ISSU0053 Data Science and Big Data Analytics **Report ID: 1043336**

UCL International Summer School for Undergraduates 2019

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

|  |  |  |
| --- | --- | --- |
| **Task** | **Mark** | Comments |
| T1: Explore the dataset | A-B |  |
| T2: Linear **/** logistic regression **/** LDA | A+ | Implemented LDA AND LR |
| T3: Decision Tree methods | A+ | Implemented KNN |
| **Overall task:** | A |  |
| Use of Figures | B |  |
| Introduction, Commentary, Summary | B |  |
| Coding Technique | A |  |
| Writing Standard | A |  |
| Structure & Presentation | A-B |  |
| **Overall:** | A  (70%) |  |

**Strengths:**

**-** one can see that you understood everything that was covered in class and implemented all the required algorithms and seem to understand them well

- implemented more algorithms than asked for and found best one (KNN) – well done!

**Focus on improvement:**

- you don’t explain what you’re testing in the t-test. What is the 0 Hypothesis and what is the Alternative Hypothesis?

- I remember you saying that there were a lot of missing values (soil type?) - where did they go?

- you are not engineering any new features – section should be just called EDA

- you have a lot of plots that you dont explain/describe

- some sections of your report are hard to follow (disable warning messages!)

- not enough comments and explanations throughout the report on what exactly you are doing

- summary at the end that sums up your findings and the most important variables would have been helpful

- your kaggle score would have been interesting too