PANIMALAR ENGINEERING COLLEGE

(An Autonomous Institution, Affiliated to Anna University, Chennai)
Bangalore Trunk Road, Varadharajapuram,
Poonamallee, Chennai – 600123

Minor Degree

DATA SCIENCE AND ANALYTICS

Curriculum & Syllabus

DEPARTMENT OF

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

REGULATION 2023

PANIMALAR ENGINEERING COLLEGE

Department of Artificial Intelligence and Data Science

MINOR DEGREE
on
DATA SCIENCE AND ANALYTICS

S. NO.	COURSE CODE	COURSE TITLE	CATE GORY	PERIODS PER WEEK			TOTAL CONTACT PERIODS	CREDITS
				L	T	P		
1.	23AD4001	Introduction to Data Science	PE	3	0	0	3	3
2.	23AD4002	Data Mining	PE	3	0	0	3	3
3.	23AD4003	Big Data Analytics	PE	3	0	0	3	3
4.	23AD4004	Data Visualization	PE	3	0	0	3	3
5.	23AD4005	Predictive Modelling	PE	3	0	0	3	3
6.	23AD4006	Image and Video Analytics	PE	3	0	0	3	3
7.	23AD4007	Text Analytics	PE	3	0	0	3	3
8.	23AD4008	Business Analytics	PE	3	0	0	3	3

23AD4001	INTRODUCTION TO DATA SCIENCE	L	T	P	C
		3	0	0	3

COURSE OBJECTIVE:
- To acquire skills in data preparatory and preprocessing steps.
- To understand the mathematical skills in statistics.
- To learn the tools and packages in Python for data science.
- To gain knowledge to understand classification and Regression Model.
- To acquire knowledge in data interpretation and visualization techniques.

UNIT- I   INTRODUCTION   9

Need for data science – benefits and uses – facets of data – data science process – setting the research goal – retrieving data – cleansing, integrating, and transforming data – exploratory data analysis - build the models - presenting and building applications.

UNIT- II   DESCRIBING DATA I   9

Frequency distributions – Outliers – relative frequency distributions – cumulative frequency distributions – frequency distributions for nominal data – interpreting distributions – graphs – averages - mode - median - mean - averages for qualitative and ranked data - describing variability - range - variance - standard deviation - degrees of freedom - inter quartile range.

UNIT- III   DESCRIBING DATA II   9

Normal distributions – z scores – normal curve problems – finding proportions – finding scores – correlation - scatter plots - correlation coefficient for quantitative data - computational formula for correlation coefficient - regression - regression line - least squares regression line - standard error of estimate - multiple regression equations - regression toward the mean.

UNIT- IV   PYTHON FOR DATA HANDLING   9

Basics of Numpy arrays - aggregations - computations on arrays - comparisons, masks, boolean logic – fancy indexing – structured arrays – Data manipulation with Pandas – data indexing and selection - operating on data - missing data – combining datasets - aggregation and grouping.

UNIT- V   PYTHON FOR DATA VISUALIZATION   9

Visualization with matplotlib - line plots - scatter plots - visualizing errors - density and contour plots - histograms, binnings and density - three dimensional plotting - geographic data – graph plotting using Plotly.

TOTAL: 45 PERIODS

COURSE OUTCOME(S):
Upon successful completion of the course, the students will be able to

CO1	Demonstrate the skills of data inspecting and cleansing.
CO2	Determine the relationship between data dependencies using statistics.
CO3	Construct data using primary tools for data science in Python.
CO4	Represent the useful information using mathematical skills.
CO5	Apply the knowledge for data describing the Data.
CO6	Implement real time application using visualization tools.

TEXTBOOKS:
1. David Cielen, Arno D. B. Meysman, and Mohamed Ali, Introducing Data Science, Manning Publications, 2016.
2. Robert S. Witte and John S. Witte, Statistics, Eleventh Edition, Wiley Publications, 2017.
3. Jake VanderPlas, Python Data Science Handbook, O'Reilly, 2016.

REFERENCE BOOKS:
1. Allen B. Downey, Think Stats: Exploratory Data Analysis in Python, Green Tea Press, 2014.

CO – PO-PSO MAPPING

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PSO1	PSO2	PSO3
CO1	3	2	–	2	3	–	–	–	–	–	2	2	3	2
CO2	3	3	–	2	2	–	–	–	–	–	2	2	3	2
CO3	3	2	2	2	3	–	–	–	–	–	2	2	2	3
CO4	3	2	1	2	3	–	–	–	–	–	2	2	3	2
CO5	3	2	1	2	2	–	–	–	–	–	2	2	3	2
CO6	2	2	2	2	3	2	–	2	2	1	3	3	3	3

Internal Assessment				End Semester Examinations
Assessment I (100 Marks)		Assessment II (100 Marks)		
Individual Assignment / Case Study / Seminar / Mini Project	Written Test	Individual Assignment / Case Study / Seminar / Mini Project	Written Test	Written Examinations
40	60	40	60	100
40%				**60 %**

23AD4002	DATA MINING	L	T	P	C
		3	0	0	3

COURSE OBJECTIVE:
- To learn the basic tasks in data mining and how they are different from knowledge discovery.
- To understand simple techniques like decision trees, neural networks, and genetic methods used to find patterns in data.
- To learn how to measure similarity and differences in data for grouping and classification.
- To compare different methods to find the best way to group or classify data.
- To understand how to find useful patterns (rules) in data and check their quality.

UNIT – I   INTRODUCTION TO DATA MINING   9
Basic Data Mining Tasks - Data Mining Versus Knowledge Discovery in Data Bases - Data Mining Issues - Data Mining Matrices - Social Implications of Data Mining - Data Mining from Data Base Perspective.

UNIT – II   DATA MINING TECHNIQUES   9
Data Mining Techniques – a Statistical Perspective on data mining – Similarity Measures – Decision Trees - Neural Networks - Genetic Algorithms.

UNIT – III   CLASSIFICATION METHODS   9
Classification: Introduction – Statistical – Based Algorithms - Distance Based Algorithms – Decision.

UNIT – IV   CLUSTERING TECHNIQUES   9
Clustering Tree - Based Algorithms - Neural Network Based Algorithms - Rule Based Algorithms - Combining Techniques: Introduction - Similarity and Distance Measures - Outliers - Hierarchical Algorithms. Partitioned Algorithms.

UNIT – V   ASSOCIATION RULES   9
Association Rules: Introduction - Large Item Sets – Basic Algorithms – Parallel & Distributed Algorithms – Comparing Approaches – Incremental Rules – Advanced Association Rules Techniques - Measuring the Quality of Rules.

COURSE OUTCOME(S):
Upon successful completion of the course, the students will be able to

CO1 Explain what data mining is and how it is different from knowledge discovery.
CO2 Apply basic statistics and techniques like decision trees and neural networks to find patterns in data.
CO3 Classify data using different methods such as statistical, distance-based, and rule-based techniques.
CO4 Group similar data using clustering methods and detect unusual data points.
CO5 Discover useful patterns or rules from large data sets using suitable data mining algorithms.
CO6 Identify the social and ethical issues involved in using data mining in real-life applications.

TEXT BOOK:

1. Data Mining: Concepts and Techniques (4th Edition, 2022) by Jiawei Han, Jian Pei, and Hanghang Tong.

REFERENCE BOOK:
1. Margaret H.Dunbam, "*Data Mining Introductory and Advanced Topics*", Pearson Education 2003.

CO – PO-PSO MAPPING

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PSO1	PSO2	PSO3
CO1	3	2	1	1	2	1	1	-	-	-	2	2	3	2
CO2	3	3	2	2	3	-	-	-	-	-	2	2	3	3
CO3	3	3	2	2	3	-	-	-	-	-	2	2	3	3
CO4	3	3	2	2	3	-	-	-	-	-	2	2	2	3
CO5	3	3	2	2	3	-	-	-	-	-	2	2	3	3
CO6	2	2	1	1	2	3	3	2	2	1	3	3	2	2

Internal Assessment				End Semester Examinations
Assessment I (100 Marks)		Assessment II (100 Marks)		
Individual Assignment / Case Study / Seminar / Mini Project	Written Test	Individual Assignment / Case Study / Seminar / Mini Project	Written Test	Written Examinations
40	60	40	60	100
40%				60 %

23AD4003	BIG DATA ANALYTICS	L	T	P	C
		3	0	2	4

OBJECTIVES:
- To understand about big data.
- To learn and use NoSQL big data management.
- To learn mapreduce analytics using Hadoop and related tools.
- To work with map reduce applications
- To understand the usage of Hadoop related tools for Big Data Analytics

UNIT I   UNDERSTANDING BIG DATA   9

What is big data – why big data – convergence of key trends – unstructured data – industry examples of big data – web analytics - big data and marketing – fraud and big data – risk and big data – credit risk management – big data and algorithmic trading – big data and healthcare – big data in medicine - advertising and big data - big data technologies .

UNIT II   NOSQL DATA MANAGEMENT   9

Introduction to NoSQL – aggregate data models – aggregates – key-value and document data models – relationships – graph databases – schemaless databases – materialized views – distribution models – sharding – master-slave replication – peer-peer replication – sharding and replication – consistency .

UNIT III   BASICS OF HADOOP   9
Data format – analyzing data with Hadoop – scaling out – Hadoop streaming – Hadoop pipes – design of Hadoop distributed file system (HDFS) - HDFS concepts – Java interface – data flow - Hadoop I/O - data integrity - compression - serialization - Avro - file-based data structures.

UNIT IV   MAPREDUCE APPLICATIONS   9
MapReduce workflows - unit tests with MRUnit - test data and local tests - anatomy of MapReduce job run - classic Map-reduce - YARN - failures in classic Map-reduce and YARN - job scheduling - shuffle and sort - task execution - MapReduce types - input formats - output formats.

UNIT V   9
HADOOP RELATED TOOLS
Hbase - data model and implementations - Hbase clients - Hbase examples - praxis.Cassandra - cassandra data model - cassandra examples - cassandra clients - Hadoop integration. Pig - Grunt – pig data model – Pig Latin – developing and testing Pig Latin scripts. Hive – data types and file formats - HiveQL data definition - HiveQL data manipulation - HiveQL queries.

TOTAL:45 PERIODS

OUTCOMES:
After the completion of this course, students will be able to:

CO1 Describe big data concepts and characteristics.
CO2 Identify and explain real-world use cases of big data from selected business domains.
CO3 Explain NoSQL database systems used in big data management.
CO4 Install, configure, and operate Hadoop and HDFS.
CO5 Perform map-reduce programming for big data analytics using Hadoop.
CO6 Use Hadoop ecosystem tools such as HBase, Cassandra, Pig, and Hive for data analytics.

TEXT BOOKS:

1. Michael Minelli, Michelle Chambers, and AmbigaDhiraj, "Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses", Wiley, 2013.
2. Eric Sammer, "Hadoop Operations", O'Reilley, 2012.

REFERENCES:

1. E. Capriolo, D. Wampler, and J. Rutherglen, "Programming Hive", O'Reilley, 2012.
2. Lars George, "HBase: The Definitive Guide", O'Reilley, 2011.
3. Eben Hewitt, "Cassandra: The Definitive Guide", O'Reilley, 2010.
4. Alan Gates, "Programming Pig", O'Reilley, 2011.

CO-PO-PSO MAPPING

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PSO1	PSO2	PSO3
CO1	3	2	1	-	2	1	-	-	1	-	2	2	2	2
CO2	2	3	2	-	2	2	-	-	1	1	2	3	3	2
CO3	3	2	2	2	3	-	-	-	-	-	2	2	2	3
CO4	3	2	2	2	3	-	-	-	-	-	2	2	2	3
CO5	3	3	3	3	3	1	-	-	1	2	3	2	3	3
CO6	3	3	3	2	3	2	-	-	2	2	3	3	3	3

Internal Assessment				End Semester Examinations
Assessment I (100 Marks)		**Assessment II (100 Marks)**		
Individual Assignment / Case Study / Seminar / Mini Project	**Written Test**	**Individual Assignment / Case Study / Seminar / Mini Project**	**Written Test**	**Written Examinations**
40	60	40	60	100
40%				**60 %**

23AD4004	DATA VISUALIZATION	L	T	P	C
		3	0	0	3

COURSE OBJECTIVES

- To understand the basics of Data Explorations
- To understand the basic concepts of handling variables
- To study the visualization and integration of processing with java
- To explore the data visualization using R language
- To apply various data visualization techniques for a variety of tasks

UNIT- I   INTRODUCTION TO DATA EXPLORATION   9

EDA fundamentals - Introduction to Single variable: Distribution Variables -Numerical Summaries of Level and Spread - Scaling and Standardising – Inequality - Smoothing Time Series

UNIT-II   INTRODUCING TWO VARIABLE AND THIRD VARIABLE   9

Relationships between Two Variables - Percentage Tables - Analysing Contingency Tables - Handling Several Batches - Scatterplots and Resistant Lines – Transformations - Introducing a Third Variable - Causal Explanations - Three-Variable Contingency Tables and Beyond.

UNIT-III   BASICS OF DATA VISUALIZATION   9

The Seven Stages of Visualizing Data - Getting Started with Processing - Mapping - Time Series - Connections and Correlations - Scatterplot Maps - Trees, Hierarchies, and Recursion - Networks and Graphs – Acquiring Data – Parsing Data - Integrating Processing with Java.

UNIT- IV   DATA EXPLORATION AND DATA VISUALIZATION IN R   9

Introduction to R and R Studio - The Basics of Data Exploration - Loading Data into R - Transforming Data - Creating Tidy Data- Basic Analysis- Data Visualization with ggplot2.

UNIT- V   TECHNIQUES AND APPLICATIONS OF DATA EXPLORATION
**   AND VISUALIZATION IN R   9**

Basic Data Exploration Techniques - Basic Data Visualization Techniques - Visualizing Geographic Data with ggmap -, Creating a basemap, Adding operational data layers - R Markdown - Case Study – Wildfire Activity in the Western United States.

THEORY:45 PERIODS

TEXT BOOKS
1. Catherine Marsh, Jane Elliott, Exploring Data: An Introduction to Data Analysis for Social Scientists, Wiley Publications, 2nd Edition, 2008.
2. Visualizing Data: Exploring and Explaining Data with the processing Environment, O Reily Publications, 2007.
3. Eric Pimpler, Data Visualization and Exploration with R, Geo Spatial Training service, 2017.
4. Authors: Xiang Zhou, Sean, Yong Rui, Huang, Thomas S., Exploration of Visual Data, Springer Publications, 2003
5. Claus.O.Wlike, Fundamentals of Data Visualization, A primer on making informative and compelling Figures, O'Reily Publications, 2019
6. Learn R for Applied Statistics: With Data Visualizations, Regressions, and Statistics by Eric Goh Ming Hui, Apress

COURSE OUTCOMES

At the end of the course, the student should be able to:

CO1 Recall the fundamental concepts of Data Exploration.
CO2 Analyze Univariate and Multivariate data for effective exploration.
CO3 Explain different methods of Data Visualization.
CO4 Illustrate the integration process of Processing with Java.
CO5 Create data visualizations using R language on different datasets.
CO6 Evaluate various datasets using Data Visualization techniques.

CO-PO-PSO MAPPING

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PSO1	PSO2	PSO3
CO1	3	1	3	3	-	-	-	2	3	3	3	2	2	2
CO2	2	2	2	1	1	-	-	3	2	3	1	2	3	2
CO3	2	1	2	1	1	-	-	3	2	1	2	2	2	3
CO4	2	2	2	1	-	-	-	1	2	1	3	2	1	3
CO5	3	1	1	2	1	-	-	3	2	1	2	2	3	3
CO6	2	1	2	2	1	-	-	2	2	2	2	2	3	3

Internal Assessment				End Semester Examinations
Assessment I (100 Marks)		**Assessment II (100 Marks)**		
Individual Assignment / Case Study / Seminar / Mini Project	**Written Test**	**Individual Assignment / Case Study / Seminar / Mini Project**	**Written Test**	**Written Examinations**
40	60	40	60	100
40%				**60 %**

23AD4005	PREDICTIVE MODELLING	L	T	P	C
		3	0	0	3

COURSE OBJECTIVES

- To explain terminology, technology and applications of predictive analysis
- To apply data preparation techniques and generate appropriate association rules.
- To discuss various descriptive models, their merits, demerits and application.
- To describe various predictive modelling methods
- To introduce the text mining tools, technologies and case study which is used in day-today analytics cycle
- To learn about advanced text visualization techniques

UNIT - I   INTRODUCTION TO PREDICTIVE ANALYTICS   9

Overview of Predictive Analytics - Setting up the problem - Data understanding - Single variable and multi-dimensional visualization - Statistical significance - Data audit -**Case Study: Churn prevention**.

UNIT - II   DATA PREPARATION AND ASSOCIATION RULES   9

Data preparation – Variable cleaning – Feature creation – Association rules: item sets, parameters, data structure - Interestingness measures - Deploying rules - Classification from association rules -**Case Study: Hospital readmission**.

UNIT - III   MODELLING   9

Descriptive modeling - Principal Component Analysis - Clustering algorithms - Interpretation of descriptive models – Standard cluster model.

UNIT - IV   PREDICTIVE MODELLING   9

Decision trees - Logistic regression - Neural networks - K-Nearest Neighbours - Naive Bayes - Regression models: Linear and others -**Case Study: Predictive web analytics**.

UNIT - V   TEXT MINING   9

Structured vs unstructured data - Predictive modeling for text - Text mining challenges - Data preparation steps – Regular expressions – Modeling with text features – **Case Study: Survey analysis**.

TOTAL: 45 PERIODS

COURSE OUTCOMES

CO1 Explain terminology, technology and applications of predictive analysis

CO2 Apply data preparation techniques to effectively interpret big data

CO3 Discuss various descriptive models, their merits, demerits and application.

CO4 Describe principles of predictive analytics and apply them to achieve real, pragmatic solutions.

CO5 Illustrate the features and applications of text mining.

CO6 Apply advanced text visualization techniques.

TEXT BOOKS

1. Dean Abbott, "Applied Predictive Analytics-Principles and Techniques for the Professional Data Analyst", Wiley, 2014
2. Jiawei Han and Micheline Kamber, Data Mining Concepts and Techniques, Third Edition, Elsevier, 2012

REFERENCES

1. Conrad Carlberg, "Predictive Analytics: Microsoft Excel", 1st Edition, Que Publishing, 2012.
2. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. An Introduction to Statistical Learning with Applications in R Springer 2013
3. Alberto Cordoba, "Understanding the Predictive Analytics Lifecycle", Wiley, 2014

CO-PO-PSO MAPPING

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PSO1	PSO2	PSO3
CO1	3	2	2	2	2	1	-	-	1	1	2	2	2	2
CO2	3	3	2	2	3	2	-	-	1	-	3	2	3	3
CO3	3	3	2	2	3	2	-	-	1	-	3	2	3	2
CO4	3	3	3	3	3	2	-	-	2	2	3	3	3	3
CO5	3	2	2	2	3	1	-	-	2	-	2	2	2	3
CO6	2	2	2	2	3	2	-	-	3	2	3	2	3	3

Internal Assessment				End Semester Examinations
Assessment I (100 Marks)		Assessment II (100 Marks)		
Individual Assignment / Case Study / Seminar / Mini Project	Written Test	Individual Assignment / Case Study / Seminar / Mini Project	Written Test	Written Examinations
40	60	40	60	100
40%				60 %

23AD4006	IMAGE AND VIDEO ANALYTICS	L	T	P	C
		3	0	0	3

COURSE OBJECTIVES

- To understand the basics of image processing techniques for computer vision
- To learn the techniques used for image pre-processing.
- To discuss the various object detection techniques
- To understand the various Object recognition mechanisms.
- To elaborate on the video analytics techniques.
- To implement real time applications.

UNIT - I   **INTRODUCTION**   **9**

Computer Vision - Image analysis tasks - Digitization and Color Images - Data structures for image analysis - Levels of image representation - Traditional and hierarchical structures - T-pyramid – Quad tree representation using intensity homogeneity.

UNIT - II   **IMAGE PRE-PROCESSING**   **9**

Local pre-processing – Image smoothing – Edge detection: Zero-crossings, Canny edge detection – Parametric edge models – Frequency domain preprocessing – Line detection – Image restoration - Geometric transformations - **Case study: MNIST dataset**.

UNIT - III   **OBJECT DETECTION USING MACHINE LEARNING**   **9**

Object detection methods – Deep Learning framework: R-CNN, Faster R-CNN, YOLO – Bounding box & IoU - Salient features - Loss functions - YOLO architecture - Motion analysis with moving edges - **Case study: Geospatial object detection**.

UNIT - IV   **FACE RECOGNITION AND GESTURE RECOGNITION**   **9**

Face Recognition: Applications, process - DeepFace, FaceNet - Implementation using FaceNet – Gesture Recognition – Static hand gestures – **Implementation of facial detection and recognition**.

UNIT - V   **VIDEO ANALYTICS**   **9**

Video processing – Use cases – Vanishing/exploding gradients – ResNet architecture, skip connections - Inception & GoogleNet - Improvements in Inception v2 and v3 - **Case study: Event detection in airport video surveillance**.

TOTAL: 45 PERIODS

COURSE OUTCOMES

CO1 Understand the basics of image processing techniques for computer vision and video analysis.

CO2 Explain the techniques used for image pre-processing.

CO3 Develop various object detection techniques

CO4 Understand the various face recognition mechanisms

CO5 Elaborate on deep learning-based video analytics.

CO6 Implement in real time applications.

TEXT BOOKS

1. Milan Sonka, Vaclav Hlavac, Roger Boyle, "Image Processing, Analysis, and Machine
   Vision", 4nd edition, Thomson Learning, 2013
2. Vaibhav Verdhan,(2021, Computer Vision Using Deep Learning Neural Network
   Architectures with Python and Keras,Apress 2021(UNIT-III,IV and V)

REFERENCES

1. Richard Szeliski, "Computer Vision: Algorithms and Applications", Springer Verlag London
   Limited,2011
2. Caifeng Shan, FatihPorikli, Tao Xiang, Shaogang Gong, "Video Analytics for Business
   Intelligence", Springer, 2012
3. D. A. Forsyth, J. Ponce, "Computer Vision: A Modern Approach", Pearson Education, 2003

CO-PO-PSO MAPPING

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PSO1	PSO2	PSO3
CO1	3	2	2	2	3	1	-	-	-	-	3	2	2	2
CO2	3	2	2	2	3	-	-	-	-	-	2	2	2	3
CO3	3	3	3	2	3	2	-	-	1	2	3	3	2	3
CO4	3	2	2	2	3	2	1	-	1	-	3	3	2	3
CO5	3	3	3	3	3	2	-	-	2	2	3	3	3	3
CO6	3	3	3	3	3	3	-	2	2	3	3	3	3	3

Internal Assessment				End Semester Examinations
Assessment I (100 Marks)		Assessment II (100 Marks)		
Individual Assignment / Case Study / Seminar / Mini Project	Written Test	Individual Assignment / Case Study / Seminar / Mini Project	Written Test	Written Examinations
40	60	40	60	100
40%				60 %

23AD4007	TEXT ANALYTICS	L	T	P	C
		3	0	0	3

COURSE OBJECTIVES

- To understand the methods for keyword extraction from documents
- To learn clustering methods for grouping of documents
- To explore the methods for classification of documents and E-mails
- To explore text visualization techniques and anomaly detection.
- To learn about Events and trends in text streams
- To learn about advanced text visualization techniques

UNIT – I   **TEXT EXTRACTION**   **9**

Introduction - Keyword extraction techniques: candidate keywords, keyword scores, adjoining and extracted keywords - Benchmark evaluation: precision, recall, efficiency - Stoplist generation - Intelligent text extraction.

UNIT – II   **DOCUMENT CLUSTERING**   **9**

Multilingual document clustering: LSA, Tucker1, PARAFAC2, LMSA with/without term alignments - Constrained clustering with k-means – Comparison: Document Clustering vs Topic Models.

UNIT – III   **CONTENT BASED CLASSIFICATION**   **9**

Document classification algorithms – Spam email classification – Email classification using Nonnegative Matrix Factorization – Content-based SMS classification.

UNIT - IV   **ANOMALY AND TREND DETECTION**   **9**

Text visualization: Tag clouds, authorship tracking, sentiment tracking – Visual analytics: FutureLens, scenario discovery – Anomaly detection with adaptive thresholds – Experimental insights.

UNIT - V   **TEXT STREAMS**   **9**

Text streams: Feature extraction, event and trend detection, semantic descriptions – Topic Modeling : LSA, pLSA, LDA – Semantic embedding from Wikipedia – Dynamic sampling in text analysis.

TOTAL: 45 PERIODS

COURSE OUTCOMES

CO1	Design text extraction techniques.
CO2	To apply clustering methods for grouping of documents.
CO3	Design classification techniques for text mining
CO4	Apply visualization techniques and perform anomaly & trend detection.
CO5	Perform Event operations in Text streams
CO6	Apply advanced text visualization techniques.

TEXT BOOKS

1. Michael W. Berry & Jacob Kogan,"Text Mining Applications and Theory" Wiley publications, 2010.
2. Aggarwal, Charu C., and ChengXiangZhai, eds., "Mining text data", Springer Science & Business Media, 2012.

REFERENCES

1. Gary Miner, John Elder, Thomas Hill, Dursun Deller, Andrew Fast, Robert A. Nisbet, "Practical text mining and statistical analysis for non-structured text data applications",Academic Press, 2012.
2. Srivastava, Ashok N., and MehranSahami, "Text mining: Classification, clustering, and applications", Chapman and Hall/CRC, 2009.
3. Buitelaar, Paul, Philipp Cimiano, and Bernardo Magnini, eds., "Ontology learning from text: methods, evaluation and applications", Vol. 123. IOS press, 2005.

CO-PO-PSO MAPPING

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PSO1	PSO2	PSO3
CO1	3	2	3	2	3	-	-	-	-	-	2	2	2	3
CO2	3	3	2	2	3	-	-	-	-	-	2	2	3	3
CO3	3	3	3	2	3	-	-	-	-	-	2	2	3	3
CO4	3	2	2	2	3	1	-	-	2	1	3	2	3	3
CO5	3	2	2	2	3	-	-	2	2	2	3	3	2	3
CO6	3	2	2	2	3	2	-	-	2	1	3	2	3	3

Internal Assessment				End Semester Examinations
Assessment I (100 Marks)		Assessment II (100 Marks)		
Individual Assignment / Case Study / Seminar / Mini Project	Written Test	Individual Assignment / Case Study / Seminar / Mini Project	Written Test	Written Examinations
40	60	40	60	100
40%				60 %

23AD4008	BUSINESS ANALYTICS	L	T	P	C
		3	0	0	3

COURSE OBJECTIVES
- To understand the Analytics Life Cycle.
- To understand various types of analytics for Business Forecasting
- To model the supply chain management for Analytics.
- To apply analytics for different functions of a business
- To apply visualization tools for Business Analytics

UNIT- I   **INTRODUCTION TO BUSINESS ANALYTICS**   **9**

Analytics and Data Science – Analytics Life Cycle – Types of Analytics – Business Problem Definition - Data Collection & Preparation - Hypothesis Generation - Modeling - Validating and Presenting Results - Interpretation - Deployment and Iteration.

UNIT-II   **BUSINESS INTELLIGENCE**   **9**

Data Warehouses and Marts - Knowledge Management - Types of Decisions - Decision Support Systems - Business Intelligence - OLAP - Analytic Functions.

UNIT- III   **BUSINESS FORECASTING**   **9**

Business Forecasting & Predictive Analytics – Data Mining Models: Linear Regression, Clustering, CART, Neural Networks – Data Visualization: Bar, Pie, Line, Scatter, Map, Bubble, Box & Whisker, Tree map, Heat map – Dashboard and Storyboard creation.

UNIT- IV   **HR & SUPPLY CHAIN ANALYTICS**   **9**

HR: Planning, Recruitment, Training & Development - Supply Chain: Demand, Inventory, Supply, Logistics – Applications of Analytics in HR and Supply Chain – Predicting demand for hourly employees.

UNIT- V   **MARKETING AND SALES ANALYTICS**   **9**

Marketing Strategy - Marketing Mix - Customer Behavior - Sales Planning - Selling Process - Applications of Analytics in Marketing and Sales.

                                                                    TOTAL   : 45 PERIODS

COURSE OUTCOMES
Upon successful completion of the course, students will be able to:

CO1 Understand real world business problems and model with analytical solutions

CO2 Illustrate business processes for extracting Business Intelligence

CO3 Apply predictive analytics for business fore-casting

CO4 Analyze analytics for supply chain and logistics management

CO5 Evaluate analytics for marketing and sales

CO6 Use Analytics for Visualization Techniques

REFERENCES

1. R. Evans James, Business Analytics, 2017
2. R N Prasad , Seema Acharya , Fundamentals of Business Analytics
3. Philip Kotler and Kevin Keller, Marketing Management, 15th edition, PHI, 2016
4. VSP RAO, Human Resource Management, 3rd Edition, Excel Books, 2010
5. Mahadevan B, "Operations Management -Theory and Practice",3rd Edition, Pearson Education, 2018.

CO-PO-PSO MAPPING

COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PSO1	PSO2	PSO3
CO1	3	3	2	2	3	2	-	-	1	1	3	3	3	2
CO2	2	2	3	2	3	1	-	-	2	2	2	3	3	2
CO3	3	3	3	3	3	2	-	-	1	2	3	3	3	2
CO4	3	3	2	2	3	3	-	-	2	2	2	3	3	2
CO5	2	2	2	2	3	2	-	-	2	2	2	3	3	2
CO6	2	2	2	2	3	2	-	-	3	1	3	2	3	3

Internal Assessment				End Semester Examinations
Assessment I (100 Marks)		Assessment II (100 Marks)		
Individual Assignment / Case Study / Seminar / Mini Project	Written Test	Individual Assignment / Case Study / Seminar / Mini Project	Written Test	Written Examinations
40	60	40	60	100
40%				60 %