ATHARVA EDUCATIONAL TRUST'S
ATHARVA COLLEGE OF ENGINEERING
(Approved by AICTE, Recognized by the Government of Maharashtra
& Affiliated to University of Mumbai)
Department of Computer Engineering
Academic Year 2020-21

"Crop It: Crop Prediction Using ML"

Submitted in partial fulfillment of the requirements of the degree
BACHELOR OF ENGINEERING IN COMPUTER ENGINEERING

By
Shriya Bijam Roll No: 10
Bharat Choudhary Roll No: 16
Nikita Dubey Roll No: 64

Supervisor

Prof. Shweta Sharma

CERTIFICATE

This is to certify that the Mini Project entitled **" Crop It : Crop Prediction Using ML"** is a bonafide work of **Shriya Bijam (10), Bharat Choudhary (16), Nikita Dubey (64) ,** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **"Bachelor of Engineering"** in **"Computer Engineering" .**

(Prof. Shweta Sharma)
Supervisor

(Prof. Dr. SUVARNA PANSAMBAL)
Head of Department

(Prof. Dr. Shrikant Kallurkar)
Principal

Mini Project Approval

This Mini Project entitled **" Crop It : Crop Prediction Using ML "** by **Shriya Bijam (10), Bharat Choudhary (16), Nikita Dubey (64)** is approved for the degree of **Bachelor of Engineering** in **Computer Engineering.**

Examiners

1……
(Internal Examiner Name & Sign)

2……
(External Examiner name & Sign)

Date:

Place:

Contents

ABSTRACT:

Agriculture is one of the most essential and widely practiced occupations in India and it has a vital role in the development of our country. Around 60 percent of the total land in the country is used for agriculture to meet the needs of 1.2 billion people, so improving crop production is therefore seen as a significant aspect of agriculture. Basically if we have a piece of land, we need to know what kind of crop can be grown in this area. Agriculture depends on the various soil properties. Production of crops is a difficult task since it involves various factors like soil type, temperature, humidity etc. If it is possible to find the crop before sowing it, it would be of great help to the farmers and the other people involved to make appropriate decisions on the storage and business side. The proposed project would solve agricultural problems by monitoring the agricultural area on the basis of soil properties and recommending the most appropriate crop to farmers, thereby helping them to significantly increase productivity and reduce loss. Our project is a recommendation system which makes use of different machine learning techniques such that it recommends the suitable crops based on the input soil parameters. This system thus reduces the financial losses faced by the farmers caused by planting the wrong crops and also it helps the farmers to find new types of crops that can be cultivated in their area.

Keywords : Random Forest Classifier, Machine Learning, Soil, Crop Prediction, Django, Html, Css, Agriculture.

List Of Figures:

1. INTRODUCTION:

1.1. INTRODUCTION TO: Crop It

Agriculture has a major role in the lives of every individual. From the olden times itself agriculture is considered to be one of the main practices practiced in India. In olden times, people used to cultivate crops in their own land in order to meet their requirements.With the rapid population growth, these innovations are very much needed to meet the needs of every person.

Our country had undergone several fluctuations in the price of onions last year. This problem occurred due to many unfavorable conditions that prevented the growth of onions. A continuous shortage in the production of onions again in the next few months had a very bad effect on the lives of the common people. This happened because the middle-class people were not able to afford the huge price of onion which is a frequently used commodity in their houses.

The example above shows us that a decision of a farmer regarding which type of crop to grow in his land is generally depends on his intuition and many other factors such as making huge profits within a short period of time, lack of awareness about the demand in the market and when he overestimates a soil's potential to support the growth of a particular type of crop and many more. A wrong decision that is taken on the farmer's side could put a much bigger pressure on the financial condition of his family resulting in severe loss. For all this reason we can see a farmer's pressure regarding which crop should be grown in his land. So now the most important aspect is to design a recommendation system that predicts the type of crop that can be grown in a particular land and thereby helping the farmers. With this aim in mind we have decided to develop a system that takes in the soil parameters like N, P, K (Nitrogen, Phosphorus, Potassium) and the pH values and predicts the most suitable crop that can be grown in that region. The dataset already contains the NPK and pH values of the soil and the appropriate or the most suitable crop that can be grown in given soil.

1.2. MOTIVATION:

The motivation behind developing Crop It was to study application of Machine Learning techniques to solve real-life problems. With the increasing use of technology in the agriculture sector and availability of facilities to test the soil for NPK and pH values, this application aims to help farmers make an informed decision based on Machine Learning algorithms.

The team members have a good grasp over Python programming language and are interested in exploring Machine Learning.We picked Crop It:Crop Prediction Using ML as our Mini Project-A for semester 5. As we researched more about the concept of machine learning, we came across its types, techniques and applications. Hence we took up this concept as the main implementation in our Project – "Crop It". Not only it helped strengthen our knowledge of programming but also helped us in exploring the new field of Machine Learning and its practical applications.

1.3. PROBLEM STATEMENT AND OBJECTIVE:

Problem Statement:

India is an agrarian economy, hence farming and related activities are of great importance. Many times farmers are unaware of which crops are best to grow in their region which leads to excessive use of fertilizers and irrigation. This may lead to degradation of soil and affect profits earned by the farmers. Profitability of a crop majorly depends on weather conditions, yield of the crop and costs of cultivation and production. Economic welfare of farmers depends on not only yield of the crop but also demand for the crop. The application must accept soil parameters (N-P-K and pH) and climate parameters (Temperature, Rainfall and Humidity). Using these parameters, the model should predict the crop that is suitable for the given conditions. The main challenge in this project is to implement, analyze a number of classification algorithms and select the best algorithm for the problem which gives the most accurate results.

Objectives:

The foremost objective is to design a robust machine learning model that would predict the optimal crop for the given soil and climate conditions. The model should be trained with a dataset having adequate information about a reasonable variety of crops over a good range of parameters. The algorithm that we choose to use needs to have high accuracy for the purpose of the project. The selected algorithm must consider all the parameters for choosing the optimal crop.

2. LITERATURE SURVEY:

2.1. SURVEY OF EXISTING SYSTEM.

[1] S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika and J. Nisha, "Crop recommendation system for precision agriculture"
The crops that were considered in the model for prediction include coriander, pulses, cotton, paddy, sorghum, groundnut, sugarcane, banana and vegetables. Different attributes of the soil were considered in order to predict the crop, which included pH, depth, erosion, permeability, texture, drainage, dater holding and soil color. The technique used was ensembling, which combined the power of using two or more different models for better prediction. The ensembling technique used was called the Majority Voting Technique.

[2] R. Kumar, M. P. Singh, P. Kumar and J. P. Singh, "Crop Selection Method to maximize crop yield rate using machine learning technique"
The crops were inspected and graded depending on an examination to estimate crop yielding. This categorisation is found from different data mining algorithms. This paper provides a perception into various grouping rules, such as K-Nearest Neighbour and Naive Bayes. By making use of this document, we evaluated the classification rules and established which all will match the set of data we will be using in our project.

[3] T.R. Lekhaa, "Efficient Crop Yield and Pesticide Prediction for Improving Agricultural Economy using Data Mining Techniques"
The paper hypothesizes analysis of Explorative Data and considers the design of different types of predictive models. A data set is taken as a sample data set, and different regression techniques are tried to recognise and examine each property. Specific regression methods discussed here are Multiple Linear, Linear, Non-Linear, Polynomial, Ridge regression and Logistic. Using this article, we obtain a comparative study of the different algorithms in data analytics. This helped in determining which algorithm is most appropriate to the proposed system.

[4] Viviliya, B. and Vaidhehi, V., "The Design of Hybrid Crop Recommendation System using Machine Learning Algorithms"
The attributes in the dataset included the soil type, groundwater level, rainfall, water availability, temperature of one dataset and the other dataset included the potassium, phosphorus, and nitrogen values, fertilizers, soil pH and organic carbon value. The dataset was preprocessed using basic preprocessing tasks. Naive Bayes and J48 classifiers were used for the crop recommendation. The final recommendation was done using association rules based on the results obtained from the classifiers. The model was trained using 10-cross validation. The testing was done based on different metrics like the Accuracy, ROC Area, Recall, Precision, F-Measure etc.

2.2. LIMITATIONS OF EXISTING SYSTEMS:

The existing system predicts the crop through analysis made on just Climatic attributes

(rainfall, temperature, humidity). This may not give the most accurate result as a number of parameters are not considered. As the application is aimed towards farmers and related people, it is necessary to have a simple yet effective user interface for such users. Various projects use different algorithms for prediction, in this project the team has used Random Forest algorithm.

3. PROPOSED SYSTEM:

3.1. INTRODUCTION:

Following best practices for crop production is essential, but prior to that, the selection of the right crop suitable for the fertile land must be ensured to make the best use of the practices followed. Hence the System focuses on the crop selection as it directly contributes to the crop production rate.The project proposes a model which can predict the crop based on the soil nutrient values (N-P-K values), pH, Rainfall, Temperature and Humidity given as the input. The first step of designing the system is Data Cleaning and Data Analysis. The dataset had information of 22 crops, each crop having 100 instances. While analysing the data, we found that all attributes except Nitrogen contain outliers. We study various algorithms, namely Decision Tree, K-nearest neighbors, Random Forest and Support Vector Machine. Maximizing true positive (TP) and minimizing false negative (FN) is very important which is why it was decided to experiment with several algorithms and choose the best one.

3.2. ARCHITECTURE and FRAMEWORK:

The architecture of the application is divided into two parts. The first part is the Machine Learning part and the second part is the web application.

1. **Machine learning:**
   Machine learning models work very effectively on features that are least correlated. It is important to extract features that are relevant and yet having least information overlap. Relevant and important features are generally known by gaining domain knowledge of the problem that is being solved.

   The ML algorithms are implemented using pandas, numpy and sci-kit learn. The data visualization is powered by seaborn and matplotlib. We implemented Decision Tree, K-nearest neighbors, Random Forest and Support Vector Machine algorithms. After analyzing their accuracies, we concluded that iit is best to use the Random Forest algorithm as its accuracy is **99.3%** which is the highest among all algorithms.

2. **Web application:**
   The Web Application is required for interaction between users and ML model. It is developed using Django framework for back-end and HTML, CSS and JavaScript framework for front-end.

3.3. ALGORITHM AND PROCESS DESIGN:

Data Analysis:

Data Cleaning: Data cleaning is one of the important parts of machine learning. It plays a significant part in building a model. The success or failure of a project relies on proper data cleaning. If we have a well-cleaned dataset, there are chances that we can achieve good results with simple algorithms also, which can prove very beneficial at times especially in terms of computation when the dataset size is large.

Data Visualization: Data visualization is the representation of data or information in a graph, chart, or other visual format. It communicates relationships of the data with images. This is important because it allows trends and patterns to be more easily seen. With the help of data visualization, we can see how the data looks like and what kind of correlation is held by the attributes of data. It is the fastest way to see if the features correspond to the output.

Data Partitioning: The Entire dataset is partitioned into 2 parts, 80% of the dataset is used for training the model and 20% of the data is set aside to test the model.

Machine Learning Algorithms:

Supervised learning: Supervised machine learning algorithms can apply what has been learned in the past to new data using labelled examples. After Sufficient training the system can provide targets for any new input. In order to change the model accordingly the learning algorithm can also differentiate its results with the correct, intended output and find errors.

Random Forest Classifier : Random forest is the most popular and powerful supervised machine learning algorithm capable of performing both classification and regression tasks, that operate by constructing a multitude of decision trees at the time of training and generating outputs of the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. The more trees in a forest the more robust the prediction.

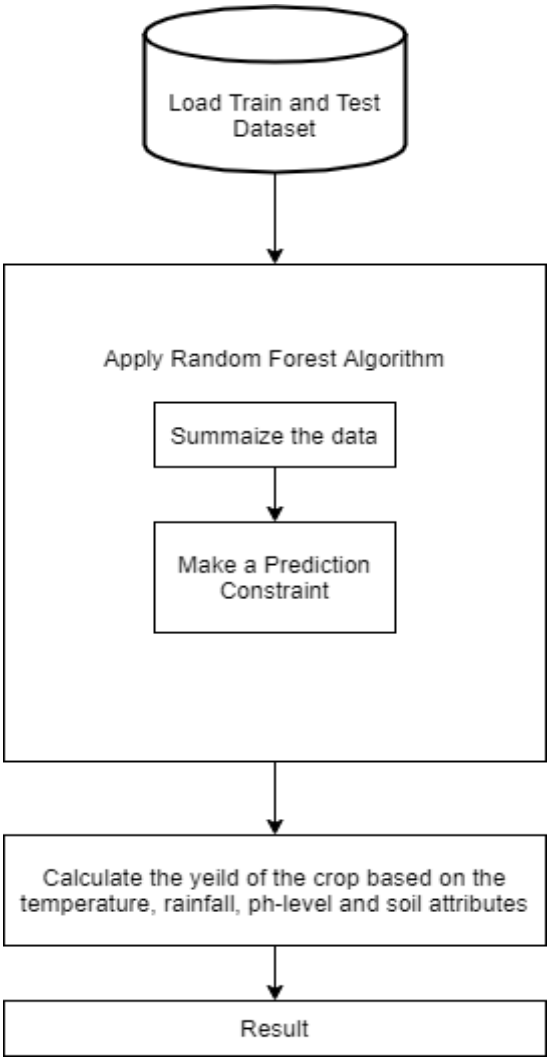

Fig. 1. Flow of Website

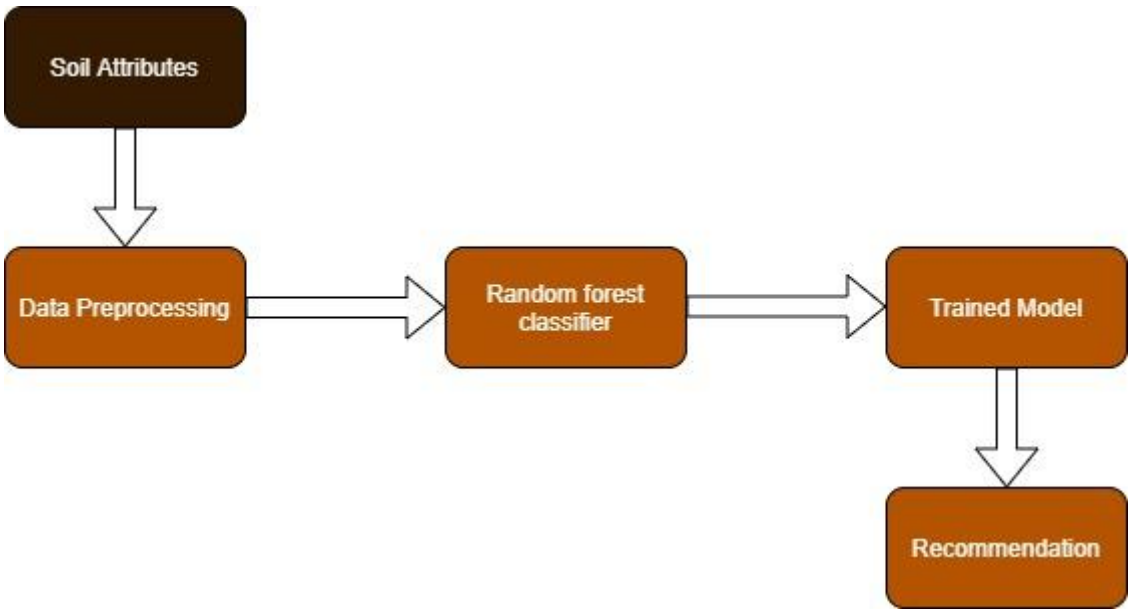

Fig. 2. Flow of Machine Learning model

3.4. DETAILS OF HARDWARE AND SOFTWARE:

Hardware :

- **Memory: 8 GB RAM**
- **Storage: 500 GB internal storage drive**

Software :

- **Django :** Django is a high-level Python web framework that encourages rapid development and clean, pragmatic design
- **Numpy :** NumPy is an open-source numerical Python library. NumPy contains a multi-dimensional array and matrix data structures.
- **Pandas** : Pandas is a software library written for the Python programming language for data manipulation and analysis.
- **Scikit-Learn :** Scikit-Learn provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistent interface in Python.
- **Scipy :** SciPy in Python is an open-source library used for solving mathematical, scientific, engineering, and technical problems.

3.5. EXPERIMENT AND RESULT:

3.5.1. Experimentation:

3.5.1.1. Decision Tree:

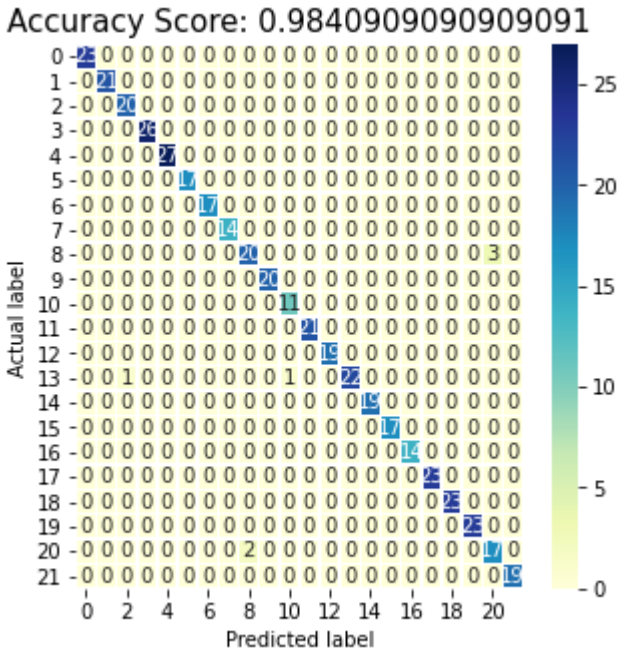

Fig. 3. Heatmap for Decision Tree Algorithm

3.5.1.2. Random Forest (Highest Accuracy):

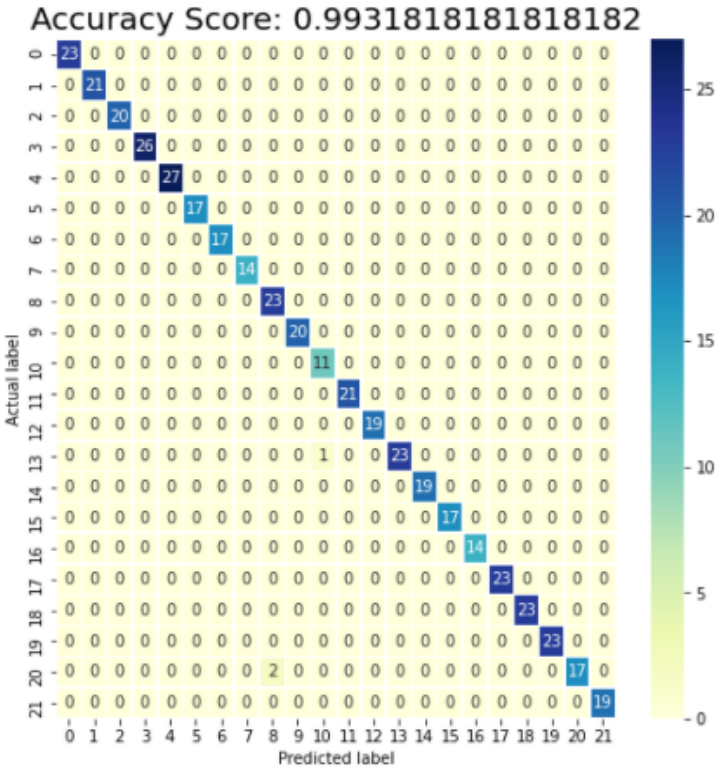

Fig. 4. Heatmap for Random Forest Algorithm

3.5.1.3. Support Vector Machine:

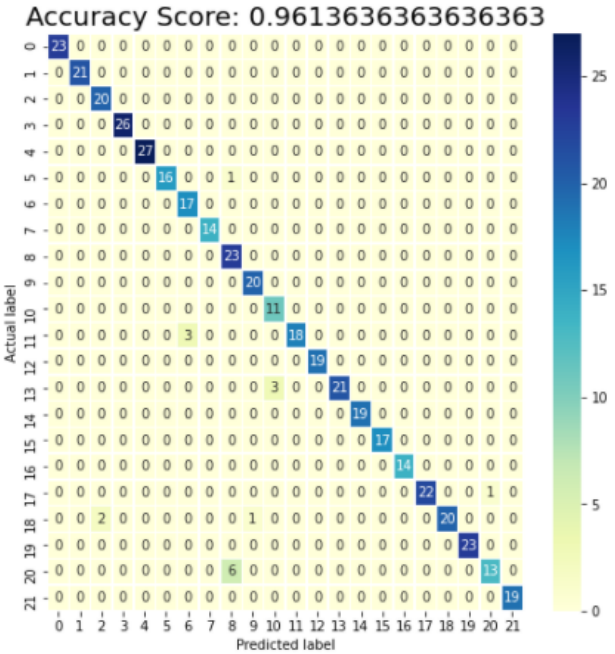

Fig. 5. Heatmap for Support Vector Machine Algorithm

3.5.1.4. K-nearest neighbor:

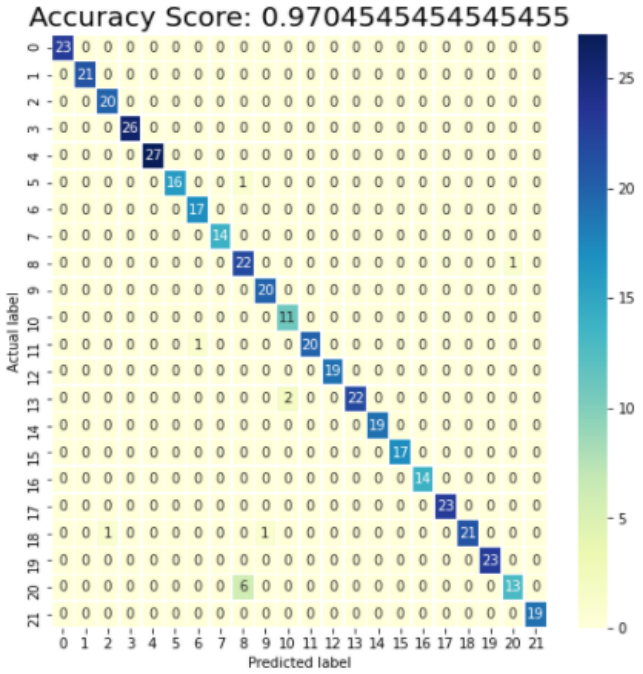

Fig. 6. Heatmap for K-nearest neighbor Algorithm

3.5.2. Web Application:

3.5.2.1. Home Page:

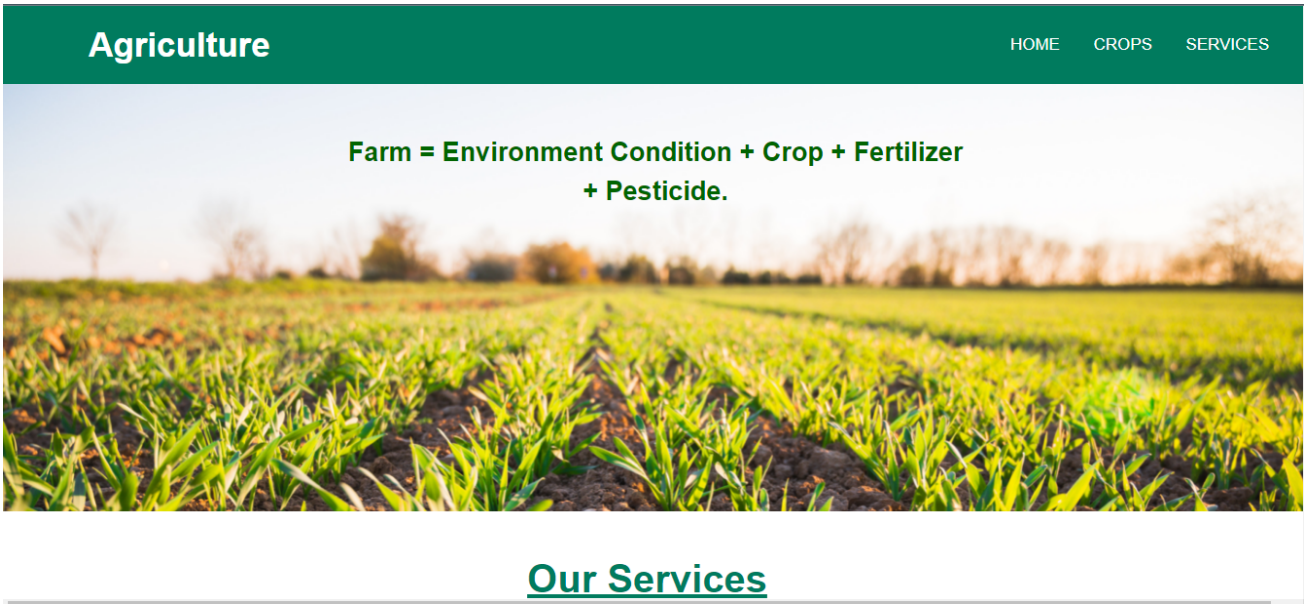

Fig. 7. Screenshot of Home Page

3.5.2.2. Prediction Page:

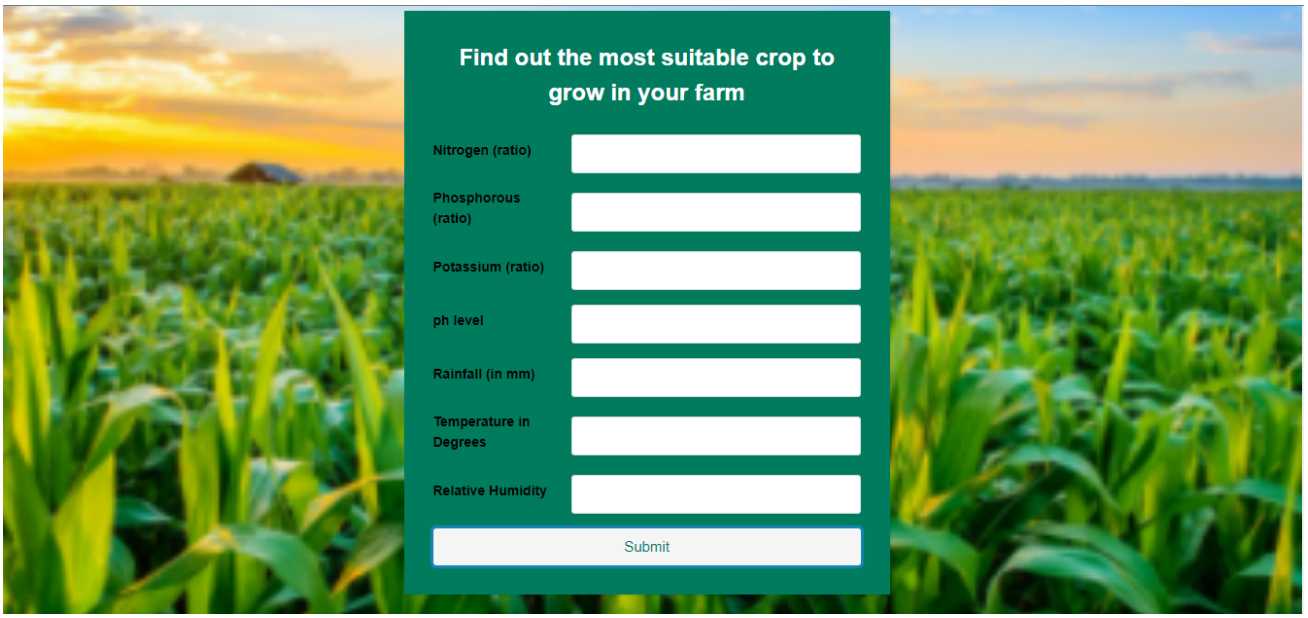

Fig. 8. Screenshot of Prediction Page

3.5.2.3. Crop Output:

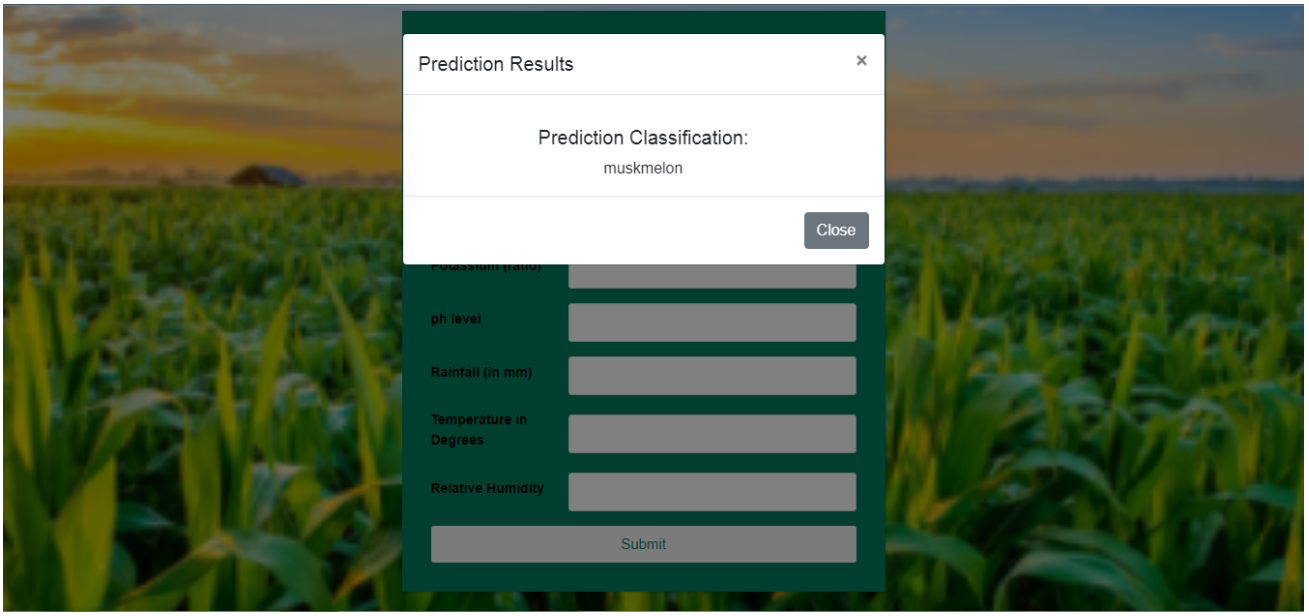

Fig. 9. Screenshot of Crop Output

4. CONCLUSION:

India is a nation in which agriculture plays a prime role.In prosperity of the farmers, prospers the nation. Thus our work would help farmers in sowing the right seed based on soil requirements to increase productivity and acquire profit out of such a technique. The proposed system takes the soil N, P, K, and pH values into consideration and determines which is the most productive crop that can be grown in that suitable soil and climate conditions. This system thus helps the farmer to decide on the maximum profitable crop and also helps in finding new crops that can be cultivated which have not been cultivated till that time by the farmer. By growing the most suitable crop, farmers can prevent soil degradation in cultivated land, and reduce chemical use in crop production and efficient use of water resources.

5. REFERENCES:

- Crop Recommendation System to Maximize Crop Yield using Machine Learning Technique Rohit Kumar Rajak, Ankit Pawar, Mitalee Pendke, Pooja Shinde, Suresh Rathod, Avinash Devare
- Crop Recommendation System through Soil Analysis Using Classification in Machine Learning Dr.A.K.Mariappan, C. Madhumitha, P. Nishitha, S. Nivedhitha
- Crop Plantation Recommendation using Feature Extraction and Machine Learning Techniques, Soumya Sri Attaluri, Nowshath K Batcha, Raheem Mafa
- Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique Nidhi H Kulkarni; G N Srinivasan; B M Sagar; N K Cauvery
- Crop Recommendation System for Precision Agriculture S.Pudumalar, E.Ramanujam
- Crop Recommendation Dataset
- Crop Recommendation System Using Machine Learning for Digital Farming
- FarmEasy: Crop Recommendation for Farmers made easy | by Darshan Gandhi