SILVER OAK UNIVERSITY

Engineering and Technology (M.Tech.)

Computer Engineering (Software Engineering)

Subject Name: Data Science

Subject Code:

Semester: II

Prerequisite:

Data Structures, Basics of Probability and Statistics, python, basics of R, number theory.

Objective:

- Provide the knowledge and expertise to become a proficient data scientist.
- Demonstrate an understanding of statistics and machine learning concepts that are vital for data science.
- Produce code to statistically analyses a dataset.
- Critically evaluate data visualizations based on their design and use for communicating stories from data.

Teaching and Examination Scheme:

Teaching Scheme			Credits	Evaluation Scheme				Total Marks
L	T	P	C	Internal		External		
				Th	Pr	Th	Pr	
3	0	2	4	70	30	30	20	150

Content:

Unit No.	Course Contents	Teaching Hours	Weightage %
1	**Introduction to core concepts and technologies:** Introduction, Terminology, data science process, data science toolkit, Types of data, Example applications.	5	10
2	**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	15
3	**Data analysis: Introduction, Terminology and concepts:** Introduction to statistics, Central tendencies and distributions, Variance, Distribution properties and arithmetic, Samples/CLT, Basic machine learning algorithms, Linear regression, SVM, Naive Bayes.	9	25

4	**Data visualization:** Introduction, Types of data visualization, Data for visualization: Data types, Data encodings, Retinal variables, Mapping variables to encodings, Visual encodings.	9	25
5	**Applications:** Applications of Data Science, Technologies for visualization, Bokeh (Python)	6	15
6	**Recent trends**: Recent trends in various data collection and analysis techniques, various visualization techniques, application development methods of used in data science.	4	10

Course Outcome:

Sr. No.	CO statement	Unit No
CO-1	Define the concepts of data science,types of data and their application.	1,5
CO-2	Understand the concept of machine learning algorithm and their Implementation.	3
CO-3	Understand various tools and techniques for data manipulation and visualization and management.	2, 3
CO-4	Apply data science concepts and methods to solve problems in real-world contexts and will communicate these solutions effectively.	3, 4, 5
CO-5	Develop the ability to build and assess data-based models.	5, 6

List of Experiments/Tutorials:

Sr. No.	Data Science (Practical)
1	AS CALCULATOR APPLICATION a.Using with and without R objects on console b.Using mathematical functions on console c. Write an R script, to create R objects for calculator application and save in a specified location in disk.
2	DESCRIPTIVE STATISTICS IN R a. Write an R script to find basic descriptive statistics using summary, str, quartile function on mtcars& cars datasets. b. Write an R script to find subset of dataset by using subset (), aggregate () functions on iris dataset.
3	READING AND WRITING DIFFERENT TYPES OF DATASETS a. Reading different types of data sets (.txt, .csv) from web and disk and writing in file in specific disk location. b. Reading Excel data sheet in R. c. Reading XML dataset in R.

4	VISUALIZATIONS a. Find the data distributions using box and scatter plot. b. Find the outliers using plot. c. Plot the histogram, bar chart and pie chart on sample data
5	CORRELATION AND COVARIANCE a. Find the correlation matrix. b. Plot the correlation plot on dataset and visualize giving an overview of relationships among data on iris data. c. Analysis of covariance: variance (ANOVA), if data have categorical variables on iris data.
6	REGRESSION MODEL Import a data from web storage. Name the dataset and now do Logistic Regression to find out relation between variables that are affecting the admission of a student in an institute based on his or her GRE score, GPA obtained and rank of the student. Also check the model is fit or not. require (foreign), require (MASS).

Major Equipment:

Modern System with related software
(Anaconda navigator)

Books Recommended:-

1. Doing Data Science, Cathy O'Neil and Rachel Schutt, Straight Talk From The Frontline. O'Reilly.
2. Introduction to Data Science, Davy Cielen, Arno D B Meysman and Mohamed Ali, Manning, dreamtech press
3. Practical Data Science, Nina Zumwl and John Mount, Manning, dreamtech press
4. The Data Science Handbook, Field Cady, Wiley
5. Getting Started with Data Science, Murtaza, Haider, Pearson
6. Data Science and Big Data Analytics, EMC Education Services, Wiley
7. Data Science, John D Kellehar, MIT Press
8. Mining of Massive Datasets. v2.1, Jure Leskovek, AnandRajaraman and Jeffrey Ullman, Cambridge University Press

List of Open Source Software/learning website:
https://www.analyticsvidhya.com/blog/2016/01/complete-tutorial-learn-data-science-python-scratch-2/
https://www.rstudio.com/online-learning/