

# Diabetes Prediction with Logistic Regression

Spring 1401 - 1402

Deadline: 15th of Tir

---

## Introduction:

Diabetes is a prevalent and chronic disease that affects millions of people worldwide. Early detection and accurate prediction of diabetes can significantly improve patient outcomes and guide proactive healthcare interventions. In this project, we aim to utilize logistic regression, a popular machine learning algorithm, to develop a predictive model for diabetes.

Logistic regression is a powerful tool for binary classification problems, where the objective is to predict one of two possible outcomes. In the case of diabetes prediction, we will train our logistic regression model on a dataset containing various patient attributes such as age, body mass index (BMI), blood pressure, and glucose levels, among others. These attributes will serve as input features, and our model will learn to distinguish between diabetic and non-diabetic individuals based on these characteristics.

## Phase One: Data Preparation

In the first phase of the project, the focus is on preparing the required data for training and evaluating the logistic regression model for diabetes prediction. To accomplish this, you can utilize [this](#) website. Ensure that the data obtained, is in a suitable format and ready for training the logistic regression model.

## Phase Two: Training a Logistic Regression Model for Diabetes Prediction

In the second phase of the project, the primary objective is to utilize the prepared data from Phase One and train a logistic regression model for predicting the probability of having diabetes based on a person's features. This phase involves the following key steps:

---

**Splitting Data:** Divide the prepared dataset into two subsets: a training set and a validation set (or test set). The training set will be used to train the logistic regression model, while the validation set will be utilized to assess the model's performance.

**Model Training:** Apply the logistic regression algorithm to the training data. During this step, the model learns the relationships between the input features (such as age, BMI, blood pressure, glucose levels) and the target variable (diabetic or non-diabetic). The objective is to estimate the coefficients (weights) for each feature, which will be used to calculate the probability of having diabetes.

**Model Evaluation:** Assess the performance of the trained logistic regression model using the validation set. Common evaluation metrics for binary classification models, such as accuracy, can be employed to gauge how well the model predicts the probability of having diabetes based on the person's features.

By completing Phase Two, the project establishes a trained logistic regression model capable of predicting the probability of having diabetes based on a person's features.

**Note:** You are required to implement the logistic regression algorithm yourself. (using pre-existing libraries isn't acceptable)

### **Phase Three:** Developing a User Interface for Diabetes Prediction

In the third phase of the project, you must develop a user interface that allows users to input their features, and the logistic regression model predicts whether they are diabetic or non-diabetic.

**Good Luck**