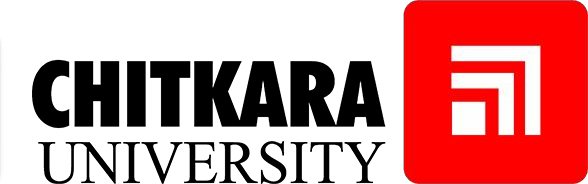
Artificial Intelligence and Machine Learning

Project Report Semester-IV (Batch-2022)

Case Study - Train Dataset With MYSQL CONNECTION

URL:-

<https://drive.google.com/file/d/1-mt8d-8a1GnUZFY8Gz_FxIxnLNJlCGX9/view?usp=sharing>

**Supervised By: Submitted By:**

Rajeev Bhardwaj Suyash Kumar Sharma

2210991000

**G-13**

**Department of Computer Science and Engineering**

Chitkara University Institute of Engineering & Technology,

**Chitkara University, Punjab**

**Description about Case Study:-**

* Display Top 5 Rows of the Dataset:

How can we view the top 5 rows of our dataset?

* Check Last 3 Rows of the Dataset:

How can we check the last 3 rows of our dataset?

* Find Shape of Our Dataset (Number of Rows & Number of Columns):

Can you provide the shape of our dataset? How many rows and columns does it have?

* Get Information About Our Dataset:

What information can you provide about our dataset, such as the total number of rows, total number of columns, data type of each column, and memory requirement?

* Get Overall Statistics About the DataFrame:

How can we get overall statistics about the dataframe, such as mean, standard deviation, minimum, maximum, and quartiles for numerical columns?

* Data Filtering:

How can we filter the data to select specific rows based on certain conditions?

* Check Null Values in the Dataset:

Are there any null values present in our dataset? If yes, in which columns and how many null values are there?

* Drop the Column:

How can we drop a specific column from our dataset?

* Handle Missing Values:

How can we handle missing values in our dataset, for example, by replacing them with the mean of each column?

* Categorical Data Encoding:

How can we encode categorical data into numerical format, for example, using one-hot encoding?

* What is Univariate Analysis:

What does univariate analysis entail, and how can we perform it on our dataset?

* How Many People Survived and How Many Died Plot on Graph:

Can you provide a visualization showing the number of people who survived and the number of people who died?

* How Many Passengers Were in Each Passenger Class Plot on Graph:

Can you visualize the number of passengers in each passenger class?

* Number of Male and Female Passengers:

How can we determine the number of male and female passengers in our dataset?

* Bivariate Analysis:

What is bivariate analysis, and how can we perform it to analyze relationships between variables in our dataset?

* How Has Better Chance of Survival: Male or Female:

Can you analyze and visualize the survival rate by gender to determine if males or females had a better chance of survival?

* Which Passenger Class Has Better Chance of Survival (First, Second, or Third Class):

How can we analyze and visualize the survival rate by passenger class to determine which class had a better chance of survival?

* Feature Engineering:

How can we create new features from existing ones to improve our analysis, for example, by combining 'SibSp' and 'Parch' to create a 'total\_family\_members' feature?

# Library:-.

* Pandas
* Mysql Connector
* Matplotlib

# Methods:-

# Display Top 5 Rows of the Dataset:

# head() method from pandas DataFrame.

# Check Last 3 Rows of the Dataset:

# tail() method from pandas DataFrame.

# Find Shape of Our Dataset (Number of Rows & Number of Columns):

# shape attribute from pandas DataFrame.

# Get Information About Our Dataset:

# info() method from pandas DataFrame.

# Get Overall Statistics About the DataFrame:

# describe() method from pandas DataFrame.

# Data Filtering:

# Boolean indexing or using methods like query() from pandas DataFrame.

# Check Null Values in the Dataset:

# isnull().sum() method from pandas DataFrame.

# Drop the Column:

# drop() method from pandas DataFrame.

# Handle Missing Values:

# Methods like fillna() from pandas DataFrame.

# Categorical Data Encoding:

# get\_dummies() method from pandas DataFrame.

# What is Univariate Analysis:

# Analyzing a single variable at a time. Various statistical methods and visualizations can be used for this purpose.

# How Many People Survived and How Many Died Plot on Graph:

# Visualization using methods like value\_counts() followed by plot().

# How Many Passengers Were in Each Passenger Class Plot on Graph:

# Visualization using methods like value\_counts() followed by plot().

# Number of Male and Female Passengers:

# value\_counts() method from pandas Series.

# Bivariate Analysis:

# Analyzing two variables simultaneously. Grouping and aggregation functions like groupby() followed by statistical analysis and visualizations are common methods.

# How Has Better Chance of Survival: Male or Female:

# Statistical analysis and visualization using methods like groupby() followed by plot().

# Which Passenger Class Has Better Chance of Survival (First, Second, or Third Class):

# Statistical analysis and visualization using methods like groupby() followed by plot().

# Feature Engineering:

# Creation of new features from existing ones. This can involve arithmetic operations, combining columns, or creating dummy variables.