WALCHAND COLLEGE OF ENGINEERING, SANGLI

(Government-Aided Autonomous Institute)

A

Project Report

on

"Time Series Prediction for Chosen Stocks Under NIFTY-50"

Under the Scheme of Minor Certificate Programme on
"Artificial Intelligence and Machine Learning"
Offered by
Department of Information Technology

Submitted By

Yutesh Shriram Mohadikar (2018BTEEN00055)
Sumedh Mukundraj Wasnik (2018BTEEN00056)
Mihir Mahesh Bedekar (2018BTEEN00060)
Pradyumna Santosh Akolkar (2018BTEEN00065)
Ashutosh Dattatraya Kodavade (2018BTEEN00066)
Piyush Prakash Kulkarni (2018BTEEN00068)

Under the guidance of

Prof. M. B. Narnaware

Information Technology Dept,

WCE, Sangli.

2021:2022

WALCHAND COLLEGE OF ENGINEERING, SANGLI

(Government-Aided Autonomous Institute)

DEPARTMENT OF ELECTRONICS ENGINEERING

CERTIFICATE

This is to certify that the Project Report entitled

'Time Series Prediction for Chosen Stocks Under NIFTY-50'

Submitted by

Yutesh Shriram Mohadikar (2018BTEEN00055)
Sumedh Mukundraj Wasnik (2018BTEEN00056)
Mihir Mahesh Bedekar (2018BTEEN00060)
Pradyumna Santosh Akolkar (2018BTEEN00065)
Ashutosh Dattatraya Kodavade (2018BTEEN00066)
Piyush Prakash Kulkarni (2018BTEEN00068)

in partial fulfilment for the award of the Minor Certification Programme of

Artificial Intelligence and Machine Learning

Offered by

Information Technology Department.

is a record of students own work carried out by them under our supervision and guidance during the academic year 2021-2022

Date: __/__/2022
Place: Sangli

Prof. M. B. Narnaware

Project Guide

Abstract

This report presents a model of LSTM based neural network which can be used to provide insights on time series data. The neural network is trained on certain stocks data under NIFTY 50. This dataset is publically available and taken via - https://www.kaggle.com/datasets/rohanrao/nifty50-stock-market-data

The model is trained on Google Colab. The goal of the model is to gain insights from tome series data provided by input features and then predict the value of VWAP that is, Volume Weighted Average Price.

Model implemented using Tensorflow module in python, consists of number of LSTM layers, Dropout layers and Dense layers. Mean Squared Error and Mean Absolute Error are considered for error analysis. The activation functions used are tanh, ReLU and sigmoid. In case of optimizer, RMSProp is selected. All of this is combined in such a way that it can give optimum performance, details of which are presented and explained further in the report.

Acknowledgement

We would like to express our gratitude to our project guide, Prof. M. B. Narnaware for his patient guidance, enthusiastic encouragement and useful tips for this project. We would also like to thank Prof. Dr. U. B. Chavan for their valuable support and important constructive suggestions throughout the process. We also want to thank HOD IT department Dr. A. J. Umbarkar ,Dean Academics Dr. V.B. Dharmadhikari and Dean Exam Cell Dr. P.K. Kharat for offering and and conducting the Minor Certification course very effectively. Finally, we would like to thank our college Walchand College of Engineering, Sangli for giving us the opportunity for carrying out our project.

Yutesh S. Mohadikar
(2018BTEEN00055)

Sumedh M. Wasnik
(2018BTEEN00056)

Mihir M. Bedekar
(2018BTEEN00060)

Pradyumna S. Akolkar
(2018BTEEN00065)

Ashutosh D. Kodavade
(2018BTEEN00066)

Piyush P. Kulkarni
(2018BTEEN00068)

Contents

Chapter 1

Introduction

- ## Project Idea:

Stock market prediction aims to determine the future movement of the stock value of a financial exchange. The accurate prediction of share price movement will lead to more profit investors can make. Predicting how the stock market will move is one of the most challenging issues due to many factors that involved in the stock prediction, such as interest rates, politics, and economic growth that make the stock market volatile and very hard to predict accurately. The prediction of shares offers huge chances for profit and is a major motivation for research in this area; knowledge of stock movements by a fraction of a second can lead to high profits

Machine Learning can be used to analyze time series data. Different methods such as RNNs, LSTMs can be effectively used for such purpose. Here, we use LSTMs for analyzing a dataset containing NIFTY 50 indices of various stocks.

• Proposed Work:

❖ Objectives:

- Understanding the data and pre-processing it to feed to the Model.
- Representing the data in graphical format for better insight.
- Finding patterns in the historical data.
- Training and using the model for predicting the future value of chosen stocks in NIFTY 50 index

❖ Methodology:

- First, we divide the project into small sub-parts such as creating useful dataset, plot the dataset to understand the insights, figure out training & testing data percentage, creating actual ML model ,etc.
- We find the existing ways or technologies for designing these small parts and will try to learn from the existing projects.
- We try to modify the existing system for better performance and better feasibility.
- After finalizing the ML model, we tried to optimize the parameters and some hyperparameters to increase the accuracy of our model.
- We implement the model for different 13 stocks and evaluate the result.

❖ Expected Outcomes:

- Achieve Insights such as high, low, mean etc through pre-processing of data.
- Build model that can predict the VWAP with good accuracy.

Chapter 2

Literature Survey

2.1 Stock Market Index Forecasting of Nifty 50 Using Machine Learning Techniques with ANN Approach -

Authors: Gaurav Kumar

Year of Publication: 2016

Keywords: Machine Learning, Artificial Neural Network (ANN), Feed Forward Neural Network, Multi-Layer Perceptron (MLP), National Stock Exchange (NSE), Stock Market Prediction

Description: This paper presents an ANN based approach to forecast the Nifty 50 Index. A feed-forward neural network using multiple back propagation algorithms has been used to forecast next day's OHLC data. This model has used the pre-processed dataset of Open price (O), High price (H), Low price (L), Close price (C), Volume Traded (V) and Turnover (T).In this research Multiple Back -Propagation (MBP version 2.2.4) software has been used to predict the future stock prices and their performance statistics have been evaluated

2.2 Machine Learning Models in Stock Market Prediction -

Authors: Gurjeet Singh

Year of Publication: 2022

Keywords: Artificial Neural Network, Stock, Market Prediction, Supervised Machine Learning Models, Time Series Data

Description: The paper focuses on predicting the Nifty 50 Index by using 8 Supervised Machine Learning Models. The techniques used for empirical study are Adaptive Boost (AdaBoost), kNearest Neighbors (kNN), Linear Regression (LR), Artificial Neural Network (ANN), Random Forest (RF), Stochastic Gradient Descent (SGD), Support Vector Machine (SVM) and Decision Trees.

Chapter 3

Design of System

- **Pre-processing the Data:**

NIFTY 50 dataset consists of total 52 csv files. Out of which 50 contain stock data, one file contains metadata and last one contains all data aggregated in a single file. Each of the 50 stock data files has 15 columns containing various information about each such as Date, High, Low, Volume, VWAP (Volume Weighted Average Price) etc.

For our interest, out of 50, we have selected one stock from each sector by carefully reviewing stock_meatadata.csv file. The list of stocks taken into consideration for this project are as follows-

1. Automobile- Maruti Suzuki (MARUTI)
2. Cement and Cement Products- UltraTech (ULTRACEMCO)
3. Construction- Larsen and Turbo (LT)
4. Consumer Goods- Nestle (NESTLEIND)
5. Energy- Bharat Petroleum (BPCL)
6. Fertilisers and Pesticides- United Phosphorus Limited (UPL)
7. Financial Services- State Bank of India (SBIN)
8. IT- Infosys (INFY)
9. Media and Entertainment- Zee (ZEEL)
10. Metals- Tata Steel (TATASTEEL)
11. Pharma- Dr Reddy (DRREDDY)
12. Services- Adani Ports (ADANIPORTS)
13. Telecom- Bharati Airtel (BHARATIAIRTEL)

In each csv file, as stated earlier there are 15 columns. They are as follows-

1. Date
2. Symbol
3. Series
4. Prev Close
5. Open
6. High
7. Low
8. Last
9. Close
10. VWAP (Volume Weighted Average Price)
11. Volume
12. Turnover
13. Trades
14. Deliverable
15. % Deliverable

Out of all these, the important ones for our purposes are –

- VWAP
- Volume
- Turnover

Thus, we format our data-frame to only focus on these columns and use min-max normalization.

- **Model Architecture:**

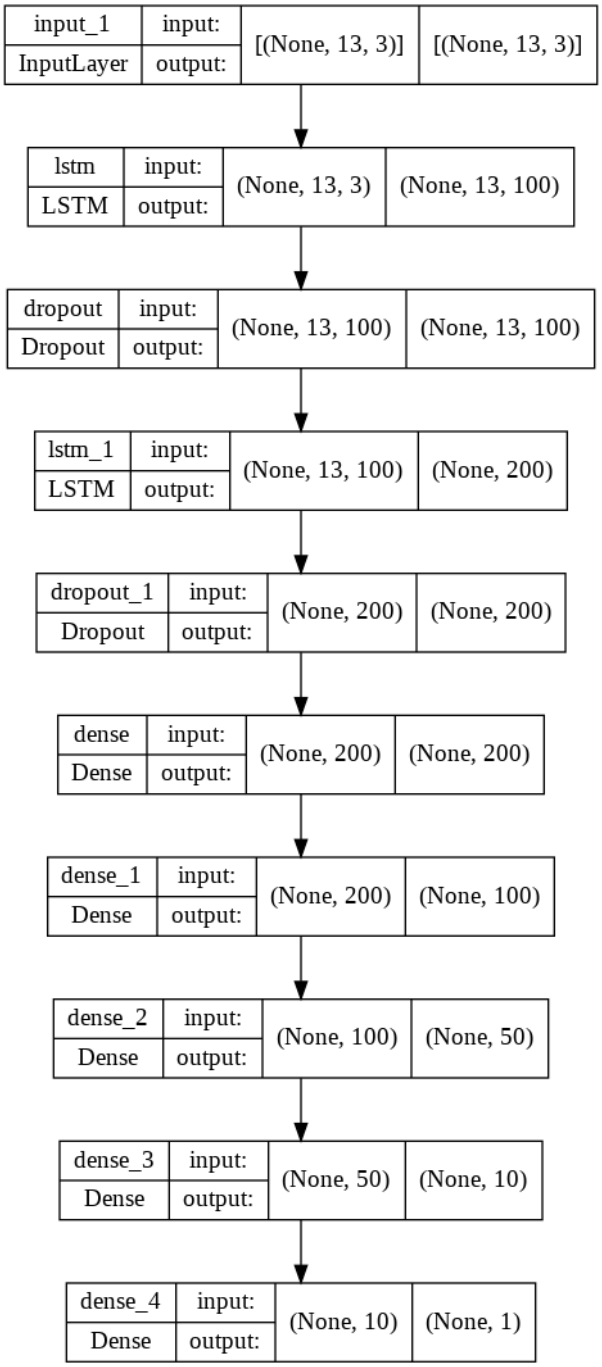


Fig. Model Architecture

The important part of a deep learning model aside from hyperparameters are the selection of optimizer, loss function and activation function for layers. For optimizer, we have selected ‘RMSprop’ as it enables us for a faster learning rate in one direction. In case of loss function, we have selected ‘mse’ as it gives simple yet effective performance. For analyzing time series data, often RNNs or LSTMs are used. Here, we use architecture containing LSTMs and Dense layers. Activation functions used are ‘tanh’, ‘ReLU’ and ‘sigmoid’- these are chosen considering their range of outputs and nature of waveform.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm (LSTM)                 (None, 13, 100)           41600

 dropout (Dropout)           (None, 13, 100)           0

 lstm_1 (LSTM)               (None, 200)               240800

 dropout_1 (Dropout)         (None, 200)               0

 dense (Dense)               (None, 200)               40200

 dense_1 (Dense)             (None, 100)               20100

 dense_2 (Dense)             (None, 50)                5050

 dense_3 (Dense)             (None, 10)                510

 dense_4 (Dense)             (None, 1)                 11

=================================================================
Total params: 348,271
Trainable params: 348,271
Non-trainable params: 0
_________________________________________________________________
```

Fig. Model Summary

Chapter 4

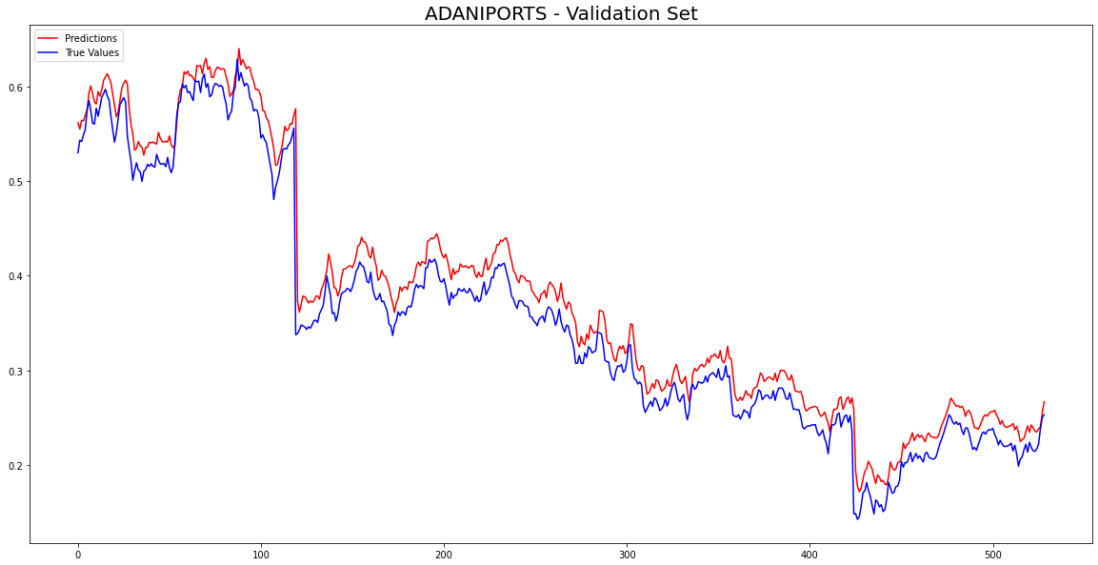
Implementation

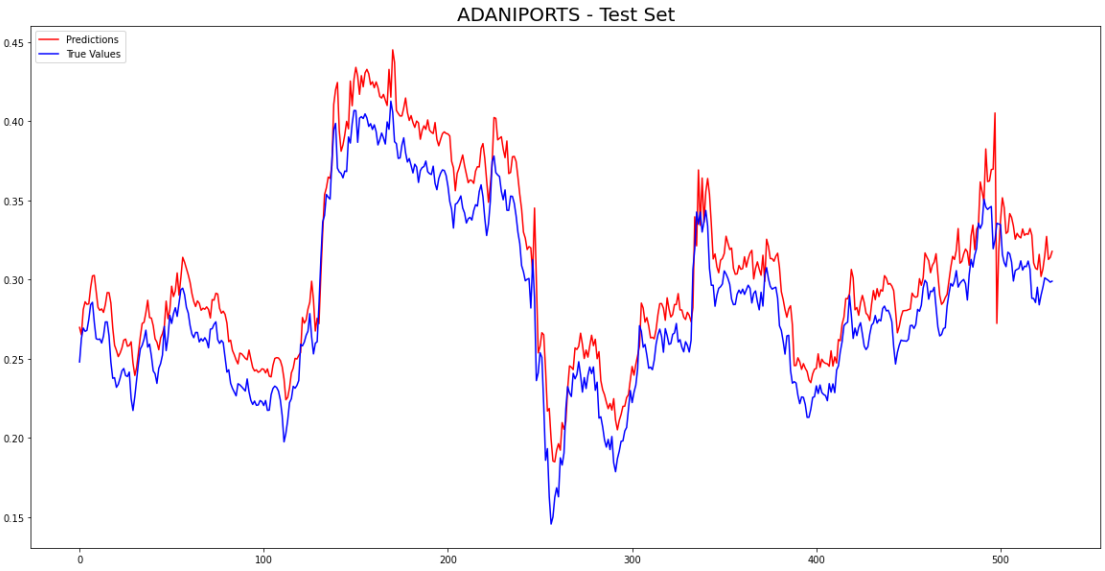
- **Code Link –**

  Time Series Prediction for Chosen Stocks Under NIFTY-50

- **Graphs of predicted vs actual values for Validation dataset and Test dataset:**

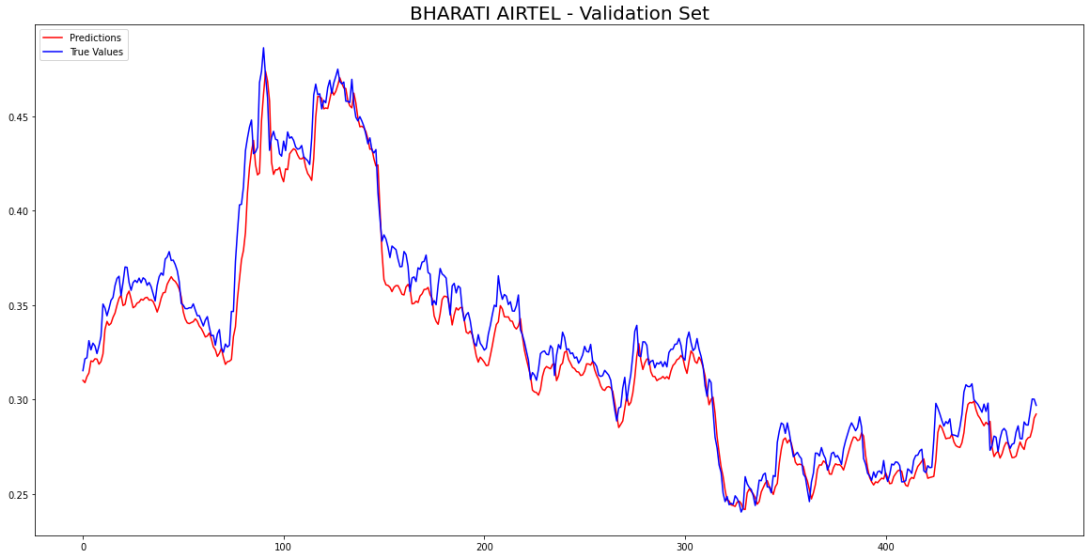
1. ADANI PORTS –

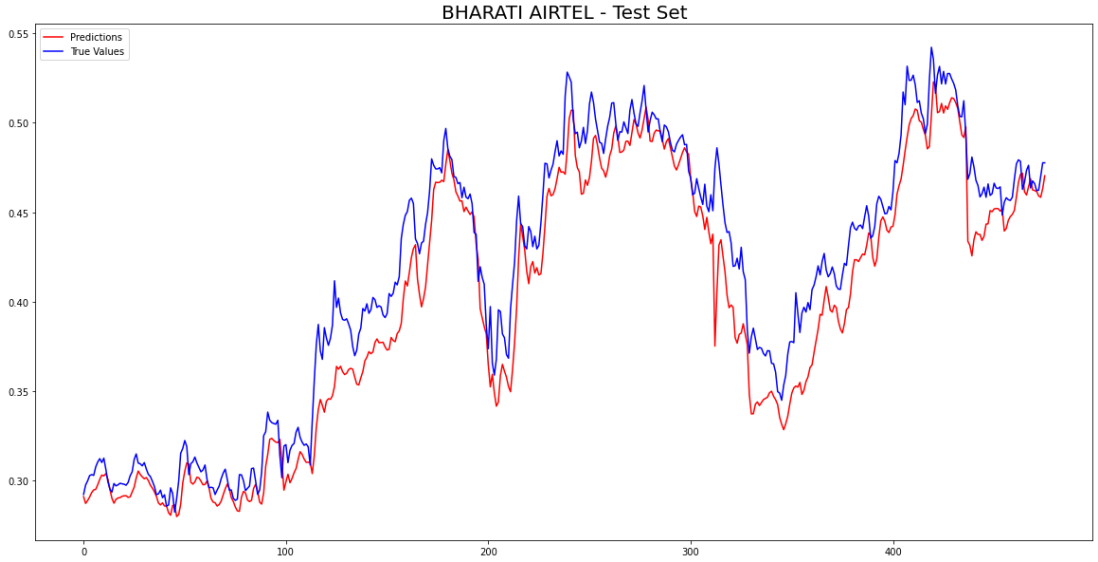




% Loss: 0.055
% Mean Squared Error: 0.055
% Mean Absolute Error: 2.1099

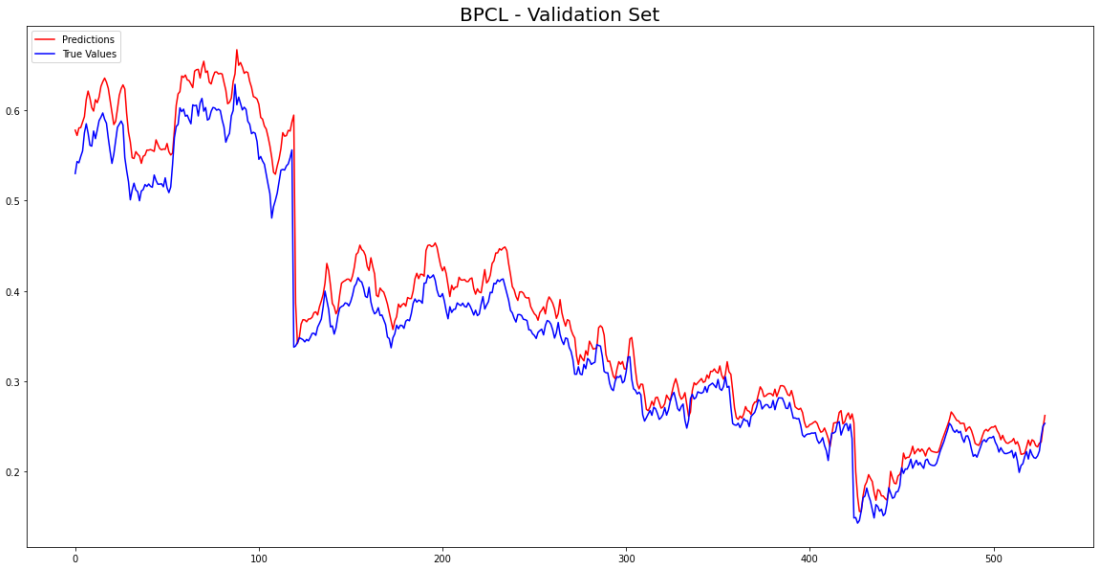
2. BHARATI AIRTEL –

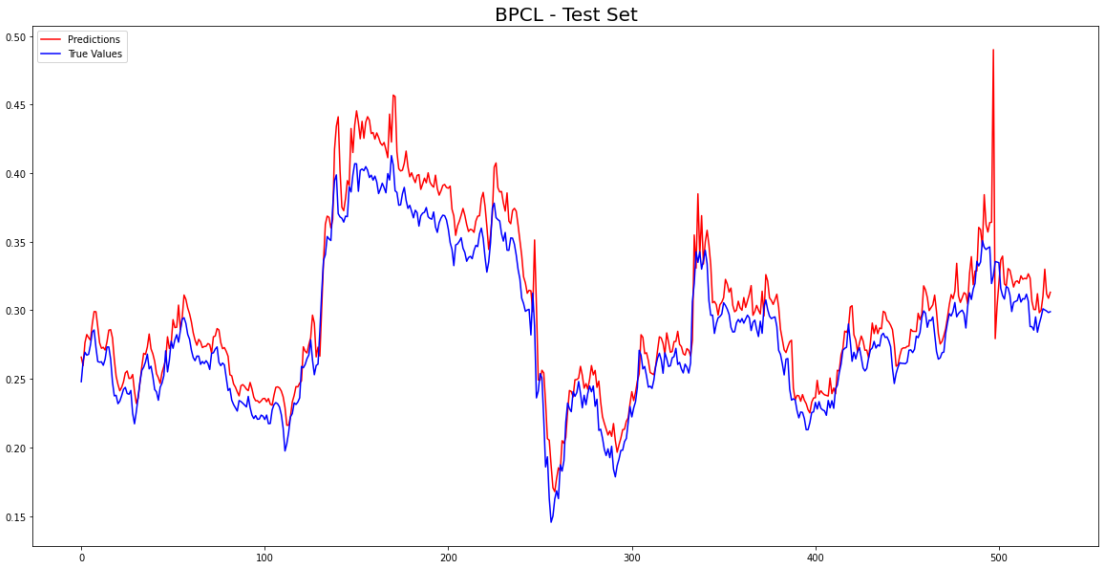




% Loss: 0.0509
% Mean Squared Error: 0.0509
% Mean Absolute Error: 1.812

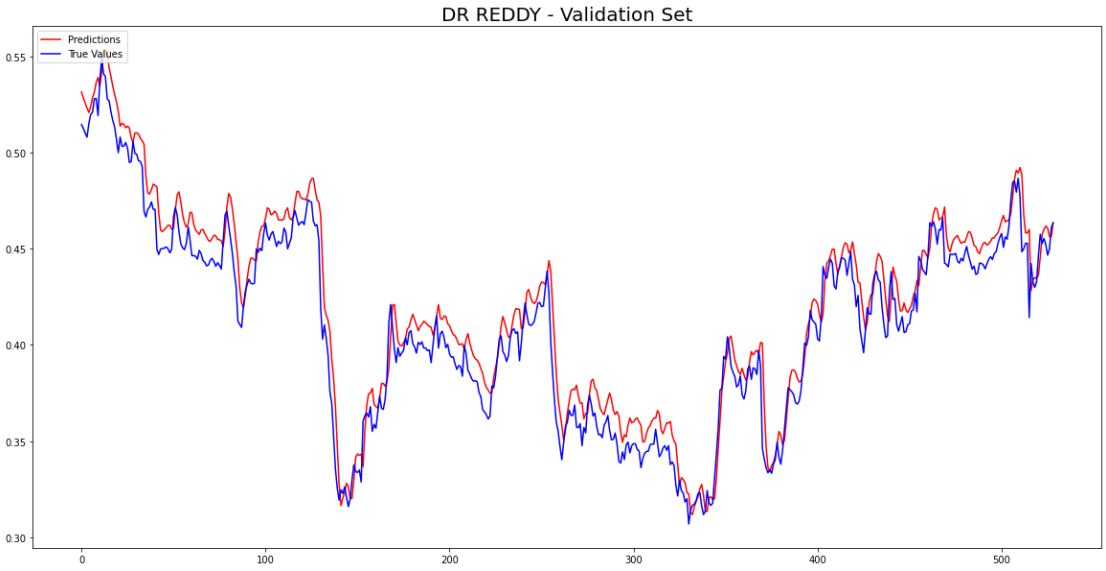
3. BPCL –

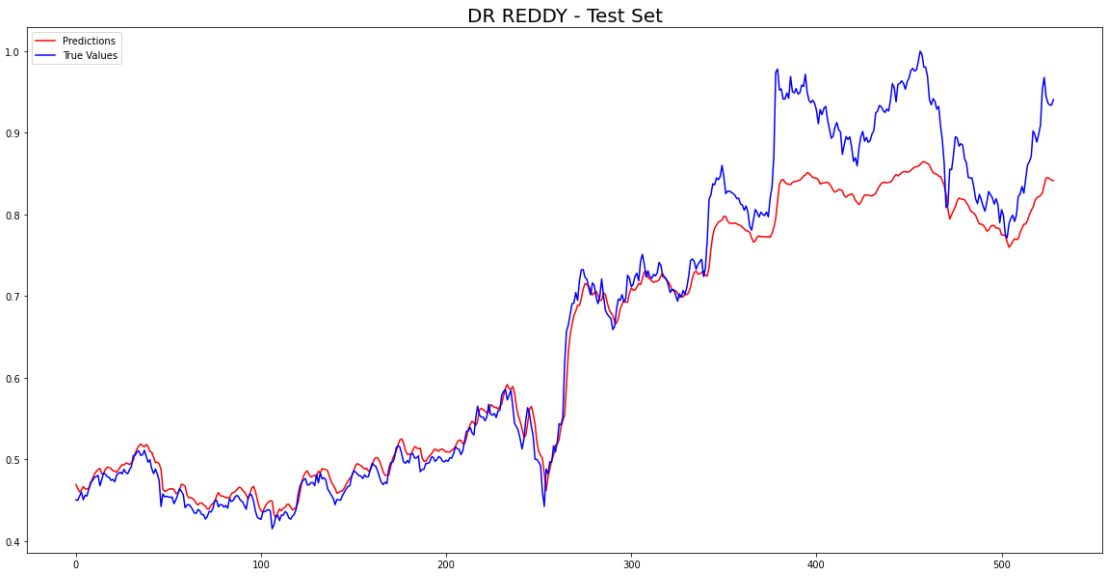




% Loss: 0.0461
% Mean Squared Error: 0.0461
% Mean Absolute Error: 1.7026

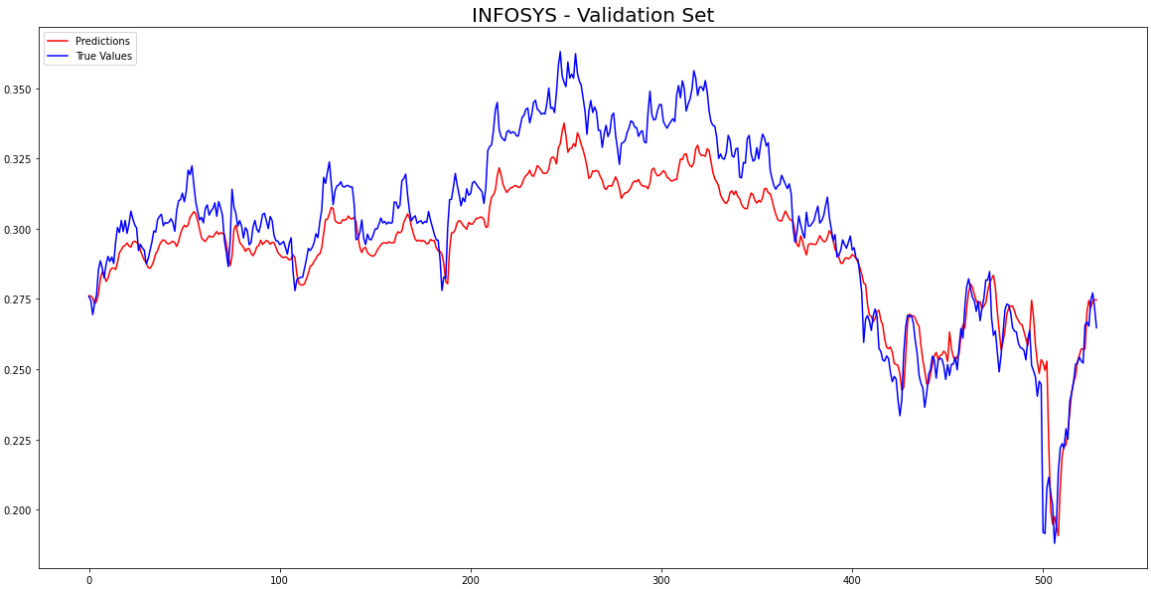
4. DR REDDY –

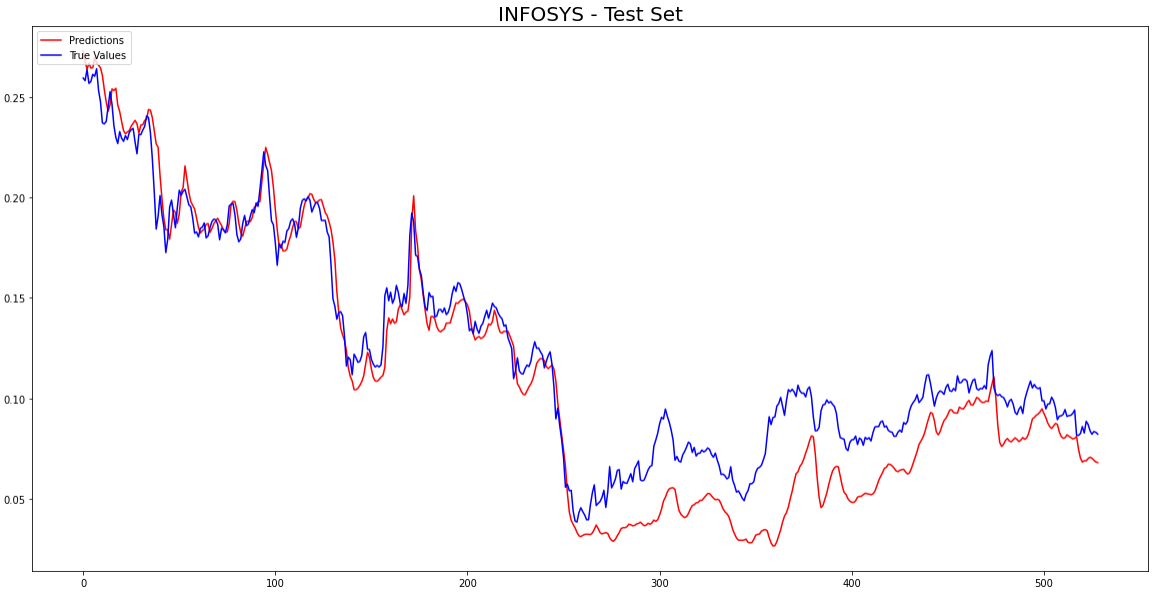




% Loss: 0.2256
% Mean Squared Error: 0.2256
% Mean Absolute Error: 3.1875

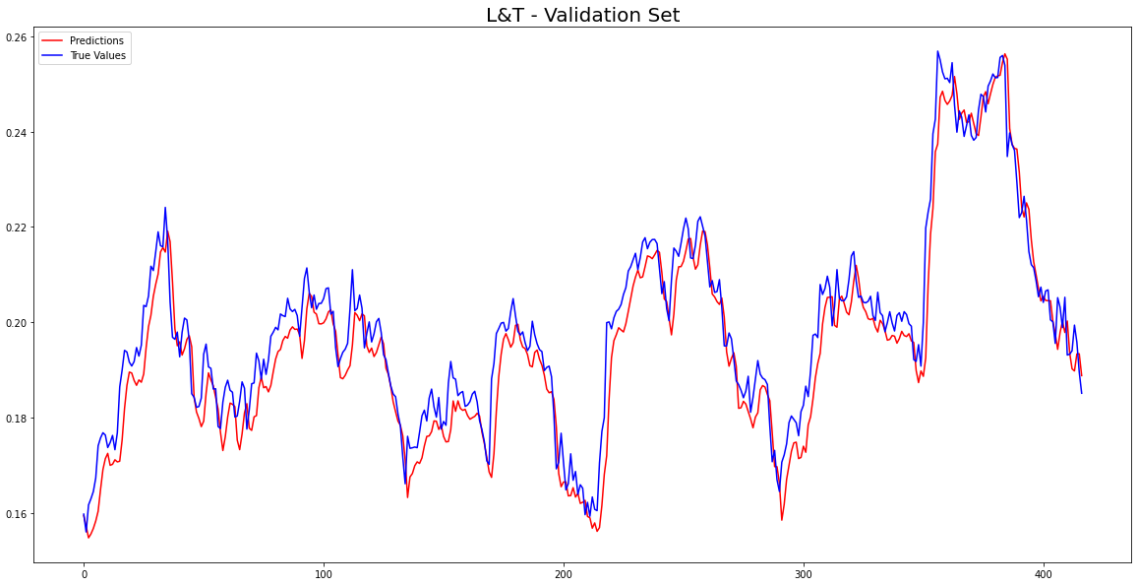
5. INFOSYS –

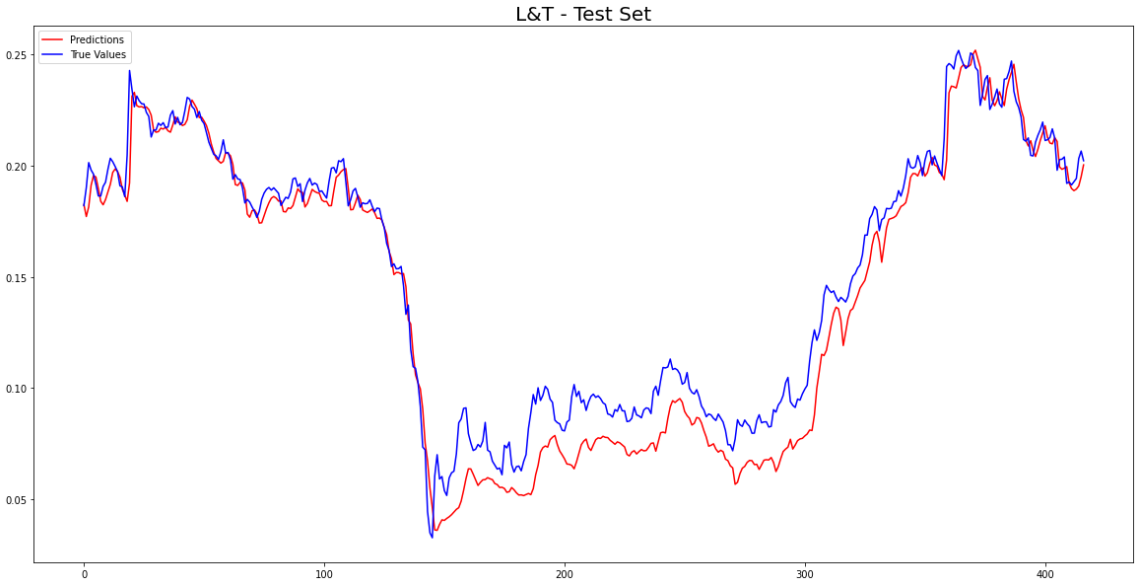




% Loss: 0.0419
% Mean Squared Error: 0.0419
% Mean Absolute Error: 1.5959

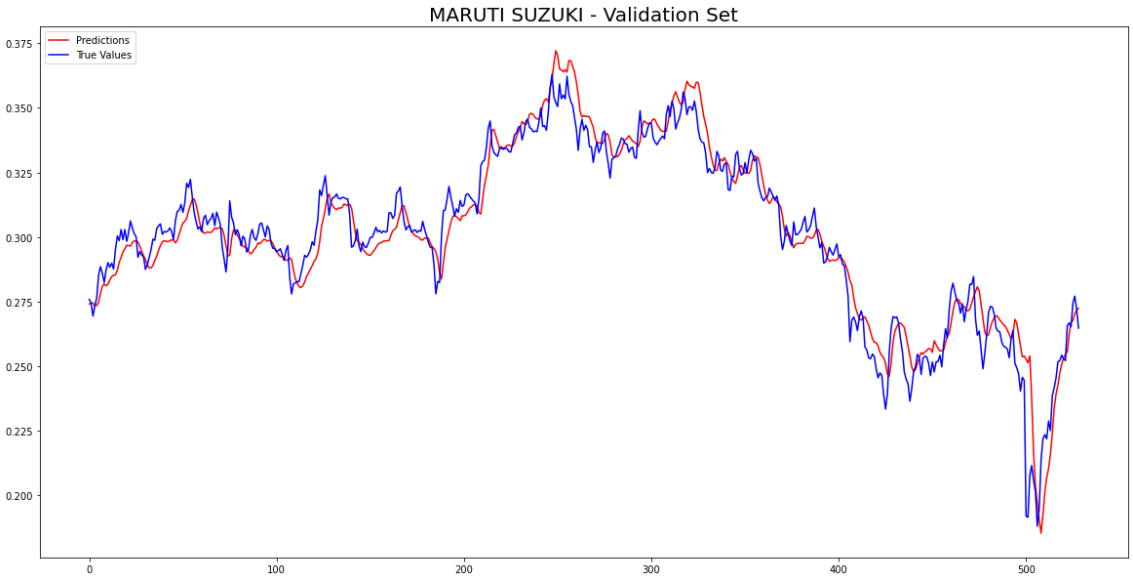
6. L & T –

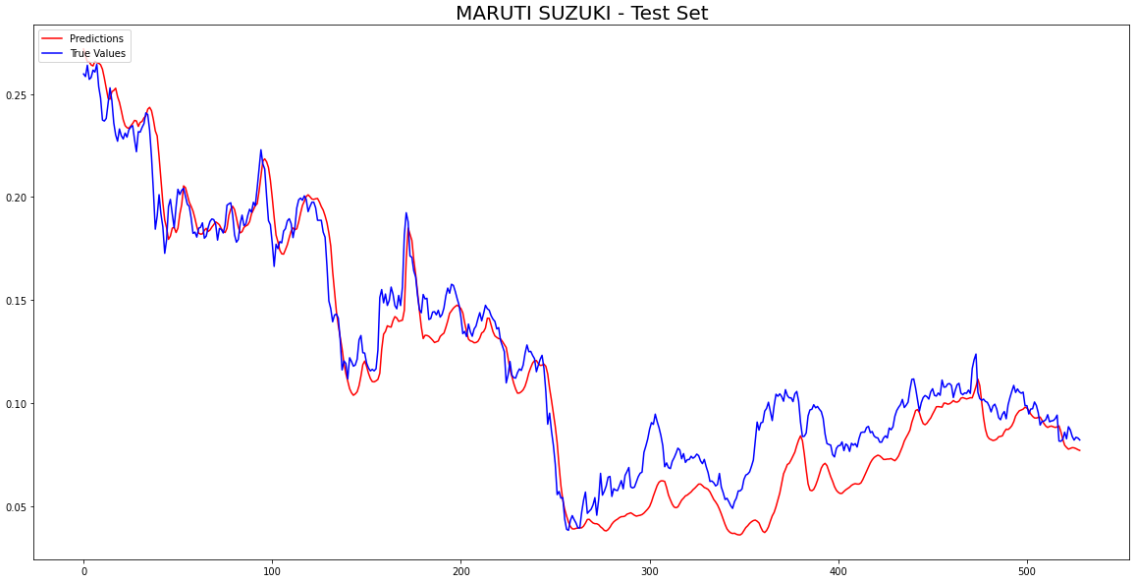




% Loss: 0.021
% Mean Squared Error: 0.021
% Mean Absolute Error: 1.1266

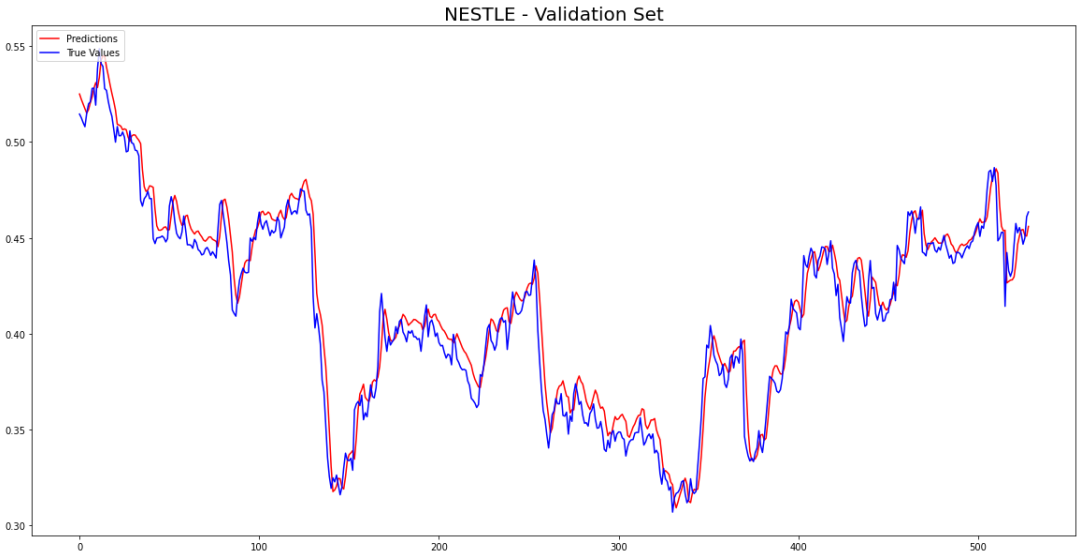
7. MARUTI SUZUKI –

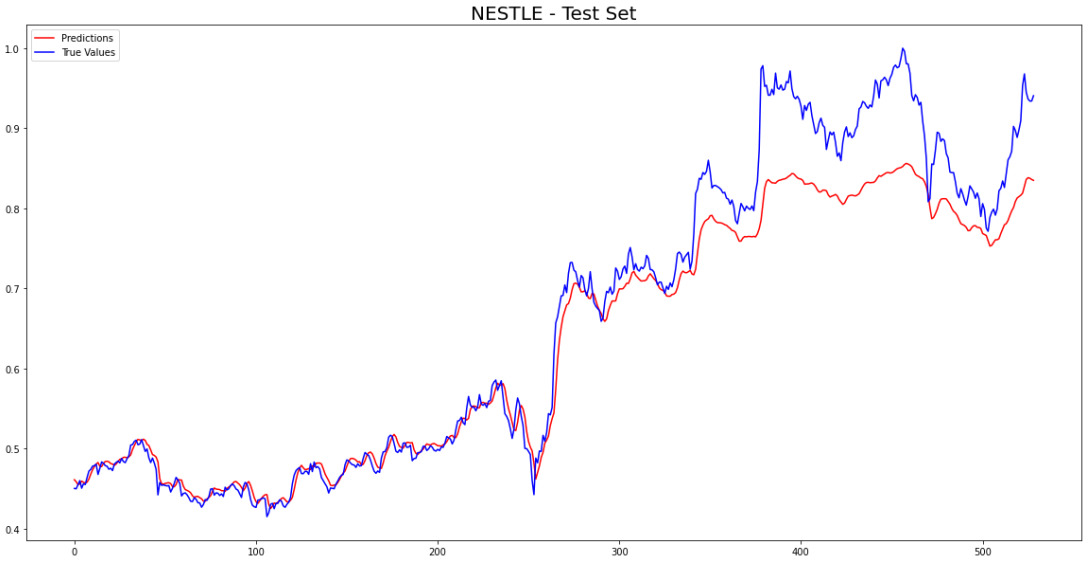




% Loss: 0.0291
% Mean Squared Error: 0.0291
% Mean Absolute Error: 1.3077

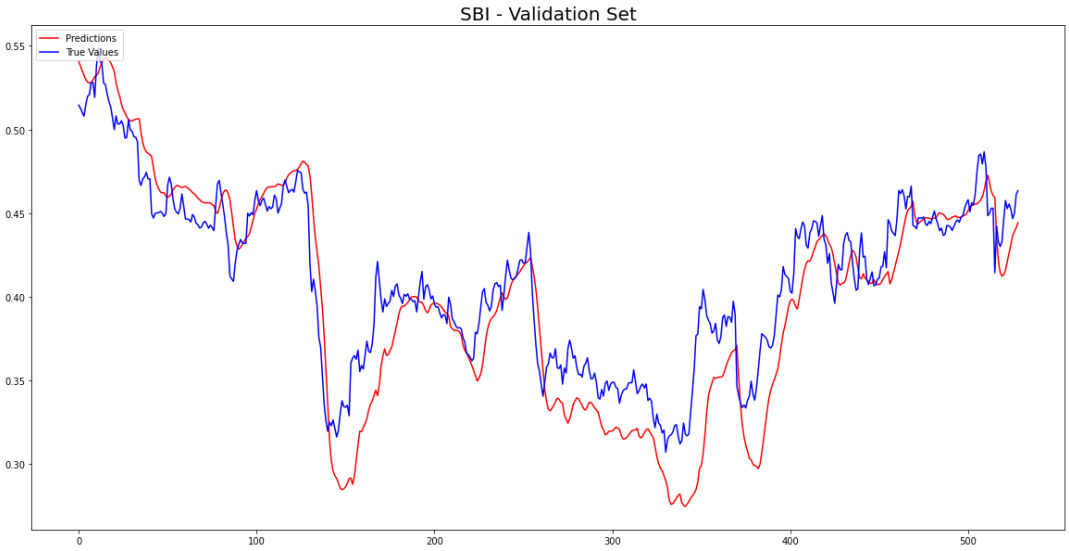
8. NESTLE –

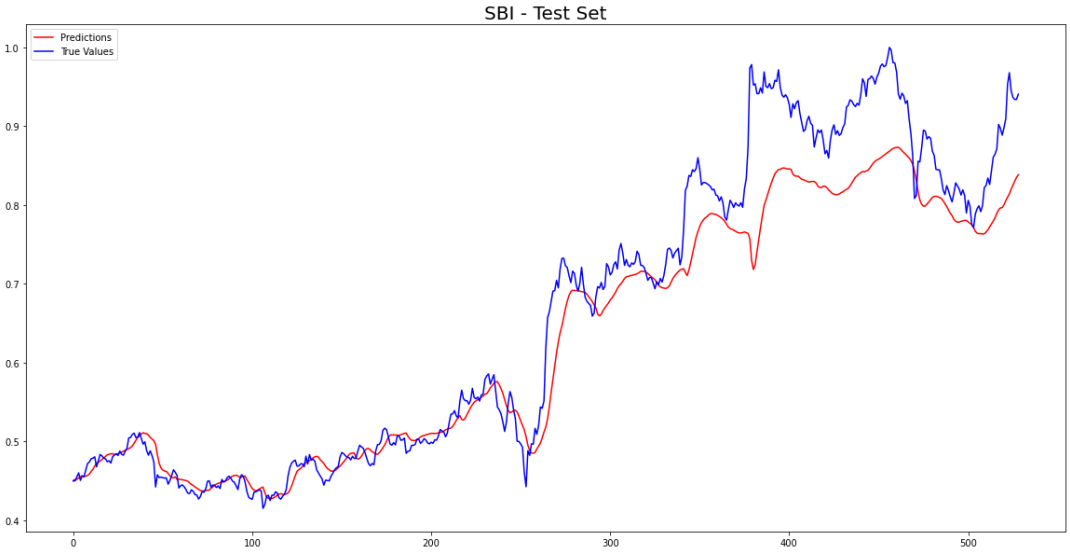




% Loss: 0.2673
% Mean Squared Error: 0.2673
% Mean Absolute Error: 3.4037

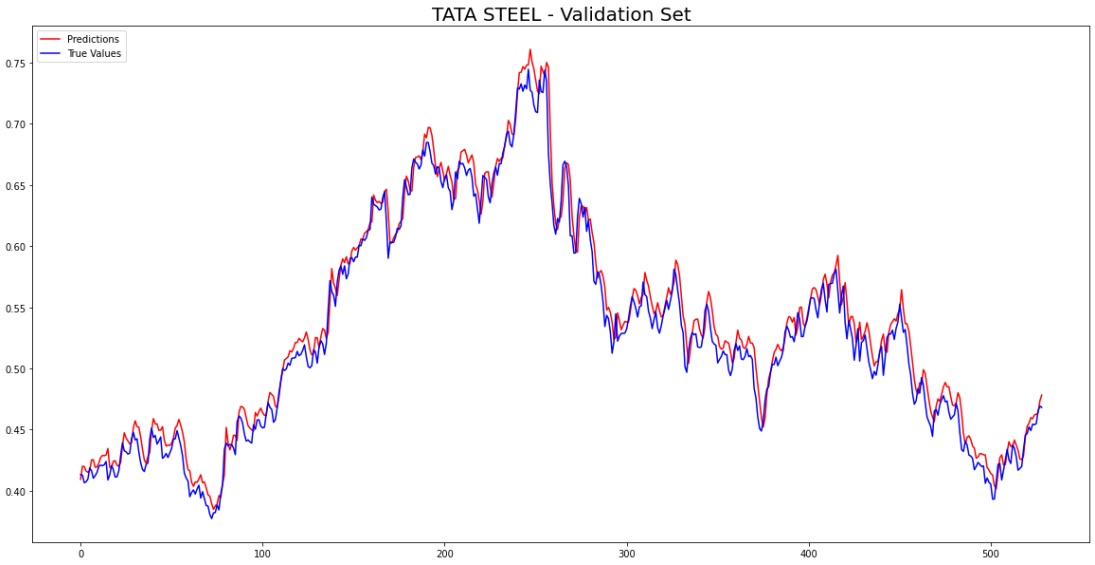
9. SBI –

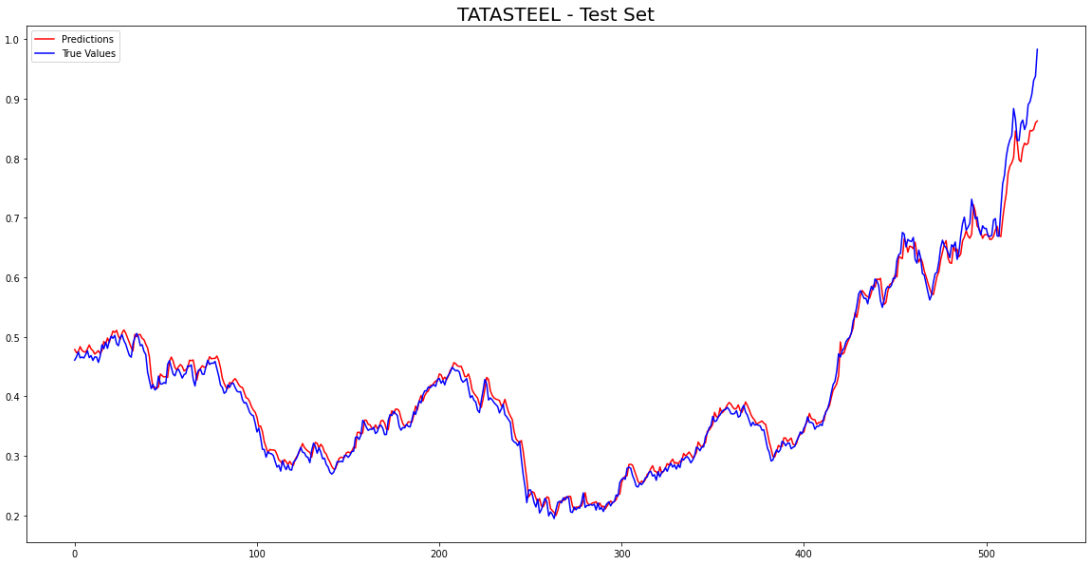




% Loss: 0.3241
% Mean Squared Error: 0.3241
% Mean Absolute Error: 3.7719

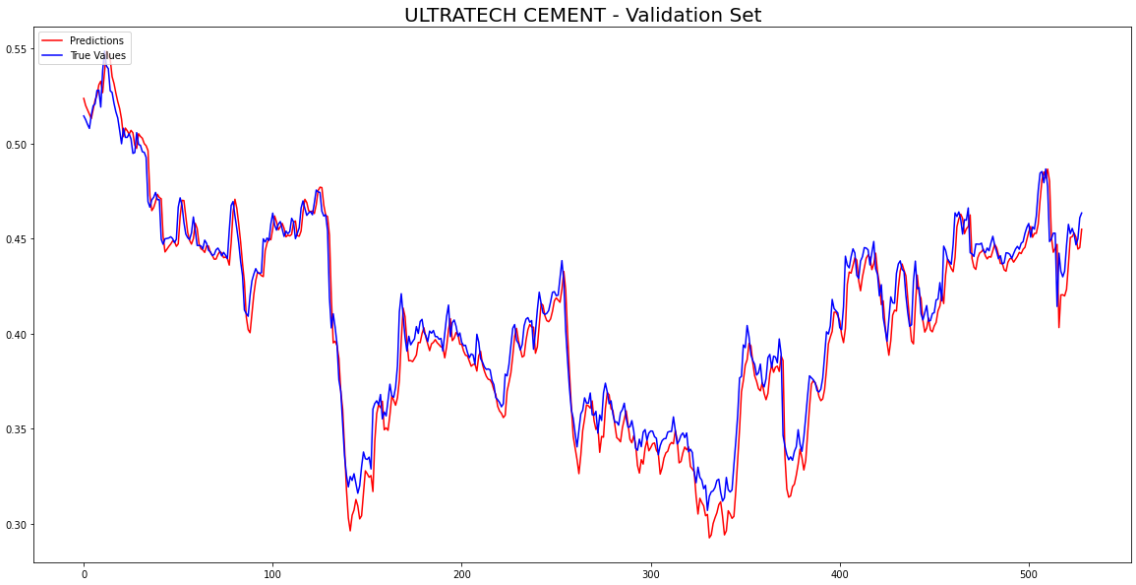
10. TATA STEEL –

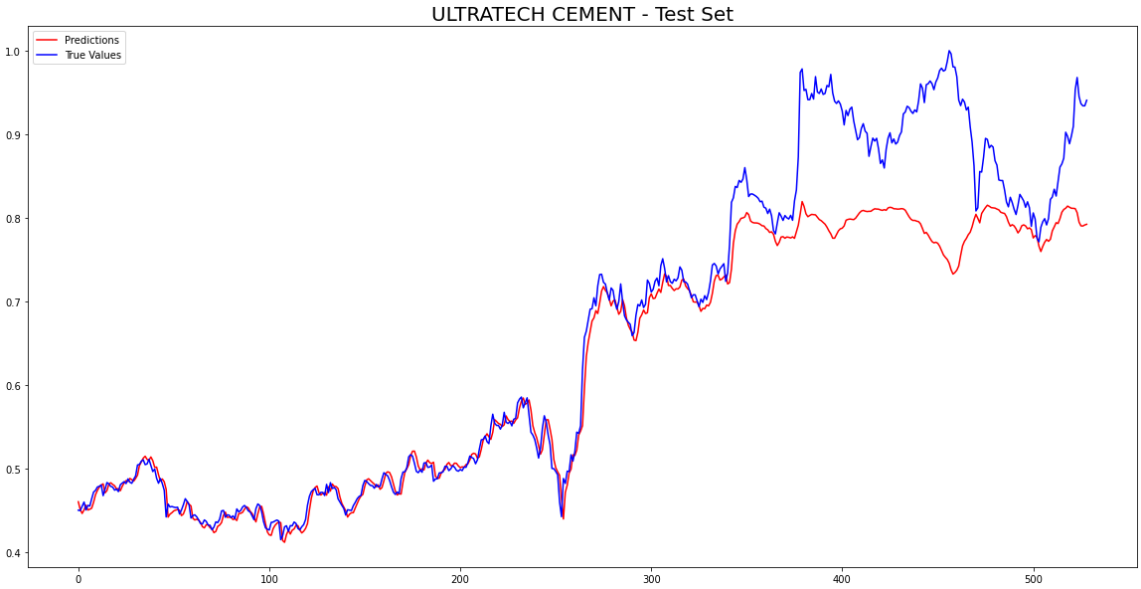




% Loss: 0.0314
% Mean Squared Error: 0.0314
% Mean Absolute Error: 1.2336

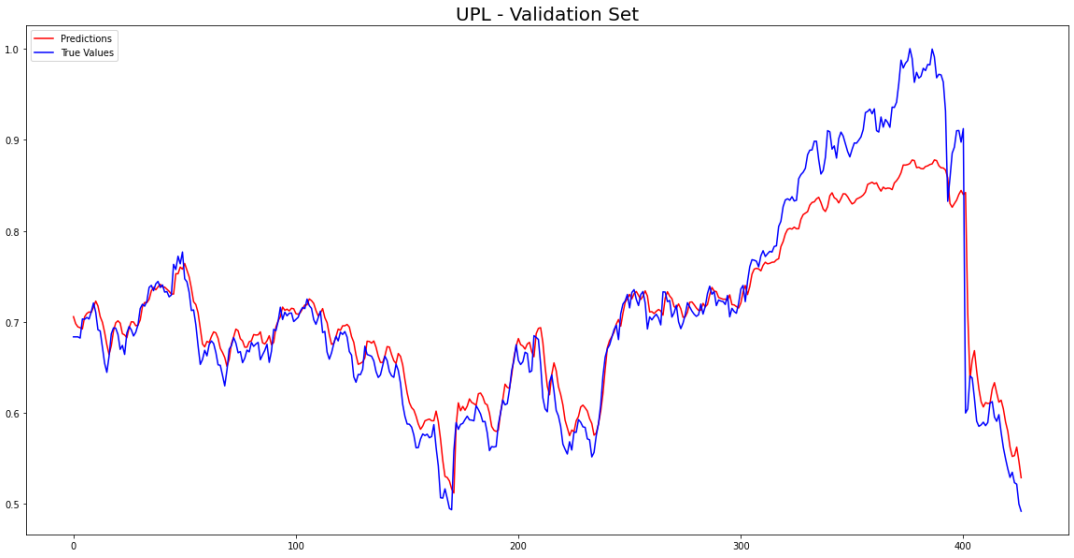
11. ULTRATECH –

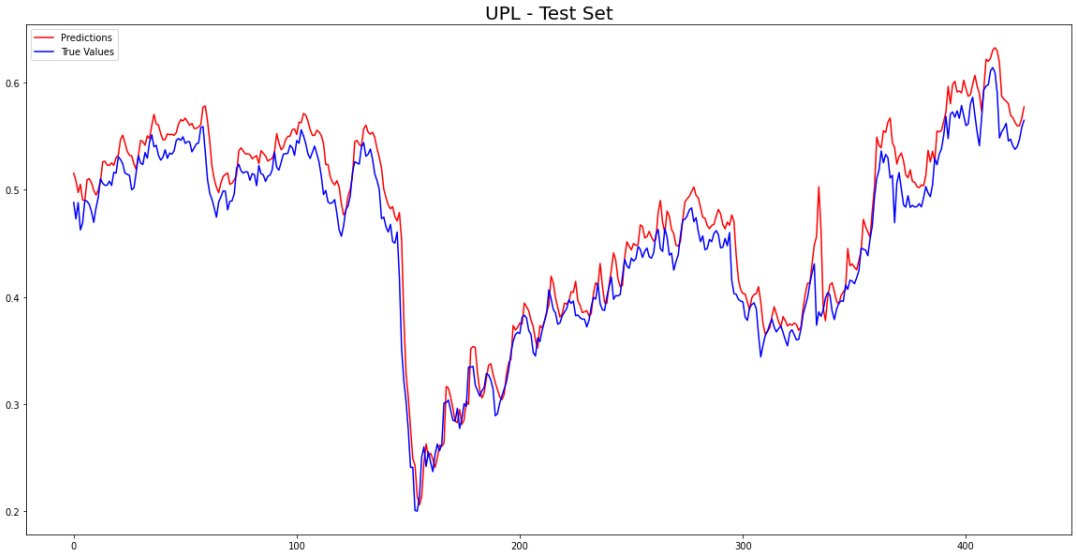




% Loss: 0.4631
% Mean Squared Error: 0.4631
% Mean Absolute Error: 3.878

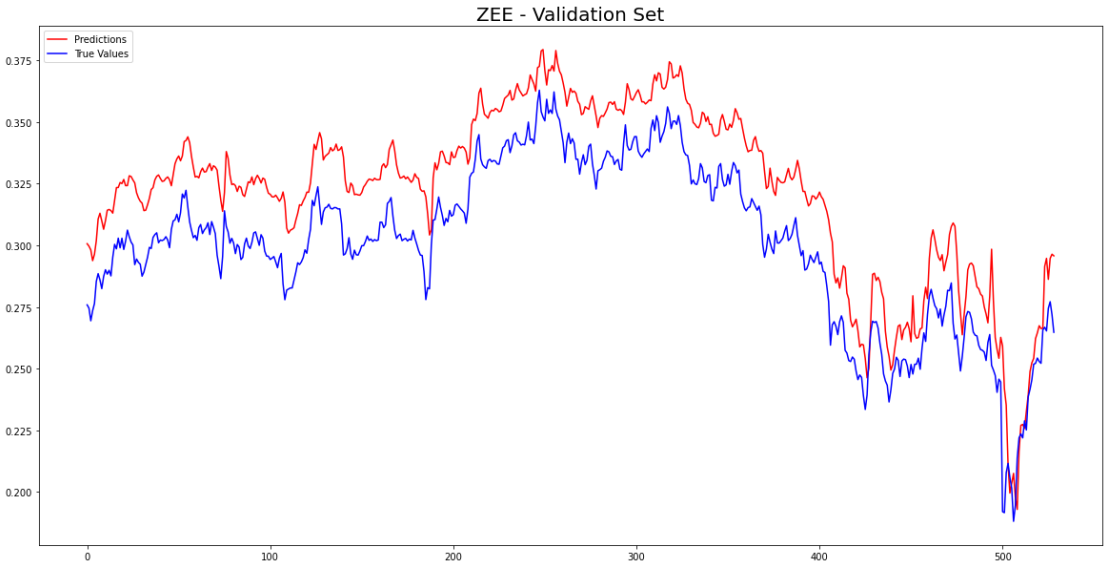
12. UPL –





% Loss: 0.0574
% Mean Squared Error: 0.0574
% Mean Absolute Error: 1.896

13. ZEE –



% Loss: 0.0231
% Mean Squared Error: 0.0231
% Mean Absolute Error: 1.1763

Chapter 5

Results And Discussion

Sr. No.	Stock	% Loss	% MAE
1	ADANI PORTS	0.0078	0.7238
2	BHARATI AIRTEL	0.0477	1.6145
3	BPCL	0.0266	1.1065
4	DR REDDY	0.1306	2.0093
5	INFOSYS	0.0046	0.5082
6	L&T	0.0091	0.6914
7	MARUTI SUZUKI	0.1115	2.7622
8	NESTLE	1.4014	11.0247
9	SBI	0.0114	0.8542
10	TATA STEEL	0.0193	1.1036
11	ULTRATECH CEMENT	0.2306	3.8662
12	UPL	0.1019	2.1644
13	ZEE	0.0221	1.1311

As we can analyse from above table the average mean absolute error of the model based on performance on Validation Set as well as Test Set comes out to be 15 %. That is, we can evaluate out our average accuracy to be 85 %.

Chapter 6

Conclusion And Future Scope

- Insights such as high, low, mean etc are achieved through pre-processing of data.
- Built model using number of LSTMs and Dense layers works with around 85% accuracy.
- In almost all cases, graphs of the predicted and true values have similar nature suggesting that model is extracting features correctly.
- The bias and variance present in case of few stocks can be minimized by including more features and fine tuning the model by varying parameters such as sequence length.

Chapter 7

References

[1] https://distill.pub/2016/augmented-rnns/

[2] https://distill.pub/2019/memorization-in-rnns/

[3] https://www.kaggle.com/raoulma/ny-stock-price-prediction-rnn-lstm-gru

[4] https://www.tensorflow.org/tutorials/structured_data/time_series

[5] https://www.coursera.org/learn/tensorflow-sequences-time-series-and-prediction

[6] https://towardsdatascience.com/a-look-at-gradient-descent-and-rmsprop-optimizersf77d483ef08b

[7] https://arxiv.org/ftp/arxiv/papers/2202/2202.09359.pdf