

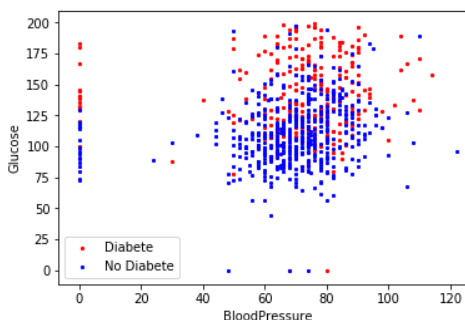
CONTROLE N° 1

Cas : prédiction du diabète

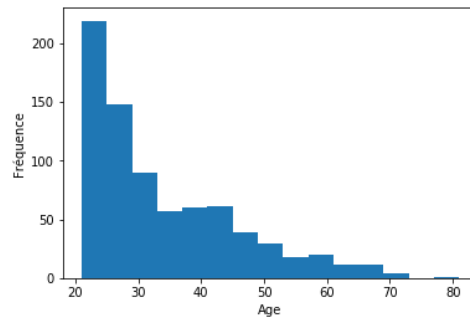
L'objectif fondamental est de prédire si oui ou non un patient est atteint du diabète à partir de certains attributs : âge, nombre de grossesses, taux d'insuline, etc...

Travail à faire :

1. Importer le module pandas
2. Importer le fichier [diabetes.csv](#) et le stocker dans une variable DataFrame du module pandas (*Df*)
3. Afficher les dimensions de *Df*
4. Afficher les variables de *Df*
5. Afficher le typage des variables
6. Calculer la moyenne le min et le max de chaque variable
7. Afficher les 5 premières lignes de *Df*
8. Afficher les 3 dernières lignes de *Df*
9. Afficher un échantillon aléatoire de 10 lignes
10. Sélectionner uniquement les colonnes : Outcome, BMI et Age
11. Faire une sélection des enregistrements où le glucose est entre 90 et 130
12. Sélectionner les données qui ont Outcome = 1 et Pregnancies>0
13. Combien de personnes présentant un diabète ont une tension artérielle supérieur à 70?
14. Quel est le maximum du "BMI" lorsque la patiente n'est pas atteinte par le diabète ?
15. Vérifier si notre *Dataset* possède des champs vides
16. Créer un graphique de type « scatter plot » Glucose vs BloodPressure :



17. Créer un histogramme des valeurs pour la colonne « âge » :



18. Diviser l'ensemble de données en un ensemble d'entraînement et un ensemble de test (TRAIN/TEST) (utiliser 20% pour l'ensemble test)

19. Utiliser l'ensemble d'entraînement pour construire un modèle pour les arbres de décision. Puis évaluer son score sur l'ensemble de test.

20. Ecrire le code permettant de visualiser l'arbre

21. Construire deux modèles KNN et SVM et afficher les précisions

22. Tracer un graphique pour comparer les performances des trois modèles