CS3481: FUNDAMENTALS OF DATA SCIENCE

Effective Term
Semester A 2024/25

Part I Course Overview

Course Title
Fundamentals of Data Science

Subject Code
CS - Computer Science
Course Number
3481

Academic Unit
Computer Science (CS)

College/School
College of Computing (CC)

Course Duration
One Semester

Credit Units
3

Level
B1, B2, B3, B4 - Bachelor's Degree

Medium of Instruction
English

Medium of Assessment
English

Prerequisites
CS2204 Fundamentals of Internet Applications Development

Precursors
Nil

Equivalent Courses
Nil

Exclusive Courses
CS4483 Data Warehousing and Data Mining

Part II Course Details

Abstract
This course aims to explore the important field of data science. The syllabus covers the main techniques in statistical data modelling, and algorithms in data science, which include predictive modelling, cluster analysis, association rule mining and

text mining. In addition, different applications of data science techniques in the real world such as web mining, business analytics and health informatics will be discussed.

Course Intended Learning Outcomes (CILOs)

	CILOs	Weighting (if app.)	DEC-A1	DEC-A2	DEC-A3
1	Identify the main characteristics of different techniques in data science through observation of their operations			x	
2	Perform a critical assessment of current techniques in data science.		x	x	
3	Design the main algorithms in data science in a computationally efficient way.			x	
4	Propose new solutions for real world information analytics problems by improving and combining current data science techniques.				

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	Lecture	Students will engage with the fundamental principles and state-of-the-art techniques in data science.	1, 2, 3, 4	3 hours/week
2	Tutorial	Students will work on a set of take-home exercises on the principles and applications of data science, and introduce their solutions in the class.	1, 2	8 hours/semester

3	Assignments/Projects	Students will apply the principles of data science to design computationally efficient algorithms for information analytics systems in these assignments/projects.	3, 4	After class

Assessment Tasks / Activities (ATs)

	ATs	CILO No.	Weighting (%)	Remarks (e.g. Parameter for GenAI use)
1	Assignments/Projects	3, 4	30	
2	Mid-term Examination	1, 2	20	

Continuous Assessment (%)

50

Examination (%)

50

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

Assignments/Projects

Criterion

1.1 Capacity for effectively implementing data science algorithms in a computationally efficient way.

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/Projects

Criterion

1.2 Capability to create new solutions for real world information analytics problems by improving and combining different data science techniques.

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
Mid-term Examination

Criterion
2.1 Ability to explain in detail the principles of different data science techniques.

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
Mid-term Examination

Criterion
2.2 Capability to correctly apply a suitable data science technique to solve an information analytics problem

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

Criterion
3.1 Capacity for understanding the main characteristics of different data science techniques in depth.

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

Criterion
3.2 Capability to perform a critical assessment of current data science techniques.

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

Criterion
3.3 Ability to integrate different data science techniques for addressing real world information analytics problems.

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

Part III Other Information

Keyword Syllabus
Data pre-processing, statistical data modelling, predictive modelling, classifier evaluation, cluster analysis, association rule mining, text mining.

Syllabus

- Knowledge discovery process
  Introduction of the knowledge discovery process in three stages: data pre-processing, data mining, and knowledge representation. Basic data pre-processing techniques including data cleaning, selection, integration, transformation and reduction will be discussed.
- Statistical data modelling
  Introduction of fundamental concepts of statistical data modelling, which include random variables, probability distribution functions, probability density functions, covariance matrix, correlation coefficient, linear regression, sampling, statistical inference and multivariate statistical analysis.
- Predictive modelling
  Introduction of the main predictive modelling techniques for data science, which include decision tree, nearest neighbour classifier, probabilistic classification, and connectionist models. In addition, the issues of classification performance evaluation and model selection will be discussed.
- Cluster analysis
  Introduction of the main clustering techniques: partitional, hierarchical, and density-based clustering. Important algorithms such as k-means, agglomerative hierarchical clustering, and DBSCAN will be discussed. Related issues in outlier analysis and detection will be introduced.
- Association rule mining
  Introduction of the Apriori algorithm for frequent pattern mining and association rule mining, and the comparison of different measures for evaluating the association patterns. Mining of frequent patterns in data streams will also be discussed.
- Text mining
  Introduction of the vector space model for document representation, the term frequency-inverse document frequency (tf-idf) approach for term weighting, and proximity measures such as cosine similarity for document comparison. Different algorithms in text mining such as document clustering and text classification will also be discussed.

Reading List

Compulsory Readings

	Title
1	Tan P. N., Steinbach M. and Kumar V. (2018) Introduction to Data Mining. Addison Wesley, 2nd edition.

Additional Readings

	Title
1	Bramer M. (2013) Principles of Data Mining. Springer, 2nd edition.
2	Han J. and Kamber M. (2011) Data Mining: Concepts and Techniques. Morgan Kaufmann, 3rd edition.
3	Witten I., Frank E., Hall M. and Pal C. (2016) Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, 4th edition.