Diabetes_Dataset_Info

당뇨병 데이터셋 (Diabetes Dataset)

이 데이터셋은 n = 442명의 당뇨병 환자에 대한 10가지 기초 변수(나이, 성별, 체질량지수, 평균 혈압, 6가지 혈청 검사 수치)와 1년 후 당뇨병 진행 정도를 나타내는 목표 변수(타겟)를 포함하고 있다.

데이터셋 특성

- **샘플 수:** 442개
- **특징(속성) 수:** 10개 (모두 숫자형)
- **타겟 변수:** 1년 후 당뇨병 진행 정도 (수치형)

속성 정보

- **age**: 나이 (years)
- **sex**: 성별
- **bmi**: 체질량지수 (Body Mass Index)
- **bp**: 평균 혈압 (average blood pressure)
- **s1**: 총 혈청 콜레스테롤 (tc, total serum cholesterol)
- **s2**: 저밀도 지단백 (ldl, low-density lipoproteins)
- **s3**: 고밀도 지단백 (hdl, high-density lipoproteins)
- **s4**: 총 콜레스테롤 / HDL 비율 (tch, total cholesterol / HDL)
- **s5**: 혈청 트리글리세라이드의 로그 값 추정치 (ltg, possibly log of serum triglycerides level)
- **s6**: 혈당 수치 (glu, blood sugar level)

추가 정보

- 모든 특징 변수(속성)는 **평균을 0으로, 표준편차를** `√n_samples` **로 정규화**하여 변환됨 (각 열의 제곱합이 1이 되도록 스케일링됨).
- 데이터 출처: North Carolina State University