iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** IV **Month of publication:** April 2021

DOI: https://doi.org/10.22214/ijraset.2021.33855

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IV Apr 2021- Available at www.ijraset.com

Smart Crop Prediction System using Machine Learning

Prof. S. P. Gade[1], Pratik V. Sul[2], Nilesh B. Thite[3], Tejpal B. Sutar[4], Abhishek A. Phutane[5], Shubham D. Khatavkar[6]
[1]*Assistant Professor, Department of Computer Engineering, PDEA's College of Engineering, Pune, India.*
[2, 3, 4, 5, 6]*Students, Department of Computer Engineering, PDEA's College of Engineering, Pune, India.*

Abstract: ***Agriculture is the backbone of our country. Agriculture is the primary source of livelihood for about 58% of India's population. It plays an important role in Indian economy as it contributes about 17% to the total GDP and provides employment to over 60% of the population. The common problem which is faced among the Indian farmer is that they don't choose the proper crop based on their soil necessity. Because of this the production of crop is affected.***
The downside of nearly all farmers is solved by adapting latest technologies like machine learning, in this predicting the crops based on the kind of soil, climate condition, temperature, humidity, moisture, pH scale, price of soil, rainfall etc. Numerous machine learning techniques like prediction, classification, regression, and clump are used to forecast crop yield. Artificial neural network, support vector machines, linear and logistic regression, decision tree, Naïve Bayes are some of the algorithms implemented for the prediction. The idea of this paper is to implement the crop choice such that this methodology helps in working out many agriculture and farmers issues. This improves our Indian economy by maximizing the yield rate of crop production.
Keywords: Machine learning, crop selection, RandomForest.

I. INTRODUCTION

Agriculture is one of the greatest riches we have. However, in particular in developing countries, techniques used require a lot of physical effort. In recent years, thanks to information technology, agriculture has been enhanced. This leads to Precision Agriculture, which is a farming management concept based on observing, measuring and responding to inter and intra-field variability in crops. Here we use Machine Learning (ML), a branch of Artificial Intelligence (AI, to estimate an agricultural yield production. New kinds of hybrid varieties are produced day by day. However, these varieties do not provide the essential contents as naturally produced crop. These unnatural techniques spoil the soil. It all leads to further environmental harm. Most of these Unnatural techniques are used to avoid losses. But when the producers of these crops know the accurate information on the crop yield it minimizes the loss. To achieve this project is made. Using past information on weather, temperature and several other factors the information is given.

II. LITERATURE REVIEW

In [1] M. Kalimuthu, P. Vaishnavi, M. Kishore, have concluded that the seed is predicted as an output for the given input parameter. source of sunlight and crop health are monitored at regular intervals and it is also taken into the account for achieving a better crop yield. The algorithm used is Naïve Bayes Gaussian classifier. The disadvantage is accuracy is less.
In[2]Aruvansh Nigam, Saksham Garg, Archit Agrawal, Parul Agrawal, have concluded that machine learning algorithms for predicting the yield of the crop on the basis of temperature, rainfall, season and area. Experiments were conducted on Indian government dataset and it has been established that Random Forest Regressor gives the highest yield prediction accuracy. Sequential model that is Simple Recurrent Neural Network performs better on rainfall prediction while LSTM is good for temperature prediction. By combining rainfall, temperature along with other parameters like season area. The algorithms used are Random Forest Regressor, Simple Recurrent Neural Network, long short-term memory(LSTM). In [3] Potnuru Sai Nishant, Pinapa Sai Venkat, Bollu Lakshmi Avinash, B. Jabber, have concluded that the performance metric used is Root mean square error. When the models applied individually, for ENet it was around 4%, Lasso had an error about 2%, Kernel Ridge was about 1% and finally after stacking it was less than 1%. The algorithm used is stacked regression.
In [4] Shivani S. Kale, Preeti S. Patil, have concluded that the proposed model with backpropagation is trying to reduce MSE by using RELU activation function and gradient decedent. The Learning rate for each layer is kept constant i.e. 0.001. As they increase number of epoch's error will get reduced. This result is used as input for deciding success rate of the crop over another crop. The best crop will be suggested to the farmer depending on the district and weather. The algorithm used is neural network. The disadvantage is less accuracy for crop prediction.

In [5] Nidhi H Kulkarni, Dr. G N Srinivasan, Dr. B M Sagar, Dr.N K Cauvery, have concluded that the soil dataset is 1st preprocessed and so the ensembling technique performs an essential function within the classification of the four crops (Rice, Cotton, Sugarcane, Wheat). The individual base learners employed in the ensemble model are Random Forest, Naïve Bayes, and Linear SVM. Majority VotingTechnique has been used as the combinationtechnique to produce the best accuracy.
In [7] Nishit Jain, Amit Kumar, Sahil Garud,Vishal Pradhan, Prajakta Kulkarni, have concluded that this paper helps in predicting crop sequences and maximizing yield rates and creating advantages to the farmers. Also, using machine learning applicationswith agriculture in predicting crop diseases, studying crop simulations, differentirrigation patterns. The algorithms used are artificial neural networks, Support Vector Machine. The disadvantage and Exact accuracy isn't specified.

III. METHODOLOGY

Crop yield is a very useful information for farmers. It is very beneficial to know theyield which results in reduction in loss. In the past the yield prediction is done by experienced farmers. The proposed system also works in a similar way. It takes theprevious information and uses it to predict the future yield. The crop yield mainly depends on weather and pesticides. This prediction is proportional to the accuracy on information provided. Therefore, the proposed system predicts the yield anddecreases the loss. The anticipated system acts as experienced farmer. But, with more accuracy and considers many other factors. Factors like soil condition, weather prediction, yield. The more increase inaccuracy results in more profit in crop yield.To increase accuracy the data has to be perfect. With all the information providedthe proposed system process all the data using data mining methods and predicts the harvest yield. With this forecast the farmer will be able to know his requirements.

IV. RANDOM FOREST CLASSIFIER

At training situation multitude decision treesare made and the output will be divided based on number of classes i.e., classification, prediction of class i.e.,regression. The number of trees is proportional to accuracy in prediction. The dataset includes factors like rainfall, ph, temperature, humidity and crop name. These factors in dataset is used for training. Only two third of the dataset is considered. Remaining dataset is used for experimental basis.

A. Datasets

The dataset consists of factors like temperature, rainfall, humidity, ph.The datasets have been obtainedfrom the github website. The data sethas 3100 instance or data that have taken from the past historic data. It includes 5 parameters or features likethe temperature, ph., humidity, rainfall and crop name.

1) *Random Forest Algorithm:* Random Forest is a ML algorithm.At training situation multitudedecision trees are made and theoutput will be divided based onnumber of classes i.e., classification,prediction of class i.e., regression.The number of trees is proportionalto accuracy in prediction. The datasetincludes factors like rainfall, ph,temperature and humidity. These factors in dataset is used for training.Only two-third of the dataset isconsidered. Remaining dataset isused for experimental basis.

The algorithm random forest has 3parameters like:

a) n tree which describes the n number of trees which need to grow,
b) m try - mentions how many variablesneed to be taken at a node split.
c) Node size - In terminal nodes it suggestsus the number of observations need to take.

B. Decision Tree

Decision tree classifiers utilizes greedy methodology henceforth afeature picks from the start move cannot be consumed any longer, that gives us best grouping whenever utilized in further advances.Likewise, it over fit the preparation information which can give poor outcomes for inconspicuous information. In this way, to beat this confinement gathering model isutilized. In gathering model outcomes from various models areconsolidated.

C. Procedure

Dataset consists of few important attributes such as temperature, rainfall, humidity, ph., applied random forest algorithm for classification and regression tasks. Tried to train the model with decision trees but found that random forest algorithm reduced the overfitting problem and also improved accuracy and used SVR,Random forest and random forest gotmore accuracy. The dataset which used is imported from github repository.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IV Apr 2021- Available at www.ijraset.com

From the dataset, used 80% of data for training the model and 20% of data for testing the results to obtain better results and trained the model by applyingrandom forest algorithm. Then,compared the predicted result withthe original data set. Later, estimated the accuracy of the model using test samples. Likewise, predicted the accuracy of the model with different algorithms. Then finally concludedthat random forest algorithm gives usmore accuracy. Hence, used random forest algorithm to train the model.

V. RESULTS

Data visualization is the representation ofunderstanding data by showing it in a graphical context, so that the designs, inclinations and connexions can be detected and exposed.

The few prevalent plotting collections:

1) *Matplotlib:* Small level, provides lotsof liberty to user.
2) *Pandas Visualization:* comfortable touse this boundary. It can be constructed on Matplotlib
3) *Plotly:* Can create interactive plot for visualization.

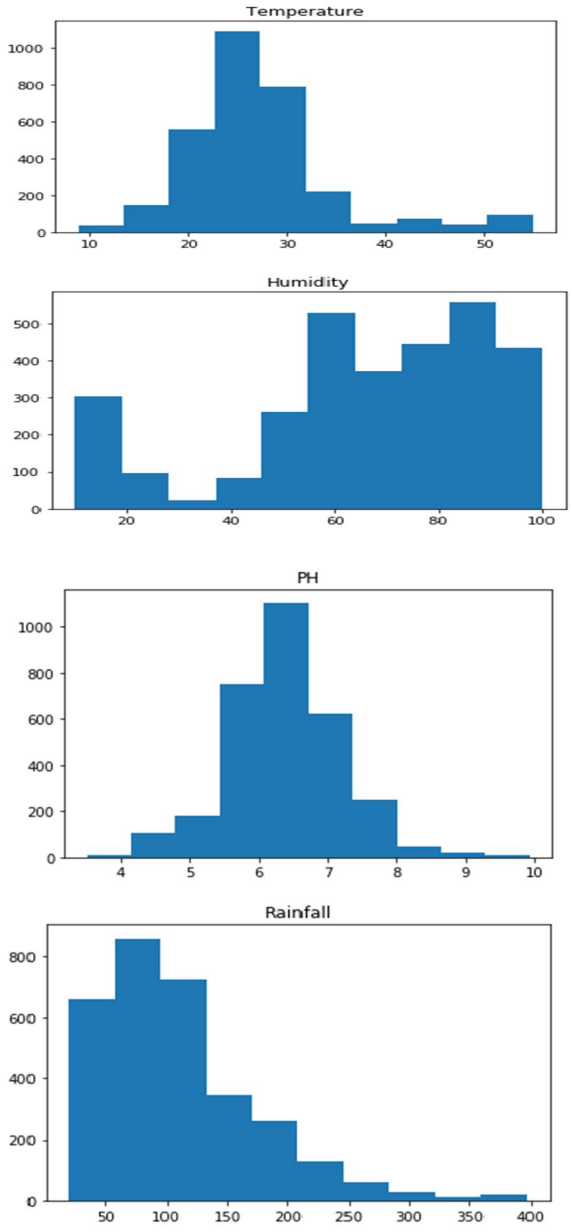


Fig. Hist Diagram

A. Correlation between the attributes helps us to figure out how strong or how weakly they are related to each other. The numeric values 1 with green colour represents the positive relationship exists between the variables. The numeric zero or negative values with darker red colour gets the more negativerelationship exists between the variables.

B. A correlation plot matrix can be formed for a collection of variables with each other variables will be plotted against each other. Here havefour columns where normallydistributed with random values and column names are: temperature, humidity, PH and rainfall.

VI. CONCLUSIONS

```
corrmat = df.corr()
top_corr_features = corrmat.index
plt.figure(figsize=(20,20))
g=sns.heatmap(df[top_corr_features].corr(),annot=True,cmap='RdYlGn')
```

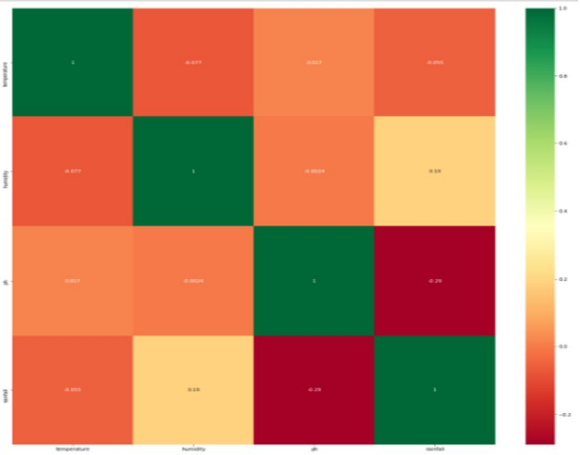


Fig. Correlation Diagram

```
import pandas
from pandas.plotting import scatter_matrix
datacorr=df.corr()
pandas.plotting.scatter_matrix(datacorr,figsize=(20,20))
plt.show()
```

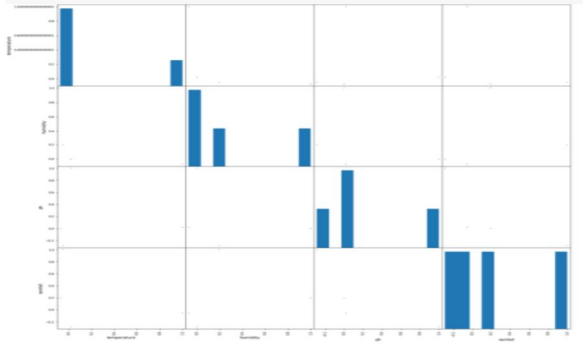
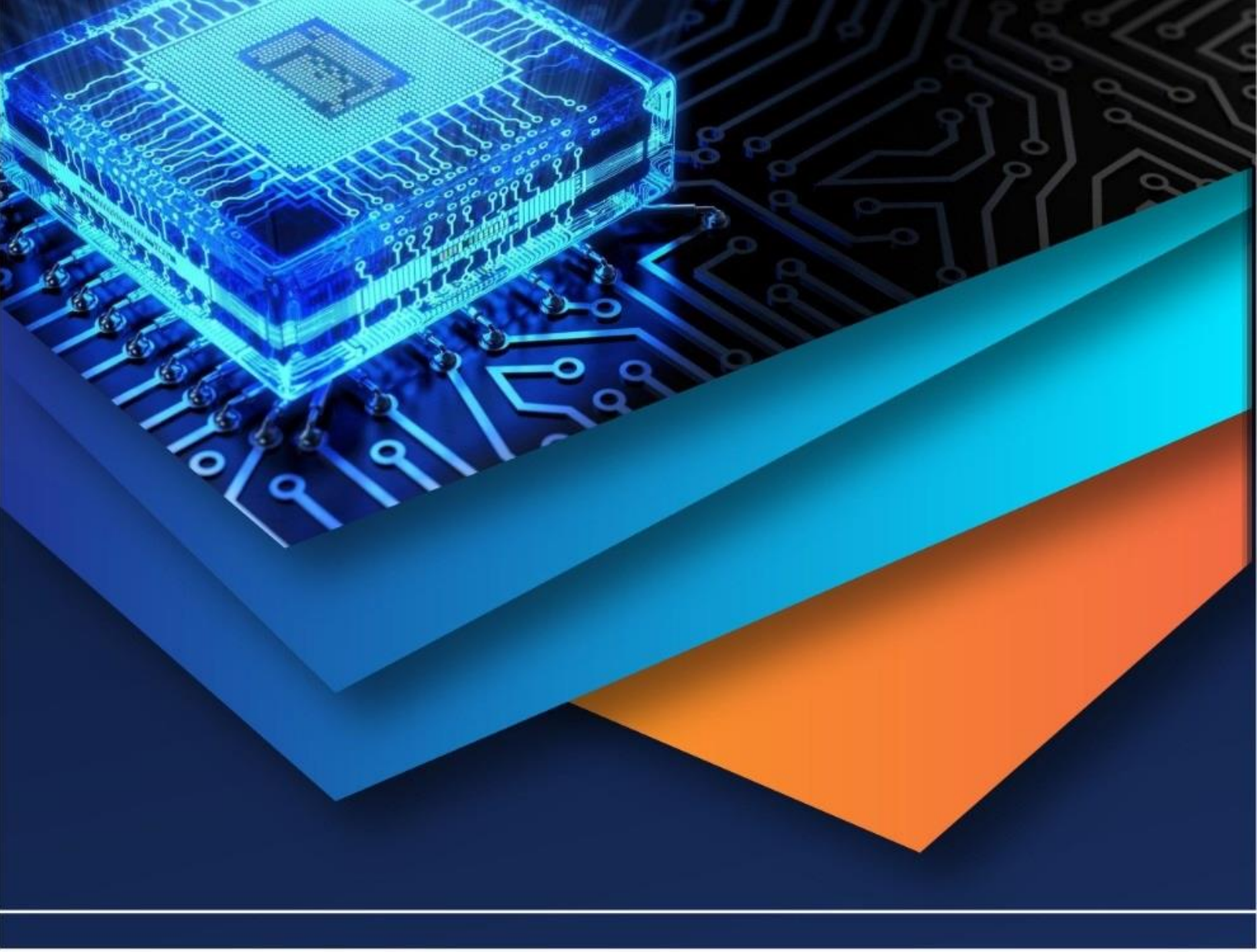


Fig. Correlation Matrix plot diagram

Implement a system to predict cropproduction from the collection of past data. Using machine learning techniques crop yield is predicted. Here, using Random Forest algorithm for predicting the best crop yield as output. In agriculture field, the crop yield prediction is mostly appropriate. The more increase in accuracy results in more profit to the crop yield. The proposedtechnique helps farmers to acquireapprehension in the requirement and price ofdifferent crops. It helps farmers in decision making of which crop to cultivate in thefield. The more increase in accuracy results in more profit to the crop yield. This work isemployed to search out the gain knowledge about the crop that can be deployed to make an efficient and useful harvesting. Under this system, maximum types of crops will becovered. The accurate prediction of different specified crops across different districts will help farmers of India.

REFERENCES

[1] M.Kalimuthu, P.Vaishnavi, M.Kishore, "Crop Prediction using Machine Learning", Third International Conference on Smart Systems and Inventive Technology | 2019.
[2] Aruvansh Nigam, Saksham Garg, ArchitAgrawal, Parul Agrawal, "Crop Yield Prediction Using Machine LearningAlgorithms", Fifth International Conference on Image InformationProcessing (ICIIP)| 2018.
[3] Potnuru Sai Nishant, Pinapa Sai Venkat, Bollu Lakshmi Avinash, B. Jabber, "Crop Yield Prediction based on Indian Agriculture using Machine Learning" International Conference for Emerging Technology (INCET)| 2020.
[4] Shivani S. Kale, Preeti S. Patil, "A Machine Learning Approach to Predict Crop Yield and Success Rate" IEEE Pune Section International Conference (PuneCon) | 2019.
[5] Nidhi H Kulkarni, Dr. G N Srinivasan, Dr. B M Sagar, Dr.N K Cauvery, "Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique" 3 rd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions | 2018
[6] Kodimalar Palanivel, Chellammal Surianarayanan, "AN APPROACH FOR PREDICTION OF CROP YIELD USING MACHINE LEARNING AND BIG DATA TECHNIQUES"International Journal of ComputerEngineering and Technology (IJCET)
Volume 10, Issue 03 | May-June 2019
[7] Nishit Jain, Amit Kumar, Sahil Garud, Vishal Pradhan, Prajakta Kulkarni, "Crop Selection Method Based on Various Environmental Factors Using Machine Learning" International Research Journal of Engineering and Technology (IRJET) Volume: 04 Issue: 02 | Feb - 2017.
[8] Dr. Y. Jeevan Nagendra Kumar, V. Spandana, V.S. Vaishnavi, K. Neha,
V.G.R.R. Devi, "Supervised Machinelearning Approach for Crop YieldPrediction in Agriculture Sector" e Fifth International Conference onCommunication and Electronics Systems (ICCES) | 2020.
[9] Monali Paul, Santosh K. Vishwakarma, Ashok Verma, "Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach"International Conference onComputational Intelligence andCommunication Networks | 2015. 39
[10] Judicael Geraud N. ZANNOU, Vinasetan Ratheil HOUNDJI, "Sorghum Yield Prediction using Machine Learning"|2019.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY