



CIV8760 - Laboratoire 05

Ordre du jour

→ Présentation TP4

→ Travail en équipe

Création de la base de données

Données GTFS sur [Transit Feed](#).

Chaque ligne de la base de données doit correspondre à une direction d'une ligne de service.

N'oubliez pas de retirer le métro de vos données !

Création de la base de données

Temps inter-arrêt : Vous allez avoir un temps entre chaque arrêt, donc vous devez choisir une statistique descriptive (mode, médiane, moyenne, etc.) afin de décrire le temps inter-arrêt pour CHAQUE direction de ligne de service.

Accessibilité : Vous avez une valeur par arrêt sur la ligne et une valeur pour la ligne, vous devez donc proposer une valeur à associer à la ligne en fonction des valeurs disponibles.

Segmentation des lignes

Tanagra

Sélectionnez vos données (incluant les titres) dans votre fichier Excel et copiez-les (évitez les accents et les espaces)

Créer un nouveau fichier texte (".txt"), ouvrez-le et collez-y vos données.

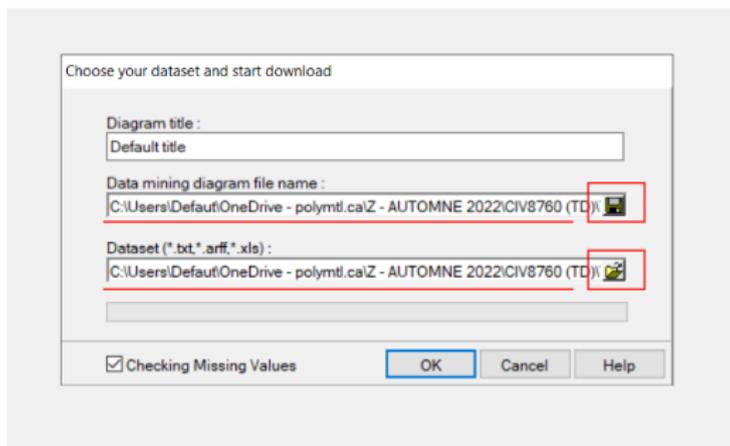
Dans Tanagra, créez un nouveau fichier projet.

Segmentation des lignes

Tanagra

Assurez-vous que le fichier sera enregistré à un endroit connu (sous "Data mining diagram file name", vous pouvez changer l'endroit où le projet sera)

Sous "Dataset (*.txt, *.arff, *.xls)", allez chercher votre fichier de données



Segmentation des lignes

Tanagra

Définissez vos attributs (*Define status*). Target ? Input ?

Plusieurs méthodes de segmentation existent → Vous avez vu en classe les k-moyenne (*K-Means*).

Pour trouver le bon nombre de classes, comment ?

Segmentation des lignes

Tanagra

Figure : Homogénéité dans les classes (WSS) et dispersion inter-classe (BSS)

Total Sum of Square (TSS) = BSS + WSS

Déterminer le nombre de classes optimal : Méthode du coude

Segmentation des lignes

Tanagra

Pour exporter les données une fois la segmentation faite, allez dans l'onglet "Data visualization" et appliquez la fonction "Export dataset"

Dans les paramètres de "Export dataset", vous pouvez préciser où sera déposé le fichier de données (format ".txt") et son nom.

Ce fichier texte contient maintenant la classe pour chaque ligne de votre base de données. Vous pouvez l'utiliser avec un autre fichier dans les données GTFS pour créer votre carte !

Segmentation des lignes

QGIS

Utiliser l'outil "Points to path" avec l'aide du fichier *shape.txt* pour afficher les lignes de la STM.

Analyse des classes

Tanagra

Vous pouvez importer de la même façon, mais cette fois-ci, prendre le fichier nouvellement créé contenant les classes.

Un arbre de décision est une méthode supervisée, elle a donc besoin des *input* et des *output* !

Quel algorithme pour construire l'arbre ? Que calcule-t-il pour mesurer sa qualité ?

Analyse des classes

Tanagra

Analyse qualitative de l'arbre (mettez une capture de l'arbre obtenu !) et de ses facteurs (variables).

Modalités

20 pages maximum

À remettre le 30 novembre 2023 avant 23:59

Vous pouvez partager un autre type de fichier (.txt ou .csv) si vous avez calculé vos indicateurs avec un autre outil que Excel.

Travail en équipe

AU TRAVAIL !