Data Mining

Assignment 1

IMPORTANT:

* Sections 1-8 should be no longer than 10 pages (minimum 8 pages), including diagrams, images, screen captures, tables, etc. Careful selection of these is needed.
  + Code does not count to this total. Code should be added to the relevant section.
  + Detailed discussion is expected. Marks are awarded based on depth of information given.

Student Details for Group Submission:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Student Name  (as per Student Card) | Student Number | Prog Code | Stream  (DS, ASD, etc) | Year  (1, 2, …) |
|  |  |  |  |  |
|  |  |  |  |  |

For group submission, all students must insure they are compliant Plagiarism policy.

For all students in group, submission of an assessment, either physically or electronically, with or without a cover sheet, acknowledges your compliance with the TU Dublin Academic Regulations and Policies.

1. **Definition of Problem**

Clearly state the problem definition, what type of data mining task is it, where was the data set sourced from, etc.

1. **Data Exploration & Descriptive Analytics**

Include any data insights discovered

1. **Data Preparation**

Include details of any data cleaning, transformations, data enrichment, feature engineering, feature reduction, etc

1. **Details of Algorithms & Configurations**
2. **Model Performance Metrics & Evaluation of Results**
3. **Identification of the most important variables**
4. **Comparison with other Research & Reflections**

Compare your results to at least three other researchers (maximum of five) who used the same dataset. What lessons did you learn from doing this? How can your work be improved? Did you include any improvements in your work and what impact did it have?

1. **Action plan to improve equality and other social issues based on your data**
2. **References**

Use Harvard referencing style.