

Ecole National Polytechnique ENP Spécialité : DSIA Enseignant : Oussama ARKI	2°Année du 2°Cycle Module : Apprentissage Automatique(AA) Année 2022-2023
---	--

TP 02 : KNN (Segmentation des clients)

Objectif :

- La création d'un modèle KNN pour un problème de classification multi-class.
- La sélection des bons hyperparamètres (le k et la distance)
- l'évaluation du modèle (score, matrice de confusion, courbe ROC ...)

Contexte

Un constructeur automobile envisage de pénétrer de nouveaux marchés avec ses produits existants (P1, P2, P3, P4 et P5). Après une étude de marché intensive, ils ont déduit que le comportement du nouveau marché est similaire à leur marché existant.

Dans leur marché existant, l'équipe commerciale a classé tous les clients en 4 segments (A, B, C, D). Cette stratégie a fonctionné exceptionnellement bien pour eux. Ils prévoient d'utiliser la même stratégie pour les nouveaux marchés et ont identifié 2627 nouveaux clients potentiels.

Vous devez aider le gestionnaire à prédire le bon groupe de nouveaux clients.

Dictionnaire de données

Variable	Definition
ID	Unique ID
Gender	Gender of the customer
Ever_Married	Marital status of the customer
Age	Age of the customer
Graduated	Is the customer a graduate?
Profession	Profession of the customer
Work_Experience	Work Experience in years
Spending_Score	Spending score of the customer
Family_Size	Number of family members for the customer (including the customer)
Var_1	Anonymised Category for the customer
Segmentation	(target) Customer Segment of the customer

Travail demandé : construire un modèle KNN de classification multi classe, qui permet de prédire (classifier) l'étiquette d'un nouveau client

Etapes :

- 1- **Pre-preprocessing de données**
- 2- **Création du modèle KNN**
- 3- **Cross Validation** : utiliser la validation croisée pour trouver le meilleur paramètre k
- 4- **Curve Validation** : elle simplifie la validation croisée
- 5- **GridSearchCV** : elle permet de trouver plusieurs paramètres à la fois
- 6- **Evaluation** : Accuracy, matrice de confusion, ROC...