

### Prédictions des diabètes

Nous allons voir un exemple de mise en œuvre d'un classificateur 'arbre de décision' pour le problème du diabète. L'objectif est de prédire sur la base de mesures diagnostiques si un patient est atteint de diabète ou non.

Nous construisons un modèle qui va faire des prédictions, nous devons donc trouver un moyen d'évaluer la qualité de ces prédictions. Étant donné que les prédictions par définition ne concernent que des données inédites, nous ne pouvons pas dépendre des données utilisées pour évaluer le modèle. Nous avons pour cela diviser le jeu de données en deux parties non croisées.

1. Importer les bibliothèques nécessaires pour l'algorithme Decision Tree (from sklearn import tree) utiliser (tree.DecisionTreeClassifier())
2. Nous utilisons l'ensemble d'entrainement pour construire notre modèle pour les arbres de decision.
3. Évaluer son score sur l'ensemble de test.
4. Visualiser l'arbres de décision.

#### Bonus ()

5. Calculer la matrice de confusion et l'accuracy score, Comparer les résultats obtenus entre l'algorithme de la régression Logistique et Les arbres de Decision. (from sklearn.metrics import accuracy\_score)
6. Afficher la matrice de confusion, comparer les résultats entre les deux algorithmes.
7. Quel algorithme est performant en termes de prédiction.

:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1
5	5	116	74	0	0	25.6	0.201	30	0
6	3	78	50	32	88	31.0	0.248	26	1
7	10	115	0	0	0	35.3	0.134	29	0
8	2	197	70	45	543	30.5	0.158	53	1
9	8	125	96	0	0	0.0	0.232	54	1
10	4	110	92	0	0	37.6	0.191	30	0
11	10	168	74	0	0	38.0	0.537	34	1
12	10	139	80	0	0	27.1	1.441	57	0
13	1	189	60	23	846	30.1	0.398	59	1
14	5	166	72	19	175	25.8	0.587	51	1
15	7	100	0	0	0	30.0	0.484	32	1
16	0	118	84	47	230	45.8	0.551	31	1