Department of Computer Science & Engineering

Final Year B. Tech. (CSE) – I : 2022-23
5CS462 : PE5 - Data Mining Lab
Assignment No. 9 : Mini Project

Sr. No.	PRN / LoginID (by excluding BTE)	Name of students	Title / Description of Mini Project
Batch B5 (PE5-DM)			
1	2017BTECS00009	Priyanka sukalal Gavit	**1] Housing Price Predictions for real estate business** Use the housing dataset which includes all the prices of the different houses. In this project, the dataset for prediction of price is added along with location, size of the house, and additional information required for it. Depending on the level of sophistication, you can follow a predictive model with simple techniques such as regressions or machine learning libraries. This project utilizes algorithms and techniques for price predictions of the houses based on different housing datasets.
2	2017BTECS00098	Naeem Akhtar Allabax Momin	
3	2018BTECS00009	Gayatri Anil Bagul	
4	2018BTECS00103	Tufel Iqbal Aattar	
5	2018BTECS00108	Sanskruti Yuvraj Chavan	**2] Smart Health Disease Prediction Using Naive Bayes** The smart health disease prediction is an end user support system that allows users to get guidance immediately with the help of an online intelligent health system. The system holds complete information about symptoms and the diseases associated with it. The system analyses diseases associated with the symptoms for the patient and advises them for X-ray, blood test or CT scan as requested by the system. Users can also directly get in touch with the specialist doctors for any ailment and share your reports. It is not just one time, rather a proper login detail is shared for future use.
6	2019BTECS00002	Parthesh Mandar Ingale.	
7	2019BTECS00003	PRAJWAL BALASAHEB YADAV	
8	2019BTECS00006	Shital Arun Solanke	

Department of Computer Science & Engineering

Final Year B. Tech. (CSE) – I : 2022-23
5CS462 : PE5 - Data Mining Lab
Assignment No. 9 : Mini Project

Sr. No.	PRN / LoginID (by excluding BTE)	Name of students	Title / Description of Mini Project
9	2019BTECS00007	Dattatray Parmeshwar Reve	**3] Product and Price Comparing tool** With the increase in popularity of e-commerce portals, shopping websites are magnifying to a great extent to enable online shoppers to purchase anything with just one click and get it delivered at your doorstep. To purchase an item, people tend to spend quite a lot of time in searching a product and comparing it with other websites by themselves. In this project, you can compare product and price of a product to buy cheap and best deal available. Also, it will track consumer demand and inform when the commodity price is lowest and notify consumers proactively.
10	2019BTECS00008	Amitkumar Ashok Khandekar	
11	2019BTECS00009	SONAL SUBODH MANE	
12	2019BTECS00011	Navjyot Netaji Sakhalkar	
13	2019BTECS00013	Mohd Nifasat Beg	**4] Solar Power generation forecaster** With the help of extracted data from two solar power plants over a period of 34- days, two pairs of files are available. Each pair includes one power generation dataset, and another is sensor reading dataset. In the power generation dataset, each inverter extracts information which has several lines of solar panels connected to it. An array of sensors optimally located at the plant collects the sensor data. In this project, you will be able to get answers of the amount of power generated in a month, any faulty performing equipment in the plant or panel cleaning/ maintenance update. In this project, the dataset is evaluated based on a transparent open box (TOB) network for data mining and predictions. It provides accurate information from the hourly data record from power generation dataset and sensor reading dataset. Expected queries : 1. To predict the power generation for the next couple of days 2. Identify the importance of panel cleaning / maintenance 3. Identify faultily or sub-optimally performing equipment
14	2019BTECS00014	Siddhi Balkrushna Lokhande	
15	2019BTECS00016	Raut Akash Sanjay	
16	2019BTECS00017	Muskan Raju Attar	

Department of Computer Science & Engineering

Final Year B. Tech. (CSE) – I : 2022-23
5CS462 : PE5 - Data Mining Lab
Assignment No. 9 : Mini Project

Sr. No.	PRN / LoginID (by excluding BTE)	Name of students	Title / Description of Mini Project
Batch B6 (PE5-DM)			
1	2019BTECS00019	Akshay Ramchandra Musmade	**5] Prediction of Adult Income based on Census Data** The following project is the classification project to predict the income level of an individual that exceeds 50K based on the census data available at the repository. The dataset that is used in the projects are variables such as age, type of work, working hours, sex and many more. It helps in understanding the standard of living of the city, benefit of setting up the business or bank loan eligibility. Also, it helps in understanding the real estate preferences by average income of the people residing in the area. In this project, you will also be able to figure out the type of tourist places that people from other countries would like to travel.
2	2019BTECS00022	Prachi Rameshkumar Chobhare	
3	2019BTECS00023	kartik khunda	
4	2019BTECS00025	Shreyash Suryakant Gaste	
5	2019BTECS00028	Smital Rajendra Patil	**6] Fraud Detection in Monetary Transactions** Detecting fraudulent transactions is a very significant use case in today's scenario of digitized monetary transactions. In order to address this problem, a Synthetic Data is generated using PaySim Simulator and it is made available at Kaggle site. The data contains transaction details like transaction type, amount of transaction, customer initiating the transaction, old and new balance in Origin i.e., before and after transaction respectively and same as in Destination Account along with the target label, is fraud. So, based on the transaction details, a Classification Model can be developed that can detect fraudulent transactions.
6	2019BTECS00041	Suyash Sanjay Chavan	
7	2019BTECS00042	Aditya Hemant Sarnobat	
8	2019BTECS00043	Krushna Mali	

Department of Computer Science & Engineering

Final Year B. Tech. (CSE) – I : 2022-23
5CS462 : PE5 - Data Mining Lab
Assignment No. 9 : Mini Project

Sr. No.	PRN / LoginID (by excluding BTE)	Name of students	Title / Description of Mini Project
9	2019BTECS00044	Prathamesh Ramdas Chavhan	**7] Intelligent Transportation System** Develop a model to predict the required number of buses for a particular route based on the passenger movement. This system helps to optimize the route by forecasting the passenger's data.
10	2019BTECS00045	Sumedh Milind Bhatkar	
11	2019BTECS00046	Aniket Ananda Vyawahare	
12	2019BTECS00047	Sayali Balasaheb Katkar	
13	2019BTECS00049	Bhavika Tanaji Ghadage	**8] FinTech Tool : Investment Predictor** An Indian user on Kaggle had collected the information about their financial investments. So, the dataset has an individuals' gender and age along with the details about their deposits in different investment options (gold bonds, PPF, Fixed deposits, etc.) With the help of above data set at Kaggle, analyse the preferences of Indians in investing their money. You can also do a gender-based analysis to understand which gender is likely to pick specific investment options. As the dataset also contains the age of the individuals, you can use it to know the bias of younger and older people for investing their money.
14	2019BTECS00052	yash deelip kalam	
15	2019BTECS00053	Miss.JADHAV KSHITIJA SHAMRAO	
16	2019BTECS00054	Ruturaj Chandrakant Shinde	

Department of Computer Science & Engineering

Final Year B. Tech. (CSE) – I : 2022-23
5CS462 : PE5 - Data Mining Lab
Assignment No. 9 : Mini Project

Sr. No.	PRN / LoginID (by excluding BTE)	Name of students	Title / Description of Mini Project
Batch B7 (PE5-DM)			
1	2019BTECS00056	Rohan Pramod Bondre	**9] Students Performance Analytics** Use the Student Performance dataset available on Kaggle. It contains information about the socio-economic background of students and their grades in various subjects. Project: You can use the dataset to analyse the significance of socio-economic factors in affecting a student's performance. You can do a gender-based analysis as well for understanding how gender relates to the student's grades.
2	2019BTECS00057	Shreyash Vinod Malu	
3	2019BTECS00059	Sanmati Subhas Sattur	
4	2019BTECS00060	Suvansh Sharma	
5	2019BTECS00067	Vaibhav Vitthal Kute	**10] Data Analytics for Food Cafes** Deciding the items and their prices on a menu card is not an easy task for cafe owners. They have to constantly analyse their customers' choices to set the optimum prices of their food items on the menu. Dataset: (https://1drv.ms/u/s!AtgrmeCPhKh7kYIacOREVS-3gUjxZw). It has three files that contain information about the cafe's sales, transactions, and time labels for each transaction. Using the dataset mentioned above, verify a few fundamental economic trends in the dataset as a first step. These trends will include analysing price trends and sales of all the items, sales on special holidays and weekends, and more such trends. Go more insights by visualising the dataset through the seaborn library of the Python Programming Language. Another metric that you must evaluate for this project is the Price Elasticity of all cafe items.
6	2019BTECS00068	Aditi Anirudha Joshi	
7	2019BTECS00072	Nikhil Purushottam Khavanekar	
8	2019BTECS00073	Abhishek Kallu Kamble	

Department of Computer Science & Engineering

Final Year B. Tech. (CSE) – I : 2022-23
5CS462 : PE5 - Data Mining Lab
Assignment No. 9 : Mini Project

Sr. No.	PRN / LoginID (by excluding BTE)	Name of students	Title / Description of Mini Project
9	2019BTECS00074	Sushant Patil	**11] Data Analytics for Amazon Reviews** Amazon Reviews are a boon for customers and Amazon itself as it can analyse the data to draw relevant inferences. The Amazon Reviews/Rating dataset has about 2 million reviews for different products. Apply cosine similarity and centred cosine similarity to understand the significance of reviews. After normalising the ratings, you can create a user-item matrix to identify similar customers.
10	2019BTECS00075	Pranav Prakash Karale	
11	2019BTECS00077	Avinash Vishnu Biradar	
12	2019BTECS00078	Om Anant Khairnar	
13	2019BTECS00079	Joshi Shantanu Anil	**12] Analytics for San Francisco Salaries Data** Use the San Francisco Salaries Dataset to understand the income inequality in San Francisco city. Analyse the factors responsible for the promotions of certain employees. To better understand the distribution of the salaries, use different visualization libraries like scatter plots, box plots, whisker plots and the density plots.
14	2019BTECS00080	Saurabh Narayan Nagre	
15	2019BTECS00083	Bhargav Girish Kulkarni	
16	2019BTECS00084	Pravin Santosh Lokhande	
Batch B8 (PE5-DM)			
1	2019BTECS00086	Rutuja Sayaji More	**13] Automated personality classifier** The automatic system analyzes the characteristics and behaviors of participants. And after observing the past patterns of data classification, it predicts a personality type and stores its own patterns in a dataset. The scope has summarized as follows: a) Store personality-related data in a database b) Collect associated characteristics for each user c) Extract relevant features from the text entered by the participant d) Examine and display the personality traits e) Interlink personality and user behaviour (There can be varying degrees of behaviour for a particular personality type) Use case : Career guidance services where a student's personality is matched with suitable career paths
2	2019BTECS00090	Udaykumar Sanjay Gadikar	
3	2019BTECS00091	Koustubh Tatikondawar	
4	2019BTECS00095	Vaishnavi Vitthal Daware	

Department of Computer Science & Engineering

Final Year B. Tech. (CSE) – I : 2022-23
5CS462 : PE5 - Data Mining Lab
Assignment No. 9 : Mini Project

Sr. No.	PRN / LoginID (by excluding BTE)	Name of students	Title / Description of Mini Project
5	2019BTECS00098	Anchit Gupta	**14] Mining of customer behaviour of any retail shop.** Consider daily billing data.
6	2019BTECS00099	Rahulkumar Bhagwansing Pardeshi	
7	2019BTECS00102	Pratik Babaso Chougule	
8	2019BTECS00103	Sanket Satish Mali	
9	2019BTECS00106	Anushka Thaware	**15] Mining Facebook Posts** Use the Facebook data to find out the most popular topics on the platform. Apply mining algorithm to predict which topics will be the most popular in the future.
10	2019BTECS00107	Abhishek Dadarao Bele	
11	2019BTECS00108	Saurabh Ashok Tribhuvan	
12	2019BTECS00111	Aditya Hanumant Sanap	
13	2019BTECS00113	SANKET RAJARAM MOTE	**16] Predicting Stock Prices with Neural Networks** Use a neural network to predict stock prices. Download the data from Yahoo! Finance and then train neural network to predict future stock prices.
14	2019BTECS00114	Swapnil Santosh Kanade	
15	2020BTECS00205	Monika Vinod Chitrakathi	
16	2020BTECS00212	Pranita Rajendra Bhosale	

Deadline : 20/11/2022

Dr. B. F. Momin
Course Teacher