**Ex No: 9 Implementation of Logistic Regression**

**Problem Scenario**

A healthcare research institute wants to develop a predictive model to identify whether a patient is likely to have diabetes based on various medical factors such as glucose level, BMI, blood pressure, age, and insulin level. The dataset being used for this study is the **Diabetes Dataset**, which contains medical records of several patients along with an outcome variable indicating whether the patient is diabetic (1) or non-diabetic (0).

The research team aims to implement a **Logistic Regression model** because the problem involves **binary classification** (diabetic or not). The objective is to train the model on patient data, predict the outcome for new patients, and evaluate the model’s performance using classification metrics.

The tasks to be performed are:

1. Load and inspect the diabetes dataset.
2. Preprocess the dataset (handle missing or invalid values if required).
3. Split the dataset into **training** and **testing** sets.
4. Train a **Logistic Regression model** using the training data.
5. Predict diabetes outcomes for the test data.
6. Evaluate the model using:
   * **Accuracy Score**: Percentage of correct predictions.
   * **F1-Score**: Balance between precision and recall.
   * **Confusion Matrix**: To analyze true positives, true negatives, false positives, and false negatives.

The hospital management expects this analysis to provide an automated decision-support tool for doctors to quickly screen patients and identify those at high risk of diabetes, thereby enabling early intervention and preventive care.