MACHINE LEARNING IN AGRICULTURE

Given Name :Yashvi saxena,Lavanya Gupta
Dept.name :Computer science
chennai,Tamil Nadu

Email address:
Yashvi.saxena2022@vitstudent.ac.in
lavanya.2022@vitstudent.ac.in

Article Information:-

Keywords:

- ***Machine Learning***
- ***Post-harvesting***
- ***Pre-harvesting***
- ***Harvesting***
- ***Precision agriculture***
- ***Decision Tree***
- ***KNN algorithm***
- ***Random forest***
- ***Logistic Regression***
- ***Naïve Bayes***
- ***Python***
- ***Numpy***
- ***Pandas***
- ***Sklearn***

Abstract*—One of the primary needs of humans is food, which can be obtained through farming. Not only does agriculture meet the basic necessities of mankind, but it is also a global source of employment. For emerging nations like India, agriculture is seen as the main driver of employment and the economy. Contributions from agriculture 15.4% of India's GDP. Three main categories can be used to group agricultural activities: pre-harvesting, harvesting, and post-harvesting. The field of machine learning has advanced, contributing to improved agricultural results. The most recent technological advancement that helps farmers reduce farming losses is machine learning, which offers insightful advice and detailed knowledge about crops. In order to address issues in the three domains of pre-harvesting, harvesting, and post-harvesting, this paper provides a thorough overview of the most recent machine learning applications in agriculture. The use of machine learning in agriculture enables higher-quality, more precise and productive farming with fewer human labourers. . The use of machine learning*

INTRODUCTION:

Food is one of the essential needs of humankind, and agriculture is regarded as a key pillar of the global economy.

It is regarded as the primary source of employment in the majority of the nations. Many nations, such as India, continue to practice traditional agriculture. Farmers are hesitant to employ cutting-edge technologies because they lack the necessary skills, the costs are prohibitive, or they are uninformed of the benefits. Irrigation issues, incorrect harvesting practices, wrong use of pesticides, ignorance of soil types, yields, crops, weather, and market trends all contributed to farmer losses or increased expenses. A lack of understanding at every level of agriculture raises the expense of farming and causes new problems to arise. The strain on the agriculture industry grows daily as a result of population growth. Overall, there are a lot of losses in the agricultural processes, from choosing crops to selling finished goods. According to the well-known proverb "Information is Power,"

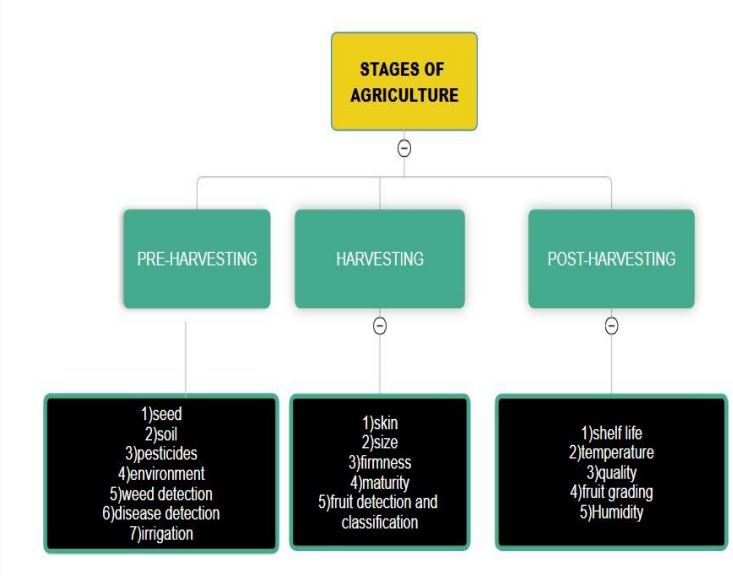
Keeping tabs on data on the market, environment, and crops could assist farmers in making better decisions and resolving agricultural-related issues. Information can be gathered and processed using technologies like blockchain, the Internet of Things, machine learning, deep learning, cloud computing, and edge computing. The use of computer vision, machine learning, and Internet of Things applications can help farmers and related industries become more profitable by increasing productivity, improving quality, and ultimately raising both. The Accurate Education in To increase the overall harvest yield, the agricultural sector is crucial.To increase the overall harvest yield, the agricultural sector is crucial.

Farmers typically follow the processes listed below when carrying out agricultural operations. The steps are:-

Step 1: Crop Selection
Step 2: Preparing the Land
Step 3: Seeding
Step 4: Fertilization and irrigation
Step 5: Crop Upkeep (using insecticides, trimming crops, etc.)
Step 6: Harvesting
Step 7: Post-Harvesting

Here is the flow chart for Agriculture Process:-

Table1:



In this we have explained about:

1.Pre-harvesting

2.Harvesting

3.Post-Harvesting

It will the basic idea for agriculture process

During pre-harvesting tasks farmers focuses on selection of crops, land preparation, seed sowing, irrigation, and crop maintenance which includes use pesticides, pruning etc. In yield estimation the farmers do the activities like yield mapping and counting the number of fruits so that they can predict the production and make the necessary arrangements required at the time of harvesting or post-harvesting.

Pre-Harvesting:

One of the most urgent and essential needs of today is ensuring food security for the rapidly growing world population and, at the same time, ensuring long-term sustainable development in the reduction of food losses. Postharvest losses significantly increase food insecurity, reduce farmer's income and enhance inefficiency in the global food system. The essential elements of postharvest losses challenge include problem of multiple points of intervention, multiple technologies, complex value chain, and poorly developed food systems [1]. In accordance with projections by FAO, food production will need to grow by 70% to feed the world population, which will reach over 10 billion by the year 2050. As efforts are being geared towards increased production, there must be corresponding efforts for an integrated and innovative approach to the global efforts to ensure sustainable food production, consumption and loss reduction. Food waste refers to food appropriate for human consumption being discarded along the food chain due to consumers' behaviour [5, 6]. Damage restricts the use of a product, whereas loss makes its use impossible. These losses occur because harvested agricultural produce consists of living tissues that respire and undergo physiological changes caused by conditions such as high temperature, low atmospheric humidity, physical injury, biotic contamination and enzyme actions. Pre-harvest refers to every activity embarked on by the producer in the production of crops before harvest, and this includes site selection, land preparations, appropriate planting date, optimum seed rate, recommended spacing, appropriate tools and equipment used, proper tillage activities and seedbed preparation, pests, disease and weed management, irrigation, mulching, staking and use of hormones. An adequate supply of potassium nutrition in tomato production enhances titratable acidity and fruit colour quality and reduces the incidence of the yellow shoulder [7, 8], while the inadequate application of potassium in aqua-phonic tomato production results in ripening disorders [9]. An increase in nitrogen supply to tomatoes grown in a controlled environment may reduce fruit quality by decreasing the sugar content of the fruits [10]. A high nitrogen supply of about 250 kg/ha can impair some important quality traits of tomato fruits, such as total soluble solids [11], glucose, fructose, and pH [12]. The quality of tomato fruit is also affected by the amount of boron used. Lower amounts of boron supply reduce fruit firmness [14]. The compositional quality of harvested produce is affected by maturity stage; Howard [15] observed that total vitamin C content of red pepper was about 30% higher compared to green pepper. Tomato fruits harvested green at table ripeness contain less vitamin C than those harvested at the full ripe stage. Tomato fruits at the 'breaker' stage contained only 69% of their vitamin C concentration. Quality refers to the state of excellence of a produce, which may be either good or bad. It refers to a property or group of properties that make a produce acceptable or desired by a consumer. It is subjective and changes according to culture, customs, environment, social status and mindset. These parameters change from one food commodity to the other. Several attributes have been used to describe quality: size, shape, colour, consistency, flavour or organoleptic properties like texture, smell and tastes. Other properties that are used to measure qualities include appearance as presentation, nutritive value, dependability and wholesomeness. Higher quality will translate to higher prices and more consumers' satisfaction.

Food spoilage is also a metabolic process that causes foods to be undesirable or The issue of food losses is of high importance in the efforts to combat hunger, raise income and improve food security in the world today. It is very important to know the pattern and scale of these losses across the world, especially in developing countries and identify their causes and possible solutions.

PRE-HARVESTING FACTORS:

Site selection: The soil properties of a given site for crop production will determine the ultimate compositional and physical quality of the harvested produce. Appropriate site selection, free from heavy metals, toxic materials and adequate fertility level is essential for maximum quality. The soil should be analysed, and the soil condition should be determined before planting.

Planting period: The quality of crops planted during the dry season differ in size, firmness, fibre content and nutritional composition compared to those cultivated in rainy season when there is adequate water availability for chemical processes necessary for plant growth and development [46].

Irrigation: Some crops are not drought resistant hence, yield decreases in terms of size and nutritional quality after short periods of water stress. Proper irrigation planning is crucial for optimum crop development and adequate nutritional composition. Efficient water management scheme is vital to maintaining quality crop and maximum yield [47]. It is observed that deficit irrigation reduced fruit water accumulation and fresh fruit yield but increased fruit total soluble solids in tomatoes [48]. Mitchell et al. observed that deficit irrigation reduced fruit water accumulation and fresh fruit yield but increased fruit total soluble solids in tomato [48]. A higher level of moisture stress affects both yield and quality by decreasing cell enlargement. Crops which have higher moisture content generally have poorer storage characteristics. Some hybrid onions give a high yield of bulbs with low dry matter content and short storage life. Fully matured banana harvested soon after rainfall or irrigation may easily split during handling operations resulting in microorganism infection and rotting. If orange is too turgid at harvest, gland in the skin can be ruptured during harvesting, releasing phenolic compounds and causing oleocellosis or oil spotting (green spot on the yellow/orange coloured citrus fruit after degreening). In green leafy vegetables, too much rain or irrigation can make leaves harder and brittle, making them more susceptible to damage and decay during handling and transportation. Generally, crops with higher moisture or low dry matter content have poorer storage characteristics. Keeping quality of bulb crops like onion and garlic will be poor if irrigation is not stopped before 3 weeks of harvesting [11].

Climatic condition: Many plants are very sensitive to environmental conditions, and thus quality will not be optimised when crop is produced under adverse conditions. Poor weather at harvesting time affects the operations and functionality of harvesting machines or human labour and usually increases the moisture content of the harvested products, consequently resulting in loss of quality and reduced shelf life [57].

Heat management: Physiological and biochemical processes involved in plant growth, yield and maturation is influenced by temperature. Higher temperature during field conditions decreases shelf life and quality of the produce. At high temperatures, plants respire at a faster rate, and stored carbohydrates in harvested produce are depleted rapidly during respiration. High temperature during the fruiting season of tomato leads to quick ripening of fruits. Orange grown in the tropics have higher sugar content and total soluble solids than those grown in the subtropics. Tropically grown oranges tend to be green in colour and peel less easily. This is due to the higher temperature that occurs in the tropics, which results in rapid maturation of fruit which halts the process of the typical temperate orange colour development [58].

Light: Light regulates several physiological processes like chlorophyll synthesis, phototropism, respiration and stomatal opening. The duration, intensity and quality of light affect the quality of fruits and vegetables at harvest. Most of the produce needs high light intensity. Absorption of red light through pigments, phytochrome, is essential for carbohydrate synthesis, which determines the shelf life of the produce. Citrus and mango fruits produced in full sun generally had thinner skin, a lower weight, low juice content and lower acidity but a higher total soluble solid. Citrus

Humidity: High humidity during the growing season results in thin rind and increased size in some horticultural produce, and this produce is more prone to a high incidence of disease during postharvest period. Humid atmosphere may cause the development of fungal and bacterial diseases, which damages produce during storage and transport. Damaged produce removes water very quickly and emits a larger ethylene concentration than healthy ones. Low humidity may cause browning of leaf edge on plants with thin leaves or leaflets. High humidity can maintain the water-borne pollutants in a condition so that they can be more easily absorbed through the cuticles or stomata. Reduced transpiration leads to calcium and other elemental deficiency [59].

Rainfall: Rainfall affects the water supply to the plant and influences the composition of the harvested plant part. This affects its susceptibility to mechanical damage and decay during subsequent harvesting and handling operations. Excess water supply to plants results in the cracking of fruits such as orange, cherries, plums and tomatoes. If root and bulb crops are harvested during heavy rainfall, the storage losses will be higher [60].

Seasons: Seasonal fluctuation and time of the day at harvest will greatly affect the postharvest quality of produce. Synthesis of higher amount of carbohydrates during the day and its utilisation through translocation and respiration at night is responsible for the variation in the longevity of some harvested produce. Roses and tuberose have been found to show longer keeping quality in the winter under ambient conditions than in the summer. Produce harvested early in the morning or in the evening hours exhibits longer postharvest life than produce harvested during hot time of the day. If long-day onion (temperate) is grown during short-day (tropics) conditions, it will result in very poor storage quality [61].

Fertilisers application: Poor fertiliser management will increase physiological disorders due to deficiencies of some minerals or increase of others, leading to toxicity. In both cases, quality will be negatively affected. The use of trace elements or the practice of soilless

tomato production can be made possible during irrigation, where fertilisers are added to the irrigation water in a form of solution and administered. These trace elements are selected depending on the specific postharvest quality traits needed in the fruits. Nutrient balance is crucial for maintaining optimal fruit texture and size: fruits from nitrogen deficient trees are usually smaller with firmer texture, while excess nitrogen leads to rapid loss of firmness and decreased storability. Potassium deficiency also leads to textural changes resulting in small, poorly coloured fruit that may not ripen, leaving fruit hard and inedible. A lack of boron can result in fruits with a mealy texture [62, 63, 64].

Pest and Disease Management: Pathogens and insects have very negative effects on quality of harvested produce. The effect of insect is more pronounced on grains but can also cause a lot of damages in fruits and vegetables. Nematodes cause various injuries to fruits and vegetables and continue the deterioration during storage. Parasites are therefore seen to be important in damage to farm produce as well as food preservation. In the case of insects, produce attacked by them in agriculture may consume over 50% of the harvest. Insects at times lay eggs in the produce, making it almost impossible to eliminate all insects pest in the produce. Parasites like nematodes and amoeba may infect the produce, and the same is true when produce comes in contact with water; this is very common in Africa. Rodents contaminate food with their urine and droppings. They also produce large litter, for example, two rats can give up to 30,000 litres per year. Through their contamination, they spread diseases like salmonellosis, plague and typhoid fever.

SEEDS	Features	Temperature conditions	Soil	Methods Used/Accuracy
Pigeons Bean	Pigeon beans, also known as Cajanus cajan or pigeon peas, are a versatile legume crop cultivated in tropical and subtropical regions worldwide. Known for their drought resistance and nitrogen-fixing abilities,	Pigeon beans thrive in warm temperatures, typically requiring temperatures between 25°C to 35°C (77°F to 95°F) for optimal growth. They are sensitive to frost and cold temperatures, which can inhibit germination and growth.	thrive in well-drained, slightly acidic to neutral soils with a sandy loam to sandy clay loam texture.	Decision Tree: Accuracy:90% Naïve Bayes Accuracy:90% Logistic Regression Accuracy:95.4% Random Forest Accuracy: 99.3%
Maize	Maize, also known as corn, is a versatile crop with widespread cultivation across the globe. It serves as a vital source of food, feed, and industrial products	Maize is a warm-season crop that thrives in temperatures between 20°C to 30°C (68°F to 86°F) during the growing season. It requires ample sunlight and is sensitive to frost, with optimal growth occurring in areas where the average(40%)	generally prefers well-drained, fertile soils with a slightly acidic to neutrals	Decision Tree: Accuracy:90% Naïve Bayes Accuracy:90% Logistic Regression Accuracy:95.4% Random Forest Accuracy: 99.3%
Chick peas	Chickpea, also known as garbanzo bean, is a versatile legume prized for its nutty flavor and high nutritional valu	Chickpeas thrive in warm climates and require temperatures between 70°F to 80°F (21°C to 27°C) during the growing season	generally prefers well-drained, fertile soils with a slightly acidic to neutral pH.	Decision Tree: Accuracy:90% Naïve Bayes Accuracy:90%

		for optimal growth and development		Logistic Regression Accuracy:95.4% Random Forest Accuracy: 99.3%
Kidney Beans	Kidney beans, scientifically known as Phaseolus vulgaris, are a nutritious legume widely consumed worldwide. These beans are renowned for their high protein and fiber content, making them a valuable addition to vegetarian and vegan diets	The optimal temperature for storing kidney beans is between 50°F and 70°F (10°C to 21°C). At this temperature range, kidney beans can be kept in a cool, dry place, such as a pantry or cupboard, away from moisture and direct sunlight	thrive in well-drained soils with a slightly acidic to neutral pH level, ideally ranging from 6.0 to 7.	Decision Tree: Accuracy:90% Naïve Bayes Accuracy:90% Logistic Regression Accuracy:95.4% Random Forest Accuracy: 99.3%
Rice	Rice stands as a vital crop globally, serving as a staple food for millions and a cornerstone of cultural heritage. Its adaptability to various growing conditions, coupled with its high caloric and nutritional value, underscores its significance.	The optimal temperature for rice cultivation typically ranges between 20°C to 35°C during the growing season. Cooler temperatures can slow growth and development, while extreme heat can stress the plants, affecting yields.	Rice cultivation benefits from slightly acidic to neutral pH levels (around 6.0 to 7.0) and moderately fertile soils, ideally sandy loam to clay loam in texture	Decision Tree: Accuracy:90% Naïve Bayes Accuracy:90% Logistic Regression Accuracy:95.4% Random Forest Accuracy: 99.3%
Moth Beans	Moth beans, also known as matki or Turkish gram,	Moth beans thrive in warm climates and are	thrive in well-drained, fertile soils with a	Decision Tree: Accuracy:90%

	are a versatile legume prized for their nutritional value and adaptability. These small, oval-shaped beans are rich in protein, dietary fiber, vitamins, and minerals,	typically cultivated in regions with temperatures ranging from 25°C to 35°C (77°F to 95°F). They require a frost-free growing season and are sensitive to cold temperatures	slightly acidic to neutral pH ranging from 6.0 to 7.0.	Naïve Bayes Accuracy:90% Logistic Regression Accuracy:95.4% Random Forest Accuracy: 99.3%
Mung Beans	Mung beans, also known as green gram or moong dal, are a nutritious legume widely cultivated in Asia and other parts of the world. These small, green beans are a rich source of plant-based protein, fiber, vitamins, and minerals	The optimal temperature requirements for mung beans typically fall within the range of 25°C to 35°C (77°F to 95°F) during the growing season. Mung beans thrive in warm climates and are sensitive to frost and cold temperatures.	thrive in well-drained sandy loam soils with a slightly acidic to neutral pH range of 6.0 to 7.5.	Decision Tree: Accuracy:90% Naïve Bayes Accuracy:90% Logistic Regression Accuracy:95.4% Random Forest Accuracy: 99.3%
Black Beans	The optimal temperature requirements for mung beans typically fall within the range of 25°C to 35°C (77°F to 95°F) during the growing season. Mung beans thrive in warm climates and are sensitive to frost and cold temperatures.	The optimum temperature for growing black seeds (Nigella sativa) ranges between 18°C to 25°C (64°F to 77°F).	thrive in well-drained, fertile soils with a slightly acidic to neutral pH	Decision Tree: Accuracy:90% Naïve Bayes Accuracy:90% Logistic Regression Accuracy:95.4% Random Forest Accuracy: 99.3%
Lentils	The optimum temperature for	The ideal temperature for	thrive in well-drained, sandy	Decision Tree:

	growing black seeds (Nigella sativa) ranges between 18°C to 25°C (64°F to 77°F).	lentil cultivation typically ranges between 18°C to 24°C (64°F to 75°F)	loam to clay loam soils with a slightly acidic to neutral pH range of 6.0 to 7.5	Accuracy:90% Naïve Bayes Accuracy:90% Logistic Regression Accuracy:95.4% Random Forest Accuracy: 99.3%
Banana	Banana, a widely cultivated fruit, is renowned for its rich nutritional profile and versatility. Packed with essential vitamins, minerals, and dietary fiber,	Bananas thrive in tropical and subtropical climates, with temperatures ideally ranging between 26°C to 30°C (78°F to 86°F) during the day and not dropping below 15°C (59°F) at night	thrive in rich, well-drained soils with a pH ranging from 5.5 to 7.0, ideally in tropical or subtropical climates	Decision Tree: Accuracy:90% Naïve Bayes Accuracy:90% Logistic Regression Accuracy:95.4% Random Forest Accuracy: 99.3%
Mango	Mango, often hailed as the "king of fruits," is prized for its deliciously sweet and tropical flavor. Cultivated in many regions with warm climates,	Mangoes thrive in warm tropical and subtropical climates, generally requiring temperatures between 25°C to 35°C (77°F to 95°F) for optimal growth and fruit productio	thrive in warm, tropical climates with well-drained soil and ample sunlight	Decision Tree: Accuracy:90% Naïve Bayes Accuracy:90% Logistic Regression Accuracy:95.4% Random Forest Accuracy: 99.3%

Source: Online Datasets

Decision Trees:

The code we have provided is a comprehensive example of using various machine learning algorithms, including Decision Trees, to predict the type of crop that can be grown based on various agricultural conditions. Let's break down the key components and explain how the Decision Tree algorithm is implemented and used in this code.

Importing Libraries:

The code starts by importing necessary libraries for data manipulation (numpy, pandas), data visualization (matplotlib, seaborn), and machine learning (sklearn). It also imports warnings to suppress warnings that might clutter the output.

Data Preparation:

The dataset is loaded from a CSV file located at PATH. The dataset is then explored to understand its structure, missing values, and the distribution of different crops based on various conditions like temperature, humidity, and rainfall. This exploratory analysis is crucial for preprocessing and feature engineering.

Data Visualization:

The code uses matplotlib and seaborn to visualize the distribution of various agricultural conditions (Nitrogen, Phosphorous, Potassium, Temperature, Humidity, PH, Rainfall) across different crops. This visualization helps in understanding the relationship between these conditions and the type of crop that can be grown.

Feature Selection and Preprocessing:

The code selects relevant features (Nitrogen, Phosphorous, Potassium, Temperature, Humidity, PH, Rainfall) and the target variable (label) for the machine learning models. It also splits the dataset into training and testing sets using train_test_split from sklearn.model_selection.

Decision Tree Algorithm:

The Decision Tree algorithm is implemented using DecisionTreeClassifier from sklearn.tree. The classifier is initialized with the following parameters:

- criterion='entropy': This specifies that the information gain (entropy) should be used to measure the quality of a split.

- max_depth=5: This limits the depth of the tree to prevent overfitting.

- random_state=2: This ensures that the splits you generate are reproducible.

The model is then trained using the fit method with the training data (X_train and y_train). After training, the model is used to predict the type of crop for the test data (X_test), and the accuracy of the predictions is calculated using accuracy_score from sklearn.metrics.

Cross-Validation:

The code also performs cross-validation using cross_val_score from sklearn.model_selection to assess the performance of the Decision Tree model. Cross-validation helps in understanding how well the model generalizes to unseen data.

Accuracy Comparison:

The code compares the accuracy of the Decision Tree model with other models (Naive Bayes, Logistic Regression, Random Forest) using a bar plot. This visualization helps in understanding the performance of different models.

Making Predictions:

Finally, the code demonstrates how to use the trained Random Forest model to make predictions on new data. It creates a new dataset with specific values for the features and uses the predict method to predict the type of crop that can be grown under those conditions.

Conclusion:

This code provides a comprehensive example of using the Decision Tree algorithm for a classification problem. It includes data preprocessing, feature selection, model training, evaluation, and prediction. The Decision Tree algorithm is a powerful tool for classification tasks, especially when the decision boundaries are not linear.

Naïve Bayes:

The code we have provided is a comprehensive example of using various machine learning models, including Naive Bayes, to predict the type of crop that can be grown based on various agricultural conditions. The Naive Bayes algorithm is one of the models used in this code. Let's break down the Naive Bayes part of the code and explain how it works within the context of this agricultural prediction task.

Naive Bayes Algorithm Overview:

Naive Bayes is a classification algorithm based on Bayes' Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. This assumption allows the algorithm to calculate the probability of a class given a set of features in a very efficient manner.

Naive Bayes in the Code:

In the provided code, the Naive Bayes algorithm is implemented using the GaussianNB class from the sklearn.naive_bayes module. This class implements the Gaussian Naive Bayes algorithm, which is suitable for classification with continuous features.

Implementation Steps:

1. Data Preparation: The code first prepares the dataset by selecting relevant features and splitting the data into training and testing sets. This is done using the train_test_split function from sklearn.model_selection.

2. Model Training: The Naive Bayes model is trained on the training data using the fit method. The model learns the distribution of the features for each class in the dataset.

3. Prediction: The trained model is then used to predict the class labels for the test data using the predict method.

4. Evaluation: The accuracy of the model is evaluated by comparing the predicted labels with the actual labels of the test data. This is done using the accuracy_score function from sklearn.metrics.

5. Cross-Validation: The code also performs cross-validation to assess the model's performance more robustly. This is done using the cross_val_score function from sklearn.model_selection.

Code Snippet for Naive Bayes:

```
python
from sklearn.naive_bayes import GaussianNB
# Initialize the Naive Bayes classifier
Naive_Bayes = GaussianNB()
# Train the model
Naive_Bayes.fit(X_train, y_train)
# Make predictions on the test data
predicted = Naive_Bayes.predict(X_test)
# Calculate the accuracy of the model
x = metrics.accuracy_score(y_test, predicted)
acc.append(x)
model.append('Naive Bayes')
# Print the accuracy
print('Naive Bayes accuracy is', x * 100)
# Perform cross-validation
score = cross_val_score(Naive_Bayes, features, target, cv=5)
```

Conclusion

The Naive Bayes algorithm is a simple yet effective method for classification tasks, especially when the features are continuous. In the context of predicting the type of crop that can be grown based on various agricultural conditions, Naive Bayes can provide a good baseline model. The code demonstrates how to implement, train, and evaluate a Naive Bayes classifier using the sklearn library in Python.

Logistic Regression:

The code we have provided is a comprehensive example of using various machine learning models, including Logistic Regression, to predict the type of crop that can be grown based on various agricultural conditions. The dataset used in this example contains information about different crops and the conditions required for their growth, such as temperature, humidity, rainfall, and soil nutrient levels.i

Logistic Regression in the Code:

Logistic Regression is a statistical model used for binary classification problems. It estimates the probability that a given input point belongs to a certain class. In this code, Logistic Regression is used to predict the type of crop that can be grown based on the provided features.

Key Steps in Implementing Logistic Regression:

1. Data Preparation: The dataset is loaded using pandas, and various exploratory data analysis steps are performed, such as checking for missing values, understanding the distribution of data, and visualizing the data.

2. Feature Selection: The features used for prediction are selected. In this case, the features include nitrogen (N), phosphorous (P), potassium (K), temperature, humidity, pH, and rainfall.

3. Splitting the Dataset: The dataset is split into training and testing sets using train_test_split from sklearn.model_selection. This is a crucial step as it allows us to evaluate the model's performance on unseen data.

4. Model Training: The Logistic Regression model is trained using the training data. The model learns the relationship between the features and the target variable (the type of crop).

5. Prediction: The trained model is used to predict the type of crop for the test data.

6. Evaluation: The accuracy of the model is evaluated using the test data. The accuracy_score from sklearn.metrics is used to calculate the accuracy. Additionally, a classification report is generated using classification_report from sklearn.metrics, which provides a detailed report on the performance of the model.

7. *Cross-Validation*: Cross-validation is performed to assess the model's performance across different subsets of the data. This helps in understanding the model's robustness.

8. Comparison of Models: The accuracy of the Logistic Regression model is compared with other models (Decision Tree, Naive Bayes, and Random Forest) to understand which model performs best for this dataset.

Code Snippet for Logistic Regression:

```
python
from sklearn.linear_model import LogisticRegression

# Initialize Logistic Regression model
LogReg = LogisticRegression()

# Train the model
LogReg.fit(X_train, y_train)

# Predict the target for test dataset
predicted = LogReg.predict(X_test)

# Calculate accuracy
x = metrics.accuracy_score(y_test, predicted)
acc.append(x)
model.append('Logistic Regression')

# Print accuracy
print("Logistic Regression Accuracy is", x * 100)

# Generate classification report
print(classification_report(y_test, predicted))

# Perform cross-validation
score = cross_val_score(LogReg, features, target, cv=5)
score
```

This code snippet demonstrates how to implement Logistic Regression for the given dataset, train the model, make predictions, and evaluate the model's performance.

Random Forest:

The code we have provided is a comprehensive example of using various machine learning algorithms, including Random Forest, for a classification task. It involves data preprocessing, exploratory data analysis, model training, and evaluation. Let's break down the key components and explain how the Random Forest algorithm is implemented and used in this context.

Data Preprocessing and Exploratory Data Analysis:

1. Importing Libraries: The code starts by importing necessary libraries for data manipulation, visualization, and machine learning tasks.

2. Reading the Dataset: It reads a CSV file containing agricultural data, which includes various features like nitrogen, phosphorous, potassium levels, temperature, humidity, pH, and rainfall, along with the crop label.

3. Data Exploration: The code explores the dataset by printing its shape, head, tail, columns, checking for null values, and understanding the distribution of different crops based on various conditions (temperature, humidity, rainfall).

4. Data Visualization: It uses seaborn and matplotlib to visualize the distribution of agricultural conditions and the impact of different conditions on crops.

Model Training and Evaluation:

1. Splitting the Dataset: The dataset is split into training and testing sets using train_test_split from sklearn.model_selection.

2. Model Training: The code trains several models, including Decision Tree, Naive Bayes, Logistic Regression, and Random Forest, on the training data.

3. Model Evaluation: Each model's performance is evaluated using accuracy score and classification report from sklearn.metrics. Cross-validation is also performed to get a more robust estimate of the model's performance.

Random Forest Algorithm:

- Random Forest Classifier: The Random Forest algorithm is implemented using RandomForestClassifier from sklearn.ensemble. It is initialized with n_estimators=29 (the number of trees in the forest), criterion='entropy' (the function to measure the quality of a split), and random_state=0 (for reproducibility).

- Training: The model is trained on the training data (X_train and y_train) using the fit method.

- Prediction: The trained model is used to predict the labels of the test data (X_test) using the predict method.

- Evaluation: The accuracy of the Random Forest model is calculated by comparing the predicted labels with the true labels (y_test) using the accuracy_score method from sklearn.metrics.

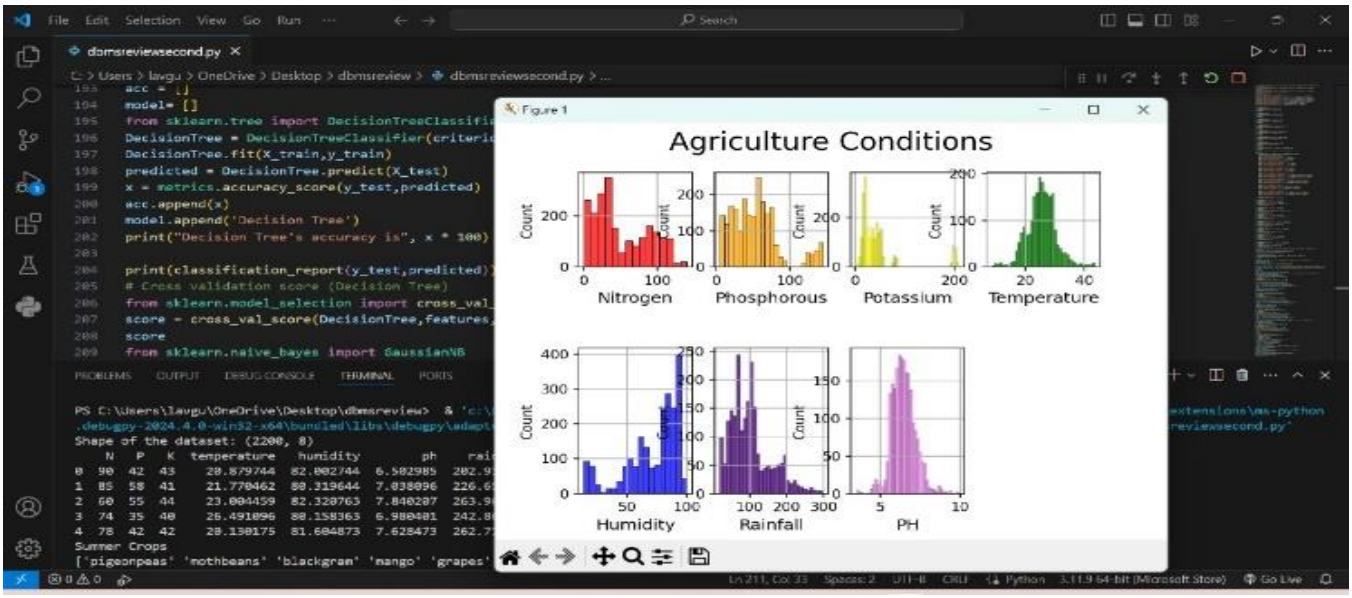
- Cross-Validation: Cross-validation is performed to assess the model's performance across different subsets of the data.

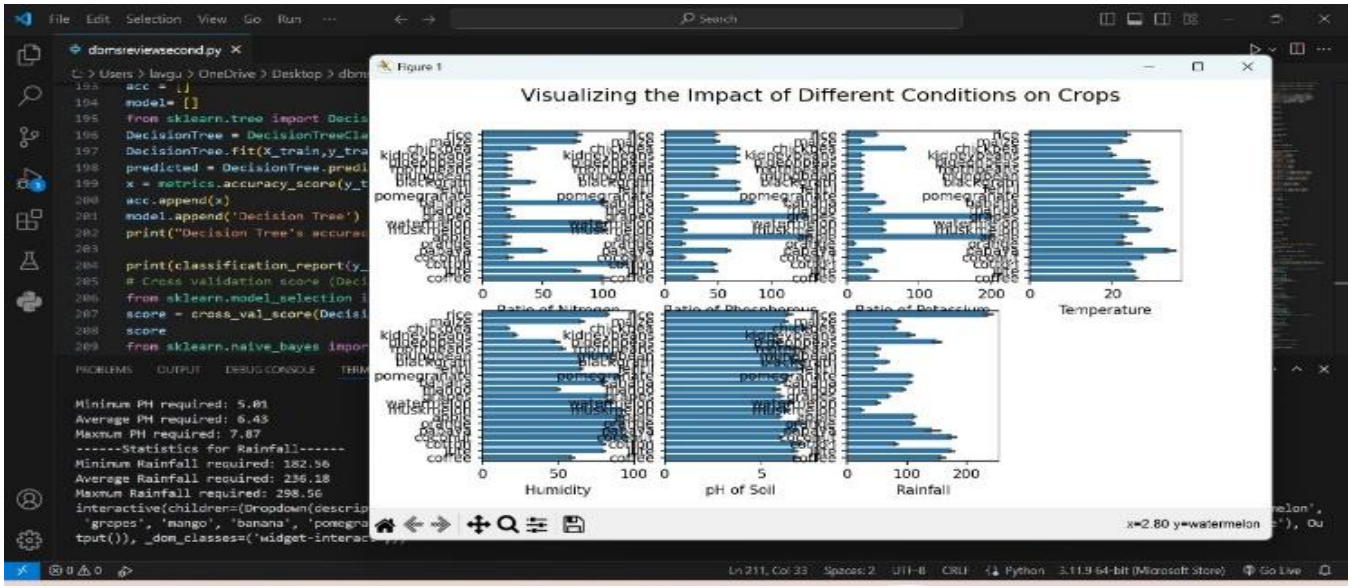
Making Predictions:

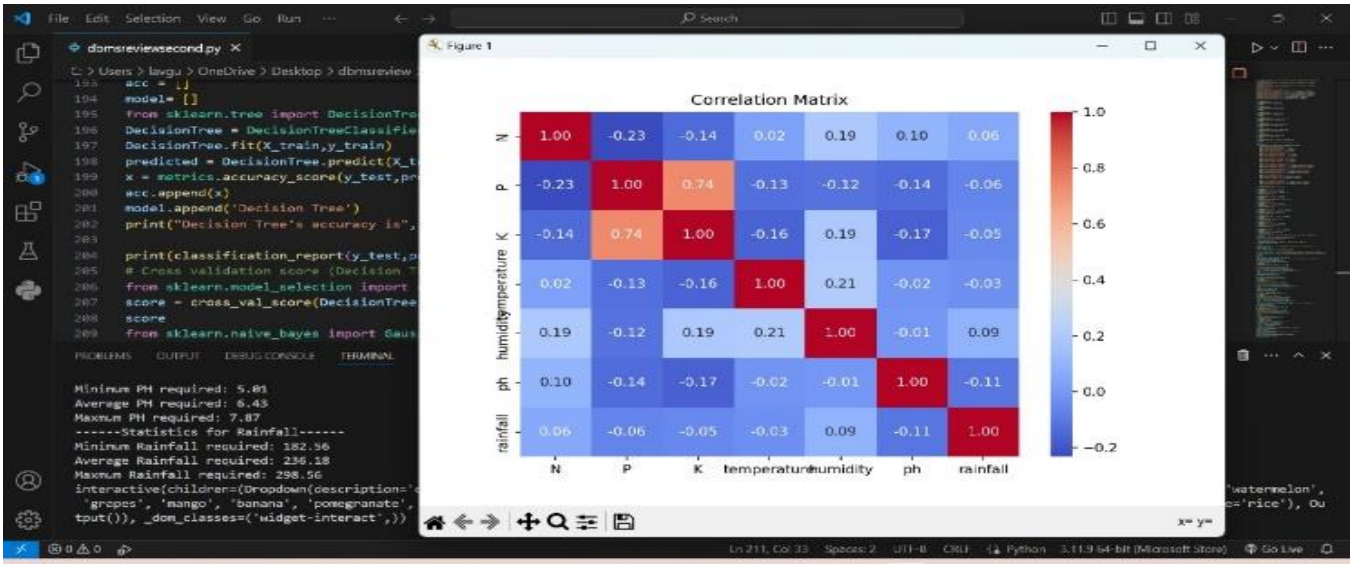
The code demonstrates how to use the trained Random Forest model to make predictions on new data. It creates three new data points and uses the predict method to classify them.

Conclusion:

This code provides a comprehensive example of how to use the Random Forest algorithm for a classification task in a real-world scenario. It includes data preprocessing, exploratory data analysis, model training, evaluation, and making predictions. The Random Forest algorithm is particularly useful for handling large datasets with many features, as it can capture complex patterns and interactions between features.


Agriculture Conditions
Nitrogen
Phosphorous
Potassium
Temperature
Humidity
Rainfall
PH
Count


Visualizing the Impact of Different Conditions on Crops
Temperature
Humidity
pH of Soil
Rainfall


Correlation Matrix
N
P
K
temperature
humidity
ph
rainfall