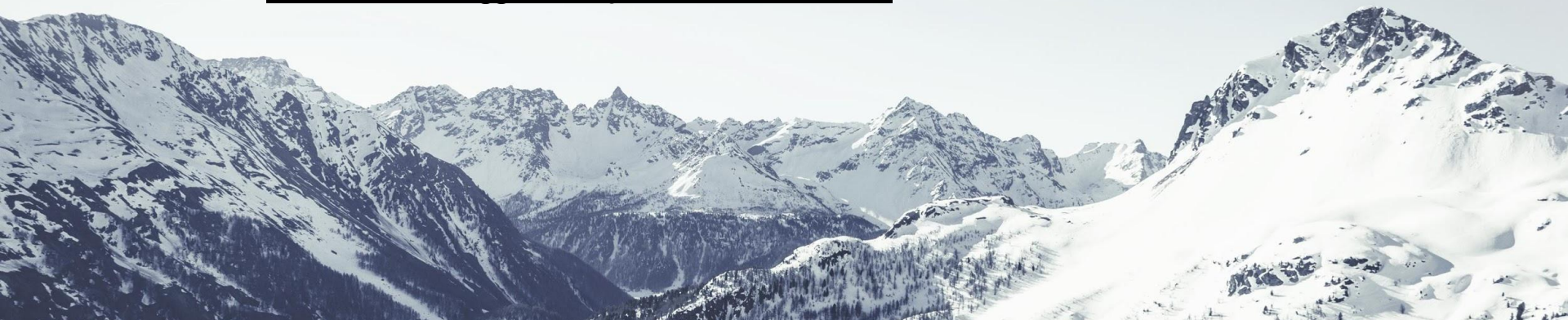
# Diabetes Risk Factors

## Case Study

Dataset : **https://www.kaggle.com/johndasilva/diabetes**



# Overview

**Diabetes** mellitus, commonly known as diabetes, is a **metabolic disease** that causes high blood sugar. People with diabetes have an **increased risk** of developing a number of **serious health problems**. They also have a **higher risk of developing infections**. In almost all high-income countries, diabetes is a **leading cause of cardiovascular disease**, blindness, kidney failure, and lower limb amputation. Although several factors are considered to lead to diabetes, it would be worth enough to **find the most predominant factors** causing this problem to gain a better understanding of the issue. We aim to apply **data mining** and statistical analysis techniques to **identify** the **dominant factors** causing diabetes in people.

# Approach

1

### Preprocessing & EDA

Univariate and Bivariate Analysis, Correlation Analysis, Missing values treatment, Handling Outliers, Analysis of Data Distribution.

2

### Data Mining & Statistical Analysis

SMOTE algorithm for balancing the dataset, Cross-validation to avoid Over-fitting, Decision Tree, Support Vector Machine for modelling and Boosting techniques to increase the prediction accuracy.

3

### Identify interesting patterns

Pose some questions over the data and apply the suitable techniques to gain insights. Visualize and provide conclusion over the risk factors of Diabetes.


Insights from
Exploratory Data Analysis

1. Imbalanced Dataset - **OVERSAMPLING** using SMOTE technique.

2. **Blood Pressure** feature has a **Normal distribution**.

3. All most **all features** except Glucose **has outliers.**

4. Features such as Pregnancies, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function is all **RIGHT-SKEWED.**

5. **Pregnancies, Blood Pressure, Skin Thickness, Glucose** increases as Age increases.

6. Blood Pressure, and **Skin Thickness increases as BMI increases.**

7. **Glucose**, Skin Thickness and BMI increases as **Insulin increases.**

8. **Blood Pressure**, Skin Thickness, **Insulin** and **BMI increases as Glucose increases.**

# Correlation Analysis

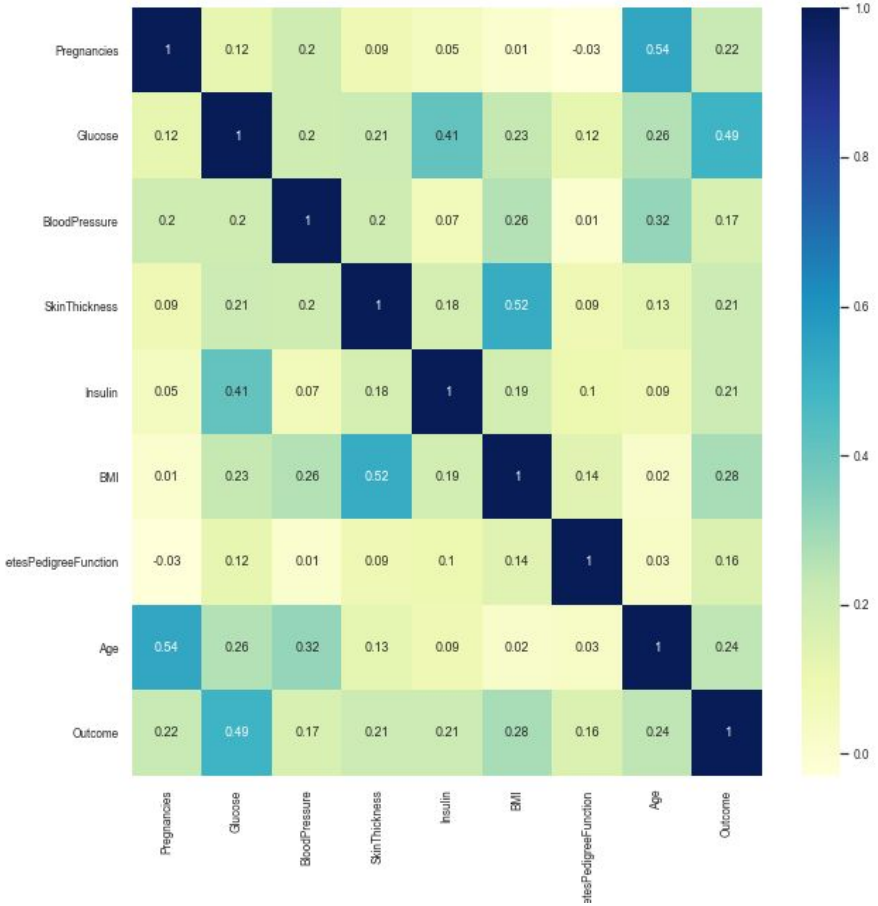
*Highly Correlated Features :*

Age - Pregnancies

Skin Thickness - BMI

Glucose - Outcome

Glucose - Insulin



# The Team!

| | |
|---|---|
| Dharini B | CB.EN.U4CSE17111 |
| Indu Varshini J | CB.EN.U4CSE17121 |
| Karthika Gurubarani M | CB.EN.U4CSE17127 |
| Mona Sweata S K | CB.EN.U4CSE17133 |

# Thank you.