



# Adventist University of Central Africa

P.O. Box 2461 Kigali, Rwanda | [www.auca.ac.rw](http://www.auca.ac.rw) | [info@auca.ac.rw](mailto:info@auca.ac.rw)

Faculty of Information Technology

Department: MSc in Big data analytics

## Midterm Exam project

Academic year: 2024-2025, Semester 2

**Course Code: MSDA9223**   **Name:** Data mining and Information retrieval

**Instructor:** Dr. Pacifique Nizeyimana

**Date:** 25 May 2024

**Exam Duration:** 3hrs

**Total max:** /100 Marks

---

### Instructions

1. You are requested to use only **python** or **R** in this project
2. Submission due date: **4th July 2025, 5 PM CAT**
3. Students **are not allowed** to work on the same project ( Same dataset with the same objectives)
4. The report should be between 12 and 18 pages

---

### Task Overview

You are required to use at least 3 different algorithms to solve real-world problems using a real dataset(choose one in EAC) and compare their performance. That problem would be in any field of your interest, such as education, health, agriculture, finance, transportation, economics, and banking, among others.

### Task details

- 1) Find the dataset (Choose one in EAC) that will allow you to perform two tasks Regression or Classification, and clustering. If one dataset does not permit these two tasks, select another dataset for a different task. Note that this can also be text dataset
- 2) Prepare data for analysis(Data cleaning)
- 3) Perform Exploratory data Analysis (EDA)
- 4) Determine the objectives depending on your dataset
- 5) Based on objective chosen use at least four different learning algorithm including deep learning and
- 6) Evaluate these models performance and compare them using performance metric results

- 7) Apply one method of improving model performance
- 8) Compare (6) and (5) and confirm if model performance has improved

### **Reporting\Submission**

- I. At the end of the project you will submit a draft of paper(only use latex ) that contains the following sections:
- II. Abstract (maximum 250 words)
- III. Introduction
- IV. Methods (including description of the dataset)
- V. Results and interpretation
- VI. Discussion & Recommendations
- VII. References

**Note: Attachment of Code(Rmarkdown file(.Rmd) or python file(.ipynb file)) is mandatory**