Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

AI and Machine Learning for Optimal Crop Yield Optimization in the USA

MD Rokibul Hasan
MBA, Gannon University, Erie, PA, USA
Corresponding Author: MD Rokibul Hasan, **E-mail**: prorokibulhasanbi@gmail.com

| ABSTRACT

The agricultural sector plays a paramount role in the economy of the United States, contributing significantly to the GDP and affirming sustainability for American residents. This study explored the application of Artificial Intelligence and Machine Learning techniques in maximizing crop yields in America. This research employed various software tools, comprising Python programming language, Pandas library for data manipulation and analysis, Scikit-learn library for machine learning models and evaluation metrics, and LIME library for explainable AI. The crop yield datasets for the current research were sourced from Kaggle. This dataset provided substantial insights regarding crop cultivation practices within the USA context. This study proposes the "XAI-CROP" algorithm, which is a novel explainable artificial intelligence (XAI) model developed particularly to reinforce the interpretability, transparency and trustworthiness of crop recommendation systems (CRS). From the experimentation, the XAI-CROP model excelled at forecasting crop yield, as demonstrated by its lowest MSE value of 0.9412, suggesting minimal errors. Besides, Its MAE of 0.9874 suggests an average error of less than 1 unit in forecasting crop yield. Furthermore, the R2 value of 0.94152 suggests that the algorithm accounts for 94.15% of the data's variability.

| KEYWORDS

Crop Yield; Machine Learning; Python; Gradient Boosting (GB); Random Forest (RF); Decision Tree (DT).

| ARTICLE INFORMATION

ACCEPTED: 02 April 2024 **PUBLISHED:** 20 April 2024 **DOI:** 10.32996/jcsts.2024.6.2.6

1. Introduction

According to the IRJET Journal (2023), the agricultural industry plays an instrumental role in the economy of America, contributing substantially to the Gross Domestic Product (GDP) and providing sustainability for American citizens. Besides, agriculture is undoubtedly a major sector in the USA, with an estimated $150 billion in exports annually. Nonetheless, unpredictable weather trends because of climate change and other environmental components have made maximizing crop yields a significant challenge. Maximizing crop yields while maintaining sustainable practices is a persistent challenge faced by farmers. This study delves into the application of Artificial Intelligence and Machine Learning techniques in maximizing crop yields in America. Particularly, the research paper explores the present state of crop yield optimization methods and the possible benefits of employing AI and ML and provides insights into future research directions.

Traditional crop yield optimization methods entail various practices, such as soil management, irrigation strategies, crop rotation, and pest control tactics. These methodologies are premised on experiential experience and knowledge attained over time. Nevertheless, they may not be capable of completely accounting for the complicated association between various factors impacting crop yields, such as soil composition, weather trends, and plant genetics (IRJET Journal, 2023). Recent innovations in precision agriculture have developed technologies such as GPS-guided machinery, remote sensing, and yield monitoring systems. These inventions have enabled the gathering of large volumes amounts of data associated with soil conditions, crop growth, and environmental aspects. Nevertheless, the efficient employment of this data for maximizing crop yields remains a noteworthy challenge, underscoring the need for advanced analytical tools and decision-support systems (Al-Adhaileh & Aldhyani, 2022).

2. Literature Review of Related Works

As per Al-Adhaileh & Aldhyani (2022), climate change comprises persistent changes in global or weather patterns and local temperatures, with greenhouse gas reduction and global warming presenting challenges because of regulatory and legal complexities. The effect of climate change is anticipated to culminate in increased food insecurity, malnutrition, and hunger, particularly affecting populations in remote areas in the USA and small islands. Therefore, local government confronts a substantial threat from climate change, such as temperatures, air quality, and weather conditions, and impacts agricultural productivity and soil composition. As such, it is fundamental for the current administration to craft strategies to combat environmental consequences and safeguard crop yields.

Researchers globally are progressively exploring crop yield prediction techniques. For instance, Suvarna (2022) proposed a deep learning framework employing remote sensing data to predict crop yields in developed countries. In particular, their technique consolidated a Convolutional Neural Network (CNN) with a Gaussian protocol component and dimensionality minimization method. By employing this technique in a soybean dataset incorporating sensing, soil, and climate data from the United States, they successfully minimized the Root Mean Square Error (RMSE) of the framework. Besides, the RMSE was enhanced from 6.27 to 5.83 on average with the Long Short-Term Memory algorithm and from 5.77 to 5.57 with the CNN algorithm.

Another research by Palanivel & Surianarayanan (2019) consolidated machine learning with agronomic rationales to tailor a baseline framework for large-scale crop yield forecasting. They ranked modularity, accuracy, and reusability in their workflow, producing features from weather data, remote sensing, crop simulation outputs, and soil information attained from the MARS Crop Yield Forecasting System (MCYFS) database.

Qin et al. (2023), in their study, adopted Support Vector Regression (SVR), Gradient Boosting, and k-nearest Neighbors to predict crop yields of distinct crops in Germany, the Netherlands, and France. They employed a multilevel deep learning framework consolidating Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) to capture spatial and temporal features. Their research targeted to examine the algorithm's efficacy in forecasting Corn Belt yields in America and examine the impacts of different datasets on the forecasting task. Experiments were performed in the US Corn Belt states utilizing time-series remote sensing data and soil property datasets as inputs.

On the other hand, Shams et al. (2024) performed research to investigate the effects of consolidating machine learning and crop modeling on enhancing corn yield forecasting in the American Corn Belt. They targeted to pinpoint the most accurate hybrid algorithm consolidation, ascertain which crop modeling component should be incorporated with machine learning, and inspect the impact of weather data and simulation crop framework variables on yield prediction accuracy. The research ascertained that enhancing simulation crop framework variables as input features to machine learning algorithm minimized yield prediction RMSE by 7-20%. The researcher recommended that for better yield forecasting, their proposed machine learning framework necessitates more hydrological inputs.

3. Methodology

This research adopted various software tools, comprising Python programming language, Pandas library for data manipulation and analysis, Scikit-learn library for machine learning models and evaluation metrics, and LIME library for explainable AI. Python was selected because of its user-friendliness, versatility, machine learning libraries, and abundance of data analysis (proAIrokibul, 2024). On the other hand, LIME was employed to reinforce the interpretability of machine learning algorithms, facilitating investigators to understand the prediction generation process.

3.1 Dataset

The crop yield datasets for the current research were sourced from Kaggle. This dataset provided substantial insights regarding crop cultivation practices within the USA context. By utilizing this dataset, a suitable model for crop recommendation was designed. Including key attributes such as production per square kilometer, location, season, area, and crop type, these datasets played a paramount role in predicting the optimal crops for cultivation in particular regions premised on historical trends (proAIrokibul, 2024). Given its extensive nature, this dataset acted as a cornerstone for machine learning software in agriculture, enabling the robust training and validation of models.

Data Pre-Processing

- ***Input:***
  - *Collected data*
- ***Output:***
  - *Preprocessed data*
- ***Steps:***

```
// Step 1: Collect input data
soil_data = read_csv('soil_data.csv')
weather_data = read_csv('weather_data.csv')
crop_yield_data = read_csv('crop_yield_data.csv')
// Step 2: Data cleaning
soil_data = remove_duplicates(soil_data)
weather_data = remove_missing_values(weather_data)
crop_yield_data = remove_duplicates(crop_yield_data)
// Step 3: Data transformation
soil_data = convert_categorical_variables(soil_data)
weather_data = convert_categorical_variables(weather_data)
// Step 4: Data integration
crop_data = join_datasets(soil_data, weather_data, crop_yield_data)
// Step 5: Data normalization
crop_data = normalize_data(crop_data)
```

The data preprocessing segment was responsible for gathering, cleaning, and processing obtained data utilizing distinct technologies such as cameras, sensors, and IoT devices to collect real-time data. Once gathered, the data went through cleaning and processing to remove disturbances or inaccuracies (proAIrokibul, 2024). In particular, the data pre-processing adhered to the following distinctive steps within the Data Preprocessing protocol

1. ***Gather input data***: Collect weather patterns, soil type, and historical crop yield data from relevant sources.
2. ***Data cleaning:*** Eliminate missing values and duplicates.
3. ***Data transformation***: Convert data into a suitable format, specifically numerical conversion of categorical variables.
4. ***Data Consolidation:*** Consolidate distinct datasets into a joint dataset for analysis.
5. ***Data normalization:*** Standardize data to foster homogenous scaling utilizing techniques such as z-score normalization or min-max scaling.

3.2 Feature Engineering and Selection

The Feature Selection segment comprised six primary stages:

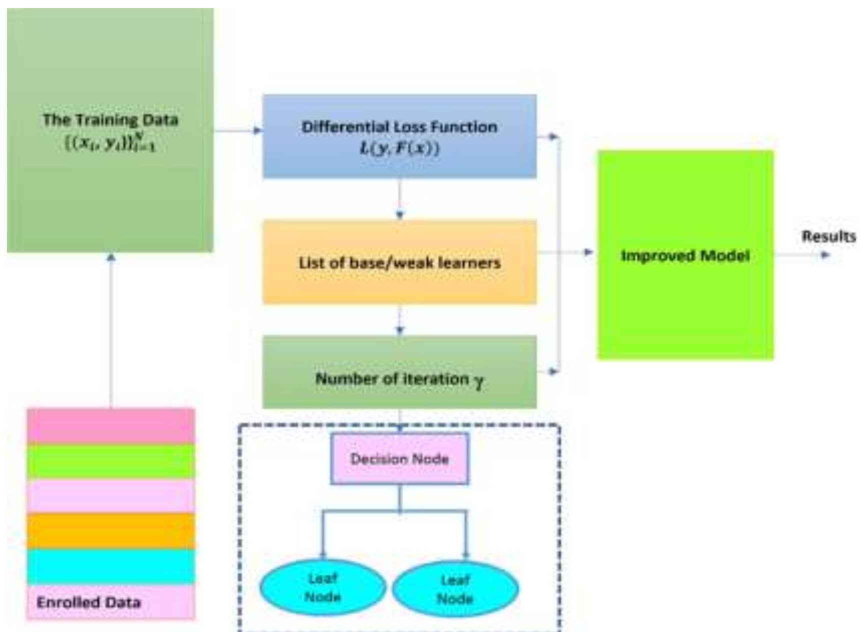
1. Inserting the preprocessed dataset with weather patterns, soil type, and historical crop production data.
2. Sub-dividing the dataset into testing and training sets.
3. Employing statistical methods such as chi-square test, correlation analysis, and ANOVA to pinpoint substantial features.
4. Applying machine learning methods such as Decision Trees, Gradient Boosting, and Random Forest, to determine significant features.
5. Ranking pinpointed features premised on significance scores produced by the selected machine learning algorithm.
6. Choosing the top n features with the greatest significance scores as input for the XAI-CROP model.

3.3 Models and Hyperparameters

The suggested model's hyperparameters were reinforced by employing Decision Tree (DT), Random Forest (RF), and Gradient Boosting (GB).

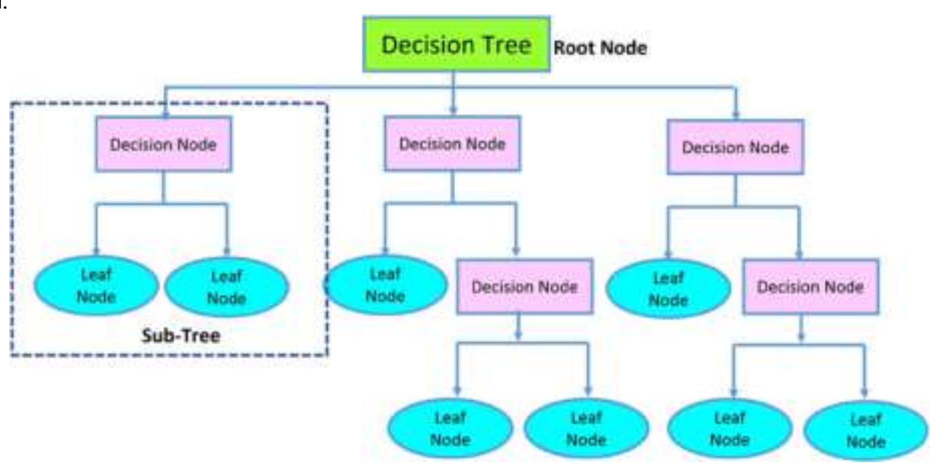
3.3.1 Gradient Boosting (GB)

As per pro-AI-Rokibul (2024), Gradient Boosting Machines are a prominent ensemble machine learning framework that transforms weak learners into solid learners via sequential training. It reduces the loss function in every repetition by calculating gradients and modifying predictions. Contrary to AdaBoost, Gradient Boosting alters weights based on gradients, making it more comprehensive and less sensitive to outliers. As regards Gradient Boosting, the algorithm is trained systematically to reduce loss operations by utilizing the Gradient Descent (GD) optimizer. The learning ratio and n estimators were two pivotal hyperparameters in the Gradient Boosting Decision Trees. The learning rate determined the contribution of every new tree, while n estimators specified the number of trees in the algorithm. The model's hyperparameters substantially enhanced the model's accuracy and performance.



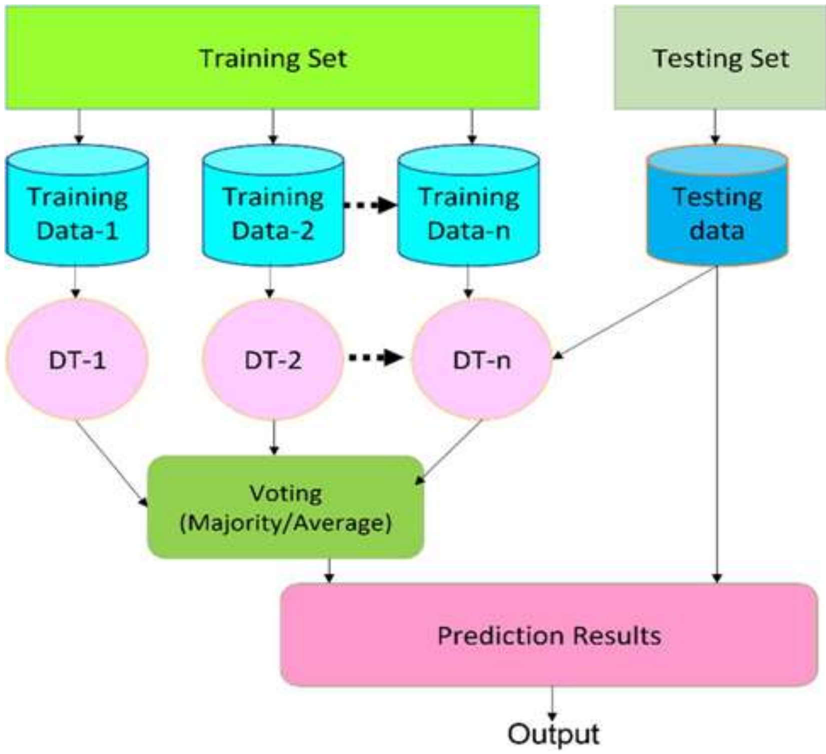
2. Decision Tree (DT)

The Decision Tree (DT) technique is a supervised learning method for regression and classification tasks attributed to its hierarchical and nonparametric structure. The Decision Tree entails an internal node, a root node, leaf nodes, and branches. The Decision Tree learning procedure comprises a comprehensive search for prime split points in a tree, applying a divide-and-conquer dimension(pro-AI-Rokibul, 2024). The subsequent splitting protocol proceeds from top to bottom until entries are classified into particular class labels. The Decision Tree structure is showcased in the following Figure 1, portraying the hierarchical division and classification of data.



3. Random Forest

Random Forest (RF) is a renowned supervised learning model used for regression and classification tasks. It functions by developing decision trees on distinct data samples and combining them via majority voting for categorization and averaging for regression. A noteworthy strength of Random Forest is its capacity to manage datasets constituting both categorical and continuous variables, making it ideal for different classification and regression challenges (Pro-AI-Rokibul, 2024). The algorithm can be illustrated as follows in Figure 2.



Experimentation

Importing Libraries

```
In [1]:
import numpy as np
import seaborn as sns
import pandas as pd
from matplotlib import pyplot as plt
```

```
In [2]:
df = pd.read_csv("crop_yield.csv")
df
```

Output:

Out[2]:

	Crop	Crop_Year	Season	State	Area	Production	Annual_Rainfall	Fertilizer	Pesticide	Yield
0	Arecanut	1997	Whole Year	Assam	73814.0	56708	2051.4	7024878.38	22882.34	0.796087
1	Arhar/Tur	1997	Kharif	Assam	6637.0	4685	2051.4	631643.29	2057.47	0.710435
2	Castor seed	1997	Kharif	Assam	796.0	22	2051.4	75755.32	246.76	0.238333
3	Coconut	1997	Whole Year	Assam	19656.0	126905000	2051.4	1870661.52	6093.36	5238.051739
4	Cotton(lint)	1997	Kharif	Assam	1739.0	794	2051.4	165500.63	539.09	0.420909
...	...	...	...	...	...	...	...	...	...	...
19684	Small millets	1998	Kharif	Nagaland	4000.0	2000	1498.0	395200.00	1160.00	0.500000
19685	Wheat	1998	Rabi	Nagaland	1000.0	3000	1498.0	98800.00	290.00	3.000000
19686	Maize	1997	Kharif	Jammu and Kashmir	310883.0	440900	1356.2	29586735.11	96373.73	1.285000
19687	Rice	1997	Kharif	Jammu and Kashmir	275746.0	5488	1356.2	26242746.82	85481.26	0.016667
19688	Wheat	1997	Rabi	Jammu and Kashmir	239344.0	392160	1356.2	22778368.48	74196.64	1.261818

19689 rows × 10 columns

3.4 Loading and Exploration

As the loading process proceeded, structural transformations were performed to match the data with the input demands of every model.

In [3]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 19689 entries, 0 to 19688
Data columns (total 10 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   Crop             19689 non-null  object
 1   Crop_Year        19689 non-null  int64
 2   Season           19689 non-null  object
 3   State            19689 non-null  object
 4   Area             19689 non-null  float64
 5   Production       19689 non-null  int64
 6   Annual_Rainfall  19689 non-null  float64
 7   Fertilizer       19689 non-null  float64
 8   Pesticide        19689 non-null  float64
 9   Yield            19689 non-null  float64
dtypes: float64(5), int64(2), object(3)
memory usage: 1.5+ MB
```

In [4]:

```
df.describe()
```

Output:

Out[4]:

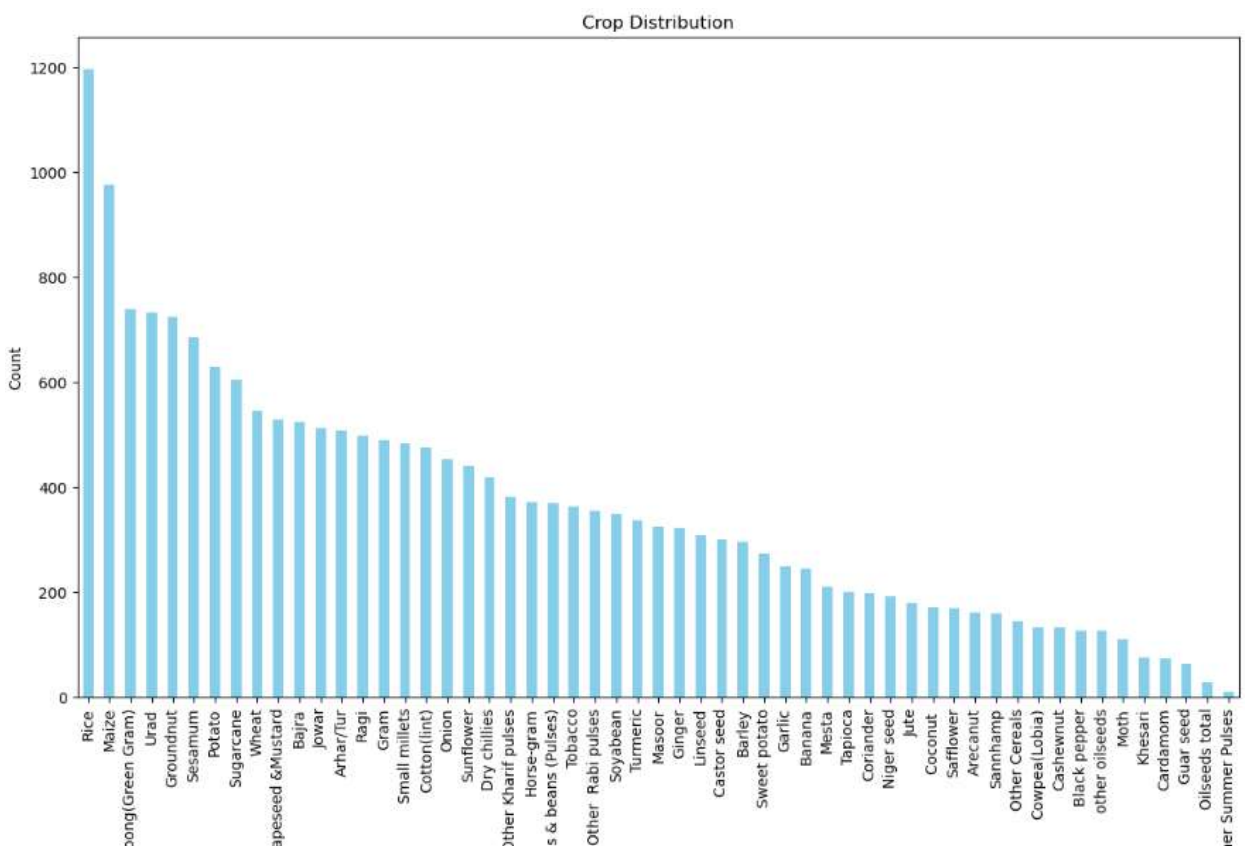
	Crop_Year	Area	Production	Annual_Rainfall	Fertilizer	Pesticide	Yield
count	19689.000000	1.968900e+04	1.968900e+04	19689.000000	1.968900e+04	1.968900e+04	19689.000000
mean	2009.127584	1.799266e+05	1.643594e+07	1437.755177	2.410331e+07	4.884835e+04	79.954009
std	6.498099	7.328287e+05	2.630568e+08	816.909589	9.494600e+07	2.132874e+05	878.306193
min	1997.000000	5.000000e-01	0.000000e+00	301.300000	5.417000e+01	9.000000e-02	0.000000
25%	2004.000000	1.390000e+03	1.393000e+03	940.700000	1.880146e+05	3.567000e+02	0.600000
50%	2010.000000	9.317000e+03	1.380400e+04	1247.600000	1.234957e+06	2.421900e+03	1.030000
75%	2015.000000	7.511200e+04	1.227180e+05	1643.700000	1.000385e+07	2.004170e+04	2.388889
max	2020.000000	5.080810e+07	6.326000e+09	6552.700000	4.835407e+09	1.575051e+07	21105.000000

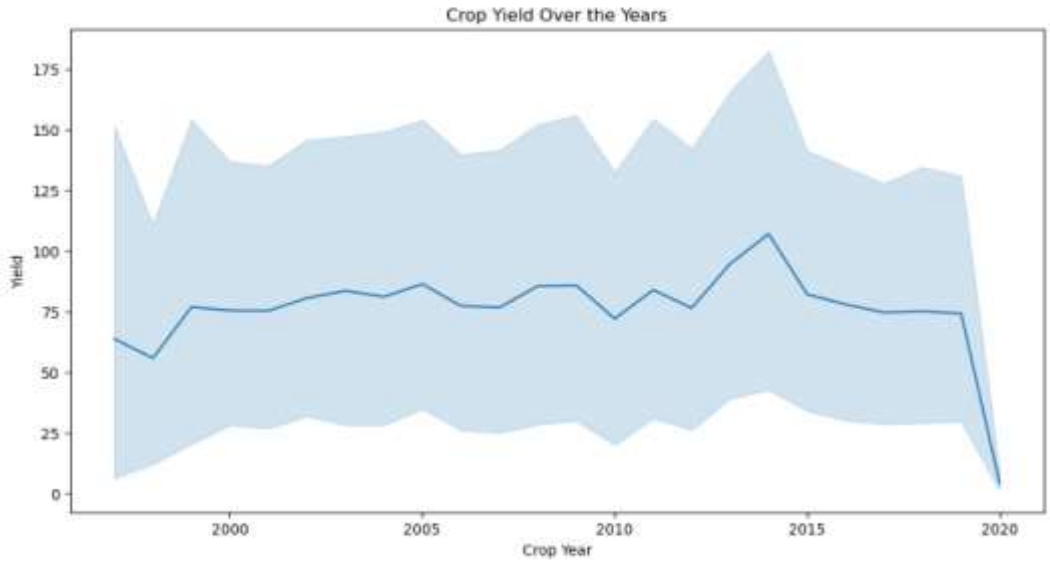
Afterwards, a code snippet was applied to create a pie chart to visualize the distribution of crops in a panda's data frame. The code snippet aimed at generating a histogram showcasing crop distribution:

In [10]:

```
# Visualization 1: Crop Distribution
plt.figure(figsize=(14, 8))
df['Crop'].value_counts().plot(kind='bar', color='skyblue')
plt.title('Crop Distribution')
plt.xlabel('Crop')
plt.ylabel('Count')
plt.show()
```

Output:





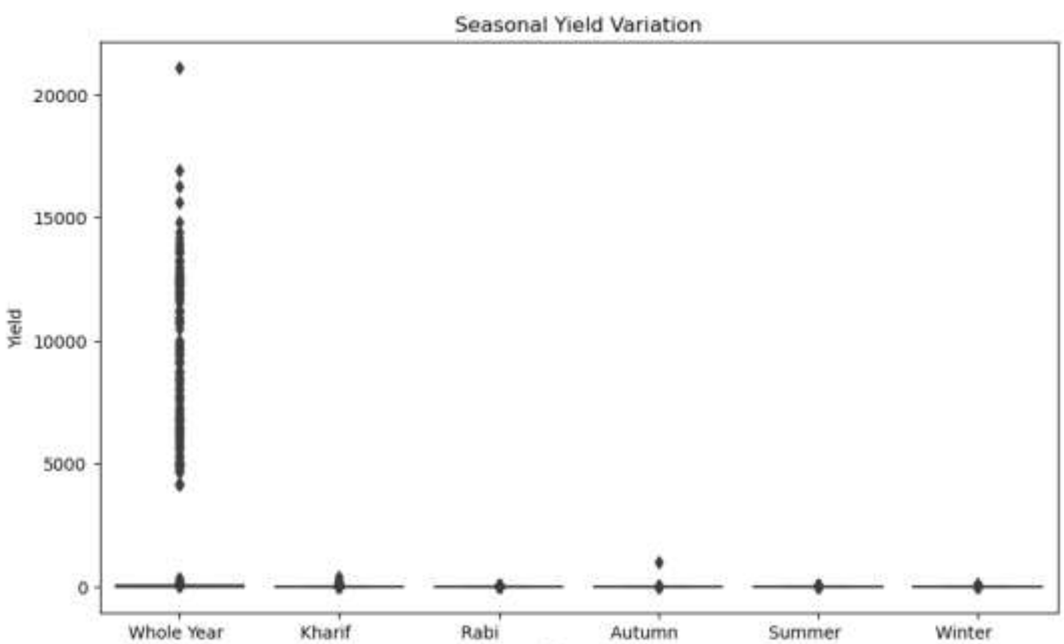
Moreover, a code snippet was applied to ascertain crop yield by year to pinpoint crop patterns and trends:

Output:

Apart from that, the researcher equally aimed to ascertain the seasonal variation of crop yield, which can be showcased by the following code snippet:

In [12]:
```
# Visualization 3: Seasonal Yield Variation
plt.figure(figsize=(10, 6))
sns.boxplot(x='Season', y='Yield', data=df)
plt.title('Seasonal Yield Variation')
plt.xlabel('Season')
plt.ylabel('Yield')
plt.show()
```

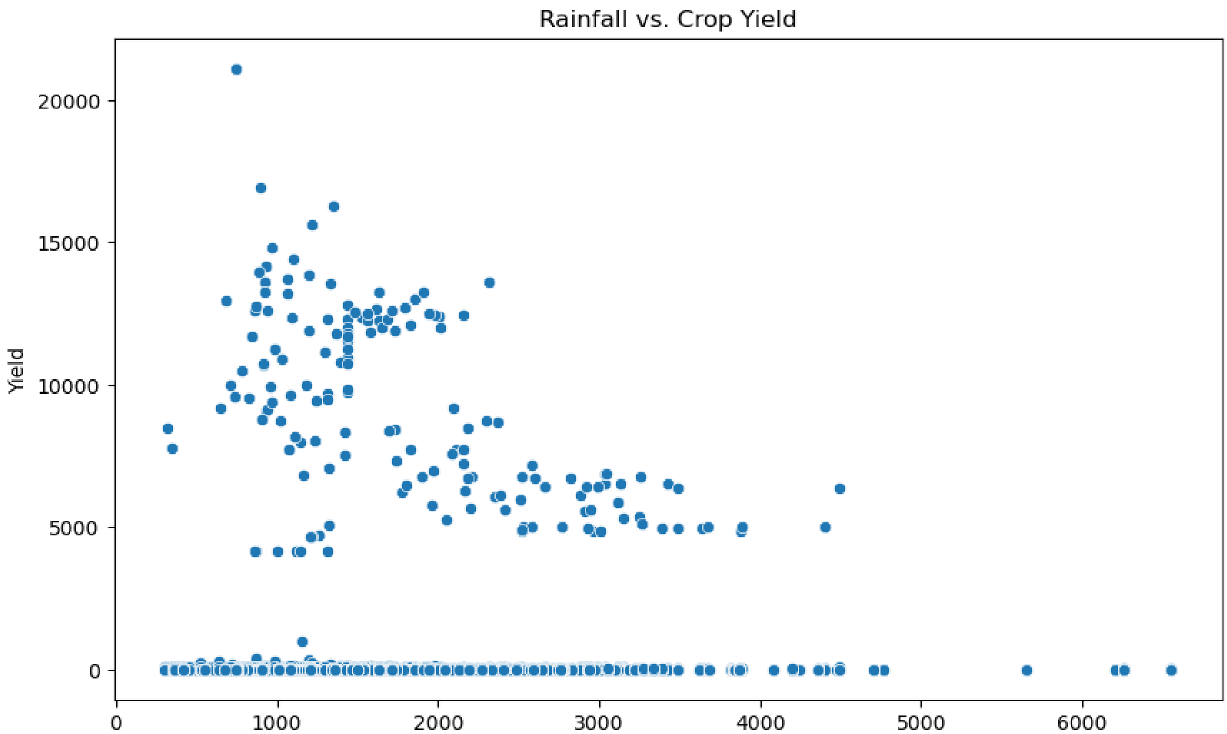
Output:



Furthermore, the researcher applied a code snippet to create a scatter plot visualization to determine the relationship between rainfall and crop yield.

```
In [15]:
# Visualization 6: Rainfall vs. Yield Scatter Plot
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Annual_Rainfall', y='Yield', data=df)
plt.title('Rainfall vs. Crop Yield')
plt.xlabel('Annual Rainfall')
plt.ylabel('Yield')
plt.show()
```

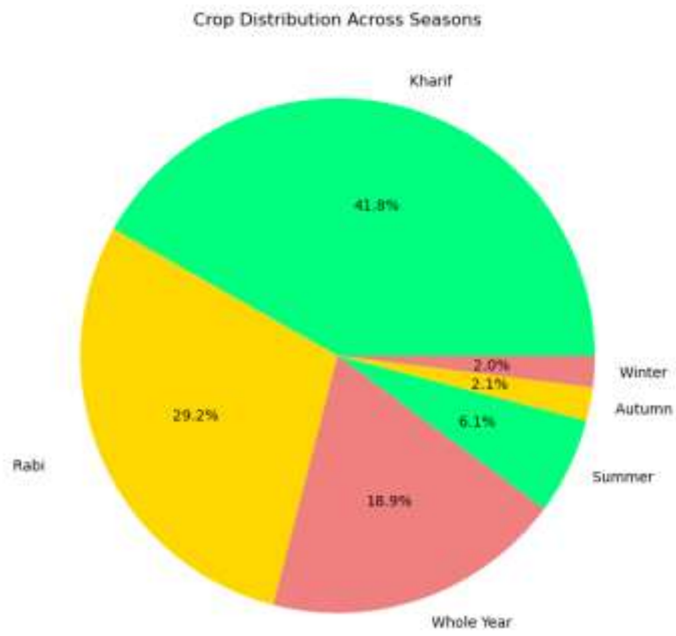
Output:



Last but not least, the analyst attempted to determine the association between the categorical distribution of crop yields across seasons, as illustrated below:

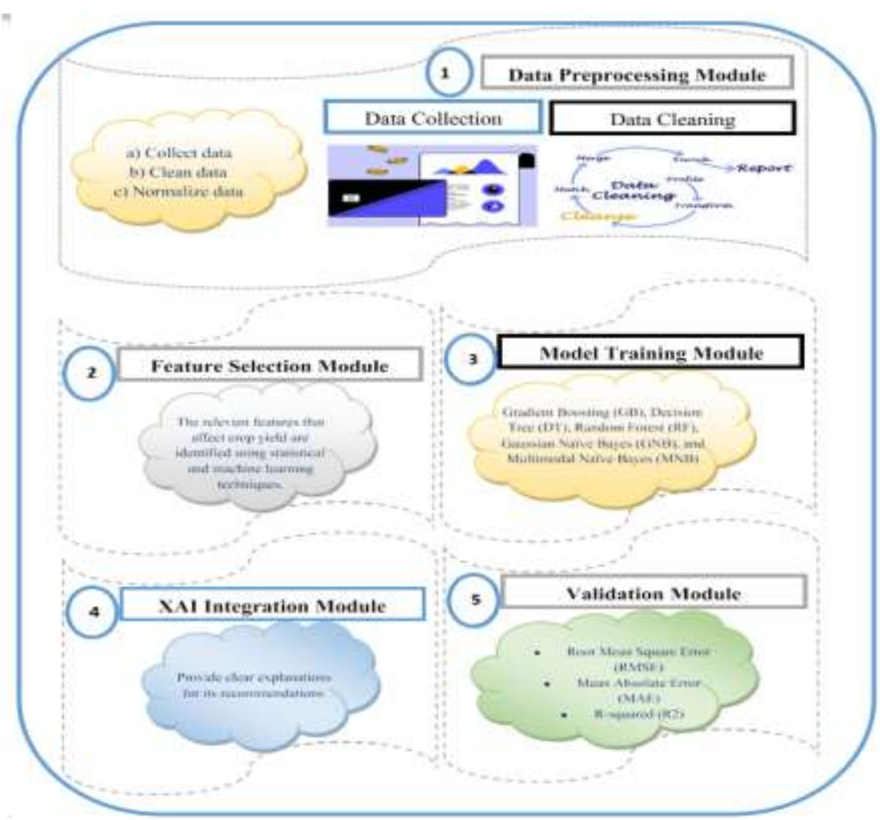
```
In [18]:
# Visualization 9: Categorical Distribution
plt.figure(figsize=(8, 8))
df['Season'].value_counts().plot(kind='pie', autopct='%1.1f%%', colors=['springgreen', 'gold', 'lightcoral'])
plt.title('Crop Distribution Across Seasons')
plt.ylabel('')
plt.show()
```

Output:



3.5 Proposed Model

This study proposes the "XAI-CROP" algorithm, which is a novel explainable artificial intelligence (XAI) model developed particularly to reinforce the interpretability, transparency and trustworthiness of crop recommendation systems (CRS). By integrating explainability correctly from algorithm design and training, XAI-CROP targets to build a new degree of transparency, engagement, and trust between farmers and automated crop recommendation frameworks - eventually facilitating enhanced on-farm decision-making for elevating yields and revenues under changing conditions. The proposed model can be showcased in the following flowchart:



3.6 XAI Consolidation

The XAI consolidation segment of the XAI-CROP module employs LIME to articulate the XAI-CROP framework recommendations using a 6-phae process: (a) load the XAI-CROP algorithm, (2) Choose a validation dataset sample, (3) Develop a raw dataset for local model training, (4) train a linear regression framework on the raw dataset, (5) compute feature weights in the local framework, and (6) produce an explanation outlining the features most impactful to the XAI-CROP prediction for the selected sample.

- ***Steps:***

```
# Load the XAI-CROP model
xai_crop_model = load_model('xai_crop_model.h5')

# Select a sample from the validation dataset for which to generate an explanation
sample = validation_dataset.sample()

# Generate perturbations of the selected sample to create a dataset for local model
training
perturbed_dataset = generate_perturbations(sample, num_perturbations)

# Train a linear regression model on the perturbed dataset
local_model = train_linear_regression(perturbed_dataset)

# Calculate the weight of each feature in the local model
weights = calculate_feature_weights(local_model)

# Generate an explanation by highlighting the features that contribute the most to the
XAI-CROP model's prediction for the selected sample
explanation = generate_explanation(sample, weights)
```

3.7 Performance Metrics

To assess the performance of the proposed model for predicting crop yield, the researcher adopted three widely adopted regression metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), and the coefficient of determination (R-squared). Mean Squared Error (MSE) is a calculation of the average squared distinction between the actual values and the forecasted values. It offers a measure of the overall magnitude of the errors, with fewer values suggesting better performance. MSE is calculated as follows:

$$MSE = (1/n)\ \Sigma\ (y\_i - y\_hat\_I)\ \hat{}2$$

Where n denotes the number of observations, y_i is the actual value for observation i, and y_hat_i is the forecasted value for observation i.

On the other hand, Mean Absolute Error (MAE) is another predominantly adopted metric that calculates the average absolute difference between the actual values and the predicted values. Contrary to MSE, MAE does not square the errors, which can make it simple to interpret. MAE is expressed as follows:

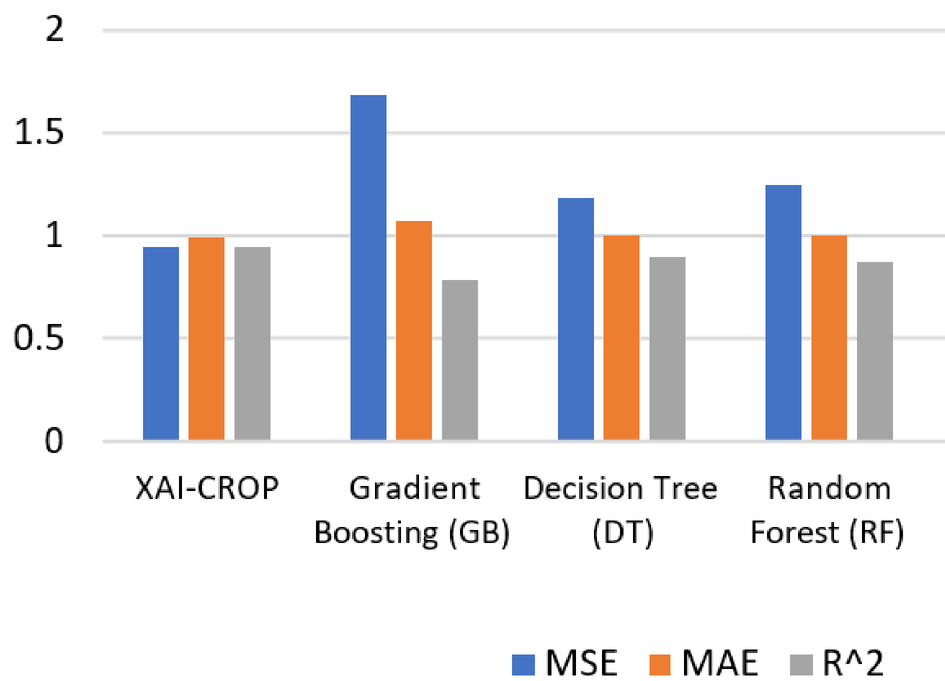
$$MAE = (1/n)\ \Sigma |\ y\_i - y\_hat\_i\ |$$

Where n is the tally of observations, y_i is the actual value for observation i, and y_hat_i is the forecasted value for observation i.

Conversely, R-squared (R^2) is a statistical computation that denotes the measure of the variance in the dependent variable (Spending Score) that can be articulated by the independent variables in the regression framework. R-squared ranges from 0 to 1, with greater values suggesting a better fit of the framework to the data.

$$R\hat{}2 = 1 - (\Sigma\ (y\_i - y\_hat\_i)\ \hat{}2)\ /\ (\Sigma\ (y\_i - y\_bar)\hat{}2)$$

Algorithm	MAE	MSE	R^2
Random Forest (RF)	1.0015	1.2487	0.8745
Decision Tree (DT)	1.0002	1.1785	0.8942
Gradient Boosting (GB)	1.0745	1.6861	0.78521
XAI-CROP	0.9874	0.9412	0.94152



By referring to the above charts, it was evident that the XAI-CROP model excelled at forecasting crop yield, as demonstrated by its lowest MSE value of 0.9412, suggesting minimal errors. Besides, Its MAE of 0.9874 suggests an average error of less than 1 unit in forecasting crop yield. Furthermore, the R2 value of 0.94152 suggests that the algorithm accounts for 94.15% of the data's variability. In contrast, the Decision Tree algorithm had the second-best performance, with an MAE of 1.0002, MSE of 1.1785, and R2 of 0.8942.

4. Business Impact

4.1 Organization

As regards Agricultural organizations, the XAI-CROP algorithm provides a competitive edge by offering better accuracy and transparency in terms of forecasting crop yields, resulting in enhanced resource allocation, better crop management, and elevated operational efficiency. This can lead to cost savings, elevated productivity, and enhanced sustainability for agricultural businesses.

By adopting the proposed XAI-CROP algorithm, companies can optimize the dissemination of resources such as water, labor, land, and fertilizers. The algorithm's insights regarding crop suitability and environmental factors enable more effective resource utilization, diminishing waste and enhancing productivity. This, in turn, results in cost savings and enhanced operational efficiency for the organization.

4.2 Benefits for the Economy of the USA:

1. ***Enhanced Agricultural Productivity***: The XAI-CROP algorithm’s capability to offer accurate and enhanced crop recommendations contributes to supreme agricultural productivity in the United States. By boosting resource allocation, crop selection, and risk mitigation, the algorithm allows farmers and agricultural companies to attain higher crop yields. This enhanced productivity not only satisfies the escalating demands for food but also boosts the export of agricultural produce, resulting in economic growth and stability.

2. ***Sustainable Agriculture***: The proposed XAI-CROP algorithm enhances sustainable agricultural measures by taking into consideration soil characteristics, environmental factors, and historical data. By suggesting crops that are suitable for particular regions, the algorithm assists in optimizing agricultural output while reducing negative effects on the environment. Sustainable practices lead to long-term soil health, water conservation, and reduced reliance on chemical inputs, fostering a more resilient and eco-friendlier agricultural sector.

4.3 How to Use the XAI-Crop Model

1. **Step 1: Define Strategic Objectives-** Businesses are required to first define the objectives of deploying the XAI-CROP algorithm. Pinpoint the specific challenges or problems they want to address, such as enhancing resource allocation, crop selection, or risk mitigation.
2. **Step 2: Data Collection and Preprocessing-** In this stage, the input raw data, which entails weather patterns, soil type, and historical crop yields, are gathered and processed for further analysis.
3. **Step 3: Feature engineering and selection.** In this phase, the pertinent features that influence crop yield are pinpointed using machine learning and statistical techniques. These features are then employed as input for the XAI-CROP algorithm.
4. **Step 4: Model Training-** The XAI-CROP algorithm should train on a dataset of crop production in the USA, which entails information on weather patterns, soil type, crop yield, and historical crop yields. The algorithm is premised on a decision tree algorithm that suggests recommendations premised on the input data such as season, location, and crop production per square kilometer, area, and crop.
5. **Step 5: XAI Consolidation-** The XAI-CROP framework employs a methodology termed ''Local Interpretable Model-agnostic Explanations'' (LIME) to offer concise explanations for its crop recommendations. LIME is a method for articulating the forecasting of machine learning algorithms by generating local frameworks that approximate the predictions of the original model.
6. **Step 6: Validation-** The XAI-CROP algorithm is validated utilizing a validation dataset to evaluate its performance in terms of forecasting crop yield. The algorithm's accuracy is computed using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared (R2).
7. **Step 7: Model Deployment and Evaluation:** After Deploying the model, businesses should progressively monitor and assess the algorithm's performance in real-world scenarios. Gather feedback from stakeholders and users to pinpoint aspects of improvement and refine the model accordingly.

5. Conclusion

This study explored the application of Artificial Intelligence and Machine Learning techniques in maximizing crop yields in America. This research employed various software tools, comprising Python programming language, Pandas library for data manipulation and analysis, Scikit-learn library for machine learning models and evaluation metrics, and LIME library for explainable AI. The crop yield datasets for the current research were sourced from Kaggle. This dataset provided substantial insights regarding crop cultivation practices within the USA context. This study proposed the "XAI-CROP" algorithm, which is a novel explainable artificial intelligence (XAI) model developed particularly to reinforce the interpretability, transparency and trustworthiness of crop recommendation systems (CRS). From the experimentation XAI-CROP model excelled at forecasting crop yield, as demonstrated by its lowest MSE value, suggesting minimal errors. Besides, its MAE was relatively low, suggesting an average error of less than 1 unit when forecasting crop yield. Furthermore, the R2 value suggested that the algorithm was effective in the data's variability.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Al-Adhaileh, M. H., & Aldhyani, T. H. H. (2022). Artificial intelligence framework for modeling and predicting crop yield to enhance food security in Saudi Arabia. PeerJ. *Computer Science, 8*, e1104. https://doi.org/10.7717/peerj-cs.1104
[2] IRJET Journal. (2020). - MACHINE LEARNING IN CROP DISEASE DETECTION AND YIELD PREDICTION. *Irjet*. https://www.academia.edu/44299667/IRJET_MACHINE_LEARNING_IN_CROP_DISEASE_DETECTION_AND_YIELD_PREDICTION?sm=b
[3] IRJET Journal. (2021). IRJET- The Best Accurate Crop Yield Prediction System Using Machine Learning Algorithms in Deep Learning. *Irjet*. https://www.academia.edu/51049185/IRJET_The_Best_Accurate_Crop_Yield_Prediction_System_Using_Machine_Learning_Algorithms_in_Deep_learning?sm=b
[4] IRJET Journal. (2022). Supervise Machine learning approach for crop yield prediction in the agriculture sector. *Irjet*. https://www.academia.edu/88382662/Supervise_Machine_Learning_Approach_for_Crop_Yield_Prediction_in_Agriculture_Sector?sm=b
[5] Suvarna, U. (2022). Application Of Machine Learning in Modern Agriculture for Crop Yield Prediction and Fertilizer Suggestions Along with The Amount of Usage. *Irjet*. https://www.academia.edu/87116320/Application_Of_Machine_Learning_in_Modern_Agriculture_for_Crop_Yield_Prediction_and_Fertilizer_Suggestions_Along_with_The_Amount_of_Usage?sm=b
[6] Palanivel, K., & Surianarayanan, C. (2019). AN} {APPROACH} {FOR} {PREDICTION} {OF} {CROP} {YIELD} {USING} {MACHINE} {LEARNING} {AND} {BIG} {DATA} {TECHNIQUES. *Iaeme*. https://www.academia.edu/40187021/AN_APPROACH_FOR_PREDICTION_OF_CROP_YIELD_USING_MACHINE_LEARNING_AND_BIG_DATA_TECHNIQUES?sm=b

[7] ProAIrokibul. (n.d.). *Crop-Yield-Project/Model/crop_yield.ipynb at main · proAIrokibul/Crop-Yield-Project*. GitHub. https://github.com/proAIrokibul/Crop-Yield-Project/blob/main/Model/crop_yield.ipynb

[8] Qin, Z., Myers, D. B., Ransom, C. J., Kitchen, N. R., Liang, S. Z., Camberato, J. J., Carter, P. R., Ferguson, R. B., Fernandez, F. G., Franzen, D. W., Laboski, C. a. M., Malone, B. D., Nafziger, E. D., Sawyer, J. E., & Shanahan, J. F. (2018). Application of machine learning methodologies for predicting corn economic optimal nitrogen rate. *Agronomy Journal*, *110*(6), 2596–2607. https://doi.org/10.2134/agronj2018.03.0222

[9] Shams, M. Y., Gamel, S. A., & Talaat, F. M. (2024). Enhancing crop recommendation systems with explainable artificial intelligence: a study on agricultural decision-making. *Neural Computing & Applications*. https://doi.org/10.1007/s00521-023-09391-2