|  |  |  |
| --- | --- | --- |
|  | **JSPM's RAJARSHI SHAHU COLLLEGE OF ENGINEERING**  **Tathawade, Pune**  **An Autonomous Institute Affiliated to Savitribai Phule Pune University**  **T.Y. B.Tech. COMPUTER ENGINEERING**  **A.Y. 2022-2023 Sem II**  **Practical Plan** |  |

**Academic Year : 2022-23 Semester :II**

**CLASS: TY B (Comp)**

**Lab: Data Mining**

**Name of Student: Roll No:**

|  |  |  |
| --- | --- | --- |
| **Sr.No.** | **Assignment Title** | **Colab Link** |
|  | For an organization of your choice, choose a set of business processes. Design star / snow flake schemas for analyzing these processes. Create a fact constellation schema by combining them. Extract data from different data sources, apply suitable transformations and load into destination tables using an ETL tool. |  |
|  | Trip History Analysis: Use trip history dataset that is from a bike sharing service in the United States. The data is provided quarter-wise from 2010 (Q4) onwards. Each file has 7 columns. Predict the class of user. | <https://colab.research.google.com/drive/1Q2IOTYqd1ZlsHGJCJ5caufm8C41A0rGE?usp=sharing> |
|  | Consider a suitable text dataset. Remove stop words, apply stemming and feature selection techniques to represent documents as vectors. Classify documents and evaluate precision, recall. | <https://colab.research.google.com/drive/10hQn6CGB6fM7XZvvemNewAwIJpioy6H4?usp=sharing> |
|  | Apply a-priori algorithm to find frequently occurring items from given data and generate strong association rules using support and confidence thresholds. | <https://colab.research.google.com/drive/15NCK3M-5UE7DjN2BCdAe9-QjhZXt6_gC?usp=sharing> |
|  | Download Boston Housing dataset. Create a Model using linear regression to predict the houses price. | <https://colab.research.google.com/drive/1nUy_ILYt7V9XPXI_GYr685ESqyeQS4vk?usp=sharing> |
|  | Predict the number of bicycle trips across Seattle's Fremont Bridge based on weather, season, and other factors and also Figure out what we can learn about people in Seattle from hourly commute data. (The daily or hourly bicycle counts can be downloaded from http://data.seattle.gov/) |  |
|  | Apply ARIMA model to perform time series analysis on COVID- India Dataser from Kaggle |  |
|  | Mini-Project Base on Data Mining and analytics. | NA |