Full Length Research Paper

Population prediction using artificial neural network

O. Folorunso[1]*, A. T. Akinwale[1], O. E. Asiribo[2] and T. A. Adeyemo[1]

[1]Department of Computer Science, University of Agriculture, Abeokuta, Ogun State, Nigeria.
[2]Department of Statistics, University of Agriculture, Abeokuta, Ogun State, Nigeria.

This study employed an artificial neural network for population prediction (ANNPP) that handles incomplete and inconsistent nature of data usually experienced in the use of mathematical and demographic models while carrying out population prediction. ANNPP uses the three demographic variables of fertility, mortality and migration which are the major dynamics of population change as the input data. The datasets were divided 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 method was also used to simulate both the validation and the test datasets as case data on the consistency of results obtained from the training session via the train data. From the sixteen different topologies tested on the basis of the mean square errors (MSE), standard deviation (STDEV) and epochs; topology 19-9-1 performed best than the rest. A comparison between the predictions based on the ANNPP derived growth rates and The cohort component method of population prediction (CCMPP) was compared. The results showed that ANNPP percentage accuracies ranged between 81.02 and 99.15% while that of CCMPP percentage accuracies ranged between 64.55 and 86.43%. These results showed that artificial neural network model performed better than the demographic model.

Key words: Supervising training, component method, population, back propagation algorithm, mean square errors, standard deviation.

INTRODUCTION

The relationship between the population dynamics and development has occupied an important position since the inception of demographic studies. Malthus drew attention to the imbalance between population and means of subsistence (Hanson, 1966). He strongly believed that the imbalance in the rate of population increase and the means of subsistence was the major cause of poverty in the society. However, there is a high relationship between the population dynamics and development. Virtually, in all facets of a nation's plans, there is need for population data to allow that nation or country to plan for the education need, health, housing and investments. Udo (2003) observed that a modern population census such as census 1991 of Nigeria is the main source of basic data about any country. According to him, it provides information on the spatial distribution of

the population between urban and rural and different political units (States, Local Government Areas, and Wards) in the country. For example, the information on the age-sex composition, occupation, education and place of birth from where information on growth rate, birth and death rates and internal migration can be obtained from population data. Government and business houses require population data for effective planning of their social programme and investments (Onibokun, 1997).

An Artificial Neural Network (ANN) has the potential to be inherently fault-tolerant or capable of robust com- putation. Its performances do not degrade significantly under adverse operating condition such as disconnection of neurons and noisy or missing data (Bernader, 2006). Since the country Nigeria is statistically underdeveloped compared with some advanced countries of the world, using Artificial Neuron Networks may tend to contribute very significantly to the improvement of accuracy of population data in the country. The issue of population data is a non-linear problem as such ANN will tend to outperform most of these other methods in use because

---

*Corresponding author. E-mail: folorunsolusegun@yahoo.com.
Tel: +234-803-564-0707.

of its ability to approximate both linear and non-linear data alike. Obviously, there are many methods of population prediction but the central objective of this work is to predict population into a distant future using artificial neural network. In order to do this, we shall conduct a general review of some methods of population prediction in use, design and train an artificial neural network for population prediction (ANNPP). This paper will estimate and predict population growth rates using the target values, train, validation and test results obtained from the training and compare the level of accuracy of ANNPP with that of The cohort component method of population prediction (CCMPP). The rest of the paper includes review of related works, methodology, data collection; system implementation and comparison of results. The paper concludes with future prediction and conclusions.

REVIEW OF RELATED WORKS

Different methods are used in population prediction. These methods are distinguishable from each other by what they hold to be constant into the future. Mathematical projection methods are usually based on extrapolation of past trends into the future. In the use of mathematical method for instance, there is an assumption of the presence of a high correlation between the population changes in successive periods.

The method hardly allows for anticipated deviations from the past trend. Furthermore, mathematical method does not explicitly include components of population changes such as fertility, mortality and migration. Since the making of assumptions is a precarious matter, from the stand point of accuracy, mathematical methods are to be avoided when possible (Boon and Kok, 1995). For example, in the component methods, the predictions/ projections are usually carried out under the assumption of a closed population where the effect of migration to population change is almost zero and we know that this is not correct (NPC and IRD, 2000). Bishop (1995) said that the multilayer feed forwards are one of the most important and most popular classes of ANNs in real world applications. According to him, a multilayer perceptron has three distinctive characteristics:

(1) The model of each neuron in the network includes usually a non-linear activation function, sigmoid or hyperbolic.
(2) The network contains one or more layers of hidden neurons that are not part of the input or output of the network to learn complex and highly nonlinear tasks by extracting progressively more meaningful features from the input patterns.
(3) The network exhibits a high degree of connectivity from one layer to the next one.

Figure 1 illustrates a schematic diagram of an I – J – K (two layer perceptron) according to (NACF, 1998).

Where
I – 5 (five input parameters)
J – 3 (three hidden parameters)
K – 3 (three output classes)

Note that this is a two layer perceptron. For a K – class problem, we need k output units instead of just the single unit required for a 2 – class problem. The error function according to NCAF becomes

$$E = \frac{1}{2} \sum_{p=1}^{p} \sum_{k=1}^{k} (y^{p}{}_{k} - t^{p}{}_{k})^{2} \; :$$

$$= \frac{1}{2} \sum_{p} \sum_{k} (g \sum_{j} w_{jk} y_{j}^{p} - t_{k}^{p})^{2}$$

$$y_{k} = g(a_{k}) = g(\sum w_{jk} y_{i})$$

Since $y_{i} = g(a_{j}) = g(\sum w_{ij} x_{i})$

The equation above can be explained thus:

$$E = \frac{1}{2} \sum_{p} \sum_{k} (g \sum_{j} w_{jk} g(\sum_{i} w_{ij} x^{p}{}_{i}) - t^{p}{}_{k}))^{2}$$

$$a = \sum_{w_{ij} x_i)} \qquad \Delta w_{jk} = -\eta \frac{\partial E}{\partial w_{jk}} = -\eta \partial_{j} y_{i}$$

Where

$$\partial_{k} = \frac{\partial E}{\partial a_{k}} = (y_{k} - t_{k}) y_{k} (1 - y_{k})$$

For the input – to – hidden layer weights:

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} = -\eta \partial_{j} y_{i}$$

Where

$$\partial_{j} = \frac{\partial E}{\partial a_{j}} = \sum_{k} \partial_{k} w_{jk} y_{i} (1 - y_{j})$$

The robustness of Artificial Neural Network and its fault tolerance even in the face of missing data has resulted into its use in various fields of human endeavors. Tay and Cao (2002) observed that Neuro computing has rapidly gained in importance in a multitude of different application areas including sensor processing, pattern recognition, data analysis and control. Halmari et al. (1992) also said that producing accurate forecasts is a

$$y_k = g(a_k) = g(\sum w_{jk} y_i)$$

$$\text{since } y_i = g(a_j) = g(\sum w_{ij} x_i)$$

The equation above can be explained thus:

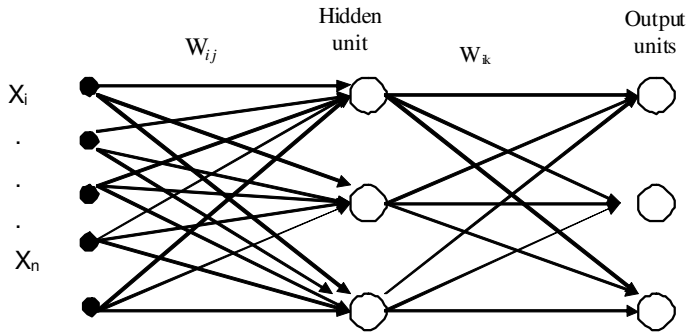$$E = \frac{1}{2} \sum_p \sum_k (g \sum_j w_{jk} g(\sum_i w_{ij} x^p_i) - t^p_k))^2$$

Figure 1. 5-3-3 multilayer perception.

is a critical activity for any marketing group or organization, inaccurate or misleading forecasts can result in missed targets, improperly allocated resources, and many other problems. Microsoft Excel according to them provides a built-in tool for predictions, but the accuracy of its results is significantly reduced when non-linear relationships or missing data are present which often, is the case when analyzing marketing data.

METHODOLOGY

Component method

Component method involves a summation of population growth components such as births, deaths, migration, employment opportunities, available housing and other reflecting potential changes in the population. The cohort component population projection method (CCPPM, 2009) can be represented by the following formula:

Pf: = Pb + (F$_1$ + F$_2$ + F$_3$ +… Fn) where
Pf = future population
Pb = base population

F$_1$, F$_2$, F$_3$,…,F$_n$ = population changes expected during the selected time period due to specific factors such as births, deaths, etc. While this method may require significantly more effort than methods such as linear growth, exponential growth, decreasing growth, correlation growth, etc, due to the amount of data needed, it may result in a more accurate projection if the data is accurate, hence the use of ANNPP.

Back propagation algorithm

The process of determining the magnitude of the weight factors that result in accurate output is called training. Several methods are

available to accomplish this but the back propagation algorithm method is the most commonly used. The method according to (Caudill and Butler, 1992) is based on the determining the error between the predicted output variables and the known values of the training data set. The error parameter is commonly defined as the root mean square of the errors for all the data points used in the training. The weight factors are adjusted by determining the effect of changing each weight on the error in the predicted output. This process takes the form of determining the partial derivatives of the errors with respect to each of the weights. The algorithm used to propagate the error correction back into the network according to (Caudill and Butler, 1992) is generally of the form:

$$W_{ij} \text{ new } - W_{ij} \text{ old } = -\eta \frac{\partial E}{\partial w_{jk}}$$

Where E is the error parameter, η is the proportional factor called the learning rate. The process of adjusting weights is continued until the error is less than some desired limit after which the network is considered trained. Once the network is trained, it can receive new input data that were not used for training and apply the weight factors obtained during training (Doraisamy et al., 2000).

Gradient descent algorithm

Input vectors are applied to the network and calculated gradients at each training sample are added to determine the change in weights and biases. Traingd function is used for the training of the network. The network is created using newff. The newff creates feedforward back-propagation network.

Syntax

net = newff (PR,[S$_1$, S$_2$,…,S$_n$] {TF$_1$ TF$_2$, TF$_3$,…,TF$_n$}, BTF, BLF, BPF)
newff takes several arguments.
PR = R x 2 matrix of min and max values for R input elements
S$_1$ = size of ith layer, for N1 layer
TF$_1$ = Transfer function of first layer (tansig)
TF$_2$ = Transfer function of second layer (log sig)
TF$_3$ = Transfer function of third layer (purelin)
BTF = Back propagation network training function (traingd)
BLF = (Backpropagation weight/bias learning function (learngdm) and returns N – layer feedforward back-propagation network.
BPF = Back propagation network training forward

The connection of the input to hidden layer, then the first hidden layer to output layer is automatically achieved when newff function is called. And as each layer has its own transfer function, the newff provides a means of specifying the transfer function of the layers in its syntax. The five parameters associated with traingd are epochs, show, goal, time, min-grad and max-fail. The ANNPP uses the tansig and logsig for the first and second hidden layers. The aim of using these is to get outputs in those two hidden layers of values between -1 and 1. Since the target values are greater than the range of values between -1 and 1, purelin is used at the output layer. This enables the network to output values of any magnitude. The algorithm for ANNPP is described in Figure 2.

Data collection

The datasets used for this paper were obtained from the

Begin
1. initialize the weights that connect inputs to hidden layer 1
2. Multiply the input vectors with their connecting weights
3. Compute the total weighted input
4. Threshold the total weighted input by tansig to get output for the first hidden layer
5. Use the output for the first hidden layer as the input for second hidden layer
6. Initialize the weights that connect hidden layer 1 to hidden layer 2
7. Repeat step 2 and 3
8. Threshold the total weighted input by logsig to get output and input for second hidden layer and output layer respectively
9. Initialize the weights that connect hidden layer 2 to output layer
10. Repeat step 2 and 3 to get the output values
11. Threshold the total weighted output by purelin to get the actual output values
12. If the output values are equivalent to the target values Then Go To Stop Else
13. Compute EA // EA is the difference between the actual values and the target values.
14. convert EA to EI// EI is the rate at which error changes as the total input received by a unit is changed.
15. Compute EW // EW is the error derivatives of the weights. That is how the error changes as each weight is increased
16. Multiply those EAs of those output units and add the products
17. compute EAs for other layers by repeating step 12 to 15 // movingrom layer to layer in a direction opposite the way activities to propagate.
18. Repeat 2 to 13
19. Stop
20. End

Figure 2. ANNPP algorithm.

Table 1. Distribution of sample data.

Types of datasets	Size of datasets
Train data	720
Validation data	720
Test data	720
Total	2160

1990/1995, 1995/2000, 2000/2005…2045/2050, 2050/20-55 and 2055/2060 projected rates produced by the National Population Commission (NPC, 1994). The datasets comprised of fertility rates, survival ratios (males and females) and migration (males and females). The rates were produced by the NPC in triplicates- low, medium and high variants. Each of the period's dataset was produced in age groups ranging form age (10 -14), (15-19)… (60 - 64) and (65+). Mortality rates were derived from survival ratios that were given. The total data points collected were 2160 in numbers. These data points were divided into 3 thus: train data (720), validation data (720), test data (720). The information is presented in Table 1.

The train data that serves as input data, 720 in size, was presented to artificial neural network for population projection (ANNPP) in batch mode with the feed forward architecture and backward propagation algorithm. The

training was done on a supervised manner. Each of the 720 data points were compressed into 144 data points by adding together the fertility, mortality (males and females) and Migration (males and females) of each of the age groups. This preprocessing was done using the Microsoft Excel. The first sets of data known as train data were presented along with other two datasets known as validation and test data. These last two datasets were used to simulate the network so as to see the generalization power of the network when introduced to the data it has not been previously exposed to. In order to discover the most appropriate topology for the network, 16 sets of topologies for each of the 3 training functions (48 topologies altogether) were used in the network one after the other.

SYSTEM IMPLEMENTATION

The algorithm in Figure 2 of ANNPP and Cohort component method formula were developed on Microsoft Windows XP using Microsoft Visual Basic 6.0 for the front and back ends implementation of population projection / prediction. MATLAB 6.5 tool box was incorporated for the training of the neural network.

ANALYSIS OF THE TRAINING

The datasets were presented to the network using three training functions thus Resilient Back Propagation (trainrp); Batch Gradient Descent (traingd) and Batch Gradient with Momentum (traingdm). The train data were used for the training while the validation and test data were used to simulate the network as tests on the consistency of the results obtained from the training session. Saenz and Pingitore (1989) said that the entire process of designing a network, train it, optimizing its performance entail considerable trials and errors. No foolproof rules exist for selecting the number of middle layers or the number of neurons in them. Klusman (1993) observed the same when he said that the best transfer function and the number of connections are also matter of choice as its learning rate parameter. According to Jacobs (2003), it is not possible to say which amount of neurons in a hidden layer is the best for the problem at hand, trials and errors will give this solution. Therefore, in order to find the most appropriate training function and the topology; different topological series namely, 20, 19, 18, 17, 16 to 10 (each of them contains 4 topologies) were simulated in turn one after the other for each of the whole topologies.

The following factors determine the rating of performance of a training function in order of their importance:

1. Ability to meet the performance goal;

2. Minimum sum of square errors;
3. Ability to approximate the target values as closely as possible as it is revealed by their maximum correlation coefficients;
4. Minimum number of epochs which determine the time taken for the training.

The first and third factors are the principal determinant of the performances while the rest are just the supportive ones. Consequently, the results obtained from the training sessions were correlated with the target values. For example, at topology 15 -10 -1, all the training methods performed differently. The train data has correlation coefficients of 0.98785, 0.9913 and 0.7922; validation data has 0.98811, 0.98415 and 0.96088 while test data has 0.99919, 0.9991 and 0.95516 for traingdm, traingd and trainrp respectively. When their various mean square errors (MSEs) were compared; traingdm has 0.0009999-22, traingd has 0.000999576 and trainrp has 0.0009775-61. However, there were significant differences in the number of epochs or passes and hence on the time used by various training functions during the training and simulation of the network. Traingdm has 15969 epochs, traingd 10393 and trainrp has 119. In summary, the traingd performs best in the area of correlation coefficients but have high number of epochs (relative to trainrp); followed closely to it is traingdm with high correlation coefficients but highest number of passes when compared with traingd. The trainrp has the minimum training time and number of passes but lowest in terms of correlation coefficients. This process was performed to all the topologies from 10 to 20 by changing their neurons in second hidden layer. The result shows that the traingd has the best performances followed by traingdm. In order to discover which of the topologies within the traingd function performs best; all the train data's results of all the topologies with the best highest values were shown in Table 2. The Table 2 reveals that topology 19-9-1(with mse of 0.089942) performs best followed by topology 17-7-1(with mse of 0.100876). The performances were measured on the basis of the minimum mean square errors and low standard deviations. Hence, topology 19-9-1 was used by the ANNPP.

CROSS COMPARISON OF ANNPP AND COHORT COMPONENT METHOD

Topology 19-9-1 was used by the ANNPP for population prediction while the formula described above was used by the component method. The data from national population commission served as input to the ANNPP and component method. Table 3 illustrates target population of both component and ANNPP results with their respective percentage accuracies. For component method, at the age group (10-14), the percentage accuracy is

Table 2. Topologies with highest values.

Topologies	Means square errors	Standard deviation
20-10-1	0.1000513	0.800694286
20-9-1	0.433279	0.799466723
20- 8-1	0.4333373	0.797018033
20-7-1	0.4333592	0.800272506
19-10-1	0.1000177	0.806359306
19-9-1	0.089942	0.640019678
19-8-1	0.0999622	0.802632117
19-7-1	0.1000161	0.799829934
17-10-1	0.1000513	0.79896227
17-9-1	0.0999927	0.80057469
17-8-1	0.0999967	0.79855447
17-7-1	0.100876	0.70080768
15-10-1	0.1000513	0.79510201
15-9-1	0.0999947	0.799466723
15- 8-1	0.099979	0.797018033
15-7-1	0.099998	0.800272551

84.95%. As we come to age (15 -19) the percentage accuracy goes up to 86.43% while the following age group (20 - 24) rose to 85.63%. At age groups (25 - 29) and (30 - 34) percentage accuracies are 80.96 and 77.71 respectively. From age group (35 - 39) to the last age group (65 - 69), there is a down ward trend in the percentage level of accuracy that ranges from 82.83% down to 64.55% respectively.

The level of accuracy in the component method of population prediction is higher at lower age groups than at the higher age groups. For ANNPP method, at the age group (10-14) and (15-19), we have 98.91 and 99.15% accuracies whereas at age group (20-24), we have 97.84%. At age groups (25-29) and (30-34) we have 87.05 and 96.56 percentage accuracies respectively. For the rest age groups 35-39, 40-44, 45-49, 50-54, 55-59, 60-64 and 65-69, we have 87.82, 84.82, 90.12, 91.76, 87.93, 81.02 and 85.46% respectively.

From the Table 3, it can be observed that the least percentage accuracy for ANPP is 81.02% whereas; the least for CMPP is 64.55%. The highest percentage accuracies for the two are 99.15 and 86.43% respectively. The ranges between the best accurate predictions of the two methods are 18.13 and 21.88% for ANNPP and CCMPP respectively. Apart from the fact that the ANNPP has the highest percentage accuracy, it also performs more evenly among all the age groups having a range of 18.13% compared with 21.88% of the CCMPP. This evidence shows that Artificial Neural Net-work performs better than components method.

PREDICTION OF POPULATION INTO DISTANCE FUTURE

The actual predictions of population were done using the visual basic 6.0 as the interface. The predictions made use of the target growth rates, train data rates, validation data rates and the test data rates. The aim of using all these rates is to produce different magnitude of population figures from where each of them could be compared together with the actual census count in the future when that is done. This is in line with the normal practice where population is predicted/projected under the assumption of low, medium and high variant (NPC, 1998).

The prediction is done taken the population of 2005 as the base year. The growth rates were varied from year to year at certain interval to reflect certain demographic thought of growing, stable then decline growth. The predictions that range from year 2006 to year 2400 were done on a yearly basis such that any year in between these two terminals can be selected. Figure 3 is an example of such a prediction of year 2015 taken year 2005 as the base population. The generated results for every age group distribution and total population are shown in Figure 3. It is possible to select the predicted year and see the results.

Conclusion

From the various tests performed on the results of the train, validation and test results, it is confirmed that Artificial Neural Network performs quit impressible in estimating the rates of population growth. Both the percentage accuracies and correlation coefficients are good evidences of the fact that given enough data at its disposal, the ANNPP can ensure population projection accuracy and eliminate errors embedded in planning for socio economic activities such as education, employment

Table 3. Comparison results between ANNPP and component method.

Age group	Target	Component results	ANNPP results	Percentage accuracy component	Percentage accuracy ANNP
10-14	14772010	14026740	14610970	84.95	98.91
15 -19	14026400	13525610	14019080	86.43	99.15
20- 24	13358770	14110820	13872060	85.63	97.84
25 -29	13113710	11928810	11415420	80.96	87.05
30 -34	10747090	10501330	10377200	77.71	96.56
35- 39	9716698	8484023	8533386	77.31	87.82
40 -44	8104330	6499021	6874384	70.19	84.82
45 -49	6386300	5248137	5760676	82.18	90.12
50- 54	5233478	4334665	4802386	82.83	91.76
55- 59	4334740	3332585	3811558	76.88	87.93
60-64	3332619	2263355	2700196	67.92	81.02
65-69	2263390	1461064	1934320	64.55	85.46
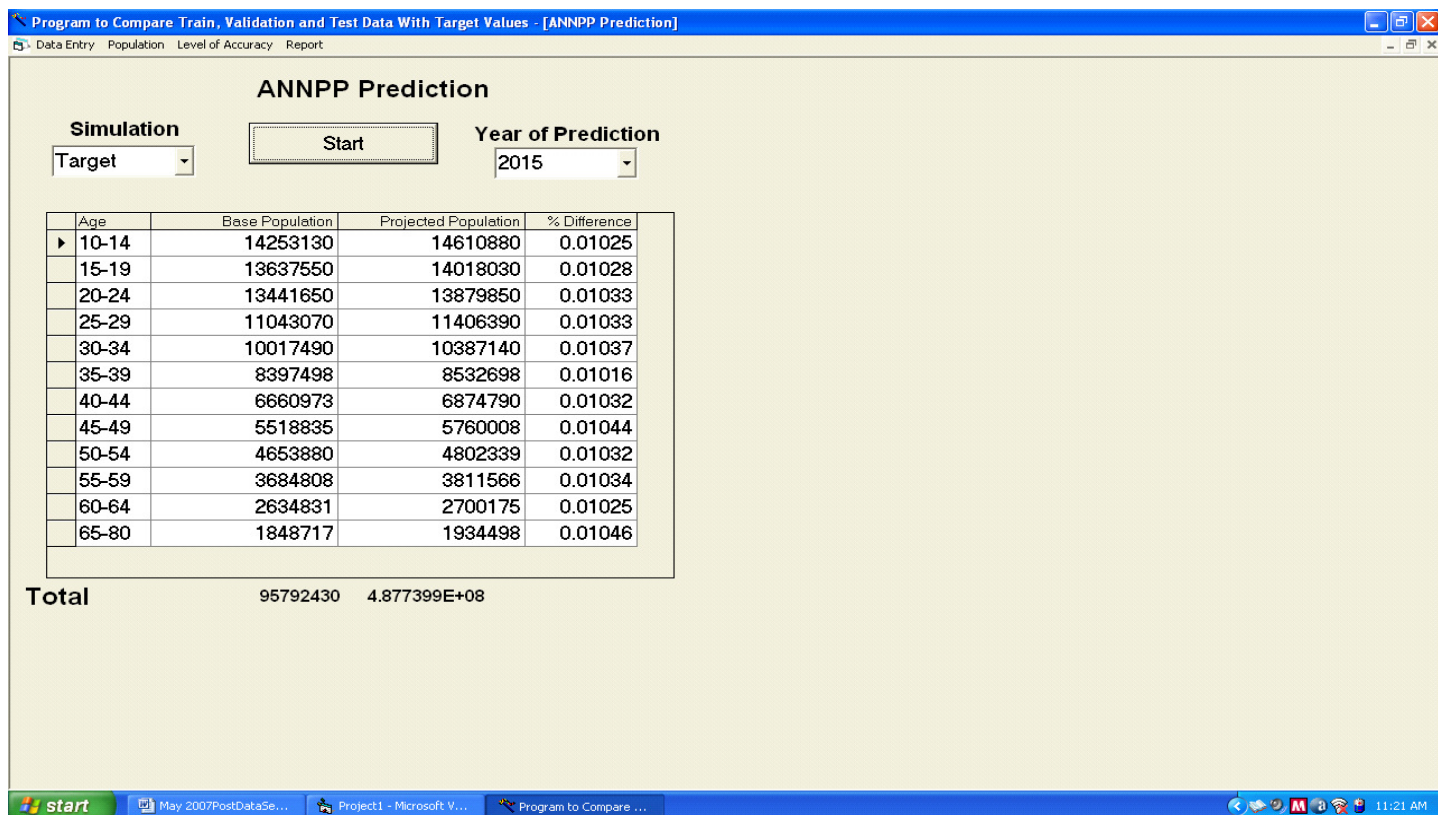
Figure 3. Prediction interface.

ment, health and other areas of a nation's life. It can also be observed that when the performance of Artificial Neural Network is compared with the Cohort Component Method of population projection, the formal performs better than the latter. Hopefully, as the method is adopted in population matter, the accrued benefit of accuracy will help both the government, public and private organizations to make reliable and enduring plans that will benefit both the planners and the beneficiaries. The accrued benefits will result into more profitability and

improvement in the standard of living of the entire people.

REFERENCES

Bernander O (2006). Neural Network." *Microsoft® Encarta®* 2006 [DVD]. Redmond, Microsoft ® Encarta ® © Microsoft Corporation. pp.1993-2005

Bishop CM (1995). Neural Networks for Pattern Recognition, Oxford University Press, Oxford.

Boon ME, Kok LP (1995), Classification of cells in cervical smears, In Applications rule of Neural Networks, Murray Af, ed. Klawer. pp. 113-131,

Cynthia J (2003). Average Speed Prediction Using Artificial Neural Networks: Published M.Sc., Project submitted to the Faculty of Information Technology and System, Delft University of Technology.

Caudill M, Butler C (1992). Understanding Neural Networks:, v1 and 2, Cambridge, Massachusetts MIT Press, p. 354.

CCMPP (2009). The Cohort Component Population Projection Method Retrieved 7th August 2009 from www: http://www.cpc.unc.edu/measure/training/mentor/population-research/pap/lesson-8.

Doraisamy H, Daniel H, Halleck PM (2000). The American Association of Geologists. AAPG Bull., 84(12): 1895-1904.

Halmari PH (1992). Using Neural Network to Forecast Automobiles Sales in Voge. W.G and Mickle M. H. the Twenty third annual Pittsburgh Conference on Modelling and Simulation, University of Pittsburgh, U.S.A.

Hanson JL (1966). A textbook of Economics, Macdonald and Evans Ltd, London

Klusman RW (1993). Soil gas and related methods for natural resource exploration, New York, John Wiley and Sons, p. 483.

NCAF (1998), Neural Computing Forum Co Published in North Central and South America by John Wiley and Sons Inc. 605 Third street, New York, NY 10158-0012.

National Population Commission (1994), Sentinel Survey of the National Programme Baseline Report. Abuja, Planning and Research Department, NPC.

National Population Commission (1998). 1991 Population Census of Federal Republics of Nigeria, Analytical Report at the National Level, United Nations, NY.

National Population Commission and IRD/Macro (2000). Nigeria Demographic and Health, Survey, 1999, IRD / MACRO International.

Onibokun AG (1997). Male Sexuality and Family Life in Nigeria-Synthesis of Findings from Empirical Research and National Workshop.

Saenz G, Pingitore NE Jr (1989). Surface Organic geochemical prospecting for hydrocarbons, Multivate analysis, J. Geochem. Exploitation, 34: 337-349.

Tay FEH, Cao LJ (2002). Modified support vector machines in financial time series forecasting, Neurocomputing, 48: 847-861

Udo RK (2003). On the use and application of Census data in public Administration, A Publication of National Population Commission, Nigeria.