

Pima Indians Diabetes Dataset

Context

Bộ dữ liệu được lấy từ **National Institute of Diabetes and Digestive and Kidney Diseases**.

Mục tiêu: dự đoán chẩn đoán bệnh tiểu đường dựa trên các chỉ số y tế.

Đối tượng: nữ giới ≥ 21 tuổi, người gốc Pima (Arizona, Mỹ).

Content

- Gồm nhiều biến dự đoán y khoa (medical predictors) và một biến mục tiêu **Outcome**.
 - Các biến dự đoán bao gồm:
 - Số lần mang thai
 - Nồng độ glucose
 - Huyết áp
 - Độ dày da (Skin Thickness)
 - Insulin
 - BMI
 - Chỉ số di truyền (Diabetes Pedigree Function)
 - Tuổi
 - Biến mục tiêu (**Outcome**):
 - 0 = không mắc tiểu đường
 - 1 = mắc tiểu đường
-

```
In [1]: import pandas as pd
```

```
columns = [
    "Pregnancies",
    "Glucose",
    "BloodPressure",
    "SkinThickness",
    "Insulin",
    "BMI",
    "DiabetesPedigreeFunction",
    "Age",
    "Outcome"
]
```

```
df = pd.read_csv("C:\\\\Users\\\\sonvuuu\\\\Desktop\\\\DA\\\\Lab03\\\\pima-indians-diabetes.csv")
df.head(5)
```

Out[1]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction
0	6	148	72	35	0	33.6	0
1	1	85	66	29	0	26.6	0
2	8	183	64	0	0	23.3	0
3	1	89	66	23	94	28.1	0
4	0	137	40	35	168	43.1	2



In []: