Data Warehousing LAB - 11

Chaitanya Sai Nutakki

AP21110010253

Code:

```
# Import necessary libraries import pandas as pd from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB from sklearn.metrics import accuracy_score,
classification_report from sklearn.preprocessing import LabelEncoder, StandardScaler

df = pd.read_csv('C:\\Users\\Vijay\\Saved Games\\breast.csv')

label_encoder = LabelEncoder()
categorical_columns = ['type_of_breast_surgery', 'cancer_type', 'cancer_type_detailed', 'cellularity',
'pam50_+_claudin-low_subtype',
'cohort', 'er_status_measured_by_ihc'] for column in categorical_columns:
    df[column] = label_encoder.fit_transform(df[column])
X = df[['age_at_diagnosis', 'chemotherapy']]

y = df['er_status_measured_by_ihc']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test) classifier
= GaussianNB()

classifier.fit(X_train, y_train) predictions
= classifier.predict(X_test)

accuracy = accuracy_score(y_test, predictions) print("Accuracy:", accuracy) report =
classification_report(y_test, predictions) print("Classification Report:\n", report)
```

Output:

```
Accuracy: 0.1732283464566929
Classification Report:
              precision    recall  f1-score   support

           0       0.69      0.69      0.69        84
           1       0.00      0.00      0.00       289
           2       0.03      1.00      0.05         8

    accuracy                           0.17       381
   macro avg       0.24      0.56      0.25       381
weighted avg       0.15      0.17      0.15       381
```