

Chapitre III

Méthodes de DM pour la GRC dans les banques

III.1 Introduction

Avant de chercher des techniques à appliquer dans la gestion des relations avec les clients. Il faut étudier les données des clients afin de déterminer les informations de base qui peuvent être utilisées. Ensuite on va parler sur chaque problème et la technique de data mining la plus utile pour résoudre le problème.

III.2 Données du client

III.2.1 Type de données

Peu importe d'où elles viennent, les données sont réparties en deux types de base : démographiques et comportementales.

III.2.1.1 Données personnelles [RUD01]

Les données démographiques décrites généralement des caractéristiques personnelles ou domestiques. Il comprend des caractéristiques telles que le sexe, l'âge, l'état matrimonial, le revenu, la fonction, le type de logement, le niveau de scolarité, et la présence d'enfants. Les données démographiques ont un certain nombre de points forts. Il est très stable, ce qui le rend attrayant pour une utilisation dans la modélisation prédictive. Les caractéristiques comme l'état matrimonial, la propriété, le niveau d'éducation, et le type de logement ne sont pas susceptibles de changer aussi souvent que les données comportementales telles que les soldes bancaires.

Les données démographiques est généralement moins coûteux que les données d'attitude et de comportement, en particulier lorsqu'ils sont achetés au niveau du groupe. Une des faiblesses de données démographiques, c'est qu'il est difficile d'obtenir sur une base individuelle avec un degré élevé de précision. Sauf si elle est exigée en contrepartie d'un produit ou d'un service, beaucoup de gens résistent à partager ce type d'information ou de fournir de fausses informations.

III.2.2 Données comportementales [RUD01]

Ce sont les caractéristiques qui doivent être surveillés au fil du temps, tels que la fréquence et le volume des transactions dans une période (hebdomadaire, quotidien, etc). Par exemple, un suivi quotidien des opérations de crédit et de débit pourrait être utile pour le déclenchement des campagnes.

Les données comportementales fait habituellement mieux de prédire le comportement futur que les autres types de données. Il est, cependant, généralement des données les plus difficiles et coûteux à obtenir à partir d'une source extérieure. Ceci sera discuté plus en détail dans la le titre suivante.

III.2.3 Sources de données [RUD01]

Les données pour le data mining peuvent être générés à partir d'un certain nombre de sources. Ces sources se répartissent en deux catégories : internes ou externes. Les sources internes sont ceux qui sont générés par l'activité de l'entreprise tels que les dossiers des clients, site Web, des bandes de publicité, de mail, ou par téléphone, ou les bases de données et/ou des entrepôts de données qui sont spécifiquement conçus pour accueillir des données de l'entreprise. Les sources de données externes comprennent les données des sociétés telles que les bureaux de crédit, les courtiers et les collecteurs de données, aussi les sociétés avec de grandes bases de données clients comme les éditeurs et les catalogueurs.

III.2.3.1 Sources internes

Les sources internes sont des sources qui sont inclus dans les données d'une entreprise ou un établissement. Ils sont souvent les données les plus prédictifs pour la modélisation, car ils représentent des informations spécifiques à un produit ou service de l'entreprise.

Certaines sources typiques sont la base de données client, base de données des transactions, la base de données de l'histoire de l'offre, les bandes de sollicitation, et des entrepôts de données.

III.2.3.2 Sources externes

La pression est sur de nombreuses entreprises à augmenter leurs profits, soit par l'acquisition de nouveaux clients ou par l'augmentation des ventes à des clients existants. Ces deux initiatives peuvent être améliorées grâce à l'utilisation de sources externes.

Les sources externes sont principalement constituées de la liste des vendeurs et des collecteurs. Comme on peut s'y attendre, la liste des vendeurs sont des entreprises qui vendent des listes de données.

Peu d'entreprises ont cependant la vente de listes que leur seule activité. Beaucoup d'entreprises ont une activité principale comme les ventes de magazines ou les ventes par catalogue, avec la vente des listes de données comme une activité secondaire. Selon le type d'entreprise, ils collectent et vendent des noms, adresses et numéros de téléphone, ainsi que des informations démographiques, comportementales, et/ou psychographiques. Parfois, ils effectuent un nettoyage au but d'améliorer la valeur de la liste. Beaucoup achètent leurs listes à des collecteurs de données et/ou des courtiers.

Chapitre III Méthodes de DM pour la GRC dans les banques

Les collecteurs sont des entreprises qui vendent une variété de listes de données simples et compilés. Certaines entreprises commencent avec une base comme le livre de téléphone ou les données d'enregistrement des permis de conduire. Ensuite, ils achètent des listes de données, les fusionner, et imputer les valeurs manquantes. Beaucoup collecteurs des listes utilisent la recherche par sondage afin d'améliorer et de valider leurs listes.

III.3 Cas d'acquisition des clients

Nous avons dit dans le premier chapitre que d'abord le client est juste un prospectif. Afin de devenir un client. Les fonctionnaires ont utilisé des méthodes statistiques, traditionnelle et à la main pour identifier les clients potentiels à attirer. Mais nous avons constaté que ces méthodes sont insuffisantes, étant donné leur lenteur et la grande concurrence entre les banques. Ces obstacles nécessitent l'utilisation d'autres méthodes plus efficace et plus rapide tel que les techniques de data mining.

Dans ce cas la banque ne dispose que des informations sur ses clients et ne dispose pas des informations sur les clients qui ne veulent pas profiter de ses services.

Par conséquent, l'apprentissage et premièrement supervisé puisque la banque dispose de certaines informations sur des exemples. Deuxièmement, ces informations appartenant à un seul class qui est la classe des clients ce qui conduit à l'utilisation de l'apprentissage supervisé mono-classe.

Plusieurs méthodes d'apprentissage supervisé mono-classe peut être utilisé tel que les réseaux de neurones, les réseaux bayesiens, les svm et même les technique de clustering.

III.4 Cas de rétention des clients

Dans ce cas, la banque dispose de deux types d'informations :

Des informations sur les clients qui continuent leur relation avec la banque.

Des clients qui ont déjà coupé leur relation avec la banque.

Pour pouvoir retenir ses clients, la banque doit surveiller en permanence leur comportement, une fois le comportement d'un client commence à se rapprocher de comportement des clients qui ont quitté la banque, cette dernière lance les mesures nécessaire à sa rétention tale que, des promotions, des services privilèges, des bonus,...etc.

La construction du modèle permettant de détecter de tels clients utilise à la fois les informations des clients fidèles ceux non fidèles ce qui justifie l'utilisation de la classification binaire.

III.5 Cas de segmentation des clients

Le responsable de la banque préfère de savoir les catégories de leur clients. Cela aide à déterminer la meilleure façon de communication avec eux. Par exemple gérer sa relation avec des particuliers n'est pas comme avec des entreprises. Donc la technique nécessaire est une technique de segmentation. Telles que K-means, classification hiérarchique, ...etc.

III.6 Choix des techniques

Le choix de la technique qui nous utilisons est une opération très importante. Avant de cette opération, nous allons essayer de connaître les techniques utilisées dans la stratégie de la gestion des relations avec les clients. Avec ce tableau (table 3.1) publié par Ngai et al [NGA05]. Qui parle sur les articles qui proposent des techniques de data mining utilisées pour cette stratégie.

Technique de data mining	Nombre
Réseaux de neurones	30
Arbres de décision	23
Règles d'association	28
Régression	10
Algorithmes génétique	4
Chaînes de Markov	4
Analyse de survie	4
K-means	3
KPPV	3
Réseaux Bayésien	2
Machine à vecteurs support	1

Tableau 3.1 : Répartition des articles par des techniques de data mining

Après l'observation du tableau, nous trouvons qu'il existe un groupe se constitué de réseaux de neurones, arbres de décision, règles d'association, régression a plus d'intérêt que les autres technique comme réseaux bayésien, SVM...etc.

Pour cela nous orientons a choisi l'un des technique qui n'est pas fréquemment étudiée, plus de ça on va intégrer les critères de ces technique comme la rapidité de décision, simplicité...etc.

Donc pour la technique d'apprentissage SVM possède les caractéristiques suivantes :

- Les SVM possèdent des fondements mathématiques solides.
- Les exemples de test sont comparés juste avec les supports vecteur et non pas avec tout les exemples d'apprentissage.
- Une décision rapide parce que la classification d'un nouvel exemple consiste à voir le signe de la fonction de décision $f(x)$.

Chapitre III Méthodes de DM pour la GRC dans les banques

- Nous n'avons pas le problème de multi-classes, parce que le processus d'apprentissage est mono-classe pour le problème d'acquisition et binaire pour le problème de rétention.

Pour les conditions précédente on à choisi SVM pour les deux problèmes d'acquisition et de rétention des clients.

Pour le problème de segmentation des clients nous trouvons que l'algorithme de segmentation k-means :

- Est simple à de mise en œuvre.
- Il partitionne les données en K clusters. Contrairement à d'autres méthodes dites hiérarchiques, qui créent une structure en « arbre de clusters » pour décrire les groupements, k-means ne crée qu'un seul niveau de clusters.
- Il renvoie une partition des données, dans laquelle les objets à l'intérieur de chaque cluster sont aussi proches que possible les uns des autres et aussi loin que possible des objets des autres clusters.
- Chaque cluster de la partition est défini par ses objets et son centroïde [FRE05].

Pour ce la on à choisi cette algorithme comme technique de segmentation des clients.

III.7 Conclusion

Dans ce chapitre nous avons expliqué les données nécessaires des clients pour l'utilisation dans la résolution des problèmes d'acquisition, rétention et de la segmentation. Ensuite, nous avons mentionné les sources de ces données et nous avons trouvé que sont des sources internes et externes. Enfin, nous choisissons les techniques nécessaires pour résoudre ses problèmes que nous allons expliquer en détail dans le chapitre suivant.