IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 5, May 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005317 |

Crop Yield Prediction Using Machine Learning

Ms. M. Rajya Lakshmi[1], Kurra Dedeepya[2], Marapaka Madhurilatha[3], Nelakurthi Sri Lakshmi[4]

Associate Professor, Department of IT, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur, Andhra Pradesh, India [1]

UG Student, Department of IT, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur, Andhra Pradesh, India[2]

UG Student, Department of IT, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur, Andhra Pradesh, India[3]

UG Student, Department of IT, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur, Andhra Pradesh, India[4]

ABSTRACT: Agriculture is the best utility region especially inside the developing worldwide areas like India. Usage of records age in agriculture can substitute the circumstance of decision making and Farmers can yieldin higher manner. About portion of the number of inhabitants in India relies upon on farming for its occupation however its commitment towards the GDP of India is just 14 percent. One suitable explanation behind this is the deficiency of adequate decisionmaking by farmers on yield prediction. There isn't any framework in location to suggest farmer what plants to grow. The proposed machine learning approach aims at predicting the best yielded crop for a particular region by knowing various atmospheric factors like rainfall, temperature, humidity etc., and land factors like soil pH, soil type including past records of crops grown. Finally, our system is expected to predict the best yield based on dataset we have collected.

KEYWORDS: Key word 1: Crop yield prediction, Key word 2: Demand-based crops, Key word 3: Machine learning techniques, Key word 4:Random Forest, Key word 5: Polynomial regression, Key word 6: Supervised Learning

I. INTRODUCTION

Agriculture is the backbone of every economy. In a country like India, which has ever increased demand offood due to rising population, advances in agriculture sector are required to meet the needs. From ancientperiod, agriculture is considered as the main and theforemost culture practiced in India. Ancient people cultivate the crops in their own land and so they have been accommodated to their needs. Therefore, the natural crops are cultivated and have been used by many creatures such as human beings, animals and birds. The greenish goods produced in the land which have been taken by the creature leads to a healthy and welfare life. Since the invention of new innovative technologies and techniques the agriculture field is slowly degrading. Due to these, abundant invention people are been concentrated on cultivating artificial products that is hybrid products where there leads to an unhealthy life. Nowadays, modern people don't have awareness about the cultivation of the crops in a right time and at a right place. Because of these cultivating techniques the seasonal climatic conditions are also being changed against the fundamental assets like soil, water and air which lead to in security of food. By analyzing all these issues and problems like weather, temperature andseveral factors, there is no proper solution and technologies to overcome the situation faced by us. In India there are severalways to increase the economic growth in the field of agriculture. There are multiple ways to increase andimprove the crop yield and the quality of the crops. Data mining also useful for predicting the crop yield production. Generally, data mining is the process of analyzing data from different perspectives and summarizing it into useful information.

Data mining software is an analytical tool that allows users to analyze data from many different dimensions or angles, categorize, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. The patterns, associations, or relationships among all this data can provide information. Information can be converted into knowledge about historical patterns and future trends.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005317 |

For example, summary information about crop production can help the farmers identify the crop losses and prevent it in future. Crop yield prediction is an important agricultural problem. Each and Every
farmer is always tries to know; how much yield will get from his expectation. In the past, yield prediction was calculated by analyzing farmer's previous experience on a particular crop. The Agricultural yield is primarily depending on weather conditions, pests and planning of harvest operation. Accurate information about history of crop yield is an important thing for making decisions related to agricultural risk management. Therefore, this paper proposes an idea to predict the yield of the crop. The farmer will check the yield of the crop as per the acre, before cultivating onto the field.

II. MOTIVATION

This project focuses on predicting the yield of the crop by applying various machine learning techniques. The outcome of these techniques is compared on the basis of mean absolute error. The prediction made by machine learning algorithms will help the farmers to decide which crop to grow to get the maximum yield by considering factors like temperature, rainfall, area, etc.

III. LITERATURE SURVEY

G Rasul, Q. Z. Chaudhry, A. Mahmood, K. W. Hider, "Effect of. 28–40Temperature Rise on Crop Growth & Productivity", *Pakistan Journal of Meteorology*, vol. 8, no. 15, pp. 7-8, 2011.

This paper discusses an analysis of ERS SAR imagery of agricultural crops in Flevoland, The Netherlands, over a four-fear period (1993 to 1996) to study the stability of multitemporal radar signatures from one year to the next. Direct comparisons of the multitemporal profiles of crop signatures are made to derive limits on their stability and to examine the differences between them from one year to the next. Sharp rises (of several dB) in temporal crop signatures are linked to variations in rainfall, freezing, and incident angle (due to imaging passes from different orbit tracks). Model simulations confirm the plausibility of these mechanisms and emphasize their importance for quantitative monitoring of agricultural crop development. The possibility of timing critical phases of the crop growth cycle is highlighted using field-to-field variations with particular regard to the emergence and closure of sugar beet. The interfere comparison also enables generalized comments to be made regarding the performance and stability of crop classification algorithms from one year to another. Only summer months are consistently identified as helping to distinguish broad-leaved crops from cereals. There is some evidence that other times of the year assist in distinguishing specific crops, but this evidence is not stable from one year to another.

AnupamaMasato, "Climate Change and its Impact on Agriculture", *International Journal of Scientific and Research Publications*, vol. 4, no. 4, pp. 4-5, April 2014, ISSN 2250-3153.

Due to the complex climate types and various meteorological disasters in Gansu province, with the background of climate change dominated by warming, the losses caused by meteorological disasters are increasing. The understanding of disaster risk characteristics and its response to climate change need to be solved urgently. In view of this, this study based on the data of the disaster situation of major meteorological stations in Gansu province since 1961, constructed the disaster risk index-comprehensive loss rate, and systematically analysed the space-time variation characteristics of drought, rainstorm and flood, wind and hail and low temperature disasters in Gansu province, and discussed the impact of climate change on them. The results show that since 1961, the risk of agricultural drought disaster is the greatest, followed by wind hail and rainstorm flood, and the risk of cold disaster is relatively minimal. Among them, drought disaster, disaster and loss rate (25.2%, 14.1% and 2.2 %) were significantly higher than the national average (15.0%, 8.1 % and 1.7%), and the increase rate (0.16%/10a, 0.15%/10a and 0.05%/10a) was also higher than the national average. The comprehensive loss rates of storm hails, rainstorm and flood disasters and cold and cold disasters also showed an increasing trend, with the increase rates of 0.29%/10a, 0.45%/10 and 0.72%/10a, respectively. The characteristics of inter-decadal disaster risk change have obvious differences, and the drought disaster risk increases uniformly. After climate change, the risk of major meteorological disasters in Gansu increased.

Janet Kaur, "Impact of Climate Change on Agricultural Productivity and Food Security Resulting in Poverty in India", *University Ca' FoscariVenezia*, vol. 23, pp. 16-18, 2017.

Ensuring food productivity is an important issue for the developing countries like India, where more than one third of the people is live in poverty. Season crop production estimations are vastly identified as an important input for investigating food balance sheets and production shortfalls. Crop production estimation and assessment is done worldwide on a regional basis to enable high production and cost reduction in crop yield estimation. The aim of this

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005317 |

method is to the productivity of the crop, using various factors like soil type, season type, water availability and risk factor. In this scenario, Parallel Layer Regression (PLR) along with Deep Belief Network (DBN) strategy is proposed to perform crop productivity estimation. Here, DBN strategy is generated for top five growing crops in Karnataka namely, rice, ragi, and pulses. The proposed methodology forecasts each area in the applicable database into one of the five crops. Finally, the experimental results show that the method has strong potential for accurate crop productivity prediction in terms of accuracy (ACC), sensitivity (SEN) and specificity (SPE) and also this method performance has verified in real time data and people interactions.

Pratap S. Barthel, Md. Tajuddin Khan, Digvijay S. Negi, Shailyn Agarwal, "Impact of Climate Change on Yields of Major Food Crops in India: Implications for Food Security", *Agricultural Economics Research Review*, vol. 27, no. 2, pp. 145-155, July-December 2014.

Crop productivity is a major concern all over the world to provide food security, resulting in the green revolution. It is noteworthy that the fertilizer implemented to farmland leads to more desirable cropping patterns. Utilization of agricultural land efficiently for the crop production requires the knowledge of the nutrient inconsistency. This paper has presented the power of geomatics, to retrieve the synoptic and substantial changes in cropping pattern. Results and interpretations lead to the evaluation of the contemporaneous cropping systems. After a major yield parameter scrutiny for crops (rice, wheat, sugarcane, and onion), the magnificent accelerations were suggested. Results demonstrated a correlation r2 value of 0.834 with the estimated crop yield and normalized difference vegetation index. The Rice Equivalent Yield (REY) is highest at the range of 17-21 t/ha in the North, central and southern lower part, lowest at the western part ranging from 7-12 t/ha, with some part with 12-14 t/ha, while the most of the eastern part of the study site has shown the REY values ranging from 14 to 17 t/ha. The surveyed information, such as pH, electrical conductivity, and organic carbon of the soil specimen, was used to examine the spatial discrepancies of rice-based cropping system's productivity. Ultimately, the spatial temporal maps of fertilization pattern, yield parameters (e.g., N, F, and K), and relational REY observation were illustrated using spatial interpolation.

B. Dumont, V. Lemans, Salvador Ferrand is, Bernard Bedson, Jean-PerrieDistain, "Assessing the potential of an algorithm based on mean climatic data to predict wheat yield", *Precision Agriculture*, vol. 15, no. 3, pp. 255-272, June 2014.

Bangladesh, a nation renowned for its rich fertile land and a population around 160 million, earns most of its living from agriculture. The nutrient rich lands help us providing year-round crop yields that play a crucial role for the economy of Bangladesh. Thus, this is important to deliberately work on agricultural planning and prediction models to ensure economic prosperity. The advancement of crop yields is significantly dependent on soil factors like Ph, nutrients and organic substances along with climatic factors like rainfall, temperature and humidity. Data of such factors are recorded to serve the purpose of scientific and statistical analysis. With the help of applying different data mining techniques on them, we are able to determine effective parameters to predict crop yield from different locations. This paper mainly focuses on the analysis to predict Bangladesh's four most yielding crops; wheat, jute, T-Aman and mustard. To carry out the whole experiment, we have analysed soil properties of medium high land and high land from different sub districts of Bangladesh and also their respective climatic data and crop production of the last 6 years. For our analysis, we have applied different data mining techniques such as K-means, PAM, CLARA and DBSCAN for clustering and four linear regression methods to predict crop yields.

B Basso, B Bodson, V. Leemans, B. Bodson, J-P Destain, M-F Destain, "A comparison of within season yield predictions algorithm based on crop model behaviour analysis", *Agricultural and Forest Meteorology*, vol. 204, pp. 10-21, May 2015.

The behaviour of crops can be accurately predicted when all the parameters of the crop model are well known, and assimilating data observed on crop status in the model is one way of estimating parameters. Nevertheless, the quality of the estimation depends on the sensitivity of model output variables to the parameters. In this paper, we quantify the link between the global sensitivity analysis (GSA) of the soil parameters of the mechanistic crop model STICS, and the ability to retrieve the true values of these parameters. The Global sensitivity indices were computed by a variance-based method (Extended FAST) and the quality of parameter estimation (RRMSE) was computed with an importance sampling method based on Bayes theory (GLUE). Criteria based on GSA were built to link GSA indices with the quality of parameters estimation. The result shows that the higher the criteria, the better the quality of
parameters estimation and GSA appeared to be useful to interpret and predict the performance of the estimation parameters process.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005317 |

Betty. J, Shem G Juma, Everline. O, "On the Use of Regression Models to Predict Tea Crop Yield Responses to Climate Change: A Case of Nandi East Sub-County of Nandi County Kenya", ***Assesing the Value of Systematic Cycling in a Polluted Urban Environment Climate*****, vol. 5, no. 3, pp. 5, July 2017.**

Data mining is an emerging field of research in Information Technology as well as in agriculture. The present study focuses on the applications of data mining techniques in tea plantations in the face of climatic change to help the farmer in taking decision for farming and achieving the expected economic return. This paper presents an analysis using data mining techniques for estimating the future yield prediction in tea cultivation with climatic change trends observed in last 30 years (1977-2006). The patterns of crop production in response to the climatic (rainfall, temperature, relative humidity, evaporation and sunshine) effect across the four tea growing regions (South Bank, North Bank, Upper Assam and Cachar) of Assam were developed using Multiple Linear Regression (MLR) technique. The tea production estimation equations developed for the regions were validated for the future yield prediction (2007, 2009 and 2010) and were found to be significant. Thus, it is suggested that the planters/farmers could use the technique to predict the future crop productivity and consequently adopt alternative adaptive measures to maximize yield if the predictions fall below expectations .

VI. EXISTING SYSTEM

The agriculture sector needs a huge up gradation in order to survive the changing conditions of Indian economy. Along with the advances in machines and technologies used in farming, useful and accurate information about different matters also plays a significant role in it. This information is being gathered by the use of remote sensors, satellite images, surveys etc. This information along with the knowledge of subject experts and researchers should be readily available to the farmers in order to exploit its potential worth.

VII. PROPOSED MODEL

This scenario mainly concentrates on weather forecasting, crop yield prediction and crop costforecasting. These factors help the farmers to cultivate the best food crops and raise the right animals with accordance to environmental components. Also, the farmers can adapt to climate changes to some degree by shifting planting dates, choosing varieties with different growth duration, or changing crop rotations. For experimental analysis, the statistical numeric data related to agriculture is undertaken. Whereas, the clustering-based techniques and supervised algorithms are utilized for managing the collected statistical data. Additionally, the suitable classification methods like Support Vector Machine (SVM), neural networks are employed for better classification outcome.

VIII. METHODOLOGY

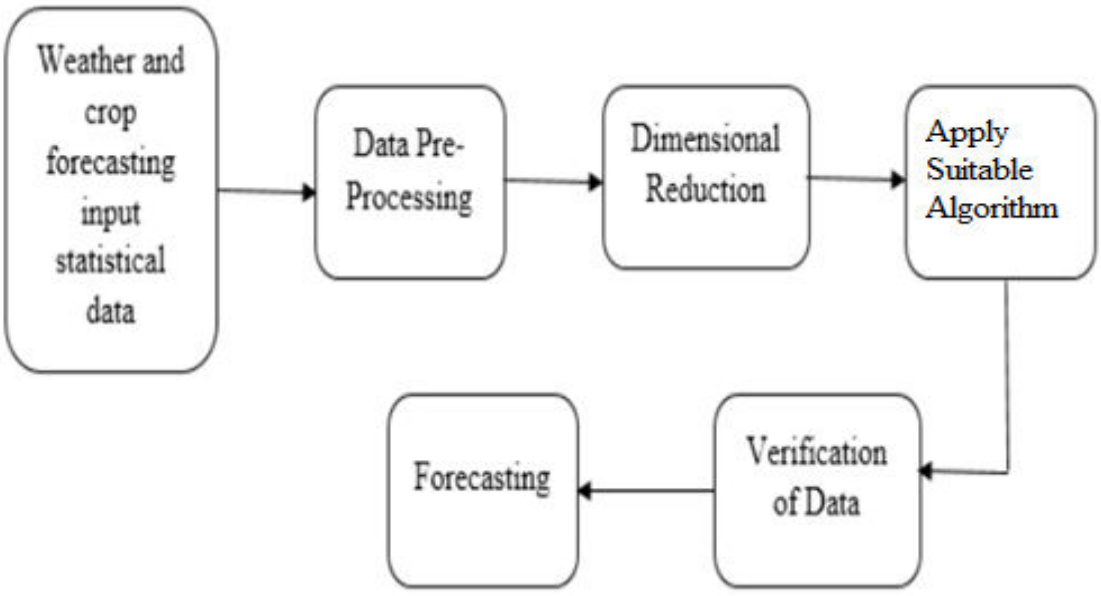


fig-1 System Architecture

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005317 |

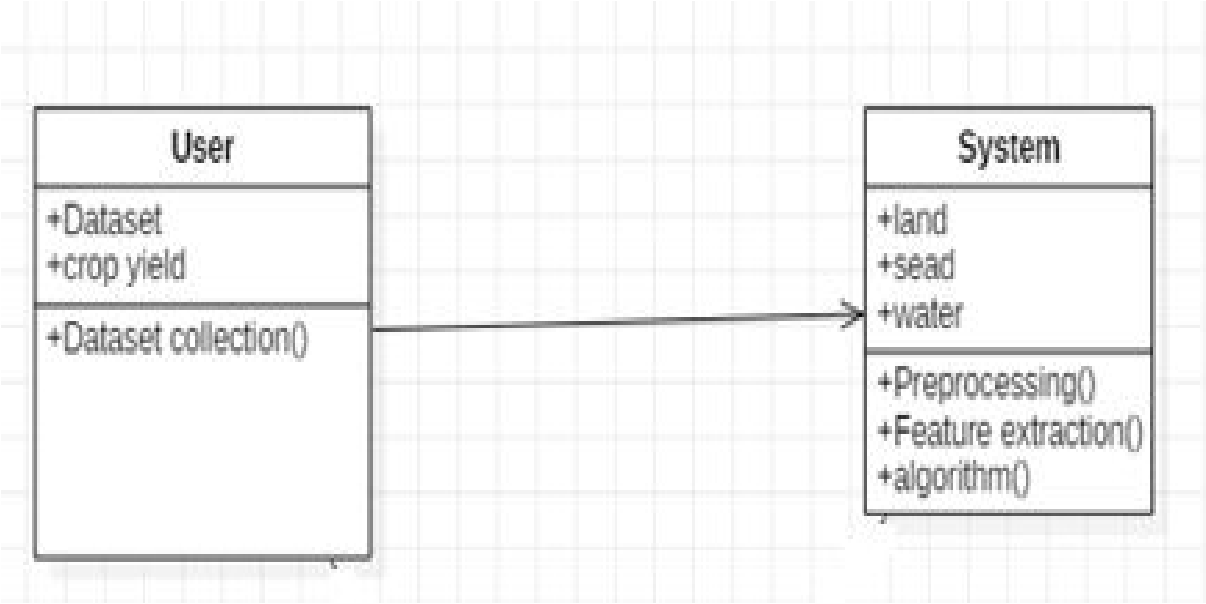


fig-2 Class Diagram

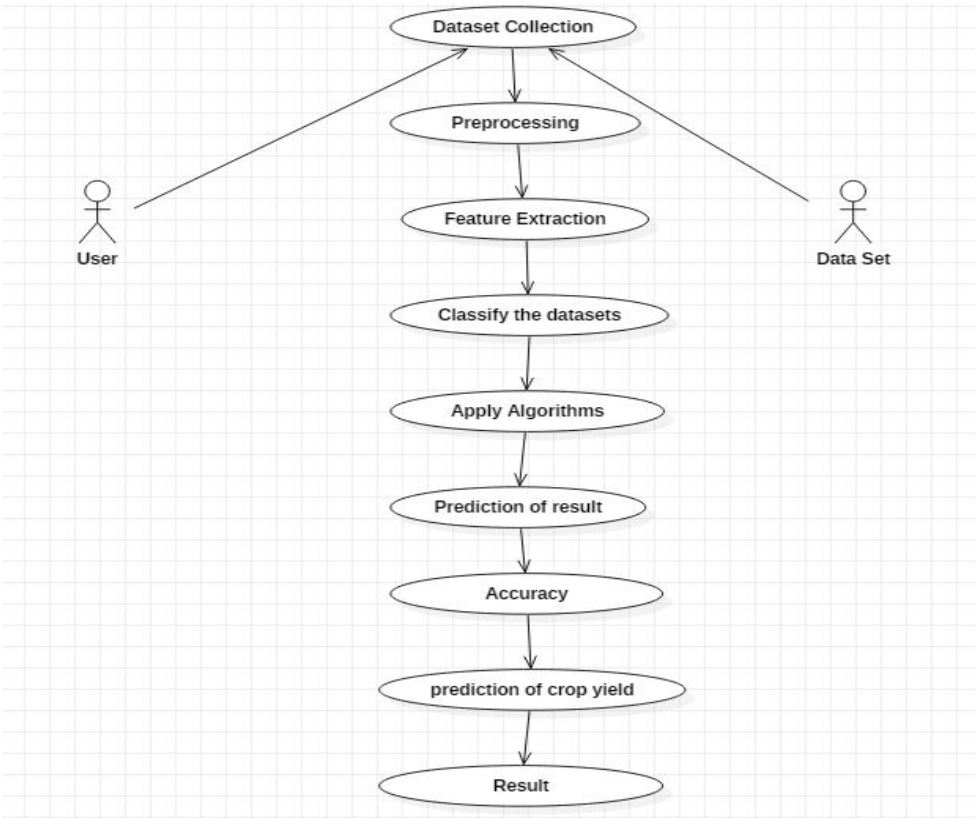


fig-3 Use Case Diagram

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005317 |

IX. RESULTS & DISCUSSION

Test case 1:

SL #	TEST CASE NAME	DESCRIPTION	STEP NO	ACTION TO BE TAKEN (DESIGN STEPS)	EXPECTED (DESIGN STEP)	Test Execution Result (PASS/FAIL)
1	Excel Sheet verification	**Objective:** There should be an excel sheet. Any number of rows can be added to the sheet.	Step 1	Excel sheet should be available	Excel sheet is available	Pass
			Step 2	Excel sheet is created based on the template	The excel sheet should always be based on the template	Pass
			Step 3	Changed the name of excel sheet	Should not make any modification on the name of excel sheet	Fail
			Step 4	Added 10000 or above records	Can add any number of records	Pass

table – 1

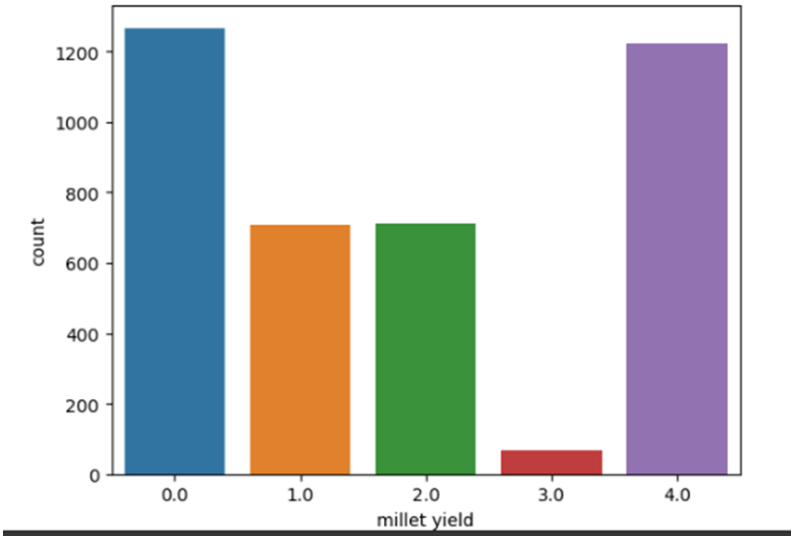
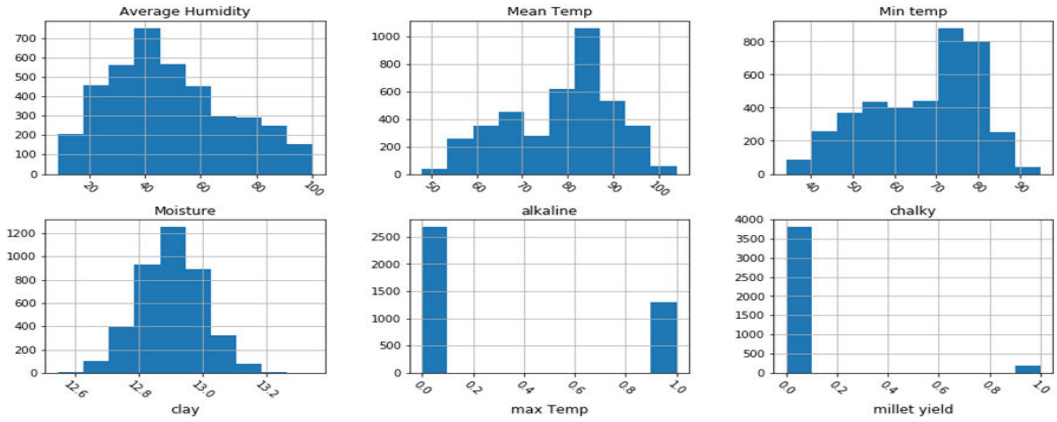


fig-4



| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 5, May 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1005317 |

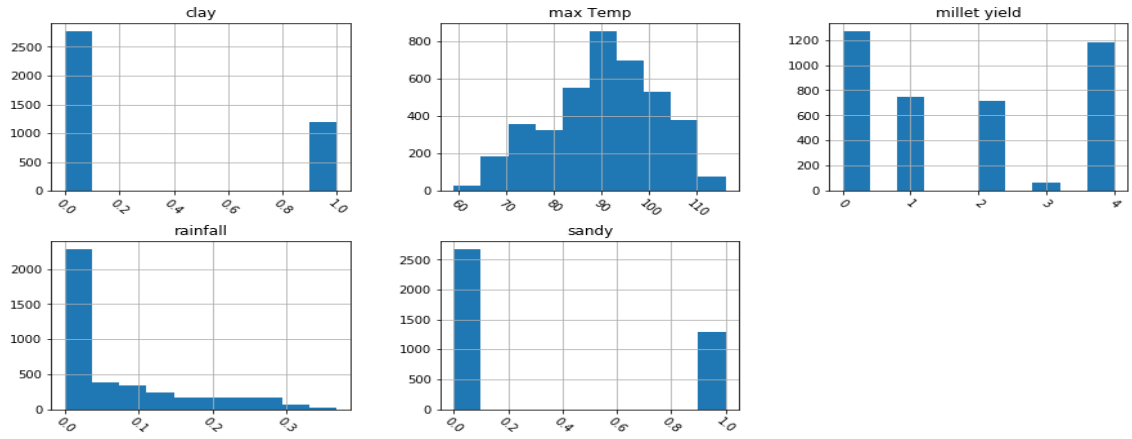


fig-5

Predication Using Algorithm

```
In [123]: from sklearn.tree import DecisionTreeClassifier
          from sklearn.ensemble import RandomForestClassifier

In [124]: rf = RandomForestClassifier(n_estimators=140)
          rf.fit(X_train,y_train)
          print (rf.score(X_test,y_test))

          C:\Users\NS\Anaconda3\lib\site-packages\ipykernel_launcher.py:2: DataConversionWarning: A column-vector y was passed when a 1d
          array was expected. Please change the shape of y to (n_samples,), for example using ravel().

          0.9962311557788944
```

X. CONCLUSION

The Results shows that we can attain an accurate crop yield prediction using the Random Forest algorithm. Random Forest algorithm achieves a largest number of crop yield models with a lowest model. It is suitable for massive crop yield prediction in agricultural planning. This makes the farmers to take the right decision for right crop such that the agricultural sector will be developed by innovative ideas.

REFERENCES

1. D. R. Legates, R. Mahmood, D. F. Levia, T. L. DeLiberty, S. M. Quiring, C. Houser, and F. E. Nelson, “Soil moisture: A central and unifying theme in physical geography,” Progress in Physical Geography, vol. 35, no. 1, pp. 65–86, 2011.
2. Y. H. Kerr, P. Waldteufel, J.-P. Wigneron, J. Martinuzzi, J. Font, and M. Berger, “Soil moisture retrieval from space: The soil moisture and ocean salinity (smos) mission,” IEEE transactions on Geoscience and remote sensing, vol. 39, no. 8, pp. 1729–1735, 2001.
3. E. G. Njoku, T. J. Jackson, V. Lakshmi, T. K. Chan, and S. V. Nghiem, “Soil moisture retrieval from amsr-e,” IEEE transactions on Geoscience and remote sensing, vol. 41, no. 2, pp. 215–229, 2003.
4. R. H. Reichle, R. D. Koster, J. Dong, and A. A. Berg, “Global soil moisture from satellite observations, land surface models, and ground data: Implications for data assimilation,” Journal of Hydrometeorology, vol. 5, no. 3, pp. 430–442, 2004.
5. S. Lambot, E. C. Slob, I. van den Bosch, B. Stockbroeckx, and M. Vanclooster, “Modeling of ground-penetrating radar for accurate characterization of subsurface electric properties,” IEEE Transactions on Geoscience and Remote Sensing, vol. 42, no. 11, pp. 2555–2568, 2004.
6. M. S. Dawson, A. K. Fung, and M. T. Manry, “A robust statisticalbased estimator for soil moisture retrieval from radar measurements,” IEEE transactions on geoscience and remote sensing, vol. 35, no. 1, pp. 57–67, 1997. [7] G. Satalino, F. Mattia, M. W. Davidson, T. Le Toan, G. Pasquariello, and M.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details