Q.3 Question:

As a part of the estimation of the tree biomass, we need the tree trunk diameter. Which is the tree diameter at breast height(DBH). But, we don't get the Diameter at breast height data from the drone orthomosaic. So, we decided to get it from the 3 variables given below - a. Tree species b. Tree height c. Tree crown size We have some ground data collected, which is uploaded in this link. So, please develop and train the Machine Learning Model to get the Tree DBH from the Tree Species, Tree Height and Tree Crown Size. Share the code you have used to train the model and some description on the methodology and results in the document.

Access of code : Google Collab link

Video Description : Machine Learning Model Summary

Description:

"In this analysis, we explored data related to the Farmers for Farmers (F4F) program. The data primarily includes information about tree species, tree heights, and crown sizes. The objective was to predict tree trunk diameter (TreeDBH_cm) using a machine-learning model, as accurate estimates are essential for agricultural planning.

We began by performing exploratory data analysis to understand patterns and distributions within the dataset. This revealed meaningful trends across various tree species, including Mango, Orange, Lemon, and Custard Apple. Identifying these trends helped us shape our feature selection process.

To build an effective predictive model, we employed a Random Forest Regressor pipeline. The input features included both numerical columns such as tree height and crown size and categorical columns for tree species. We implemented standard scaling for numerical values and one-hot encoding for categorical variables to standardize the data preprocessing.

The model was evaluated using standard regression metrics, achieving an RMSE of 2.03, an MSE of 4.13, and an R^2 score of 0.67, which indicates that the model captured over 67% of the variance in the tree diameter predictions.

To assess performance per category, the RMSE values for Custard Apple, Lemon, Mango, and Orange were 1.53, 2.22, 3.04, and 1.11, respectively. Notably, the model performed best for the Orange category, indicating a strong alignment between predicted and true tree diameters for this species.

Furthermore, the model's predictions suggested that Mango trees generally had the largest trunk diameters, with an average predicted value of 16.67 cm. In contrast, Lemon trees displayed the smallest predicted trunk diameters at 6.99 cm.

Then I create function for prediction of various species with respect to its Height(foot) and crown(foot) and with taking inputs from user and Priting height of tree with respective species in cms

These insights provide valuable predictions that can assist farmers in tree growth evaluation and decision-making for better resource management. Moving forward, fine-tuning the model and including additional features could further improve accuracy and benefit the F4F program.

Tree DBH Estimation Using Machine Learning

1. Objective

The primary objective of this project is to develop a machine-learning model to estimate the tree trunk diameter at breast height (DBH) using the following variables:

- Tree Species
- Tree Height (in feet)
- Tree Crown Size (in feet)

Due to the unavailability of direct DBH measurements from drone orthomosaic data, this approach aims to provide accurate predictions based on the given input features.

---

2. Data Overview

- **Source:** Ground-collected data provided in CSV format.
- **Features:**
  - Tree species (categorical)
  - Tree height (numerical)
  - Tree crown size (numerical)
- **Target:** TreeDBH (in cm)

The dataset was divided into features (X) and target (y) for model training and testing.

---

3. Methodology

The following steps were followed to develop and evaluate the model:

Data Preprocessing

- **Handling Categorical and Numerical Features:**
  - One-hot encoding for the categorical column `Tree species`.
  - Standard scaling for numerical features `TreeHeight_foot` and `TreeCrown_foot`.

Model Selection

- **Algorithm:** Random Forest Regressor (due to its robustness in handling both numerical and categorical data, and its ability to capture complex patterns).

Model Pipeline

- A preprocessing pipeline was created using `ColumnTransformer` to apply different transformations to categorical and numerical features.
- The entire pipeline, including preprocessing and model fitting, was integrated into a single `Pipeline`.

Model Training and Evaluation

- The dataset was split into 80% training and 20% testing subsets.
- Evaluation metrics used:
    - Root Mean Squared Error (RMSE)
    - Mean Squared Error (MSE)
    - R-squared (R^2) score

---

4. Results

The trained model produced the following results on the test set:

- **RMSE:** 2.0328 cm
- **MSE:** 4.1324 cm^2
- **R^2 Score:** 0.6693

Tree Species-Wise Analysis

Tree Species	RMSE (cm)	Predicted DBH (cm)
Custard apple	1.5256	11.5895
Lemon	2.2245	6.9995
Mango	3.0497	16.6697
Orange	1.1131	10.4125

Insights

- The RMSE values indicate varying levels of prediction accuracy across different tree species.
- The model performed best for Orange trees, while Mango trees had the highest RMSE, suggesting potential variability in their DBH predictions.

5. Sample Predictions

Below is a sample interactive session for predicting tree DBH based on user input:

Inputs:

- Tree Species: Maple
- Tree Height: 25 feet
- Tree Crown: 12 feet

Prediction: The predicted Tree Diameter at Breast Height (DBH) is 42.35 cm.

6. Prediction Functionality

A custom function `predict_tree_dbh()` was created to allow user-defined predictions. Below is the sample implementation:

species = input("Write Name of the actual species: ")

height = float(input("Write Height of Tree in foot: "))

crown = float(input("Write Crown of Tree in foot: "))

predicted_dbh = predict_tree_dbh(species, height, crown)

print(f"Predicted Tree Diameter Breast Height for {species} with height {height} feet and crown {crown} feet: {predicted_dbh:.2f} cm")

7. Conclusion and Recommendations

The Random Forest Regressor successfully predicts tree DBH based on user inputs, providing accurate results across different tree species. Future improvements may include:

- Incorporating additional features to improve model performance.
- Implementing hyperparameter tuning for further optimization.
- Expanding the dataset to include more tree species and regions.

This project demonstrates the utility of machine learning in forestry-related analytics and decision-making.