**TER : Scoring**

1. **Introduction**

L’idée générale du scoring est d’affecter une note (un score) globale à un individu à partir de plusieurs descripteurs, quantitatifs ou qualitatifs. A partir de cette note, on affecte l’individu à un groupe préexistant. Un score peut donc être définit comme un outil statistique ou probabiliste de détection de risque. Le scoring peut également être vu comme l’application au monde de l’entreprise de plusieurs techniques de classement. Nous en aborderons 2 dans ce rapport.

Nous pouvons déjà citer plusieurs types de score :

→ Les scores de risque : - risque de crédit ou credit scoring : prédire le retard de remboursement de crédit.

- risque financier : prédire la bonne ou mauvaise santé d’une entreprise.

- risque médical.

→ Les scores en marketing : - score d’attrition : prédire le risque qu’un client passe à la concurrence ou résilie son abonnement.

- score d’appétence : prédire l’appétence d’un client à acheter tel ou tel type de produit.

La création d’un score se fait en fonction des objectifs recherchés et des moyens techniques disponibles. Par exemple, le développement d’un score comportemental nécessite de disposer de données sur au moins un an, si l’on a moins d’historique, il vaut mieux partir sur un score générique ou un score d’octroi.

Il faut aussi également choisir l’utilisation qui sera faite du score : outil d’aide à la décision ou outil de ciblage pour le marketing direct par exemple. C’est en fonction de l’utilisation que l’on en fera que la règle de décision sera ajustée.

Pour construire un score, il faut dans un premier temps disposer d’un échantillon suffisamment conséquent pour pouvoir tester plusieurs modèles prédictifs. De plus, pour éviter des problèmes de surestimation de la qualité du modèle, il est préférable de séparer l’échantillon d’étude en deux sous-échantillons : un échantillon d’apprentissage à partir duquel sera créé le modèle, et un échantillon test sur lequel sera testé la qualité du modèle par rapport à l’objectif recherché et au risque que l’on est prêt à prendre.

Ensuite, il faut élaborer un modèle prédictif à l’aide de techniques prédictives : analyse discriminante et régression logistique en l’occurrence.

Enfin, les notes de score sont découpées en plusieurs classes de valeur. Dans le domaine financier, on aura tendance à découper les notes de score en trois classes : faible, moyen, fortes. Dans le milieu médical, on préfèrera 2 classes : à risque, non à risque. La règle de classement (seuil comparatif du score) se décide en fonction du risque d’erreur que l’on souhaite prendre.

Nous présentons dans ce rapport 2 des techniques prédictives les plus utilisées en scoring : la régression logistique et l’analyse discriminante. Pour illustrer ce qu’est le scoring, nous avons utilisé ces 2 techniques sur 2 jeux de données différents. Nous présentons dans la suite la théorie de chaque technique ainsi que l’étude des données associée.