

STATISTICAL MACHINE LEARNING

B Rithwik

2303A52330

ASSIGNMENT 6

Question 1

Code –

```
from google.colab import drive
```

```
drive.mount('/content/drive')
```

```
import pandas as pd
```

```
from sklearn.model_selection import train_test_split
```

```
from sklearn.ensemble import RandomForestClassifier
```

```
from sklearn.metrics import accuracy_score
```

```
data = pd.read_csv("/content/drive/MyDrive/SML Dataset/Assignment_6.csv")
```

```
X = data.drop('Pass/Fail', axis=1)
```

```
y = data['Pass/Fail']
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
```

```
clf = RandomForestClassifier(random_state=42)
```

```
clf.fit(X_train, y_train)
```

```
y_pred = clf.predict(X_test)
```

```
accuracy = accuracy_score(y_test, y_pred)
```

```
print(f"Model Accuracy: {accuracy * 100:.2f}%")
```

OUTPUT –

Model Accuracy: 100.00%

Question 2

Code –

```
import pandas as pd

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.linear_model import Perceptron
from sklearn.metrics import accuracy_score

df = pd.read_csv("/content/drive/MyDrive/SML Dataset/diabetes_data_upload.csv")
df = df.dropna()

label_encoder = LabelEncoder()
df['Gender'] = label_encoder.fit_transform(df['Gender'])

for column in df.columns:
    if df[column].isin(['Yes', 'No']).any():
        df[column] = label_encoder.fit_transform(df[column])

X = df.drop('class', axis=1)
y = df['class']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

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

perceptron = Perceptron(max_iter=1000, tol=1e-3, random_state=42)
perceptron.fit(X_train, y_train)

y_pred = perceptron.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
print(f"Model accuracy: {accuracy * 100:.2f}%")
```

OUTPUT –

Model accuracy: 89.74%