International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 1, ISSN No. 2455-2143, Pages 187-193
Published Online May 2020 in IJEAST (http://www.ijeast.com)

CROP YIELD AND FERTILIZERS PREDICTION USING DECISION TREE ALGORITHM

T. Nikhil Yadav
Department of CSE
Institute of Aeronautical Engineering,
Hyderabad, Telangana, INDIA.

Dr. R Obulakonda Reddy
Department of CSE,
Institute of Aeronautical Engineerning,
Hyderabad, Telangana, INDIA.

Ram Chandra Prasad
Department of CSE,
Institute of Aeronautical Engineering,
Hyderabad, Telangana, INDIA.

Pradeep Gopal
Department of CSE,
Institute of Aeronautical Engineering,
Hyderabad, Telangana, INDIA.

***Abstract*—In India most of the people ultimately depends on agriculture, some percentage of people directly occupied by farming while other does trading with final products. India is producing huge amount of food grains which have great effect on Indian economy. Data Mining is sky rocketing in the field of crop production. Every farmer is interested in knowing about the yield he is about to get. This study will attempt to quantify past weather and technological factors, analyze their effects on yields and predict future crop yield based on season, area, temperature, humidity, moisture, amount of nutrients value of potassium, nitrogen and phosphorous. Then all these factors are divided into training and test datasets and tested with different machine learning models. By this model we can accurately predict the crop yield and fertilizer to be used in the area of land, as a result it will help the agronomist to increase their income and improve country's economy.**

Keywords— **Random forest Regression, Decision tree Regression, definite prediction.**

I. INTRODUCTION

Indian agriculture sector gives huge employment in different fields of cultivation industry. India cultivates many different kinds of crops, but produces more amounts of rice and wheat. Indian agronomist also farm urad, jowar, small millets, sugarcane and non food crops like hemp, jute, cotton. But nowadays the weather conditions of the atmosphere are not stable like previous days, because of global warming and other man made calamities. The weather conditions are highly unstable and are very difficult for the farmers to rely upon the traditional agricultural techniques.

The farmers need more effective measures for predicting the weather conditions which in order gives the farmers a better crop yield. Indian agriculture mainly depends on rainfall and ground water availability. According to Rahul tangia (2019) crop production of India is decreasing per capital compared to other countries as show in fig 1.

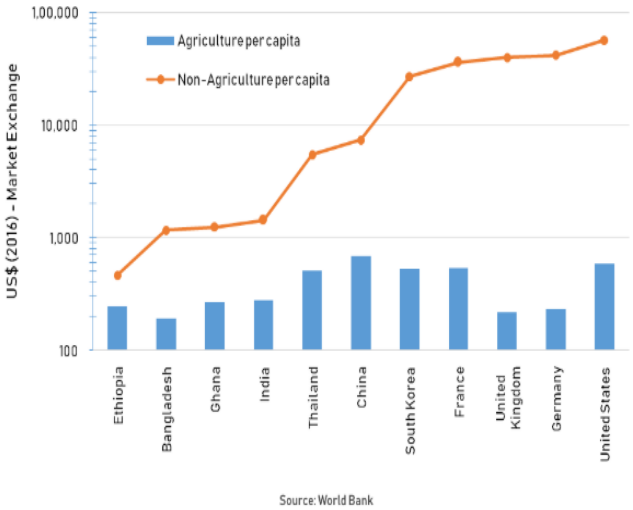

Fig. 1 GDP per capita

If we are able to predict the specific crop production at a particular land it will be very useful to the farmer to choose a specific crop to cultivate, as a result crop production capital will increase. In India there are fifteen agro climatic regions and these agro climatic regions are separated into groups based on season, type of land, temperature, soil type and its fertility. By considering these factors we can help the agronomist to predict the yield of a crop and suggest fertilizers to framer which he can use in the field. Generally, framer predicts the production by using traditional methods and experience, but due to climatic changes agronomist are unable

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 1, ISSN No. 2455-2143, Pages 187-193
Published Online May 2020 in IJEAST (http://www.ijeast.com)

to predict the yield. By using different machine learning models we can predict the most accurate output of the yield.

II. LITERATURE SURVEY

Niketa Gandhi et al. (2016) have used different classifiers and showed that LADTree and J48 classifiers have showed sky-high accuracy, however SMO classifiers are less definite than other family of classification like naïve bayes. Here mainly four factors that were taken are precipitation, average temperature, maximum temperature and minimum temperature to predict crop yield. This system has only analyzed rice crop data which made the model limited to use, as shown in fig 2.

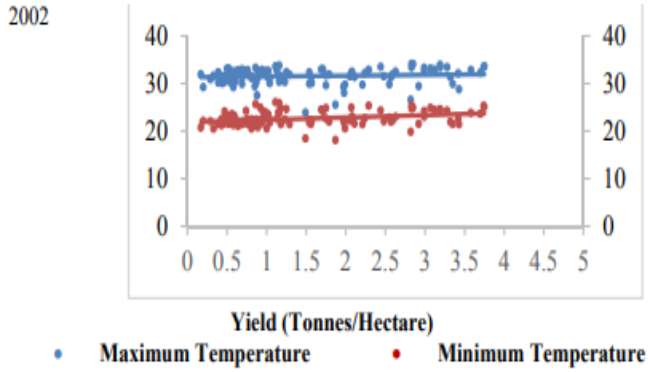

Fig. 2 Prediction of rice crop yield

K.E. Eswari et al. (2018) has used Baysian Network to predict the yield and proved that Naive Bayes classifier has more accurate rate compare to sequential minimal optimization classifier. In model author has taken seven factors as show in fig 3.

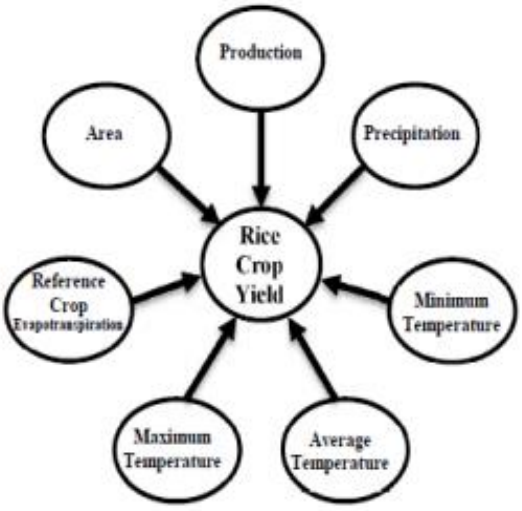

Fig. 3 Dataset attributes

Additional factor consider in this model is evapotranspiration, which is combination of evaporation as well as transpiration and responsible for absorption water vapors from plants. In this classification model prediction is done by using two classes which are true class and false class and further divided into true negative, true positive, false negative and false positive as shown in fig 4.

Predicted	Observed		
		True	False
	True	True Positive (TP)	False Positive (FP)
	False	False Negative (FN)	True Negative (TN)

Fig. 4 Classification classes used in Baysian Network

Shruti Mishra et al. (2018) used various data mining classifiers such as Instance Based Learner (IBK), Locally Weighted Learning (LWL), LADTree and J48 to predict the production of crop. They have considered several factors like name of district, year of crop, name of the season, name of crop, area and these factors are pre-processed and classified in WEKA explorer. They have predicted error rate by using Root Mean Squared Error, Mean Absolute Error and Relative Absolute Error as shown in fig 5.

Algorithm	RMSE	MAE	RAE (%)
J48	0.2773	0.1101	37.9755
LWL	0.3209	0.2213	76.3471
LAD Tree	0.4127	0.1997	68.8888
IBK	0.3057	0.104	35.8648

Fig. 5 Error Rate

As per author [1] J48 classifier have got higher accuracy but in this proposed model Instance Based Learner classifier (IBK) got higher accuracy as shown in fig 6.

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 1, ISSN No. 2455-2143, Pages 187-193
Published Online May 2020 in IJEAST (http://www.ijeast.com)

Algorithm	Accuracy (%)
J48	78.145
LWL	66.225
LAD Tree	62.251
IBK	80.794

Fig. 6 Accuracy percentage

III. IMPLEMENTATION METHODOLOGY

Estimating the production of crop using the potent algorithm and prediction of which type of fertilizer to be used on crop

A. Data Set Explanation

In this model we are utilizing two datasets. Fig 7 is for prediction of crop yield, this data frame consists of six factors which are trained by various machine learning techniques to predict the production of crop and from fig 8 we can predict the fertilizer name which should be used for better yield of a crop. Fertilizer prediction is based on six factors; these factors are temperature, humidity, moisture, nitrogen, potassium and phosphorous.

State_Name	District_Name	Crop_Year	Season	Crop	Area
Andaman and Nicobar Isla…	NICOBARS	2000	Kharif	Arecanut	1254.000000
Andaman and Nicobar Isla…	NICOBARS	2000	Kharif	Other Kharif pulses	2.000000
Andaman and Nicobar Isla…	NICOBARS	2000	Kharif	Rice	102.000000
Andaman and Nicobar Isla…	NICOBARS	2000	Whole Year	Banana	176.000000
Andaman and Nicobar Isla…	NICOBARS	2000	Whole Year	Cashewnut	720.000000
Andaman and Nicobar Isla…	NICOBARS	2000	Whole Year	Coconut	18168.000000

Fig. 7 Sample crop data

Temparature	Humidity	Moisture	Nitrogen	Potassium	Phosphorous
26	52	38	37	0	0
29	52	45	12	0	36
34	65	62	7	9	30
32	62	34	22	0	20
28	54	46	35	0	0

Fig.8 Sample fertilizer data

B. Necessary Packages

- Numpy
- Pandas
- Tkinter
- Scikit-learn
- Matplotlib

C. Architecture

The Fig. 9 shows the architecture diagram used in this model.

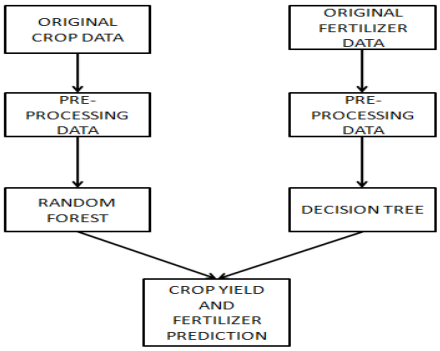

Fig. 9 Architecture

D. Metadata

In this stage we are going to encode all the categorical data (state name, district name, season and crop) into a unique number that is one unique number is given to one state, district, season and crop as shown in fig 10. The same number is not given to the other elements. By converting original data into metadata we can easily process the dataset. This converted data consists of more than 35 states, 600 districts, 5 seasons and around 100 crops all over India.

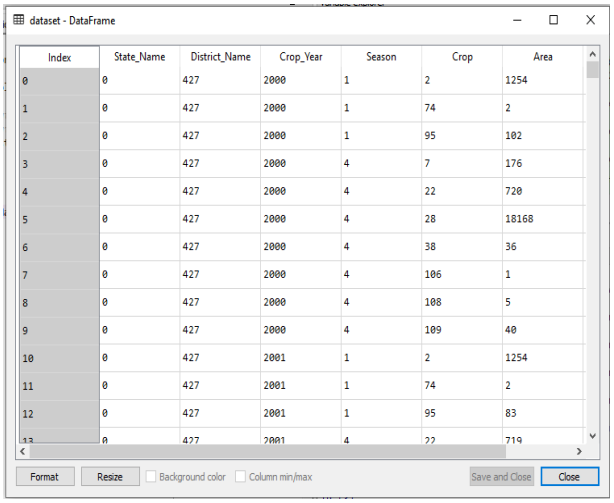

Fig. 10 Metadata

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 1, ISSN No. 2455-2143, Pages 187-193
Published Online May 2020 in IJEAST (http://www.ijeast.com)

E. Data Pre-processing

In this stage the original crop data set is preprocessed (like taking care of missing data) and the converted data (metadata) is added to it by removing all the factors that are converted to the integer (state, district, season and crop) as shown in fig 10. So, by doing this we can easily train the data.

For splitting the dataset into training set and test set we need to use package sklearn model selection and import train_test_split, this helps to successfully split the training data and test data according to the given split value. The split size of the test data should be 20% and training data should be 80%.

F. Random Forest Regression

Random Forest regression is a flexible and easy algorithm which provides a potent prediction. Random Forest regression is a machine learning technique used for prediction.

Working of random forest regression:-

- First takes random sampling from the training dataset, depending upon the value given in estimators (n) and random state.
- Constructs n decision trees from sampled data.
- Each decision tree returns some predicted value, on averaging all prediction we get random forest Prediction.

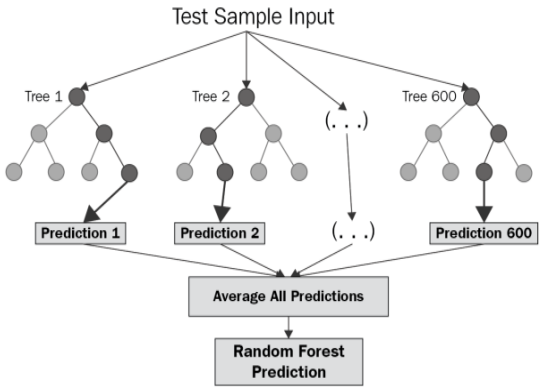

Fig.11 Random forest working model

G. Fertilizer prediction using Decision Tree Algorithm

The Fertilizer dataset which is in the CSV format is cleaned and made ready to train the data frame. Here split size of training dataset is 80% and test dataset is 20%. Decision Tree algorithm used to train the dataset. Decision tree algorithm consists of supervised learning algorithm for predicting output target features data into incrementally smaller nodes. Top node is root node, internal nodes are decision nodes and terminal nodes are leaf nodes. Tree pruning and time series cross-validation are used for lowering variance error source generated by a greater model complexity. This algorithm consists of greedy top-down approach for finding optimal recursive binary node splits by locally minimizing variance at terminal nodes measured mean square error function at each stage.

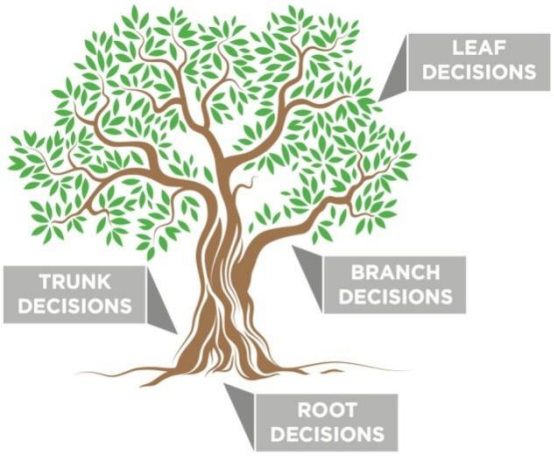

Fig. 12 Decision tree structure

IV. Comparison With Other Algorithm

Here we are comparing Random Forest Regression with other algorithms like Multiple Linear Regression and Polynomial Regression.

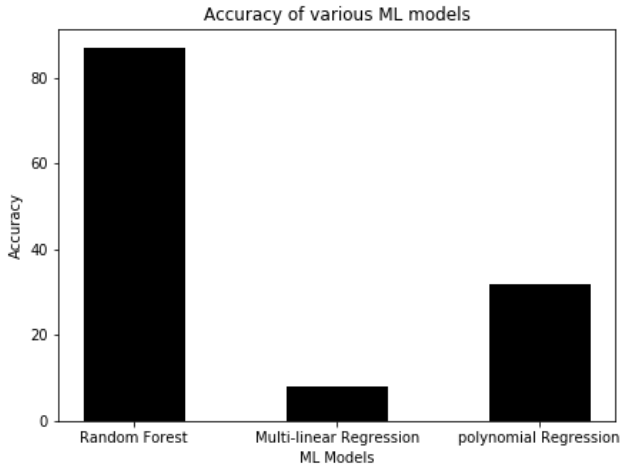

Fig.13 Accuracy of various ML models

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 1, ISSN No. 2455-2143, Pages 187-193
Published Online May 2020 in IJEAST (http://www.ijeast.com)

As shown in fig 13 the accuracy of Random Forest Regression is 87, Multiple Linear Regression is 8 and Polynomial Regression is 32. Accuracy is calculated by cross validation scores (mean squared error). From the above accuracy figures it is proved that random forest regression has higher accuracy and performance compare to other algorithms.

V. PERFORMANCE ANALYSIS & RESULT

Here result is represented by using tkinter user interface. Crop yield prediction is done by Random Forest regression and fertilizer prediction is done Decision Tree algorithm. Random Forest model was experimented with different types of attributes like state, district, year, season, crop and area in various regions across India to predict the result. Fertilizer dataset is train with six attributes to predict the fertilizer used on the crop land. For crop yield prediction the user should enter seven fields (state name, district name, year, season name, crop name and area) to get the output is shown in Fig. 14. The result of the crop yield prediction is shown in Fig. 15. The input of the fertilizer data takes six factors; temperature, humidity, moisture, nitrogen value, potassium value and phosphorous value as shown in fig 16 and the result of the fertilizer data is shown in fig 17.

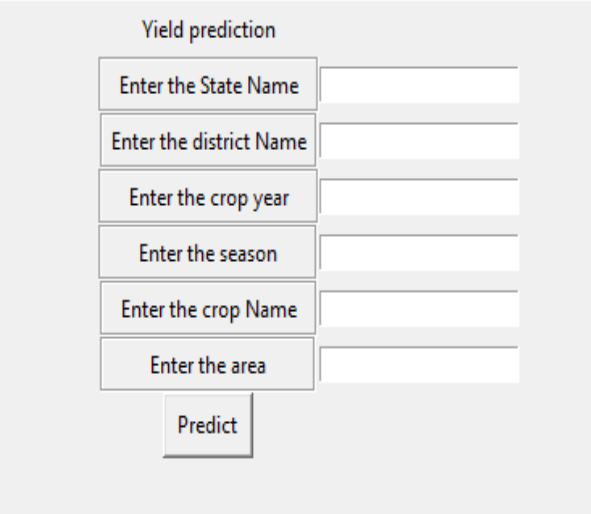

Fig. 14 Crop data input

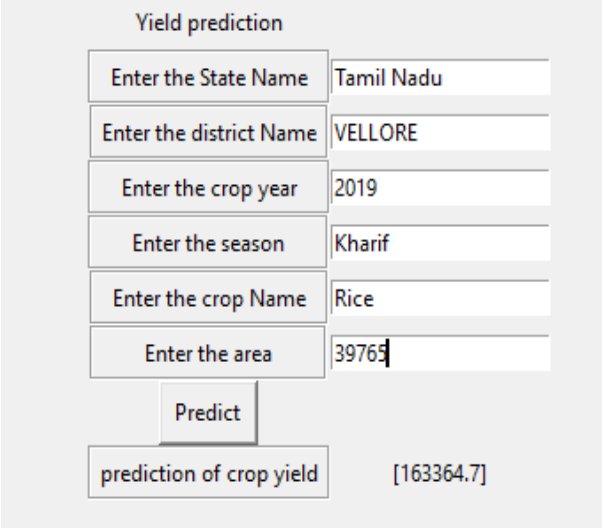

Fig. 15 Result for the given crop input

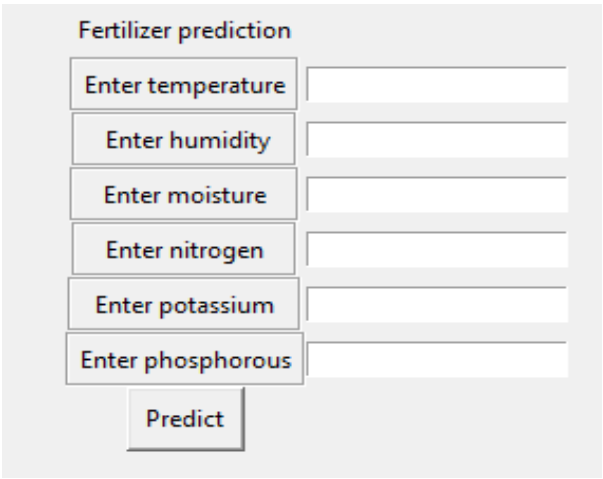

Fig. 16 Fertilizer input

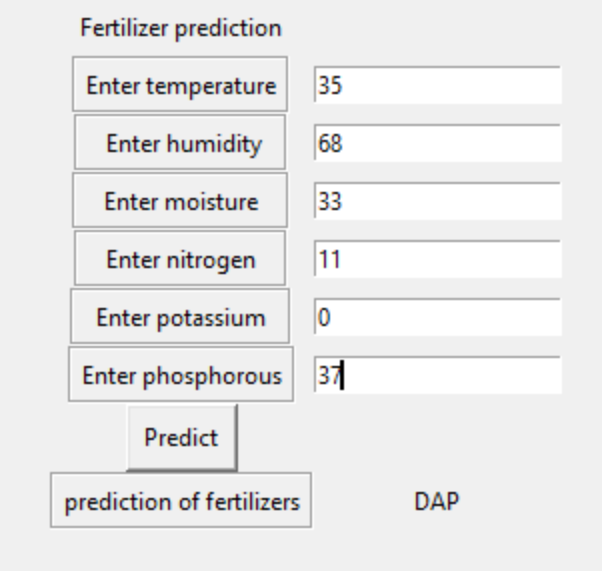

Fig. l7 Result for the fertilizer input

VI. CONCLUSION

Prediction of crop yield and fertilizer prediction is successfully predicted by using efficient machine learning algorithm (Random Forest Regression and Decision Tree Algorithm). The experimental output tells that Random Forest Regression got highest accuracy percentage and performance compared to other algorithms used. The result is developed using python Tkinter which is a GUI tool kit. In future we can develop using the app application which makes the user to use this application more effectively.

VII. ACKNOWLEDGEMENT

I wish to acknowledge Dr. R Obulakonda Reddy and Mohan Kalyan for their supervision. I also wish to thank Ram Chandra Prasad and Pradeep Gopal for their contribution in making this research published.

VIII. REFFERENCE

[1] Niketa Gandhi et al. (2016)," Rice Crop Yield Forecasting of Tropical Wet and Dry Climatic Zone of India Using Data Mining Techniques",IEEE International Conference on Advances in Computer Applications (ICACA) .

[2] K.E. Eswari. L.Vinitha. (2018) "Crop Yield Prediction in Tamil Nadu Using Baysian Network ", International Journal of IntellectualAdvancementsand Research in Engineering Computations, Vol-6,Issue-2,ISSN: 2348-2079.

[3] Shruti Mishra, Priyanka Paygude, Snehal Chaudhary, SonaliIdate. (2018) "Use of Data Mining in Crop Yield Prediction" IEEE Xplore Compliant - Part Number:CFP18J06-ART, ISBN:978-1-5386-0807-4; DVD PartNumber: CFP18J06DVD ,ISBN:978-1-5386-0806-7.

[4] Anna Chlingaryana, Salah Sukkarieha, Brett Whelanb(2018) — Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review , Computers and Electronics in Agriculture 151 61–69,Elisver.

[5] Dakshayini Patil et al (2017),"Rice Crop Yield Prediction using Data Mining Techniques:An Overview", International Journal of Advanced Research in Computer Science and Software Engineering ,Volume 7, Issue 5.

[6] A.T.M Shakil Ahamed, NavidTanzeem Mahmood, Nazmul Hossain, Mohammad Tanzir Kabir, Kallal Das, Faridur Rahman, Rashedur M Rahman (2015)"Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in Bangladesh" 978-1-4799-8676-7/15/$31.00 copyright 2015 IEEE SNPD .

[7] SnehalS.Dahikar, Dr.SandeepV.Rode(2014) ,"Agricultural Crop Yield Prediction Using Artificial Neural Network Approach ", International journal of innovative and research in electrical, instrumentation and control engineering, volume 2,Issue2.

[8] Ramesh A. Medar (2014) "A survey on data mining techniques for crop yield prediction", International Journal of advance in computer science and management studies, ISSN:2231-7782, volume 2, Issue 9.

[9] S.kanagaSubba Raju et al.(2017), Demand based crop recommender system for farmers, International Conference on Technological Innovations in ICT For Agriculture and Rural Development.

[10] S. Bhanumathi, M. Vineeth, N. Rohit. "Crop Yield Prediction and Efficient use of Fertilizers", (2019) International Conference on Communication and Signal Processing (ICCSP).

[11] E Manjula and S. Djodkitachoumy(2017)International journal of computation Intelligence and informatics,Vol 6:No 4.

[12] ShriyaSahu et al.(2017)," An Efficient Analysis of Crop Yield Prediction Using Hadoop framework based on random Forest approach, International Conference on Computing, Communication and Automation.

[13] Jig Han Jeong et al. (2016), "Random Forests for Global and Regional Crop YieldPredictions", PLOS-ONE DOI: 10.1371/journal.pone.0156571.

[14] Pooja M C et al. (2018) —Implementation of Crop Yield Forecasting Using Data Mining, International Research

Journal of Engineering and Technology (IRJET),p-ISSN: 2395-0072.

[15] Rahul Tangia. (2019, OCTOBER 4). India's Biggest Challenge: The Future of Farming. Retrieved from https://www.theindiaforum.in/article/india-s-biggest-challenge-future-farming