Module COM662 January, 2021 **Data Analytics**Dr R Bond & Dr Jose Santos

**Final Year - Assignment 1**

Assignment hand-out: Week 1, Hand-in: Tuesday **4th May 2021** (week 13) **at 5.30pm**

You will receive feedback ~ 28th May 2021

(Note: No paper-based materials will be required – all materials and administration will be electronic.)

Contribution towards course module: 100% (and scored out of 100 marks)

Keep a copy of all submitted coursework – i.e. your computer files. When you submit your coursework do take a screenshot of the completion screen as your receipt. Remember plagiarism is a serious offense – ensure that you use your own words even when referencing a source. Also reference libraries, tutorials or code that you used.

**Learning Outcomes:** To recognise and illustrate understanding of data analytics, modelling, data science workflows and pipelines. To operate integrated development environments for statistical computing as well as the practical use of statistical programming languages. Use intuition and complex reasoning skills for synthesising, exploring and understanding concepts in data analytics. To practice exercising research skills such as report writing and the articulation of complex ideas

**Practical Data Analytics Report**

**Assignment Brief**

Using R Studio, you are to carry out data analytics of an openly available dataset e.g. from UCI or an Open Data repository such as Open Data NI, and write a report detailing your methodology and analysis of this open dataset showing results and data visualisations.

The aim of this assignment is to demonstrate your learning of key elements of data analytics and to build your knowledge of what is possible using R. Each week you should be able to incorporate additional components into your report in line with your practical work, lectures and directed reading. As the weeks progress, retain the analysis you have undertaken in earlier weeks, so the report and R script is your cumulative work. **You will receive feedback within 20 working days.**

**Report**

You are required to submit a data analytics report not exceeding **3500 words. The report will consist of 5 sections: 1) Introduction, 2) Methods, 3) Results, 4) Discussion and 5) Conclusions.** Within each section of the report you should describe the technical details of the analysis and the techniques used, and in some cases describing the R code you developed. Further detail and examples are provided in Table 1. Please include references to any research articles and/or online resources that helped you during the data analysis.

**Table 1: Example details for each section in report**

|  |
| --- |
| **Introduction**  Provide anintroduction to the chosen domain and dataset. For example, where the data was obtained from, what data is contained in the set. The size of the dataset in terms of instances and features. What are the classes of the dataset. Descriptive statistics, data visualisation, exploratory data analysis etc. |
| **Methods**  Provide detail on the data analytics process you have undertaken, e.g. data cleansing and preparation, feature selection, data modelling techniques (such as K-NN, Bayesian) and your rationale for selecting these along with code snippets. |
| **Results**  Application of methods and metrics for model evaluation (e.g. k-fold cross validation, sensitivity, specificity, F-measure, receiver operator characteristic analysis and area under the curve). Visualization of results. |
| **Discussion**  Provide a summary of the various results that you have obtained from the analysis and what these tell us. |
| **Conclusion**  Overview of the overall data analytical process undertaken and the interpretation of the results. Any limitations noted or changes you could make in future analysis e.g. boosting algorithms etc. |

**Submission**

The report will be electronically submitted online using Blackboard along with your R script file as an appendix (script file/s do **NOT** count towards the word limit)*.* You will be given an individual mark based on the assessment criteria.

|  |  |
| --- | --- |
| **Assessment Criteria** | **Marks** |
| **Introduction to the chosen domain and description of the dataset**  **Excellent (25-30 marks):** Introduction to domain and dataset were very detailed and clearly described.  **Good (19-24 marks):** Introduction to domain and dataset were detailed and clearly described.  **Satisfactory** **(13-18 marks)**: Introduction to domain and dataset were adequately detailed and described.  **Poor (7-12 marks):** Introduction to domain and dataset were poorly detailed and described.  **Very Poor (0-6 marks):** Introduction to domain and dataset were extremely basic. | **30 Marks** |
| **The rigor, detail and complexity of the data analytics process**  **Excellent (25-30 marks):** An impressive data analytics project showing clear processes, very complex techniques and detailed results  **Good (19-24 marks):** A good data analytics project showing complexity, a clear process and some detailed results.  **Satisfactory** **(13-18 marks)**: An adequate data analytics project showing little complexity, little regard for data analytics processes and showing basic results.  **Poor (7-12 marks):** A poor data analytics project showing no complexity and very basic results.  **Very Poor (0-6 marks):** A very poor data analytics project showing no complexity and no clear results. | **30 Marks** |
| **Overall quality, completeness and innovation in the data analytics project**  **Excellent (17-20 marks):** High quality innovative project that is complete and uses impressive data visualisations.  **Good (13-16 marks):** Good quality project showing some innovation and incorporates good data visualisations and is somewhat complete.  **Satisfactory** **(9-12 marks)**: Decent quality project but shows basic innovation and incorporates basic data visualisations and is only adequately completed**.**  **Poor (5-8 marks):** Poor project showing little innovation and incorporates no interesting data visualisations and is far from being completed.  **Very Poor (0-4 marks):** Verypoor project showing no innovation and incorporates no interesting data visualisations and is very far from being completed. | **20 Marks** |
| **Quality of the referencing style, grammar and punctuation and also the evidence of synthesizing research and making conclusions**  **Excellent (17-20 marks):** Written report is excellent with logical formatting, flow and layout. An academic referencing style with excellent research articles cited. Excellent grammar and writing style.  **Good (13-16 marks):** A well written report with logical formatting, flow and layout. An academic referencing style with some good research articles cited. Good grammar and writing style.  **Satisfactory** **(9-12 marks)**: An adequately written report with some formatting and layout. Some academic referencing with only a few adequate research articles cited. Grammar and writing style is fine but there are some errors.  **Poor (5-8 marks):** A poorly written report with little formatting and layout. Very little academic referencing and grammar and writing style is not adequate.  **Very Poor (0-4 marks):** A very poorly written report with no formatting or proper layout. No academic referencing and grammar and writing style is very poor. | **20 Marks** |
| Total | 100 |