Assessment Submission Cover Sheet

This Assessment Cover Sheet **must** be included on all Assessment submissions.

|  |  |
| --- | --- |
| Assignment Title | CA2 – Association Rules |
| Module | Data Mining |
| Student Name  (same as Student Card) | Ciaran Finnegan |
| Student Number | D21124026 |
| Programme |  |
| Part-Time/Full-Time |  |
| Year of Study  (First Year, Second Year, etc) |  |

Late Submissions: Assessment submitted after the deadline will have a late penalty applied.

**Academic Integrity for assessment in TU Dublin Programmes**

Each student is responsible for knowing and abiding by TU Dublin Academic Regulations and Policies. Any student in breach of these regulation/policies will be subject to action in accordance with the University’s procedures for breaches of assessment regulations. Please refer to the General Assessment Regulations at

<https://tudublin.libguides.com/c.php?g=674049&p=4794713>

<https://www.tudublinsu.ie/advice/exams/breachesofregulations/>

All students are expected to complete their courses/programmes in compliance with University regulations. No student shall engage in any activity that involves attempting to receive a grade by means other than honest effort, for example:

1. No student shall complete, in part or in total, any examination or assessment for another person.
2. No student shall knowingly allow any examination or assessment to be completed, in part or in total, for themselves by another person.
3. No student shall plagiarise or copy the work of another and submit it as their own work.
4. No student shall falsify any data. Falsification is the invention of data, its alteration, its copying from any other source, or otherwise obtaining it by unfair means, or inventing quotations and/or references.
5. No student shall use aids or devices excluded by the lecturer in undertaking course work or assessments/ examinations.
6. No student shall knowingly procure, provide, or accept any materials that contain questions or answers to any examination or assessment to be given at a subsequent time.
7. No student shall provide their assignments, in part or in total, to any other student in current or future classes of this module/ programme unless authorised to do so by the lecturer.
8. No student shall submit substantially the same material in more than one module/programme without prior authorization.
9. No student shall alter graded assignments or examinations and then resubmit them for regrading, unless specifically authorised to do so by the lecturer.
10. All programming code and documentation, unless correctly referenced, submitted for assessment or existing in the student’s computer accounts must be the students’ original work or material specifically authorized by the lecturer.
11. Collaborating with other students to develop, complete or correct course work is limited to activities explicitly authorized by the lecturer.
12. For all group assignments, each member of the group is responsible for the academic integrity of the entire submission. Consequently, all group members must satisfy themselves that all elements of their submission adhere to the academic integrity statement points above.

By submitting coursework, either physically or electronically, you are confirming that it is your own work (or, in the case of a group submission, that it is the result of joint work undertaken by members of the group that you represent) and that you have read and understand the University’s Regulations and Policies covering Academic Integrity (see General Assessment Regulations)*.*

Coursework may be submitted to an electronic detection system in order to help ascertain if any plagiarised material is present. If you have queries about what constitutes plagiarism, please speak to your lecturer.

|  |  |
| --- | --- |
| Student Signature |  |
| Date |  |

IMPORTANT:

* Complete the required number of tasks as defined in Assessment Handout
* The sections listed below are an example of the section headings for each task. You can use alternative headings
* Tasks 1-3: Sub-Sections 1-7 should be no longer than 8 pages (minimum 6 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.
* Marks are awarded based on complexity of problem and depth of work.

# TASK 2 - *Association Rules – Grocery Market Basket Analysis*

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. **Comparison with Other Research**

*Enhance the decision-making process around business rules*

A paper published by Faron and Chakraborty at the SAS Global Forum in 2012**[1]** suggested a way to add the Pearson’s Chi Squared statistic to the output from the SAS EM *Association* node.

The addition of this test would enhance the quality of the data analysis by showing which rules are statistically significant, therefore adding an additional metric the evaluate the importance of given association rules.

**Did you do it?**

*Representing Rules: Visual Clutter*

The graph output in SAS EM is a useful at a glance tool to visually represent the associations rules in our grocery database. However, research on similar grocery datasets, such as that by Hahsler, M., Hornik, K. and Reutterer, T., 2016**[2]**, highlights the challenge that these graphs easily become cluttered as the number of roles grow. (This would be true of any rule set, not just the typical grocery basket analysis being conducted in this section of the assignment).

If a visual representation of the association rules is an important outcome, then there are tools such as ***arulesViz*** that offer more sophisticated interactive functionality. For the purposes of this assignment the SAS EM cluster tools were deemed sufficient.

*What to do with these Association Rules?*

This dataset is a relatively straightforward representation of grocery purchases, and the common benefit is often considered to be a more effective physical retail shop layout.

A key point made in a 2018 Towards Data Science article**[3]** is that Association Rules look at lists of items with unique transaction ID from many users, and studies these lists as a block. This is not an approach that generates a recommendation for one *specific* user.

That said, research I found on the role of Association Data Mining and E-Commerce website structure**[4]** shows how the selection of *antecedent* lists can be used to meaningfully direct a user to different pages, where they are most likely to find the products for which they are looking.

A follow-on challenge for me would be to repeat this exercise with a similar but extended dataset that looked at product groupings in the *itemsets*, and also considered user profiles.

1. **References**

[1] Faron, M. and Chakraborty, G., 2012. *Easily Add Significance Testing to your Market Basket Analysis in SAS® Enterprise Miner*. [online] Citeseerx.ist.psu.edu. Available at: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.421.1785&rep=rep1&type=pdf> [Accessed 5 January 2022].

[2] Hahsler, M., Hornik, K. and Reutterer, T., 2016. Implications of Probabilistic Data Modeling for Mining Association Rules. *From Data and Information Analysis to Knowledge Engineering*, [online] pp.598-605. Available at: <https://link.springer.com/content/pdf/10.1007/s11573-016-0822-8.pdf> [Accessed 5 January 2022].

[3] Garg, A., 2018. *Complete guide to Association Rules (1/2)*. [online] TowardsDataScience. Available at: <https://towardsdatascience.com/association-rules-2-aa9a77241654> [Accessed 31 December 2021].

[4] Omari, A., Conrad, S. and Alcic, S., 2007. Designing a Well-Structured E-Shop Using Association Rule Mining. *2007 Innovations in Information Technologies (IIT)*, [online] Available at: <https://ieeexplore.ieee.org/abstract/document/4430429> [Accessed 5 January 2022].