Yaman Salman 11926532
Mahmoud Suwwan 12011679
Ameed Malhis 12011326

Olive Tree Peacock Disease Detection, Analysis, and Prediction, using Machine Learning.

Summary:

This project aims to study real world and local data to develop a machine learning model application. The application is developed to have a few predictions and analysis. Some of these functionalities is to predict the susceptibility of olive trees to Peacock disease. Utilizing a dataset comprising over 44,000 entries from various regions, including Nablus, Tulakrem, Qalqilya, and more the study uses historical data obtained from Prof. Mazen Salman's research at Al-Khadourie University. The model analyzes key dates, times, and tree-related features to determine the likelihood of disease incidence, offering a valuable tool for early detection and preventative measures in olive cultivation. It also offers a tool to help predict the effectiveness and time duration needed for treatment.

These three functionalities were implemented using three different machine learning models in python. Many libraries such as pandas, numpy and sklearn were used too. The data that was used was split into two sets. A training set, and a validation set. Finally, the models were trained, their results were compared using performance measures.

Introduction:

The biggest and greatest motivation behind this project was the chance of getting a hold onto real world data. This data was not just any data, but data about Olive Trees. This made our team think of the importance of olive trees and what they stand for in regards toPalestine. This was too good of a chance to miss out on. The data was obtained from Prof. Mazen Salman’s Research from Al-Khadourie University. The vast amount of data needed a lot of cleaning and refining. Yet the thought of having so much data available to use made us excited.

This project gave us a chance to understand what the disease is and how it works with olive trees. It also gave us the chance to study a real life problem and its solution. Important results were derived from the project too. Finally, merging our culture with machine learning to solve problems is a very rewarding opportunity.

Data Set:

Source: The dataset was acquired from the research of Prof. Mazen Salman, Al-Khadourie University.

Features:

Trip Num: Identifier for the data collection trip.
Date: Date of observation.
City: Location of the olive tree.
Species: Species of the olive tree.
Tree Num: Tree identifier.
Leaf Type: Type of leaf on the olive tree.
Leaf Num: Number of leaves observed.
OLS Before: Olive Leaf Spot (OLS) measurement before treatment.
OLS After: OLS measurement after treatment.
Non Visible Lesions: Count of lesions not visible.
Over 44,000 Lines of Data entered.

Methodology:

The methodology involves training various machine learning models, including Linear Regression, Random Forest, and Gradient Boosting Regressors, to predict the treatment effectiveness and susceptibility to Peacock disease. The models are evaluated based on Mean Squared Error (MSE) and R2 Score to determine the best-performing model. Further, the approach includes splitting the data into training and validation sets to ensure the model's generalizability and accuracy.

The clean-up of the data was the most time consuming due to its huge nature. The models were all trained on both the training set and then the validation set. After training the models, the models were compared and analysed. Furtheron, more results for the user are printed to be read and analysed too.

Results:

1. In the first method called: Analyze Treatment Effectiveness Over Time
    a. Model Performance:
        - Model: Linear Regression, MSE: 3303.642804641096, R2: 0.10166868736388268
        - Model: Random Forest Regressor, MSE: 1168.7723088616706, R2: 0.6821857493438929
        - Model: Gradient Boosting Regressor, MSE: 1467.2410713123586, R2: 0.6010256933061964
        - With The Random Forest Regressor yielding the best results of MSE: 1168.7723088616706 and R2: 0.6821857493438929
    b. Prediction & Analysis Output:
    Average Reduction in Olive Leaf Spot (OLS): 0.4872916666666667%
    Percentage of Trees Showing Improvement: 18.516666666666666%

2. In the second method called: Predict Most Susceptible Months to OLS
    a. Model Perfomance:
        - Model: Linear Regression, MSE: 4.361460590231976, R2: 0.05739030319250393
        - Model: Random Forest Regressor, MSE: 3.8569220228798136, R2: 0.16643243166309862
        - Model: Gradient Boosting Regressor, MSE: 3.6916869816868356, R2: 0.20214343921632816
        - With Best Model by MSE: Gradient Boosting Regressor (MSE: 3.6916869816868356)
        Best Model by R2: Gradient Boosting Regressor (R2: 0.20214343921632816)
    b. Predictions & Analysis:
        Months Most Susceptible to Olive Leaf Spot (Ranked by Percentage):
        1. May: Average Predicted OLS Susceptibility = 100.00%
        2. December: Average Predicted OLS Susceptibility = 44.19%
        3. March: Average Predicted OLS Susceptibility = 8.12%
        4. January: Average Predicted OLS Susceptibility = 3.18%
        5. February: Average Predicted OLS Susceptibility = 1.02%
        6. June: Average Predicted OLS Susceptibility = -1.52%
        7. April: Average Predicted OLS Susceptibility = -2.90%
        8. July: Average Predicted OLS Susceptibility = -4.02%
        9. October: Average Predicted OLS Susceptibility = -5.97%
        10. August: Average Predicted OLS Susceptibility = -6.00%
        11. November: Average Predicted OLS Susceptibility = -7.06%
        12. September: Average Predicted OLS Susceptibility = -7.10%

3. In the third method called: Estimate Time Duration For Effective Treatment
    a. Model Performance:
        - Model: Linear Regression, MSE: 0.8662006268894454, R2: -0.007161740935504701
        - Model: Random Forest Regressor, MSE: 1.0378985397737397, R2: -0.2068009047589332
        - Model: Gradient Boosting Regressor, MSE: 0.9813392621398591, R2: -0.14103745601570794
    b. Analysis & Predictions:

Average Treatment Duration for Each Month:
Month: January, Average Duration: 0 days, 3 hours, 42 minutes
Month: February, Average Duration: 0 days, 0 hours, 38 minutes
Month: March, Average Duration: 0 days, 1 hours, 2 minutes
Month: April, Average Duration: 0 days, 0 hours, 51 minutes
Month: May, Average Duration: 0 days, 1 hours, 6 minutes
Month: June, Average Duration: 0 days, 4 hours, 24 minutes
Month: July, Average Duration: 0 days, 0 hours, 54 minutes
Month: August, Average Duration: 0 days, 0 hours, 33 minutes
Month: October, Average Duration: 0 days, 2 hours, 11 minutes
Month: December, Average Duration: 0 days, 1 hours, 40 minutes