Working with the Copilot for Machine Learning

Duration: 45 minutes

Working with Copilot for machine learning involves leveraging GitHub Copilot, an AI-powered code completion tool developed by GitHub in collaboration with OpenAI. Here's a quick summary of the key steps and considerations:

Installation: Ensure you have GitHub Copilot installed as an extension in your integrated development environment (IDE), such as Visual Studio Code.

GitHub Integration: Link your IDE to your GitHub account to enable seamless integration. This allows Copilot to access your code repositories and provide context-aware suggestions.

Machine Learning Frameworks: Copilot supports various Machine Learning libraries and frameworks like TensorFlow, PyTorch, scikit-learn, and more. It can assist with code generation for tasks like data preprocessing, model building, and evaluation.

In this exercise, you will be cloning the Git repository with the required dataset into your environment. Here, you will be working with Copilot for Machine Learning, which involves leveraging GitHub Copilot.

**Disclaimer**: GitHub Copilot will automatically suggest an entire function body or code in grayed text. Examples of what you'll most likely see in this exercise, but the exact suggestion may vary.

**Note**: Before proceeding with the exercise, make sure you have installed Python and pip packages.

Task 1: Prerequisites and Injecting the Required Dataset into Your Environment

1. Open VS Code Terminal by clicking on **Ellipsis (...)** **(1)**, selecting **Terminal** **(2)**, and clicking on **New Terminal** **(3)**.

A screenshot of a computer

Description automatically generated

1. Clone the below git repository in your environment.
2. git clone https://github.com/CloudLabsAI-Azure/ml-copilot-workshop.git

git clone https://github.com/CloudLabsAI-Azure/ml-copilot-workshop.git

1. Change the directory in the terminal by running the below command:
2. cd ml-copilot-workshop

cd ml-copilot-workshop

1. To install all the required Python dependencies in your environment before working with Copilot run the below command in your terminal:
2. pip install -r requirements.txt

pip install -r requirements.txt

Task 2: Auto\_Completion of Code with Different Experiments

1. From the VS Code explorer window, right-click on the folder named **ml-copilot-workshop** **(1)** and click on **New File**. Name the file Experiments.ipynb **(2)** and verify that your new file looks as shown below:

A screenshot of a computer

Description automatically generated

1. Type the below comments to import all the libraries where the Copilot automatically prompts all the libraries; press "tab" and then press enter to get output.
2. # Import libraries with respect to loading data and creating a random forest model

# Import libraries with respect to loading data and creating a random forest model

A screenshot of a computer

Description automatically generated

1. Accept all the suggestions for importing libraries as shown in the below screenshot **(1)** and click the **Run** **(2)** button to execute the cell. Click on +Code **(3)** to add the new cell.

A screenshot of a computer program

Description automatically generated

**Note**: While running the cell, you may need to install the required packages and select the kernel.

1. Type the below comments to load the data using the Copilot prompt. Press "enter" to get into the next line and review the suggestion, and press "tab" to accept the suggestion.
2. # Load the data from a csv file, and the name of the file is diabetes.csv

# Load the data from a csv file, and the name of the file is diabetes.csv

A screenshot of a computer

Description automatically generated

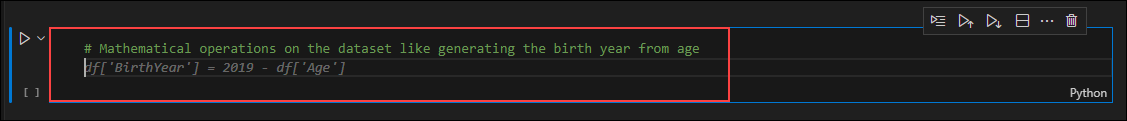
1. Now click on **+Code** and move on to the next task.

Task 3: Mathematical and Machine Learning with Different Examples

Task 3.1: Mathematical Operations

1. Type the below comments to performe the first mathematical experiment would be to generate the birth year from the age column present in the dataset, press "tab", and then press "enter".
2. # Mathematical operations on the dataset, like generating the birth year from age

# Mathematical operations on the dataset, like generating the birth year from age



1. Type the below comment, press "tab", and then press "enter" to get the output similar to the below image.
2. # Show the new column

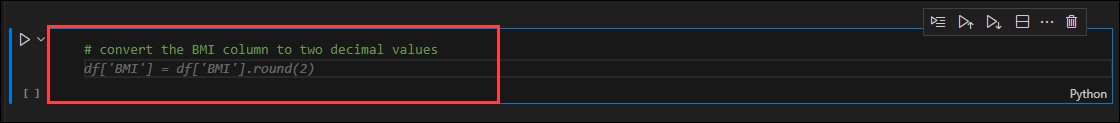
# Show the new column

A screenshot of a computer

Description automatically generated

1. Type the below comments to convert the BMI column up to two decimal values, press "tab", and then press "enter".
2. # convert the BMI column to two decimal values

# convert the BMI column to two decimal values



1. In the same code cell, type the below comment, press "tab", and then press "enter" to get the output similar to the below image.
2. # Show the new column only

# Show the new column only

A screenshot of a computer

Description automatically generated

Task 3.2: Machine Learning

1. Type the below comments to performe the data analysis and summary statistics on dataset, press "tab", and then press "enter" to get the output.
2. # Perform count, min, max, std, mean, 25%, 50%, and 75% on the dataset

# Perform count, min, max, std, mean, 25%, 50%, and 75% on the dataset

A screen shot of a computer

Description automatically generated

1. Once the cell run is completed, you will get an output similar to the below image.

A screenshot of a computer

Description automatically generated

1. Before building the model, the main frame is to split the data into train tests and splits and this would be done by Copilot itself. Type the below comments, press "tab" and run the cell.
2. # Split the data into training and testing data and the column name Diabetic is the target column

# Split the data into training and testing data and the column name Diabetic is the target column

A screenshot of a video game

Description automatically generated

A screenshot of a computer

Description automatically generated

Task 4: Data Visualization and Data Transformation

Task 4.1: Data Visualization

1. Click on **+Code** to open the new cell, and type the below comments to performe the basic operations on dataset, press "tab", and then press "enter" to get the output.
2. # Perform univariate analysis on the dataset and plot the graphs

# Perform univariate analysis on the dataset and plot the graphs

A screen shot of a computer

Description automatically generated

1. Once the cell run is completed, you will be getting a graphical representation output similar to the below image.

A screenshot of a computer

Description automatically generated

1. Now, type the below comments for specifying certain commonly used plots for visualization, press the "tab", and then press "enter" to get the output as shown in the below-given image.
2. # Perform scatter plot on the dataset and plot the graphs

# Perform scatter plot on the dataset and plot the graphs

A screenshot of a computer

Description automatically generated

1. Type the below comments for performing Joint Plot or Grid (rarely used plots) for visualization, press "tab", and then press "enter" to get the output as shown in the below-given image.
2. # Perform Joint Grid plot on the dataset and plot the graphs

# Perform Joint Grid plot on the dataset and plot the graphs

A screenshot of a computer

Description automatically generated

A graph with blue dots

Description automatically generated

1. Type the below comments for performing visualization on all features of the dataset, press "tab", and then press "enter" to get the output as shown in the below-given image.
2. # Perform comparison on all features of the dataset and plot the graphs in a single plot using heatmap

# Perform comparison on all features of the dataset and plot the graphs in a single plot using heatmap

A screen shot of a computer

Description automatically generated

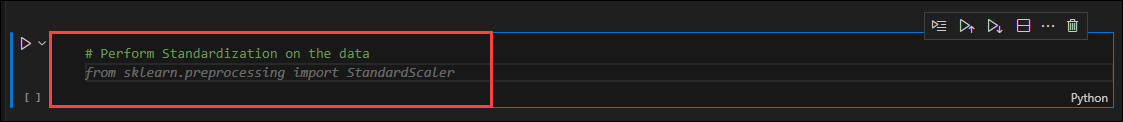
A screenshot of a computer

Description automatically generated

Task 4.2: Data Transformations

1. Click **+Code** to add a new cell, and type the below comments for Standardization which is the process of scaling and centering numeric features to have a mean of 0 and a standard deviation of 1, making them comparable and suitable for certain algorithms, press "tab", and then press "enter".
2. # Perform standardization on the data

# Perform standardization on the data



A screen shot of a computer

Description automatically generated

1. You will get the output as shown in the below-given image once the code cell finishes running.

A screen shot of a computer

Description automatically generated

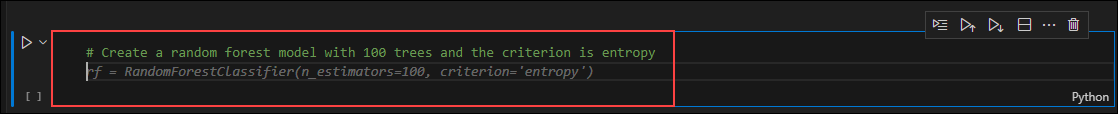
Task 5: Training the sample model

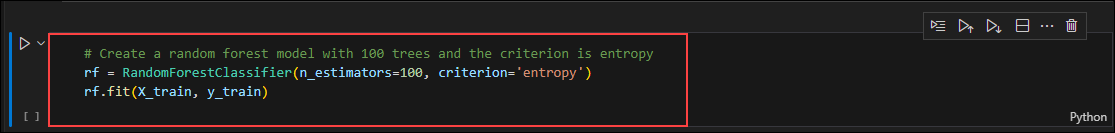
Training a sample model using Random Forest.

**Note**: The model training is a continuation of the train test split step to train the model; run the train test split step first and then continue with the model building.

1. Click **+ Code** to add a new cell, and type the below comments for creating the random forest model, press "tab", and then press "enter".
2. # Create a random forest model with 100 trees, and the criterion is entropy

# Create a random forest model with 100 trees, and the criterion is entropy





A screen shot of a computer screen

Description automatically generated

1. In a new cell, type the below comments for calculating the accuracy of the model, press "tab", and then press "enter" to see the output.
2. # Calculate the accuracy of the model

# Calculate the accuracy of the model

A black screen with a red border

Description automatically generated

Summary

In this exercise, you have successfully leveraged the GitHub Copilot for Machine Learning.