A Predictive Model For Improving Cereals Crop Productivity Using Supervised Machine Learning Algorithm

[1]O.E. Oduntan and [2]M. Hammed

Department of Computer Science, The Federal Polytechnic, Ilaro, Ogun State

[1]odunayo.oduntan@federalpolyilaro.edu.ng, [2]mudasiru.hammed@federalpolyilaro.edu.ng

ABSTRACT

In a developing country, the gross domestic productivity of the nation is a function of the level of its production. Agriculture is regarded as the bedrock of the nation; hence farmers are faced with the challenges of determining the quantity of its product each year, leading insufficient produce. This study focuses on the developing of a predictive model that provides a cultivation plan for farmers to be able to produce cereals crops that are measurable with the population and the rate of consumption using a decision tree algorithm. The methodology employed involves the use of two variables: Independent and dependent variables. The dependent variables are the consumption rate and production rate while the independent variable is the population rate. The Decision Tree Algorithm makes the prediction of rice and beans produced based on the population. The outcome of the prediction aids the farmer to plan adequately in cultivation of cereals. The final rules extracted from this study, are useful for farmers and the government to make proactive decisions.

Keywords: Prediction Model, Agriculture Product, Cereal, Consumption rate, Production rate, Population rate, Decision Tree Algorithm

1.0 INTRODUCTION

Agricultural productivity is measured as the ratio of agricultural outputs to agricultural inputs. The output is usually measured as the market value of final output, which excludes intermediate products such as corn feed used in the meat industry. This output value may be compared to many different types of inputs such as labour and land (yield). These are called partial measures of productivity (Dharmasiri, 2007)

Agricultural productivity may also be measured by what is termed total factor productivity (TFP). This method of calculating agricultural productivity compares an index of agricultural inputs to an index of outputs. This measure of agricultural productivity was established to remedy the shortcomings of the partial measures of productivity; notably that it is often hard to identify the factors cause them to change. Changes in TFP are usually attributed to technological improvements (Lydia Zepeda, 2001).

Increases in agricultural productivity lead also to agricultural growth and can help to alleviate poverty in poor and developing countries, where agriculture often employs the greatest portion of the population. As farms become more productive, the wages earned by those who work in agriculture increase. At the same time, food prices decrease and food supplies become more stable. Labourers therefore have more money to spend on food as well as other products. This also leads to agricultural growth. People see that there is a greater opportunity to earn their living by farming and are attracted to agriculture either as owners of farms themselves or as labourers (OECD, 2006)

Machine learning means to give the knowledge to the machine. There are various types of machine learning techniques such as supervised and unsupervised learning. Supervised learning means there is one supervisor to supervise the thing that is the program is trained by training examples and then that can be used to find the accurate conclusion for new data.

Examples of supervised learning are: Artificial neural network, Bayesian network, decision tree, support vector machines, ID3, k-nearest neighbor, hidden markov model etc The unsupervised machine learning means a vast amount of data is given to the program and the program will find the patterns and the relations between them. So hidden patterns in the data can be discovered using unsupervised learning. Some examples of unsupervised learning algorithms are k-nearest neighbor, self organizing map, and partial based clustering, hierarchical clustering, k-mean clustering.

Machine learning is used in Computer Science and Statistics to improve the prediction power. It is mainly used by data scientists, data analysts and also for them who wants to use the raw data to predict or find trends in data. As there are vast amount of data in the agriculture and also increased day to day, so machine learning techniques can be used for agriculture and agricultural production to find the accurate prediction of crop production.

The rate at which farmers produce crops especially cereals is dependent on the consumption rate, this is so because, some of the produce are wasted as a result of poor preservative means. This study focuses on the developing of a predictive model that provides a cultivation plan for farmers to be able to produce cereals crops that are measurable with the population and the rate of consumption using a decision tree algorithm.

2.0 LITERATURE REVIEW

Prediction of crop can be performed by using various machines learning algorithms such as mathematical and statistical method etc. In applied machine learning (ML) techniques for Maize breeding as revealed those ML algorithms are promising and can be used in statistical techniques applied in maize, alike the more newly popularize linear mixed models. Among the current technology available for expedite the releasing for new genotypes there is an emerging subject of ML.

Several strategic uses of ML in maize breeding, quantitative trait loci mapping heterotic group assignment and the popular genome-wide selections are few of the main areas presently address by the literature. Corn is one of the most important cereals in the world and a primary source of calories for human being along with rice and wheat the evolution of genotypes adapted to aggravating climate, particularly drought situation which has to be grown in marginal law and changing climatic condition for crop production (Ornella *et. al*., 2012)

Decision Tree Model

The decision tree models include the concepts as nodes, branches, terminal values, strategy, payoff distribution, certain equivalent, and the rollback method. There are three kinds of nodes and two kinds of branches. The decision node which is represented as square is a point where a choice must be made. The decision branches are extending from a decision node. Each terminal node has an associated terminal value, sometimes called a payoff value, outcome value, or endpoint value. The result of a scenario or the sequences of decisions are measured by each terminal value. There are two step processes for the construction of a decision tree algorithm- first, growth of large decision tree then reduction of size and over fitting the data, in the second step, and tree is pruned. The pruned decision tree that is used for classification purposes is called the classification tree described (Vijaysinh, 2014)

A *decision tree* is a map of the possible outcomes of a series of related choices. They can be used either to drive informal discussion or to map out an algorithm that predicts the best choice mathematically. A *decision tree* typically starts with a single node, which branches into possible outcomes.

3.0 METHODOLOGY

3.1 Data

The data used in this study was collected from Ogun State Agricultural Development Project and Population data was retrieved from population commission website during the period from 1991 - 2016.

3.2 Supervised Machine Learning (Decision Tree) with Regression Analysis

In this study, the regression analysis is used to estimate rice cultivation(Y) based on data population (X) to improve cereals crop production. Linear regression was used to model the linear relationship between a dependent variable Y and one or more independent variables X as shown in equation 1 .

$$Y = a + b \times X \tag{1}$$

Where a and b in the equation forms regression line which are estimated from the values of Y and X to predict the value of Y from X.

Since the detail data were not available for population (X) and rice cultivation (Y), this study modeled the set of data (d_i) as it is shown in equation 2 and 3.

$$d_i = \{d_1, d_2, ..., d_n\} \tag{2}$$

$$\therefore d_i = \{\mathrm{x}_i, y_i\} \tag{3}$$

This studyclassified population corresponded to cultivation of rice for five years as it is in equation 4.

$$x_i = \begin{Bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{Bmatrix} \quad y_i = \begin{Bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{Bmatrix} \tag{4}$$

$\therefore$ Through continuous learning process the techniques predicted quantities at which rice should be produced to cater for population by mapping $f : \mathrm{X} \rightarrow Y$ as it shown in equation 5

$$y_i = f(\mathrm{x}_i) \; \forall i = 1, 2, ..., \mathrm{n} \tag{5}$$

Error (e) that may occur in prediction was calculated using equation 6

$$e = y - \hat{y} \tag{6}$$

$$\therefore error(\mathrm{x}_i, \mathrm{y}_i) = \begin{cases} 1 \\ \\ 0 \end{cases} \tag{7}$$

i.e. If $f(\mathrm{x}_i) \neq \mathrm{y}_i$ error is 1and if $f(\mathrm{x}_i) = y_i$ error is 0 as it was expressed in equation 7.

Where;

x is independent variable (predictor)

y is dependent variable (desired output)

$\hat{y}$ is target prediction

Figure 1; gives an illustration of the decision tree model in predicting the classification process.

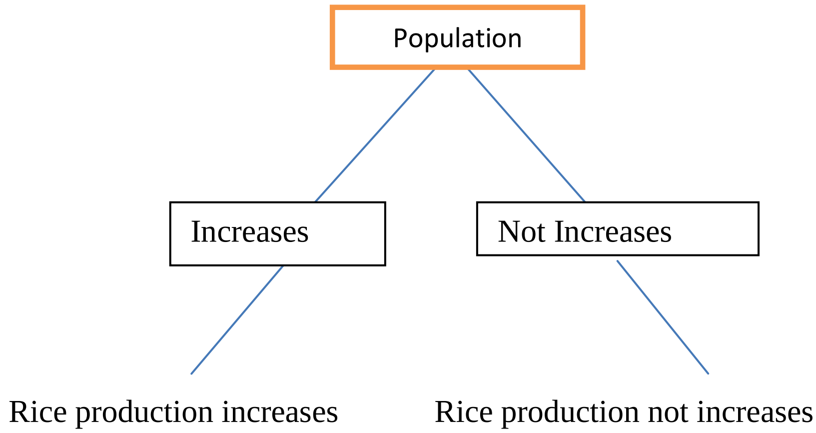

Figure 1: Decision Tree for Rice crop production

The decision tree predicting algorithm can be expressed as follows:

```
 If (population of ogun state increases) THEN

              (production of rice in ogun state increases)

     ELSE

       (no increases in rice production)

     ENDIF

     END
```

3.3 Data Analysis

In this study, 1991 was used as the initial population of occupants in Ogun State and initial rate for rice production. The data collected was analyzed using regression analysis to determine the growth rate and the production rate at the interval of ten(10) years. The prediction was implemented using Python version 3.50.

4.0 RESULTS AND DISCUSSION

The calculations in this study are efficient and are proportional to the instances that are observed. The decision tree algorithm used in this study predicted the production of rice cultivation in

Ogun State in proportional to the population of ogun state between 1991 – 2031 as it shown in the table 1 and the growth rate which used for prediction is shown in table 2

Table 1 shows the extracted data from the prediction implementation, 2,117tonnes of rice was produced in the year 1991, when the population was 2,333,726. In 2001, the decision tree using regression model predicted 2,974tonnes of rice production and a population of 3,278,668. Furthermore, in the year 2011, the decision tree using regression model predicted 4,068tonnes of rice production and a population of 4,484,420. Hence, in 2018, 5,000tonnes of rice was produced as predicted for a population 5,511,012. Therefore, in 2021, if 5,586tonnes of rice is produced by the farmers and the population is predicted to be 6,157,394.

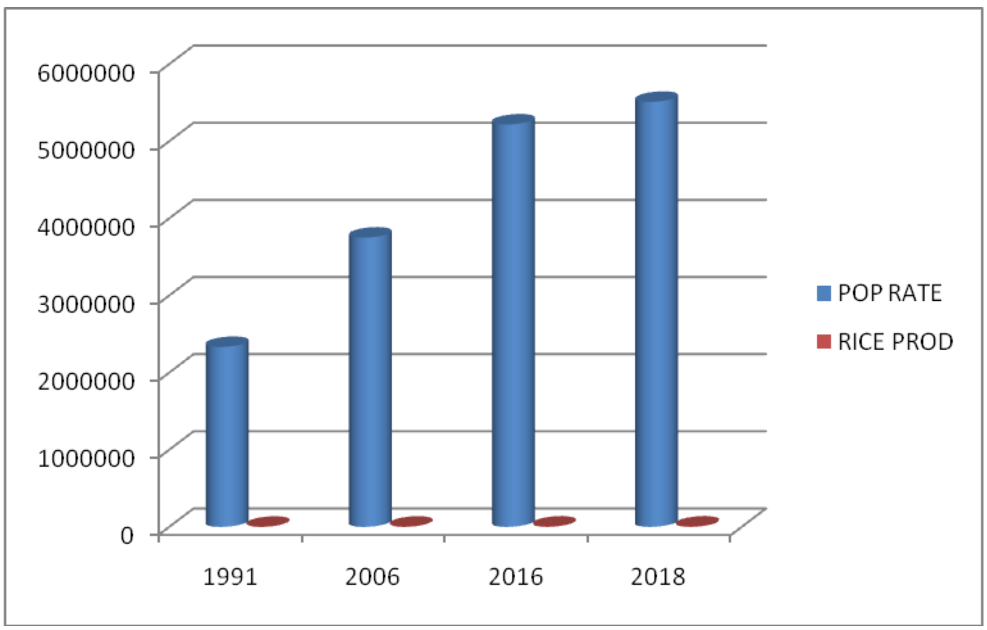

Figure 2: Comparison of Rice Production with Population rate at 10 years interval

Table 1: Ogun state population proportional to rice production prediction

year	population	production
1991	2333726	2117
1992	2428220	2203
1993	2522714	2288
1994	2617208	2374
1995	2711702	2460
1996	2806197	2545
1997	2900691	2631
1998	2995185	2717
1999	3089679	2803
2000	3184174	2888
2001	3278668	2974
2002	3373162	3060
2003	3467656	3146
2004	3562151	3231
2005	3656645	3317
2006	3751140	3403
2007	3897796	3536
2008	4044452	3669
2009	4191108	3802
2010	4337764	3935
2011	4484420	4068
2012	4631076	4201
2013	4777732	4334
2014	4924388	4467

2015	5071044	4600
2016	5217700	4733
2017	5364356	4866
2018	5511012	5000
2019	5726472	5195
2020	5941933	5390
2021	6157394	5586
2022	6372854	5781
2023	6588315	5977
2024	6803776	6172
2025	7019237	6368
2026	7234697	6563
2027	7450158	6759
2028	7665619	6954
2029	7881080	7150
2030	8096540	7345
2031	8312001	7541

Table 2: Growth Rate Versus Rice Production Increment

Period	Growth rate in Population (%)	Increase in Rice Production (tonnes)
1991 - 2000	40.49070	857
2001- 2010	36.77566	1094
2011 - 2020	32.50171	1322
2021 - 2030	31.49150	1759

As shown in Table 2, it was observed that between 1991 and 2000, the growth rate in population was 40.49%, while rice production was 857tonnes, as the growth rate in population decreases there exist a slight increase rice cultivation.

5.0 CONCLUSION

In this study, it has been observed that increase in the cultivation of rice and its production will aid the government agencies in making decision. This prediction enhanced the production of rice in Ogun state and also helps the farmers to produce in large quantity. This could stop illegal importation of rice from foreign countries.

REFERENCES

Dharmasiri L. M.(2007) Measuring Agricultural Productivity Using the Average Productivity Index (API)

Lydia Zepeda (2001) Agricultural Investment and Productivity in Developing Countries, FAO Economic and Social Development Paper No. 148, ed., FAO Corporate Document Repository, 12 July 2007, http://www.fao.org/docrep/003/X9447E/x9447e00.HTM.

Ornella L, Cervigni G, Tapia E., (2012) Applications of Machine Learning for Maize Breeding. In: Venkateswarlu B, Shanker AK, Shanker C. Book chapter of Crop stress and its management: Perspectives and Strategies, Springer, New York, USA. 1–29.)

OECD (2007) Promoting Pro-Poor Growth: Agriculture, DAC Guidelines and Reference Series, Paris: OECD, 2006,.org, 13 July 2007 *"Archived copy" (PDF). Archived from the original (PDF) on 2007-08-17. Retrieved* 2007-07-13. (10).

Vijaysinh A, Solanki S., (2014) Data mining techniques using WEKA classification for Sickle Cell Disease. (IJCSIT) International Journal of Computer Science and Information Technologies. 5(1):1–26.)

Appendix: A screen shot of the Prediction Implementation