**DATA MINING PROJECT - ISM6136.005F23.80128**

# Prof. Arindam Ray

**Team: Data Wranglers**

**Team members:**   
  
1. Pavan Teja Gupta Allenki - U44608698  
2. Manvitha Nagandla - U45845647  
3. Gayatri Mohan Kshirsagar - U49498997  
4. Sneha Panjala - U80132269  
5. Pradeep Chand Potturi - U63754285

**Project Proposal:**

**Data set:** <https://archive.ics.uci.edu/dataset/891/cdc+diabetes+health+indicators>

**Data Set Characteristics:**

Associated Task: Classification

Number of Instances: 253,680

Number of Features: 21

**Problem Statement:**

Developing a predictive model to find the association between lifestyle factors and diabetes prevalence in the United States to provide data-driven insights for informed decision-making in efforts to combat diabetes.

**Objective:**

Diabetes is a growing public health concern in the United States, affecting millions of individuals and imposing a significant economic burden on the healthcare system. Lifestyle factors, such as age, sex, and mental health, physical health, are known to influence the risk of developing diabetes. However, the intricate interplay between these factors and the prevalence of diabetes remains a complex and insufficiently explored issue.

This data mining project aims to investigate and model the relationship between lifestyle factors and the prevalence of diabetes in the United States. The primary objectives are as follows

This project aims to:

1. Compile and refine a comprehensive dataset containing lifestyle-related variable .

2. Utilize advanced data mining techniques to uncover hidden patterns and associations within this dataset, focusing on how specific lifestyle factors interplay with the prevalence of diabetes.

3. Develop predictive models that leverage machine learning algorithms to estimate diabetes rates based on various lifestyle attributes, offering insight into the factors most strongly linked to the disease.

4. Translate the data-driven findings into actionable insights, providing healthcare practitioners, policymakers, and individuals with clear and targeted recommendations for preventative measures and interventions.

**Closing Insights:**

This project seeks to offer a detailed analysis of the complex connection between lifestyle decisions and diabetes prevalence in the United States, with the goal of equipping stakeholders with the knowledge to make well-informed decisions in the fight against diabetes.