**Introduction:**  
Diabetes is one of the most widespread chronic diseases, affecting millions globally and costing billions of dollars in healthcare expenses each year. Identifying individuals at risk and predicting the onset of diabetes is critical for early intervention and better disease management. This project aims to leverage health-related survey data to classify individuals based on their diabetes risk, providing actionable insights for public health officials and healthcare providers.

**Data Description:**  
For this project, we will use the **Diabetes Health Indicators Dataset** from Kaggle. This dataset contains survey responses from the Behavioral Risk Factor Surveillance System (BRFSS) conducted by the CDC.

We will use this data to build models that can classify individuals into one of these three categories based on their health and lifestyle indicators.

**Problem Statement:**  
The goal of this project is to develop predictive models using classification algorithms to accurately classify individuals into one of three diabetes-related categories.

**Dataset Description:**

We will be using second-hand data: **Diabetes Health Indicators Dataset**

The dataset includes 253,680 responses with 21 feature variables such as **BMI, Smoking Status, Physical Activity**, and others. The target variable for classification is **Diabetes\_012**, with three possible values (0, 1, 2).

**Link for the dataset**:  
[Diabetes Health Indicators Dataset](https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset)  
**Source of the dataset**: Kaggle

**Group Members:**

Group 9

* Aishwarya Balmoori
* Krishna Priyanka Challa
* Peihua Tsai
* Tara Canugovi