

Projet de la SAé

Vous devez choisir un jeu de données lié à une problématique de classification supervisée. Ce jeu de données devra comporter des prédictors quantitatifs et qualitatifs (données mixtes). Vous devez présenter et contextualiser les données ainsi que la problématique que vous souhaitez traiter. L'objectif est de comparer les trois approches k -NN, LDA et les arbres de décision. Vous traiterez dans une première analyse uniquement les prédictors quantitatifs et dans une deuxième l'ensemble des prédictors après avoir réalisé les transformations nécessaires pour appliquer k -NN (avec la distance euclidienne) et LDA. Vous étudierez l'importance des prédictors.

Livrable : Un rapport (**par étudiant**) à rendre **par mail** avant le 31/01/2025 contenant :

1. une présentation des données et de la problématique
2. Les deux analyses
3. une étude sur l'importance des prédictors dans la prédiction de la variable cible
4. une conclusion

Evaluation.- Le rapport sera évalué vis-à-vis des compétences suivantes :

1. rédiger un rapport
2. traiter des prédictors de nature différente
3. calibrer, valider choisir un modèle
4. analyser l'importance de chaque variable
5. analyser les résultats et défendre ses choix d'analyse