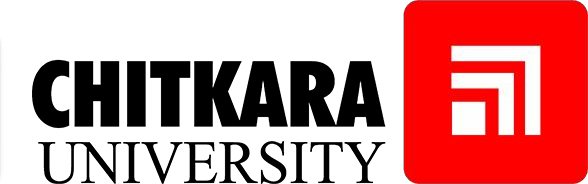
Artificial Intelligence and Machine Learning

Project Report Semester-IV (Batch-2022)

Case Study - Adult Dataset

URL:-

https://drive.google.com/file/d/1BmIyOArZrFJPXwc2PPZN637QeozO4cQq/view?usp=sharing



**Supervised By: Submitted By:**

Rajeev Bhardwaj Tanisha

2210990891

**G-13**

**Department of Computer Science and Engineering**

Chitkara University Institute of Engineering & Technology,

**Chitkara University, Punjab**

**Description about Case Study:-**

* Display Top 10 Rows:

How can we view the top 10 rows of our dataset?

* Display Last 10 Rows:

How can we check the last 10 rows of our dataset?

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

Where can we find the shape of our dataset, i.e., the number of rows and columns?

* Getting Information About Our Dataset:

How do we obtain information about our dataset, including the total number of rows, total number of columns, data type of each column, and memory requirements?

* Fetch Random Samples from Dataset (50%):

How do we randomly sample 50% of the data from our dataset?

* Check Null Values in Dataset:

How do we identify and handle null values in our dataset?

* Perform Data Cleaning (Replace '?' with NaN) and Plot in Graph with Seaborn Library:

How can we replace occurrences of '?' with NaN in our dataset as part of data cleaning, and then plot the results using the Seaborn library?

* Drop All Rows Having Missing Values:

How do we remove all rows from our dataset that contain missing values?

* Check for Duplicate Data and Drop Them:

How can we identify and drop duplicate rows in our dataset?

* Get Overall Statistics About the DataFrame:

How do we obtain overall statistical information about the entire DataFrame?

* Bivariate Analysis on Graph:

How can we perform bivariate analysis and visualize relationships between two variables using graphs?

* Replace Salary Values ['<=50k', '>50k'] with 0 and 1:

How do we convert salary values '<=50k' and '>50k' to numerical values 0 and 1 in our dataset?

* Which Workclass is Getting the Highest Salary?:

How can we determine which workclass has the highest average salary in our dataset?

* Who has a Better Chance to Get Salary >50k, Male or Female?:

How do we analyze and compare the chances of getting a salary >50k between male and female individuals?

* Convert Workclass Column Datatype to Category Data:

How can we convert the data type of the 'Workclass' column to the category data type?

# Library:-.

* Pandas
* Seaborn
* Matplotlib

# Methods:-

1. Display Top 10 Rows:

Method: Use the head() function.

1. Display Last 10 Rows:

Method: Use the tail() function.

1. Find Shape of Our Dataset (Number of Rows and Number of Columns):

Method: Use the shape attribute.

1. Getting Information About Our Dataset:

Method: Use the info() function.

1. Fetch Random Samples from Dataset (50%):

Method: Use the sample() function.

1. Check Null Values in Dataset:

Method: Use the isnull() function.

1. Perform Data Cleaning (Replace '?' with NaN) and Plot in Graph with Seaborn Library:

Method: Use the replace() function for data cleaning and Seaborn library for plotting.

1. Drop All Rows Having Missing Values:

Method: Use the dropna() function.

1. Check for Duplicate Data and Drop Them:

Method: Use the duplicated() function and drop\_duplicates() function.

1. Get Overall Statistics About the DataFrame:

Method: Use the describe() function.

1. Bivariate Analysis on Graph:

Method: Use the Seaborn or Matplotlib library for plotting bivariate analysis.

1. Replace Salary Values ['<=50k', '>50k'] with 0 and 1:

Method: Use the replace() function.

1. Which Workclass is Getting the Highest Salary?:

Method: Use the groupby() function and calculate the mean.

1. Who has a Better Chance to Get Salary >50k, Male or Female?:

Method: Use the groupby() function and calculate the percentage.

1. Convert Workclass Column Datatype to Category Data:

Method: Use the astype() function.