**Sanjay Mahabir – 301231274**

**COMP 216**

Networking for software Developers

Assignment 3

**Description:**

For this assignment, you will do basic python programming. You will be provided with a dataset that describes the factors. **Your task is to predict the values, with the provided features**. You may find it necessary to drop irrelevant features, or to combine it.

**Instructions:**

**Google colab link for this assignment:**

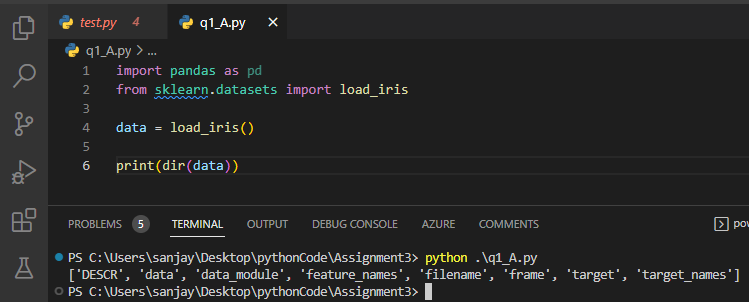
* https://colab.research.google.com/drive/1dG06cjdI3P8LHE0AYfgBXkXY0q6ctOXL?usp=sharing
* From the sklearn. Datasets – use datasets other than breast\_cancer, digits

(You will lose marks to work on the above-mentioned two datasets)

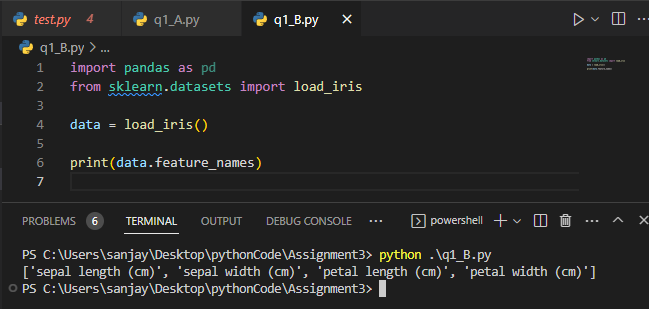
* Import pandas, matplotlib, sea born, scikitlearn/ other necessary libraries.
* Take the snapshot of the executed results along with code, paste it under each question of the word document – this file needs to be submitted.
* Marks are reduced - If the results are unclear, out of order, only graphs without code or pasted incorrectly
* Understand the machine learning concepts prior to the assignment work
* Watch the broadcasted video on machine learning.
* DO NOT SUBMIT ZIPPED FOLDERS – marks you get will be 0.
* Avoid late submission – Check at the due date.
* Assignments submitted after the due date are evaluated for 70%

**Questions**

1. Import the dataset, and examine the data
2. Use dir() function on the dataset and find out the list of files



1. Find out the encoded labels for each data



1. A screenshot of a computer

   Description automatically generated with medium confidenceUse the proper Matplotlib tool for visualization and plot the images

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

Chart

Description automatically generated

1. Using pandas, create the dataframe to manipulate the data and add a new column to the data frame, show it in the screen shot.

Text

Description automatically generated

1. Use the proper sklearn function to create the training data and test data , Use the Random forest classifier to predict the values of the model (use the minimum of 30 trees).

A screenshot of a computer

Description automatically generated

1. Plot the Confusion matrix (8,5) to evaluate the accuracy of the classification. Compare the predicted and actual values and show it in graph.

Text

Description automatically generated

Chart

Description automatically generated