Data Mining Assignment – 2020

# Members:

## Member 1 – Team Leader

Full name: Gokhan Arkan

Student number: 33493223

Email: [garka002@gold.ac.uk](mailto:garka002@gold.ac.uk)

## Member 2

Full name: Juan David Diaz Duran

Student number: 33551395

Email: [jdiaz001@gold.ac.uk](mailto:jdiaz001@gold.ac.uk)

# Contribution

The project split into half, and the members shared equal work. The collaboration has been done on Github and Google Colab. The communication has done with Skype.

### Part 1:

Implementation of KNN was done by following pair programming practice with Skype Screen Sharing. The observer and driver roles occasionally exchanged, and at at the end, both members shared the equal work.

### Part 2:

In order to get better predictions, the datasets were prepared by Juan as the first function of Part 2. Then both team members worked on nested cross-validation together to make sure all data is accurate. There are four other features written. Model and visual evaluations are done by Juan, printing confusion matrices and finding the best parameter function done by Gokhan.

We went further and individually created all the models one by one. Initial two models, KNN and Decision Tree, done by Juan. The rest of the models, Random forest, AdaBoosting, Bagging and SVM, done by Gokhan.

Each member has written their own comments, and we cross-checked each other's comments to comply with the same commenting style.