ASSIGNMENT-5

Student Name : Pavan Gandavarapu

Student Id : 700741173

Video link : https://drive.google.com/file/d/1DqdIsdKxJFGxLrKXK3ajd6rwaeRKjZuq/view

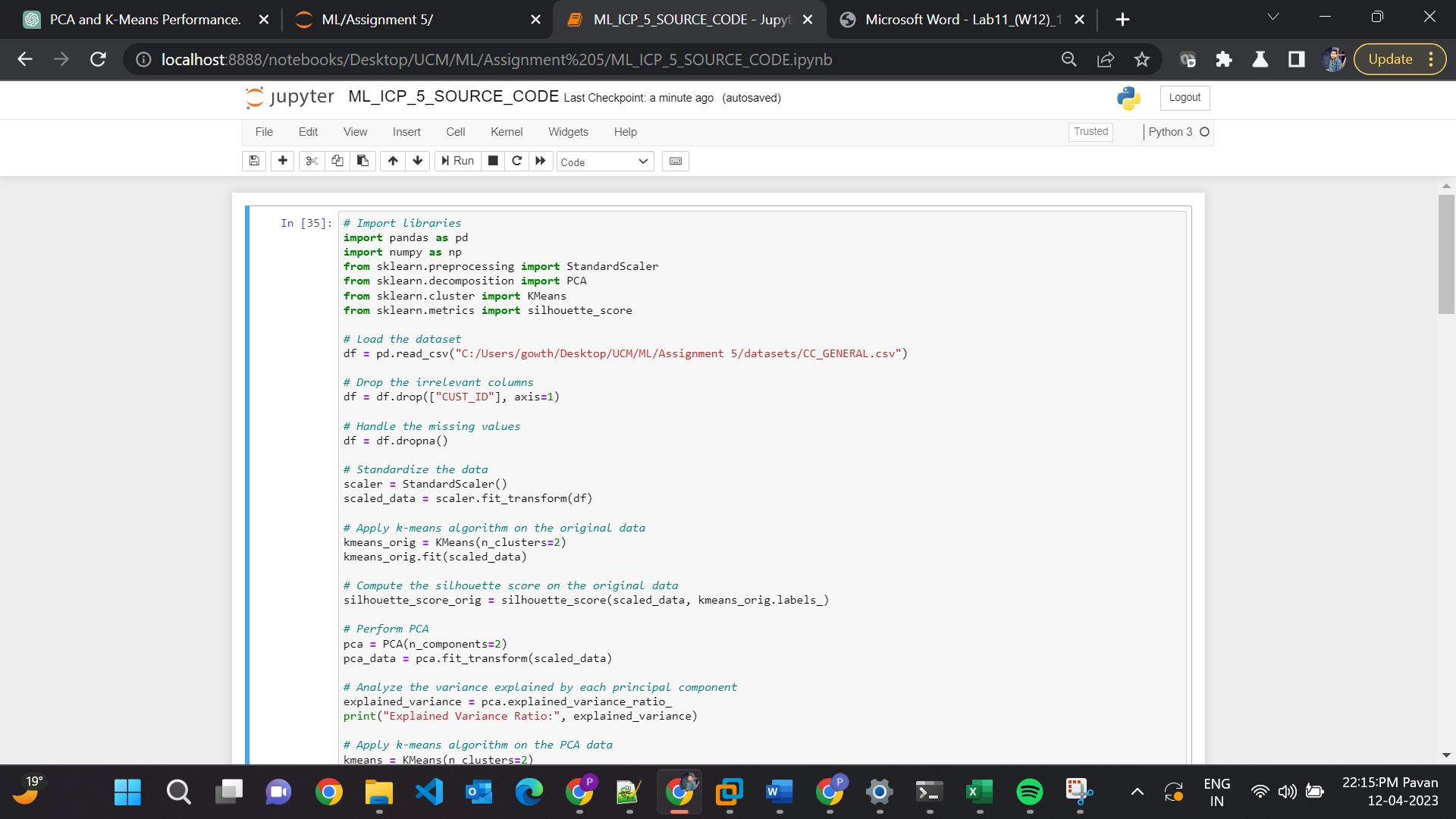
GIT Hub URL: https://github.com/PavanGandavarapu/ML\_Assignment\_5

1. Principal Component Analysis

a. Apply PCA on C:/Users/gowth/Desktop/UCM/ML/Assignment 5/datasets/CC\_GENERAL.csvdataset.

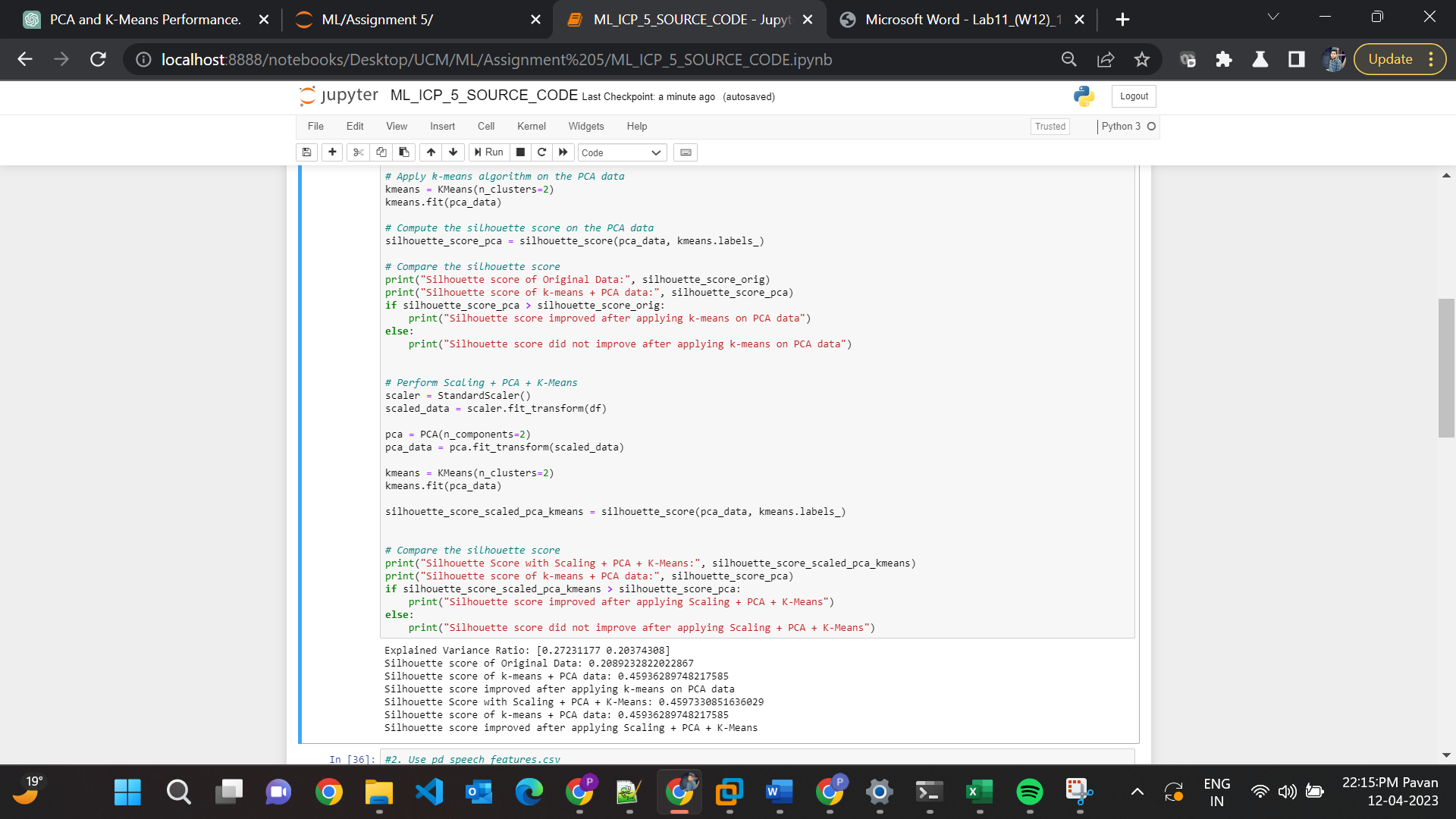
b. Apply k-means algorithm on the PCA result and report your observation if the silhouette score has improved or not?

c. Perform Scaling+PCA+K-Means and report performance.



**Explanation:**

* Load the "CC\_GENERAL.csv" dataset, drop the irrelevant columns, and handle the missing values.
* Standardize the data using StandardScaler to make sure all features have the same scale and are centered around zero.
* Apply k-means algorithm on the original data and on the PCA-transformed data, and calculate the silhouette scores for each.
* Compare the silhouette score obtained from applying k-means on PCA data with the silhouette score obtained from the original data to see if there is any improvement.
* Compare the silhouette score obtained from applying Scaling + PCA + K-Means with the silhouette score obtained from k-means + PCA data to see if there is any improvement.

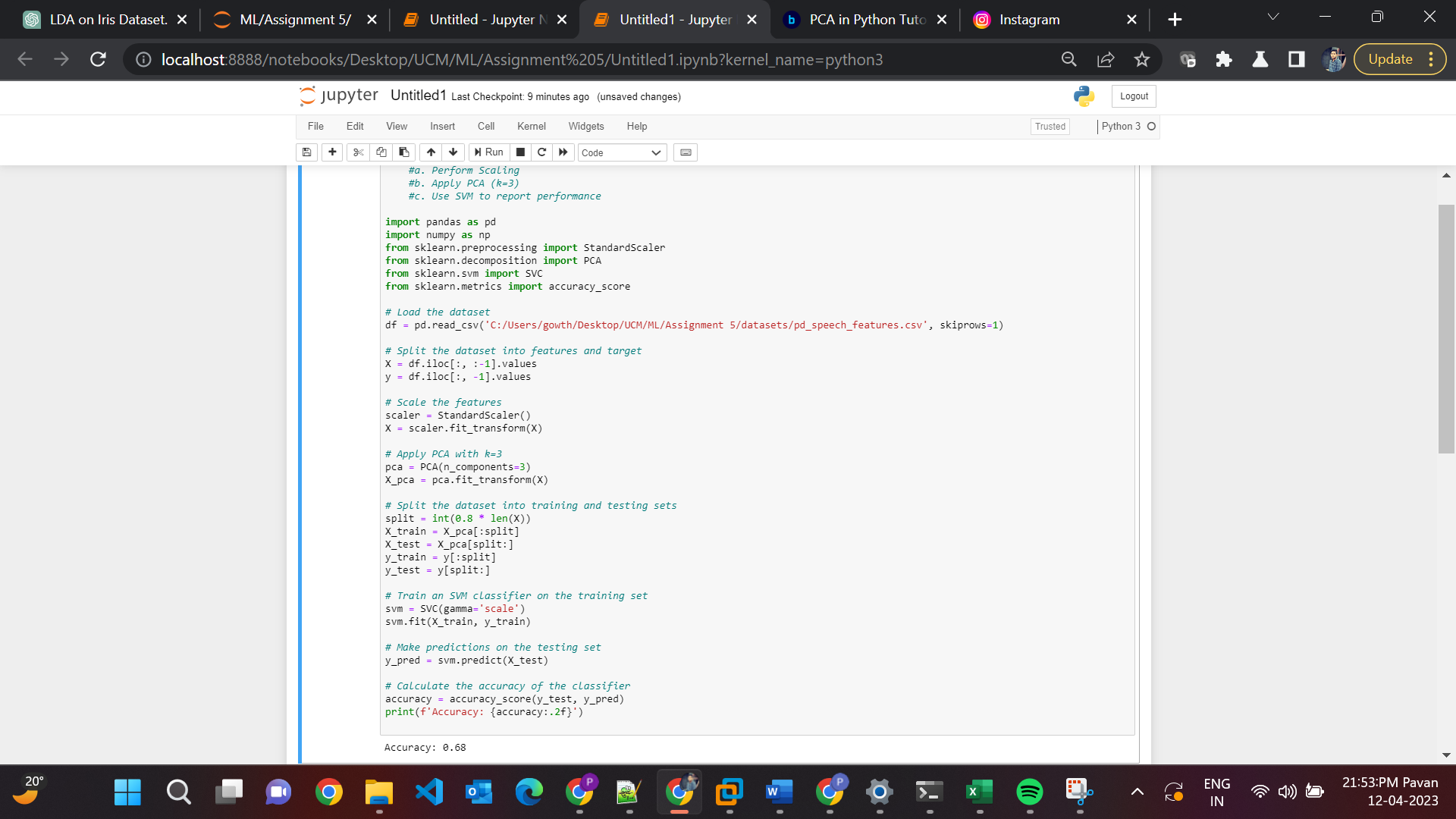


2. Use pd\_speech\_features.csv

a. Perform Scaling

b. Apply PCA (k=3)

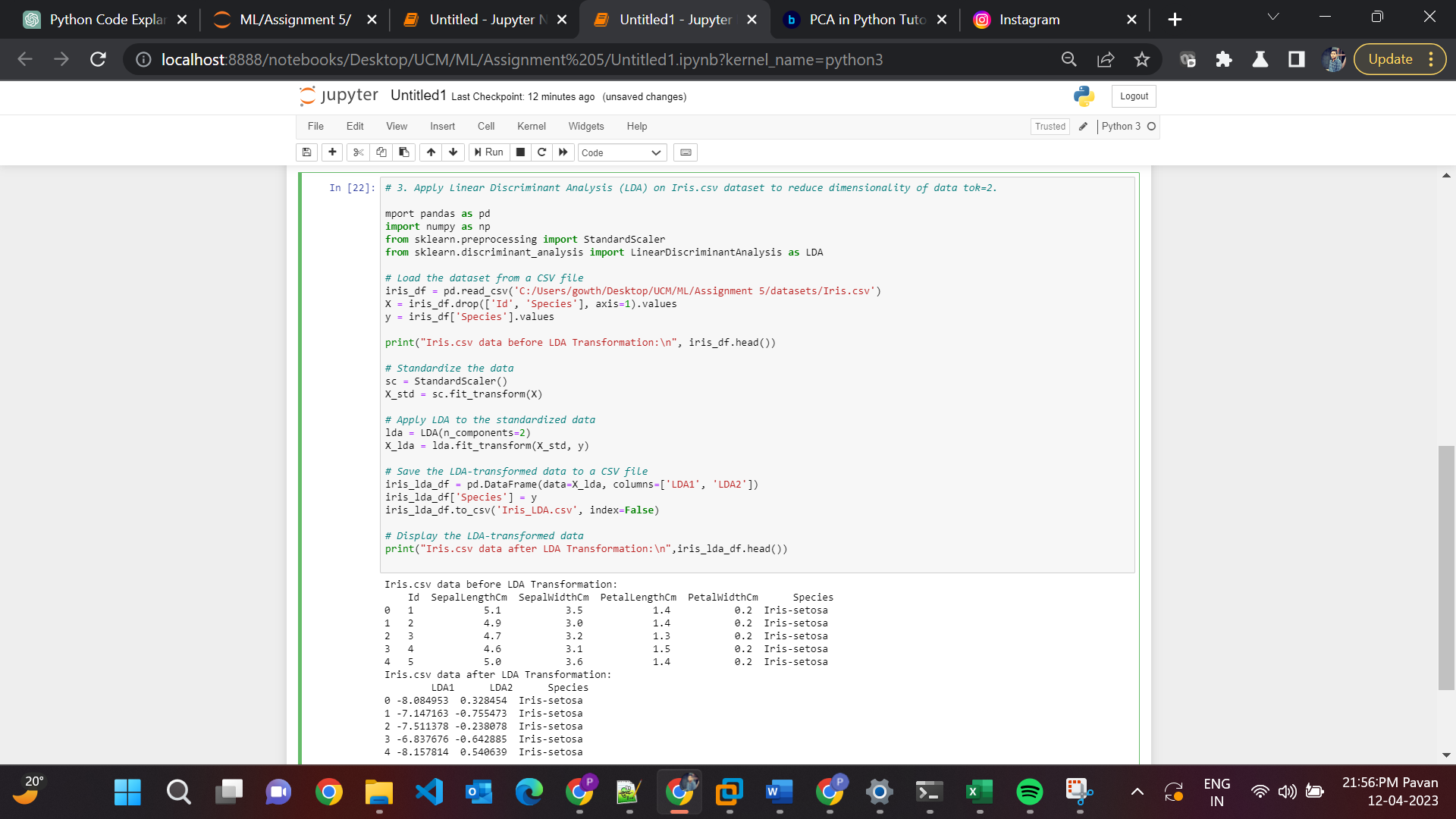
c. Use SVM to report performance



**Explanation:**

* Load the "pd\_speech\_features.csv" dataset using Pandas and split it into features (stored in "X") and target variable (stored in "y").
* Scale the features using the "StandardScaler()" function from scikit-learn.
* Apply Principal Component Analysis (PCA) to reduce the dimensionality of the dataset to 3 principal components using the "PCA()" function from scikit-learn.
* Train a Support Vector Machine (SVM) classifier on the training set using the "SVC()" function from scikit-learn.
* Evaluate the performance of the classifier on the testing set using the "accuracy\_score()" function from scikit-learn and print the accuracy.
* Top of Form

3. Apply Linear Discriminant Analysis (LDA) on Iris.csv dataset to reduce dimensionality of data tok=2.



**Explanation:**

* Load the Iris.csv dataset: The code loads the Iris.csv dataset using Pandas and separates it into features (stored in "X") and target variable (stored in "y").
* Standardize the data: The code uses the "StandardScaler()" function from scikit-learn to standardize the features by subtracting the mean and scaling to unit variance.
* Apply Linear Discriminant Analysis (LDA): The code uses the "LinearDiscriminantAnalysis()" function from scikit-learn with the argument "n\_components=2" to reduce the dimensionality of the dataset to 2 components.
* Save the LDA-transformed data to a CSV file: The code saves the LDA-transformed data to a new CSV file named "Iris\_LDA.csv".
* Display the LDA-transformed data: The code displays the LDA-transformed data in a new Pandas dataframe and prints it to the console.

4. Briefly identify the difference between PCA and LDA

| PCA | LDA |
| --- | --- |
| Unsupervised learning method | Supervised learning method |
| Aims to maximize the variance in the data | Aims to maximize the separation between different classes |
| Doesn't use class labels or any other external information | Uses class labels to guide the transformation process |
| Primarily used for data compression, noise reduction, and visualization purposes | Mainly used for feature extraction, dimensionality reduction in the context of classification, and visualization of high-dimensional data |
| Sensitive to the scale of the variables; standardizing the data is recommended | Less sensitive to the scale of the variables, but standardizing the data is still recommended |
| Doesn't consider the class structure of the data, so it might not always lead to the best separation between different classes | Assumes that the data is normally distributed and the classes have identical covariance matrices |
| May be less effective than LDA in reducing dimensionality in the context of classification | Can lead to better separation between different classes and improved classification accuracy compared to PCA |