Francis Xavier Engineering College

(An Autonomous Institution)
Tirunelveli 627 003
Department of Computer Science and Business Systems

(Specialization in Big data Analytics)

Curriculum and Syllabi – R 2021-UG

CHOICE BASED CREDIT SYSTEM AND OBE

Vision of the Department

To become a center of excellence in Computer Technology and to generate young Engineers with enriched knowledge to serve industries with high values and social responsibilities..

Mission of the Department

- To provide world class teaching learning environment and to offer computing education programs.
- To inculcate varied skill sets that meets global industry standards and to practice moral values.
- To enrich moral and ethical values to lead and serve the society.

Programme Educational Outcomes (PEOs)

PEO1**:** To apply problem solving skills in Computer science and Business Management by applying Engineering fundamentals.
PEO2: To improve communication skills, business management skills, professional ethics, team work and to innovate technologies for the betterment of society.
PEO3: Toexhibit leadership qualities, interpersonal skills and adapting to a rapidly changing environment by applying knowledge in technology abstraction and common business principles.
PEO4: To develop professional and ethical attitude, effective communication skills, moral values and an ability to relate engineering issues to social welfare in contemporary areas in Computer Science and Business systems.

Programme Specific Objectives (PSOs)

PSO1: Enriched knowledge in Business Management and human ethics.
PSO2: The students will have effective knowledge in software engineering principles and solving scientific and business problems.
PSO3: The students will explore **e**merging technologies in Information and Communication Technologies (ICT), Business Analytics and Machine Learning to innovate ideas and solutions to existing/novel Business applications.

Programme Outcomes(POs)

Engineering Graduates will be able to:

1. Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

2. Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

3. Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

4. Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

5. Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

6. The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

7. Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

8. Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

9. Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

10. Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

11. Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

12. Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Mapping with PO Vs PEO, PSO

PO	PEO1	PEO2	PEO3	PEO4
1	3			3
2	3			3
3	3			2
4	3			2
5	2			3
6	3	3	3	3
7	3	3	3	3
8	3	3	3	3
9	3	3	3	3
10	3	3	3	3
11	3	3	3	3
12	3	3	3	3
PSO1	3	3	3	3
PSO2	3	2	3	3
PSO3	3	3	2	3

Summary

S.No	Course Code	Course Name	L	T	P	C	H
Theory Course							
1	21CB4S01	Big Data Frameworks	3	0	0	3	3
	21CB5S0X	Elective	3	0	0	3	3
Theory cum Practical Courses							
1	21CB6S01	Mining Massive Data	2	0	4	4	4
2	21CB7S01	Big data computing for Business Analytics	2	0	4	4	4
Practical Course							
1.	21CB8S11	Project work	0	0	8	4	8
		Total	10	0	16	18	22
Elective Courses							
1	21CB5S01	Exploratory Data Analysis	3	0	0	3	3
2	21CB5S02	Information Visualization	3	0	0	3	3
3	21CB5S03	Predictive Analytics in Business	3	0	0	3	3

Syllabus

21CB4S01	BIG DATA FRAMEWORKS	L	T	P	C
		3	0	0	3

Prerequisites for the course

- Nil

Objectives

1. To understand the need of Big Data, challenges and different analytical architectures

2.Installation and understanding of Hadoop Architecture and its ecosystems

3. Processing of Big Data with Advanced architectures like spark.

4. Describe graphs and streaming data in Spark.

5.Explore data analysis to process BigData

UNIT I	INTRODUCTION TO BIG DATA	9

Data Storage and Analysis - Characteristics of Big Data – Big Data Analytics - Typical AnalyticalArchitecture – Requirement for new analytical architecture – Challenges in Big Data Analytics –Need of big data frameworks

UNIT II	HADOOP FRAMEWORK	9

Hadoop – Requirement of Hadoop Framework - Design principle of Hadoop –Comparison with other system - Hadoop Components – Hadoop 1 vs Hadoop 2 – Hadoop Daemon's – HDFS Commands –Map Reduce Programming: I/O formats, Map side join, Reduce Side Join, Secondary sorting

UNIT III	HADOOP ECOSYSTEM	9

Introduction to Hadoop ecosystem technologies: Serialization: AVRO, Co-ordination: Zookeeper, Databases: HBase, Hive, Scripting language: Pig, Streaming: Flink, Storm.

UNIT IV	SPARK FRAMEWORK	10

Overview of Spark – Hadoop vs Spark – Cluster Design – Cluster Management – performance,ApplicationProgramming interface (API): Spark Context, Resilient Distributed Datasets, Creating RDD, RDD Operations, Saving RDD - Lazy Operation – Spark Jobs.

UNIT V	DATA ANALYSIS WITH SPARK SHELL	8

Writing Spark Application - Spark Programming in Scala, Python, R, Java - Application Execution.SQL Context – Importing and Saving data – Data frames – using SQL – GraphX overview – Creating Graph .

Total Periods	45

Suggestive Assessment Methods

Continuous Assessment Test (30 Marks)	Formative Assessment Test (10 Marks)	End Semester Exams (60 Marks)
1. DESCRIPTIVE QUESTIONS	1.ASSIGNMENT 2. ONLINE QUIZZES	1.DESCRIPTIVE QUESTIONS

Outcomes

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

CO1 Discuss the challenges and their solutions in Big Data

CO2 Understand and work on Hadoop Framework and eco systems.

CO3 Analyse the Big Data using Map-reduce programming in both Hadoop and Spark framework.

CO4 Demonstrate spark programming with different programming languages.

CO5Demonstrate the graph algorithms and live streaming data in Spark

Text Books

1. Mike Frampton, "Mastering Apache Spark", Packt Publishing,2015.
2. TomWhite,"Hadoop:TheDefinitiveGuide",O'Reilly,4thEdition,2015.

Reference Books

1. Nick Pentreath, Machine Learning with Spark, Packt Publishing, 2015.
2. Mohammed Guller, Big Data Analytics with Spark, Apress,2015
3. Donald Miner, Adam Shook, "Map Reduce Design Pattern", O'Reilly, 2012

Web Resources

1. https://www.oracle.com/big-data
2. https://nptel.ac.in/courses/106104189
3. https://www.javatpoint.com › java-big-data-frameworks

CO Vs PO Mapping and CO Vs PSO Mapping

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
1	3	3				1							3	3	1
2	3				2	1			1	1		1	3	1	1
3	3	2	1		2				1			1	3	3	2
4	3	2		1	1		1		1	1		1	3	3	3

5	3	2		1	1		1		1		1	1	2	2	1

BLOOMS LEVEL ASSESSMENT PATTERN

BLOOMS CATEGORY	CAT 1	CAT 2	FAT 1	FAT 2	END SEM EXAM
REMEMBER	20	10	5	5	10
UNDERSTAND	40	20	10	10	20
APPLY	40	50	5	5	50
ANALYZE		20	5	5	20
EVALUATE					
CREATE					

COURSE LEVEL ASSESSMENT QUESTIONS

COURSE OUTCOME 1: Students will be able to Predict the suitable method for...(Apply)

Course Outcome 1 (CO1):

1. Describe the Characteristics of Big Data. (Understand)
2. What is the significance of big data frameworks? (Understand)
3. Summarize the Challenges in Big Data Analytics. (Evaluate)

Course Outcome 2 (CO2):

1. Discuss the functions of Hadoop Components. (Understand)
2. Analyze the uses of HDFS Commands. (Analysis)

Course Outcome 3 (CO3):

1. Write elaborately on Hadoop ecosystem technologies. (Understand)
2. Identify Scripting languages used for Hadoop ecosystem technologies. (Remember)

Course Outcome 4 (CO4):

1. Analyse on Hadoop vs Spark. (Analyse)
2. List out the Datasets used for Spark. (Remember)

3. Give some steps in Lazy Operation. (Understand)

Course Outcome 5 (CO5):

1. Write a simple Spark Application. (create)
2. How do you create a graph using spark shell? (Understand)

Theory cum Practical Courses

21CB6S01	MINING MASSIVE DATA	L	T	P	C
		2	0	4	4

Pre requisites for the course

NIL

Objectives

1. To provide comprehensive knowledge on developing
2. To apply machine learning algorithms for massive real-world datasets in distributed frameworks.
3. To demonstrate the use of big data analytics tools like Spark and Mahout for mining massive datasets.
4. To impart in depth knowledge on Deep Learning and Extreme Learning concepts

UNIT I	MapReduce Based Machine Learning	7

K-Means, PLANET, Parallel SVM, Association Rule Mining in MapReduce, Inverted Index, Page Ranking, Expectation Maximization, Bayesian Networks

UNIT II	Classification and Regression models	5

linear support vector machines - Naive Bayes model- Decision Trees – Least square regression Decision trees for regression.

UNIT III	Clustering in Spark and Mahout	6

Hierarchical Clustering in a Euclidean and Non-Euclidean Space - The Algorithm of Bradley, Fayyad, and Reina - Processing Data in BFR Algorithm CURE algorithm - Clustering models with Spark - Spectral clustering using Mahout

UNIT IV	Mining Social-Network Graphs	6

Clustering of Social-Network Graphs - Direct Discovery of Communities - Partitioning of Graphs Finding Overlapping Communities - Counting Triangles using MapReduce Neighborhood Properties of Graphs

UNIT V	Semi-Supervised Learning , Deep Learning	6

Introduction to Semi-Supervised Learning, Semi-Supervised Clustering, Transductive Support Vector Machines, Deep Neural Networks, Deep Belief Networks

S.No	List of Experiments	CO

1	K-means implementation in MapReduce	CO1
2	Association Rule Mining with MapReduce	CO1
3	Decision trees in Spark	CO2
4	Naive bayes classification using Spark	CO2
5	Advanced text processing with Spark	CO3
6	Representing social-network data using Graphs	CO4
7	Implementing Semi-supervised Clustering	CO4
8	Predictive analysis using H2O tool	CO4
9	SVM Classification using Mahout	C05
10	Building a recommendation engine with Sparkling water	CO5
	Total Periods	**30Theory+30Lab**

Laboratory Requirements

- 60 Systems with windows/LINUX operating system with Hadoop, Mahout, Spark and H2O tool.

Suggestive Assessment

Continuous AssessmentTest (30Marks)	**Lab Components Assessments (20Marks)**	**End Semester Exams (50Marks)**
1. DESCRIPTIVE QUESTIONS	1. LAB EXPERIMENTS 2. MODEL EXAMINATION	**1. DESCRIPTIVE QUESTIONS**

Outcomes

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

CO1. Identify right machine learning / mining algorithm for handling massive data
CO2. Apply classification and regression models with Spark and Mahout
CO3. Implement clustering models using Spark and Mahout
CO4. Mine social Network graphs using MapReduce
CO5. Apply semi supervised learning for clustering and classification

TextBooks

1. Joao Gama, "Knowledge Discovery from Data Streams", CRC Press, 2010.
2. David Luckham, "The Power of Events: An Introduction to Complex Event Processing in Distributed Enterprise Systems", Addison Wesley, 2002.
3. Charu C. Aggarwal, "Data Streams: Models And Algorithms", Kluwer Academic Publishers, 2007.

ReferenceBooks

1. GuidovanRossum,FredL.DrakeJr.,"AnIntroductiontoPython–RevisedandUpdatedforPython3.2", NetworkTheoryLtd., 2011.
2. JohnVGuttag,"IntroductiontoComputationandProgrammingUsingPython",RevisedandExpandedEdition,MITPress , 2013
3. CharlesDierbach,"IntroductiontoComputerScienceusingPython",WileyIndiaEdition,2016.
4. TimothyA.Budd,"ExploringPython",Mc-GrawHillEducation(India)PrivateLtd.,2015.
5. KennethA.Lambert,"FundamentalsofPython:FirstPrograms",CengageLearning,2012.

WebResources

1.https://nptel.ac.in/courses/106/106/106106182/

CO Vs PO Mapping and CO Vs PSO Mapping

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
1	3	2	3	2	3	1	3	1	1			2	2		
2		3	3	3	2	3				3	3	2	2	1	2
3	3	2	1	2	1	1	2	1	2			1	1	1	1
4	2	3	2	3	2	2	2	1		3	3	2			1
5	1	3	2				2	3	2	3	2	1			2

BLOOMS LEVEL ASSESSMENT PATTERN

BLOOMS CATEGORY	CAT 1	CAT 2	FAT 1	FAT 2	END SEM EXAM
REMEMBER	20	10	5	5	10
UNDERSTAND	40	20	10	10	20
APPLY	40	50	5	5	50
ANALYZE		20	5	5	20
EVALUATE					
CREATE					

COURSE LEVEL ASSESSMENT QUESTIONS

COURSEOUTCOME 1:

1.Define data mining.

2.what are the different tasks of data?

COURSEOUTCOME2:

1. Explain linear support vector machines
2. Describe Naive Bayes model
3. Discuss about Decision Trees

COURSEOUTCOME3:

1. Differentiate Hierarchical Clustering in a Euclidean and Non-Euclidean Space
2. Implement a variant of K-means algorithm
3. Demonstrate a Processing Data in BFR

COURSEOUTCOME4:

1. Examine Clustering of Social-Network Graphs
2. Test the Overlapping Communities
3. Compare the properties of Graphs.

COURSEOUTCOME5:

1. Design Semi-Supervised Learning,
2. Construct Semi-Supervised Clustering,
3. Develop a Deep Neural Networks

21CB7S01	**BIG DATA COMPUTING FOR BUSINESS ANALYTICS**	**L**	**T**	**P**	**C**
		2	0	4	4

Prerequisites for the course

- **NIL**

Objectives

1. Providing fundamental concepts and significance of big data analytics
2. To know how organizations can leverage information to gain competitive advantage
3. Providing an understanding of the application of Big data analytics methods and techniques
4. To address strategic business problems

UNIT I	**INTRODUCTION TO BIG DATA ANALYTICS**	**4**

Big Data Overview – Characteristics of Big Data –Business Intelligence v/s Data Analytics – Need of Data Analytics – Data Analytics in Industries – Role of the Data Scientist – Data Analytics Life Cycle– Main phases of the lifecycle

UNIT II	**PREDICTIVE AND DESCRIPTIVE ANALYTICS WITH BIG DATA**	**6**

Linear Regression – Logistic Regression – Decision Trees – Support Vector Machines – Ensemble Methods – Multi-class Classification Techniques – Evaluating Predictive Models- Association Rules – Sequence Rules – Segmentation – Visualization Charts		
UNIT III	**BATCH ANALYSIS, REAL-TIME ANALYSIS AND SOCIAL NETWORK ANALYTICS**	**10**
Batch Analysis –with Hadoop MapReduce – Sensor Data – New articles – Real-time analysis with Streaming – Sensor data and social media data -Social Network Metrics – Social Network Learning – Relational Neighbour Classifier –Collective Inferencing – Egonets - Bigraphs.		
UNIT IV	**GRAPH ANALYTICS FOR BIG DATA**	**6**
What is a Graph?- Why Graphs?-What are the impact of Big Data's V's on Graphs?- Focusing on Graph Analytics Techniques- Path Analytics-Applying Dijkstra's Algorithm- Inclusion and Exclusion Constraints- Connectivity Analytics- Disconnecting a Graph- Use cases and Case studies		
UNIT V	**COMMUNITY AND CENTRALITY ANALYTICS FOR BIG DATA**	**4**
Community Analytics and Local Properties- Global Property: Modularity- Centrality Analytics.		
S.No	**List of Experiments**	**CO**
1	Setting up Hadoop environment and Hadoop cluster	CO1
2	Working with Hadoop, spark	CO2
3	Implementation of Machine learning, algorithms using graph analytics.	CO3
4	Mapreduce Programs in Hadoop Environment	CO4
5	Design, Develop and implement Machine Learning algorithms in Big Data environment using SPARK architecture	CO5
6	Design, Develop and implement Graph analytics algorithms using GraphX in SPARK architecture	CO4
Total Periods		**30 Theory +30 Lab**
Laboratory Requirements		
• Hadoop		
Suggestive Assessment Methods		

Continuous Assessment Test (30Marks)	Lab Components Assessments (10 Marks)	End Semester Exams (50 Marks)
1. DESCRIPTIVE QUESTIONS	1. LAB EXPERIMENTS 2. MODEL EXAMINATION	**1. DESCRIPTIVE QUESTIONS**

Outcomes

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

CO 1 Assess the role of big data analytics within an organization and the challenges
CO 2 Apply Big data analytics methods and techniques in addressing strategic business problems
CO 3 Acquire an understanding of machine learning algorithms and how it can be applied in addressing strategic business problems
CO 4 Acquire an understanding of graph analytics in the context of big data
CO 5 Use Hadoop, spark architecture, machine learning, graph analytics and other big data tools for the model development and interpreting the outputs

Text Books

1. Thomas Davenport et.al, (2010), Analytics at Work: Smarter Decisions, Better Results , 3rd edition, Harvard Business School Press, Boston, Massachusetts.
2. Zikopoulos P, Eaton C, (2011),Understanding big data: Analytics for enterprise class Hadoop and streaming data, McGraw-Hill Osborne Media.
3. Viktor Mayer-Schönberger, Kenneth Cukier (2014),Big Data: A Revolution That Will Transform How We Live, Work, and Think, Mariner Books

Reference Books

1. Pramod J. Sadalage, Martin Fowler, (2012), NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence, Addison-Wesley.
2. Sammer E, (2012),Hadoop Operations, 1st edition, O'Reilly Media, Inc.
3. Marz N, Warren J, (2015), Big Data: Principles and best practices of scalable real-time data systems, Manning Publications Co.
4. Miner D, Shook A, (2012),MapReduce Design Patterns: Building Effective Algorithms and Analytics for Hadoop and Other Systems, O'Reilly Media, Inc.
5. Rajaraman A, Ullman J. D, (2014),Mining of massive datasets, Cambridge: Cambridge University Press.

Web Resources

1. https://www.iare.ac.in/sites/default/files/NEW%20LECHURE%20NOTES.pdf
2. https://mrcet.com/downloads/digital_notes/CSE/IV%20Year/(R17A0528%20)%20Big%20Data%20Analytics%20Digital%20notes.pdf
3. https://www.aalimec.ac.in/wp-content/uploads/2020/01/CS8091-BIGDATA-ANALYTICS-
4. https://www.iare.ac.in/sites/default/files/lecture_notes/BDBA-

Question%20bank.pdf

CO Vs PO Mapping and CO Vs PSO Mapping

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
1	3	2	3	2	3	1	3	1	1			2	2		
2		3	3	3	2	3				3	3	2	2	1	2
3	3	2	1	2	1	1	2	1	2			1	1	1	1
4	2	3	2	3	2	2	2	1		3	3	2			1
5	1	3	2				2	3	2	3	2	1			2

BLOOMS LEVEL ASSESSMENT PATTERN

BLOOMS CATEGORY	CAT 1	CAT 2	FAT 1	FAT 2	END SEM EXAM
REMEMBER	20	10	5	5	10
UNDERSTAND	40	20	10	10	20
APPLY	40	50	5	5	50
ANALYZE		20	5	5	20
EVALUATE					
CREATE					

COURSE LEVEL ASSESSMENT QUESTIONS

Course Outcome 1 (CO1):

1. What are the various applications of big data analytics? (Understand)
2. Enumerate the terms a. OLAP b. OLTP c. RTAP(Understand)
3. Define streaming data? (Remember)

Course Outcome 2 (CO2):

1. Express the term bucketing data?(Understand)
2. Discuss Why Hadoop came into an existence in processing big data?(Apply)

3. Implement the processing data with Hadoop? (Apply)

Course Outcome 3 (CO3):

1. List out the basic Filesystem Operations?(Understand)
2. Implement the Master-Slave architecture?(Analyse)
3. Extrapolate the Master components: Name node, Secondary Node and JobTracker? (Create)

Course Outcome 4 (CO4):

1. How to explore the Scale-out architecture? (Analyse)
2. Design Reducer Phase?(Create)
3. Can MapReduce be used to solve any kind of computational problems? if not, explain the cases where MapReduce is not applicable?(Evaluate)

Course Outcome 5 (CO5):

1. Discuss the use of the FOREACH and ASSERT operator in Pig Latin?(Evaluate)
2. Write a shell command in Hive to list all the files in the current directory? (Create)

Electives:

21CB5S01	**Exploratory Data Analysis**	**L**	**T**	**P**	**C**
		3	**0**	**0**	**3**
Prerequisites for the course					
NIL					
Objectives					
1. To introduce the methods for data preparation and data understanding. 2.Covers essential exploratory techniques for understanding multivariate data by summarizing it through statistical methods and graphical methods. 3.To Summarize the insurers use of predictive analytics, data science and DataVisualization 4.Know about outlier analysis.					
UNIT I	**Introduction To Exploratory Data Analysis**				**9**
Data Analytics lifecycle, Exploratory Data Analysis (EDA)– Definition, Motivation, Steps in data exploration, The basic data types Data Type Portability					
UNIT II	**Preprocessing-Traditional Methods andMaximum Likelihood Estimation**				**9**

Introduction to Missing data, Traditional methods for dealing with missing data, Maximum Likelihood Estimation – Basics, Missing data handling, Improving the accuracy of analysis		
UNIT III	**Preprocessing Bayesian Estimation**	**9**
Introduction to Bayesian Estimation ,Multiple Imputation-Imputation Phase, Analysis and Pooling Phase,Practical Issues in Multiple Imputation, Models for Missing Notation Random Data		
UNIT IV	**Data Summarization & Visualization**	**10**
Statistical data elaboration, 1-D Statistical data analysis, 2-D Statistical data Analysis, N- D Statistical data analysis		
UNIT V	**Outlier Analysis**	**8**
Introduction, Extreme Value Analysis, Clustering based, Distance Based and Density Based outlier analysis, Outlier Detection in Categorical Data		
Total Periods		**45**

Suggestive Assessment Methods

Continuous Assessment Test (30 Marks)	**Formative Assessment Test (10 Marks)**	**End Semester Exams (60 Marks)**
1. DESCRIPTIVE QUESTIONS	1.ASSIGNMENT 2. ONLINE QUIZZES	1.DESCRIPTIVE QUESTIONS

Outcomes

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

CO1 Handle missing data in the realworld data sets by choosing appropriate methods.

CO2 Summarize the data using basic statistics. Visualize the data using basic graphs and plots.

CO3 Identify the outliers if any in the data set.

CO4Choose appropriate feature selection and dimensionality reduction

CO5 Techniques for handling multi-dimensional data

Text Books

1. Roger S.Pressman, “Software Engineering: A Practitioner's Approach”, 8th Edition, Tata McGraw Hill Edition, 2015.

Reference Books

1. Charu C. Aggarwal, "Data Mining The Text book", Springer, 2015.
2. Craig K. Enders, "Applied Missing Data Analysis", The Guilford Press, 2010.
3. Inge Koch, "Analysis of Multivariate and High dimensional data", Cambridge University Press, 2014.
4. Michael Jambu, "Exploratory and multivariate data analysis", Academic Press Inc. 1990.
5. Charu C. Aggarwal, "Data Classification Algorithms and Applications", CRC press, 2015

Web Resources

1. https://www.tutorialspoint.com/software_engineering/index.htm
2. https://nptel.ac.in/courses/106/105/106105182/
3. https://www.javatpoint.com/software-engineering-tutorial
4. www.mhhe.com/pressman

CO Vs PO Mapping and CO Vs PSO Mapping

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
1	3	3				1							3	3	1
2	3				2	1			1	1		1	3	1	1
3	3	2	1		2				1			1	3	3	2
4	3	2		1	1		1		1	1		1	3	3	3
5	3	2		1	1		1		1		1	1	2	2	1

BLOOMS LEVEL ASSESSMENT PATTERN

BLOOMS CATEGORY	CAT 1	CAT 2	FAT 1	FAT 2	END SEM EXAM
REMEMBER	20	10	5	5	10
UNDERSTAND	40	20	10	10	20
APPLY	40	50	5	5	50
ANALYZE		20	5	5	20
EVALUATE					
CREATE					

COURSE LEVEL ASSESSMENT QUESTIONS

COURSE OUTCOME 1: Students will be able to Predict the suitable method for...(Apply)

Course Outcome 1 (CO1):

1. Define Data Analytics lifecycle
2. Describe basic data types

Course Outcome 2 (CO2):

1. Explain Missing data,
2. Discuss Traditional methods for dealing with missing data
3. Summarize Maximum Likelihood Estimation

Course Outcome 3 (CO3):

1.Analysis Phase,Practical Issues in Multiple Imputation

2.Draw Models for Missing Notation

3.Describe Random Data

Course Outcome 4 (CO4):

1. Compare 1-D Statistical data analysis, 2-D Statistical data Analysis
2. Examine N- D Statistical data analysis
3. Examine Statistical data elaboration

Course Outcome 5 (CO5):

1. Design Extreme Value Analysis
2. Plan a Clustering based
3. Design Distance Based and Density Based outlier

21CB5S02	INFORMATION VISUALIZATION	L	T	P	C
		3	0	0	3

Prerequisites for the course

- Nil

Objectives

1. To understand the various types of data, apply and evaluate the principles of data visualization.

2. Acquire skills to apply visualization techniques to a problem and its associated dataset.

3. To apply structured approach to create effective visualizations.

4. To learn how to bring valuable insight from the massive dataset using visualization.

5. To learn how to build visualization dashboard to support decision making.		
UNIT I	**Introduction to Data Visualization**	**9**
Overview of data visualization - Data Abstraction - Task Abstraction - Analysis: Four Levels for Validation, Human Visual Perception		
UNIT II	**Visualization Techniques – I**	**9**
Scalar and point techniques – vector visualization techniques – matrix visualization		
UNIT III	**Visualization Techniques – II**	**10**
Visualization Techniques for Trees, Graphs, and Networks, Multidimensional data		
UNIT IV	**Visual Analysis of data from various domains**	**10**
Time-oriented data visualization – Spatial data visualization and case studies Text data visualization – Multivariate data visualization, and case studies		
UNIT V	**Designing Effective Visualizations**	**7**
Guidelines for designing successful visualizations, Data visualization dos and don'ts		
Total Periods		**45**

Suggestive Assessment Methods		
Continuous Assessment Test (30 Marks)	**Formative Assessment Test (10 Marks)**	**End Semester Exams (60 Marks)**
1. DESCRIPTIVE QUESTIONS	1.ASSIGNMENT 2. ONLINE QUIZZES	1.DESCRIPTIVE QUESTIONS

Outcomes
Upon completion of the course, the students will be able to:

CO 1. Identify the data types and its associated visualization mechanisms.

CO2. Apply the various scalar and vector visualization techniques to create suitable visualization for real life applications.

CO 3. Handle and analyse multidimensional data and hierarchical data for visualization.

CO4. Perform multivariate data analysis and visualization.

CO5. Apply the visualization guidelines for effective information visualization.

Text Books

1. Matthew O. Ward, Georges Grinstein, Daniel Keim"Interactive Data Visualization: Foundations, Techniques, and Applications", CRC Press, Second Edition, 2015.

2.Dr.Chun-hauh Chen, W.K.Hardle, A. Unwin, "Handbook of Data Visualization",

Reference Books

1. Tamara Munzer, "Visualization Analysis and Design", CRC Press, 2014.

2. Stephen Few, "Now You See It", Analytics Press, 2009.

3. Stephen Few, "Information Dashboard Design: the effective visual communication of data", Oreilly, 2006.

4. Ben Fry, "Visualizing Data", O'Reilly Media, 2008

5ki. Winston Chang, "R Graphics Cookbook", O'Reilly, 2012.

Web Resources

http://www.fusioncharts.com/whitepapers/

CO Vs PO Mapping and CO Vs PSO Mapping

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
1	3	3				1							3	3	1
2	3				2	1			1	1		1	3	1	1
3	3	2	1		2				1			1	3	3	2
4	3	2		1	1		1		1	1		1	3	3	3
5	3	2		1	1		1		1		1	1	2	2	1

BLOOMS LEVEL ASSESSMENT PATTERN

BLOOMS CATEGORY	CAT 1	CAT 2	FAT 1	FAT 2	END SEM EXAM
REMEMBER	20	10	5	5	10
UNDERSTAND	40	20	10	10	20
APPLY	40	30	5	5	30
ANALYZE		20	5	5	20
EVALUATE		20			20
CREATE					

COURSE LEVEL ASSESSMENT QUESTIONS

COURSE OUTCOME 1: Students will be able to Predict the suitable method for.(Apply)

Course Outcome 1 (CO1):

1. Describe data visualization (create)
2. Define Data Abstraction (Remember)
3. List Four Levels for Validation(understand)

Course Outcome 2 (CO2):

1. Discuss Scalar and point techniques(Remember)
2. Analysematrix visualization(understand)

Course Outcome 3 (CO3):

1. Write elaborately visualization Techniques for Trees(Remember)
2. Write elaborately Multidimensional data. (understand)

Course Outcome 4 (CO4):

1. Analyse on Time-oriented data visualization(create)
2. List out the Spatial data visualization(Remember)
3. Give some steps toMultivariate data visualization (Understand)

Course Outcome 5 (CO5):

1. Write aGuidelines for designing successful visualizations. (create)
2. Write a Data visualization dos and don'ts? (Understand)

21CB5S03	PREDICTIVE ANALYTICS IN BUSINESS	L	T	P	C

		3	**0**	**0**	**3**

Prerequisites for the course

- NIL

Objectives

1. To introduce theoretical foundations, algorithms, methodologies
2. Study in Risk Management and Operational Hedging
3. Know about financial time series analytics
4. To analyze data in various domains such Retail, Risk and Healthcare.

UNIT I	**RETAIL ANALYTICS**	**9**

Understanding Customer: Profiling and Segmentation, Modelling Churn. Modelling Lifetime Value, Modelling Risk, Market Basket Analysis.

UNIT II	**RISK ANALYTICS**	**9**

Risk Management and Operational Hedging: An Overview, Supply Chain Risk Management, A Bayesian Framework for Supply Chain Risk Management, Credit Scoring and Bankruptcy Prediction

UNIT III	**FINANCIAL DATA ANALYTICS**	**9**

Financial News analytics: Framework, techniques, and metrics, News events impact market sentiment, Relating news analytics to stock returns

UNIT IV	**FINANCIAL TIME SERIES ANALYTICS**	**9**

Financial Time Series and Their Characteristics, Common Financial Time Series models, Autoregressive models, Markov chain models, Time series models with leading indicators, long term forecasting

UNIT V	**HEALTH CARE ANALYTICS**	**9**

Introduction to Healthcare Data Analytics, Electronic Health Records, Privacy-Preserving Data Publishing Methods in Healthcare, Clinical Decision Support Systems

Total Periods	**45**

Suggestive Assessment Methods

Continuous Assessment Test (30 Marks)	**Formative Assessment Test (10 Marks)**	**End Semester Exams (60 Marks)**
1. DESCRIPTIVE QUESTIONS	1.ASSIGNMENT 2. ONLINE QUIZZES	1.DESCRIPTIVE QUESTIONS

Course Outcomes

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

CO 1 Recognize challenges in dealing with data sets in domains such as finance, risk and healthcare.

CO 2 Identify real-world applications of machine learning in domains such as finance, risk and healthcare

CO 3 Identify and apply appropriate algorithms for analyzing the data for variety of problems in finance, risk and healthcare

CO 4 Make choices for a model for new machine learning tasks based on reasoned argument

Text Books

1. Chris Chapman, Elea McDonnell Feit "R for Marketing Research and Analytics", Springer, 2015.
2. Olivia Parr Rud "Data Mining Cookbook: Modeling Data for Marketing, Risk, and Customer Relationship Management", Wiley, 2001.

Reference Books

1. Chandan K. Reddy, Charu C. Aggarwal "Healthcare Data Analytics", CRC Press,2015.
2. Rene Carmona "Statistical Analysis of Financial Data in R", Springer, 2014.
3. James B. Ayers "Handbook of Supply Chain Management" Auerbach Publications, 2006.
4. PanosKouvelis, Ling xiu Dong, Onur Boyabatli, Rong Li "The Handbook of Integrated Risk Management in Global Supply Chains", Wiley, 2012.

Web Resources

1. **https://www.predictiveanalyticsworld.com/book/notes.php**
2. **https://www.slideshare.net/machinepulse/predictive-analytics-an-overview**
3. **https://nptel.ac.in/courses/110104086**
4. https://www.vskills.in/certification/big-data/predictive-analytics-certification

CO Vs PO Mapping and CO Vs PSO Mapping

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
1	2	2	3	1	1	2							3	1	
2	3	3	3	3				3	3	3	2		3		2
3	2	3	2			2	2	2	2					2	2
4	3	3	2	2	2					2	2	3		3	3

BLOOMS LEVEL ASSESSMENT PATTERN

BLOOMS CATEGORY	CAT 1	CAT 2	FAT 1	FAT 2	END SEM EXAM

REMEMBER	20	10	5	5	10
UNDERSTAND	40	20	10	10	20
APPLY	40	50	5	5	50
ANALYZE		20	5	5	20
EVALUATE					
CREATE					

COURSE LEVEL ASSESSMENT QUESTIONS

Course Outcome 1 (CO1):

1. What are the two most prominent open-source tools for predictive analytics?(Understand)
2. Which important measure do we gain by using PERT? (Analyse)

Course Outcome 2 (CO2):

1. What is the primary role of statistics in predictive analytics? (understand)
2. When dealing with a continuous variable, what is the appropriate statistics calculation? (apply)
3. When gathering data from noncentralized data, when should regular extraction activity take place? (anlyze)

Course Outcome 3 (CO3):

1. Explain - Financial News analytics (Understand)
2. How can you apply techniques, and metrics in financial news analytics? (apply)

Course Outcome 4 (CO4):

1. List the Characteristics of Financial Time Series. (Apply)
2. Compare and contrast Autoregressive models, Markov chain models (Analyse)
3. Construct long term forecasting model (create)

Course Outcome 5 (CO5):

1. Explain Healthcare Data Analytics(Remember)
2. List Privacy-Preserving Data Publishing Methods in Healthcare (apply)