

MS4252: BIG DATA ANALYTICS

Effective Term

Semester B 2024/25

Part I Course Overview

Course Title

Big Data Analytics

Subject Code

MS - Department of Decision Analytics and Operations

Course Number

4252

Academic Unit

Department of Decision Analytics and Operations (DAOS)

College/School

College of Business (CB)

Course Duration

One Semester

Credit Units

3

Level

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

Medium of Instruction

English

Medium of Assessment

English

Prerequisites

Nil

Precursors

IS2240 Python Programming for Business, MS3251 Analytics Using SAS, MS3252 Regression Analysis

Equivalent Courses

Nil

Exclusive Courses

Nil

Part II Course Details

Abstract

This course delves into the essential concepts and techniques of big data analytics, social network analysis, and advanced data mining within the framework of business administration. Students will explore the issues and success factors in

big data analytics, understanding both structured and unstructured data management, including tools like MapReduce and Hadoop. The curriculum covers information retrieval, web search, vector space and statistical language models, and singular value decomposition. Social network analysis topics include network measures, graph theory, centrality, prestige, and network propagation models. The course also examines recommendation systems, advanced data mining techniques such as Naïve Bayes, Support Vector Machines, and Random Forests, and text mining methods like NLP, TF-IDF, and sentiment analysis. Students will gain hands-on experience with professional software packages (e.g., SAS/DIS, SAS/EM, Python, R), enhancing their ability to solve real-world business problems. Additionally, the course prepares students to excel in interpersonal communication and teamwork.

Course Intended Learning Outcomes (CILOs)

CILOs		Weighting (if app.)	DEC-A1	DEC-A2	DEC-A3
1	Apply a wide range of big data, social network analytics, and advanced data mining techniques, recognizing their characteristics, strengths, and weaknesses	10		x	
2	Evaluate and assess emerging and newly-adopted methodologies and technologies to facilitate knowledge discovery.	40	x		
3	Formulate and prepare statistical-oriented data for business solutions.	20		x	
4	Perform big data, social network analysis, and advanced data mining techniques using professional software such as SAS/DIS, SAS/EM, Python, and R.	30			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	Engage in Lectures	Students will attend lectures where fundamental concepts and techniques of big data, social network analytics, and advanced data mining are explained using tools like SAS, Python, or R. This will help me understand and apply these methods effectively (CILO 1, CILO 2).	1, 2	
2	Participate in Tutorials	Students will actively participate in hands-on practice sessions to enhance my skills in big data and social network analytics using SAS, Python, or R. During tutorials, I will identify and tackle learning difficulties, examine business case issues, and practice different analytics techniques to analyze big data and generate various levels of statistical reporting (CILO 1, CILO 3, CILO 4).	1, 3, 4	

Assessment Tasks / Activities (ATs)

ATs		CILO No.	Weighting (%)	Remarks (e.g. Parameter for GenAI use)
1	Mid-term Test The mid-term test will assess students' understanding of big data concepts and my ability to apply big data techniques to solve practical problems. This test will provide evidence of students' capability to recognize the characteristics, strengths, and weaknesses of various big data techniques (CILO 1).	1	20	

2	Group project In this group project, students will collaborate with peers to complete a comprehensive big data project. The project will be assessed at various stages, including the design of project objectives, data modeling, project presentation, and final project report. This task will provide evidence of students' ability to evaluate methodologies, formulate data for business solutions, and perform big data analytics using professional software (CILO 2, CILO 3, CILO 4).	1, 3, 4	20	
3	Individual presentation and Q&A Students will provide an individual presentation on the group project, answer questions, and give recommendations to address the problems, justified by the data modeling. This task will provide evidence of students' ability to communicate findings, defend my analysis, and recommend solutions based on data-driven insights (CILO 3, CILO 4).	3, 4	20	

Continuous Assessment (%)

60

Examination (%)

40

Examination Duration (Hours)

3

Additional Information for ATs

Examination The final examination will assess students' comprehensive professional knowledge and ability to apply big data techniques to solve business problems. This exam will provide evidence of students' understanding and application of all the concepts covered in the course (CILO 1, CILO 2, CILO 3, CILO 4).

Assessment Rubrics (AR)**Assessment Task**

Mid-term Test

Criterion

1.1 ABILITY to EXPLAIN the key concepts and fundamental knowledge of big data and social network analysis

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

Group project

Criterion

2.1 CAPACITY for COLLABORATING with students to carry out problem-based activities based on real world problems.

2.2 ABILITY to EXPLAIN in DETAIL and with ACCURACY methods in analysing the relationship between business and sustainability solutions.

2.3 CAPACITY for SELF-DIRECTED LEARNING to find solutions to the problems and make recommendations

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

Individual presentation and Q&A

Criterion

3.1 ABILITY to UNDERSTAND the knowledge of big data and social network analysis

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

4.1 ABILITY to EXPLAIN the key concepts and fundamental knowledge of big data and social network analysis

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

a. Issue of big data analytics for business

Success factors for big data and social network analytics, The Analysis Process, Business Point of view in big data, Analytic Complexity;

b. Structured and unstructured big data management

Unstructured and structured big data management; Probabilistic matching for unstructured data; Map Reduce and Hadoop; Analytics Process Model;

c. Big data, Social Network Analytics, and Advanced Data Mining technique

i. Information Retrieval and Web Search:

Vector Space Model; Statistical Language Model; Singular Value Decomposition;

ii. Social Network Analysis:

Network Measures and Structure; Graph Theory; Centrality (Degree, Eigenvector, etc); Prestige (Degree, Proximity, etc.); Network propagation and models (Epidemic, Threshold, Stochastic); Link Prediction; Path Analysis; Link Analysis; etc.

- iii. Statistical foundations of recommendation systems:
Content-based filtering; neighbourhood-based collaborative filtering; Model-based Collaborative Filtering;
- iv. Advanced Data Mining Tools and Technologies
Statistical foundations of Naïve Bayes classification; Support Vector Machine; Memory-based Reasoning; Ensemble and Random Forests;
- v. Text Mining and Modelling:
Natural Language Processing; Wordcloud; Term Frequency - Inverse Document Frequency (TF-IDF); Maximum Entropy; Text-Clustering; Expectation-Maximization; Text Association Rule; Opinion Mining and Sentiment Analysis
- vi. Lifetime Modelling:
Lifetime value model; Survival Analysis; A/B testing;

Reading List

Compulsory Readings

Title	
1	EMC Education Services, 2015. Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data. WILEY
2	Jennifer Golbeck 2013. Analyzing the Social Web. Elsevier Inc.
3	Goutam Chakraborty et al, 2013. Text Mining and Analysis: Practice Methods, Examples, and case studies using SAS. Cary, NC: SAS Institute Inc.
4	David Ghan, 2016. Introduction to SAS and Hadoop: Essentials Course Notes. Cary, NC: SAS Institute Inc.
5	Tan, P.N., Steinbach, M. and Kumar, V., 2014 Introduction to Data Mining. Pearson.
6	Marc Huber & Michael Patetta, 2013 Survival Analysis Using the Proportional Hazards Model Course Notes, SAS Institute.

Additional Readings

Title	
1	Lin, Jimmy. 2010, Data-Intensive Text Processing with MapReduce, Morgan & Claypool Publishers.
2	Svolba, Gerhard. 2006. Data preparation for analytics using SAS. Cary, NC: SAS Institute Inc.
3	Michael Berry, & Gordon Linoff, 2004. Data mining techniques: For marketing, sales, and customer support, John Wiley & Sons.
4	Madhavan Samir 2015. Mastering Python for Data Science. Packt Publishing.
5	Marco Bonzanini 2016. Mastering Social Media Mining with Python. Packt Publishing.
6	Bing Liu 2013, Web Data Mining. Springer