Introduction:

Our team at IBM, inspired by the vision of "Naan Mudhalvan," has transformed the design of our AI Diabetes prediction system into an innovative solution to tackle this pressing problem. Recognizing the urgent need for accurate and accessible predictive tools in healthcare, we embarked on this mission to make a difference.

Methodology:

In the core of our innovation lies the Logistic Regression Model, a robust analytical tool that serves as the foundation of our prediction system. This advanced statistical technique enables us to meticulously analyze complex datasets, identifying patterns and correlations that might otherwise go unnoticed. By utilizing this model, we not only predict the likelihood of diabetes but also empower individuals and healthcare professionals with valuable insights for early intervention.

By leveraging cutting-edge technology and the power of data analytics, our project aims to revolutionize diabetes prediction, paving the way for a healthier future.

Sigmoid Function:

The core of logistic regression is the sigmoid function (also called the logistic function). The sigmoid function maps any real-valued number to the range [0, 1]. It has an S-shaped curve and is defined as follows:

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

Where z is a linear combination of the input features and their respective weights, plus a bias term. In mathematical terms, for a binary classification task with n features:

$$z = b + \sum_{i=1}^{n} (w_i \times x_i)$$

Here, b represents the bias term, w_i represents the weights associated with each feature x_i.

Probability Prediction:

The sigmoid function takes the linear combination of features and weights as input and squashes it between 0 and 1. This output represents the probability of the instance belonging to the positive class (class 1). For example, if the output is 0.7, it means there is a 70% chance that the instance belongs to the positive class.

Decision Boundary:

To make a binary decision (class 0 or class 1), a threshold is set (usually 0.5). If the predicted probability is greater than or equal to 0.5, the instance is classified as class 1; otherwise, it is classified as class 0. The decision boundary is the line (or hyperplane in higher dimensions) where the sigmoid function equals the threshold value.

Training the Model:

During the training process, the model adjusts the weights and the bias term to minimize the difference between the predicted probabilities and the actual class labels in the training data. This is typically done using optimization techniques such as gradient descent, which iteratively updates the weights to minimize the loss function.

Loss Function:

The loss function used in logistic regression is the log loss (or cross-entropy loss). It measures the dissimilarity between the predicted probabilities and the actual labels. Minimizing this loss function during training ensures that the model's predicted probabilities align well with the true labels.

Evaluation Metrics:

Common evaluation metrics for logistic regression models include accuracy, precision, recall, F1-score, and the area under the ROC curve (AUC-ROC). These metrics help assess the model's performance and its ability to correctly classify instances into their respective classes.

Team Members:

1. M.Keerthivasan
2. P.Thivahar
3. A.Pachaiyappan
4. A.K.Praveen kumar
5. S.Muralidharan