USING PREDICTIVE MACHINE LEARNING REGRESSION MODEL TO PREDICT THE POPULATION OF NIGERIA

Odunayo Olanloye [1], Esther Oduntan [2], Olawumi Olasunkanmi [3]

[1] Emmanuel Alayande College of Education - Nigeria, Department of Computer Science
[2] Federal Polytechnic, Ilaro - Nigeria, Department of Computer Science
[3] Ladoke Akintola University of Technology - Nigeria, Department of Computer Science and Engineering

Corresponding Author: Odunayo Olanloye, olanloyeo@yahoo.com

ABSTRACT: In any nation, there is always a government in place charged with the responsibly of governing the citizens of the country and for effective governance, there should be plan based on the number of citizens (population) in such country. Government therefore spends a lot on census exercises which is meant to count the number of people living in the country. Census exercises in Nigeria were characterized by various abnormalities resulting into inaccuracies in the results obtained. Series of research work have been published on how to solve this problem. In this research, different types of predictive models to characterize the population of Nigeria were developed using machine learning regression method. The best of the models was selected and used to predict the population of Nigeria up to the year 2050. By 2050, all things being equal, the population was predicted to be 400,000,000. The instrument used for the implementation is Matlab modeling toolbox. The research work will serve as a very useful tool in the area of population prediction and assist the government in her future plan.

KEYWORDS: Algorithm, Artificial Intelligence, Government, Machine Learning, Population, Prediction.

1. INTRODUCTION

The Artificial Intelligence (AI) is one of the major areas in computer that has been used to solve various type of real life problems. Its application is becoming so wide to extent that it now cut across various disciplines. For instance, AI has been found to be a useful tool in solving problem in the area of Medicine, Engineering, Natural Sciences, Applied Sciences, Agriculture, Language etc. Hardly can we identify a particular field or discipline where AI is not found useful. One major branches of AI is Machine Learning (ML).

John, Brian and Aoife [JBA14] defined ML as an automated process that extract pattern from data. Bontempi [Bon13] defined machine learning as the domain of computational intelligence which is concerned with the question of how to construct computer program that automatically improves with experience. To have a functioning ML algorithm, there must be sufficient data from where the algorithm acquires its experience to solve similar problems in future.

So, Machine Learning is that field of AI where the program acquires series of experience from the historical data and applies such experience to solve future problems. It is obvious that ML algorithm depends on the historical data for its optimum performance.

One characteristics of man is that he has natural desire to make enquiries about likely future occurrences. In fact it is a natural human desire to predict the future events. After decades of research and development, Computer Science is now a point where predictive algorithm as one of the areas of ML is becoming more important and indispensable. Hence, many technology including fast computer power, cloud computing, voice recognition, mobile computation are coming together to make this possible ([Hal16])**.**

Predictive algorithm is a scientific idea of empirically establishing a relationship between the historical set of data which can then be used to make future decision in an attempt to solve some real life problems. Research has therefore shifted to the area of forecasting or predicting the likelihoods future events making use of the series of available data (big data) which are made available through modern technologies. Hal ([Hal16]) stated that the action of predictive algorithm on big data will generate predictive model. Hence predictive model is the product of predictive algorithm and big data.

Big data + predictive algorithm = predicted model.

Predictive modeling has been applied to solve various problems by various researchers. For instance, it has been applied to predict the weather condition, the rate of spread of diseases, birth rate, death rate, rate of accidents on roads etc and in population prediction.

Our research work is to use machine learning predictive regression model to predict the population of Nigeria. Though some authors have applied

different predictive model in population prediction but their works were based on statistical approach or other types of predictive model.

1.2 Statement of the Problem

There are many countries all over the world and for a country to continue to be in existence there must always be a government that has the sole responsibility of governing the citizens of the country. In an attempt to govern effectively, one of the major steps that should be taken by the government is to have a comprehensive idea about the number of people they are governing i.e the population of the country must be known by the government. This gave birth to the census activity which has to do with counting the number of heads in a particular country in a particular time. Odewumi ([Ode00]) as quoted in ([Ban09]) while referring to 1970 United Nation National Draft recommendation on population census defined census as the total process of collecting, compiling, analyzing or publishing or otherwise disseminating demographic, economic and social data at specified time to all persons in a country or in a well delimited part of a country. Bangbose ([Ban09]) defined population census as a terminology now restricted to complete enumeration of human population legally conducted at regular intervals, often every ten years. Unfortunately there has always been one problem or the other in arriving at accurate census data or census result despite the billions of naira that is being spent on the exercise. How do we now solve or minimize this problem such that government will be able to predict the population of her citizens without having to spend millions or billions of naira on census exercise which eventually produce inaccurate result.

2. LITERATURE REVIEW

John ([Joh17]) presents a research work titled models and reliable projection. In his research work, projection of Nigeria population was made for the range of years 1991 to 2050 using experimental growth model (EGM) and Logistic Growth Model (LGM). The two models were combined using average projection (AP). The result obtained shows that the AP is better than the actual or official projection. The intent of this research work was to compare the various models. Machine learning algorithm was not used.

Olatayo and Adeboye ([OA13]) were able to predict the population of Nigeria through births and deaths. With the use of regression analysis, the research work establishes the fact that death and the birth rate

have significant effect on the population growth in Nigeria. The approach did not use any machine learning algorithm but rather a statistical approach.

Folorunso et al ([F+10]) also used Neural Network to predict the population of Nigeria.The method used training, validation and test data set. Though it was an AI based method but quite different from regression model.

Adebayo, Nneama and Rotimi ([ANR09]) presented a research work on spatial predictive model for malaria in Nigeria. A predictive model was actually developed by making use of Geographical Information System. The model will actually help in prevention of Malaria and also reduce large amount of money being spent by government, industries, private and public organization, schools as well as individuals in the diagnosis and treatment of malaria. A predictive model was developed by the authors but not actually to predict the population of Nigeria but only serve as a decision support system for all the stakeholders in the health sector. The method used has nothing to do with Machine Learning but rather a GIS based method.

Andre et al [A+17] was able to compare machine learning algorithm to build a predictive model for detecting undiagnosed diabetes. They were able to develop predictive model for detecting undiagnosed diabetes and were able to compare the performance of different machine learning algorithm. The instrument used was artificial neutral network and logistic repression. The research developed a machine learning predictive model for diabetics and not in the area of population prediction.

Aditi and Anjali ([AA14]) develop a smart home using predictive algorithm. Smart home according to the authors is a working specie that interact in a natural way and adapts to the immediate enviroment. The system was able to automate activities that would otherwise be manually performed by the inhabitants. A predictive algorithm was developed for smart home but the work has nothing to do with the prediction or forecasting of population

Vilalta et al ([V+02]) developed prediction algorithm in management of computer system. The system was so developed to alert the user of particular failure. The result shows that predictive algorithm will be able to identify critical events. The system could not predict the population in any form.

Going by the literature, it is obviously clear that seasonal authors though have worked in the area of predictive algorithm yet; they have not done enough in the area of population prediction. The few that worked in the area did not actually use machine learning prediction regression model/ algorithm but rather statistical or other models. For the purpose of

this work, we used predictive machine learning regression model.

3. METHODOLOGY

We developed machine leaning regression model with actual population data from the year 1900 to 2006, to obtain different models and the best of them was selected and extrapolated to predict population growth till 2050. Matlab Machine Learning toolbox was used for implementation.

4. RESULTS AND DISCUSSIONS

Figure 1 shows the actual and the predicted population side by side. The model shows that the predicted population for Nigeria in 2050 will be about 400,000,000.
Figure 2 shows the predicted population according to the age group. The age distribution group 0 to 15 years is the largest and the age distribution 105 to 120 years has the least.
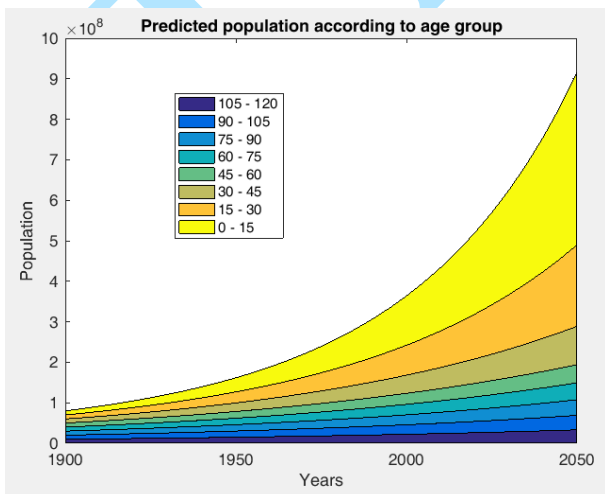

Fig. 1. Predicted and actual population

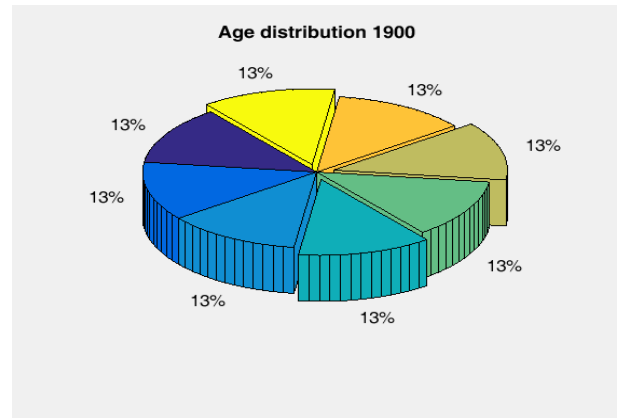**Fig. 2. Predicted population according to age group**

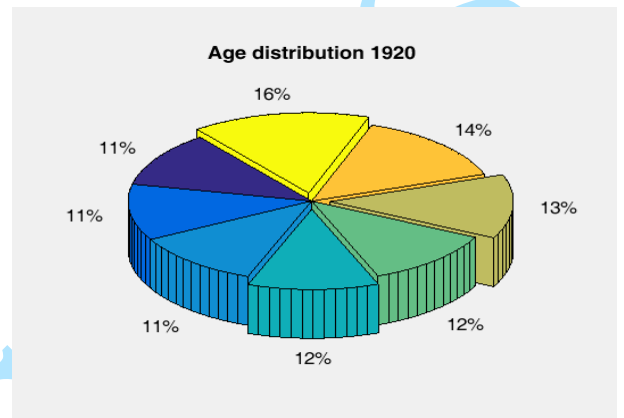**Fig. 3a. Age distribution in the year 1900**

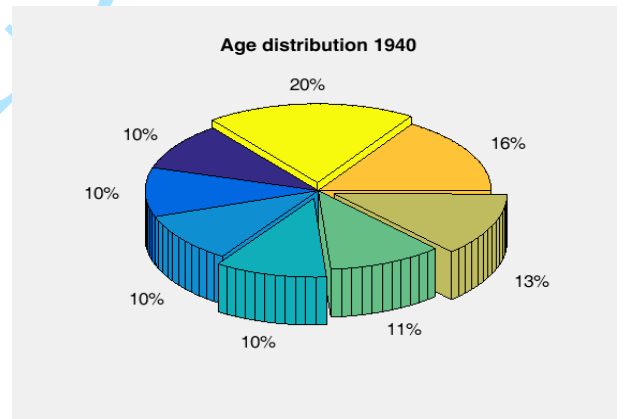**Fig. 3b. Age distribution in the year 1920**

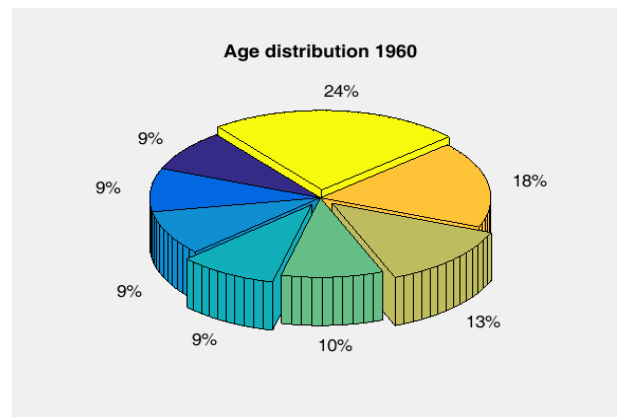**Fig. 3c. Age distribution in the year 1940**

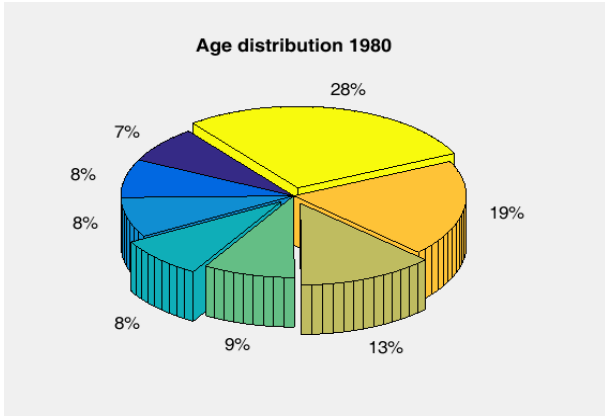**Fig. 3d. Age distribution in the year 1960**

Fig. 3e. Age distribution in the year 1980

Fig. 3f. Age distribution in the year 2000
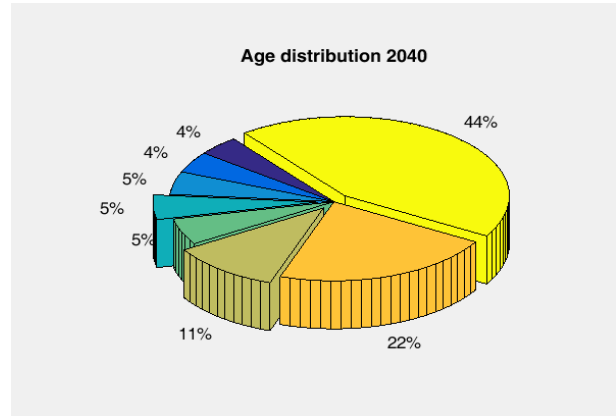

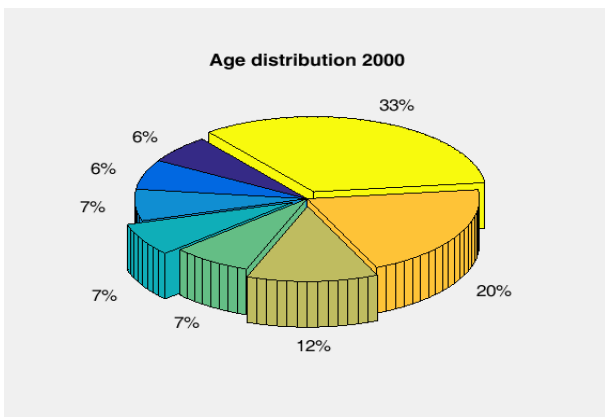**Fig. 3g. Age distribution in the year 2020**

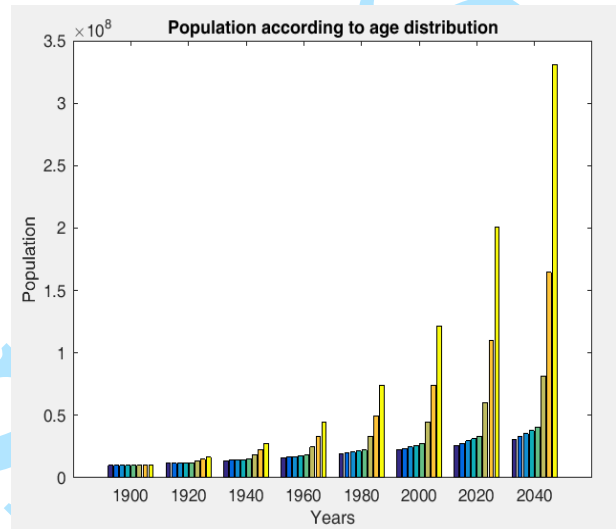**Fig. 3h. Age distribution in the year 2040**

Fig. 4. Population according to age distribution
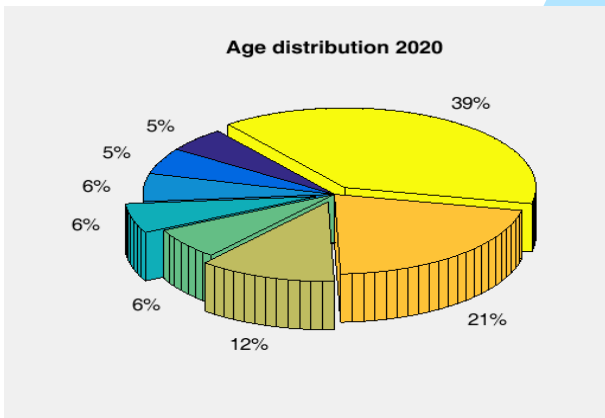
Figure 3 (a to h) explain further the comparison of age distribution in term of percentage on yearly basis ranging from year 1900 to 2040 with an interval of 20 years. For each of the year, the children under the age of 15 years has the largest population.

Figure 4 explain the age distribution further. The actual and the predicted population data were compared in table 1 and the differences (error) appears to be very insignificant. This shows that the error recorded were almost zero.

Table 1: Comparism of actual and predicted population data (Sample)

Year	Population (actual)	Population (predicted)	Error	% Error
1921	18,720,000	18,720,02	- 20.0000	1.06837e-5
1931	20,956,000	20,956,015	- 15.0000	7.157855e-7
1952	30,402,000	30,402,017	- 17.0000	5.591737e-7
1965	55,670,000	55,669,990	10.0000	1.7963e-7
1972	78,927,000	78,926,992	8.0000	1.013595e-7
1984	96,684,000	96,683,995	5.0000	5.171486e-8
1994	101,900,000	101,900,010	- 10.0000	9.83543e-8
1999	110,650,000	110,650,012	- 12.0000	1.084501e-7
2006	140,431,000	140,431,000	0.0000	00000000
2030	-------------	260,000,000	---------	00000000
2050	-------------	400,000,000	---------	00000000

Table 2: Sample Error obtained in [Jhn17]

Year	Projected Population	Actual population	Error	% Error
1995	108.425	102.066	6.419	5.92
2002	139.611	136.443	3.16	2.27
2030	262.599	257.438	5.161	1.97
2050	398.588	404.375	5.787	1.45

When our result is compared with recent models - in Folorunso ([F+10]), the percentage errors obtained lies between 0.01016 and 0.01034 for Nigeria population prediction in based on age distribution. This is quite significant compared to errors obtained in the new model as shown in table I. Also in John ([Joh17]), the predicted population is very close to that obtained from the new model. The research work finally predicted that by the year 2050, the population of Nigeria will be 404,375,000. This is very close to our prediction of 400,000,000 despite the fact tha thet methods used were quite different. Again, it was observed that the percentage errors obtained in this research work as shown table II were quite significant compared to that of the new model shown in table 1.

The new model therefore performs better than the existing ones.

5. CONCLUSIONS

In this research work, an attempt was made to model the population of Nigeria using Machine Learning regression model. Different models were developed and the best that characterized the population of Nigeria was selected. Our model reveals that by the year 2050, the population of Nigeria, will be about 400,000,000 with the children of under the age of 15 years having the highest population and the very old population group of age 105 – 120 having the least population. The research work will provide very useful information for the government and assist in preventing billions of money being used in carrying out population census that will not eventually yield any useful result. Hence, the research work will go a long way to contribute to the national development.

This research work will sensitize the government at all level to make adequate provision for the growing population. The population of children under the age of 15 years is on high side. Hence there should be reasonable plans to provide good schools, pipe born water and well equipped hospitals. Creation of jobs for young school leavers should be given serious attention. Again, the few that fall into old age group (above 70 years) should be properly taken care of and likewise those that belong to very old age group (above 100 years) should enjoy special incentives from the government.

REFERENCES

[AA14] **D. Aditi, N. Anjali** – *Use of prediction algorithm in smart homes.* International Journal of Machine Learning and Computing, 4(2), 2014.

[ANR09] **P. I. Adebayo, O. Nneama, E. A. Rotimi -** *Spatial predictive model for Health Informatics in Developing Countries*, vol 3, 2009. www.jhidc.org.

[A+17] **R. O. Andre, R. Valter, L. Cirano, I. S. Maria, V. Alvaro, M. B. Sandhi, B. D. Bruce** – *Comparison of machine learning algorithms to build a predictive model for detecting undiagnosed diabetes.* Sao Paulo Med. Journal, 135(3), 234-246, 2017.

[Ban09] **J. A. Bangbose -** *Falsification of population census data in an heterogeneous Nigeria state in the fourth republic,* African Journal of Political Science and International Relations, vol 3: 331-319, 2009. www.academicjournals.org/AJPSIR

[Bon13] **G. Bontempi -** *Machine learning strategies for time series prediction,* Machine Learning Group, Computer Science Department. Boulevard de Triomphe – CP212, 2013. http://www.ulb.ac.be/di

[F+10] **S. Folorunso, A. T. Akinwale, O. E. Asiribo, T. A. Adeyemo -** *Population prediction using Artificial Neural Network*, African Journal of Mathematics and Computer Science Research, vol 3: 155-162, ISSN 2006-9731,2010. www.academicjournals.org/AJMCSR

[Hal16] **K. Hal -** *A simple framework for building predictive model: A little data science business guide squared consultancy*, A Salam Edward Company, 2016.

[Joh17]   **N. I. John -** *The population question in Nigeria*, Asian Research Journal of Mathematics, vol 5: 1 – 10, ISSN: 2456-477X, 2017.

[JBA14]   **D. K. John, M. N. Brian, D. Aoife** – *Fundamentals of machine learning for predictive data analytics.* The MIT Press, Cambridge, England, 2014.

[Ode00]   **S. Odewumi -** *Problem of census in Nigeria,* in T. Odumosu, W. Alere, F. Adewumi (Eds) Social Problem and Planning Studies in Nigeria. Lagos Centre for Planning Studies, Lagos State University, 2000.

[OA13]   **T. O. Olatayo, N. O. Adeboye -** *Predicting population growth through birth and death rate in Nigeria*, Mathematical Theory and Modelling, vol 3, ISSN: 2224-5804 (print), 2225-0522 (online). 2013.

[V+02]   **R. Vilalta, C. V. Abte, J. L. Hellerstein, S. Ma, S. M. Weiss -** *Predictive algorithm in the management of computer system,* IBM System Journal, vol 41, 2002.