NEP and Learning Outcomes based Curriculum Framework (LOCF)

For

Postgraduate Programme

Master of Business Administration

(Applicable from the Academic Session 2022-23)

Department of Management
Gurugram University, Gurugram (Haryana)
(A State Govt. University Established Under Haryana Act 17 of 2017)

MBA 2 Year: 3rd Semester

Course Code	Course Title	External Marks (Theory)	Internal Assessment Marks/ Workshop/Practical	External/ Practical Marks	Total Marks	Credit (L-T-P)
Discipline Specific Elective Courses						
223MGDSEBA1	Basic Data Analytics using R and Python	50	25	-	75	3-0-0
Specialization: Business Analytics (Two Electives from Specialization offered by Department in the semester)						
223MGDSEBA2	Predictive Modeling	50	25	-	75	3-0-0
223MGDSEBA3	Statistics with R	50	25	-	75	3-0-0
				Total	**225**	**9**

MBA 2 Year: 4th Semester

Course Code	Course Title	External Marks (Theory)	Internal Assessment Marks/ Workshop/ Practical	External/ Practical Marks	Total Marks	Credit (L-T-P)
Discipline Specific Elective Courses						
224MGDSEBA1	Relational Database Management and SQL	50	25	-	75	3-0-0
Specialization Specific Electives Courses (specialization areas offered under dual specialization scheme)						
Specialization: Business Analytics (Two Electives from Specialization offered by Department in the semester)						
224MGDSEBA2	Text Mining using NLP and Machine Learning	50	25	-	75	3-0-0
224MGDSEBA3	Big Data Analysis	50	25	-	75	3-0-0
				Total	225	9

Detailed Syllabus

Third Semester

Basic Data Analytics using R and Python

223MGDSEBA1

L-T-P
3-0-0

External Marks: 50
Internal Marks: 25
Time Allowed: 2 Hrs.

Type of Course: Discipline Specific Elective Course **(Business Analytics)**

Core Courses	General Elective Course	Ability Enhancement Courses	Skill Enhancement Course	Discipline Specific Elective Courses
				✓

Introduction to the Course: This course presents a gentle introduction into the concepts of business analysis, the role of a business Analyst, and the tools that are used to perform daily functions. The participants will gain an understanding of the data ecosystem and the fundamentals of data analysis through R and Python, such as data gathering or data mining. The participants will then learn the soft skills that are required to effectively communicate the data to stakeholders, and how mastering these skills can give them the option to become a data driven decision maker.

Course Outcomes:

The Objective of this course is to provide students with a conceptual understanding of business analysis, R and Python and the skills required in applying knowledge for decision making. Following learning objectives will be emphasized. This course attempts to:

CO1: Sharpen participant's ability to understand and analyze data.

CO2: Provide students with a conceptual understanding of R and Python.

CO3: Understand the need & relevance of application of R and Python knowledge in managerial decision making.

CO4: Emphasize on the interpretation and use of creation and accumulation.

DETAILED SYLLABUS:

UNIT I

Understanding data: Importing, plotting, understanding and cleaning the data. Understand Univariate and multivariate, categorical and quantitative data, visual presentations of data, descriptive statistics, data tables, interpretation from graphical charts-bar plots, box plots, scatter diagrams. Hands on case study using software.

UNIT II

Introduction to R: R Data Types (Character, Numeric, Integer, Logical Complex), Different Data Structures in R, Basics of R Syntax, install R / RStudio, RStudio interface, import, export, and view files, save files. Data types in R and its appropriate uses, Program Structure in R, Flow Control: For loop, If condition, While conditions, Concatenation of Data, Combining Vars, cbind, rbind, Sapply, apply, tapply functions, Introduction to Data Structure in R, Vectors, Lists, Scalars, Data Frames, Matrices, Arrays,

Factors.

UNIT III

Data Visualization using R: Introduction to ggplot, Univariate Graph, Bivariate Graph, Multivariate Graph, Customizing Graphs, Saving Graphs. Descriptive Analysis in R: Using Summary Command, Using Name Command, Summary command: Single value result, Summary command: Multiple Result cumulative commands, Descriptive Statics for R Data Frames, Descriptive statistics in R for Matrix Objects.

UNIT IV

Introduction to Python: Python Data Types: Functions, String and List, Python Data Types: Tuples and Dictionaries, Files and Exceptions, Types of Operators, Classes and Objects, Reading files with Open, writing files with Open, loading data with Pandas, working with and saving with Pandas, Array oriented Programming with Numpy, Data cleaning and preparation, Plotting and Visualization, data Aggregation and Group Operations.

SUGGESTED READINGS:

1. Levin & Rubin, Statistics for Business, Prentice Hall of India, Delhi.
2. Anderson, Quantitative Methods in Business, Thomson Learning, Bombay.
3. Anderson, Statistics for Business & Economics, Thomson Learning, Bombay.
4. Kothari C.R., Quantitative Techniques, Vikas Publishing House, New Delhi
5. Andy Field, Discovering Statistics Using SPSS, Pearson Press.
6. Damodar Gujrati, Basic Econometrics, McGraw Hill Education, 5th Edition
7. Joseph F. Hair Jr, William C. Black, Barry J. Babin, Rolph E. Anderson, Multivariate Data Analysis, Pearson Press.

NOTE: The duration of all the end term theory examinations shall be 3hours

Instructions for External Examiner: The question paper shall be divided in two sections as follows:

Section A	Seven (7) short answer type questions from whole of the syllabus carrying two marks each, This section will be compulsory	7*2=14 marks
Section B	8 questions (2 questions from each unit). The students will be required to attempt four questions	14*4= 56 marks
	selecting one question of 14 marks from each unit.	
	Total Marks	70 marks

Instructions for Internal Examiner: The internal assessment should be spread evenly throughout the semester and must include at least 3 independent components including a mid-term exam. Below are the suggested components for 30 marks. A teacher has a choice to change these components as per the need except the mid-term exam.

S. No.	Course Assessment Components	Marks/Weightage (%)
1	Assessment 1 : Class Participation(**CP**) And Individual Assessment	10
2	Assessment 2 : Mid Term Exam (**MTE**)	10
3	Assessment 3 : Case Analysis / Presentation (**CAP**)/ Group Project (**GP**) / Role Play / Live Projects/ Simulation / Worksheet Assessment	10
	Internal Assessment (**IA**) **(1+2+3)**	30 (30%)
	End-Term Examination (**EE**)	70 (70%)
Total Marks (IA+EE)		**100**

Mapping Matrix of Course: 223MGDSEBA1

Table 1: CO-PO Matrix for the Course 223MGDSEBA1

COURSE OUTCOMES	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	3	3	3	3	2	3	2	2
CO2	3	3	3	3	3	3	3	3
CO3	3	3	3	3	3	3	3	3
CO4	2	3	3	3	3	3	3	3
Average	2.75	3	3	3	2.75	3	2.75	2.75

Table 2: CO-PSO Matrix for the Course 223MGDSEBA1

CO	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	3	3	3	3	3	3
CO2	3	3	3	3	**3**	3
CO3	**3**	3	3	2	3	**3**
CO4	3	**3**	3	3	**3**	3
Average	3	3	3	2.75	**3**	3

Predictive Modelling
223MGDSEBA2

L-T-P
3-0-0

External Marks: 50
Internal Marks: 25
Time Allowed: 2 Hrs.

Type of Course: Discipline Specific Elective Course **(Business Analytics)**

Core Courses	General Elective Course	Ability Enhancement Courses	Skill Enhancement Course	Specialization Specific Elective Courses
				✓

Introduction to the Course: Predictive modeling is a name given to a collection of mathematical techniques having in common the goal of finding a mathematical relationship between a target, response, or "dependent" variable and various predictor or "independent" variables with the goal in mind of measuring future values of those predictors and inserting them into the mathematical relationship to predict future values of the target variable. Because these relationships are never perfect in practice, it is desirable to give some measure of uncertainty for the predictions, typically a prediction interval that has some assigned level of confidence like 95%. Another task in the process is model building.

Course outcomes:

After completing the course, students would be able to:

CO1: Draw inferences from data and computer output and its application for business decision.

CO2: Understand the use of statistics from modelling perspective

CO3: Use regression and model fit for better decision making

CO4: Understand the use of variables and their roles in regression models

DETAILED SYLLABUS:

UNIT I

Regression: Introduction to regression, simple regression, method of least square, goodness of fit: R2 , interpretation of regression coefficients, testing the significance of coefficients. Hands on case study using software.

UNIT II

Multiple regression model, sample size in regression, estimating partial regression coefficients, testing of the model fit, interpretation of ANOVA results of regression, testing of regression coefficients, method of regression-hierarchical (block wise entry), forced entry, stepwise regression, forward pass, backward pass, Interpretation of beta values. Interpretation of computer output and report writing

UNIT III

Assumptions of multiple regression-linearity, normality, autocorrelation, multi-co linearity-VIF and tolerance, condition index, homogeneity of variance (heteroscedasticity), regression plots, accuracy of regression modal-checking outliers, leverage and influence (case wise diagnosis). Interpretation of computer output and report writing

UNIT IV

Curvilinear regression-method of fitting the model, checking the assumptions and interpretation of computer output. Concept of Dummy variable in regression, use of dummy variables in seasonal analysis, piecewise linear regression, interaction effect of predictive variables and interpretation of results. Interpretation of computer output and report writing

SUGGESTED READINGS:

1. Levin & Rubin, Statistics for Business, Prentice Hall of India, New Delhi.
2. Anderson, Quantitative Methods in Business, Thomson Learning, Bombay.
3. Anderson, Statistics for Business & Economics, Thomson Learning, Bombay.
4. Kothari C.R., Quantitative Techniques, Vikas Publishing House, New Delhi
5. Andy Field, Discovering Statistics Using SPSS, Pearson Press.
6. Damodar Gujrati, Basic Econometrics, McGraw Hill Education, 5th Edition
7. Joseph F. Hair Jr, William C. Black, Barry J. Babin, Rolph E. Anderson, Multivariate Data Analysis, Pearson Press.

NOTE: The duration of all the end term theory examinations shall be 3hours

Instructions for External Examiner: The question paper shall be divided in two sections as follows:

Section A	Seven (7) short answer type questions from whole of the syllabus carrying two marks each, This section will be compulsory	7*2=14 marks
Section B	8 questions (2 questions from each unit). The students will be required to attempt four questions selecting one question of 14 marks from each unit.	14*4= 56 marks
	Total Marks	70 marks

Instructions for Internal Examiner: The internal assessment should be spread evenly throughout the semester and must include at least 3 independent components including a mid-term exam. Below are the suggested components for 30 marks. A teacher has a choice to change these components as per the need except the mid-term exam.

S. No.	Course Assessment Components	Marks/Weightage (%)
1	Assessment 1 : Class Participation(**CP**) And Individual Assessment	10
2	Assessment 2 : Mid Term Exam **(MTE)**	10
3	Assessment 3 : Case Analysis / Presentation **(CAP)**/ Group Project **(GP)** / Role Play / Live Projects/ Simulation / Worksheet Assessment	10
	Internal Assessment **(IA) (1+2+3)**	30 (30%)
	End-Term Examination **(EE)**	70 (70%)
Total Marks (IA+EE)		**100**

Mapping Matrix of Course: 223MGDSEBA2

Table 1: CO-PO Matrix for the Course 223MGDSEBA2

COURSE OUTCOMES	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	3	3	3	3	2	3	2	2
CO2	3	3	3	3	3	3	3	3
CO3	3	3	3	3	3	3	3	3
CO4	2	3	3	3	3	3	3	3
Average	2.75	3	3	3	2.75	3	2.75	2.75

Table 2: CO-PSO Matrix for the Course 223MGDSEBA2

CO	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	3	3	3	3	3	3
CO2	3	3	3	3	**3**	3
CO3	**3**	3	3	2	3	**3**
CO4	3	**3**	3	3	**3**	3
Average	3	3	3	2.75	**3**	3

Statistics with R
223MGDSEBA3

220-DE-3000013BA

L-T-P
3-0-0

External Marks: 50
Internal Marks: 25
Time Allowed: 2 Hrs.

Type of Course: Discipline Specific Elective Course **(Business Analytics)**

Core Courses	General Elective Course	Ability Enhancement Courses	Skill Enhancement Course	Specialization Specific Elective Courses
				✓

Introduction to the Course: The following module comprises of R programming basics and application of several Statistical Techniques using it. The module aims to provide exposure in terms of Statistical Analysis, Hypothesis Testing, Regression and Correlation using R programming language. The objective of this module to make students exercise the fundamentals of statistical analysis in R environment. They would be able to analysis data for the purpose of exploration using Descriptive and Inferential Statistics. Students will understand Probability and Sampling Distributions and learn the creative application of Linear Regression in multivariate context for predictive purpose.

Course Outcomes: On completion of this course, the students will be able to

CO1: Install, Code and Use R Programming Language in R Studio IDE to perform basic tasks on Vectors, Matrices and Data frames.

CO2: Describe key terminologies, concepts and techniques employed in Statistical Analysis.

CO3: Define Calculate, Implement Probability and Probability Distributions to solve a wide variety of problems.

CO4: Conduct and interpret a variety of Hypothesis Tests to aid Decision Making.

Detailed Syllabus:

Unit -I

R and R Studio, Logical Arguments, Missing Values, Characters, Factors and Numeric, Help in R, Vector to Matrix, Matrix Access, Data Frames, Data Frame Access, Basic Data Manipulation Techniques, Usage of various apply functions – apply, lapply, sapply and tapply, Outliers treatment.

Unit II

Types of Data, Nominal, Ordinal, Scale and Ratio, Measures of Central Tendency, Mean, Mode and Median, Bar Chart, Pie Chart and Box Plot, Measures of Variability, Range, Inter-Quartile- Range, Standard Deviation, Skewness and Kurtosis, Histogram, Stem and Leaf Diagram, Standard Error of Mean and Confidence Intervals. Experiment, Sample Space and Events, Classical Probability, General Rules Of Addition, Conditional Probability, General Rules For Multiplication, Independent Events, Bayes' Theorem, Discrete Probability Distributions: Binomial, Poisson, Continuous Probability Distribution, Normal Distribution & t-distribution, Sampling Distribution and Central Limit Theorem.

Unit III

Population and Sample, Null and Alternate Hypothesis, Level of Significance, Type I and Type II Errors, One Sample t Test, Confidence Intervals, One Sample Proportion Test, Paired Sample t Test, Independent Samples t Test, Two Sample Proportion Tests, One Way Analysis of Variance and Chi Square Test.

Unit IV

Analysis of Relationship, Positive and Negative Correlation, Perfect Correlation, Correlation Matrix, Scatter Plots, Simple Linear Regression, R Square, Adjusted R Square, Testing of Slope, Standard Error

of Estimate, Overall Model Fitness, Assumptions of Linear Regression, Multiple Regression, Coefficients of Partial Determination, Durbin Watson Statistics, Variance Inflation Factor.

Suggested Readings:

1. Ken Black, 2013, Business Statistics, New Delhi, Wiley.
2. Lee, Cheng. et al., 2013, Statistics for Business and Financial Economics, New York: Heidelberg Dordrecht.
3. Anderson, David R., Thomas A. Williams and Dennis J. Sweeney, 2012, Statistics for Business and Economics, New Delhi: South Western.
4. Waller, Derek, 2008, Statistics for Business, London: BH Publications.
5. Levin, Richard I. and David S. Rubin, 1994, Statistics for Management, New Delhi: Prentice Hall.

Instructions for External Examiner: The question paper shall be divided in two sections as follows:

Section A	Seven (7) short answer type questions from whole of the syllabus carrying two marks each, This section will be compulsory	7*2=14 marks
Section B	8 questions (2 questions from each unit). The students will be required to attempt four questions selecting one question of 14 marks from each unit.	14*4= 56 marks
	Total Marks	70 marks

Instructions for Internal Examiner: The internal assessment should be spread evenly throughout the semester and must include at least 3 independent components including a mid-term exam. Below are the suggested components for 30 marks. A teacher has a choice to change these components as per the need except the mid-term exam.

S. No.	Course Assessment Components	Marks/Weightage (%)
1	Assessment 1: Class Participation **(CP)** And Individual Assessment	10
2	Assessment 2: Mid Term Exam **(MTE)**	10
3	Assessment 3: Case Analysis / Presentation **(CAP)**/ Group Project **(GP)** / Role Play / Live Projects/ Simulation / Worksheet Assessment	10
	Internal Assessment **(IA) (1+2+3)**	30 (30%)
	End-Term Examination **(EE)**	70 (70%)
Total Marks (IA+EE)		**100**

Mapping Matrix of Course: 223MGDSEBA3

Table 1: CO-PO Matrix for the Course 223MGDSEBA3

COURSE OUTCOMES	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	3	3	3	3	2	3	2	2
CO2	3	3	3	3	3	3	3	3

CO3	3	3	3	3	3	3	3	3
CO4	2	3	3	3	3	3	3	3
Average	2.75	3	3	3	2.75	3	2.75	2.75

Table 2: CO-PSO Matrix for the Course 223MGDSEBA2

CO	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	3	3	3	3	3	3
CO2	3	3	3	3	**3**	3
CO3	**3**	3	3	2	3	**3**
CO4	3	**3**	3	3	**3**	3
Average	3	3	3	2.75	**3**	3

Fourth Semester

Relational Database Management and SQL 220-DE-400191BA

224MGDSEBA1

L-T-P
3-0-0

External Marks: 50
Internal Marks: 25
Time Allowed: 2 Hrs.

Type of Course: Discipline Specific Elective Course **(Business Analytics)**

Core Courses	General Elective Course	Ability Enhancement Courses	Skill Enhancement Course	Discipline Specific Elective Courses
				✓

Introduction to the Course: In this course, the students will learn the basics of the SQL/No SQL and the Relational Databases. They will learn about the Relational Model and Relational Model concepts and constraints. The students will get exposure to key concepts with regards to SQL Language and DBMS such as Normalization, Transaction Processing along-side an exposure to No SQL programming.

Course Outcomes:
CO1: To understand the basic concepts and the applications of Database Systems.
CO2: To master the basics of SQL and construct queries using SQL.
CO3: To become familiar with the basic issues of Transaction Processing and Concurrency Control.
CO4: To become familiar with NO SQL Programming Language and Explain the architecture, define objects, load and query data within No SQL databases.

DETAILED SYLLABUS:
UNIT I
Introduction to Database Management Systems
Introduction-Database System Applications, Purpose of Database Systems, Views of Data, Data Abstraction, Instances and Schemas, Data Models, Database Languages, DDL, DML, Database Architecture, Database Users and Administrators, Database Design, ER Diagrams, Entities, Attributes and Entity Sets, Relationships and Relationship sets, Integrity Constraints, Views.

UNIT II
SQL Operators and Relational Theorems Relational Algebra and Calculus, Selection and Projection, Set Operations, Renaming, Joins, Division, Relational calculus, Tuple Relational Calculus, Domain Relational Calculus, Forms
of Basic SQL Query, Nested Queries, Comparison Operators, Aggregate Operators, NULL values, Logical connectives, AND, OR and NOT, Outer Joins, Triggers.

UNIT III
Normalization: Problems Caused by Redundancy, Decompositions, Functional Dependencies, Normal Forms, First, Second, Third Normal forms, BCNF, Properties of Decompositions, Loss less Join.

Decomposition, Dependency Preserving Decomposition, Multi Valued Dependencies, Fourth Normal Form, Join Dependencies, Fifth Normal Form.

UNIT IV

Transactions: Transaction Management, Transaction Concept, Transaction State, Implementation of Atomicity and Durability, Concurrent, Executions, Serializability, Recoverability, Implementation of Isolation, testing for Serializability, Concurrency Control, Lock, Timestamp Based Protocols, Validation Based Protocols, Recovery, Failure Classification, Storage Structure, Atomicity, Log Based Recovery, Remote Backup Systems.

SUGGESETD READINGS:

1. Guy Harrison, 2015, Next Generation Databases: No SQL and Big Data, Apress.
2. RamezElmasri, ShamkatB.Navathe, 2013, Database Systems, Pearson.
3. Pramod J. Sadalage, Martin Fowler, 2012, No SQL Distilled, Addison Wesley.
4. A.Silberschatz, H.F. Korth, S.Sudarshan, 2006, Database System Concepts, McGraw Hill.
5. Raghurama Krishnan, Johannes Gehrke, 2003, Database Management Systems, McGraw Hill.

NOTE: The duration of all the end term theory examinations shall be 3hours

Instructions for External Examiner: The question paper shall be divided in two sections as follows:

Section A	Seven (7) short answer type questions from whole of the syllabus carrying two marks each, This section will be compulsory	7*2=14 marks
Section B	8 questions (2 questions from each unit). The students will be required to attempt four questions selecting one question of 14 marks from each unit.	14*4= 56 marks
	Total Marks	70 marks

Instructions for Internal Examiner: The internal assessment should be spread evenly throughout the semester and must include at least 3 independent components including a mid-term exam. Below are the suggested components for 30 marks. A teacher has a choice to change these components as per the need except the mid-term exam.

S. No.	Course Assessment Components	Marks/Weightage (%)
1	Assessment 1 : Class Participation(**CP**) And Individual Assessment	10
2	Assessment 2 : Mid Term Exam (**MTE**)	10

3	Assessment 3 : Case Analysis / Presentation **(CAP)**/ Group Project **(GP)** / Role Play / Live Projects/ Simulation / Worksheet Assessment	10
	Internal Assessment **(IA) (1+2+3)**	30 (30%)
	End-Term Examination **(EE)**	70 (70%)
Total Marks (IA+EE)		**100**

Mapping Matrix of Course: 224MGDSEBA1

Table 1: CO-PO Matrix for the Course 224MGDSEBA1

COURSE OUTCOMES	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	3	3	3	3	2	3	2	2
CO2	3	3	3	3	3	3	3	3
CO3	3	3	3	3	3	3	3	3
CO4	2	3	3	3	3	3	3	3
Average	2.75	3	3	3	2.75	3	2.75	2.75

Table 2: CO-PSO Matrix for the Course 224MGDSEBA1

CO	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	3	3	3	3	3	3
CO2	3	3	3	3	**3**	3
CO3	**3**	3	3	2	3	**3**
CO4	3	**3**	3	3	**3**	3
Average	3	3	3	2.75	**3**	3

Text Mining using NLP and Machine Learning 220-DE-400192BA
224MGDSEBA2

L-T-P
3-0-0

External Marks: 50
Internal Marks: 25
Time Allowed: 2 Hrs.

Type of Course: Discipline Specific Elective Course **(Business Analytics)**

Core Courses	General Elective Course	Ability Enhancement Courses	Skill Enhancement Course	Specialization Specific Elective Courses
				✓

Introduction to the Course: Data mining is the process of extracting patterns from data. Data mining is becoming an increasingly important tool to transform the data into information. It is commonly used in a wide range of profiling practices, such as marketing, surveillance, fraud detection and scientific discovery. Data mining can be applied on a variety of data types. Data types include structured data (relational), multimedia data, free text, and hypertext. Nowadays, text is the most common and convenient way for information exchange. This is due to the fact that much of the world's data is contained in text documents (newspaper articles, emails, literature, web pages, etc.). The importance of this way has led many researchers to find out suitable methods to analyze natural language texts to extract the important and useful information. In comparison with data stored in structured format (databases), texts stored in documents is unstructured and to deal with such data, a preprocessing is required to transform textual data into a suitable format for automatic processing.

Course Outcomes:

CO1: To develop an understanding of the fundamentals of data and text miming and statistical pattern recognition.

CO2: To gain an insight into the various components of machine learning.

CO3: To Understand supervised learning, unsupervised learning, learning theory, reinforcement learning and adaptive control.

CO4: To acquire skills that can be applied to various components of machine learning to applications like robotic control, data mining, autonomous navigation, bioinformatics, speech recognition, and text and web data processing.

DETAILED SYLLABUS:

Unit I

Introduction to Text Mining: Basics of Text Mining, Natural Language Content Analysis, Core Text Mining Operations, Associations, Using Background Knowledge for Text Mining, Domain Ontologies, Domain Lexicons. Text Mining Pre-processing Techniques, Task Oriented Approaches, NLP Tasks, Tokenization, Part-of-Speech Tagging, Syntactical Parsing and Shallow Parsing.

Unit II

Text Categorization and Clustering: Applications of Text Categorization, Document Representation, Knowledge Engineering Approach to Text Categorization, Machine Learning Approach to Text Categorization, Evaluation of Text Classifiers. Clustering Tasks in Text Analysis, Clustering Algorithms and Clustering of Textual Data. Latent Dirichlet Allocation, Word Topic Probabilities, Per-Document Classification, Bywords Assignments, Alternative LDA Implementations. Hidden Markov models, Stochastic Context Free Grammar, Conditional Random fields, Parallel Learning Algorithms.

Unit III

Machine Learning and Cognitive Intelligence: Introduction to Machine Learning- History and Evolution, Machine Learning categories: Supervised, Unsupervised and Reinforcement learning. Framework for building ML Systems- KDD process model, CRISP-DM & SEMMA, Machine learning Python packages, Machine Learning Core Libraries. Introduction to Cognitive Intelligence, Features of Cognitive Intelligence

Unit IV

Supervised and Unsupervised Learning: Supervised Learning: Introduction to classification, Linear Regression, Metrics for evaluating linear model, Multivariate regression, Non-Linear Regression, K-Nearest Neighbour, Decision Trees, Logistic Regression, Support Vector Machines, Model Evaluation, Applications of supervised learning in multiple domains. Unsupervised Learning: Clustering, Hierarchical clustering, Partitioning Clustering- K-mean clustering, Applications of unsupervised learning in multiple domains.

SUGGESTED READINGS:

1. Richard Duda, Peter Hart and David Stork, "Pattern Classification," John Wiley & Sons.
2. Tom Mitchell, "Machine Learning." McGraw-Hill.
3. Richard Sutton and Andrew Barto, "Reinforcement Learning: An introduction," MIT Press.
4. Trevor Hastie, Robert Tibshirani and Jerome Friedman, "The Elements of Statistical Learning," Springer.

NOTE: The duration of all the end term theory examinations shall be 3hours

Instructions for External Examiner: The question paper shall be divided in two sections as follows:

Section A	Seven (7) short answer type questions from whole of the syllabus carrying two marks each, This section will be compulsory	7*2=14 marks
Section B	8 questions (2 questions from each unit). The students will be required to attempt four questions selecting one question of 14 marks from each unit.	14*4= 56 marks
	Total Marks	70 marks

Instructions for Internal Examiner: The internal assessment should be spread evenly throughout the semester and must include at least 3 independent components including a mid-term exam. Below are the suggested components for 30 marks. A teacher has a choice to change these components as per the need except the mid-term exam.

S. No.	Course Assessment Components	Marks/Weightage (%)
1	Assessment 1 : Class Participation(**CP**) And Individual Assessment	10
2	Assessment 2 : Mid Term Exam (**MTE**)	10

3	Assessment 3 : Case Analysis / Presentation **(CAP)**/ Group Project **(GP)** / Role Play / Live Projects/ Simulation / Worksheet Assessment	10
	Internal Assessment **(IA) (1+2+3)**	30 (30%)
	End-Term Examination **(EE)**	70 (70%)
Total Marks (IA+EE)		**100**

Mapping Matrix of Course: 224MGDSEBA2

Table 1: CO-PO Matrix for the Course 224MGDSEBA2

COURSE OUTCOMES	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	3	3	3	3	2	3	2	2
CO2	3	3	3	3	3	3	3	3
CO3	3	3	3	3	3	3	3	3
CO4	2	3	3	3	3	3	3	3
Average	2.75	3	3	3	2.75	3	2.75	2.75

Table 2: CO-PSO Matrix for the Course 224MGDSEBA2

CO	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	3	3	3	3	3	3
CO2	3	3	3	3	**3**	3
CO3	**3**	3	3	2	3	**3**
CO4	3	**3**	3	3	**3**	3
Average	3	3	3	2.75	**3**	3

Big Data Analytics
224MGDSEBA3

L-T-P
3-0-0

220-DE-400193BA
External Marks: 50
Internal Marks: 25
Time Allowed: 2 Hrs.

Type of Course: Discipline Specific Elective Course **(Business Analytics)**

Core Courses	General Elective Course	Ability Enhancement Courses	Skill Enhancement Course	Specialization Specific Elective Courses
				✓

Introduction to the course: This module forms an introduction as well an in-depth study in the field of Big Data and Hadoop. It comprises of the fundamentals as well as advanced topics needed to progress in this technology. The students will learn about the applications, usage and several use case scenarios pertaining to Big Data- Hadoop where they can use the knowledge and progress ahead. This course will help students gain knowledge and understanding about Big Data Technology, Hadoop Ecosystem and various tools related to it. The students will learn about the HDFS File System, Map Reduce Framework, Analysing data using Hbase and Hive along with the Integration of R with Hadoop.

Course Outcomes: On completion of this course, the students will be able to
CO1: Understand the fundamentals of Big Data and its Applications in various Domains.
CO2: Conceptualize and Incorporate the Technologies behind Big Data.
CO3: Understand HDFS File Structure, Map Reduce Framework, the architectures related to them CO4: Integrate R with Hadoop and solve analytical problems.

Detailed syllabus:

Unit-I
What Is Big Data? History of Data Management, Evolution of Big Data, Structuring of Big Data, Elements of Big Data, Application of Big Data in the Business Context, Careers in Big Data. Business Applications of Big Data: The Significance of Social Network Data, Financial Fraud and Big Data, Fraud Detection in Insurance, Use of Big Data in the Retail Industry. Distributed and Parallel Computing for Big Data, Understanding Hadoop, Cloud Computing, Grid Computing and In-Memory Technology for Big Data. VMWare Installation of Hadoop, Linux and its Shell Commands, Different Hadoop Distributions and their advantages, Hortonworks, Cloudera, MapR.
Unit-II
The Hadoop Ecosystem, Storing Data with HDFS, Design of HDFS, HDFS Concepts, Command Line Interface to HDFS, Hadoop File Systems, Java Interface to Hadoop, Anatomy of a file read, Anatomy of a file write, Replica placement and Coherency Model. Parallel Copying with distcp, keeping an HDFS Cluster Balanced.

Unit-III

Origins of Map Reduce, How Map Reduce Works, Optimization Techniques for Map Reduce Jobs, Applications of Map Reduce, Java Map Reduce classes (new API), Data flow, combiner functions, running a distributed Map Reduce Job. Configuration API, setting up the development environment, Managing Configuration.

Unit-IV

Understanding R-Hadoop, Integration Procedure, Packages needed for R under Hadoop Ecosystem, Text Mining for Deriving Useful Information using R within Hadoop, Introduction to Hive & Hbase, Hive and Hbase Architecture, Understanding Queries, Mining Big Data with Hive & Hbase.

Suggested Readings:

1. Arshdeep Bahga, 2016, Big Data Science & Analytics: A Hands-On Approach.
2. Tom White, 2012, Hadoop: The Definitive Guide, O'Reilly.
3. Adam Shook and Donald Miner, 2012, Map Reduce Design Patterns: Building Effective
4. Algorithms and Analytics for Hadoop and Other Systems, O'Reilly.
5. Dean Wampler, Edward Capriolo & Jason Rutherglen, 2012, Programming Hive, O'Reilly.
6. Lars George, 2011, HBase - The Definitive Guide: Random Access to Your Planet- Size Data, O'Reilly.

Instructions for External Examiner: The question paper shall be divided in two sections as follows:

Section A	Seven (7) short answer type questions from whole of the syllabus carrying two marks each, This section will be compulsory	7*2=14 marks
Section B	8 questions (2 questions from each unit). The students will be required to attempt four questions selecting one question of 14 marks from each unit.	14*4= 56 marks
	Total Marks	70 marks

Instructions for Internal Examiner: The internal assessment should be spread evenly throughout the semester and must include at least 3 independent components including a mid-term exam. Below are the suggested components for 30 marks. A teacher has a choice to change these components as per the need except the mid-term exam.

S. No.	Course Assessment Components	Marks/Weightage (%)
1	Assessment 1 : Class Participation(**CP**) And Individual Assessment	10
2	Assessment 2 : Mid Term Exam (**MTE**)	10
3	Assessment 3 : Case Analysis / Presentation (**CAP**)/ Group Project (**GP**) / Role Play / Live Projects/ Simulation / Worksheet Assessment	10
	Internal Assessment (**IA**) (**1+2+3**)	30 (30%)
	End-Term Examination (**EE**)	70 (70%)
Total Marks (IA+EE)		**100**

Mapping Matrix of Course: 224MGDSEBA3

Table 1: CO-PO Matrix for the Course 224MGDSEBA3

COURSE OUTCOMES	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	3	3	3	3	2	3	2	2
CO2	3	3	3	3	3	3	3	3
CO3	3	3	3	3	3	3	3	3
CO4	2	3	3	3	3	3	3	3
Average	2.75	3	3	3	2.75	3	2.75	2.75

Table 2: CO-PSO Matrix for the Course 224MGDSEBA3

CO	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	3	3	3	3	3	3
CO2	3	3	3	3	**3**	3
CO3	**3**	3	3	2	3	**3**
CO4	3	**3**	3	3	**3**	3
Average	3	3	3	2.75	**3**	3