



**RECOMMANDATION  
DE VÉHICULES BASÉE SUR  
L'ANALYSE DES DONNÉES CLIENTS**

# **DATA MINING**

Réalisé par :  
**ANOADA NOHAYLA**

Présenté à :  
**EL MOUKHTAR ZEMMOURI**

# 1. Introduction

## 1.1 Contexte et Objectifs

Dans un contexte où les ventes de véhicules sont de plus en plus influencées par la personnalisation de l'offre aux besoins des clients, il devient essentiel d'utiliser des outils d'intelligence artificielle pour recommander des véhicules adaptés. Ce projet vise à mettre en place un système capable de recommander un véhicule à un client en fonction de ses caractéristiques personnelles (âge, sexe, capacité d'endettement, situation familiale...).

## 1.2 Problématique

Les recommandations de véhicules doivent prendre en compte plusieurs critères variables, et la difficulté réside dans la capacité à prédire précisément quel type de véhicule correspond à un utilisateur donné. L'objectif est de développer un modèle qui utilise ces critères pour classer un client dans une catégorie de véhicule spécifique.

## 1.3 Objectifs

- **Créer un modèle de machine learning** qui prédit la catégorie de véhicule la plus appropriée pour un client.
- **Développer une interface utilisateur interactive avec Streamlit**, permettant aux utilisateurs de saisir leurs informations et de recevoir des recommandations personnalisées de véhicules.

# 2. Méthodologie

## 2.1 Préparation des Données

### 2.1.1 Source des Données

Les données utilisées pour ce projet proviennent de deux ensembles principaux :

- **Données des clients** : Informations relatives aux clients, y compris des variables comme l'âge, le sexe, le taux d'endettement, la situation familiale, et le nombre d'enfants à charge.

- **Données des véhicules** : Information sur les véhicules disponibles, incluant la marque, le modèle, la puissance, la longueur, et la couleur.

### 2.1.2 Nettoyage des Données

- **Suppression des doublons** : Une première étape consiste à éliminer les doublons dans les ensembles de données.
- **Traitement des valeurs manquantes** : Les valeurs manquantes dans certaines colonnes ont été supprimées ou imputées avec des valeurs par défaut appropriées.
- **Encodage des variables catégoriques** : Les variables catégorielles, telles que le sexe, la situation familiale, et la présence d'une deuxième voiture, ont été encodées en valeurs numériques à l'aide du **LabelEncoder** de scikit-learn.
- **Standardisation des données** : Les données ont été normalisées avec un **StandardScaler** pour garantir que toutes les variables aient la même échelle et éviter que certaines dominent le modèle.

## 2.2 Modélisation

### 2.2.1 Choix des Modèles

Cinq modèles de classification ont été utilisés pour prédire la catégorie de véhicule :

- **Arbre de Décision** : Modèle simple et interprétable.
- **Forêt Aléatoire** : Utilise un ensemble d'arbres de décision pour améliorer la performance.
- **Support Vector Machine (SVM)** : Efficace dans les espaces de haute dimensionnalité.
- **Perceptron Multi-couche (MLP)** : Un modèle de réseau de neurones pour explorer des architectures plus complexes.
- **Deep Learning avec TensorFlow/Keras** : Un modèle de deep learning plus complexe pour observer les performances dans des situations non linéaires.

### 2.2.2 Entraînement et Évaluation

- **Entraînement** : Les modèles ont été entraînés sur les données d'entraînement (80 % du dataset) et évalués sur les données de test (20 % restantes).

- **Évaluation** : Les performances des modèles ont été mesurées à l'aide des critères suivants :
  - **Accuracy** : Précision globale du modèle.
  - **Rapport de classification** : Pour observer la précision, le rappel, et le score F1 de chaque modèle.
  - **Matrice de confusion** : Pour comprendre les erreurs de prédiction.

### 2.2.3 Sélection du Modèle

Parmi les modèles testés, le modèle de **forêt aléatoire** a montré la meilleure performance avec une accuracy de **0.51**.

## 2.3 Prédictions

Une fois le modèle choisi, nous avons utilisé des informations saisies par l'utilisateur pour effectuer des prédictions. Le modèle prend en entrée des données telles que : Âge, Sexe, Capacité d'endettement, Situation familiale, Nombre d'enfants à charge, Possession d'une deuxième voiture.

Ces informations sont standardisées et passées dans le modèle pour prédire la catégorie de véhicule la plus adaptée.

## 2.4 Interface Utilisateur avec Streamlit

- **Streamlit** a été utilisé pour créer une interface simple où les utilisateurs peuvent entrer leurs données et obtenir une recommandation instantanée.
- L'interface permet de saisir :
  - L'âge, le sexe, la capacité d'endettement, la situation familiale, et les autres informations pertinentes.
  - Après avoir saisi les informations, l'utilisateur reçoit la recommandation du véhicule adapté.

## 3. Résultats

### 3.1 Performance des Modèles

- **Arbre de décision** : Accuracy = 0.49
- **Forêt aléatoire** : Accuracy = 0.51
- **SVM** : Accuracy = 0.63
- **Réseau de neurones (MLP)** : Accuracy = 0.68

- **Deep Learning (TensorFlow/Keras)** : Accuracy = 0.68

Le modèle **Forêt Aléatoire** a montré la meilleure précision avec un score de **0.51**, ce qui en fait le modèle sélectionné pour la recommandation.

### 3.2 Analyse des Résultats

Le modèle de forêt aléatoire a bien performé, avec des résultats solides sur les différentes métriques de classification. En particulier, la capacité à classer correctement la catégorie de véhicule a été élevée, avec une précision globale de 0.51 %.

### 3.3 Prédiction pour un Client

L'exemple d'un client avec un âge de 30 ans, une capacité d'endettement de 20 000 €, et d'autres caractéristiques a permis d'obtenir une recommandation fiable et détaillée concernant le véhicule le plus adapté.

## 4. Conclusion

Ce projet a permis de développer un modèle de machine learning pour recommander des véhicules aux clients en fonction de leurs caractéristiques. La forêt aléatoire a montré les meilleures performances avec une précision de 0.51%. Bien que le modèle soit efficace, il pourrait être amélioré en ajoutant plus de variables et en explorant des modèles plus avancés comme les réseaux de neurones. Ce système peut être utilisé pour des plateformes de vente de véhicules et offre des perspectives d'amélioration pour une personnalisation accrue des recommandations...