

MA6633: STATISTICAL MODELLING FOR DATA MINING

Effective Term

Semester B 2024/25

Part I Course Overview

Course Title

Statistical Modelling for Data Mining

Subject Code

MA - Mathematics

Course Number

6633

Academic Unit

Mathematics (MA)

College/School

College of Science (SI)

Course Duration

One Semester

Credit Units

3

Level

P5, P6 - Postgraduate Degree

Medium of Instruction

English

Medium of Assessment

English

Prerequisites

MA5617 Statistical Data Analysis

Precursors

Nil

Equivalent Courses

Nil

Exclusive Courses

Nil

Part II Course Details

Abstract

This course will aim to provide the student with mathematical foundation and statistical knowledge of data mining and machine learning techniques. Particular focus will be on the fundamental statistical properties and analysis of many popular techniques for learning, classification and prediction. The topics covered in this course will include elements of the following: Bayesian decision theory, model selection, linear Models for regression and classification, Bayesian networks, decision trees, association rule mining, and clustering.

Course Intended Learning Outcomes (CILOs)

CILOs		Weighting (if app.)	DEC-A1	DEC-A2	DEC-A3
1	Explain the basic problems and mathematical challenges in data mining	20	x	x	
2	Ability to use various algorithms for supervised and unsupervised learning and understand their underlying principles	50	x	x	
3	Ability to choose a suitable combination of different models to explore data sets and solve real problems	30	x	x	x

A1: Attitude

Develop an attitude of discovery/innovation/creativity, as demonstrated by students possessing a strong sense of curiosity, asking questions actively, challenging assumptions or engaging in inquiry together with teachers.

A2: Ability

Develop the ability/skill needed to discover/innovate/create, as demonstrated by students possessing critical thinking skills to assess ideas, acquiring research skills, synthesizing knowledge across disciplines or applying academic knowledge to real-life problems.

A3: Accomplishments

Demonstrate accomplishment of discovery/innovation/creativity through producing /constructing creative works/new artefacts, effective solutions to real-life problems or new processes.

Learning and Teaching Activities (LTAs)

LTAs	Brief Description	CILO No.	Hours/week (if applicable)	
1	teaching	Students will gain knowledge and participate in discussions on data mining techniques	1, 2, 3	3 hours/week

Assessment Tasks / Activities (ATs)

ATs	CILO No.	Weighting (%)	Remarks (e.g. Parameter for GenAI use)	
1	Lab assignments	1, 2, 3	15	
2	Mid-term quiz	1, 2, 3	25	

Continuous Assessment (%)

40

Examination (%)

60

Examination Duration (Hours)

2

Additional Information for ATs

For a student to pass the course, at least 30% of the maximum mark for the examination must be obtained.

Assessment Rubrics (AR)

Assessment Task

1. Lab assignments (for students admitted before Semester A 2022/23 and in Semester A 2024/25 & thereafter)

Criterion

Software usage and data analysis ability

Excellent

(A+, A, A-) Reliably utilizes statistical models and algorithms for data mining to tackle intricate problems using R

Good

(B+, B, B-) Sufficiently employs statistical models and algorithms for data mining to address moderately complex problems

Fair

(C+, C, C-) Utilizes statistical models and algorithms for data mining to address problems of intermediate complexity

Marginal

(D) Applies statistical models and algorithms for data mining with limited effectiveness to solve simple problems

Failure

(F) Incapable or inept at applying statistical models and algorithms for data mining to problem solving

Assessment Task

2. Mid-term quiz (for students admitted before Semester A 2022/23 and in Semester A 2024/25 & thereafter)

Criterion

Problem solving based on comprehensive understanding

Excellent

(A+, A, A-) Consistently demonstrates a thorough understanding of statistical models for data mining and can always apply them to explore data set and solve complex problems

Good

(B+, B, B-) Adequately demonstrates an understanding of statistical models for data mining and can usually apply them to explore data set and solve problems

Fair

(C+, C, C-) Demonstrates some understanding of statistical models for data mining and has some ability to explore data set and solve simple problems

Marginal

(D) Demonstrates limited understanding of statistical models for data mining and has limited ability to explore data set and solve simple problems

Failure

(F) Demonstrates little understanding of statistical models for data mining and can rarely or never apply them to explore data set and solve complex problems

Assessment Task

3. Examination (for students admitted before Semester A 2022/23 and in Semester A 2024/25 & thereafter)

Criterion

Problem solving based on comprehensive understanding

Excellent

(A+, A, A-) Consistently demonstrates a thorough understanding of statistical models for data mining and can always apply them to explore data set and solve complex problems

Good

(B+, B, B-) Adequately demonstrates an understanding of statistical models for data mining and can usually apply them to explore data set and solve problems

Fair

(C+, C, C-) Demonstrates some understanding of statistical models for data mining and has some ability to explore data set and solve simple problems

Marginal

(D) Demonstrates limited understanding of statistical models for data mining and has limited ability to explore data set and solve simple problems

Failure

(F) Demonstrates little understanding of statistical models for data mining and can rarely or never apply them to explore data set and solve complex problems

Assessment Task

1. Lab assignments (for students admitted from Semester A 2022/23 to Summer Term 2024)

Criterion

Software usage and data analysis ability

Excellent

(A+, A, A-) Reliably utilizes statistical models and algorithms for data mining to tackle intricate problems using R

Good

(B+, B) Sufficiently employs statistical models and algorithms for data mining to address moderately complex problems

Marginal

(B-, C+, C) Applies statistical models and algorithms for data mining with limited effectiveness to solve simple problems

Failure

(F) Incapable or inept at applying statistical models and algorithms for data mining to problem solving

Assessment Task

2. Mid-term quiz (for students admitted from Semester A 2022/23 to Summer Term 2024)

Criterion

Problem solving based on comprehensive understanding

Excellent

(A+, A, A-) Consistently demonstrates a thorough understanding of statistical models for data mining and can always apply them to explore data set and solve complex problems

Good

(B+, B) Adequately demonstrates an understanding of statistical models for data mining and can usually apply them to explore data set and solve problems

Marginal

(B-, C+, C) Demonstrates limited understanding of statistical models for data mining and has limited ability to explore data set and solve simple problems

Failure

(F) Demonstrates little understanding of statistical models for data mining and can rarely or never apply them to explore data set and solve complex problems

Assessment Task

3. Examination (for students admitted from Semester A 2022/23 to Summer Term 2024)

Criterion

Problem solving based on comprehensive understanding

Excellent

(A+, A, A-) Consistently demonstrates a thorough understanding of statistical models for data mining and can always apply them to explore data set and solve complex problems

Good

(B+, B) Adequately demonstrates an understanding of statistical models for data mining and can usually apply them to explore data set and solve problems

Marginal

(B-, C+, C) Demonstrates limited understanding of statistical models for data mining and has limited ability to explore data set and solve simple problems

Failure

(F) Demonstrates little understanding of statistical models for data mining and can rarely or never apply them to explore data set and solve complex problems

Part III Other Information

Keyword Syllabus

Bayesian decision theory; Linear and nonlinear models; Discriminant analysis; Clustering models; Naïve Bayes; Bayesian networks; Decision trees; Association rule.

Reading List

Compulsory Readings

Title	
1	Course materials provided

Additional Readings

Title	
1	Pattern classification 2nd edition by Duda, Hart and Stork, Wiley-Interscience; 2nd edition
2	Data Mining: Concepts, Models, Methods, and Algorithms by Mehmed Kantardzic, Wiley-IEEE Press; 2nd edition
3	Pattern Recognition and Machine Learning by Bishop, Springer, 2006