



Financial Time Series Analysis with Machine Learning

Submitted as Research Report / Honours in SIT723

30-September-2022
T2-2021

Satvik Sharma
STUDENT ID 218595095
COURSE - Bachelor of Software Engineering Honours (S464)

Supervised by: Prof. Maia Angelova

Abstract

This article surveys the publications available on the financial time series analysis with machine learning and aims to answer what machine learning models are most commonly used, how accurate the models can be and what is the future like for the field. Different machine learning techniques have been discussed. A number of papers have been read and a technical analysis on the basis of each paper has been done and according to which a conclusion has been drawn, which is, each of the machine learning model has its own pros and cons, but still is not precise to rely on the forecast. The paper follows the workings of the Statistical machine learning, Deep learning and Hybrid models, a total of 10 different algorithms to find the best among them. The data set for the stock has been chosen from the top three stocks from the Nifty-50, and their results for 10-years, 5-years and 1-year has been discussed. The model was trained for 70% of dataset and gave prediction for the rest. The optimizer used for the deep learning models was Adamax. For the short term results, it was seen that Statistical models are the best performing models, but as the dataset increases, the performance of them decreased, especially for ARIMAX model. For the deep learning models, they are working well for all types of dataset. The bidirectional and the hybrid models did not perform well for the 1-year dataset, but as the dataset increased, the performance increased as well. Overall, most reliable model were CNN and SARIMAX, which showed consistent results.

Contents

1	Introduction	1
1.1	Introduction to Financial Time Series	1
1.2	Objectives	1
1.3	Resources used for the research project	2
2	Literature Review	3
2.1	Literature Review Analysis	3
2.2	Common methodologies used	9
2.3	Gaps in current literature	10
3	Research design and methodology	11
3.1	Some common Time series evaluating techniques	12
3.2	Experimentation design	12
3.3	Evaluating techniques used in the paper	14
3.4	Goals of the research	14
4	Artefact Development	15
4.1	Data set and Data pre-processing	15
4.2	Description of models used	16
5	Results	18
5.1	Comparing resulting graphs for test data and the predicted data for different stocks	18
6	Evaluation	31
6.1	evaluating stocks and the algorithms used	32
7	Discussion	36
8	Conclusion	40
8.1	Conclusion	40
8.2	Future for Financial Time series	41

List of Figures

1	Development Phases for the Project	11
2	HDFC 1-year graphical results	19
3	HDFC 5-years graphical results	21
4	HDFC 10-years graphical results	22
5	Infosys 1-year graphical results	24
6	Infosys 5-years graphical results	25
7	Infosys 10-years graphical results	26
8	Hindustan Unilever 1-year graphical results	28
9	Hindustan Unilever 5-years graphical results	29
10	Hindustan Unilever 10-years graphical results	31

List of Tables

1	Table reviewing literature.	3
2	Methodologies used in the reviewed literature.	9
3	HDFC stock evaluation	33
4	Infosys stock evaluation	34
5	Hindustan Unilever stock evaluation	36

1 Introduction

Time Series can be defined as the study of dynamic consequences over a period of time. This can be represented with a simple first order linear equation, for example in the equation

$$y_t = \Theta y_{t-1} + c$$

y_t represents the value to be taken out at time t in respect to the changes (Θ), external factors(c) and previous value at time $t-1$ [13]. This article focuses on the financial time series, where the study is more concerned with the financial assets like stocks, shares, currency evaluation, etc. Even though the study of financial time series is a part of time series, but it is highly logical area, where the uncertainty is extremely high, let's say for example the asset returns in the stocks' time series cannot be observed directly. The addition of uncertainty, statistical theory, methods and high volatile market makes financial time series analysis different from regular time series analysis [34].

1.1 Introduction to Financial Time Series

Financial time series has always been of interest of business and financial analysts and when machine learning started to gain popularity, more and more publications have kept adding to the finance literature as determining more effective ways of prediction is important for right investments. This paper will survey the publications available used for financial time series analysis with machine learning and will try to answer the following questions

- What machine learning models are used for financial time series analysis?
- Difference between the most commonly used machine learning algorithms
- Future of machine learning in the field of financial time series analysis

1.2 Objectives

After answering those questions in the literature review, we can start working on the machine learning models and learn how the different models work. For the project, I am choosing the data set from the stock from the Nifty-50 for the time period of ten years, five years and one year between the time period of 2011 to 2020 and working

on the previous close value for each of the stock chosen. The previous close value is defined as the price of a stock or market index was when the market closed on the previous trading day.

The models that have been compared for the research analysis are from Statistical, Deep learning and Hybrid models, and see how they are working and how accurate its prediction is. The accuracy will be determined with the evaluation criteria based on root mean squared error, absolute mean error and R2 score and will try to answer the following questions.

- What can be the similarities and differences in the algorithms used for financial time-series analysis?
- How does it tackle the problem of customers who want to invest in that particular stock
- What can be done in the future in order to tackle the problem of higher accuracy for the algorithms?

1.3 Resources used for the research project

- Google Scholar: most of the literature reviewed has been found on the google scholar and reading different published articles.
- Kaggle: the data set has been chosen from the Kaggle website.
- Jupyter Notebook/Anaconda: Jupyter notebook available on anaconda is used for all the statistical machine learning algorithms
- Google Colab: Deep learning algorithms are more compatible to work on the Google colab.
- OneDrive: All the files are stored as a backup on OneDrive
- Trello: Trello board is used for managing all the tasks of the project.
- Teams' folder: Microsoft teams is used as a cloud folder for all the tasks, documents etc.

2 Literature Review

There are hundreds of publications on stock market, trading systems, Forex, etc that use the machine learning technologies like Artificial Neural Network, Evolutionary Computations, Genetic programming, Hybrid techniques or some other [30]. Some of the papers also studied the traditional time series models and compared them with the machine learning techniques, for example in one of the papers written by Xin-Yao Qian, ARIMA was used in comparison with the Logical regression, SVM, and De-noising Auto encoders and it was found that the other machine learning models performed better than ARIMA because the other machine learning models took external factors into account as well [27]. One of the papers, compares the performance of LSTM and ARIMA, and in conclusion it was found that LSTM based models run much better than the ARIMA based models even if the data set was for one month [32].

2.1 Literature Review Analysis

There are number of papers which have been mentioned in the table below, that have been reviewed and thus analysed what machine learning algorithms outperforms others. The table will show a small summary of what machine learning models were used and what was the conclusion of the papers.

Table 1: Table reviewing literature.

Model Name	Paper	Performance Criteria	feature data-set	time period	environment	conclusion
ANN, LSTM, MLR	[20]	MSE, r, RMSE	Prediction of irrigation groundwater quality	-	Python	ANN and MLR model have highest accuracy in multiple scenarios

Continuation of Table 1						
Model Name	Paper	Performance Criteria	feature data-set	time period	environment	conclusion
ARIMA, LSTM	[32]	percentage Reduction in RMSE	Dow Jones Index	1-month	Python with Keras and Theano	LSTM based models work better than ARIMA
ARIMA, LSTM	[11]	RMSE	DAX30, S and P500	10-years (2004-2014)	-	KNN outperforms other models
KNN, SVM, Gaussian progress	[2]	-	M3 time series competition data	Monthly for thousand time series	-	Best two models are MLP and gaussian progression
ANN, ARMAX, 3-D Hydrody - namic model	[38]	MAE, RMSE, R	Water level fluctuation in alpine lake	1-year (2009-2010)	Python Statsmodels	The best working models are ANN and ARIMAX

Continuation of Table 1						
Model Name	Paper	Performance Criteria	feature data-set	time period	environment	conclusion
ARIMA, ANN, LSTM	[23]	P values, graphical observations	Dell's stock price	1 year (2010)	-	ANN model is better than ARIMA model and LSTM, but the hybrid of ARIMA-GARCH model can be used for more accuracy.
ANN, LSTM	[37]	nRMSE, MAPE, R2	Prediction of Solar irradiance	Hourly	Google Colab, Python scikit-learn, Keras	Due to small dataset, the ANN model works better than the LSTM, but both works much better than the persistence model

Continuation of Table 1						
Model Name	Paper	Performance Criteria	feature data-set	time period	environment	conclusion
ARIMA, GARCH, regime-switching	[8]	RMSE	Home price indices by the OFHEO	20-years (1980-2000)	-	Regime switching models perform better than ARIMA and GARCH
ARIMA-ANN, ARIMA-Kalman	[22]	MAE, MAPE, MSE	Wind speed prediction	-	-	Both hybrid models have good forecasting accuracy and suitable for wind samplings.
ARIMA-SVM, ARIMA-ANN, ARIMA-Random Forest	[21]	-	indian stock trend	5-years (2004-2009)	MATLAB6.1, SPSS13.0	The hybrid model ANN-ARIMA, was able to predict great values than other models

Continuation of Table 1						
Model Name	Paper	Performance Criteria	feature data-set	time period	environment	conclusion
AR, ARDL, KNN, SVR, Naïve, VAR, MLP	[41]	RMSE, R2	Inflation forecasting	30- years (1984- 2014)	-	SVR and ARDL outperforms other models and machine learning models work best with more volatile and irregular series.
Hybrid ARIMA models	[9]	-	Canadian lynx time Series, sunspot time series, airline and star data	Different time periods		The hybrid system leads to a higher accuracy in prediction
ANN, KNN	[26]	-	Recorded EEG signals	Patient data set	-	ANN classifier's accuracy and sensitivity was higher than that of KNN classifier.

Continuation of Table 1						
Model Name	Paper	Performance Criteria	feature data-set	time period	environment	conclusion
ANN, LSTM, MLR	[20]	MSE, r, RMSE	Prediction of irrigation groundwater quality	-	Python	ANN and MLR model have highest accuracy in multiple scenarios
LR,LSTM	[18]	RMSE, MAPE	For hire vehicles and yellow taxi	-	-	LR is used to select the important variables and LSTM helps to improve the accuracy.
ARIMA	[4]	-	New York Stock exchange and Nigeria stock exchange	-	python	ARIMA has strong potential for short term prediction

Continuation of Table 1						
Model Name	Paper	Performance Criteria	feature data-set	time period	environment	conclusion
MLR, KNN, ANN, ANFIS	[17]	Mash-Sutcliff Coefficient	Stream flow prediction	Monthly		The accuracy of each of the model depends on the condition, but the hybridisation was effective.
End of Table						

2.2 Common methodologies used

After observing the table above, it can be clearly seen that the ARIMA is not a good option for the analysis of financial time series even though it is one of the most common methods used. Multiple machine learning methods like LSTM and ANN are also most commonly used as well as have good prediction accuracy as well. hybrid models have the highest accuracy among the machine learning models. The table below enlists the popular models from the journals and books published.

Table 2: Methodologies used in the reviewed literature.

Model Name	Total Papers	Paper
ARIMA	5	[27], [32], [23], [8], [4]
ANN	7	[11], [38], [23], [37], [26], [20], [17]
LSTM	5	[32], [23], [37], [20], [18]
LR	2	[27], [18]

Continuation of Table 2		
Model Name	Total Papers	Paper
SVM	3	[27], [11], [2]
KNN	4	[2], [41], [26], [17]
Hybrid Models	3	[22], [21], [9]
others	9	[27], [2], [38], [8], [22], [21], [41], [20], [17]
End of Table		

These are the most commonly used algorithms that have been reviewed and used for the analysis of the financial time series.

2.3 Gaps in current literature

Financial time series is a very popular and complex branch of time series analysis. The machine learning techniques have taken the branch to new levels. The aim of the paper was to answer what machine learning techniques that are being used for financial time series analysis, finding the best one and what void still needs to be filled. The common algorithms that are used for the financial time series analysis are ARIMA, ANN, LR, LSTM, SVM, KNN, MLP, decision tree, random forest and the hybridisation of these. Even though the study is not just limited to these machine learning algorithms, but these are the most commonly used ones due to their higher accuracy than the other algorithms or their ease of implementation. Moreover, the accuracy of each of the model also varies, depending on the data set as well, for example if the data set is huge, LSTM will work better. All the models have their own pros and cons, but after review of number of papers, hybrid models outperform the basic machine learning models. The gap that needs to be filled for the betterment of the topic is to study and develop more machine learning algorithms, whether it be by creating new ones, or creating new hybrids of existing models for higher accuracy, thus a great opportunity for the researchers in this field. The next part of the dissertation will consist of the development of the artefact, it's evaluation and future work regarding the topic.

3 Research design and methodology

The main objective of the project is to model financial time series with machine learning algorithms available. The first part of the project is done with the help of existing machine learning algorithms available and working on them. each of the model tested will be evaluated with RMSE, R2, and mean absolute error. Since, as mentioned in the literature review that each of the model has its own pros and cons, it is a better idea to compare the models in the first phase of the research. For the research I have chosen six models, which will be compared to find out the best. the benefit of doing this would come in handy when hybridising them with the other models. The second phase of the research will be to hybridise them and work on the client portfolios.



Figure 1: Development Phases for the Project

3.1 Some common Time series evaluating techniques

This part of the literature review will give succinct details about some of the most popular machine learning models used for the analysis of financial time series.

- **ARIMA:** ARIMA model is basically the integration of AR (Auto regressive) and MA (Moving Average) and is capable of working with the non-stationary data. It is also referred to as box Jenkins models as it was popularized by George Box and Gwilym Jenkins [25]. It can be represented mathematically as shown below
$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \epsilon_t$$
where y_t stands for the goal variable, with which the values of y_{t+1} , y_{t+2} , and so on can be determined [15]. It is one of the most common statistical methods used for financial time series analysis as from the reviewed literature five of the papers were on ARIMA.
- **ANN:** ANN or Artificial Neural Network are inspired from the human brain's neurological functions and made in such a manner that it replicates its decisions similar to humans and can be created by programming computer to behave like neurons. Mathematically, it can be represented simply as
$$h(\Theta(x)) = 1 / (1 + e^{-\Theta(x)})$$
where $h(\Theta(x))$ is the output, x is the input and Θ is the parameter vector [24].
- **LSTM:** LSTM or long short-term memory belongs to recurrent neural network architecture and consists of memory cells that store information which can be updated from time to time by input, output and forget gate [7].
- **Hybrid models:** Hybrid models are the combination of two or more algorithms, and this is a major approach towards more accurate and reliable methods, because it benefits from the two methods and thus reach higher performance. The examples of the hybrid models are DTFNN, ARIMA-ANN, etc. [3].

3.2 Experimentation design

For the analysis of financial time series, the project was divided into number of phases, that has been followed in order to create an optimal research thesis. The steps taken are as follow

- Read and review a number of literature, articles, book, journals etc to find out the current approach and gap in the analysis of financial time series.
- Finding and researching about different machine learning algorithms that can be used to fix the gap in the current learning.
- Create a number of machine learning models and evaluate them and find out if they cover the gaps or not.

For the project, ten years of the Nifty-50 data from 2011-2020 has been chosen [29]. The whole project works on the previous close value which will be extremely helpful in finding out if the valuation is performing according to the curve or not. The statistical models have been chosen for this part of the project and are down as follow

- Auto Regressive integrated moving average with Exogenous variables (ARIMAX)
- Seasonal Auto Regressive integrated moving average with Exogenous variables (SARIMAX)
- Artificial Neural Network (ANN)
- Convolutional Neural Network (CNN)
- Long Short Term Memory (LSTM)
- Gated Recurrent Unit (GRU)
- Bidirectional-LSTM
- Bidirectional-GRU
- Hybridized LSTM+CNN
- Hybridized LSTM+GRU

All of these models have been tested and compared against each other and evaluated. this is the basis on what the whole research design has been created.

3.3 Evaluating techniques used in the paper

For evaluating the financial time series, I have used three most commonly used methods, i.e., mean absolute error, root mean squared error, and R2 score each of which I will be discussing succinctly so as to understand what the results mean.

- Mean absolute error (MAE): the magnitude of difference between prediction and the actual value is mean absolute error. It can be easily understood quantifiable measurement of errors [5]. Even though it can range from 0 to infinity, Lower the value of mean absolute error is, the better the model is working. The mathematical formula for mean absolute error is $MAE = (\sum_{i=1}^n |\text{predicted value} - \text{actual value}|) / n$ Where n is the total number of data points.
- Root mean squared error (RMSE): root mean squared error is, as the name suggests, basically the square root of the mean of the square of all the error. It is extremely helpful to have a single number to judge a model's performance [6]. Same as MAE, RMSE can range from 0 to infinity and lower the value, the better the model performs. $RMSE = \sqrt{(\sum_{i=1}^n (\text{predicted value} - \text{actual value})^2) / n}$ Where n is the total number of data points.
- R2 score: R2 is the proportion of variance in the dependent variable that can be explained by the independent variable meaning it is a measure how well the model fits the data. The values range from 0 to 100 and higher the value is, the better the model fits. If the score is 100 percent, it means the data set is perfectly correlated [12]. And if the score is negative or bigger than 100, then it means that the model does not fit the prediction and should be rejected. The mathematical formula is $R^2 = 1 - (\text{predicted value} / \text{initial value}) \times 100$

3.4 Goals of the research

The goals that this research report needs to answer are:

- Find the current gaps by review the existing literature.
- Design a solution to fix the gaps in the existing literature.
- Test the solution proposed.
- Evaluate and compare the models and find out the best one.

4 Artefact Development

The artefact is divided in two parts and the first part is to review literature, find gaps in it, and then create machine learning models, and evaluate them. This will be helpful in finding out what machine learning algorithm works most accurately and what next steps of the project could be. The second part of the model will be more focused in creating deep learning models and Hybrid models. The main focus is to find out which model outshines the other one and what is the ideal period of time for each of the algorithm.

4.1 Data set and Data pre-processing

The data set required for the analysis of financial time series analysis should be related to something finance or money related. This could have included the stock market prices, gold prices, crypto currency evaluation etc. For this project I have chosen Nifty-50 stock data from 2011 to 2020 and divided them in three parts for the testing and prediction for 30 percent towards the end of the series. The data consists of top-50 stocks from the national stock exchange, India. All of the data is of format comma separated values, which is ideal format for machine learning and is stored in the local file, but later on during the second iteration of the research project, the data could be shifted to more reliable cloud database like google drive.

Each of the CSV file contains a number of heading but only one of the headings are most important for the project, that is, Previous close value. The index of the data set is date starting from 01/01/2011 to 31/12/2020 with the frequency of Business days, since the stock market is only opened on business days. Furthermore, there are a number of Null values, which have been fixed by assuming the value of close value from last observation.

Below are the three stocks that have been chosen as data set for training and testing, with a small description about each.

- **HDFC:** HDFC or Housing Development Finance Corporation Limited is an Indian financial services company's stock. it was founded in 1977 and is one of the top performing stock in Nifty-50. This company also has presence in banking, life and general insurance, asset management, education etc. the company is evaluated at 19 Billion U.S. dollars. [14]
- **Infosys:** Infosys limited is an Indian multinational IT company. Infosys has 16.3

Billion U.S. dollars in revenue and was founded in 1981. The highest annual growth for the stock was seen in the last decade with 19.7 percent. [16]

- Hindustan Unilever: Hindustan Unilever is a consumer goods company, and is a subsidiary of a British company Unilever. It was founded in 1993 and has a revenue of 5.4 billion U.S. dollars. It has many subsidiaries, some which are also enlisted in National Stock Exchange, India. [35]

4.2 Description of models used

For the project, i have used ten learning models, and here is a description of each below.

- Auto Regressive integrated moving average with Exogenous variables (ARIMAX): This model performs to predict the future of a time series with an auto-regression model. What makes it different is that it takes external factors into consideration while making a prediction. This uses same libraries as the ARIMA model but while fitting the model, we need to declare the exogenous factors along with training data set and the order of p, d and q. [28]
- Seasonal Auto Regressive integrated moving average with Exogenous variables (SARIMAX): it is just similar model like the ARIMAX except for the fact that it takes Seasonality into consideration while making a prediction. While fitting the model we need to tell the exogenous values along with training data set, order of p, d and q and seasonal order. [36]
- Artificial Neural Network (ANN): ANN is the most basic model in deep learning and it mimics the way the human brain cells work. ANN is consisted of layers interconnected, which sends data to deeper layers, which in turn sends the final output to the last output layer, meaning ANN takes data samples rather than entire data sets to arrive at solutions. [1]
- Convolutional Neural Network (CNN): Convolutional neural networks or CNN have one or more Convolutional layers, and are mainly used for image processing, classification. It requires a very little pre processed data and it offers dilated convolutions, in which filters can be used to compute dilations between cells. [40]

- Long Short Term Memory (LSTM): LSTM is a one of the type of Recurrent Neural network or RNN, and have the capability of long term dependencies, especially in prediction problems. Unlike the standard feed-forward neural networks, LSTM has feedback connections and its structure is based on the foundation of short term memory processes to longer term memory. [39]
- Gated Recurrent Unit (GRU): The gated recurrent unit is a type of RNN and may or may not have advantage over LSTM in very specific cases. GRU uses less memory and is faster than LSTM. While performing the experiments, I could actually see the difference in time between the two. GRU uses two gates namely the update gate and the reset gate, and these gates decide what information is allowed to pass to output. [33]
- Bidirectional-LSTM: Bi-LSTM is the process of making the LSTM to have a sequence information in both the directions, that is backwards and future. This is basically an upgrade on the LSTM. The final output is a combination of the outputs of the forwards and backwards directions. [31]
- Bidirectional-GRU: Bi-GRU is the process of making the GRU to split the normal GRU into two directions, one forwards with historical data and one backward with future data, and thus give a final output with [10]
- Hybridized LSTM+CNN: This is the hybridized or combination model of LSTM and CNN. For this particular mode, LSTM was stacked on top of the CNN to predict the outcome.
- Hybridized LSTM+GRU: This is the hybridized or combination model of LSTM and GRU. For this particular mode, LSTM was stacked on top of the GRU to predict the outcome.

All of these models have been tested and compared against each other and evaluated. This will be covered in the next part of the thesis. I have chosen Adamax as the optimizer for the deep learning models as it replaces second order moment with the infinite order moment. The update rule is $u(t) = \max(b(2) u(t-1), |g(t)|)$ $g(t)$ will be completely ignored when it is close to zero and $u(t)$ will be influenced by fewer gradients and this makes the algorithm more robust to noise in the gradients. [19]

5 Results

There are ten models in total, as defined above, each of them have been compared against each other. The training and testing between the data has been split in the ratio of 70 and 30, meaning 70 percent of the data is trained to predict the 30 percent of the data. There are three time frames which are short term, 1-year, mid term 5-years and long term 10-years. To keep things simple, the graphs are combined based on the type they belong to, for example there are four deep learning algorithms and these four have been combined in one. Let's look at the next part of the thesis where the graphs of the algorithms are compared against each other.

5.1 Comparing resulting graphs for test data and the predicted data for different stocks

There are three stocks in total and all the algorithms' predicted and actual graphs will be compared for the results, and find what model closely follows the trend of the actual graph.

- **HDFC:** HDFC is a financial corporation stock from the Nifty 50, National Stock exchange, India.

1-year (Short-term prediction)

For the statistical models, ARIMAX and SARIMAX, these two closely follow each other as can be seen in graph. They are both following the similar trend as actual. What's interesting to see over here is that these both overlap each other and thus it is difficult to see which one is performing better through graph, and if there is any difference between them, it will be seen while evaluating them.

The next graph is for four of the deep learning models. By looking at the graph, it can be seen all the models follow the basic trend of going upwards and following the actual trend. The trend is most closely followed ANN and CNN as can be seen from the figure below. The most deviation can be seen from LSTM model. Further results would be clarified with the evaluation and thus it would be found out which one performs the best.

There are two Bidirectional algorithms that have been used in the paper, and these are Bi-LSTM and Bi-GRU. It can be seen from the graph that there is a very high deviation in the graph. The higher deviation can be seen with Bi-LSTM, making Bi-GRU a better performing model. Further evaluation will

reveal which performs better, and it will be covered in the next part of the thesis. The next results are for the Hybrid models of LSTM-CNN and LSTM-GRU. Both the models follow the trend of the actual figures, but it can be clearly seen that there is a higher deviation in the LSTM-GRU model than LSTM-CNN model and has a negative deviation meaning it goes below the actual trend. Thus it means that graph shows that LSTM-CNN model performs better than LSTM-GRU model, but further evaluation will reveal the actual performance. Overall, for 1-year, the best performing results that follow the trend most closely are Statistical and deep learning models. by looking at the graphs, it looks like the deviation for them is least and and would be beneficial for prediction of 1-year of a stock based for a financial company. The Bidirectional models have the highest deviation, meaning it is not advisable to use bidirectional algorithm for short term prediction for a stock for a financial company.

HDFC 1 year graphical results

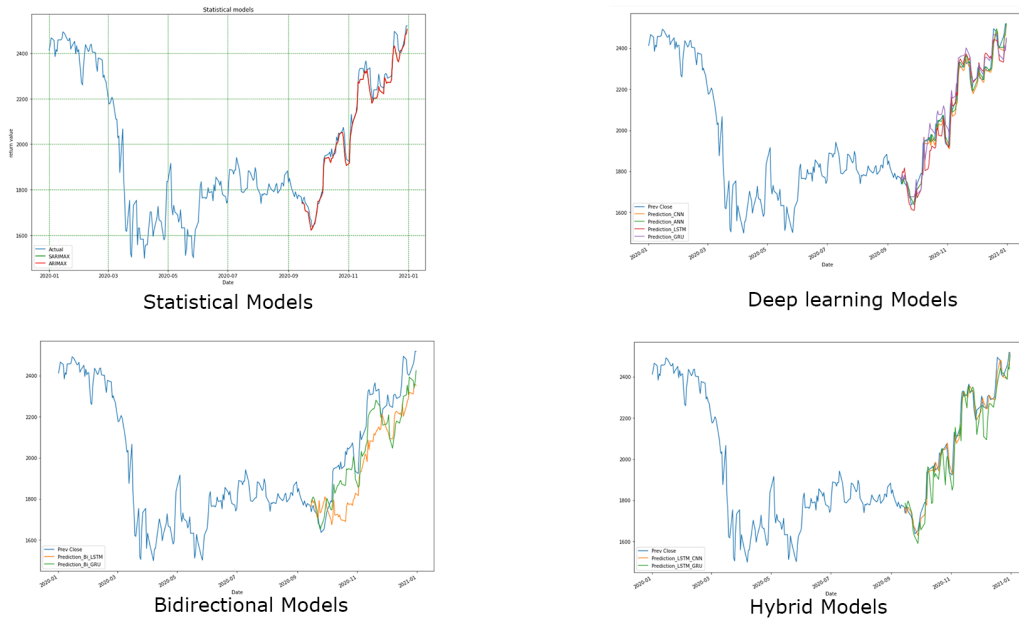


Figure 2: HDFC 1-year graphical results

5-years (Mid-term predictions)

For the statistical models, here the results start to get interesting. Even though the SARIMAX model follows the actual trend closely, the ARIMAX model shows extreme deviation. This just shows that with the addition of a seasonality, the whole performance of the model can change. It is clear from the graph that

SARIMAX is performing better than ARIMAX.

While getting the results for the deep learning models, the graph was getting extremely congested, but if the graph is zoomed in, it can be seen that CNN follows the trend most closely followed by ANN, and then LSTM and GRU. This shows that CNN is a better performing algorithm than others for mid term prediction. Further evaluation will reveal the actual results for it.

The next graph is for the Bidirectional models, which in theory should perform better than the normal deep learning algorithms. Both Bi-LSTM and Bi-GRU follows the actual trend very closely, but there are a few deviations, which interestingly, both the algorithms show. On further examination of the graph, it can be seen that Bi-GRU performs better than Bi-LSTM, but it is not a very big difference, but a very minor one. Further evaluation will determine, which one performs better.

The next results are for the Hybrid models. The two models, LSTM-CNN and LSTM-GRU follow the actual trend closely, and can be considered as good algorithms to use if predicting for mid term. As can be seen from the graph, the LSTM-GRU shows more deviations than LSTM-CNN, meaning graphically LSTM-CNN is better performer than LSTM-GRU, but further evaluation can verify the results.

Overall, it can be seen the best working models for the mid term predictions would be Hybrid models, followed by Deep learning, bidirectional and statistical models. It can be seen that as the dataset is increasing the, prediction accuracy for deep learning, bidirectional and hybrid models are increasing and the accuracy for statistical models is decreasing, more specifically for ARIMAX model. It will be interesting to see how the results come for the 10-year period and as well as how evaluation will determine the best performing models.

HDFC 5 years graphical results

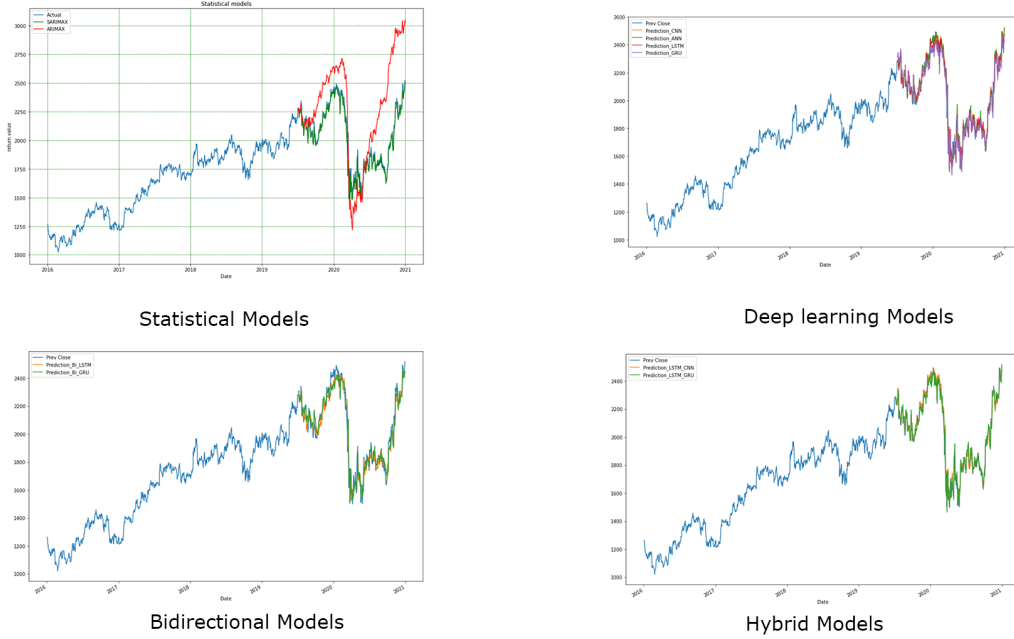


Figure 3: HDFC 5-years graphical results

10-years (Long-term predictions)

Firstly for the Statistical model, the SARIMAX model is performing well and is following the actual trend quite closely. The ARIMAX model is showing even higher deviation than before, which directly means that bigger the data set, less accurate the ARIMAX model would be, even though it will show the general trend, but it is clear that it is not a reliable model to use when the data sets are bigger. Further evaluation will show similar results.

The next turn is for the Deep learning models. On looking at the graph it can be seen that all the algorithms perform equally good and follow the general actual trend of the stock, but when the graph is zoomed in it can be seen that there is a higher deviation in LSTM and GRU as compared to CNN and ANN, meaning CNN and ANN are performing better than LSTM and GRU. Further evaluation is required so as to find out the best one as there is a very minor difference between each of them.

The next part is for the Bidirectional models, both Bi-LSTM and Bi-GRU are following the actual trend, even though there is a little bit of deviation here and there. By examining the graph, it looks like Bi-LSTM is more aggressive than Bi-GRU, meaning Bi-LSTM is showing more deviations. This makes Bi-

GRU a better performing model than Bi-LSTM, but the trend is so close that it is extremely hard to find which one performs better. Further evaluation will determine, which one is better.

Now the results for the hybrid models will be discussed. It can be clearly seen from the graph that both the models are following the actual trend values. On a closer look, it can be seen that the deviation for LSTM-GRU is higher than LSTM-CNN, meaning the LSTM-CNN performs better than LSTM-GRU. Further evaluation will determine if it is true.

Overall the best working models belong from deep learning part and these were closely following the trend with most minimal deviations from the actual trend. Further evaluation is required to determine which model performs best, but it is clear that ARIMAX is the worst choice for the prediction.

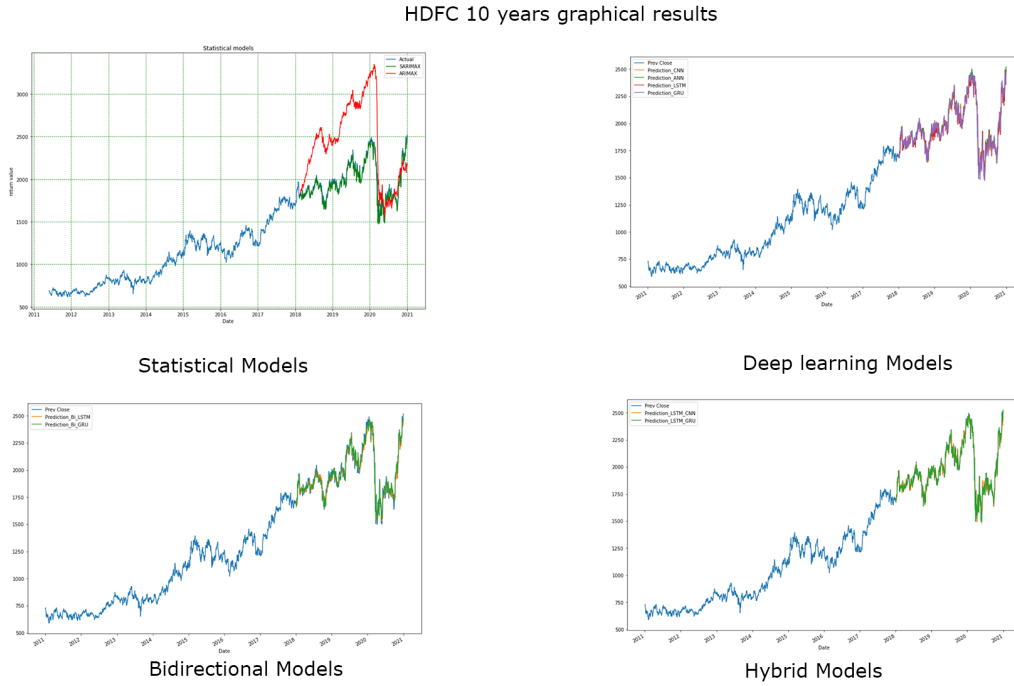


Figure 4: HDFC 10-years graphical results

- Infosys: Infosys is a stock for an information technology company.
1-year or short term

As was the case with HDFC, it was similar with Infosys. The ARIMAX and SARIMAX model were following the trend very closely. If there was any difference it could hardly be seen by the eyes. Theoretically, SARIMAX should

be performing better than ARIMAX but further evaluation was required so as to find out which model was performing the best.

The next results are for the four deep learning models. It was interesting to see that the predictions were following the basic trend, but the deviation was increasing as it went further. The actual trend was most closely followed by CNN, then ANN, then with the highest deviation with LSTM and GRU. This should be the expected results after looking at the graph, and further evaluation is required.

Now it is turn for the Bidirectional models. These showed very interesting results as the the basic trend was followed but with different deviations. The Bi-LSTM showed a negative deviations, meaning it was below the actual trend line and Bi-GRU showed positive deviation, meaning it was above the actual trend line. It was difficult to determine which performed better by actually looking at the graphs and further evaluation is required.

By examining the graphs for the Hybrid models, it can be seen that the both the models are following the actual trend. On a further examination, it can be seen the the deviation for LSTM-GRU was higher than LSTM-CNN, meaning LSTM-CNN performed better than the other. Further evaluation will determine if it is correct or not.

Overall, the best working models for this scenario belongs to Statistical and deep learning methods. As was the case with HDFC, it can be seen similar results with Infosys and the Hybrid and Bidirectional models, do not perform well with short term dataset.

Infosys 1 year graphical results

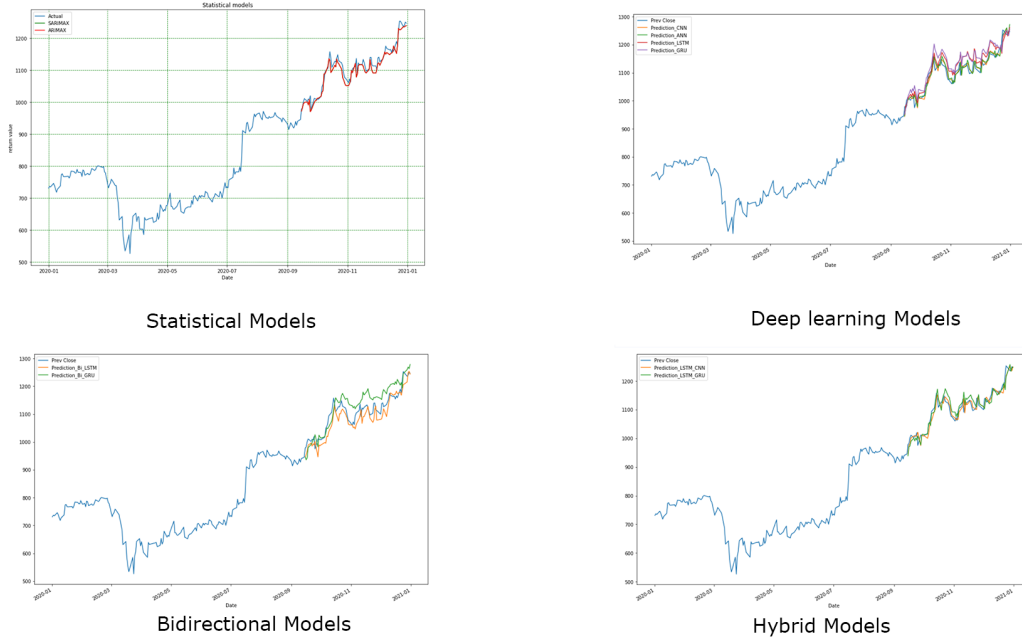


Figure 5: Infosys 1-year graphical results

5-years or mid term

The graph clearly indicates that the dataset has been increased. The SARIMAX model is closely following the actual trend, but the ARIMAX model shows a very extreme and high deviation. It falls down extremely and then increases rapidly, meaning this model is extremely unreliable and should not be used for mid-term predictions.

The deep learning models very closely follow the actual trend and overlaps one another. It is extremely hard to determine which is the best one with just a look of it and needs further evaluation. There is a little bit of deviation that can be seen and there would be a very little differences in each.

The results for the Infosys bidirectional models are interesting as at the first look, it is seen that the models are following the trends closely, but the Bi-LSTM model is more aggressive in showing deviations than Bi-GRU and thus Bi-GRU should perform better than Bi-LSTM. This will be seen while evaluating the models.

Theoretically, the hybrid models should be the best performing out of all the models. Both LSTM-CNN and LSTM-GRU closely follow the actual trend. On a closer examination, it can be seen the LSTM-GRU deviates from the trend

multiple times, which would make LSTM-CNN a better performing model than LSTM-GRU. Further evaluation will determine if these results are correct. Overall, the ARIMAX is the least favourable model for Infosys and Hybrid model and specifically LSTM-CNN is the most favourable. The deep learning models are difficult to analyse from the graphs and require further evaluation and Bidirectional models have a very high chance of deviation, meaning further evaluation will clear the results.

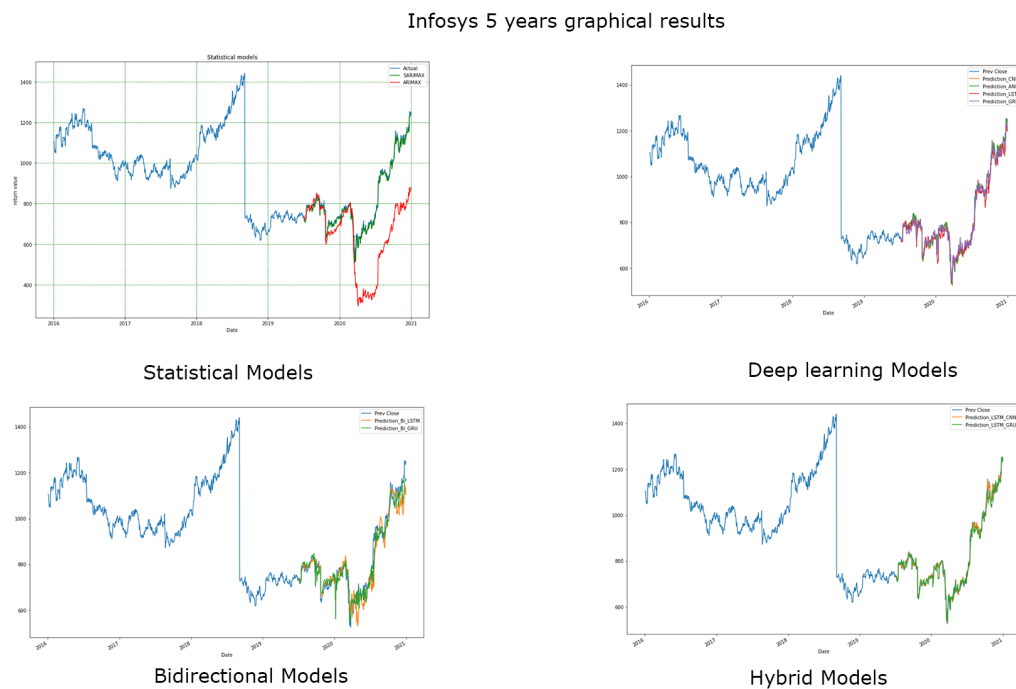


Figure 6: Infosys 5-years graphical results

10-years or long term predictions.

The results for the 10-years prediction for Statistical model was somewhat expected by looking at the trend, but it was interesting to see that ARIMAX was following the trend as well, except for the fact if the value of the stock was expected to fall, the value predicted for the stock fell more than the actual value. SARIMAX was following the trend closely, and out of these two statistical model, SARIMAX would be preferable.

Overall, the deep learning models give a great predictions, and predict similar values as the actual values with minor differences. On a closer examination, it can be seen that GRU was giving a little more deviation than others and thus making it the least preferable of the all four. Further evaluation will reveal which

one is the best.

The bidirectional models theoretically should perform better than the regular deep learning models. Both Bi-LSTM and Bi-GRU are following the actual trend as well as giving deviations from the trend. What was interesting to see here was when there was a positive deviation for one model, there was negative deviation for the other one and vice versa. Further evaluation is needed to find out the better performing model.

The hybrid models were showing the expected results and following the trend closely. Further evaluation will be needed to find out which one is performing better, but overall these models can be reliable while analysis of long term models.

Overall, it was found out that the most unreliable model would be ARIMAX, followed by the bidirectional models as they showed a very high deviation. Deep learning and Hybrid models were the most optimum and they were following the trend as close as possible, but further evaluation is needed to find out which one is the best among them.

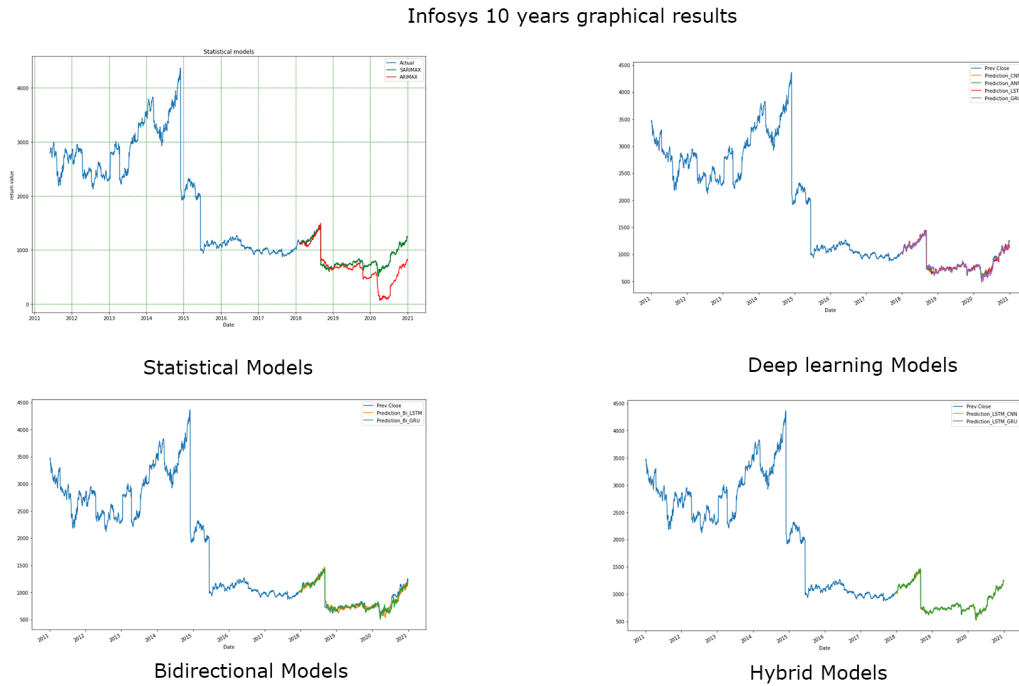


Figure 7: Infosys 10-years graphical results

- Hindustan Unilever: Hindustan Unilever is a consumer goods company.
1-year or short term

For short term prediction for statistical model, it was expected that the ARIMAX and SARIMAX model would perform equally well and almost trace the actual data set. Further evaluation will show that which one is performing better, but for short term this model can be used.

For the deep learning models, there were four of them as mentioned above. At first look, it looks like the models it performing well and following the trend. On a closer observation it can be seen that LSTM and GRU are showing a lot of deviations and least deviation is shown by CNN, making CNN the best among the deep learning models. Further evaluation will be used to compare it with other sets of models.

For the Bidirectional models, the results were not as good. There is a very High deviation among them and thus it is not a good idea to use them for shor term predictions. In comparison, by looking at the graph, it can be seen that Bi-GRU performed better than Bi-LSTM as there were higher deviations for Bi-LSTM. Further evaluation will determine if this model can be used for financial time series analysis.

For the Hybrid models, it was expected to see the best results of the lot, but still it was not the best for short term predictions. While looking at the graph and analysing it, it can be seen that LSTM-GRU is showing deviations, while LSTM-CNN is performing better and closely following the trend. it is clear that LSTM-CNN is a better model, which will be shown by further evaluation.

Overall, the Statistical and deep learning models are performing well. Out of those, CNN looked like the best performing model. It was expected that Hybrid and Bidirectional models should perform well, but there was a very high deviation for the Bidirectional models and Hybrid models. Further evaluation is needed to verify the results.

Hindustan Unilever 1 year graphical results

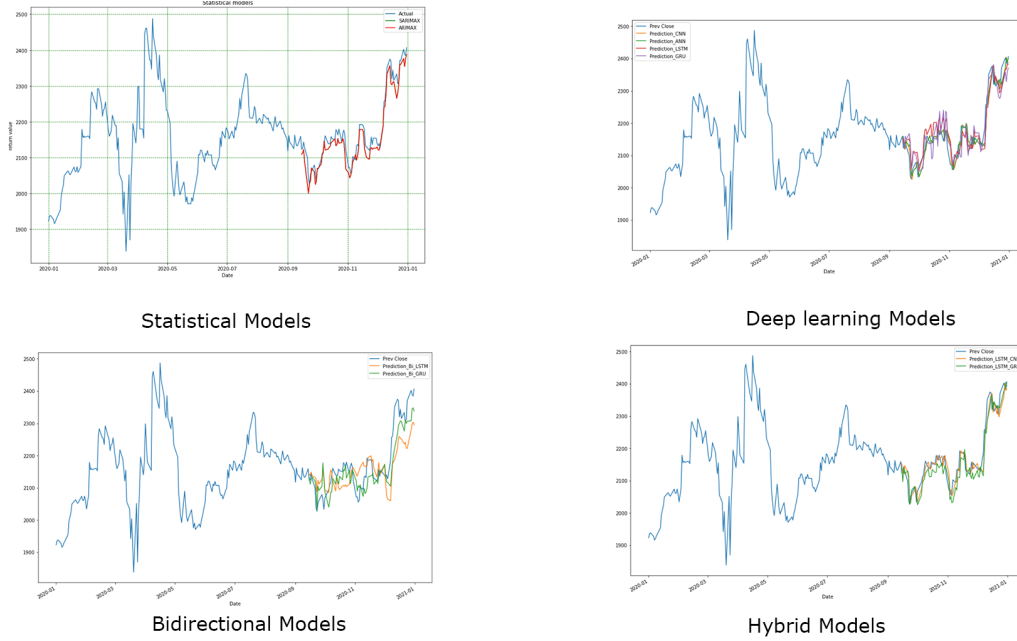


Figure 8: Hindustan Unilever 1-year graphical results

5-years or mid term

As it was expected for the statistical models, ARIMAX showed very high deviation and thus it is clear that as the data set increases, the reliability on ARIMAX decreases. Just with the addition of seasonality, the model becomes a lot reliable and starts giving a proper results. It is clear that the ARIMAX for mid term dataset should be discarded, which will be seen in the further evaluation.

The Deep learning models are following the trends very closely and thus it can be regarded as a reliable method. On further examination it can be seen that the models are following the basic actual trend, but on a further inspection it can be seen all the models show some deviations and it is extremely difficult to find the best one out of the four. Further evaluation is required so as to find which one works best.

Once the size of the dataset starts to increase, it can be seen that bidirectional models start to come together. Both Bi-LSTM and Bi-GRU are following the trend and can be regarded as reliable models for mid term prediction. On a closer inspection, it can be seen that Bi-GRU is a little more aggressive, making Bi-LSTM a better model. Further evaluation is required to find out what difference

between each of them is and how much accurate these are.

Overall for the Hindustan Unilever, the performance can be seen as extremely well and both LSTM-CNN and LSTM-GRU are following the actual trend.

Since it is extremely close and following the actual trend of the stock. Further evaluation is required to find out the which one is performing better.

For the Hindustan Unilever mid term prediction, Statistical models were the least reliable especially the ARIMAX model. The bidirectional models started to look like it is performing better for mid term than for short term. The last two which are the Deep learning and hybrid models are very congested and require further evaluation, there would be just minor differences.

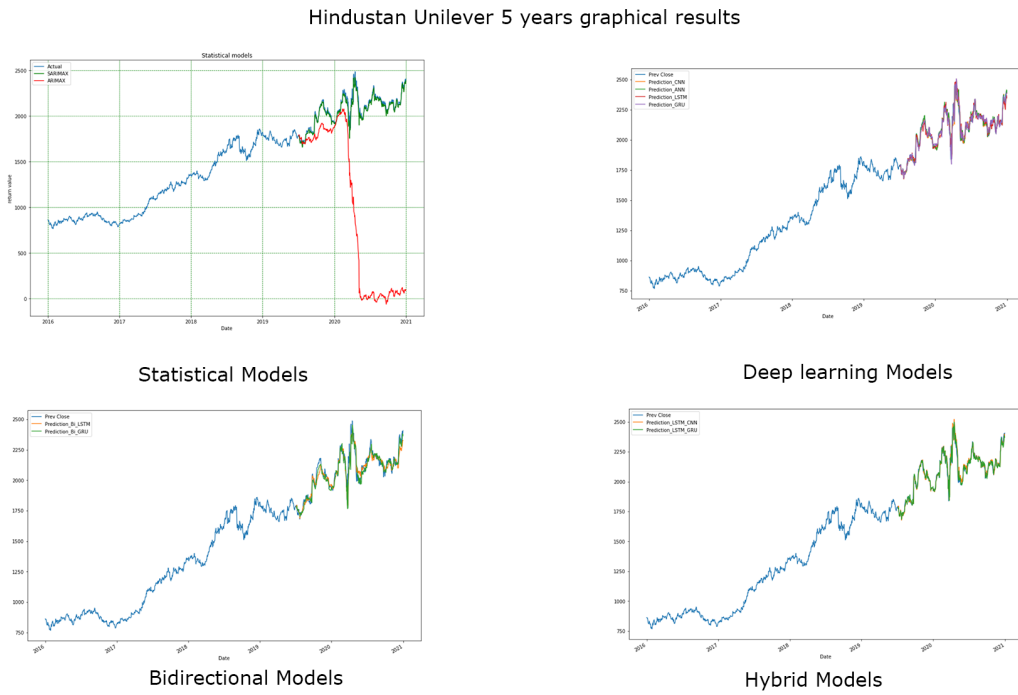


Figure 9: Hindustan Unilever 5-years graphical results

10-years or long term

The model's accuracy and prediction was expected earlier as it has already been established that as the dataset increases, the ARIMAX model starts performing worse and worse. What's interesting to see here was that for mid term prediction, ARIMAX showed negative deviation, but for long term it is showing it positive deviation. SARIMAX model was closely following the actual trend and thus makes it a reliable model as compared to ARIMAX.

Overall, by looking at the graph it can be seen that all the deeplearning models

are following the trend and will be a very reliable model for it. By looking at different colors, it can be seen a very minimal deviation, but it is extremely difficult to find out the best model just by looking at the graph and requires further evaluation.

By looking at the Bidirectional models, which are Bi-LSTM and Bi-GRU, it can be seen that both the models are performing well. Since this was a long term prediction, it was already expected to follow the actual trend line. While observing both, it can be seen it is difficult to find out the better one and further evaluation is required to find the best model among them.

When analysing the graphs for Hybrid models, it can be seen that both LSTM-CNN and LSTM-GRU perform equally well and can be extremely reliable for the prediction of time series. Different colors in the graph shows that the models are deviating from each other, but it is extremely difficult to find out the best among them and further evaluation is required.

For the ten years prediction for Hindustan Unilever, it was extremely difficult to analyse which would be best just by looking at the graph. It was clear that there was some deviations in the model, but it was difficult to predict and find the best one among all of them. Further evaluation, which would be done in the next part of the thesis, will be seen which is performing best. One thing was clear that ARIMAX is unreliable for the long term prediction and should be used only for short term predictions.

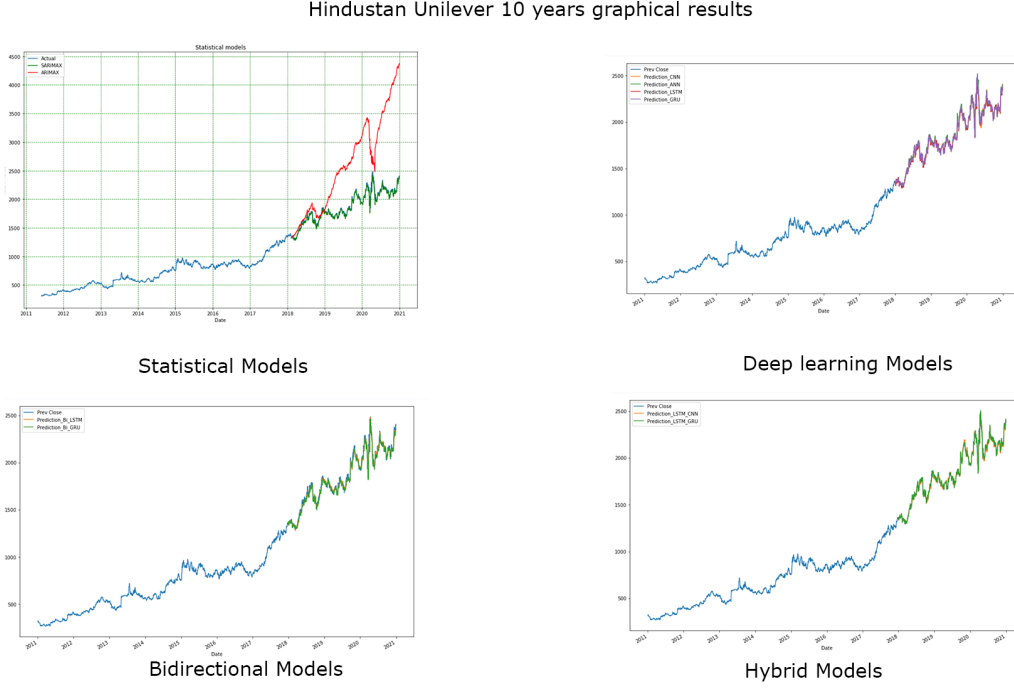


Figure 10: Hindustan Unilever 10-years graphical results

It was interesting to see the results of how each of the model performs with different datasets. For the Short term period, the Statistical models were showing great results, but as the datasets increased, their reliability decreased. Furthermore it was interesting to see the similar thing with Hybrid and Bidirectional models, where lower the dataset was lower its prediction accuracy was. The most complex results were for Deep learning models as they were congested and closely following the trend. Overall, the graphs just show how each of the models were performing in a big picture but further evaluation is required to find out which method should be applicable with what size of data which will be done in the next step.

6 Evaluation

There are a number of ways to evaluate the performance of machine learning models like accuracy, confusion matrix, F1 score, etc. But those methods cannot be used to evaluate the time series models, but can be evaluated by methods like mean absolute error, root mean squared error, etc.

6.1 evaluating stocks and the algorithms used

As mentioned before, this paper consists of three stocks, ten machine learning algorithms and three time frames for datasets and each of them will be evaluated using the techniques discussed above.

- **HDFC:** First let's start with the one year or short term evaluation. It was interesting to see that ARIMAX model surpasses all other models when evaluating. Since this is a statistical model, it was not expected to show such results but overall it is the best working model. The worst model of the lot is Bi-LSTM which is very high RMSE, MAE and R2 score. While evaluating, it was interesting to see that ANN was performing better than CNN, LSTM or GRU, and Bidirectional models were not reliable at all. Further more it was interesting to see that hybrid model for LSTM-CNN performed better than both LSTM and CNN, but the hybrid model for LSTM-GRU performed worse than both LSTM and GRU.

The table below is for the evaluation of the mid term values for the HDFC stock. For the ARIMAX, it was earlier seen that it was the best model but now it is the worst model, where the value for MAE, RMSE are extremely high and the value for the R2 was out of bounds, meaning the model should be discarded.

What was interesting to see was that SARIMAX was showing the best values as a statistical model. Now the results for the deep learning models looks somewhat that has been mentioned in a number of papers, but still it was interesting to see ANN working better than LSTM. The Bidirectional and the Hybrid models were still not the best choice, with the exception of LSTM-CNN which showed better results than both LSTM and CNN.

For the long term evaluation, it can be seen that ARIMAX was obviously the worst model with extremely high RMSE and MAE, and out of bounds R2 score. Surprisingly, just with the addition of seasonality, the SARIMAX model performs extremely well. Furthermore, it was interesting to see that LSTM was performing better than Bi-LSTM, while GRU was performing worse than Bi-GRU. Finally Hybrid models performed better than the deep learning models. It shows that the bigger the data set, the more hybrid models will perform well.

Table 3: HDFC stock evaluation

	Short term			Mid term			Long term		
	MAE	RMSE	R2 score	MAE	RMSE	R2 score	MAE	RMSE	R2 score
ARIMAX	26.420	34.473	98.158%	294.900	369.842	value out of bounds	484.01	568.858	value out of bounds
SARIMAX	25.441	34.497	98.157%	32.097	47.516	96.913%	25.972	38.676	96.900%
ANN	31.861	42.957	97.108%	37.443	53.540	96.074%	28.324	41.004	96.484%
CNN	35.110	47.171	96.513%	36.024	52.417	96.237%	29.758	42.552	96.214%
LSTM	48.970	61.953	93.986%	42.333	57.347	95.496%	95.132	48.253	95.132%
GRU	45.003	56.373	95.021%	47.831	63.014	94.562%	32.923	45.670	95.639%
Bi-LSTM	158.69	178.180	49.971%	50.874	64.82	94.288%	40.152	52.540	94.228%
Bi-GRU	91.400	105.813	82.458%	47.098	62.636	94.627%	36.645	49.006	94.979%
LSTM- CNN	32.584	44.175	96.942%	35.268	51.668	96.344%	29.692	42.192	96.278%
LSTM- GRU	54.121	44.175	91.632%	40.004	51.668	95.784%	30.356	42.192	96.050%
End of Table									

- Infosys: The evaluation for Infosys will be done in a similar way. it will slowly ascend from short term to long term evaluation. It was interesting to note that both SARIMAX and ARIMAX model performed similar to each other but SARIMAX model surpassed all other models during the evaluation. Since SARIMAX is a statistical model and we also have Bi-directional and hybrid models, it was not expected to show much results, but it surprisingly turned out to be the best working model in short term. The worst model of the group is Bi-GRU with the highest MAE, RMSE AND R2 score. It is surprising to see that both the Bi-directional models performed inferior as compared to their individual statistical models and they performed worst among the lot, hence they weren't reliable at all. It was interesting to see that CNN performed better than ANN, LSTM AND GRU Models. Furthermore it was interesting to see that the hybrid model for LSTM-GRU performed better than both LSTM and GRU, but the hybrid model for LSTM-CNN performed better than CNN but worse than LSTM

model.

The table below is for the evaluation of the mid term values for the Infosys stock. For the ARIMAX model, it was earlier seen that it was the second best model but now it's quite surprising to see it is the worst model for mid term evaluation, where the value for MAE, RMSE are extremely high and the value for the R2 score was out of bounds, meaning the model should be discarded. Like before, the SARIMAX model turned out to be the best model even in mid term evaluation. Like before, ANN and CNN showed similar results but this time ANN beat CNN to be the best model among CNN, ANN LSTM AND GRU models. This time too, both the bidirectional models performed worst and clearly indicates that they should be avoided in mid term too. However what's more interesting to see is this time Hybrid models for LSTM- CNN AND LSTM- GRU performed better than the individual statistical models.

For the long term evaluation, it can be seen that ARIMAX was obviously the worst model with extremely high MAE AND RMSE, and out of bounds R2 score. The SARIMAX model turned out to be the best model even in long term evaluation. The ANN and CNN model showed similar results but ANN turned out to be the best among ANN, CNN LSTM and GRU models. Furthermore it was interesting to see that LSTM model performed better than Bi-LSTM and GRU performed better Bi- GRU and both the Bi-directional models performed worst in long-term evaluation and hence indicates that Bi-directional models should be avoided. The hybrid model for LSTM and GRU performed better than both LSTM and GRU models and the hybrid model for LSTM and CNN performed better than both LSTM and CNN models. It clearly shows that bigger the data set, better the hybrid models perform.

Table 4: Infosys stock evaluation

	Short term			Mid term			Long term		
	MAE	RMSE	R2 score	MAE	RMSE	R2 score	MAE	RMSE	R2 score
ARIMAX	20.478	26.223	93.254%	1110.37	1464.35	Value out of bounds	721.407	935.055	Value out of bounds
SARIMAX	20.474	26.219	93.256%	25.916	37.243	94.853%	25.916	30.536	94.853%
ANN	23.549	30.025	90.922%	28.091	42.726	92.992%	22.030	33.732	98.548%

Continuation of Table 4									
	Short term			Mid term			Long term		
	MAE	RMSE	R2 score	MAE	RMSE	R2 score	MAE	RMSE	R2 score
CNN	21.085	27.232	92.532%	26.676	41.309	93.449%	20.973	32.628	98.642%
LSTM	30.958	38.342	85.197%	31.077	44.317	92.461%	25.182	36.854	98.333%
GRU	31.957	39.747	84.093%	29.636	43.805	92.634%	23.260	33.854	98.538%
Bi-LSTM	56.504	68.780	52.367%	37.821	54.203	88.723%	28.299	39.289	98.021%
Bi-GRU	38.789	49.238	75.588%	33.509	51.154	89.956%	28.787	40.709	97.886%
LSTM-CNN	21.737	27.970	92.122%	27.590	41.883	93.266%	21.216	32.304	98.668%
LSTM-GRU	27.648	27.970	88.542%	26.534	41.883	93.691%	23.005	32.304	98.546%
End of Table									

- Hindustan Unilever: As it had been with the Infosys and HDFC, firstly short term predictions will be evaluated then mid term and then finally long term. Overall the worst model for this stock would be the Bi-LSTM as it is having a high MAE, and RMSE and very low R2 score. The interesting thing for this is that both Bidirectional models perform worse than the deep learning counterparts. Further more it is a similar case with the Hybrid models, which shows if they predict the results individually, it will better results. The best working model is SARIMAX, which is the statistical model, meaning the statistical models perform best with short dataset.

For the mid term evaluation, it can be clearly seen from the table that ARIMAX is showing values which are out of bounds and thus this model should be rejected. Further ARIMAX is having extremely high value for MAE and RMSE. It is interesting to see that the SARIMAX model just performs as the best model of the lot. Same as short term predictions, the Bidirectional models are performing poorer than their counterpart. This might be the first model where LSTM-GRU is outperforming LSTM-CNN and is performing better than LSTM and GRU. After SARIMAX, LSTM-GRU is the best performing model, which theoretically as mentioned in the literature review should be correct.

The table for evaluating the Hindustan Unilever for long term predictions is below. It can be clearly seen from the table that ARIMAX was the rejected

model as it is having very high MAE and RMSE, along with an out of bounds value for R2 score. Furthermore, in the whole lot, it can be seen that SARIMAX was also the least preferable of all the models use. The bidirectional models are still not showing better results than their deep learning counter parts, but Hybrid models have performed extremely well in the case. Overall best working model is LSTM-CNN, even better than the normal CNN and LSTM. it was not expected to see ANN performing better than LSTM and GRU, but the results state otherwise.

Table 5: Hindustan Unilever stock evaluation

	Short term			Mid term			Long term		
	MAE	RMSE	R2 score	MAE	RMSE	R2 score	MAE	RMSE	R2 score
ARIMAX	13.240	17.687	93.285%	188.739	246.314	Value out of bounds	206.347	287.515	Value out of bounds
SARIMAX	13.196	17.630	93.332%	11.432	17.563	98.876%	11.608	30.631	98.161%
ANN	15.882	19.478	91.447%	12.276	18.213	98.791%	12.278	30.220	98.216%
CNN	14.925	18.944	91.909%	12.908	18.832	98.707%	12.460	30.195	98.219%
LSTM	24.989	30.494	79.037%	23.817	32.072	96.252%	24.886	40.817	96.745%
GRU	31.811	37.104	68.964%	17.845	24.236	97.859%	26.357	43.356	96.328%
Bi-LSTM	25.552	30.478	79.059%	35.637	51.031	90.512%	43.473	60.166	92.929%
Bi-GRU	34.929	40.054	63.834%	28.126	37.992	94.741%	35.332	54.532	94.191%
LSTM- CNN	15.217	19.491	91.435%	12.553	18.441	98.761%	12.442	30.244	98.213%
LSTM- GRU	17.718	19.491	88.490%	17.274	18.441	97.885%	15.108	30.244	97.901%
End of Table									

7 Discussion

After looking at the resulting graphs of different models and evaluating them gives us a clear understanding of how each of the model works and what kind of results can be

expected. Let's discuss each of the stock's performance.

- HDFC:

For the short term predictions, the graphs tell that the statistical models and the deep learning models are very reliable. The Hybrid and the Bidirectional models were showing very high deviations and thus were not reliable to use, with just the exception of LSTM-CNN which was following the trend fairly closely. By evaluation, it is seen that the worst models are Bi-LSTM as it was having very high RMSE, MAE and R2. The best model of all was ARIMAX. The deep learning models were showing great but interesting results as ANN was performing better than LSTM and GRU.

For mid term predictions, the ARIMAX model stopped being the best and started to show higher deviations. The SARIMAX and deep learning models were performing well according to the graphs. Now that the dataset has increased, the performance of the Bidirectional and Hybrid models have increased as well. It can be seen that there is some deviation, but the overall trend is fitting the actual trend. When evaluating, it was seen that ARIMAX showed a very high MAE and RMSE as well as an out of bounds value for the R2 score. Interestingly, CNN model was the best model out of all and it was interesting to see this result as it was expected that other models will perform better. Furthermore, the bidirectional models were performing lesser than the regular deep learning models.

For long term predictions, it was interesting to see that for the ARIMAX model, initial predicted values did not fit the graph, but for later on it predicted the values more accurate and closer to the actual values. The SARIMAX and the deep learning models fit the model well and follow the actual values. Since it was for long term predictions, the predictions for long term Hybrid and Bidirectional models was showing great results. After seeing the results for mid term prediction, it was obvious that due to increased dataset, the results for the ARIMAX will not be good. The deep learning models were performing equally well, but it was seen that ANN was performing much better than other deep learning models as well as Hybrid models. The bidirectional and the hybrid models were expected to perform better, but these models are not performing as well as deep learning models. The lowest value for MAE was for SARIMAX, as well as the lowest value for RMSE. For the long term predictions, it was obvious that the best results was for SARIMAX, which was interesting as the ARIMAX model had to be discarded due to very extreme values.

- Infosys:

For the short term predictions, Statistical model were expected to perform well and follow the actual trend and that can be seen from the graph. The predicted values for both the model was a little lower than the actual values, but it can be seen as a reliable method. The graph for the deep learning models clearly tells which one is the best working of them, as there is an observable gap between the predicted values, making CNN the best. Since this is a short term prediction, the prediction accuracy would be very less for the Bidirectional models and Hybrid models, just with the exception of LSTM-GRU. It was expected to see as the dataset increases, the accuracy for these models increase as well. The lowest value for the MAE was for ARIMAX, but for the RMSE was for SARIMAX. The last evaluation criteria was R2 score, which was lowest for ARIMAX. Overall it can be seen that the best working model for Infosys for short term prediction was ARIMAX. The overall lowest value for MAE, RMSE was for Bi-GRU as well as it had the highest value for R2 score, making this the worst performing model for the stock.

For the mid term predictions, it can be seen that ARIMAX was the worst model of all, but it was interesting to see that initially it was following the actual trend but later on showed a very high deviation, making this model unreliable to use. Interestingly, SARIMAX turned out to be the best model, and it is just an extension of the ARIMAX model. Deep learning models as was expected and seen with the case of HDFC, they performed well, but they were packed very congested. Furthermore, since this was a mid term predictions, rest of the model's prediction accuracy increased by a huge difference if compared with the short term predictions. By just having one look at the table, it was obvious that ARIMAX is the worst model for predictions as its MAE, RMSE was extremely high and the R2 score was out of bounds. The model with the lowest MAE was SARIMAX, which also had the lowest RMSE and the highest R2 score. It was seen that hybrid models performed better, but the bidirectional models were not living upto their expectations.

For the long term predictions, As was the case with mid term predictions ARIMAX model, it showed similar results that initially it was very close to the actual trend but later on showed a very high deviation. When we move on to the next models, which are deep learning, hybrid and bidirectional models. Some deviations can be seen in those models, but it is extremely hard to make a firm decision just by looking at the graphs. As was seen with the graphs that ARIMAX, this was proved by the evaluation as well. The values for MAE, and

RMSE was extremely high as well as the R2 score was out of bounds. The best model of all ten was CNN as it was having the highest R2 score and the lowest RMSE value. The only value that was not lowest was MAE, which was for SARIMAX, but it had a higher value for RMSE and lower value for R2 score than CNN, making CNN the best performing model.

- Hindustan Unilever:

For the short term predictions, it was seen that both the ARIMAX and SARIMAX models perform equally well. The values of the predictions are a little lower than the actual values, but overall the model can be best useful. When analysing deep learning models, it was clearly seen that the CNN and ANN are performing well, but there were deviations in LSTM and GRU. Furthermore, the Bidirectional models and the hybrid models showed extremely high deviations and thus had to be rejected. When the models were evaluated, the worst model was Bi-LSTM as it had a high MAE and RMSE values, as well as had a low R2 score. The best model would be SARIMAX, with the lowest MAE and RMSE values and high R2 score. It can be seen when the dataset is small, the statistical models outperform the other models.

For the mid term predictions, as it was before, the ARIMAX model showed a very high deviation, but the SARIMAX model was almost too perfect fit for the stock. The deep learning models were tightly congested, and here the evaluation comes to play to find out which is better. In the bidirectional models, there was some deviations, but it was seen that Bidirectional GRU was much more closer to the actual trend line. The Hybrid models, like deep learning models, was very close to the actual trendline, meaning the models were performing well, but further evaluation was required in this case. As the dataset is for mid term evaluation, the ARIMAX model gave a very high MAE and RMSE values as well as out of bounds value for R2 score, making it the worst model of all used in the paper. It was interesting to see that SARIMAX was the best model, and gave the predicted values as close to the actual values and gave the lowest MAE, RMSE and R2 score.

For the long term predictions, since the dataset is now ten years long, the ARIMAX model is showing results no closer to the actual model, meaning when the dataset becomes large, ARIMAX should not be used. For deep learning, bidirectional and hybrid models, the predicted results were too close to the actual values and thus needed further evaluation to find out which one's are the best, though it does not mean that others will be unusable. As was seen with the graphical results, the ARIMAX model is unusable as it is giving very high MAE

and RMSE values and out of bounds value for R2 score. The overall lowest value for MAE, was for CNN, but the lowest value for RMSE was for hybrid LSTM-CNN as well as had the best R2 score, meaning hybrid LSTM-CNN was the best performing model for this evaluation.

8 Conclusion

This paper summarises financial time series in a manner that even a layman can understand. There is still some work that is to be done in the future for the financial time series analysis, but for now let's see what the conclusion is.

8.1 Conclusion

This paper researches about different algorithms that can be used for the financial time series analysis. Different techniques have been used. The first methodology is for statistical models, where it is seen that when the data set starts to increase, the models start to perform worse and worse. It can be concluded that statistical models are best suited for short term predictions, with short term data set and will not be reliable with mid term or long term data. When the deep learning models were analysed, it could be clearly seen that the models are performing well for all three datasets, that is short term, mid term and long term. The deep learning models are extremely reliable and can be concluded as the most reliable among all other models and among deep learning models, CNN is the most reliable. During the analysis of statistical models, it was concluded that for short term, the models work best and for big data sets, they lose their performance. The opposite thing can be said for the bidirectional models. It is clear that the model do not work best when the dataset is small as could be seen in the short term predictions, but their performance increases once the dataset has enough values for the predictions. It can be concluded that the Bidirectional models do not perform well for short term predictions and it performs well for long term predictions. The hybrid models are very interesting, as they work well and can be reliable, but at certain times, it performs better than other models, but other times it can be seen that they do not perform as well as others. It can be concluded that the models are reliable but not as reliable as others.

8.2 Future for Financial Time series

This paper only compares ten algorithms and tries to show the diversification in the machine learning methodology, but it is still not enough to cover the entire financial time series analysis.

- Different algorithms: For the paper, only 10 algorithms were used and compared against each other, but there are hundreds of algorithms and they are not just limited to these 10. It will be interesting to see how other models perform.
- Using different optimizers: Since only one optimizer was used for the analysis, it would be interesting to see what type of results come with different optimizers.
- Creating portfolios: creating portfolios and choosing the best stocks can be a real headache and it would be good idea to create portfolios on the requirements of the client's investment's period of duration.

Financial time series analysis is an interesting area for financial and business analysts. The future work is not just limited to the points above, and different approaches and different algorithms can give different results. Further research can answer the questions that will be required by the analysts in the time for the financial time series analysis with machine learning.

References

- [1] Springer, 2010.
- [2] N. K. Ahmed, A. F. Atiya, N. E. Gayar, and H. El-Shishiny, An empirical comparison of machine learning models for time series forecasting, *Econometric Reviews*, 29 (2010), pp. 594–621.
- [3] S. Ardabili, A. Mosavi, and A. R. Várkonyi-Kóczy, Advances in machine learning modeling reviewing hybrid and ensemble methods, *Lecture Notes in Networks and Systems*, (2020), pp. 215–227.
- [4] A. A. Ariyo, A. O. Adewumi, and C. K. Ayo, Stock price prediction using the arima model, 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, (2014).
- [5] C3.AI, Mean absolute error, 2022.
- [6] —, Root mean squared error, 2022.
- [7] J. Cao, Z. Li, and J. Li, Financial time series forecasting model based on ceemdan and lstm, *Physica A: Statistical Mechanics and its Applications*, 519 (2019), pp. 127–139.
- [8] G. W. Crawford and M. C. Fratantoni, Assessing the forecasting performance of regime-switching, arima and garch models of house prices, *Real Estate Economics*, 31 (2003), pp. 223–243.
- [9] D. S. de O. Santos Júnior, J. F. de Oliveira, and P. S. de Mattos Neto, An intelligent hybridization of arima with machine learning models for time series forecasting, *Knowledge-Based Systems*, 175 (2019), pp. 72–86.
- [10] Y. Deng, L. Wang, H. Jia, X. Tong, and F. Li, A sequence-to-sequence deep learning architecture based on bidirectional gru for type recognition and time location of combined power quality disturbance, *IEEE Transactions on Industrial Informatics*, 15 (2019), pp. 4481–4493.
- [11] D. Ersan, C. Nishioka, and A. Scherp, Comparison of machine learning methods for financial time series forecasting at the examples of over 10 years of daily and hourly data of dax 30 and samp;p 500, *Journal of Computational Social Science*, 3 (2019), pp. 103–133.
- [12] J. Frost, How to interpret r-squared in regression analysis, 2022.
- [13] J. D. Hamilton, *Time Series Analysis*, Princeton University Press, 2020.
- [14] HDFC, About-us, 2022.
- [15] R. J. Hyndman and G. Athanasopoulos, *Forecasting*, O Texts, 2 ed., 2018.
- [16] Infosys, About-us, 2022.

- [17] A. Khazaee Poul, M. Shourian, and H. Ebrahimi, A comparative study of mlr, knn, ann and anfis models with wavelet transform in monthly stream flow prediction, *Water Resources Management*, 33 (2019), pp. 2907–2923.
- [18] T. Kim, S. Sharda, X. Zhou, and R. M. Pendyala, A stepwise interpretable machine learning framework using linear regression (lr) and long short-term memory (lstm): City-wide demand-side prediction of yellow taxi and for-hire vehicle (fhv) service, *Transportation Research Part C: Emerging Technologies*, 120 (2020), p. 102786.
- [19] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, 2014.
- [20] S. Kouadri, C. B. Pande, B. Panneerselvam, K. N. Moharir, and A. Elbeltagi, Prediction of irrigation groundwater quality parameters using ann, lstm, and mlr models, *Environmental Science and Pollution Research*, 29 (2021), pp. 21067–21091.
- [21] M. Kumar and M. Thenmozhi, Forecasting stock index returns using arima-svm, arima-ann, and arima-random forest hybrid models, *International Journal of Banking, Accounting and Finance*, 5 (2014), p. 284.
- [22] H. Liu, H.-q. Tian, and Y.-f. Li, Comparison of two new arima-ann and arima-kalman hybrid methods for wind speed prediction, *Applied Energy*, 98 (2012), pp. 415–424.
- [23] Q. Ma, Comparison of arima, ann and lstm for stock price prediction, *E3S Web of Conferences*, 218 (2020), p. 01026.
- [24] D. A. Otchere, T. O. Arbi Ganat, R. Gholami, and S. Ridha, Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ann and svm models, *Journal of Petroleum Science and Engineering*, 200 (2021), p. 108182.
- [25] D. I. Pardoe, Regression methods, 2022.
- [26] S. S. Poorna, P. M. V. D. S. Baba, G. L. Ramya, P. Poreddy, L. S. Aashritha, G. J. Nair, and S. Renjith, Classification of eeg based control using ann and knn — a comparison, 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), (2016).
- [27] X.-Y. Qian, Financial series prediction: Comparison between precision of time series models and machine learning methods, 2017.
- [28] N. Ramana, D. V. S. R. Kumar, J. Jagdish, P. Anil Kumar, G. V. R. Rao, and V. S. Rao, Comparative assessment of auto regressive integrated moving average with explanatory variable (arimax) and neural network autoregressive models with exogenous inputs (nnarx) for forecasting the old-world bollworm, *helicoverpa armigera* (lepidoptera: Noctuidae) in india, *International Journal of Tropical Insect Science*, (2022).

- [29] R. Rao, nifty-50 stock market data (2000 - 2021), 2022.
- [30] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, Financial time series forecasting with deep learning : A systematic literature review: 2005–2019, *Applied Soft Computing*, 90 (2020).
- [31] F. Shahid, A. Zameer, and M. Muneeb, Predictions for covid-19 with deep learning models of lstm, gru and bi-lstm, *Chaos, Solitons Fractals*, 140 (2020), p. 110212.
- [32] S. Siami-Namini and A. S. Namin, Forecasting economics and financial time series: Arima vs. lstm, 2018.
- [33] Q. Tao, F. Liu, Y. Li, and D. Sidorov, Air pollution forecasting using a deep learning model based on 1d convnets and bidirectional gru, *IEEE Access*, 7 (2019), pp. 76690–76698.
- [34] R. S. Tsay, *Analysis of Financial Time Series*, John Wiley and Sons inc, 2 ed., 2006.
- [35] H. Unilever, About-us, 2022.
- [36] S. I. Vagropoulos, G. I. Chouliaras, E. G. Kardakos, C. K. Simoglou, and A. G. Bakirtzis, Comparison of sarimax, sarima, modified sarima and ann-based models for short-term pv generation forecasting, in *2016 IEEE International Energy Conference (ENERGYCON)*, 2016, pp. 1–6.
- [37] V. H. Wentz, J. N. Maciel, J. J. Gimenez Ledesma, and O. H. Ando Junior, Solar irradiance forecasting to short-term pv power: Accuracy comparison of ann and lstm models, *Energies*, 15 (2022), p. 2457.
- [38] C.-C. Young, W.-C. Liu, and W.-L. Hsieh, Predicting the water level fluctuation in an alpine lake using physically based, artificial neural network, and time series forecasting models, *Mathematical Problems in Engineering*, 2015 (2015), pp. 1–11.
- [39] Y. Yu, X. Si, C. Hu, and J. Zhang, A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures, *Neural Computation*, 31 (2019), pp. 1235–1270.
- [40] B. Zhao, H. Lu, S. Chen, J. Liu, and D. Wu, Convolutional neural networks for time series classification, *Journal of Systems Engineering and Electronics*, 28 (2017), pp. 162–169.
- [41] V. Ülke, A. Sahin, and A. Subasi, A comparison of time series and machine learning models for inflation forecasting: empirical evidence from the usa, *Neural Computing and Applications*, 30 (2016), pp. 1519–1527.