Time Series Data Mining: Comparative Study of ARIMA and Prophet Methods for Forecasting Closing Prices of Myanmar Stock Exchange

Wint Nyein Chan
Faculty of Information Science, University of Computer Studies (Hpa-an)
wintnyeinchan2012@gmail.com

Abstract

Stock price prediction is an important topic in finance and economics which has predicted the stock trends in an efficient manner can minimize the risk of loss and maximize profit. There is a challenging task in stock price prediction is owing to the complexity patterns behind time series. To solve these types of problems, the time series analysis will be the best tool for forecasting the trend or even future. In this paper, time series analysis model: Autoregressive Integrated Moving Average (ARIMA) and PROPHET has been used extensively for time series forecasting in the field of Myanmar Stock prices. Data is prepared for time series analysis by performing data preprocessing steps such as time stamp conversion, stationary identification and stationary treatment. To find the most accurate forecast model and the most suitable forecasting period, the error analysis of PROPHET and ARIMA methods are performed and compared on the same dataset (daily, weekly, monthly Myanmar stock prices). Based on the analysis results, PROPHET outperforms ARIMA for three periods and both of the two models are suitable for short-term prediction (daily and weekly prediction). The aim of this paper is able to support for Myanmar stock prediction and helps researchers in the field of time series modeling, economic analysis and investments.

1. Introduction

The stock market is one of the most important components of a market economy, because it provides companies with access to capital by allowing investors to buy shares of ownership in a company [1]. The area of stock market is constantly developing under the process of refinements. Considering the variations, it brings every day, investors need to plan intensively to make profit. Forecasting the stock exchange data includes an assumption that the information publicly available at present has some predictive relationships to the future stock returns. Stock trend forecasting is one of the most difficult tasks to achieve in finance market because of the difficulty in the highly intricate world of stock market. The investors in the stock market always find a technique that can guarantee easy profiting by forecasting the stock trends and minimize the risk of investing. This motivates the researchers in the domain field to develop new forecasting models.

Stock prices are not randomly generated values rather they can be treated as a discrete time series model which is based on a set of well-defined numerical data items collected at successive points at regular intervals of time. Since, it is essential to identify a model to analyze trends of stock prices with adequate information for decision making, it recommends that transforming the time series using ARIMA [11]. Another popular method that work on seasonal time series quiet well is PROPHAT [3]. It provides some options to handle seasonality of the dataset. These options are yearly, weekly and daily seasonality. Due to providing these options, a data analyst can choose the available time granularity for the forecast model on the dataset [5].

In this paper extensive process of building time series models of different periods for Myanmar stock price prediction is presented. This intended to find out which model is more successful for Myanmar Stock forecasting. The results obtained from real-life data demonstrated the potential strength of time series models to provide investors stock prediction that could aid investment decision making process. The rest of the paper is organized as follows. Section 2 describes the methodology used in the proposed system while section 3 discusses the experimental results obtained. The paper is concluded in section 4.

2. Proposed System

In the proposed system, Myanmar stock price is predicted by using ARIMA model and implemented with python. The historical data set of Myanmar Stock has been collected from Myanmar Stock Price Index (MYANPIX) [4]. The system consists of two major components: preprocessing and stock market prediction. Each component contains sub processes and the process flow of proposed system are illustrated in Figure 1. Detail functions of each component are presented in the following subsections.

2.1. Description of the Myanmar Stock Data

Stock data used in this research work are historical daily stock prices obtained from Myanmar Stock Price Index (MYANPIX). MYANPIX is YSX(Yangon Stock Exchange)-calculated stock price index [10] , which represents price fluctuation of an overall stock market and becomes basic tools to describe Myanmar's stock market for investors. There are multiple features in the dataset – date, open, high, low, close, trading volume, and trading value, no of listed company and market cap. The Open and Close features represent the starting and final price at which
the stock is traded on a particular day. High and Low features represent the maximum, and minimum price of the share for the day [9]. Trading volume [8] is the total quantity of shares or contracts traded for a specified security. It can be measured on any type of security traded during a trading day.

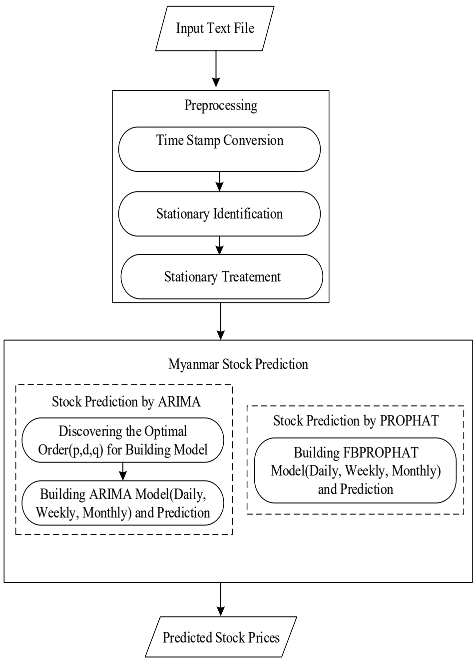


Figure 1. Process flow of Myanmar Stock Prediction System

Trading Value [3] feature represent the traded common stock on a national securities exchange, the market price of the common stock on the applicable date (calculated based on the average closing price during the preceding 20 days trading period). Currently, in Myanmar stock exchange, no of listed company is 5. Market cap [7] market capitalization feature refers to the total value of all a company's shares of stock. It is calculated by multiplying the price of a stock by its total number of outstanding shares. Trading days can be defined as the days on which a given stock exchange is open. In this study, trading date and the closing price features are chosen to represent the price of the index to be predicted. Because the profit or loss calculation is usually determined by the closing price of a stock for the day. Therefore the closing price feature is considered as the target variable. The stock data used in this study covers the period from 2.5.2016 to 31.3.2020 having a total number of 946 observations. The sample data set with five features are presented in Table 1.

Table 1. Sample Myanmar Stock Data (Daily)

Date	Open	High	Low	Close
2-5-2016	967.74	967.74	903.23	903.23
3-5-2016	919.35	935.48	919.35	935.48
5-5-2016	903.23	903.23	903.23	903.23
….	…	….	…	...
25-3-2020	437.1	437.1	425.78	425.78
26-3-2020	424.36	424.36	420.14	421.57
30-3-2020	420.63	420.63	415.41	418.25
31-3-2020	418.25	419.19	417.3	419.19

2.2. Preprocessing

Data preprocessing techniques have significant influence on prediction system, therefore are essential in a prediction model. Three steps are included and detail functions of each steps are described as follows:

2.2.1. Time Stamp Conversion

The first step is to read the text file with date format using pd.datetime.strptim function. strptime() functions in Python's built-in datetime module can be used. As stock trader generally have access to daily, weekly, monthly price data for a stock or a stock portfolio, the proposed prediction system focus on daily, weekly and monthly data. The original input data contains daily time stamp, it can be processed as the daily data. For creating weekly and monthly time series data, resample dataframe is used to select rows whose weekday is not include Saturday or Sunday. For weekly time series data, "w" parameter of resample dataframe is utilized and the daily data is aggregated for each week. The mean (average) value are conducted for corresponding week. For monthly time series data, "m" parameter of resample dataframe is utilized and the daily data is aggregated for each month. The mean (average) value are conducted for corresponding month.

2.2.2. Time Stationary Identification

Stationary identification is important for time series analysis because this analysis only works with stationary data. The Augmented Dickey-Fuller (ADF) test is one of the most popular statistical tests. It can be used to determine the presence of unit root in the series, and hence help to check whether the series is stationary or not. If the test statistic is less than the critical value, the null hypothesis is rejected (aka the series is stationary). When the test statistic is greater than the critical value, rejecting is fail for the null hypothesis (which means the series is not stationary). The result of ADF test for daily time series prediction is described in Table 2. In this case, the test statistic of daily, weekly, monthly data is greater than critical value, which implies that the series is not stationary.

Table 2. Result of ADF Test for daily time series prediction

	Value
Test Statistics	-3.275227
p-value	0.016023
No. of lags used	8
Number of observations used	937
Critical Value (1%)	-3.437348
Critical Value (5%)	-2.864630
Critical Value (10%)	-2.568415

2.2.3. Stationary Treatment

The important step in the forecasting process is typically to do some transformation to convert a non-stationary series to stationary. The most common and convenient method to stationarize the series is taking the log of the series. In this work, numpy package is used for convert the data to log data. Transformations are used to stabilize the non-constant variance of a series. Log transformation are used for treating the stationary.

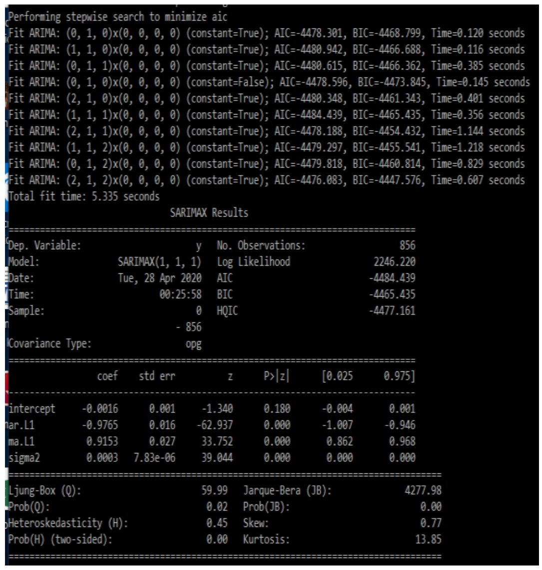
2.3. ARIMA

ARIMA is stand for Autoregressive Integrated Moving Average. An ARIMA (p, d, q) model consists of three parts namely auto regression AR (order p), moving average MA (order q) and the degree of ordinary differencing (order d). Adopting an ARIMA model for a time series, the optimal parameter for building model is former vital step. And ARIMA statistical model is constructed for analyzing and forecasting stock data. Detail functions of each step are presented as follows:

2.3.1. Discovering the optimal order (p, d, q) for Building Model

Automatically discover the optimal order function: auto_arima is used for constructing an ARIMA model. The auto_arima function seeks to identify the most optimal parameters for an ARIMA model, and returns a fitted ARIMA model. This function works by conducting Augmented Dickey-Fuller (ADF) to determine the order of differencing, d, and then fitting models within ranges of defined start_p, max_p, start_q and max_q ranges. In this case, the Auto ARIMA model provided the value of p, d, and q as 1, 1 and 1 respectively. The results of Auto ARIMA model for daily stock prices can be seen in Figure 2. According to the results, ARIMA (1, 1, 1) is relatively the best model. The model returned the smallest AIC of -4484.439. Orders that are suitable for daily, weekly, and monthly series are ARIMA(1,1,1), ARIMA(0,1,0) and ARIMA(1,2,2) respectively.

Figure 2. The results of Auto ARIMA model (for daily stock prediction)



```
Performing stepwise search to minimize aic
Fit ARIMA: (0, 1, 0)x(0, 0, 0, 0) (constant=True); AIC=-4478.301, BIC=-4468.799, Time=0.120 seconds
Fit ARIMA: (1, 1, 0)x(0, 0, 0, 0) (constant=True); AIC=-4480.942, BIC=-4466.688, Time=0.116 seconds
Fit ARIMA: (0, 1, 1)x(0, 0, 0, 0) (constant=True); AIC=-4480.615, BIC=-4466.362, Time=0.385 seconds
Fit ARIMA: (0, 1, 0)x(0, 0, 0, 0) (constant=False); AIC=-4478.596, BIC=-4473.845, Time=0.145 seconds
Fit ARIMA: (2, 1, 0)x(0, 0, 0, 0) (constant=True); AIC=-4480.348, BIC=-4461.343, Time=0.401 seconds
Fit ARIMA: (1, 1, 1)x(0, 0, 0, 0) (constant=True); AIC=-4484.439, BIC=-4465.435, Time=0.356 seconds
Fit ARIMA: (2, 1, 1)x(0, 0, 0, 0) (constant=True); AIC=-4478.188, BIC=-4454.432, Time=1.144 seconds
Fit ARIMA: (1, 1, 2)x(0, 0, 0, 0) (constant=True); AIC=-4479.297, BIC=-4455.541, Time=1.218 seconds
Fit ARIMA: (0, 1, 2)x(0, 0, 0, 0) (constant=True); AIC=-4479.818, BIC=-4460.814, Time=0.829 seconds
Fit ARIMA: (2, 1, 2)x(0, 0, 0, 0) (constant=True); AIC=-4476.083, BIC=-4447.576, Time=0.607 seconds
Total fit time: 5.335 seconds
                           SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  856
Model:               SARIMAX(1, 1, 1)   Log Likelihood                2246.220
Date:                Tue, 28 Apr 2020   AIC                          -4484.439
Time:                        00:25:58   BIC                          -4465.435
Sample:                             0   HQIC                         -4477.161
                                - 856
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept     -0.0016      0.001     -1.340      0.180      -0.004       0.001
ar.L1         -0.9765      0.016    -62.937      0.000      -1.007      -0.946
ma.L1          0.9153      0.027     33.752      0.000       0.862       0.968
sigma2         0.0003   7.83e-06     39.044      0.000       0.000       0.000
===================================================================================
Ljung-Box (Q):                       59.99   Jarque-Bera (JB):              4277.98
Prob(Q):                              0.02   Prob(JB):                         0.00
Heteroskedasticity (H):               0.45   Skew:                             0.77
Prob(H) (two-sided):                  0.00   Kurtosis:                        13.85
===================================================================================
```

2.3.2. Building Model and Prediction

To build the model, the data is split into train and test data. The appropriate order from step 1 of ARIMA model construction was used for constructing the three ARMA model on training data (daily, monthly, weekly data) and forecasting time series (daily, monthly, weekly test data). In this work, alpha 0.05 means 95% confidence are considered for forecasting. The predicted prices are compared with the actual prices and detail analysis results are described in section 3. The sample comparative result of actual and predicted daily stock prices by ARIMA are illustrated in Table 3.

Table 3. Comparative Result of actual and prediction stock Prices by ARIMA model (daily)

2.4. PROPHET

Time	Prediction by ARIMA	Actual
13.11.2019	459.62	460.58
14.11.2019	460.45	470.48
15.11.2019	458.96	469.58
18.11.2019	459.66	477.54
19.11.2019	458.29	479.52
…	…	
24.3.2020	430.30	449.84
25.3.2020	429.95	425.78
26.3.2020	429.62	421.57
30.3.2020	429.27	418.25
31.3.2020	428.94	419.19

PROPHET is an open source software that is available in Python and R for forecasting time series data. PROPHET is published by Facebook's Core Data Science team. It depends on a contribution model where non-linear trends are fit with weekly and yearly seasonality and plus holidays. PROPHET is strong to missing data, capturing the shifts in the trend and large outliers. Adopting PROPHET model for developing the system, fbprophet version 0.6 is utilized.

2.4.1. Building Model and Prediction

The two split data: train and test data are used to build the model. PROPHET model fit own special data frame to handle time series and seasonality. The data frame contains two basic columns: "ds" column that stores date time series and the other column "y" that stores the corresponding values of the time series in the data frame. The three different parameter for period in the model are used to forecast a time series data (daily, weekly, monthly). The predicted prices are compared with the actual prices and detail analysis results are described in section 3. The sample comparative result of actual and predicted daily stock prices by PROPHET are illustrated in Table 4.

Table 4. Comparative Result of actual and prediction stock Prices by PROPHET model (daily)

Time	Prediction by PROPHET	Actual
13.11.2019	469.46	460.58
14.11.2019	469.10	470.48
15.11.2019	470.79	469.58
18.11.2019	470.36	477.54
19.11.2019	470.27	479.52
…	…	
24.3.2020	447.07	449.84
25.3.2020	445.20	425.78
26.3.2020	443.80	421.57
30.3.2020	441.16	418.25
31.3.2020	440.24	419.19

3. Result Analysis

In this research, ARIMA and PROPHET models are used for Myanmar stock price prediction. The stock price dataset has 946 transactions from 2.3.2016 to 31.3.2020. The time series data was divided into two groups. The first group was training dataset which contain data for construction the forecasting models. The second group is test dataset which contain data for forecasting time series data. Detail description for training and test data for three different periods are presented in Table 5.

Table 5. Different Periods of Train and Test Data

Period	Train Data	Test Data
Daily	856 days(2.3.2016 -8.11.2019)	90 days(13.11.2019 -31.3.2020)
Weekly	191 weeks(1st week Jan, 2016- last week Dec, 2020)	14 weeks((1st week January- last week March, 2020)
Monthly	46 months(Jan, 2016-Dec, 2020)	3 months(Jan, Feb, March, 2020)

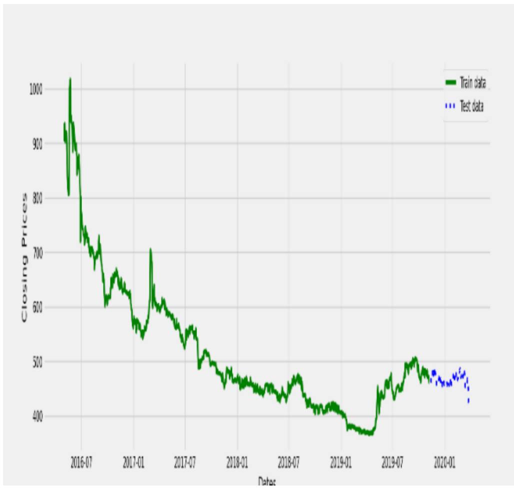


Figure 3. Train and Test Data on Matplot (for Daily Prediction)

For finding the most suitable forecasting period, daily, monthly and weekly prediction are performed. The train and test data set for daily stock data are set on Matplot and the matplot can be seen in following Figure 3. Error analysis is done for evaluating the time series models. The mean absolute percentage error (MAPE) is used to measure for finding the most accurate forecast model. Lewis [2] stressed that the MAPE is the most useful measure to compare the accuracy of time series forecasting methods as it measures relative performance.

The comparative running result of daily models is shown in Figure 4. According to the prediction result, the solutions of PROPHET model is closer with the actual prices. Especially, PROPHET model prediction result is significantly closer than ARIMA model with actual stock prices on February and March 2020.

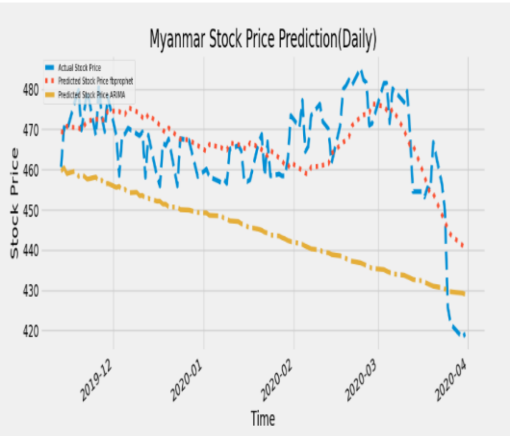


Figure 4. Comparative Predicted Prices of Two models and Actual Prices (Daily Prediction)

Then, the comparative running results of time period for forecasting in weekly series result on matplot are shown in Figure 5. The sample prediction values of time period for forecasting in weekly series are presented in Table 6. During the 14 weeks, the stock has fluctuation and the prediction results of two models are slightly difference over actual prices.

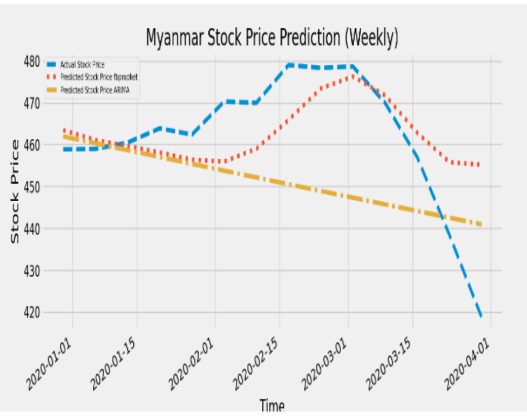


Figure 5. Comparative Predicted Prices of Two models and Actual Prices (Weekly Prediction)

Table 6. Sample Prediction Results of Two models and Actual Prices (Weekly Prediction)

Time	Prediction by ARIMA	Prediction by PROPHET	Actual
1st week	461.95	463.44	458.85
2nd week	460.30	461.06	458.96
3rd week	458.65	459.64	460.45
4th week	457.01	458.11	463.87
5th week	455.38	456.35	462.38
6th week	453.75	455.92	470.26
7th week	452.13	458.98	469.96
8th week	450.51	465.87	478.99
9th week	448.90	473.38	478.31
10th week	447.30	476.25	478.66
11th week	445.70	471.75	469.95
12th week	444.11	462.80	456.95
13th week	442.52	455.79	438.38
14th week	440.94	455.18	418.72

Then, the running results of time period for forecasting in monthly series result on matplot are shown in Figure 6. ARMA model is worse for monthly stock prediction. According to the results, ARIMA is more appropriate for short term prediction such as daily and weekly prediction.

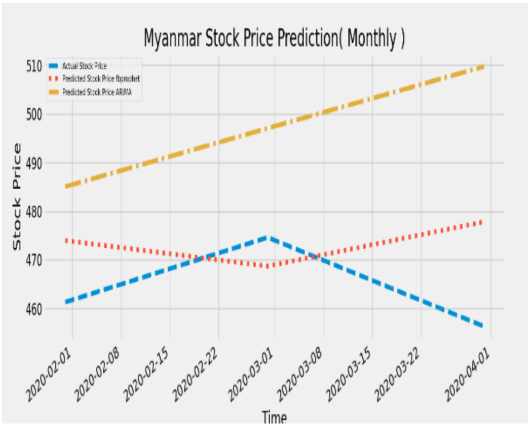


Figure 6. Comparative Predicted Prices of Two models and Actual Prices (Monthly Prediction)

Mean Absolute Percent Error (MAPE) is a very commonly used metric for forecast accuracy. Since MAPE is a measure of error, high numbers are bad and low numbers are good. The comparative results and MAPE of each model are shown in Table 7.

TABLE 7. Stock Prediction Error Rate

Prediction Models	Time	MAPE
ARIMA	Daily(90 days)	0.046
	Weekly(14 weeks)	0.032
	Montly(3 months)	0.071
PROPHET	Daily(90 days)	0.015
	Weekly(14 weeks)	0.020
	Montly(3 months)	0.029

The error rate of ARIMA stock forecasting for daily, monthly, weekly series are 0.046, 0.032 and 0.071 respectively. The error rate of PROPHET models for monthly time series are greater than other period and the lowest error rate is 0.32 for weekly prediction. The error rate of PROPHET stock forecasting for daily, monthly, weekly series are 0.015, 0.020 and 0.29 respectively. The error rate of PROPHET models for three periods are not significantly difference and the lowest error rate is 0.15 for daily prediction. According to the analysis result, the PROPHET model outperform over ARIMA model for three periods.

4. CONCLUSIONS

This paper presents constructing and forecasting of ARIMA model and PROPHET model for Myanmar stock price prediction. For finding the most suitable forecasting period, daily, monthly and weekly prediction are performed. For finding the most accurate model, MAPE is used to measure for comparing the error analysis results of forecasting methods on different three periods. The experimental results demonstrated the PROPHET model makes predictions quite close to reality, that is the error rate of ARIMA stock forecasting for daily, monthly, weekly series are 0.046, 0.032 and 0.071 respectively while the error rate of PROPHET stock forecasting for daily, monthly, weekly series are 0.015, 0.020 and 0.29 respectively. The experimental results obtained with best PROPHET model for daily series is 0.015 error rate. Therefore PROPHET outperforms ARIMA for three periods and both of the two models are suitable for short-term prediction (daily and weekly prediction). Actually, in the proposed system, the ARIMA model is developed by auto.arima function (auto parameter calculation function) does not produce the seasonal pattern at all. For further testing, the ARIMA with fine turned parameters should be compared the results against the one from PROPHET with no parameter setting. Moreover, more technologies should be considered for obtaining the high level performance. In this work, time series data is divided into randomly split and develop the prediction model by random split training and testing data set. For a better predictive models to avoid over-fitting, model selection should be performed by using cross validation. The proposed system forecast the stock prices (closing prices) by using only two features: date and closing prices. Additional features should be considered for developing more reliable stock prediction system. And the best features selection should be operated.

References

[1] A. A. Adebiyi, A. O. Adewumi, C. K. Ayo "Stock Price Prediction Using the ARIMA Model", UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, pp.106-112.

[2] C. D. Lewis, "Industrial and Business Forecasting Methods",London, Butterworths, 1982.

[3] Definition of Trading value, https://www.lawinsider.com/dictionary/trading-value , [Online; accessed May 2020]

[4] I. Yenidogan, A. Cayir, O. Kozan, et al. "Bitcoin Forecasting Using ARIMA and PROPHET, "3rd International Conference on Computer Science and Engineering, 2018, pp. 621-624.

[5] Myanmar Stock Price Index (MYANPIX), https://ysx-mm.com/mktdata/market-summary/, [Online; accessed March 2020]

[6] PROPHET Forecasting At Scale, https://research.fb.com/PROPHET-forecasting-at-scale/, [Online; accessed March 2020].

[7] Understanding market capitalization https://www.fidelity.com/learning-center/trading-investing/fundamental-analysis/understanding-market-capitalization, [Online; accessed May 2020]

[8] Volume of trade, https://www.investopedia.com/terms/v/volumeoftrade.asp, [Online; accessed May 2020]

[9] What is Open High Low Close in Stocks? https://analyzingalpha.com/open-high-low-close-stocks, [Online; accessed May 2020]

[10] Yangon Stock Exchange, https://ysx-mm.com/mktdata/index-data/ [Online; accessed March 2020]

[11] Y. W. Lee, K. G. Tay, Y. Y. Choy, "Forecasting Electricity Consumption Using Time Series Model", International Journal of Engineering and Technology, Volume 7, 2018, pp. 218-223