**Title:** Diabetes Prediction Using Logistic Regression

**1. Introduction** Diabetes is a chronic disease that affects millions worldwide. Early prediction of diabetes can help in taking preventive measures. This project aims to develop a machine learning model using Logistic Regression to classify whether a person is diabetic (1) or not (0) based on their health parameters.

**2. Dataset Description** The dataset used for this project contains health-related attributes such as Age, BMI, Glucose Level, and Blood Pressure. Each row represents a patient, and the target variable (Diabetic) indicates whether the person has diabetes (1) or not (0).

**3. Data Preprocessing**

* **Loading the Data:** The dataset is loaded using Pandas.
* **Handling Missing Values:** The dataset is checked for missing values and handled accordingly.
* **Feature Scaling:** Standardization or normalization is applied to scale the data properly.
* **Splitting Data:** The dataset is divided into training (80%) and testing (20%) sets.

**4. Model Implementation**

* Import necessary libraries:
* import pandas as pd
* 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

* Load the dataset:

df = pd.read\_csv("diabetes.csv")

* Define features (X) and target (y):
* X = df[['Age', 'BMI', 'Glucose', 'BloodPressure']]

y = df['Diabetic']

* Split the dataset:

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

* Standardize the data:
* scaler = StandardScaler()
* X\_train = scaler.fit\_transform(X\_train)

X\_test = scaler.transform(X\_test)

* Train the Logistic Regression model:
* model = LogisticRegression()

model.fit(X\_train, y\_train)

* Make predictions:

y\_pred = model.predict(X\_test)

**5. Model Evaluation**

* **Accuracy Score:** Measures the percentage of correct predictions.
* accuracy = accuracy\_score(y\_test, y\_pred)

print("Accuracy:", accuracy)

* **Classification Report:** Provides Precision, Recall, and F1-score.

print(classification\_report(y\_test, y\_pred))

**6. Results and Conclusion**

* The Logistic Regression model provides an accuracy score of approximately **X%** (depending on dataset variations).
* The classification report shows the precision, recall, and F1-score for both diabetic and non-diabetic classes.
* The model can help in predicting diabetes efficiently and can be improved with more data and feature engineering.

**7. Future Work**

* Using advanced algorithms like Random Forest or Neural Networks.
* Collecting more diverse data for better generalization.
* Implementing feature selection techniques to improve accuracy.

**8. References**

* Machine Learning with Python Documentation
* Scikit-learn Library
* Kaggle Diabetes Datasets