

Credit Scoring and Credit Risk Measurement Assignment

Assignment 2: DSA 8304 Risk Management Analytics

Master of Science in Data Science and Analytics

Strathmore University

Due Date: Midnight, 7 days to End of Semester Exams

Objective:

The goal of this group assignment is to expose students to the practical and theoretical aspects of credit scoring and credit risk measurement. The assignment will require candidates to write a technical report, perform data analytics, and apply machine learning techniques to build a credit scoring model. The assignment consists of two parts: a technical report and a practical data analytics project.

Part I: Technical Report (50 Marks)

Requirements:

1. A comprehensive literature review of the following:
 - Credit scoring models (e.g., Logistic Regression, Decision Trees, Random Forests, Gradient Boosting, Neural Networks, Support Vector Machines, Nonlinear regression, structural models, econometric models etc). Discuss at the different methodologies or models used in credit scoring and credit risk measurement. For each method, explain how it works, its strengths, and its limitations.
 - Credit risk measurement techniques (e.g., Value at Risk (VaR), Expected Shortfall (ES), Conditional Value at Risk (CVaR), Probability of Default (PD), Loss Given Default (LGD), and Exposure at Default (EAD) etc)
 - Discuss the strengths and weaknesses of the various models and techniques you review.

- **Regulatory and Ethical Considerations:** Discuss the regulatory environment related to credit scoring and credit risk measurement, focusing on any guidelines or rules set by financial authorities. Then, discuss ethical considerations such as fairness and discrimination in credit scoring.
2. A technical report summarizing your findings from the literature review, including an overview of the evolution of credit scoring and credit risk measurement techniques over time.

Part II: Practical Data Analytics (50 Marks)

Requirement:

1. Use a publicly available dataset (include references to this) relevant to credit scoring or create a synthetic dataset. Clean and preprocess the data as needed for analysis.
2. Conduct an exploratory data analysis. Identify and describe the characteristics of the dataset, such as the distribution of various features, the presence of outliers, and any patterns or correlations and relationships that might be evident.
3. Develop a credit scoring model using a machine learning algorithms of your choice. You could use logistic regression, decision trees, random forests, classification, support vector machines or any other suitable method. Split your data into a training set and a testing set, and use the training set to train your model. You are required to compare different techniques.
4. Validate your model using the testing set. Calculate and report metrics such as accuracy and precision. Also, perform a sensitivity analysis to understand how changes in the input variables affect the outcome.
5. Apply your credit scoring model to assess the credit risk of a new, unseen set of customers. Discuss the results and the implications.

Deliverables:

1. A technical report detailing the principles of credit scoring and credit risk measurement, various methodologies, and the regulatory and ethical considerations. The report should be between 3000-5000 words.
2. A Code notebook (or equivalent) containing the code for data preprocessing, exploratory data analysis, credit scoring model development, and model validation. Include comments and markdown cells to explain your code and findings.
3. A brief report (1000-1500 words) summarizing your data analytics project, explaining your methodology, and discussing your findings.