IJCSPUB.ORG ISSN : 2250-1770

INTERNATIONAL JOURNAL OF CURRENT SCIENCE (IJCSPUB)

IJCS PUBLICATION (IJCSPUB.ORG) An International Open Access, Peer-reviewed, Refereed Journal

Ref No : IJCSPUB/Vol 13 / Issue 3 / 247

To,
Suneel Chandra Vanamu

Subject: Publication of paper at International Journal of Current Science.

Dear Author,

With Greetings we are informing you that your paper has been successfully published in the International Journal of Current Science - IJCSPUB (ISSN: 2250-1770). Thank you very much for your patience and cooperation during the submission of paper to final publication Process. It gives me immense pleasure to send the certificate of publication in our Journal. Following are the details regarding the published paper.

About IJCSPUB:		Scholarly open access journals, Peer-reviewed, and Refereed Journals, Impact factor 8.17 (Calculate by google scholar and Semantic Scholar | AI-Powered Research Tool) , Multidisciplinary, Monthly, Indexing in all major database & Metadata, Citation Generator, Digital Object Identifier(DOI)
Registration ID	:	IJCSP_217470
Paper ID	:	IJCSP23C1247
Title of Paper	:	Efficient Techniques of Harvesting
Impact Factor	:	8.17 (Calculate by Google Scholar) | License by Creative Common 3.0
Publication Date:		2023-09-25 09:08:00
DOI	:	
Published in	:	Volume 13 | Issue 3 | September 2023
Page No	:	108-121
Published URL	:	http://www.ijcspub.org/viewfull.php?&p_id=IJCSP23C1247
Authors	:	Suneel Chandra Vanamu
Notification	:	

Thank you very much for publishing your article in IJCSPUB.

Editor In Chief
International Journal of Current Science - IJCSPUB
(ISSN: 2250-1770)

International Journal of Current Science (IJCSPUB) ISSN 2250-1770 IJCSPUB

IJCSPUB.ORG ISSN : 2250-1770
INTERNATIONAL JOURNAL OF CURRENT SCIENCE (IJCSPUB)
IJCS PUBLICATION (IJCSPUB.ORG) An International Open Access, Peer-reviewed, Refereed Journal

The Efficient Techniques of Harvesting

Suneel Chandra Vanamu
Student
Department Of Computer Science Engineering,
GITAM Institute Of Technology, Visakhapatnam, India.

Abstract: Farmers are essential to the nation's economy, yet their efforts are often overlooked. Pesticides can be beneficial when used correctly, but too much can damage the entire crop. Farmers must accept losses if their crops fail, so predicting the harvest outcome is essential. Machine learning algorithms can be used to analyze historical data such as pesticide usage, soil parameters, and crop yields. This project seeks to forecast the harvest outcome and the crop type that should be grown in a given field based on existing data. The prediction will help farmers plan crops and forecast yields before planting. The project uses Random Forest, Decision Tree, Artificial Neural Networks, and Logistic Regression algorithms to compare accuracy and performance and select the best model.

1. Introduction

Recently, the concept of Machine Learning has evolved in every sector, including agriculture, where farmers are making efforts to ensure a smooth harvest season and a healthy plantation at the end of the season. Machine Learning can also help farmers choose the type of crop to cultivate based on the soil fertility. However, there are several other factors in the agriculture sector that need further development, including the improvement of irrigation technologies. The use of pesticides is crucial during the harvest period, as they can damage the entire crop if the measured amounts are not limited. Predicting the status of the crop in the early stages of harvesting can help farmers take appropriate action and prevent crop damage. Machine Learning techniques can also take into account the amount of chemicals in the soil and predict which plants can be cultivated. We used several datasets that contain information on previous crop plantations, including the amount of pesticides, insects, soil type, crop type, and the crop damage status of specific plantations. We also used another dataset that includes the amount of phosphorus, ammonia, and other soil nutrients to predict which crop would yield better results.

2. Literature Review

In our study, data from various latitudes is used to predict crop yields, aligning with the objectives of forecasting before the harvest season. A notable reference is the use of Decision Trees for crop yield prediction, as demonstrated by research at Minnesota University. The study found Random Forest to be a highly accurate machine learning approach for global and regional crop yield prediction. Other techniques, such as Support Vector Machines, were also employed in related research. Notably, Chlingaryan and Sukkarieh explored nitrogen estimation using machine learning, emphasizing advances in sensing technology. Additionally, the literature reveals the evolving role of machine learning in agriculture, offering promising opportunities for computational enhancements and digitalization. Various algorithms have been applied to predict agricultural production, with meteorological variables playing a significant role. However, it is essential to widen the search to consider additional factors affecting crop yields, as indicated in recent literature. Finally, research focused on fruit maturity anticipation highlights the broader applicability of machine learning in agriculture.

3. Problem Identification and Objectives

3.1. Problem Statement

Pesticides, while essential for protecting crops, can become detrimental when applied in excess. This study leverages data from various farmers' end-of-season crop harvests, supplemented by soil composition and humidity data, to address this issue. We assume control over other variables like farming techniques for simplicity. The core challenge involves predicting harvest outcomes, distinguishing between healthy crops, those damaged by excessive pesticide use, and those harmed by other factors. Additionally, the study aims to identify the optimal crop to cultivate based on soil conditions.

3.2. *Objectives*

- Develop an efficient model to guide farmers in adopting best practices.
- Improve crop management throughout the harvest season.
- Accurately predict crop health, distinguishing between pesticide damage and other causes.
- Recommend crop varieties tailored to the soil's fertility.
- Provide a user-friendly interface for farmers to access the predictive model and receive crop recommendations.

4. SYSTEM METHODOLOGY

This Section describes the proposed system, working methodology, performance metrics, and software and hardware details.

4.1. *Proposed System*

In this study, two approaches were explored to improve the harvest model's predictions. Firstly, a thorough analysis of the continuous data points in all columns was conducted, with 8 independent variables and a dependent variable that is a predictor with 3 unique data points: (0,1,2) which represents [Crop is alive, damaged by pesticides, damaged by some reason]. For the crop prediction model, 7 independent variables and 1 dependent variable with 22 different types of crops that can be classified into were used. After careful visualization of the data, missing values and outliers were corrected with the mean of the columns, and the data was subjected to data preprocessing before being fed into several machine learning algorithms, including Random Forest, KNN, Decision Trees, AdaBoost, XGBoost, Gaussian Naive Bayes, and LightGBM. The algorithm that provided the best fit with a minimal amount of accuracy, precision, recall, and F1-Score was then selected. To further increase efficiency, a second approach was introduced where columns were extended with all possible ways of grouping columns and then all algorithms were applied again with the extended dataset, resulting in high accuracy and better results.

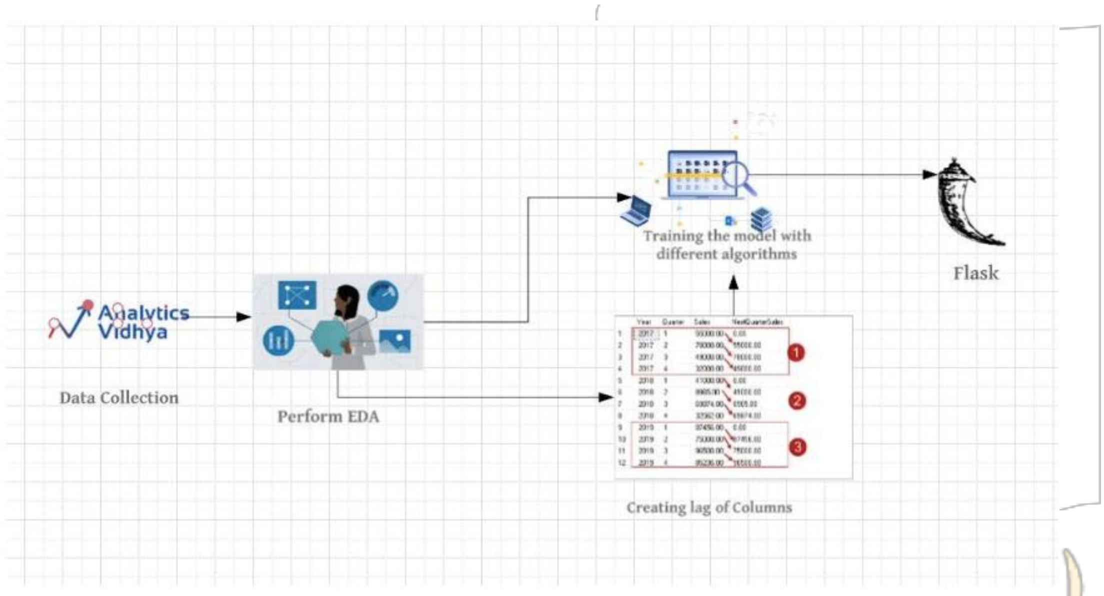


figure 1 : system block diagram

4.2. *Working Methodology*

We sourced our data from an Analytical Vidya hackathon. We pursued two distinct approaches to improve prediction accuracy. Firstly, we performed a thorough data analysis, focusing on columns with continuous data. Our dataset included 8 independent variables and a dependent variable with 3 unique values [0, 1, 2] representing crop status (alive, damaged by pesticides, damaged by other causes). We addressed missing data, identified outliers, and replaced them with column means. We then applied machine learning algorithms, prioritizing models that offered enhanced accuracy, precision, recall, and F1-Score with minimal complexity.

To further boost accuracy, we explored an alternative approach involving column grouping to create extended datasets. Applying machine learning algorithms to this augmented dataset yielded improved accuracy compared to the previous approach.

For the crop prediction model, we had a clean dataset, requiring minimal preprocessing. Consequently, we proceeded directly to model training, identifying the most suitable model for the task.

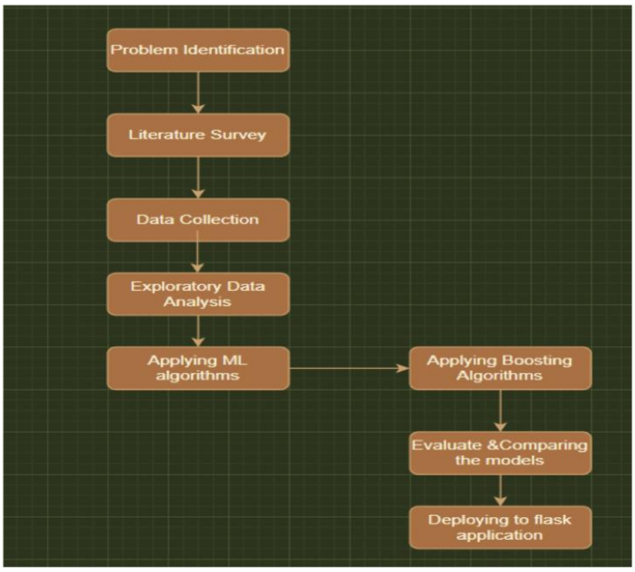


figure 2 : flowchart

4.3 Datasets

4.3.1 Harvest Prediction Dataset

	A	B	C	D	E	F	G	H	I	J
1	ID	Estimated_Insects_Count	Crop_Type	Soil_Type	Pesticide_Use_Category	Number_Doses_Week	Number_Weeks_Used	Number_Weeks_Quit	Season	Crop_Damage
2	F00000001	188	1	0	1	0	0	0	1	0
3	F00000003	209	1	0	1	0	0	0	2	1
4	F00000004	257	1	0	1	0	0	0	2	1
5	F00000005	257	1	1	1	0	0	0	2	1
6	F00000006	342	1	0	1	0	0	0	2	1
7	F00000008	448	0	1	1	0		0	2	1
8	F00000009	448	0	1	1	0		0	2	1
9	F00000010	577	1	0	1	0	0	0	1	2
10	F00000012	731	0	0	1	0	0	0	2	0
11	F00000020	1132	1	0	1	0	0	0	1	2
12	F00000021	1212	1	0	1	0		0	3	0
13	F00000023	1575	0	0	1	0	0	0	1	1
14	F00000024	1575	0	1	1	0	0	0	2	1
15	F00000028	1575	1	1	1	0	0	0	2	1
16	F00000029	1575	1	1	1	0	0	0	2	2
17	F00000030	1785	1	1	1	0	0	0	2	1
18	F00000035	2138	0	1	1	0	0	0	1	1
19	F00000037	2401	0	1	1	0		0	1	1
20	F00000038	2401	1	1	1	0	0	0	2	1
21	F00000039	2401	1	1	1	0	0	0	2	1
22	F00000045	2999	0	1	1	0	0	0	3	1

figure 3 : harvest dataset

- ID – unique harvest ID
- Estimated Insect Count – Number of insects estimated in particular Harvest Season
- Crop Type – Type of the crop in the harvest
- Soil Type – Type of the Soil in the Harvest
- Pesticide Use Category – Category of the pesticides used in harvest season
- Number Doses Week – No: of Doses with respective to pesticides
- Number Weeks Used – No: of weeks that pesticides are used
- Number week Quit – Number of Weeks that the plantation quits
- Season – Type of season
- Crop Damage – (Dependent variable) shows the status of crop

4.3.2 Crop Prediction Dataset

191	92	60	23	[illegible]	[illegible]	[illegible]	[illegible]	maize
192	79	59	17	20.37999665	63.73849998	6.644205485	108.5054416	maize
193	91	55	15	18.09300227	72.61024172	6.376651091	78.96159541	maize
194	76	51	18	26.16985907	71.96246617	6.247040422	79.84925393	maize
195	87	48	25	18.65396672	61.37879671	6.656730007999999	93.62039175	maize
196	71	60	22	26.07470121	59.37147589	6.2048017	85.75692395	maize
197	90	57	24	18.92851916	72.80086137	6.158860284	82.34162918	maize
198	67	35	22	23.30546753	63.24648023	6.3856842139999985	108.7603001	maize
199	60	54	19	18.74826712	62.49878458	6.417820493	70.23401597	maize
200	83	58	23	19.74213321	59.66263104	6.381201909	65.50861389	maize
201	83	57	19	25.73044432	70.74739256	6.877869005	98.73771338	maize
202	40	72	77	17.02498456	16.98861173	7.485996067	88.55123143	chickpea
203	23	72	84	19.02061277	17.13159126	6.920251378	79.92698081	chickpea
204	39	58	85	17.88776475	15.40589717	5.9969320370000005	68.54932919	chickpea
205	22	72	85	18.86805647	15.65809214	6.391173589	88.51048983	chickpea
206	36	67	77	18.36952567	19.56381041	7.152811172000001	79.26357665	chickpea
207	32	73	81	20.45078582	15.40312102	5.988992796000002	92.68373702	chickpea
208	58	70	84	20.6543203	16.60820843	6.231049027999999	74.6631118	chickpea
209	59	70	84	17.3348681	18.74926979	7.550808267000001	82.61734721	chickpea
210	42	62	75	18.17912258	18.90426935	7.010570541	81.84997529	chickpea
211	28	74	81	18.01272266	18.30968112	8.753795334	81.98568791	chickpea
212	58	66	79	20.99373558	19.33470387	8.718192847000001	93.55280105	chickpea
213	43	66	79	19.46233971	15.22538951	7.976607593	74.58565097	chickpea

figure 4 : crop dataset

●N - Nitrogen value in the soil.
●P - Phosphorus value in the soil.
●K - Potassium value in the soil.
●Temperature - The temperature of the soil.
●Humidity - Humidity levels in soil.
●Ph - ph value of the soil
●Rainfall - The amount of rainfall the soil receives every year. • Label - The type of the crop that can be grown in that soil.

5. Overview Of Technologies

5.1. Numpy

NumPy, like a trusty toolbox, equips Python with powerful tools for handling data, arrays, and a slew of mathematical and logical operations, making it the go-to for scientists and data enthusiasts.

5.2. Pandas

Pandas acts as a data wizard for Python, simplifying data analysis with its user-friendly data structures and extensive support for data manipulation, helping data scientists unravel insights effortlessly.

5.3. Matplotlib

Matplotlib transforms Python into an artist's canvas, allowing users to create visually appealing data visualizations and plots, making data storytelling an engaging experience.

5.4. Seaborn

Seaborn, like an artist's palette, enhances Matplotlib by providing a range of customizable charting options and simplifying data visualization, enabling data explorers to paint vivid insights.

5.5. SciPy

SciPy is Python's scientific Swiss Army knife, extending NumPy with powerful functions for scientific computing, statistics, and signal processing, making it indispensable for researchers.

5.6. Scikit-learn

Scikit-learn is the Python companion for aspiring machine learning enthusiasts, offering a treasure trove of machine learning algorithms, empowering data scientists to build intelligent models with ease.

5.7. NLTK(Natural Language ToolKit)

NLTK acts as a language guide for Python, aiding developers in natural language processing tasks with its robust text processing libraries, making language understanding a breeze.

5.8. LightGBM

LightGBM is the speedster of machine learning, excelling in handling large datasets with lightning-fast training, minimal memory usage, and exceptional accuracy, catering to data scientists' need for efficiency.

6. Implementation

6.1. Data Visualization

Various visualizations used in this project are :

I. Correlation between the attributes

Correlation states the strong and weak bond between the columns respectively. So that it helps us while we are filling the missing values, to do feature engineering etc.

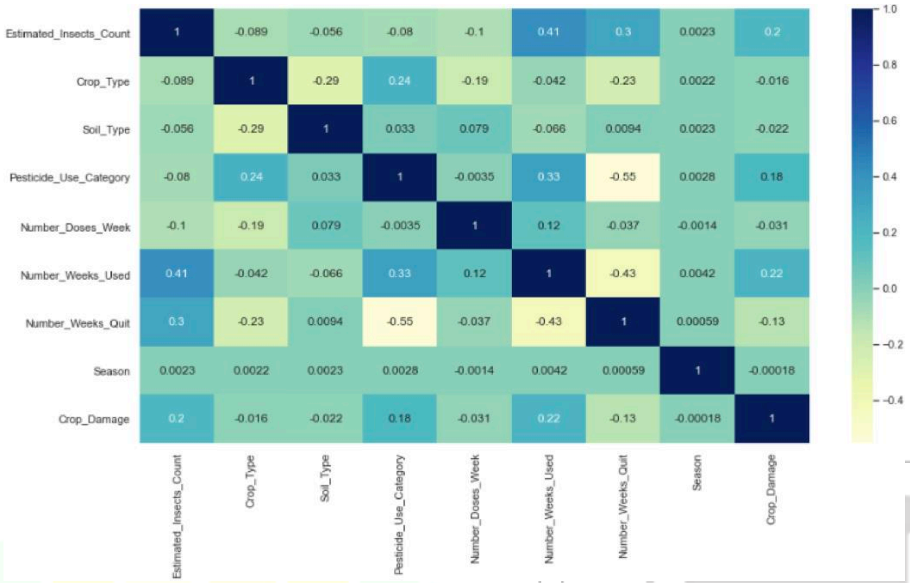


figure 5 : correlation heatmap

II. Univariate Analysis

A. Crop Damage Grouped Count

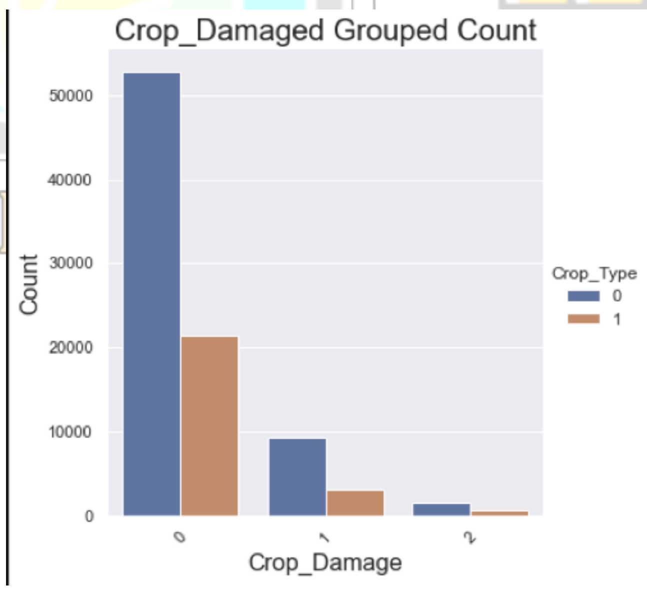


figure 6 : crop_damaged group count

Looking at the grouped count graph for crop_type in relation to crop damage, we discern that crop_type 0 experiences more damage compared to crop_type 1.

B. *Estimated Insects Grouped Count*

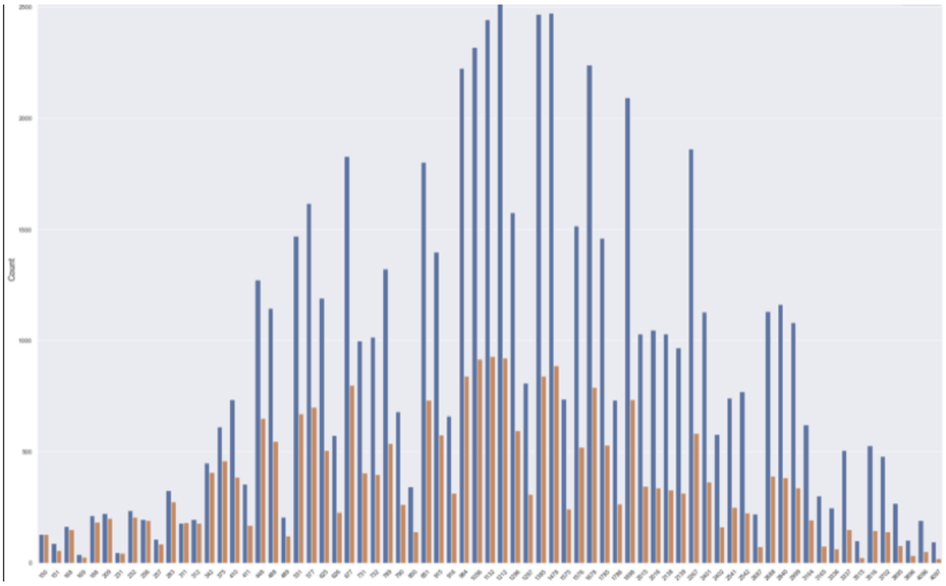


figure 7 : estimated insects grouped count

Moving on to the Estimated Insects Grouped Count graph, we notice that crop_type 0 hosts a larger insect population than crop_type 1.

C. *Soil Type Grouped Count*

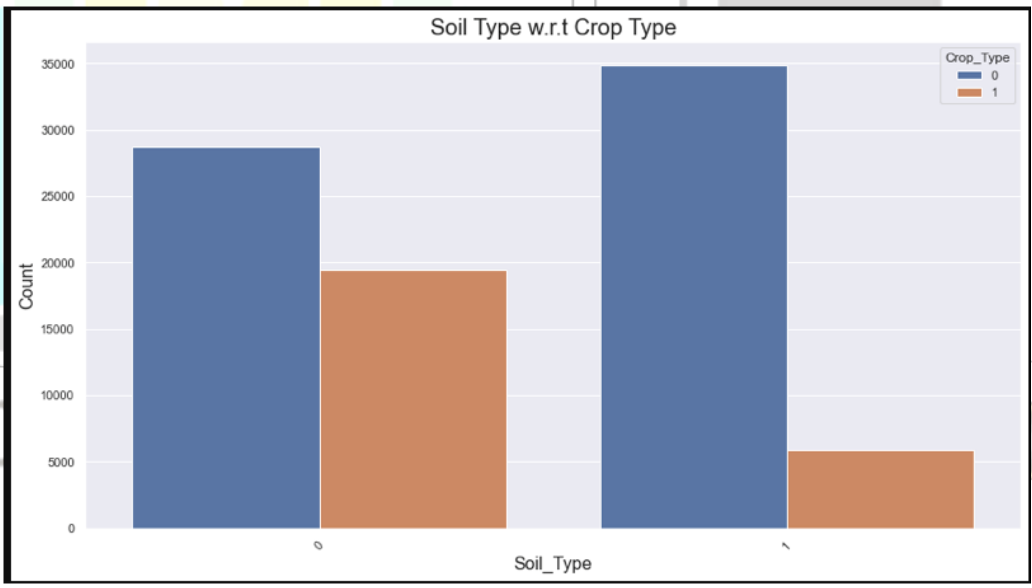


figure 8 : soil typed grouped count

In the Soil Type Grouped Count graph, the results indicate that crop_type 0 performs well in both soil types, unlike crop_type 1.

D. Season Grouped Count

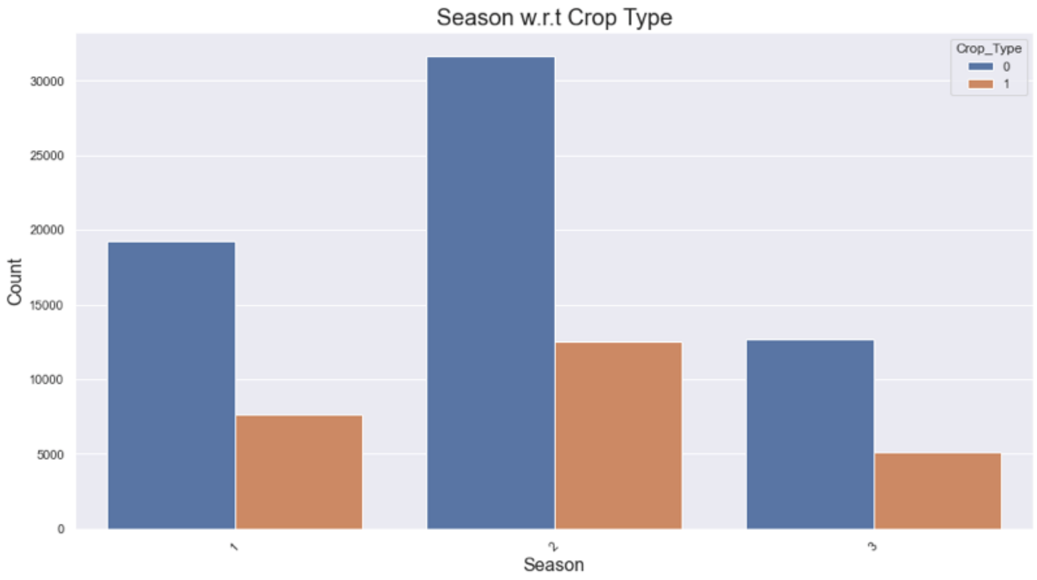


figure 9 : season grouped count

Lastly, the Season Grouped Count graph reveals that type 2 season boasts the highest production for both crop types.

II. Bivariate Analysis

A. Crop Damage Vs Pesticide Use Category

figure 10 : crop damage vs pesticide use category

The above bar graph shows us the crop damage vs Pesticide use category.

B. *Distplot*

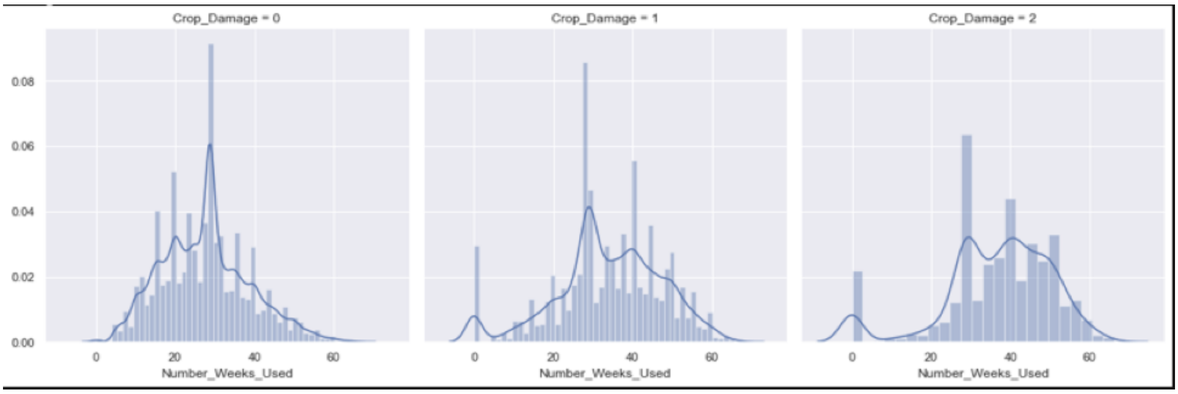


figure 11 : distplot

The above graph shows the distribution of Number weeks used with respect to the dependent class variable.

C. *Boxplot with outliers present*

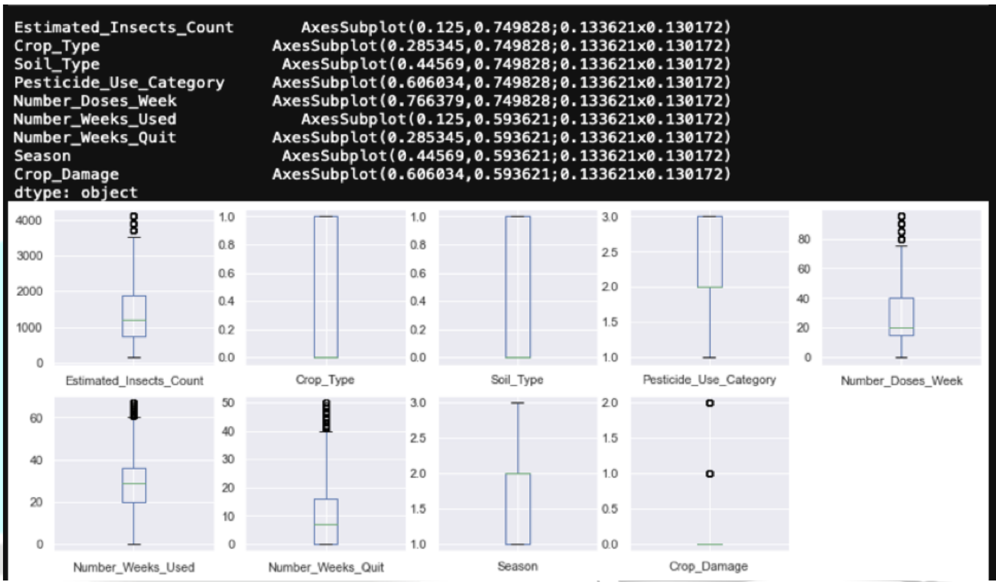
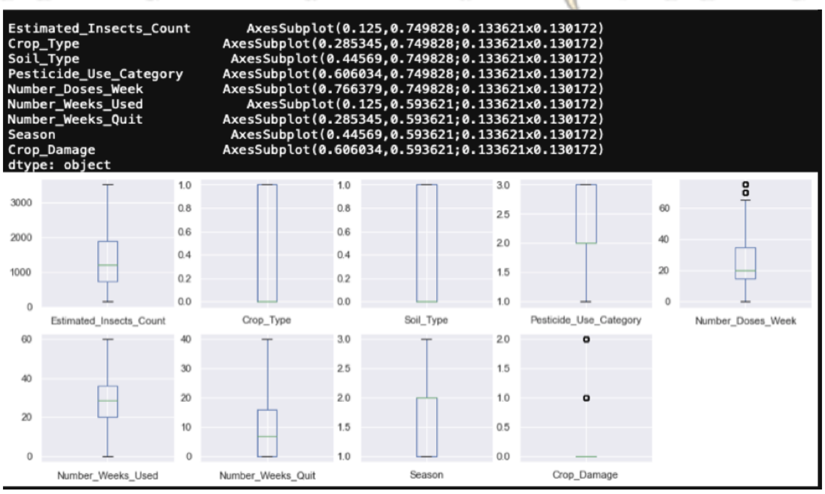


figure 12 : boxplot with outliers present



D. *Boxplot after removing the outliers*

figure 13 : boxplot after removing the outliers

6.2. Skew Analysis

Skewness refers to the deviation from a symmetrical, bell-shaped curve, known as a normal distribution, in a dataset. When the curve shifts to the left or right, it indicates skewness. It quantifies how much a distribution deviates from a standard normal distribution, where zero skewness signifies perfect symmetry, while a lognormal distribution, for instance, demonstrates a degree of right-skewness.

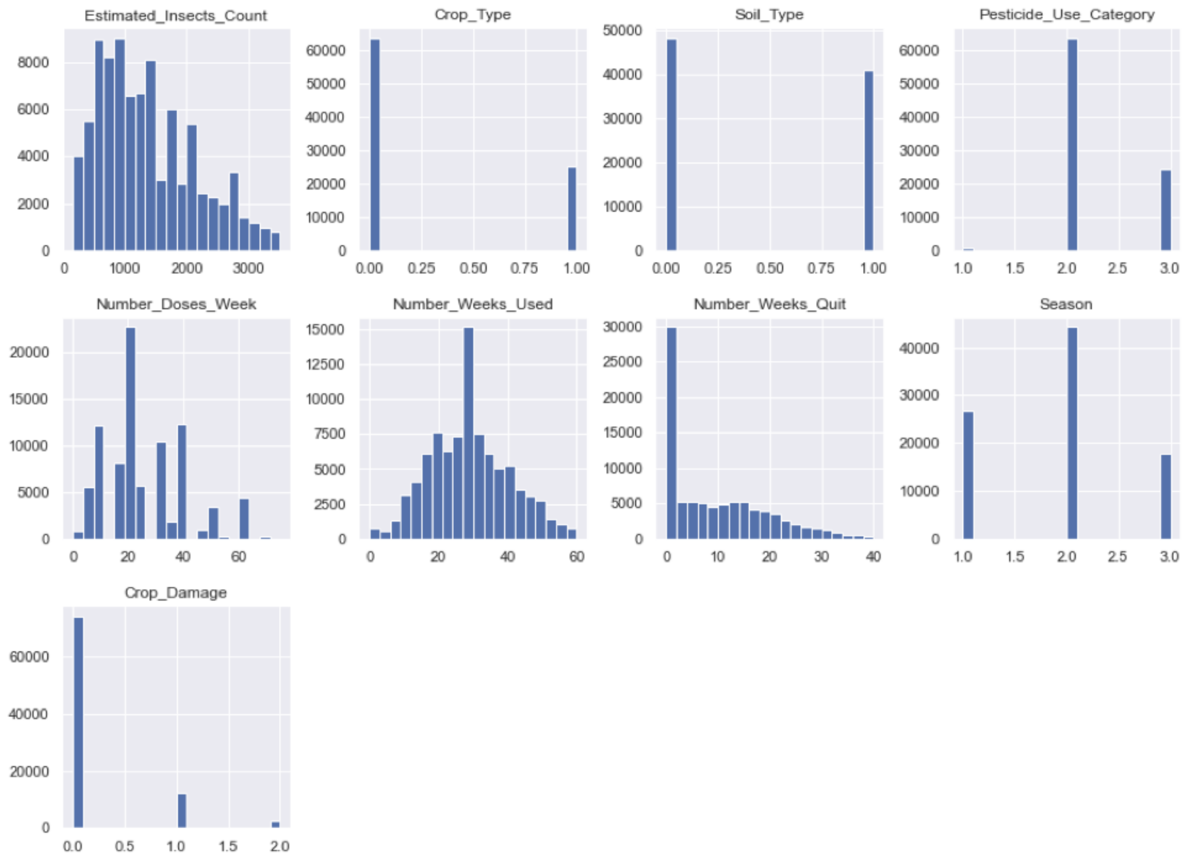


figure 14 : skew analysis

After preprocessing the data and removing the outliers it is clear that the data is normally distributed. We can now move further to train the model.

6.3. Training Harvest Prediction Model

The harvest model is trained in two approaches :

Approach 1 :

In this initial approach, we utilize K-nearest neighbors (KNN), Light GBM, and AdaBoost.

KNN (K-nearest neighbors): KNN is a supervised learning algorithm used for both regression and classification. It predicts the class of test data by measuring the distance between the test data point and all the training points, then selecting the K closest points.

```
Results for model :  KNeighbors Classifier
max accuracy score is 0.8442913654344564
Mean accuracy score is :  0.8420400913398536
Std deviation score is :  0.0015047495269905472
Cross validation scores are :   [0.84396804 0.84256133 0.84115463 0.8428901  0.83962636]
```

figure 15 : knn results

Light GBM: Light GBM is a high-speed, distributed gradient boosting framework based on decision tree algorithms. It finds applications in ranking, classification, and various machine learning tasks.

```
0.8471190636957011
[[14611   237     0]
 [ 2018   443     0]
 [  328   134     1]]
              precision    recall  f1-score   support

           0       0.86      0.98      0.92     14848
           1       0.54      0.18      0.27      2461
           2       1.00      0.00      0.00       463

    accuracy                           0.85     17772
   macro avg       0.80      0.39      0.40     17772
weighted avg       0.82      0.85      0.81     17772
```

figure 16 : lightgbm results

AdaBoost: AdaBoost is an ensemble learning method designed to enhance the performance of binary classifiers. It employs an iterative process to learn from the weaknesses of weak classifiers, ultimately transforming them into strong ones.

```
0.8407607472428539
[[14827    21     0]
 [ 2346   115     0]
 [  427    36     0]]
              precision    recall  f1-score   support

           0       0.84      1.00      0.91     14848
           1       0.67      0.05      0.09      2461
           2       0.00      0.00      0.00       463

    accuracy                           0.84     17772
   macro avg       0.50      0.35      0.33     17772
weighted avg       0.80      0.84      0.78     17772
```

figure 17 : adaboost results

Approach 2 :

The second approach involves data manipulation with lags and training the model using Light GBM, offering greater precision compared to the previous approach.

Lags: Lags are created using the Pandas library, allowing us to shift a column's values up or down by a specified number of positions.

```
df_data['Crop_Damage_lag1'] = df_data['Crop_Damage'].shift(fill_value=-999)
df_data['Estimated_Insects_Count_lag1'] = df_data['Estimated_Insects_Count'].shift(fill_value=-999)
df_data['Crop_Type_lag1'] = df_data['Crop_Type'].shift(fill_value=-999)
df_data['Soil_Type_lag1'] = df_data['Soil_Type'].shift(fill_value=-999)
df_data['Pesticide_Use_Category_lag1'] = df_data['Pesticide_Use_Category'].shift(fill_value=-999)
df_data['Number_Doses_Week_lag1'] = df_data['Number_Doses_Week'].shift(fill_value=-999)
df_data['Number_Weeks_Used_lag1'] = df_data['Number_Weeks_Used'].shift(fill_value=-999)
df_data['Number_Weeks_Quit_lag1'] = df_data['Number_Weeks_Quit'].shift(fill_value=-999)
df_data['Season_lag1'] = df_data['Season'].shift(fill_value=-999)

df_data['Crop_Damage_lag2'] = df_data['Crop_Damage'].shift(periods=2,fill_value=-999)
df_data['Estimated_Insects_Count_lag2'] = df_data['Estimated_Insects_Count'].shift(periods=2,fill_value=-999)
df_data['Crop_Type_lag2'] = df_data['Crop_Type'].shift(fill_value=-999)
df_data['Soil_Type_lag2'] = df_data['Soil_Type'].shift(fill_value=-999)
df_data['Pesticide_Use_Category_lag2'] = df_data['Pesticide_Use_Category'].shift(periods=2,fill_value=-999)
df_data['Number_Doses_Week_lag2'] = df_data['Number_Doses_Week'].shift(periods=2,fill_value=-999)
df_data['Number_Weeks_Used_lag2'] = df_data['Number_Weeks_Used'].shift(periods=2,fill_value=-999)
df_data['Number_Weeks_Quit_lag2'] = df_data['Number_Weeks_Quit'].shift(periods=2,fill_value=-999)
df_data['Season_lag2'] = df_data['Season'].shift(periods=2,fill_value=-999)
```

```
clf = lgb.LGBMClassifier(**params)

clf.fit(df_train[feature_cols], df_train[label_col], eval_metric='multi_error', verbose=False, categorical_feature=cat_

# eval_score_auc = roc_auc_score(df_train[label_col], clf.predict(df_train[feature_cols]))
eval_score_acc = accuracy_score(df_train[label_col], clf.predict(df_train[feature_cols]))

print('ACC: {}'.format(eval_score_acc))

ACC: 0.9911881878952936
```

figure 18 : approach 2 results

6.4. Training Crop Prediction Model

Various models are employed to train the crop prediction model.

Model 1: Decision Tree
Decision Tree is a robust and widely-used algorithm in supervised learning. It accommodates both continuous and categorical output variables, making it versatile for various scenarios.

```
from sklearn.tree import DecisionTreeClassifier

DecisionTree = DecisionTreeClassifier(criterion="entropy",random_state=2,max_depth=5)

DecisionTree.fit(Xtrain,Ytrain)

predicted_values = DecisionTree.predict(Xtest)
x = metrics.accuracy_score(Ytest, predicted_values)
acc.append(x)
model.append('Decision Tree')
print("DecisionTrees's Accuracy is: ", x*100)

print(classification_report(Ytest,predicted_values))

DecisionTrees's Accuracy is:  90.0
```

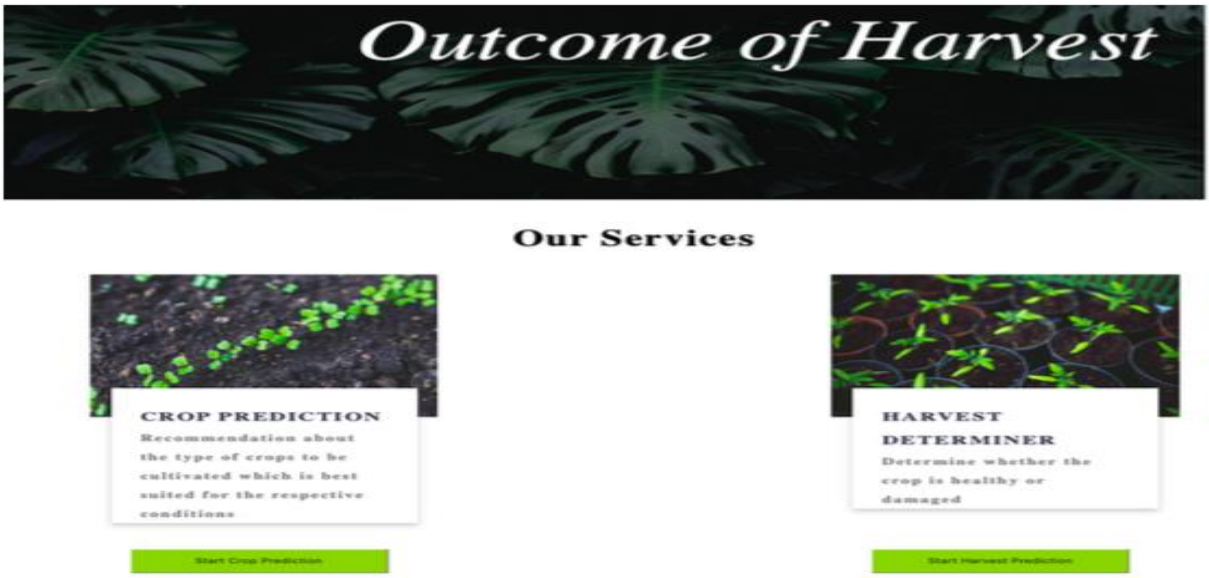
figure 19 : model 1 result

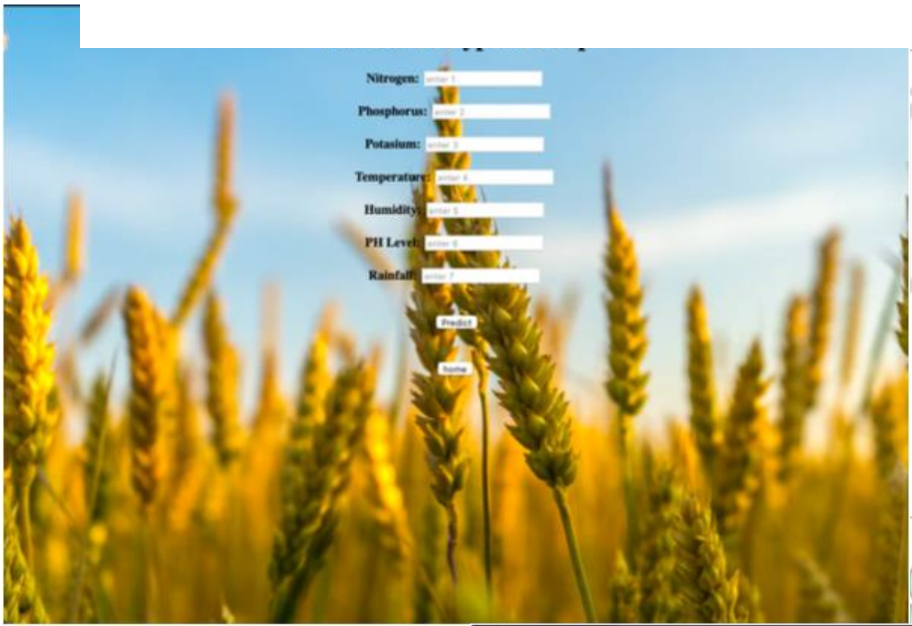
Model 2: Random Forest
Random Forest is a powerful supervised machine learning algorithm suitable for classification and regression tasks. It operates by constructing decision trees on diverse samples and combining their results through majority voting or averaging.

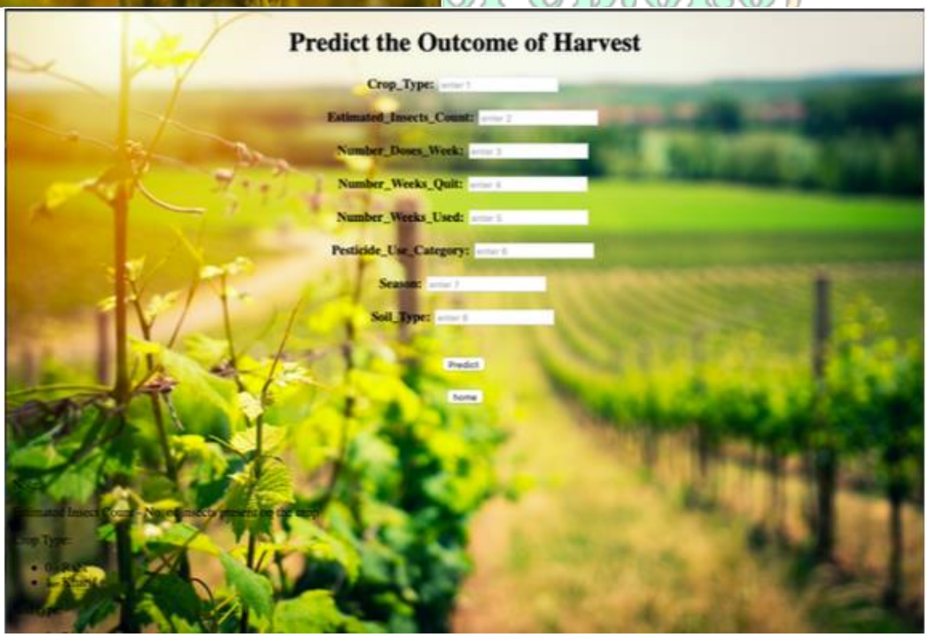
```
from sklearn.ensemble import RandomForestClassifier

RF = RandomForestClassifier(n_estimators=20, random_state=0)
RF.fit(Xtrain,Ytrain)

predicted_values = RF.predict(Xtest)

x = metrics.accuracy_score(Ytest, predicted_values)
acc.append(x)
model.append('RF')
print("RF's Accuracy is: ", x)

print(classification_report(Ytest,predicted_values))

RF's Accuracy is:  0.990909090909091
```

figure 20 : model 2 result

6.5. Results


Outcome of Harvest
Our Services
CROP PREDICTION
Recommendation about the type of crops to be cultivated which is best suited for the respective conditions
HARVEST DETERMINER
Determine whether the crop is healthy or damaged
Start Crop Prediction
Start Harvest Prediction


Nitrogen:
Phosphorus:
Potassium:
Temperature:
Humidity:
PH Level:
Rainfall:
Predict
home


Predict the Outcome of Harvest
Crop_Type:
Estimated_Insects_Count:
Number_Doses_Week:
Number_Weeks_Quit:
Number_Weeks_Used:
Pesticide_Use_Category:
Season:
Soil_Type:
Predict
home

7. Results and Discussions

7.1. Harvest Model Prediction

Predicting the outcome of the harvest season involves determining the status of crop plantations at the season's end. We tackled this challenge through machine learning, conducting a comparative analysis using various algorithms: Random Forest Classifier, K-Nearest Neighbor, Decision Tree Classifier, Gaussian NB, Ada-Boost, LightGBM, XgBoost. After evaluating each algorithm's performance metrics, we found that LightGBM outperformed the others, achieving the highest accuracy at 84.6%. However, we sought further improvement.

Table 1 : Results of Approach 1

S.no	Algorithm	Accuracy	Precision	Recall	F1-Score
1	Random Forest Classifier	82.2%	0.74	0.81	0.73
2	K-Nearest Neighbors	84.3%	0.79	0.84	0.80
3	Decision Tree Classifiers	75.3%	0.66	0.71	0.78
4	Gaussian NB	82.4%	0.75	0.80	0.74
5	Adaboost	84%	0.79	0.84	0.77
6	Lightgbm	84.6%	0.82	0.85	0.81
7	Xgboost	80%	0.72	0.70	0.65

In our second approach, we enriched the dataset by introducing new columns. These included lagged data columns to highlight trends and some variations in the data. We then repeated the process of exploratory data analysis, feature engineering, and algorithm testing. This time, our model exhibited significantly improved performance, with LightGBM again leading with an accuracy of 97.2%.

Table 2 : Results of Approach 2

S.No	Algorithm	Accuracy
1	Lightgbm	97.2%
2	Random Forest Classifier	92.6%
3	K-Nearest Neighbors	93.2%
4	Gaussian NB	77.4%
5	Decision Tree	90.3%
6	Ada Boost	95.4%

7.2. Crop Model Prediction

Crop prediction revolves around determining the most suitable crop to plant based on factors like soil pH, chemical composition, and soil humidity. This information aids farmers in making informed planting decisions. Our dataset comprised 7 independent variables and 1 dependent variable.

Through data inspection, exploratory analysis, and model development, we achieved an impressive final accuracy of 99% with an average weighted precision, recall, and F1-Score of 0.99.

Table 3 : Results of Crop Model

S.No	Algorithm	Accuracy
1	Random Forest Classifier	99.1%
2	Logistic Regression	95%
3	Naive Bayes	99%
4	Decision Tree	90%
5	Support Vector Machine	10.6%
6	Xgboost	98%

8. Conclusions and Future Scope

The harvest season's outcome prediction empowers farmers to optimize pesticide usage and address potential insect issues, ultimately enhancing crop health and efficiency. In our dataset, the dependent variable has three class labels: 0 for a live crop, 1 for damage due to pesticides, and 2 for damage from other causes. After thorough data inspection, exploratory analysis, and model building, we achieved an accuracy of 97% with an average weighted precision, recall, and F1-Score of 0.97.

Looking ahead, our research aims to explore additional methods for crop protection, such as managing field water levels to prevent flooding and reduce farmer debts. Farmers are the backbone of our country, and by supporting their livelihoods, we contribute to a smoother life for everyone.

9. References

[1] Sk Al Zaminur Rahman, S.M. Mohidul Islam, Kaushik Chandra Mitra,‖ Soil Classification using Machine Learning Methods and Crop Suggestion Based on Soil Series‖,2018 21st International Conference of Computer and Information Technology (ICCIT), 21-23 December, 2018.
[2] S. Panchamurthi. M.E.,M.D. Perarulalan,A. Syed Hameeduddin,P. Yuvaraj,‖Soil Analysis and Prediction of Suitable Crop for Agriculture using Machine Learning‖, International Journal for Research in Applied Science & Engineering Technology(IJRASET),ISSN:2321-9653;IC Value:45.98;SJ Impact Factor:6.887,Volume 7 Issue III,Mar 2019.
[3] D Ramesh,B Vishnu Vardhan,―Data Mining Techniques and Applications to Agricultural Yield Data,‖ International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 9, September 2013.
[4] https://lightgbm.readthedocs.io/en/latest/pythonapi/lightgbm.LGBMClassifier.html
[5] https://machinelearningmastery.com/boosting-and-adaboost-for-machine-learning/
[6] https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.GaussianNB.html
[7] https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html
[8] https://xgboost.readthedocs.io/en/latest/