PRACTICAL NO 5

Code :

```
import pandas as pd
import numpy as np
import statsmodels.api as sm
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, confusion_matrix,
classification_report

# Load the dataset
url = "https://stats.idre.ucla.edu/stat/data/binary.csv"
admissions_data = pd.read_csv(url)

# Preprocess the data
# For simplicity, we'll assume there are no missing values and the data is
clean
# Define predictor (X) and target (y) variables
X = admissions_data[['gre', 'gpa', 'rank']]
y = admissions_data['admit']

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
random_state=42)

# Perform logistic regression
# Add intercept term
X_train = sm.add_constant(X_train)
X_test = sm.add_constant(X_test)

# Fit logistic regression model
logit_model = sm.Logit(y_train, X_train)
result = logit_model.fit()

# Print summary of the model
print(result.summary())

# Evaluate the model
# Predict on test set
y_pred = result.predict(X_test)
y_pred_class = np.where(y_pred > 0.5, 1, 0)  # Convert predicted
probabilities to binary classes

# Model accuracy
accuracy = accuracy_score(y_test, y_pred_class)
print("\nAccuracy:", accuracy)

# Confusion matrix
conf_matrix = confusion_matrix(y_test, y_pred_class)
print("\nConfusion Matrix:")
print(conf_matrix)

# Classification report
class_report = classification_report(y_test, y_pred_class)
print("\nClassification Report:")
print(class_report)
```

Output :

```
         Current function value: 0.577245
         Iterations 6
                           Logit Regression Results
==============================================================================
Dep. Variable:                  admit   No. Observations:                  280
Model:                          Logit   Df Residuals:                      276
Method:                           MLE   Df Model:                            3
Date:                Tue, 13 Feb 2024   Pseudo R-squ.:                 0.07268
Time:                        08:47:20   Log-Likelihood:                -161.63
converged:                       True   LL-Null:                       -174.30
Covariance Type:            nonrobust   LLR p-value:                 1.314e-05
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         -3.4637      1.368     -2.532      0.011      -6.145      -0.782
gre            0.0016      0.001      1.218      0.223      -0.001       0.004
gpa            0.8822      0.390      2.265      0.024       0.119       1.646
rank          -0.5205      0.151     -3.457      0.001      -0.816      -0.225
==============================================================================

Accuracy: 0.725

Confusion Matrix:
[[78  3]
 [30  9]]

Classification Report:
              precision    recall  f1-score   support

           0       0.72      0.96      0.83        81
           1       0.75      0.23      0.35        39

    accuracy                           0.73       120
   macro avg       0.74      0.60      0.59       120
weighted avg       0.73      0.72      0.67       120
```