SARVAJANIK UNIVERSITY
INCLUSIVE | INTEGRATED | INNOVATIVE

SARVAJANIK UNIVERSITY
Sarvajanik College of Engineering and Technology
Master of Technology

Year: M. Tech. I (Semester – I)

Subject Name: Data Science **Subject Code:** MTCO24103
Type of course: Professional Elective - II
Prerequisite (if any): Python Programming, Statistics and Probability

List of Courses where this course will be prerequisite: --

Rationale: This course provides insights into inferential and descriptive statistics. Students can gain a thorough understanding of data science life cycle, data analytics and data visualization techniques and tools. Students will be able to solve predictive and categorization problems on real time data after learning this course.

Teaching and Examination Scheme:

Teaching Scheme				Theory Marks			Practical Marks		Total
L	T	P	C	TEE	CA1	CA2	TEP	CA3	
3	0	2	4	60	25	15	30	20	150

CA1: Continuous Assessment (assignments/projects/open book tests/closed book tests CA2: Sincerity in attending classes/class tests/ timely submissions of assignments/self-learning attitude/solving advanced problems TEE: Term End Examination TEP: Term End Practical Exam (Performance and viva on practical skills learned in course) CA3: Regular submission of Lab work/Quality of work submitted/Active participation in lab sessions/viva on practical skills learned in course

Content:

Sr.No.	Description	No. of Hours
1	Unit 1: Introduction to Data Science • Overview of Data Science • Importance and applications • Data Science lifecycle • Data Science tools and technologies • Data Types	4
	Data collection and management: • Introduction, • Sources of data, • Data collection and APIs, • Exploring and fixing data, • Data storage and management, using multiple data sources	6

2	Numpy and Pandas • NumPy Basics: Arrays and Vectorized Computation; The NumPy ndarray; Universal Functions: Fast Element-Wise Array Functions; Linear Algebra; Pseudorandom Number Generation • Pandas Basics: Dataframes Basics, Pandas Data structures -Series, Dataframe, Index Objects; creating, updating, exporting, importing, Applying functions over Series and Dataframes row wise and column wise. Aggregate, Transform, Filter, groupby, Multiindex, converting, downcasting, DateTime	8
3	Exploratory Data Analysis • Descriptive Statistical Analysis- Measures for Centrality, Spread, Normality- Skewness and Kurtosis; • Univariate and bivariate analysis • Correlation and covariance analysis	6
5	Estimation: • Inferential Statistics • Classical Methods of Estimation: Point estimation, Interval Estimation, • Estimating the Mean, Standard Error of a Point Estimate, • Confidence Interval, Prediction Interval • Estimating the Variance, • Estimating a Proportion, • Difference between Two Means, between Two Proportions for Two Samples	10
6	Data Visualization • Introduction to Data Visualization • Types of data visualization, Data for visualization: Data types, Data encodings, Retinal variables, mapping variables to encodings, Visual encodings. • Matplotlib: Basic and Advanced Plotting • Seaborn: Statistical Data Visualization • Plotly: Interactive Visualization • Best Practices in Data Visualization	7
7	Data Modeling and Evaluation using Metrics: Case Study : Prediction Model and Classification Model	4

Reference Books:

Sr No	Title of book /article	Author(s)	Publisher and details like ISBN	Year of publication	Publication Edition
1	Python for Data Analysis - Data Wrangling with Pandas, NumPy, and IPython	Wes McKinney	O'Reilly	2017	2nd Edition
2	Pandas for Everyone - Python Data Analysis	Daniel Y. Chen	Pearson Addison Wesley -Data & Analytics Series	2020	5th Edition
3	Mastering python for data science	Samir Madhavan	Ingram short title, ISBN : 978-1784390150	2015	1st Edition
4	Applied Data Science with Python and Jupyter	Alex Galea	Packt Publications, ISBN : 978-1-78995-817-1	2018	1st Edition
5	Data Science for Dummies	Lillian Pierson	by John Wiley & Sons, ISBN : 978-1-119-32763-9	2017	2nd Edition

Course Outcomes:

Sr. No.	CO statement	Marks % weightage
CO-1	Demonstrate how to generate, collect, clean and process data.	10
CO-2	Conduct exploratory data analysis using statistics and relevant visual methods and ability to handle missing data, outliers, and integrate disparate datasets effectively.	20

CO-3	Estimate population parameters, intervals, and proportions from sample data to make informed statistical inferences from data using classical estimation methods.	25
CO-4	Distinguish among common visual presentations of data used for effective communication.	20
CO-5	Design and develop practical algorithms for solving real life problems.	25

List of Open learning website:

1. NPTEL Course on Data Analytics using Python : ttps://onlinecourses.nptel.ac.in/noc21_cs45/course
2. NPTEL Course on Python for Data Science : https://onlinecourses.nptel.ac.in/noc21_cs33/

List of Open Source Software:

1. Open Source Python Libraries : NumPy, SciPy, Pandas, SCIKIT-Learn, TensorFlow, BeautifulSoup, Matplotlib, Seaborn
2. IDEs: Jupyter Notebook, Spyder, WIDE -Google Colab

FOR LAB SESSIONS:

List of Experiments:

Sr. No	Practical
1	Write a program to collect the Twitter data using Twitter API and Python libraries.
2	Manipulate arrays in Python using Numpy.
3	Practice data manipulation with Pandas. (Practice data series, data frames, data selection, sorting, searching and statistics with Pandas Python library.)
4	Data Collection and Management Practical: • Collect data from multiple sources, including APIs, and explore the collected data. • Use Pandas to clean and fix data inconsistencies. • Store and manage the cleaned data in multiple formats and sources.
5	Exploratory Data Analysis (EDA) Practical:

	• Perform descriptive statistical analysis on a dataset, including measures for centrality, spread, skewness, and kurtosis. • Conduct univariate and bivariate analysis to explore relationships between variables. • Calculate correlation and covariance to understand the strength and direction of relationships between variables.
6 7	Data Visualization Practical: • Create various types of data visualizations using Matplotlib, Seaborn, and Plotly libraries. • Choose appropriate visual encodings based on data types and variables to effectively communicate insights. Data Modeling and Evaluation Practical: • Build prediction and classification models using machine learning techniques. • Evaluate model performance using relevant metrics such as accuracy, precision, recall, and F1-score. • Conduct a case study to develop and evaluate a prediction or classification model on a real-world dataset.