International Journal of Computer Trends and Technology
ISSN: 2231 – 2803 / https://doi.org/10.14445/22312803/IJCTT-V68I12P109

Volume 68 Issue 12, 40-43, December 2020

Original Article

Population Forecasting System Using Machine Learning Algorithm

Dr. Nwanze Ashioba [1], Ndubuife Nonso Daniel [2]

[1]*Delta State Polytechnic – Nigeria, Computer Science Department*
[2]*National Open University – Nigeria, Information Technology Department*

Received Date: 14 October 2020
Revised Date: 22 November 2020
Accepted Date: 24 November 2020

Abstract *- In every nation, there has been a platform to ascertain its citizens' exact number, population growth rate and make plans and decisions using the population information. The government spent a lot of resources on census enumeration. Unfortunately, in Nigeria, census enumeration has been embroiled in controversies. To overcome these problems, the existing systems face, the researchers have designed and developed a population forecasting system using a machine learning algorithm. The researchers adopted the Object-Oriented analysis and design methodology in developing the Population Forecasting System. The results have shown that Linear Regression Model has lower percentage error margins (between 0.76% and 1.09%) than the Average Projection Model and the Nature Fund Growth model with a percentage error margin between 4.73% - 1.43% and 0.9% - 1.89%, respectively.*

Keywords – *Population, Population Forecasting, Population Estimation, Machine Learning Algorithm.*

I. INTRODUCTION

A key component of successful countries is the ability to plan flexibly for the short and long term. Planning must be based on good information about the present situation, how important variables will change in the future, and how much faith is placed in future prediction. Members of a population rely on the same resources and are subject to similar environmental constraints, and depend on other members' availability to persist over time. Scientists study population by examining how individuals in that population interact with their environment. The information obtained from the population studies can be a determining factor in future planning.

A. Statement of Problem

Forecasting the trends in the human population is a complex problem. There are a number of uncertainties associated with the population of any country. Traditionally, statistical approaches have been used to forecast the demography of a country. But these approaches are not very suitable for predicting a chaotic system like the population. Firstly, the approaches make some assumptions, which are sometimes found unrealistic [1]. Secondly, statistical population forecast procedures cannot deal with the intrinsic chaos.

Additionally, population growth is difficult to forecast because of some events that can alter a location's population profile within the shortest period. Events like birthrate, death rates, migration. Migration can speedily alter a population, especially when it is a result of life-changing outcomes like war. In this case, an area's demographic setting is affected while another place's population is increased. Some mathematical models like the geometrical and arithmetical increase models may accurately forecast the population of either new cities or old cities over a decade using historical data while assuming some factors remain constant. However, they are widely accepted but are limited in their areas of application. Optimizing these models or adopting a more effective model is crucial so that a more accurate result can be achieved at the lowest cost.

II. LITERATURE REVIEW

[2] used controlled Artificial Neural Network to predict the population of Nigeria. Their prediction was done using training, validation, and test dataset in their research, Population Prediction Using Artificial Neural Network. Their study employed the artificial neural network for population prediction (ANNPP), which handles the inconsistent and incomplete nature of data, often experienced in mathematical and demographical models while carrying out population prediction. Three major dynamics of population changes (fertility, mortality, and migration) were considered in the input dataset. They divided the datasets into train, validation, and test data. The train data was presented to the supervised artificial network to approximate some known twelve target values of population growth rates. The results showed that ANNPP percentage accuracy ranged between 81.02 and 99.15%, while that of CCMPP percentage accuracy ranged between 64.55 and 86.43%. These results showed that the artificial neural network model performed better than the

demographic model. However, the model used by these authors is different from the model employed in this study, which will give greater accuracy.

[3] population and its impact on the level of unemployment in the least developed countries. The ordinary least square method was applied. The result clearly states that employment problems can be created only if the government does not set up adequate measures to control the population and distribute resources equitably, giving citizens a sense of belonging. Therefore, to provide the teeming population with the needed employment opportunities, there is a need to allocate resources and initiate employment-generation projects efficiently. However, the authors focused on the impact of population on unemployment in developing countries rather than population forecasting or designing a population forecasting system.

[4] in their research, population forecasting using geometric increase method. The authors critically analyze the geometric increase in population forecasting and the basic models for geometric change in population size. According to this model, it is assumed that the rate of increase in population growth is proportional to the present population. This method accurately describes the continuous and cumulative nature of population growth. However, this method has a major drawback because it is time-consuming and has a higher percentage error.

[5] in the paper "Fine-Scale Population Estimation by 3D Reconstruction of Urban Residential Buildings aimed to obtain fine-scale population distribution based on the 3D reconstruction of urban residential buildings with morphological operation using optical High-Resolution (HR) image from Chinese No.3 Resources Satellite (XY-3). In this research, building extraction and height retrieval were combined to construct a 3D model and estimate the population. This study shows that the fine-scale population estimation in the complicated urban landscape. Though the building distribution is accurately structured, population estimation is a major error because of the housing occupancy rate. There is an uneven distribution of people at the multi-layered building. Though the research can be applied to some enclosed locations, it cannot be used for a wider projection due to its complexity.

[6] investigated the relationship between population dynamic and economic growth in Nigeria using time-series data spanning from 1970-2014. The data analysis was done using the least square estimation technique from variables they considered [e.g., fertility, mortality, and net-migration]. They proposed that these variables are inversely related to economic growth during the period under investigation. As the concern about our growing population draws more economic attention, the government is advised to check the alarming fertility rate in Nigeria so that the population will not be a problem in our already shrinking economy. They solely focused on the population dynamics and economic growth in Nigeria but not on population forecasting.

[7] the paper Population Estimation in Singapore Based on remote Sensing and Open Data, investigate the applicability and effectiveness of using satellite images from google earth to extract building footprint and population estimation. Volunteered Geographic Information (VGI) was utilized as ancillary data for building extraction, while open data such as Open Street Map was employed to enhance the extraction process. The result shows that the method discussed can produce a high accuracy result for estimating building heights, area, and volume. However, the population's demographic profile was not considered, and the satellite imagery of the area under investigation cannot be used to estimate the population of people living in that area under review.

[8] presented a paper titled "Model and Reliable Projection." In his research, Nigeria's projection was made from 1991 to 2050 using the Experimental Growth Model (EGM) and Logistic Growth Model (LGM). These two models were combined using Average Projection (AP). The result obtained shows that AP is better than actual or official projection. Comparatively, several models were analyzed, but Machine Learning was not used.

[9] adopted the use of Predictive Machine Learning Regression, attempted to model the Nigeria population. They analyzed and compared different models that predicted the population of Nigeria. In this research, different predictive models to characterize Nigeria's population were developed using the machine learning regression method. The best of the models was selected and used to predict Nigeria's population up to the year 2050. By 2050, all things being equal, the population was predicted to be 400,000,000. MATLAB toolbox was used in the implementation of this research work. Though the various models analyzed yielded expected results, but in conclusion, predictive regression was adopted. However, they did not design any system that can accurately forecast a population-based on their adopted model. Rather, the only analysis was done.

[10] developed a population projection using mathematical curves. They developed a simple mathematical approach to predict India's population in the long term without an absolute justification of a particular assumption such as fertility, mortality, and migration. The researchers adopted the logistic and Guassian curves while utilizing the decadal population dataset from 1901 to 2001. They compared the output of the two curves output, which indicates that the Gaussian curve is more reasonable than the logistic curve. This study established that the population projection of male and female obtained separately shows that the female population will be more than males in the future.

[11] presented a paper on using Machine Learning Algorithm in Population Forecasting: Turkey Example. In

their research, different machine learning algorithms were used to forecast the population. Models were trained using 1595 different demographic indicators of 262 different countries between 1960 and 2017. When the performance of algorithms was compared, the ARIMA model was the most successful among all models. After analyzing each, their result showed that Machine Learning performs better than other demographic models. They concluded that Machine Learning Algorithms give more meaningful results with the big dataset. Though the ARIMA model gave the best result from their study, they proposed that the regression model can be used in future studies because it can better predict a larger dataset.

[13] developed an intelligent demographic forecasting system using fuzzy time series. The author analyzed things that affect economic changes and how they are interconnected with demographic processes. According to the author, the study's main aim is to develop an intelligent demographic forecasting system that supports demographic situation management decisions based on predictions. To achieve this aim, the author modeled and develop a demographic prediction technique. To achieve The smallest AIC and the diagnostic test performed on the model using the Box-pierce test show that the model is fit enough for the data. The authors conclude that with the current population growth rate in Nigeria, her net population my surpass the United States by the year 2050. This is because the forecast was found to have a gradual, constant increase in the population growth rate trend up to the year 2050.

]12] studied the development of a small area population estimation model for the densely populated metropolitan area and its application using Metro Manila as a case study. Their main objective was to develop a linear regression-based small area population estimation model using recent census data of Metro Manila. The authors the above objectives, the author adopted a fuzzy time series for estimating model parameters. The analysis of this series involves constructing a mathematical model of time series of observation of real processes. In conclusion, the author developed an intelligent system that can forecast a population, and the forecast result for different indicators can be obtained. However, the fuzzy time series is not a machine learning model, and some demographic indicators were omitted.

[14] carried out the study to find a trend in demographic changes, fit a model, and forecast Nigeria's population. They analyzed that population dataset obtained from the World Bank Data. The authors used both time and autocorrelation plots to assess the stationarity of the data. They also used the Fuller test to test for the unit root, while the Ljung box test was used to check for the fitted model fit. Time series were used in the data analysis. From the model's analysis, the authors agreed that ARIMA (2,2,0) is considered the best model since it has the smallest AIC(Akaike Information Criteria). The Arima function gave a population growth because it was the best fit model out of all the other selected models because it concluded that the population would increase exponentially up till 2050, and the model can project the population based on age, gender, and income. However, the study was not conducted using machine learning.

III. METHODOLOGY

We adopted the machine learning Linear Regression Model. The actual population data starting from 2006 as obtained from the Nigeria Bureau of Statistics and National Population Commission was used. The analysis, design, and implementation were done using Python Programming Language.

IV. RESULTS AND DISCUSSIONS

Table 1. Comparison between the actual and predicted population (Linear Regression Model)

Year	Population (Actual)	Population (predicted)	Error	%Error
2006	140431790	139368400	1063390	0.7630065
2009	154581566	154942650	-361084	0.2335880
2010	159608173	160134066	-525893	0.3294900
2014	181403148	180899733	503415	0.2775117
2015	187301926	186091150	1210776	0.6464301
2016	193392517	191282566	2109951	1.0910199
2020		212048233	----------	00000000
2030		263962399		00000000
2050		367790732	----------	00000000

Table 2. Result obtained From Nature Fund Growth Model

Year	Population (Actual)	Population (Predicted)	Error	%Error
2006	140,003,542	142,663,609	-2660067	1.899
2011	153,818,985	155,215,573	-1396588	0.908
2016		168,997,727	0000000	0000
2021		185,674,205	0000000	0000
2026		203,996,494	0000000	0000
2031		224,126,714	0000000	0000
2036		246,243,368	0000000	0000

Table 3. Sample Error obtained from Average Projection Model

Year	Population (Actual)	Population (Projected)	Error	% Error
2000	117.332	122.877	5.545	4.73
2005	136.443	139.611	3.168	2.27
2015	177.723	182.202	4.479	2.46
2030	257.438	262.599	5.161	1.97
2050	404.375	398.588	5.787	1.45

When the result (as shown in Table 1) is compared with recent models in (Augustine 2014) and in (John 2017), the percentage errors obtained in the two types of research lies between 0.9% and 1.8% and 4.73% and 1.45% respectively for some selected years. This is very significant compared to the new model's errors, as shown in Table 1, which varied between 0.76% and 1.09%. From Table 1, this study demonstrates a strong correlation between the actual and predicted population in each of the years under consideration. The result partly agreed with the previous studies but gave qualitative support to a linear regression model, especially when solved using a machine learning algorithm. The new model, therefore, performs better than the existing models.

V. CONCLUSION

The researchers have modeled the population of Nigeria using Machine Learning Linear Regression Model. The model was compared with some already existing models. The results that by 2050 the population of Nigeria will be about 365 million.

REFERENCES

[1] N. Keilman, D. Pham, and A. Hetland. Why population forecasts should be probabilistic – illustrated by the case of Norway. Demographic Research, 6, (2002) 409–453.

[2] S.O. Fulurunso, A.T. Akinwale, O.E. Asiribo, and T. A. Adeyemo. Population Prediction Using Artificial Neural Network. African Journal of Mathematical Computer Science Reserve, 3(2010) 155-162.

[3] I.An Ilegbinosa, O.L Moses, and Praise. Population and its Impact on the Level of Unemployment in the least Developed Countries; An Appraisal of the Nigerian Economy. Arts and Social Sciences Journal. 6, (2014) 277-289.

[4] B. Lovindraj, M. Boyjoo, D, Bundhoo, O. Ramgoolam, and I. Seburn (2015) Population Forecasting Using Geometric Increase Method. https://slideplayer.com/slide/5903511 Retrieved 21st August 2020.

[5] C. Lin, W. Liu, Y. Tian and S. Wang Fine-Scale Population Estimation by 3D Resolution of Urban Residential Buildings. Sensors (Basel)16(10) (2016) 1755.

[6] H. Aidi, C. Emecheta, and I.M. Ngwudiobi (2016). Population Dynamic and Economic Growth in Nigeria; Journal of Economic and Sustainable Development 7(15) 1-23.

[7] C. Lin, W. Liu, Y. Tian and S. Wang Fine-Scale Population Estimation by 3D Resolution of Urban Residential Buildings. Sensors (Basel)16(10) (2016) 1755.

[8] N.I John. The Population Question in Nigeria. Asian Research Journal of Mathematics, 5 (2017) 1-10.

[9] O. Odunayo, E. Oduntan, and O. Olawumi, (2018). Using predictive Machine Learning Regression Model to Predict the Population of Nigeria. Anale. Seria Informatica.. XVI Fasc. 2-2018. Annals. Computer Science Series. 16th Tome 2nd Fasc.-2018.

[10] P. Brijesh, Gunjan Singh and Sandeep Singh. Population Projection using Mathematical Curves: Will Indian Population Stabilize by 2050. Demography India, (Special Issue) (2018) 25-35.

[11] V. S. Fatih, T.T. Ahmet and C. Ferhan Machine Learning Algorithm to Forecast the Population: Turkey Example. International Engineering and Technology Management Summit 2019 – ETMS. www.researchgate.net/publication/33714439

[12] R. Monorom, J. L. Soliman, N. S. A. Lopez, A. M. Fillone, & J. B. M. Biona Development of Small Area Population Estimation Models for a Developing, Densely Populated Metropolitan Area and its Applications: A case study of Metro Manila. Engineering and Applied Science Research, 47(2), (2020) 206-215.

[13] Z. Jabrayilova Development of Intelligent Demographic Forecasting System. Eastern-European Journal of Enterprise Technologies. 101 (2019) 18-25.

[14] O.M. Olayiwola, G. O. Lawal, A. A. Amalare, S. A. Agboluaje and J. O. Fantola, Modeling Nigeria Population Growth Rate. Journal of Advances in Mathematics (2015) 10 (6).