

Mini-projet : Développement d'un modèle de prédiction des prix de voitures pour le marché marocain

(A réaliser par binôme)

1. Introduction :

La prédiction des prix des voitures est un problème courant dans le marché de l'automobile. Un modèle de Machine Learning bien conçu peut aider les acheteurs et vendeurs à estimer un prix juste en fonction de divers critères. Ce projet vise à développer un modèle prédictif basé sur des données réelles du marché marocain.

2. Objectifs

Les objectifs fixés pour ce projet sont les suivants :

- Collecter et préparer un jeu de données représentatif du marché automobile marocain.
- Explorer les données pour identifier les facteurs influençant le prix des véhicules.
- Expérimenter différentes méthodes de modélisation pour prédire le prix.
- Valider et améliorer la précision du modèle.
- Déployer une interface permettant une estimation facile du prix.

3. Déroulement du projet

Pour organiser efficacement ce projet, il est préférable d'avancer étape par étape. Ci-après l'ensemble des étapes à suivre de la collecte des données à la mise en production du modèle. Ce cheminement est donné à titre indicatif et peut être enrichi par l'ajout de tâches supplémentaires en fonction des besoins rencontrés.

- **Etape 1 : Collecte et Préparation des Données**

La première étape consiste à recueillir et structurer les données nécessaires à la modélisation. Les étudiants doivent rechercher des sources fiables et mettre en place un processus de scraping pour extraire les informations pertinentes. Une fois les données collectées, elles doivent être nettoyées et préparées pour garantir leur qualité et leur cohérence.

Tâches à réaliser :

- **Sources de données :** Rechercher et extraire les annonces marocaines. Ci-après quelques-uns : Avito.ma, Moteur.ma, Wandaloo Maroc.
- **Extraction :** Utiliser les techniques de web scraping à l'aide des bibliothèques BeautifulSoup, Requests, Selenium ou Scrapy.
- **Nettoyage :**
 - Supprimer les valeurs aberrantes et incohérentes.
 - Gérer les valeurs manquantes (remplissage ou suppression).

- **Encodage des variables** : Transformer les variables catégoriques (ex : marque, type de carburant, modèle) en valeurs numériques.
- **Normalisation** : Standardiser les variables continues (ex : kilométrage, puissance, année).
- **Etape 2 : Analyse Exploratoire des Données (EDA)**

L'analyse exploratoire permet de mieux comprendre la distribution des données et d'identifier les facteurs ayant un impact significatif sur les prix. Cette phase est essentielle pour détecter d'éventuelles erreurs et orienter la sélection des modèles de prédiction. Cette étape peut bien être réalisée en amont ou en aval de la préparation de données.

Tâches à réaliser :

- Étudier la distribution des prix et d'autres variables pour comprendre la répartition des valeurs.
- Identifier les corrélations entre les caractéristiques et le prix.
- Détecter et gérer les valeurs aberrantes.
- Générer des visualisations avec :
 - Histogrammes pour observer la distribution des prix.
 - Boxplots pour visualiser la répartition des prix par marque ou type de carburant.
 - Heatmaps pour identifier les corrélations entre les variables
- **Etape 3 : Étude des Tendances du Marché**

Cette étape vise à analyser en profondeur les tendances du marché automobile marocain. Il s'agit d'identifier les modèles de voitures les plus populaires, les écarts de prix entre différentes marques et de repérer les véhicules sous-évalués ou surestimés.

Tâches à réaliser :

- Analyser la distribution des prix pour détecter les éventuelles anomalies du marché.
- Identifier les facteurs ayant le plus d'impact sur le prix (âge du véhicule, marque, carburant, kilométrage, etc.).
- Comparer les tendances des marques et modèles les plus populaires afin d'identifier des modèles surévalués ou sous-évalués.
- Visualiser les tendances à travers des graphiques comparatifs et des analyses statistiques.
- **Etape 4 : Modélisation & Entraînement**

Une fois les données préparées, il est temps de tester différents algorithmes pour prédire les prix des véhicules. Les étudiants devront expérimenter plusieurs modèles et évaluer leur performance afin de sélectionner la meilleure approche.

Tâches à réaliser :

- Expérimenter plusieurs modèles de Machine Learning.
- Diviser les données en ensembles d'entraînement et de test.
- Entraîner et comparer les modèles selon leur performance sur le jeu de test.

- **Etape 5 : Validation et Optimisation**

L'étape de validation permet de s'assurer que le modèle généralisera bien sur de nouvelles données. Il s'agit d'évaluer la qualité des prédictions en utilisant différentes métriques et d'ajuster les paramètres pour améliorer la performance globale.

Tâches à réaliser :

- Évaluer la performance des modèles en utilisant différentes métriques telles que:
 - MAE (Mean Absolute Error) : Moyenne des erreurs absolues.
 - RMSE (Root Mean Squared Error) : Écart-type des erreurs de prédiction.
 - R^2 (coefficient de détermination) : Indicateur de qualité d'ajustement.
- Comparer les modèles et sélectionner le plus performant.
- Utiliser la validation croisée pour éviter l'overfitting.

- **Etape 6 : Déploiement & Interface Utilisateur**

Une fois un modèle performant obtenu, il doit être déployé pour permettre une utilisation pratique. L'objectif est de concevoir une interface simple et intuitive permettant aux utilisateurs d'entrer les caractéristiques d'un véhicule et d'obtenir une estimation de prix en temps réel.

Tâches à réaliser :

- Créer une API avec Flask ou FastAPI permettant d'envoyer des requêtes et de recevoir des prédictions.
- Concevoir une interface utilisateur avec Streamlit pour une utilisation interactive.
- Permettre aux utilisateurs d'entrer les caractéristiques d'un véhicule et d'obtenir une estimation de prix en temps réel.

4. Résultats et livrables attendus

Une fois toutes les précédentes tâches effectuées et les résultats de la prédiction obtenus, il est nécessaire de formaliser votre travail. La pertinence de votre approche, la performance du modèle ainsi que la clarté et la fonctionnalité du code et de l'interface utilisateur seront évalués à travers les livrables suivants :

- Un rapport d'analyse détaillant les tendances et les choix méthodologiques.
- Un code Python propre et bien documenté.
- Un modèle de Machine Learning opérationnel.
- Une interface interactive pour tester les prédictions en temps réel.