



**VR Siddhartha Engineering College**  
**Department of Information Technology**



**Link to download the dataset corresponding to the task:**

<https://drive.google.com/drive/folders/1Y5stBLqJpB3uFICmDsXYfuMasK30fo5F?usp=s>  
here link

**Note:** In the folder, Dataset CSV file is also named with Task number. For classification task file is named like Cx. name and for Regression task file is named like Rx.name. So download the file which is named with your task number.

Bat ch No	Question	CO	BTL
1	<p><b>TASK 1:</b> <b>CLASSIFICATION: STROKE PREDICTION</b> According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths. This dataset is used to predict whether a patient is likely to get stroke based on the input parameters like gender, age, various diseases, and smoking status. Each row in the data provides relevant information about the patient.</p> <p><b>REGRESSION: WIND SPEED PREDICTION</b> High precision and reliable wind speed forecasting is a challenge for meteorologists. Severe wind due to convective storms, causes considerable damages (large scale forest damage, outage, buildings/houses damage, etc.). Convective events such as thunderstorms, tornadoes as well as large hail, strong winds, are natural hazards that have the potential to disrupt daily life, especially over complex terrain favoring the initiation of convection. Even ordinary convective events produce severe winds which causes fatal and costly damages. Therefore, wind speed prediction is an important task to get advanced severe weather warning. This dataset contains the responses of a weather sensor that collected different weather variables such as temperatures and precipitation.</p> <p>Perform the following tasks on the dataset given:</p> <ol style="list-style-type: none"> <li>1. Apply the required preprocessing on the dataset to apply classification / regression.</li> <li>2. Construct the classification/regression model with any 2 algorithms by writing the code in your preferable language.</li> <li>3. Evaluate the model with suitable measures.</li> <li>4. Perform the comparison analysis of your results.</li> <li>5. Discuss what your observations and conclusions.</li> </ol>	CO 1-4	Apply, Analyze and Evaluate
2	<p><b>TASK 2:</b> <b>CLASSIFICATION: HEART FAILURE PREDICTION</b> Cardiovascular diseases (CVDs) are the number 1 cause of death globally,</p>	CO 1-4	Apply, Analyze and

	<p>taking an estimated 17.9 million lives each year, which accounts for 31% of all deaths worldwide. Four out of 5 CVD deaths are due to heart attacks and strokes, and one-third of these deaths occur prematurely in people under 70 years of age. Heart failure is a common event caused by CVDs and this dataset contains 11 features that can be used to predict a possible heart disease. People with cardiovascular disease or who are at high cardiovascular risk (due to the presence of one or more risk factors such as hypertension, diabetes, hyperlipidaemia or already established disease) need early detection and management wherein a machine learning model can be of great help.</p> <p><b>REGRESSION: FUEL CONSUMPTION RATINGS</b> Dataset provides model-specific fuel consumption ratings and estimated carbon dioxide emissions for new light-duty vehicles for retail sale in Canada in 2022.</p> <p>Perform the following tasks on the dataset given:</p> <ol style="list-style-type: none"> <li>1. Apply the required preprocessing on the dataset to apply classification / regression.</li> <li>2. Construct the classification/regression model with any 2 algorithms by writing the code in your preferable language.</li> <li>3. Evaluate the model with suitable measures.</li> <li>4. Perform the comparison analysis of your results.</li> <li>5. Discuss what your observations and conclusions.</li> </ol>		Evaluate
3	<p><b>TASK 3:</b> <b>CLASSIFICATION: HEPATITIS C PREDICTION</b> An infection caused by a virus that attacks the liver and leads to inflammation. The virus is spread by contact with contaminated blood; for example, from sharing needles or from unsterile tattoo equipment. Most people have no symptoms. Those who do develop symptoms may have fatigue, nausea, loss of appetite and yellowing of the eyes and skin. Hepatitis C is treated with antiviral medication. In some people, newer medicines can eradicate the virus. Hence your task predicting the HEPATITIS C with given features.</p> <p><b>REGRESSION: DAILY GOLD PRICE PREDICTION</b> Historically gold coinage was widely used as currency; when paper money was introduced, it typically was a receipt redeemable for gold coin or bullion. In a monetary system known as the gold standard, a certain weight of gold was given the name of a unit of currency. For a long period, the United States government set the value of the US dollar so that one troy ounce was equal to \$20.67 (\$0.665 per gram), but in 1934 the dollar was devalued to \$35.00 per troy ounce (\$0.889/g). By 1961, it was becoming hard to maintain this price, and a pool of US and European banks agreed to manipulate the market to prevent further currency devaluation against increased gold demand.</p> <p>Perform the following tasks on the dataset given:</p> <ol style="list-style-type: none"> <li>1. Apply the required preprocessing on the dataset to apply classification / regression.</li> <li>2. Construct the classification/regression model with any 2 algorithms by writing the code in your preferable language.</li> <li>3. Evaluate the model with suitable measures.</li> <li>4. Perform the comparison analysis of your results.</li> </ol>	CO 1-4	Apply, Analyze and Evaluate

	Discuss what your observations and conclusions.		
4	<p><b>TASK 4:</b></p> <p><b>CLASSIFICATION: PREDICTING THE ABSENTEEISM AT WORK</b></p> <p>Workplace absences always have a personal component to them. This can make it hard for a company to discern the particular cause without having an employee divulge information about his personal life. In specific cases, that may eventually become necessary, but in general companies can fight absenteeism by making it more appealing for their employees to come to work and by showing understanding and lenience when absences occur.</p> <p>In particular, companies should strive to provide their workers with enough pay, enough dignity and a positive enough workplace so that merely showing up each day doesn't wear a worker down. These are all major factors affecting absenteeism. Absenteeism costs a lot in terms of bottom lines and productivity, according to AIHR Digital.</p> <p><b>REGRESSION: PREDICT SALES OR PROFIT AT SUPERMARKET STORE</b></p> <p>With growing demands and cut-throat competitions in the market, a Superstore Giant is seeking your knowledge in understanding what works best for them. They would like to understand which products, regions, categories and customer segments they should target or avoid. You can even take this a step further and try and build a Regression model to predict Sales or Profit.</p> <p>Go crazy with the dataset, but also make sure to provide some business insights to improve.</p> <p>Perform the following tasks on the dataset given:</p> <ol style="list-style-type: none"> <li>1. Apply the required preprocessing on the dataset to apply classification / regression.</li> <li>2. Construct the classification/regression model with any 2 algorithms by writing the code in your preferable language.</li> <li>3. Evaluate the model with suitable measures.</li> <li>4. Perform the comparison analysis of your results.</li> <li>5. Discuss what your observations and conclusions.</li> </ol>	CO 1-4	Apply, Analyze and Evaluate
5	<p><b>TASK 5:</b></p> <p><b>CLASSIFICATION: IN-VEHICLE COUPON RECOMMENDATION</b></p> <p>Coupon systems have been widely used to enhance customers' engagement in digital-based platforms. By offering users a challenge and a corresponding reward, companies' services become not only more attractive, but most importantly it can lead users to become frequent customers, thus enhancing a brand's impact on its customers. However, knowing which coupon to provide can be a rather complex task since each customer profile responds differently to each offer, and frequently offering them bad deals might drag them away from your business. To overcome this problem, machine learning techniques can be used to build data-driven customer profiles and develop better coupon recommendations.</p> <p><b>REGRESSION: CAR PRICE PREDICTION</b></p> <p>Imagine a situation where you have an old car and want to sell it. You may of course approach an agent for this and find the market price, but later may have to pay pocket money for his service in selling your car. But what if you can know your car selling price without the intervention of an agent. Or if you are</p>	CO 1-4	Apply, Analyze and Evaluate

	<p>an agent, definitely this will make your work easier. Yes, this system has already learned about previous selling prices over years of various cars.</p> <p>Perform the following tasks on the dataset given:</p> <ol style="list-style-type: none"> <li>1. Apply the required preprocessing on the dataset to apply classification / regression.</li> <li>2. Construct the classification/regression model with any 2 algorithms by writing the code in your preferable language.</li> <li>3. Evaluate the model with suitable measures.</li> <li>4. Perform the comparison analysis of your results.</li> <li>5. Discuss what your observations and conclusions.</li> </ol>		
6	<p><b>TASK 6:</b></p> <p><b>CLASSIFICATION: STUDENT PERFORMANCE</b></p> <p>Research on the educational field involving machine learning techniques has recently taken a steep growth trajectory. A new term called “Educational Data Mining” has come into existence, i.e., the application of data mining techniques in an educational background aiming to discover hidden trends and patterns about student’s performance.</p> <p>This project aims to develop a prediction model for students’ academic performance based on machine learning techniques. The resultant model can be used to identify any student’s performance for a particular subject.</p> <p><b>REGRESSION: FLIGHT PRICE PREDICTION</b></p> <p>The objective of the study is to analyse the flight booking dataset obtained from “Ease My Trip” website and to conduct various statistical hypothesis tests in order to get meaningful information from it. The 'Linear Regression' statistical algorithm would be used to train the dataset and predict a continuous target variable. 'Easemytrip' is an internet platform for booking flight tickets, and hence a platform that potential passengers use to buy tickets. A thorough study of the data will aid in the discovery of valuable insights that will be of enormous value to passengers.</p> <p>Perform the following tasks on the dataset given:</p> <ol style="list-style-type: none"> <li>1. Apply the required preprocessing on the dataset to apply classification / regression.</li> <li>2. Construct the classification/regression model with any 2 algorithms by writing the code in your preferable language.</li> <li>3. Evaluate the model with suitable measures.</li> <li>4. Perform the comparison analysis of your results.</li> <li>5. Discuss what your observations and conclusions.</li> </ol>	CO 1-4	Apply, Analyze and Evaluate
7	<p><b>TASK 7:</b></p> <p><b>CLASSIFICATION: CAMPUS RECRUITMENT</b></p> <p>Placements hold great importance for students and educational institutions. It helps a student to build a strong foundation for the professional career ahead as well as a good placement record gives a competitive edge to a college/university in the education market. This study focuses on a system that predicts if a student would be placed or not based on the student’s qualifications, historical data, and experience.</p> <p><b>REGRESSION: BIKE RENTS FOR THE DAY</b></p> <p>The objective of the study is to analyse the flight booking dataset obtained from</p>	CO 1-4	Apply, Analyze and Evaluate

	<p>“Ease My Trip” website and to conduct various statistical hypothesis tests in order to get meaningful information from it. The 'Linear Regression' statistical algorithm would be used to train the dataset and predict a continuous target variable. 'Easemytrip' is an internet platform for booking flight tickets, and hence a platform that potential passengers use to buy tickets. A thorough study of the data will aid in the discovery of valuable insights that will be of enormous value to passengers.</p> <p>Perform the following tasks on the dataset given:</p> <ol style="list-style-type: none"> <li>1. Apply the required preprocessing on the dataset to apply classification / regression.</li> <li>2. Construct the classification/regression model with any 2 algorithms by writing the code in your preferable language.</li> <li>3. Evaluate the model with suitable measures.</li> <li>4. Perform the comparison analysis of your results.</li> <li>5. Discuss what your observations and conclusions.</li> </ol>		
8	<p><b>TASK 8:</b>  <b>CLASSIFICATION: HR ANALYTICS: PREDICT WHO WILL MOVE TO A NEW JOB</b>  A company which is active in Big Data and Data Science wants to hire data scientists among people who successfully pass some courses which conduct by the company. Many people signup for their training. Company wants to know which of these candidates are really wants to work for the company after training or looking for a new employment because it helps to reduce the cost and time as well as the quality of training or planning the courses and categorization of candidates. Information related to demographics, education, experience are in hands from candidates signup and enrollment.  This dataset designed to understand the factors that lead a person to leave current job for HR researches too. By model(s) that uses the current credentials, demographics, experience data you will predict the probability of a candidate to look for a new job or will work for the company, as well as interpreting affected factors on employee decision.  target: 0 – Not looking for job change, 1 – Looking for a job change</p> <p><b>REGRESSION: MOTOR CYCLE PRICE PREDCITION</b>  This dataset contains information about used motorcycles. This data can be used for a lot of purposes such as price prediction to exemplify the use of linear regression in Machine Learning. <i>The columns in the given dataset are as follows:</i> name, selling price, year, seller type, owner, km driven, ex showroom price</p> <p>Perform the following tasks on the dataset given:</p> <ol style="list-style-type: none"> <li>1. Apply the required preprocessing on the dataset to apply classification / regression.</li> <li>2. Construct the classification/regression model with any 2 algorithms by writing the code in your preferable language.</li> <li>3. Evaluate the model with suitable measures.</li> <li>4. Perform the comparison analysis of your results.</li> <li>5. Discuss what your observations and conclusions.</li> </ol>	CO 1-4	Apply, Analyze and Evaluate
9	<p><b>TASK 9:</b></p>	CO	Apply,

	<p><b>CLASSIFICATION: MOBILE PRICE CLASSIFICATION</b>  Bob has started his own mobile company. He wants to give tough fight to big companies like Apple,Samsung etc. He does not know how to estimate price of mobiles his company creates. In this competitive mobile phone market you cannot simply assume things. To solve this problem he collects sales data of mobile phones of various companies. Bob wants to find out some relation between features of a mobile phone(eg:- RAM,Internal Memory etc) and its selling price. But he is not so good at Machine Learning. So he needs your help to solve this problem. In this problem you do not have to predict actual price but a price range indicating how high the price is. The last column is the target variable with value of 0(low cost), 1(medium cost), 2(high cost) and 3(very high cost).</p> <p><b>REGRESSION: AUTO INSURANCE PREDICTION</b>  The famous sonar dataset for quick debugging and new application testing purpose Auto Insurance in Sweden. In the following data  X = number of claims  Y = total payment for all the claims in thousands of Swedish Kronor for geographical zones in Sweden</p> <p>Perform the following tasks on the dataset given:</p> <ol style="list-style-type: none"> <li>1. Apply the required preprocessing on the dataset to apply classification / regression.</li> <li>2. Construct the classification/regression model with any 2 algorithms by writing the code in your preferable language.</li> <li>3. Evaluate the model with suitable measures.</li> <li>4. Perform the comparison analysis of your results.</li> <li>5. Discuss what your observations and conclusions.</li> </ol>	1-4	Analyze and Evaluate
10	<p><b>TASK 10:</b>  <b>CLASSIFICATION: DIABETES HEALTH INDICATORS</b>  Diabetes is among the most prevalent chronic diseases in the United States, impacting millions of Americans each year and exerting a significant financial burden on the economy. Diabetes is a serious chronic disease in which individuals lose the ability to effectively regulate levels of glucose in the blood, and can lead to reduced quality of life and life expectancy. After different foods are broken down into sugars during digestion, the sugars are then released into the bloodstream. This signals the pancreas to release insulin. Insulin helps enable cells within the body to use those sugars in the bloodstream for energy. Diabetes is generally characterized by either the body not making enough insulin or being unable to use the insulin that is made as effectively as needed. The target variable Diabetes_binary has 2 classes. 0 is for no diabetes, and 1 is for prediabetes or diabetes. This dataset has 21 feature variables.</p> <p><b>REGRESSION: STUDENT GRADE PREDICTION</b>  This data approach student achievement in secondary education of two Portuguese schools. The data attributes include student grades, demographic, social and school-related features) and it was collected by using school reports and questionnaires. Two datasets are provided regarding the performance in two distinct subjects: Mathematics (mat) and Portuguese language (por). In [Cortez and Silva, 2008], the two datasets were modeled under binary/five-level classification and regression tasks. Important note: the target attribute G3 has a</p>	CO 1-4	Apply, Analyze and Evaluate

	<p>strong correlation with attributes G2 and G1. This occurs because G3 is the final year grade (issued at the 3rd period), while G1 and G2 correspond to the 1st and 2nd period grades. It is more difficult to predict G3 without G2 and G1, but such prediction is much more useful.</p> <p>Perform the following tasks on the dataset given:</p> <ol style="list-style-type: none"> <li>1. Apply the required preprocessing on the dataset to apply classification / regression.</li> <li>2. Construct the classification/regression model with any 2 algorithms by writing the code in your preferable language.</li> <li>3. Evaluate the model with suitable measures.</li> <li>4. Perform the comparison analysis of your results.</li> <li>5. Discuss what your observations and conclusions.</li> </ol>		
11	<p><b>TASK 11:</b></p> <p><b>CLASSIFICATION: BANK LOAN DEFAULTER PREDICTION:</b>  Predict if a person will be a loan defaulter or not  Banks run into losses when a customer doesn't pay their loans on time. Because of this, every year, banks have losses in crores, and this also impacts the country's economic growth to a large extent. In this hackathon, we look at various attributes such as funded amount, location, loan, balance, etc., to predict if a person will be a loan defaulter or not.  To solve this problem, MachineHack has created a training dataset of 67,463 rows and 35 columns and a testing dataset of 28,913 rows and 34 columns.</p> <p><b>REGRESSION: BODYFAT PREDICTION</b>  Lists estimates of the percentage of body fat determined by underwater weighing and various body circumference measurements for 252 men. This data set can be used to illustrate multiple regression techniques. Accurate measurement of body fat is inconvenient/costly and it is desirable to have easy methods of estimating body fat that are not inconvenient/costly.</p> <p>Perform the following tasks on the dataset given:</p> <ol style="list-style-type: none"> <li>1. Apply the required preprocessing on the dataset to apply classification / regression.</li> <li>2. Construct the classification/regression model with any 2 algorithms by writing the code in your preferable language.</li> <li>3. Evaluate the model with suitable measures.</li> <li>4. Perform the comparison analysis of your results.</li> <li>5. Discuss what your observations and conclusions.</li> </ol>	CO 1-4	Apply, Analyze and Evaluate
12	<p><b>TASK 12:</b></p> <p><b>CLASSIFICATION: PREDICTION OF MUSIC GENRE: CLASSIFY MUSIC INTO GENRES</b>  The task is classification of music files based on the genres. Generally, people carry their favorite songs on smartphones. Songs can be of various genres. With the help of deep learning techniques, we can provide a classified list of songs to the smartphone user.</p> <p><b>REGRESSION: HOUSE PRICE PREDICTION</b>  House Price Index (HPI) is commonly used to estimate the changes in housing</p>	CO 1-4	Apply, Analyze and Evaluate

	<p>price. Since housing price is strongly correlated to other factors such as location, area, population, it requires other information apart from HPI to predict individual housing price.</p> <p>Perform the following tasks on the dataset given:</p> <ol style="list-style-type: none"> <li>1. Apply the required preprocessing on the dataset to apply classification / regression.</li> <li>2. Construct the classification/regression model with any 2 algorithms by writing the code in your preferable language.</li> <li>3. Evaluate the model with suitable measures.</li> <li>4. Perform the comparison analysis of your results.</li> <li>5. Discuss what your observations and conclusions.</li> </ol>		
13	<p><b>TASK 13:</b></p> <p><b>CLASSIFICATION: FRAUDULENT CLAIM ON CARS PHYSICAL DAMAGE:</b></p> <p>Identifying First-Party Physical Damage Fraudulence</p> <p>Team is concerned about the fraud detection accuracy as well as the key drivers that cause fraudulence. Tasked with identifying first-party physical damage fraudulence and explaining the indicators of fraudulent claims.</p> <p><b>REGRESSION: SONG POPULARITY PREDICTION</b></p> <p>Humans have greatly associated themselves with Songs &amp; Music. It can improve mood, decrease pain and anxiety, and facilitate opportunities for emotional expression. Research suggests that music can benefit our physical and mental health in numerous ways. Lately, multiple studies have been carried out to understand songs &amp; its popularity based on certain factors. Such song samples are broken down &amp; their parameters are recorded to tabulate. Predicting the Song Popularity is the main aim.</p> <p>Perform the following tasks on the dataset given:</p> <ol style="list-style-type: none"> <li>1. Apply the required preprocessing on the dataset to apply classification / regression.</li> <li>2. Construct the classification/regression model with any 2 algorithms by writing the code in your preferable language.</li> <li>3. Evaluate the model with suitable measures.</li> <li>4. Perform the comparison analysis of your results.</li> <li>5. Discuss what your observations and conclusions.</li> </ol>	CO 1-4	Apply, Analyze and Evaluate
14	<p><b>TASK 14:</b></p> <p><b>CLASSIFICATION: LUMPY SKIN DISEASE DETECTION</b></p> <p>Skin types of diseases are most common among the globe, as people get skin disease due to inheritance, environmental factors. In many cases people ignore the impact of skin disease at the early stage. In the existing system, the skin disease is identified using biopsy process which is analyzed and medicinal prescribed manually by the physicians. To overcome this manual inspection and provide promising results in short period of time.</p> <p><b>REGRESSION: DIAMOND PRICE PREDICTION</b></p> <p>The data is scrapped from Australian Diamond Importers website on 24th Feb 2022. It includes diamond features, like shape, size, colour, cut, clarity, and other features and the price in US dollars (excluding GST)</p>	CO 1-4	Apply, Analyze and Evaluate



	<p>The data is a good candidate for regression models to predict the price based on diamond features.</p> <p>Perform the following tasks on the dataset given:</p> <ol style="list-style-type: none"> <li>1. Apply the required preprocessing on the dataset to apply classification / regression.</li> <li>2. Construct the classification/regression model with any 2 algorithms by writing the code in your preferable language.</li> <li>3. Evaluate the model with suitable measures.</li> <li>4. Perform the comparison analysis of your results.</li> <li>5. Discuss what your observations and conclusions.</li> </ol>		
15	<p><b>TASK 15:</b></p> <p><b>CLASSIFICATION: TOUR &amp; TRAVELS CUSTOMER CHURN PREDICTION</b></p> <p>Predict Tour &amp; Travels Customer Churn</p> <p>Churn prediction is probably one of the most important applications of data science in the commercial sector. The thing which makes it popular is that its effects are more tangible to comprehend and it plays a major factor in the overall profits earned by the business. Churn is defined in business terms as ‘when a client cancels a subscription to a service they have been using.’ A common example is people cancelling Spotify/Netflix subscriptions. So, Churn Prediction is essentially predicting which clients are most likely to cancel a subscription i.e ‘leave a company’ based on their usage of the service. From a company point of view, it is necessary to gain this information because acquiring new customers is often arduous and costlier than retaining old ones. Hence, the insights gained from Churn Prediction helps them to focus more on the customers that are at a high risk of leaving.</p> <p><b>REGRESSION: ENGINEERING GRADUATE SALARY PREDICTION</b></p> <p>Engineering is a broad discipline that is often broken down into several sub-disciplines. Although an engineer will usually be trained in a specific discipline, he or she may become multi-disciplined through experience. Engineering is often characterized as having four main branches: chemical engineering, civil engineering, electrical engineering, and mechanical engineering. India has a total 6,214 Engineering and Technology Institutions in which around 2.9 million students are enrolled. Every year on an average 1.5 million students get their degree in engineering, but due to lack of skill required to perform technical jobs less than 20 percent get employment in their core domain. A relevant question is what determines the salary and the jobs these engineers are offered right after graduation. Various factors such as college grades, candidate skills, the proximity of the college to industrial hubs, the specialization one have, market conditions for specific industries determine this. On the basis of these various factors, your objective is to determine the salary of an engineering graduate in India.</p> <p>Perform the following tasks on the dataset given:</p> <ol style="list-style-type: none"> <li>1. Apply the required preprocessing on the dataset to apply classification / regression.</li> <li>2. Construct the classification/regression model with any 2 algorithms by writing the code in your preferable language.</li> <li>3. Evaluate the model with suitable measures.</li> </ol>	CO 1-4	Apply, Analyze and Evaluate

	<p>4. Perform the comparison analysis of your results.</p> <p>5. Discuss what your observations and conclusions.</p>		
16	<p><b>TASK 16:</b></p> <p><b>CLASSIFICATION: DRUG CLASSIFICATION</b></p> <p>The task of predicting the interactions between drugs and targets plays a key role in the process of drug discovery. There is a need to develop novel and efficient prediction approaches in order to avoid costly and laborious yet not-always-deterministic experiments to determine drug–target interactions (DTIs) by experiments alone. This database contains information about certain drug types.</p> <p><b>REGRESSION: UBER FARES PREDICTION</b></p> <p>The project is about on world's largest taxi company Uber inc. In this project, we're looking to predict the fare for their future transactional cases. Uber delivers service to lakhs of customers daily. Now it becomes really important to manage their data properly to come up with new business ideas to get best results. Eventually, it becomes really important to estimate the fare prices accurately.</p> <p>Perform the following tasks on the dataset given:</p> <ol style="list-style-type: none"> <li>1. Apply the required preprocessing on the dataset to apply classification / regression.</li> <li>2. Construct the classification/regression model with any 2 algorithms by writing the code in your preferable language.</li> <li>3. Evaluate the model with suitable measures.</li> <li>4. Perform the comparison analysis of your results.</li> <li>5. Discuss what your observations and conclusions.</li> </ol>	CO 1-4	Apply, Analyze and Evaluate
17	<p><b>TASK 17:</b></p> <p><b>CLASSIFICATION: TRAVEL INSURANCE PREDICTION</b></p> <p>A Tour &amp; Travels Company Is Offering Travel Insurance Package To Their Customers. The New Insurance Package Also Includes Covid Cover. The Company Requires To Know Which Customers Would Be Interested To Buy It Based On Its Database History. The Insurance Was Offered To Some Of The Customers In 2019 And The Given Data Has Been Extracted From The Performance/Sales Of The Package During That Period. The Data Is Provided For Almost 2000 Of Its Previous Customers And You Are Required To Build An Intelligent Model That Can Predict If The Customer Will Be Interested To Buy The Travel Insurance Package Based On Certain Parameters Given.</p> <p><b>REGRESSION: REVISITING A CONCRETE STRENGTH</b></p> <p>Concrete is the most important material in civil engineering. The concrete compressive strength is a highly nonlinear function of age and ingredients. Your task is to measure the concrete strength based on the attributes given.</p> <p>Perform the following tasks on the dataset given:</p> <ol style="list-style-type: none"> <li>1. Apply the required preprocessing on the dataset to apply classification / regression.</li> <li>2. Construct the classification/regression model with any 2 algorithms by writing the code in your preferable language.</li> </ol>	CO 1-4	Apply, Analyze and Evaluate

	<p>3. Evaluate the model with suitable measures.</p> <p>4. Perform the comparison analysis of your results.</p> <p>5. Discuss what your observations and conclusions.</p>		
18	<p><b>TASK 18:</b></p> <p><b>CLASSIFICATION: HR ATTRITION DATA BASED ON IBM ATTRITION</b></p> <p>The dataset contains the same features as the IBM dataset with a few added features to help us with the project, which includes the survival analysis and prediction of an employee, whether he/she would attrite or not.</p> <p><b>REGRESSION: INSURANCE PREMIUM PREDICTION</b></p> <p>All of us wish to achieve financial freedom at some point in our life, and when it comes to doing that, we tend to believe that savings are enough to be financially stable. But, if you look at life from a practical perspective, you would understand that savings alone are not enough to achieve financial freedom; insuring your assets with general insurance policies is equally important. Hence predict the insurance premium amount to be paid of the persons based on their data.</p> <p>Perform the following tasks on the dataset given:</p> <ol style="list-style-type: none"> <li>1. Apply the required preprocessing on the dataset to apply classification / regression.</li> <li>2. Construct the classification/regression model with any 2 algorithms by writing the code in your preferable language.</li> <li>3. Evaluate the model with suitable measures.</li> <li>4. Perform the comparison analysis of your results.</li> <li>5. Discuss what your observations and conclusions.</li> </ol>	CO 1-4	Apply, Analyze and Evaluate
19	<p><b>TASK 19:</b></p> <p><b>CLASSIFICATION: PREDICTION SALARY IS GREATER THAN 50K</b></p> <p>The dataset provided predictive feature like education , employment status , marital status to predict if the salary is greater than \$50K.</p> <p><b>REGRESSION: AVACADO PRICE PREDICTION</b></p> <p>Avocado is a fruit consumed by people heavily in the United States.This data was downloaded from the Hass Avocado Board website in May of 2018 &amp; compiled into a single CSV. The table below represents weekly 2018 retail scan data for National retail volume (units) and price. Retail scan data comes directly from retailers' cash registers based on actual retail sales of Hass avocados. Starting in 2013, the table below reflects an expanded, multi-outlet retail data set. Multi-outlet reporting includes an aggregation of the following channels: grocery, mass, club, drug, dollar and military. The Average Price (of avocados) in the table reflects the as the class which we need to predict.</p> <p>Perform the following tasks on the dataset given:</p> <ol style="list-style-type: none"> <li>1. Apply the required preprocessing on the dataset to apply classification / regression.</li> <li>2. Construct the classification/regression model with any 2 algorithms by writing the code in your preferable language.</li> <li>3. Evaluate the model with suitable measures.</li> <li>4. Perform the comparison analysis of your results.</li> <li>5. Discuss what your observations and conclusions.</li> </ol>	CO 1-4	Apply, Analyze and Evaluate

20	<p><b>TASK 20:</b></p> <p><b>CLASSIFICATION: PANCREATIC CANCER DETECTION</b></p> <p>Pancreatic cancer is an extremely deadly type of cancer. Once diagnosed, the five-year survival rate is less than 10%. However, if pancreatic cancer is caught early, the odds of surviving are much better. Unfortunately, many cases of pancreatic cancer show no symptoms until the cancer has spread throughout the body. A diagnostic test to identify people with pancreatic cancer could be enormously helpful. They gathered a series of biomarkers from the urine of three groups of patients: Healthy controls, Patients with non-cancerous pancreatic conditions, like chronic pancreatitis, Patients with pancreatic ductal adenocarcinoma. When possible, these patients were age- and sex-matched. The goal was to develop an accurate way to identify patients with pancreatic cancer. The goal in this dataset is predicting diagnosis, and more specifically, differentiating between 3 (pancreatic cancer) versus 2 (non-cancerous pancreas condition) and 1 (healthy).</p> <p><b>REGRESSION: CALORIE, EXERCISE AND WEIGHT CHANGES</b></p> <p>When losing weight, more physical activity increases the number of calories your body uses for energy or “burns off.” The burning of calories through physical activity, combined with reducing the number of calories you eat, creates a “calorie deficit” that results in weight loss. Here your task to predict the change in the weight with activities done.</p> <p>Perform the following tasks on the dataset given:</p> <ol style="list-style-type: none"> <li>1. Apply the required preprocessing on the dataset to apply classification / regression.</li> <li>2. Construct the classification/regression model with any 2 algorithms by writing the code in your preferable language.</li> <li>3. Evaluate the model with suitable measures.</li> <li>4. Perform the comparison analysis of your results.</li> <li>5. Discuss what your observations and conclusions.</li> </ol>	CO 1-4	Apply, Analyze and Evaluate
21	<p><b>TASK 21:</b></p> <p><b>CLASSIFICATION: SYMPTOMS AND COVID PRESENCE</b></p> <p>Corona Virus disease (COVID-19) is an infectious disease caused by a newly discovered virus, which emerged in Wuhan, China in December of 2019. Most people infected with the COVID- 19 virus will experience mild to moderate respiratory illness and recover without requiring special treatment. Older people and those with underlying medical problems like cardiovascular disease, diabetes, chronic respiratory disease, and cancer are more likely to develop serious illness. The COVID-19 virus spreads primarily through droplets of saliva or discharge from the nose when an infected person coughs or sneezes, so you might have heard caution to practice respiratory etiquette (for example, by coughing into a flexed elbow). The purpose of this dataset is to provide symptoms as input and it should be able to predict if COVID is possibly present or not.</p> <p><b>REGRESSION: PIZZA PRICE PREDICTION</b></p> <p>Task is predicting the price of pizza based on company, diameter, topping, variant size, extra_sauce, and extra_cheese.</p> <p>Perform the following tasks on the dataset given:</p>	CO 1-4	Apply, Analyze and Evaluate

	<ol style="list-style-type: none"> <li>1. Apply the required preprocessing on the dataset to apply classification / regression.</li> <li>2. Construct the classification/regression model with any 2 algorithms by writing the code in your preferable language.</li> <li>3. Evaluate the model with suitable measures.</li> <li>4. Perform the comparison analysis of your results.</li> <li>5. Discuss what your observations and conclusions.</li> </ol>		
22	<p><b>TASK 22:</b></p> <p><b>CLASSIFICATION: CIRRHOSIS PREDICTION</b>  Cirrhosis is a late stage of scarring (fibrosis) of the liver caused by many forms of liver diseases and conditions, such as hepatitis and chronic alcoholism. The following data contains the information collected from the Mayo Clinic trial in primary biliary cirrhosis (PBC) of the liver.</p> <p><b>REGRESSION: CRAB AGE PREDICTION</b>  Crab is very tasty and many countries of the world import huge amount of crabs for consumption every year. The main benefits of crab farming are, labor cost is very low, production cost is comparatively lower and they grow very fast. Commercial crab farming business is developing the lifestyle of the people of coastal areas. By proper care and management we can earn more from crab farming business than shrimp farming. You can raise mud crabs in two systems. Grow out farming and fattening systems. For a commercial crab farmer knowing the right age of the crab helps them decide if and when to harvest the crabs. Beyond a certain age, there is negligible growth in crab's physical characteristics and hence, it is important to time the harvesting to reduce cost and increase profit. The goal of the dataset is to Build a regression model to predict the age of the Crab.</p> <p>Perform the following tasks on the dataset given:</p> <ol style="list-style-type: none"> <li>1. Apply the required preprocessing on the dataset to apply classification / regression.</li> <li>2. Construct the classification/regression model with any 2 algorithms by writing the code in your preferable language.</li> <li>3. Evaluate the model with suitable measures.</li> <li>4. Perform the comparison analysis of your results.</li> <li>5. Discuss what your observations and conclusions.</li> </ol>	CO 1-4	Apply, Analyze and Evaluate
23	<p><b>TASK 23:</b></p> <p><b>CLASSIFICATION: AIRLINE PASSENGER SATISFACTION</b>  This dataset contains an airline passenger satisfaction survey. What factors are highly correlated to a satisfied (or dissatisfied) passenger? Can you predict passenger satisfaction?</p> <p><b>REGRESSION: AUTOMOBILE MILEAGE PREDICTION</b>  Transportation is an important factor that affects energy consumption, and driving behavior is one of the main factors affecting vehicle fuel consumption. Mileage of a vehicle is very important for efficient fuel consumption. Hence the task is predicting the mileage of a vehicle based on no.of cylinders, displacement, horsepower, weight, acceleration, model year, and origin.</p> <p>Perform the following tasks on the dataset given:</p> <ol style="list-style-type: none"> <li>1. Apply the required preprocessing on the dataset to apply classification /</li> </ol>	CO 1-4	Apply, Analyze and Evaluate

	<p>regression.</p> <ol style="list-style-type: none"> <li>2. Construct the classification/regression model with any 2 algorithms by writing the code in your preferable language.</li> <li>3. Evaluate the model with suitable measures.</li> <li>4. Perform the comparison analysis of your results.</li> <li>5. Discuss what your observations and conclusions.</li> </ol>		
24	<p><b>TASK 24:</b></p> <p><b>CLASSIFICATION: DOG BREED CLASSIFICATION</b></p> <p>The dataset contains 97 unique dog breeds and their various attributes. The first column represents the breed name. The following 2 columns indicate the general life expectancy. Then the following 8 columns are the weight and height of the male and female gender. After that, there are 12 columns that represent different attributes of dog breeds like coat length, playfulness, and trainability. These columns have values ranging from 0 to 5. Example: A breed with energy value 5 shows that the corresponding breeds have very high energy. Build a classifier to predict the breed of the dogs based on the remaining features</p> <p><b>REGRESSION: SALARY PREDICTION</b></p> <p>This dataset contains information about the salaries of employees at a company. Each row represents a different employee, and the columns include information such as age, gender, education level, job title, years of experience, and salary. Train a model in order to predict the salary to predict the paid to a person</p> <p>Perform the following tasks on the dataset given:</p> <ol style="list-style-type: none"> <li>1. Apply the required preprocessing on the dataset to apply classification / regression.</li> <li>2. Construct the classification/regression model with any 2 algorithms by writing the code in your preferable language.</li> <li>3. Evaluate the model with suitable measures.</li> <li>4. Perform the comparison analysis of your results.</li> </ol> <p>Discuss what your observations and conclusions.</p>		