SDSC3011: SOCIAL DATA PROCESSING AND MODELLING

Effective Term
Semester A 2024/25

Part I Course Overview

Course Title
Social Data Processing and Modelling

Subject Code
SDSC - Data Science
Course Number
3011

Academic Unit
Data Science (DS)

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
SDSC1001 Introduction to Data Science* and SDSC2001 Python for Data Science

*Pre-requisite SDSC1001 will be exempted for students who are enrolled in Minor in Data Science

Precursors
Nil

Equivalent Courses
Nil

Exclusive Courses
Nil

Part II Course Details

Abstract

This course provides students with an extensive exposure to the elements of data processing and modelling for social media. Topics include human error detection, missing data handling, record aggregation, data integration, categorical variable modelling, multivariate data modelling, multilevel data modelling, latent data modelling, temporal data modelling, and spatial data modelling.

Course Intended Learning Outcomes (CILOs)

	CILOs	Weighting (if app.)	DEC-A1	DEC-A2	DEC-A3
1	Explain clearly fundamental principles and methods of social media data processing and modelling	20	x		
2	Classify various properties of social media data and the corresponding modelling methods	20	x	x	
3	Evaluate existing practices in processing and modelling of social media data and seek ways to improve the existing practices	30	x	x	x
4	Apply appropriate processing/modelling methods to solve given practical problems in social media data	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	Lecture	Students will engage in formal lectures to gain knowledge about social data processing and modelling.	1, 2, 3, 4	39 hours in total
2	Case studies	Students will describe and critique classic cases of social media data processing and modelling.	2, 3, 4	in or after classes

Assessment Tasks / Activities (ATs)

	ATs	CILO No.	Weighting (%)	Remarks (e.g. Parameter for GenAI use)
1	Test	1, 2, 3, 4	20	Questions are designed for basic characteristics of social data to see how well the students have learned fundamental concepts and methods, and applications of social data processing. (20-40%)
2	Hands-in assignments	3, 4	30	These are skills based assessment to enable students to demonstrate the basic concepts and methods of social data modelling, and applications of the models in some applications. (0-30%)

Continuous Assessment (%)

50

Examination (%)

50

Examination Duration (Hours)

2

Additional Information for ATs

Note: To pass the course, apart from obtaining a minimum of 40% in the overall mark, a student must also obtain a minimum mark of 30% in both continuous assessment and examination components.

Assessment Rubrics (AR)

Assessment Task

Test

Criterion

Ability to understand and apply fundamental concepts and methods of social media data processing.

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
Hands-in assignments

Criterion
Ability to learn the basic concepts, apply methods and algorithms of social data modelling, and develop applications of modelling algorithms.

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
Ability to solve learning tasks using social media data modelling 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

Part III Other Information

Keyword Syllabus

Human error detection, missing data imputation, data transformation, record aggregation, data integration, multivariate data modelling, multilevel data modelling, latent data modelling, temporal data modelling, and spatial data modelling

Reading List

Compulsory Readings

	Title
1	Müller, H., & Freytag, J. C. (2005). Problems, methods, and challenges in comprehensive data cleansing. Professoren des Institute für Informatik.
2	Osborne, J. W. (2013). Best practices in data cleaning: A complete guide to everything you need to do before and after collecting your data. Sage.
3	Buttrey, S. E., & Whitaker, L. R. (2017). A Data Scientist's Guide to Acquiring, Cleaning, and Managing Data. John Wiley & Sons.

Additional Readings

	Title
1	Nil