**Machine Learning in Agriculture**

Recently we have observed the emerging concept of smart farming that makes agriculture more efficient and effective with the help of high-precision algorithms. The mechanism that drives it is Machine Learning — the scientific field that gives machines the ability to learn without being strictly programmed. It has emerged together with big data technologies and high-performance computing to create new opportunities to unravel, quantify, and understand data intensive processes in agricultural operational environments.

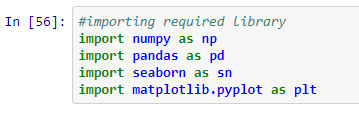
**Agriculture Dataset:**

Farmer's job is real test of endurance and determination. Once the seeds are sown, he works days and nights to make sure that he cultivates a good harvest at the end of season. A good harvest is ensured by several factors such as availability of water, soil fertility, protecting crops from rodents, timely use of pesticides & other useful chemicals and nature. While a lot of these factors are difficult to control for, the amount and frequency of pesticides is something the farmer can control.

We will determing the outcome of the harvest season i.e whether harvest will be healthy(Alive) or damaged by pesticides or any other reason.

We have columns Estimated\_Insects\_Count, Crop\_Type, Soil\_Type,Pesticide\_Use\_Category , Number\_Doses\_Week, Number\_Weeks\_Used, Number\_Weeks\_Quit,Season to find out the outcome.

**Importing the Required Libraries :**



**Getting the Data:**

Loading the data from local



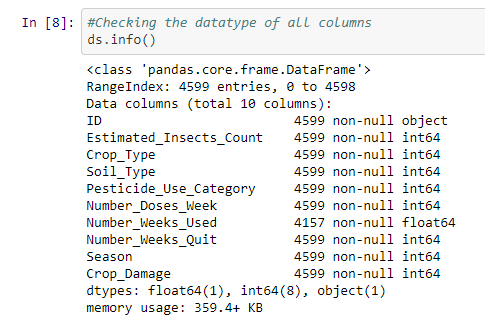
**Data Exploration/Analysis:**

**Dataset Information:**

Reading all the basic information of dataset.

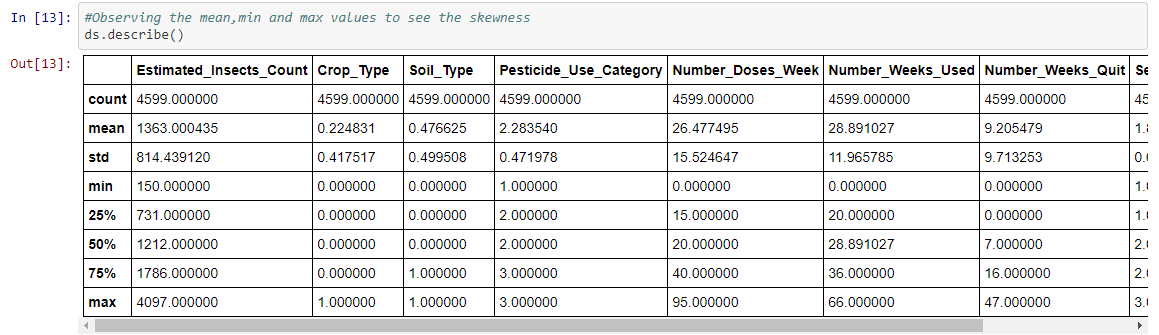
We can see except Columns “ID” all other columns are int type.

There are total 4599 entries and 10 columns

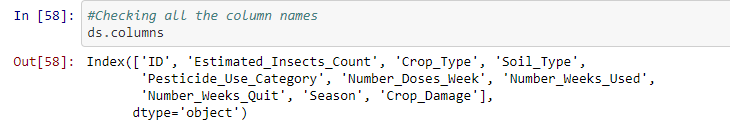


**Describe :**

We can find the mean,max,min values for all columns

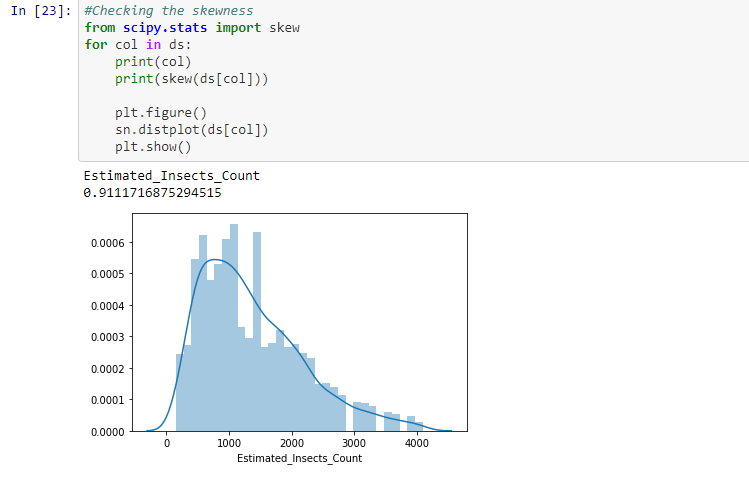


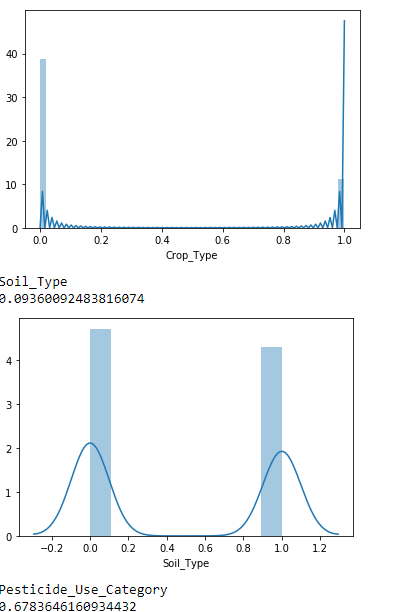
**All the Columns in dataset:**



**Checking Skewness :**

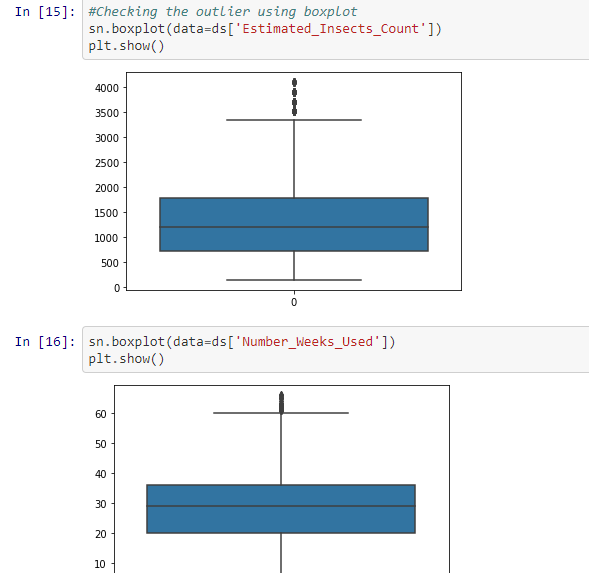
There is not much skewness in any columns





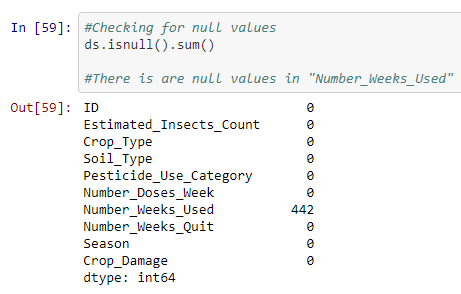
**Checking the outlier :**

There are outliers in Estimated insect count and Number Weeks used



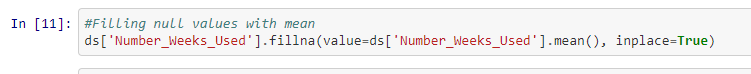
**Checking Null values:**

There are null values in columnsNumber\_Weeks\_Used



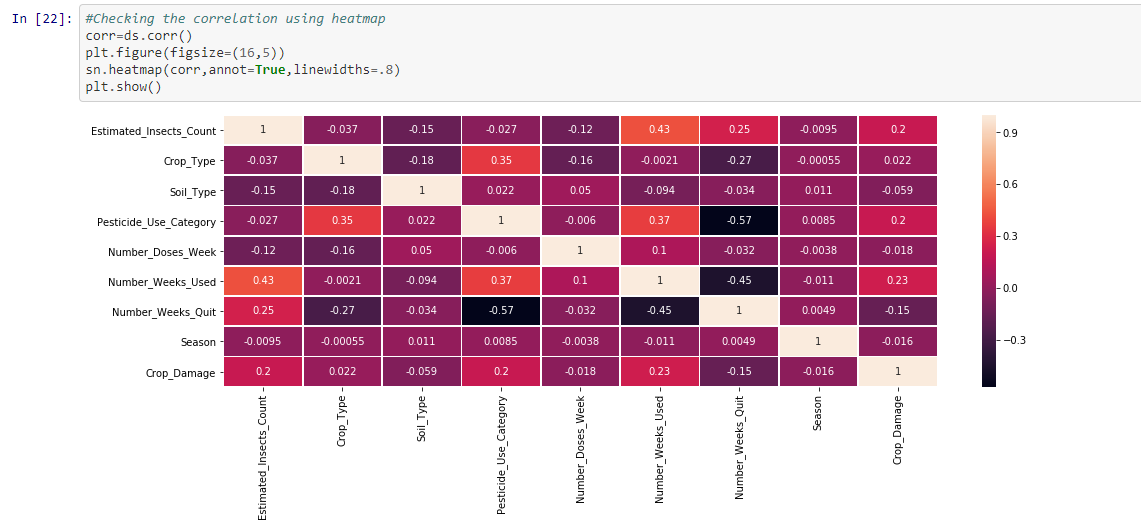
**Removing the null values :**

Replaced the null values in column Number\_Weeks\_Used



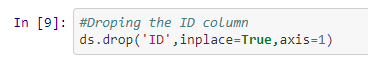
**Checking the correlations :**

Pesticide\_use\_category has the highest correlation followed by soil type. Where as crop type has the lowest correlation



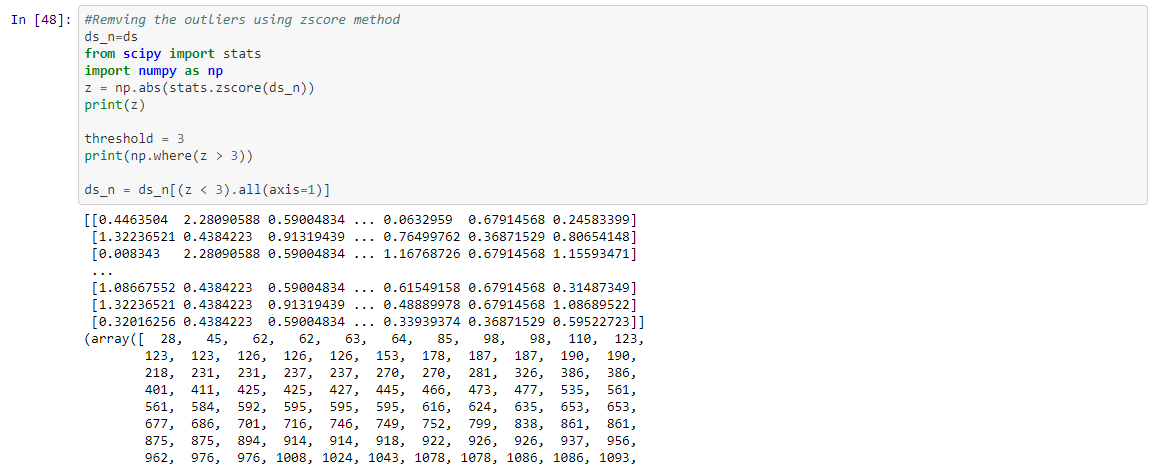
**Droping the Column:**

Droping the ID columns as it has no relevance



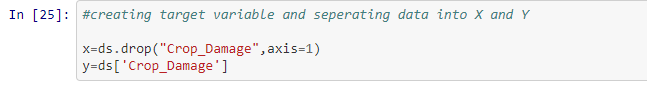
**Removing Outlier:**

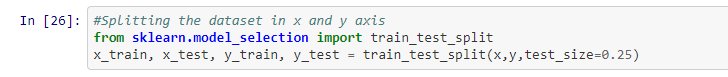
Removing outliers using Zscore



**Spliting Dataset :**

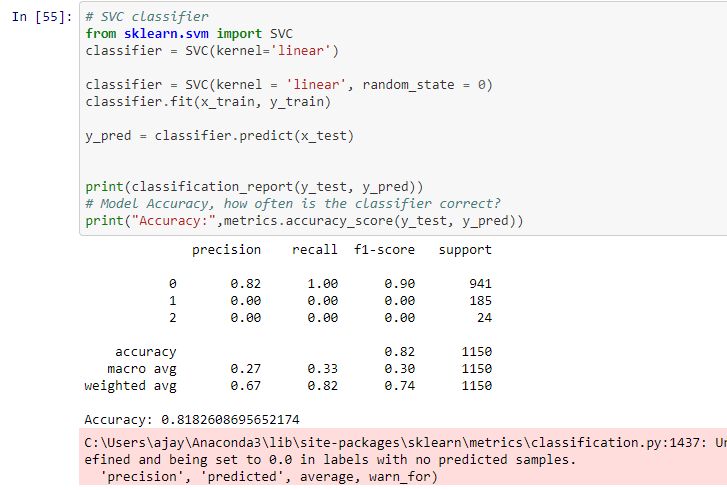
Crop\_Damage is our target variable. And we will split our data into x and y





**Training Model:**

Using SVC, Gaussian NB,Random Forest Classifier

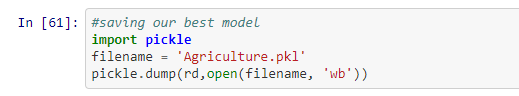


**HyperParameter Tuning :**

Hypertuning SVC model

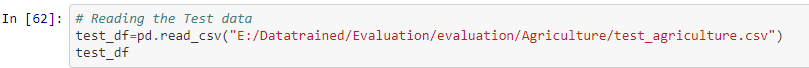


**Saving Mode :**

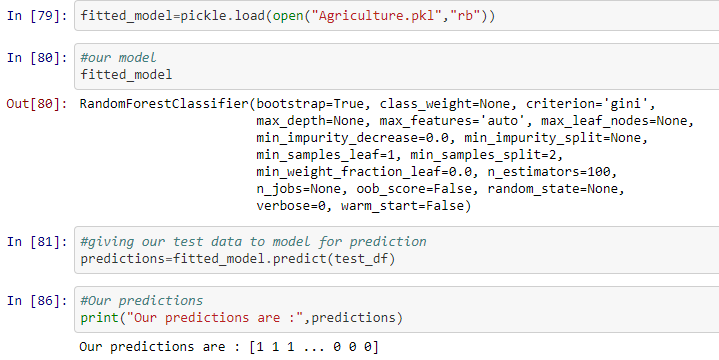


**Reading the Test Dataset:**

**We will again perform the same Preprocessing for Test dataset which we did for train dataset.Once the data is processed we will pass it to our model.**



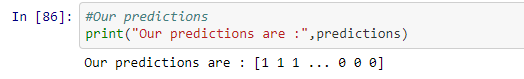
**Loading Saved Model :**



**Giving our dataset to Model we build:**



**Data Predicted from Test Data**

:

**Summary :**

We have used 4 models( Logistic Regression,SVC,Gaussian NB and Random Forest ) for predictions.

During preprocessing we found there are null values in columns” columnsNumber\_Weeks\_Used” which we replaced using its mean value.There were few outliers in Estimated insect count and Number Weeks used for whichh we applied Zscore to remove the outliers. Pesticide\_use\_category has the highest correlation followed by soil type. Where as crop type has the lowest correlation.

We dropped the columns “ID” as it was not relevant.

Out of all the 4 models that we used we got best accuracy with SVC which is 81.82%.So we can say that SVC best fitted out model. Therefore we used SVC for hypertuning. After hypertuning the parameters of SVC our accuracy increased to 83%