**Population Dynamics: Comparative Analysis of Mathematical Models and Machine Learning Techniques for Predicting Future Population Growth**

Sheena K J,

Research Scholar,

VIT Bhopal University

[sheena.kj2021@vitbhopal.ac.in](mailto:sheena.kj2021@vitbhopal.ac.in)

&

Jyoti Badge,

Assistant Professor,

VIT Bhopal University

[jyoti.badge@gmail.com](mailto:jyoti.badge@gmail.com)

**ABSTRACT**

The global population is increasing at a very fast rate, and it is likely to increase another two billion by 2050. India has also seen a prominent increase in demographical numbers across the states. India has already surpassed China in the Population this year. Keeping in view of the current situation, population dynamics has a high scope and importance. It deals with changes in population growth rates, age structures, distribution of people and these factors are closely linked to national and global development challenges and their solutions. Many population growth models have been developed in the past by demographers and scientists to study and predict the future population. Exponential, Logistic and Fibonacci population models are few among them. Fibonacci models are based on Fibonacci numbers in Mathematics and its research study was based on Rabbit population. Later it is generalized and extended to human population. Exponential model created a significant place among the population models and predicted that the world population will increase exponentially. Contradicting to this, Logistic population model said that the population can grow either increasing or decreasing based on the carrying capacity- a new term introduced by Verhulst. Recently, machine learning techniques are introduced to predict the population growth and it gives a more approximate and reliable predictions as compared to differential equation based mathematical techniques. This chapter goes deeply into modeling the human population growth of Fibonacci, Exponential and Logistic population models, and machine learning techniques. The data set is collected from 2011 census of 5 major Indian states, analyzed, compared, and predicted the future population till 2031 using the machine learning techniques and Mathematical population models. Among the machine learning techniques, Random Forest method has been used here to predict the population.

**Keywords:** Population growth, Exponential model, Logistic model, Fibonacci model, Random Forest method, Future population

**1. INTRODUCTION**

Many research studies have been conducted using Exponential, Logistic and Fibonacci population models mathematically. These mathematical predictions are accurate to some extent. This research study is an attempt to attain a more accurate and reliable prediction by comparing mathematical and machine learning methods.

Machine learning models are gaining utmost importance in the modern world. There are several machine learning methods available for predicting the human population – Linear regression, decision trees, artificial neural network, k-Nearest neighbors – to name a few. There have been studies in the recent past that talks about the comparison of 17 machine learning models[3], forecasting small area populations with Long Short- Term memory networks[2,5], study in Nepal that estimates the post-earthquake human migration within Nepal by tracing the human movement using their mobile phone data[4], another paper that explores supervised ML algorithms to identify the most suitable algorithm for predicting the population growth rate of a particular area[1], population prediction of Turkey with six different machine learning algorithms were done in 2017[6], population forecasting system using machine learning algorithm in Nigeria[8], a model that is used to predict the disease severity and outcome in Covid-19 patients[11] etc.

The data set is taken from 2011 census Govt.of India that deals with the population of first 5 most populated Indian States – namely Uttar Pradesh, Bihar, Maharashtra, West Bengal and Madhya Pradesh. As per the 2011 census, it is observed that 50% of India’s demographic growth is expected to take place in these 5 states and hence the projections of these particular states are of great importance. Hence the data set of these Indian states have been taken in this research study. Here Python programming and MATLAB programming are used to calculate the predictions of these 5 states from 2012 to 2031 using mathematical methods as well as machine learning methods.

**2. RESEARCH METHODOLOGY**

**2. 1 Exponential Growth Model**

Exponential growth model developed by Malthus [16] consider the rapid growth of human populations against the backdrop of limited resources. In fact, Malthus predicted a society plagued by misery and hunger.

This growth model is described by the differential equation = rN, where r = b – d (Fertility – Mortality)

On integrating, we get N(t) = N0 , r is the Malthusian factor which is a constant.

**2.2 Logistic Population Growth Model**

Formulated by Pierre Verhulst, this model [17] introduced a term called carrying capacity. Verhulst found that population growth not only depends on the population size, but also on how far this size is from the carrying capacity. As per Verhulst, population cannot always increase because there will be scarcity of resources like water, food etc. So, he predicted that population will decrease as and when the resources will run out of short. To summarize, the population growth rate decreases as the number of individuals increases. He formulated the equation as follows.

---------------(1)

a and b are vital coefficients.

On integration and applying the initial values, we get the following equations and with the use of these equations, we have created a MATLAB code to calculate the numbers.

Nmax =

N =

=

**3. DATA TABLES AND DIAGRAMS**

My first study was based on the Exponential and Logistic population models of 2011 census data set of 20 Indian states. The population of the year 2011 has been taken as the base year and predicted the population for the years 2012 to 2031 using both the Exponential and Logistic population models. Table 1 and table 2 depict the comparison of actual and predicted population using both the models for the said years whereas Table 3 tells us the population predictions for the years 2023 to 2031of all the 20 Indian states. The graphs have been plotted to show the comparison between Exponential and Logistic model and we can observe that the exponential model is more accurate and almost matches with the actual data in all the cases. Fig 1, Fig 2 and Fig 3 portray the actual, exponential and logistic curves of each state for the years 2012 to 2022. Fig 4 shows the population predictions of all 20 states for the upcoming years from 2023 to 2031.

**(Table 1- Population Predictions – Exponential Model vs Logistic Model – (2012- 2016)**

A screenshot of a chart

Description automatically generated

**[Actual(A) vs Exponential Predicted (EXP-P), Logistic Predicted(LOGI-P) (In thousands)]**

In the exponential model, the growth rate of each state is calculated using the 2011 census data and performed the MATLAB code to get the predicted population for each year for 20 different states. When we look at the growth rate, Tamil Nadu and Kerala are at the lowest growth rate of 0.0068 and 0.0069 respectively whereas the highest growth rate of 0.0213 and 0.0185 are recorded by NCT of Delhi and Bihar respectively.

**(Table 2- Population Predictions – Exponential Model vs Logistic Model – (2017- 2022)**

A table with numbers and lines

Description automatically generated with medium confidence

In the logistic model, the carrying capacity is calculated for each state based on the census data 2011 and performed the MATLAB code to get the predictions. When we calculate the carrying capacity, we can observe that Tamil Nadu and Kerala are at a rate of 0.02 whereas Jammu & Kashmir has a carrying capacity of 0.03. It varies from 0.02 to 0.04 depending upon the state population of the base year.

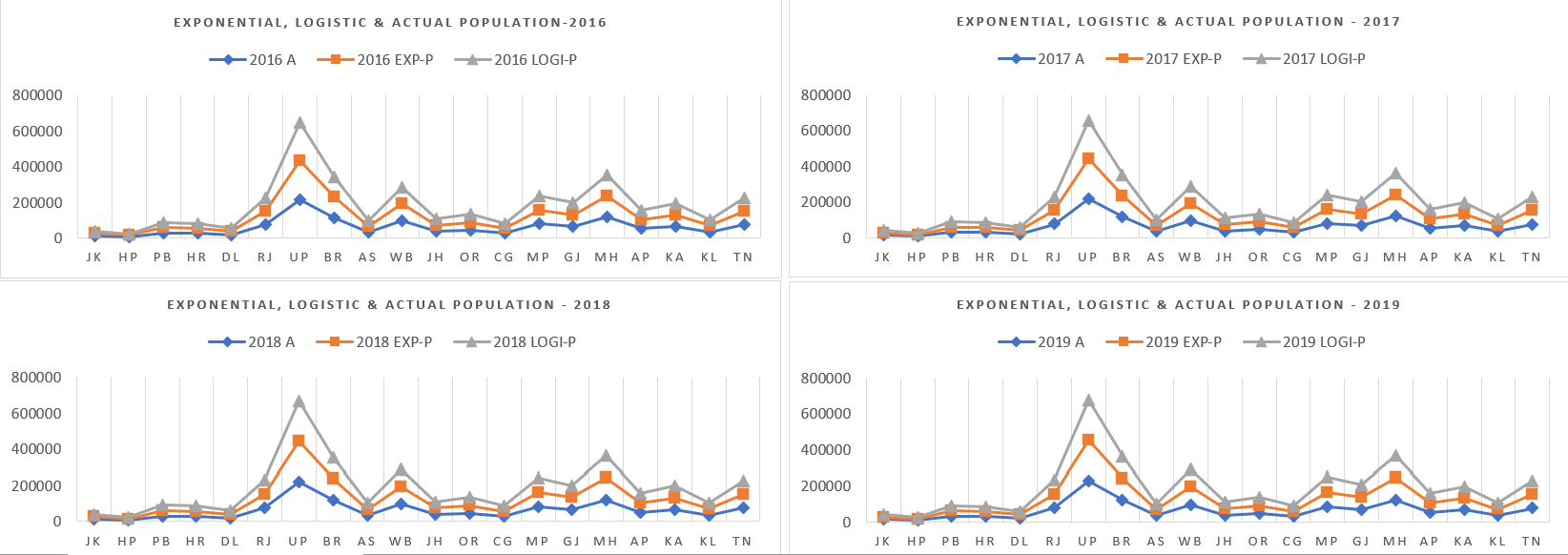
The following figures give us a detailed picture of exponential and Logistic data mapped against the actuals. Fig 1 shows the comparison of both the models with the actuals for the years 2012 to 2015 and there we can see the data is almost matching with the actuals in both the models, out of which exponential model is closer to the actual. On the other hand, fig 2 gives us a comparison of both the models with actuals for the years 2016 to 2019 and there is a noticeable difference in both the models from the actual data, Exponential data being the more accurate. When we observe fig 3, that is the comparison of the models with the actual for the years 2020 to 2022 and it is seen that the data varies significantly from the actual in both the cases, Exponential being the more accurate.

**(Fig 1-Comparison between Exponential, Logistic and Actual curves – 2012 -2015)**

A graph of a number of people

Description automatically generated with medium confidence

**(Fig 2- Comparison between Exponential, Logistic and Actual curves –2016-2019**)



**(Fig 3- Comparison between Exponential, Logistic and Actual curves –2020-2022**)

A graph of a graph

Description automatically generated with medium confidence

Based on the above observations and comparative study till 2022, the population is predicted for the upcoming years 2023 to 2031 using both the models and is given below in table 3. Fig 4 and 5 give a pictorial representation of the table 3 data. It is very clear that logistic population prediction is on a higher side than the exponential population.

**(Table – 3 – Prediction of Population – Exponential vs Logistic – 2023 - 2031)**

A table with numbers and symbols

Description automatically generated with medium confidence

**[Exponential Predicted (EXP-P) vs Logistic Predicted (LOGI-P), Numbers in Thousands]**

**(Fig 4 – Prediction of Exponential and Logistic Population curves – 2023 - 2026)**

**A group of graphs showing different sizes and shapes

Description automatically generated with medium confidence**

**(Fig 5 - Prediction of Exponential and Logistic Population curves – 2027 - 2031)**

**A group of graphs showing the growth of a company

Description automatically generated with medium confidenceA graph of different types of numbers and graphs

Description automatically generated with medium confidence**

Indian states abbreviations in the graphs:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Jammu &Kashmir | Himachal Pradesh | Punjab | Haryana | Delhi | Rajasthan | Uttar Pradesh | Bihar | Assam | West Bengal |
| JK | HP | PB | HR | DL | RJ | UP | BR | AS | WB |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Jharkhand | Orissa | Chhattisgarh | Madhya Pradesh | Gujrat | Maharashtra | Andhra Pradesh | Karnataka | Kerala | Tamil Nadu |
| JH | OR | CG | MP | GJ | MH | AP | KA | KL | TN |

**4. STATISTICAL TESTS**

**4.1 Chi-Square Test**

Chi-Square test of independence **[21]** has been conducted on the exponential and logistic models separately for each year starting from 2012 till 2022. The null hypothesis for the test - “There is no significant difference in the predicted and the actual values” and the alternate hypothesis - “There is a significant difference in the predicted and actual values”. The level of significance has been taken as 0.05 and degree of freedom is calculated as (c-1) (r-1) and that is 19. The critical value for the degree of freedom 19 is found to be 30.14 from the chi-square table. The following table gives us the overview of the chi-square values calculated per year for both the models. This test shows that the predicted population is very close to the actual population from 2012 to 2016, and then from 2016 to 2022 it starts to vary significantly from the actual in the case of exponential model. Whereas, in the case of logistic model, from 2012 to 2014 it shows a match with the actual population but starts to vary apart from the actual as we move from 2015 to 2022. The logistic model shows more fluctuation from the actual as compared to the exponential model in this research study.

**(Table 4 – Chi-Square Values – Exponential vs Logistic Models – 2012 - 2022)**

|  |  |  |  |
| --- | --- | --- | --- |
| Year | Chi Square value (Exponential model) | Chi-square value (Logistic model) | Critical Value |
| 2012 | 1.09 | 4.50 | 30.14 |
| 2013 | 3.97 | 10.62 |
| 2014 | 8.37 | 21.14 |
| 2015 | 14.36 | 35.77 |
| 2016 | 22.27 | 54.57 |
| 2017 | 39 | 79.47 |
| 2018 | 69.59 | 112.19 |
| 2019 | 116.13 | 149.33 |
| 2020 | 182.14 | 198.19 |
| 2021 | 269.41 | 350.99 |
| 2022 | 437.67 | 317.92 |

**4.2 Mean Absolute Percentage Error Test**

The Mean Absolute Percentage Error (MAPE) test **[1]** is calculated to understand the error percentage for each model. It is calculated using the formula (Actual – Predicted)/Actual \* 100 for the years 2012 to 2022 in the case of both the models. The data calculated is displayed in Table 5 state wise. This table tells a fair idea about the correctness of the model. Exponential model produces less error as compared to the Logistic model.

**(Table 5 – MAPE % Values – State Wise - Exponential vs Logistic Models)**

|  |  |  |
| --- | --- | --- |
| **Indian States** | **MAPE % (Exponential)** | **MAPE % (Logistic)** |
| Jammu & Kashmir | 0.73 | 3.18 |
| Himachal Pradesh | 0.32 | 15.46 |
| Punjab | 0.35 | 1.98 |
| Haryana | 0.35 | 4.15 |
| NCT of Delhi | 0.37 | 0.72 |
| Rajasthan | 0.55 | 0.47 |
| Uttar Pradesh | 0.50 | 0.13 |
| Bihar | 0.69 | 1.02 |
| Assam | 1.27 | 0.66 |
| west Bengal | 0.38 | 0.81 |
| Jharkhand | 0.39 | 1.31 |
| Odisha | 0.33 | 0.28 |
| Chhattisgarh | 0.38 | 0.77 |
| Madhya Pradesh | 0.46 | 0.44 |
| Gujrat | 0.26 | 0.16 |
| Maharashtra | 0.33 | 0.79 |
| Andhra Pradesh | 2.46 | 0.74 |
| Karnataka | 0.40 | 0.59 |
| Kerala | 0.34 | 0.77 |
| Tamil Nadu | 0.43 | 1.23 |

**5. FIBONACCI POPULATION GROWTH MODEL**

In 1202, Leonardo Fibonacci researched the breeding of a pair of rabbits under specific circumstances and he recorded his observations. According to him, the predictions can be made based on the previous year’s population of rabbits and it follows Fibonacci number series in Mathematics and hence the name Fibonacci Population Growth [18]. Later it is observed that the human population also can be predicted using this population model.

Let X(n) and Y(n) be the no. of adults and juveniles of a hypothetical population with the dynamics is given by the recursive equations, X(n+1) = X(n) + Y(n), Y(n+1) = X(n) with X(0) = 1 and Y(0) = 0.

The system of equations can be written in the form, X(n+1) = X(n) + X(n-1)

**6. RANDOM FOREST MODEL**

There are variety of Machine Learning methods that is been used in predicting the data recently. Out of that, Linear regression is the most commonly used method predicting the population growth that has been used very often in almost all research papers. Based on rigorous study of research papers and comparison of machine learning methods, it has been observed that Random Forest method gives 95% accuracy in population prediction as compared to the other methods.

Random Forest [5] is an ensemble machine learning algorithm. Ensemble learning is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. It belongs to the supervised learning technique. Random Forest is a classifier that contains several decision trees [7] on various subsets of the given data set and takes the average to improve the predictive accuracy of the data set. The advantages of Random Forest are it takes less training time as compared to other algorithms, it predicts the output with high accuracy, and it can also maintain accuracy when a large proportion of data is missing.

<https://www.javatpoint.com/machine-learning-random-forest-algorithm>

**7. COMPARISON OF ALL THE MODELS**

The exponential, Logistic, Fibonacci and Random Forest models have been used here to predict the human population from 2012 to 2031. Here we used mathematical methods and calculated the projections using MATLAB programming language for the first three models and Python programming language for the machine learning based Random Forest model. Tables given below shows the projections using three models and the actual projection by Govt.of India census of all the 5 states.

(<https://nhm.gov.in/New_Updates_2018/Report_Population_Projection_2019.pdf>)

The comparison between all four models and actual projections can be clearly seen in the below figures. It is observed that the actual projections and exponential model predictions are very close to each other and in fact in some cases they completely overlap with each other. Hence the future projections made are more accurate in exponential model than the other models.

|  |  |
| --- | --- |
|  | A graph with red line and blue line  Description automatically generated |

 A graph of a number of people

Description automatically generated with medium confidence

 A graph with red line and blue line

Description automatically generated

 A graph with red and blue lines

Description automatically generated

 A graph with red and blue lines

Description automatically generated

 A graph with numbers and lines

Description automatically generated

 A graph with red line and blue line

Description automatically generated

 A graph with numbers and lines

Description automatically generated

 A graph with red and blue lines and numbers

Description automatically generated

 A graph with numbers and lines

Description automatically generated

When we compare all the above models, we observe that exponential model gives a more accurate prediction as compared to the other models. The below table shows a comparison based on the behaviour of all the models.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sl.no | Properties | Exponential | Logistic | Fibonacci | Random Forest (ML model) |
| 1 | Model Behavior | Population will show an increasing trend | It shows increasing and decreasing trends | It depends on the previous year population | It is a random generated method and so behavior changes randomly. |
| 2 | Accuracy | Very good | Pretty good | Not so good | It is good |
| 3 | Reliability | Reliable | Reliable | Not reliable | Reliable to an extent |

Based on my two studies as described above, exponential models always give a more accurate and reliable prediction and hence can be considered as the best model.

**REFERENCES**

# [1]. Brintha Rajkumari S, P. C. (2020). Prediction of Population Growth Using Machine Learning Techniques. *European Journal of Molecular & Clinical Medicine*, 1871-1879.

[2]. Chian -Yue Wang, S. J. (2021). Regional Population Forecast and Analysis Based on Machine Learning Strategy. *Entropy*, 1-12.

[3]. Otoom, M. M. (2021). Comparing the Performance of 17 Machine Learning Models in Predicting Human Population Growth of countries. *International Journal of Computer Science and Network Security*, 220-225.

[4]. R Mahesh. R.Priyanga, S. (2019). Population Prediction Using MAchine Learning. *Journal of Emerging Technologies and Innovative Research*, 189-192.

[5]. J. Brownlee, “Random Forest for Time Series Forecasting,” November 2, 2020, <https://machinelearningmastery.co5]m/random-forest-for-time-series-forecasting/>

[6]. F.V. Şahinarslan, A.T. Tekin, and F. Çebi, “Application of machine learning algorithms for population forecasting,” International Journal of Data Science, vol. 6, no. 4, pp. 257–270, 2021.

[7]. E. A. Rady, H. Fawzy, and A.M. Abdel Fattah, “Time Series Forecasting Using Tree Based Methods,” Journal of Statistics Applications & Probability, vol.10, no. 1, pp. 229-244, 2021

[8]. Dr. N. Ashioba and N. N. Daniel, “Population Forecasting System Using Machine Learning Algorithm,” International Journal of Computer Trends and Technology, vol. 68, no. 12, pp. 40-43, December 2020.

[9]. F.V. Şahinarslan, A.T. Tekin, and F. Çebi, “Application of machine learning algorithms for population forecasting,” International Journal of Data Science, vol. 6, no. 4, pp. 257–270, 2021

[10]. E. A. Rady, H. Fawzy, and A.M. Abdel Fattah, “Time Series Forecasting Using Tree Based Methods,” Journal of Statistics Applications & Probability, vol.10, no. 1, pp. 229-244, 2021

[11]. F.V. Şahinarslan, A.T. Tekin, and F. Çebi, “Application of machine learning algorithms for population forecasting,” International Journal of Data Science, vol. 6, no. 4, pp. 257–270, 2021

[12]. C.Y. Wang and S.J. Lee, “Regional Population Forecast and Analysis Based on Machine Learning Strategy,” Entropy, vol. 23, no. 656, pp. 1-12, 2021

[13]. O. Folorunso, A. Akinwale, O. Asiribo, and T. Adeyemo, “Population prediction using artificial neural network,” African Journal of Mathematics and Computer Science Research, vol. 3, no. 8, pp. 155–162, 2010.

[14]. F. Rustam, A. A. Reshi, A. Mehmood, S. Ullah, B.-Won On, W. Aslam, and G. S. Choi, “COVID-19 Future Forecasting Using Supervised Machine Learning Models,” IEEE Access, vol.8, pp.101489-101499, 2020

[15]. M. Sh. Hajirahimova, L. R. Yusifova, “Experimental Study of Machine Learning Methods in Anomaly Detection,” Problems of Information Technology, vol. 13, no. 1, pp. 9-19, 2022.

[16]. Hironmoy Mondal, U. K. (2018). Mathematical Modeling and Predicting the Current Trends

of Human Population Growth in Bangladesh. Modelling, Measurement and Control D, 1 -7.

[17]. Bacaer, N. (2011). Verhulst and the logistic equation (1838). London: Springer - Verlag

London Limited.

[18]. Sigler, L. (2002). *Fibonacci's Liber Abaci.* Lewisburg: Springer, Newyork.