

MS6711: DATA MINING

Effective Term

Semester B 2024/25

Part I Course Overview

Course Title

Data Mining

Subject Code

MS - Decision Analytics and Operations

Course Number

6711

Academic Unit

Decision Analytics and Operations (DAOS)

College/School

College of Business (CB)

Course Duration

One Semester

Credit Units

3

Level

P5, P6 - Postgraduate Degree

Medium of Instruction

English

Medium of Assessment

English

Prerequisites

MS5217 Statistical Data Analysis

Precursors

Nil

Equivalent Courses

Nil

Exclusive Courses

FB6711 Data Mining

Part II Course Details

Abstract

This course introduces students to a range of popular and practical data mining and machine learning algorithms relevant to business applications. Students will perform data analysis using the Python programming language. Upon successful

completion of this course, students will have acquired core foundational knowledge in the field and be well-prepared for a wide variety of careers in data analytics.

Course Intended Learning Outcomes (CILOs)

CILOs		Weighting (if app.)	DEC-A1	DEC-A2	DEC-A3
1	Demonstrate knowledge of state-of-the-art machine learning and data mining algorithms.	30	x	x	x
2	Define and formulate real-world data mining problems, prepare data for data mining projects, execute algorithms, and interpret outputs.	30	x	x	x
3	Analyze real-world data using Python.	40	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	Lectures	Students will be guided through an exploration and discussion of data mining concepts and algorithms, along with instructions on assessing the quality of extracted information.	1, 2, 3

2	In-class activities	Students will observe Python demonstrations in class on designated subjects. They will be expected to engage in individual or group projects involving simulated or modest real-world datasets using the programming language. These in-class activities will enable the identification of common challenges faced by students, facilitating additional explanations where necessary. Moreover, students will have the opportunity to recognize their errors and acquire the skills to rectify them.	1, 2, 3	
3	Out-of-Class assignments	Students will engage with targeted challenges derived from substantial business datasets as assignments outside of class. They will have the option to work in small groups on these assignments, facilitating collaborative problem-solving discussions and solution development.	1, 2, 3	
4	Project	Students will receive a large dataset and a specified business scenario, tasked with extracting valuable insights aligned with predefined data mining objectives. Collaboration is encouraged through small group work, allowing students to collectively analyze data and operate software, seeking guidance from the instructor as needed throughout the semester.	1, 2, 3	

ATs		CILO No.	Weighting (%)	Remarks (e.g. Parameter for GenAI use)
1	Assignments	1, 2, 3	10	
2	Project	1, 2, 3	50	

Continuous Assessment (%)
60

Examination (%)
40

Examination Duration (Hours)
3

Assessment Rubrics (AR)

Assessment Task
Assignments (for students admitted before Semester A 2022/23 and in Semester A 2024/25 & thereafter)

Criterion
Core concepts and ideas; use of appropriate statistical methods

Excellent
(A+, A, A-) High

Good
(B+, B, B-) Significant

Fair
(C+, C, C-) Moderate

Marginal
(D) Basic

Failure
(F) Not even reaching marginal levels

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

Criterion
Core concepts and ideas; use of appropriate statistical methods

Excellent
(A+, A, A-) High

Good
(B+, B, B-) Significant

Fair
(C+, C, C-) Moderate

Marginal
(D) Basic

Failure

(F) Not even reaching marginal levels

Assessment Task

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

Criterion

Core concepts and ideas; use of appropriate statistical methods

Excellent

(A+, A, A-) High

Good

(B+, B, B-) Significant

Fair

(C+, C, C-) Moderate

Marginal

(D) Basic

Failure

(F) Not even reaching marginal levels

Assessment Task

Assignments (for students admitted from Semester A 2022/23 to Summer Term 2024)

Criterion

Core concepts and ideas; use of appropriate statistical methods

Excellent

(A+, A, A-) High

Good

(B+, B) Significant

Marginal

(B-, C+, C) Moderate

Failure

(F) Fail to understand the core concepts

Assessment Task

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

Criterion

Core concepts and ideas; use of appropriate statistical methods

Excellent

(A+, A, A-) High

Good

(B+, B) Significant

Marginal

(B-, C+, C) Moderate

Failure

(F) Fail to understand the core concepts

Assessment Task

Project (for students admitted from Semester A 2022/23 to Summer Term 2024)

Criterion

Core concepts and ideas; use of appropriate statistical methods

Excellent

(A+, A, A-) High

Good

(B+, B) Significant

Marginal

(B-, C+, C) Moderate

Failure

(F) Fail to understand the core concepts

Part III Other Information

Keyword Syllabus

1. Introduction to data mining

- Explain the background of data mining and machine learning.
- Provide a brief introduction to the history, current developments, and future directions of these fields.

2. Data Mining Procedure

- Demonstrate the steps of data mining, including problem definition, data preparation and cleaning, model execution, and result interpretation.
- Show examples using Python.

3. Data Mining Algorithms

- Cover various algorithms, including clustering (K-means and nearest neighbor), linear regression, logistic regression, decision trees, tree ensembles (random forest, boosting), neural networks, and Bayesian approaches such as Markov chain Monte Carlo.
- Understand the fundamental problem of bias-variance trade-off and the importance of cross-validation in machine learning.
- Briefly introduce recent advanced deep learning models, including large language models.

4. Ethics in Data Mining

- Discuss potential ethical issues in data mining applications, and possible solutions.

5. Case Studies

- Present applications using real data.

6. Data Mining Software

- Focus on using Python.

Reading List

Compulsory Readings

Title	
1	Müller, Andreas C., and Sarah Guido. Introduction to machine learning with Python: a guide for data scientists. " O'Reilly Media, Inc.", 2016.

Additional Readings

Title	
1	Han, Jiawei, Kamber Micheline and Pei, Jian. Data Mining: Concepts and Techniques 3rd edition. Elsevier, 2011. (This is a great overview book of data mining, from basic concept, data cleaning to algorithms and database. It covers many problems you might face in the real world.)
2	An Introduction to Statistical Learning with Applications in R, Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani. (This book provides a great introduction to machine learning algorithms. Concepts are well explained without too much technical details.)
3	AIQ: How People and Machines Are Smarter Together. Nicholas Polson, James Scott. St. Martin' s Press 2018. (This book contains many interesting stories about how machine learning and data mining influence the business world. It is not a technical book.)
4	Introduction to Data Mining, Tan P N, Steinbach M, Kumar V. Pearson Education. 2016 (This is another great book about data mining. It focuses more on algorithms.)