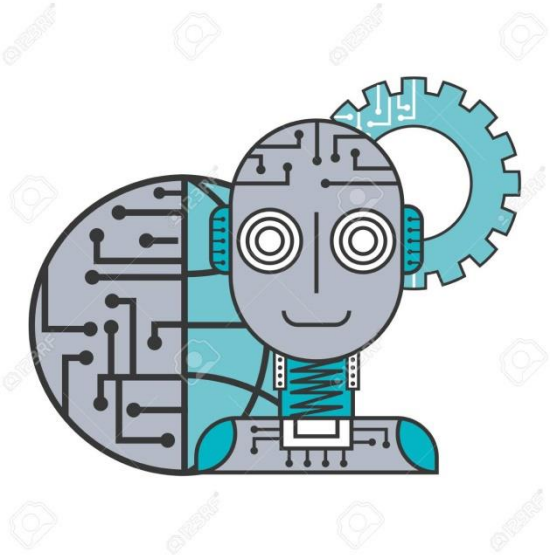
Data Mining and Foundation of AI

6CC555

Assessment 1



Dr. Aaisha Makkar

Contents

Module Leader

Dr. Aaisha Makkar
Email: a.makkar@derby.ac.uk
Office Hours: Wednesday 10:00 pm to 11:30 pm (Pre-booking via email is required)
Room: MS312

Key dates and details

Assessment Type:	**Data Mining and Decision Tree Algorithms**
Assessment weighting:	50%
Page Count	4 pages (Font: Times New Roman, Font Size: 11)
Learning Outcomes:	1
Submission Method:	Blackboard online submission point
Date Set:	1pm UK time, 04/02/2025 in a synchronous of the class
Submission Date:	12:00 noon UK time, 24/03/2025
Provisional Feedback Release Date:	12:00 noon UK time, 14/04/2025

Description of the assessment

You will be working on one of the following datasets related to:

(1) Patient appointment request dataset: Check below link for details:
Medical Appointment Scheduling System

(2) Diabetes, see dataset on Kaggle
(https://www.kaggle.com/datasets/nanditapore/healthcare-diabetes).

Tasks include:

1. Select one dataset from the above listed datasets and define the problem to be solved using the selected dataset.

2. Implementing pre-processing techniques (handle categorical data, missing data and noisy data) and perform EDA (Exploratory Data Analysis) for the dataset.

3. Perform further analysis, either Regression or Classification depending upon the problem selected.

4. Local Jupiter Notebook, Cloud Kaggle or Google Colab can be used for implementation (We will cover the different tools in lectures and tutorial sessions).

5. Your final submission should contain two file; one file of ipython notebook and the other file (docx/pdf) with well written comments regarding your selecting dataset, analysis problems, EDA process, models' selection, evaluation, comparing discussions and conclusion.

Assessment Content

You can download the dataset yourself from Kaggle.

The docx/pdf file structure include:

1. Cover page (Only student number and your course name)
2. Indicating CSV file that you selected and your analysis of problem definition (Introduction)
3. EDA codes and explanations and discoveries related to your analysis problem
4. Algorithm codes with explanations
5. Evaluations of your work and possible comparing with different solutions
6. Conclusion comments.

To complete this assessment, you need to have the skill of basic python programming that we practised in our practice sessions and knowledge of first 6-week lectures.

Assessment Rubric

The assessment rubric on the next page shows the complete criteria of the CW and how you will be assessed. We will explore the content of the rubric together in a synchronous session on the first-week lab session.

When the assessment is returned you will receive a digital version of the rubric showing how you performed against each criterion. There will be an opportunity to discuss these with your personal academic tutor.

Criteria	5-39%	40-49%	50-59%	60-69%	70-79%	80-89%	90-100%
Introduction (10)	No introduction (0) Introduction is not too relevant to the requirement (20-39)	An introduction provided with required contents but not in details	An introduction provided with required contents with some degree of details	A good introduction provided with required contents with good degree of details.	A great introduction provided with required contents with great degree of details. The defined analysis problem is a research question in the research field (90-100)		
EDA and pre-processing (40%)	Incorrect submission. No evidence of EDA or pre-processing	EDA provided but not using correct or understandable methods or the presentations of the EDA are not understandable No justification on pre-pre-processing	EDA provided with reasonable methods but with some inconsistent or wrong results or without clear explanations on the EDA results No justification on pre-pre-processing	EDA provided with reasonable methods that produce consistent and explainable results but not explained in-depth in the document. Justification on pre-pre-processing is made	EDA provided with reasonable methods that produce consistent and explainable results and explained in-depth in the document. Justification on pre-pre-processing is made	EDA provided with reasonable and many different types of methods that produce consistent and explainable results and explained in-depth in the document. Justification on pre-pre-processing is made with understandable arguments.	EDA provided with reasonable and many different types of methods that produce consistent and explainable results and explained in-depth in the document. Justification on pre-pre-processing is made with understandable arguments. The solution ideas have been discussed based on EDA
Regression or classification analysis and solutions (30%)	No evaluation evidence in the document	Only one solution has been evaluated and no critical analysis	More than one solution have been evaluated but no critical analysis and no preference conclusion	More than one solution has been evaluated with basic critical analysis and preference conclusion	More than one solution has been evaluated with well-structured critical analysis and convinced preference conclusion	More than three solutions have been evaluated with well-structured critical analysis and convinced preference conclusion	More than three solutions have been evaluated with well-structured critical analysis and some improvements or new thought on the solutions.
Documentation 20%	The document is almost none.	Little effort is put on the documentation and not following the required structure.	The documentation is simply comments on the EDA results or evaluations and not all required sections are provided	The document is structured well to cover all required sections, but some contents are not explained well.	The document is structured well to cover all required sections with good explanation and arguments but very basic	The document is structured well to cover all required sections with good explanation and arguments in-depth	The document is structured in style to cover all required sections with good explanation and arguments in-depth. Moreover, the whole report shows good self-learning and

							independent thinking skills
% Mark 1-4%	**Nothing of Merit:** Nothing of value is contained in the submitted work. The work presents information that is irrelevant and unconnected to the task; no evident awareness of appropriate principles, theories, evidence or techniques						
NS	**Non-submission:** No work has been submitted.						

Anonymous Marking

You must submit your work using your **student number** to identify yourself, not your name. You must not use your name in the text of the work at any point. When you submit your work in Turnitin you must submit your student number within the assignment document and in the *Submission title* field in Turnitin.

Assessment Regulations

The University's regulations, policies and procedures for students define the framework within which teaching and assessment are conducted. Please make sure you are familiar with these regulations, policies and procedures.