600092 Data Mining and Decision Systems - CRG
ACW - Data Mining of Legacy Data

Learning Outcome	Criterion	Pass	2:2	2:1	1st	Upper 1st
[LO1] Demonstrate a critical knowledge and understanding of data warehousing, mining and reclamation, data mining systems and expert, decision support systems.	Outline of how the methodology was applied to the process of Legacy Data Mining with evidence supported by artefact. (20%)	An outline of the methodology followed is presented, but may be shallow in depth, or contain some factual errors. No evidence supporting methodology phases is presented.	A complete outline of the methodology followed is presented, with some sections evidenced back to code.	In addition to previous. Each phase of the methodology followed is supported by empirical evidence taken from code. Methodology phases are contextualised to the problem domain.	In addition to previous. Methodology is well outlined, with sections broken down into sub-sections with strong evidencing of process. Contextualisation of the process as applied to the problem domain is well-presented.	In addition to previous Deployment and Business Understanding phases are presented as a hypothetical scenario and contextualised to the given domain. Alternative and/or additional methodologies are evidenced, and fully justified, alongside their interactions with the current methodology.
[LO2] Critically analyse, research and report on the concepts of data, information and knowledge within a decision support system.	Presentation of results tables, and any accompanying figures. (10%)	A basic results section is presented. This may cover basic data relationships or metrics.	In addition to previous. Single model classification accuracy is presented.	In addition to previous. Multiple models are evaluated, with TP, TN, FP, FN metrics. Additional figures and/or tables may be present.	In addition to previous. Additional fit-for-purpose metrics are reported, with results tables enabling comparison of different input sets for each model evaluated. Additional figures are presented appropriately to highlight any additional data patterns.	In addition to previous. Results presented utilise a variety of fit-for-purpose metrics, and contain interesting entries to invoke in-depth discussion and comparison.

[LO4] Select, justify and use appropriate approaches, including some at the forefront of the subject / profession, to identity the impact that data mining and decision systems have on an organisation.	Critical evaluation of results with discussion surrounding the process.

(30%) | Results are interpreted correctly in the context of the assignment. | In addition to previous.

Results are interpreted correctly, with some discussion relating back to the domain. Rationale may be provided but not fully justified. | In addition to previous.

Rationale is provided for chosen metrics, with correct justification, relating these metrics back to the domain.

Comparisons between models, and their results, is provided but may not be fully justified. | In addition to previous.

Results are thoroughly evaluated, referencing several metrics, and comparing various data preprocessing approaches towards the objective.

Discussion critically reflects on the application of the methodology towards the assignment. | In addition to previous.

Discussion is thoughtful and critically reflects on aspects of the methodology highlighting advantages and disadvantages of certain stages.

Alternative Methodologies and practices are considered and contextualised to the problem task. |
| [LO3] Develop an appropriate decision support tool knowledge using a rigorous data mining methodology for a complex information scenario. | Demonstrate the use of Python Jupyter Notebooks in a Data Mining Context towards Mining of Legacy Data

(40 %) | Basic data cleaning, and transformation is achieved, but may be erroneous or incomplete in places.

A single model is created, but may be incorrectly utilised. | In addition to previous.

All Data errors are repaired, with data cleaned. Transformations may not be most appropriate.

A single model is created and trained on the processed data. | In addition to previous.

Multiple models are trained on the processed data.

Evaluation metrics are generated by predicting on a trained model. This is done utilising basic data partitioning.

Notebook is laid out logically following the phases of the methodology followed. | In addition to previous.

Markdown cells are used to better differentiate sections and introduce code segments.

Advanced Data partitioning is performed with data-driven metrics considered for the partition.

Python code is well-documented, with reference to documentation where appropriate.

Multiple methods of data preprocessing for model creation are implemented, alongside advanced model training techniques. | In addition to previous.

Data is processed efficiently and effectively, showing a clear understanding of library use.

Model hyper-parameters are explored and expertly determined from data-driven processes. |
| Weighting | All criteria are weighted as shown by the percentages indicated in the relevant criterion box. | | | | | |