Crop Yield Prediction based on Indian Agriculture using Machine Learning Hi-Tech Agri

1st Tushar Yadav
Electronics and telecommunication
Symbiosis Institute Of Technology
Pune, India
tushar.yadav.btech2019@sitpune.edu.in

2nd Arnav Wankhade
Electronics and telecommunication
Symbiosis Institute Of Technology
Pune, India
arnav.wankhade.btech2019@sitpune.edu.in

3rd Varun Sarathy
Electronics and telecommunication
Symbiosis Institute Of Technology
Pune, India
varun.sarathy.btech2019@sitpune.edu.in

4th Aditya Vijay Singh
Electronics and telecommunication
Symbiosis Institute Of Technology
Pune, India
aditya.singh.btech2019@sitpune.edu.in

Abstract* — India is an agricultural country with a thriving economy. Agriculture directly employs 70% of the Indian population. The most common issue among young Indian farmers is selecting the appropriate crop depending on soil requirements. They are consequently seeing a marked decline in productivity. Our work offers to assist farmers in determining soil quality by analyzing its many factors and recommending crops based on the results obtained using a data mining approach. To boost the efficiency of the Crop Recommendation System, the system employs the Random Forest Classification algorithm.The technology correlates soil and crop data to forecast a list of crops suitable for the soil as well as details on the nutrients in the soil that are insufficient for the particular crop. The choice of the crop to be sown is therefore left to the user. As a result, the method contributes to the education of inexperienced farmers. The dataset for our crop prediction experiment was obtained from kaggle.com. Crop forecast criteria include location, temperature, humidity, ph, rainfall, N, P, K, and so on. These criteria are now taken into account while producing a given crop on a specific type of soil. To choose the optimum type of crop to be grown, a crop recommender system takes into account the numerous soil factors. This particular recommender system model will take temperature, humidity, and soil moisture content into account.

***Keywords*— crop prediction, soil parameters, crop price prediction, crop yield, effects of climate changes on agriculture, agriculture land grading, efficient use of fertilizers model etc.**

I. INTRODUCTION

In terms of agricultural output, India is among the world's leaders. Despite being a significant producer of agricultural products, India's agricultural productivity remains low. Farmers earn extremely little money due to low farm yield. Productivity must grow in order for farmers to earn more money. To maximize productivity, farmers should understand which crop would be suitable for the particular piece of land. If the correct type of crop is grown on that plot of land, the crop's yield will immediately increase. As a result, crop suggestion systems can be quite valuable to farmers. The growth of crops is impacted by many different things. Temperature, humidity, pH, rainfall, potassium, nitrogen, and phosphorus levels in the soil are all elements that influence yield. Many farmers have no notion which crops should be cultivated in certain areas to maximize production and profit. As a result, in this study, we will demonstrate how machine learning algorithms can be used to estimate agricultural yield and pricing.

Agriculture is India's backbone. As we all know, food is the most basic essential for living; therefore, the agriculture sector should be prioritized in development. The Indian agriculture sector contributes for 18 percent of the country's agriculture gross domestic product (GDP) and employs 50% of the workforce. The primary reason for examining the Agriculture sector is that it is critical to the development of the country's economy. Crop Selection is the area of research for the suggested system since it is the first and most essential phase in the process of agricultural development, and its success assures the outcome of production. Agriculture development provides crop growers with aid through the use of diverse agricultural resources. As a result, it offers high productivity while consuming few resources.

II. LITERATURE REVIEW

1)

Project Title : Crop Yield Prediction and Efficient use of Fertilizers
Author Name : S.Bhanumathi, M.Vineeth and N.Rohit
Year of Publish : 2019
Abstract :

Because India is an agricultural country, its economy is heavily reliant on agricultural yield increases and agroindustry products. Data mining is a new area of study in crop production analysis. In agriculture, yield prediction is a critical issue. Any farmer would like to know how much harvest he may expect. Analyze the different associated attributes such as location, pH value, and alkalinity of the soil. It is utilized in conjunction with third-party apps such as APIs for weather and temperature, kind of soil, nutrient value of the soil in that region, amount of rainfall in that region, and soil composition. All of the data properties will be examined, and the data will be trained using multiple machine learning methods to create a model. The system includes a model that is exact and accurate in predicting crop production and provides the end user with suitable suggestions about required fertilizer ratios based on atmospheric and soil parameters of the land, hence increasing crop yield and farmer revenue.

2)

Project Title : Crop Cultivation Information System on Mobile
Devices
Author Name : Vikas Kumar_, Vishal Dave_, Rohan Nagrani_, Sanjay Chaudhary
Year of Publish : 2017
Abstract :

People utilize mobile devices widely for communication, music, entertainment, the Internet, and social networking. There is a scarcity

of software that can help professionals boost their working capabilities. People in rural areas use mobile phones, yet there are few useful applications to help them enhance their productivity. We proposed and implemented an information system for farmers that can be accessed via mobile phones in this research. To process spatial data and knowledge base, the system is built with Service Oriented Architecture (SOA). Ontologies are used to preserve the knowledge base. The system is an attempt to bridge the knowledge gap between farmers and agricultural professionals. A farmer can give agricultural inputs as well as location-specific information to receive precise ideas, alerts, and recommendations to boost productivity. The knowledge base will be used to generate it. When a farmer notices unusual crop or weather behavior, the system can offer recommendations based on the inputs provided. As part of ongoing work, we have handled some of the queries, and the findings are displayed on Android-based mobile devices for system demonstration.

3)

Project Title : Crop Yield Forecasting Methods
Author Name : 1R.K. Singh, 2T.R. Singh and 2U. Kaushal
Year of Publish : 2019
Abstract :

Increasing the accuracy of agricultural forecasts through the use of earth observation. According to the review study, the capacity to correctly forecast crop output and quality is vital for economic planning and commodity forecasting, as well as guaranteeing global food security. This paper discusses studies on an overview of existing crop yield forecasting methods, including strategies to apply crop yield forecasting methodologies to improve agriculture production forecasting.

4)

Project Title : Crop Prediction
Author Name : Mayank Champaneri, Chaitanya Chandvidkar
Year of Publish : 2020
Abstract :

The influence of climate change in India has had a negative impact on the performance of most agricultural products over the previous two decades. Predicting crop yields ahead of harvest would assist policymakers and farmers in taking proper marketing and storage measures. This research will assist farmers in determining the yield of their crop prior to cultivating on the agricultural field, allowing them to make informed decisions. It makes an attempt to answer the problem by creating a prototype of an interactive prediction system. The machine learning algorithm and an easy-to-use web-based graphic user interface will be implemented in such a system. The farmer will be informed of the outcome of the prediction. Thus, there are various strategies or algorithms for such types of data analytics in crop prediction, and with the help of those algorithms, we may predict crop production. The random forest method is employed.

There is no correct method or technology to overcome the predicament we confront after studying all of these challenges and problems such as weather, temperature, humidity, rainfall, and moisture. There are numerous approaches in India to boost economic growth in the agricultural sector. Data mining can also be used to forecast crop yield production. In general, data mining is the process of examining data from diverse perspectives and distilling it into useful information. Random forest is the most popular and powerful supervised machine learning algorithm, capable of performing both classification and regression tasks. It works by constructing a large number of decision trees during training time and generating class output that is the mode of the classes (classification) or

mean prediction (regression) of the individual trees.

5)

Project Title : Crop yield predictions - high resolution statistical model for intra-season forecasts applied to corn in the US
Author Name : Yiqing Cai, Kristen Moore, Adam Pellegrini, Aymn Elhaddad
Year of Publish : 2017
Abstract :

Accurate crop yield forecasting has far-reaching ramifications for economic trading, food production monitoring, and global food security. The range of environmental variables over time and geography, on the other hand, provides obstacles and limits the precision of yield models. Using data from the United States from 2001 to 2016, we construct a machine-learning-based model for estimating end-of-season, or final, maize yields. To obtain high precision in our intra-season projections, we mix many algorithms in a multi-level model and leverage understanding of physiological processes in temporal feature selection.

6)

Project Title : Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh
Author Name : A.T.M Shakil Ahamed, Navid Tanzeem Mahmood, Nazmul Hossain
Year of Publish : 2015
Abstract :

Agricultural crop production is influenced by a variety of elements including biology, climate, economy, and geography. A variety of factors have varying effects on agriculture, which can be assessed using suitable statistical approaches. Using such procedures and techniques on historical crop yields, farmers and government organizations can acquire information or knowledge that can assist them make better decisions and policies that lead to improved production. The focus of this study is on the use of data mining techniques to extract knowledge from agricultural data in order to estimate crop production for key cereal crops in major districts of Bangladesh.

7)

Project Title : Recommendation System for Crop Identification and Pest Control Technique in Agriculture
Author Name : Avinash Kumar, Sobhangi Sarkar and Chittaranjan Pradhan
Year of Publish : 2019
Abstract :

Most crops are lost in agriculture as a result of incorrect crop selection for a certain land holding. Farmers are frequently unaware of the crop's needs, such as minerals, soil moisture, and other soil requirements. Farmers may experience mental and financial distress as a result of this. Another issue that farmers commonly face is pests and illnesses that can impact the crops they raise and that they are generally ignorant of at an early stage. This farmer dilemma is addressed in our research, and we attempted to solve it using a Recommendation System. We use our model to estimate the optimum crop for the farmer, detect pests that may impact the crop, and provide pest control strategies. In this paper, we used the SVM classification algorithm, the Decision Tree algorithm, and the Logistic Regression algorithm, and we discovered that the SVM classification model outperforms the other algorithms in terms of accuracy.

8)

Project Title : Soil Classification using Machine Learning Methods and Crop Suggestion Based on Soil Series
Author Name : Sk Al Zaminur Rahman, Kaushik Chandra Mitra, S.M. Mohidul Islam
Year of Publish : 2018
Abstract :

Soil is an essential component of agriculture. There are various types of soil. Distinct types of soil have different characteristics, and different types of crops grow on different types of soils. To understand which crops grow best in which soil types, we need to grasp the traits and characteristics of distinct soil types. In this scenario, machine learning approaches may be useful. It has made significant progress in recent years. In agricultural data analysis, machine learning is still an emerging and hard study subject. In this research, we created a model that can predict soil series based on land type and then recommend suitable crops based on the prediction. Soil categorization employs a variety of machine learning methods, including weighted k-Nearest Neighbor (k-NN), Bagged Trees, and Gaussian kernel based Support Vector Machines (SVM). The experimental results reveal that the suggested SVM-based strategy outperforms numerous other methods.

9)

Project Title : Soil Health in Cropping Systems
Author Name : Subhadip Paul, Neha Chatterjee, J. S. Bohra, S. P. Singh
Year of Publish : 2019
Abstract :

Soil health has existed as an integrative feature that indicates the soil's ability to respond to agricultural interference in such a way that it persistently supports both agricultural production and the provision of other ecosystem services. The main conflict in sustainable soil management is to protect ecosystem services while optimizing agricultural production. Soil health is expected to be dependent on the preservation of four primary functions: carbon changes, biogeochemistry-mediated nutrient cycles, soil structure maintenance, and pest and disease control via cropping systems. Each of these activities is identified as a combination of several biological processes carried out by numerous interdependent soil organisms, all of which are governed by the abiotic soil environment, which governs the assessment and maintenance of soil health.

10)

Project Title : Soil Quality for Sustainable Agriculture
Author Name : Duraisamy Vasu, Pramod Tiwary, Padikkal Chandran
Year of Publish : 2020
Abstract :

Being integral to all _unctions o_ terrestrial ecosystem, soil is intended to produce _ood _or _eeding the growing population o_ the world. However, _ood security is _acing threat _rom soil degradation occurring worldwide. Soils degrade due to the exerting pressure _rom various sectors o_ the society including urbanization and industrialization. The major driving _orces o_ soil degradation are de_orestation, change in land use, soil erosion, uncontrolled grazing, waste disposal, and unscientific land management. Globally, 24% (350 lakh km2) o_ the land has degraded which is increasing at the rate o_ 50–100 lakh ha a year−1 and poses threat to the livelihood o_ more than 1500 million people. In this scenario, sustaining soil quality (SQ) is the major challenge to meet the increasing _ood demand. Hence, evaluating and monitoring SQ is crucial to sustain agricultural production and to overcome the vagaries o_ climate change on soil _unctions. However, soil quality per se is complex and site-speci c because o_ the larger variety o_ soil

usage, and its evaluation is di_ cult due to the subjectivity. Nonetheless, soil quality can be quanti ed in the _orm o_ an index _or temporal and spatial comparison o_ various land use and management systems. In this chapter, we discuss the concept and importance of o_ SQ, indicators o_ SQ, minimum data set (MDS) _or evaluating SQ, methods o_ MDS selection, and indexing o_ the soil quality. It will bring out the e__ect o_ soil and crop management practices such as tillage, cropping systems, cover crops, and nutrient management on soil quality and crop production _ocusing in tropical environments. We conclude that principal component analysis is an e__ective method to select MDS _rom a large set o_ soil properties and weighted index method o_ quanti_ying SQ proved to be e_ cient in predicting changes in SQ under various crop production systems. Conservation tillage methods coupled with integrated nutrient management sustains or aggrades the soil quality in di__erent agroecosystems.

III. METHODOLOGY

Dilettante farmers today find it challenging to comprehend crop types, cultivation methods, climatic change, etc. The backbone of every country's economy is agriculture. Dilettante farmers are crucial to the future of agriculture. However, young farmers lack experience, therefore machine learning assists in resolving their issues. The current technique uses soil factors to forecast crop yield and uses machine learning to provide fertilizer. It helps end users choose the crop to be sowed by providing them with information about crop yield. As a result, novice farmers are unable to understand the system.

A specific Tamil Nadu district's major crop production was forecasted using the method. The client must create an account when they first log in to the flask-created Web application. The SQLite database contains the login information. Once a user logs in, they have full access to the system and can anticipate crop production by using inputs like location, nitrogen, phosphorus, potassium, and pH values, which rely on the land's developing environment. By providing the input as the crop name, we can also determine the main nutrients in the soil. The controller receives the various inputs and utilizes Random Forest to classify the data. We recommend to the former how much fertilizer is required in ratio based on soil parameters and the crop price using machine learning techniques.

Module Description:

- Dataset collection
- Implementation
- Prediction

Dataset Collection:

The data for our crop prediction study was obtained from kaggle.com. Following the acquisition of numerous records, data is preprocessed. The dataset includes a greater number of records, some of which have missing values. The dataset has been cleaned up of those missing records, and pre-processing is done on the remaining records. Following the acquisition of numerous records, data is preprocessed. There are 103 records in the dataset overall, 6 of which have some missing values. The remaining 101 records are used for pre-processing after those 6 records were eliminated from the dataset.

Applying machine learning algorithm:

Classification and clustering algorithms are needed by the recommendation system in order to map datasets. Random Forest Algorithm is the suggested system for performing machine learning. Based on an investigation of several algorithms, it was discovered that the Random Forest provided the maximum efficiency and

precision in comparison to Decision tree, Random Forest, etc. As a result, the suggested system uses the Random Forest algorithm to identify the appropriate crop list.

Listing the suitable crops for the soil provided as input:

The technology makes more accurate and effective crop recommendations using a supervised machine learning algorithm. The system provides a list of suitable crops based on the soil, but it is up to the farmers to choose which crop to seed.

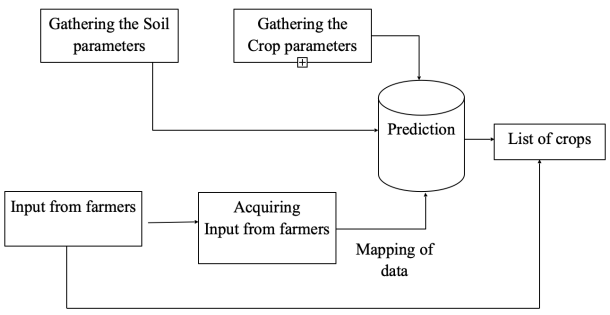
Prediction:

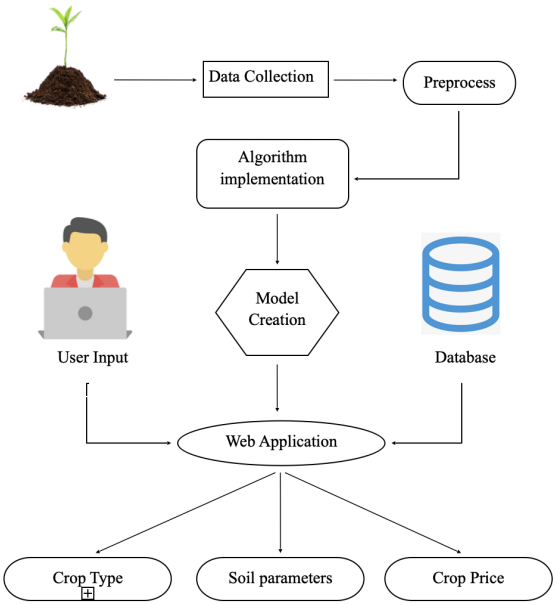
The user's input is added to the training dataset once preprocessed data have been trained. The predicted value is output on a web application after prediction (Flask Framework).

System Feasibility and Requirement Analysis.

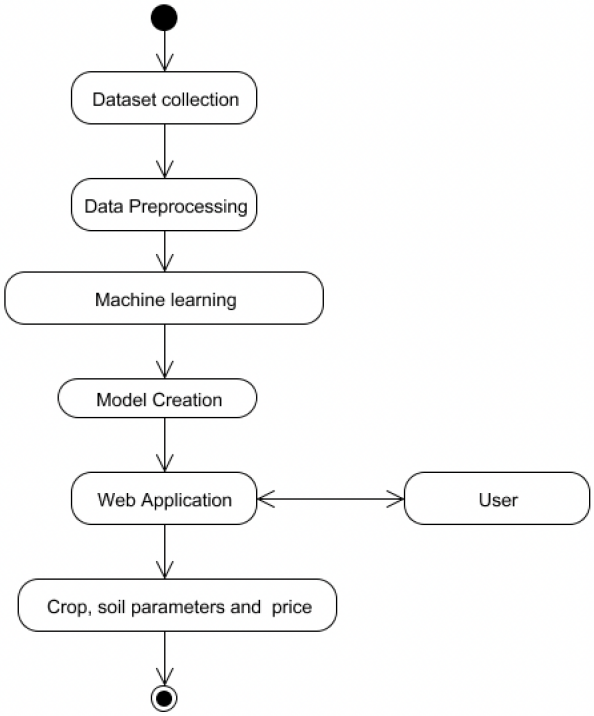
1. Operating System : Windows 7 , 8, 10 (64 bit)
2. Software : Python and Anaconda
3. Framework : Flask, HTML
4. Database : SQLite
5. Hard Disk : 500GB and Above
6. RAM : 4GB and Above
7. Processor : I3 and Above

Architectural Diagram

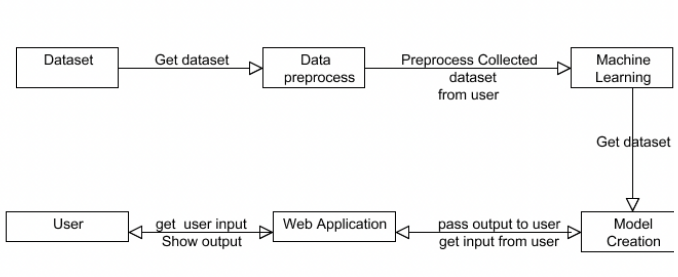




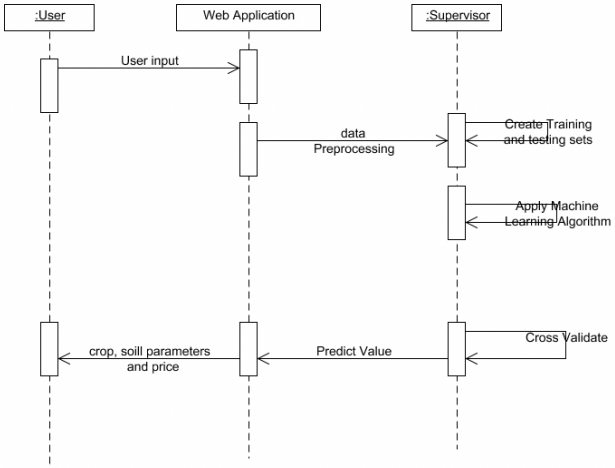
Algorithm



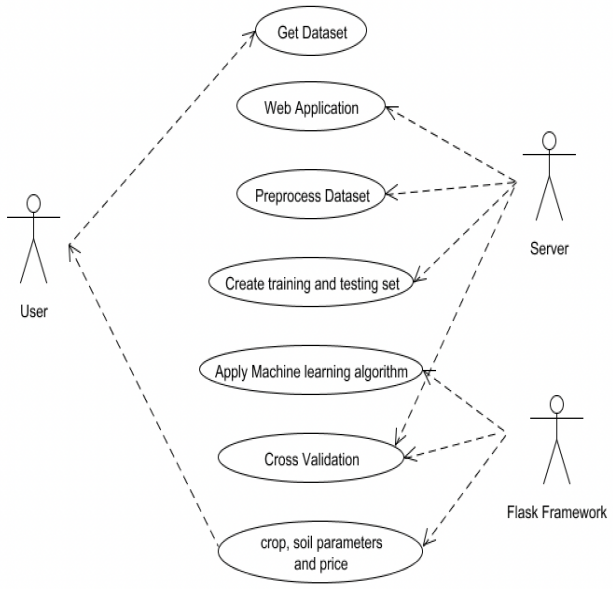
Collaboration



Sequence

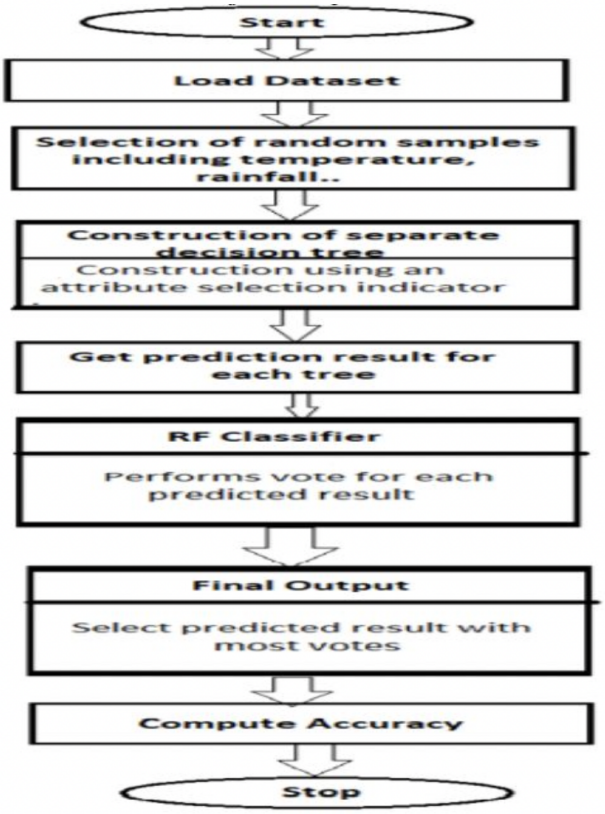


Use Case



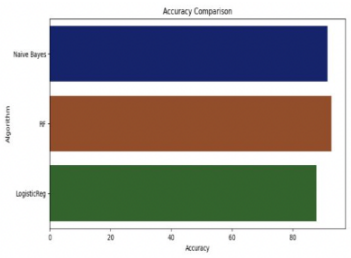
Use of Random Forest Classification

ML-based algorithm for crop prediction. The Because RF incorporates a variety of decision tree algorithms that use the average to improve the forecast in an analysis carried out within various algorithms. the most effective classification algorithms. For the dataset's highest level of accuracy, we are employing Random Forest. As a result, the suggested system uses the RF algorithm to identify the appropriate crop. When compared to Decision trees, it was discovered that Random Forest offered the most efficiency and precision.



Random Forest Model for Crop Prediction

Comparison of random forest, naive bayes and logistic regression.



ALGORITHM	ACCURACY
RANDOM FOREST	92.81407991690006
NAÏVE BAYES	91.49621790098573
LOGISTIC REGRESSION	87.82982929223341

Accuracy Table & comparison of Different Algorithm

IV. FUTURE SCOPE

To collect information about local variety for recommendations: If all the kinds are routinely evaluated at various sites, varieties for which there was no prior data are frequently suggested for those locations. The Attempts to use mixed models (treating some predictors as random variables). Crop illnesses can be found via image processing, and users can upload photos of infected crops to get advice on pesticides. Using a smart irrigation system to automatically adjust watering schedules based on weather and soil conditions, plant water needs, etc.

Here are some additional system features, such as a list of all the new and cutting-edge tools used to grow the predicted best crop and based on location, a list of the various state and federal government policies and programmes that offer financial assistance to farmers, and an examination of how to further develop the system. By utilizing the Internet of Things to connect various farm equipment and sensors to the internet, you may give real-time crop and disease monitoring.

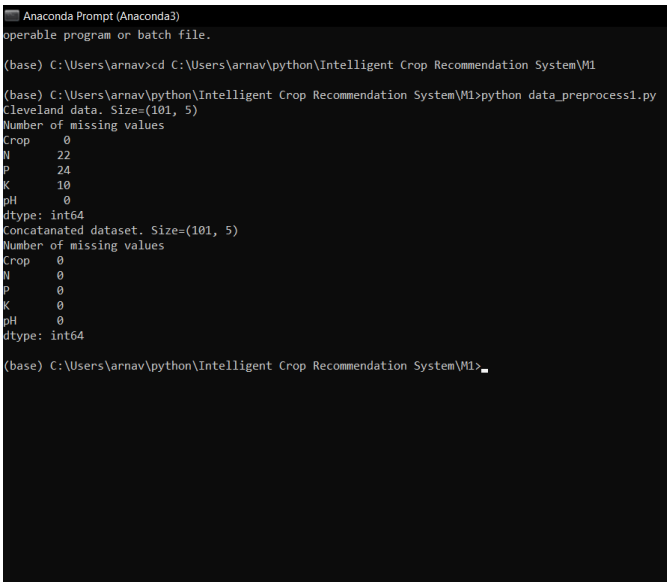
Merits:

1. A deep neural network can reasonably handle some variations in scales due to its deep convolutional layers.
2. It is worth noting that we choose this threshold as smaller than the ImageNet challenge due to the relatively small fruit size with respect to the image resolution

Demerits:

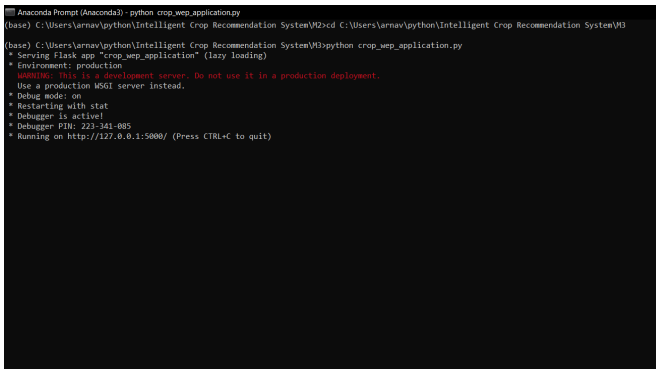
1. If an object size in a testing image is significantly less than that of a training set, it misses the detection.
2. Longer computation time is required.

V. EXPERIMENTAL RESULTS

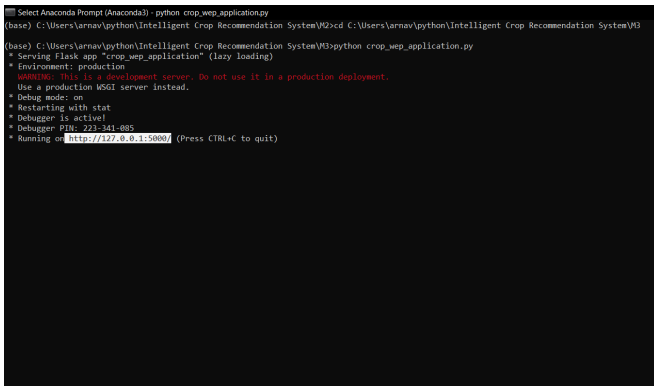


Overview of dataset 1

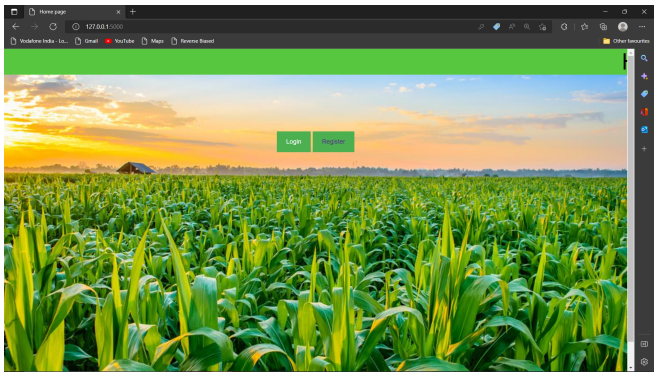
Overview of dataset 2



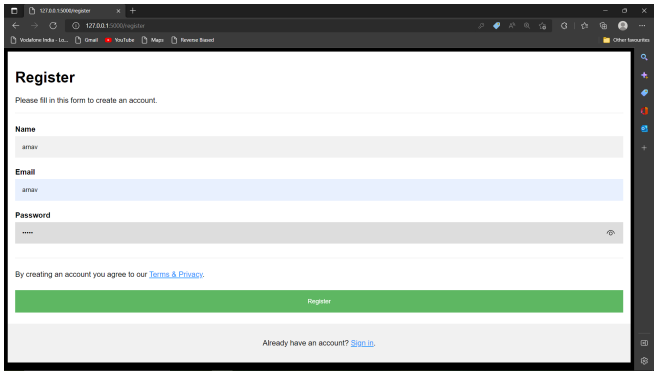
Generating Web Page for further Implementation



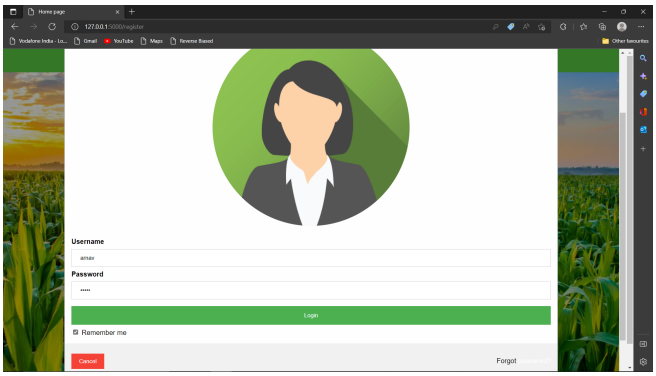
Copying the Website Link Address



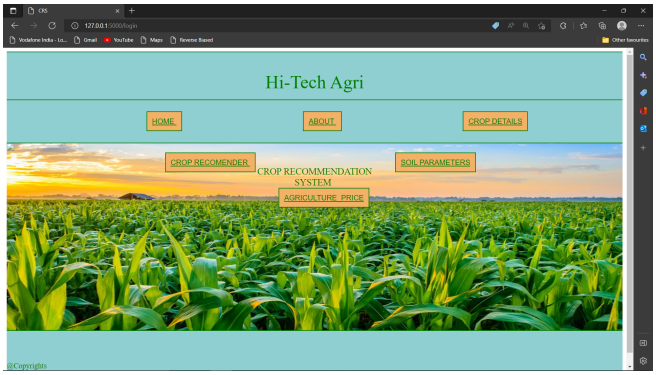
Login/Register Page



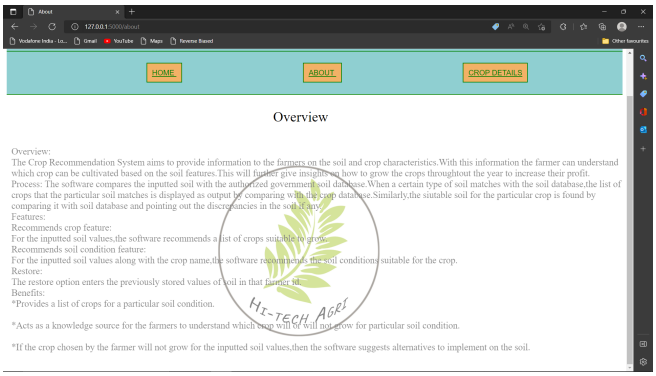
Register Page



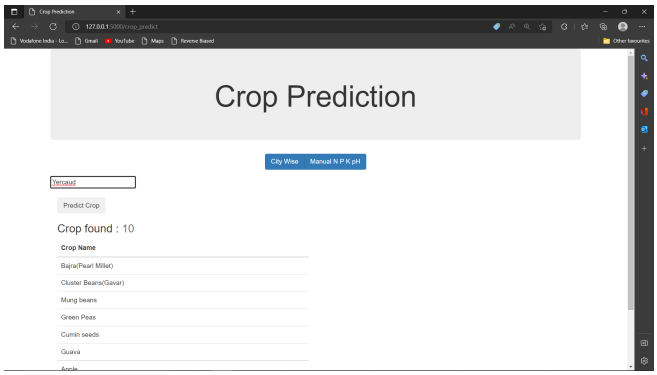
Login Page



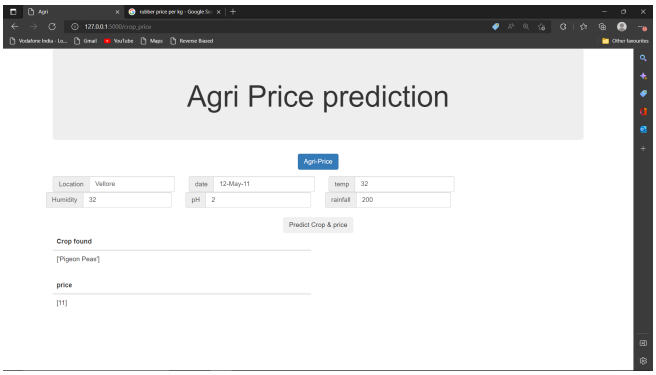
Main Homepage



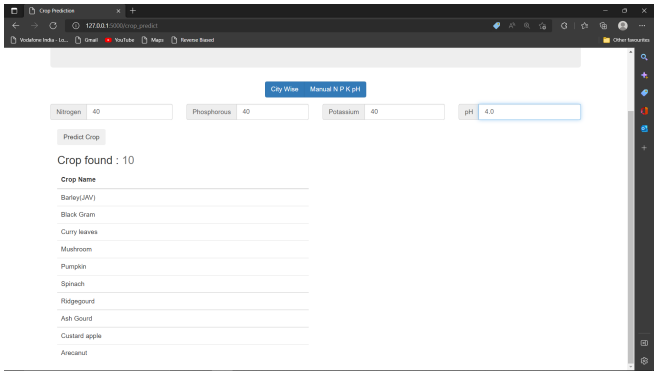
About Page Overview



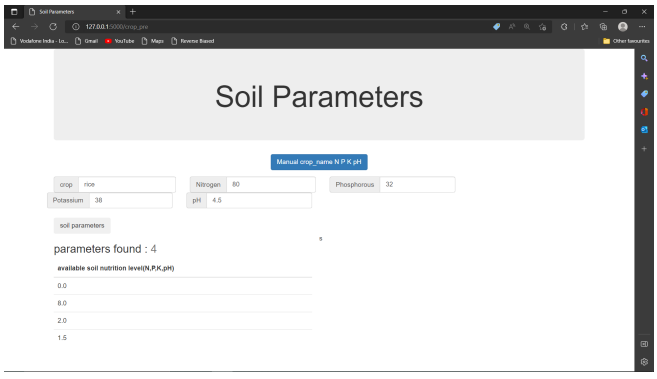
Implementing Crop Prediction by Location Input



Implementing Crop & Price Prediction



Implementing Crop Prediction by giving manual spoil parameters input



Implementing Soil Nutrition Parameters

VI. CONCLUSION

The importance of crop management was extensively addressed in this research. Farmers require modern technologies to help them raise their crops. The features chosen are determined by the dataset's availability and our goal. The developed webpage is user friendly and the accuracy of predictions are above 75 per cent in all the crops and districts selected in the study indicating higher accuracy of prediction. By providing climatic data of that place the user-friendly web page developed for predicting crop yield can be used by any user of their choice of crop. The crop which was predicted by the Random Forest Classifier was mapped to the production of the predicted crop. Then the area entered by the user was divided from the production to get crop yield. Crop name predicted with their respective yield helps farmers to decide the correct time to grow the right crop to yield maximum result.

VII. REFERENCES

1. A CNN-RNN Framework for Crop Yield Prediction Saeed Khaki 1*, Lizhi Wang1 and Sotirios V. Archontoulis 2 1 Industrial and Manufacturing Systems Engineering Department, Iowa State University, Ames, IA, United States, 2 Department of Agronomy, Iowa State University, Ames, IA, United States

2. Snehal S. Dahikar, India Prof. Dr. Sandeep V. Rode An Artificial Neural Network Approach for Agricultural Crop Yield Prediction Based on Various Parameters,International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume 4, Issue 1, January 2019

3. Mayank Champaneri , Darpan Chachpara. Crop Yield Prediction Using Machine Learning. Department of Information Technology, K.J. Somaiya Institute of Engineering and Information Technology, University of Mumbai, India

4. Miss.Snehal S.Dahikar , Dr.Sandeep V.Rode2. Agricultural Crop Yield Prediction Using Artificial Neural Network Approach. INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING Vol. 2, Issue 1, January 2018

5.]D. Ramesh and B. Vardhan, “Analysis of crop yield prediction using data mining techniques”, International Journal of Research in Engineering and Technology, vol. 4, no. 1, pp. 47-473, 2019.

6. Yethiraj N G, ” Applying data mining techniques in the field of Agriculture and allied sciences”, Vol 01, Issue 02, December 2017.

7. Zelu Zia (2018). An Expert System Based on Spatial Data Mining used Decision Trees for Agriculture Land Grading. Second International Conference on Intelligent Computation Technology and Automation. Oct 10-11, China

8. S. Djodiltachoumy, A Model for Prediction of Crop Yield, International Journal of Computational Intelligence and Informatics, March 2017.

9. Nishit Jain, Amit Kumar, Sahil Garud, Vishal Pradhan, Prajakta Kulkarni, Crop Selection Method Based on Various Environmental Factors Using Machine Learning, Feb -2017.

10. J.P. Singh, M.P. Singh, Rakesh Kumar and Prabhat Kumar Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique, International Journal on Engineering Technology, May 2015.

11. Subhadra Mishra, Debahuti Mishra and Gour Hari Santra, Applications of Machine Learning Techniques in Agricultural Crop Production, Indian Journal of Science and Technology, October 2016.

12. Karan deep Kaur, Machine Learning: Applications in Indian Agriculture, International Journal of Advanced Research in Computer and Communication Engineering, April 2016.

13. D.Sindhura, B.Navya Krishna, K.Sai Prasanna Lakshmi,B.Mallikarjun Rao, Dr. J Rajendra Prasad, Effects of Climate Changes on Agriculture International Journal of Advanced Research in Computer Science and Software Engineering, March 2016.

14. Potnuru Sai Nishant1 , Pinapa Sai Venkat2 , Bollu Lakshmi Avinash3 , B. Jabber Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P., India,March 2019.

15. Veenadhari, S., Misra, B., & Singh, C. D. (2014, January). The Machine learning approach for forecasting crop yield prediction supported climatic parameters. In 2014 an International Conference on Computer Communication and Informatics (pp. 1-5). IEEE.

16. T.Giri Babu, Dr.G.Anjan Babu, The Big Data Analytics to Produce Big Results in the Agricultural Sector, in March 2016.

17. D Ramesh , B Vishnu Vardhan, Analysis Of Crop Yield Prediction by Using the Data Mining Techniques in an International Journal of Research in Engineering and Technology, on Jan-2015.

18. Dr. R. Jegadeesan, Ravindar Mogili, Dr. G.Narsimha, Mahesh Nagamalla 2019 ”Rough Set Attribute Reduction with Support Vector Machine for Prediction of Heart Disease” Journal of Emerging Technologies and Innovative Research. Volume.6 Issue. 3 Pages 652-656.

19. Villanueva, M.B., Louella, M., Salenga, M., 2018 Bitter Melon Crop Yield Prediction using Machine Learning Algorithm with techniques. IJACSA) International Journal of Advanced Computer Science.

20. Ananthara, M.G., Arunkumar, T., Hemavathy, R., 2013 CRY- improved crop yield prediction model using bee hive clustering approach for agricultural datasets. In: By Proceedings of the 2013 an International Conference on Pattern Recognition, on Informatics and Mobile Engineering.