Crop Yield Prediction based on Indian Agriculture using Machine Learning

This project predicts the yield of almost all kinds of crops that are planted in India.In India, there are more than a hundred crops planted around the whole country. These crops are categorized for better understanding and visualization. The data for this project has been acquired from the Indian Government Repository

Link for the data set:

<https://drive.google.com/file/d/1betvWRQ3Xi31KBupLZpToH04BUIO4HCW/view?usp=sharing>

The data consists of attributes – State, District, Crop, Season, Year, Area and Production with around 2.5 Lakh observations

We used advanced regression techniques – Lasso, ENet and Kernel Ridge and further we used stacking of these models to minimize the error and to obtain better predictions.

Pre-processing of data

For the given data set, there are quite a few ‘NA’ values which are filtered in python. Furthermore, as the data set consists of numeric data, we used robust scaling, which is quite similar to normalization, but it instead uses the interquartile range whereas normalization is something which normalization shrinks the data in terms of 0 to 1.

Stacked Regression:

This is a kind of ensembling but a little of enhancement of averaging. In this, we add a meta model and use the out of fold predictions of the other models used to train the main meta model. Step-1: the total training set is again divided into two different sets. (train and holdout) Step-2: train the selected base models with first part (train). Step-3: Test them with the second part. (holdout) Step-4: Now, the predictions obtained from test part are inputs to the train higher level learner called meta-model.

Step-1: the total training set is again divided into two different sets. (train and holdout)

Step-2: train the selected base models with first part (train).

Step-3: Test them with the second part. (holdout)

Step-4: Now, the predictions obtained from test part are inputs to the train higher level learner called meta-model.

Output:

The performance metric used in this project is Root mean square error. When the models applied individually, for ENet it was around 4%, Lasso had an error about 2%, Kernel Ridge was about 1% and finally after stacking it was less than 1%. The user or the farmer can enter the following details over the web application to get the prediction

We have created a webpage in html ,which the farmer can give the inputs and predictions can occur

Graphical user interface

Description automatically generated

Graphical user interface

Description automatically generated

Conclusion:

When we apply stacked regression, the result has been so improvised than when those models were applied individually. The output which has been shown in figure is currently a web application

By this project the farmer can himself predict the crop of the yield and can avoid the losses which occur due to unpredicted weather condition.