NATIONAL UNIVERSITY OF MODERN LANGUAGES

ISLAMABAD

MACHINE LEARNING (LAB)

Assignment: 02

Submitted to
Ms. Qurat Raja

Submitted By
Junaid Asif
(BSAI-144)

Submission Date: October 22, 2024

PART-1

Dataset Import and Exploration

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score
```
✓ 0.0s

```
# Load dataset
df = pd.read_csv('titanic.csv')
df.head()
```
✓ 0.0s

	PassengerId	Survived	Pclass	Name	Gender	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

```
# description
df.describe()
```
✓ 0.0s

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

```
# features
df.dtypes
```
✓ 0.0s

```
PassengerId      int64
Survived         int64
Pclass           int64
Name            object
Gender          object
Age            float64
SibSp            int64
Parch            int64
Ticket          object
Fare           float64
Cabin           object
Embarked        object
dtype: object
```

Data Preprocessing

Certainly! Here's how we can properly organize the data preprocessing steps for the Titanic dataset before splitting and training our model.

```
# Dropping irrelevant columns
df = df.drop(['PassengerId', 'Name', 'Ticket', 'Cabin'], axis=1, errors='ignore')

# Handling missing values
df['Age'].fillna(df['Age'].median(), inplace=True)
df['Embarked'].fillna(df['Embarked'].mode()[0], inplace=True)

# Encoding categorical data
df['Gender'] = df['Gender'].map({'male': 0, 'female': 1})
df = pd.get_dummies(df, columns=['Embarked'], drop_first=True)  # One-hot encoding for Embarked
```

PART-2

Data Splitting

```
# Features and target variable
x = df.drop('Survived', axis=1)
y = df['Survived']

# Splitting the data into training and testing sets (80% training, 20% testing)
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=42)
```

✓ 0.0s

PART-3

Naïve Bayesian Implementation

```
# Initialize the Naive Bayes classifier
gnb = GaussianNB()

# Train the classifier on the training data
gnb.fit(x_train, y_train)

# Predict the test data
y_pred = gnb.predict(x_test)
```

✓ 0.0s

```
# Calculate the accuracy
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy of the Naive Bayes classifier: {accuracy * 100:.2f}%')
```

✓ 0.0s

```
Accuracy of the Naive Bayes classifier: 77.09%
```