

Lab 4: Arbres de décision

Objectifs:

- Installer les packages «rpart» et «rpart.plot»;
- Importer la base de données «ptitanic» du package rpart.plot;
- Prédire le modèle à l'aide des arbres de décision;
- Visualiser les arbres de décision.



Exercice 1 : Titanic

Le 15 avril 1912, lors de son premier voyage, le Titanic a coulé après avoir heurté un iceberg, tuant 1502 sur 2224 victimes. Cette tragédie a choqué la communauté internationale et a conduit à de meilleures règles de sécurité pour les navires. Bien qu'il y ait eu une certaine chance de survivre au naufrage, certains groupes de personnes étaient plus susceptibles de survivre que d'autres, comme les femmes, les enfants et les classes supérieures. Dans cet exercice, nous vous demandons de compléter l'analyse de ce genre de personnes susceptibles de survivre. En particulier, nous vous demandons d'appliquer les outils du data mining pour prédire quels passagers ont survécu à la tragédie.

La base de données Titanic décrit 1046 passagers selon 6 variables:

- **pclass**: donne la classe tarifaire sur le bateau;
- **survived**: indique si le passager a survécu;
- **sex**: donne le sexe du passager;
- **age**: donne l'âge du passager exprimé en années;
- **sibsp**: donne le nombre de frères, sœurs, mari ou épouse à bord;
- **parch**: donne le nombre d'enfants ou de parents à bord.

- 1) Importez la base de données «ptitanic» du package rpart.plot, puis affichez le résumé de ces données à l'aide de la commande `summary()`.
- 2) Divisez la base de données en 75% pour la partie apprentissage et 25% pour la partie test.
- 3) Construisez l'arbre de décision à l'aide de la commande `rpart()` qui prendra en argument les données de la partie apprentissage, puis affichez le résultat.
- 4) Essayez de minimiser l'erreur pour définir le niveau d'élagage.
- 5) Effectuez l'élagage de l'arbre avec le `cp` optimal, puis réaffichez l'arbre.

- 6) Faites une prédiction du modèle sur les données de test.
- 7) Affichez la matrice de confusion, puis calculez l'erreur de classement.
- 8) Affichez les règles de construction de l'arbre.