

SCHOOL OF ARCHITECTURE, COMPUTING, & ENGINEERING

Submission instructions

- Cover sheet to be attached to the front of the assignment when submitted
- All pages to be numbered sequentially
- All work has to be presented in a ready to submit state upon arrival at the ACE Helpdesk. Assignment cover sheets or stationery will **NOT** be provided by Helpdesk staff

Module code	CN6005
Module title	Artificial Intelligence
Module leader	Dr Nadeem Qazi
Assignment tutors	Dr Nadeem Qazi/Dr Mahmud Ahmad
Assignment title	AI coursework
Assignment number	
Weighting	50%
Handout date	3/11/2025
Submission date	20/12/2025
Learning outcomes assessed by this assignment (see course handbook)	Learning Outcomes: 1 – 9
Turnitin submission requirement	yes
Additional information	<ul style="list-style-type: none"> • ASSESSMENT FEEDBACK - Feedback on your assessment will be available in four working weeks from the submission date. Please refer to the module pages on UEL+ for assessment specific details.

Form of assessment:

☐

Individual work

☒

Group work



For **group work** assessment which requires members to submit both individual and group work aspects for the assignment, the work should be submitted as:

- ☒ Consolidated single document ☐ Separately by each member

Number of assignment copies required:

- ☒ 1 ☐ 2 ☐ Other

Assignment to be presented in the following format:

- ☒ On-line submission
☐ Stapled once in the top left-hand corner
☐ Glue bound
☐ Spiral bound
☐ Placed in a A4 ring bound folder (not lever arch)

Note: To students submitting work on A3/A2 boards, work has to be contained in suitable protective case to ensure any damage to work is avoided.

Soft copy:

- ☐ CD (to be attached to the work in an envelope or purpose made wallet adhered to the rear)
☐ USB (to be attached to the work in an envelope or purpose made wallet adhered to the rear)
☒ Soft copy not required

Note to all students

Assignment cover sheets can be downloaded from UEL Plus via the following pathway.

Home Page → **ACE Information** → **ACE Helpdesk** → **Assignment Front Sheets**

All work has to be presented in a ready-to-submit state upon arrival at the ACE Helpdesk. Assignment cover sheets or stationery (including staplers) will **NOT** be provided by Helpdesk staff. This will mean students will not be able to staple cover sheets at the Helpdesk.

Group Coursework:

You must work in a Group of a maximum of 5 students. Stickly, No Individual Work. You will be provided a Peer Review Assessment Form to evaluate each member of the group's working participation. Your final project marks will be calculated based on the peer review Assessment Form.

CourseWork Title :Classification and Clustering Analysis on Real-World Datasets

This coursework is divided into **two parts**.

Coursework Overview

This coursework is divided into **two distinct parts**, each focusing on a different machine learning paradigm and dataset type. Students are expected to demonstrate practical implementation, critical analysis, and explainability in both tasks.

Part 1: Multiclass Classification Using Deep Learning and Machine Learning

Students will perform **multiclass classification** on an **image dataset** using the following models:

- Two deep learning models: **Artificial Neural Network (ANN)** and **Convolutional Neural Network (CNN)**
- One traditional machine learning algorithm of their choice: **Logistic Regression**, **Naïve Bayes**, or **Decision Tree**

Requirements:

- Preprocess the image dataset appropriately (e.g., normalization, resizing, encoding labels)
 - Train and evaluate each model using suitable metrics (e.g., accuracy, precision, recall, F1-score)
 - Perform a **critical analysis** comparing the performance of the models
 - Include **Explainability** techniques (e.g., Grad-CAM, LIME, or SHAP) to interpret model predictions
 - Provide well-commented code and visualizations to support findings
-

Part 2: Regression or Clustering Analysis Using Numerical Data

This part focuses on applying either **Linear Regression** or **KMeans Clustering** on a **numerical dataset**. Students must use a **different dataset** from Part 1.

Option A: Linear Regression

- Conduct **Exploratory Data Analysis (EDA)** to understand feature distributions and relationships
- Implement **Linear Regression** and evaluate using metrics such as:
 - **Mean Absolute Error (MAE)**
 - **Mean Squared Error (MSE)**
 - **R-squared (R²)**
- Apply **SHAP** for model explainability to interpret feature contributions
- Discuss model performance and limitations

Option B: Clustering with KMeans

- Perform **EDA** to explore data structure and feature characteristics
- Apply **KMeans Clustering** and determine optimal number of clusters
- Evaluate clustering quality using **Silhouette Score**
- Compare clustering results and provide insights into cluster characteristics

Suggested Data set for Part 1

FashionMNIST <https://www.kaggle.com/datasets/zalando-research/fashionmnist>

Or any other image dataset .

Suggested Data set for Part 2

1. Air Quality Data in the U.S. (EPA): <https://www.kaggle.com/datasets/thedevastator/epa-air-quality-data>
2. Iris Data set <https://archive.ics.uci.edu/dataset/53/iris>
3. Wine Quality Dataset <https://archive.ics.uci.edu/dataset/186/wine+quality>
4. Breast Cancer Dataset <https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic>

Part One: Multiclass Classification with Image Dataset (35 Marks)

Objective

To implement multiclass classification using ANN, CNN, and logistic regression, evaluate model accuracy, and critically analyze the results.

Dataset Requirements

- **Dataset:** Use a publicly available image dataset suitable for multiclass classification, Do not use s CIFAR-10, MNIST as these are already used in Labs.
- **Size:** Minimum of 500 images across multiple classes.

Tasks

1. **Data Pre-Processing (5 Marks)**
 - Pre-process the image dataset by resizing, normalizing, and applying any other required transformations.
 - Visualize sample images and explore class distributions.
2. **Model Implementation and Training (Total 20 Marks)**
 - Implement the following models for multiclass classification:
 - One Machine Learning Model (5 marks)
 - Deep Learning Model Artificial Neural Network (ANN) using Keras (5 marks)
 - Convolutional Neural Network (CNN) using Keras (10 marks)
 - Document hyperparameters and architecture choices for each model.
3. **Model Evaluation (5Marks)**
 - Evaluate each model using accuracy, confusion matrix, and multiclass metrics (e.g., precision, recall, F1-score).
 - Compare model performances and highlight strengths and weaknesses.
4. **Critical Analysis of Results (5 Marks)**
 - Provide a thorough analysis comparing the performance of logistic regression, ANN, and CNN.

- Discuss the impact of model complexity on accuracy, computational efficiency, and model suitability for image-based multiclass classification.

Part Two : Clustering Analysis (15 Marks)

Objective

To apply KMeans or KNN clustering techniques, visualize the dataset, and evaluate the clustering performance.

Dataset Requirements

- **Dataset:** Select a publicly available dataset suitable for clustering tasks, with a mix of categorical and numerical features.
- **Size:** Minimum of 500 instances with at least 8 columns.
- **Example Datasets:** UCI Machine Learning Repository, Kaggle datasets, etc.

Tasks

1. **Exploratory Data Analysis (EDA) (5 Marks)**
 - Analyze each feature with summary statistics and visualizations (e.g., histograms, scatter plots).
 - Discuss trends, distributions, and any interesting observations.
2. **Clustering Analysis (10 Marks)**
 - Implement KMeans or KNN clustering on the dataset.
 - Use silhouette score to evaluate and compare the performance of both clustering algorithms.

Provide a comparative analysis of KMeans or KNN clustering and insights from the clustering performance.

OR

Linear Regression Analysis Task 15 Marks

Objective

To apply **Linear Regression** techniques, visualize the dataset, and evaluate the regression model's performance.

Dataset Requirements

- **Dataset:** Select a publicly available dataset suitable for regression tasks, ideally with a continuous target variable and a mix of categorical and numerical features.
 - **Size:** Minimum of **500 instances** with at least **8 columns**.
 - **Example Sources:** UCI Machine Learning Repository, Kaggle datasets, OpenML, etc.
-

Tasks

1. Exploratory Data Analysis (EDA) – 5 Marks

- Analyze each feature using **summary statistics** and **visualizations** (e.g., histograms, box plots, scatter plots).
- Identify **correlations** between features and the target variable.
- Discuss **trends, distributions**, and any **interesting observations** or potential data quality issues (e.g., missing values, outliers).

2. Regression Analysis – 10 Marks

- Implement **Linear Regression** on the selected dataset.
- Evaluate model performance using metrics such as:
 - **Mean Absolute Error (MAE)**
 - **Mean Squared Error (MSE)**
 - **R-squared (R^2)**
- Visualize:
 - **Regression line** (for simple linear regression)
 - **Residual plots**
 - **Predicted vs Actual values**

3. Comparative Insights

- If applicable, compare **Linear Regression** with other regression models (e.g., Ridge, Lasso, Decision Tree Regressor).
- Discuss:
 - Which model performed better and why
 - The impact of feature selection or transformation
 - Any limitations or assumptions of linear regression observed in the dataset

Report Format 3000 Words only

R

Structure

1. **Introduction(200 words)** : Brief overview of the coursework objectives and chosen datasets.
2. **Part One (1500words) - Multiclass Classification:**
 - Data pre-processing steps and sample visualizations
 - Models architecture, implementation, and evaluation
 - Critical analysis of model performance and recommendation on the best-suited model
 - Mention challenges and techniques to overcome Overfitting.
- i. **Machine Learning algorithm (400 words)**
 - a. **Implementation Steps:**
 - i. **Data Preparation:** Describe how you prepared the dataset, including feature selection and target variable definition.

- ii. **Model Building:** Explain how you built the model using a programming language (e.g., Python with Scikit-learn).
 - iii. **Training Process:** Detail the training process, including splitting the dataset into training and testing sets and fitting the model.
 - iv. **Code Snippets:** Include relevant code snippets for clarity.
 - ii. **Artificial Neural Networks (ANN) (400 words)**
 - a. **Implementation Steps:**
 - i. **Tools and Libraries:** Mention the tools and libraries used (e.g., TensorFlow, Keras).
 - ii. **Data Preprocessing:** Explain any preprocessing steps taken, such as normalization or encoding.
 - iii. **Network Architecture:** Describe the architecture of the ANN, including the number of layers, neurons, and activation functions.
 - iv. **Training Process:** Explain the training process, including epochs, batch size, and any optimization techniques used.
 - v. **Code Snippets:** Provide relevant code snippets to illustrate the implementation.
 - iii. **Convolutional Neural Networks (CNN) (400 words)**
 - a. **Implementation Steps:**
 - i. **Framework:** Specify the framework used to build the CNN (e.g., TensorFlow, PyTorch).
 - ii. **Data Pipeline:** Describe how you managed the data pipeline, including any data augmentation techniques.
 - iii. **CNN Architecture:** Outline the architecture of the CNN, detailing the layers involved (convolutional, pooling, dropout).
 - iv. **Training Process:** Explain how you trained the model, including learning rate adjustments and optimizer choices.
 - v. **Code Snippets:** Include relevant code snippets for clarity.
 - iv. **Comparison of Algorithms (300 words)**
 - a. Description of accuracy measurement metrics (Confusion Matrix, etc.)
 - b. Summary of findings from the implemented models
 - c. Insights gained from the analysis
3. **Part Two - Clustering Analysis(600words:**
- o EDA and insights
 - o Clustering implementation and results
 - o Comparative analysis of KMeans OR KNN using silhouette scores

Exploratory Data Analysis and Pre-processing (300 words)

- Importance of data exploration
- Description of pre-processing techniques used

Algorithms implementation(300)

- **Data Preparation:** Describe how you prepared the dataset for clustering.
- **Model Building:** Explain how you implemented the K-Means algorithm using a programming language (e.g., Python with Scikit-learn).
- **Training Process:** Detail how you determined the optimal number of clusters (e.g., using the elbow method) and how you trained the model.
- **Code Snippets:** Include relevant code snippets for clarity.
- **Interpretation of clustering model**
- **Accuracy of the Model**

Regression Analysis

- **Correlation Analysis of the chosen dataset**
 - **Explain the model and accuracy measures**
 - **Explain how SHAP interprets the model**
- **Conclusion (200 words)**
 - Recap of the main points discussed
 - Future directions or implications of the work

Submission Requirements

- **Jupyter Notebook:** Code for data processing, clustering, and classification models with clear comments.
- **Report:** A structured report in PDF format, following the outlined format.

AUTHOR GUIDELINES FOR PROCEEDINGS MANUSCRIPTS IN
16 POINT TIMES NEW ROMAN, FULLY CAPITALISED AND
CENTRED AND ONE BLANK LINE AFTER THE TITLE

Author(s) Name(s) in 14 point times New Roman & Centred

Author Affiliation(s), Italic in 12 point times New Roman & Centred

*E-mail Italic in 12 point **times** New Roman, Centred and give one blank line before starting
the abstract*

Abstract: Type abstract in, 11 point times New Roman, single-spaced type with zero spacing before
and after and the word abstract in bold.

All manuscripts must be in English.

All text after Abstract must be in a two-column format.

Give two blank lines before starting the introduction

1. Formatting your page:

Top & Bottom Margins: 2.5cm

Left & Right Margins: 2.5cm

All text after Abstract must be in a two-column format, single spaced in 12 point times New Roman.

Please do not place any additional blank lines between paragraphs.

Columns are to be 7.6 cm wide, with a 0.8cm space between them. Text must be fully justified.

2. First-order headings:

For example, "**1. Introduction**", should be 14 Times New Roman boldface, initially capitalised, flush left, with one blank line before, and one blank line after. Use a period (".") after the heading number, not a colon.

2.1. Second-order headings:

As in this heading, they should be 12 Point Times New Roman boldface, initially capitalised, flush left, with one blank line before, and one after.

2.1.1. Third-order headings. Third-order headings, as in this paragraph, are discouraged. However, if you must use them, use 12 Points Times New Roman boldface, boldface, initially capitalised, flush left, preceded by one blank line, followed by a period and your text on the same line.

3. Page numbering and Footnotes:

Do not exceed eight pages; including graphs, illustrations, references etc.

No page numbering and Do not use any footnotes.

4. Illustrations, Figures, photographs and tables:

All should have captions below and centred 11 Points Times New Roman within TWO columns at the top or bottom of the page with NO Bold face or Italics

5. References:

List all bibliographical references alphabetically in 12 point Times New Roman, single-spaced and one blank line after each reference at the end of your paper. When referenced in the text, enclose the citation like for example, (Smith, 2004).

Smith S., Smith A., Roberts A., "Article Title", *Journal*, Publisher, Location, Date, pp. 1-10.

Smith S., Smith A., Roberts A., *Book Title*, Publisher, Location, Date.