1. Dataset: - There are three different datasets are used for the prediction of the model. i) Soil Dataset ii) Crop Dataset iii) Yield Dataset
2. Soil Dataset: It consists of 15 attributes like PH, EC, OC, OM, N, P, K, Zn, Fe, Cu, Mn, Sand, Silt, Clay, CaCo3, and CEC. From this all attributes we classified and analyzed by applying ML model whether the Soil is fertile or not.
3. Crop Dataset: It consists of 4 attributes like temperature, humidity, PH, rainfall. Crop Prediction is performed by using a different algorithm. iii) Yield Dataset: It consists of 6 attributes are Nitrogen (N), Phosphorous (P), Potassium (K), Organic Care (Og), PH, temp. Yield Prediction is performed on these different attributes using an ML algorithm.
4. Method and Experimentation: - Datasets are collected from different resources then classified the data and group into two sets: -
5. Training Dataset,
6. Testing Dataset

a) Implementation using ML algorithm: Different Supervised algorithm and compared the result and accuracy with the models.

1. Classifying if the soil is Fertile or not:

1. In this model, we take various attributes of the soil such as Ph., EC, OC, OM, N, P, K, etc. to determine if the soil is fertile or not.
2. To classify, we have first taken SVC and the accuracy has come around 0.82.
3. Next, we tried Multinomial Naïve Baye’s model and the accuracy has come around 0.815.
4. Finally, we tried Random Forest and the accuracy came up to 1 with an F-1 score of 1. This is no ordinary accomplishment, because we may get high Accuracy for imbalanced datasets, but in this case, the F-1 score along with accuracy is 1 implying the best possible model to classify if the soil is fertile or not.

Challenges

Challenges are the major basis which imminent the negative impacts on current project. Some of the challenges faced during crop yield prediction are:

* Choosing appropriate dataset, after choosing dataset tuning of the parameters which makes project more efficient to get the desired results.
* Model must be trained by taking consideration of less computational efficiency and power.
* Increase of error rate due to dynamically changing the environment.