



UNIVERSITE DE MAROUA

Ecole Nationale Supérieure Polytechnique de Maroua  
(ENSPM)

Informatique et Télécommunication

M1 / S8 – Data Sciences & Master Informatique

Année académique : 2022-2023

Distribution : Mardi le 23 Mai 2023.

Par : M. Touza Isaac

### DATA MINING (DSC 438 – MIF 438)

#### FICHE DE TRAVAIL PERSONNEL ENCADRE - TPE

#### Consignes :

- Le travail se fera par groupe de quatre étudiants
- Le travail est à remettre au tard le 03 juin 2023 à 23h59. Délai de rigueur.
- Les fichiers numériques (script python, fichier csv, rapport TPE en PDF) doivent être compressé dans un fichier et nommé TPE\_DataMining\_Groupe\_NumeroGroupe (Exemple : TPE\_DataMining\_Groupe\_1)
- Le fichier ci-dessus doit être envoyé par mail à l'adresse [isaac\\_touza@outlook.fr](mailto:isaac_touza@outlook.fr) au plus tard à la date indiquée plus haut.
- Le code source (fichier .ipynb) doit être commenté

#### TRAVAIL A FAIRE :

Vous disposez d'un ensemble de données contenant des commentaires d'utilisateurs sur un site de commerce électronique. Chaque commentaire est étiqueté comme positif (+) ou négatif (-) en fonction du sentiment exprimé. Votre tâche est de construire un modèle de classification de texte pour prédire le sentiment des commentaires non étiquetés. Il vous est demandé d'utiliser l'ensemble de données d'entraînement pour construire votre modèle en répondant aux questions suivantes :

1. Télécharger les données présent à l'adresse suivante qui représentant les commentaires : [https://drive.google.com/file/d/1BsN1IX4i1AnUDE8iOsSb93VNeIqYln4c/view?usp=share\\_link](https://drive.google.com/file/d/1BsN1IX4i1AnUDE8iOsSb93VNeIqYln4c/view?usp=share_link)
2. Réaliser le prétraitement de ces données
3. Subdiviser l'ensembles des données en données d'entraînement (80%) et en données de test (20%)
4. Construire un modèle de classification en utilisant l'algorithme de votre choix (Naïve Bayes ou Knn).
5. Générer la matrice de confusion du modèle de classification construit
6. Imprimer les rapports de la classification puis évaluer la performance du modèle utilisé.
7. Utiliser votre modèle entraîné pour prédire le sentiment des commentaires non étiquetés suivants :
  - "Ce produit est incroyable !"
  - "Je suis déçu de mon achat."
  - "Le service client est exceptionnel."
  - " Vous êtes le meilleur