**CIS-660 Lab 3**

**Submitted By:**

A person standing in front of a bridge

Description automatically generated with medium confidence

**Name: Mohamed Gani Mohamed Sulthan**

**CSU ID: 2811619**

**Grail Id: momohame@grail.eecs.csuohio.edu**

**Master’s in computer science**

**Designing and Building a Prediction Model for Bike Buyer Data with a Classifier**

Choose any two classifiers covered in class and apply to your Bike Buyer data SetPlan your experiment with:

1. Determine Data preprocessing methods required to apply for each of your Classifiers

**Solution:**

We have already selected the features and done with preprocessing in the labParts. We’re going to get the dataset and consider the input features given by professor.

**Process:**

**Step 1:** Import the required modules

**Step 2:** read the csv file

**Step 3:** Apply the preprocessing techniques

**Step 4:** Done Binarization for Gender, converting to binarization for NumberCarsOwned, TotalChildren, NumberChildrenAtHome, One hot encoding for EnglishEducation, EnglishOccupation, CommuteDistance, Region and normalization for YearlyIncome and Age

**Step 5:** Display dataFrame

**Code:**

Graphical user interface, text, application

Description automatically generated

**Applying models:**

**KNN Model:**

**Step 1:** Allocating the values without BikeBuyer as X and only BikeBuyer as Y

**Step 2:** Splitting the X and Y to train and test the dataset

**Step 3:** Scaling the training dataset

**Step 4:** Calling the KNeighborsClassifier with the groups and fit that in the train dataset

**Step5:** Calling the confusion\_matrix and pass the y\_test and y\_pred

**Step 6:** Print the accuracy score

Graphical user interface, text, application

Description automatically generated

**Decision Tree:**

**Step 1:** Allocating the values without BikeBuyer as X and only BikeBuyer as Y

**Step 2:** Splitting the X and Y to train and test the dataset with the test size

**Step 3:** Scaling the training dataset with StandardScaler()

**Step 4:** Fitting the transformation in to X\_train and X\_test

**Step5:** Calling the decision tree classifier and train the dataset

**Step 6:** Predicting the final accuracy

Graphical user interface, text, application

Description automatically generated

1. **For each classifier,**
2. **Compare the accuracy of the classifier with two different sets of input parameters if applicable**

**Solution:**

**Trying with different sets of input parameters for KNN:**

Tried test\_size with 40% of data and neighbors of 8

Graphical user interface, text, application

Description automatically generated

Tried test\_size with 25% of data and neighbors of 5

Graphical user interface, text, application

Description automatically generated

**Trying with different sets of input parameters for Decision Tree:**

Tried test\_size with 25% of data

**Graphical user interface, text, application, email

Description automatically generated**

Tried test\_size with 40% of data

Graphical user interface, text, application, email

Description automatically generated

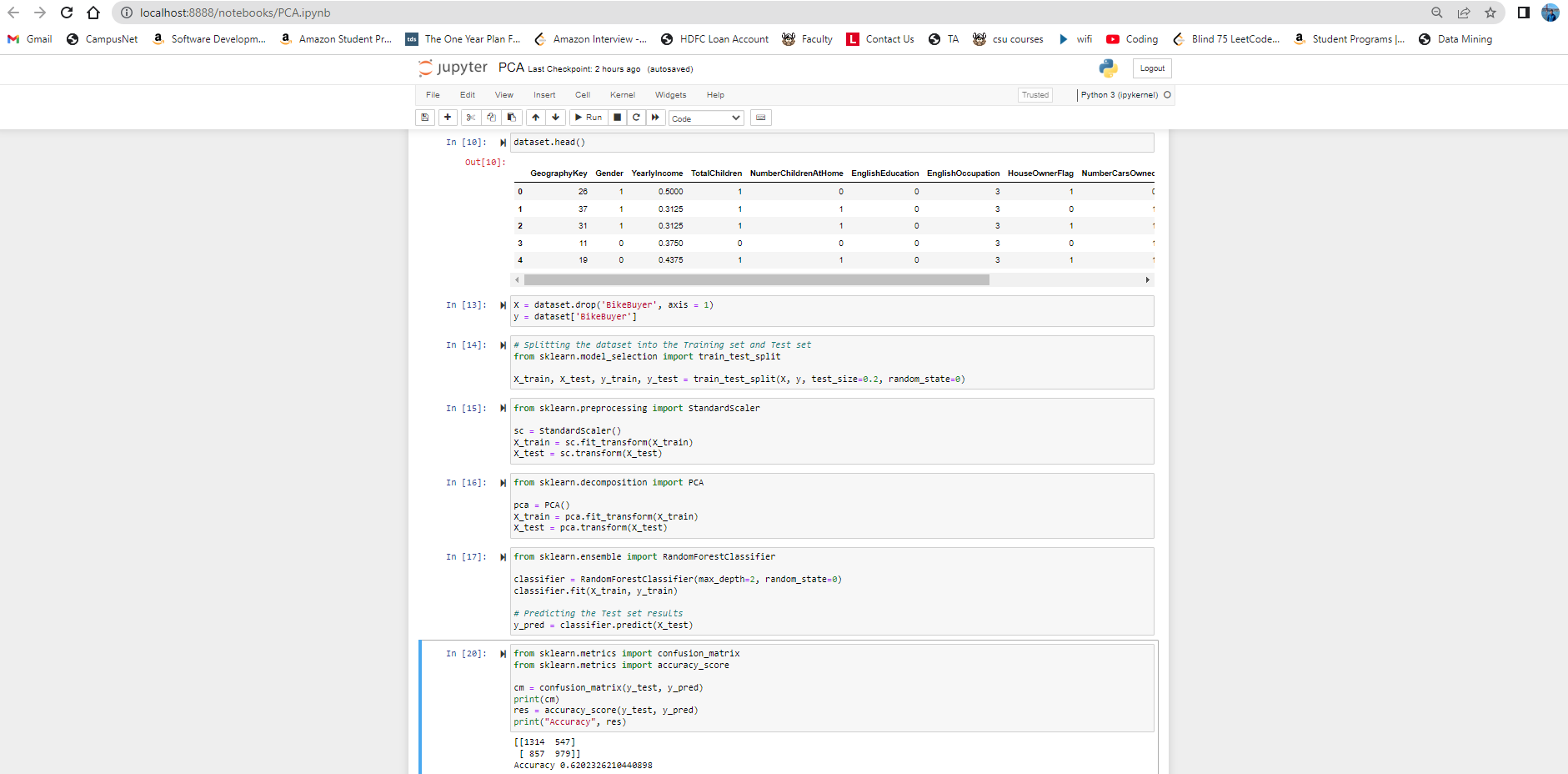
**Extra Credit:**

**PCA Implementation:**

**Step 1:** Initialize the PCA class by passing the number of components to the constructor.

**Step 2:** Call the fit and then transform methods by passing the feature set to these methods.

**Step 3:** The transform method returns the specified number of principal components.

****

1. **Compare the accuracy of each test of the classifiers**

**Solution:**

The accuracy of KNN score is 0.72 and decision tree score is 0.73. In the KNN algorithm it groups the values with the parameter we passed but in decision tree it was trained with the test size and fetched the correct results.

1. **Validate your result with your Test Set to compare the Accuracy of your models for each classifier with different Parameter settings or different transformation method.**

**Solution:**

I already tried with different input parameter to validates the result

Tried test\_size with 40% of data and neighbors of 8

Graphical user interface, text, application

Description automatically generated

Tried test\_size with 25% of data and neighbors of 5

Graphical user interface, text, application

Description automatically generated

Tried test\_size with 25% of data

**Graphical user interface, text, application, email

Description automatically generated**

Tried test\_size with 40% of data

Graphical user interface, text, application, email

Description automatically generated

**4. Discuss about your results:**

**a. Why your inducted model is different for the same training data as you change the parameter values or the classifier.**

**Solution:**

Whenever we are changing the values of the test size and clustering the group, it changes a measure of how sensitive the algorithm is to the specific data used during training. If we get a different value then variance happened in our algorithm

**b. Why a certain parameter setting or a classifier shows with better accuracy than the others that you tried**

**Solution:**

When I’m trying to change the test\_size of the data set in KNN and reduced the neighbors I got a low accuracy compared to the smaller set of groups.

When I’m dealing with the decision tree classifier I tried to keep the constant test\_size dataset to maintain standard of the train data set should be 60 and it provides me the good accuracy

**c. Anything you observed**

**Solution:**

From the selected features, I observed there will be an at least 70% of people will buy the bikes. By using the two different types of algorithms, both ended up greater than 70% to when passed the test set.