Harvestify

**Project Synopsis**

**<Version 1.0>**

Project Work Phase- I (EAI753)

**BACHELOR OF TECHNOLOGY (CSE)**

|  |  |
| --- | --- |
| PROJECT GUIDE:  **MR. Ashish Bishnoi**  **Assistant Professor**  **CCSIT, TMU** | SUBMITTED BY:  **Atishay Jain (TCA1959018)**  **Charu Saxena (TCA1959050)** |
|  |  |
|  |  |

August, 2022



**FACULTY OF ENGINEERING & COMPUTING SCIENCES**

**TEERTHANKER MAHAVEER UNIVERSITY, MORADABAD**

Table of Contents

[1 Project Title 3](#_Toc120925503)

[2 Domain 3](#_Toc120925504)

[3 Problem Statement 3](#_Toc120925505)

[4 Project Description 4](#_Toc120925506)

[4.1 Scope of the Work 5](#_Toc120925507)

[4.2 Project Modules 5](#_Toc120925508)

[5 Implementation Methodology 7](#_Toc120925509)

[5.1 Need of Technical Environment 8](#_Toc120925510)

[5.2 Data Source 8](#_Toc120925511)

[5.3 Data Cleaning 8](#_Toc120925512)

[5.4 Tool 1 Jupyter Lab/Notebook 8](#_Toc120925513)

[5.5 Implementation 9](#_Toc120925514)

[5.6 Evaluation 9](#_Toc120925515)

[6 Technologies to be used 10](#_Toc120925516)

[6.1 Software Platform 10](#_Toc120925517)

[6.2 Hardware Platform 10](#_Toc120925518)

[6.3 Tools 10](#_Toc120925519)

[7 Advantages of this Project 11](#_Toc120925520)

[8 Future Scope and further enhancement of the Project 11](#_Toc120925521)

[9 Team Details 12](#_Toc120925522)

[10 Conclusion 12](#_Toc120925523)

[11 References 13](#_Toc120925524)

# Project Title

Harvestify

# Domain

The domain of this project is to creating a web application using HTML, CSS, JavaScript, and Machine Learning and Deep Learning for model building. The dataset is taken from authorized website Kaggle. We are also using Python Flask Server which basically work between our web application and the machine learning model. For Deployment of this website, we basically use Heroku which basically give a URL to access it over the internet.

# Problem Statement

Currently, agriculture is facing a hideous problem. Where in spite of all right knowledge consumed, an agriculture sector is facing a huge loss. Why? To put simply the crops aren’t supervised properly. classification analysis can help you with finding the right area for your crop, further resulting to control damage and more revenue generation. To be able to successfully yield crops, foremost and major key role is proper irrigation functionality. Machine learning algorithm can help with better irrigation resulting in following ways – maintain a desired soil water range in the root zone that is optimal for plant growth, low labor input for irrigation process management.

Farming is about risk calculation – But what if the risk can be calculated and cured beforehand. Anomaly analysis can help you with identifying the weakness and strength of the soil, resulting in more revenue generation and saving ample amount of time.

“Climate is now a data problem,” says Claire Monteleone. Earlier, Improper weather predictions lead to many crops lost- resulting in loss of money and time invested. But technology has evolved over years leading businesses to higher stable growth. Regression analysis will help you with better production forecasting using weather conditions.

Farming is one of the major sectors that influences a country’s economic growth. In country like India, majority of the population is dependent on agriculture for their livelihood. Many new technologies, such as Machine Learning and deep Learning, are being implemented into agriculture so that it is easier for farmers to grow and maximize the yield.

In this project, I present a website in which the following applications are implemented; Crop recommendation, Fertilizer recommendation and plant disease prediction, respectively.

* In the crop recommendation application, the user can provide the soil data from their side and the application will predict which crop should the user grow.
* For the Fertilizer recommendation application, the user can input the soil data and the type of the crop they are growing, and the application will predict what the soil lacks or has excess of and will recommend improvements.
* For the last application, that is the plant disease prediction application, the user can input an image of a diseased plant leaf, and the application will predict what disease it is and will also give a little background about the disease and suggestion to cure it.
* The works done till now only concentrated on crop prediction using different soil properties and Data Mining Techniques. Fertilizer Recommendation is not taken into consideration. So, it is necessary to develop crop yield prediction and fertilizer recommendation system which predicts crop yield based on soil nutrients crop yield data and recommend fertilizer for selected crop based on different datasets like fertilizer data, location data and crop yield data.

# Project Description

The aim of proposed system is to help farmers to cultivate crop for better yield. The crops selected in this work are based on important crops from selected location. The selected crops are Rice, Jowar, Wheat, Soyabean, and Sunflower, Cotton, Sugarcane, Tobacco, Onion, Dry Chili etc. The dataset of crop yield is collected from last 5 years from different sources.

There are 3 steps in proposed work.

* Soil Classification: Soil classification can be done using soil nutrients data. Two Machine learning algorithms used for soil classification are Random Forest and Support Vector Machine. The two algorithms will classify, and display confusion matrix, Precision, Recall, f1-score and average values, and at the end accuracy in percentage as output.
* Crop Yield Prediction: Crop Yield Prediction can be done using crop yield data, nutrients and location data. These inputs are passed to Random Forest and Support Vector Machine algorithms. These algorithms will predict crop based on present inputs.
* Fertilizer Recommendation: Fertilizer Recommendation can be done using fertilizer data, crop and location data. In this part suitable crops and required fertilizer for each crop is recommended.
* Third Party applications are used to display Weather information, Temperature information as well as Humidity, Atmospheric Pressure and overall description.

## Scope of the Work

Harvestify is important to drive farming efficiency. As earlier the soil erosion, Manures, Fertilizer and Biocides, etc. is less used by the farmers but now these products will increase the nitrogen, phosphorus, potassium and also increases the ph. level of the soil. So, this project will help those farmers who are not aware from it so they simply give their soil details, and also upload their crop their image for best crop recommendation, fertilizer and disease prediction with prevention and solution.

## Project Modules

1. **Data Cleaning:**

It is the very first module of this module. In this firstly we download the dataset of crop prediction, fertilizer prediction and disease prediction from Kaggle. We load this the CSV file in our model and see the head to know how the data look like and group it by area type. Data cleaning process is start with handling the NA values in which we use is null function. In this we have used different function like drop NA values, unique value function so that we can differentiate between them and lambda function.

1. **Feature Engineering:**

This module is also called as dimensionality reduction. The main motive of this model is used remove the redundant features or the error from the give dataset.

In feature engineering we create new feature which is helpful in future for outlier detection and removal in later stage. In our dataset location is the categorial feature which cause problem if we have so many locations at the same time. It creates problem in searching the best crop, fertilizer.

1. **Outlier Detection:**

Outlier are the data points which are the data errors or sometimes they are not the errors but they represent the extreme variations in the dataset. So, it is better to remove them on time otherwise it creates problem in further processing.

For removing the outlier, we can use different technique like standard deviation or we can use simple or domain knowledge. In this we set some threshold value and compare these values by using standard deviation method.

1. **Model Building:**

In this module, firstly we train different-different machine learning algorithm model on the same data and check their accuracy and select the best fit algorithm which give high accuracy result and prediction as a crop recommendation algorithm and use for further process.

For Fertilizer Recommendation we use the same algorithm which is used by crop recommendation as they need same data as an input like Nitrogen, Potassium, Phosphorus, Ph value, rainfall in mm, location to predict the crop and fertilizer.

For Disease Prediction which is a part of image classification we use the CNN and RS-Net to process the image and predict the result for this we use GPU to train our model for this GPU is not provided on every platform to train the model again and again so we train the model only once’s and save that model for the future uses.

1. **Application Module:**

This module is divided in three parts i.e., Crop Recommendation, Fertilizer Recommendation and Plant Diseases Prediction. For crop recommendation we take several inputs of soil testing and taking location as an input and through weather API we fetch the rainfall condition in that area and recommend the best crop based upon the input. For Fertilizer recommendation we take soil testing result values as an input and the crop that you want to grow and then recommend the fertilizer. For Diseases it will take image as an input and then classify the result and predict the diseases and crop and the prevention method and solutions.

# Implementation Methodology

A framework for training a Machine Learning Model in real time is shown in figure 1.

Data Collection

Data Preprocessing

Remove Outliers

Training Dataset

Jupyter Lab

Decision Tree Classifier

Random Forest

Support Vector Machine

XGboost

Gaussian Naïve Bayes

CNN and RS-Net

Model Building

Predicted Value

The main aim of this project to be implemented is to recommend best fit crop and fertilizer and predict the disease via image classification and give the prevention and solution method. Here step by step process involved is represented below.

1. Scientific Environment
2. Source of Data
3. Excel 2016: The first process to store the data
4. Loading data into notebook.
5. Normalizing of data
6. Detecting outliers
7. Machine Learning model are built, and various algorithm used for predicting as listed in methodologies.
8. Split the data sets as testing and train data for cross validator process.
9. Fitting the data into different machine learning for prediction.
10. Finding the Error to find the accuracy percentage.
11. Visualization: Making a web page and deploy it so that everyone can use it and for user friendly environment.

## Need of Technical Environment

1. Microsoft Excel.
2. Jupyter Notebook/Google Collab for creating the Script.
3. Essential Libraries to be installed.
4. Regressing and Classification
5. Appropriate Function to be selected for algorithm
6. Web Designing.

## Data Source

As I mentioned before the main data sets were collected from the Kaggle machine learning repository which is open data resource for the data mining and for the predictive analytics purposes. The acquired data source was a text file for finding the errors in the data source, the text file is connected to the EXCEL and they are separated using the commas and saved as a CSV file.

## Data Cleaning

The data in the CSV files need to be checked whether it has any missing values, as the data source have missing values every attribute have checked using the filters and null values are removed which helps to increase the accuracy level.

## Tool 1 Jupyter Lab/Notebook

The Crop and fertilizer data is processed with the help of python program in Jupyter Notebook and the numerical data is normalized using the script so the data will equally distributed and variables cannot dominate each other. It helps to bring the values of the attribute on a common scale.

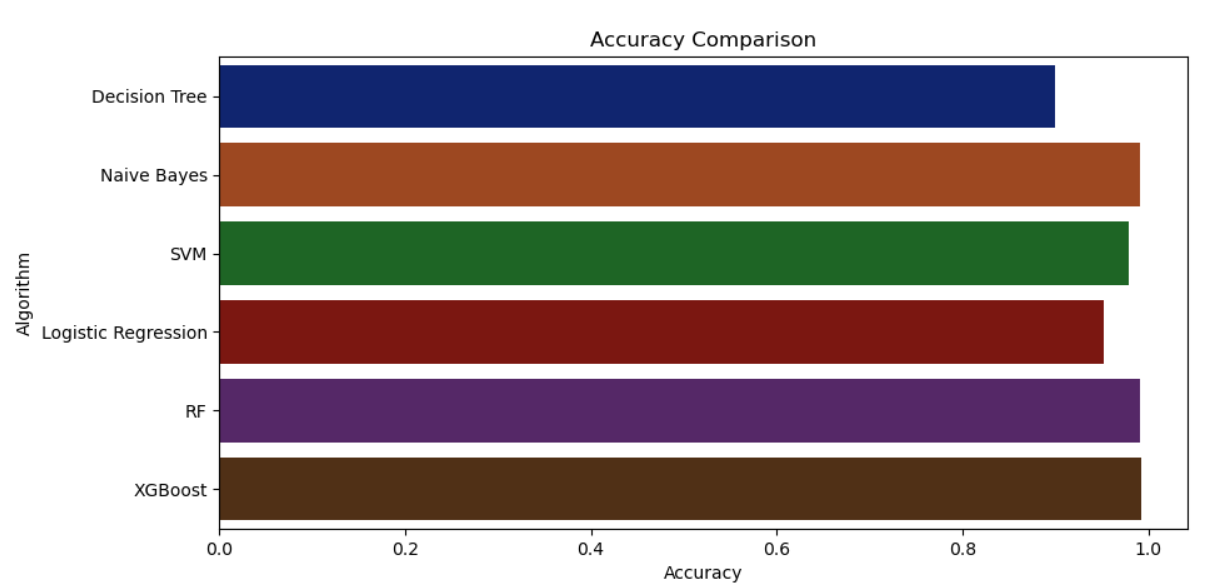
## Implementation

The data is been split into test and train 75 percent data is used for the test data and remaining for the train data. Consider the standard recursive partitioning algorithm will start it searches all the data and in-depth search is made for all the variables and the best split point is taken and the process gets repeated for the right and left leaves. The method has been followed for different assumption here the dependent variable is values of the soil and the location and attribute which is going to be predicted and the other attributes are independent variables and the attributes are given into the model and with the help of coefficients an equation has been made for regression and the predicted prices are produced.

## Evaluation

The idea of a regression is to predict a real value which means number in regression model we can compute the several values the most common terms are explained below.

As we train different-different algorithm model based on their accuracy for same data the best fit algorithm with the high accuracy is selected for the recommendation and analysis.



# Technologies to be used

## Software Platform

1. **Front-end**

* HTML (Hyper Text Markup Language)
* JavaScript
* CSS (Cascading Style Sheet)
* Bootstrap

1. **Back-end**

* Python
* TensorFlow
* Flask
* Heroku
* Machine Learning Model
* Deep Learning Model

## Hardware Platform

* RAM
* Hard Disk
* OS
* Editor
* Browser, etc.

## Tools

|  |  |  |
| --- | --- | --- |
| **Name** | **Version** | **Purpose** |
| Python | 3.7 | Primary Programming Languages |
| Google Collab | 1.0.0 | Editor or IDLE |
| JavaScript | 13th Edition | For Making responsive |
| Sklearn | .21.0 | Machine Learning Libraries |
| Matplotlib | 3.0 | Used for graph and plotting |
| Pandas | 1.3.5 | Data Analysis Tool |
| Bootstrap | 5.2 | Web Development tool |
| Pytorch | 1.21.1 | Used for Deep learning neural network |
| TourchVision | 0.13.1 | Neural Network |
| Pickle | 3.5 | For Saving Model in file/data structure |

# Advantages of this Project

* The proposed model predicts the crop yield for the data sets of the given region. Integrating agriculture and ML will contribute to more enhancements in the agriculture sector by increasing the yields and optimizing the resources involved. The data from previous years are the key elements in forecasting current performance.
* The proposed system uses recommender system to suggest the right time for using fertilizers.
* The methods in the proposed system includes increasing the yield of crops, real-time analysis of crops, selecting efficient parameters, making smarter decisions and getting better yield.

# Future Scope and further enhancement of the Project

* More data can be collected manually via web scrapping to make the system more accurate.
* Additional plant images can be collected to make the disease detection part more robust and generalized.

# Team Details

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Project Name** | **Course Name** | **Student ID** | **Student Name** | **Role** | **Signature** |
| Harvestify | Project work phase- I  (EAI753) | TCA1959018 | Atishay Jain | Designer and Tester |  |
| TCA1959050 | Charu Saxena |  |

# Conclusion

The system uses supervised and unsupervised Machine learning algorithms and gives best result based on accuracy. The results of the two algorithms will be compared and the one giving the best and accurate output will be selected. Thus, the system will help reduce the difficulties faced by the farmers and stop them from attempting suicides. It will act as a medium to provide the farmers efficient information required to get high yield and thus maximize profits which in turn will reduce the suicide rates and lessen his difficulties.

The prediction of crop yield based on location and proper implementation of algorithms have proved that the higher crop yield can be achieved. From above work I conclude that for soil classification Random Forest is good with accuracy 86.35% compare to Support Vector Machine. For crop yield prediction Support Vector Machine is good with accuracy 99.47% compare to Random Forest algorithm. The work can be extended further to add following functionality. Web application can be built to help farmers by uploading image of farms. Crop diseases detection using image processing in which user get pesticides based on disease images. Implement Smart Irrigation System for farms to get higher yield.

We have worked on a sample dataset from Kaggle which has taken into consideration records obtained from a broad agricultural demography. Farmers generally use hit and trial method which leads to wastage of land and resources or even disproportionate growth of crops. We are trying to break all such taxing walls by providing them with an accurate and justified model made by machine learning using random forest classifier to identify the correct crop to be grown in their farms. This will help them in improving their crop production both qualitatively and quantitatively. This will also help them to maintain the quality and nutrition contents of the soil

# References

* <https://www.youtube.com/>
* <http://cs229.stanford.edu/>
* <https://ieeexplore.ieee.org/abstract/document/8697639>
* <https://linuxhint.com/>
* [https://towardsdatascience.com/](https://towardsdatascience.com/predicting-house-prices-with-linear-regression-machine-learning-from-scratch-part-ii-47a0238aeac1)
* [https://www.kaggle.com/](https://www.kaggle.com/code/ashydv/housing-price-prediction-linear-regression)
* <https://studygyaan.com/data-science-ml/>
* <https://www.jetir.org/papers/JETIR2110302.pdf>
* Bhagat, N., Mohokar, A., & Mane, S. (2016). International Journal of Computer Applications, 152(2), 23–26.
* N. N. Ghosalkar and S. N. Dhage,” 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-5, Doi: 10.1109/ICCUBEA.2018.8697639.
* C. R. Madhuri, G. Anuradha, and M. V. Pujitha, A Comparative Study,” 2019 International Conference on Smart Structures and Systems (ICSSS), Chennai, India, 2019, pp. 1-5, Doi: 10.1109/ICSSS.2019.8882834
* T. D. Phan: The Case of Melbourne City, Australia,” 2018 International Conference on Machine Learning and Data Engineering (iCMLDE), Sydney, Australia, 2018, pp. 35-42, Doi: 10.1109/iCMLDE.2018.00017.
* Asia: Survey. (2020). Press Trust of India. https://www.business-standard.com