

# MS4252: BIG DATA ANALYTICS

---

## Effective Term

Semester A 2022/23

## Part I Course Overview

### Course Title

Big Data Analytics

### Subject Code

MS - Management Sciences

### Course Number

4252

### Academic Unit

Management Sciences (MS)

### 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 aims to :

- Provide fundamental concepts and techniques of using big data, social network analysis and advanced data mining techniques in the context of enterprise strategic decision making;
- Develop students' analytical ability to identify, formalize and solve the real world problem with big data, social network analytics and advanced data mining techniques;
- Develop students' hands-on experience of construction of big data, social network analysis, and advanced data mining techniques using professional software packages;
- Prepare students to demonstrate generic skills in interpersonal interaction, communication, working both individually and in teams.

### Course Intended Learning Outcomes (CILOs)

CILOs		Weighting (if app.)	DEC-A1	DEC-A2	DEC-A3
1	Understand and apply a wide range of big data, social network analytics and advanced data mining techniques, and recognize their characteristics, strengths and weaknesses	10		x	
2	Evaluate a wide range of emerging and newly-adopted methodologies and technologies to facilitate the knowledge discovery.	40		x	x
3	Formulate and prepare statistical-oriented data for business solutions	20		x	x
4	Perform big data, social network analysis and advanced data mining techniques using professional software (e.g. SAS/DIS, SAS/EM, Python, R)	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.

### Teaching and Learning Activities (TLAs)

TLAs	Brief Description	CILO No.	Hours/week (if applicable)
1	Lecture	Concepts and techniques of big data and social network analytics using SAS or Python or R are explained;	1, 2, 3

2	Tutorial	Hands-on practice to enhance their skills in big data and social network analytics using SAS or Python or R so that learning difficulties can be identified and tackled. Identify the business case issues regarding how to enhance the data retrieval and preparation for big data and social network analytics, performing different analytics technique to analyze the big data and generate different levels of statistical reporting.	1, 2, 3, 4	
---	----------	--	------------	--

**Assessment Tasks / Activities (ATs)**

	ATs	CILO No.	Weighting (%)	Remarks (e.g. Parameter for GenAI use)
1	Mid-term Test The test is designed to assess students' big data knowledge and ability in applying big data techniques to solve the problems	1, 2, 3	20	
2	Group project Students work together in groups to complete a big data project. Assessments/Evaluations are carried out at various stages of the project, including the design of the project objectives, data modelling, project presentation and project report.	1, 2, 3, 4	20	
3	Individual presentation and Q&A Students have to provide a project presentation, answer the questions, and give recommendations to address the problems, justified by the data modelling, in relation to the project at hand.	1, 2, 3, 4	20	

**Continuous Assessment (%)**

**Examination (%)**

40

**Examination Duration (Hours)**

3

**Additional Information for ATs**

**Examination**

The exam is designed to assess students' professional knowledge and ability in applying big data techniques to solve business problems

**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