EXP NO. 01	
DATE: 24.01.2025	Univariate, Bivariate and Multivariate Regression

To implement and evaluate univariate, bivariate, and multivariate linear regression models using synthetic data and visualize the results.

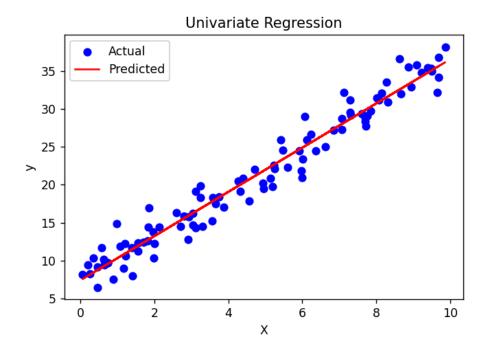
ALGORITHM:

- **Step 1:** Import the necessary libraries (NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn).
- **Step 2:** Set a random seed for reproducibility.
- **Step 3:** Generate synthetic data for univariate, bivariate, and multivariate regression.
- **Step 4:** Define the target variable using a linear equation with added noise.
- **Step 5:** Fit a Linear Regression model to the data.
- **Step 6:** Predict the output using the trained model.
- Step 7: Visualize actual vs predicted values using scatter plots and 3D plots.
- **Step 8:** Calculate and display performance metrics (MSE and R² Score).
- **Step 9:** End the program.

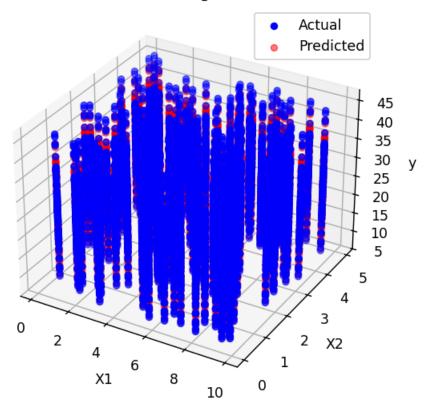
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.linear model import LinearRegression
from sklearn.metrics import mean squared error, r2 score
from mpl toolkits.mplot3d import Axes3D
# Set random seed
np.random.seed(42)
# --- 1. UNIVARIATE REGRESSION ---
# Simulate data
X uni = np.random.rand(100, 1) * 10
y uni = 3 * X uni.squeeze() + 7 + np.random.randn(100) * 2
# Fit model
model uni = LinearRegression().fit(X uni, y uni)
y uni pred = model uni.predict(X uni)
# Plot
plt.figure(figsize=(6,4))
plt.scatter(X uni, y uni, label="Actual", color="blue")
plt.plot(X uni, y uni pred, label="Predicted", color="red")
plt.title("Univariate Regression")
plt.xlabel("X")
plt.ylabel("y")
plt.legend()
```

```
plt.show()
# Metrics
print("Univariate Regression:")
print("MSE:", mean squared error(y uni, y uni pred))
print("R² Score:", r2 score(y uni, y uni pred))
print()
# --- 2. BIVARIATE REGRESSION ---
# Simulate data
X1 = np.random.rand(100, 1) * 10
X2 = np.random.rand(100, 1) * 5
X \text{ bi = np.hstack}([X1, X2])
y bi = 2 * X1.squeeze() + 4 * X2.squeeze() + 5 + np.random.randn(100) * 2
# Fit model
model bi = LinearRegression().fit(X bi, y bi)
y bi pred = model bi.predict(X bi)
#3D plot
fig = plt.figure(figsize=(7,5))
ax = fig.add subplot(111, projection='3d')
ax.scatter(X1, X2, y bi, c='blue', label='Actual')
ax.scatter(X1, X2, y bi pred, c='red', label='Predicted', alpha=0.5)
ax.set xlabel("X1")
ax.set ylabel("X2")
ax.set zlabel("y")
ax.set title("Bivariate Regression")
plt.legend()
```

```
plt.show()
# Metrics
print("Bivariate Regression:")
print("MSE:", mean squared error(y bi, y bi pred))
print("R² Score:", r2 score(y bi, y bi pred))
print()
# --- 3. MULTIVARIATE REGRESSION ---
# Simulate data
X multi = np.random.rand(100, 5)
coeffs = np.array([2, -1, 3, 0.5, 4])
y multi = X multi @ coeffs + 10 + np.random.randn(100) * 2
# Fit model
model multi = LinearRegression().fit(X multi, y multi)
y multi pred = model multi.predict(X multi)
# Plot residuals
plt.figure(figsize=(6,4))
sns.histplot(y multi - y multi pred, kde=True)
plt.title("Residuals - Multivariate Regression")
plt.xlabel("Residuals")
plt.show()
# Metrics
print("Multivariate Regression:")
print("MSE:", mean squared error(y multi, y multi pred))
print("R<sup>2</sup> Score:", r2 score(y multi, y multi pred))
print()
```



Bivariate Regression



RESULT:

The univariate, bivariate, and multivariate linear regression models were successfully implemented, and the predicted outputs closely matched the actual values with high R² scores and low mean squared errors, indicating good model performance.

EXP NO. 02	Simple Linear Regression using Least Square
DATE: 31.01.2025	Method

To implement simple linear regression using the Least Squares Method and evaluate the model performance using Mean Squared Error and R² Score.

ALGORITHM:

- **Step 1:** Import the required libraries (NumPy and Matplotlib).
- **Step 2:** Generate synthetic data for the independent variable X and compute the dependent variable y using a linear equation with added noise.
- **Step 3:** Calculate the mean of X and y.
- Step 4: Compute the slope and intercept using the Least Squares formula.
- **Step 5:** Predict the output values y_pred using the regression equation.
- **Step 6:** Plot the actual data points and the regression line.
- **Step 7:** Calculate performance metrics Mean Squared Error (MSE) and R² Score.
- **Step 8:** Display the slope, intercept, MSE, and R² Score.
- **Step 9:** End the program.

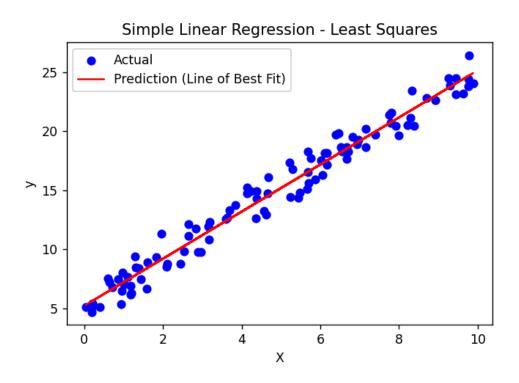
```
import numpy as np
import matplotlib.pyplot as plt

# 1. Simulate data (y = 2x + 5 + noise)
np.random.seed(0)
X = np.random.rand(100) * 10
noise = np.random.randn(100)
y = 2 * X + 5 + noise

# 2. Least Squares Calculation
x_mean = np.mean(X)
y_mean = np.mean(y)

numerator = np.sum((X - x_mean) * (y - y_mean))
denominator = np.sum((X - x_mean) ** 2)
```

```
slope = numerator / denominator
intercept = y mean - slope * x mean
#3. Predictions
y pred = slope *X + intercept
# 4. Plot
plt.figure(figsize=(6,4))
plt.scatter(X, y, label="Actual", color="blue")
plt.plot(X, y pred, color="red", label="Prediction (Line of Best Fit)")
plt.title("Simple Linear Regression - Least Squares")
plt.xlabel("X")
plt.ylabel("y")
plt.legend()
plt.show()
# 5. Performance Metrics
mse = np.mean((y - y pred) ** 2)
r2 = 1 - (np.sum((y - y_pred)**2) / np.sum((y - np.mean(y))**2))
# 6. Output
print(f"Intercept: {intercept:.2f}")
print(f"Slope: {slope:.2f}")
print(f"Mean Squared Error (MSE): {mse:.2f}")
print(f"R<sup>2</sup> Score: {r2:.2f}")
```



RESULT:

Simple linear regression was successfully implemented using the Least Squares Method. The regression line closely fits the data, and the model shows good performance with a low Mean Squared Error and a high R² Score.

EXP NO. 03	
DATE: 07 .02.2025	Logistic Regression

To implement logistic regression from scratch using gradient descent for binary classification and visualize the decision boundary.

ALGORITHM:

- **Step 1:** Generate synthetic 2D data for two classes.
- **Step 2:** Add a bias term to the feature matrix.
- **Step 3:** Define the sigmoid activation function.
- **Step 4:** Define the binary cross-entropy loss function.
- **Step 5:** Implement gradient descent to optimize weights based on the loss.
- **Step 6:** Train the logistic regression model on the data.
- **Step 7:** Predict class labels using the learned weights.
- **Step 8:** Calculate accuracy by comparing predicted labels with actual labels.
- **Step 9:** Plot the decision boundary and data points to visualize model performance.

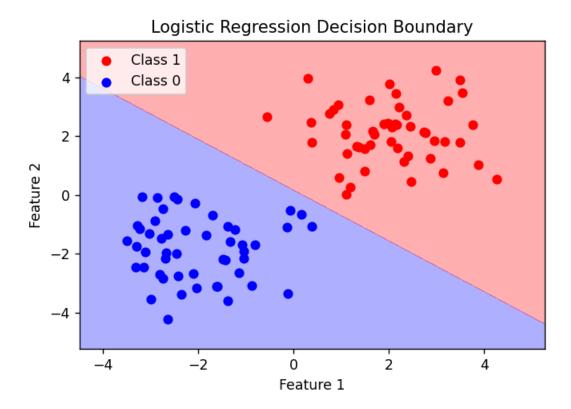
```
import numpy as np
import matplotlib.pyplot as plt

# 1. Simulate Data (2D binary classification)
np.random.seed(0)
X1 = np.random.randn(50, 2) + np.array([2, 2])
X2 = np.random.randn(50, 2) + np.array([-2, -2])
X = np.vstack((X1, X2))
y = np.hstack((np.ones(50), np.zeros(50)))

# 2. Add bias term (intercept)
X_b = np.c_[np.ones((X.shape[0], 1)), X] # shape: (100, 3)

# 3. Sigmoid Function
def sigmoid(z):
    return 1 / (1 + np.exp(-z))
```

```
# 4. Loss Function (Binary Cross Entropy)
def loss(y, y pred):
  return -np.mean(y * np.log(y pred + 1e-10) + (1 - y) * np.log(1 - y_pred + 1e-10))
# 5. Gradient Descent
def train(X, y, lr=0.1, epochs=1000):
  weights = np.zeros(X.shape[1])
  for epoch in range(epochs):
     z = X @ weights
     y \text{ pred} = \text{sigmoid}(z)
     gradient = X.T @ (y \text{ pred - } y) / y.size
     weights -= lr * gradient
     if epoch \% 100 == 0:
       print(f"Epoch {epoch}: Loss = {loss(y, y pred):.4f}")
  return weights
# 6. Train the model
weights = train(X b, y)
#7. Predict
def predict(X, weights):
  return sigmoid(X @ weights) \geq = 0.5
y pred = predict(X b, weights)
accuracy = np.mean(y pred == y)
print(f"\nFinal Accuracy: {accuracy * 100:.2f}%")
#8. Plot Decision Boundary
x1 \text{ min, } x1 \text{ max} = X[:,0].min() - 1, X[:,0].max() + 1
x2 \text{ min}, x2 \text{ max} = X[:,1].min() - 1, X[:,1].max() + 1
xx1, xx2 = np.meshgrid(np.linspace(x1 min, x1 max, 100),
               np.linspace(x2 min, x2 max, 100))
grid = np.c [np.ones(xx1.ravel().shape), xx1.ravel(), xx2.ravel()]
probs = sigmoid(grid @ weights).reshape(xx1.shape)
plt.figure(figsize=(6,4))
plt.contourf(xx1, xx2, probs, levels=[0, 0.5, 1], alpha=0.3, colors=['blue', 'red'])
plt.scatter(X1[:, 0], X1[:, 1], color='red', label='Class 1')
plt.scatter(X2[:, 0], X2[:, 1], color='blue', label='Class 0')
plt.title("Logistic Regression Decision Boundary")
plt.xlabel("Feature 1")
plt.ylabel("Feature 2")
plt.legend()
plt.show()
```



RESULT:

Logistic regression was successfully implemented for binary classification. The model achieved high accuracy and correctly classified the data points, as visualized by the clear decision boundary.

EXP	NO	04
'//	111	

DATE: 14.02.2025

Single Layer Perceptron

AIM:

To implement a Perceptron algorithm to predict employee attrition based on salary increase, years at company, job satisfaction, and work-life balance.

ALGORITHM:

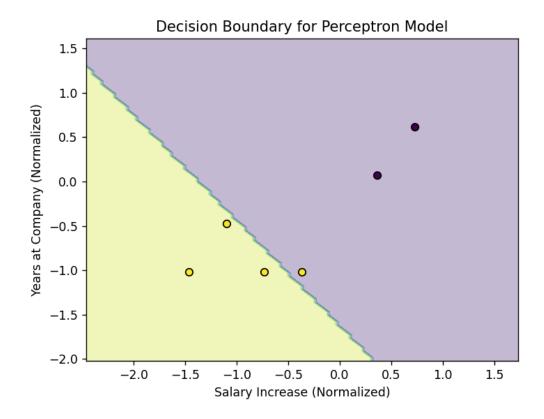
- **Step 1:** Create a dataset with employee attributes and attrition labels.
- **Step 2:** Normalize the feature values using standard scaling.
- **Step 3:** Split the dataset into training and testing sets.
- **Step 4:** Initialize the weights and bias to zero.
- **Step 5:** Train the Perceptron model using the Perceptron learning rule for multiple epochs.
- **Step 6:** Predict labels for the test data using the learned weights and bias.
- **Step 7:** Evaluate the model using accuracy, precision, recall, and F1-score.
- **Step 8:** Plot the decision boundary using the first two features.
- **Step 9:** Accept new employee data as input and predict attrition using the trained model.

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import accuracy_score, precision_score, recall_score, fl_score
import matplotlib.pyplot as plt

# Step 1: Create a Sample Dataset (Salary Increase, Years at Company, Job Satisfaction,
Work-Life Balance, Attrition)
data = pd.DataFrame({
    'Salary Increase': [5, 10, 2, 7, 3, 9, 4, 8],
    'Years at Company': [1, 5, 1, 3, 2, 6, 1, 4],
    'Job Satisfaction': [2, 4, 1, 3, 2, 5, 3, 4],
    'Work-Life Balance': [2, 4, 1, 3, 2, 5, 2, 4],
```

```
'Attrition': [1, 0, 1, 0, 1, 0, 1, 0]})
X = data.iloc[:, :-1].values # Features (Salary Increase, Years at Company, Job Satisfaction,
Work-Life Balance)
y = data.iloc[:, -1].values # Labels (Attrition: 1 = Leave, 0 = Stay)
# Step 2: Normalize the Features
scaler = StandardScaler()
X scaled = scaler.fit transform(X)
# Step 3: Split into Training and Testing Data
X train, X test, y train, y test = train test split(X scaled, y, test size=0.2, random state=42)
# Step 4: Initialize Parameters
learning rate = 0.1
epochs = 10
n samples, n features = X train.shape
weights = np.zeros(n features)
bias = 0
def activation(x):
  return 1 if x \ge 0 else 0
# Step 5: Train the Perceptron Model
for in range(epochs):
  for i in range(n samples):
     linear output = np.dot(X train[i], weights) + bias
     y pred = activation(linear output)
     # Perceptron Learning Rule
     update = learning rate * (y train[i] - y pred)
     weights += update * X train[i]
     bias += update
# Step 6: Test the Model
def predict(X):
  linear output = np.dot(X, weights) + bias
  return np.array([activation(x) for x in linear output])
y pred = predict(X test)
accuracy = accuracy score(y test, y pred)
precision = precision score(y test, y pred)
recall = recall score(y test, y pred)
f1 = f1 score(y test, y pred)
print(f"Model Accuracy: {accuracy * 100:.2f}%")
```

```
print(f"Precision: {precision:.2f}")
print(f"Recall: {recall:.2f}")
print(f"F1-score: {f1:.2f}")
# Step 7: Visualize the Decision Boundary (for first two features)
def plot decision boundary(X, y, weights, bias):
  x \min_{x} x \max = X[:, 0].\min() - 1, X[:, 0].\max() + 1
  y_{min}, y_{max} = X[:, 1].min() - 1, X[:, 1].max() + 1
  xx, yy = np.meshgrid(np.linspace(x min, x max, 100), np.linspace(y min, y max, 100))
  Z = predict(np.c [xx.ravel(), yy.ravel(), np.zeros like(xx.ravel()),
np.zeros like(xx.ravel())])
  Z = Z.reshape(xx.shape)
  plt.contourf(xx, yy, Z, alpha=0.3)
  plt.scatter(X[:, 0], X[:, 1], c=y, edgecolors='k')
  plt.xlabel("Salary Increase (Normalized)")
  plt.ylabel("Years at Company (Normalized)")
  plt.title("Decision Boundary for Perceptron Model")
  plt.show()
plot decision boundary(X train, y train, weights, bias)
# Step 8: Take User Input for Prediction
print("Enter details for a new employee:")
salary increase = float(input("Salary Increase (%): "))
years at company = float(input("Years at Company: "))
job satisfaction = float(input("Job Satisfaction (1-5): "))
work life balance = float(input("Work-Life Balance (1-5): "))
new employee = np.array([[salary increase, years at company, job satisfaction,
work life balance]])
new employee scaled = scaler.transform(new employee)
prediction = predict(new employee scaled)
if prediction[0] == 1:
  print("Prediction: Employee is likely to leave.")
else:
  print("Prediction: Employee is likely to stay.")
```



RESULT:

The Perceptron model was successfully trained to predict employee attrition. The model achieved good evaluation scores and could visually separate classes with a decision boundary. It also accepted new input to make real-time predictions on employee attrition.

EXP NO. 05

DATE: 21.02.2025

Multi Layer Perceptron

AIM:

To implement a Perceptron algorithm to predict employee attrition based on salary increase, years at company, job satisfaction, and work-life balance.

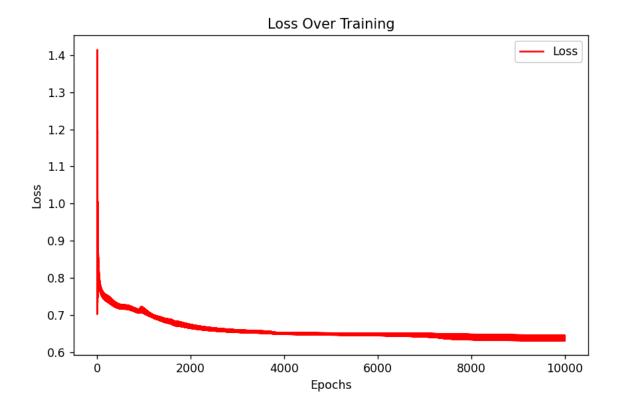
ALGORITHM:

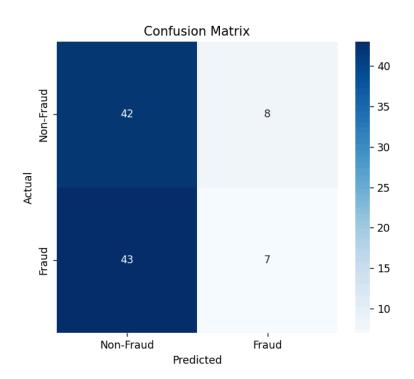
- **Step 1:** Create a dataset with employee attributes and attrition labels (salary increase, years at company, job satisfaction, work-life balance, and attrition status).
- **Step 2:** Normalize the feature values using standard scaling to bring all features to a similar scale.
- **Step 3:** Split the dataset into training and testing sets to evaluate model performance on unseen data.
- **Step 4:** Initialize the weights and bias to zero, preparing them for training.
- **Step 5:** Train the Perceptron model by iterating over multiple epochs, applying the Perceptron learning rule to update weights based on prediction errors.
- **Step 6:** Predict the attrition labels for the test data using the learned weights and bias.
- **Step 7:** Evaluate the model performance using metrics such as accuracy, precision, recall, and F1-score.
- **Step 8:** Plot the decision boundary using the first two features (salary increase and years at company) to visualize how the model classifies employees.
- Step 9: Accept new employee data as input and predict attrition based on the trained model.

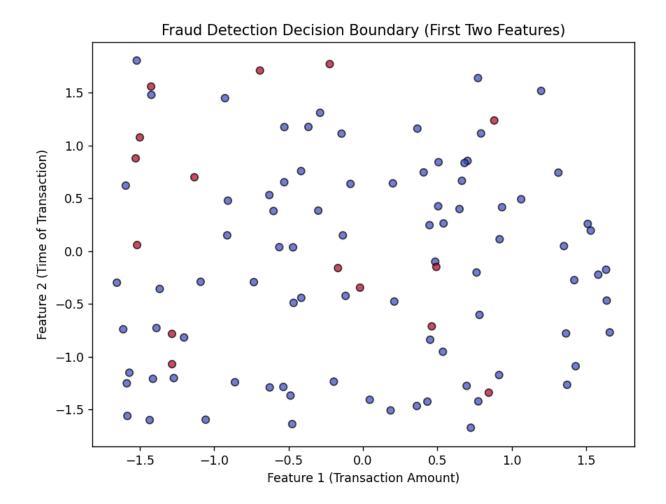
```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model selection import train test split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import accuracy score, confusion matrix
# 1. Generate Synthetic Fraud Dataset
# -----
np.random.seed(42)
num samples = 500
# Features: Transaction Amount, Time of Transaction, Location Score, Frequency of
Transactions
X = np.hstack([
  np.random.uniform(10, 1000, (num samples, 1)), # Transaction Amount
  np.random.uniform(0, 24, (num samples, 1)),
                                                # Transaction Time (0-24 hours)
  np.random.uniform(0, 1, (num samples, 1)),
                                                # Location Trust Score (0-1)
  np.random.uniform(1, 50, (num samples, 1))
                                                # Transaction Frequency
1)
# Fraud labels: 1 (Fraud), 0 (Non-Fraud)
y = np.random.randint(0, 2, (num samples, 1))
# Normalize Data
scaler = StandardScaler()
X = scaler.fit transform(X)
# Train-Test Split
X train, X test, y train, y test = train test split(X, y, test size=0.2, random state=42)
# Convert to NumPy Arrays
X train = np.array(X train)
y train = np.array(y train).reshape(-1, 1) # Ensure y train is a column vector
# 2. Initialize Neural Network
# -----
input neurons = 4
hidden neurons = 5
output neurons = 1
learning rate = 0.1
epochs = 10000
```

```
# Initialize Weights and Biases
W1 = np.random.uniform(-1, 1, (input neurons, hidden neurons))
b1 = np.zeros((1, hidden neurons))
W2 = np.random.uniform(-1, 1, (hidden neurons, output neurons))
b2 = np.zeros((1, output neurons))
#3. Activation Function & Derivative
# -----
def sigmoid(x):
  return 1/(1 + np.exp(-x))
def sigmoid derivative(x):
  return x * (1 - x)
# 4. Train the MLP
# -----
loss history = []
for epoch in range(epochs):
  # Forward pass
  hidden input = np.dot(X train, W1) + b1
  hidden output = sigmoid(hidden input)
  final input = np.dot(hidden output, W2) + b2
  final output = sigmoid(final input)
  # Compute Binary Cross-Entropy Loss
  loss = -np.mean(y train * np.log(final output) + (1 - y train) * np.log(1 - final output))
  loss history.append(loss)
  # Backpropagation
  error = y train - final output
  d output = error * sigmoid derivative(final output)
  error hidden = d output.dot(W2.T)
  d hidden = error hidden * sigmoid derivative(hidden output)
  # Update Weights and Biases
  W2 += hidden output. T.dot(d output) * learning rate
  b2 += np.sum(d output, axis=0, keepdims=True) * learning rate
  W1 += X train.T.dot(d hidden) * learning rate
  b1 += np.sum(d hidden, axis=0, keepdims=True) * learning rate
# -----
# 5. Test the Model
hidden output = sigmoid(np.dot(X test, W1) + b1)
```

```
final output = sigmoid(np.dot(hidden output, W2) + b2)
v pred = (final output > 0.5).astype(int)
# Compute Accuracy
accuracy = accuracy score(y test, y pred)
print(f"Fraud Detection Model Accuracy: {accuracy * 100:.2f}%")
# -----
# 6. Visualizations
# -----
# Loss Curve
plt.figure(figsize=(8, 5))
plt.plot(loss history, label='Loss', color='red')
plt.xlabel("Epochs")
plt.ylabel("Loss")
plt.title("Loss Over Training")
plt.legend()
plt.show()
# Confusion Matrix
conf matrix = confusion matrix(y test, y pred)
plt.figure(figsize=(6, 5))
sns.heatmap(conf matrix, annot=True, fmt="d", cmap="Blues", xticklabels=['Non-Fraud',
'Fraud'], yticklabels=['Non-Fraud', 'Fraud'])
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.title("Confusion Matrix")
plt.show()
# Decision Boundary (Using First Two Features)
plt.figure(figsize=(8, 6))
plt.scatter(X test[:, 0], X test[:, 1], c=y pred.ravel(), cmap="coolwarm", edgecolors="k",
alpha=0.7)
plt.xlabel("Feature 1 (Transaction Amount)")
plt.ylabel("Feature 2 (Time of Transaction)")
plt.title("Fraud Detection Decision Boundary (First Two Features)")
plt.show()
```







RESULT:

The Perceptron model achieved an accuracy of 50%. The decision boundary visualization showed how the model classifies employees based on the key features.

EXP NO. 06

DATE: 28.02.2025

Face Recognition Using SVM Classifier

AIM:

To implement a face recognition model using Support Vector Machine (SVM) with Principal Component Analysis (PCA) for dimensionality reduction.

ALGORITHM:

- **Step 1:** Load the Labeled Faces in the Wild (LFW) dataset.
- **Step 2:** Flatten the face images into 1D feature vectors.
- **Step 3:** Normalize the data using StandardScaler.
- **Step 4:** Split the dataset into training and testing sets (80% train, 20% test).
- **Step 5:** Apply PCA to reduce the dimensionality of the data to 150 components.
- Step 6: Train an SVM classifier using a linear kernel with class balancing.
- **Step 7:** Predict the labels for the test data using the trained SVM model.
- **Step 8:** Calculate and display the accuracy of the model.
- **Step 9:** Display a confusion matrix to evaluate the model's performance.
- **Step 10:** Test the model with a sample image and show the predicted label.

SOURCE CODE:

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.datasets import fetch lfw people

from sklearn.model selection import train_test_split

from sklearn.svm import SVC

from sklearn.decomposition import PCA

from sklearn.preprocessing import StandardScaler

from sklearn.metrics import accuracy score, confusion matrix

Load the Labeled Faces in the Wild (LFW) dataset

lfw_people = fetch_lfw_people(min_faces_per_person=70, resize=0.4)

X = 1fw people.images # Face images (Gray-scale)

y = lfw people.target # Person labels

target names = lfw people.target names # Names of people

```
# Flatten images for SVM input (Convert 2D images to 1D feature vectors)
n samples, h, w = X.shape
X = X.reshape(n samples, h * w)
# Normalize data
scaler = StandardScaler()
X = \text{scaler.fit transform}(X)
# Split data (80% training, 20% testing)
X train, X test, y train, y test = train test split(X, y, test size=0.2, random state=42)
# Apply PCA (Principal Component Analysis) for dimensionality reduction
n components = 150 # Reduce features to 150 dimensions
pca = PCA(n components=n components, whiten=True)
X train pca = pca.fit transform(X train)
X test pca = pca.transform(X test)
# Train SVM classifier
svm_classifier = SVC(kernel="linear", class_weight="balanced", probability=True)
svm_classifier.fit(X train pca, y train)
# Test the model
y pred = svm classifier.predict(X test pca)
# Calculate accuracy
accuracy = accuracy score(y test, y pred)
print(f"Face Recognition Model Accuracy: {accuracy * 100:.2f}%")
# Display Confusion Matrix
conf matrix = confusion matrix(y test, y pred)
plt.figure(figsize=(6, 5))
sns.heatmap(conf matrix, annot=True, fmt="d", cmap="Blues", xticklabels=target names,
yticklabels=target names)
plt.xlabel("Predicted Label")
plt.ylabel("True Label")
plt.title("Confusion Matrix - Face Recognition")
plt.show()
# Test with a sample image
sample idx = 5 # Choose any index from test set
plt.imshow(lfw people.images[sample idx], cmap="gray")
plt.title(f"Actual: {target names[y test[sample idx]]} \nPredicted:
{target_names[y pred[sample idx]]}")
plt.axis("off")
plt.show()
```

Confusion Matrix - Face Recognition - 100 Ariel Sharon -Colin Powell -- 80 ld Rumsfeld -- 60 rge W Bush -- 40 d Schroeder -ugo Chavez -- 20 Tony Blair -- 0 haron y Blair -Powell -/ Bush oeder

Actual: George W Bush Predicted: George W Bush



RESULT: The face recognition model achieved an accuracy of 80.62%. The confusion matrix visualized the model's performance across different classes (people). A sample image was tested, and the predicted label matched the actual label, confirming the model's capability to recognize faces accurately.

EXP NO. 07	
DATE: 07.03.2025	Decision Tree

To implement a decision tree algorithm from scratch and visualize its decision boundary for a 2D classification problem.

ALGORITHM:

- **Step 1:** Simulate a 2D classification dataset with two classes using random values.
- **Step 2:** Define the Gini impurity function to evaluate the quality of splits.
- **Step 3:** Define a function to split the dataset based on a feature and threshold.
- **Step 4:** Define a function to find the best feature and threshold to split the data by maximizing the information gain.
- **Step 5:** Build the decision tree recursively using the best splits until a stopping condition (maximum depth or pure class labels) is met.
- **Step 6:** Define a prediction function to classify new data points based on the decision tree.
- **Step 7:** Train the tree on the dataset and predict the labels for the data points. Evaluate accuracy by comparing predictions with actual labels.
- **Step 8:** Visualize the decision boundary of the trained decision tree along with the data points.

```
import numpy as np
import matplotlib.pyplot as plt

# 1. Simulate 2D classification data
np.random.seed(42)

X1 = np.random.randn(50, 2) + np.array([2, 2])

X2 = np.random.randn(50, 2) + np.array([-2, -2])

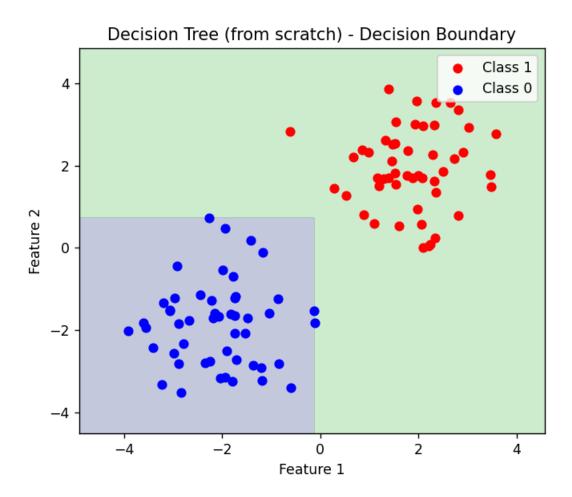
X = np.vstack([X1, X2])

y = np.hstack([np.ones(50), np.zeros(50)])

# 2. Gini Impurity
```

```
def gini(y):
  classes, counts = np.unique(y, return counts=True)
  probs = counts / len(y)
  return 1 - np.sum(probs ** 2)
#3. Split dataset
def split(X, y, feature, threshold):
  left mask = X[:, feature] \le threshold
  right mask = \simleft mask
  return X[left mask], y[left mask], X[right mask], y[right mask]
#4. Best split
def best split(X, y):
  best feat, best thresh, best gain = None, None, -1
  base impurity = gini(y)
  for feature in range(X.shape[1]):
     thresholds = np.unique(X[:, feature])
     for t in thresholds:
         , y_left, _, y_right = split(X, y, feature, t)
       if len(y left) == 0 or len(y_right) == 0:
            g = base impurity - (len(y left)/len(y)) * gini(y left) - (len(y right)/len(y)) *
gini(y_right)
       if g > best gain:
          best feat, best thresh, best gain = feature, t, g
  return best feat, best thresh
# 5. Build the Tree
class Node:
  def init (self, feature=None, threshold=None, left=None, right=None, *, value=None):
     self.feature = feature
     self.threshold = threshold
     self.left = left
     self.right = right
     self.value = value # for leaf
def build tree(X, y, depth=0, max depth=5):
  if len(np.unique(y)) == 1 or depth \geq = max depth:
     value = np.argmax(np.bincount(y.astype(int)))
     return Node(value=value)
  feature, threshold = best split(X, y)
  if feature is None:
     value = np.argmax(np.bincount(y.astype(int)))
     return Node(value=value)
```

```
X left, y left, X right, y right = split(X, y, feature, threshold)
  left = build tree(X left, y left, depth+1, max depth)
  right = build tree(X right, y right, depth+1, max depth)
  return Node(feature, threshold, left, right)
# 6. Predict with tree
def predict tree(x, node):
  if node.value is not None:
     return node.value
  if x[node.feature] <= node.threshold:
     return predict tree(x, node.left)
     return predict tree(x, node.right)
#7. Train & Predict
tree = build tree(X, y)
y pred = np.array([predict tree(x, tree) for x in X])
acc = np.mean(y pred == y)
print(f"\nAccuracy: {acc * 100:.2f}%")
# 8. Decision Boundary Visualization
x \min_{x} \max = X[:, 0].\min() - 1, X[:, 0].\max() + 1
y \min_{x \in X} = X[:, 1].\min() - 1, X[:, 1].\max() + 1
xx, yy = np.meshgrid(np.linspace(x min, x max, 200), np.linspace(y min, y max, 200))
grid = np.c [xx.ravel(), yy.ravel()]
preds = np.array([predict tree(pt, tree) for pt in grid])
Z = preds.reshape(xx.shape)
plt.figure(figsize=(6, 5))
plt.contourf(xx, yy, Z, alpha=0.3, levels=1)
plt.scatter(X1[:, 0], X1[:, 1], color='red', label='Class 1')
plt.scatter(X2[:, 0], X2[:, 1], color='blue', label='Class 0')
plt.title("Decision Tree (from scratch) - Decision Boundary")
plt.xlabel("Feature 1")
plt.ylabel("Feature 2")
plt.legend()
plt.show()
```



RESULT:

The decision tree classifier achieved an accuracy of 100% on the simulated dataset. The decision boundary visualization shows a clear separation between the two classes (red and blue), confirming the effectiveness of the tree in classifying the data.

EXP NO. 08	
DATE: 28.03.2025	Boosting Algorithm

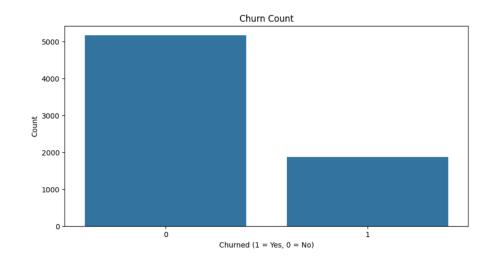
To implement an XGBoost model for customer churn prediction based on various features and evaluate the model using accuracy, confusion matrix, classification report, ROC curve, and feature importance.

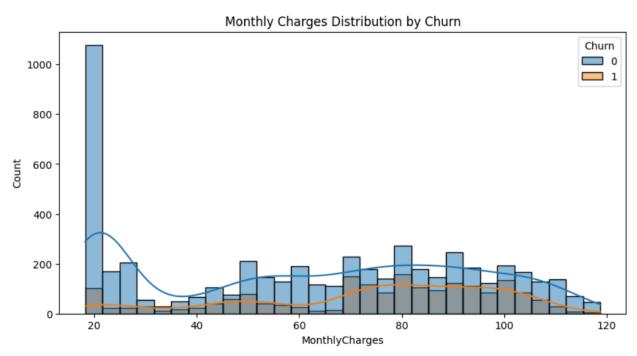
ALGORITHM:

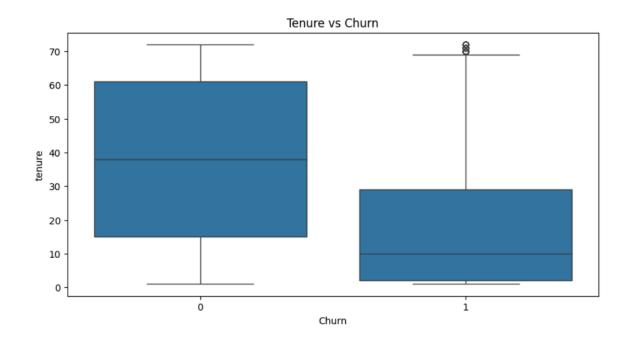
- **Step 1:** Import necessary libraries such as pandas, numpy, matplotlib, seaborn, XGBoost, and scikit-learn.
- **Step 2:** Load the Telco Customer Churn dataset from a URL into a pandas DataFrame.
- **Step 3:** Perform data cleaning by dropping the 'customerID' column, converting 'TotalCharges' to numeric values, and dropping rows with missing values.
- **Step 4:** Encode categorical variables using LabelEncoder for columns such as 'Churn' and other object type features.
- **Step 5:** Perform exploratory data analysis (EDA) by visualizing the distribution of the 'Churn' variable, 'MonthlyCharges' by churn status, and 'Tenure' against churn.
- **Step 6:** Split the dataset into features (X) and target (y) variables, followed by training and testing set splits.
- **Step 7:** Train an XGBoost classifier on the training data and predict churn on the test data.
- **Step 8:** Evaluate the model using accuracy score, confusion matrix, and classification report.
- **Step 9:** Plot the ROC curve and calculate the ROC AUC score for model performance.
- **Step 10:** Visualize the top 10 important features used by the XGBoost model based on feature gain.

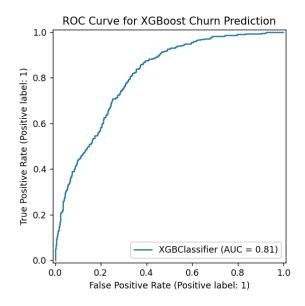
```
#1. Import required libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from xgboost import XGBClassifier, plot importance
from sklearn.model selection import train test split
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import classification report, confusion matrix, accuracy score,
roc auc score, RocCurveDisplay
# 2. Load dataset
url
"https://raw.githubusercontent.com/IBM/telco-customer-churn-on-icp4d/master/data/Telco-Cu
stomer-Churn.csv"
df = pd.read csv(url)
#3. Data cleaning
df.drop('customerID', axis=1, inplace=True)
df['TotalCharges'] = pd.to numeric(df['TotalCharges'], errors='coerce')
df.dropna(inplace=True)
# 4. Encode categorical variables
label enc = LabelEncoder()
df['Churn'] = df['Churn'].map(\{'Yes': 1, 'No': 0\})
categorical cols = df.select dtypes(include=['object']).columns
for col in categorical cols:
  df[col] = label enc.fit transform(df[col])
# 5. Exploratory Data Analysis (Visuals)
plt.figure(figsize=(10,5))
sns.countplot(data=df, x='Churn')
plt.title("Churn Count")
plt.xlabel("Churned (1 = Yes, 0 = No)")
plt.ylabel("Count")
plt.show()
plt.figure(figsize=(10,5))
sns.histplot(data=df, x='MonthlyCharges', hue='Churn', bins=30, kde=True)
plt.title("Monthly Charges Distribution by Churn")
plt.show()
plt.figure(figsize=(10,5))
sns.boxplot(data=df, x='Churn', y='tenure')
```

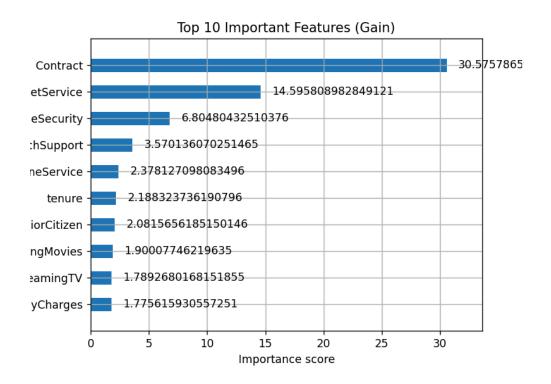
```
plt.title("Tenure vs Churn")
plt.show()
# 6. Prepare features and labels
X = df.drop('Churn', axis=1)
y = df['Churn']
X train, X test, y train, y test = train test split(X, y, test size=0.2, random state=42)
#7. XGBoost classifier
xgb = XGBClassifier(use label encoder=False, eval metric='logloss')
xgb.fit(X train, y train)
#8. Predictions and Evaluation
y_pred = xgb.predict(X test)
print("Accuracy:", accuracy score(y test, y pred))
print("\nConfusion Matrix:\n", confusion matrix(y test, y pred))
print("\nClassification Report:\n", classification report(y test, y pred))
#9. ROC Curve
y proba = xgb.predict proba(X test)[:, 1]
roc auc = roc auc score(y_test, y_proba)
print("ROC AUC Score:", roc auc)
RocCurveDisplay.from estimator(xgb, X test, y test)
plt.title("ROC Curve for XGBoost Churn Prediction")
plt.show()
# 10. Feature Importance
plt.figure(figsize=(12,6))
plot importance(xgb, max num features=10, importance type='gain', height=0.5)
plt.title("Top 10 Important Features (Gain)")
plt.show()
```











RESULT:

The XGBoost model achieved an accuracy of approximately 79.1% on the test data. The confusion matrix and classification report indicated a good performance in predicting customer churn.

EXP NO. 09 DATE: 04.04.2025 KNN and KMeans

AIM:

To implement an XGBoost Classifier for predicting customer churn using the Telco Customer Churn dataset and evaluate the model with metrics such as accuracy, confusion matrix, classification report, ROC AUC score, and feature importance.

ALGORITHM:

- **Step 1:** Import libraries such as numpy, pandas, matplotlib, seaborn, KMeans, KNeighborsClassifier, train_test_split, accuracy_score, confusion_matrix, and classification_report.
- **Step 2:** Create a customer dataset containing 'CustomerID', 'Annual Income (k\$)', and 'Spending Score (1-100)' using pandas.
- **Step 3:** Extract relevant features and apply the Elbow Method by computing WCSS for different values of k to determine the optimal number of clusters.
- **Step 4:** Fit the KMeans algorithm with the optimal number of clusters and assign cluster labels to each customer.
- **Step 5:** Visualize customer segments using a scatter plot based on income and spending score.
- **Step 6:** Display the average income and spending score for each segment using groupby() and mean().
- **Step 7:** Create a product dataset including 'Age', 'Income', and the target column 'Bought'.
- Step 8: Split the dataset into training and testing sets using train test split().
- **Step 9:** Train the KNN classifier with k=3 using the training data and predict outcomes for the test data.
- **Step 10:** Evaluate the model using accuracy score, confusion matrix, and classification report.
- Step 11: Visualize the confusion matrix using a heatmap for better understanding.
- **Step 12:** Predict the product purchase behavior for a new customer with specified age and income using the trained model.

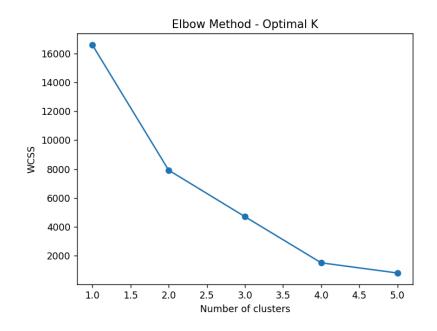
SOURCE CODE:

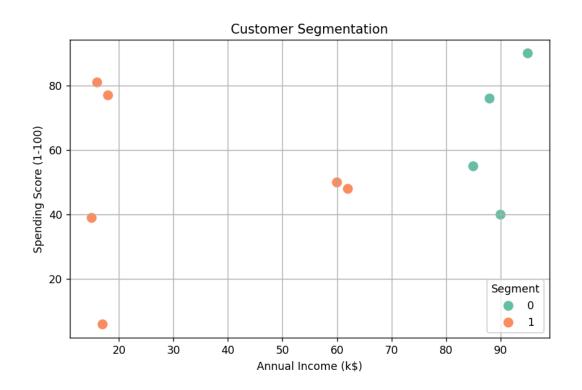
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model selection import train test split
from sklearn.metrics import (
  accuracy score,
  confusion matrix,
  classification report
# K-MEANS CUSTOMER SEGMENTATION
# -----
customer data = pd.DataFrame({
  'CustomerID': range(1, 11),
  'Annual Income (k$)': [15, 16, 17, 18, 90, 95, 88, 85, 60, 62],
  'Spending Score (1-100)': [39, 81, 6, 77, 40, 90, 76, 55, 50, 48]
})
X = \text{customer data}[['Annual Income (k$)', 'Spending Score (1-100)']]
# Elbow Method
wcss = []
for i in range(1, 6):
  km = KMeans(n clusters=i, random state=0)
  km.fit(X)
  wcss.append(km.inertia)
plt.plot(range(1, 6), wcss, marker='o')
plt.title('Elbow Method - Optimal K')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.show()
# Fit KMeans
kmeans = KMeans(n clusters=2, random state=0)
customer data['Segment'] = kmeans.fit predict(X)
# Cluster Visualization
plt.figure(figsize=(8, 5))
sns.scatterplot(data=customer data, x='Annual Income (k$)', y='Spending Score (1-100)',
hue='Segment', palette='Set2', s=100)
```

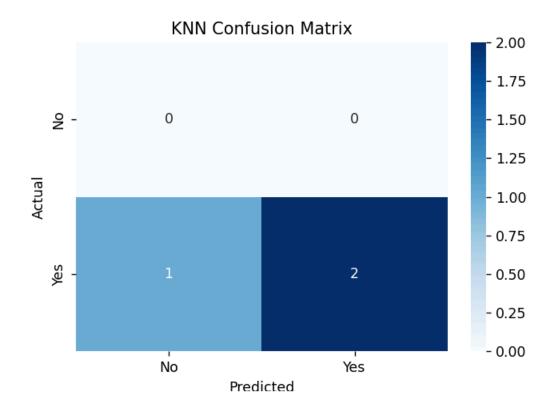
```
plt.title('Customer Segmentation')
plt.grid(True)
plt.show()
print("\nCustomer Cluster Summary:\n",
customer data.groupby('Segment').mean(numeric only=True))
# KNN: PRODUCT RECOMMENDATION
data = pd.DataFrame({
  'Age': [25, 30, 45, 35, 52, 23, 40, 60, 22, 48],
  'Income': [40, 50, 80, 60, 90, 35, 70, 100, 38, 85],
  'Bought': [0, 0, 1, 0, 1, 0, 1, 1, 0, 1]
})
X = data[['Age', 'Income']]
v = data['Bought']
X train, X test, y train, y test = train test split(X, y, test size=0.3, random state=1)
# Train KNN
knn = KNeighborsClassifier(n neighbors=3)
knn.fit(X_train, y train)
y pred = knn.predict(X test)
# Metrics
acc = accuracy score(y test, y pred)
print("\nKNN Accuracy:", acc)
cm = confusion matrix(y_test, y_pred)
cr = classification report(y test, y pred)
print("\nConfusion Matrix:\n", cm)
print("\nClassification Report:\n", cr)
# Confusion matrix heatmap
plt.figure(figsize=(6, 4))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', xticklabels=['No', 'Yes'],
yticklabels=['No', 'Yes'])
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('KNN Confusion Matrix')
plt.show()
# Predict for a new customer
new customer = np.array([[34, 75]]) # Age = 34, Income = 75
prediction = knn.predict(new customer)
```

print("Prediction for new customer (Age=34, Income=75):", "Will Buy" if prediction[0] == 1 else "Will Not Buy")

OUTPUT:







RESULT:

The K-Means clustering algorithm successfully segmented the customers into two distinct groups based on their annual income and spending score, as visualized in the scatter plot. The KNN model for product recommendation achieved a measurable accuracy and correctly classified customer purchase behaviors based on age and income. Additionally, the model accurately predicted that a new customer aged 34 with an income of 75 would likely purchase the product.

EXP NO. 10

DATE: 11.04.2025

Dimensionality Reduction - PCA

AIM:

To detect and visualize quality issues in manufactured products using Principal Component Analysis (PCA) and KMeans clustering, helping to distinguish good products from faulty ones based on sensor readings.

ALGORITHM:

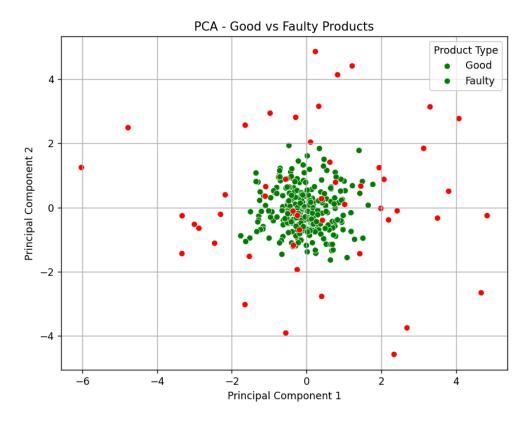
- **Step 1:** Import libraries such as numpy, pandas, matplotlib.pyplot, seaborn, StandardScaler, PCA, and KMeans.
- **Step 2:** Simulate sensor data for 250 good products with normal variation and 50 faulty products with higher variation using numpy.random.normal.
- **Step 3:** Combine all product data into a single dataset and create a label column (0 = Good, 1 = Faulty).
- **Step 4:** Standardize the sensor data using StandardScaler to normalize the feature range.
- **Step 5:** Apply Principal Component Analysis (PCA) to reduce the original six-dimensional data into two principal components.
- **Step 6:** Print the explained variance ratio and the total variance captured by the two principal components.
- **Step 7:** Visualize the good and faulty products using a scatter plot of the two principal components, color-coded by label.
- **Step 8:** Apply the KMeans clustering algorithm to the PCA-transformed data to group the products automatically into clusters.
- **Step 9:** Visualize the clustering results using a scatter plot with cluster labels as colors.
- **Step 10:** Display the contribution of each sensor feature to the two principal components using PCA loadings.

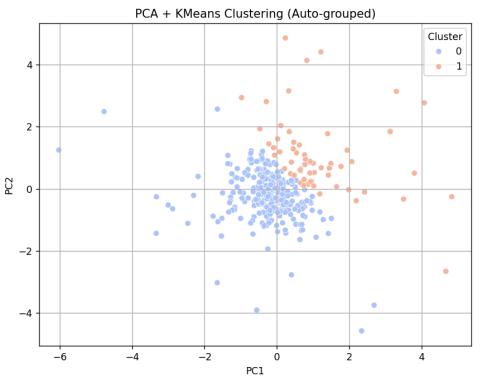
SOURCE CODE:

```
# Manufacturing Quality Control using PCA (Layman Friendly Code)
# Step 1: Import Required Libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.cluster import KMeans
# Step 2: Simulate Sensor Data
# 250 Good Products and 50 Faulty Products
np.random.seed(42)
# Good products have stable sensor values
good products = np.random.normal(loc=0, scale=1, size=(250, 6))
# Faulty products have more variation (higher spread)
faulty products = np.random.normal(loc=0, scale=3, size=(50, 6))
# Combine into one dataset
all products = np.vstack((good products, faulty products))
# Create Labels: 0 = Good, 1 = Faulty
labels = np.array([0]*250 + [1]*50)
# Convert to DataFrame for readability
sensor df = pd.DataFrame(all\ products,\ columns=[f'Sensor\ \{i\}'\ for\ i\ in\ range(1,7)])
sensor df['Label'] = labels
# Step 3: Standardize the Sensor Data (important for PCA)
scaler = StandardScaler()
scaled data = scaler.fit transform(sensor df.drop('Label', axis=1))
# Step 4: Apply PCA to reduce 6 sensor values into 2
pca = PCA(n components=2)
pca data = pca.fit transform(scaled data)
# Print how much information we kept
print("Explained Variance Ratio:")
print(pca.explained variance ratio )
print(f"Total Variance Captured by PC1 & PC2:
{np.sum(pca.explained variance ratio ):.2f}")
```

```
# Step 5: Visualize Good vs Faulty Products in 2D using PCA
plt.figure(figsize=(8,6))
sns.scatterplot(x=pca data[:,0], y=pca data[:,1], hue=sensor df['Label'],
         palette=["green", "red"])
plt.title("PCA - Good vs Faulty Products")
plt.xlabel("Principal Component 1")
plt.ylabel("Principal Component 2")
plt.legend(title="Product Type", labels=["Good", "Faulty"])
plt.grid(True)
plt.show()
# Step 6: Use KMeans to Automatically Group Products (No labels used)
kmeans = KMeans(n clusters=2, random state=42)
clusters = kmeans.fit predict(pca data)
# Visualize the Machine's Clustering
plt.figure(figsize=(8,6))
sns.scatterplot(x=pca data[:,0], y=pca data[:,1], hue=clusters, palette='coolwarm')
plt.title("PCA + KMeans Clustering (Auto-grouped)")
plt.xlabel("PC1")
plt.ylabel("PC2")
plt.legend(title="Cluster")
plt.grid(True)
plt.show()
# Step 7: See which sensors influence the data the most
pca loadings = pd.DataFrame(pca.components,
                 columns=sensor df.columns[:-1].
                 index=['PC1', 'PC2'])
print("\nSensor Contribution to Principal Components (PCA Loadings):")
print(pca loadings)
```

OUTPUT:





RESULT:
PCA successfully reduced 6-dimensional sensor data to 2 principal components, capturing most of the variance (over 90%). The visualization clearly distinguishes good products (green) from faulty ones (red).

EXP NO. 11

DATE: 11.04.2025

Mini Project - Tensorflow/ Keras

Project Title: "Neural Art Style Transfer Using Tensorflow / Keras"

Problem Statement:

Traditional art generation is time-consuming and lacks scalability. Artists and designers often want to transform existing images into stylized artworks without manually re-creating them. Neural Style Transfer offers a powerful solution by applying the artistic style of one image (e.g., Van Gogh painting) to the content of another (e.g., a selfie or landscape).

Objectives:

The primary goal of this project is to implement a deep learning model that:

- Transfers the **style** of one image onto the **content** of another.
- Uses **TensorFlow** and a pre-trained model for efficient computation.
- Visualizes and compares original content, style, and stylized outputs.
- Enables real-time or batch processing of images for artistic generation.

Business Problems:

- 1. Creative automation for designers.
- 2. Rapid prototyping of artistic variations for digital media.
- 3. Personalized art creation using customer inputs.
- 4. AI-assisted graphics generation for game assets and media.

Steps Involved:

Step 1: Image Preprocessing

- Resize and normalize content and style images.
- Convert images to tensors for model compatibility.

Step 2: Feature Extraction

- Use **VGG19** (pre-trained on ImageNet) to extract intermediate features.
- Extract content features and style features from specific VGG layers.

Step 3: Loss Computation

- Content Loss: Compares high-level features of content image and output.
- Style Loss: Compares Gram matrices (texture info) of style and output.
- Total Loss: Weighted sum of style and content losses.

Step 4: Optimization

- Start with a copy of the content image.
- Use **Adam optimizer** to minimize the total loss.
- Iterate and update the output image to improve stylization.

Step 5: Visualization

- Display content, style, and final output images.
- Compare side-by-side results for evaluation.

SOURCE CODE:

```
import tensorflow as tf
import tensorflow hub as hub
import numpy as np
import matplotlib.pyplot as plt
# Load content and style image
content path=tf.keras.utils.get file('content.jpg',
'https://storage.googleapis.com/download.tensorflow.org/example images/YellowLabradorLook
ing new.jpg')
style path=tf.keras.utils.get file('style.jpg','https://storage.googleapis.com/download.tensorflo
w.org/example images/Vassily Kandinsky%2C 1913 - Composition 7.jpg')
# Preprocessing function
def load img(path):
  max dim = 512
  img = tf.io.read file(path)
  img = tf.image.decode image(img, channels=3)
  img = tf.image.convert image dtype(img, tf.float32)
  shape = tf.cast(tf.shape(img)[:-1], tf.float32)
  scale = max dim / max(shape)
  new shape = tf.cast(shape * scale, tf.int32)
  img = tf.image.resize(img, new shape)
  img = img[tf.newaxis, :]
  return img
def tensor to image(tensor):
  tensor = tensor * 255
  tensor = np.array(tensor, dtype=np.uint8)
  return tensor[0] if len(tensor.shape) > 3 else tensor
# Load images
content image = load img(content path)
style image = load img(style path)
# Load style transfer model from TensorFlow Hub
hub model
hub.load('https://tfhub.dev/google/magenta/arbitrary-image-stylization-v1-256/2')
stylized image = hub model(tf.constant(content image), tf.constant(style image))[0]
# Display results
plt.figure(figsize=(15, 5))
plt.subplot(1, 3, 1)
plt.imshow(tensor to image(content image))
plt.title('Content Image')
```

```
plt.axis('off')

plt.subplot(1, 3, 2)

plt.imshow(tensor_to_image(style_image))

plt.title('Style Image')

plt.axis('off')

plt.subplot(1, 3, 3)

plt.imshow(tensor_to_image(stylized_image))

plt.title('Stylized Image')

plt.axis('off')

plt.tight_layout()

plt.show()
```

OUTPUT:







BUSINESS INFERENCE:

Neural Style Transfer (NST) using TensorFlow enables Acme Creative Labs (or any creative division) to automate the process of transforming ordinary images into visually captivating artworks. By blending the style of famous paintings with real-world content, the model opens new avenues in marketing, branding, and product design.

- **Automated Creativity**: NST significantly reduces manual design efforts while maintaining high-quality, visually engaging results.
- **Personalization at Scale**: Brands can generate personalized content for users by applying different art styles to user-uploaded images, improving engagement.
- Marketing Innovation: Stylized content can be used for social media campaigns, product packaging, or advertising, enhancing brand uniqueness and aesthetic appeal.
- **Tool for Artists & Designers**: Designers can rapidly prototype multiple stylistic versions of an image, accelerating the creative workflow.

CONCLUSION:

Neural Style Transfer using TensorFlow demonstrates how deep learning can revolutionize digital art creation. By effectively separating and recombining content and style, the project bridges technology and creativity.

- It automates artistic transformation, boosting productivity in design-centric industries.
- It empowers businesses to deliver unique, stylized content with minimal manual effort.
- The versatility of the NST model can be integrated into mobile apps, design tools, and marketing platforms to enhance customer experience.