# B8IS119: Big Data: Achieving Scale

## Module Code:

## Continuous Assessment One(CA 1)

| | |
|---|---|
| **Assessment Title & Description:** | **Big Data Analysis using Python Libraries** |
| **Lecturer** | Nitya Govindaraju |
| **Individual/Group:** | Individual |
| **Task:** | Utilize Python libraries for data pre-processing and machine learning modelling on a big dataset |
| **MIMLOs being assessed:** | 1,2,3,4,6 |
| **Issue Date** | 29 November 2024 |
| **Submission Date** | 05 Jan 2025 |
| **Assessment Weighting** | 20% |

Overview:

In this assessment, you will utilize Python libraries such NumPy, Pandas, Matplotlib, seaborn, Sci-kit Learn, Kera etc., powerful framework for big data processing, to perform machine learning modelling on a large dataset. You will be required to apply machine learning algorithms and techniques to gain insights and make predictions from the data.

Task:

Your task is to:

1. Load and Preprocess the given big data set using Python libraries.
2. Apply appropriate machine learning algorithms to the preprocessed data for analysis.
3. Evaluate the performance of the machine learning models using suitable metrics.
4. Interpret the results and provide insights based on the analysis.

Task requirements:

1. Preprocessing: Cleanse and transform the data as necessary for machine learning analysis.
2. Modeling: Implement at least two different machine learning algorithms (e.g., classification, regression, clustering) using Scikit Learn, Keras etc.
3. Evaluation: Assess the performance of each model using appropriate evaluation metrics (e.g., accuracy, F1-score, RMSE).
4. Interpretation: Provide insights and conclusions based on the analysis results.

File to be submitted: Two file (1 PDF + Zip file containing the dataset and python notebook)

1. Python code files (. ipynb or .py) containing the scripts used for preprocessing, modeling, and evaluation.
2. A report documenting the analysis process, including explanations of the preprocessing steps, the choice of machine learning algorithms, model evaluation metrics, and interpretation of results. (1500+/- 10% words)

## Assessment Rubric:

| Grade Criteria | Fail | Pass | Good | Very Good | Excellent |
|---|---|---|---|---|---|
| **Data Pre-processing** | Incomplete or incorrect pre-processing steps, resulting in data inconsistencies or errors. | Basic pre-processing steps applied but with some errors or inefficiencies. | Correct pre-processing steps applied, ensuring data quality and consistency. | Thorough pre-processing applied, effectively addressing data cleaning and transformation requirements. | Advanced pre-processing techniques applied, optimizing data quality and usability for analysis. |
| **Modelling** | correct implementation of machine learning algorithms, or models do not perform adequately. | Basic implementation of machine learning algorithms with some errors or limitations in performance. | Correct implementation of machine learning algorithms, with models demonstrating satisfactory performance. | Effective implementation of machine learning algorithms, with models performing well and meeting performance criteria. | Advanced implementation of machine learning algorithms, with models achieving high performance and accuracy. |
| **Evaluation** | Inadequate or incorrect evaluation metrics used, or evaluation process is incomplete. | Basic evaluation metrics used, but interpretation of results is limited or unclear. | Appropriate evaluation metrics used, with clear interpretation of results and model performance. | Comprehensive evaluation metrics used, providing thorough assessment of model performance and insights. | Advanced evaluation techniques used, with in-depth analysis of model performance and implications. |
| **Interpretation** | Limited or unclear insights provided based on analysis results. | Basic insights provided, but lacking depth or clarity in interpretation. | Clear and insightful interpretation of analysis results, demonstrating understanding of implications. | Comprehensive interpretation of analysis results, offering valuable insights and recommendations. | Advanced interpretation of analysis results, with profound insights and actionable recommendations for stakeholders. |

## a) What is referencing and why is it necessary?

Please follow this link to the Harvard Style Referencing Guide - all referencing is required in this forma: https://libguides.dbs.ie/skills/referencing