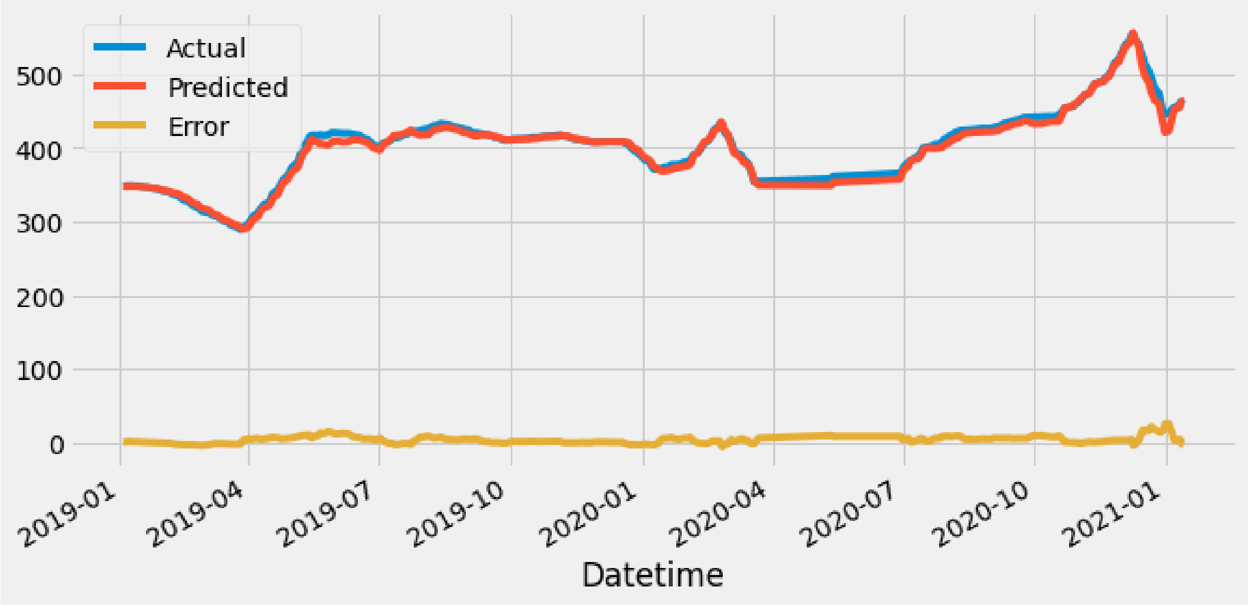

Figure 4.7 (a): Plot for LSTM model

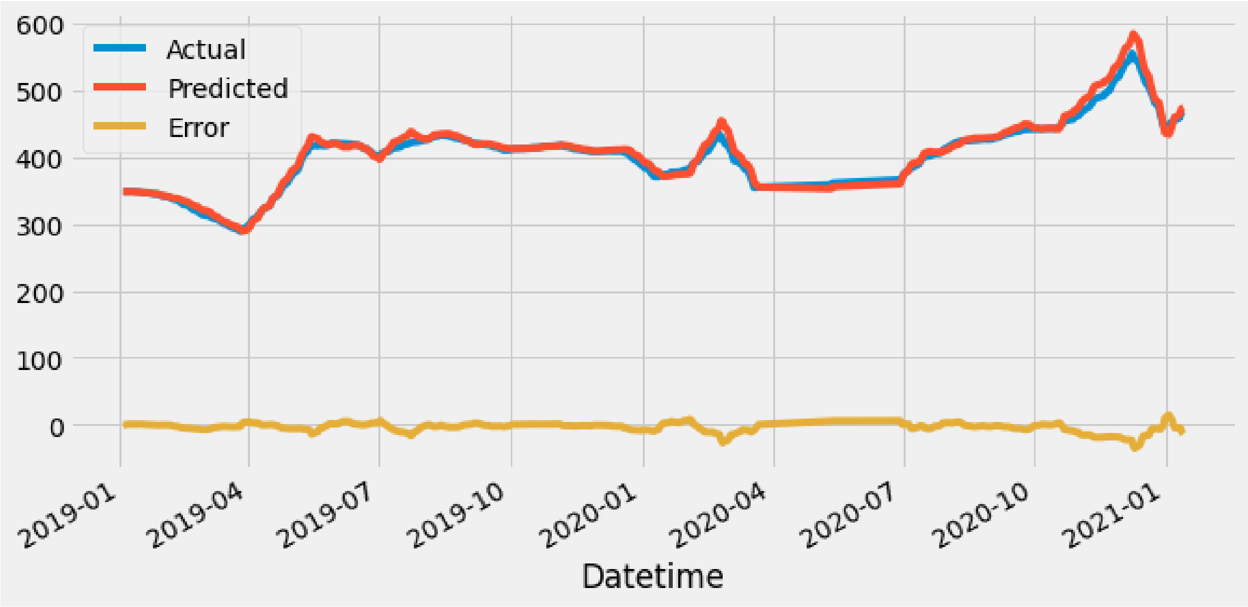

Figure 4.7(b): Plot for the GRU model

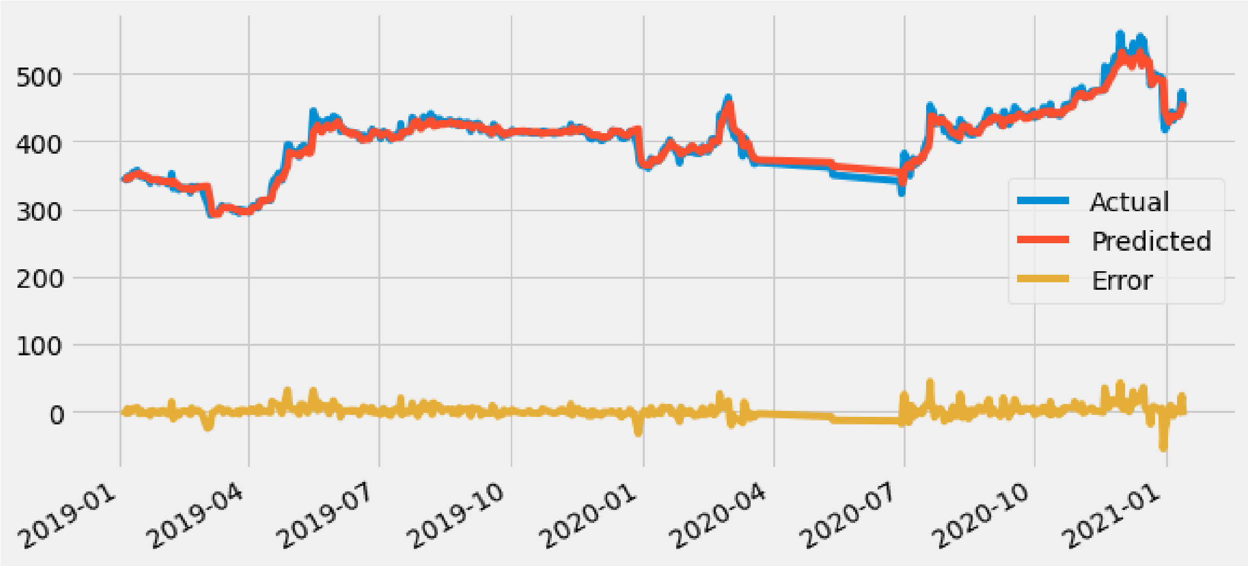

Figure 4.7(c): Plot for LSTM-AT model

4.3.6 Performance comparison of GRU, LSTM and LSTM-AT for denoised features with human sentiment

Table 4.6 shows the comparison between the GRU,LSTM and LSTM-AT for the dataset that is denoised with the Wavelet Transform and the addition of the sentiment.

Table 4.6 Comparison of GRU, LSTM and LSTM-AT with denoised features and sentiment

Performance Metrics/Models	LSTM	GRU	LSTM-AT
RMSE	4.783	5.935	2.35
MAE	3.78	4.37	2.14
R-squared	0.941	0.93	0.956
MDA	0.731	0.662	0.79

Figure 4.8 shows the plot of all the three models showing the actual and corresponding predicted value .

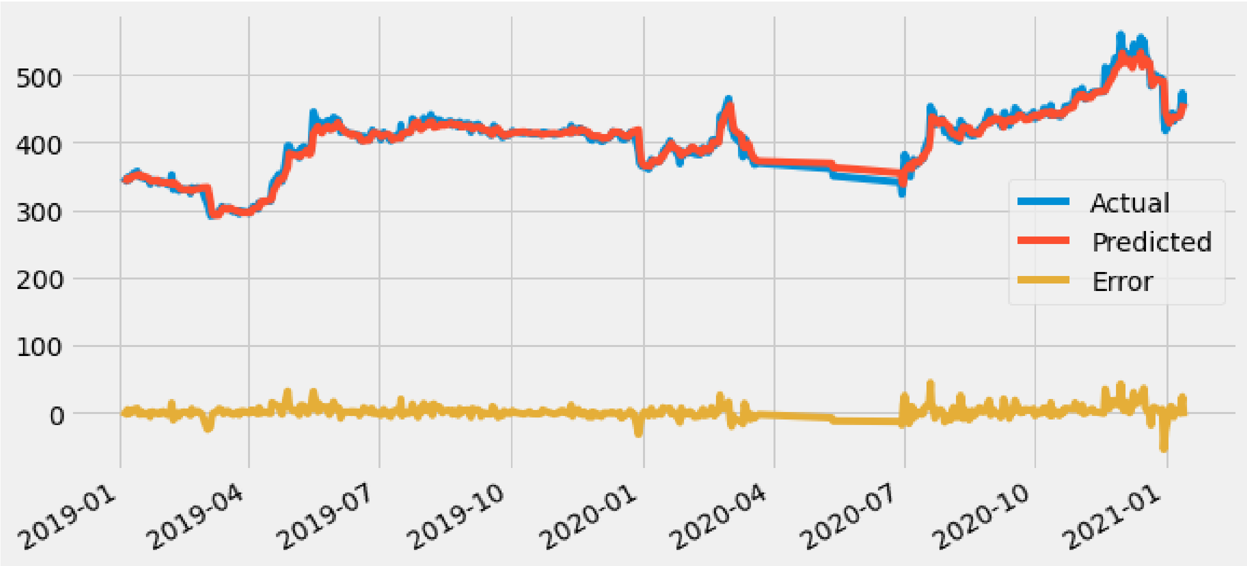

Figure 4.8(a): Plot for GRU

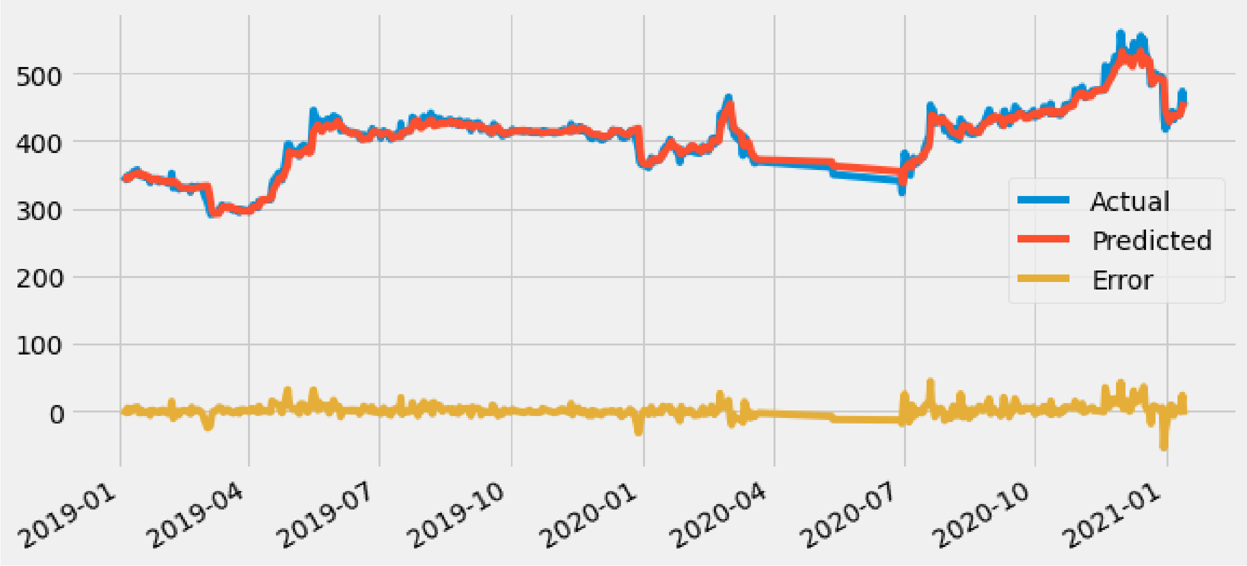

Figure 4.8(b): Plot for LSTM

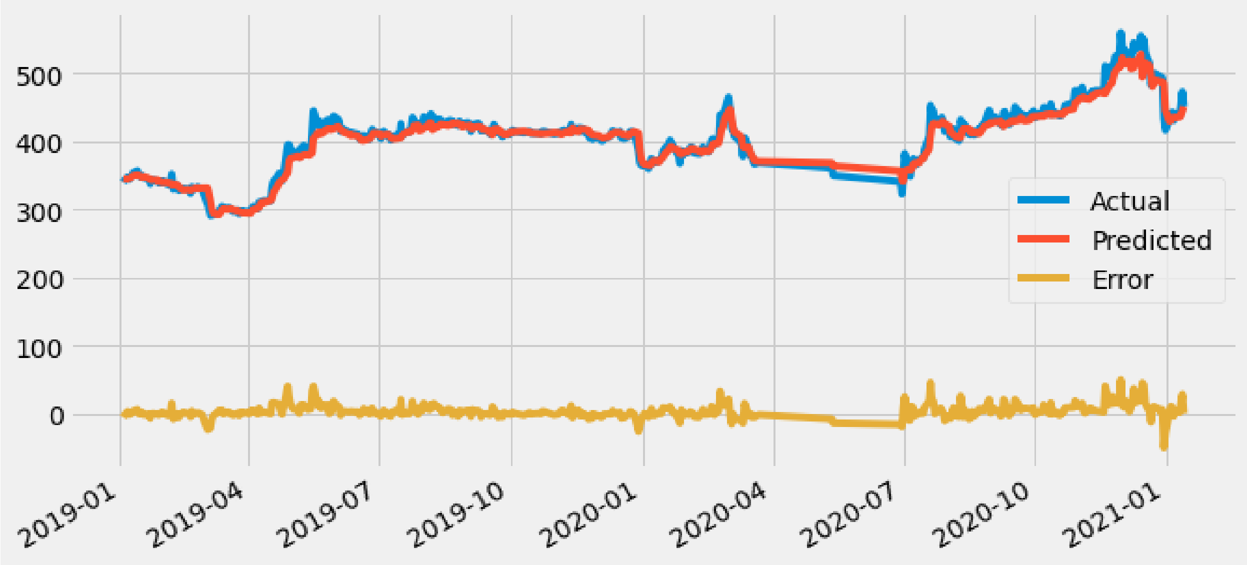

Figure 4.8 (c): Plot for LSTM-AT

The comined plot of all the three model's prediction and the actual value is shown in figure 4.9 for the ADBL dataset.

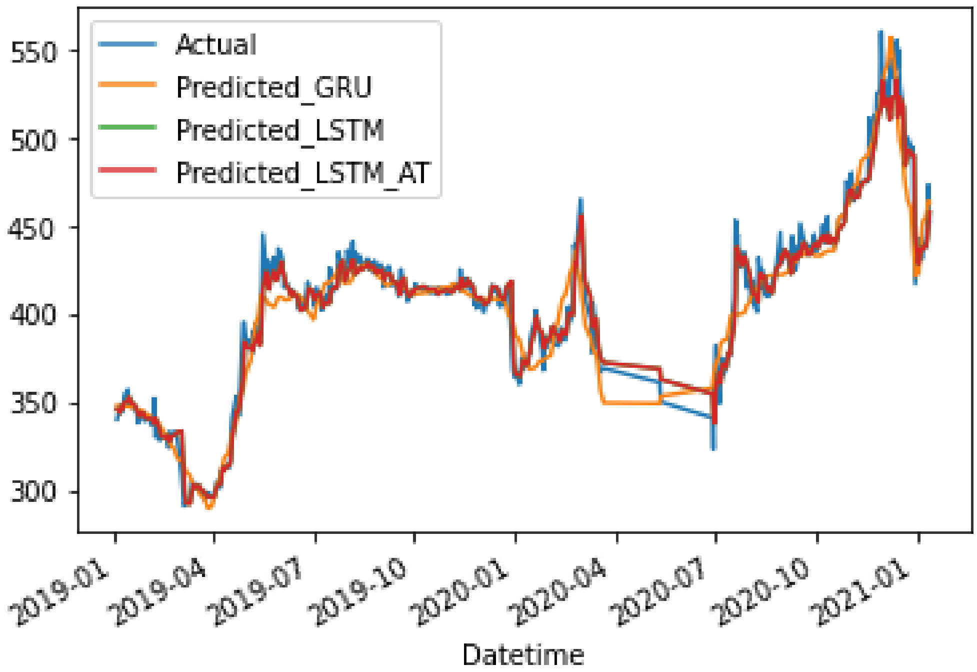

Figure 4.9: Comparison plot of GRU, LSTM and LSTM_AT for ADBL dataset having denoised and sentiment features

4.4 Comparsion of the models on different stocks

For the comparison of the model, four stocks are choosen namely ADBL, NIB, SCB and NABIL. Figure 4.10 shows the comparison plot of NABIL dataset with and without denoising on LSTM and LSTM_AT models.

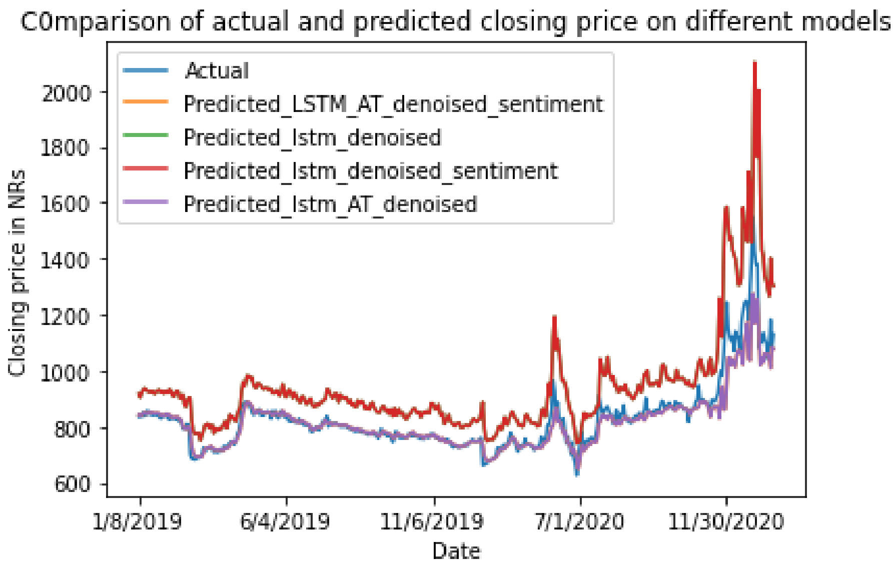

Figure 4.10: Comparison plot of NABIL dataset for LSTM and LSTM_AT models

For the NABIL stock dataset the LSTM_AT model with denoised data prediction price seems very close to the actual data, that is the attention model with denoising as the preprocessing technique outperform the other models.

For NIB and SCB stock the plot for actual versus predicted on different models for rawdata and denoised data are shown in figure 4.11 and figure 4.12.

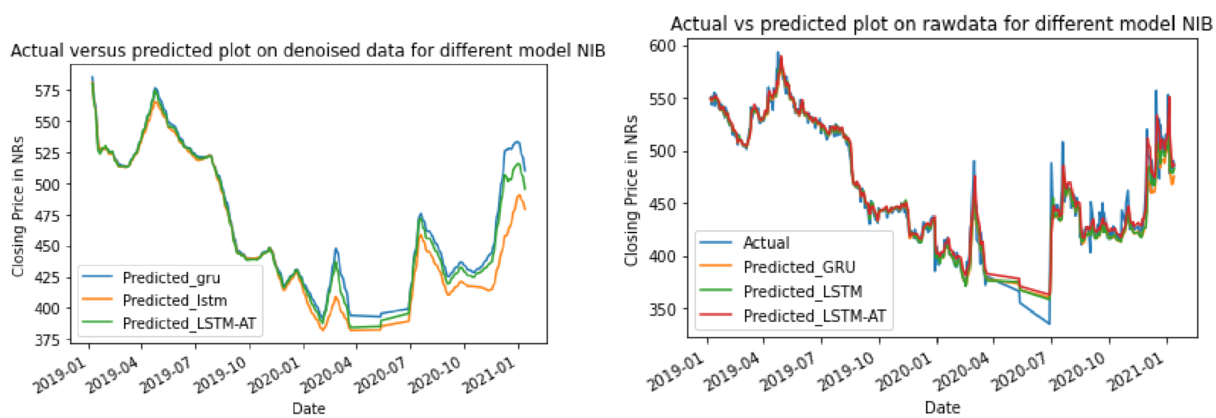

Figure 4.11: Actual versus predicted plot for different models on rawdata and denoised data on NIB stock

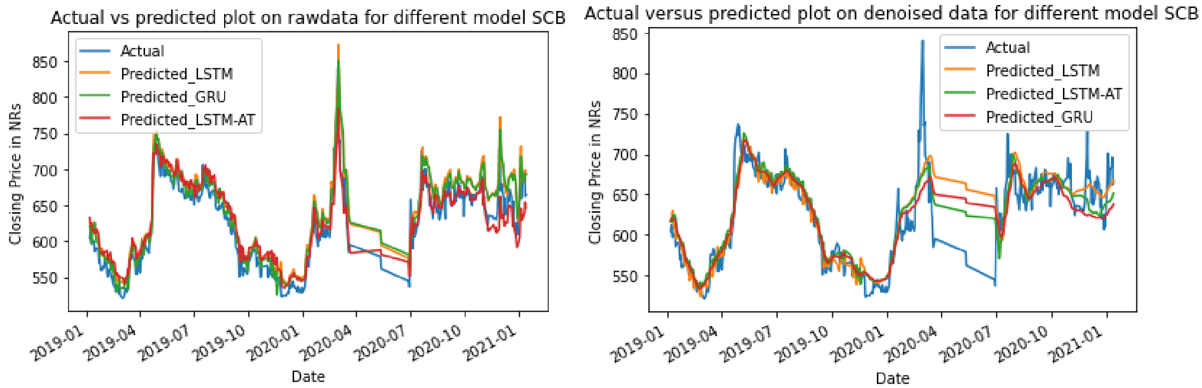

Figure 4.12: Actual versus predicted plot for different models on rawdata and denoised data on SCB stock

4.5 Residual plot for different models for different stocks

The residual plot of different models on both rawdata and denoised data for ADBL stock is shown in figure 4.13.

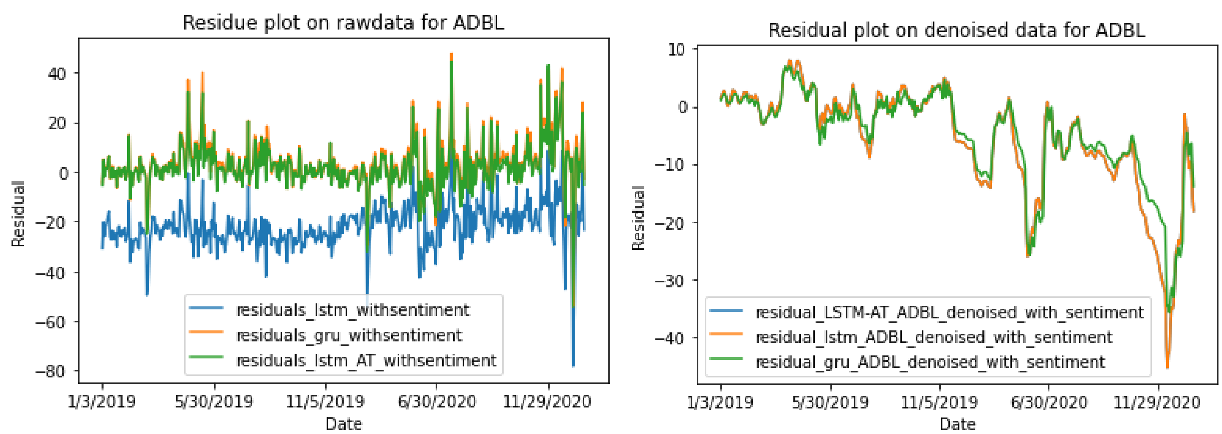

Figure 4.13: Residual plot for both rawdata and denoised data on ADBL stock

For the NIB stock the plot for residual on rawdata and denoised data on different model is hown in figure 4.14.

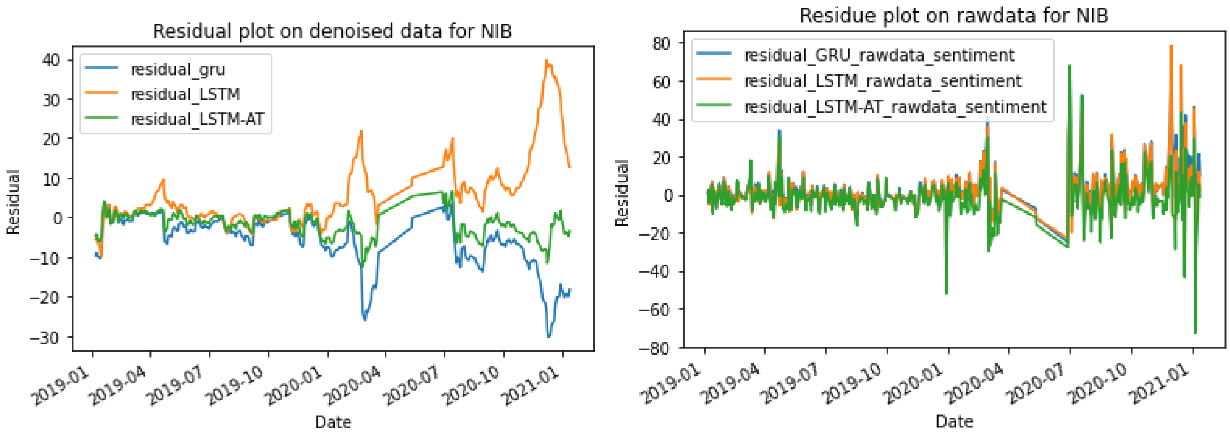

Figure 4.14:Residual plot on NIB stock for different model for both raw data and denoised data

Again for the SCB stock the residual plot is shown in figure 4.15.

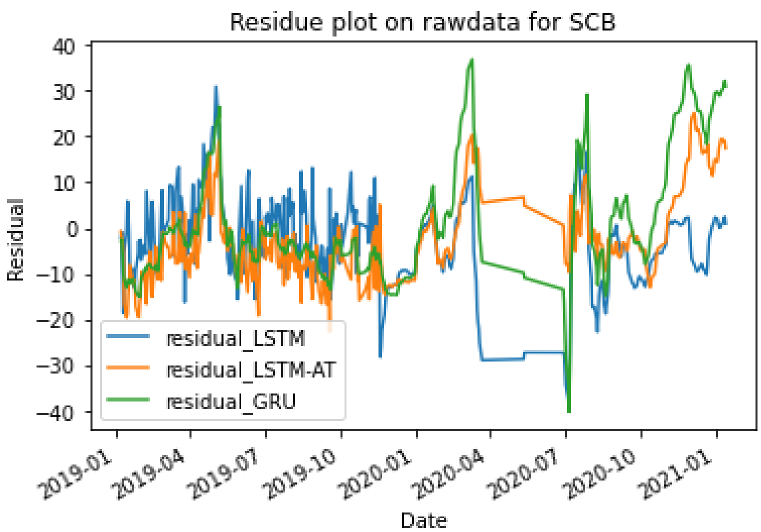

Figure 4.15: Residual plot for different model on rawdata for DCB dataset

From the above residual plot, it can be analysed that model perform well on ADBL and NIB dataset and show weaker performance on SCB dataset.

CHAPTER 5
CONCLUSION AND RECOMMENDATION

Stock price prediction is the way of forecasting the future stock price. The share market is very volatile and shows noisy characteristics and is non-stationary in nature. Because of the noisy behavior it is very crucial to denoise the data. Here in this research work the multi-resolution Wavelet Transform using db3 based denoising technique is used as the data preprocessing, and comparing different basis wavelets Db3 outperforms the other mother wavelets. Therefore, Db3 is used as the basis function for wavelet based denoising. The dataset contains the historical daily transaction data OHLCVT, some technical indicators and the daily sentiment polarity score. The attention mechanism based LSTM is used as the deep neural network to predict the next day closing price of the stock. The denoised dataset has RMSE, MAE, R^2 and MDA equals to 3.75, 3.54, 0.98, and 0.78 respectively whereas the dataset with sentiment has the value of RMSE, MAE, R^2 and MDA equal to 2.35, 2.54, 0.956, and 0.79 respectively. The results from the experiments on all the stock datasets shows that RMSE, and MAE is less than 2.5 and 2.2 respectively and R2 is greater than 0.94 and MDA greater than 0.79.

The experimental results shows that compared with the other widely used model like GRU and LSTM, wavelet transform based data denoising and attention based LSTM perform better on both type of dataset namely only historical data and historical data with human sentiment on all the performance metrics.

The research work can be extended in several ways. The research work has found that the attention mechanism with data denoising has more predictive accuracy than other models. However, adding more technical indicators and the advance sentiment rendering method can be applied. In this research work VADER is used to obtain the sentiment polarity for each day but other machine learning and deep learning approach can also be used to improve the performance and stability of the financial forecasting system.

REFERENCES

[1]. A. N. Refenes, A. Zapranis and G. Francis, "Stock performance modeling using neural networks: a comparative study with regression models," *Neural Networks,* vol. 7, no. 2, pp. 375-388, 1994.

[2]. R.Vishwanath, C. Krishnamurti, "Investment Management: A modern Guide to Security Analysis and Stock Selection", *Springer- Verlag Berlin Heidelberg,* 2009.

[3]. H. G. J. H. Yuling Lin, "An SVM-based approach for stock market trend prediction," in *Proceedings of the International Joint Conference on Neural Networks, Dallas, TX,* 2013.

[4] W. Bao, J. Yue and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and long short term memory," *PLOS ONE,* vol. 12, no. 7, 2017.

[5]. J. Qin, B.Wang, C.Zhou, "Forecasting stock prices with long short term memory neural network based on attention mechanism", *PLOS ONE,* 2020.

[6]. Z.Jin, Y.Yang, Y.Liu, "Stock closing price prediction based on sentiment analysis and LSTM" in *Neural Computing and Applications,* 2019.

[7]. A. Porshnev, I. Redkin, A. Shevchenko, "Machine learning in prediction of stock market indicators based on historical data and data from Twitter sentiment analysis" *in International conference on Data Mining Workshops,* 2013

[8]. I. Alaoui, Y. Gahi, R. Mesoussi, Y. Chaabi, A. Todoskoff, "A novel adaptable approach for sentiment analysis on big social data" *in Journal of big data*, 2018.

[9]. B. D. T. Yahya Eru Cakra, "Stock Price Prediction using Linear Regression based on Sentiment Analysis", *International Conference on Advanced Computer Science and Information Systems,* pp. 147-154, 2016

[10]. T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions", *European Journal of Operational Research*, vol. 270, pp. 654-669, 2018

[11]. J. Liu, et al., "Attention-Based Event Relevance Model for Stock Price Movement Prediction in China", *Conference on Knowledge Graph and Semantic Computing*, 2017.

[12]. G. P. Cheong Fung, J. Xu Yu, and W. Lam, "News sensitive stock trends prediction," in *6th Pacific-Asia Knowledge Discovery in Data Mining*, Beijing, 2002.

[13]. V. Lavrenko , M. Schmill , D. Lawrie , P. Ogilvie , D. Jensen , J. Allan, "Mining of concurrent text and time series," *Workshop of 6th International Conference on Knowledge Discovery and Data Mining,* 2000.

[13]. J. Li, H. Bu, J. Wu, "Sentiment-aware stock market prediction: A deep learning method. In Proceedings", *International Conference on Service Systems and Service Management, Dalian, China*, vol. 16, pp. 1–6, June 2017.

[14]. A. S. Saud and S. Shakya, “Analysis of look back period for stock price prediction with RNN variants: A case study on banking sector of NEPSE”, *Procedia Computer Science*, vol. 167, pp. 788–798, 2020.

[15]. T.B Shahi, A. Shrestha, A Neupane, W. Guo, “Stock Price Forecasting with Deep Learning: A Comparative Study”, *Multidisciplinary Digital Publishing Institute* (MDPI), August 2020.

[16]. P-F Pai, C.-S Lin, “A hybrid ARIMA and support vector machines model in stock price forecasting”, The International Journal of Management Science(Omega), vol. 33, pp. 497–505, 2005.

[17]. D. Lv., Z. Huang, M. Li, Y. Xiang,“Selection of the optimal trading model for stock investment in different industries” *PLOS ONE* , vol. 14, 2019.

APPENDIX