

Nguồn DB

[Database]<https://archive.ics.uci.edu/dataset/34/diabetes>

Data Sources

- **Electronic device:** ghi dữ liệu với timestamp thực.
- **Paper records:** chỉ có mốc thời gian logic (ăn sáng, trưa, tối, đi ngủ).
 - Gán giờ cố định:
 - Breakfast → 08:00
 - Lunch → 12:00
 - Dinner → 18:00
 - Bedtime → 22:00

👉 Như vậy, **hồ sơ giấy có thời gian giả định**, còn **thiết bị điện tử có timestamp thực tế**.

File Structure

Mỗi bản ghi có **4 trường**, phân cách bằng tab, mỗi dòng là một record:

1. **Date** – MM-DD-YYYY
 2. **Time** – HH:MM
 3. **Code** – mã sự kiện/hoạt động
 4. **Value** – giá trị đo lường
-

Code Definitions

Code	Ý nghĩa
33	Liều insulin Regular
34	Liều insulin NPH
35	Liều insulin UltraLente
48	Đo glucose (không xác định)
57	Đo glucose (không xác định)
58	Đo glucose trước bữa sáng
59	Đo glucose sau bữa sáng

Code	Ý nghĩa
60	Đo glucose trước bữa trưa
61	Đo glucose sau bữa trưa
62	Đo glucose trước bữa tối
63	Đo glucose sau bữa tối
64	Đo glucose trước bữa phụ (snack)
65	Triệu chứng hạ đường huyết
66	Ăn uống bình thường
67	Ăn nhiều hơn bình thường
68	Ăn ít hơn bình thường
69	Vận động bình thường
70	Vận động nhiều hơn bình thường
71	Vận động ít hơn bình thường
72	Sự kiện đặc biệt (không xác định)

In [5]: `import pandas as pd`

```
columns = [
    "Date",
    "Time",
    "Code",
    "Value"
]
```

```
df = pd.read_csv("C:\\Users\\sonvuuu\\Desktop\\DA\\Lab03\\data\\diabetes-data\\data")
df.head(10)
```

Out[5]:

	Date	Time	Code	Value
0	04-21-1991	9:09	58	100
1	04-21-1991	9:09	33	9
2	04-21-1991	9:09	34	13
3	04-21-1991	17:08	62	119
4	04-21-1991	17:08	33	7
5	04-21-1991	22:51	48	123
6	04-22-1991	7:35	58	216
7	04-22-1991	7:35	33	10
8	04-22-1991	7:35	34	13
9	04-22-1991	13:40	33	2