

Diabetes Prediction

A.I. Project (Spring 1401 – 1402)

Kiarash Kiani – 9832095

Here is the description:

The goal of this project is to build a diabetes prediction tool using logistic regression. Logistic regression is a commonly used algorithm for binary classification tasks. The dataset used for this project is the "diabetes_data.csv" file, which contains various health-related features such as glucose level, blood pressure, body mass index (BMI), and age.

The project starts with data preprocessing steps. Zero values in relevant columns are replaced with NaN to handle missing values. Rows with missing values are then dropped from the dataset. The data is split into features (X) and the target variable (y). A train-test split is performed to create training and validation sets, with 80% of the data used for training and 20% for validation.

Feature scaling is applied using the StandardScaler from scikit-learn to ensure that all features are on a similar scale. This step is crucial for the logistic regression algorithm to perform optimally.

The logistic regression algorithm is implemented from scratch in the LogisticRegression class. The class provides methods to fit the model to the training data, make predictions on new data, and calculate prediction probabilities using the sigmoid function. The weights of the logistic regression model are updated using gradient descent during the training process.

The model is trained using the training data and evaluated on the validation set. The accuracy of the model is calculated using the accuracy_score metric from scikit-learn.

The program prompts the user to enter information about a patient, such as the number of pregnancies, glucose level, blood pressure, etc. The user input is preprocessed to match the format of the training data, using the same StandardScaler object. The preprocessed input is then fed into the trained logistic regression model to make predictions on whether the patient is diabetic or not.

The program allows the user to make multiple predictions and provides the prediction result as "Diabetic" or "Non-diabetic". The user can choose to continue making predictions or exit the program.

This project demonstrates the implementation of logistic regression, data preprocessing techniques, and the use of scikit-learn libraries for data splitting, feature scaling, and model evaluation. It provides an accurate prediction tool for diabetes based on the given features, which can be used for healthcare and medical purposes.

Some test-case :

Accuracy: 0.75

=====

=== Welcome to the diabetes prediction tool! ===

=====

Enter the following information about the patient:

- + Number of pregnancies: 1
- + Glucose level: 122
- + Blood pressure: 90
- + Skin thickness: 51
- + Insulin level: 220
- + Body mass index (BMI): 49.7
- + Diabetes pedigree function: 0.325
- + Age: 31

Prediction: Diabetic

Would you like to make another prediction? (y/n): y

=====

Enter the following information about the patient:

- + Number of pregnancies: 1
- + Glucose level: 163
- + Blood pressure: 72
- + Skin thickness: 0
- + Insulin level: 0
- + Body mass index (BMI): 39
- + Diabetes pedigree function: 1.222
- + Age: 33

Prediction: Diabetic

Would you like to make another prediction? (y/n): y

=====

Enter the following information about the patient:

+ Number of pregnancies: 1

+ Glucose level: 151

+ Blood pressure: 60

+ Skin thickness: 0

+ Insulin level: 0

+ Body mass index (BMI): 26.1

+ Diabetes pedigree function: 0.179

+ Age: 22

Prediction: Non-diabetic

Would you like to make another prediction? (y/n): y

=====

Enter the following information about the patient:

+ Number of pregnancies: 0

+ Glucose level: 125

+ Blood pressure: 96

+ Skin thickness: 0

+ Insulin level: 0

+ Body mass index (BMI): 22.5

+ Diabetes pedigree function: 0.262

+ Age: 21

Prediction: Non-diabetic

Would you like to make another prediction? (y/n): n

Take care!