

DiabetesUFO

ML simplon sur du dataset medical pour prévoir la maladie du diabète; avec Anna, Olivier et Fidel :)

Dataset utilisé

Le dataset utilisé provient de la source:

<https://archive.ics.uci.edu/dataset/529/early+stage+diabetes+risk+prediction+dataset>

Il contient des informations médicales et des symptômes rapportés par des patients, ainsi qu'un diagnostic final indiquant s'ils sont atteints de diabète ou non.

Attention: nous utilisons les données de [data/train_with_id.csv](#) pour l'apprentissage automatisé..

Description du dataset

Champ	Signification médicale / métier	Intérêt métier
ID	Identifiant unique du patient	Juste technique
Age	Âge du patient	Le risque de diabète augmente avec l'âge
Gender	Sexe du patient	Le diabète de type 2 est légèrement plus fréquent chez les hommes, mais dépend aussi des habitudes de vie
Polyuria	Urines fréquentes	Signe typique du diabète (le sucre attire l'eau dans les urines)
Polydipsia	Soif excessive	Autre symptôme clé (le corps se déshydrate à cause de la polyurie)
sudden weight loss	Perte de poids rapide	Indique un diabète mal contrôlé (le corps brûle les graisses)
weakness	Fatigue ou faiblesse générale	Résulte d'un manque d'énergie dû au glucose non utilisé
Polyphagia	Faim excessive	Le corps "pense" manquer d'énergie malgré le sucre sanguin élevé
Genital thrush	Mycose génitale	Très fréquent chez les diabétiques à cause du sucre dans les urines
visual blurring	Vision trouble	Causée par des fluctuations du taux de sucre

Champ	Signification médicale / métier	Intérêt métier
Itching	Démangeaisons	Liées à des infections cutanées ou mycoses
Irritability	Irritabilité accrue	Conséquence possible des variations de glycémie
delayed healing	Cicatrisation lente	Caractéristique du diabète (affecte les vaisseaux et nerfs)
partial paresis	Faiblesse musculaire partielle	Peut signaler une neuropathie diabétique
muscle stiffness	Raideur musculaire	Parfois associée à un mauvais métabolisme du glucose
Alopecia	Perte de cheveux	Effet secondaire possible d'un déséquilibre hormonal
Obesity	Obésité	Facteur de risque majeur du diabète de type 2
class	Diagnostic final (Positive/Negative)	Cible à prédire

Notes sur les modifications des données

- Les noms de colonnes ont été modifiés pour respecter le snakecase ainsi que tout en minuscules.
- Les valeurs des données ont été modifiées en "booléens" à valeur entière (0/1) pour faciliter l'apprentissage automatique.

Bibliothèques ajoutées

Librairie	Description courte	Commande d'installation	Utilisation principale
Matplotlib	Bibliothèque de base pour créer des graphiques 2D (courbes, histogrammes, scatter plots, etc.)	pip install matplotlib	Visualisation personnalisée et fine des données
Pandas	Outil essentiel pour la manipulation, le nettoyage et l'analyse de données tabulaires (DataFrames)	pip install pandas	Chargement, transformation et agrégation de données
Seaborn	Extension de Matplotlib qui	pip install seaborn	Visualisations statistiques (heatmaps, boxplots,

Librairie	Description courte	Commande d'installation	Utilisation principale
	simplifie la création de graphiques statistiques attrayants		pairplots, etc.)