ISSU0053 Data Science and Big Data Analytics **Report ID:** **1043334**

UCL International Summer School for Undergraduates 2019

**Assessment I: Computer Practical Work and Write-up (50%)**

|  |  |  |
| --- | --- | --- |
| **Task** | **Mark** | Comments |
| T1: Explore the dataset | A+ |  |
| T2: Linear **/** logistic regression **/** LDA | B |  |
| T3: Decision Tree methods | A |  |
| **Overall task:** | A |  |
| Use of Figures | A+ |  |
| Introduction, Commentary, Summary | A |  |
| Coding Technique | A |  |
| Writing Standard | A |  |
| Structure & Presentation | A |  |
| **Overall:** | A+  (80%) |  |

**Strengths:**

- overall a good report with clear comments and easy to follow

- implemented all the main tasks

- you seem to have understood the main concepts covered in class

- also implemented a classification task (although the problem was a bit primitive – so not surprising accuracy was so high)

**Focus on improvement:**

- it’s a regression task. Why do you use logistic regression? I am not sure whether this was a typo or whether you didnt understand the difference between linear and logistic regression

- why do you tune the length of your tree? Caret::train can select the best tune – I think 50 is too overfitted for this dataset