Assi-5 Data Analytics II

CSV File / Dataset - Social_Network_Ads.csv

Required Libraries-

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import standardscalar
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import Confusion_matrix,
classification report, accuracy score, precision_score,
recall score, F1_score.
import Warnings
```

Functions Used-

```
df = pd.read_data csv("Social_Network_Ads.csv")
df.head()
df.shape
df.info()
df.describe()
df.isnull().sum()
histplot = sns.histplot()
plt.show()
Draw histogram for each column
df["column name"].value_counts()
Countplot = sns.Countplot()
sns.heatmap()
```

- Data preparation
- Model building
- Evaluation

Oral Questions With Answers

Recall.

① Explain confusion Matrix with Accuracy, Error Rate, Precision & Recall.

→ - It contain Actual value & Predicted value.
- Terms included in confusion ~~Matrix~~ Matrix are
  i) True Negative (TN)
  ii) True positive (TP)
  iii) false Positive (FP)
  iv) False Negative (FN)

eg. Patient have ~~Dis~~ Disease with samples 165.

Confusion Matrix

Actual \ Predicted	NO	Yes	
NO	50 [TN]	10 [FP]	60
Yes	5 [FN]	100 [TP]	105
	55	110	

Actual value - which are already true, it's reality
Predicted values - After some experiment ~~tests~~ took some observations & this is the prediction it will happed or not

Accuracy = $\frac{TP + TN}{Total} = \frac{100 + 50}{165} = 0.91$

Error rate = 1 - Accuracy or $\frac{FP + FN}{Total} = 0.09$

Precision = $\frac{TP}{predicted\ yes} = \frac{100}{110} = 0.64$

Recall = $\frac{TP}{actual\ yes} = \frac{100}{105} = 0.95$