



"Investing in Africa's Future"

COLLEGE OF ENGINEERING AND APPLIED SCIENCES (CEAS)

NCSE 412: DATA MINING AND BUSINESS INTELLIGENCE

END OF SECOND SEMESTER EXAMINATIONS

APRIL 2025

LECTURER: MR BRAITON U. MUKHALELA

TIME: 3 HOURS

INSTRUCTIONS

- All tasks must be completed using **Orange**. Datasets are to be provided during examination.
- Submit a final report with screenshots and explanations of each step.
- Ensure that all steps, from data preprocessing to model evaluation, are documented.
- Interpretation of results is crucial for full marks.

Section A: Data Preprocessing

(20 Marks)

1. Load the dataset into **Orange**.
2. Perform the following preprocessing steps:
 - Handle missing values using appropriate techniques.
 - Normalize numerical features.
 - Select the most relevant features using a feature selection method.
 - Provide a **summary report** on data cleaning decisions.

Section B: Exploratory Data Analysis (EDA) & Visualization

(15 Marks)

1. Generate the following visualizations:
 - A **scatterplot** showing relationships between two numerical variables.
 - A **histogram** of a key feature.
 - A **boxplot** comparing a categorical variable's distribution.
2. Provide insights from each visualization.

Section C: Classification Task

(20 Marks)

1. Train a classification model using **Decision Tree, k-NN, and Naïve Bayes**.
2. Compare the accuracy of these models using cross-validation.
3. Choose the best model and justify your selection.
4. Provide a **confusion matrix** and **ROC curve** for the chosen model.

Section D: Clustering Task

(15 Marks)

1. Apply **k-Means clustering** on the dataset.
2. Determine the optimal number of clusters using the **Elbow Method**.
3. Visualize the clusters using a scatterplot.
4. Interpret the clusters and their business implications.

Section E: Association Rule Mining

(15 Marks)

1. Use the **Apriori algorithm** to generate association rules.
2. Set a minimum support of 5% and a minimum confidence of 50%.
3. Identify the top 3 rules and explain their significance in a business context. (use a word file and save it as Section E3 in your folder)

Section F: Business Intelligence Dashboard

(15 Marks)

1. Create a **dashboard in Orange** to summarize key insights.
2. Include at least three different types of charts/graphs.
3. Provide a brief **report on business insights** derived from the dashboard.

END OF EXAMINATION