International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 6, Issue 2, pp 80-85, February-2022

https://doi.org/10.47001/IRJIET/2022.602014

Predicting NEPSE Index Using ARIMA Model

Avinash Maskey

Department of Graduate Studies, Nepal College of Information Technology, Pokhara University, Lalitpur, Nepal

Email: avinash.202802@ncit.edu.np

***Abstract* - Since its inception, the stock market, one of the most financially turbulent markets, has captivated the hearts of thousands of investors. Predicting stock indexes or prices has long been a prominent subject of research in the domain of financial information because of the reward and risk. Also, it has a tremendous beauty as everyone wants to benefit from it. Many approaches are used to forecast stock indexes or prices in the stock market, including technical analysis, fundamental analysis, statistical analysis, and so on, but no single method has shown to be a reliable forecasting tool. This research work contributes to the subject of Time Series Analysis, which seeks to forecast stock market indexes using historical data. In this research work, ARIMA model for NEPSE index has been examined. Published data of daily closing stock index was obtained from Nepal Stock Exchange (NEPSE). With 1798 observations, the data set covers the NEPSE index for approximately eight years, from February 02, 2014 to January 13, 2022. Furthermore, based on the findings, the ARIMA model has a high potential for anticipating short-term market swings, which might be beneficial to short-term traders or investors.**

Keywords: NEPSE, Forecasting, ARIMA, Time series analysis, Stock Index Prediction, Financial Market, Moving Average, Big Data.

I. INTRODUCTION

The Nepal Stock Exchange, often known as NEPSE, was established in 2006 under the Companies Act and is governed by the Securities Act of 2007 [2]. NEPSE main objective is to provide government and corporate sector with free asset commercialization and liquidity by facilitating transactions on its trading floor through members, financial intermediaries, such as market makers and brokers. NEPSE commenced its trading floor on January 13, 1994 [2].

When people talk about market swings, they're referring to a strong or weak performance, turning bullish or bearish, and also talking about the stock market from the standpoint of indices. A stock market, also referred to as share market, equity market, or financial market, is a gathering place for buyers and sellers to buy and sell stocks they own or want to own. A stock market indices is a quantification that estimates the worth of a part of the stock market based on the pricing of specific stocks. Investors use it to define the market and compare the returns on different investments [1]. In most cases, a weighted average method is utilized to create an index for a certain stock exchange. Because an index is a mathematical concept, it cannot be built directly. Many mutual funds and dedicated financial institutions, strive to track the stock index in order to build specialized investment vehicles such as Index Funds (IF) and Exchange Traded Funds (ETF). Those funds and investment vehicles may not be compared to the overall market. From a macro viewpoint, the stock index can be viewed as a gauge of the whole economy, indicating how well the economy is performing [3].

The subject of stock market index prediction is well-known among stock traders. Stock market indexes are extremely volatile due to a wide range of contributing factors. A accurate prediction of a stock's future index could yield a large reward. It can be accomplished by confirming that time series patterns have significant forecasting capacity for a high probability of lucrative trades and a high likelihood of successful returns in a competitive business environment. Moreover, using enormous amounts of historical market data to reflect different situations and showing that time series patterns have statistically significant predictive potential investors can make their investments fecund and gain confidence for the future trades. Because failing to track stock market movements can result in considerable losses for investors, numerous strategies and approaches for predicting stock prices and trends have been developed and employed since the stock market's founding. These techniques range from basic to technical to quantitative analysis. The focus of this research work is on the technical aspects of index prediction using the ARIMA model. This study outlines the arduous process of constructing ARIMA model for predicting short-term indexes. The results derived from real-world data proved the ARIMA model's potential strength to short-term investors, which might help them make better investment decisions.

The rest of the paper is organized as follows. The related work forecasting time series is presented in Section 2. Section 3 provides a quick summary of the ARIMA model. The methodology adopted is described in section 4, and the experimental findings are discussed in section 5. Finally, section 6 brings the paper to a close.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 6, Issue 2, pp 80-85, February-2022

https://doi.org/10.47001/IRJIET/2022.602014

II. RELATED WORK

Several ARIMA models and approaches developed in recent years to anticipate market indices or stock prices are discussed below. The following literatures explain previous time series analysis research experiments.

Ayodele A. Adebiyi et al. (2014) developed a method for forecasting stock price utilizing the ARIMA model in their study. According to their results of short-term prediction, the ARIMA model's final outcome can compete relatively well with new forecasting techniques. Published data was obtained from the NYSE [8].

Using the Box-Jenkins approach, Mohamed Ashik A and Senthamarai Kannan K (2017) analyzed and forecasted Nifty 50 index for the next few trading days. For their fitted model, the R-Square value had a 94 percent influence, and the Mean Absolute Percentage Error (MAPE) had quite small influence. As a result, the prediction accuracy was better for the Nifty 50. Also, in their study of closing price Nifty 50 revealed a slow falling fluctuation tendency for the following trading days [4].

In Nepalese context, Hom Nath Gaire (2017) reported the results of the empirical investigation, the NEPSE index included all of time series components such as AR, MA, and seasonal component. The best model for forecasting future values of daily NEPSE Index based on previous behavior was determined to be the univariate ARIMA model with seasonality, i.e., SARIMA. The author also used GARCH (EGARCH) to represent the NEPSE Index's volatility. As a result, it was discovered to be suitable for predicting NEPSE Index volatility. The author then stated on the effects of both observed and random factors, which would be repeated every five days (because NEPSE is only open five days a week) and would last until the second week. Hence, he concluded that individual company prices could not follow the same pattern as the NEPSE Index [5].

Addition to Nepalese context, Prakash Chandra Prasad et al. (2018) contributed to the sub field of time series analysis, which seeks to forecast stock market prices using historical data. It showed how the moving average approach could be used to find undiscovered and hidden patterns in the stock market, using SARIMA as an example. Authors suggested a system that entails creating and training a model utilizing historical data from the chosen stock, as well as comparing the model's findings to real-world data to determine model accuracy. The outcome added to the creation of more reliable forecasts for qualitative and quantitative information on stock prices. Hence, the Mean Square Error (MSE) obtained using the SARIMA model was 30.01, whereas the ARIMA model produced an MSE of 205.82 [6].

Mohankumari C et al. (2019) explained a detailed procedure for creating an ARIMA model. Based on historical data forecasting, the ARIMA model generated a prediction in which data was applied by first order difference to eliminate random walk pattern issues. In addition, the experimented results for stock prices on a short-term basis were satisfactory which could help investors make better investing choices. As a result of their analysis, authors concluded that ARIMA model has a strong prospect for short term price prediction. [7].

III. ARIMA MODEL

ARIMA models are a type of time series forecasting model. The two most common methodologies for forecasting time series are exponential smoothing and ARIMA models. Also, they can be considered as complementary methodology to the problem of study. While exponential smoothing methods attempts to explain the data's trend and seasonality, ARIMA models strive to explain the data's autocorrelations. In 1970, Box and Jenkins introduced the ARIMA model. It is also referred as the Box-Jenkins methodology, which consists of a series of steps for identifying, estimating, and diagnosing ARIMA models with time series data. In financial forecasting, the model is one of the most widely used methods. ARIMA models have demonstrated their capacity to deliver accurate short-term projections. In terms of short-term prediction, it consistently beat complicated structural models. In an ARIMA model, a variable's predicted value is a linear mixture of prior values and past errors. This acronym entails the key aspects of the model itself. Briefly, they are Autoregression (AR) a model that makes advantage of a contingent relationship between a single observation and a set of lagged observations. Integrated (I) to make a time series stationary by differencing raw observations (e.g., we can subtract an observation from the previous observation in a time series). Moving Average (MA) it refers to the dependency between an observation and a residual error. Moving average applied to lagged observations is used in this model. Every component achieved are provided as a parameter in this model. ARIMA (p, d, q) is a standard notation in which the parameters are replaced by integer values to immediately indicate the ARIMA model being utilized. The ARIMA model's parameters are (p) it defines the number of lagged observations in the model, also referred as the lag order, (d) it can be understood as number of raw observations differenced, also referred as the degree of differencing, and (q) it is referred as the moving average order or the size of window in moving average [8, 9, 10].

Mathematically,

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + \epsilon_t + \emptyset_1 \epsilon_{t-1} + \emptyset_2 \epsilon_{t-2} + \dots + \emptyset_q \epsilon_{t-q}$$

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 6, Issue 2, pp 80-85, February-2022

https://doi.org/10.47001/IRJIET/2022.602014

Where, Y_t is the predicted value, Y_{t-1} is the lag 1 of the series, β_1 is the coefficient of lag 1, α is the constant or the intercept term, ϵ_t is the random error at t, ϵ_{t-1} is the lagged error term at t-1, $\emptyset$ is the coefficient of lagged error at t-1.

IV. METHODOLOGY

In order to achieve the set of objectives of the study an analytical research design has been adopted. The subsections below outline the method used in this study to create the ARIMA model for stock index prediction. The tools, used for implementation are Jupyter Notebook, Python language and their libraries. Fig. 7 below explains the research design in such a way that the collection, analysis and interpretation of the secondary data related to the study may be easier and efficient while drawing conclusions.

4.1 Data Import

Historical daily stock indices were used as data in this study which was obtained from NEPSE's website. The data in series is made up of four components, namely: date, index, absolute change and percentage change respectively. The daily closing index is used to predict time series data in this study because it estimates the weighted value of all the shares listed on the exchange in a trading day and reflects all of the index's activities. The data was collected from February 02, 2014 to January 13, 2022.

4.2 Data Visualization

The data was represented in the shape of a line chart to check if it was distributed evenly and shows the same trend as of original.

4.3 Stationarity Test

A stationary time series is one whose values are not affected by time. That is to say, the series' statistical features such as mean, variance, and autocorrelation remain constant across time. For stationarity test, we opted for Dickey Fuller Test (DFT). The results obtained from stationarity test are shown below in Fig. 1. So, from figure we can conclude that p-value is greater than 0.05 and the test statistics also exceeds the critical values. As, a result the data is not stationary. So, we must isolate our data series from seasonality and trend before conducting a time series analysis.

```
Results of Dickey Fuller Test:
Test Statistics                  0.085856
p-value                          0.965082
No. of lags used                19.000000
Number of observations used   1778.000000
critical value (1%)             -3.434033
critical value (5%)             -2.863167
critical value (10%)            -2.567636
```

Figure 1: Results of Dickey Fuller Test (DFT)

So, in order to achieve the data or series as stationary, we must decompose it as follows:

- Firstly, the log of the series should be taken to reduce the magnitude and growing trend in series.
- Secondly, the rolling average of the series was calculated (average of 12 days were taken).
- Lastly, the mean consumption value at each subsequent point in the series was calculated.

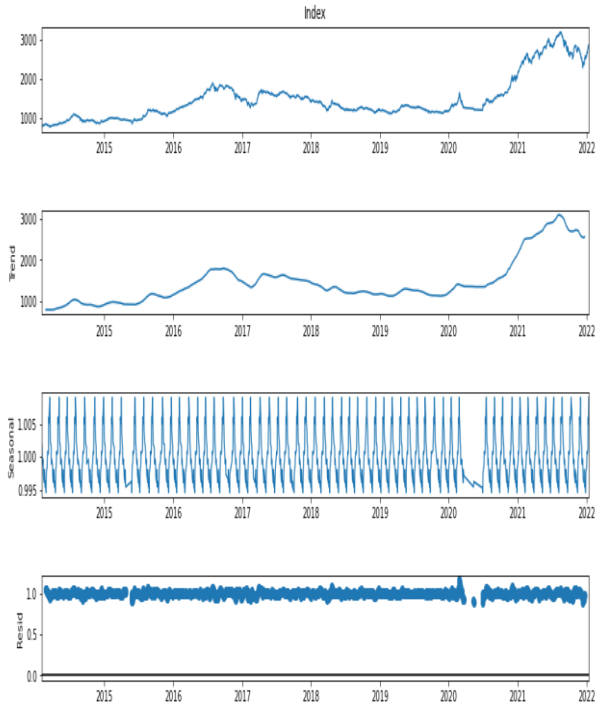

Figure 2: Isolating time series from trend and seasonality

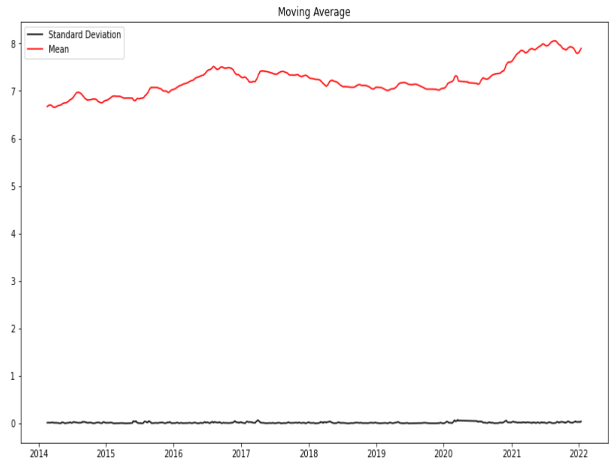

Figure 3: Visualizing series after stationarity test

Fig. 2 and 3 above, depicts the result of stationarity test. In Fig. 2, the seasonality and trend were isolated and in Fig. 3, it visualizes the results obtained after rolling mean and standard deviation operations were performed.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 6, Issue 2, pp 80-85, February-2022

https://doi.org/10.47001/IRJIET/2022.602014

4.4 Split Data and Develop model

Fig. 4 below visualizes the data by dividing it into training and testing sets. 90% of the data in series was used for the training set, while the remaining 10% was used for the testing set. Now, the ARIMA model can be developed for further operations.

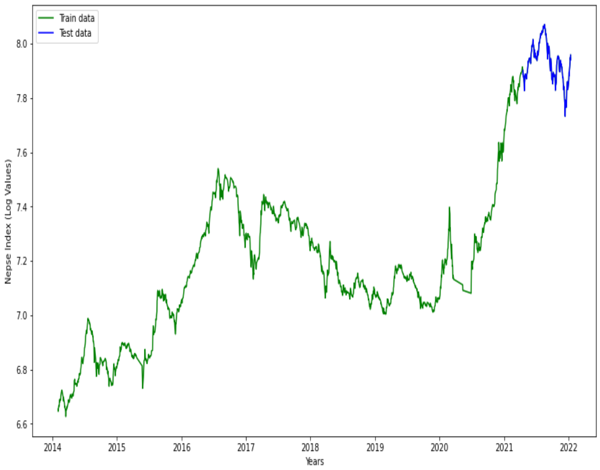

Figure 4: Developing a train and test model for ARIMA

4.5 Discover Optimal ARIMA Model

In order to choose the best optimal p, d, q parameters for ARIMA model the Auto ARIMA was utilized instead of Autocorrelation (ACF) and Partial Autocorrelation (PACF) charts. After, identification of the best suitable parameters for ARIMA model, the auto_arima() function provides a fitted ARIMA model values. auto_arima() uses a step-by-step method to find multiple possibilities of p, d, q parameters and chooses the best model that has the least Akaike Information Criteria (AIC). Hence, Fig. 5 below explains the Auto ARIMA model assigned values (0, 1, 1) to (p, d, q).

```
Performing stepwise search to minimize aic
 ARIMA(0,1,0)(0,0,0)[0] intercept   : AIC=-9563.437, Time=0.51 sec
 ARIMA(1,1,0)(0,0,0)[0] intercept   : AIC=-9608.863, Time=0.57 sec
 ARIMA(0,1,1)(0,0,0)[0] intercept   : AIC=-9610.389, Time=2.14 sec
 ARIMA(0,1,0)(0,0,0)[0]             : AIC=-9559.213, Time=0.35 sec
 ARIMA(1,1,1)(0,0,0)[0] intercept   : AIC=-9608.370, Time=1.58 sec
 ARIMA(0,1,2)(0,0,0)[0] intercept   : AIC=-9608.415, Time=1.09 sec
 ARIMA(1,1,2)(0,0,0)[0] intercept   : AIC=-9607.305, Time=3.55 sec
 ARIMA(0,1,1)(0,0,0)[0]             : AIC=-9607.745, Time=0.39 sec

Best model:  ARIMA(0,1,1)(0,0,0)[0] intercept
Total fit time: 10.215 seconds
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                 1615
Model:               SARIMAX(0, 1, 1)   Log Likelihood                4808.194
Date:                Tue, 01 Feb 2022   AIC                          -9610.389
Time:                        17:38:20   BIC                          -9594.229
Sample:                             0   HQIC                         -9604.391
                               - 1615
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.0008      0.000      2.130      0.033    6.18e-05      0.001
ma.L1          0.1757      0.013     13.221      0.000       0.150      0.202
sigma2         0.0002   3.03e-06     49.863      0.000       0.000      0.000
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):              1291.53
Prob(Q):                              1.00   Prob(JB):                         0.00
Heteroskedasticity (H):               1.64   Skew:                             0.29
Prob(H) (two-sided):                  0.00   Kurtosis:                         7.34
===================================================================================
```

Figure 5: Developing parameters for ARIMA model

Further, Fig. 6 demonstrated below, helps to better understand the hyper parameters of ARIMA model. Following model diagnosis is performed to interpret the residual plots in ARIMA model:

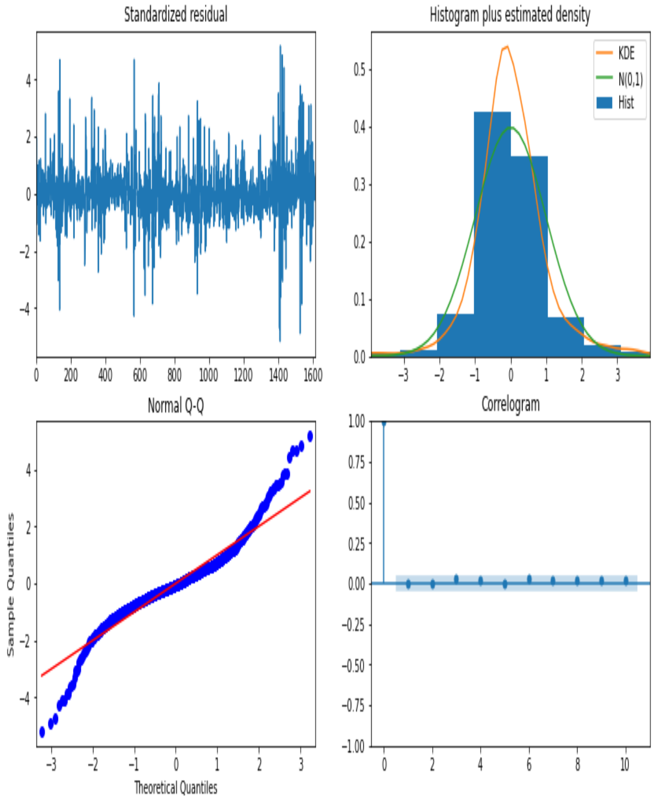

Figure 6: ARIMA model diagnostic

- Top left: The residual errors appear to be uniformly distributed and oscillate around a mean of zero.
- Top right: A normal distribution with a mean of zero is inferred from the density diagram. In the top right plot, the red Kernel Density Estimate (KDE) line closely follows the N (0, 1) line. N (0, 1) is the traditional notation for a normal distribution with mean 0 and standard deviation 1. This shows that the residuals are normally distributed over data.
- Bottom left: All of the dots should be completely aligned with the red line. A skewed distribution would be shown by any large deviations. The ordered distribution of residuals (blue dots) follows the linear trend of samples drawn from a standard normal distribution with N (0, 1) in the QQ-plot on the bottom left. This indicates that the residuals are regularly dispersed once more.
- Bottom right: The Correlogram, often known as the ACF plot, shows that the residual errors are not autocorrelated. Any autocorrelation implies that the residual errors have a pattern that the model doesn't explain. As a result, the model will require more predictors.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 6, Issue 2, pp 80-85, February-2022

https://doi.org/10.47001/IRJIET/2022.602014

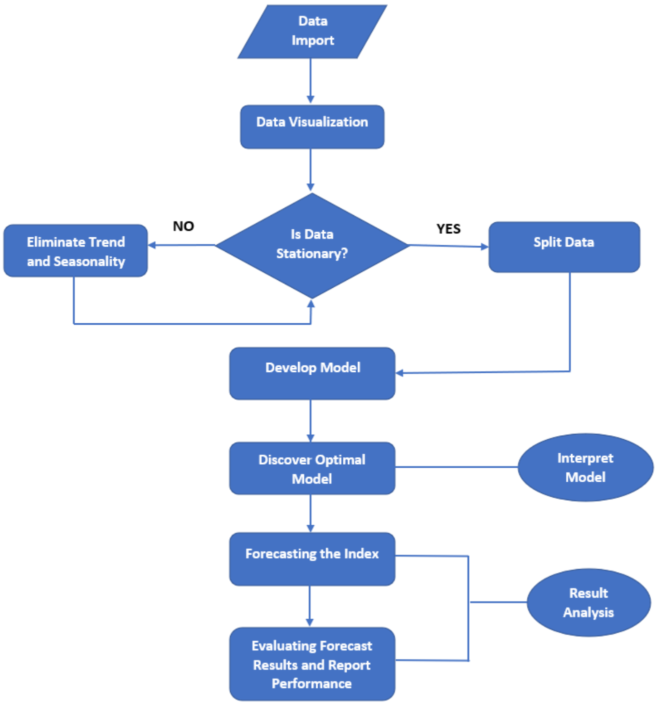

Figure 7: Block diagram (Proposed methodology)

V. RESULT ANALYSIS AND DISCUSSIONS

After the above procedures, the auto_arima() function suggests the p, d, and q parameters for building the ARIMA model. In the subsection below, the experimental findings of NEPSE index are discussed.

5.1 Result of ARIMA Model for NEPSE Index

In this instance, ARIMA (0, 1, 1) was selected as the best model for predicting NEPSE Index after series of adjustments. Fig. 8 below demonstrates the results obtained after the values were fitted into the model.

```
                             ARIMA Model Results
==============================================================================
Dep. Variable:                D.Index   No. Observations:                 1614
Model:                 ARIMA(0, 1, 1)   Log Likelihood                4808.195
Method:                       css-mle   S.D. of innovations              0.012
Date:                Wed, 09 Feb 2022   AIC                          -9610.390
Time:                        13:47:52   BIC                          -9594.231
Sample:                             1   HQIC                         -9604.392

================================================================================
                   coef    std err          z      P>|z|      [0.025      0.975]
--------------------------------------------------------------------------------
const            0.0008      0.000      2.159      0.031    7.16e-05       0.001
ma.L1.D.Index    0.1765      0.025      7.130      0.000       0.128       0.225
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
MA.1           -5.6666           +0.0000j            5.6666            0.5000
```

Figure 8: ARIMA model results

5.2 Forecasted Result of NEPSE Index

The Fig. 9 below explains the forecasting of NEPSE Stock Index on the test dataset with a 95% confidence level. As, there are certain instances of closely related actual and forecasted values, the experimented results of ARIMA model used for prediction was fairly impressive.

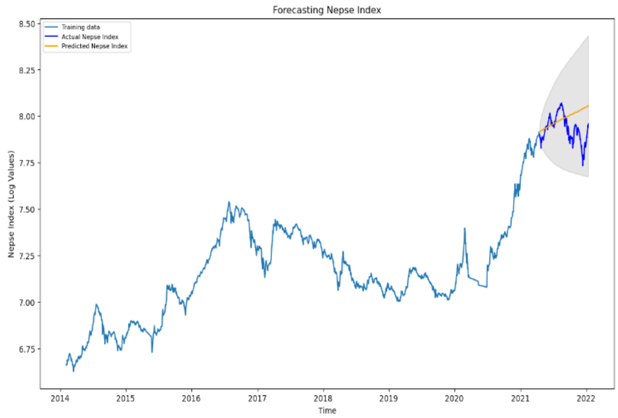

Figure 9: Actual NEPSE index Vs Predicted NEPSE index

5.3 Evaluating Forecast Results of NEPSE Index

In Table 1 below, the accuracy metrics for evaluating forecasts results were reported. Since, other error metrics measures in quantities; it cannot properly be compared between the forecasts of two differently scaled time series. So, the best forecast methodology was considered to be Mean Absolute Percentage Error or MAPE. Hence, the achieved result was 0.0104 ~= 1.04%, implying that the model was around 98.96% accurate in forecasting the next fifteen observations.

Table 1: Accuracy Metrics for NEPSE Index

Error Metrics	Obtained Values
MSE	0.0116
MAE	0.0819
RMSE	0.1076
MAPE	0.0104

VI. CONCLUSION

According to the aforementioned empirical research, the NEPSE index includes all time series components such as the Autoregressive (AR), Integrated (I), and Moving Average (MA). This paper details the steps involved in developing an ARIMA model for prediction of stock indices. With the results obtained, ARIMA model was discovered to be dependable, efficient, and capable of predicting short-term stock market changes and has been widely used in banking and economics. The identified models could be used by investors, stock analysts, and policymakers to forecast the daily NEPSE index and make policy changes based on it. This report has given clear signal to investors and analysts that the daily NEPSE

ISSN (online): 2581-3048
Volume 6, Issue 2, pp 80-85, February-2022
https://doi.org/10.47001/IRJIET/2022.602014

index will be affected by its prior values as well as random errors in the past. This indicates that both observed and random influences in the past will have an impact on the NEPSE index's future value. The proposed method from the above study is straightforward and can be used to any time series analysis. The fitted model's MAPE was relatively minimal, as it can be seen from the data above. Hence, it can be concluded that the present study of daily closing NEPSE index shows a gradual increase for the upcoming trading days.

ACKNOWLEDGEMENT

The author wishes to express his paramount gratitude to Mr. Madan Kadariya, Assistant Professor (HoD, Department of Information Technology Engineering), Nepal College of Information Technology (NCIT), Lalitpur, Nepal. His scholarly guidance and constructive feedback have been a source of inspiration and motivation to complete this paper.

REFERENCES

[1] (2019). In R. B. Paudel, K. J. Baral, R. R. Gautam, & S. B. Rana, Basic Finance (p. 397). Kathmandu: Asmita Books Publisher & Distributors (P) Ltd.

[2] Nepal Stock Exchange Ltd. (n.d.). Introduction. Retrieved from Nepal Stock Web site: http://www.nepalstock.com/about-us/introduction

[3] Levine, R., & Zervos, S. (1998). Stock Markets, Banks, and Economic Growth. The American Economic Review, 537-558.

[4] A, M. A., & K, S. K. (2017). Forecasting National Stock Price Using ARIMA Model. Global and Stochastic Analysis, 77-81.

[5] Gaire, H. N. (2017). Forecasting NEPSE Index: An ARIMA and GARCH Approach. NRB: Economic Review Article.

[6] Prasad, P. C., Manandhar, D., Maharjan, L., Rajkarnikar, L., & Jaiswal, A. (2018). Stock Market Forecast Using Time Series Analysis. 9th National Students' Conference on Information Technology (NaSCoIT), 9-13.

[7] C, M., M, V., & Chillale, N. R. (2019). Analysis of Daily Stock Trend Prediction Using ARIMA Model. International Journal of Mechanical Engineering and Technology (IJMET), 1772-1792.

[8] Adebiyi, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Stock Price Prediction Using the ARIMA Model. UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, 105-111.

[9] Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice. OTexts.

[10] Brownlee, J. (2017). Introduction to time series forecasting with python: how to prepare data and develop models to predict the future. Machine Learning Mastery.

AUTHOR'S BIOGRAPHY

Avinash Maskey has completed Bachelor's degree in Information Management (B.I.M.) from KIST college of Management, Tribhuvan University, Nepal in 2018 and is currently pursuing Master's degree in Computer Information System (M.C.I.S.) from NCIT college, Pokhara University, Nepal. His research interest areas are time series analysis, machine learning, econometrics etc.

Citation of this Article:

Avinash Maskey, "Predicting NEPSE Index Using ARIMA Model" Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 6, Issue 2, pp 80-85, February 2022. Article DOI https://doi.org/10.47001/IRJIET/2022.602014

****

