

Projet - Stats-Proba - M1 FD&IA - UVB

Enseig : Armel SOUBEIGA

Introduction

Cet examen prend la forme d'un devoir maison, à réaliser en groupe de 4 personnes au maximum pour le jeudi 15 juillet.

Vous devrez me rendre par mail (armel.soubeiga@yahoo.fr):

- 1. un rapport de 4 pages maximum (sans le code) décrivant les analyses réalisées,
- 2. un notebook Jupyter ou un r Rarkdown (contenant vos codes, et commenté un minimum) permettant de reproduire vos analyses.

Projet

Dans cet exercice vous travaillerez sur une problématique d'analyse multidimensionnelle visant à fournir une typologie ou segmentation des clients d'une entreprise afin de d'aider à des prises de décision marketing. Plus précisément, vous devrez identifier des groupes ou classes homogènes de clients en fonction de leurs comportements bancaires.

Les algorithmes de classification que nous avons étudiée sont adaptés à des variables quantitatives ou des matrices de distances, et on a également mentionné que l'intégration d'informations qualitatives peut se faire par recodage ou à l'aide d'une analyse des correspondances multiples.

L'objectif de cette étude sera donc de comparer deux approches: classification à partir des seules variables quantitatives ou classification à partir des scores issues d'une analyse factorielle des composantes multiples (afcm).

Présentation des données

Il s'agit d'une base de données retraçant l'historique mensuel bancaire et les caractéristiques de tous les clients de l'entreprise.

Un premier nettoyage des données d'origines raw data uncleaned a été effectué pour permettre que tous les étudiants aient les mêmes data de départ. Les scripts de nettoyages peuvent êtres consultés ici.

Ainsi vous dispose :

- **visa_raw_cleaning.xlsx** : les données
- Les libellés des variables

Ressources

[1] https://mtes-mct.github.io/parcours_r_module_analyse_multi_dimensionnelles/

[2] <https://lbelzile.github.io/math60602/>