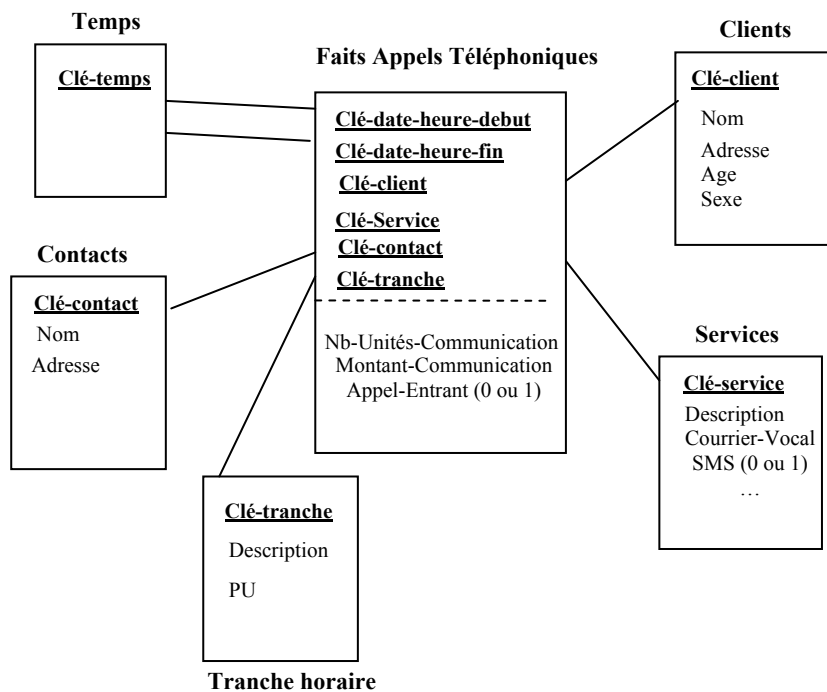


Concepteur : Ahlame DOUZAL
Durée : 3 heures

TD SUR LES METHODES DU DATA MINING

Exercice 1 : Extraction de données à partir d'un entrepôt pour le Data Mining.

Une compagnie de télécommunication dispose d'un entrepôt de données multidimensionnel dont le schéma est indiqué ci-dessous. Cet entrepôt permet le suivi de toutes les transactions " Appels Téléphoniques ".



Afin d'analyser les données relatives aux profils des consommations des clients on génère le rapport mensuel suivant :

Clients	Nb-Appel	Durée Comm	Nb-Appel-Entrant	Courrier-Vocal	Nb-Contact
C1	35	500	41	Non	16
C2	9	170	25	Non	13
C3	7	210	45	Oui	3
C4	12	220	5	Non	17
C5	28	580	39	Non	37
C6	11	180	30	Oui	5
C7	20	110	10	Oui	30
C8	40	600	50	Oui	12

Donnez la requête OLAP pour la génération du rapport mensuel ci-dessus à partir de l'entrepôt " Appels Téléphoniques ".

Exercice 2 : De l'objectif décisionnel vers la méthode en Data Mining

La compagnie de télécommunication souhaite augmenter sa rentabilité afin de rester concurrentielle. La direction demande au responsable du service technique « Base de données décisionnelles » la génération d'un rapport en réponse à chacune des requêtes ci-dessous.

- a) Existe-t-il un lien entre les services utilisés et l'âge des clients ?
- b) Nos concurrents ont établi que 80% des clients de sexe féminin âgés de 15 à 20 ans utilisent le service SMS. De combien cette assertion est-elle vraie rapportée à nos clients ?
- c) Les données de l'entrepôt contiennent-elles assez d'information pour être exploitées dans un but décisionnel ?
- d) Afin de relancer les ventes notre compagnie souhaite lancer un nouveau package incluant un ensemble d'offres à tarif préférentiel. Comment peut-on exploiter les données de l'entrepôt afin de définir la meilleure composition du package ?
- e) Les packages en question ont été mis sur le marché. La proportion de clients ayant acheté ce package est en deçà des estimations faites par la compagnie. Une nouvelle stratégie fondée sur du marketing direct auprès des clients potentiels va être lancée. Pour ce faire on a besoin de déterminer les profils clients présentant le plus de potentiel à l'acquisition des packages.

Indiquez pour chacune de ces requêtes la ou les méthodes en Data Mining nécessaires à la génération du rapport. Illustrez puis justifiez votre choix.

Exercice 2 : Mesure de similarité entre nuplets

En vous basant sur la description des profils des consommations mensuelles des clients donnée à l'exercice 1 :

- a) Évaluez la similarité entre les clients C1, C2 et C5.
- b) Quel est le couple de clients le plus similaire ?
- c) Donnez la description dans le même espace multidimensionnel du nuplet central correspondant aux clients C1 et C2.
- d) En utilisant la distance euclidienne, évaluez de combien les profils des clients C3 et C7 sont différents.

Exercice 3 : Classification hiérarchique

On considère la description des clients dans l'espace à deux dimensions défini par les attributs "Nb-Appel" et "Nb-Contact". On souhaite obtenir une succession de partitions (partitionnement horizontal) sur l'ensemble des clients. On rappelle qu'à chaque partition est associée un seuil de similarité (de distance) au delà (respect. en deçà) duquel deux nuplets sont considérés comme similaires et appartiennent donc à la même classe.

- En vous basant sur la distance euclidienne, donnez la succession de partitions et précisez pour chacune d'elle le seuil de similarité (distance) associé. Fournir une représentation graphique de la classification des nuplets.

Exercice 4 : Classification par Centres-Mobiles

En considérant toujours la description des clients dans l'espace défini par les attributs "Nb-Appel" et "Nb-Contact", on souhaite extraire trois principaux profils de consommation.

Pour ce faire utilisez la méthode des centres mobiles en prenant comme nuplets centraux les clients C1, C6 et C8. Retrouve-t-on la même partition avec une classification hiérarchique ?

Exercice 5 : extraction de règles par arbres binaires de décisions

On considère le tableau ci-dessous donnant la description d'un ensemble de cinéphiles par leur âge, leur activité, leur sexe ainsi que le dernier film vu.

Cinéphile	Age	Source	Titre	Sexe
1	M	Universitaire	Independence day	F
2	M	Universitaire	L'Armée des douze singes	M
3	P	Universitaire	The Bridcage	M
4	P	Universitaire	Trainspotting	F
5	M	Universitaire	I shot Andy Warhol	F
6	P	Universitaire	Chain Recation	F
7	P	Industriel	La liste de Schindler	M
8	M	Universitaire	Super Cop	M
9	M	Universitaire	Eddie	F
10	M	Universitaire	Le Phénomène	F
11	P	Libérale	The Bridcage	F
12	M	Universitaire	Kingpin	M
13	P	Libérale	T2 Judgment Day	M
14	P	Libérale	Independence day	M
15	M	Libérale	Trainspotting	F

M : < 30 , P : >= 30

- Procédez à l'extraction des règles de décisions permettant de déterminer le sexe du cinéphile en fonction de son âge, son activité et du dernier film vu.

Exercice 6 : Extraction de règles d'associations

Le tableau suivant donne la description de cinq transactions au niveau d'un point de vente dans une épicerie. Chaque transaction identifie le sous-ensemble de produits achetés simultanément.

Transactions	Articles
1	jus d'orange, eau minérale
2	lait, jus d'orange, nettoyant de vitre
3	jus d'orange, détergent
4	jus d'orange, détergent, eau minérale
5	nettoyant de vitre, eau minérale

- Donnez la matrice de co-occurrence associée à l'ensemble des transactions.

- Procédez à l'extraction des règles d'associations les plus pertinentes en entachant chacune d'elle de son degré de confiance et de support.