Prediction System of Crops with Machine Learning

Saimanoj Bera [1] Mythresh Neerugattu [1] Varshitha Sannareddy [1]

[1]University of Michigan-Dearborn

Overview of Project

This project is to help farmers with potential crop selection. Few Machine learning models are applied to past research data of various features(soil types, climatic conditions) to identify the best crop type to increase the yield of crops. After performing different models, the best model is evaluated based on accuracy and gives the best results.

The description of research data consists of the count and mean of all values in each factor. Figure that contains Data Description

	nitrogen	phosphorus	pottasium	temperature	humidity	ph	rainfall
count	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000
mean	50.551818	53.362727	48.149091	25.616244	71.481779	6.469480	103.463655
std	36.917334	32.985883	50.647931	5.063749	22.263812	0.773938	54.958389
min	0.000000	5.000000	5.000000	8.825675	14.258040	3.504752	20.211267
25%	21.000000	28.000000	20.000000	22.769375	60.261953	5.971693	64.551686
50%	37.000000	51.000000	32.000000	25.598693	80.473146	6.425045	94.867624
75%	84.250000	68.000000	49.000000	28.561654	89.948771	6.923643	124.267508
max	140.000000	145.000000	205.000000	43.675493	99.981876	9.935091	298.560117

Figure 1. Data Description of crop data

Introduction

The agricultural business experiences numerous difficulties during the farming process because it is mostly impacted by the climatic conditions of its surrounds. Instead of taking into account factors like soil quality, sustainability, etc. when sowing crops, farmers usually base their decisions on the possible market value and financial returns. Thus, they perceive a big decrease in output. Precision agriculture is a cutting-edge farming method that makes use of research data on soil types, characteristics, and crop yields to recommend the best crop to farmers based on site-specific factors. As a result, crop selection errors are decreased and crop productivity is increased Building an intelligent system in this project will help farmers choose the best crop to produce based on the sowing season, the location of their farms, and the qualities of their soil

Research Questions

1. Is soil type parameter a best significant for crop recommendation ?
2. Can we visualise data in the best way and can understand it ?
3. Can we build a efficient model to predict the crop based on the soil type and climatic conditions ?
4. Which Classification gives the best accuracy ?

Understanding the Data

After Performing the basic checking we came to know that label is the only categorical column in the data set and the data is clean as there are no null values and outliers in the data set.

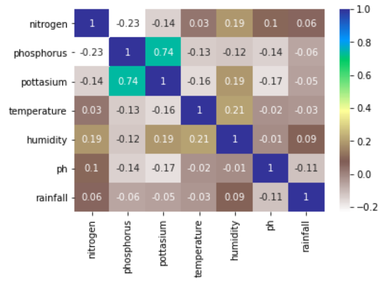
From the data we also came to know that there are 22 different crops in the data so based on the parameters like soil features and climatic conditions we can predict the suitable crop among 22 different crops.

After understanding data we just renamed the columns for better understanding the data.

Then we encoded the label column which is in object datatype into numerical

Data Visualization

Below images are the some of the visualizations of our project .



(a) Correlation of each attribute

(b) Cat plot of nitrogen, phosphorus and potassium with soil

Finding best predictors and Model scaling

After visualization we followed the following procedures.

1. By applying the Ordinary least squares method we get the significant predictors.
2. The significant predictors are nitrogen,phosphorus,pottasium,ph and humidity
3. Then we performed Model scaling using minmax scaler we do this model scaling for better accuracy.
4. Later we divided data into two parts one part has 75 percent of data and other part has 25 percent of data. So the data which has 75 percent of data is used for training of data and the other is used for testing the train model

Equation of Multiple regression model

After finding the best predictors the equation of the best model is

$$Y = \beta_0 + \beta_1 \cdot x1 + \beta_2 \cdot x2 + \beta_3 \cdot x3 + \beta_4 \cdot x4 + \beta_5 \cdot x5 + \epsilon$$

OLS Equation for best predictors is

$$Y = 16.9995 + -0.0305 \cdot nitrogen + -0.0888 \cdot phosphorus + -0.0098 \cdot pottasium$$
$$+0.0530 \cdot humidity + -0.5786 \cdot ph$$

Models used in our project

In our project we have used different models they are

- KNN Algorithm,
- Decision Tree,
- Random Forest,
- Naive Bayes,
- XG boost and
- Support Vector Machine,

Best two models among all the models

Among all the 8 models, Random forest and XG boost are more efficient models as they are giving 95 percentage of accuracy

Random Forest :

The classification system known as the random forest uses several decision trees. In order to try to develop an uncorrelated forest of trees whose forecast by committee is more accurate than that of any individual tree, it employs bagging and feature randomness when generating each individual tree.

XG Boost :

Gradient boosting is a supervised learning process that combines the predictions of a number of weaker, simpler models to attempt to properly predict a target variable. The gradient boosted trees approach is widely used and well implemented in XG Boost.

In XG boost calculation is quicker and the precision is more precise than with Random Forest because the gradient of the data is taken into account for each tree.

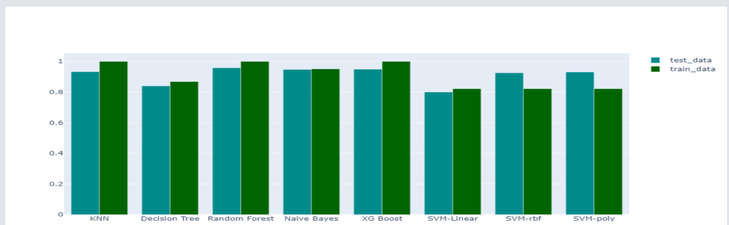


Figure 3. Training and Testing Accuracies of each model

Scores of each models performed on the data

Below tabel shows the Scores of all the models

Models	Scores of models
KNN Model	0.910455
Decission Tree	0.865909
Random Forest	0.954091
Naive Bayes	0.945455
XG boost	0.955909
SVM linear	0.920455
SVM rbf	0.888182
SVM poly	0.908636

Table 1. Cross Validation Scores of all the models.

From this scores we can say that XG boost model gives the best accuracy so this is most efficient model among all the models

References

[1] Pradeepa Bandara, Thilini Weerasooriya, Ruchirawya T.H., W.J.M. Nanayakkara, Dimantha M.A.C, and Pabasara M.G.P. Crop recommendation system. *International Journal of Computer Applications*, 175(22):22–25, 10 2020.

[2] SHILPA MANGESH PANDE, PREM KUMAR RAMESH, ANMOL ANMOL, B. R AISHWARYA, KARUNA ROHILLA, and KUMAR SHAURYA. Crop recommender system using machine learning approach. pages 1066–1071, 2021.