Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

CROP YIELD PREDICTION AND FERTILIZER RECOMMENDED SYSTEM

Ms.ARAVAPALLI CHATHURYA[1], Ms.NADIPALLI TEJASWINI[2], Ms.RANKIREDDY SAI SWETHA[3], Ms.REPALLE JAHNAVI[4], Ms.PAPPULA MADHAVI[5]

1 .BTech, Vijaya Institute of Technology For Women, Enikepadu, Vijayawada, Andhra Pradesh, India. Email : aravapallichathurya@gmail.com
2. BTech, Vijaya Institute of Technology For Women, Enikepadu, Vijayawada, Andhra Pradesh, India.
3. BTech, Vijaya Institute of Technology For Women, Enikepadu, Vijayawada, Andhra Pradesh, India.
4. BTech, Vijaya Institute of Technology For Women, Enikepadu, Vijayawada, Andhra Pradesh, India.
5. Assistant Professor, Computer Science and Engineering, Vijaya Institue of Technology For Women, Enikepadu, Vijayawada, Andhra Pradesh, India. Email : pappulamadhavi06@gmail.com

ABSTRACT

India's economy, which is largely dependent on agricultural products and yield growth, is an agricultural nation. An emerging area of study in crop yield analysis is data mining. The ability to predict yield is a critical issue in agriculture. Any farmer is curious to know how much yield he can anticipate and what crop is best for his particular plot of land. Examine the different factors that are associated, such as the location and the pH level used to calculate the soil's alkalinity. Location is used in conjunction with the use of third-party applications like APIs for weather and temperature, type of soil, nutrient value of the soil in that region, amount of rainfall in that region, and soil composition. Percentages of nutrients like Nitrogen (N), Phosphorous (P), and Potassium (K) are also used. In order to build a model, all of these data attributes will be examined. The data will then be trained using a variety of appropriate machine learning techniques, including SVM, Random-Forest, KNN, and Voting Classifier. The system has a model built in to be exact and accurate in predicting crop output and to provide the end user with the right advice about the required fertiliser ratio based on atmospheric and soil factors of the land which enhance to increase crop yield and increase farmer revenue. The data on soil quality and weather-related information are thus inputs into the proposed system. The soil's composition, including its levels of nitrogen, phosphorus, potassium, and pH. To forecast a better harvest, use weather data like rainfall, temperature, and humidity. We are using the datasets from the Kaggle website in our study.

1 INTRODUCTION

One of the most significant professions in our nation is agriculture. It is the most diverse economic sector and is crucial to the overall growth of the nation. To meet the demands of 1.2 billion people, almost 60% of the country's land is used for agriculture. Therefore, modernising the agricultural sector is crucial and will help our nation's farmers make money. The process of studying data sets to make inferences about the information they contain is known as data analysis (DA), and it is increasingly done with the aid of specialised hardware and software. In the past, yield prediction was done by taking into account the farmer's prior experience with a certain field and crop. Farmers are compelled to grow an increasing number of crops, though, as the environment is changing very quickly day by day. Due to the existing circumstances, many of them are unaware of the advantages of growing the new crops and lack sufficient understanding about them. Understanding and predicting crop performance under various environmental situations can also boost farm output. The data on soil quality and weather-related information are thus inputs into the proposed system. The soil's composition, including its levels of nitrogen, phosphorus, potassium, and pH. information about the weather, such as temperature, humidity, and rainfall. We are using the datasets from

the Kaggle website in our study.

2. RELEATED WORK

[1] VIRENDRA PANPATIL ET:
By creating a framework for effective yield proposal, it had done enormous good for Indian ranchers. They developed a framework using classifier models including the Naive Bayes Classifier, KNN, and Decision Tree Classifier. The suggested framework can be used to determine the ideal time for planting, plant growth, and plant harvesting. They used unique classifiers to get improved accuracy, for instance: Decision trees exhibit lower precision when a dataset has more variety, however Naive Bayes exhibits preferable accuracy over choice trees for these datasets. The best feature of the framework is that it can easily be flexible all things considered and used to test on different yields.
[2] MAYANK ET:
It has been assumed that this paper will create an extemporaneous framework for crop yield using managed AI calculations with the goals of providing an easy-to-use user interface, increasing the accuracy of crop yield forecasting, and examining various climatic boundaries, such as overcast cover, precipitation, temperature, and so on. They chose the MAHARASHTRA State for the proposed framework's implementation, and they used a government website, such www.data.gov.in, to gather information. They used formulas, such as the Random Forest Algorithm, to anticipate crop yields, and they produced a website page that was user-friendly for everyone to use. The suggested framework's main advantage is its precision rate, which is greater than 75% overall for the yields and regions examined.
[3] SHWETA ET:
It has been implied that this article will examine the various ways that AI is being used in agricultural settings. Furthermore, by using these methods, it is possible to agree on the right crop, season, and location. K-Nearest Neighbor and Naive Bayes are both used in the calculations. The calculations make use of execution precision.
[4] AMIT KUMAR ET:
It has been assumed that this study aids in predicting crop arrangements, increasing yield rates, and benefiting ranchers. Additionally, applying machine learning to farming involves assessing crop duplicates, predicting crop diseases, and developing various water system designs. False neural networks were used in the calculations. The significant problem with brain organisation is that it takes experimenting to achieve the right organisation that works best for the arrangement. The second problem with brain organisation is the reliance on equipment; as more computations are performed both backwards and forwards, more preparation is required. It takes knowledge and patience to ensure that the organisational structure is appropriate. The suggested approach also emphasises crop selection using both natural and economical factors. The framework also makes use of the financial factor On the off chance that the yields have the same yield but different yield costs, it is the crop cost that assumes a significant role. The framework also makes use of another technique that is crop sequencing, which provides a complete yield arrangement that may be created throughout the season. The suggested methodology also focuses on crop selection using only ecological and financial factors. The crop cost, which plays a large role if two crops have the same yield but different yield costs, is another financial component that the framework uses. The framework also makes use of a different technique called crop sequencing, which provides a comprehensive yield plan that may be created throughout the season.
[5] MANJULA ET:
By using rule-based mining, it has been assumed that this article will aid in enhancing crop yield rates. The crop's yield is forecast in the article via affiliation rule mining. The k-Means algorithm, bunching method, and derived

affiliation rule mining are used in the calculations. The fact that the paper uses affiliation rule digging to predict crop productivity is a big barrier. The problem with affiliation decision mining is that it occasionally generates too many rules, which reduces the precision of the expectation. Similar to how the concepts will generally vary depending on the dataset, so will the results. The suggested framework focuses primarily on the issue of crop yield expectations, which play a significant role in crop selection since ranchers can select crops with the highest yield. The frameworks mine affiliation rules to find the rules and crops that produce the most. This framework is focused on creating an expectation model that might be used to anticipate agricultural yield in the future.

[6] RAKESH KUMAR ET:

By using order procedures and focusing on boundaries, it is hoped that this research may assist in accelerating crop yields. The paper explains how several calculations are used to achieve the equivalent. The computations that are suggested include support vector machines, clustering algorithms, Bayesian calculations, and K-implies calculations. The challenge is that, when using the suggested calculations, there may not be adequate precision and execution referenced in the study. The paper is a study and only suggests using the calculations; however, there is no evidence of their use provided in the publication. The method used in this research for agricultural decision-making is specifically focused on the potential for plants to develop according to season. The suggested method resolves crop choice, which is mostly based on anticipated yield costs supported by constraints (such as environment, soil type, water thickness, and crop type). It identifies a succession of plants whose creation with regard to day are greatest over season using crop, their planting time, estate days, and anticipated yield charge as information.

[7] RAJSHEKHAR ET:

In India, a variety of agricultural crops are produced, and those crops depend on a limited number of different factors. The illustrates and provides us with details for a rundown of the strategies used. It is possible to obtain knowledge or knowledge that can help ranchers and government organisations make wise decisions and improve rules that contribute to increased creation, for instance, natural science, economy, and furthermore the geological variables covering such procedures and strategies on memorable yield of disparate yields. In this paper, we focus on using information mining techniques to extract data from horticultural records in order to evaluate improved crop yields for primary yields in important regions of India. In our research, we discovered that Indian ranchers will benefit from the precise expectation of different indicated crop yields across distinct locations. Indian ranchers will use this to produce a variety of crops in various areas.

[8] VISHNUVARDHAN ET:

They looked at a few developments in India that are addressing serious issues in order to maximise crop productivity. Over 60 out of every 100 plants genuinely depend on rainy precipitation. Rapid advancements in the field of information technology for agriculture have created a fascinating area for speculation about crop productivity. The risk associated with yield expectations is a significant problem that must be addressed in light of the facts now available. The more accurate assessments are made via information mining techniques. In agriculture, several data mining techniques are used to estimate crop production for the upcoming year. With the help of the Multiple Linear Regression (MLR) technique and the Density-based grouping procedure, this paper provides a succinct analysis of crop yield forecasting for a specific district, such as the East Godavari region of Andhra Pradesh in India. This study makes an effort to determine the precise crop yield analysis for the area, and it is generated by using both the Multiple Linear Regression technique and the Density-based bunching strategy. These models were examined in relation to the diverse regions of Andhra Pradesh; hence, the evaluation technique is abandoned with only the East Godavari region of Andhra Pradesh in India.

3 Implementation Study

Here, the crop data's raw data is cleaned, and metadata is added by omitting the items that are converted to integers. So, training using the data is simple. Hear all the information. This pre-processing involves loading the metadata first, after which it is associated to the data and takes the place of the modified data. Then, this data will be transferred further, the unnecessary data in the list will be removed, and the data will be divided into train and test data. To facilitate this data splitting into train and test, we must import train test split. In the scikit-learn, this will assist the preprocessed data in being divided into train and test sets of data in accordance with the specified weights in the code. The test and train are divided in half, or 0.2 and 0.8, or 20 and 80%, respectively.
The following steps to be followed for the implementation
1) Modal Creation
2) Modal Evaluation
3) Prediction.

4 **PROPOSED WORK AND ALOGRITHAM**

In the suggested system, we create crop prediction utilising an effective algorithm.The task at hand is to create an effective model to foretell a better crop. The Voting classifier, which is nothing more than a hybrid classification/ensemble of models, is one of the machine learning techniques we utilise in this project. The voting classifier in our project is a collection of models derived from SVM, Random-Forest, and KNN. It can improve forecast accuracy and provide a better system.

4.1 BENEFITS OF THE PROPOSED SYSTEM

The project's ultimate goal is to forecast the best crop.
Farmers can improve crop productivity by using early problem identification and management techniques.
With the aid of a machine learning algorithm, we must examine vast amounts of data to better understand crop yields. This will allow the algorithm to make accurate crop predictions and advise farmers on how to grow better crops.

5 METHODOLOGIES

Give the temperature, rainfall, humidity, PH, nitrogen, phosphorus, and potassium values. The dataset has already been trained. Our value will be compared to the dataset, and the final result will show what seed we planted in that specific location.
In this research, we employ machine learning techniques like the voting classifier, which is simply an ensemble of models for hybrid classification. The voting classifier in our project is a collection of models derived from SVM, Random-Forest, and KNN. It can improve forecast accuracy and provide a better system.

SVM(Support Vector Machine) Algorithm:

An SVM model is a mapping of the examples as points in space with as much space between the examples of the various categories as possible. SVMs may effectively do non-linear classification in addition to linear classification by implicitly translating their inputs into large feature spaces.An SVM training method creates a model that categorises fresh examples according to one of two categories given a series of training examples that have each been tagged as belonging to one of the categories. This makes the algorithm a non-probabilistic binary linear classifier. Let's move on after you have a basic knowledge of this subject. Here, I'll go over an example of how machine learning methods were used to classify cancer using SVM utilising UCI datasets.

Random Forest Regression Algorithm:

Random forest is also a widely-used algorithm for non-linear regression in Machine Learning. Unlike decision tree

regression (single tree), a random forest uses multiple decision trees for predicting the output. Random data points are selected from the given dataset (say k data points are selected), and a decision tree is built with them via this algorithm. Several decision trees are then modelled that predict the value of any new data point. Since there are multiple decision trees, multiple output values will be predicted via a random forest algorithm. You must find the average of all the predicted values for a new data point to compute the final output. This happens due to the large number of decision trees mapped under this algorithm, as it requires more computational power. It's a bagging technique not a boosting technique trees run parallel. i.e no interaction between these tree while building trees.

Classification Algorithms:

Classification is an algorithm that finds functions that help divide the dataset into classes based on various parameters. a computer program gets taught on the training dataset and categorizes the data into various categories depending on what it learned.

- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier
- KNN - K Nearest Neighbour

Random Forest Algorithm:

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. Random Forest is a classifier that contains several decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. It predicts output with high accuracy, even for the large dataset it runs efficiently. It can also maintain accuracy when a large proportion of data is missing.

K-Nearest Neighbour (KNN) Algorithm:

K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm. K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems. KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much like the new data.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

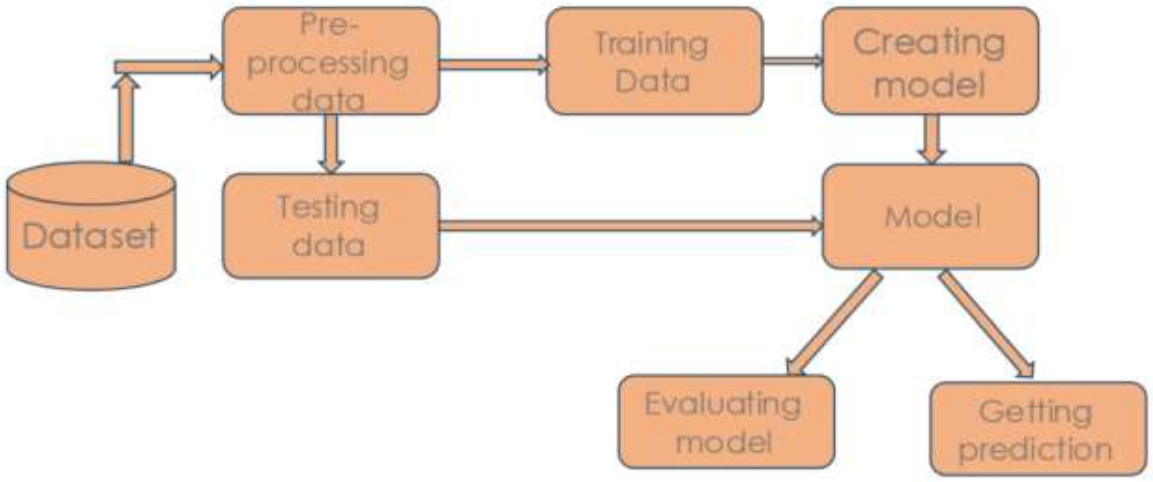


Fig. 1: propose Architecture

6 **RESULTS AND DISCUSSION**
SCREENSHOTS

Fig-6.1 Screenshots of some of the pest images for which we are making pest predictions

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

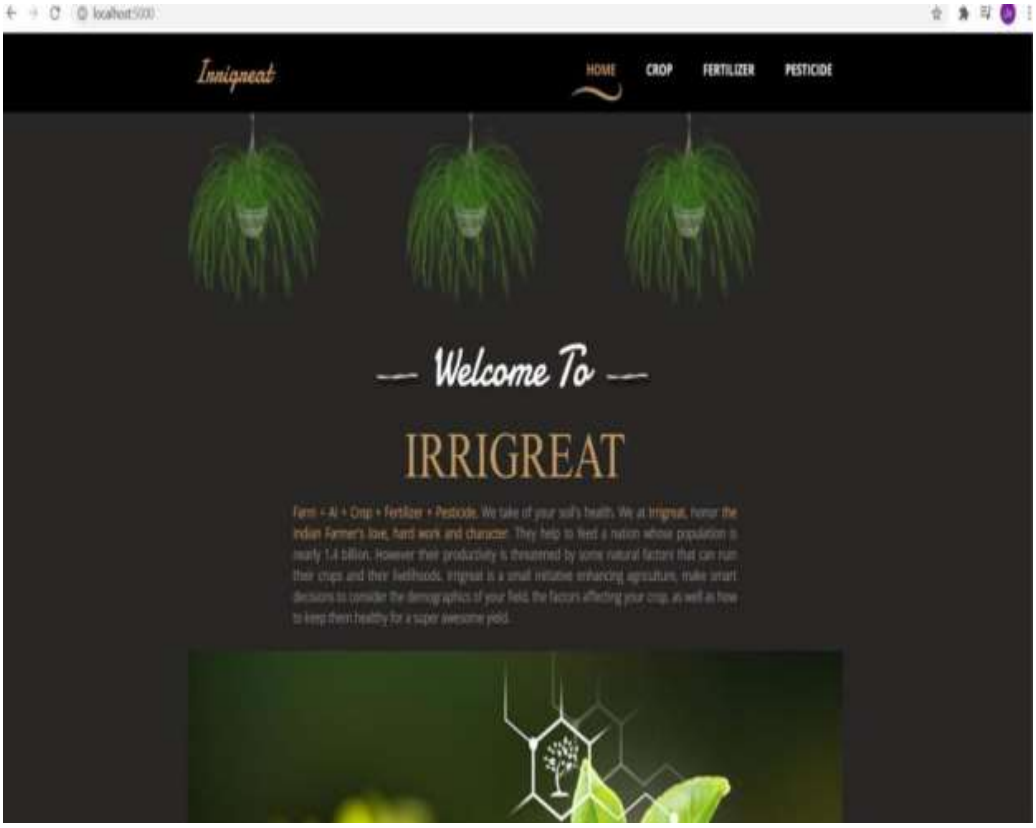


Fig-6.2 Home Page

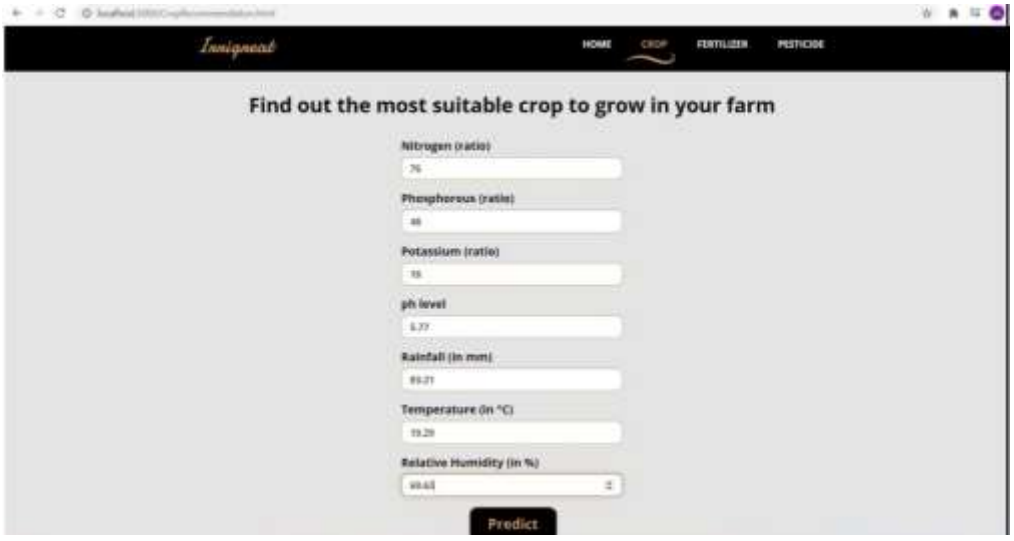


Fig-6.3 RECESSION OF DATA FOR CROP PREDICTION

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

Fig-6.4 Crop Predicted

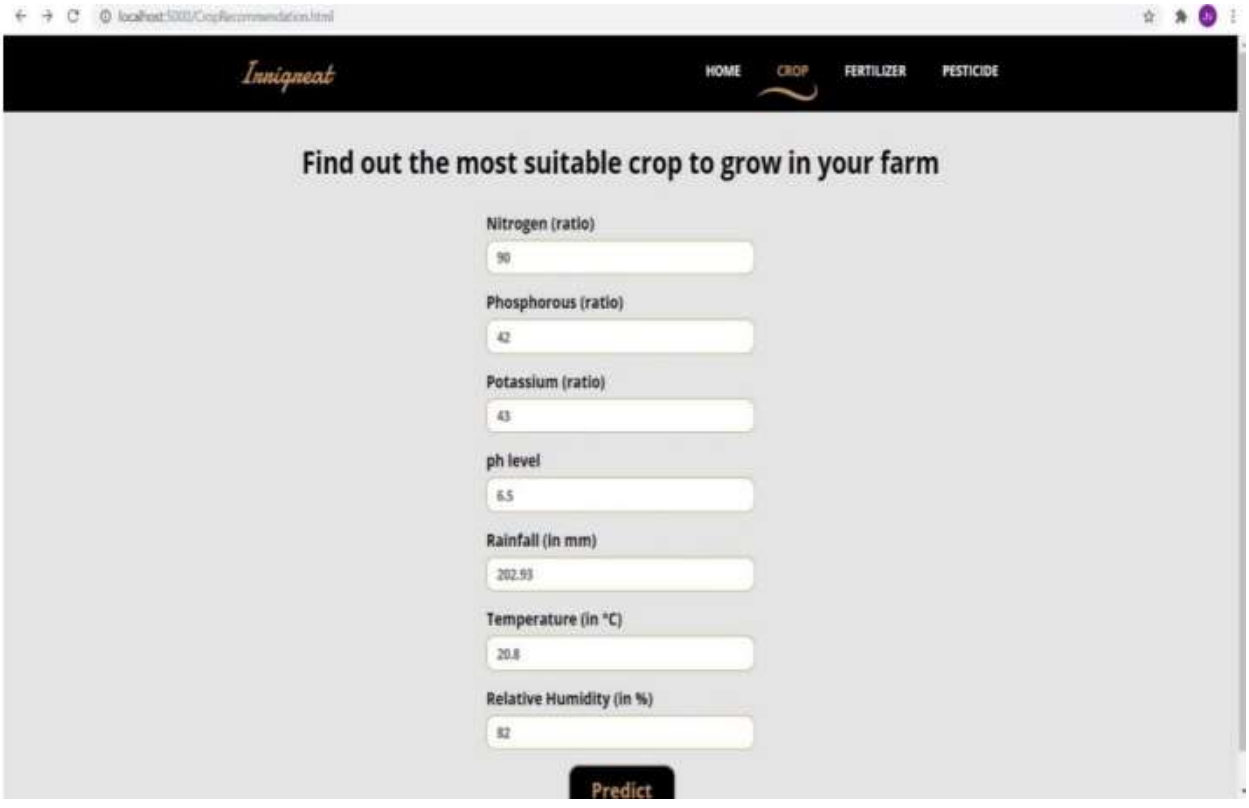
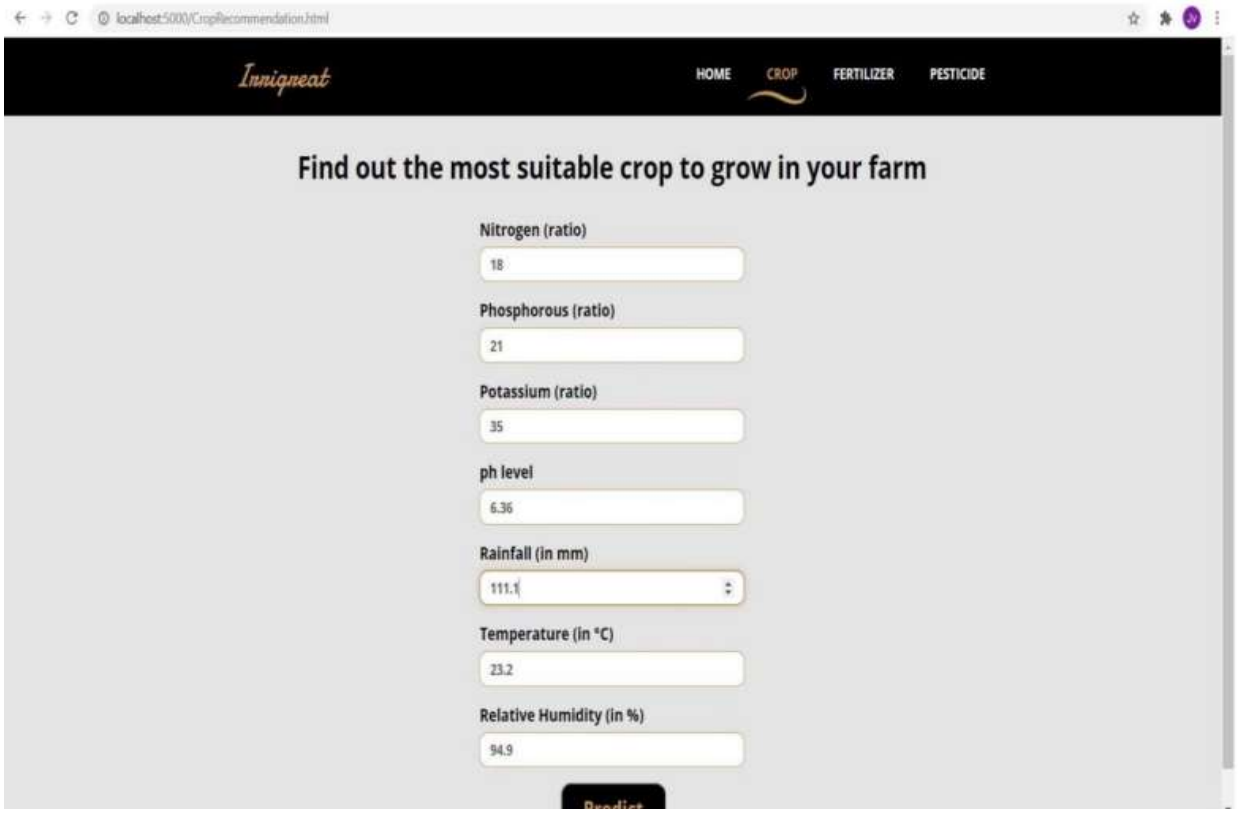


Fig-6.5 RECESSION OF DATA FOR CROP PREDICTION

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

Fig-6.6 Crop Predicted



Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

Fig-6.7 RECESSION OF DATA FOR CROP PREDICTION

Fig-6.8 Crop Predicted

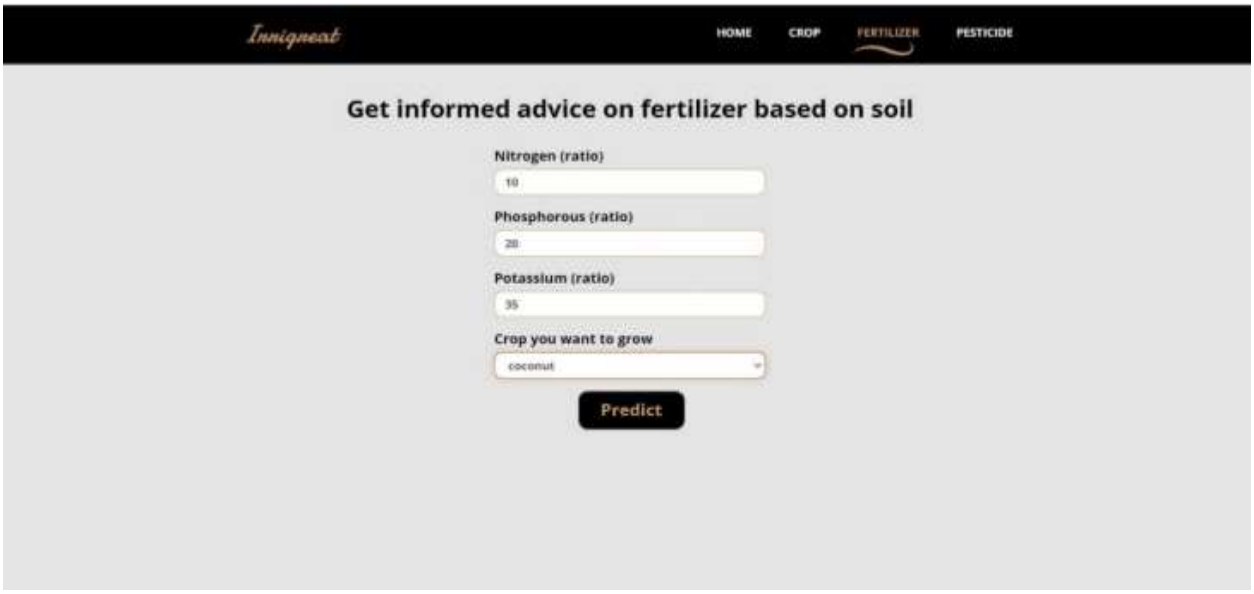


Fig-6.9 PREDICTING FERTILIZERS BY USING INPUTS

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

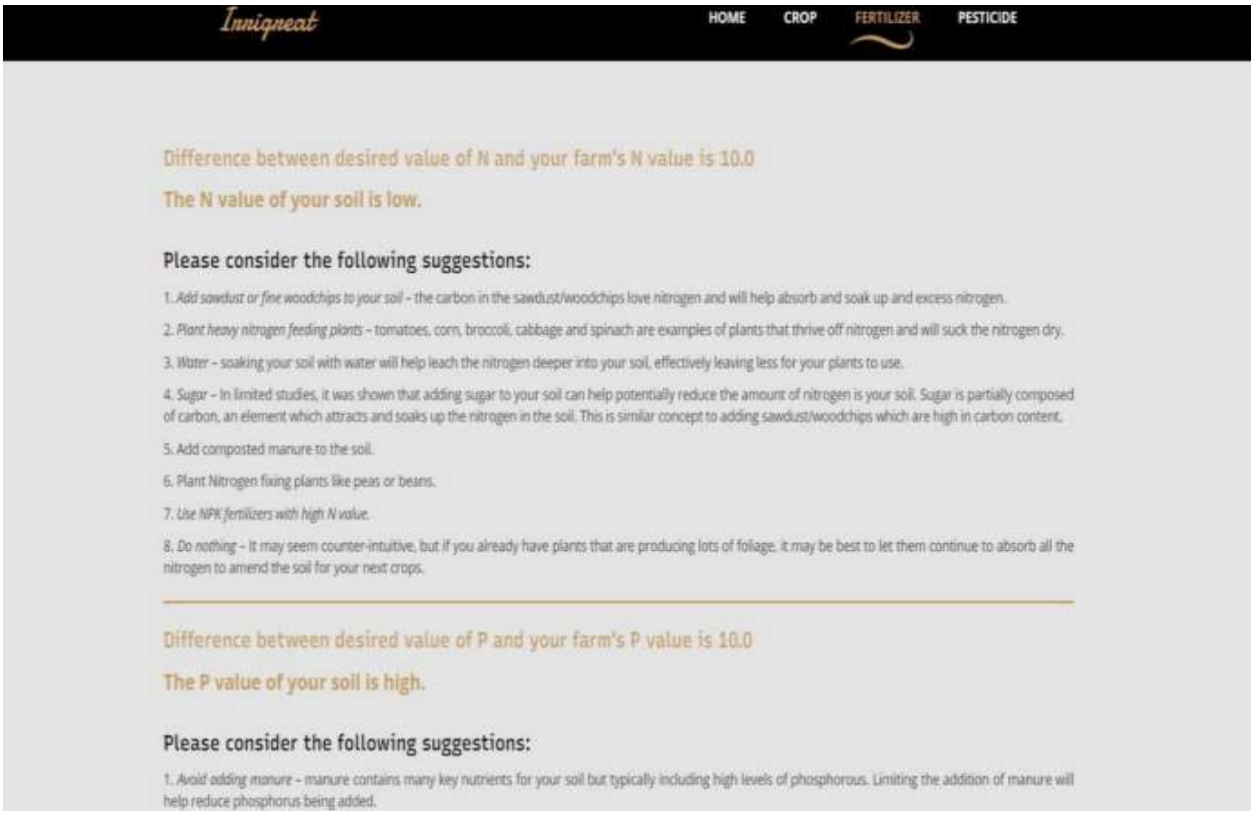


Fig-6.10 Suggestions

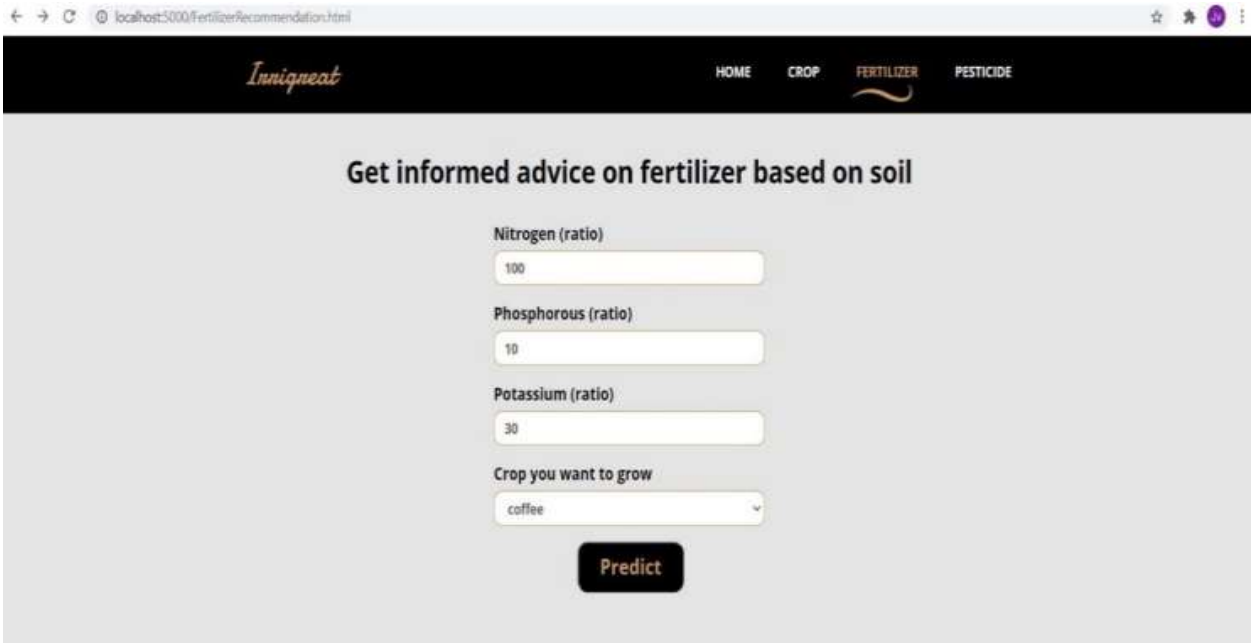


Fig-6.11 Detailed information on fertilizers

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

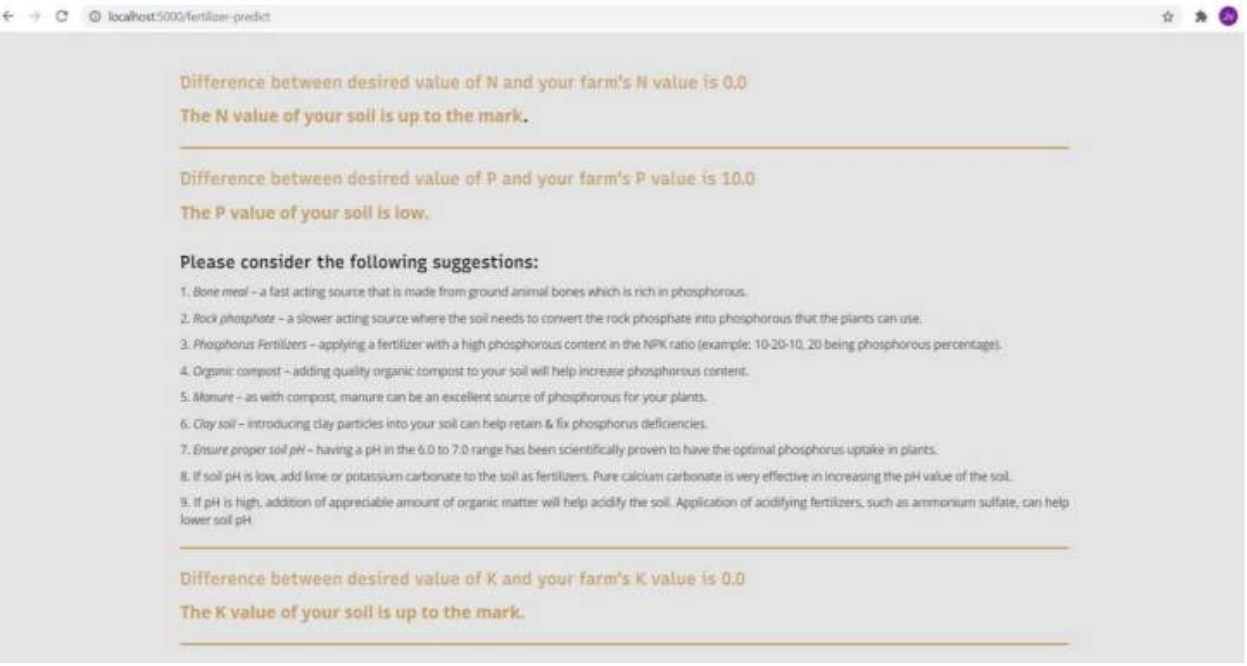


Fig-6.12 Recommendations for fertilisers.

Fig-6.13 Predicting pesticides

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

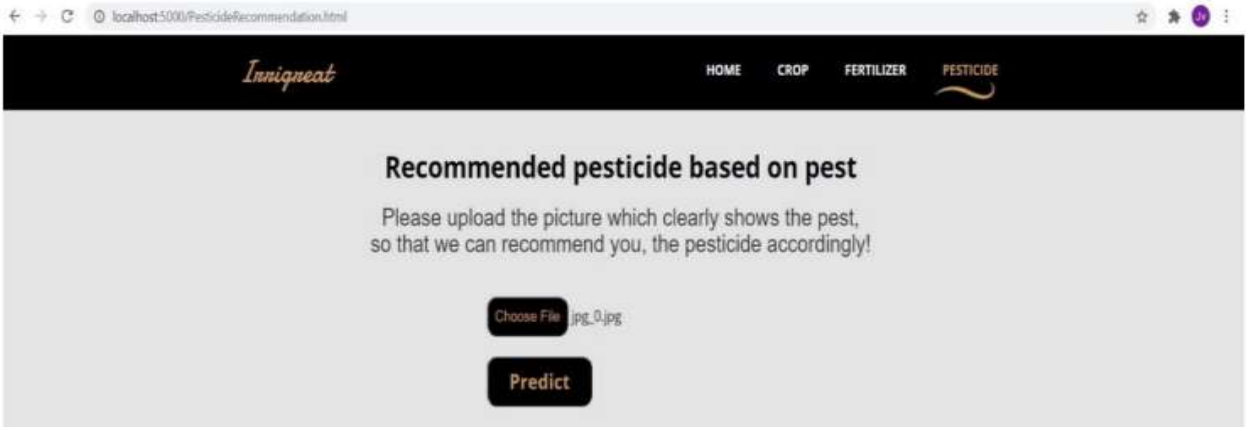


Fig-6.14 GETTING COMMENTS TO PROJECT PESTICIDES

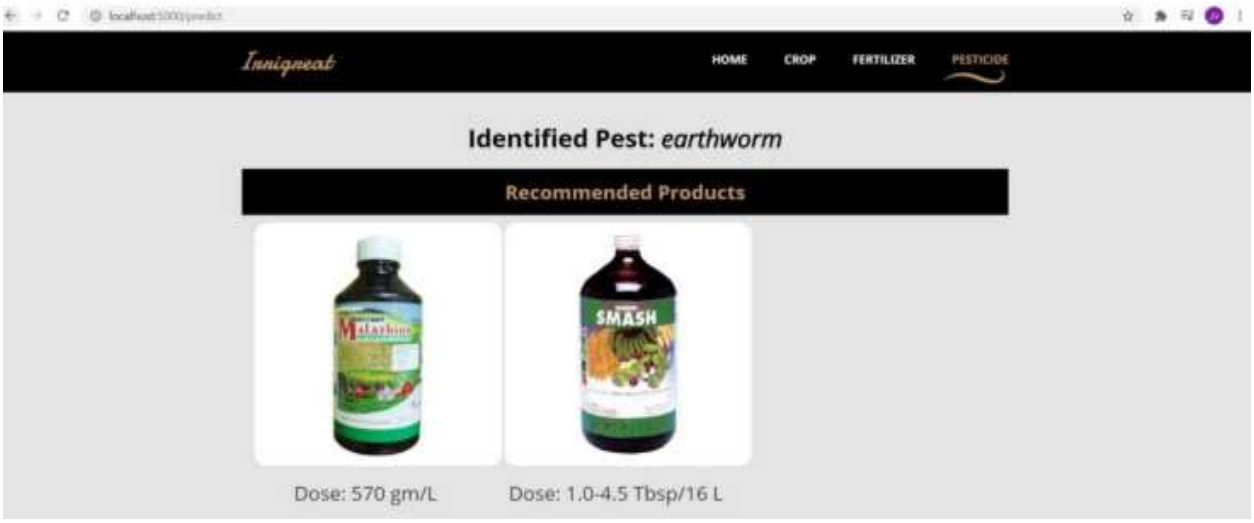


Fig-6.15 PESTISIDE IDENTIFIED

Fig-6.16 INPUT TO PREDICT PESTICIDES

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

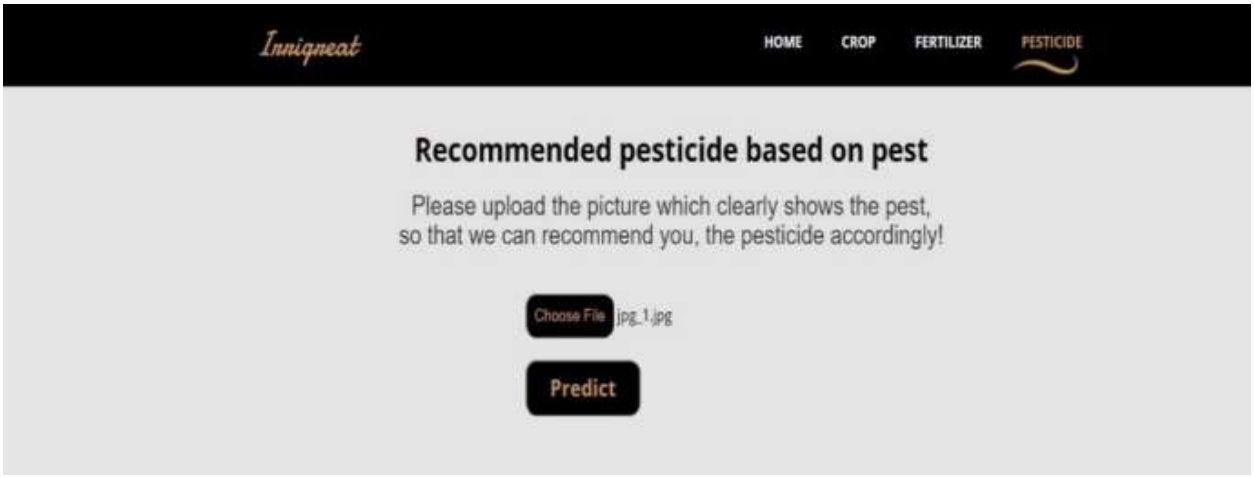


Fig-6.17 Based on the pest, a recommended pesticide

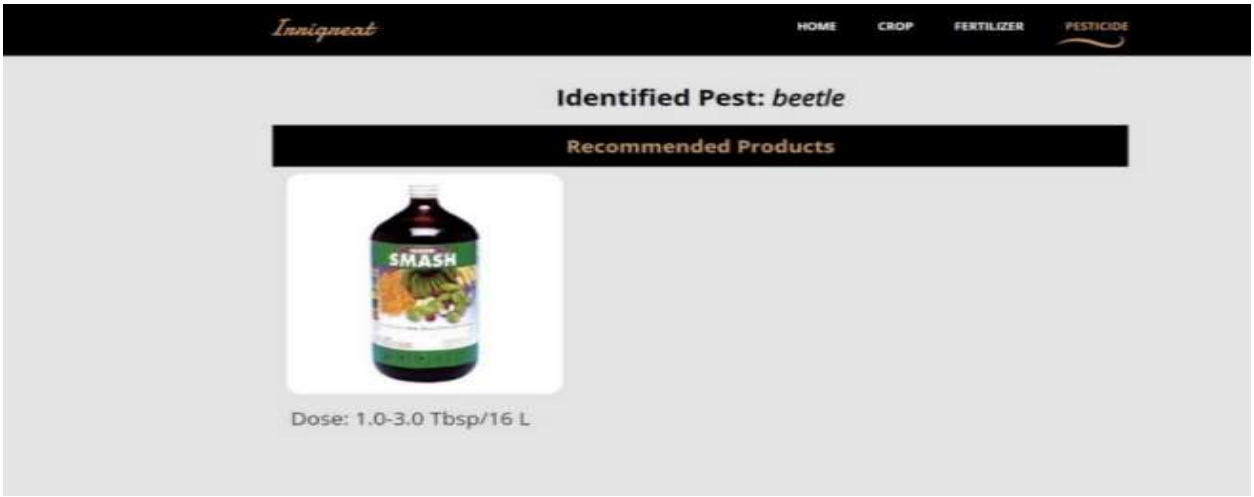


Fig-6.18 Pest Identified

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 4, April : 2023

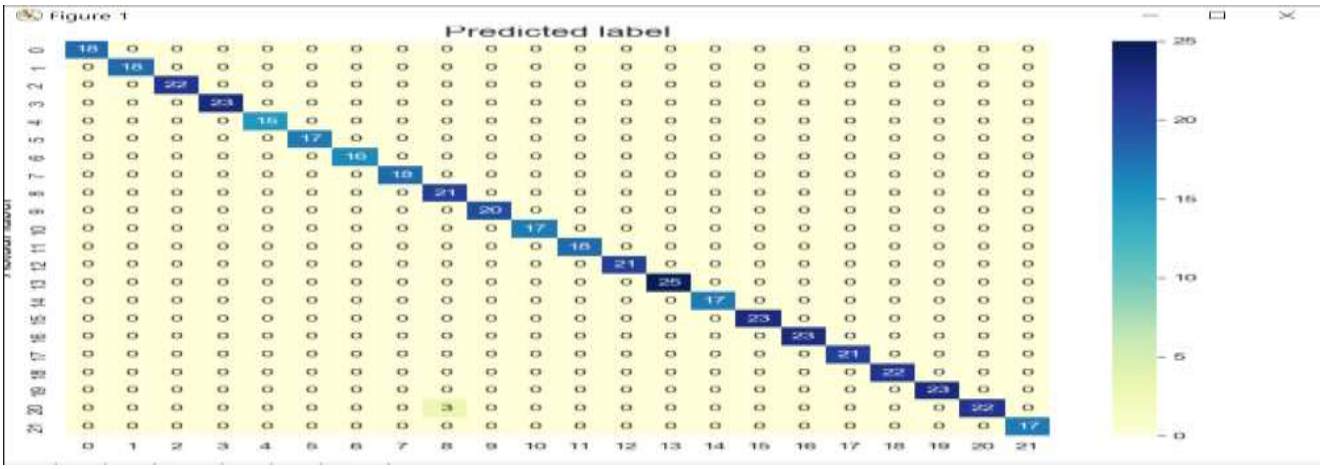


fig 6.19 predict label

7. CONCLUSION AND FUTURE WORK

The suggested work uses a voting classifier, which is nothing more than an ensemble of models, to present a crop prediction framework. Here, voting classifier ensembles are used in our research. Models derived from Random-Forest, KNN, and SVM. Our project makes more accurate crop predictions. The framework will assist in lowering the difficulties faced by the farmers and preventing them from attempting suicide. It will serve as a conduit for providing farmers with the useful information they need to produce high returns and subsequently increase benefits, which will lower rates of self-destruction and lessen their difficulties.

The overall profit of the country has increased as a result. In our experiment, we discovered that farmers will benefit from accurate predictions of various specified crop yields across several districts. Farmers will use this to sow various crops in various districts. Geospatial analysis can be introduced in the near future to boost accuracy and implement better geographic data.

8. REFRENCES

[1] Mayank Champaneri, Chaitanya Chandvidkar , Darpan Chachpara, Mansing Rathod, "Crop yield prediction using machine learning" International Journal of Science and Research ,April 2020.
[2] Pavan Patil, Virendra Panpatil, Prof. Shrikant Kokate, "Crop Prediction System using Machine Learning Algorithms", International Research Journal of Engineering and Technology, Feb 2020.
[3] Ramesh Medar , Shweta, Vijay S. Rajpurohit, "Crop Yield Prediction using Machine Learning Techniques", 5th International Conference for Convergence in Technology, 2019.
[4] Trupti Bhange, Swati Shekapure, Komal Pawar, Harshada Choudhari, "Survey Paper on Prediction of Crop yield and Suitable Crop", International Journal of Innovative Research in Science, Engineering and Technology, May 2019.
[5] E. Manjula, S. Djodiltachoumy, "A Modal for Prediction of Crop Yield "International Journal of Computational Intelligence and Informatics, March 2017.
[6] 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), Feb 2017.
[7] Rakesh Kumar, M.P. Singh, Prabhat Kumar, J.P. Singh, "Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique", 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM),Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, T.N., India., May 2015.

[8] Rajshekhar Borate., "Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in India", International Journal of Novel Research in Computer Science and Software Engineering,Vol. 3, Issue 1, pp: (34-37), April 2016. 79

[9] D Ramesh, B Vishnu Vardhan, "Analysis of Crop Yield Prediction using Data Mining Techniques", International Journal of Research in Engineering and Technology (IJRET),Vol.4, 2015.

[10] S.Veenadhari, Dr Bharat Misra, Dr CD Singh.2019."Machine learning approach for forecasting crop yield based on climatic parameters.".978-1-4799-2352- 6/14/$31.00 ©2014 IEEE.

[11] I gor Oliveira, Renato L. F. Cunha, Bruno Silva, Marco A. S. Netto.2018."A Scalable Machine Learning System for PreSeason Agriculture Yield Forecast.".978-1-5386-9156- 4/18/$31.00 ©2018.

[12] Neha Rale, Raxitkumar Solanki, Doina Bein, James Andro-Vasko, Wolfgang Bein."Prediction of Crop Cultivation".978-1-7281-0554-3/19/$31.00©2019 IEEE.

[13] Md. Tahmid Shakoor, Karishma Rahman, Sumaiya Nasrin Rayta, Amitabha Chakrabarty.2017."Agricultural Production Output Prediction Using Supervised Machine Learning Techniques".978-1-5386-3831-6/17/$31.00 ©2017 IEEE.

[14] G Srivatsa Sharma, Shah Nawaz Mandal, Shruti Kulkarni, Monica R Mundada, Meeradevi.2018."Predictive Analysis to Improve Crop Yield Using a Neural Network Model".978-1-5386-5314-2/18/$31.00 ©2018 IEEE.

[15] Rashmi Priya, Dharavath Ramesh.2018."Crop Prediction on the Region Belts of India: A Naïve Bayes MapReduce Precision Agricultural Model". 978-1-5386-5314- 2/18/$31.00 ©2018 IEEE