

Branch Name:	Bachelor of Science (Information Technology)
Program Code:	ET110
Semester/Year:	Semester - V/ Third Year
Subject Title:	Data Analytics & Reporting
Subject Code:	1ET1100505
Pre-requisite:	Basic knowledge of programming and Data Structures Algorithms, Mathematics and Statistics

Course Objective:

To provide strong foundation for data analytics and application area related to it and understand the underlying core concepts and emerging technologies in data analytics.

Teaching and Examination Scheme:

Teaching Scheme (Hours per week)				Evaluation Scheme (Marks)				
Lecture (L)	Tutorial (T)	Practical (P)	Credit	Theory (Marks)		Practical (Marks)		Total (Marks)
				University Assessment	Continuous Assessment	University Assessment	Continuous Assessment	
03	01	02	05	70	30	30	20	150

Subject Contents			
Sr. No	Topic	Total Hours	Weight (%)
1	Introduction to Data Science and Analytics: - Data, features - Pre-processing on data - Cleaning of data - Feature selection techniques like Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) - Components of Analytics - reporting and analysis	6	10
2	Handling Data Sources : - Different types of data sources: structured, unstructured and semi-structured data - Relational databases: normal forms, transactional data, Structured Query Language (SQL) - NoSQL databases and its types - Handling semi-structured data with JSON, CSV files and XML	6	20
3	Exploratory Data Analysis (EDA), Models and Techniques Working with trend detection, outlier detection, summarization, association rule mining, missing distribution and imputation technique, spurious relationship or spurious correlation, concept of performance window, missing trends or percentile distribution from	8	25

	time perspective and concept of winsorization or flooring - Regression models: linear and non-linear, logistic, variable transformation, spinning of variables, population stability index and characteristic analysis - Regularization, overfitting and under fitting, mean square error, root mean square error, mean absolute percentage error - Decision tree classification, support vector machine, k-means clustering, usage of clustering techniques in variable selection		
4	Reporting Fundamentals - Anatomy and types of reports - Top-down approach: Drill down reports and dashboards - Bottom-up approach: analysis and prediction with ad-hoc queries - Strategies and techniques for effective reporting: best practices	8	20
5	Reports for Data Analysis - Descriptive analysis and its reports: Key Performance Indicator (KPI) dashboards and periodic reports - Diagnostic analysis and detailed drill down reports - Predictive analysis and reports based on predictive models - Prescriptive analysis and reports based on AI/ML models	6	15
6	Data Reporting Tools - Graphs and Charts: types and implementation - Tables: varieties and its usage in standard reports - Dashboards and drill down reports - Interactive reports - Report generation best practices based on case studies	6	10

List of Reference Books:

1. Davy Cielen, Arno D. B. Meysman, Mohamed Ali, Introducing Data Science, Manning Publications Co., 1st edition, 2016.
2. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, An Introduction to Statistical Learning: with Applications in R, Springer, 1st edition, 2013.
3. Bart Baesens, Analytics in a Big Data World: The Essential Guide to Data Science and its Applications, Wiley 2014 .
4. D J Patil, Hilary Mason, Mike Loukides, Ethics and Data Science, O' Reilly, 1st edition, 2018.
5. Dr Anil Maheshwari, Data Analytics Made Accessible, Publisher: Amazon.com Services LLC 2024.
6. Joel Grus, Data Science from Scratch: First Principles with Python, O'Reilly, 1st edition, 2015.
7. Cathy O'Neil, Rachel Schutt, Doing Data Science, Straight Talk from the Frontline, O' Reilly, 1st edition, 2013.
8. Jure Leskovec, Anand Rajaraman, Jeffrey David Ullman, Mining of Massive Datasets, Cambridge University Press, 2nd edition, 2014.

9. Eric Siegel, Predictive Analytics The Power to Predict Who Will Click, Buy, Lie, or Die, 2nd Ed., Wiley. Kevin Night, Elaine Rich, and Nair B., “Artificial Intelligence”, McGraw Hill, 2017

List of Suggested titles of Experiments:

1. Install, configure and run python, numPy and Pandas.
2. Install, configure and run Hadoop and HDFS.
3. Visualize data using basic plotting techniques in Python.
4. Implement NoSQL Database Operations: CRUD operations, Arrays using MongoDB).
5. Implement Functions: Count – Sort – Limit – Skip – Aggregate using MongoDB.
6. Implement word count / frequency programs using MapReduce.
7. Implement a MapReduce program that processes a dataset.
8. Implement clustering techniques using SPARK.
9. Implement an application that stores big data in MongoDB / Pig using Hadoop / R.

Course Learning Outcomes (CLO): On completion of this course, the students will be able to:

CLO	Description	Bloom's Taxonomy Level
CLO1	Explore the fundamental concepts of data analytics	1 Remembering 2. Understanding
CLO2	Understand data analysis techniques for applications handling large data	2 Understanding 3 Applying 4 Analyze
CLO3	Understand various machine learning algorithms used in data analytics process	2 Understanding 3 Applying 4 Analyze
CLO4	Visualize and present the inference using various tools	3 Applying 5 Evaluate
CLO5	Learn to think through the ethics surrounding privacy, data sharing and algorithmic decision-making	2. Understanding 3 Applying

Mapping of CLOs with POs & PSOs

Course Learning Outcomes	Program Outcomes (POs)												Program Specific Outcomes(PSOs)			
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO 1	PSO 2	PSO 3	PSO 4
CLO1	H	M	L	L	M								M	L	L	H
CLO2	H	M	M	L	H								H	M	M	L
CLO3	M	M	L	M	H								M	L	M	M
CLO4	H	L	M	L	M								H	M	M	M
CLO5	M	M	M	H	M								M	M	M	M

H: High, M: Medium, L: Low