UGANDA CHRISTIAN UNIVERSITY
A Centre of Excellence in the Heart of Africa

FACULTY OF ENGINEERING, DESIGN AND TECHNOLOGY
DEPARTMENT OF COMPUTING AND TECHNOLOGY
ADVENT 2025 SEMESTER **EXAMINATION**

PROGRAM: *[BACHELOR OF SCIENCE IN COMPUTER SCIENCE]*

YEAR: 3 SEMESTER: 1

COURSE CODE: *[DSC3108]*

COURSE NAME: *[BIG DATA MINING AND ANALYTICS]*

EXAMINATION TYPE: *PROJECT-BASED EXAM*

PROJECT DURATION: DECEMBER 2025

TIME ALLOWED: *[14 DAYS]*

Examination Instructions

1. The general Uganda Christian University examination guidelines and academic & financial policies apply to this examination. Violating any of the policies by the student automatically makes this examination attempt void, even if you have completed and submitted the answer booklet.
2. This exam consists of a project to be executed in [14] days.
   i. There are three parts to this project-based exam; Part A carries 30 Marks, Part B carries 40 Marks and Part C carries 30 Marks.
   ii. Assessment of the project shall be based on three milestones.
   iii. To achieve all the milestones, students should submit their work as links to github or kaggle jupyter notebooks under the students' personal accounts.
   iv. A comprehensive final report (PDF) accompanied by Presentation Slides should also be submitted.
3. Every student has a responsibility to prove their contribution towards every milestone, and marks will be awarded to every student individually.

Project-based assessment guidelines (This project counts for 30% of your final grade)

S/N	Milestone Description	Maximum Marks
1	**MILESTONE ONE** [Project design and implementation]	**30 %**
2	**MILESTONE TWO** [Model development and analysis]	**40 %**
3	**MILESTONE THREE** [Project report, Ethics, and Presentation]	**30 %**
	TOTAL MARKS	**100 %**

PROJECT SCENARIOS: DATA SOURCING

Students must select **ONE** of the following scenarios and justify how it qualifies as a **Big Data problem:**

1. **Social Media Trend Analysis:** Analyze a high-volume stream of real-time social media data (simulated or archived public datasets) to identify emerging trends or sentiment shifts (e.g., product launch, political event).
2. **Large-Scale Retail Recommendation:** Process historical transaction data (thousands of records) for a large e-commerce platform to develop a product recommendation system.
3. **IoT Sensor Anomaly Detection:** Analyze time-series data from a massive network of industrial IoT sensors to detect and predict equipment failure.

PART A: Big Data Platform Setup and Data Preprocessing [TOTAL 30 MARKS]

1. **Justify Big Data:** Provide a compelling argument (maximum 50 words) on why your chosen scenario's dataset necessitates a **Big Data platform** over a traditional relational database.

2. **Tool Selection:** Select and set up a suitable **Big Data Tool/Platform** using a lightweight distributor (e.g. Google Collab or Kaggle for Spark)

3. **Data Acquisition:** Source or simulate an appropriate dataset for the chosen project scenario.
4. **Distributed Processing:** Ingest the raw data and perform **data cleaning and transformation**.

PART B: Data Modelling and Analytics [TOTAL 40 MARKS]

1. Technique Selection: Select and justify one or more core Data Mining Techniques suited to the dataset selected.
2. Model Scalability: Implement the chosen algorithm using the selected Big Data platform. The model must be designed for scalability.
3. Model Execution and Optimisation: Run the model on the full dataset and document the execution time. Propose and implement suitable optimization techniques and show the impact on model performance.
4. Result Interpretation: Analyse the model's output.

PART C: Final Report and Ethics [TOTAL 30 MARKS]

Submit a single PDF report (2-3 pages, excluding code) covering your findings in Sections A and B and include the following aspects:

1. Propose a concrete **business application** for the analytical results. Explain how the insights can drive value for the organization in the chosen scenario.
2. Dedicate a section to discuss the **ethical implications** and **privacy concerns** of your solution.

Submit presentation slides: 8-10 slides summarizing the entire project for a final defence.

~END OF PROJECT EXAM~