AI in AgriTech

Surabhi Lambe
IIT Jodhpur

Date: 20-05-2022

Abstract

AgriTech or “Agriculture Technology” is the use of technological innovations in agriculture to increase the yield of grown crops, efficiency, and profitability of the company or startup. The main idea of this paper is to develop an AI ML model-based product that will profit small AgriTech companies, Agro companies, and Agricultural business companies. The product will help the AgriTech companies in predictions and decision-making related to the outcomes of the harvest season. The model will be capable of predicting whether a particular crop would turn out to be healthy or damaged by pesticides or damaged by any other reasons. It will also predict crop diseases. It will help these companies connect to farmers.

“Predicting the future isn’t magic, it’s artificial intelligence.” ~Dave Waters

Now it's time to expand the applications of artificial intelligence and machine learning in agricultural businesses. AI ML will act as a bridge that connects the farmers to these AgriTech companies which would result in revolutionary growth in the farming sector.

1. Problem Statement

To develop a machine learning model which would determine the outcome of the harvest season by making predictions about whether the crop would be healthy or damaged due to pesticides or damaged due to any other reasons. To connect farmers to AgriTech companies, using AI systems to boost production and predict crop diseases.

2. Market/Customer/Business Need Assessment

In the era of digital agriculture, technologies like artificial intelligence (AI), cloud machine learning, satellite imagery, and advanced analytics are enabling small-holder farmers to boost their revenue by increasing crop productivity and controlling prices. Farmers in India have long relied on rain for their livelihoods, and climate change has rendered them increasingly vulnerable to crop loss. AI insights will help minimize uncertainty and risk in farm operations across the life cycle. AI in agriculture has the potential to change the lives of millions of farmers in India and around the world.

However, the model that has been described here, not only helps the farmers to deliver high-quality produce, but it mainly helps the budding small agricultural-based companies and startups who have no exposure to agriculture but wish to enter this arena, explore and grow their business.

The role of these companies would be :

1. Communicating quality
2. Creating the safety net
3. Creating quality specific supply lanes

The first step to achieving the above goals is to **connect with the farmers**. The ml model will suggest which farmers the company should connect with based on the dataset of the farmer's location from farming land, their skills and expertise, and payment demands.

The second step is to analyze the data such as pesticide use, soil type, climate, soil moisture content, etc to make predictions on the survival of crops after harvest season. The information (whether the crop is healthy or damaged by pesticides or damaged due to any other reason) is useful to plan the business in order to create quality specific supply lanes and in managing the growth of crops so as to reduce damage in the future.

The third step is crop disease prediction using ml.

3. Target Specifications and Characterization

A farmer is an expert in cultivation with limited resources and protections. Food and agriculture corporations, on the other hand, are experts at transforming agricultural commodities into value-added products for the right customers; nevertheless, they have no control over production and have found it difficult to engage directly with farmers.

The problem emerges during communication and involvement between the two parties, notwithstanding their expertise in their respective fields. As a result, despite various governmental attempts, poor farmer earnings and unmet consumer requirements persist. If we can provide the appropriate resource options and construct a safety net around what the farmer cultivates, the farmer will be able to deliver high-quality fruit with ease. At the same time, if we can help agribusiness companies build predictability in supply quality, agribusinesses will be able to give the best prices for the products they buy.

The model will be able to :

1. Connect to farmers
2. Determine the outcome of the harvest season (crop is damaged or healthy)
3. Predict disease with the help of crop images
4. Perform aerial surveys and imaging

4. External Search

The dataset used for predicting the outcome of the harvest season whether the crop is healthy or damaged can be found on analyticsvidhya.com.

	ID	Estimated_Insects_Count	Crop_Type	Soil_Type	Pesticide_Use_Category	Number_Doses_Week	Number_Weeks_Used	Number_Weeks_Quit
0	F00000001	188	1	0	1	0	0.0	0
1	F00000003	209	1	0	1	0	0.0	0
2	F00000004	257	1	0	1	0	0.0	0
3	F00000005	257	1	1	1	0	0.0	0
4	F00000006	342	1	0	1	0	0.0	0

Season	Crop_Damage
1	0
2	1
2	1
2	1
2	1

The dataset contains 11 categories and 88858 entries. Categories are insect count, crop type, soil type, pesticides used, doses of pesticides per week, number of weeks used, number of weeks quit, and season. The target variable is crop damage.
More information on these categories can be found on the government of India's website https://www.india.gov.in/topics/agriculture.

The ml model is trained using sklearn and libraries such as NumPy, pandas, matplotlib, and seaborn have been used.

5. Benchmarking Alternate Products

Krish-e

Krish-e, a popular Mahindra & Mahindra agriculture app, provides a personalized crop calendar for farms as well as useful agriculture information like land preparation, crop sowing, crop planning, fertilizer management, seed treatment, pest, and disease management, crop diagnosis and treatment, weed treatment, and irrigation.

Shetkari

Shetkari Mitra is a multi-purpose smartphone app for Indian farmers. It gives information and knowledge of government programs, crop management, agricultural business and guidelines and market rates.

In comparison to these already existing products, the model described in this paper uses data analysis and machine learning to predict the outcome of harvest uses AI techniques for aerial survey and imaging as well as it tries to connect to suitable farmers.

6. Applicable Patents

U.S. Provisional Patent Application Ser. No. 60/987,883, entitled, “Method and Apparatus of Taking Aerial Surveys,” filed on Nov. 14, 2007, and naming Elaine S.

Acree as an inventor, the disclosure of which is incorporated herein, in its entirety, by reference.

Already existing technologies can be integrated into this product model and can be implemented as per the requirements of the Agri company.

7. Applicable Regulations

The Insecticides Act, 1968 and Insecticides Rules, 1971 regulate the import, registration process, manufacture, sale, transport, distribution and use of insecticides (pesticides) with a view to prevent risk to human beings or animals and for all connected matters, throughout India.

Fertilizer (Control) Order, 1985 which is administered by Deptt. of Agriculture Cooperation, Govt. of India has been issued under the Essential Commodities Act, 1955. The FCO lays, down what substances qualify for use as fertilizers in the soil, product-wise specifications, methods for sampling and analysis of fertilizers, procedure for obtaining a license/registration as a manufacturer/dealer in fertilizers and conditions to be fulfilled for trading thereof, etc.

8. Applicable Constraints

The Department of Agricultural Research and Education has been allocated about Rs 8,513 crore in the Union Budget 2021. In comparison to 2020, the Department of Agriculture, Cooperation, and Farmers' Welfare's budget have been cut by moreover Rs 10,000 crore.

9. Business Opportunity

The following are important aspects influencing the AI market in agriculture:

1. Increasing need for agricultural production due to population growth.

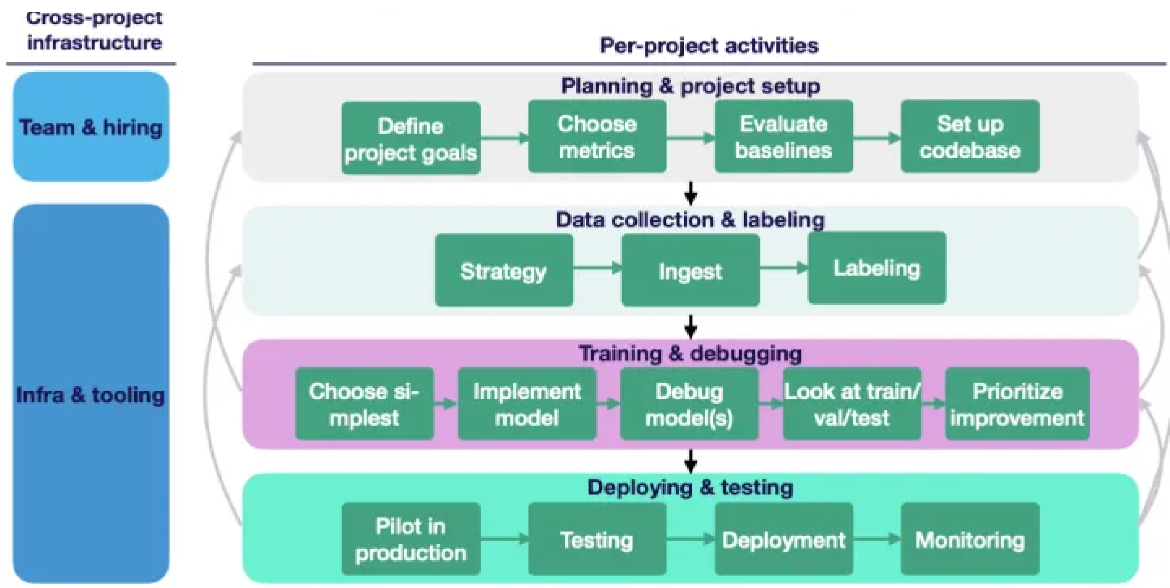
2. Crop productivity can be improved by expanding information management systems and adopting more technological technology.
3. Integration of deep learning techniques to boost crop output.
4. Governments all across the world are stepping up to assist the adoption of new agricultural techniques.

10. Concept Generation

To create a machine learning model that can forecast the outcome of the harvest season by predicting whether the crop will be healthy, harmed by pesticides, or damaged for any other reason. To connect farmers with AgriTech firms, and to use AI systems to increase crop productivity and predict crop illnesses.

11. Concept Development

Perform data analysis on the collected database, train ml models to make predictions, use the information obtained from the above analysis to increase business. Design an ai ml algorithm which can suggest the companies which farmers they should connect with according to their specifications. Use ai ml techniques and computer vision to predict crop disease by just taking a picture of the crop as input. Make use of unmanned drones to perform aerial surveys and imaging.
For ml model:



12. Final Product prototype and details of its working

The libraries used:

Libraries

In [1]:
```
# importing libraries

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

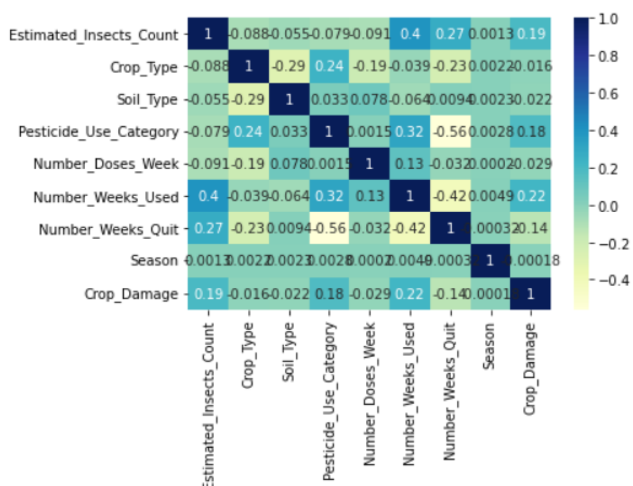
:
```
#Importing libraries

from sklearn.metrics import accuracy_score,classification_report,confusion_matrix,roc_auc_score,roc_curve
from sklearn.neighbors import KNeighborsClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import cross_val_score
from scipy.stats import zscore
from sklearn.preprocessing import LabelEncoder,StandardScaler
from sklearn.model_selection import train_test_split,GridSearchCV
```

Heatmap

In [17]:
```
sns.heatmap(ag.corr(),annot=True,cmap='YlGnBu')
```

Out[17]:
```
<AxesSubplot:>
```



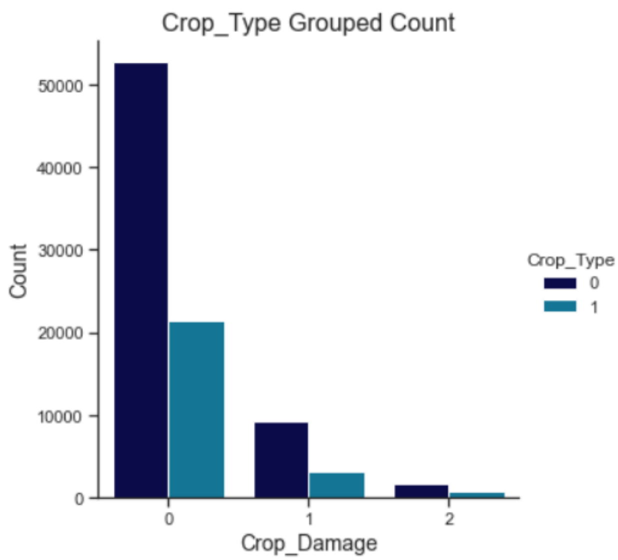
We need to find correlation with Crop_Damage

The categories which are positively correlated with crop damage are found to be :

Estimated_Insects_Count, Pesticide_Use_Category, Number_Weeks_Used

Catplot

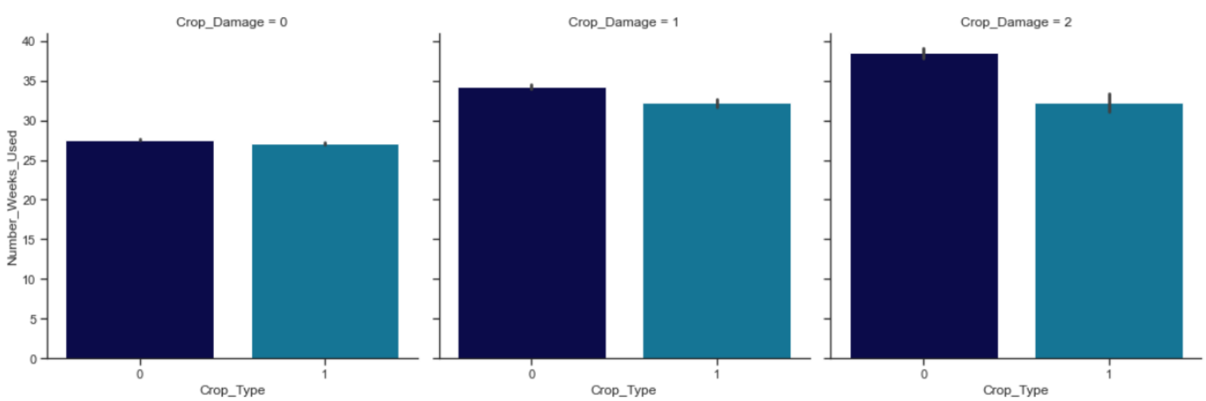
In [30]:
```
sns.catplot(x='Crop_Damage',data=ag,palette="ocean",kind='count',hue='Crop_Type')
plt.title("Crop_Type Grouped Count",fontsize=16)
plt.xlabel("Crop_Damage",fontsize=14)
plt.ylabel("Count", fontsize=14)
plt.show()
```



From the above catplot, we can observe that crop 0 has higher chances of survival compared to crop 1
Damage due to pesticides is less comapred to others

```
plt.figure(figsize=(12,5))
sns.catplot(x='Crop_Type',y='Number_Weeks_Used',data=ag, palette="ocean",kind='bar',col='Crop_Damage')
plt.show()
```

```
<Figure size 864x360 with 0 Axes>
```



From the above 3 graphs, it is observed that :
Crop 0 is more affected by pesticides and gets more damaged as compared to crop 1.
Duration of pesticide damage is lower for crop 1 than crop 0.

Countplot

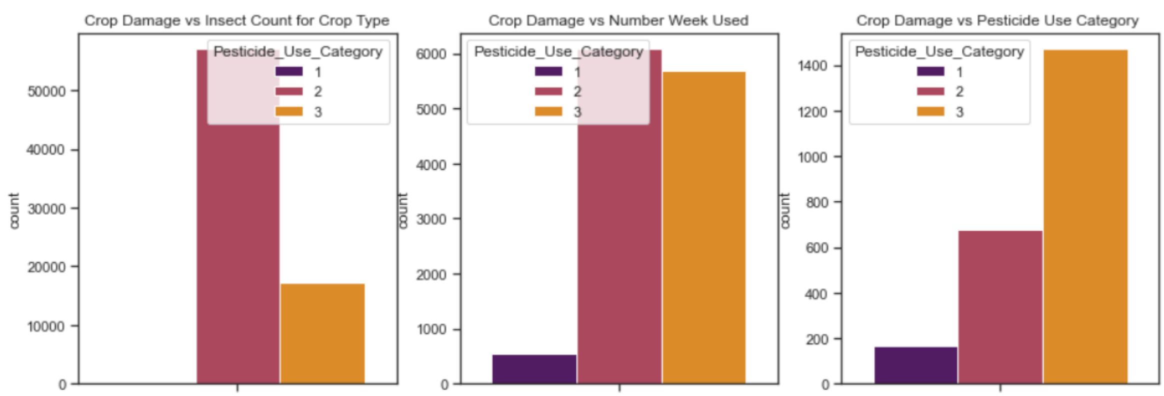
```
In [70]: fig, [ax1,ax2,ax3] = plt.subplots(nrows=1,ncols=3,figsize=(15,5))

ax1=sns.countplot(x="Crop_Damage" ,hue="Pesticide_Use_Category",data=ag[ag["Crop_Damage"]==0],ax=ax1,palette="inferno")
ax1.set_title("Crop Damage vs Insect Count for Crop Type")

ax2=sns.countplot(x="Crop_Damage" ,hue="Pesticide_Use_Category",data=ag[ag["Crop_Damage"]==1],ax=ax2,palette="inferno")
ax2.set_title("Crop Damage vs Number Week Used")

ax3=sns.countplot(x="Crop_Damage" ,hue="Pesticide_Use_Category",data=ag[ag["Crop_Damage"]==2],ax=ax3,palette="inferno")
ax3.set_title("Crop Damage vs Pesticide Use Category ")
```

```
Out[70]: Text(0.5, 1.0, 'Crop Damage vs Pesticide Use Category ')
```

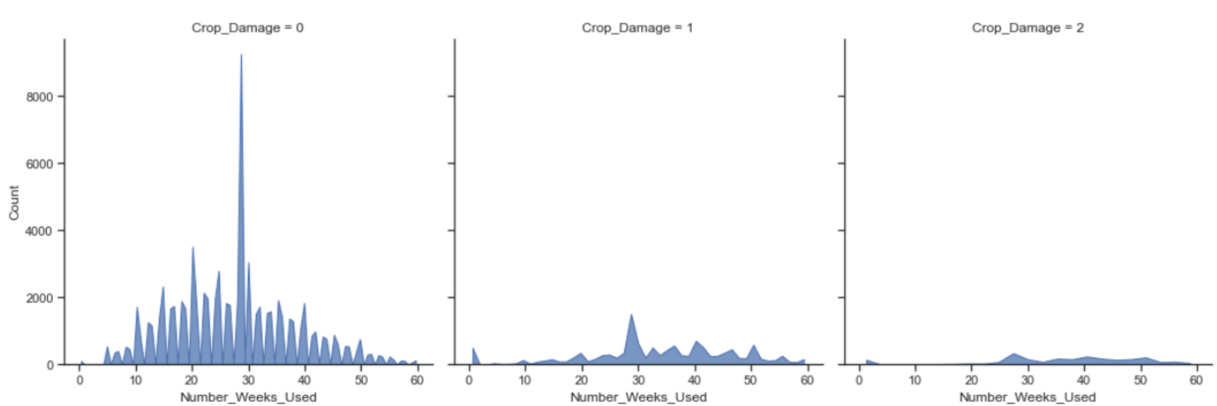


From the above plots, type 2 pesticide is safer to use than type 3.

Histplot

```
In [58]: plt.figure(figsize=(20,10))
grp= sns.FacetGrid(ag, col='Crop_Damage',height=5)
grp = grp.map(sns.histplot, "Number_Weeks_Used",element="poly")
plt.show()
```

```
<Figure size 1440x720 with 0 Axes>
```

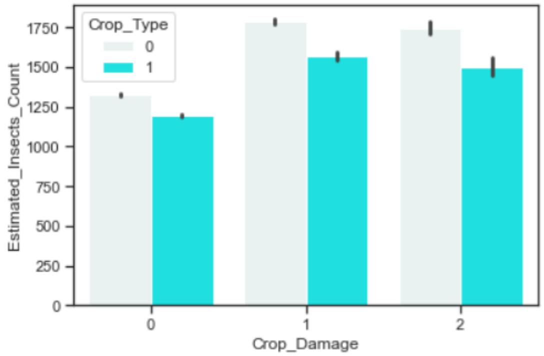


Graph 1 :
Till week 25 approx, damage due to pesticides is comparitively less
Graph 2 and 3 :
After week 25, damage increases significantly

Barplot

```
In [71]: sns.barplot(x="Crop_Damage",y="Estimated_Insects_Count",hue="Crop_Type",data=ag,color="cyan")
```

```
Out[71]: <AxesSubplot:xlabel='Crop_Damage', ylabel='Estimated_Insects_Count'>
```



From the barplot, it's observed that insects attack more on crop 0 compared to crop 1

SKEW ANALYSIS

```
In [72]: ag.skew()
```

```
Out[72]: Estimated_Insects_Count    0.699465
         Crop_Type                  0.955978
         Soil_Type                  0.166914
         Pesticide_Use_Category     0.779488
         Number_Doses_Week          0.744487
         Number_Weeks_Used          0.213554
         Number_Weeks_Quit          0.831183
         Season                     0.145228
         Crop_Damage                2.367816
         dtype: float64
```

Data: Estimated_Insects_Count, Crop_Type, Pesticide_Use_Category, Number_Doses_Week, Number_Weeks_Quit: moderately skewed
Data: Soil_Type, Number_Weeks_Used, Season: fairly symmetrical

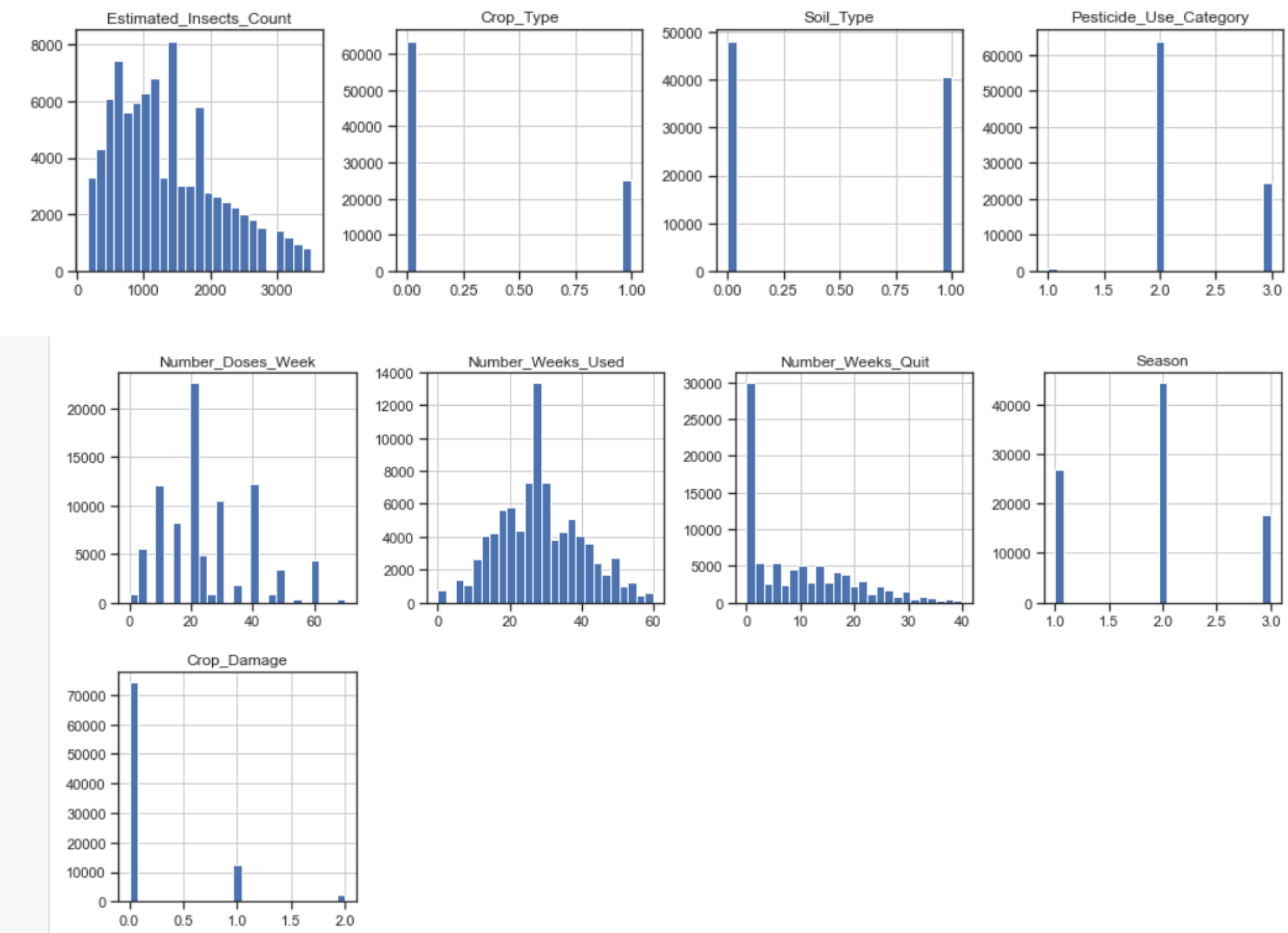
```
In [85]: ag.kurt()
```

```
Out[85]: Estimated_Insects_Count    -0.229338
         Crop_Type                  -1.086131
         Soil_Type                  -1.972184
         Pesticide_Use_Category     -0.630011
         Number_Doses_Week          -0.007164
         Number_Weeks_Used          -0.228283
         Number_Weeks_Quit          -0.206514
         Season                     -0.967584
         Crop_Damage                 4.978762
         dtype: float64
```

Based on the obove obtained kurtosis, we can observe that this data is a platykurtic(Kurtosis <3) distribution i.e. a flat distribution where the values are moderately spread out

```
In [84]: ag.hist(figsize=(16,16),layout=(4,4),bins=25)
```

```
Out[84]: array([[<AxesSubplot:title={'center':'Estimated_Insects_Count'}>,
                 <AxesSubplot:title={'center':'Crop_Type'}>,
                 <AxesSubplot:title={'center':'Soil_Type'}>,
                 <AxesSubplot:title={'center':'Pesticide_Use_Category'}>],
                [<AxesSubplot:title={'center':'Number_Doses_Week'}>,
                 <AxesSubplot:title={'center':'Number_Weeks_Used'}>,
                 <AxesSubplot:title={'center':'Number_Weeks_Quit'}>,
                 <AxesSubplot:title={'center':'Season'}>],
                [<AxesSubplot:title={'center':'Crop_Damage'}>, <AxesSubplot:>,
                 <AxesSubplot:>, <AxesSubplot:>],
                [<AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>, <AxesSubplot:>]],
               dtype=object)
```



evenly spread distribution is observed

3 Machine Learning Classifier models are used

1. KNN
2. Decision Tree
3. Random Forest

```
Results for model :  KNN
max accuracy score is 0.8307870686127404
Mean accuracy score is :  0.8278939507402254
Std deviation score is :  0.0016142378230195366
Cross validation scores are :   [0.82995724 0.82793158 0.82511816 0.82887851 0.82758427]
---------------------------------------------------------------------------------------------
Results for model :  Decision Tree Classifier
max accuracy score is 0.7546378393125085
Mean accuracy score is :  0.7487001990829059
Std deviation score is :  0.0015754305634090766
Cross validation scores are :   [0.7498312  0.74859329 0.745386   0.75021102 0.75111136]
---------------------------------------------------------------------------------------------

Results for model :  Random Forest
max accuracy score is 0.8285022507161369
Mean accuracy score is :  0.8235837157176809
Std deviation score is :  0.0016785887237225283
Cross validation scores are :   [0.82635607 0.8241616  0.82027909 0.82403916 0.82499578]
---------------------------------------------------------------------------------------------
```

Maximum accuracy is shown by KNN.

Now, on training the model again using the n_parameter found from GridSearchCV (n is found to be **20**)

```
{'n_neighbors': 20}
```

```
Results for model :  KNeighbors Classifier
max accuracy score is 0.8442913654344564
Mean accuracy score is :  0.842141373623542
Std deviation score is :  0.001602548634088536
Cross validation scores are :   [0.84396804 0.84289894 0.8413797  0.84305892 0.83940127]
---------------------------------------------------------------------------------------------
```

The final accuracy obtained is **84.429%**

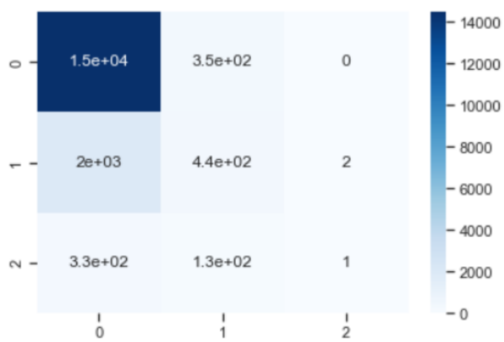
Heat map obtained after training KNN on 80-20 train-test data

```
accuracy score is :  0.8405919423812739
classification report
              precision    recall  f1-score   support

           0       0.86      0.98      0.91     14848
           1       0.48      0.18      0.26      2461
           2       0.33      0.00      0.00       463

    accuracy                           0.84     17772
   macro avg       0.56      0.39      0.39     17772
weighted avg       0.79      0.84      0.80     17772

<AxesSubplot:>
```



Final model is KNN

13. Code Implementation

1. Necessary libraries are imported such as NumPy, pandas, matplotlib, sklearn, etc.
2. The dataset contains 88858 rows and 11 columns. Info of the dataset is found.
3. Preprocessing of the data is done i.e. cleaning the data: checking null values, checking datatypes, checking unique values, replacing null values with the mode of the data
4. Outliers are removed and boxplots are used for outliers detection.
5. Exploratory Data Analysis is done. The following plots are drawn: heatmap, catplot, countplot, histplot, and barplot. Using these, univariate and bivariate analysis is done.
6. Skew and Kurtosis Analysis is done and histplots are made.
7. Scaling of the data is done using StandardScalar.
8. 3 ml classifier models are trained: KNN, random forest, and decision trees. KNN is found to generate the maximum accuracy score.
9. GridSearchCv is used to obtain n_neighbors value. This value is 20.
10. KNN model is trained again with n=20. Accuracy is found to increase a bit.

[Github link](#) to code implementation.

14. Observations

After analyzing the dataset, the following conclusions can be obtained:

1. From the **heatmap**:
    1. The categories which are positively correlated with crop damage are found to be: Estimated_Insects_Count, Pesticide_Use_Category, Number_Weeks_Used.
2. From the **Catplot**:
    1. Crop 0 has a higher chance of survival compared to crop 1.
    2. Damage due to pesticides is less compared to others.
    3. Crop 0 is more affected by pesticides and gets more damaged as compared to crop 1.
    4. Duration of pesticide damage is lower for crop 1 than crop 0.
3. From the **Countplot**:
    1. Type 2 pesticide is safer to use than type 3.
4. From the **histplot**:
    1. Till week 25 approx, damage due to pesticides is comparatively less
5. From the **barplot**:
    1. Insects attack more on crop 0 compared to crop 1
6. **Skew and kurtosis** Analysis: data is moderately skewed with a flat (platykurtic) distribution.
7. **KNN** gives the highest accuracy score after performing GridSearchCV with n as 20 which is **0.844**.

15. Conclusion

KNN can be used to train data. The predictions obtained will surely help the AgriTech companies to plan their resources and plan their future actions. It will help them to decide which pesticides to use and for what duration.

The real purpose of data collection and production is to use it. Data analytics in agriculture can result in large productivity gains and significant cost reductions. Farmers can acquire credible recommendations based on well-sorted real-time information on crop needs by merging AI and big data. As a result, guesswork will be eliminated, allowing for more exact farming methods such as irrigation, fertilization, crop protection, and harvesting.

Future scopes can be extending this technology to vertical farming and organic farming. An app can be created which would incorporate all these features and models. The app will be able to make predictions, analyze data, connect farmers and thus increase overall connectivity, and help in transportation by suggesting low-cost transport options.

References:

S. Y. Liu, "Artificial Intelligence (AI) in Agriculture," in IT Professional, vol. 22, no. 3, pp. 14-15, 1 May-June 2020, doi: 10.1109/MITP.2020.2986121.

https://www.feedough.com/what-is-agritech/

https://www.cnbctv18.com/agriculture/how-can-agritechs-and-fpos-build-a-bridge-between-agribusiness-corp-and-small-farmers-6972361.htm

https://news.microsoft.com/en-in/features/ai-agriculture-icrisat-upl-india/

https://www.mahindra.com/news-room/press-release/mahindra-rolls-out-krish-e-centres-in-maharashtra

https://opengeekslab.com/blog/ai-in-agriculture-ways-enhance-business/

https://ipca.org.in/rescources/pesticide-regulations/#:~:text=Pesticide%20Regulations%20in%20India,all%20connected%20matters%2C%20throughout%20India.

https://www.kdnuggets.com/2020/02/deploy-machine-learning-model.html