

Studi Kasus Data Set Dan Pemilihan Pengenalan Pola Yang Tepat



Nama : Linggar Maretva Cendani
NIM : 24060117120031

Dataset

Dataset

Dataset yang digunakan pada studi kasus pengenalan pola adalah data **Pima Indians Diabetes Database** (<https://www.kaggle.com/uciml/pima-indians-diabetes-database>), yang merupakan data milik *National Institute of Diabetes and Digestive and Kidney Diseases* (NIDDK).

Dataset ini berisi data pasien wanita berusia minimal 21 tahun yang memiliki darah suku Pima Indian.

Dataset ini merupakan salah satu jenis structured data, yang sudah berada dalam satu bentuk file csv, dan memiliki informasi berupa : kehamilan yang dialami pasien, kadar glukosa, tekanan darah, ketebalan kulit, kadar insulin, BMI, keterangan silsilah diabetes, usia, dan keterangan mengidap diabetes atau tidak (Outcome).

Pola Yang Didapatkan

Pola Klasifikasi

Pada dataset yang tersedia, terdapat kolom target (Outcome) yang menerangkan bahwa pasien mengidap diabetes atau tidak. Pasien pengidap diabetes memiliki nilai “1”, sedangkan yang tidak memiliki nilai “0”.

Karena target yang akan digunakan untuk prediksi hanya memiliki 2 kelas (1 dan 0), maka jenis pola yang dapat dikenali adalah **binary classification** (klasifikasi biner).

# Pregnancies	# Glucose	# BloodPres...	# SkinThickn...	# Insulin	# BMI	# DiabetesP...	# Age	# Outcome
6	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	0
8	183	64	0	0	23.3	0.672	32	1
1	89	66	23	94	28.1	0.167	21	0
0	137	40	35	168	43.1	2.288	33	1

Referensi Yang Digunakan

Referensi

- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2245318/>
- <https://ieeexplore.ieee.org/document/6399663>
- <https://www.sciencedirect.com/science/article/abs/pii/S0140673671923038>
- <https://www.cabdirect.org/cabdirect/abstract/19921441489>

Terima Kasih