# Dédicaces

# Remerciement

# ***Table des matières***

[Dédicaces 2](#_Toc194753862)

[Remerciement 3](#_Toc194753863)

[Table des matières 4](#_Toc194753864)

[Liste des abréviations 5](#_Toc194753865)

[Liste des figures 6](#_Toc194753866)

[Introducion générale 7](#_Toc194753867)

[Chapitre 1 : Contexte du projet 9](#_Toc194753868)

[I. Introduction 10](#_Toc194753869)

[II. Marché automobile marocain 10](#_Toc194753870)

[III. Objectifs du modèle de prédiction. 11](#_Toc194753871)

[1. Prédiction précise des prix : 11](#_Toc194753872)

[2. Analyse des déterminants de prix : 11](#_Toc194753873)

[3. Détection d’anomalies : 11](#_Toc194753874)

[4. Outil d’aide à la décision : 11](#_Toc194753875)

[IV. Problématique et enjeux. 12](#_Toc194753876)

[Chapitre 2 : Collecte et Préparation des Données 14](#_Toc194753877)

[I. Sources de données 15](#_Toc194753878)

[II. Nettoyage des données 15](#_Toc194753879)

[III. Prétraitement des données … 15](#_Toc194753880)

# Liste des abréviations

**SOMACA** : Société Marocaine de Construction Automobile

# Liste des figures

# Introducion générale

Le marché automobile marocain, en plein essor, se caractérise par une demande croissante pour les véhicules neufs et d’occasion, notamment dans un contexte de diversification des offres et d’évolution des comportements d’achat. Toutefois, fixer un prix juste pour une voiture reste un défi complexe, tant pour les vendeurs que pour les acheteurs, en raison de la multiplicité des facteurs influençant les prix (âge du véhicule, kilométrage, marque, etc.) et des disparités observées sur le marché.

Face à cette problématique, ce projet propose de développer un **modèle prédictif basé sur le Machine Learning**, capable d’estimer avec précision le prix des véhicules commercialisés au Maroc. L’objectif principal est de fournir un outil fiable et accessible, s’appuyant sur des données réelles collectées auprès de plateformes locales (Avito.ma, Moteur.ma, etc.), pour guider les utilisateurs dans leurs transactions et contribuer à un marché plus transparent. Ce travail s’articule autour de plusieurs enjeux clés :

**Techniques** : Collecte et nettoyage de données hétérogènes, sélection d’algorithmes adaptés aux spécificités du marché marocain.

**Économiques** : Réduction des asymétries d’information et facilitation des échanges équitables.

**Pratiques** : Déploiement d’une interface utilisateur intuitive permettant une estimation instantanée.

Structuré en étapes claires (collecte des données, analyse exploratoire, modélisation, validation, déploiement), ce rapport détaille la méthodologie employée, les résultats obtenus et les limites rencontrées. Il met également en lumière l’innovation apportée par l’application de techniques avancées d’apprentissage automatique à un contexte local, offrant ainsi des perspectives concrètes pour l’optimisation des transactions automobiles au Maroc.

# Chapitre 1 : Contexte du projet

## Introduction

Le marché automobile marocain connaît une forte croissance, notamment dans le secteur des véhicules d’occasion. Cependant, évaluer correctement le prix d’un véhicule reste complexe à cause de nombreux facteurs. Ce chapitre présente le contexte économique et technique du projet, ainsi que les problématiques liées à la prédiction de prix.

## Marché automobile marocain

Le secteur automobile marocain, initié dans les années 1950 avec la création de la **SOMACA**, s’est transformé en un pilier stratégique de l’économie nationale. Aujourd’hui, le Maroc compte plus de 50 entreprises spécialisées dans la fabrication de composants automobiles (faisceaux électriques, pièces de rechange, etc.), attirant des investisseurs internationaux comme Renault, PSA, ou Lear Corporation. Ces entreprises alimentent principalement les chaînes d’approvisionnement européennes et africaines, avec 85% de la production destinée à l’exportation.

##### Caractéristiques clés du marché :

**Dynamisme industriel** : Le secteur génère 80 000 emplois directs et contribue à 10% des exportations nationales (source : AMDIE, 2023).

**Exigences logistiques strictes** : Les donneurs d’ordre internationaux imposent des livraisons juste à temps, avec des cycles de commande réduits à une semaine dans certains cas.

**Défis logistiques** : Le transport aérien d’urgence coûte 7 fois plus cher que le routier, tandis que les retards administratifs entraînent des stocks tampons de 35 jours.

**Compétitivité internationale** : Le Maroc accuse un retard face à la Tunisie ou la Pologne, avec des délais de transport de 4-6 jours vers l’Europe.

### La prédiction des prix des voitures d’occasion

Le marché des véhicules d’occasion au Maroc est en pleine expansion, notamment grâce à la digitalisation des plateformes de vente. Néanmoins, la détermination du juste prix d’un véhicule reste une tâche complexe en raison de nombreux facteurs influents : marque, modèle, année, kilométrage, type de carburant, état général, etc.

Cette complexité génère une forte hétérogénéité des prix pour des véhicules parfois très similaires, ce qui peut entraîner des déséquilibres dans les transactions. C’est dans ce contexte qu’intervient la nécessité de développer des modèles de prédiction automatisés et fiables pour estimer le prix réel d’un véhicule d’occasion.

### Problématique et défis de la prédiction

La prédiction du prix d’un véhicule d’occasion est un problème complexe en raison :

* Du grand nombre de variables impliquées, parfois redondantes ou corrélées.
* Des relations non linéaires entre certaines caractéristiques et le prix.
* De l’existence de valeurs aberrantes.
* De la variabilité du marché dans le temps.
* Du manque de standardisation des données collectées (descriptions non structurées, formats différents selon les sources).

### Techniques de prédiction utilisées : Regression

Pour répondre à cette problématique, le Machine Learning offre une panoplie d’approches adaptées à la prédiction de valeurs numériques. Parmi les plus utilisées dans ce contexte, on retrouve :

* **Régression linéaire multiple** : Modèle de base permettant de quantifier l’impact de chaque variable sur le prix.
* **Régression Ridge et Lasso** : Versions régularisées qui réduisent le sur-apprentissage et sélectionnent les variables les plus pertinentes.
* **Arbres de décision et forêts aléatoires (Random Forest)** : Méthodes puissantes capturant les interactions complexes entre variables.
* **Gradient Boosting Machines (XGBoost, LightGBM)** : Modèles très performants pour des données hétérogènes.
* **Réseaux de neurones (Deep Learning)** : Utilisés pour modéliser des relations très complexes, surtout avec de grandes quantités de données.

L’utilisation de ces techniques permet non seulement de prédire les prix, mais aussi d’analyser les facteurs déterminants, de détecter des anomalies, et de fournir un outil d’aide à la décision fiable et transparent.

## Objectifs du modèle de prédiction.

Ce modèle de prédiction des prix des véhicules d’occasion a pour objectif de répondre à plusieurs besoins complémentaires. Tout d’abord, il vise à fournir une estimation précise du prix de vente, en prenant en compte ses caractéristiques techniques et historiques. Cela permet de proposer une fourchette de prix réaliste, offrant ainsi une orientation tant pour les acheteurs que pour les vendeurs. Ensuite, le modèle facilite une analyse détaillée des déterminants de prix, en identifiant les variables les plus influentes et en mesurant l’impact de chaque facteur. Par exemple, il peut quantifier l’effet d’une augmentation du kilométrage ou de l’âge du véhicule sur sa valeur.

Parallèlement, le modèle joue un rôle crucial en détectant les anomalies, en repérant les annonces dont le prix est soit nettement inférieur, soit supérieur à la valeur estimée. Enfin, ce modèle constitue un véritable outil d’aide à la décision, permettant aux utilisateurs de simuler différents scénarios (comme une réduction du kilométrage) et de disposer d’une base objective pour faciliter les négociations.

## Problématique et enjeux.

### Problématique

Le marché marocain des véhicules d’occasion est caractérisé par plusieurs défis majeurs. Tout d’abord, une **asymétrie d’information** persiste entre acheteurs et vendeurs, ces derniers ne disposant pas toujours des mêmes données, ce qui engendre des déséquilibres dans les transactions. Ensuite, on observe une **forte hétérogénéité des prix**, avec des écarts significatifs entre des véhicules pourtant similaires, principalement à cause de l’absence de standardisation dans l’évaluation. Enfin, le marché souffre de **pratiques opaques**, où certains revendeurs profitent du manque de transparence pour surévaluer leurs véhicules, tandis que des annonces frauduleuses, notamment concernant le kilométrage, induisent les acheteurs en erreur.

### Enjeux

#### Économiques :

Pour les acheteurs, l’objectif est d’éviter les surcoûts liés à des prix injustifiés, en ayant accès à une estimation fiable du prix d’un véhicule. Pour les vendeurs, il s’agit d’optimiser leurs marges sans risquer de décourager les acheteurs, en trouvant un équilibre entre compétitivité et rentabilité. En ce qui concerne les plateformes, l’enjeu est d’augmenter la confiance des utilisateurs, ce qui peut entraîner une plus grande fidélité des clients et attirer davantage de trafic sur leurs sites.

#### Sociaux :

L’un des enjeux sociaux majeurs est de réduire les risques de conflits lors des négociations en fournissant une estimation neutre du prix, ce qui permet de faciliter les échanges entre acheteurs et vendeurs. De plus, rendre le marché plus accessible aux primo-accédants, souvent désavantagés par leur manque d’expertise technique dans l’évaluation des prix. Cela permet de niveler les inégalités et de favoriser une participation plus large à ce marché.

#### Techniques :

L’un des enjeux techniques majeurs réside dans la gestion de la complexité des données, qui incluent des variables hétérogènes telles que des données numériques et catégorielles, ainsi que des corrélations non linéaires. De plus, il est essentiel d’assurer la scalabilité du modèle afin qu’il puisse s’adapter à l’évolution constante du marché, en prenant en compte de nouveaux modèles de véhicules ainsi que les effets de l’inflation.

#### Éthiques :

Les enjeux éthiques du modèle incluent la nécessité de garantir sa neutralité, en évitant tout biais, comme la survalorisation de certaines marques. Il est également crucial de protéger la vie privée des utilisateurs, en veillant à anonymiser les données utilisées lors de l’entraînement du modèle afin de respecter les principes de confidentialité et de sécurité des informations personnelles.

## Conclusion

La prédiction des prix de voitures d’occasion est un défi justifié par la diversité des données et l’absence de références claires. Les techniques de régression du Machine Learning offrent des solutions adaptées, ouvrant la voie à une modélisation efficace que les chapitres suivants vont explorer.

# Chapitre 2 : Collecte et Préparation des Données

## Sources de données

## Nettoyage des données

## Prétraitement des données …