

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	Diabetes_012	HighBP	HighChol	CholCheck	BMI	Smoker	Stroke	HeartDiseaseorAttack	PhysActivity	Fruits	Veggies	HvyAlcoholConsump	AnyHealthcare	NoDocbcCost	GenHlth	MentHlth	PhysHlth	DiffWalk	Sex	Age	Education	Income
2	0.0	0	1.0	1	15.0	1.0	0.0	0.0	0	1	1	0	1	0.0	5.0	10.0	20.0	0.0	0	11	4.0	5.0
3	2.0	1	0.0	1	28.0	0.0	0.0	1.0	0	1	0	0	1	0.0	2.0	0.0	0.0	0.0	0	11	4.0	3.0
4	2.0	1	1.0	1	33.0	0.0	0.0	0.0	1	1	1	0	1	0.0	2.0	10.0	0.0	0.0	0	9	4.0	7.0
5	2.0	0	1.0	1	29.0	0.0	1.0	1.0	1	1	1	0	1	0.0	5.0	0.0	30.0	1.0	1	12	3.0	4.0
6	0.0	0	0.0	1	24.0	1.0	0.0	0.0	0	0	0	0	1	0.0	3.0	0.0	0.0	1.0	1	13	5.0	6.0
7	0.0	1	0.0	1	40.0	1.0	0.0	0.0	1	1	1	0	1	0.0	3.0	5.0	25.0	1.0	0	10	4.0	8.0
8	0.0	1	1.0	1	27.0	1.0	0.0	0.0	0	0	1	0	1	0.0	4.0	25.0	0.0	0.0	0	10	5.0	3.0
9	2.0	0	0.0	1	24.0	0.0	0.0	1.0	0	0	1	0	1	0.0	4.0	0.0	0.0	0.0	1	12	6.0	7.0
10	0.0	1	1.0	1	30.0	0.0	0.0	0.0	0	1	1	0	1	0.0	2.0	0.0	0.0	0.0	0	7	4.0	6.0
11	0.0	1	1.0	1	36.0	1.0	0.0	0.0	0	0	0	0	1	0.0	4.0	0.0	0.0	0.0	1	10	4.0	8.0
12	2.0	0	0.0	1	33.0	1.0	0.0	0.0	1	0	1	0	1	0.0	4.0	0.0	0.0	0.0	1	6	5.0	2.0
13	2.0	1	1.0	1	30.0	0.0	0.0	0.0	0	1	1	0	1	0.0	3.0	2.0	4.0	0.0	0	11	5.0	4.0
14	2.0	0	0.0	1	27.0	0.0	0.0	0.0	0	0	1	0	1	0.0	4.0	0.0	30.0	0.0	1	13	6.0	5.0
15	2.0	0	0.0	1	31.0	1.0	0.0	0.0	0	1	1	0	1	0.0	3.0	10.0	10.0	0.0	0	8	4.0	8.0
16	0.0	1	1.0	1	27.0	1.0	0.0	0.0	0	1	1	0	1	0.0	3.0	0.0	2.0	1.0	0	13	6.0	6.0
17	0.0	1	1.0	1	35.0	1.0	0.0	0.0	1	1	0	0	1	0.0	3.0	5.0	3.0	0.0	0	11	4.0	5.0
18	0.0	1	1.0	1	30.0	0.0	0.0	0.0	1	1	0	0	1	0.0	2.0	0.0	10.0	0.0	0	10	6.0	8.0
19	0.0	0	0.0	1	31.0	0.0	0.0	0.0	1	1	1	0	1	0.0	1.0	0.0	2.0	0.0	1	11	6.0	8.0
20	1.0	1	1.0	1	33.0	0.0	0.0	0.0	0	1	1	0	1	0.0	2.0	0.0	0.0	0.0	0	11	5.0	5.0
21	0.0	0	0.0	1	31.0	1.0	0.0	0.0	1	1	1	0	1	0.0	4.0	5.0	0.0	0.0	0	10	5.0	6.0
22	0.0	1	0.0	1	29.0	0.0	0.0	0.0	1	1	1	0	1	0.0	4.0	0.0	0.0	0.0	1	11	4.0	5.0
23	0.0	1	0.0	1	27.0	1.0	1.0	0.0	1	0	0	0	1	0.0	3.0	7.0	15.0	0.0	0	7	5.0	5.0
24	2.0	1	1.0	1	22.0	0.0	0.0	1.0	1	0	1	0	1	0.0	3.0	0.0	0.0	0.0	0	12	4.0	5.0
25	0.0	1	0.0	1	29.0	1.0	0.0	0.0	1	1	1	0	1	0.0	3.0	0.0	30.0	0.0	1	12	6.0	8.0
26	0.0	0	0.0	1	22.0	0.0	0.0	0.0	1	1	1	0	1	0.0	1.0	0.0	0.0	0.0	0	9	6.0	8.0
27	0.0	1	1.0	1	26.0	1.0	0.0	0.0	0	1	1	0	1	0.0	3.0	0.0	0.0	1.0	0	13	5.0	8.0
28	0.0	1	1.0	1	26.0	1.0	0.0	0.0	0	1	1	0	1	0.0	3.0	0.0	0.0	1.0	0	12	5.0	2.0
29	2.0	0	1.0	1	28.0	1.0	0.0	0.0	1	0	0	0	1	0.0	3.0	0.0	0.0	0.0	1	11	5.0	9.0
30	0.0	0	0.0	1	28.0	0.0	0.0	0.0	1	1	1	0	1	0.0	2.0	0.0	0.0	0.0	1	8	6.0	10.0

Context:

Diabetes adalah kondisi kesehatan kronis yang memengaruhi cara tubuh Anda mengubah makanan menjadi energi. Ada tiga jenis utama diabetes: tipe 1, tipe 2, dan diabetes gestasional.

- **Diabetes tipe 1** adalah penyakit autoimun yang menyebabkan tubuh Anda menyerang sel-sel dalam pankreas yang memproduksi insulin. Insulin adalah hormon yang membantu tubuh Anda menggunakan glukosa sebagai energi.
- **Diabetes tipe 2** adalah jenis diabetes yang paling umum. Ini terjadi ketika *tubuh Anda tidak merespons secara normal terhadap insulin*, atau ketika tubuh Anda tidak memproduksi cukup insulin.
- **Diabetes gestasional** adalah jenis diabetes yang berkembang selama kehamilan. Biasanya akan hilang setelah bayi lahir.

Prevalensi Diabetes

Menurut CDC BRFSS 2021, 34,1 juta orang dewasa di Amerika Serikat mengidap diabetes, atau 10,5% dari populasi orang dewasa. Jumlah ini terus meningkat dari waktu ke waktu. Pada tahun 2010, 29,1 juta orang dewasa di Amerika Serikat mengidap diabetes, atau 9,3% dari populasi orang dewasa.

Isi Dataset

Sistem Pengawasan Faktor Risiko Perilaku (BRFSS) adalah survei telepon berbasis negara bagian yang sedang berlangsung yang mengumpulkan data tentang perilaku berisiko terkait kesehatan, kondisi kesehatan kronis, dan penggunaan layanan pencegahan di antara orang dewasa berusia 18 tahun ke atas yang tinggal di Amerika Serikat. Dilaksanakan setiap tahun oleh Pusat Pengendalian dan Pencegahan Penyakit (CDC), BRFSS telah memberikan wawasan yang berharga tentang status kesehatan dan perilaku orang dewasa AS sejak didirikan pada tahun 1984.

Untuk dataset ini, digunakan csv dari dataset BRFSS 2021 yang tersedia di Kaggle. diabetes_012_indikator_kesehatan_BRFSS2021.csv adalah set data bersih dari 236.378 tanggapan survei terhadap BRFSS2021 CDC. Variabel target Diabetes_012 memiliki 3 kelas:

- 0 untuk tidak ada diabetes atau hanya selama kehamilan,
- 1 untuk pradiabetes, dan
- 2 untuk diabetes.

Terdapat ketidakseimbangan kelas dalam dataset ini. Dataset ini memiliki 21 variabel fitur. Mungkin aku bakal pakai SMOTE buat imbalanced Data biar seimbang.

Penting untuk ditegaskan kembali bahwa saya tidak membuat dataset ini, ini hanyalah dataset yang telah dibersihkan dan dikonsolidasikan yang dibuat dari dataset BRFSS 2021 yang sudah ada di Kaggle. Dataset tersebut dapat ditemukan [di sini](#) dan buku catatan yang saya gunakan untuk membersihkan data dapat ditemukan [di sini](#).

Inspirasi

[Alex Teboul](#) yang membersihkan dataset untuk penggunaan Machine Learning dengan menggunakan BRFSS 2015 menjadi inspirasi untuk membuat dataset ini dan mengeksplorasi BRFSS secara umum.