

**CHAROTAR UNIVERSITY OF SCIENCE & TECHNOLOGY**  
**FACULTY OF TECHNOLOGY AND ENGINEERING**  
**Smt. Kundanben Dinsha Patel Department of Information Technology**

**Subject Name:** Data Science  
B.Tech. VII  
**Subject Code:** IT448  
**Year:** 2022-23

**Semester:**

**Academic**

Note: The laboratory will emphasize Data acquisition, Data preprocessing, Data Storage and Retrieval, Statistical tools for Data Science and its applications.

Instructions:

1. **All Practical must be performed individually and all experimental results must be uploaded on your respective data science blog/github/word.**
2. All Practical must be evaluated regularly in the laboratory by concern Lab Teacher.
3. Each practical answer would be evaluated as learning outcome.

**Practical List**

Sr. No.	Aim of the Practical	Hrs
Pre Req.	Introduction to Data Science and tools. How python/R used in Data Science? Discuss on real life applications and usage of Data Science.	-
1	Perform data collection by web scrapping with python. <b>Perform following tasks for Web scrapping with python:</b>  1. Find the URL that you want to scrape 2. Inspecting the Page 3. Find the data you want to extract 4. Write the code 5. Run the code and extract the data 6. Store the data in the required format	2
2	Perform following Data Pre-processing tasks in Python using Scikit-learn. standardization, normalization, encoding, discretization, imputation of missing values. Use your own dataset to perform all pre-processing tasks as suggested in given reference. 1. <a href="https://www.analyticsvidhya.com/blog/2016/07/practical-guide-data-preprocessing-python-scikit-learn/">https://www.analyticsvidhya.com/blog/2016/07/practical-guide-data-preprocessing-python-scikit-learn/</a> 2. <a href="https://scikit-learn.org/stable/modules/preprocessing.html">https://scikit-learn.org/stable/modules/preprocessing.html</a> Answer the following question in your blog (As per dataset taken by you): Dataset Description: Task to be performed:	4

	<p>How to decide variance threshold in data reduction?</p> <p>Code Snapshot</p> <p>Output Snapshot</p> <p>Task-2</p> <p>Code Snapshot</p> <p>Output Snapshot</p>	
3	<p>Perform following Data Pre-processing tasks using python</p> <p>Data reduction using variance threshold, univariate feature selection, recursive feature elimination, PCA, correlation</p> <p>Reference:</p> <ol style="list-style-type: none"> <li>1. <a href="https://medium.com/analytics-vidhya/feature-selection-using-scikit-learn-5b4362e0c19b">https://medium.com/analytics-vidhya/feature-selection-using-scikit-learn-5b4362e0c19b</a></li> <li>2. <a href="https://machinelearningmastery.com/rfe-feature-selection-in-python/">https://machinelearningmastery.com/rfe-feature-selection-in-python/</a></li> <li>3. <a href="https://towardsdatascience.com/pca-using-python-scikit-learn-e653f8989e60">https://towardsdatascience.com/pca-using-python-scikit-learn-e653f8989e60</a></li> <li>4. <a href="https://towardsdatascience.com/feature-selection-using-python-for-classification-problem-b5f00a1c7028">https://towardsdatascience.com/feature-selection-using-python-for-classification-problem-b5f00a1c7028</a></li> <li>5. <a href="https://www.analyticsvidhya.com/blog/2016/01/guide-data-exploration/">https://www.analyticsvidhya.com/blog/2016/01/guide-data-exploration/</a></li> </ol> <p>Answer the following question in your blog (As per dataset taken by you):</p> <p>Dataset Description:</p> <p>Task to be performed:</p> <p>Why feature selection is important?? Its advantages/disadvantages.</p> <p>Code Snapshot</p> <p>Output Snapshot</p> <p>What is the impact on accuracy, with or without data reduction?</p> <p>Code Snapshot</p> <p>Output Snapshot</p> <p>Amongst all methods, which method avoids overfitting and improves model performance?</p>	4
4	<p><b>Perform following task in the orange tool.</b></p> <p><b>How to use workflows in orange?</b></p> <p><b>How to do basic data exploration (like data distribution, data information).</b></p> <p><b>How to load your data in Orange and how to load external data from API in Orange?</b></p> <p><b>Learn all widgets in Orange tool.</b></p> <p><a href="https://orange.biolab.si/widget-catalog/">https://orange.biolab.si/widget-catalog/</a></p>	2
5	<p><b>What is the effect of discretization, continuization, Normalization, Randomization on the data w.r.t. Orange?</b></p> <p><b>How to work with Orange in Python and vice-versa?</b></p> <p><b>Text / Data Preprocessing with Orange tool.</b></p>	2

	<a href="https://docs.biolab.si//3/data-mining-library/reference/preprocess.html">https://docs.biolab.si//3/data-mining-library/reference/preprocess.html</a> <a href="https://orange-data-mining-library.readthedocs.io/en/latest/#tutorial">https://orange-data-mining-library.readthedocs.io/en/latest/#tutorial</a>	
6	Introduction to PowerBI and Get started with PowerBI, Prepare data for analysis and Model data in Power BI.	2
7	Consume data with Power BI and How to build a simple dashboard.	2
8	Perform Data Analytics using PowerBI using the given dataset. Generate a report that contains various visualizations.	2
9	Execute queries in the Neo4j graph database and in Gephi tool perform the operations of loading csv data, running graph statistics scripts and displaying various graphical layouts.	2
10	Using image data, predict the gender and age range of an individual in Python. Test the data science model using your own image.	4
11	Perform Exploratory Data Analysis using Matplotlib and Seaborn library of given dataset	2
12	Explore weka tool and explain various task can be perform with this tool.	2

Prepared By: Jalpesh Vasa, Hemant Yadav