



GANPAT UNIVERSITY

FACULTY OF COMPUTER APPLICATIONS

Programme	Master of Computer Applications					Branch/Spec.	Computer Application		
Semester	II					Version	1.0.0.0		
Effective from Academic Year			2026-27			Effective for the batch Admitted in		July 2026	
Subject Code	P12E5EDP		Subject Name			Exploratory Data Analysis & Predictive Analytics			
Teaching scheme						Examination scheme (Marks)			
(Per week)	Lecture (DT)		Practical (Lab.)		Total		CIA	SEE	Total
	L	TU	P	TW					
Credit	2	0	2	0	4	Theory	40	60	100
Hours	2	0	4	0	6	Practical	20	30	50

Objective:

The primary objective of this course is to equip MCA students with the knowledge and practical skills required to analyze, interpret, and visualize data for intelligent decision-making. Upon completion of the course, students will be able to:

- Understand the complete data analytics lifecycle and its applications in real-world business problems.
- Apply statistical and mathematical techniques to analyze structured and unstructured datasets.
- Perform data cleaning, preprocessing, feature engineering, and exploratory data analysis (EDA).
- Design effective data visualizations and interactive dashboards for data-driven decision making.
- Develop and evaluate predictive analytics models using machine learning algorithms.
- Utilize Python-based data science libraries for solving analytical problems.

Pre-requisites:

Students should have basic knowledge of:

- Programming using Python or any high-level programming language
- Statistics and Probability
- Database Management Systems
- Basic Mathematics and Linear Algebra

Course Outcomes:

- 1 = Slight (Low); 2 = Moderate (Medium); 3 = Substantial (High); “-” = No Correlation

Name of CO	Description
CO1	Explain the data analytics lifecycle and apply statistical methods to analyze real-world datasets.
CO2	Perform data preprocessing, feature engineering, and exploratory data analysis using Python-based tools.
CO3	Design effective visualizations and interactive dashboards to communicate analytical insights.
CO4	Develop, evaluate, and interpret predictive machine learning models for solving business and research problems.

Mapping of CO and PO								
Cos	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	3	2	-	2	-	-	-	2
CO2	3	3	2	3	2	-	-	2
CO3	2	2	3	3	2	1	1	2
CO4	3	3	3	3	2	2	2	3

Content:		
Unit	SECTION-I	Hours
01	Introduction to Data Science and Statistical Foundations: Introduction to Data Science, Data Analytics Process and Data Science Life Cycle, CRISP-DM Methodology, Business Understanding and Problem Definition, Data Collection and Data, Acquisition Techniques, Types of Data: Structured Data, Semi-Structured Data, Unstructured Data, Data Sources and Data Integration, Data Quality Assessment, Descriptive Statistics, Measures of Central Tendency, Measures of Dispersion, Correlation and Covariance, Basics of Probability, Sampling Methods, Statistical Inference.	8
02	Data Preprocessing and Exploratory Data Analysis: Data Cleaning Techniques, Handling Missing Values, Duplicate Detection and Removal, Outlier Detection and Treatment, Data Transformation, Data Normalization and Standardization, Feature Engineering, Feature Selection, Correlation Analysis, Exploratory Data Analysis (EDA), Univariate Analysis, Bivariate Analysis, Multivariate Analysis, Pattern Discovery using EDA, Data Profiling.	7
SECTION-II		
03	Data Visualization and Business Intelligence: Fundamentals of Data Visualization, Principles of Visual Perception, Choosing Appropriate Charts, Line Charts, Bar Charts, Pie Charts, Scatter Plots, Histograms, Box Plots, Heat Maps, Pair Plots, Correlation Visualization, Interactive Data Visualization, Dashboard Design Principles, Storytelling with Data, Introduction to Business Intelligence Dashboards.	7
04	Predictive Analytics using Machine Learning: Introduction to Predictive Analytics, Machine Learning Workflow, Data Splitting and Model Training, Regression Techniques, Classification Techniques, Decision Tree, Random Forest, Naïve Bayes, K-Nearest Neighbour (KNN), Clustering Fundamentals, Model Validation, Cross Validation, Performance Evaluation Metrics, Confusion Matrix, Precision, Recall and F1-Score, ROC Curve and AUC	8

Practical Content:	
List of programs specified by the subject teacher based on above mentioned topics.	
Text Books:	
1	Joel Grus, Data Science from Scratch , 2nd Edition, O'Reilly Media.
2	Foster Provost & Tom Fawcett, Data Science for Business , O'Reilly Media.
3	Cathy O'Neil & Rachel Schutt, Doing Data Science , O'Reilly Media
4	Wes McKinney, Python for Data Analysis, 3rd Edition, O'Reilly Media.
5	
Reference Books:	
1	Gareth James, Daniela Witten, Trevor Hastie & Robert Tibshirani, An Introduction to Statistical Learning

	Springer.
2	Trevor Hastie, Robert Tibshirani & Jerome Friedman, The Elements of Statistical Learning , Springer.
3	Aurélien Géron, Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow , 3rd Edition, O'Reilly Media.
4	Jake VanderPlas, Python Data Science Handbook , O'Reilly Media.
5	Nina Zumel & John Mount, Practical Data Science with R , Manning Publications.
MOOC/Certification Courses:	
1	NPTEL – Data Analytics with Python
2	Google Data Analytics Professional Certificate
3	Coursera – IBM Data Science Professional Certificate
Question Paper Scheme:	
University Examination Duration: 3 Hours Note for Examiner: - (I) Questions 1 and 4 are compulsory with no options. (II) Internal options should be given in questions 2, 3, 5 and 6. SECTION – I Q.1 –8 Marks Q.2 –11 Marks Q.3 –11 Marks SECTION - II Q.4 –8 Marks Q.5 –11 Marks Q.6 –11 Marks	

