Ann. For. Res. 65(1): 7422-7432, 2022
ISSN: 18448135, 20652445

ANNALS OF FOREST RESEARCH
https://www.e-afr.org/

EXPERT SYSTEM FOR CROP YIELD FORECASTING USING MACHINE LEARNING TECHNIQUES

Meghna Gupta[1], Dr.Chatrapathy K[2], Dr. Naveen Kumar B[3], Dr Bhagavant.K.Deshpandey[4]
[1]Assistant Professor, Department of computer Applications, ABES Engineering College, Ghaziabad, Uttar Pradesh, India, meghna.gupta04@gmail.com
[2]Professor, School of Computing & Information Technology, Reva University, Bangalore, India, pathykc@gmail.com
[3]Associate Professor, Sahyadri College of Engineering & Management, Mangalore, Karnataka, India, navkan24@gmail.com
[4]Professor, Department of Information Science and Engineering, East Point College of Engineering and Technology, Bangaluru, India, deshcapricorn@gmail.com

ABSTRACT
The economy of an agricultural country is primarily dependent on the growth of agricultural yields and agroindustry products. Agricultural crop production depends on seasonal, organic, and monetary causes. Farm yield forecasting is a task that is both of challenging and rewarding for every nation. Due to scarcity of water resources and erratic climate changes, farmers are extremely hostile to producing a yield. Proposed system provides a solution for farmers by tracking field parameters and increase productivity to a large extent by applying integrated data from various sources on Machine Learning techniques to forecast the most cost-effective harvest based on current soil and weather conditions, the productivity of crop yield will be increased by the predictive analysis. As a result of this forecast, the most suitable crop is chosen, increasing the value and benefit of the farming region.

Keywords: Crop Recommender, Crop Forecasting, Machine Learning, Crop productivity,Yield Prediction.

1. INTRODUCTION
As the population of the world continues to increase, scientists have given greater attention to food security. In order to preserve grain security, the distribution pattern and variation of agricultural production must be studied, as this can improve the agricultural management and agricultural planning of the economy. The regional estimate of crop yield was based primarily on vegetation indices derived from remote sensing data. The techniques can be split into single index-based and multiple indices. In the primary method of quantification of crop prediction based on an empirical model at the principal phonological era, vegetation indexes, such as the Normalized Difference Water Index (NDWI), Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) were used. The other approach, which was typically used to forecast crop yields in advance, used time series vegetation indices to optimize the crop process model. As a result, the multiple indices approach will be faster more efficient than the single one.
Crop yield forecasting is an important issue in agriculture. Every farmer always tries to know how much his expectation would yield. In the past, yield forecasts have been calculated by assessing a

farmer's previous experience on a particular crop. Agricultural yield is mainly dependent on weather, pests and crop planning for decision making on the operations of the harvest. For agricultural risk managers, specific information on the history of crop yield is important for decision-making. The research focuses primarily on the development of a prediction model for forecasting crop yield production. The proposed methods forecast yield using data mining techniques based on association rules.

Precision Agriculture or Site-Specific Crops (SSCM) is a crop management concepts based on observation, measures, and response to field variability in crops and aims at increasing crop quantity by implementing latest technology. Primary objectives of Precision Agriculture are sustainability and profitability. Latest technologies and data analysis techniques must be used to identify or recognize crop parameters and consequently direct farmers to cultivate crops and use of fertilizers, as one of the key factors for crop cultivation.

By conducting research on various Machine Learning algorithms, built in functions and libraries in python language are implement to finalize a model that has least error and efficient modules to predict crop yield. In this paper, the generated model output of predicting the crop yield and the performance is compared with existing algorithms and results are represented graphically.

2. LITERATURE REVIEW

Aakunuri Manjula, 2015 suggested a Flexible and Extensible Framework for Predicting Agricultural Crop Yield, indicate that various data mining algorithms like Naïve Bayes and KNN are used to analyze and predict the class of soil dataset, which helps the soil analysis and the farmers to aware about the soil properties to identify best crop that suits to sow, and helps to predict crop yield.

Niketa, 2016 illustrated that crop yield is affected by seasonal climate. In India, the weather changes unconditionally. Farmers are facing serious difficulties in a time of drought. As a result, Expert system for predicating crop yield using Machine Learning techniques is developed to assist farmers in selecting a crop for increased yield. To estimate future data, a series of analysis is conducted using the classification tools like WEKA and SMO to analyze present situation using previous year's real-time data about cultivation, crop, and soil conditions. The crop yield forecasting using SMO-tool classifies earlier data into two classes of high yield and low yield, which shows less accuracy when compared with multilayer perceptron Bayes Network.

Dakashayini Patil 2017 states that rice crops are a major contributor to the country's economy. Different techniques of data mining are used to predict rice yield as it contributes 40% of total yield to reflect the sustainability of India. The yield of crop depends on favorable conditions of the weather; the yield of the crop can be improved by learning a better strategy for crop growth in accordance with climate conditions. The reports employ different mining techniques based on earlier data of yield and various climatic conditions. To predict crop yield, author used data sets from 5 states with known regions and conditions in India.

Chlingaryana 2017, have shown that soil nitrogen levels are the common factor to forecast yield of the crop. Real time data from deployed sensors from fields are primarily used in making-

decision, these enormous remote sensing data helps farmers to understand, how to increase the yield and the steps necessary during cultivation. Nitrogen is used to enhance crop yield and fertile the land, so these algorithms help in taking timely decisions.

Key factors like Soil type, Nitrogen percentage, climate and past data about crop yield are taken into consideration, which are helpful in making accurate decisions, predicting yields in assisting the farmer. Precision farming is aimed at increasing yields and at making suggestions for farmers. It uses modern technology to simplify the task for farmers when they make their decision in due course. It illustrates how to optimize crop production and crop health. The result is the achievement of various vegetarian incidents. The network is used for reverse spreading. Conventional long-term memory neural networks are used for data prediction.

Shruti Mishra 2018 proposed a variety of predictions using data-mining techniques to improve crop productivity with reference to crop production and historical climate conditions. A decision support system was implemented to help farmers make informed decisions about the soil and crops to cultivate.

The data set was collected and analyzed using a variety of flaws, including crop, season, area, and production attributions. These outputs are compared using FOUR techniques with the resulting data. J48, IBK, LAD tree, LWL in WEKA, IBK was found to be accurate when compared with other data depends upon the type of data set and its nature. Eswari 2018, reported that the results of crops are based on perception, average, minimum and high temperatures. In addition, they have taken an additional attribute called crop evapotranspiration. Both the developing stages of the crop and the weather depends on the evapotranspiration of the crop. This attribute is considered to make an effective decision on the group's performance. The data set with these attributes are transmitted to the Bayesian Network as an input and classified into two groups known as true and false, the uncertainty and accuracy matrix of these classifications are compared and concluded that crop yield predictions by Naïve Bays and Bayesian networks are more accurate than SMO classification.

3. METHODOLOGY

India is now making fast progress towards technological growth, modern technologies related to agriculture sector, will improve crop production and country economy. The proposed model provides a solution for smart agriculture, which will allow farmers to increase productivity to a great extent through surveillance of the agricultural sector. Weather forecast data from the IMD (Indian Metrological Department) such as rainfall depository, temperature, soil parameters provide an insight about crops which can be grown in a specific region.

The proposed method incorporates the data sensed and transmitted by the deployed sensory devices, the weather department and Statistics data from Agriculture department and then by applying the Machine Learning algorithms like Multiple Linear Regress, Polynomial regression techniques, to predict crop yield for a particular year on the basis of data set from the past years of a specific state to calculate and visualize the crop yield data of all the specified which state will

have more yield for a specific crop from the considered FIVE states. The performance is assessed based on published information from the Government of India which offers the farmer a variety of crops that can be grown and examines performance on the basis of predicted accuracy, trying to improve profit margins.

3.1 Description of the Data Set

The data set is taken from the Andhra Pradesh State statistics published by the Indian government, which is one of 5 state data analyzed in this paper. Table-I data are used to predict crop output by 9 factors. These 9 factors include the conditions, crops, production, soil, temperature, and the environment, and to predict the crop yield, these machine learning models are trained and tested with past data sets.

TABLE I
Data Set of CROPs and Environmental Condition

Year	Avg. High Temp.	Avg. Low Temp.	Precipitation (Annual Totals) (in mm)	Potential Evotranspiration (Annual Totals) (in mm)	Cloud Cover (in%)	Rice Yield (in kg/hectare)	Wheat Yield (in kg/hectare)	Maize Yield (in kg/hectare)
2005	25	16	1028	60.5	46	2509	2109	2915
2006	26	17	1056	62.8	48	2593	2282	2829
2007	25	16	1098	65.4	45	2573	2602	2924
2008	26	18	1125	66.3	44	2633	2490	3024
2009	28	16	1046	65.8	46	2647	2680	3125
2010	27	19	1280	69.2	39	2739	2760	3450
2011	29	20	1064	67.9	48	2888	2765	3258
2012	30	20	1156	66.4	45	3105	2786	2986
2013	29	18	1087	66.2	48	3059	2791	3057
2014	28	19.5	1198	65.8	52	2946	2807	3152
2015	27	16	1201	67.8	35	2987	2825	3256
2016	26	17.8	1058	66.5	45	2984	2682	3429
2017	28	19.4	1147	68.4	55	2926	2667	3465
2018	29	20.2	1056	69.2	58	3215	2785	3525
2019	30	20	1028	70	61	3358	2895	3475

3.2 Data Implementation

The objective of this regression analysis is to analyze and determines the interaction between the response variable and the explanatory variable. During this analysis, the variables considered for analysis were annual estimation and area under cultivation. Crop yield will be influenced by a variable that is affected by all these environmental factors. Polynomial regression analysis is mainly used for prediction functions as a function of dependent entities it provides a predicted entity. A prediction is referred to predict the value of reaction variable for explanatory variable.

To analyze the data set, we need to import train split that permits to split the pre-processed data into training and testing sets in accordance with the weight defined in the code (scicit-learn). Polynomial Regression applies to trained sets. The entire data set is organized to train the model

and to perform the testing, validation of the model with these data sets in the ration of 0.7 and 0.3, which is respectively 70 and 30 percent, so that the efficiency of the generated model based on the error rate can be certified for successful utilization of this model.

3.3 Use Case Diagram

For modelling an application system/subsystem the use case diagram will be helpful. A single case diagram captures the functionality of the proposed system, so that the whole system in case diagrams can be modelled. A diagram of a Use Case is the interaction of the user with the system is represented in a simple way, and the specifications of an application case are described. A diagram can be used to depict different types of system users and cases, and other forms of diagrams can also be used.

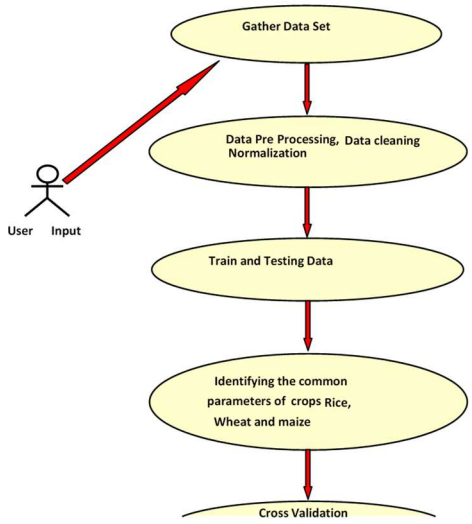


Figure 1: Use Case Diagram of the proposed Model

3.4 Sequence Diagram

An interaction block diagram in a Unified Modelling Language (UML) illustrates how various processes are communicating with each other, and the order in which they build a Message Sequence Chart. Also called sequence diagrams or Event diagrams, Event scenarios and time schedules.

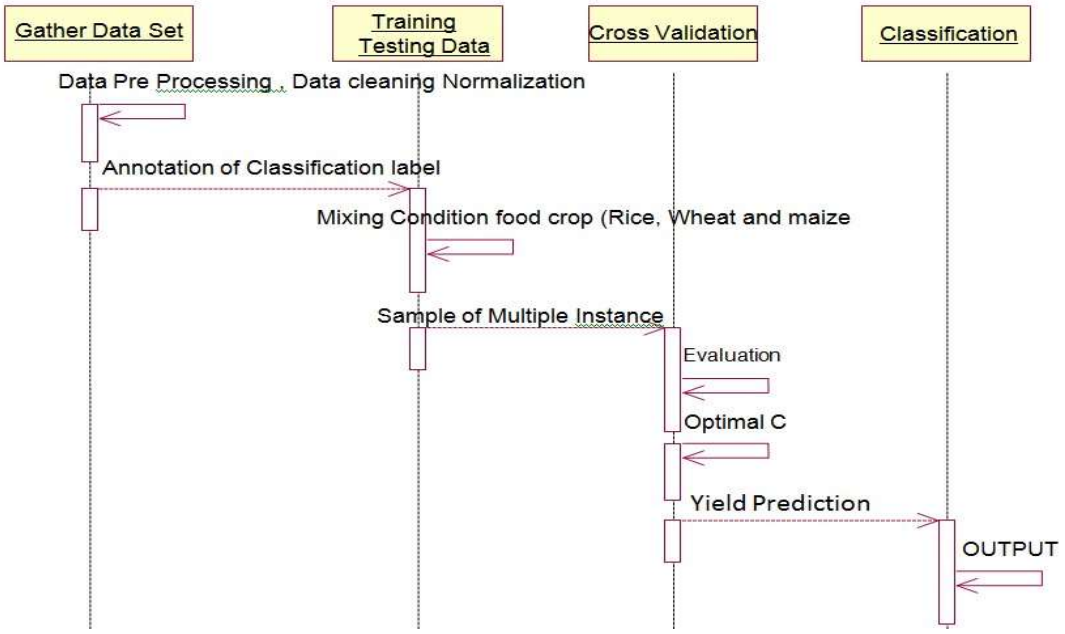


Figure 2: Sequence Diagram of the proposed Model

3.5 Mathematical Model of Proposed Methodology

A Multiple Linear Regression is a statistical model that is used to predict the outcome of a series of events. it uses several independent variables to identify one predictor variable. The objective of this machine learning technique is to shape the linear relation among independent and dependent variables. The following model used a Multi Linear regression technique with two X_1 and X_2 predictor variables.

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$, Where$\beta_0, \beta_1, \beta_2$ ... .are the coefficients of Multiple Linear Regression andX_1, X_2, ... are categorized as independent variables.

The model is linear since the parameters$\beta_0, \beta_1, \beta_2$ ... . are linear. Within the 3D space of Y, X1, and X2, it represents a plane. This plane's intercept is provided by the parameter β_0. Parameters β_1 and β_2 are partial regression coefficients. Parameter $\beta_1$represents the mean response change corresponding to a unit change in X_1, where X_2 is maintained constantly. Parameter $\beta_2$shows the average response corresponding for each unit change X_2 where X_1 is kept constant.

Train data

$Y_1 = \beta_0 + \beta_1 X_{i2} + \ldots\ldots\ldots\ldots\ldots + \beta_p X_{ip} + \epsilon$

For i = 1, 2, 3, 4,……., n

Where β_0 , $\beta_1, \beta_2, \ldots\ldots$ ……are coefficients of Multiple

Linear Regression $X_{i1}, X_{i2} \ldots\ldots\ldots\ldots, X_{ip}$ are independent variables

X {Attributes of soil and weather}

Y {Crop Production}

Y = X*β + E

X = Attribute matrix, □= coefficient matrix
E= Error rate control, Y= Production matrix
$\beta = (X' + X)^{-1}X'Y$ …….. Least Square Estimate
Prediction: $\hat{Y} = X.\hat{\beta}$
Result: $Y - \hat{Y}$

3.6 Functions developed in python Function to normalize the data of a particular required parameter from the data set.

```
def featureNormalize(x):
mean = np.mean(x)
stddev=np.std(x)
X_norm = (np.array(x)-mean) / stddev
X_norm = X_norm.tolist()
return(X_norm)
```

Function to predict or forecast the yield of a specific crop for a selected year by application of polynomial regression technique on the collected data set by considering the learning rate to be 0.001, the maximum number of iterations to be 10000, the penalty to be 0.1 and the tolerance rate to be 1e-5.

```
        x=featureNormalize(x)
y=featureNormalize(y)
reg = Regression()
reg.set_learning_rate(0.001)
reg.set_max_iterations(10000)
reg.set_l1_penalty(0.1)
reg.set_l2_penalty(0.1)
reg.set_tolerance(1e-5)
theta, cost, it = reg.polynomial_regression(x, y, 5)
z = np.linspace(-1.9, 2.1, 4/0.01)
prediction = reg.predict(z)
x = np.array(x) * stddevx+meanx
y = np.array(y) * stddevy+meany
z = np.array(z) * stddevx+meanx
prediction=np.array(prediction)*stddevy+meany
yieldpredict2=NormalEquation([hightemppredict,
lowtemppredict],yieldlist,hightemplist,lowtemplist)
yieldpredict3=NormalEquation([cloudpredict],yieldlist,cloudlist)
yieldpredict4=NormalEquation([prepppredict,evappredict],y ieldlist,preplist,evaplist)
yieldpredict=(yieldpredict1+yieldpredict2+yieldpredict3+y ieldpredict4)/4
```

4. EXPERIMENTAL RESULTS

The objective of this paper is to clarify crops research and process by implementing polynomial and Multi linear regression algorithms to predict the effort to understand the prediction procedure of crop yield and required environmental conditions. Regression analysis is used to evaluate these models with different crops in specific regions across India in order to forecast yield. While training the models soil and atmospheric data were considered and evaluated to determine the required percentage of nitrogen and phosphorus for the fertile land. In predicting the performance of different parameters to compute the error rate, both models with respective to crop production were compared. The performances of these systems are validated based on the obtained error rate generated by the system, while comparing the Multiple Linear Regression and polynomial regression, where polynomial regression generate less error rate when compared with multiple linear regression, Fig.3 depicts the projected out of both the models and the compassion, forecasting Rice yield in Andhra Pradesh state for the year 2022, Fig.4 forecasting of Wheat yield in Indian states for the year 2020, Fig. 5 forecasting the yield of three crops in West Bengal for the year 2020, Fig. 6 forecasting of Rice yield in West Bengal for the year 2021 and Fig. 7 required Atmospheric condition to predict Rice yield in West Bengal for the year 2021.

Figure 3: Forecasting the Yield of Rice crop in Andhra Pradesh statefor the year 2022

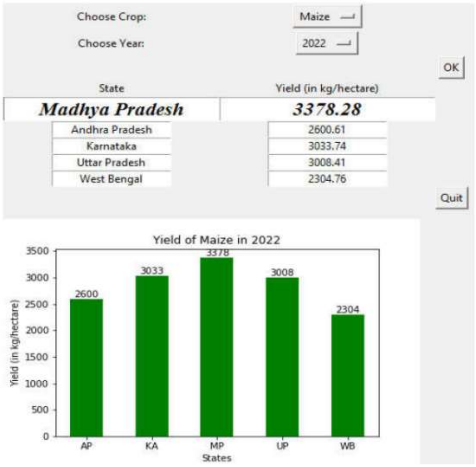
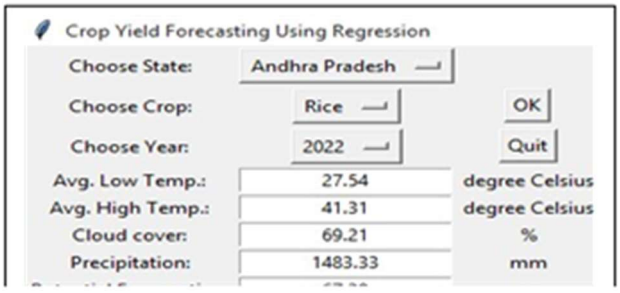


Figure 4: Forecasting the Yield of Maize crop in four statesfor the year 2022



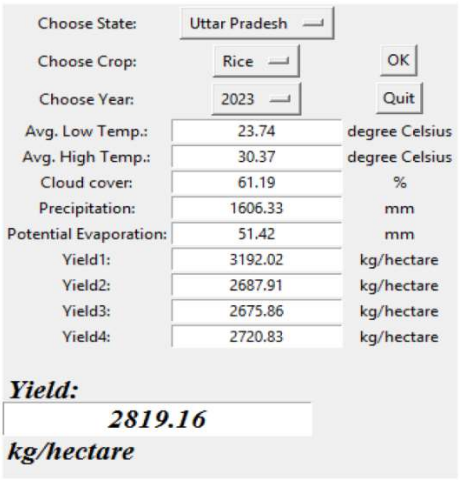


Figure 6: Forecasting the Yield of Rice crop in Uttar Pradesh statefor the year 2023

5. CONCLUSION

The proposed framework forecasts the yields of different crops that are feasible in a specific area and allows farmers in decision making of which harvest to develop. During the investigation process, careful analysis of soil, pH, weather, and crop production data of last year was considered and the planned method during investigation stage and the designed system proposes which yields are the most productive that can be cultivated under the appropriate ecological conditions. The system examines previous production data, to enable farmers to know the market demand and costs of different crops.

After successful implementation of effective algorithms, we reviewed the findings of the forecast as well as the use these predictions in agriculture. This research will advance to the next step, allowing farmers to make more informed decisions about in a particular season which crop should be sown by observing the sensor data from the farm lands and assisting in early detection of crop diseases, so by taking effective actions with performances, they will gain more profit.

TABLE II

Efficiency of the model compared with other Machine Learning Techniques

RICE YIELD PREDICTION IN KG/Hectare				
YEAR	SVM	KNN	PROPOSED MODEL	GOVT. DATA
2020	2874	2741	2604	2578
2021	2946	2842	2685	2798
2022	3018	3145	2946	2980
WHEAT YIELD PREDICTION IN KG/Hectare				
YEAR	SVM	KNN	PROPOSED MODEL	GOVT. DATA
2020	3454	3359	3342	3371
2021	3685	3422	3374	3428
2022	3728	3629	3571	3616
MAIZE YIELD PREDICTION IN KG/Hectare				
YEAR	SVM	KNN	PROPOSED MODEL	GOVT. DATA
2020	3140	3198	3168	3032
2021	3298	3340	3147	3248
2022	3624	3458	3392	3461

REFERENCES

[1] *SnehalS.Dahikar, Dr.Sandeep,V.Rode, Agricultural Crop Yield Prediction Using Artificial Neural Network Approach", International journal of innovative and research in electrical, instrumentation and control engineering, vol. 2, no. 2,(2014).*

[2] *D.S.Zingade, Omkar Buchade, Bilesh Mehta, ShubhmGhodekar, Chandan Mehta, Crop prediction System using Machine Learning, International Journal of Advance Engineering and Research Development, vol. 4, No. 5, (2017), pp. 1-6.*

[3] *K. Samundeeswari, K. Srinivasan, Crop yield prediction and soil data analysis using data mining techniques in krishnagiri district, International Journal of Computer Sciences and Engineering, Vol.06, No.08,(2018), pp. 49-55.*

[4] *Aditya Shastry, H.A. sanjay, Prediction of crop yield using Regression Techniques, International Journal of Soft Computing, Vol.3, No. 12(2),(2017), pp. 96-102.*

[5] *BhanuprakashDudi., Medicinal plant recognition based on CNN and machine learning, International journal of advanced trends in computer science and engineering, vol. 8 No.4, (2019), pp. 999-1003.*

[6] *Balram, G., Kiran Kumat, K. , Smart farming: Disease detection in crops,,International Journal of Engineering& Technology, Vol.7, No. 27, (2018), pp.33-36.*

[7] *Srinivasan, K., Sharmila B, Devasena D, Murugavelrajan., Development of seed sowing machine and monitoring soil nutrients for farmers, Journal of Green Engineering, vol. 11, No. 2, , (2021), pp. 2018-2028.*

[8] *X. K. Chen and C.H. Yang, Characteristic of agricultural complex giant system and national grain output prediction, System Engineering Theory and Practice, vol. 6, No.6 (2002), pp. 120-125.*

[9] *Mann, M. L., Warner, J. M., & Malik, A. S. Predicting high-magnitude, low-frequency crop losses using machine learning: an application to cereal crops in Ethiopia. Climatic Change, Vol.154, No.2,(2019), pp.211-227.*

[10] *Anna Chlingaryan, Salah Sukkarieh, Brett Whelan, Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review, Computers and Electronics in Agriculture, Vol. 151,(2018), pp.61-69*

[11] *D.S.Zingada, OmakarBauchade, Machine Learning Based Crop Prediction System using Multi-Linear Regression, Vol.3, No.2, (2018), pp. 31-37.*

[12] *Sahu, S., Chawla, M., &Khare, N, An efficient analysis of crop yield prediction using Hadoop framework based on random forest approach, International Conference on Computing, Communication and Automation (ICCCA) May 2017, IEEE.*

[13] *Georg Ruß, Rudolf Kruse, Martin Schneider, and Peter Wagner. Estimation of neural network parameters for wheat yield prediction. In Max Bramer, editor, Artificial Intelligence in Theory and Practice II, volume 276 of IFIP International Federation for Information Processing,. Springer, July (2008).*

[14] *Niketa Gandhi et al , 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) ,(2016).*

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

[16] *RanadheerDonthi, S. Vijay Prasad, Estimation methods of nonlinear regression models, 2019, AIP Conference Proceedings, (2019).*