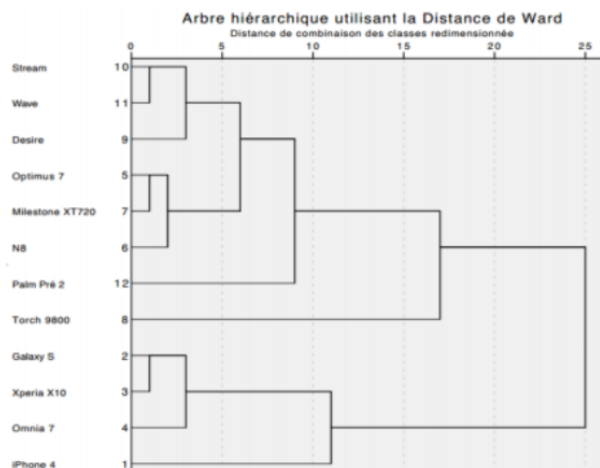


DATA MINING

Test : SEGMENTATION

Exercice 1 :

Un magazine français a publié un comparatif des 12 principaux Smartphones disponibles sur le marché en novembre 2010.



- 1- Citer l'avantage de la classification hiérarchique ascendante
- 2- Combien de groupes suggérez-vous de retenir selon le dendrogramme ?

Exercice 2 : On désire appliquer la méthode CAH sur les données suivantes : 5 individus caractérisés par deux variables X_1 et X_2 . Déterminer la matrice des distances et tracer le dendrogramme

| | X_1 | X_2 |
|-------|-------|-------|
| I_1 | 12 | 5 |
| I_2 | 8 | 16 |
| I_3 | 14 | 5 |
| I_4 | 8 | 10 |
| I_5 | 2 | 20 |

NB : Utiliser la distance de Manhattan

$$d(I, J) = |X_1(I) - X_1(J)| + |X_2(I) - X_2(J)|$$

Exercice 3 : Un chef d'entreprise souhaite avoir une idée sur les profils de ses clients afin de réaliser une action marketing « one to one ». Pour ce faire, un expert data mining se donne la mission de séparer l'ensemble des clients de l'entrepôt en deux groupes mais sans critères connus à l'avance.

1. A quelle famille d'analyse appartient la segmentation?
2. Expliquer l'utilité d'une métrique de la distance dans ces méthodes
3. Dans un premier temps, l'expert a appliqué la méthode K-Means. En croisant les résultats obtenus sur un échantillon de clients caractérisé par une variable qui identifie les clients réellement à tendance fidèle par le caractère F et les autres clients par le caractère NF, on obtient la table de contingence suivante. Donner le taux des individus bien classés.

| | 1 | 2 |
|----|-----|----|
| F | 120 | 15 |
| NF | 9 | 46 |