Journal of Applied Artificial Intelligence
https://jaai.sabapub.com
ISSN: 2709-5908

2023 Volume 4, Issue 1 : 9 – 20
DOI : 10.48185/jaai.v4i1.728

Seasonal Crop Yield Prediction in Nigeria Using Machine Learning Technique

Abdulbasit Ahmed[1,*], Sunday Eric Adewumi[2], Victoria Yemi-peters[2]

[1]Department of Computer Science, Division of Agricultural Colleges, Ahmadu Bello University, Zaria, Nigeria
[2]Department of Computer Science, Faculty of Sciences, Federal University Lokoja, Kogi, Nigeria

Received: 09.03.2023 • Accepted: 20.06.2023 • Published: 30.06.2023 • Final Version: 30.06.2023

Abstract: The old methods adopted in the past by were very slow, undependable and sizable quantity of crops are damaged in fields because bacterial attacks and lack of adequate information. automating agriculture processes may likely be the solution to feed the nation in the future. Though there is still debate on its application to agriculture. The importance of food security in any society cannot be over emphasized, therefore balancing the inputs and outputs on a farm is fundamental to its success and profitability. With the increase in population index food production need to meet population growth, creating a wide gap between demand and supply of food. Data Mining is emerging research field in crop yield analysis. In the past farmers make use past yield to predict what they may likely have when farming in the current season The yield prediction is a major issue that remains to be solved based on available data. Data mining are the better choice for this purpose. Three (3) Different Data Mining techniques will be used for predicting crop yields during rainy and dry season. This research proposes and implements a system to predict crop yield from previous data. This can be done by using association rule mining on agriculture data. This research focuses on creation of a prediction model which may be used to future prediction of crop yield. It also shows that South East has the best in terms of accuracy for rainy season farming with model performance evaluation 138.9 using Decision Tree Classifier.

Keywords: Machine Learning, Support Vector Machine, Random Forest, Decision Tree Classifier and Algorithm

1. Introduction

Timely advice for predicting yields and analysis are made to help farmers maximize farm outputs. Prediction yield is a problem in agriculture. In the past, farmers predict yields from previous year yields. Thus, there are various algorithm or techniques and with the help of these techniques crop yield can be predicted. Effective use of the algorithms. By analysing climatic factors like weather and other factors, there is no techniques to overcome the situation faced [1].

For good crop yield, farmers require guidance timely to help predict crop productivity and analysis is to be done to help farmers utilize the full capacity in crop production. Yield prediction is a great problem. Farmer's previous experience in a particular crop can be used to make predictions for crops. Data Mining are methods of loading, predicting and extracting meaningful information from large data to extract some patterns and also change it into meaningful format for further use. Applying data mining approach on crop data and historical climate, numerous predictions can be modelled on the basis of data [2].

* Corresponding Author: abdulbasitahmedcs@gmail.com

This work aimed at developing a machine learning model that will help to predict seasonal crop yield using machine learning techniques.

The Objectives are to:

i. To predict crop yield by learning from previous data of the farming land.
ii. To compare different techniques for predicting seasonal farming
iii. To evaluate the efficiency of the techniques
iv. To build a system that predict yields
v. To recommend the most suitable technique among the three (3)

2. Literature Review

[3] presented the adoption of smart farming by understanding the Nitrogen, Phosphorous and Potassium (NPK) and leaf Chlorophyll status of soil in agricultural field as a means of improving yield and to equally have production costs. A link between a dependent variable, Y yield and three predictor variables, X were used using a multiple linear regressions (MLR) model (soil nutrient and crop healthiness). It was created a model to using soil nutrient variable, crop healthiness and interaction variables, predict rice yield with R^2 of 0.6403.

[1] focused on yield prediction by means of random forest only base on the existing data. Real data were adopted for testing and building the models. It made use of a test version of an interactive prediction system. It was put into practice to implement such graphic user interface, user friendly and the machine learning algorithm. It employs the random forest algorithm. There are no suitable solutions or technologies to deal with the scenario we are in right now, despite the analysis of all these concerns and problems including moisture, humidity, temperature, weather, temperature and rainfall.

[4] found a new established field of study known as embedded intelligence (EI). By analysing the digital footprints people leave behind when engaging with the Internet of Things, this field seeks to shed light on individual social patterns, spatial settings behaviours and urban dynamics (smart cards, smart cars and cameras). Knowledge extraction was included from big data and mining of already existing technologies to provide high-depth ontologies or metadata for reasoning and proactive decision-making using open data information. It also analyses the general architecture, research history, significant applications, characteristics and research challenges of EI.

[5] validated root-mean-square-error (RMSE) of 12% of the average yield and 50% of the standard deviation for the validation dataset utilizing anticipated meteorological data, the generated model was determined to have high accuracy. Computational results model performed better than other approaches including Lasso, regression tree (RT) and shallow neural networks (SNN). Result showed environmental factors affect crop production than genetics.

[6] used statistical model and machine learning to build and provide precise and accurate decision which help to select appropriate crops to be grown depending on area and season with minimal risk.

[7] used historical data such as weather, soil conditions and previous crop yield. His work centered on random forest to forecast the yield based on available data. The models were constructed using actual Tami lnadu data and were tested using multiple samples before planting seeds in an agricultural land.

[8] suggested hybrid of sensing technologies and machine learning will result to cost effective way to agricultural problem. It also conducted systematic literature review to identify and compile the characteristics that have been employed in studies to predict crop production. We obtained 567 pertinent papers from six internet resources using our search parameters. We thoroughly examined and evaluated the techniques, features applied and offered recommendations for additional study. Soil type, rainfall and temperature were most used features while artificial neural networks were the most widely used technique models. Remark based on analysis of 50 publications using machine learning, we searched other electronic resources for studies using deep learning. We found 30 such papers and retrieved data from them.

[2] provide a user interface for farmers that analysis the scenario of crop production forecast based on easily available datasets by applying vector support system, data mining and bayesian network. Numerous researchers are engaged in this field and have introduced the procedures of data mining and its uses in allied sectors and agriculture. It further explained the methods for analyzing historical crop output and climatic data on the basis of the data acquired, a range of forecasts was produced and in turn help to increase crop productivity. It further suggested decision support system (DSS) to reduce the burden of making decisions about the soil and crop to be grown.

[9] developed a model using artificial neural networks and have shown to be useful model and prediction tools by boosting their efficiency. Utilized variety of soil and climatic parameters in order to ensure the sustainability technology predicts the best crop to plant. parameters such as soil type, PH, nitrogen, rainfall, phosphate, potassium, calcium, magnesium, temperature and humidity were considered

3. Materials and Method

3.1. Description of Existing System

[10] farmers were worried during farming season and a curious to know the expected yield based on farmers experience for certain crops and climatic condition. This variation in productivity of expected yield have been generating rigorous stress. There is need to get another ways of solving such problems.

3.1.1 Review of Existing System

[6] adopted machine learning and a statistical model to build a precise and accurate decision which help farmers to pick the right crop to plant to avoid shortage at the end of the farming season.

[11] used decision tree to classify crop from available dataset factoring crop to be cultivated and location because of the soil content

3.1.2 Problem with the existing system

[12] due to traditional ways of farming, it is associated with financial and time loss because right machineries were not used.

[13] the methods is capital intensive, subjective and inability to react to differences in of farming.

3.2. Description of the Proposed System

The proposed system will work in such away that input dataset harvested from various agricultural colleges across the country will be modelled. The new system also has provisions for farmers to enter the size of the farmland, number of seeds and number of bags which will be modelled to give accurate prediction. Adoption of machine learning has help with lots of process that has been carried out manually. This models help farmers in the prediction of crop yield within short time. In the implementation we used support vector machine, random forest and decision tree model to make the classifier strong.

3.3. HIGH LEVEL MODEL OF THE NEW SYSTEM

Figure 1 present High Level Model of the System that consists of Prediction System, Farming Parameters, Analysis of Result and Result Display.

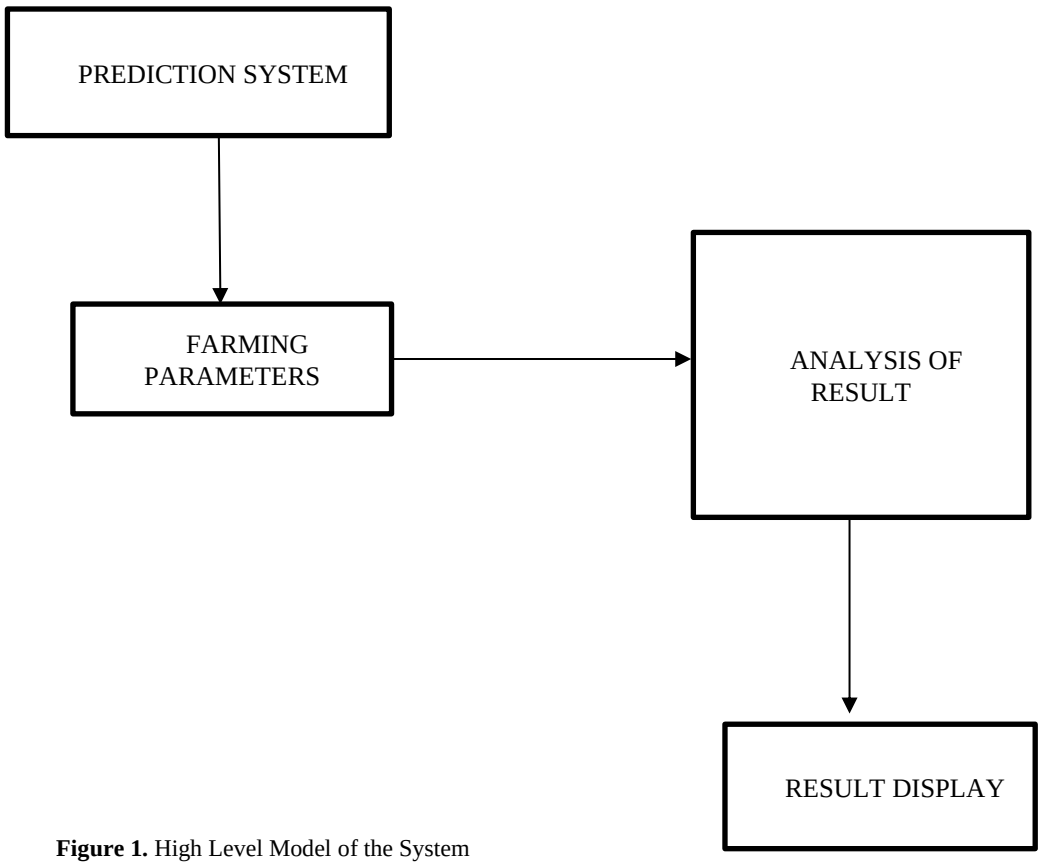

Figure 1. High Level Model of the System

3.4. Methodology

The process of submitting, extracting and storing was repeated several times. The methodology adopted for the modelling was Cross Industry Standard Processing for Data Mining meaning (CRISP-DM) and it is the most widely used in research institute and industry and it has six (6) stages:

i. Business Understanding
ii. Data Understanding
iii. Data Preparation
iv. Modelling
v. Evaluation
vi. Deployment

The methodology adopted was to study and understand the dataset. That is, the objectives of developing the predictive model. At the stage, research is made to understand what is to be predicted. Then I advanced to identification and understanding of the datasets, the relationships between each contents of the dataset and the characteristics of the attributes of each contained in the dataset. This stage allows us to determine the quality of the data contained in the dataset and also describe the steps in documenting the project. The dataset was analysed and prepared for modelling. Data cleaning was performed to identify errors in our dataset. Unwanted noise was also removed and was converted for modelling, the dataset was splitted into training and testing.. After a successful training model, we then evaluate the performance of the model. This is the stage the performance evaluation of the model was done by evaluating the result to know how accurate and precise the model is, at this stage we used the Root Mean Square Error and R^2 to evaluate the accuracy of the model. This is the stage where we deployed the prediction model into the computer by entering values for prediction.

4. Result

After thorough evaluation of the value gotten from each zone using different algorithm, comparison was made between the zones using random forest, support vector and decision tree classifier. Results shows Decision Tree Classifier for South East with the least value give the best output based on their performance and accuracy. The result also shows that South East has the best in terms of accuracy for rainy season farming with model performance evaluation of 138.9 using Decision Tree Classifier while South South has the worst in terms of accuracy for rainy season farming with model performance evaluation of 2400.8 as shown in Table 1. It is recommended for farmers within South East region to adopt Decision Tree Classifier so at to get good yields at the end of the farming season.

Table 1. Model Performance Evaluation

S/no	Zone/Season	Support Vector Machine	Random Forest	Decision Tree Classifier
1	North Central/Rainy	1412.6	432.8	332.2
2	North Central/Dry	739.6	318.8	259.1
3	North East/Rainy	825.5	260.2	637.7
4	North East/Dry	675.6	322.1	659.9
5	North West/Rainy	739.6	241.0	247.8
6	North West/Dry	699.3	182.1	175.3
7	South East/Rainy	699.3	187.2	138.9
8	South East/Dry	844.8	187.4	147.6
9	South West/Rainy	923.2	353.2	183.7
10	South West/Dry	923.2	353.2	496.2
11	South South/Rainy	2400.8	623.9	620.0
12	South South/Dry	2086.4	704.1	644.2

Figure 2 present accuracy for dry season farming in North Central.

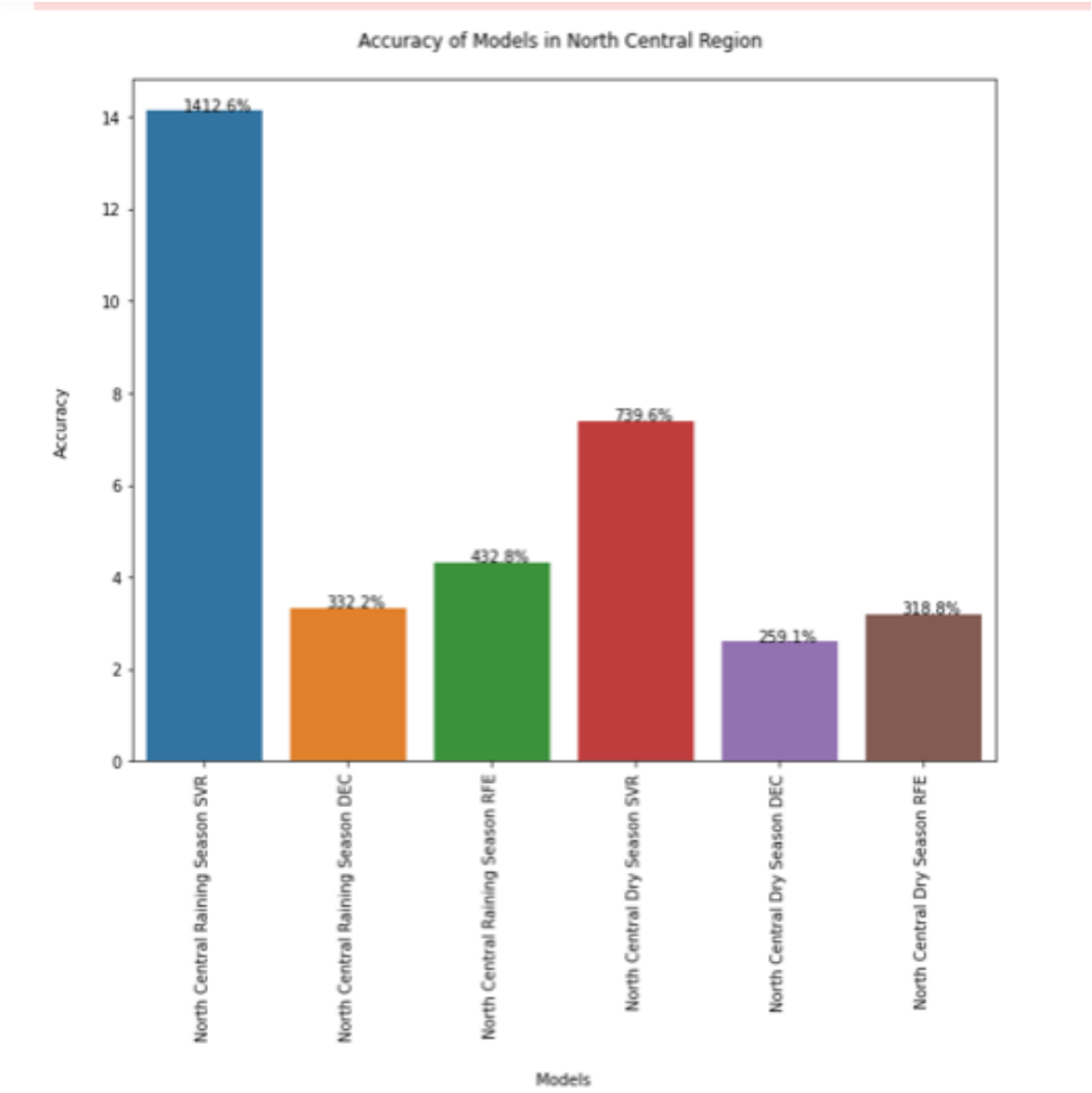

Figure 2. Decision Tree Classifier has the best accuracy during dry season for North Central.

Figure 3 present accuracy for dry season farming in North East.

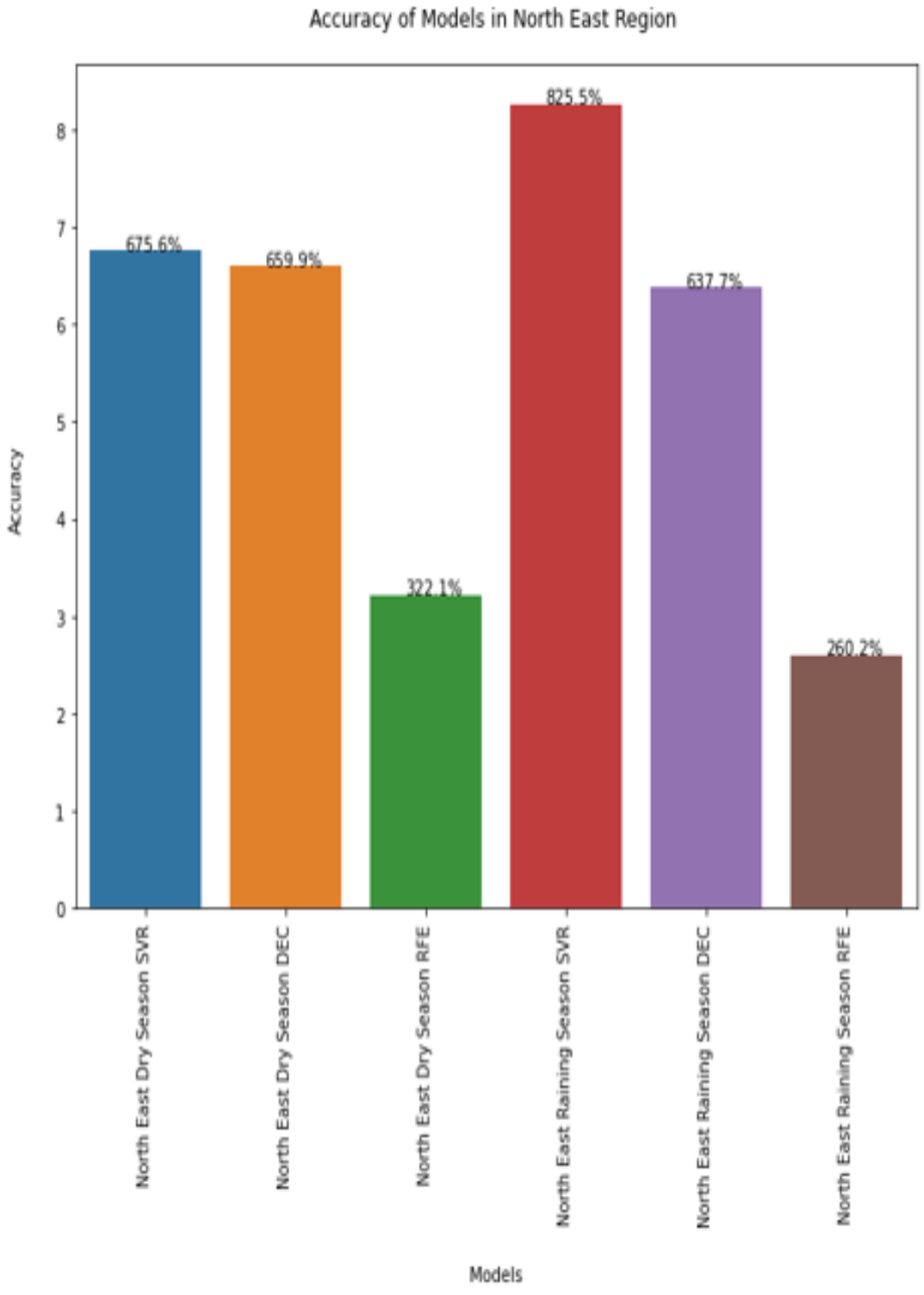

Figure 3. Decision Tree Classifier has the best accuracy during dry season for North East

Figure 4 present accuracy for dry season farming in North West.

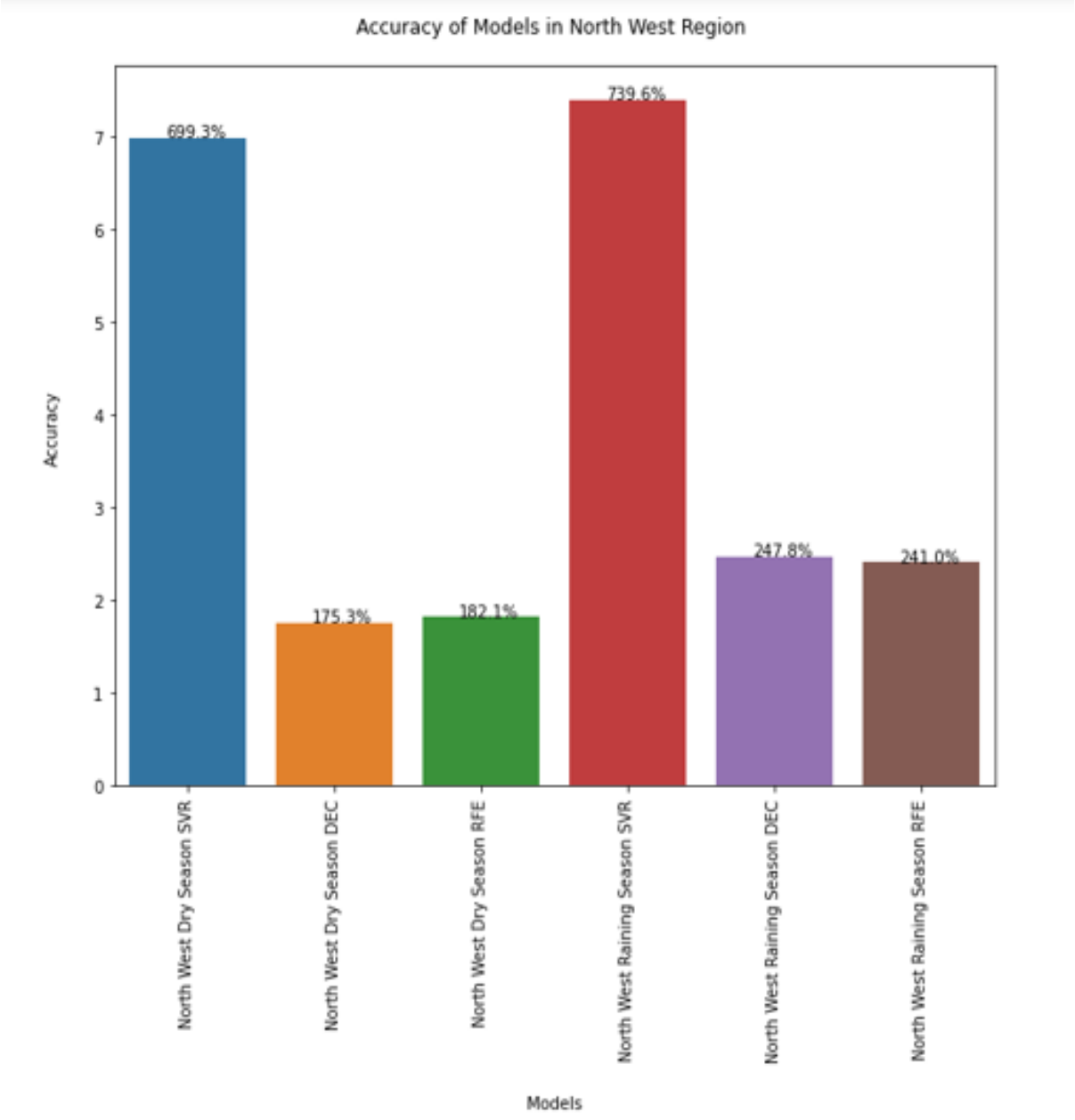

Figure 4. Decision Tree Classifier has the best accuracy during dry season farming for North West

Figure 5 present accuracy for raining season farming in South East.

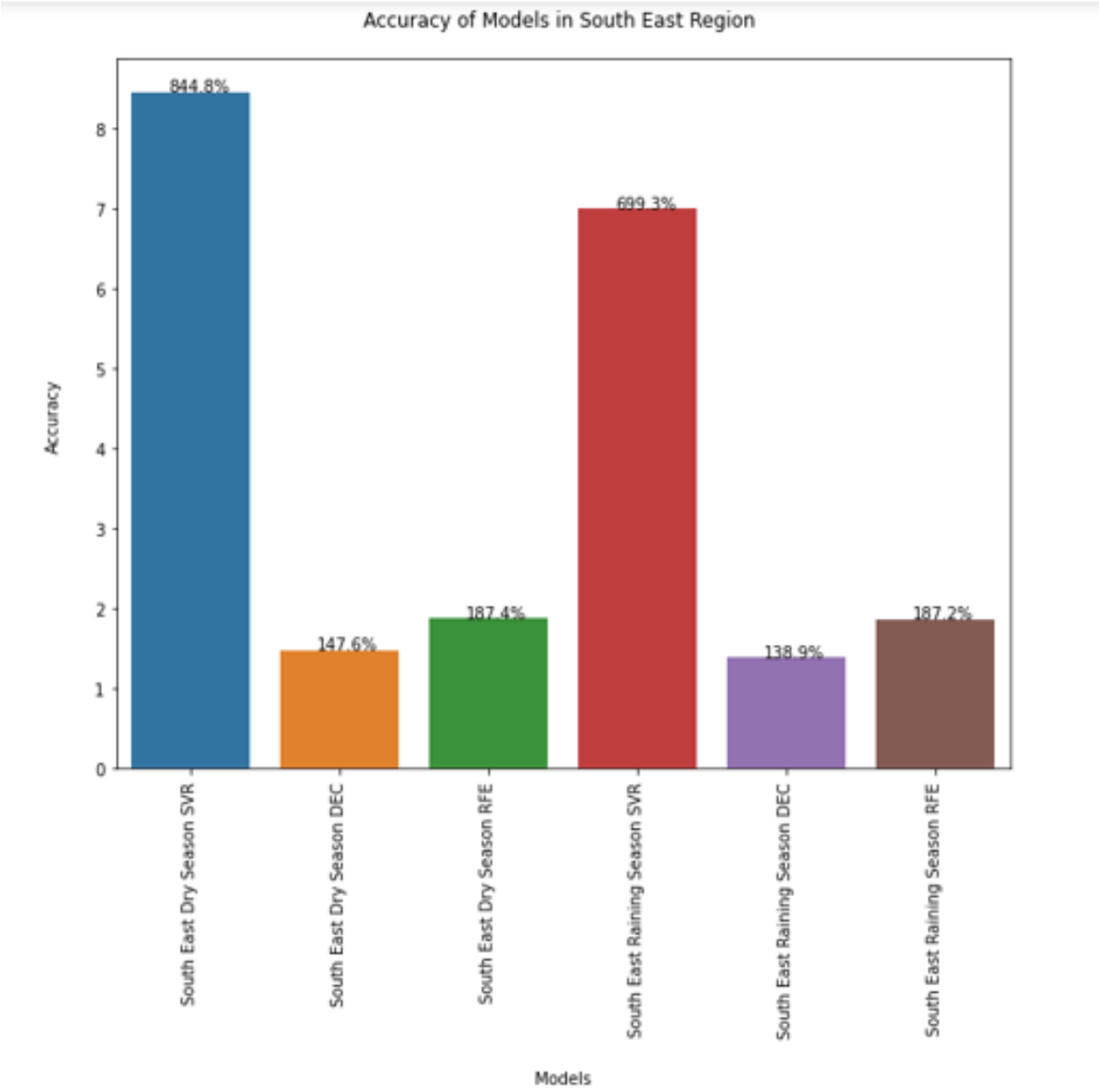

Figure 5. present Decision Tree Classifier has the best accuracy during raining season farming for South East

Figure 6 present accuracy for raining season farming in South West.

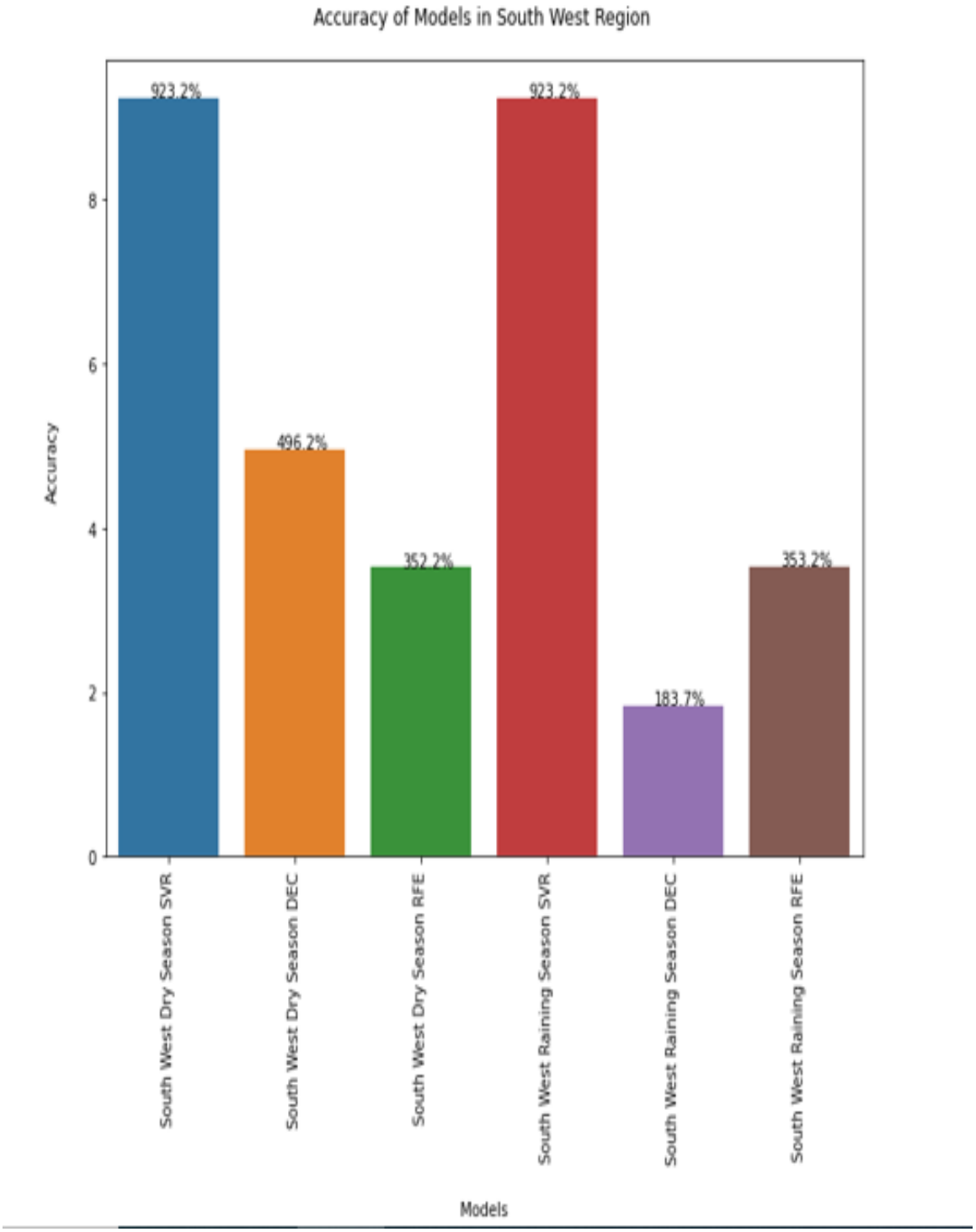

Figure 6. Decision Tree Classifier has the best accuracy during raining season farming South West

Figure 7 present accuracy for raining season farming in South South.

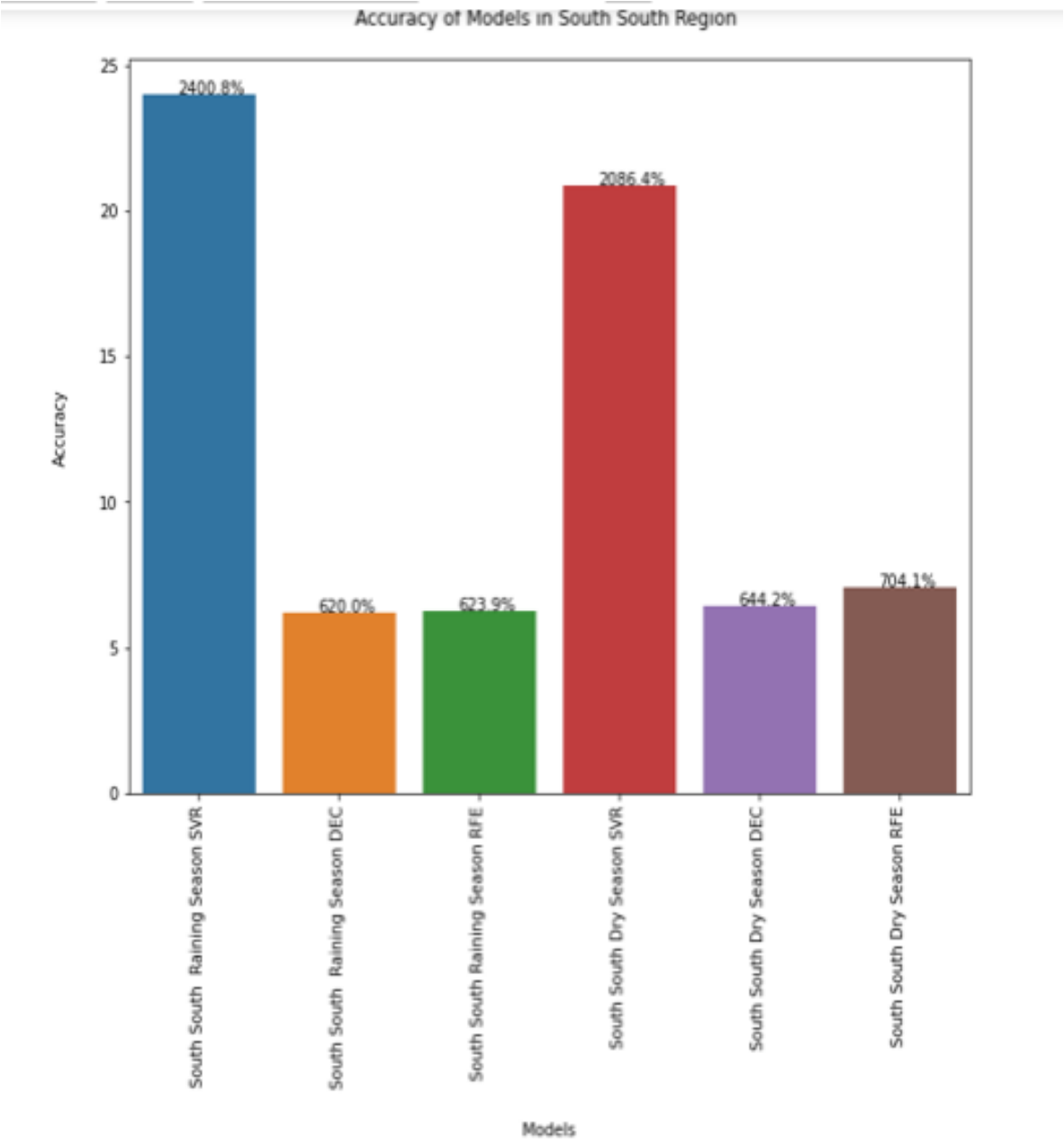

Figure 7. Decision Tree Classifier has the best accuracy during raining season farming South South

5. Conclusion

The research has shown wide gap between traditonal ways of predicting and machine learning techniques using softwares to run modelling and prediction. It can be concluded that the new system will go a long in proffering solutions to problems faced using traditional methods and also provide solutions to challenges faced when manual prediction is used. It can be concluded by saying that South East has the best in terms of accuracy for rainy season farming with model performance evaluation 138.9 using Decision Tree Classifier and South South has the worst in terms of accuracy for rainy season farming with model performance evaluation 2400.8

References

[1] Mayank, C., Darpan, C., Chaitanya, C., & Mansing, R. (2020). Crop Yield Prediction Using Machine Learning. *International Journal of Science and Research (IJSR)*, 645.

[2] Kusum, L., & Bhushan, C. (2020). Crop yield prediction using data mining techniques and machine learningmodels for decision suppport system. *Journal of Emerging Technologies and Innovative Research (JETIR) Volume 6, Issue 4*, 391, 393-395.

[3] Putri, R. E., Yahya, A., Adam, N. M., & Abd Aziz, S. (2019). *Rice yield prediction model with respect to crop healthiness and soil fertility.* Indonesia: Food Research.

[4] Wu, Y., Li, S., Li, L., Li, M., Li, M., Arvanitis, K. G., . . . Sigrimis, N. (2018). Smart Sensors from Ground to Cloud and Web Intelligence. *International Federation of Automatic Control Conference Paper* (p. 32). China: Elsevier.

[5] Saeed, K., & Lizhi, W. (2019). *Crop Yield Predictions using Deep Neural Network.* United States of America: Frontiers in Plant Science.

[6] Manoj, G. S., Prajwal, G. S., Ashoka, U. R., Prashant, K., & Anitha, P. (2020). Prediction and Analysis of Crop Yield using Machine Learning Techniques. *International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181*, 19.

[7] Priya, P., Muthaiah, U., & Balamurugan, M. (2018). Predicting yield of the Crop using Machine Learning Algorithm. *INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY*, 1.

[8] Thomas, V. K., Ayalew, K., & Cagatay, C. (2021). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 2.

[9] Snehal, S. D., & Sandeep, V. R. (2014). Agricultural Crop Yield Prediction Using Artificial Neural Network Approach. *INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING*, 684.

[10] Shreya, S. B., Kalyani, A. B., Aarti, G. D., & Bhagyashree, R. G. (2016). Crop and Yield Prediction Model. *INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH AND ENGINEERING TRENDS*, 23.

[11] Aksheya, S., Monisha K., Pavithra, R., & Marish, H. B. (2020). Crop Selection and it's Yield Prediction. *International Journal of Recent Technology and Engineering (IJRTE)*, 4127.

[12] Deepak, G., Deepika, J., Dharshini, M., & Vanathi, B. (2020). Crop Yield Prediction Based on Ensemble Model Using Historical Data. *International Journal of Advance Research and Innovative Ideas in Education (IJARIIE)*, 670.

[13] Aakunuri, M., & Narsimha, G. (2016). Crop Yield Prediction with Aid of Optimal Neural Network in Spatial Data Mining: New Approaches. *International Journal of Information & Computation Technology*, 28.