Forecasting Mobile Teledensity Growth in India

Chinmay Raut, Ashutosh Pandya, Sukanya Kulkarni

Department of Electronics and Telecommunication Engineering

Sardar Patel Institute of Technology
Mumbai, India

***Abstract*— Telecommunication industry plays a very vital role for the overall economic growth of any country, by providing means to promote economic and regional development as well as international trade in the country. The number of mobile subscribers has seen a very rapid growth in the past decade. Today India has around 1172.44 million mobile phone users making it one of the most booming sectors. The purpose of this research is to analyze the various factors that affect the growth of the sector in India and predict the future rate at which the mobile teledensity is going to increase in our country throughout the decade (2020 - 2029). Various non-linear models have been experimented under assumed saturation level and accordingly, the best model for time series forecasting is selected. The model selected is ARIMA (Autoregressive Integrated Moving Average) and with the help of it the projected rate at which the base of the mobile subscribers is going to increase has been successfully implemented. These results will have very important implications for telecom companies providing them with the idea of how much the infrastructure is to be developed to match the increasing requirements.**

Keywords—ARIMA; Forecasting; Investment; MAPE; Teledensity; Timeseries

I. Introduction

Mobile Teledensity is the number of mobile connections for every hundred individuals living in a particular area. Mobile teledensity varies widely across the nations with the context of their demographic and geographic conditions. Teledensity in any country will have a significant correlation with the infrastructure present in the country. The number of mobilephone subscribers is increasing rapidly day by day, therefore it is very important to forecast the number of users in the next decade so that the companies that provide mobile services will have a clear idea about the market size and the role that is to be played by the companies to meet the corresponding infrastructural needs. Model ARIMA is chosen to forecast the telecommunication data [1]. The paper begins with the selection data sets and identification of features for the model ARIMA [2] and its forecasting technique [3] to predict the numbers of subscribers and foreign investment. Periodic data is mostly a set of data which we get after observation over a large period of time. Therefore, we can say that, a large period of time is required to form periodic data. Thus, we can say that the data we get after periodic observation can be taken into account for the process by which decisions are made and also as a very efficient tool for time forecasting [4]. ARIMA is fully capable of not taking into consideration the independent variables that cause errors in time forecasting. The ARIMA model is an excellent choice to predict the results for a time-based series which is mostly based on data which is of variable type, whose behavior is observed for a sufficient amount of time. ARIMA model is completely capable for ignoring the independent variable for the model which we are using where we are making use of the present as well as the values obtained in the past to get a result that is an accurate short term forecasting the method [5]. Other techniques which can be used for time based series forecasting include algorithms that are based on Neural Network and are able to analyse patterns of data to give us the most efficient result from that data about which we do not have any idea, although the problem with neural networks is that it is difficult to say how many hidden layers will be required and also the time required to give the configuration of the parameters is too high, besides a lot of training data is required for good outputs and with increasing number of data sets the required time will also increase significantly [6]. The increase in teledensity also implies that the spectrum requirement will see a very sharp rise, these increase in spectrum [7]. Thus, different factors that affect the teledensity are taken into account and we obtained a result that matches the requirements. A very important aspect that affect the telecommunication sector is the internet usage thus it is very important to know the growth trend in the internet usage. The method by which networks are expanded is internet plus [8].The paper gives us an idea on the growth rate in teledensity in a very effective way and thus it also points on the need to increase the quality as well as the quantity of infrastructure so as to improve the service to the customer thus leading to improvement in the sector.

II. Methodology

A. Preparation of Dataset

The required data in this case is the data for Growth in teledensity in India from the year 2001 to 2019 is obtained from the official website of the Telecom Regulatory Authority of India (TRAI) and Open Data Government Platform. The data which we have got is in the form of a time-based series. The dataset includes attributes such as growth rate of teledensity, investment in the sector, classification on the basis of urban and rural areas, Private and public service

TABLE I: DATASET GROWTH IN TELEDENSITY FROM THE YEAR 2001

Year	Rural Area	Urban Area	Wireless	Wireline	Public sector	Private sector	Overall Teledensity
2001	0.93	10.37	3.23	0.35	3.2	0.38	3.58
2002	1.21	12.2	3.65	0.64	3.64	0.65	4.29
2003	1.49	14.32	3.87	1.24	4.04	1.07	5.11
2004	1.55	20.79	3.76	3.26	4.27	2.75	7.02
2005	1.73	26.88	3.77	5.18	4.74	4.21	8.95
2006	2.34	38.28	3.61	9.13	5.48	7.26	12.74
2007	5.89	48.1	3.61	14.61	6.32	11.9	18.22
2008	9.46	66.39	3.44	22.78	6.94	19.28	26.22
2009	15.11	88.84	3.27	33.71	7.71	29.27	36.98
2010	24.31	119.45	3.14	49.6	8.99	43.75	52.74

providers and also on the basis of the transmission medium. Table I shows the small instance from the teledensity dataset.

Table II shows the investment in the sector from the foreign territories. Thus, giving a quantitative idea of what should be the approach of the Indian brands should be to make a decent profit share in the sector.

TABLE II: INVESTMENT IN TELECOM SECTOR

Year	Total Investment in million dollars	Investment in Telecomm Sector	Share of Telecomm Sector in%
2001	2463	178	7.12
2002	4065	873	21.48
2003	2705	192	7.08
2004	2108	86	3.96
2005	3219	118	3.68
2006	5540	618	11.15
2007	12,492	477	3.82
2008	25,575	1261	5.13
2009	31,960	2549	8.12
2010	25,834	2539	9.83

B. Pre-processing of the data

The Once the data set is ready the pre-processing of the data set will start. This pre-processing of data has five steps (a) Attribute reduction, (b) Cleaning of data (c) data loading (d) Numeric factors, (e) Date factors. The process of attribute reduction is the process of elimination of the unwanted attributes. In this case, we have only considered the year wise growth in teledensity. The attribute year is chosen to be time stationary. Thus, the attributes after reduction is observed in Table III.

TABLE III: INVESTMENT IN TELECOM SECTOR

Year	Growth in Teledensity
2001	3.58
2002	4.29
2003	5.11
2004	7.02
2005	8.95
2006	12.74
2007	18.22
2008	26.23
2009	36.98
2010	52.74

C. ARIMA model

ARIMA is very important and efficient technique to analyse time-based series. The ARIMA model makes use of values obtained in the past and also the current values to accurately forecast the values that can be obtained in the future. A time-based series is often regarded as a sequence which is random in nature and then it predicts the function that will best suit the purpose. The ARIMA model is based predominantly on the functions p, d and q. These functions are mostly explained as (a) p stands for the order in which autoregression is present, d is the order of difference in the sequence (c) q stands for moving average and its order and is the value which is observed at a given instances of time. From pyramid package we have used auto_arima() to detect these parameters automatically.

$$(1- \sum_{k=1}^{p} \alpha_k L^k) (1-L^d)X_t=(1+ \sum_{k=1}^{q} \beta_k L^k)\varepsilon_t \quad (1)$$

Where

ε_t= is an error term, usually a white noise with intensity

L= Time lag operator.

The formula for developing the model of the ARIMA model which has been utilized for this forecast is given in equation (2). The formation of different layers of arima model is mostly based on hidden layers [9]. The equation of the ARIMA model helps us to see the values of Xt which has a major dependence on the values of the past errors for the variable itself.

D. Error parameters of the model

MAPE stands for “mean absolute percentage error”, it is also described as “mean absolute percentage deviation” (MAPD). It is a measure of how accurate the prediction for time forecasting is done. The mean absolute percentage error (MAPE) is the most usual method that we use to determine error, and works best if no extremes are attached to the data (and no zeros). In statistics, test accuracy is used to indicate how close the predict values can be to the true or actual values. MAPE can be a substitute technique to analyze the error. MAPE is also capable of measuring the divergence between the actual value and the value that is forecasted and afterwards it will measure the mean of percentages which is governed by the equation (2). It measures the accuracy and is calculated as the absolute percent error for each period of time subtracted from actual values divided by actual values.

$$MAPE=\frac{\frac{\sum |A-F|}{A}}{N}\times 100 \quad (2)$$

Where A = Actual
F = Forecasted
N = Number of observations

The use of the MAPE which includes it's use as a loss function for Regression analysis is possible in both cases that is on a practical frame of reference and also from the frame of reference. This is possible because the existence of an optimal model can be proved for MAPE. Considerations of factors that impact the accuracy [10] is very important for the model to forecast the future values accurately and to minimize the error.

Another error parameter is RMSE. It stands for "Root Mean Square Error" is described as the deviation for the residuals. Residuals are the measure of by how much a regression line will vertically miss a data point. RMSE points towards the residuals' spreadness. In order for the model to have a good accuracy the value of RMSE is quite critical and it must be reduced to bare minimum.

$$RMSE=\sqrt{\frac{(Predicted-Actual)^2}{N}} \quad (3)$$

III. Results and Discussion

A. *Teledensity Forecasting*

The test predictions are carried out by stating the total number of years which are going to be predicted. In this scenario, we have forecasted the growth for this decade (2020-2029) whose results is effectively depicted in Fig.1 and Table IV depicts the details of the graph and also the predicted outcome for the growth of teledensity in India using ARIMA method.

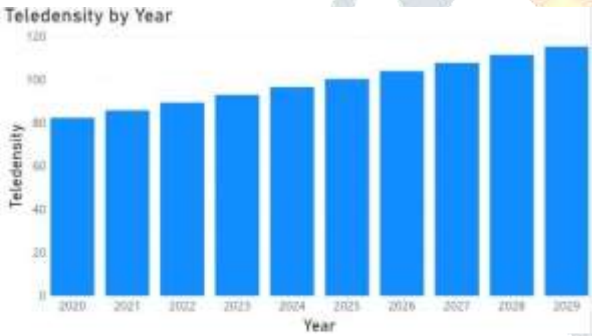

Fig 1: Next decade predictions % for growth rate in teledensity

B. *Factors contributing Teledensity*

There are various sectors contributing towards the teledensity growth in India. Contribution of such factors are depicted in the Fig. 2 effectively. It is quite clear from the Fig. 2 that population from the urban area with wireless technology are majorly contributing. Thus, the telecom companies will target the urban customers as the customer base is increasing day by day.

TABLE IV: NEXT DECADE PREDICTIONS % FOR GROWTH RATE IN TELEDENSITY

Year	Percentage growth in Tele density
2020	82.49
2021	85.91
2022	89.44
2023	93.05
2024	96.71
2025	100.42
2026	104.14
2027	107.88
2028	113.63
2029	115.39

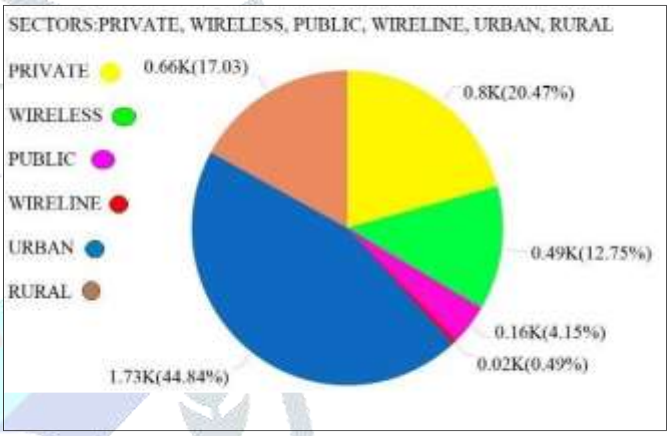

Fig.2: Sectors contributing to the growth in Teledensity

C. *Investments*

Investment in a particular sector plays a very important role, for that sector to excel. It is very important that the investment in that sector maintains a particular pace so that it can adjust to inflation. In case of telecommunication sector in India the investment from local market and brands was lacking till the middle of the previous decade as to when compared with similar countries. The investment from foreign territories thus becomes very crucial. The foreign players have their presence in India for almost 10 years now and they have seen a tremendous growth however it has till now not attained the level which it had intended at their arrival mainly because of lack of vibrant technologies. The foreign direct investment in India especially the telecom sector has been on a very steady part because of the removal of restrictions in the government policies the total investment in the sector is around 74% of the total investment in the sector hence to know how it will trend in the future is of a large importance.

Fig. 3 shows the forecasted total investment in the telecommunication sector which has been made from foreign territories. It shows that the investment from the foreign market will reach around 60,000 million dollars by the end of the decade thus showing that the telecom sector can be increasing and thus it creates employment opportunities.

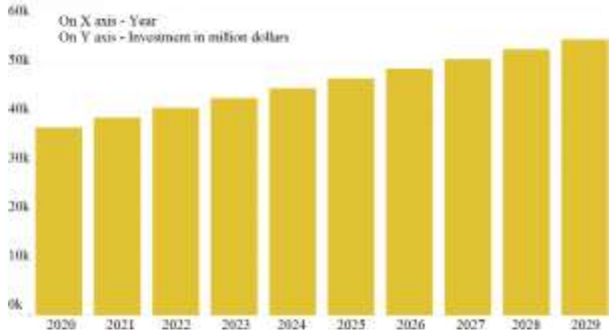

Fig. 3: *Foreign Investment (US $ Million)*

D. Internet usage

International telecommunication union (ITU) has made an estimate that the number internet user in the world is nearing 6 billion people buy the end of 2019.The the need to track the usage of internet in minutes become very important. The prediction which we have made shows that internet usage drastically differs between regions. Fig. 4 shows the increasing rate of internet usage in different countries. It is around 1600 minutes per person in developed nations. In Africa the average usage per user is expected to reach 400 minutes and in India it is going to reach 1000 minutes per person by the end of this decade thus we know that there will be 6.67% rise per year in India. Thus, this tells us how much development in infrastructure is required to support the requirement.

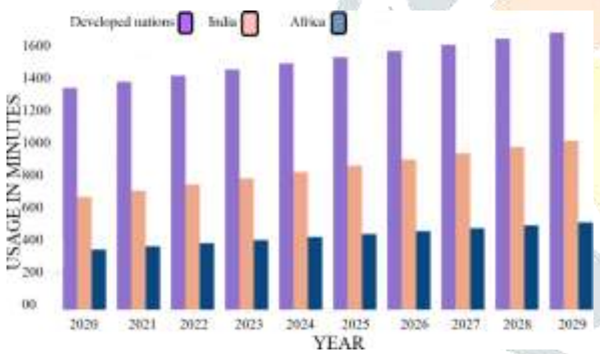

Fig. 4: Internet Usage in Minutes in different regions of the world

E. Accuracy of the model

The models accuracy will largely depend on the value of "Mean absolute percentage error" MAPE. The final value of MAPE is found to be 3.3580. So, the accuracy score of results obtained in this study is found to be 96.642%. The outcome of MAPE is shown in the following Fig.5.

```
def mean_absolute_percentage_error(y_true, y_pred):
    y_true, y_pred = np.array(y_true), np.array(y_pred)
    return np.mean(np.abs((y_true - y_pred) / y_true)) * 100
MAPE= np.mean(np.abs((y_true - y_pred) / y_true)) * 100
MAPE

3.3580346676923236
```

Fig. 5: The MAPE score

Another important error parameter for the model is "Root mean square Error" RMSE. In this model the RMSE score was found to be 2.3270% as depicted in Fig. 6. Thus, we can say that this model is sufficient to predict the future values as the value of the RMSE is also quite less.

```
RMSE=(np.sqrt(metrics.mean_squared_error(y_true, y_pred)))
print(RMSE)

2.327017932156626
```

Fig. 6: The RMSE score

IV. Conclusion

One of the important key aspects in the study of teledensity is time series forecasting of different data sets. In this paper we have a teledensity growth rate forecasting model that is implemented using the ARIMA model also we have proposed an approach that is based on applying the estimation of different errors so that the model becomes much more accurate. The results of the experiment confirm that the forecasting model had good accuracy, thus we can say that the model is successful implemented.

V. Future Scope

The research work was done using a data set that had samples of the previous fourteen years. The outcome of the project helped us to envision the future of teledensity growth in India. The results of the research will help us to understand the variation in mobile teledensity growth for the upcoming decade i.e. from 2020 to 2029. This research work can turn out crucial in making business decisions for telecommunication companies as they now have the future count so they can develop the infrastructure in accordance.

Acknowledgment

We take this opportunity to thank "Sardar Patel Institute of Technology, Mumbai " for the constant support and encouragement without which this research work would not be possible.

References

[1] N. S. Nalawade and M. M. Pawar, "Forecasting telecommunications data with Autoregressive Integrated Moving Average models," 2nd International Conference on Recent Advances in Engineering & Computational Sciences (RAECS), Chandigarh, 2015, pp. 1-6.

[2] A. Qonita, A. G. Pertiwi and T. Widiyaningtyas, "Prediction of rupiah against US dollar by using ARIMA," 2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), Yogyakarta, 2017, pp. 1-5.

[3] B. Siregar, E. B. Nababan, A. Yap, U. Andayani and Fahmi, "Forecasting of raw material needed for plastic products based in income data using ARIMA method," 2017 5th International Conference on Electrical, Electronics and Information Engineering (ICEEIE), Malang, 2017, pp. 135-139.

[4] J. Hanke and D. Wichern, Business forecasting, 1st ed. [Chennai]: Pearson, 2015.

[5] Ramesh Srivastava, Bhavya Sharma, Nand Kumar, Ranganath , "Analyzing and Forecasting the Mobile Teledensity Growth in India," 2015 International Conference of Advance Research and Innovation (ICARI-2015).

[6] Putu Sugiartawan, Sri Hartati,Time Series Data Prediction using Elman "Recurrent Neural Network on Tourist Visits in Tanah Lot Tourism Object, "October 2019 International Journal of Engineering and Advanced Technology (IJEAT),Volume no-9 Issue no-1.

[7] I. Ridwany and Iskandar, "Forecast of Spectrum Requirement for Mobile Broadband," 2018 12th International Conference on Telecommunication Systems, Services, and Applications (TSSA), Yogyakarta, Indonesia, 2018, pp. 1-5.

[8] Z. Jianqiu and Y. Mengke, "Internet plus and networks convergence," in China Communications, vol. 12, no. 4, pp. 42-49, April 2015.

[9] S. Xing and Y. Lou, "Hydrological time series forecast by ARIMA+PSO-RBF combined model based on wavelet transform," 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, China, 2019, pp. 1711-1715.

[10] X. Wang, W. Lee, H. Huang, R. L. Szabados, D. Y. Wang and P. Van Olinda, "Factors that Impact the Accuracy of Clustering-Based Load Forecasting," in IEEE Transactions on Industry Applications, vol. 52, no. 5, pp. 3625-3630, Sept.-Oct. 2016.