United States International University-Africa

Education to take you places

SCHOOL OF SCIENCE AND TECHNOLOGY

DEPARTMENT OF DATA SCIENCE AND ANALYTICS

SPRING 2024 – QUIZ 3

COURSE CODE: STA 3020A

UNIT NAME: MULTIVARIATE ANALYSIS

DATE: 14TH FEBRUARY 2024

TOTAL MARKS: 20 MARKS

INSTRUCTIONS:

For this exercise:

1. ANSWER ALL QUESTIONS
2. Do all your working in the RMarkdown (.rmd).
3. Submissions should be in a **`.rmd` file**
4. NO SUBMISSIONS SHOULD BE DONE VIA EMAIL

QUESTIONS:

For this task, use the `german.credit` data set available in the fairml package.

This discriminant analysis task aims to create a predictive model that assists in assessing the credit risk of consumers based on their financial attributes, providing valuable insights for decision-making in the lending process. Given the German Credit Data set, the task is to perform a discriminant analysis to develop a model that predicts credit risk (Credit_risk) based on various features in the dataset. Specifically, the objective is to assess the discriminative power of the available variables in distinguishing between "BAD" and "GOOD" credit risks.

1. Explore and understand the distribution of the target variable, "Credit_risk."
2. Identify and select relevant predictor variables that may contribute to the discrimination between "BAD" and "GOOD" credit risks.
3. Split the dataset into training and testing sets to assess the model's performance on unseen data.
4. Apply linear discriminant analysis (LDA) or quadratic discriminant analysis (QDA) to build a model predicting "Credit_risk" based on the selected variables.
5. Evaluate the discriminant model's performance using appropriate metrics such as accuracy, precision, recall, and F1 score. Visualize the discriminant functions and decision boundaries to interpret how the model separates the two classes.
6. Analyze the importance of each variable in predicting credit risk by examining the discriminant loadings.
7. Validate the discriminant model using cross-validation techniques to ensure robustness.
8. Interpret the results and provide insights into which variables are significant in determining credit risk in the German market.