Minor Project

Domain – Machine Learning

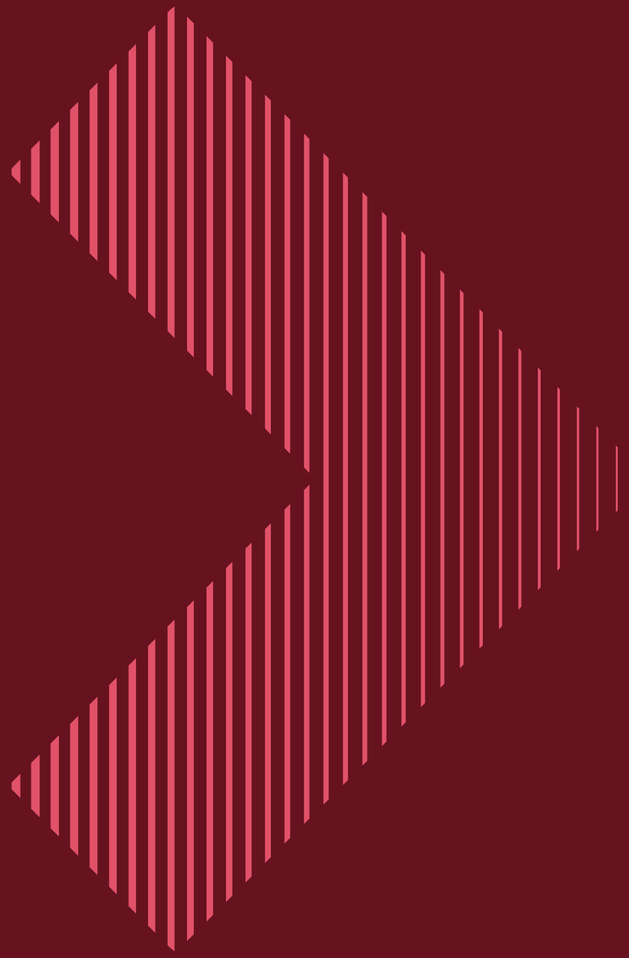
Intelligent Crop Recommendation System

Guided by–
Dr. Ambika Prasad Mishra

Overview

- Introduction
- Problem Statement
- Literature Review
- Methodology
- Results
- Conclusion and Future Scope

Introduction



- The agriculture industry is crucial to India's economy, but there is a need for more scientific and precise approaches to improve efficiency and productivity.
- The project aims to address this need by offering a crop recommendation system based on the area and soil type.
- This can help farmers make better decisions for the type of crop to be grown.
- These features can benefit both novices and experts in the field, and ultimately contribute to the growth and modernization of India's agriculture sector.

Problem Statement

Intelligent Crop Recommendation System

Objectives

- Collect agricultural datasets with soil chemistry and location-related information for different crops.
- Build a crop predictor system using Random Forest algorithm and soil chemistry dataset.
- Perform dataset analysis on various ML models for better recommendations.

Literature Review

The project addresses to the important issue of crop recommendation systems

- It has the potential to revolutionize India's agriculture industry by providing personalized and optimized crop suggestions to farmers.

The dataset has been collected from four Indian states for analyzing crops.

- Used dataset of four states: Uttar Pradesh, Bihar, Jharkhand and Madhya Pradesh.

Decision tree and random forest classifier ML algorithms are used.

- These algorithms are capable of handling high-dimensional datasets and can provide accurate predictions of crop yields under different environmental conditions,

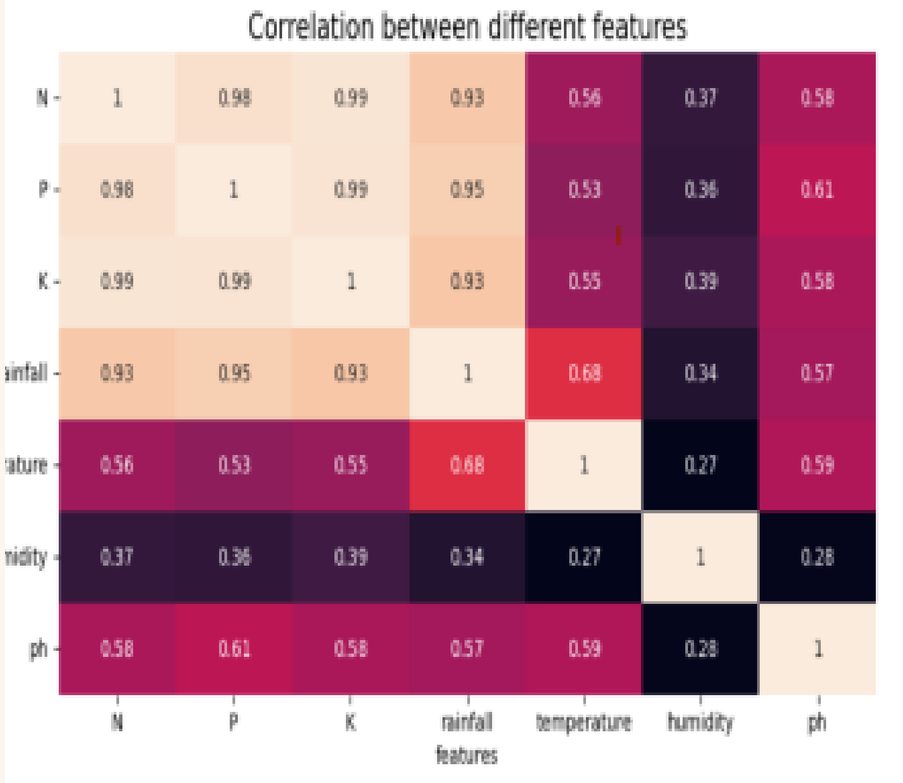
It helps in better crop productivity and enables the farmers to earn profit .

- Soil parameters can help farmers make more informed decisions about crop management practices, leading to higher yields and increased profitability.

Summarize Details of Dataset

Number of samples in test and train data	Total Samples :- 391 Train Data- 311 Test Data- 81
Number of Features	8 Features- (Nitrogen, Potassium, Phosphorous, Rainfall, Tempearture, Humidity, Label)
Target Features	CROP(Label)

Methodology (EDA)



We imported necessary Python libraries for data handling and visualization purposes such as Pandas, Numpy, Matplotlib, and Sklearn.

The dataset was loaded into a Pandas data frame and examined to understand the values assigned to each independent and dependent feature.

We scrutinized the dataset for any NULL values and determined the datatype of each column to avoid errors during data analysis.

We checked for balanced classes to enhance model performance using various sampling methods.

A heatmap was plotted for features to identify variables that had high correlations with each other.

Correlation between features were reduced to decrease model complexity and improve its accuracy.

Methodology

```
[38] rt = RandomForestClassifier()
     rt.fit(Xtrain,Ytrain)
     rtpred = rt.predict(X1test)
```

```
[39] r1 = metrics.accuracy_score(Y1test, rtpred)
     print("Random forest's Accuracy is: ", r1)
```

```
Random forest's Accuracy is:  0.9247311827956989
```

Looking at the dataset we first saw that that there are multiple algorithm that can be used for our model such as Logistic Regression, Naive Bayes, Decision Tress, Random Forest.

But after the analysis of our Dataset, we find out that:-

We might have chosen logistic regression if we were interested in predicting a binary outcome, such as whether a particular crop will grow well or not.

Decision trees may suffer from overfitting if the model is too complex, leading to poor generalization on new data.

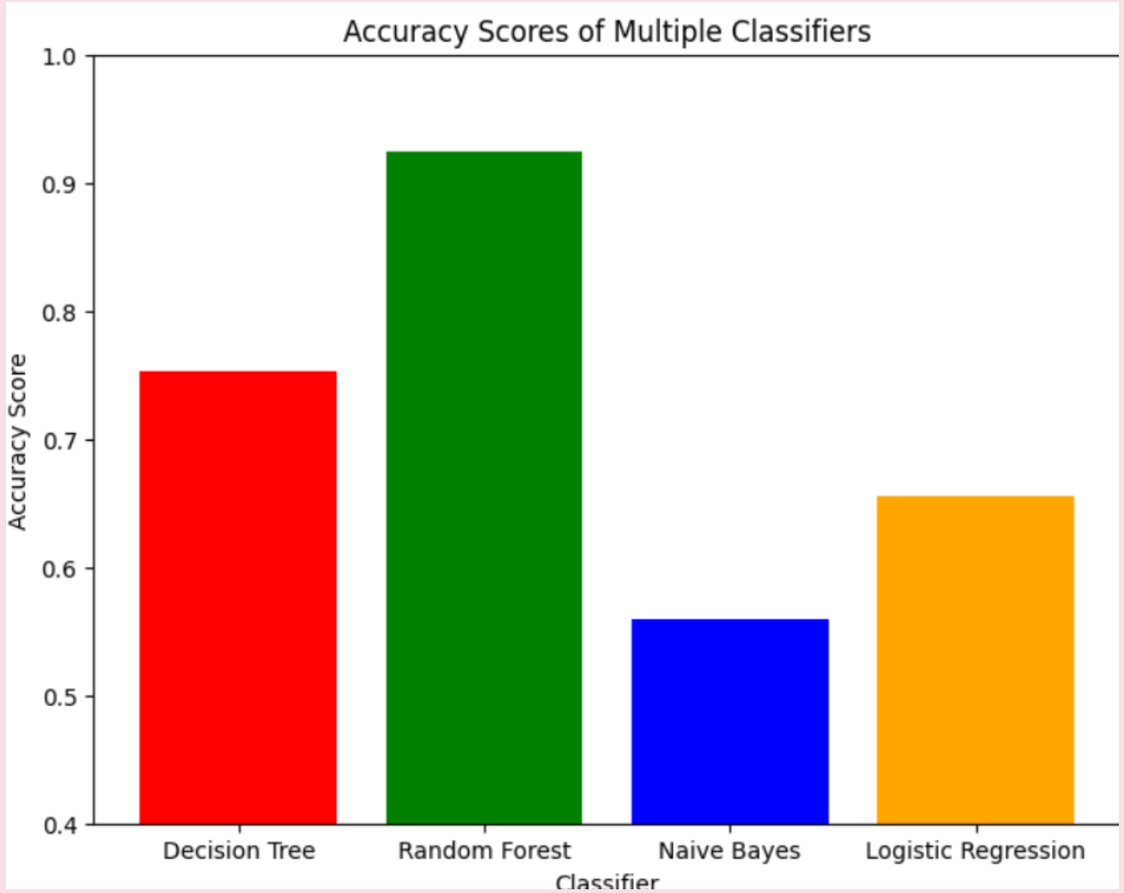
We might have chosen naive Bayes because it is a simple and fast algorithm that can handle large amounts of data. However, we have rejected it as features were not independent of each other.

We trained each model using the dataset of 3 states-Bihar, Uttar Pradesh and Jharkhand and tested our model using the datset of Madhya Pradesh.

After carefully analyzing our Dataset with Random Forest Algorithm, we recieved the highest accuracy(92.47%).Random forest is an ensemble learning algorithm that combines multiple decision trees to improve the accuracy of predictions. Random forest works well when there is a large number of features and the data is noisy. It can handle both categorical and continuous data and is less prone to overfitting than decision trees. Random forest is also computationally efficient and can handle large datasets.

Results

After evaluating several machine learning models in our crop recommendation system, we found that the random forest algorithm gave the highest accuracy rate in predicting crop yields. Despite testing other models, including naive bayes, logistic regression, and decision tree, the random forest model was the most effective approach. Although these models showed promising results, they were not as accurate as the random forest model. Therefore, we concluded that the random forest model is the most suitable for our recommendation system with the accuracy of 92.47%.



Conclusion

Data Gathering	• Based on the data gathered from four states, it can be concluded that specific crop recommendations would depend on various factors.
Data Preprocessing	• The dataset needs to be cleaned and normalized to remove any outliers or inconsistencies before performing crop recommendation analysis.
Data Analysis	• After analyzing the dataset, it can be concluded that specific crops are suited for particular regions based on various factors such as soil type, climate conditions and availability of resources.
Random Forest Classifier	• The random forest classifier works by creating multiple decision trees and combining their predictions to produce a final output, which can be used to recommend the most suitable crop for a given set of environmental parameters.
Accuracy	• The model gives an accuracy of 92.4731 %
Crop Recommendation	• The recommended crop can empower farmers in India by providing personalized and data-driven suggestions for optimal crops to grow based on their specific environmental conditions.

Future Scope

Presently our farmers are not effectively using technology and analysis, so there may be a chance of wrong selection of crop for cultivation that will reduce their income.

To reduce those type of loses we can developed a farmer friendly system with GUI, that will predict which would be the best suitable crop for particular land and this system will also provide information about required nutrients to add up, required seeds for cultivation, expected yield and market price. So, this makes the farmers to take right decision in selecting the crop for cultivation such that agricultural sector will be developed by innovative idea.

Group Members

Chandrika Papneja	Debashree Chakraborty	Aditi Singh
20051001	20051139	20051629