Scoil na Faisnéisíochta & na nEalaíon Cruthaitheach
School of Informatics & Creative Arts

Department of Visual and Human Centred Computing

Assignment Brief

Name of module	Machine Learning		
Module code	COMPC8130		
Lecturer(s)	Sargam Yadav		
No. of credits for module	10		
Breakdown of module	**CA - 100%**	**Exam**	**Total 100%**
Percentage of overall module CA for this assignment	40% - Cross Module with Data Mining		
Name of assignment	Machine Learning Project		
Indicate if the assignment is recoverable or non-recoverable	Recoverable		
Due date	*21st December 2025 TBC*		
Assignment description	See description below.		
Extent of AI use narrative[1] (This will indicate if AI use is unrestricted, restricted, prohibited by the lecturer(s))	**AI use is only allowed to be used for the Data Acquisition stage of the assessment. If it is used at this stage then this should be explained in a markdown block at the start of the Machine Learning Jupyter Notebook. AI is not permitted used for any other elements of the assessment.**		
Format of submission (if relevant) Deliverables (What must be included in submission)	See description below		
Mode of submission	Moodle submission		

[1] Consult the Generative Artificial Intelligence Guidance for Staff for narrative to be included here.

(If submission is in hardcopy a student must also provide evidence via Moodle (photo/video) of the submission by the due date for submission)	
Special requirements as per ACS (if relevant). **Indicate if reattendance of the module is required if the overall module is failed.**	
Stages of delivery of proposals / draft versions / subcomponents with staggered delivery (where applicable) **Percentages of marks for items delivered in this manner (where applicable)**	See description below
Marking criteria (details on rubric / criteria)	See description below
Method of feedback to be provided to students	Marks for each section through Moodle.

Notes to Students

This information on this assignment brief may represent only one component of the overall module. Please ensure that you contact the other lecturers teaching on this module for the information on assessments relating to the other components of the module. It is your responsibility to ensure that you complete all assessments for every module.

- Please note that all work submitted as partial fulfilment of this module must adhere to the Institute Academic Integrity Policy. The Generative Artificial Intelligence (AI) and Your Assessments A Guide for Students provides useful support information for students.

- If you have been permitted to use AI in the preparation of your assignment you should include a Declaration of Use acknowledging that you have used generative artificial intelligence in the creation of material for the assessment. Material that has not been adapted/modified should be referenced using existing reference styles. The Declaration of Use[2] should be included in an appendix in the assessment submitted by the learner and should:
    - Provide a written acknowledgment of the use of generative artificial intelligence.
    - Specify which technology/technologies were used.
    - Identify the prompts used.
    - Provide the resulting outputs.
    - Explain how the output was used in the submitted work (used directly or modified).

- If generative artificial intelligence (AI) is permitted in the assessment and the learner has chosen not to use it, the following disclosure in the learner submission is recommended: "*No content generated by AI technologies has been used in this assessment*".
- All work submitted must be accompanied by a Continuous Assessment Cover Sheet.
- We recommend all students consult the DkIT Guide to Harvard Referencing which provides useful information on referencing and how to avoid plagiarism. https://www.dkit.ie/documents/dkit-referencing-guidelines-and-how-avoid-plagiarism
- All work must be submitted by the due date indicated by your lecturer(s). Any work submitted after the due date is subject to the penalties outlined in the Institute Continuous Assessment Procedures
- The terms recoverable and non-recoverable are detailed in the Institute Continuous Assessment Procedures.

[2] Consult The Generative Artificial Intelligence (AI) and Your Assessments A Guide for Students for example of wording to be used for Declaration of Use statement.

Machine Learning– Project

PROJECT (40%) – Machine Learning PROJECT 2025– VERSION 1 (PROVISIONAL)

Summary

For this assessment, you are required to apply machine learning and deep learning techniques to the dataset that you previously collected and prepared during your Data Mining project. The aim of this project is to develop, evaluate, and interpret predictive and analytical models that provide meaningful insights into a real-world industry problem.

The dataset used should contain approximately 10 variables (a mixture of numerical and categorical features) and include a clearly defined outcome (target) variable suitable for prediction. The data should also be structured in a way that supports the application of clustering and data reduction methods alongside supervised learning. You are expected to explore, select, and implement appropriate machine learning models (for example, regression, classification, or clustering algorithms), evaluate their performance using suitable metrics, and interpret the results in the context of the identified business or research problem.

Your work should be documented clearly in a Jupyter Notebook, showing the complete workflow — from feature preparation and model training to validation, evaluation, and interpretation of results. Visualisations should be used to support your discussion of model performance and findings. This continuous assessment contributes 40% towards the final module grade. The project is due for submission by December 21st, 2025.

A first-stage submission is required by October 31th, 2025, where you must confirm the dataset you intend to use (from your Data Mining project) and briefly describe the predictive problem to be addressed.

An individual interview or presentation may be scheduled after submission to discuss your work, modelling choices, and conclusions.

Cross-module Project

Note that the data you collect in this project will also be used in your Data Minings module. The exact specification for that element of the work will be provided by the lecturer for that module.

Description

In this project, you are required to apply machine learning techniques to a real-world dataset that you have previously sourced and prepared during your Data Mining project. The objective is to use appropriate supervised and unsupervised learning methods to address a specific prediction or classification problem, as well as to extract additional insights through clustering and data reduction. The dataset should ideally include at least 10 columns containing a mix of numerical and categorical variables, and a minimum of 10,000 rows to allow for meaningful model development and evaluation.

There must be a clearly defined outcome (response) variable, representing the target for prediction or classification. The data should also be suitable for the application of clustering and data reduction methods.

Your final deliverable will be a Jupyter Notebook that documents all stages of your analysis — from problem definition to model evaluation and interpretation — using appropriate statistical and visual techniques.

Step 1 – Business Understanding: Clearly identify and define the problem(s) that your dataset can help solve. Explain the business or industry relevance of these problems, and why a machine learning solution would be valuable. You should describe both the predictive task (for example, predicting customer churn, product demand, or disease outcome) and any secondary analytical goals such as segmentation or dimensionality reduction. The business problem may be expressed as a set of objectives or research questions. You may also include a brief discussion of how predictive modelling, clustering, or data reduction could be used to support decision-making in the selected domain.

Step 2 – Data Preparation: Load the dataset that you sourced in your earlier project. Ensure that the data is complete, clean, and ready for analysis. You may include minor additional cleaning or integration steps if needed to make the dataset suitable for machine learning. Clearly document all preprocessing steps and save the cleaned dataset as a CSV file for reproducibility. All activities and decisions in this stage should be clearly documented using Markdown cells.

Step 3 – Feature Engineering: Identify the outcome (target) variable and specify the input (predictor) features to be used in the modelling process. Perform any necessary feature engineering, such as:

- Encoding categorical variables
- Scaling or normalising numerical features
- Creating new derived variables where appropriate
- Detecting and handling multicollinearity
- Include a correlation matrix or other feature-selection visualisation to justify your chosen features.

Step 4 – Model Development: Apply suitable machine learning algorithms to your dataset, selecting methods that fit the type of prediction problem (e.g., regression, classification, or clustering).

Possible models include (but are not limited to): Linear or logistic regression, Decision trees and random forests, Support Vector Machines (SVMs), k-Nearest Neighbours (kNN), Clustering algorithms (e.g., K-Means, Hierarchical Clustering), Neural networks (Recurrent Neural Networks, Convolutional Neural Networks).

You are encouraged to apply multiple models and compare their performance. All code and parameter choices should be accompanied by concise explanations.

Step 5 – Model Evaluation: Assess model performance using appropriate metrics, depending on the problem type. Examples include:

- Classification: accuracy, precision, recall, F1-score, ROC/AUC
- Regression: RMSE, MAE, R^2

You should also include relevant visualisations, such as confusion matrices, ROC curves, feature importance plots, or cluster visualisations. Discuss the strengths and limitations of each model and identify the one that performs best in practical terms.

Step 6 – Data Reduction Analysis: Apply data reduction techniques such as Principal Component Analysis (PCA) or Factor Analysis to simplify your dataset and uncover hidden structure. Explain the proportion of variance captured by the principal components and interpret their meaning where possible. Compare how dimensionality reduction affects the results of your machine learning models (e.g., performance, interpretability).

Step 7 – Findings and Discussion: Summarise your overall findings and highlight the key insights from your analysis.

Your discussion should:

- Clearly link back to the business or research problem defined in Step 1
- Present key visualisations and model results
- Explain the practical implications of your findings
- Reflect on model limitations and possible improvements

Data Agreement Stage

It is very important that you identify and mine the data you intend to use as soon as possible. With that in mind you must have sourced your data by October 31st. You should have met with the Data Mining and the Machine Learning lecturers before then and agreed the data set with them.

Final Deliverables

You are required to submit the following **three** files only through Moodle:

1. Data File

A single csv data file named ***dataProject.csv*** which contains the data used for analysis in the project. This will be the same as the output from you data mining project.

2. Model Training Jupyter Notebook

This notebook should include the structured analysis of your work, following the steps outlined in the Machine Learning project brief. You may include subsections as appropriate, but the core sections should follow the defined project stages (e.g., Business Understanding, Data Preparation, Feature Engineering, Model Development, Model Evaluation, Data Reduction, Findings).

Formatting Guidelines:

- Each major step must be labelled in a Markdown block using # headers (e.g., # Step 1 – Business Understanding).
- Subsections should use ### formatting.

- All interpretations and explanations must appear in Markdown cells — not as comments within code.
- Code comments should only describe the function or purpose of the code itself.
- All visualisations (graphs, tables, performance metrics) must be generated as outputs from code cells.

3. HTML Output Document:

Filename: DataAnalysis.html

This HTML file should be an exported version of your main analysis notebook, showing all outputs, visualisations, and Markdown commentary. It will serve as a readable summary report of your project and must include:

- All text, results, and graphical outputs
- Clear explanations and interpretations of your findings
- A well-structured, visually coherent layout

Indicative Marking Scheme (Out of 100%) – this may be tailored to each project.

1. Business Understanding: 5%
2. Data Preparation:7.5%
3. Exploration of data: 10%
4. Feature Engineering: 10%
5. Model Development: 25%
6. Model Evaluation: 25%
7. Data Reduction Analysis: 7.5%
8. Overall document quality: 5%
9. Ongoing Progress: 5%

Interview

Note that you must attend and pass an interview after submitting the report. Failure to pass the interview will result in a grade of 0.

Rubric

Grade (%)	Description
80+	Project demonstrates mastery of subject matter with novel/original work applied to a complex problem. Analysis of data and model development is accurate, complete, thoroughly explained, and appropriate with clear link to aims. Analysis shows original thinking and implementation beyond what was learned in course. All documentation is well-structured and very well-written. Student has presented with full knowledge of both the problem and the solution including appropriate critical analysis.

70-79	Project demonstrates thorough understanding of subject matter. Analysis of data and model development is accurate, complete, thoroughly explained, and appropriate with clear link to aims. Functionality is well developed with only minor issues. Document is well-structured, with complete version history. Student has presented with excellent knowledge of both the problem and the solution including appropriate critical analysis.
60-69	Project demonstrates good understanding of subject matter. Analysis of data and model development is accurate, substantially complete, well explained, and appropriate with good link to aims. May be missing some appropriate analysis. Document is reasonably structured, substantially complete and well delivered. Student has presented with reasonably good knowledge of problem and solution but lacks some critical analysis and in depth understanding.
50-59	Project demonstrates reasonable understanding of subject matter. Data analysis and model development is accurate, partially complete, reasonably well explained, and appropriate with some link to aims. May be missing some appropriate analysis and links to aims and objectives may not be complete. Document is adequately structured, mostly complete and reasonably well delivered. Student has presented problem and solution satisfactorily, but with some issues relating to problem and solution. Lacks critical understanding.
40-49	Project demonstrates partial understanding of subject matter. Analysis and model training is relatively simple allowing very basic exploration and model development. Basic but incomplete analysis based entirely on functionality provided in class with significant omissions. Poor link to aims and objectives and weak understanding of logic and findings of the model. Document missing some key elements. Student has presented poorly. While presentation has some relevant structure, it lacks completeness and coherence.
30-39	Project demonstrates little understanding of subject matter. Little evidence of exploration of problem and model development. Incomplete analysis based entirely on functionality provided in class with weak links to aims. Documentation poor and missing significant elements. Presentation is very confusing and lacks clarity and focus. No critical analysis.
0-29	Project demonstrates almost no understanding of subject matter. No serious attempt to address the problem.

Additional Notes

You are encouraged to explore applications and data sets which may be significantly different to the ones we investigated in the course. In this case it may be necessary to adjust the marking scheme to reflect the project application chosen.

Plagiarism

PLEASE PAY SPECIAL ATTENTION TO THE ISSUE OF PLAGIARISM. The DkIT policies are available at https://www.dkit.ie/system/files/academic_integrity_policy_and_procedures.pdf

In summary, all work submitted by learners for assessment purposes, or for written or oral publication, must be their own work. Where this is informed by the work of others, the source must be properly referenced using the accepted norms and formats of the appropriate academic discipline.

Late Submission

The policy for late submission will apply and is available at the link below. Any legitimate late submission must be accompanied by explanation and supporting documentation as per the policy.

https://www.dkit.ie/system/files/continuous_assessment_procedures_document_v4.pdf