THE UNIVERSITY OF HONG KONG
MSc in E-Commerce and Internet Computing

ECOM7126: Machine Learning for Business and E-Commerce
(2022-2023)

Miaoling Mai   Student ID: 3036043761   17, March 17, 2023

Assignment 2 – Diamond Classification

1. Process

1) Load the data from the CSV file

2) Take a Quick Look at the Data Structure using head, info, describe and shape function

3) Remove outlier according to pairplot

4) Convert categorical data to numerical data

5) Create a test set

6) Build a OneVsRestClassifier model

7) Evaluate the model using Cross- Validation

2. Advice to the management of the company

The accuracy of this model is just about 70%. The company should feed more data on a regular basis to improve accuracy. Also, they could use a semi-supervised model to improve the accuracy furtherly.