Proc. of Int. Conf. on Intelligence Computing and Information Technology , ICIT

Online Support System for Smart Agriculture

Manasa M R[1] and Dr.Yuvaraju B.N[2]
[1-2]The National Institute of Engineering, Department of Information Technology (Department of Computer Science), Mysuru, Karnataka, India
Email: manasa.31198ravi@gmail.com, yuvarajubn@nie.ac.in

***Abstract*—As we all know, India is the world's second most populous country, and most of its people rely on agriculture and work in agriculture. Agriculture is critical to the Indian economic system; it employs more than half of the population. Food is perhaps the most important aspect of life; without it, humanity cannot survive. Our current agronomical tasks include ways to advance crop production, maintain appropriate agriculture parameters, and keep appropriate chemical fertilizers on hand. Our main goal is to reduce the number of agriculturalist suicides, increase farmer profits, and so on. Machine learning applications are used in crop yield predictions and trying to improve crop yield prediction to overcome such a wide range of issues. This paper discusses how the system uses data gathered from agriculture departments. Our systems are powered by real-time data and farmer-friendly agricultural production software. Our project assists farmers in determining which crops to be grown to maximize profits. In this paper, we attempt to forecast crop yields using data science algorithms.**

***Index Terms*— Machine learning, Agriculture, Crop prediction, Yield prediction, Data Science, Agriculture.**

I. INTRODUCTION

In our country, Agriculture began in 9000 BCE in northwest India. Agriculture is the cultivation of land for effective crop growth and livestock grazing. Agriculture is considered a fundamental supply for developing crops and the backbones of the county's economic development. In rural regions, agriculture's contribution toward the supply of livelihood is more than 70%. The main contribution in the early days has been the cultivation of plant life and the domestication of vegetation and animals. Beyond decade earlier than Industrial Revolution, agriculture became practiced within the form of traditional open area farming and changed into growing plants in a massive set of fields. Before the industrial revolution agriculture employees worked six days a week from sun rise to sun set simply looking after developing crops. In positive seasons, there has been a large call for the plants, particularly for ploughing and harvesting. After Industrial Revolution, there was an improvement in farming and the development of vegetation. Major traits had been seen below enclosure of the usage of open fields turned into changed into small fields and adopting new farming techniques together with a seed drill, plow, breeding technique. Smart farming was introduced in the twenty-first century as a result of technological advancement. One of the biggest issues that farmers currently face in the agriculture industry is a lack of profit levels. Farmers grow crops but do not receive the proper yield, resulting in a lower profit level. Yield prediction is important in agriculture because yield is affected by many factors such as temperature, rainfall, soil characteristics, and so on. Earlier start yields forecasting based on these constraints is crucial. As a

Grenze ID: 02.ICIT.2022.6. 506

result, "Yield Prediction Software" is created. The current system Predicting agricultural yield is critical for dealing with crop storage issues, transportation concerns, and risk management issues. collecting data and keeping track of documents in the form of ledgers. The present system is based on a manual comparison of previous and present results. Based on previous experiences and findings, we estimate crop production. We collected raw data from the Mysore agriculture department for this study, which included agricultural parameters such as humidity, rainfall, temperature, soil, and ph values. Each research paper investigated and implemented various crops grown throughout India. In our study, we proposed two distinct methodologies: one for crop yield prediction using Machine Learning for two major crops, paddy and ragi, and another method, our system designed as an e-learning system to assist farmers, and an online form is used to engage in farming and also if any agronomists have any inquiries, they can directly contact the system administrator. Machine learning, according to previous research, is an effective classifier for crop yield prediction. Machine learning can be used aggressively to assist farmers in producing the highest possible yield. As a result, all of these supervised classification algorithms perform well with both small and large data sets, produce accurate results, and employ efficient classifier techniques. ID3, decision trees, KNN, Bayesian classification, random forest, and other techniques are examples of these techniques.

II. Related Work

[1] V. Spandana Supervised Machine Learning Approach for crop yield prediction in the agriculture sector. In this paper, the author has collected data from the Kaggle website. The data set includes 3101 instances that have beyond historic statistics and it includes five parameters inclusive of temperature, PH, humidity, rainfall, and crop name. Random forest and Data mining strategies are applied for classification and regression. For training, the model 80% is used as training and 20% is used for testing. Data visualization is used for displaying the information in a graphical layout. Matplot lib, pandas, and plot are used for visualizations. Hist Diagram is used to show the humidity and rainfall are exponential distribution, ph is Gaussian distribution attributes, and Temp is a skewed distribution. The density plot shows the distribution in numerical values. Box plot suggests exceptional for each attribute consisting of humidity indicates its miles accordingly skewed with a big set of values, the temperature is miles skewed with small values and pH is neither skewed as a small nor massive value of skewed. Correlation diagrams are used to symbolize the relationship of the attributes. Numeric value 1 positive relationship between variables. Numeric 0 represents negative variables. The correlation Plot Matrix for every characteristic of humidity and rainfall is an exponential distribution, temperature, its skewed distribution, and pH, it's the Gaussian distribution. Random forest is used for selecting a fine crop. This model is a comparison and brought with the decision tree while this model is compared and trained with a decision tree random forest has got maximum accuracy because there because of pleasant fits for decreasing overfitting.

[2] Rohit. N Crop yield prediction and efficient use of fertilizers in this work, their main goal is to produce the highest crop yield with the lowest yield. In this study, they developed a model with a set of crop information predictions based on different attributes from seven factors, including state, district, crop, year, season, and area. Another model they proposed for predicting yields for each crop is based on soil fertilizers such as nitrogen and potassium. They proposed employing two different algorithms, such as random Forest and backpropagation Algorithm, to predict effective fertilizers consumption for each crop using Machine Learning. This Random Forest Algorithm is beneficial for Indian agriculture. The backpropagation Algorithm is beneficial for predicting fertilizers amounts for each crop, such as potassium, nitrogen, and phosphorus. The error rate is predicted by comparing the outcomes of these two algorithms. When compared to the Random Forest Algorithm, the experimental analysis backpropagation algorithm is less.

[3] S Kanaga Suba Raja Demand-Based Crop Recommender System for Farmer in this paper, this method is meant to assist farmers in recommending vegetation to develop primarily based on the previous 12 months' data and marketplace pricing worth. We try and count on crop yield and charge that a farmer can achieve from his assets on this take a look at via analysing patterns in additional facts. We use a sliding window non-linear regression approach to make predictions based totally on numerous components of agricultural production, which include rainfall, temperature, marketplace costs, area of the land, and crop yields that cross beyond the common. This technique is also supposed to assist farmers with speech recognition thru textual content conversion. This generation is supposed to translate English right into a local language. So, similarly, the above system could be applied to promote rural improvement.

[4] Lavanya system to predict the paddy yield depending on agriculture data using Data Mining. In this paper, the author has collected required data from government sectors to model and predict temperature, rainfall, and paddy plant patterns. The primary goal is to discover the relationship between environmental factors and paddy

yield in the Karnataka district of Nanjangud. In this case, we are predicting paddy yield by taking temperature, rainfall, soil moisture, and ph into account, which will be useful for agricultural enterprises. The process of identifying hidden and frequent patterns in large databases is known as data mining. The methodology that we used to create the system Data is initially collected from various governmental agencies, and then noisy data is removed. Using the éclat algorithm on the pre-processed data, we can obtain the paddy prediction pattern.

III. Proposed System

The proposed system assists farmers in cultivating the appropriate crops at the appropriate time and assists farmers in increasing crop yield by suggesting appropriate profits. A significant portion of the Indian population will gain from this application. For the recommendations, supervised learning algorithms such as "Bayesian classifier" or "K nearest neighbors" and "Naive Bayes Algorithm" were used. These algorithms are preferred because they work efficiently, produce faster results, and work with all data formats. Furthermore, a small number of survey papers indicate that these algorithms are efficient and suitable for agricultural datasets.

IV. Methodology

A. Work Flow for Online Support System for Smart Agriculture

The workflow is shown with an arrow. It suggests the route of the workflow within the activity diagram.

User Requirements are a set of specifications that a user must meet to use a product.

1. **Administrator:** The administrator is the person in charge of the application as a whole. There is just one administrator in the system.
2. **In-charger/Staff at the location:** Staff is a character who uses the application's services. A person in charge of a specific location is known as a staff member. The administrator creates the staff.
3. **Visitors or Farmers:** Visitors are the users who can access basic agriculture information.

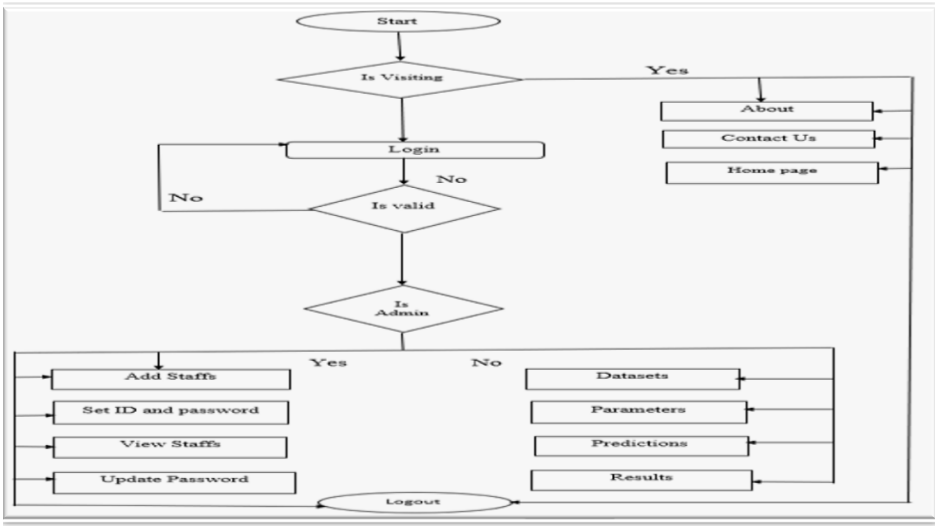


Figure 1: Workflow for Online Support System for Smart Agriculture

The Project Administrator's Modules include the following:

1. Login Module: This module allows an administrator to log into the application using credentials such as the admin id and password.
2. Designate a location in-charger/staff member: An administrator enters the staff/location in-chargers for future use of the system in this module. Staff is mostly developed for a certain place.
3. Provide staff with an id as well as a password: An administrator creates a unique staff Id and password for each employee; with these credentials, employees can log in.
4. Change your password: The system administrator can update the password for future use here.

Location In-charge or Staff has the following modules;

1. Login Module: In this module, staff gets login into the application by specifying the credentials such as admin id and password.

2. Manage Datasets based on Location [upload humidity, region, area, rainfall, temp, rice yield, and soil features data: In this module, the location in-charger manages the existing datasets for the crop yield prediction. Here staff adds the existing datasets [sample soil attributes, humidity, region, area and temp, rainfall, and crop yield details] based on daily, monthly, and yearly.

3. Crop Yield Prediction Module (Core Module - ML): In this module, the system predicts crop yield based on the previous data, here we make use of the *Machine Learning technique "classification rules"* for the prediction based on the constraints humidity, region, and sample soil attributes temp and rainfall.

4. Result Analysis: algorithms results are analyzed to find the accuracy and efficiency.

5. Update Password: Here staff can update the password for future use.

Farmers has the following modules

1. The Login Module: Farmers can access this subsystem by providing credentials such as a user id and password.
2. Review the Basic Information: Farmers can visit the website or portal to view required information such as documents about farming and some videos that have been uploaded, which they can view.
3. Online Discussion Board: If a user has a question about agriculture, they can post it in the forum, and each question will redirect to the admin portal. The answer would be provided by the administrator.
4. Password update: In this, the user to change their password for future use.

B. Methodology

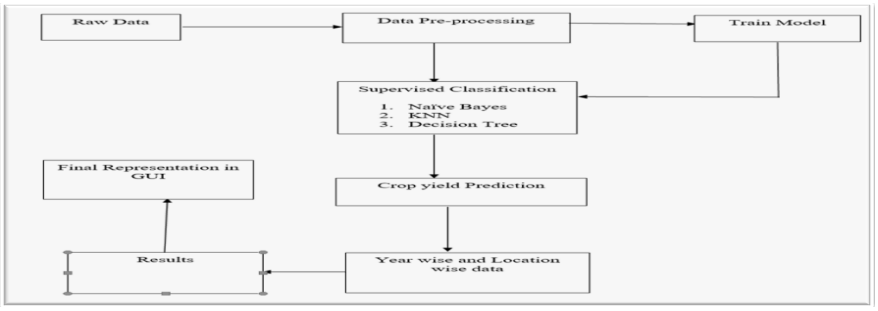


Figure 2 Methodology for Online support system for smart agriculture

1. **Data Collection:** This is the first step in the crop recommendation system, where we collect agricultural statistics. It contains agricultural parameters, crop details, farmers' information, and parameter information. Agriculture facts were gathered for the two most significant crops, paddy, and ragi, from the nearby Karnataka, Mysore district. Rainfall, temperature, and soil functions such as PH, nitrogen, potassium, iron, and many others are examples of agriculture parameters.
2. **Data Pre-processing**: Agriculture data is analyzed in this section, and only relevant information is extracted. The data needed for processing is extracted and segmented according to region. Because the entire set of agricultural data is not necessary for processing and because processing would take too long if we input all the data, data extraction is done.
3. **Training data**: Once the required data has been extracted and segmented, we must train the data, which entails converting the data into the required format, such as numerical values, binary values, strings, and so on. The algorithm type determines conversion.
4. **Machine learning Classification**: ML is concerned with the development and study of systems that can learn from data.

Supervised Learning Techniques

It is a predictive form used for the responsibilities which entail predicting one fee based on other values in the statistics set. Labels for Supervised Learning may be desired. It categorizes an item primarily based on the parameters of one of the predefined label units. We have many algorithms to build a version in supervised learning, such as KNN, Bayesian Network, Decision Tree, ID3, Random Forest, SVM, Regression strategies, and so on. We identify the suitable system of rules for predictions depending on the demands, labels, parameters, and information set. In the presence of uncertainty, the algorithm is used to construct the model that helps make predictions mainly based on evidence.

Classification Techniques

Classification is a technique for locating a model (or feature) that describes and distinguishes information, instructions, or standards. The model is created entirely by analyzing a set of training data (i.e., facts gadgets for

which the elegant labels are regarded). The version is used to predict the class label of objects with unidentified elegance labels

5. ***KNN Algorithm OR Naive Bayes Algorithm.***

For crop recommendation, the "Naive Bayes Algorithm" is used for the following reasons:

- The effective classifier
- works equally well with fewer and more parameters.
- It works for both small and large data sets.
- more precise results

6. **Crop Yield Forecasting**: Using a machine learning algorithm, the system predicts crop yield based on agricultural parameters.
7. **Based on Location and Year:** Crop yield predictions are made both regionally and annually.
8. **Outcomes:** We calculate the algorithm's accuracy by dividing the training datasets into training and testing datasets. 80% were classified as training datasets, while 20% were classified as testing datasets.
9. **Create a Visual Representation**: Crop yield is displayed on the GUI for farmers. When users log in to the application, the system recommends suitable and profitable crops for farmers on a graphical user interface.

C. Algorithms

1. Naïve Bayes Procedure

The Naive Bayes Method is a Machine Learning Algorithm. Initially, we are gathering information from the department of agriculture in Mysore and estimating the probability of each attribute. The pseudo-code below explains the procedures of the Naive Bayes classification algorithm.

Step 1: Scan the dataset (storage server)

Step 2: Calculate the probability of each attribute value. [n, n_c, m, p]

Step 3: Apply the formulae

P(attribute value(ai)/subject value(vj)=(n_c + mp)/(n+m)

Where:

- n = the quantity of training examples for which v = vj
- nc = wide variety of examples for which v = vj and a = ai
- p = a priori estimate for P(aijvj)
- m = the equal sample size

Step 4: We are going to multiply the possibilities by way of p

Step 5: Compare the values and assign the characteristic value to one of the predefined sets.

2. KNN Algorithm

The agriculture data are collected and clustered using the KNN algorithm. Initially, we are gathering information from the department of agriculture in Mysore and we calculating the value of K using agriculture parameters such as humidity, rainfall, temperature, and ph values. Here is the value of k=10 agriculture parameters. We are calculating the predicted yield result using Euclidean Distance Formula:

Euclidean distance between 2 points is square root[(x1-x2)square + (y1-y2)square + (z1 - z2)square]

Then we are calculating k which has got minimum neighbors' collecting all nearest categories of y and predicting the yield. The below pseudo-code describes the steps of the KNN algorithm.

Step 1: This works primarily based on the spatial calculation K (the number of nearest neighbours)

Step 2: Calculate the distance using the Euclidean equation

Step 3: Determine the K minimum distance in the neighbours

Step 4: Collect the nearest category y values

Step 5: Predict the value of the query instance using a simple majority of the nearest neighbour's

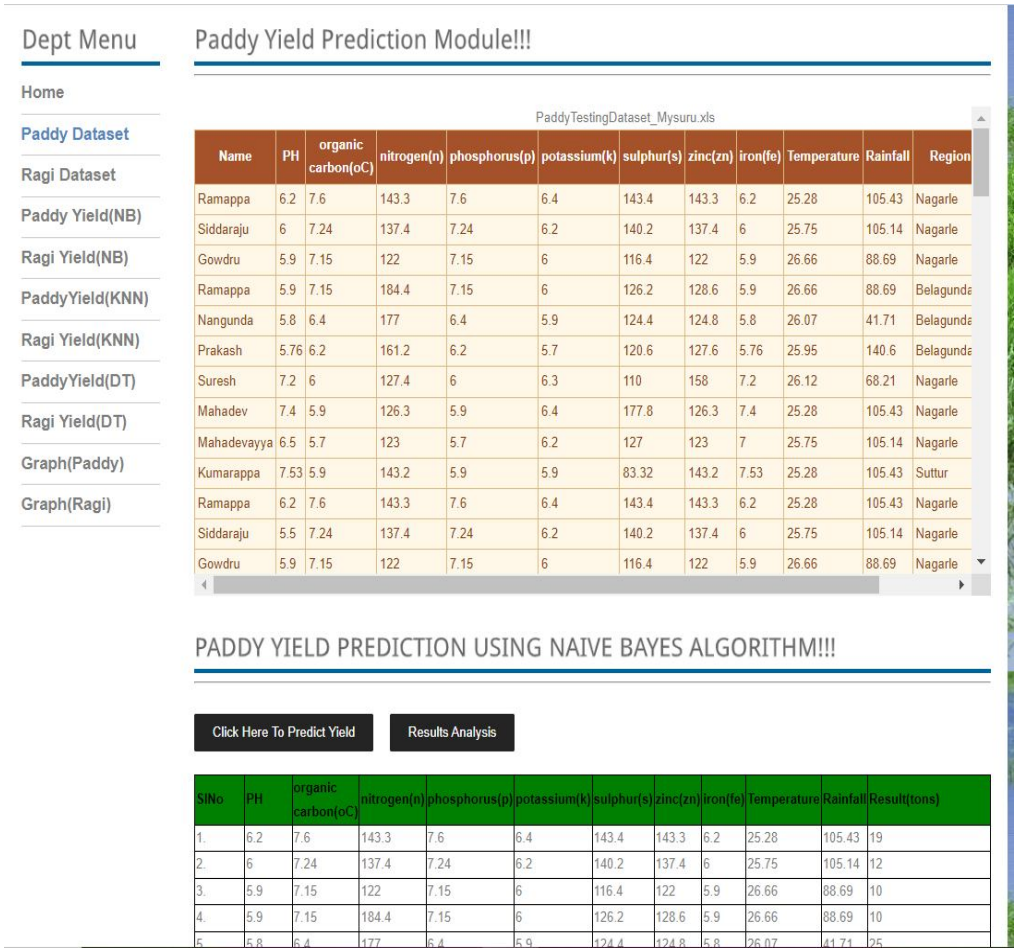


Figure 3: Paddy yield prediction using Naive Bayes Algorithm

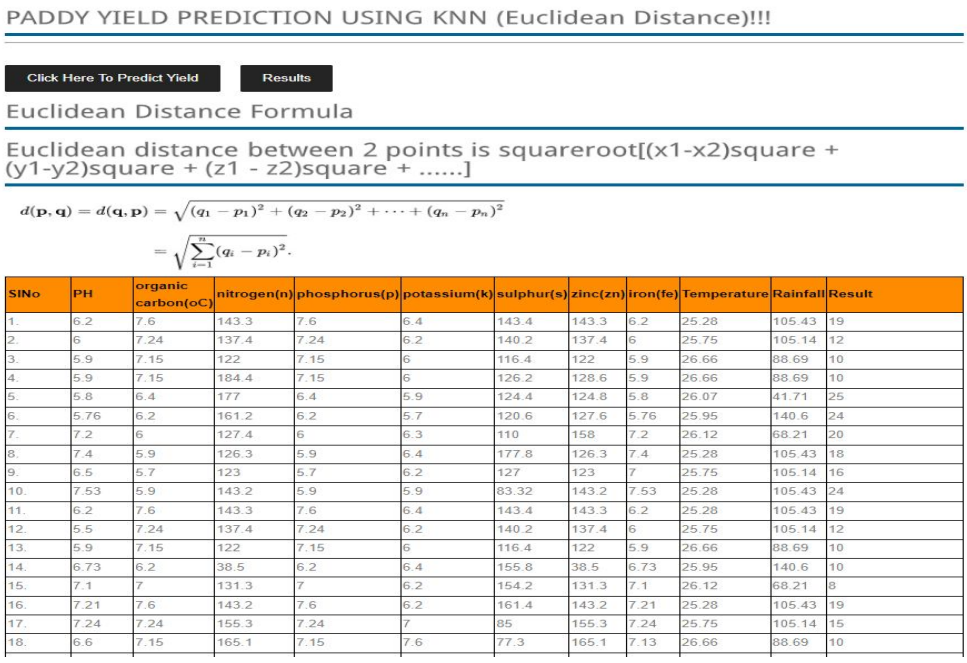


Figure 4: Paddy Yield prediction using the KNN algorithm

4.3.2 Decision Tree Algorithm

The decision tree algorithm is classified as unsupervised learning. They apply to both regression and classification problems. In the beginning, we scan the datasets from the root, calculate each attribute occurrence, calculate the internal node occurrences based on the best yield results, and recursively record the results based on each attribute. We classify the new node again using the best node. The Decision Tree algorithm is described in pseudo-code below.

Step 1: Scan the data set (storage servers)

Step 2: for each attribute a, calculate the gain [number of occurrences]

Step 3: Let a_best be the attribute of highest gain [highest count]

Step 4: Create a decision node based on a_best – retrieval of nodes[records] where the attribute

values match with a_best.

Step 5: recur on the sub-lists [list of patients] and calculate the count of outcomes[results] –

termed as sub-nodes. Based on the highest count we classify the new node.

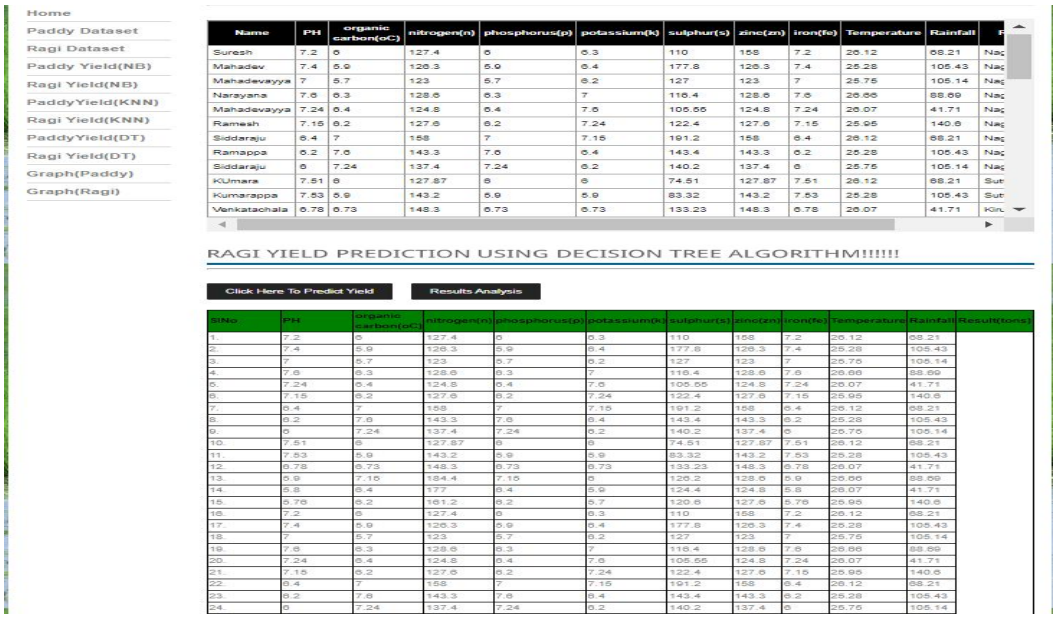


Figure 5: Ragi Yield prediction using Decision Tree Algorithm

V. EXPERIMENTAL RESULT

The outputs that we can get after the step-by-step execution of all the modules of the systems are defined by the subsequent snapshots. Figure 6 shows the home page for crop yield prediction. This page contains basic information such as services, user login, and new user registration.

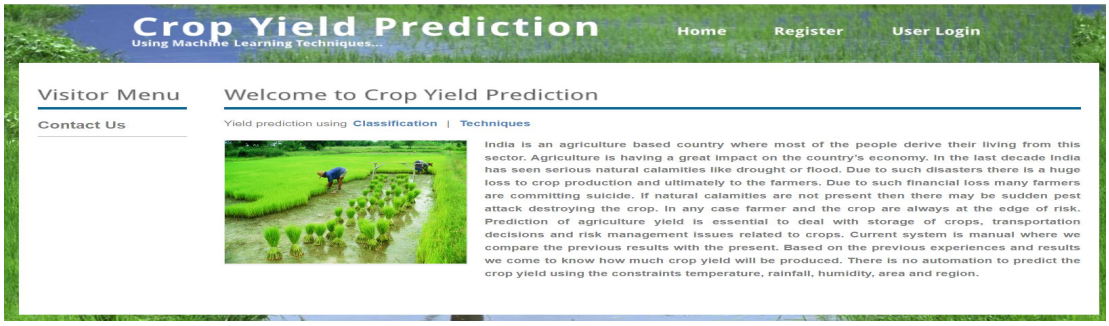
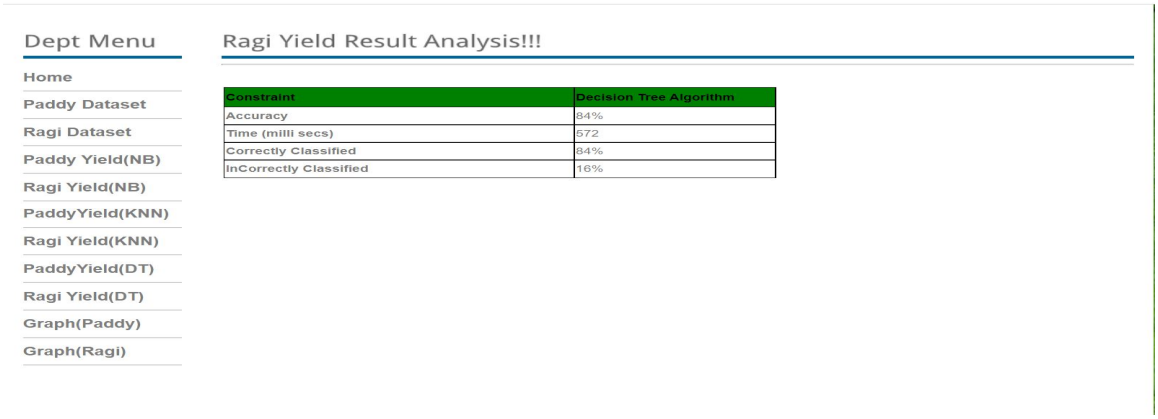


Figure 6: Home page



The above figure 7 depicts the result analysis for the ragi crop using the Decision tree algorithm.
Here we discover the precision of the procedure with the aid of dividing the datasets into training and testing datasets. 80% were taken into consideration as train datasets and 20% were considered training datasets.
Figure 8 depicts an online discussion forum. Farmers can view required agricultural documents and videos on this portal, and if they have any questions, they can post them, which will be forwarded to system administration.

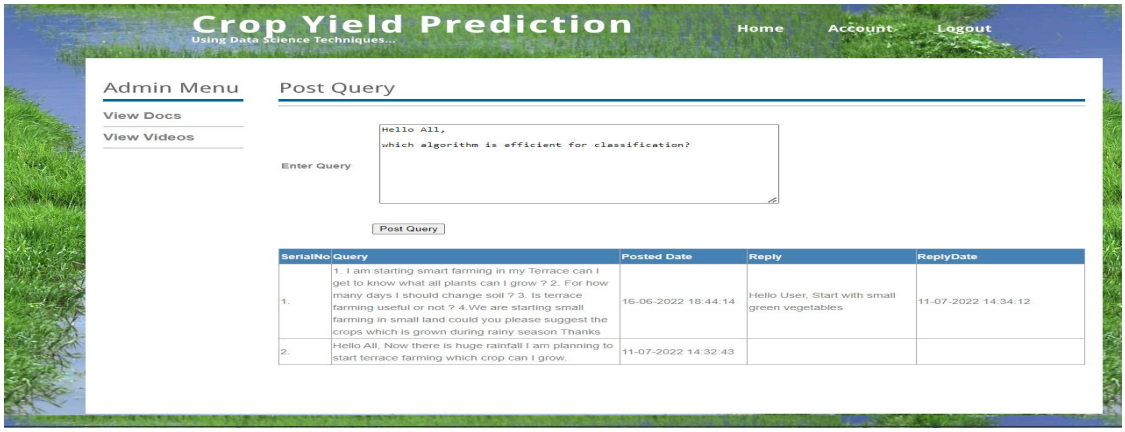


Figure 8: Farmer online discussion forum

REFERENCES

[1] Mr. A Suresh, "prediction of major crop yields of Tamil Nadu using K-Menas and Modified KNN", 978-1-5386-4765-3/18/$31:00 © 2018 *IEEE.*

[2] Lavanya," system to predict the paddy yield depending on agriculture data using Data Mining",© 2019 *JETIR* May 2019, Volume 6, Issue 5 www.jetir.org (ISSN-2349-5162)

[3] Ramesh A. Medar, Vijay S. Rajpurohit, "A Survey on Data Mining Techniques for crop yield prediction", *International Journal of Advanced Research in Computer Science and Management Studies*.Volume 2, Issue nine, Sept 2014.

[4] S.Kanaga, Rishi R, Srijit,v "Demand-based crop recommender system for farmers," *IEEE Conference on Technological Innovations in ICT for Agriculture and Rural Development (TIAR 2017)*", 978-1-5090-4437-5/17/$31.00, 2017.

[5] Maria Rossana C.De Leon and Eugene Rex L.Jalao, "A crop yield prediction model framework," Asia Pacific Industrial Engineering and Management System.

[6] D Ramesh and B Vishnu Vardhan, "Region-precise Crop Yield Analysis: A Data Mining Approach," UACEE *International Journal of Advances in Computer Science and its Applications-IJCSIA*, Volume 3: Issue 2.

[7] S. Veenadhari, Dr. Bharat Mishra, and Dr. C D Singh, "Soyabean Productivity Modeling Using Decision Tree Algorithms," *International Journal of Computer Applications*, Vol 27-No7, August 2011.

[8] George RuB, "Data Mining of Agricultural Yield Data: A Regression Comparison."

[9] Wim G.M. Bastiaanssen and Samia Ali, "A new crop yield forecasting version based on satellite measurements taken across the Indus Basin in Pakistan," Agriculture, Ecosystems, and Environment 94 (2003) 321-340.

[10] M. Gunasundari, Dr. T Arunkumar, Ms. R. Hemavathy," CRY – An improved Crop Yield Prediction model using Bee Hive Clustering 423 Approach for Agricultural data sets", Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering, February.

[11] A.T.M Shakil Ahamed, Navid Tanzeem Mahmood, Nazmul Hossain, Mohammad Tanzir Kabir, Kallal Das, Faridur Rahman, Rashedur M Rahman," 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 2015, June 1-3 2015, Takamatsu, Japan.

[12] R. Sujatha, Dr. P.Isakki, A Study on Crop Yield Forecasting Using Classification Techniques, 978-1- 4673- 8437-7/sixteen/$31.00 ©2016 IEEE.