Ensembling Technique For Crop Recommendation system

Ensemble Learning

Ensemble learning improves the predictive performance of the model than from any single machine learning algorithm

The model can be trained on different subsets of data .boosting is used to convert weak learners into keen learners

The majority voting technique dictates the class of the output in the ensemble technique

Bagging is one of its techniques

Boosting is used to convert weak learners into keen learners

Stacking combines multiple classification models via Meta-classifier

Dataset collection

We combined our datasets

https://github.com/haritha1313/AgriTech/blob/master/v0/cropdata.csv

https://github.com/haritha1313/AgriTech/blob/master/data/cost-of-cultivation.csv

https://github.com/haritha1313/AgriTech/blob/master/data/state-profit-data.csv

https://github.com/haritha1313/AgriTech/blob/master/data/state-yield-data.csv

For recent data

www.data.gov.in

www.dacnet.nic.in

www.agricoop.nic.in

http://planningcommission.gov.in/sectors/agri_html/DataBank.html

The first dataset contains four parameters month,soil type ,state and the crop

The second data set contains the cost of cultivation of a particular crop

The third dataset contain the profit of the crop according to the staste

The fourth dataset contains the staste yield according to crop and staste

After merging all the datasets, the final datasets contains the following parameters

Soil: Type of soil that can be alluvial soil,black soil, and red soil

State: the state where the crop is intended to sow

Month: the month in which we would like to sow the crop,it would generally climatic factor like rainfall and temperature

Cost of cultivation: the estimated cost of cultivation that a farmer could afford in the units of Rs/acre

Modal price : the estimated price of the crop that a farmer could expect in units of Rs/Quintal

Profit: the profit that a farmer could expect in Rs/acre

Ensemble model

Artificial neural network

Support vector machines

Decision tree

K nearest neighbors

Preprocessing of data

The dataset contains the string as input data for month ,location, and type of soil ,so we have to transform them into numerical labels.

Preprocess techniques such as label encoding ,normalization transformation, and dropping the non prime attributes were applied to improve the model performance

Splitting into training and testing data

The dataset is divided into training and testing data with split ratio 80:20. In the model

Applying machine learning algorithm on the dataset

Training the models using the novel parameter selection method

Ensemble the learners

All the above (slide 8) four models are coupled

The majority class label technique should be use for making the label of class