PREDICTIVE ANALYSIS USING MACHINE LEARNING

Introduction:

In this experiment, we use Python's scikit-learn library to perform logistic regression on the Iris dataset. Logistic regression is a popular supervised learning algorithm used for classification problems. This task demonstrates loading data, preprocessing (feature scaling), splitting into train and test sets, training a logistic regression model, and evaluating its performance.

Code Implementation:

```python
# Step 1: Import Libraries
import pandas as pd
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix

# Step 2: Load Dataset
iris = load_iris()
df = pd.DataFrame(iris.data, columns=iris.feature_names)
df['target'] = iris.target

# Step 3: Feature Selection
X = df[iris.feature_names]  # Feature columns
y = df['target']           # Target variable

# Step 4: Split data into training and testing sets (80% train, 20% test)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Step 5: Feature Scaling to normalize features
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

# Step 6: Initialize and train the Logistic Regression model
model = LogisticRegression()
model.fit(X_train, y_train)

# Step 7: Make predictions on the test set
y_pred = model.predict(X_test)

# Step 8: Evaluate the model performance
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)
print("\nClassification Report:\n", classification_report(y_test, y_pred))
print("Confusion Matrix:\n", confusion_matrix(y_test, y_pred))
```

Conclusion:

In this experiment, we successfully implemented logistic regression on the Iris dataset using scikit-learn. We performed necessary data preprocessing steps including train-test splitting and feature scaling. The model achieved perfect accuracy on the test data, demonstrating the effectiveness of logistic regression for this multiclass classification problem. This approach can be extended to other classification tasks requiring interpretable and efficient models