**RER - Naîves Bayes**

# Contexte :

Comprendre l’utilisation des algorithmes (Naïves Bayes , OneR) dans le machine learning et notamment dans la classification supervisée.

# Mots clés :

* Naîves Bayes : La classification naïve bayésienne est un type de classification bayésienne probabiliste simple basée sur le théorème de Bayes avec une forte indépendance (dite naïve) des hypothèses.
* Naïf: Un algorithme de machine learning dit naïf, est un algorithme qui assume que toutes les variables sont indépendante des unes et des autres.
* Théorème de Bayes : Le théorème de Bayes est l'un des principaux théorèmes de la théorie des probabilités. Il est aussi utilisé en statistiques du fait de son application, qui permet de déterminer la probabilité qu'un événement arrive à partir d'un autre évènement qui s'est réalisé, notamment quand ces deux évènements sont interdépendants.
* Indépendance :
* Classifieur (?) : un classifieur est un algorithme qui permet de classer dans des groupes ou classes des échantillons qui ont des propriétés similaires, mesurées sur des observations.
* Classifieur linéaire : un classifieur linéaire est un type particulier de classifieur, qui calcule la décision par combinaison linéaire des échantillons.
* Classification : classe les données (construit un modèle) en fonction de l'ensemble d'apprentissage et des valeurs (étiquettes de classe) dans un attribut de classification et l'utilise pour classer de nouvelles données
* Classification supervisée : Les données d'apprentissage (observations, mesures, etc.) sont accompagnées d'étiquettes indiquant la classe des observations
* Classification non-supervisée (clustering) : Les étiquettes de classe des données d'entraînement sont inconnues. le programme reçoit un ensemble de mesures, d'observations, etc. dans le but d'établir l'existence de classes ou de clusters dans les données
* Scikit-learn : bibliothèque libre Python destinée à l'apprentissage automatique
* Crisp-DM : modèle de processus d'exploration de données qui décrit une approche communément utilisée pour résoudre les problèmes du domaine de l'analyse, de l'extraction et des sciences des données.
* Apprentissage :
* OneR :

# Problématique(s) :

1. Comment fonctionne un classifieur Naîves Bayes ?
2. Comment l’utilisation d’un algorithme “naïf” à partir du Théorème de Bayes permet de classifier les données ?
3. Quelles sont les limites du Naîves Bayes ?

# Hypothèses :

1. Le Modèle Naîve Bayes permet d’entrainer la machine plus rapidement que d’autres modèles (Adeline)
2. CRISP-DM permet de faire du Clustering (Aude)
3. L’algorithme Naîve Bayes permet la classification des données (bien trop simple vu que Bassam a répondu) FAUX !!!!
4. L’algorithme de Naîve Bayes est l’algo de classification le plus utilisé (Axel)
5. L’algorithme Naîve Bayes peut servir pour classifier n’importe quel type de donnée (Briand)
6. L’algorithme Naîve Bayes est un algo d’IA plus intelligent que les autres (Osman)
7. On ne peut pas utiliser le Naîve Bayes classifier pour réaliser des diagnostiques médicaux (étienne)
8. Naîve Bayes est un algorithme de Data Mining qui permet de ressortir des règles d’une base de donnée (Adrien)
9. L’algorithme demande moins de ressource en entrée mais moins précis en sortie (Niko)
10. Naîve Bayes ne convient pas forcément à toutes les classifications de données (Tetyana)
11. La sortie d’un classifier Bayesien peut être une égalité (Seydou)
12. Le caractère naïf du théorème de bayes à fait sa réputation (Jean-Paul SOSSAH)
13. Il existe un algorithme bayesien non naïf ( Adrien Gémini)
14. Le théorème de Bayes étant un algorithme de probabilité, nous l’utilisons plus dans du data mining que dans le machine learning. (Panda)
15. On utilise les même métriques que règles d’association (Solenn sans le ‘e’ by Bassam)

# Plan d'action :

* Étudier les ressources
* Comparaison OneR / Bayes
* Découvrir Scikit-learn
* Répondre aux problématiques
* Réfléchir aux hypothèses
* Faire les Workshop en collaboration
* Livrer les Livrables