Forecasting Rwandan Agricultural Crop Yield Value and Identify Factors That Influence This Crop Yield

Pierre Clement CYEMEZO[*], Dr. Emmy MUGISHA[**], Dr. Papias NIYIGENA[***]

[*]Student at MSc. IT, Department of Information Technology, University of Lay Adventists of Kigali
[**]Lecture, Department of Information Technology, University of Lay Adventists of Kigali
[***]Lecture, Department of Information Technology, University of Lay Adventists of Kigali

DOI: 10.29322/IJSRP.9.10.2019.p94106
http://dx.doi.org/10.29322/IJSRP.9.10.2019.p94106

Abstract- The forecasts are prediction, projection, or estimate of some future activity derived from observations that are random variables, and they are not especially informative unless accompanied by some indication of the magnitude of prediction error. Several researchers using different tools such as SPSS, WEKA and Python with various algorithms and models, worked on forecast of agricultural areas include crop yield. However, their findings still need improvement for more accurate performance.
The main objective of this study is to forecast crop yield value, analyze predictions and reveal factors that influence the variability of that crop yield value for Rwanda. To achieve this objective, a dataset on crop (Banana, Beans, Cassava, Maize, Rice, Coffee and Tea) yield extracted from Food and Agriculture Organization (FAO) online database on Rwanda between 1990 and 2017 is forecasted using WEKA 3.8.3 Time series forecasting Multilayer Perceptron algorithm with Mean Absolute Percentage Error (MAPE) accuracy metric. From predictions, we found that there is a good crop yield in Rwanda with a growth rate of 38.42%, and we estimate to achieve 121096.66hg/ha by 2030; The best growing yields are Rice with 81.95% and Cassava with 80.9% growth rate in 41 years period and for home consumables while Tea is the best growing exported crop with 34.79% growth rate. By averaging the yield of crops from 1990 to 2030, the highest yield value is for Potatoes with 90794.51hg/ha. Political factors such Genocide against Tutsi in 1994, introduction of technology and fertilizers in agriculture; disasters and climate changes are the main factors that influenced Rwandan crop yield from 1990 to 2030.

Index Terms- Agriculture, Crop yield production, Time series, Forecasting.

I. INTRODUCTION

Agriculture prediction is a very important problem in country. Any farmer is interested in knowing how much yield he is about to expect; the training data is to be collected from some time back to the past and the gathered data is used in terms of training which has to be exploited to learn how to classify future agriculture predictions (Vinciya, P. and Valarmathi, A., 2016).
The conventional and traditional system of data analysis in agriculture is purely dependent on statistics (Jaganathan, P., Vinothini, S., and Backialakshmi, P., 2014). The information and knowledge gained can be used for applications ranging from business management, production control, and market analysis, to engineering design and science exploration. Data mining tools predict future trends and behaviours, allowing businesses to make proactive, knowledge-driven decisions (Yethiraj, N. G., 2012).
Rwanda is a landlocked country in the Great Lakes region of East Africa. Its territory is dominated by highlands, giving it the name “Land of A Thousand Hills”. Agriculture is the main economic activity in Rwanda with 70% of the population engaged in the sector, and around 72% of the working population employed in agriculture (Rwanda at a glance, 2019).
Agriculture is the art and science of growing plants (crops) and the raising of animals for food, for economic gain or other human needs, Agriculture as a business is unique crop production is dependent on many climatic, geographical, biological political and economic factors that are mostly independent of one another (Raorane, A. A, and Kulkarni, R.V., 2013). It is estimated that over 66% of small scale farmers in Africa lack access to international market for their produce. Making it more difficult is that 41% of the vulnerable farmers come from Eastern Africa and these include Rwanda (IPAR Rwanda Report August, 2009).
The effort, in this research, is made on assessing and find the best algorithmfor forecasting Rwandan agricultural yield (1990 - 2017) into next thirteen years (2018 - 2030) using WEKA Time series forecasting built-in algorithms. By analyzing predictions, I will reveal the factors that impact the variability of agriculture productivity; This will help decision and policy makers especially in the Ministry of Agriculture and Rwanda Agriculture Board to identify which crop is mostly produced and which is mostly less produced and plan how to strength the agriculture sector for lightning the welfare of the people as country well.

II. LITERATURE REVIEW

Nowadays, the research community has given more attention to the topics that is related to Agriculture and its contribution in the growth of economies in countries. There are different approaches used to study the agricultural gross production forecasting. Here after we are going to look back on some research works related to the use of data mining techniques in agriculture especially prediction and forecasting approaches.
Dhekale, B. S. et al. 2014, Attempt to study trend, growth of tea in West Bengal buy forecasting production scenario considering

various factors of production like weather and fertilizer. Developed models are compared according to the minimum values of Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Square Error (MSE) and Mean Absolute Percentage Error (MAPE) and maximum value of Coefficient of determination (R^2) and of course the significance of the coefficients of the models. Once the model satisfies the requirement, are used for forecasting purpose.

Four types of models were developed considering with and without factors of production. The models are compared for maximum R square, minimum AIC, SBC, Log likelihood, RMSE, MAE etc. As results, Univariate ARIMA (2, 1, 1) model was best fitted with R^2 (0.98), AIC of 1043.53, SBC of 1052.99, RMSE of 9570.70, MAPE of 4.40 and MAE of 6851.87. ACF and PACF of residuals are not significant.

Even though the results have been demonstrated, there is a gap, regarding the accuracy of the used model where the best performing model is ARIMA with RMSE of 9570.70, MAPE of 4.40 and MAE of 6851.87; in this study We use a dataset including eight crops and analyze this using model built with MultilayerPerceptron algorithm based on accuracy metrics such as MAE, MAPE, RMSE and MSE where the minimum values of these metrics can be brought between 0% and 1% error rate for each examined crop.

Rajasekaran, T., et al 2016, stated that Previous few year yield value, cultivated area, irrigation methods, usage of fertilizer and pesticides, rainfall level at each season, soil maturity and weather condition of particular area details must be progressed in effective manner to forecast crop yield.

WEKA can be used as any modern computing platform where the ZeroR classifier is simplest classifier. More specifically, ZeroR with Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE) and Root Relative Absolute Error (RRAE) output will always found using target attribute from given data. The results indicated that ZeroR evaluation on training data had values such as 1.9835 of MAE, 2.2275 of RMSE, 100% at RAE and 100% at RRAE.

The above research paper results, indicates that ZeroR classifier with used accuracy metrics, it gave high values of errors for RRAE and RAE where the prediction is 100% erratic and on the other side MAE and RMSE values are understandable while are around 2% of errors. For getting more accurate and estimation of forecasting yield for crop production in this study, we use MultilayerPerceptron algorithm with performance metrics such as MAE, RMSE, MSE and MAPE with which expected to get as best performing and predicting model.

Shital, H. B., and Nirav, B., 2018, in their research, the main objective of this research was to forecast the wheat crop yield for the different districts of Gujarat state. For the analysis of data SPSS **"**Statistical Package for Social Science**"** and WEKA **"**Waikato Environment for Knowledge Analysis**"** tool**s** were used. Different classification algorithms within WEKA were applied on dataset and algorithms namely Gaussian Processes (GP), Multilayer Perceptron (MLP), Kstar, Sequential Minimal Optimization (SMO), M5Rules and Additive Regression (AR) were selected in this study.

The result of actual and predicted wheat yield from 2012-13 to 2014-15 was studied for yield forecast model and found that error percent had -58.3 to 3% as minimum and -18.4 to 48.9%. The value of R^2 ranges from 0.75 to 0.88 for GP, 0.92 to 1.00 for ML, 0.79 to 0.91 for SMOReg, 0.61 to 0.91 for M5Rule and 0.96 to 1.00 for AR algorithms.

From the above results, observation shows that the error percentage is higher and there is a way of improving this accuracy through use of MultilayerPerceptron algorithm associated with accuracy metrics such as MAE, RMSE, MSE and MAPE to forecast the dataset composed of home consumables and export crops, where the error percentage is proposed to be minimized around 0% to rise up the accuracy of the model to be used while forecasting yield values.

Saeed, K. and Lizhi, W., 2019; in their paper, they trained two deep neural networks, one for yield and the other for check yield, and then used the difference of their outputs as the prediction for yield difference. The performances of the four models, Deep Neural Networks (DNN), Least Absolute Shrinkage and Selection Operator (LASSO), Shallow Neural Network (SNN) and Regression Tree (RT) on both training and validation datasets with respect to the RMSE.

The results suggest that the DNN outperformed the other three models to varying extents. The weak performance of LASSO was mainly due to its linear structure; The DNN model was particularly effective in predicting yield and check yield, with RMSE metric for the validation dataset being approximately 11% of their respective average values.

From this research made by Saeed, K. and Lizhi, W., (2019), we indicated two main gaps, which are the followings; the erratic value of Neural networks models with RMSE as accuracy metric is high where they found that the most performing model is DNN with 11% of RMSE. By this I will improve this RMSE value with use of another Linear Regression algorithm with RMSE, MAPE, MAE and MSE accuracy metrics and the expected values will be around 0%. Another issue is to prove the strength of linear models in crop yield prediction by using MultilayerPerceptron algorithm for forecasting the crop yield for future.

In this study, my scope is limited on my nation Rwanda, and I propose to forecast agricultural crop yield in next thirteen years (2018 – 2030) using MultilayerPerceptron algorithm with MAE, MAPE, RMSE and MSE metrics to measure algorithm performance. After this, I will analyze predictions to identify the factors that influence the variability of Rwandan crop yield value.

In summary, even if diverse works have been conducted on use of data mining techniques in agriculture, there still many other subjects and topics to be worked on and there is many dataset that need to be analyzed with purpose of discovering patterns that may lead to useful knowledge for decision and policy makers.

III. METHOLOGY

According to the Longman English dictionary (2012), Methodology is the set of methods and principles that used when studying a particular subject or doing particular research or kind of work. This is in line with the nature of the study undertaken.

3. 1 Dataset

The dataset is commonly known as a collection of data which represents a particular variable for a single table and also data combination in the whole entity. This data set can be organized into several characteristics of information based on the structure and properties need to be carried out (Ahmad, F. K., et al. 2017). Here, the dataset is extracted from Food and Agriculture Organization database via their online platform and was downloaded in Microsoft Excel file; data collected is about Crop yield in values of hectogram per hectare, for Rwanda, from 1990 to 2017. The dataset is composed of twenty six instances / records for nine crops (Banana, Beans, Cassava, Maize, Potatoes, Rice, Tea and Coffee) are grouped into 2 categories; the home consumables and exports. This dataset is analyzed and forecasted for thirteen years ahead (2018 – 2030) to meet this research objective.

3. 2 Data preprocessing

When dataset is downloaded in Microsoft Excel file (.xls) format, it has been converted into Attribute-Relation File Format (.arff) file such as this format can be understood and process by WEKA tool. After the dataset is obtained, the next step is data preprocessing. By building model using WEKA 3.8.3 Time Series forecasting, the future agricultural yield can be predicted and analyze predictions to identify the factors that influence the variation of this crop yield value in next thirteen years.

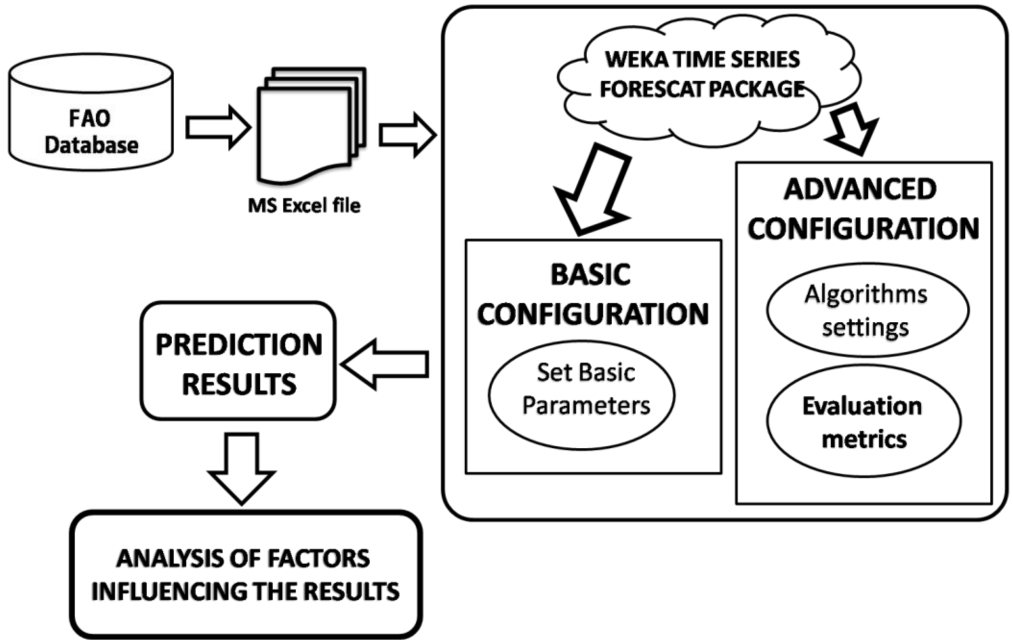

Figure 1- Study framework

The dataset is loaded into WEKA tool to be processed, after being loaded, the next step is the data forecasting for predicting the next eight years results of Rwandan agricultural gross production values and analyzing the factors that influencing the results after determining the best model to fit the dataset.

3. 3 Data analysis

As drawn in Figure 1 above, under WEKA Forecasting, there are two configuration options named "Basic configuration" and "Advanced configuration". In Basic configuration, I choose to use the following four parameters: Number of time units to forecast, Time stamp, Periodicity, and Perform evaluation.

In advanced configuration, there are other panels such as; Base learner, Lag creation, Periodic attributes, Overlay data, Evaluation and Output. These panels give the user the ability to configure the parameters for a specific selected learning algorithm, by customizing periodicity attributes, evaluation metrics and output settings.

- *Algorithms*

Under WEKA, Version 3.8.3, Time series forecast package there are 4 active algorithms such as; Multilayer Perception (MP), Sequential minimal optimization (SMOreg), Linear Regression (LR), and Gaussian Processes (GP). These all four algorithms have been tested on training set for determining which one is most accurate to be used for forecasting.

- *Performance measurement*

By default, in this version of WEKA, the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are computed but there other options that user can select to compute the performance of learning algorithm on the dataset. In my study, I choose to measure the accuracy of the algorithms with Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Mean Square Error (MSE) and Root Mean Square Error (RMSE) metrics.

Mean Absolute Percentage Error (MAPE) is a statistical metric of how accurate a forecast system is; this is measured as the average absolute percent error of actual values minus forecasted values divided by actual values. This is the most common measure used to forecast error. This is given by:

$$MAPE = \frac{\sum \frac{|A-F|}{A} \times 100}{N} \qquad (1)$$

Where;
A: Actual
N: Number of observations

F: Forecast
| |: Stands for absolute value

Mean Absolute Error (MAE) metric measurements are used to finalize the result by showing the closest quantity values based on the final result prediction. Simply, can say that the mean absolute error (MAE) is a quantity used to measure how close forecasts are to the eventual outcomes. This metric can be given by the following formula;

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - x| \quad (2)$$

Where;
n: The number of errors
∑: Summation symbol
$|x_i\text{-}x|$: The absolute errors
The standard statistical metrics of RMSE are used to measure the algorithms performance towards this study dataset, it compares a predicted value and known value. It is given by the formula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}} \quad (3)$$

With;
P: Predicted values
O: Actual values
n: Number of errors
Mean Squared Error (MSE) measures the average of squared error of predictions. That is, it calculates the square of different between the predicted valued and actual value and then averages those values.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y}_i)^2 \quad (4)$$

Where;
N: The number of data points
y_i: Represents observed values
$\bar{y}_i$: Represents predicted values
Above metrics are expressing forecasting error rate in percentage, lower values indicate that the forecasted values are good predictions and the higher they increase in percentage indicates that forecaster is doing worse predictions.

3. 4 Implementation

Our dataset is presented in .arff file format, is imported into WEKA Version 3.8.3, the time stamp is set as "year", periodicity as "detect automatically", number of time units to forecast as 8 and Perform evaluation check box is checked to evaluate the algorithm performance level. Thereafter, we test our target four algorithms to find one that best fit in the dataset using our selected three accuracy metrics.

IV. RESULTS

The findings indicate that MutilayerPerceptron algorithm fits in our dataset as presented in the table below;

Table 1 - Tested algorithms fitness on Dataset

CROPS	ALGORITHMS	METRICS			
		MAE (% Error rate)	MAPE (% Error rate)	RMSE (% Error rate)	MSE (% Error rate)
Banana	Linear Regression	8611.263	14.2905	12838.0073	164814431.5
	Multilayer Perception	1456.1263	1.9455	2102.5526	4420727.461
	Support Vector Machine	394.4967	**0.7737**	1647.7228	2714990.497
	Gaussian Processes	3565.193	5.7776	5150.7402	26530124.83
Beans	Linear Regression	496.8338	6.5767	628.756	395334.1571
	Multilayer Perception	163.4262	**1.9808**	234.7051	55086.4758
	Support Vector Machine	5.9013	4.0776	6.3861	40.7829
	Gaussian Processes	296.975	4.1367	405.7527	164635.2537
Cassava	Linear Regression	11564.1856	17.4907	14973.2004	224196730.9
	Multilayer Perception	27.0317	**0.0636**	35.1216	1233.5301
	Support Vector Machine	274.8891	0.4503	535.997	287292.7988
	Gaussian Processes	7578.6636	13.4319	9511.8192	90474704.93
Coffee	Linear Regression	649.6859	12.6668	845.5272	714916.2401
	Multilayer Perception	1.4586	**0.0231**	2.5121	6.3108
	Support Vector Machine	9.6899	0.1897	25.7941	665.334

	Gaussian Processes	423.6271	8.4684	584.6988	341872.737
Maize	Linear Regression	2158.7488	17.8553	2715.8264	7375712.912
	Multilayer Perception	1.1283	**0.0092**	1.6905	2.8579
	Support Vector Machine	22.2843	0.1986	24.7705	613.5797
	Gaussian Processes	970.1286	7.7369	1226.136	1503409.596
Potatoes	Linear Regression	9870.2418	10.5562	12365.9193	152915960.8
	Multilayer Perception	32.6444	**0.0431**	44.941	2019.6972
	Support Vector Machine	87.4621	0.105	99.2334	9847.2693
	Gaussian Processes	4871.0583	5.9252	6197.0897	38403920.27
Rice	Linear Regression	4704.9765	13.8769	6073.276	36884681.28
	Multilayer Perception	222.2063	**0.6537**	312.2077	97473.6515
	Support Vector Machine	54.368	0.7474	59.8399	3580.8143
	Gaussian Processes	2082.9131	6.0364	2608.6908	6805267.501
Tea	Linear Regression	728.367	5.3203	971.2876	943399.5816
	Multilayer Perception	0.6569	**0.005**	0.9596	0.9209
	Support Vector Machine	15.4786	0.1174	17.4841	305.6928
	Gaussian Processes	599.9179	4.748	782.0433	611591.7464

As shown in the Table 1, MP algorithm with MAPE accuracy metric, compared to other algorithms and accuracy metrics, has became the best performing one as it has smallest error rate percentage for all crops means for Banana with 0.7737%, Beans with 1.9808%, Cassava at 0.0636%, Maize at 0.0092%, for Potatoes at 0.0431%, Rice with 0.6537%, Coffee at 0.0231% and Tea with 0.005%.
The error rate percentage within Table 1, prove that MultilayerPerceptron algorithm is the best algorithm to fit our dataset with performance rate over 98.0192%, and the best metrics for accuracy is Mean Absolute Percentage Error (MAPE); this lead us to conclude that our first objective of assessing and find the best performing algorithm that fits our data is fulfilled.
Predictions refer to the results of a learning algorithm after it has been trained on a dataset, and this is applied to new data when you are trying to forecast the possibility of a particular outcome in a given time. One of the objectives of this study is to forecast the data to get future crop production value of Rwanda in next eight years; we are using Multilayer Perceptron algorithm with MAPE accuracy metric as this algorithm is proven to be the best performing on our dataset.

Table 2 - Predicted GPV for 2017 -2024

Year	Banana	Beans	Cassava	Maize	Potatoes	Rice	Coffee	Tea
2018	24115.32	12025.39	64647.46	7985.139	97721.76	36166.61	5053.928	16614.36
2019	77323.24	12885.8	89968.41	10278.08	84062.67	38889.89	4810.809	15265.25
2020	61176.5	10777.88	120936.7	21047.62	98654.6	60332.09	3421.406	18716.79
2021	52892.26	8846.305	81417.48	24689.94	112498.4	61310.23	3515.285	17635.96
2022	124098.4	7737.992	116854.7	22550.06	125951.1	72152.79	6698.244	11378.83
2023	117646.3	7431.009	115542.5	20499.2	112348.4	65285.26	6542.067	11736.96
2024	53222.83	7555.638	92354.02	13569.59	93399.19	50786.04	5226.67	16165.54
2025	70036.29	8696.031	76373.77	4607.909	96471.29	40461.66	7281.024	16496.09
2026	114756.9	9664.216	89838.53	4573.623	97731.68	39655.66	7735.346	16597.49
2027	47357.11	9884.96	101231.4	17361.74	104361.3	33417.13	4231.794	17568.9
2028	37795.82	10219.41	104721.4	19892.22	120812.6	47925	4797.873	18223.16

2029	123617.6	10072.21	118223.6	22205.57	122216.5	68982.81	6713.105	14382.19
2030	121096.7	8102.529	105359.7	22223.93	110211.8	75633.96	4704.13	16024.55

Table 2 presents predicted crop yields for thirteen years ahead. Based on our observations, we are able to identify some factors that influence its variation along the period of 1990 to 2030; the results are interpreted as follows;

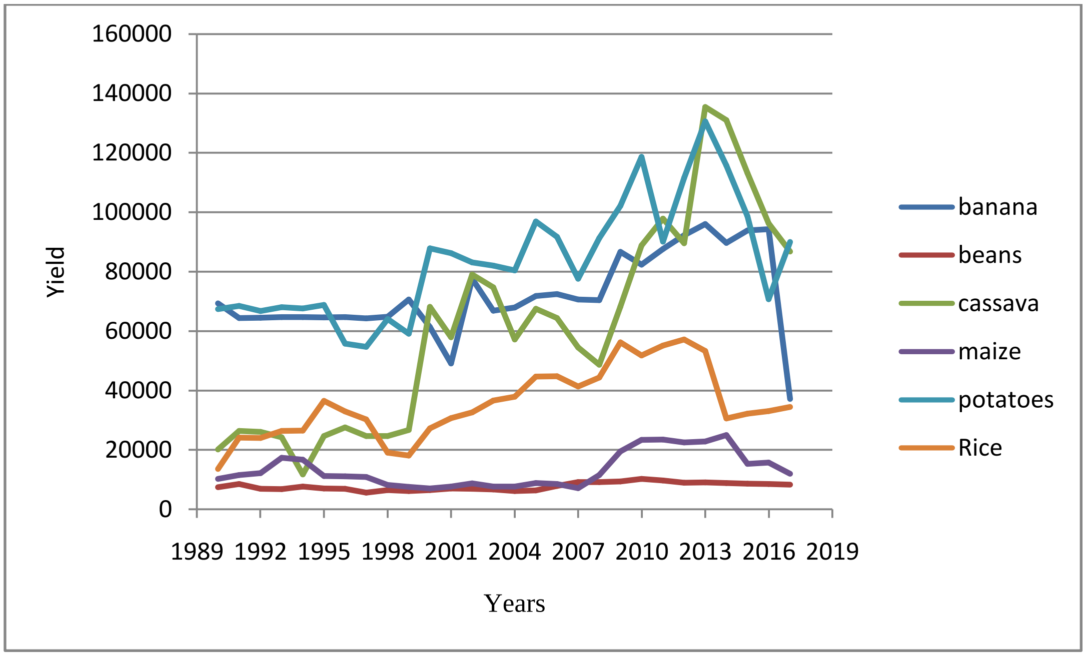

Figure 2 - Crop yield for home consumables 1990-2017

Figure 2, shows that the consumable crops selected for this research, are six: Banana, Cassava, Maize, Potatoes, Rice and Beans. With observation of the graph, we see that the yield of home consumables between 1990 and 2017 are decreasing between 1990 and 2000 due to political and security crisis, population were displaced, unstable, others were killed and in 1994 there was Genocide against Tutsi which took more than one million victims' life, whom many were farmers. This is explained by the lowest yield of Beans, Cassava, Potatoes and Rice where are respectively 5606hg/ha in 1997, 11778hg/ha in 1994, 54673hg/ha in 1997 and 13652hg/ha in 1990.
Compared to 1996, in next year, all six crops yield, were decreased because there were rebellions (Abacengezi) war in North and West of the country; Also, in the East and South of the country had climate changes that lead to drought. In 1999 and 2000 the country had lack of crop production and this caused famine in some regions like Bugesera, Nyanza, Kayonza, Ngomba, Gisagara and others. Due to security stability, political wills for mechanizing agriculture, from 2005 to 2014, there were a continuous increasing of yield for all crops as follows; Banana from 71781hg/ha to 89646hg/ha, Beans from 6378hg/ha to 8914hg/ha, Cassava from 67561hg/ha to 131057hg/ha, Maize from 8889hg/ha to 25009hg/ha and Potatoes from 96891hg/ha to 115678hg/ha respectively from 2005 to 2014.

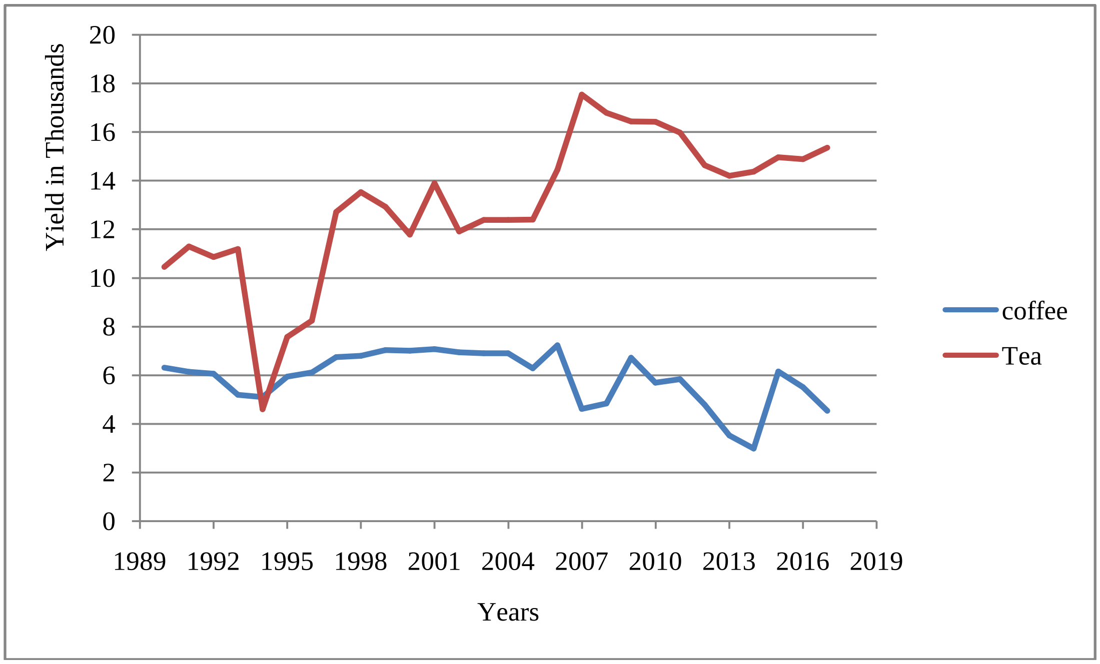

Figure 3. Crop yield for exports 1990 - 2017

The exported crops production analyzed in this research paper are two crops which are Tea and Coffee as presented in Figure 3. The yield for exports crops decrease by 1994 where Coffee yield was 5096hg/ha from 5191hg/ha in 1993 and Tea had 4596hg/ha in 1994 which is the lowest ever for the series, from 11181hg/ha in 1993 with causes to political instability, Genocide against Tutsi in 1994 and Liberation war. From the results, we see that from 1995 to 2004 the yield for export is increasing with some constraints, caused by climate changes remarkable in 2000 and 2016 affected Tea more than Coffee. Our findings show that Tea is more productive than Coffee for all years expect in 1994, this is due to the policies that prefer Tea versus Coffee.

Our analysis is going to take lead on Predictions analysis and see if the different factors analyzed above are influencing the future values of crop yield in thirteen years ahead.

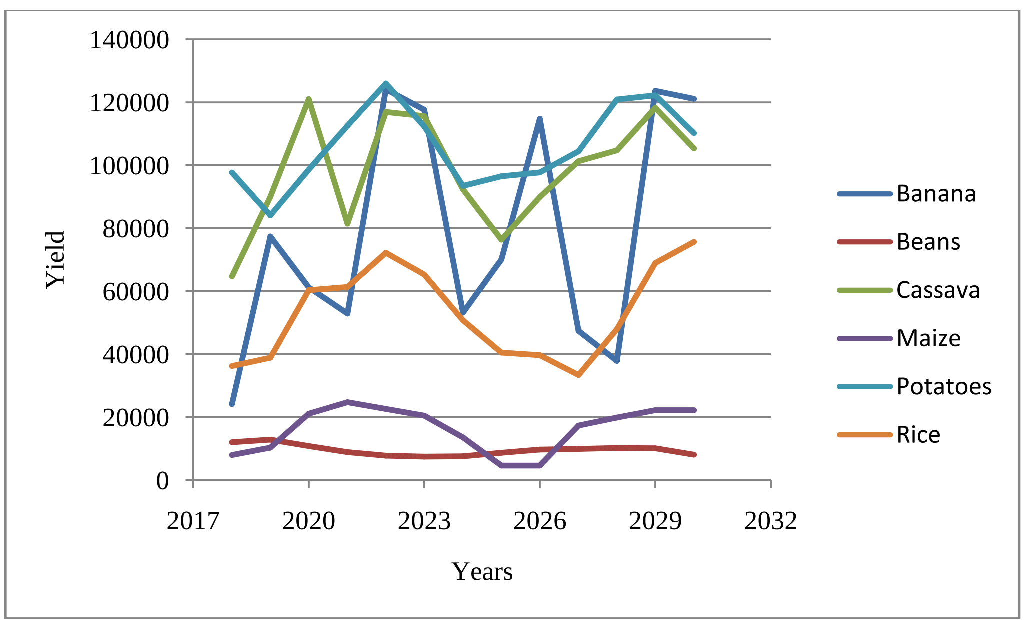

Figure 4. Predicted home consumable Crop yield

We predicted crop yield for thirteen years start by 2018 to 2030, from findings, it shows that comparing to the previous series (1990 - 2017), and the agriculture yield will increase and decrease in general. Banana is the most growing yield with an average rate around 80%, Beans will have the worst growing rate with -48.42%, 38.64%, 64.07%, 11.33% and 52.18% for

Cassava, Maize, Potatoes and Rice respectively. We found that in these coming years Beans will have a highest yield ever between 1990 and 2030 with 12885.8hg/ha in 2019, Banana 124098.36hg/ha in 2022 and Rice with 130600hg/ha in 2030; in this coming years, the predictions indicate that Banana and Maize will suffer for least productivity with 24115.31hg/ha in 2018 and 4573.08hg/ha in 2026 respectively. By 2030, the most productive home consumable crop we expect that will be Banana with 121096.66hg/ha and least productive will be 8102.53hg/ha. This new look is due to political stability, use of fertilizers, farmers are trained for amelioration of their work, investors will be welcomed and the sector in general will gain strength from different corners, and all these factors will resulting to the best change in yield to earned in future.

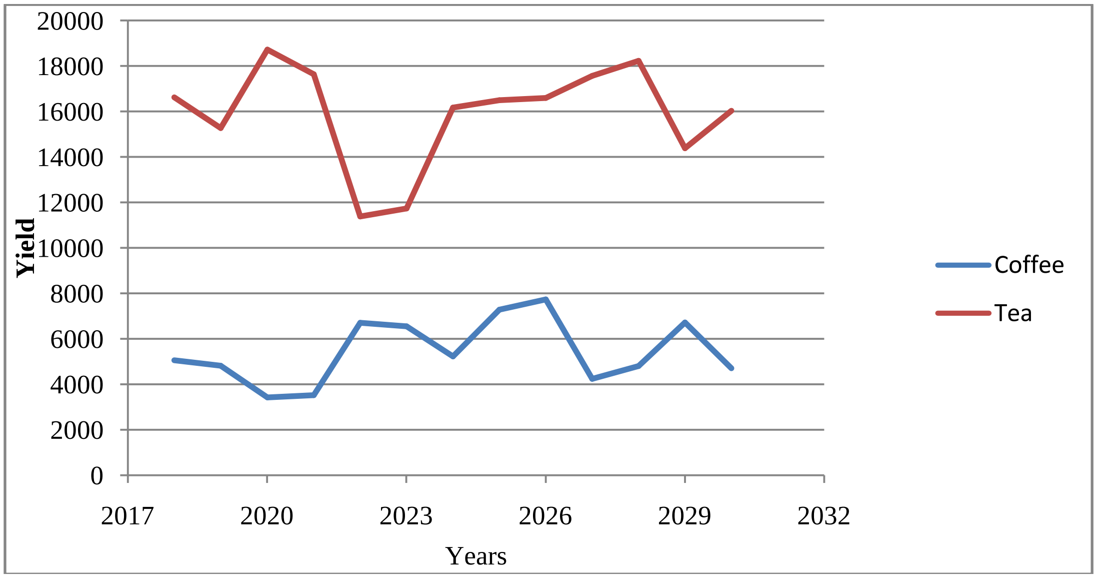

Figure 5. Predicted GPV for exported crops

Predicted GPV for exported crops as presented in Figure 5, we used two crops Tea and Coffee. From results we mark a high difference between yield of Tea and Coffee comparing to the previous years (1991 - 2017) because in this coming years for all years the difference of Tea for Coffee will be greater to 4680.5846hg/ha. Two crops will achieve their highest yield in this series of years (1990 – 2030) in this coming period with 18716.789hg/ha for Tea in 2020 and 7735.346hg/ha in 2026 for Coffee, even if the average growth rate for these crop yields will be falling up to -7.44% for Coffee and -3.68% for Tea.

In general, based crop yield growth rate, our predictions indicated that the gap between the past and predicted agricultural yield is significantly high for all crops except Potatoes and Rice with a difference of 13.74% and 8.25% respectively, then in coming years the agricultural yield is expected to be good comparing to the past years; and this due to the factors stated during predictions analysis.

V. CONCLUSION

In order to contribute to the amelioration of agriculture sector production, this study is conducted to find the best performing algorithm on present dataset on Crop yield on Rwanda (1990 - 2017) from FAO online database, forecast the crop yield in next coming thirteen years (2018 - 2030) then analyze predictions to identify factors that influence the variability of Crop yield values for Rwanda. Four algorithms, Linear regression, Gaussian Processes, Multilayer Perceptron and Support Vector Machine (SMOreg), all have been trained on dataset with four accuracy metrics, with Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Mean Square Error (MSE) and Root Mean Square Error (RMSE).

We forecasted our dataset with Multilayer Perceptron algorithm with Mean Absolute Percentage Error (MAPE) after finding this as the best to forecast our dataset; the predictions indicated that there is a good crop yield in Rwanda with a growth rate of 38.42%, and we estimate to achieve 121096.66hg/ha by 2030. The best growing yields are Rice with 81.95% and Cassava with 80.9% growth rate in whole period and for home consumables while Tea is the best growing exported crop with 34.79% growth rate.

By averaging the yield of crops from 1990 to 2030, the highest yield value is for Potatoes with 90794.51hg/ha this is explicative to the strength made for cooperatives of farmers of potatoes in Nyabihu, Rubavu and Musanze districts in Nothern province of Rwanda, and the less productivity is 5746.30hg/ha for Coffee as this crop is in way of disappearing in some corners of country by replacing it with trees and Bananas.

Our findings are in same way as our literatures as Dhekale, B. S. et al. (2014), in their research presented highest productivity of Tea in 2020, and Shital, H. B., and Nirav, B. (2018) in their research they proved that Multilayer Perceptron is the best performing algorithm for forecasting crop yield dataset.

With this coming period, policy makers must improve their politics on Coffee and Beans as these two crops' yields are falling; policy and decision makers are called to dispatch government strength in all fields as did for Rice, Potatoes and Maize. Investors too are welcomed to invest much to these crops production and famers are required to join the cooperatives and conduct visits to those who are partitioning the cultivation of Rice, Potatoes, Maize and Banana.

ACKNOWLEDGEMENT

I would like to express my deep sees of gratitude to the University Of Lay Adventists Of Kigali, UNILAK, administration, lecturers, especially Dr. Emmy MUGISHA and Dr. Papias NIYIGENA, and classmates for providing me with the knowledge to do this research project work.

REFERENCES

[1] IPAR Rwanda Report August, 2009; Rwandan agriculture sector situational analysis. An IPAR sector review

[2] Jaganathan, P., Vinothini, S., and Backialakshmi, P., 2014; A Study of Data Mining Techniques to Agriculture. IJRIT International Journal of Research in Information Technology, Volume 2, Issue 4, April 2014, Pg: 306- 313

[3] Yethiraj, N. G., 2012; Applying data mining techniques in the field of agriculture and allied sciences. International Journal of Business Intelligent. ISSN: 2278-2400

[4] Raorane, A. A, and Kulkarni, R.V., 2013; Review- Role of Data Mining in Agriculture. International Journal of Computer Science and Information Technologies, Vol. 4 (2) , 2013, 270 – 272.

[5] Vinciya, P. and Valarmathi, A., 2016; Agriculture Analysis for Next Generation High Tech Farming in Data Mining. International Journal of Advanced Research in Computer Science and Software Engineering. Volume 6, Issue 5, May 2016. ISSN: 2277 128X

[6] Ahmad, F. K., et al. 2017; Daily stream flow prediction on time series forecasting. Journal of Theoretical and Applied Information Technology. 28th February 2017. Vol.95. No 4 © 2005 – ongoing JATIT & LLS. ISSN: 1992-8645 and E-ISSN: 1817-3195.

AUTHORS

First Author – Pierre Clement CYEMEZO, Student at Masters of Science in Information Technology at University of Lay Adventists of Kigali (UNILAK), cyepi.academic@gmail.com
Second Author – Dr. Emmy MUGISHA, University of Lay Adventists of Kigali (UNILAK), emymugi@gmail.com, +250 785 330 211

Correspondence Author – Dr. Papias NIYIGENA, University of Lay Adventists of Kigali (UNILAK), papiasni@yahoo.fr, +250 788 548 000