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	2022-23	**Effective for the batch Admitted in**	June 2022
Subject Code	P12A4DS1	**Subject Name**	Data Science-I

Teaching scheme						**Examination scheme (Marks)**			
(Per week)	**Lecture (DT)**		**Practical (Lab.)**		**Total**		**CE**	**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:

- To learn the fundamentals of Data Science.

Pre-requisites:

- The student must have Basic computer knowledge.

Course Outcomes :

Name of CO	**Description**
CO1	Explain the data science process, roles, applications, and challenges, and differentiate data science from business intelligence.
CO2	Apply data collection and pre-processing techniques to clean, integrate, transform, reduce, and discretize datasets.
CO3	Analyze data using exploratory data analysis techniques and develop appropriate regression models.
CO4	Evaluate and optimize predictive models using validation techniques, performance metrics, and grid search.

Mapping of CO and PO

Cos	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	2	2	1	1	—	—	—	2
CO2	3	2	1	3	—	—	—	1
CO3	3	3	2	3	—	—	—	1
CO4	3	3	2	3	—	—	—	2

Content:

Unit	**Section- I**	**Hrs**

1.	Introduction Introduction to Data Science, Components of Data Science, Data Science process, Data Science Roles, Difference between Data Science and Business Intelligence, Applications of Data Science in various fields, Challenges of Data Science Technology.	9
2.	Data Collection and Data Pre-Processing. Data Collection Methods, Data Pre-processing, Data Cleaning, Data Integration and Transformation, Data Reduction, Data Discretization.	9
	Section - II	
3.	Exploratory Data Analysis and Model Development Exploratory Data Analysis (EDA), Types of Exploratory Data Analysis, Various Steps involved in EDA, Exploratory Data Analysis Tools (Python, R), Box Plots, Pivot Table, Heat Map, Histograms, Line Graphs. Model Development: Linear Regression (Simple and Multiple Regression), Logistic Regression, Ridge regression, Lasso Regression, Polynomial Regression, and Stepwise Regression.	9
4.	Model Evaluation Generalization Error, Out-of-Sample Evaluation Metrics, Cross Validation, Over fitting, Under Fitting, Prediction by using Ridge Regression, Testing Multiple Parameters by using Grid Search.	9

Practical Content:

List of programs specified by the subject teacher based on above mentioned topics.

Text Books:

1	"Doing Data Science: Straight Talk from the Frontline" by Cathy O'Neil and Rachel Schutt
2	"The Art of Data Science" by Roger D. Peng and Elizabeth Matsui

MOOC/Certification Courses:

1	https://www.coursera.org/specializations/introduction-data-science
2	https://www.edx.org/course/intro-to-data-science
3	https://nptel.ac.in/courses/106107220

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