TU59/TU60 - Data Mining Assignment B – Portfolio (60%) **Due Before : Wednesday 5th January @23:00**

Portfolio Assessment

This assessment builds upon the work you completed in Assignment A.

This assessment is individual [No groups allowed]

With this assessment you will complete exercises for each of the main topics covered in the module. These can be used to showcase your work as a Data Scientist to perspective employers.

Required Tasks

You are to complete a full example of applying the following approaches to a suitable dataset, to include details of analysis and reflections on findings.

There are **four** parts to this Portfolio Submission.

Select **three** of the following topics, from the box on the left, and for each select an appropriate dataset and problem, complete the task, document your work and insights. Additionally, complete the task defined in the box on the right.

- Association Rule Analysis. [SAS]
- Clustering [SAS / Python]
- Text Mining [Python]
- Forecasting [Python]
- Comparing use of Classification versus including/using Segmentation and Clustering as a step before Classification. Compare Classification model performance including/using versus not using it. [SAS / Python]

+

Select two case studies / examples, discuss the ethical and legal issues of using Data Science, Machine Learning, Artificial Intelligence, etc., how these may impact the results from Data Mining task, and how they might impact on the work of the Data Scientist.

Important: The dataset should be **non-trivial** (eg. Iris data set and similar), and should **not** include any of the datasets and examples used throughout the module i.e. use a different dataset.

Datasets: See links to various websites on the module webpage. Also check out this post with lots of links to various dataset repositories. https://oralytics.com/2019/04/18/data-sets-for-analytics/

Software: Use SAS Enterprise Miner or Python to complete all the Data Mining tasks. For Forecasting and Text Mining, you can use Python. Tool/language to use is indicated in the above box.

References & Citations: Include a References section for each Task. Each reference should be cited in your text. Use an referencing/citation method you are used to or select APA or Numeric Style.
https://tudublin.libguides.com/APA_quick_guide
https://www.dit.ie/media/library/documents/Numeric.pdf

Deliverables

You will be required to document your work. I've provided a template document, including plagiarism cover page, for this assessment on Brightspace. Use this template document. You can alter/change the suggested format/headings.

Only include details and images which are important for your findings and narrative. Do not fill your report with reports/charts/etc which do not add to your discussion.

TU59/TU60 - Data Mining Assignment B – Portfolio (60%) **Due Before : Wednesday 5th January @23:00**

Document / Report

See the sample document/report (on BrightSpace). This document contains generic sections for you to complete for each part of this assignment. These sections are just suggestions! You can modify or use your own section headings. Whatever makes sense to you for completing the assignment and telling your story.

Submission Details

The assignment is due on 5th January @23:00

You should create one document/report containing all the material for each part of the assignment. Convert this document into a PDF. It is this PDF document that should be submitted.

Make sure you complete the Plagiarism Report on the first page.

You will need to submit your assignment on **BrightSpace VLE**. You cannot submit your assignment via email.

Q&A and Support

You should commence work on this portfolio immediately. This will allow you to ask questions and gain some guidance on each task during the lab sessions each week.

Week 13 will be used as session to discuss any questions you may have.

I will be available up to 22nd December to answer any questions via email. All questions asked and their answers will be posted on the FAQ for the module. Make sure to check the FAQ regularly.

Post Christmas & New Year Break: I'll be available on 4th and 5th January (until 16:00 each day) to answer any questions via email. Again, updates will be made to FAQ.

Important: I will have no access to emails during the Christmas and New Year break. It is vital you commence your work early on this assessment and raise any questions as they arise, ideally before the Christmas break.

Marking Scheme

See Marking Rubric

The documentation for your assignment must contain the name, student number, class, course (**TU??**) and year information for each student in the group. **Failure to give this information will incur a 10% penalty.**

Each submission must be original work as plagiarism will result in a **zero** mark (0%).
There will be a 10% penalty deduction will be applied for each day the assignment is late.
There is no penalty for submitting early.

DIT Plagiarism Policy : https://tudublin.libguides.com/c.php?g=674049&p=4794713
https://www.tudublinsu.ie/advice/exams/breachesofregulations/

Assignment Feedback

Feedback will be via Brightspace VLE. This will consist of a mark for your assignment and a short comment on the assignment.

Marking Rubric

Achievement	Excellent	Satisfactory	Basic	Unsatisfactory
% of Marks Available	>75%	55-75%	40-55%	<40%

	Weighting				
Dataset, Problem Definition	10	Complex data sets selected for all tasks, well defined problem definition, justifications for selections	Mostly moderately to complex data sets, with corresponding problem definition, medium to well defined	Trivial, Simple and Moderate complex datasets chosen, Problem defined, missing details and impact	Trivial dataset chosen, poorly defined problem, etc
Data Insights & Data Preparation	20	Good focused insights from dataset select, good explanation, no trivial data analysis, selection of appropriate data preparation, explanations given,	Useful insights with explanations, impact of these on problem solution, selection of some appropriate data preparation, explanations given	Some insights given, limited details, limited data preparations, appropriateness of data prep, minimum explanations given	No or poorly selected data insights, limited or no data preparation
Application of Algorithms	15	Suitable algorithms selected, good details on these and why, good details on experimentation, insights from experimentation, reflections, and discussion	Suitable algorithms selected, some detail of selection and why, some details of algorithm experimentation, some discussion of experimentation	Suitable algorithms selected, limited details of selections given, limited details of application of algorithms given, limited details of algorithm settings and tuning	Limited or no details of selection and application of algorithms for data and problem. No explanations
Analysis of Results	20	Excellent detailed analysis of results and excellent insights of these. Clearly demonstrates impact and outcomes	Good detailed analysis of results, good level of insights on these, what then mean, their impact and outcomes	Some discussion of results, at a basic level with little insights	Little, no or very limited analysis of results and outcomes from tasks
Learning for work	15	Excellent level of insights, brings together details through work, 4-8 citations comparing related work in each task, clear identification of improvements, reflection on learning outcomes from tasks	Good level of discussion of results, identify some areas for improvement, 3-5 citations used to compare results in each task, identification of some improvements with limited discussion	Some discussion and evaluation of work, some comparison with related research, limited number of citations used	Little or no discussion or work, no comparison with related research
Ethics & Legal	20	Excellent level of discussion and insights for both case studies. Clear well defined ethical and legal issues. Clear well defined impact on role and tasks	Good level of discussion of ethical and legal issues for both case studies. Clearly reflection on these and the roles and tasks	Some discussion of ethical and/or legal aspects. Limited discussion with little or no reflection, and impact on roles and tasks	Little or no discussion, simple overviews given, Little or no ethical and legal aspects considered