



Strathmore
UNIVERSITY

STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES
BACHELOR OF BUSINESS SCIENCE IN FINANCIAL ENGINEERING
END OF SEMESTER EXAMINATION
BSF 3216: APPLIED ANALYTICS IN FINANCE

DATE: 19TH DECEMBER 2024

Time: 2 Hours

Instructions

1. This examination consists of **FIVE** questions.
2. Answer **Question ONE (COMPULSORY)** and any other **TWO** questions.

Question 1 (30 Marks)

1. Differentiate between the following terms as applied in analytics for financial applications (10 Marks)
 - i) Statistical Learning vs Machine Learning
 - ii) Supervised Learning
 - iii) Semi-Supervised Learning
 - iv) Reinforcement Learning
 - v) Feature Engineering
2. In financial datasets, missing data can be problematic. Which imputation methods would be most appropriate for missing values in a time-series dataset (e.g., stock prices), and how would you implement them (6 marks)
3. In the finance industry, alternative data (e.g., social media sentiment) is increasingly used for predictions. Describe one method for transforming unstructured alternative data into features suitable for a machine learning model (4 marks)
4. Write an R program that creates a vector of 5 random numbers between 1 and 100. Then, calculate and print:
 - The sum of these numbers
 - The average of these numbers
 - The maximum and minimum values (6 marks)
5. Draw an algorithmic flowchart for a program in R that determines whether a given number is positive, negative, or zero (4 marks).

Question 2 (20 Marks)

1. Discuss the tools available for visualization and how they can enhance understanding of complex financial data, particularly in real-time environments (5 marks)
2. Explain and differentiate the following machine learning models as applied for a binary classification problem. Provide appraisal for the foundational working of the models, benefits and challenges of each model in the classification task
 - a) Artificial Neural Network (ANN) (3 marks)
 - b) Logistic Regression (3 marks)
 - c) K-Nearest Neighbors (KNN) (3 marks)
 - d) Support Vector Machine (SVM) (3 marks)
 - e) Decision Trees (3 marks)

Question 3 (20 Marks)

1. Given a vector of stock prices: `prices <- c(100.5, 101.2, 99.8, NA, 102.4, 103.1, 101.8)`
Calculate the daily returns, handling any NA values appropriately (3 marks)
2. Explain the concept of "data cleaning" in the context of data analysis. Discuss three common issues that data cleaning addresses and how R can be used to resolve these issues (4 marks).
3. Write R code to compare the summary statistical features of 5 stocks for a rolling window holding period of 1 month, 3 months, 6 months and 1 year (5 marks)
4. Discuss how descriptive, predictive, and prescriptive analytics can be applied in financial markets to improve decision-making processes. Include examples of tools that facilitate these types of analysis. (8 marks)

Question 4 (20 Marks)

A financial institution that has collected a dataset for credit scoring with the following variables:

- Age (continuous)
- Years with current employer (continuous)
- Years at current address (continuous)
- Income (continuous)
- Credit card debt (continuous)
- Auto loan debt (continuous)
- Homeowner (binary: yes or no)
- Default (binary: yes or no)

- i) Utilizing logistic regression, describe the process to develop a credit scoring model based on this data. Discuss the necessary steps for data preprocessing, feature selection, model training, and model evaluation. (4 marks)
- ii) Suppose the trained logistic regression model assigns negative coefficients to 'Age' and 'Years with current employer', and positive coefficients to 'Credit card debt' and 'Auto loan debt'. Interpret these results and explain what they mean in the context of credit scoring. (4 marks)
- iii) Analyze the 2 limitations of different machine learning algorithms (such as decision trees, neural networks, and support vector machines) in predicting credit default risk (4 marks)
- iv) Write R code that will simulate a geometric Brownian motion process with 100 paths, risk free rate of 10%, and volatility of 20% and apply it to pricing of a European call option with a strike price of 10 (8 marks).

Question 5 (20 Marks)

1. Explain the use of the following model validation techniques. (5 Marks)
 - i) F1 Score
 - ii) ROC-AUC
 - iii) Precision
 - iv) Recall
 - v) Type I vs Type II Errors
2. Identify and explain three emerging technologies in artificial intelligence in Finance (6 Marks).
3. In financial context, computer programming is useful in building information systems that will provide data analytics solutions for the organization. Describe the data analytics lifecycle (9 marks)