**U18ISI6204 – Machine Learning Techniques**

**LAB- EXPERIMENT 9**

**NAME:** Navina G A

**ROLL\_NO:** 20BIS028

Using Principal component Analysis as Dimensionality reduction component implement Logistic Regression for detecting credit card frauds

# OBJECTIVE OF THE EXERCISE/EXPERIMENT

To perform logistic regression along with PCA on the given dataset, using scikit library

# STEP 2: ACQUISITION PROCEDURE:

**STEP-1:** Start the program.

**STEP-2:** import all the necessary libraries

1. Numpy – array manipulation
2. Pandas – dataframe manipulation
3. Matplotlib and seaborn – for data visualization
4. Sklearn.model\_selection – train test data split
5. Sklearn.metrics –f1 score.
6. Sklearn,linear\_model– for logistic regression
7. Sklearn.decomposition – for PCA
8. Sklearn.preprocessing – for Normalisation

**STEP-3:** Loading the dataset using read\_csv method in pandas module.

**STEP-4:** Analyze the dataset using info method, which gives its data types and number of non- null values in each columns.

**STEP-5:** Perform basic statistic operation using describe() method.

**STEP-6:** Use heatmaps, correlation matrix, regression plots and pairplots in seaborn to find the relationship between features.

**STEP-7**: Normalize the data points

**STEP-8**: Using selective feature, perform PCA in order to reduce number of feature from 30 to 11.

**STEP-9:** Implement logistic regression with 11 PCA variable and calculate f1 score.

**STEP-10:** Stop the program.

# PROGRAM:

# from sklearn.preprocessing import StandardScaler

# from sklearn.decomposition import PCA

# from sklearn.metrics import f1\_score

# from sklearn.linear\_model import LogisticRegression

# from sklearn.model\_selection import train\_test\_split, cross\_val\_score

# import pandas as pd

# import matplotlib.pyplot as plt

# import seaborn as sns

# import numpy as npA screenshot of a computer code Description automatically generated with low confidence

# Loading dataset

# df=pd.read\_csv("C:/Users/Sankamethra/Documents/3rdYear/ML/LAB/archive (9)/breast-cancer.csv")

# df.head()

# 

# print(df.info())

# df.describe()

# A screenshot of a computer Description automatically generated

# A screenshot of a computer Description automatically generated with medium confidence

# Correlation between columns

# sns.heatmap(df.corr(),annot=True)

# 

**Missing rows for the values in age.**

# df[df['Age'].isnull()]

# 

# Filling the missing values in age with median of corresponding data values in pclass and sex.

# A picture containing text, screenshot, font Description automatically generated

**Filling missing values in fare with median of datas corresponding to pclass Sibsp and Parch columns.**

