Agriculture Project Blog

Agriculture is the main supporting sector of Indian economy and most of the rural population’s livelihood depends on it. Mostly, machine learning techniques are used in crop management processes, following with farming conditions management and livestock management. In farming, they are used to predict yield and quality of crops as well as livestock production.

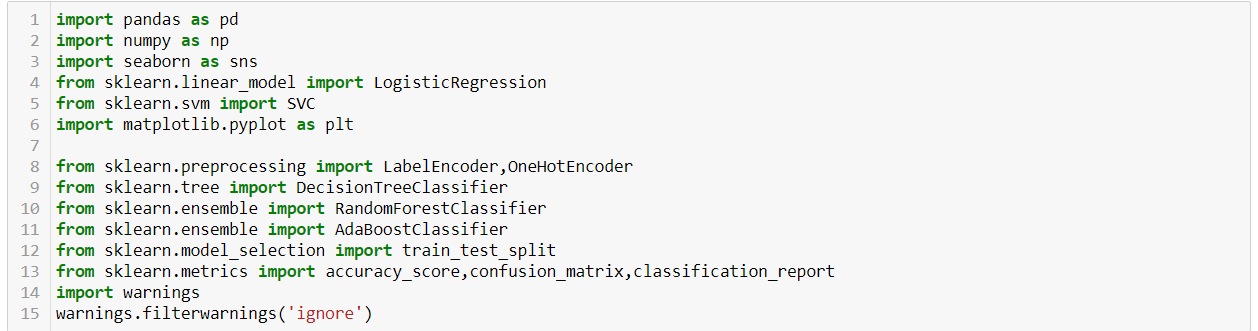
1.Problem Definition:

Agriculture requires the knowledge of right crop, Pesticide, soil quality, etc.  Further, the crop growth also depends on soil parameters, which varies due to ground contamination. I have a dataset which crop type, Soil type, Season etc. In this Dataset I n**eed to determine the outcome of the harvest season, i.e. whether the crop would be healthy (alive), damaged by pesticides or damaged by other reasons.**

2.Data Analysis:

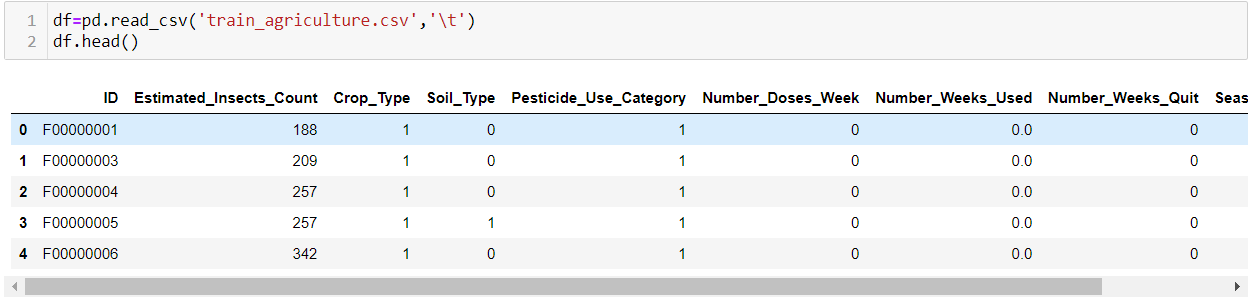
The dataset contains 4599 rows and 10 columns, Where I see the uniqueID, Estimated insect count, crop type, Soil type. The Soil yield also depends on the Number of doses per week and Seasons affecting the growth of the harvest.

I am starting by importing the libraries which I will require to do the EDA process, prediction, visualization and all the findings from the dataset. Instead of adding the libraries when required adding them in the beginning ease the process of EDA.



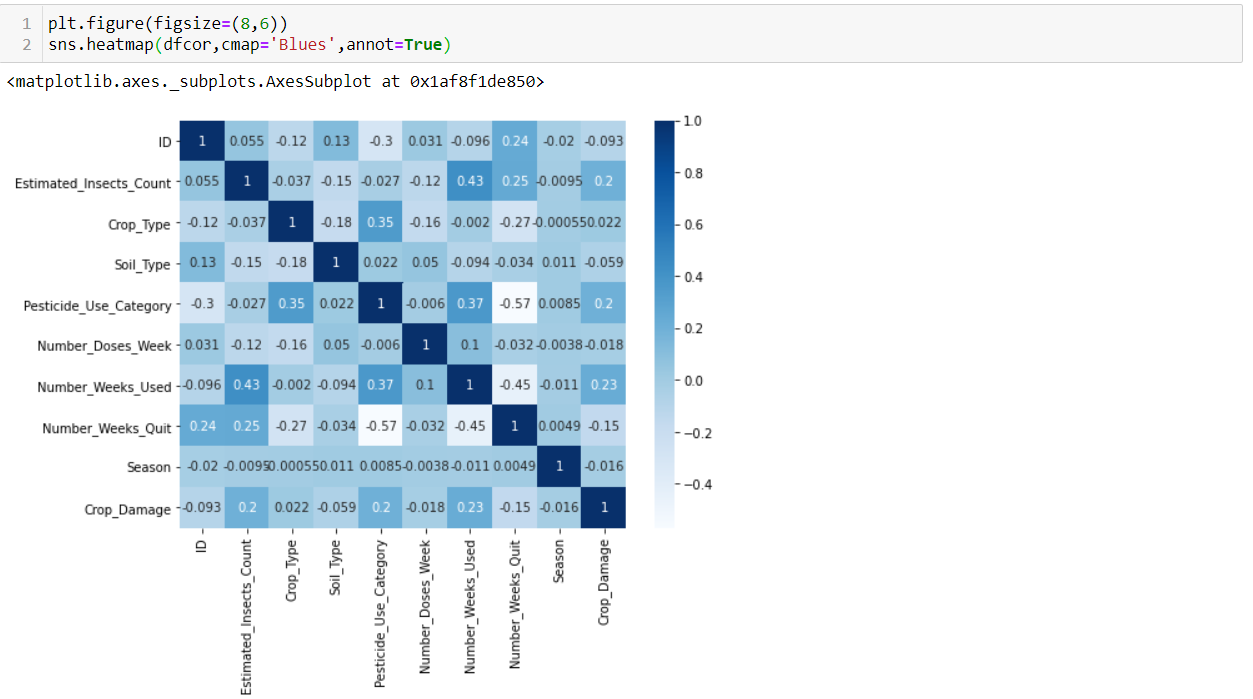
Loading the Dataset into the Variable:

I could see that all the dataset has interger datatype except “number\_of\_weeks” used, hence I applied the label encoder bringing all the columns into one type. The dataset shows that “number\_of\_weeks” has 442 null values, hence adding them with mean values.



3.Exploratory Data Analysis concluding Remarks:

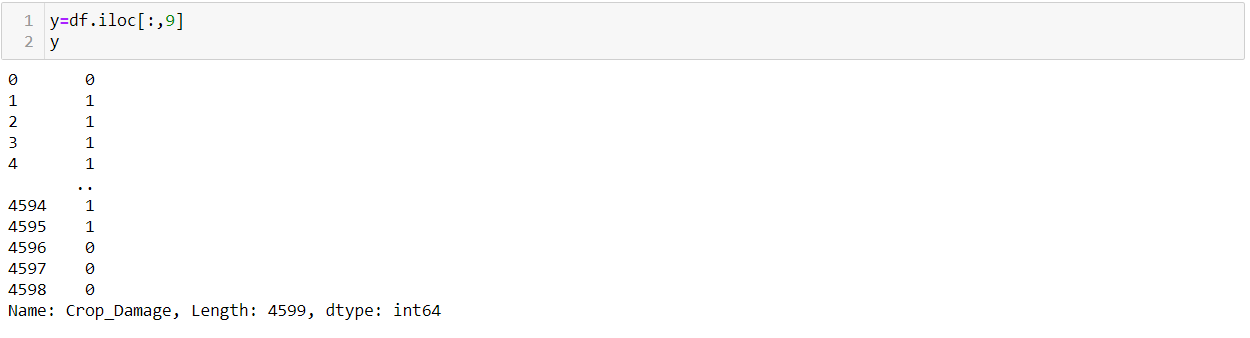
The dataset now has no null values and has same datatype, now I could proceed with the EDA process. Heatmap shows the co-relation between the columns, where Number\_Doses\_Week and Number\_Weeks\_Used are having a good co-relation. Pesticide\_USe\_category and Number\_Weeks\_Quit are having a bad Co-relation.

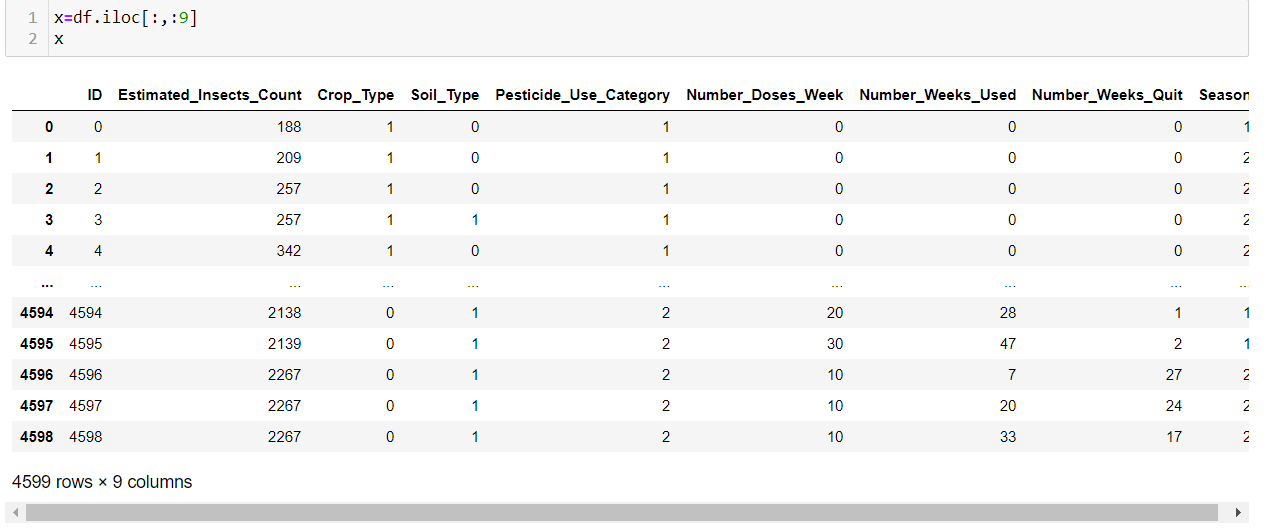


To confirm the outliers, I checked the Univariate, Bivariate Analysis, pair plot and count plot. Checked the skewness with distribution plot.

4.Pre-Processing Pipeline:

After passing label encoder for ID and changing the datatype as interger, though the Number\_Weeks\_Quit and Pesticide\_User\_catergory has a bad co-relation it does contribute the dataset in its best way. Now we split the dataset into x and y variables to predict the dataset after splitting it into train and test data. Y Variable has a crop\_Damage and rest of the dataset in X variable.

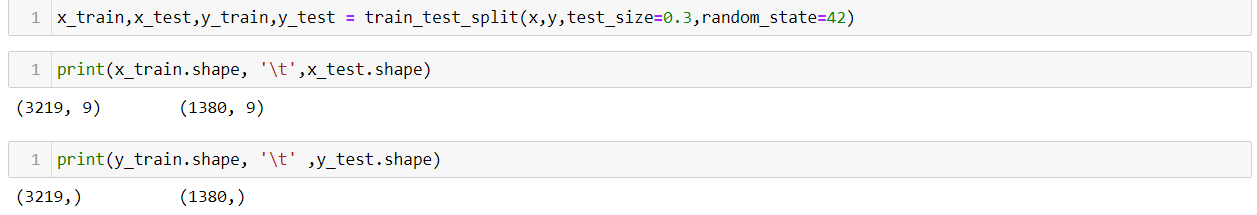


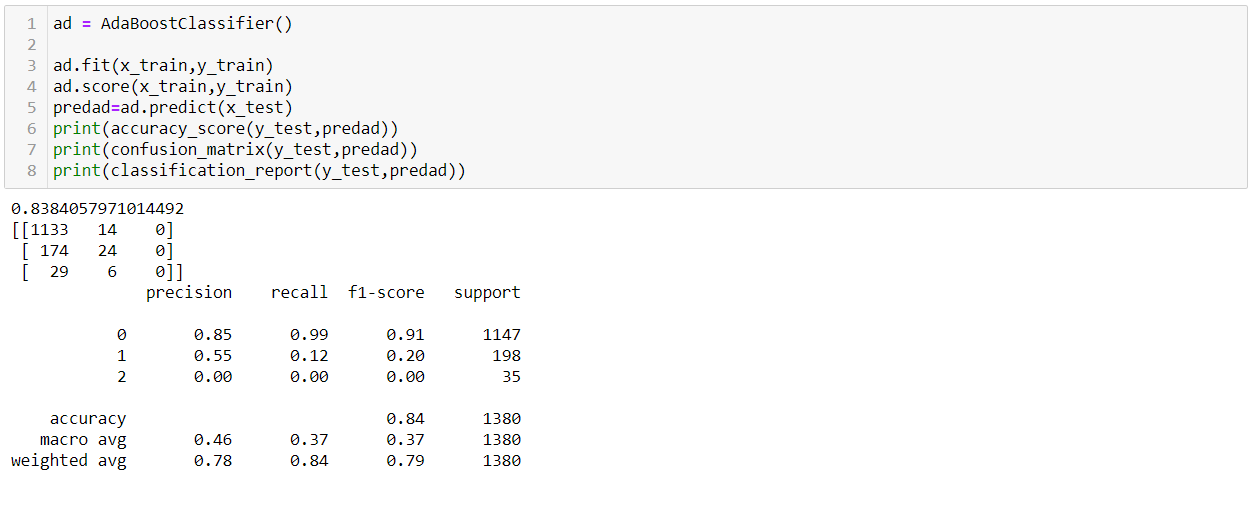


5.Building Machine Learning Models:

The dataset has been split into the x train and y train, which is 3219 for train and 1380 for test data , I used Adaboost classifier , logistic Regression, etc.

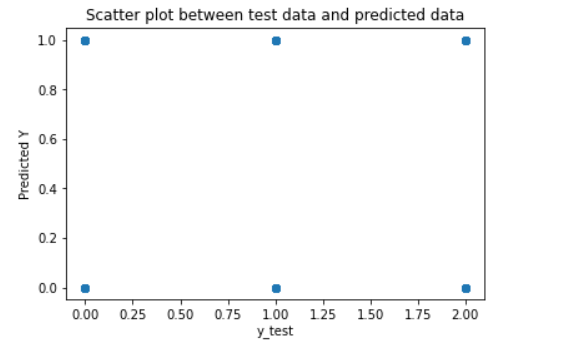
And found that Adaboost classifier gives 83 percent score.





6.Concluding Remarks:

Since I selected the Adaboost classifier, we could see its performance via scatter plot and the results are as shown below



I could also check the predictions over the fitted model as follows.

