ISSU0053 Data Science and Big Data Analytics **Report ID: 1043358**

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 | A |  |
| T3: Decision Tree methods | A | Tunelength in trees not fully explored but Implemented KNN and NB |
| **Overall task:** |  |  |
| Use of Figures | B |  |
| Introduction, Commentary, Summary | A-B |  |
| Coding Technique | A |  |
| Writing Standard | A |  |
| Structure & Presentation | A |  |
| **Overall:** | A  (75%) |  |

**Strengths:**

- good and clear report, easy to follow

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

- good intro and conclusion which presents all the built models and their accuracies

- implemented extra models (KNN and NB)

**Focus on improvement:**

- no comments throughout the code

- In the t-test 0 Hypothesis and Alternative Hypothesis not stated (this should be a report – not me digging into your code, trying to understand it)

- trees are not understandbale – why didnt you leave tunelength parameter out to see what caret::train() selects as best length? Tuning the length to 100 is probably no necessary for this dataset and results in an overfitted model (that’s probably why your decision tree model was not that great on the test set)