



ECOLE MAROCAINE DES  
SCIENCES DE L'INGENIEUR  
Membre de  
HONORIS UNITED UNIVERSITIES

---

# HomeHunt

## Prédiction des prix immobiliers

---

ABBAS Hamza  
BOULAICH Ismail  
CHAHID Anass  
ECHBERAAL Ahmed Redouane

*Encadrant:*

M. BAZZA Houssam

*Membres du jury:*

M. BAZZA Houssam  
M. TBATOU Zakariae  
Mme. BENADDI Hafsa

2 ème année Ingénierie Informatique et Réseaux (4IIR)

2024 - 2025



## Résumé

Ce projet a pour objectif la création d'une plateforme web intelligente permettant d'estimer le prix d'un bien immobilier à partir de ses caractéristiques. Pour cela, une chaîne complète de traitement de données a été mise en place, incluant le scraping d'annonces immobilières depuis plusieurs sites francophones et marocains, le nettoyage, l'intégration et l'analyse de ces données, puis l'entraînement de modèles de machine learning pour la prédition des prix. Le système repose sur une architecture technique moderne combinant une interface Angular, un backend Node.js, un microservice Python (FastAPI) et une base de données MySQL/MongoDB. L'ensemble est orchestré avec Airflow et déployé via Docker. Un tableau de bord Power BI permet enfin de visualiser les tendances du marché en temps réel. Ce projet illustre l'interconnexion entre collecte de données, intelligence artificielle, visualisation et génie logiciel, appliqués au secteur immobilier.

---

## Abstract

This project aims to develop an intelligent web platform that estimates the price of a real estate property based on its characteristics. A complete data processing pipeline has been implemented, including the scraping of property listings from multiple French and Moroccan websites, data cleaning and integration, and training of machine learning models to predict prices. The system is based on a modern architecture combining an Angular frontend, a Node.js backend, a Python microservice (FastAPI), and MySQL/MongoDB databases. The entire platform is orchestrated with Airflow and deployed using Docker. A Power BI dashboard allows real-time visualization of market trends. This project illustrates the integration of data collection, artificial intelligence, visualization, and software engineering applied to the real estate sector.



# **Remerciements**

Nous tenons à exprimer nos sincères remerciements à toutes les personnes qui ont contribué, de près ou de loin, à la réussite de ce projet de fin d'études.

Nous adressons tout particulièrement notre gratitude à notre encadrant Monsieur Houssam Bazza, pour son accompagnement constant, ses conseils avisés et sa disponibilité tout au long du projet. Son encadrement technique et pédagogique a été essentiel dans la concrétisation de cette plateforme.

Nous remercions également Madame Hafsa Benaddi et Monsieur Zakariae Tbatou, membres du jury, pour leur intérêt porté à notre travail et pour le temps qu'ils consacrent à l'évaluation de ce projet. Leurs observations et retours d'expérience seront pour nous une source précieuse d'amélioration.

# Table des matières

<b>Introduction Générale</b>	<b>7</b>
<b>1 Introduction et vue globale</b>	<b>9</b>
<b>2 État de l'art</b>	<b>11</b>
1 Études similaires . . . . .	11
2 Problématique . . . . .	12
2.1 Étude préalable de la problématique . . . . .	12
2.2 Les résultats de l'étude . . . . .	13
3 Méthodes existantes . . . . .	16
<b>3 Analyse des besoins et conception</b>	<b>18</b>
1 Identification des besoins . . . . .	18
2 Description des besoins fonctionnels . . . . .	18
3 Besoins non fonctionnels . . . . .	19
4 Conception . . . . .	20
4.1 Diagramme de cas d'utilisation . . . . .	20
4.2 Diagramme de séquence . . . . .	22
<b>4 Réalisation et résultat</b>	<b>27</b>
1 La répartition des tâches 'Diagramme de GANTT' . . . . .	27
2 Concepts clés . . . . .	29
3 Architecture du système . . . . .	29
4 Mise en œuvre . . . . .	31
4.1 Couche de Présentation (Front-End) . . . . .	31
4.2 Couche Métier (Business Logic) . . . . .	33
4.3 Couche de Données (Data Access Layer) . . . . .	34
4.4 Scraping et traitement des données . . . . .	35
4.5 Entraînement et déploiement des modèles de Machine Learning	36
4.6 Visualisation des résultats via Power BI . . . . .	37
4.7 Conteneurisation et déploiement avec Docker . . . . .	38
5 Tests et validation . . . . .	38
5.1 Outil utilisé pour tester les performances . . . . .	38

5.2	Les tests . . . . .	39
6	Visualisation des marchés . . . . .	41
6.1	Le marché Marocain . . . . .	41
6.2	Le marché Français . . . . .	44
7	Illustration de l'interface graphique du site . . . . .	47
7.1	L'interface registration/authentification . . . . .	47
7.2	L'interface d'utilisateur . . . . .	48
7.3	L'interface d'administrateur . . . . .	52
8	Problèmes rencontrés et solutions apportées . . . . .	56
9	Limitations et améliorations . . . . .	58
<b>Conclusion Générale</b>		<b>61</b>
1	Récapitulation . . . . .	61
2	Points clés . . . . .	61
3	Retour d'expérience . . . . .	62
4	Conclusion . . . . .	62

# Table des figures

2.1	Résultat de la première enquête . . . . .	13
2.2	Résultat de la deuxième enquête . . . . .	13
2.3	Résultat de la troisième enquête . . . . .	14
2.4	Résultat de la quatrième enquête . . . . .	14
2.5	Résultat de la cinquième enquête . . . . .	15
2.6	Résultat de la sixième enquête . . . . .	15
2.7	Résultat de la septième enquête . . . . .	16
3.1	Diagramme de cas d'utilisation — Fonctionnalités Client de la plate-forme HomeHunt . . . . .	20
3.2	Diagramme de cas d'utilisation — Fonctionnalités Administrateur de la plateforme HomeHunt . . . . .	21
3.3	Diagramme de séquence — Inscription d'un utilisateur sur HomeHunt	22
3.4	Diagramme de séquence — Navigation dans les annonces immobilières	23
3.5	Diagramme de séquence — Prédiction de prix immobilier . . . . .	24
3.6	Diagramme de séquence — Validation d'un bien par l'administrateur	25
3.7	Diagramme de séquence — Inscription d'un utilisateur . . . . .	25
4.1	Le diagramme de Gantt . . . . .	28
4.2	Cycle de vie des données . . . . .	31
4.3	Logo du Framework Angular . . . . .	32
4.4	Logo de TypeScript . . . . .	32
4.5	Logo de Bootstrap . . . . .	33
4.6	Logo de Node.js . . . . .	33
4.7	Logo d'Axios . . . . .	34
4.8	Logo de MySQL . . . . .	34
4.9	Logo de HeidiSQL . . . . .	35
4.10	Logo de MongoDB . . . . .	35
4.11	Logo de l'outil Lighthouse . . . . .	39
4.12	Les tests de performance, accessibilité, bonnes pratiques et SEO . . . . .	39
4.13	1 <sup>ère</sup> visualisation du marché Marocain . . . . .	41
4.14	2 <sup>ème</sup> visualisation du marché Marocain . . . . .	42
4.15	3 <sup>ème</sup> visualisation du marché Marocain . . . . .	43

4.16 4 <sup>ème</sup> visualisation du marché Marocain . . . . .	44
4.17 1 <sup>ère</sup> visualisation du marché Français . . . . .	44
4.18 3 <sup>ème</sup> visualisation du marché Français . . . . .	45
4.19 3 <sup>ème</sup> visualisation du marché Français . . . . .	46
4.20 La page de registration d'un utilisateur . . . . .	47
4.21 La page d'authentification d'un utilisateur . . . . .	47
4.22 La page d'accueil du site . . . . .	48
4.23 Liste des articles du marché Marocain . . . . .	49
4.24 Liste des articles du marché Français . . . . .	49
4.25 Liste des immobiliers d'un utilisateur . . . . .	49
4.26 Liste des articles favoris d'un utilisateur . . . . .	50
4.27 Les informations d'un profil utilisateur . . . . .	50
4.28 Formulaire d'ajout d'un bien à prédire . . . . .	51
4.29 Résultat de la prédition du prix d'un bien saisi . . . . .	51
4.30 Tableau de bord d'administration du site . . . . .	52
4.31 Liste des différents utilisateurs du site . . . . .	52
4.32 Ajout d'un compte utilisateur . . . . .	52
4.33 Modification des propriétés d'un utilisateur . . . . .	53
4.34 Suppression d'un compte utilisateur . . . . .	53
4.35 Liste des différents immobiliers publié dans le site . . . . .	53
4.36 Ajout d'un immobilier avec ses propriétés . . . . .	54
4.37 Modification des propriétés d'un immobilier . . . . .	55
4.38 Suppression d'un immobilier . . . . .	55
4.39 Affichage des différentes propriétés d'un immobilier . . . . .	56

# Introduction Générale

Le secteur immobilier représente un pilier stratégique de l'économie, tant au niveau national qu'international. Il est marqué par une forte variabilité des prix selon la localisation, le type de bien, les caractéristiques du logement ou encore la conjoncture économique [1]. Face à cette complexité, l'évaluation juste et objective du prix d'un bien immobilier constitue un véritable enjeu pour les particuliers, les investisseurs et les professionnels du secteur.

Dans ce contexte, les récentes avancées technologiques en matière de science des données, d'intelligence artificielle et de visualisation interactive offrent de nouvelles perspectives. Grâce à des techniques telles que le web scraping, l'analyse statistique, les modèles de machine learning et la visualisation dynamique, il est désormais possible de capter et d'exploiter à grande échelle les informations publiées en ligne pour prédire, analyser et comprendre les tendances du marché immobilier.

Le projet que nous présentons dans ce rapport s'inscrit dans cette dynamique. Il a pour objectif principal la conception et le développement d'une plateforme web complète capable de prédire les prix immobiliers à partir des caractéristiques saisies par un utilisateur, en se basant sur des données réelles collectées automatiquement depuis plusieurs sites spécialisés, tant au Maroc qu'en France.

Pour y parvenir, le projet s'est articulé autour de plusieurs volets techniques interdépendants :

- La collecte automatisée de données immobilières issues de sites tels que Le-BonCoin [2], Le Figaro [3], Avito [4] ou Mubawab [5], à l'aide de bibliothèques telles que `Playwright`, `Selenium` et `BeautifulSoup`<sup>4</sup>.
- Le nettoyage, la transformation et la consolidation de ces données dans une base relationnelle (MySQL) via des scripts Python, orchestrés quotidiennement grâce à `Apache Airflow`.
- La conception de modèle de machine learning performant (`Random Forest`), entraîné sur les données traitées pour fournir des estimations de prix fiables.
- Le développement d'une plateforme web intuitive basée sur la stack **MEAN** (MongoDB, Express.js, Angular, Node.js), permettant aux utilisateurs d'interagir avec le système, de saisir les caractéristiques de leur bien et d'obtenir une prédiction instantanée.
- La création de tableaux de bord interactifs via `Power BI`, à destination des

administrateurs de la plateforme, afin d'assurer un suivi en temps réel du marché et des performances du système.

Ce rapport présente l'ensemble des étapes de réalisation du projet, de l'étude préliminaire à la mise en production, en passant par l'analyse des besoins, la conception, l'implémentation, les difficultés rencontrées et les solutions apportées. Il met également en lumière les choix techniques opérés, les résultats obtenus ainsi que les perspectives d'évolution du système.

Ce travail s'inscrit dans le cadre d'un projet de fin d'années mobilisant quatre étudiants, chacun ayant contribué à des volets techniques spécifiques, complémentaires et indispensables à la réussite du projet.

# Chapitre 1

## Introduction et vue globale

- **Contexte générale :**

L'évaluation des prix de l'immobilier est un processus complexe influencé par une multitude de facteurs tels que la localisation, la superficie, l'ancienneté du bien, ainsi que la dynamique du marché. Dans un contexte où les annonces immobilières en ligne se multiplient, la capacité à exploiter efficacement ces données représente un levier stratégique pour les particuliers et les professionnels. Le présent projet vise à concevoir une plateforme web intelligente de prédiction des prix immobiliers, s'appuyant sur des données scrappées automatiquement depuis plusieurs sites spécialisés, tant en France qu'au Maroc. Ces données, une fois nettoyées et structurées, sont utilisées pour entraîner des modèles de machine learning capables de fournir une estimation fiable du prix d'un bien en fonction de ses caractéristiques.

- **Méthodologie adoptée :**

Le projet a été conduit selon une approche agile basée sur la méthode "Scrum", permettant un travail itératif, collaboratif et évolutif. Le développement s'est déroulé en sprints de deux semaines, chacun étant précédé d'une planification des tâches à réaliser et suivi d'une rétrospective permettant d'ajuster l'organisation. Des réunions quotidiennes ont permis de maintenir une communication continue au sein de l'équipe. Cette démarche a favorisé une répartition claire des responsabilités : extraction des données, modélisation, développement web et visualisation analytique. Elle a également permis une intégration progressive des différents modules dans un système cohérent et fonctionnel.

- **Structure du rapport :**

- **Chapitre 2** : Présente un état de l'art détaillé des études similaires, des méthodes existantes et des concepts clés mobilisés.
- **Chapitre 3** : Décrit l'analyse des besoins fonctionnels et non fonctionnels, la conception et l'architecture du système proposé.
- **Chapitre 4** : Expose la réalisation technique du projet, les problèmes rencontrés et les solutions adoptées.
- **Conclusion Générale** : Résume les principaux apports du projet, le retour d'expérience et les perspectives d'amélioration.

# Chapitre 2

## État de l'art

### 1 Études similaires

Dans le domaine de l’immobilier, plusieurs travaux ont été menés afin de mieux comprendre les dynamiques du marché et d’automatiser l’estimation du prix des biens immobiliers. Ces initiatives combinent généralement l’analyse de données issues d’annonces immobilières avec des méthodes statistiques ou d’apprentissage automatique. Nous présentons ici un panorama des études similaires les plus pertinentes, à la fois dans la littérature scientifique et dans le monde professionnel.

— **Plateformes commerciales** : Des entreprises telles que Zillow (États-Unis) ou MeilleursAgents (France) ont développé des modèles propriétaires de prédiction de prix de l’immobilier. Zillow, par exemple, propose le "Zestimate", un algorithme basé sur des milliers de variables (emplacement, historique des ventes, caractéristiques du bien, etc.) et entraîné sur des millions d’annonces. Ces modèles sont très performants mais restent des boîtes noires, souvent indisponibles au grand public en tant qu’outil libre ou open source.

— **Travaux académiques** : L’article de Yu et al. (2017) [6] intitulé “*E-commerce logistics in supply chain management : Implementations and future perspective in furniture industry*” met en lumière les bénéfices d’une collecte automatisée de données couplée à une analyse prédictive dans un contexte d’e-commerce, qui peut être transposé à l’immobilier. D’autres études comme celle de Goy et al. (2007) [7], portant sur la personnalisation dans les applications de commerce électronique, insistent sur l’importance d’un système adaptatif capable de proposer une estimation en fonction du profil utilisateur.

— **Projets open-source et universitaires** : De nombreux projets académiques ont tenté de prédire le prix de l’immobilier à partir de jeux de données ouverts (ex. : Kaggle — House Prices : Advanced Regression Techniques [8]). Ces projets utilisent généralement des modèles comme Linear Regression, Random Forest, ou XGBoost, et impliquent une phase de préparation des données incluant le nettoyage, l’enco-

dage des variables catégorielles et la détection des valeurs aberrantes.

## 2 Problématique

### 2.1 Étude préalable de la problématique

Pour mieux comprendre les attitudes et les pratiques concernant le marché d'immobilier, Nous avons mené une étude auprès de 100 participants via un formulaire Google. Les questions visaient à évaluer :

- **1<sup>ère</sup> enquête - Figure 2.1 :** Expérience de difficultés lors d'une transaction immobilière (achat, location ou vente).
- **2<sup>ème</sup> enquête - Figure 2.2 :** Principaux obstacles rencontrés dans le processus immobilier.
- **3<sup>ème</sup> enquête - Figure 2.3 :** Critère jugé le plus important lors de la recherche d'un bien.
- **4<sup>ème</sup> enquête - Figure 2.4 :** Usage d'outils en ligne pour estimer la valeur d'un bien.
- **5<sup>ème</sup> enquête - Figure 2.5 :** Intérêt pour une plateforme fournissant une estimation précise des prix.
- **6<sup>ème</sup> enquête - Figure 2.6 :** Canaux principaux utilisés pour rechercher un bien.
- **7<sup>ème</sup> enquête - Figure 2.7 :** Fonctionnalités souhaitées dans une plateforme de prédiction des prix.

## 2.2 Les résultats de l'étude

1. Avez-vous déjà rencontré des difficultés lors de l'achat, de la location ou de la vente d'un bien immobilier ?

100 réponses

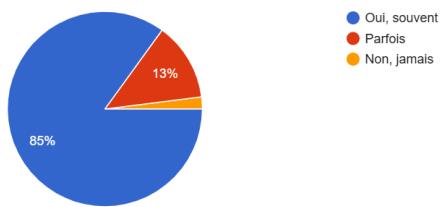


FIGURE 2.1 – Résultat de la première enquête

- **1<sup>ère</sup> enquête :** Une très large majorité des répondants (85 %) affirme rencontrer souvent des difficultés, tandis que 13 % déclarent y être confrontés parfois. Seuls 2 % n'ont jamais connu de problèmes. Autrement dit, près de 98 % du panel confirme qu'une transaction immobilière s'accompagne régulièrement d'obstacles, soulignant l'importance de solutions qui simplifient et sécurisent le parcours utilisateur.

2. Quels sont les principaux obstacles que vous rencontrez ?

100 réponses

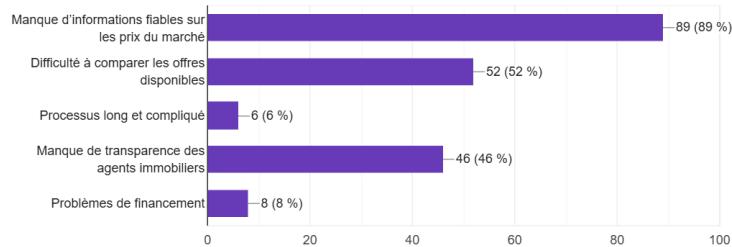


FIGURE 2.2 – Résultat de la deuxième enquête

- **2<sup>ème</sup> enquête :** Le frein principal cité est le manque d'informations fiables sur les prix du marché (89 % des réponses). Viennent ensuite la difficulté à comparer les offres (52 %) et le manque de transparence des agents immobiliers (46 %). Les aspects « processus long et compliqué » (6 %) ou « problèmes de financement » (8 %) sont nettement moins prégnants. Ces données confirment que la douleur dominante porte sur l'accès à une information de prix claire et vérifiable.

**3. Lorsque vous cherchez un bien, quel critère est le plus important pour vous ?**

100 réponses

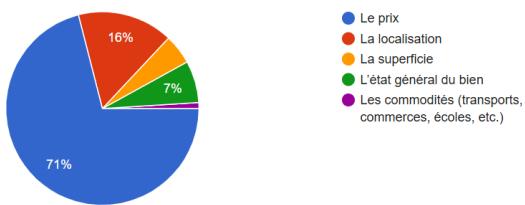


FIGURE 2.3 – Résultat de la troisième enquête

- 3<sup>ème</sup> enquête :** Le prix ressort comme critère décisif pour 71 % des répondants, loin devant la localisation (16 %). L'état général du bien (7 %), la superficie (5 %) et les commodités (1 %) pèsent beaucoup moins. Cette hiérarchie confirme que la valeur financière reste la variable première dans la prise de décision, avant même l'emplacement ou les caractéristiques physiques du bien.

**4. Utilisez-vous des outils en ligne pour estimer la valeur d'un bien immobilier ?**

99 réponses

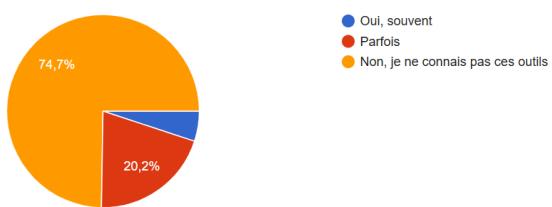


FIGURE 2.4 – Résultat de la quatrième enquête

- 4<sup>ème</sup> enquête :** Plus des trois quarts des personnes interrogées (près de 75 %) déclarent ne pas connaître ou ne pas utiliser ces outils. À l'inverse, seuls 5 % y ont recours régulièrement et 20 % occasionnellement. Le marché manque donc de solutions visibles et pédagogiques ; cela constitue une opportunité de proposer un estimateur simple et fiable.

**5. Seriez-vous intéressé par une plateforme qui vous donne une estimation précise des prix immobiliers en fonction du marché actuel ?**

100 réponses

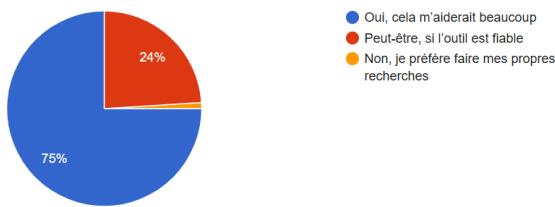


FIGURE 2.5 – Résultat de la cinquième enquête

- **5<sup>ème</sup> enquête :** L'intérêt est très fort : 75 % répondent « Oui, cela m'aiderait beaucoup », et 24 % se disent prêts à l'utiliser si l'outil est fiable. Seul 1 % préfère faire ses propres recherches. En clair, 99 % du panel est ouvert à une solution d'estimation à condition qu'elle inspire confiance.

**6. Quels moyens utilisez-vous principalement pour rechercher un bien immobilier ?**

100 réponses

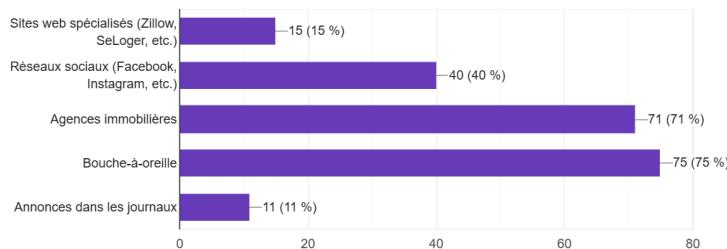


FIGURE 2.6 – Résultat de la sixième enquête

- **6<sup>ème</sup> enquête :** Les canaux traditionnels restent en tête : le bouche-à-oreille est mentionné par 75 % et les agences immobilières par 71 %. Les réseaux sociaux suivent (40 %), surpassant nettement les sites web spécialisés (15 %) et les annonces presse (11 %). La recherche immobilière demeure donc fortement relationnelle, mais les plateformes sociales prennent de l'importance, indiquant un potentiel de diffusion numérique encore peu exploité.

**7. Selon vous, quelles fonctionnalités devraient être incluses dans une plateforme de prédiction des prix immobiliers ?**

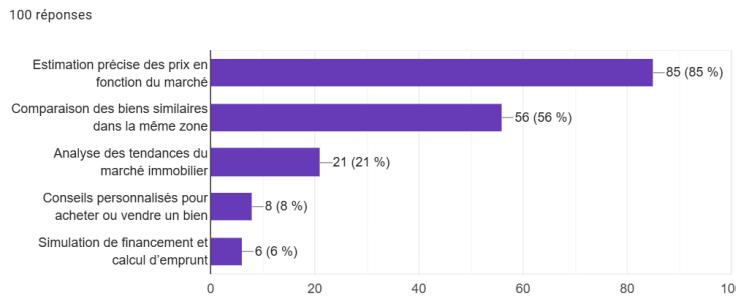


FIGURE 2.7 – Résultat de la septième enquête

- **7ème enquête :** L'attente numéro 1 concerne une estimation précise et en temps réel des prix (85 % des votes). La comparaison de biens similaires arrive ensuite (56 %), suivie par l'analyse des tendances du marché (21 %). Les conseils personnalisés (8 %) et la simulation de financement (6 %) apparaissent comme des compléments appréciés mais non prioritaires. Le cœur de valeur attendu est donc la fiabilité du calcul de prix, enrichie d'outils de benchmark et de veille marché.

### 3 Méthodes existantes

Le domaine de la prédiction des prix immobiliers a connu un développement important grâce à l'émergence de techniques avancées de traitement des données et de modélisation prédictive [9]. Les méthodes existantes peuvent être classées selon trois grandes catégories :

— **Méthodes statistiques traditionnelles** : Ces méthodes s'appuient sur des modèles paramétriques tels que la régression linéaire multiple. Bien que simples à mettre en œuvre et à interpréter, elles supposent des relations linéaires entre les variables explicatives (surface, nombre de pièces, quartier...) et la variable cible (prix). Elles peuvent être limitées dans leur capacité à modéliser des relations non linéaires et à prendre en compte des interactions complexes entre les facteurs.

— **Méthodes d'apprentissage supervisé** : Les algorithmes d'apprentissage automatique tels que les *Random Forests*, *XGBoost*, ou les *Gradient Boosting Machines* sont largement utilisés pour la prédiction de valeurs continues comme les prix. Ces méthodes permettent de modéliser des relations non linéaires et d'interagir avec des variables catégorielles, tout en fournissant des outils d'évaluation comme l'importance des variables. Elles nécessitent toutefois un traitement préalable des données rigoureux (nettoyage, normalisation, encodage).

— **Méthodes basées sur l'intelligence artificielle avancée** : Dans certains

cas, des approches plus récentes comme les réseaux de neurones profonds (Deep Learning) sont explorées, notamment pour intégrer des images ou des descriptions textuelles dans les prédictions. Cependant, leur utilisation reste limitée en raison de leur complexité, de leur besoin en données massives et de leur faible explicabilité.

Dans le cadre de notre projet, nous avons opté pour des modèles d'apprentissage supervisé (Random Forest), offrant un bon compromis entre performance, rapidité d'entraînement et interprétabilité.

# Chapitre 3

# Analyse des besoins et conception

## 1 Identification des besoins

Dans le cadre de ce projet, les besoins ont été identifiés à partir des exigences de deux types d'utilisateurs : les particuliers (utilisateurs finaux) souhaitant estimer la valeur d'un bien immobilier, et les administrateurs de la plateforme souhaitant suivre l'évolution du marché. L'analyse du besoin a également pris en compte les contraintes techniques liées à l'automatisation du processus de collecte, au traitement de données hétérogènes et à la mise à disposition de modèles de prédiction performants et accessibles.

L'objectif global est donc de concevoir une plateforme capable de :

- Offrir un service d'estimation du prix immobilier basé sur les caractéristiques fournies par l'utilisateur.
- Collecter, traiter et stocker automatiquement des données immobilières issues de sites spécialisés.
- Exploiter ces données à travers des algorithmes de machine learning fiables.
- Proposer un tableau de bord d'analyse destiné aux administrateurs.
- Garantir une bonne expérience utilisateur à travers une interface ergonomique et réactive.

## 2 Description des besoins fonctionnels

Les besoins fonctionnels correspondent aux principales fonctionnalités attendues par les utilisateurs de la plateforme. Ils sont divisés selon les rôles :

- **Utilisateur normal :**
  - Accéder à l'interface de prédiction.
  - Saisir les caractéristiques de son bien (ville, surface, nombre de pièces, type, ascenseur...).

- Recevoir une estimation du prix et une fourchette de confiance.
- Consulter l'historique de ses requêtes (optionnel).
- **Administrateur :**
  - Gérer les utilisateurs et leurs droits d'accès.
  - Accéder à un tableau de bord Power BI interactif.
  - Suivre l'évolution du marché par période, région ou type de bien.
  - Déclencher manuellement une mise à jour du scraping ou du modèle.
- **Système technique :**
  - Lancer automatiquement des processus de scraping (via Airflow).
  - Nettoyer, consolider et stocker les données dans une base relationnelle.
  - Entraîner et déployer un modèle de prédition.
  - Exposer une API REST pour la communication frontend/backend.

### 3 Besoins non fonctionnels

Les besoins non fonctionnels déterminent les critères de qualité attendus du système, en matière de performance, sécurité, accessibilité et maintenabilité :

- **Performance** : le temps de réponse de l'API de prédition doit être inférieur à 1 seconde.
- **Scalabilité** : le système doit pouvoir gérer un volume croissant d'annonces et de requêtes utilisateurs.
- **Sécurité** : les données personnelles doivent être protégées via une authentification sécurisée (JWT, Bcrypt).
- **Disponibilité** : le service doit être opérationnel de manière continue, avec une gestion des erreurs de scraping automatique.
- **Accessibilité** : l'application doit être responsive et accessible depuis un navigateur web moderne.

## 4 Conception

### 4.1 Diagramme de cas d'utilisation

Fonctionnalités Client de la plateforme HomeHunt :

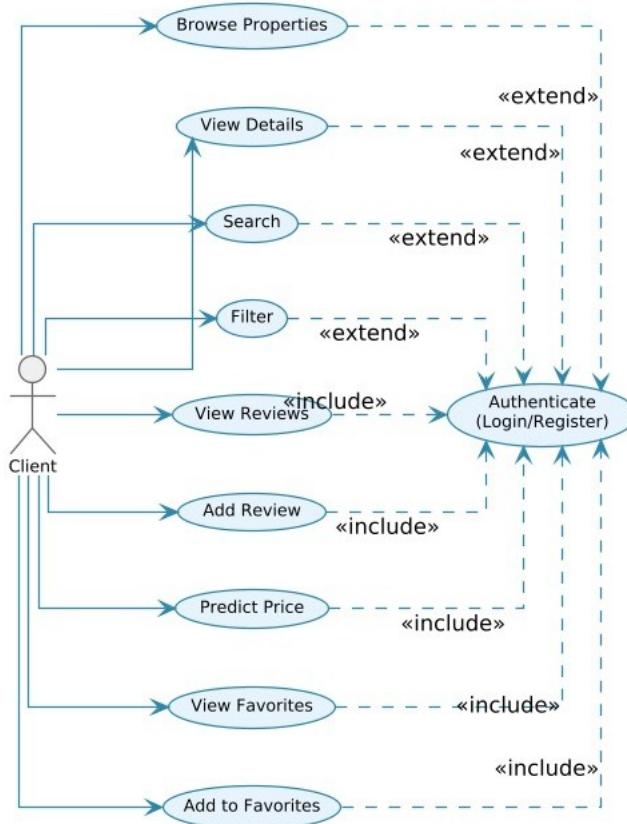


FIGURE 3.1 – Diagramme de cas d'utilisation — Fonctionnalités Client de la plateforme HomeHunt

Ce diagramme de cas d'utilisation illustre les interactions possibles entre un utilisateur client et la plateforme HomeHunt. Le client peut parcourir les biens immobiliers (Browse Properties), consulter les détails (View Details), rechercher (Search) ou filtrer les annonces (Filter). Certaines fonctionnalités comme l'ajout d'un bien aux favoris (Add to Favorites), la prédiction de prix (Predict Price), la lecture ou l'ajout de commentaires (View/Add Reviews), nécessitent une authentification préalable (Login/Register). Le diagramme met en évidence les relations « include » et

« extend » entre les cas d'utilisation, illustrant la dépendance ou l'extension de certaines actions en fonction du contexte. Ce diagramme reflète l'expérience utilisateur typique sur la plateforme du point de vue fonctionnel.

### Fonctionnalités Administrateur de la plateforme HomeHunt :

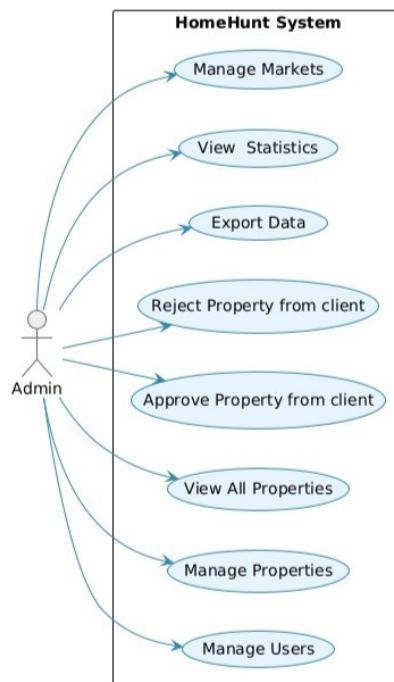


FIGURE 3.2 – Diagramme de cas d'utilisation — Fonctionnalités Administrateur de la plateforme HomeHunt

Ce diagramme présente les cas d'utilisation associés à l'acteur Administrateur dans le système HomeHunt. L'administrateur a la responsabilité de gérer les marchés (Manage Markets), visualiser les statistiques (View Statistics), et exporter des données (Export Data) pour une analyse approfondie. Il supervise également la modération des contenus proposés par les utilisateurs, avec la possibilité d'approuver ou de rejeter les biens immobiliers soumis (Approve/Reject Property from Client). Par ailleurs, l'administrateur peut consulter l'ensemble des propriétés disponibles sur la plateforme (View All Properties), gérer ces biens (Manage Properties), ainsi qu'administrer les comptes des utilisateurs (Manage Users). Ce diagramme reflète ainsi les fonctions de contrôle, de validation et de gouvernance assurées par le rôle administratif au sein de l'application.

## 4.2 Diagramme de séquence

Inscription d'un utilisateur sur HomeHunt :

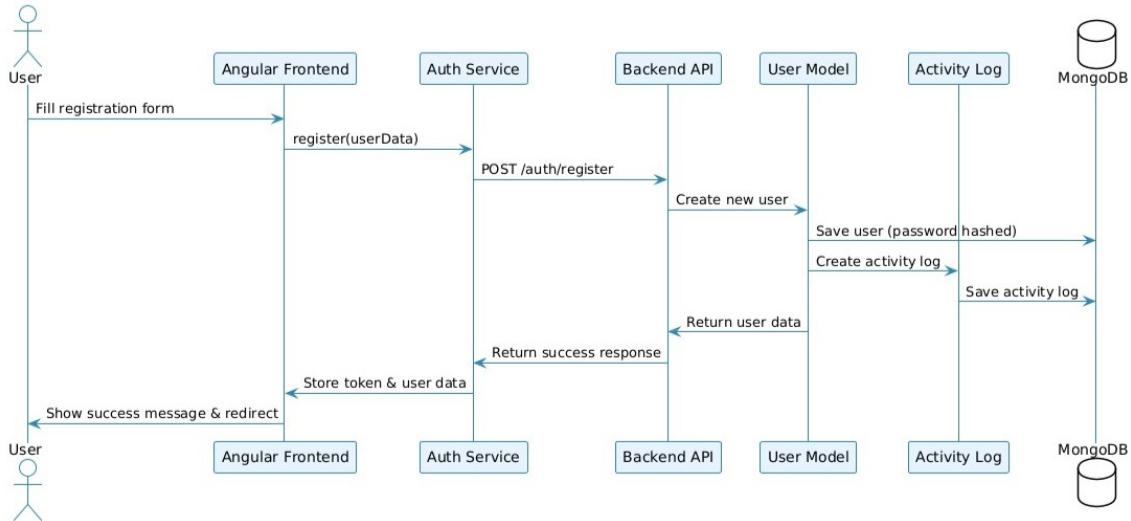


FIGURE 3.3 – Diagramme de séquence — Inscription d'un utilisateur sur HomeHunt

Ce diagramme illustre les interactions entre les composants lors de l'inscription d'un utilisateur. Le frontend Angular envoie les données au service d'authentification, qui les transmet à l'API backend. L'utilisateur est créé, son mot de passe est haché et les logs d'activité sont enregistrés dans MongoDB. Une fois validé, le token et les informations sont stockés côté client.

### Navigation dans les annonces immobilières :

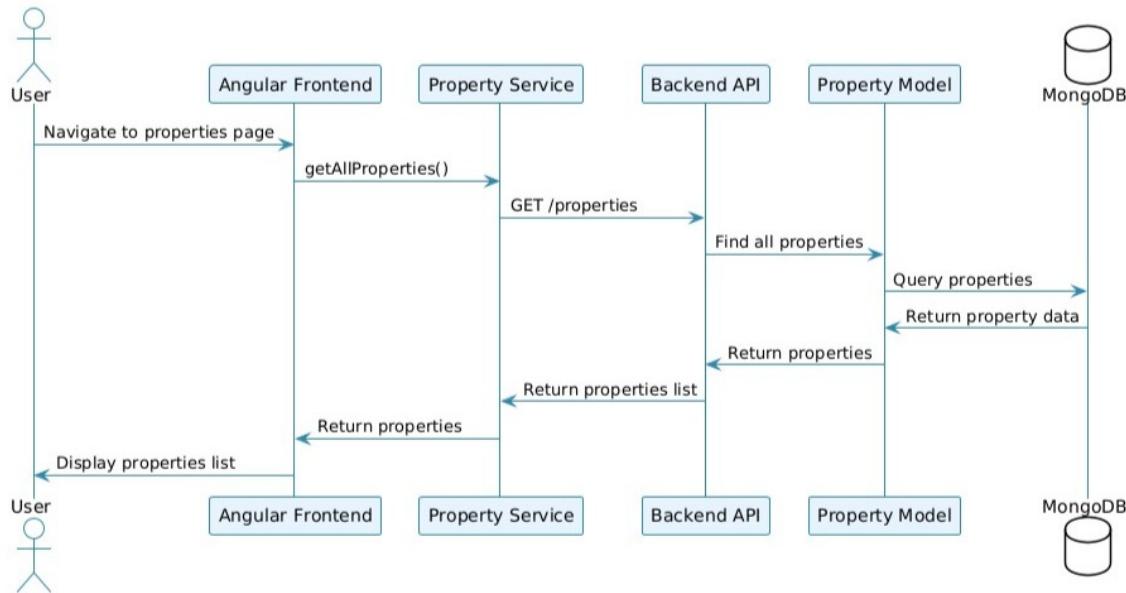


FIGURE 3.4 – Diagramme de séquence — Navigation dans les annonces immobilières

Ce diagramme décrit le processus de consultation des propriétés. Lorsqu'un utilisateur accède à la page des annonces, une requête est envoyée via le frontend Angular et le service de propriété jusqu'au backend, qui interroge MongoDB pour récupérer les données et les retourner à l'interface utilisateur.

### Prédiction de prix immobilier :

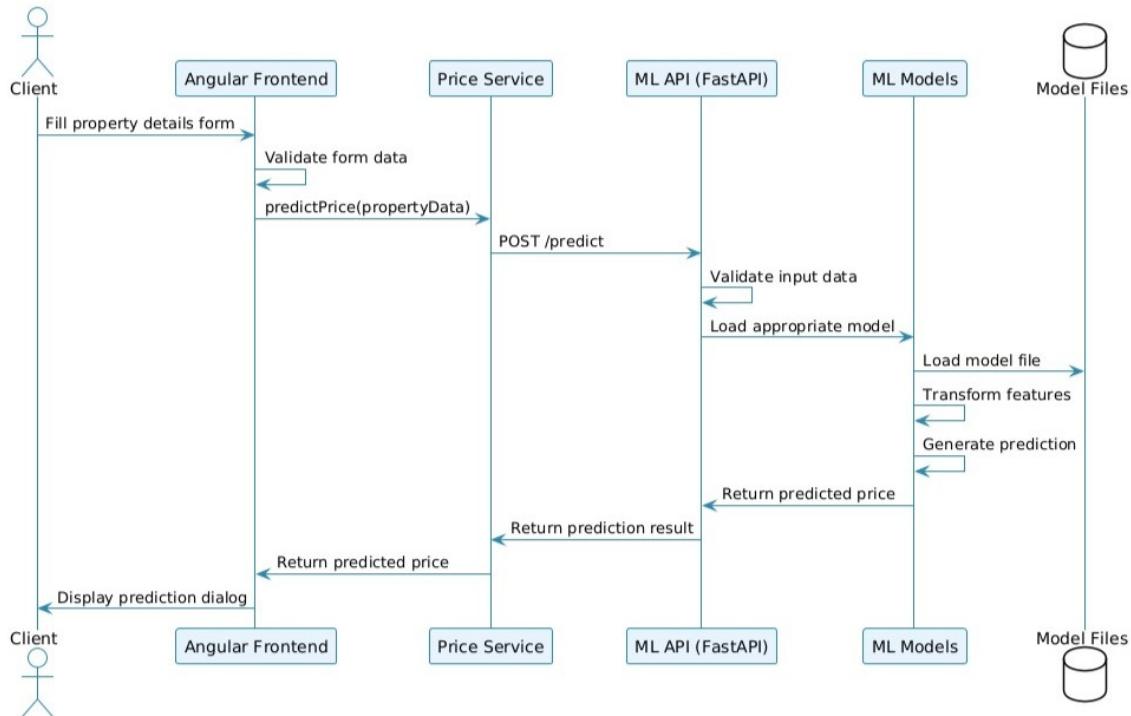


FIGURE 3.5 – Diagramme de séquence — Prédiction de prix immobilier

Ce diagramme illustre le processus de prédiction du prix d'un bien. Après que l'utilisateur soumet le formulaire de caractéristiques, les données sont envoyées à une API FastAPI qui charge le modèle adapté, effectue les transformations nécessaires et retourne le prix prédict à afficher à l'utilisateur.

### Validation d'un bien par l'administrateur :

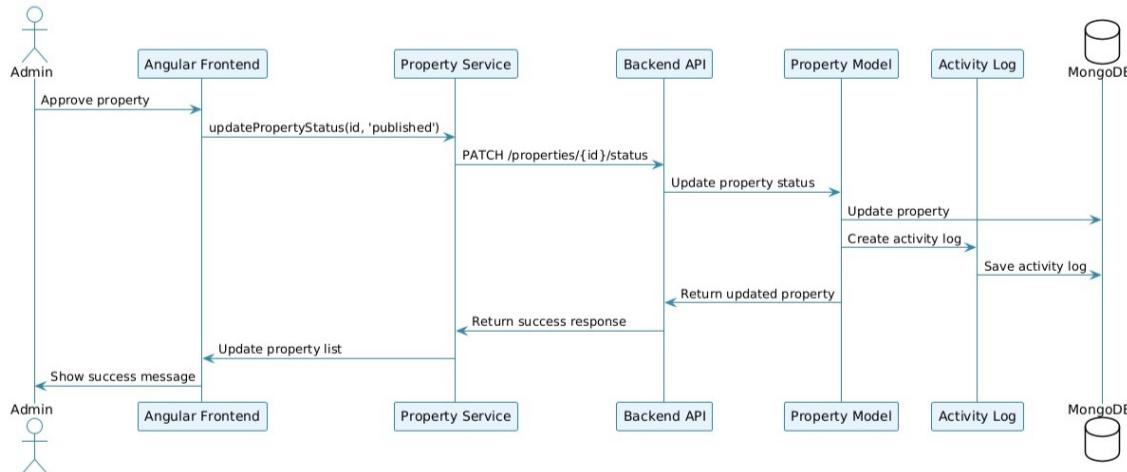


FIGURE 3.6 – Diagramme de séquence — Validation d'un bien par l'administrateur

Ce diagramme décrit le processus d'approbation d'un bien immobilier par un administrateur. Lorsqu'il valide une annonce, le statut est mis à jour via l'API, enregistré dans la base MongoDB, et un journal d'activité est créé. Le frontend est ensuite notifié du succès de l'opération.

### Inscription d'un utilisateur :

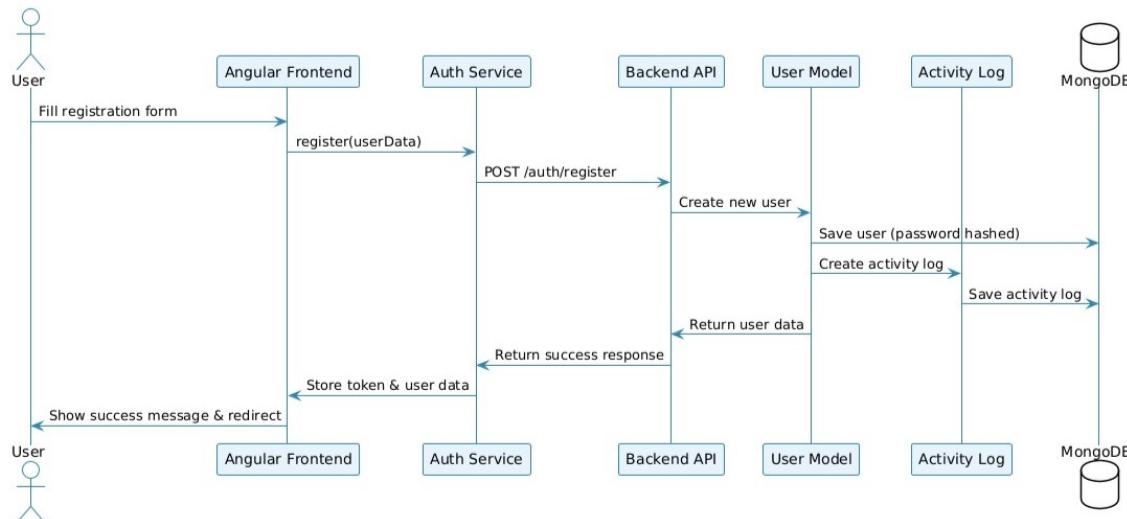


FIGURE 3.7 – Diagramme de séquence — Inscription d'un utilisateur

Ce diagramme illustre les étapes d'inscription d'un nouvel utilisateur. Après la soumission du formulaire, les données sont envoyées au service d'authentification qui appelle l'API backend. L'utilisateur est créé, les informations sont enregistrées dans MongoDB, puis le frontend affiche un message de confirmation.

# **Chapitre 4**

## **Réalisation et résultat**

### **1 La répartition des tâches 'Diagramme de GANTT'**

Pour une gestion efficace du projet de création de notre plateforme, nous avons utilisé un diagramme de Gantt. Ce diagramme nous a permis de planifier, décomposer le projet en tâches spécifiques, et suivre l'avancement.

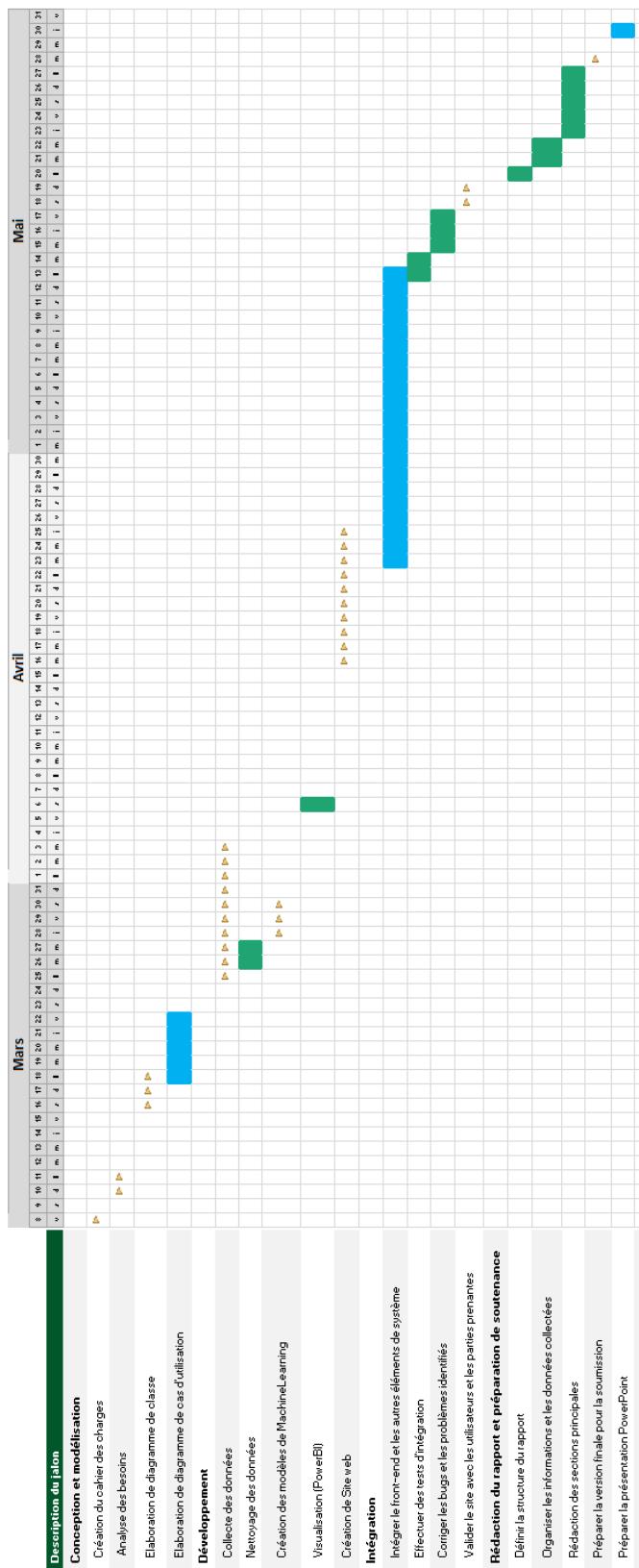


FIGURE 4.1 – Le diagramme de Gantt

## 2 Concepts clés

La réussite d'un projet de prédition immobilière repose sur la maîtrise de plusieurs concepts techniques que nous avons mobilisés tout au long du développement :

- **Web Scraping** : Technique permettant d'extraire automatiquement le contenu structuré de sites web. Elle repose sur des bibliothèques comme `BeautifulSoup`, `Playwright` ou `Selenium`. Elle nécessite la compréhension de la structure DOM et la gestion des pages dynamiques.
- **Pipeline ETL** : Processus structuré en trois étapes : Extraction, Transformation et Chargement (ETL). Dans notre projet, les données sont extraites via scraping, transformées avec `pandas` (nettoyage, typage, fusion), puis stockées dans une base relationnelle (MySQL). Ce processus est automatisé via des DAGs sous `Apache Airflow`.
- **Apprentissage supervisé** : Type d'apprentissage automatique où un modèle est entraîné à partir d'un jeu de données étiqueté. Dans notre cas, le prix du bien est la variable cible, et les caractéristiques du bien (surface, ville, type, etc.) sont les variables explicatives. L'objectif est de minimiser l'erreur de prédition (RMSE, MAE).
- **Stack MEAN** : Architecture logicielle complète constituée de MongoDB (base NoSQL), Express.js (serveur web), Angular (frontend) et Node.js (backend). Elle permet de construire des applications web monopage (SPA) performantes, interactives et entièrement JavaScript.
- **Business Intelligence (BI)** : Utilisation de tableaux de bord dynamiques via Power BI pour visualiser les tendances du marché, explorer les volumes d'annonces ou les variations de prix. Ces visualisations permettent à l'utilisateur admin de piloter efficacement l'évolution de la plateforme.

## 3 Architecture du système

L'architecture globale de notre plateforme repose sur une chaîne complète de traitement de données allant de la collecte brute d'informations sur le web jusqu'à la restitution d'une prédition de prix via une interface web conviviale. Elle intègre plusieurs briques fonctionnelles et techniques interconnectées qui assurent le bon fonctionnement du système, la modularité du code et la scalabilité de la solution.

- **Collecte des données** : Le système commence par le scraping de données immobilières depuis quatre plateformes majeures : Avito.ma et Mubawab pour le Maroc, LeBonCoin et Le Figaro Immobilier pour la France. Pour gérer des contenus HTML statiques et dynamiques, nous utilisons les bibliothèques Python `Selenium`, `Playwright` et `BeautifulSoup4`. Les données sont extraites de manière automatisée

à une fréquence régulière, puis sauvegardées localement sous forme de fichiers CSV normalisés.

— **Traitement, nettoyage et intégration** : Une fois extraites, les données sont traitées via la bibliothèque `pandas` : nettoyage des valeurs aberrantes, uniformisation des types et conversion de formats. Les données finalisées sont ensuite chargées dans une base de données relationnelle MySQL.

— **Orchestration automatisée** : Toutes les étapes du pipeline (scraping, nettoyage, chargement) sont orchestrées à l'aide de `Apache Airflow`. Des DAGs spécifiques ont été définis pour chaque flux de données (France / Maroc), permettant une exécution planifiée et une reprise en cas d'erreur. L'ensemble est conteneurisé via `Docker Compose`, garantissant l'isolation et la portabilité des services.

— **Modélisation prédictive** : Les données stockées sont ensuite utilisées pour entraîner des modèles de machine learning supervisé. Après évaluation ( $R^2$ , RMSE), les modèles sont sérialisés au format `.pkl` via `joblib`, puis intégrés dans un micro-service Python basé sur `FastAPI`. Ce service expose un endpoint RESTful (`/predict`) capable de recevoir les caractéristiques d'un bien et de retourner une estimation en temps réel.

— **Plateforme web** : Côté interface, nous avons développé une application complète selon l'architecture `MEAN` : `MongoDB` pour stocker les utilisateurs, `Express.js` pour gérer l'API `Node.js`, `Angular` pour le frontend, et `Node.js` comme moteur serveur. Cette plateforme permet :

- à l'utilisateur normal de saisir les caractéristiques d'un bien et d'obtenir une prédition ;
- à l'administrateur de gérer les comptes et d'accéder à un tableau de bord analytique.

— **Visualisation et supervision** : Pour fournir une vue synthétique du marché immobilier, nous avons développé des tableaux de bord interactifs sous `Power BI`, connectés à la base MySQL. Ces dashboards permettent à l'administrateur de visualiser l'évolution des prix, la répartition des biens par région, type ou surface, ainsi que les volumes d'annonces collectés.

L'architecture adoptée est donc modulaire, distribuée et robuste. Elle permet une collecte fiable des données, un traitement automatisé, une intégration fluide avec des modèles prédictifs, et une exposition claire des résultats à travers une interface web moderne et des outils BI puissants.

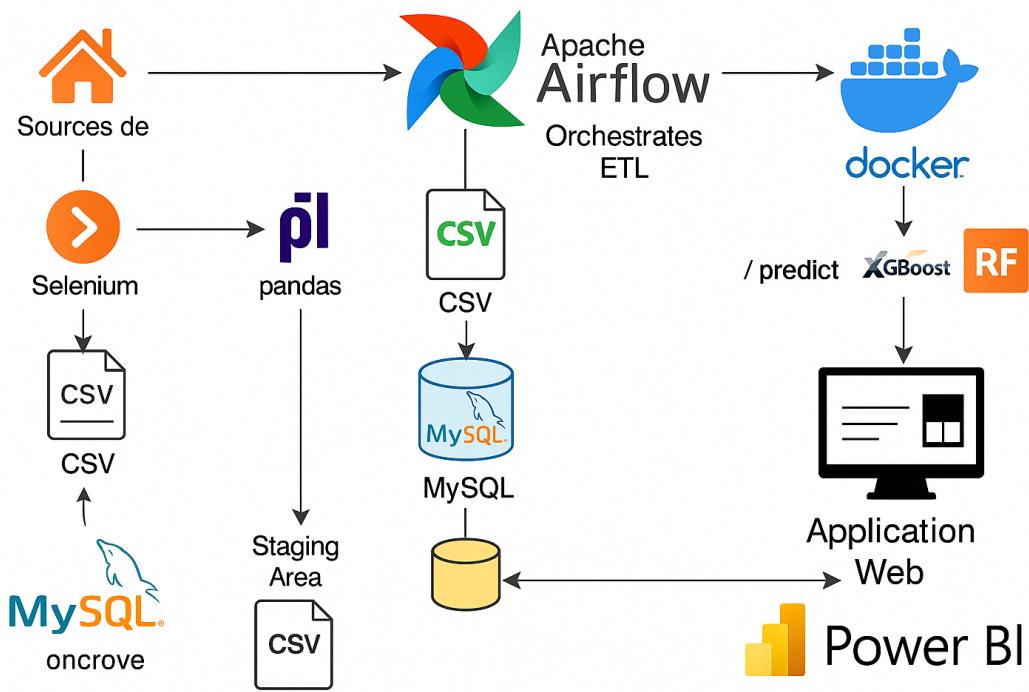


FIGURE 4.2 – Cycle de vie des données

## 4 Mise en œuvre

Le développement de notre système s'est organisé selon une architecture en couches logiques distinctes. Chaque couche est responsable d'un ensemble cohérent de tâches, facilitant la maintenance, l'évolutivité et l'intégration.

### 4.1 Couche de Présentation (Front-End)

- **Angular 17 :**

*Fonctionnalités principales :*

- Construction d'une application monopage (SPA) réactive.
- Organisation du code en **components**, **services** et **modules**.
- Routing côté client pour naviguer sans rechargement de page.

*Structure technique :*

- **components/** : formulaire de prédition, carte de résultat, barre de navigation.
- **services/** : **PredictionService**, **AuthService**.
- **guards/** : protection des routes admin.



FIGURE 4.3 – Logo du Framework Angular

- **Angular Reactive Forms :**

*Fonctionnalités principales :*

- Gestion dynamique des champs (ajout / suppression en temps réel).
- Validation synchrone et asynchrone (contraintes, regex, appels API).

- **Angular HttpClient :**

*Fonctionnalités principales :*

- Envoi des requêtes POST/GET au backend Express.
- Intercepteur pour l'injection automatique du token JWT dans les en-têtes.

- **TypeScript 5 :**

*Rôle :* langage typé garantissant la robustesse du code front.



FIGURE 4.4 – Logo de TypeScript

- **Bootstrap 5 :**

*Fonctionnalités principales :*

- Grille responsive et composants UI (cards, navbar, modals).
- Thématisation rapide sans surcharge CSS.



FIGURE 4.5 – Logo de Bootstrap

## 4.2 Couche Métier (Business Logic)

- **Node.js 20 :**

*Rôle* : environnement d'exécution JavaScript côté serveur.

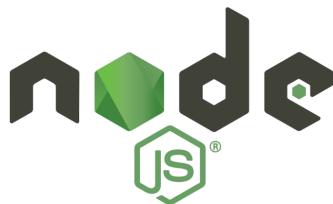


FIGURE 4.6 – Logo de Node.js

- **Express.js 5 :**

*Fonctionnalités principales* :

- Définition des routes REST : /auth, /predict, /admin.
- Middleware pour parsing JSON et gestion des erreurs.

*Structure technique* :

- `routes/` : segmentation par domaine métier.
- `controllers/` : logique associée à chaque route.
- `middlewares/` : authentification, limitation de débit, CORS.

- **JSON Web Token (JWT) :**

*Fonctionnalités principales* :

- Crédit de tokens signés (HMAC-SHA256) lors du login.
- Portage du rôle (`admin` / `user`) dans le payload.

- **bcrypt :**

*Rôle* : hachage sécurisé des mots de passe avant stockage MongoDB.

- **Axios :**

*Fonctionnalités principales :*

- Appels HTTP intra-services (Express → FastAPI /predict).
- Gestion automatique des timeouts et retries.



FIGURE 4.7 – Logo d’Axios

- **Mongoose 6 :**

*Fonctionnalités principales :*

- Schémas et validations pour les collections `users` et `logs`.
- Méthodes statiques pour la gestion des droits et des historiques.

### 4.3 Couche de Données (Data Access Layer)

- **MySQL 8.0 :**

*Fonctionnalités principales :*

- Stockage relationnel des annonces (`fr_listings`, `ma_listings`).
- Indexation composite sur (`localisation`, `date_pub`).



FIGURE 4.8 – Logo de MySQL

- **SQLAlchemy + PyMySQL :**

*Rôle :*

- ORM Python pour l’insertion bulk durant l’étape ETL.
- Mapping automatique des types (VARCHAR, FLOAT, DATE).

- **HeidiSQL :**

*Rôle :* client graphique pour requêtes ad-hoc et inspection manuelle.



FIGURE 4.9 – Logo de HeidiSQL

- **MongoDB 6 :**

*Fonctionnalités principales :*

- Stockage orienté documents des comptes et rôles utilisateurs.
- Journalisation légère des requêtes de prédition.



FIGURE 4.10 – Logo de MongoDB

- **Récapitulation :**

Chaque technologie joue donc un rôle déterminant dans sa couche : *Angular/Ty-  
peScript/Bootstrap* offrent une interface riche ; *Node/Express/JWT/bcrypt/Axios/-  
Mongoose* implémentent la logique, la sécurité et l’API ; *MySQL/SQLAlchemy/-  
MongoDB* assurent la persistance et l’alimentation analytique.

## 4.4 Scraping et traitement des données

- **Technologies utilisées :**

- *Selenium, Playwright* : Pour simuler un navigateur et extraire des pages web dynamiques (JavaScript).
- *BeautifulSoup4* : Pour parser les pages HTML et extraire les données pertinentes.
- *pandas* : Pour nettoyer, fusionner et optimiser les données.

- **Fonctionnalités principales :**

- Collecte automatisée d’annonces immobilières depuis 4 sites (Avito, Muba-  
wab, LeBonCoin, Le Figaro).

- Simulation de clics, scroll infini, attente des éléments DOM.
- Parsing HTML : extraction des champs prix, surface, localisation, type de bien, etc.
- Suppression des doublons, conversion des types, standardisation des formats.
- Fusion finale des jeux de données marocains et français dans un CSV unique.

- **Rôle dans le système :** Ce module constitue la première étape de la chaîne ETL. Il permet d'avoir une base fiable et structurée de données immobilières, actualisée régulièrement et prêtes à l'analyse ou à l'entraînement ML.

## 4.5 Entraînement et déploiement des modèles de Machine Learning

- **Technologies utilisées :**
  - scikit-learn, RandomForest : Pour l'entraînement des modèles prédictifs.
  - joblib : Pour la sérialisation des modèles entraînés.
  - FastAPI : Micro-service Python exposant l'API de prédiction.
- **Fonctionnalités principales :**
  - Nettoyage du dataset final pour l'apprentissage supervisé (features numériques et catégorielles).
  - Entraînement de modèle Random Forest Regressor.
  - Évaluation des performances via RMSE, MAE, R<sup>2</sup>.
  - Sérialisation des modèles pour déploiement.
  - Mise à disposition d'un service RESTful via FastAPI : endpoint /predict.
- **Rôle dans le système :** Ce service est consommé par le backend Node.js et permet de fournir à l'utilisateur une estimation de prix immobiliers basée sur ses données en temps réel.
- **Tableau comparatif justifiant le choix :**
  - **Le marché Français :**

TABLE 4.1 – Comparaison des modèles de régression pour le marché Français

Algorithm	Use case	RMSE	MAE	R <sup>2</sup>	MAE/RMSE
XGBoost	VentesAchats	179,234.17	94,273.73	0.8161	0.53
Random Forest	Location	390.79	225.39	0.9687	0.58
Decision Tree	VentesAchats	240,000	202,000	0.82	0.841
Linear Regression	Location	500.55	200.33	0.75	0.4

Le tableau comparatif du marché français présente les résultats de quatre algorithmes de régression : XGBoost, Random Forest, Decision Tree et Linear Regression. Le modèle XGBoost appliqué au cas d'usage « VentesAchats » se distingue par ses bonnes performances avec un RMSE de 179234,17 € et un R<sup>2</sup> de 0,8161, indiquant une forte capacité à modéliser la relation entre les caractéristiques et les prix. Pour la prédiction des loyers, Random Forest surpassé les autres modèles avec un excellent R<sup>2</sup> de 0,9687 et une faible erreur quadratique moyenne (RMSE = 390,79 €), traduisant une grande précision. En comparaison, la régression linéaire montre des résultats moins convaincants, avec un R<sup>2</sup> de seulement 0,75, ce qui suggère qu'elle est moins adaptée à la complexité du marché immobilier français.

- **Le marché Marocain :**

TABLE 4.2 – Comparaison des modèles de régression pour le marché Marocain

Algorithm	Use case	RMSE	MAE	R <sup>2</sup>	MAE/RMSE
XGBoost	VentesAchats	665,399.35	470,795.41	0.9380	0.71
Random Forest	Location	2,545.96	1,744.46	0.7812	0.69
Decision Tree	VentesAchats	800,000	640000	0.82	0.80
Linear Regression	Location	500.00	400.00	0.75	0.80

Concernant le marché marocain, les résultats révèlent une performance relativement similaire en termes de hiérarchie des modèles. Le modèle XGBoost demeure le plus performant pour le cas « VentesAchats » avec un RMSE de 665399,35 DH et un R<sup>2</sup> de 0,9380, démontrant sa robustesse même dans un contexte de données localement hétérogènes. Pour le cas « Location », le Random Forest obtient un R<sup>2</sup> modéré de 0,7812 et une erreur de 2545,96 DH, ce qui reste acceptable. La régression linéaire, quant à elle, présente une erreur relativement faible (RMSE = 500 DH), mais son R<sup>2</sup> de 0,75 indique une moindre capacité à expliquer la variance. Ces résultats montrent que, malgré des disparités de marché, XGBoost et Random Forest restent des choix cohérents, avec une supériorité marquée de XGBoost pour les ventes.

## 4.6 Visualisation des résultats via Power BI

- **Technologies utilisées :**

- Power BI Desktop : Pour la création de dashboards dynamiques.
- Power Query : Pour la connexion à MySQL et le pré-traitement.
- DAX : Pour les formules, KPI et visualisations calculées.

- **Fonctionnalités principales :**

- Tableau de bord interactif connecté à la base MySQL.
- Cartes de chaleur géographiques, histogrammes, séries temporelles.

- KPIs : nombre d'annonces, prix moyen/m<sup>2</sup>, tendances par mois ou par ville.
- Segmentation dynamique par type de bien, source, pays.

- **Rôle dans le système :** La BI permet à l'administrateur de suivre en temps réel l'évolution du marché, la qualité des données collectées, et d'orienter les futures itérations du modèle.

## 4.7 Conteneurisation et déploiement avec Docker

- **Technologies utilisées :**

- Docker, Docker Compose : Pour la conteneurisation et la coordination multi-services.
- GitHub Actions : Pour l'intégration continue et le déploiement automatique.

- **Fonctionnalités principales :**

- Chaque module (scraper, FastAPI, Node.js, Angular, MongoDB, MySQL, Airflow) est encapsulé dans un conteneur indépendant.
- Docker Compose orchestre le lancement coordonné des services.
- CI/CD : à chaque push sur la branche `main`, les images sont rebuildées, testées, puis poussées sur Docker Hub.

- **Rôle dans le système :** Docker garantit la portabilité, la reproductibilité et la rapidité du déploiement sur n'importe quel environnement local ou cloud.

## 5 Tests et validation

### 5.1 Outil utilisé pour tester les performances

- **Description de l'outil Lighthouse :**

Lighthouse est un outil automatisé et open-source développé par Google pour améliorer la qualité des pages web. Il offre un ensemble de tests et d'audits sur différents aspects d'un site web, permettant aux développeurs d'identifier les problèmes et d'optimiser les performances, l'accessibilité, les bonnes pratiques et le SEO (search engine optimization).



FIGURE 4.11 – Logo de l'outil LIGHTHOUSE

- **Analyse des Performances** : Lighthouse mesure le temps de chargement des pages, l'interactivité et la stabilité visuelle. Il simule des conditions de réseau et d'appareil pour fournir des recommandations sur l'optimisation des ressources.
- **Évaluation de l'Accessibilité** : L'outil vérifie que les pages web sont utilisables par les personnes ayant des handicaps, en testant des éléments comme les attributs ARIA (Accessible Rich Internet Applications), le contraste des couleurs, la navigation au clavier, et les descriptions textuelles des images.
- **Suivi des Standards** : Lighthouse évalue si le site respecte les meilleures pratiques en termes de développement web, sécurité et utilisation des API.
- **Optimisation pour les Moteurs de Recherche** : L'outil vérifie que le site est bien optimisé pour être indexé par les moteurs de recherche, en testant les balises meta, les en-têtes, la structure des URL (Uniform Resource Locator), et d'autres éléments pertinents.

## 5.2 Les tests

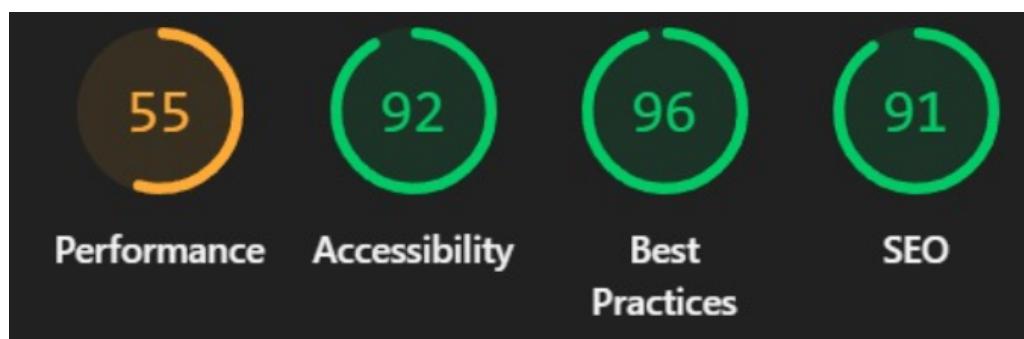


FIGURE 4.12 – Les tests de performance, accessibilité, bonnes pratiques et SEO

- **Performance (Score : 55) :**

Ce score mesure la rapidité de chargement et la fluidité globale de l'expérience utilisateur. Avec un score de 55, le site affiche une performance moyenne. Cela peut être dû à des temps de réponse longs du serveur, à des images non optimisées, à un JavaScript trop lourd ou au manque de mise en cache. Ce niveau suggère qu'un utilisateur peut rencontrer des lenteurs, en particulier sur des connexions mobiles ou des appareils peu puissants. Une optimisation des assets statiques, du lazy-loading, ou de l'utilisation de CDN est fortement recommandée pour améliorer ce score.

- **Accessibilité (Score : 92) :**

Ce test évalue dans quelle mesure ton site est accessible aux utilisateurs en situation de handicap (par exemple via lecteurs d'écran, navigation clavier, contrastes de couleurs). Un score de 92 indique que l'accessibilité est bien prise en compte. Cela signifie que les balises alt sont présentes, que les contrastes sont corrects, que les éléments interactifs sont accessibles au clavier, etc. Quelques points pourraient encore être optimisés pour atteindre un score parfait, comme améliorer certaines balises ARIA ou vérifier la hiérarchie des titres.

- **Bonnes Pratiques (Score : 96) :**

Cette métrique vérifie la conformité du site aux bonnes pratiques de développement web (sécurité, responsive design, erreurs JavaScript, etc.). Avec un score de 96, le site respecte presque toutes les recommandations : pas d'utilisation dangereuse d'API obsolètes, bonne structure du DOM, et gestion correcte des erreurs. Les 4 points manquants peuvent être dus à de petits avertissements comme l'absence de HTTPS sur certains liens externes, ou des images non responsive.

- **SEO (Score : 91) :**

Le score SEO (Search Engine Optimization) mesure dans quelle mesure le site est lisible et indexable par les moteurs de recherche. Un score de 91 est excellent et montre que le site est globalement bien structuré : les balises title, meta-description, alt, les URLs et la structure des pages sont bien définies. Les points restants peuvent concerter des optimisations fines comme l'ajout d'un fichier robots.txt, d'un sitemap.xml ou de balises Open Graph pour le partage sur les réseaux sociaux.

## 6 Visualisation des marchés

### 6.1 Le marché Marocain

- 1<sup>ère</sup> visualisation :

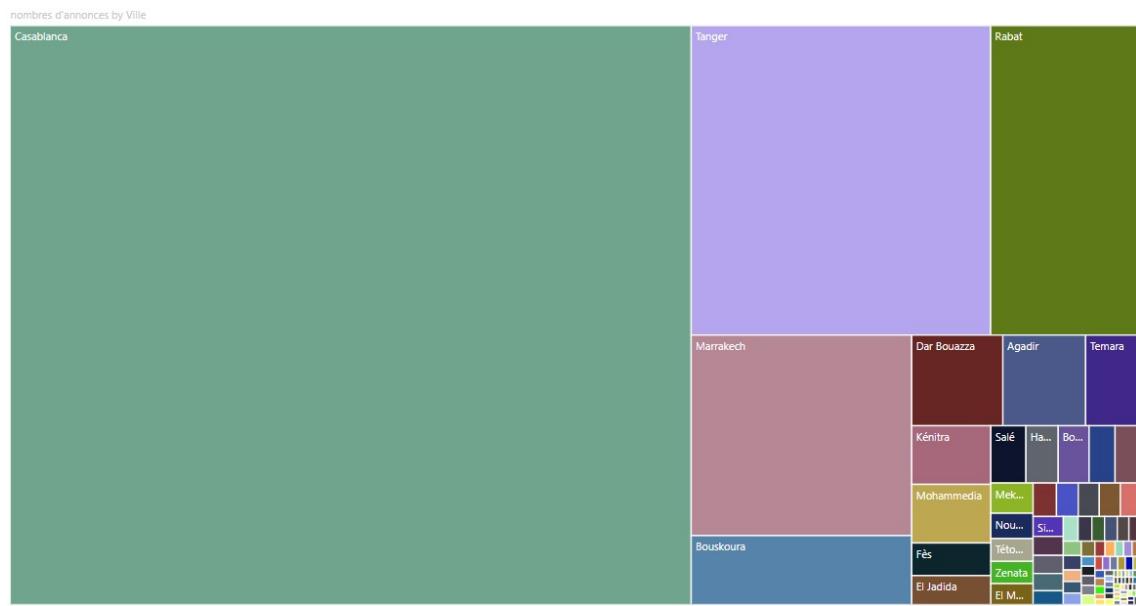


FIGURE 4.13 – 1<sup>ère</sup> visualisation du marché Marocain

Diagramme Treemap représentant la dominance des villes dans l'ensemble des annonces immobilières, qu'il s'agisse de ventes ou de locations.

- 2<sup>ème</sup> visualisation :

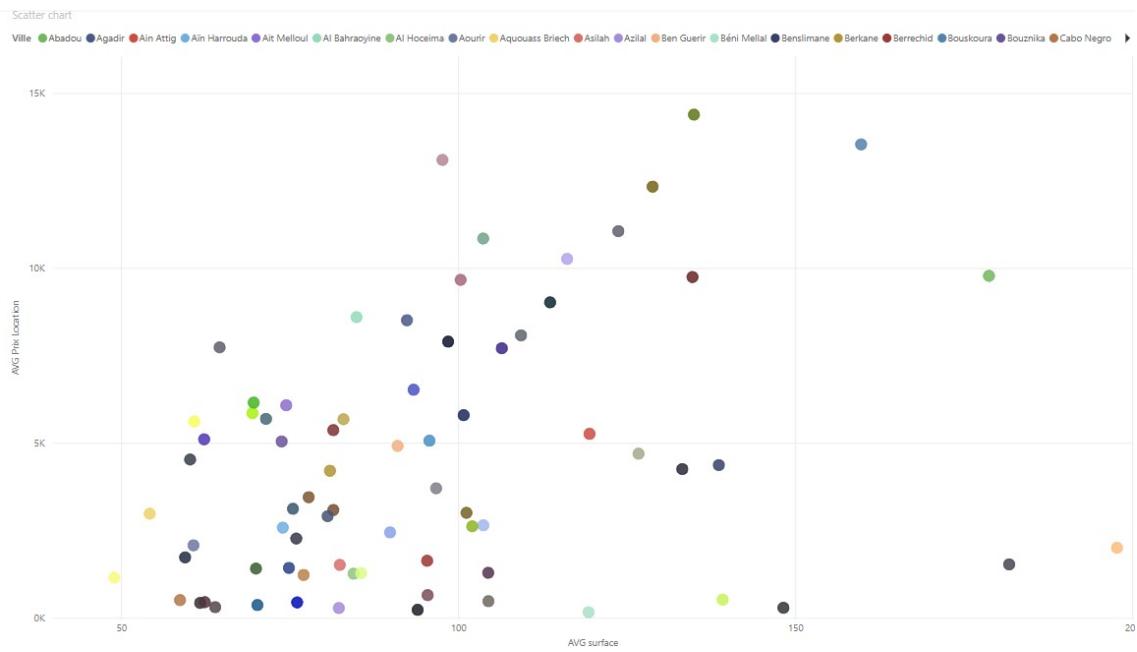


FIGURE 4.14 – 2<sup>ème</sup> visualisation du marché Marocain

Diagramme de dispersion représentant la variance de la surface Moyenne et le prix moyen appartenant aux différentes villes des annonces de location.

• 3<sup>ème</sup> visualisation :

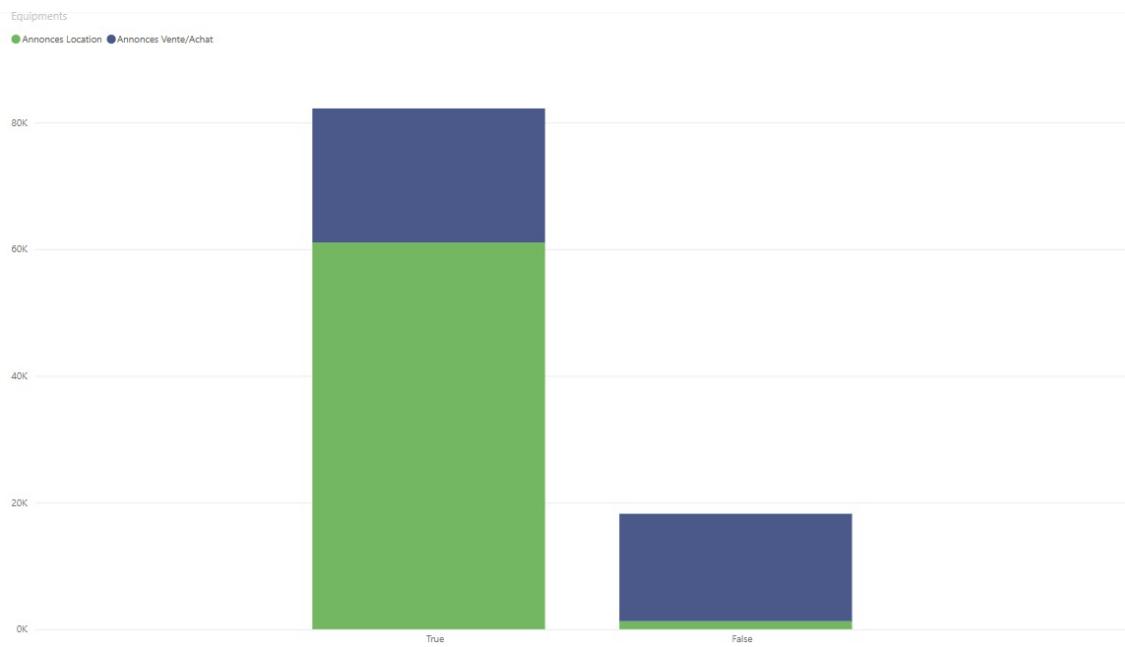


FIGURE 4.15 – 3<sup>ème</sup> visualisation du marché Marocain

Histogramme empilé illustrant le nombre et la répartition des annonces de location (en vert) et de vente (en gris) équipées ou pas équipées.

- 4<sup>ème</sup> visualisation :

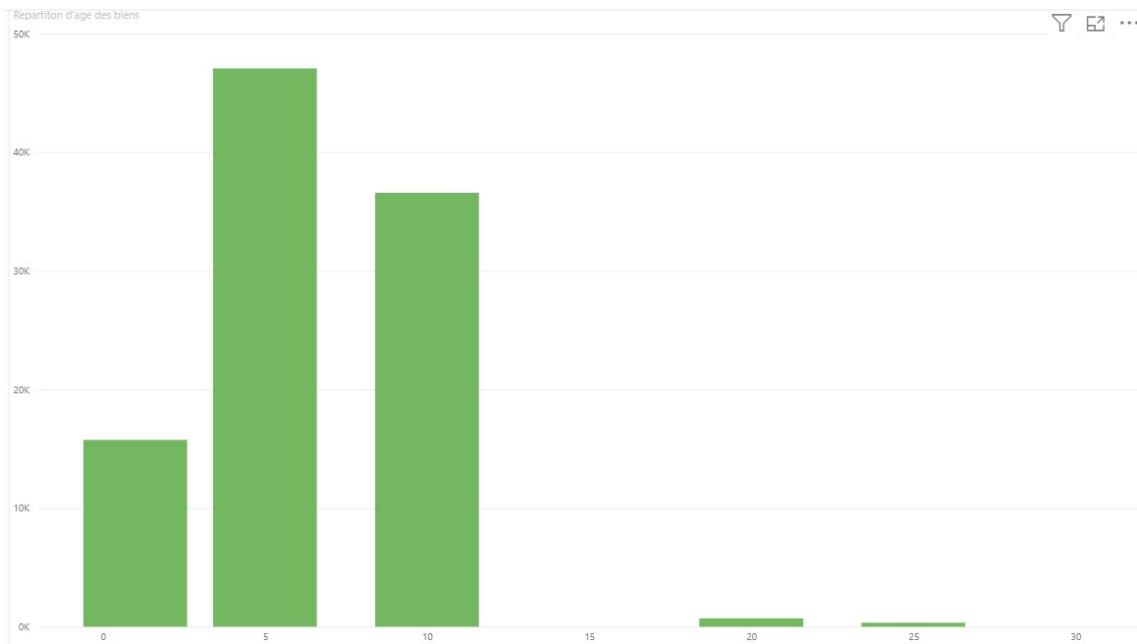


FIGURE 4.16 – 4<sup>ème</sup> visualisation du marché Marocain

Histogramme illustrant le nombre des annonces s'agissant de ventes ou de locations et appartenant à une tranche d'âge de bien immobilier spécifique.

## 6.2 Le marché Français

- 1<sup>ère</sup> visualisation :



FIGURE 4.17 – 1<sup>ère</sup> visualisation du marché Français

La figure présente un visuel Key Influencers (Power BI) visant à identifier les facteurs qui font varier le prix moyen en euros d'un bien immobilier.

Le panneau de gauche met en évidence que la surface moyenne est l'influenceur principal : lorsque la surface dépasse 48,87 m<sup>2</sup>, le prix moyen augmente en moyenne de 91 340 €.

Le nuage de points à droite détaille cette relation : chaque point correspond à un bien ; l'axe X représente la surface moyenne, l'axe Y le prix moyen. La droite de tendance en pointillés confirme une corrélation positive entre surface et prix – plus la surface est élevée, plus le prix tend à augmenter.

Cette visualisation démontre ainsi que la superficie est le principal levier explicatif des variations de prix dans l'échantillon étudié.

- **2<sup>ème</sup> visualisation :**

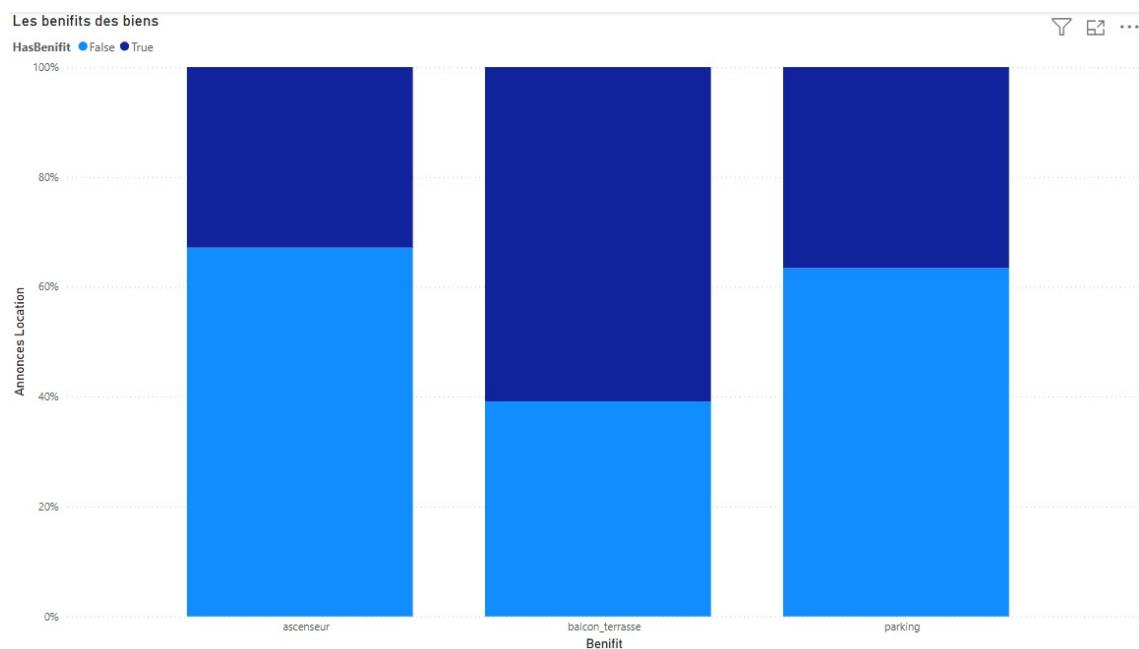


FIGURE 4.18 – 3<sup>ème</sup> visualisation du marché Français

L'histogramme à colonnes empilées en pourcentage qui visualiser la présence relative des différents types de bénéfices (Parking, Ascenseur, terrasse/balcon) dans l'ensemble des annonces immobilières, qu'il s'agisse de ventes ou de locations.

- 3<sup>ème</sup> visualisation :

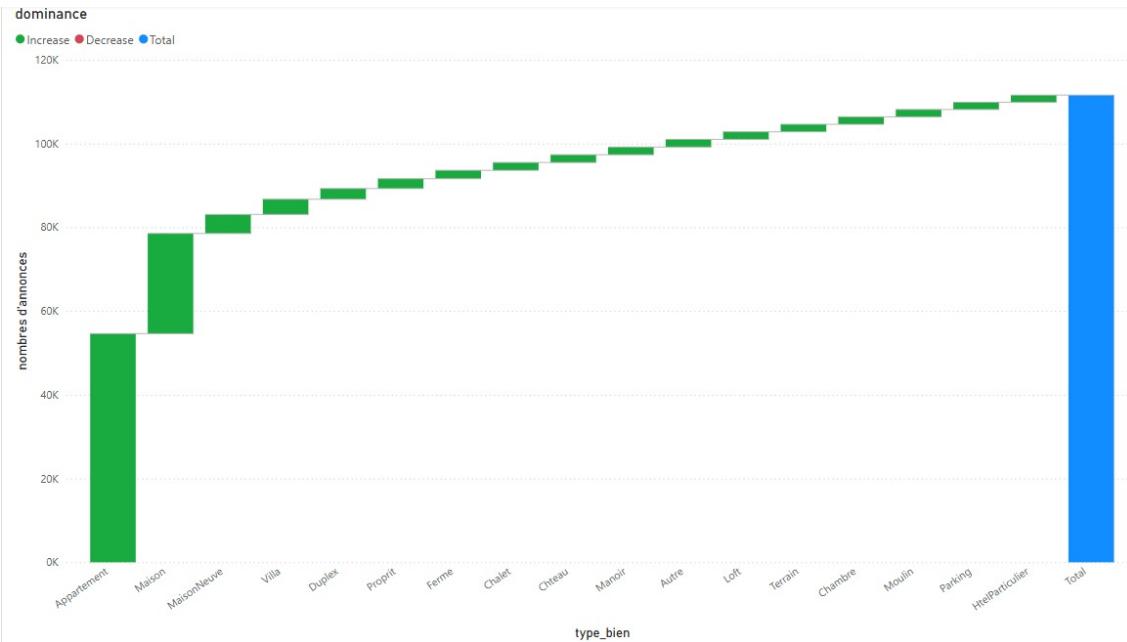


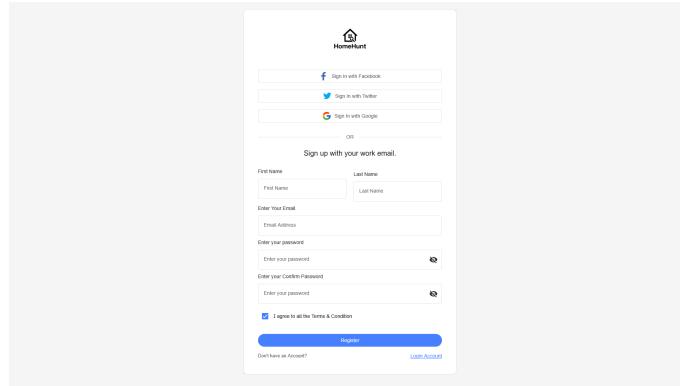
FIGURE 4.19 – 3<sup>ème</sup> visualisation du marché Français

Diagramme de cascade illustrant la dominance des différents types de biens immobiliers dans l'ensemble des annonces immobilières, qu'il s'agisse de ventes ou de locations.

## 7 Illustration de l'interface graphique du site

### 7.1 L'interface registration/authentification

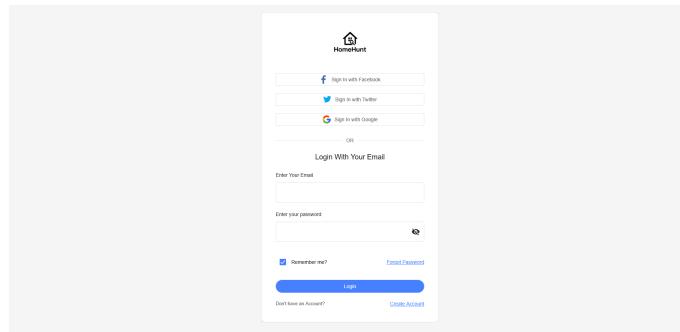
- Page de registration :



The registration page for HomeHunt features social login options (Facebook, Twitter, Google) at the top. Below them is a "Sign up with your work email" section. The form includes fields for First Name, Last Name, Email Address, and two password entries (Password and Confirm Password). A checkbox for accepting Terms & Condition is present, along with a "Register" button. Navigation links for "Don't have an Account?" and "Login Account" are located at the bottom.

FIGURE 4.20 – La page de registration d'un utilisateur

- Page d'authentification :



The login page for HomeHunt offers social logins (Facebook, Twitter, Google) at the top. It then asks for "Login With Your Email". The form contains fields for Email Address and Password, with a "Remember me?" checkbox and a "Forgot Password" link. A "Login" button is at the bottom, flanked by "Don't have an Account?" and "Create Account" links.

FIGURE 4.21 – La page d'authentification d'un utilisateur

## 7.2 L'interface d'utilisateur

- Page d'accueil :

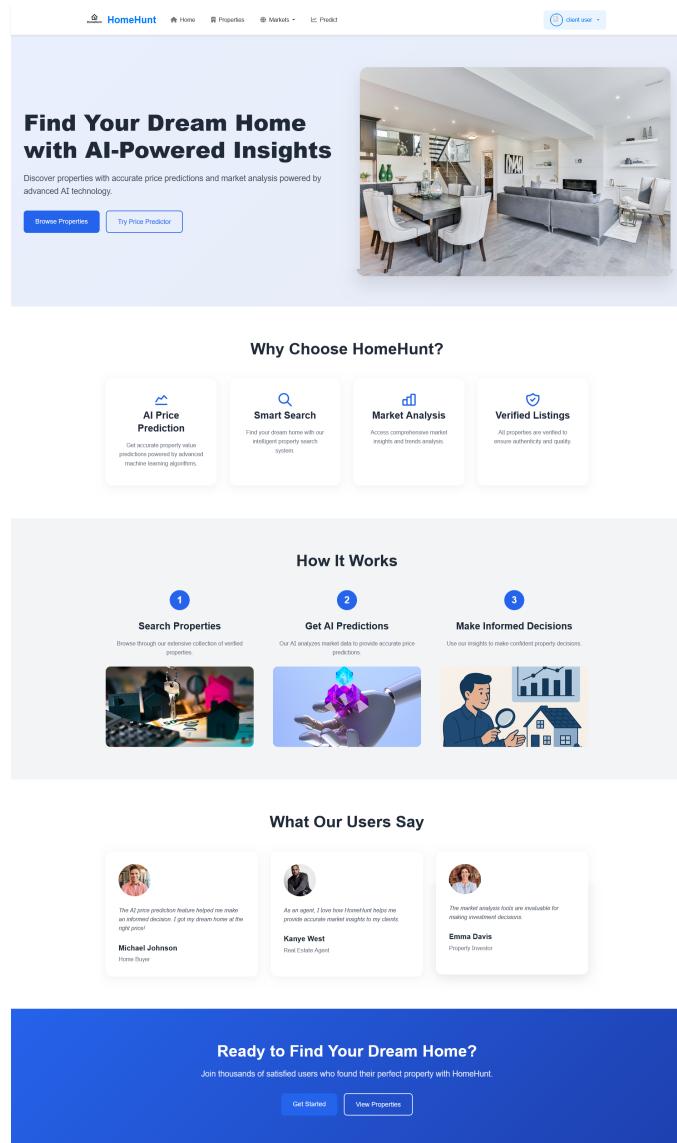


FIGURE 4.22 – La page d'accueil du site

- Page du marché Marocain :

FIGURE 4.23 – Liste des articles du marché Marocain

- Page du marché Français :

FIGURE 4.24 – Liste des articles du marché Français

- Liste des immobiliers :

FIGURE 4.25 – Liste des immobiliers d'un utilisateur

- Liste des favoris :

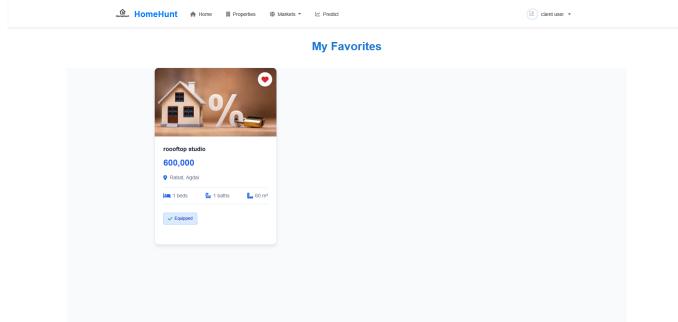


FIGURE 4.26 – Liste des articles favoris d'un utilisateur

- Profil d'un utilisateur :

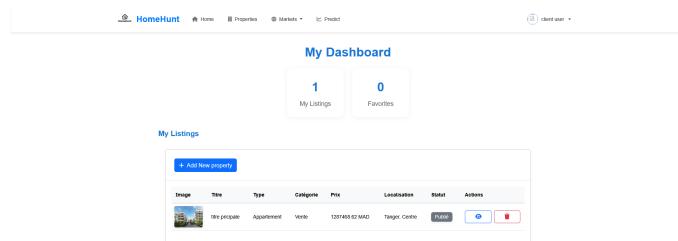


FIGURE 4.27 – Les informations d'un profil utilisateur

- Formulaire d'ajout d'un bien :

FIGURE 4.28 – Formulaire d'ajout d'un bien à prédire

- Résultat de la prédiction :

FIGURE 4.29 – Résultat de la prédiction du prix d'un bien saisi

## 7.3 L'interface d'administrateur

- Tableau de bord de l'administrateur :

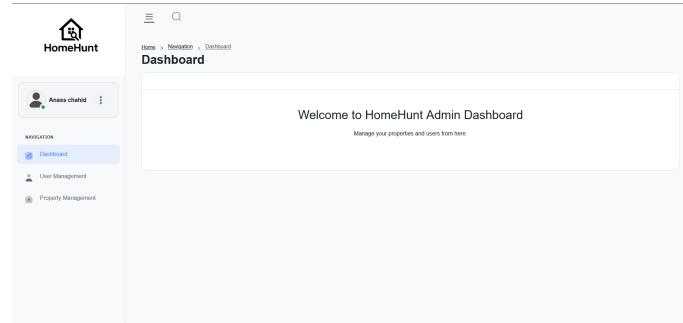


FIGURE 4.30 – Tableau de bord d'administration du site

- Liste des utilisateurs :

FIRST NAME	LAST NAME	EMAIL	ROLE	STATUS	ACTIONS
Anass	chahid	anasse@gmail.com	client	active	[Edit] [Delete]
client	user	client@gmail.com	client	active	[Edit] [Delete]
Hanza	Abdes	hanza@gmail.com	client	active	[Edit] [Delete]
test	test	test@gmail.com	client	active	[Edit] [Delete]
awdawd	wadawdawd	awdawdawd@gmail.com	client	active	[Edit] [Delete]

FIGURE 4.31 – Liste des différents utilisateurs du site

- Ajout d'un utilisateur :

The modal dialog for adding a new user contains the following fields:

- Email: anasse@gmail.com
- Profile Picture: No file selected
- Password: test
- Role: Client
- Status: active

FIGURE 4.32 – Ajout d'un compte utilisateur

- Modification d'un utilisateur :

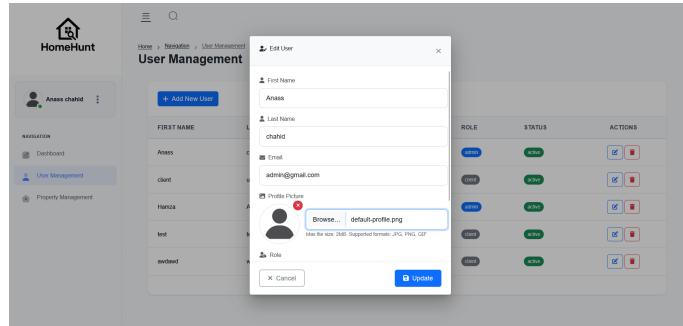


FIGURE 4.33 – Modification des propriétés d'un utilisateur

- Suppression d'un utilisateur :

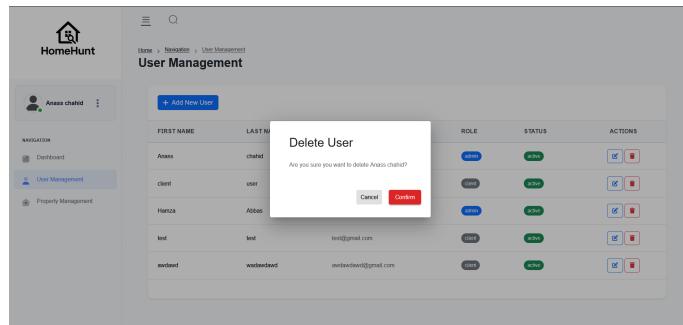


FIGURE 4.34 – Suppression d'un compte utilisateur

- Liste des biens :

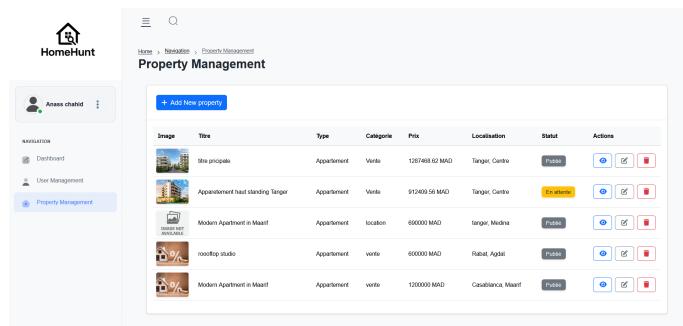
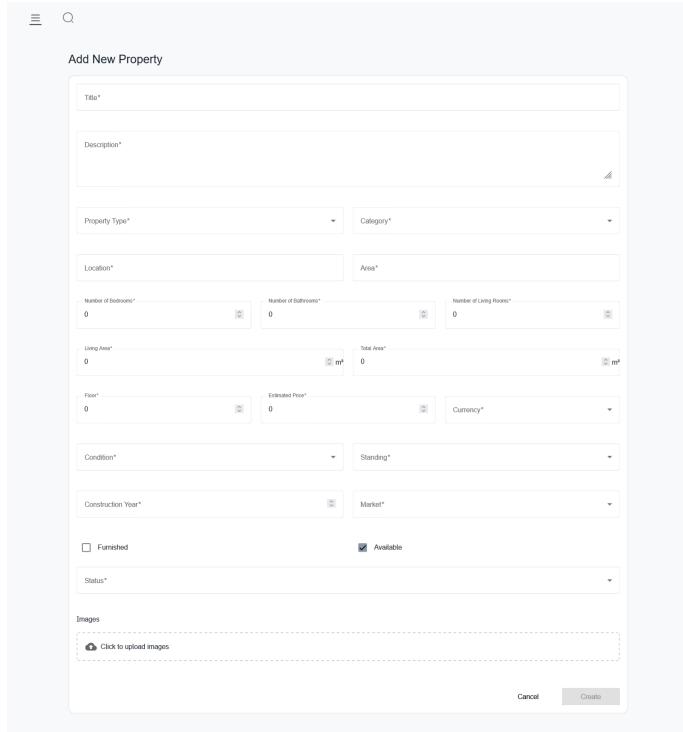


FIGURE 4.35 – Liste des différents immobiliers publié dans le site

- Ajout d'un bien :



The screenshot shows a web-based form titled "Add New Property". The form is divided into several sections:

- Title\***: A text input field.
- Description\***: A large text area for a detailed description.
- Property Type\*** and **Category\***: Two dropdown menus.
- Location\*** and **Area\***: Text input fields.
- Number of BedRooms\***, **Number of Bathrooms\***, and **Number of Living Rooms\***: Input fields with numerical values (0) and unit dropdowns (m²).
- Living Area\*** and **Total Area\***: Input fields with numerical values (0) and unit dropdowns (m²).
- Floor\*** and **Estimated Price\***: Input fields with numerical values (0) and unit dropdowns (Currency).
- Condition\*** and **Standing\***: Two dropdown menus.
- Construction Year\*** and **Market\***: Two dropdown menus.
- Furnished** and **Available**: Two checkboxes, with "Available" checked.
- Status\***: A dropdown menu.
- Images**: A section with a placeholder text "Click to upload images" and a dashed rectangular area for image uploads.

At the bottom right of the form are two buttons: "Cancel" and "Create".

FIGURE 4.36 – Ajout d'un immobilier avec ses propriétés

- Modification d'un bien :

FIGURE 4.37 – Modification des propriétés d'un immobilier

- Suppression d'un bien :

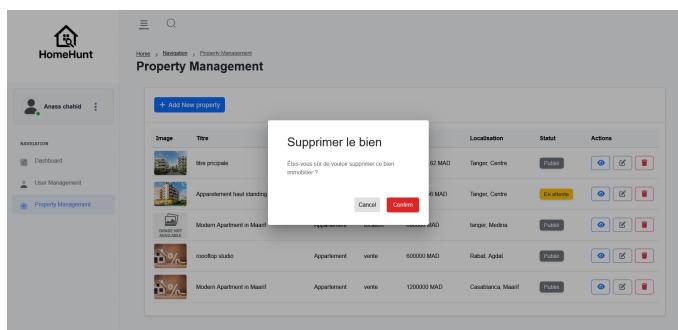


FIGURE 4.38 – Suppression d'un immobilier

- Les détails d'un bien :

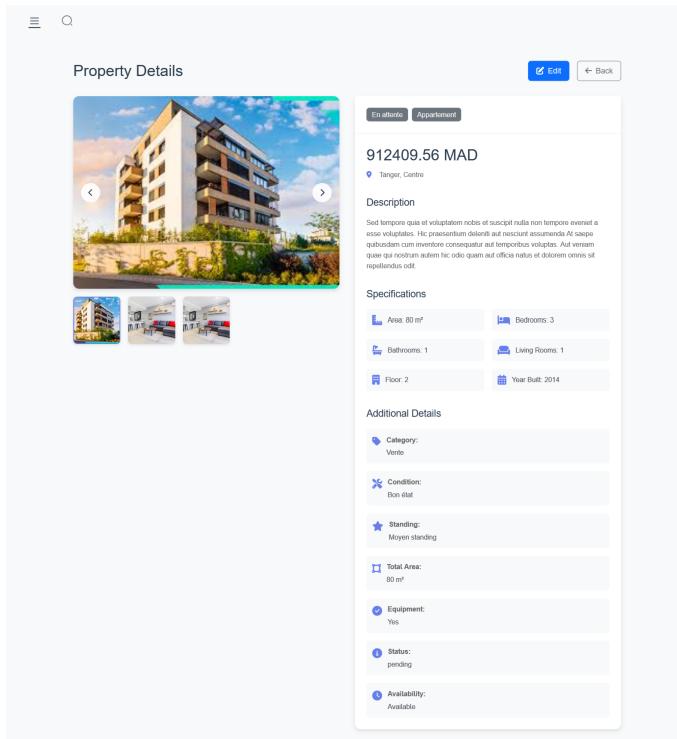


FIGURE 4.39 – Affichage des différentes propriétés d'un immobilier

## 8 Problèmes rencontrés et solutions apportées

Au cours du développement du projet, plusieurs difficultés techniques et organisationnelles ont été rencontrées. Grâce à une collaboration continue et à une approche itérative, l'équipe a pu identifier les causes racines de ces problèmes et y apporter des solutions concrètes. Cette section présente les principaux obstacles rencontrés ainsi que les stratégies adoptées pour les résoudre.

### Blocages dans le scraping dynamique (JavaScript, CAPTCHA) :

*Problème :* Certains sites comme LeBonCoin ou Mubawab chargent leur contenu de façon dynamique via JavaScript, ou présentent des obstacles de type CAPTCHA ou détection de bots.

*Solution apportée :*

- Utilisation de Playwright et Selenium en mode headless pour simuler un vrai navigateur avec interaction DOM.

- Ajout de délais aléatoires (`time.sleep`) et d'en-têtes personnalisés (User-Agent) pour limiter la détection automatique.
- Implémentation d'un système de relance automatique des sessions en cas d'échec.

## Orchestration et reprise de données avec Airflow :

*Problème :* L'enchaînement des tâches ETL (scraping → nettoyage → stockage) devait être résilient en cas d'échec d'une étape. De plus, il fallait garantir la non-duplication des données.

*Solution apportée :*

- Mise en place de DAGs robustes dans Apache Airflow avec `retry` automatique, gestion de dépendances et journalisation.
- Ajout d'un identifiant unique par annonce basé sur l'URL et le titre (`SHA-1 hash`) pour prévenir les doublons lors des insertions MySQL.
- Planification quotidienne à horaire fixe avec contrôle d'intégrité après chaque chargement.

## Déploiement des modèles ML en production :

*Problème :* L'intégration d'un modèle ML (Machine Learning) dans un système web nécessite à la fois une bonne performance (temps de réponse < 1s) et une gestion d'erreurs robuste.

*Solution apportée :*

- Sérialisation du modèle entraîné via `joblib` et encapsulation dans une API REST avec `FastAPI`.
- Containerisation du micro-service Python avec Docker et communication interne avec le backend Express via `Axios`.
- Implémentation de vérifications d'entrée et de logs côté Python pour gérer les cas d'entrée non conformes.

## Sécurisation de l'accès à la plateforme MEAN :

*Problème :* Il était crucial de protéger les routes sensibles (prédiction, tableau de bord admin) contre les accès non autorisés.

*Solution apportée :*

- Intégration complète de l'authentification via JWT côté backend et AuthGuard côté Angular.
- Hachage des mots de passe avec `bcrypt` pour protéger les données des utilisateurs.
- Mise en place de middleware Express pour filtrer les rôles (`admin` vs `user`) et restreindre l'accès aux données sensibles.

## Gestion collaborative et intégration continue :

*Problème :* La coordination entre quatre développeurs sur des modules séparés a engendré des conflits de versions, des dépendances manquantes et des pertes de cohérence.

*Solution apportée :*

- Centralisation du code sur GitHub avec gestion des branches par fonctionnalité.
- Mise en place d'un fichier `.env` partagé pour uniformiser les configurations d'environnement.
- Déploiement continu via `GitHub Actions` et `Docker Compose` pour simuler le système complet en local.
- Organisation de réunions hebdomadaires + points quotidiens informels pour synchroniser les avancées.

## 9 Limitations et améliorations

Bien que la plateforme développée remplisse ses objectifs fondamentaux, plusieurs limites ont été identifiées au cours du développement et des phases de test. Certaines sont d'ordre technique, d'autres sont liées à la nature même des données disponibles ou à l'évolution des technologies utilisées. Cette section présente ces limitations ainsi que des suggestions d'amélioration envisagées pour de futures itérations du projet.

### Dépendance aux sources externes :

Les données immobilières collectées proviennent de plateformes tierces (Avito, LeBonCoin, etc.) qui peuvent modifier leur structure HTML à tout moment. Cela peut casser les scripts de scraping et nécessiter des ajustements fréquents.

*Amélioration envisagée :* Mise en place de tests de scraping automatisés et d'un système d'alerte pour détecter les ruptures de structure. Possibilité d'enrichir la base avec des API officielles si disponibles.

### Volume de données limité pour certains segments :

Certaines combinaisons (ex. : maison avec ascenseur à Marrakech) sont peu représentées dans les annonces, ce qui diminue la fiabilité du modèle pour ces cas spécifiques.

*Amélioration envisagée :* Accumulation progressive de données historiques, data augmentation (par interpolation ou génération synthétique), ou pondération des modèles par fréquence.

## Absence d'intégration géographique avancée :

Les modèles ne tiennent pas compte de la position géographique exacte (latitude/longitude) ou de l'environnement immédiat du bien.

*Amélioration envisagée* : Intégrer une API de géocodage (Google Maps ou OpenStreetMap) pour enrichir chaque annonce avec des coordonnées géographiques et des scores de proximité (écoles, transports, hôpitaux...).

## Dispersion des données liée aux villas et impact sur la précision du modèle

Une limite importante identifiée durant la phase d'évaluation des performances du modèle de machine learning concerne l'augmentation du RMSE (Root Mean Square Error) causée par une forte dispersion dans les prix de certaines catégories de biens, notamment les villas et riads.

Ces types de biens affichent des prix nettement supérieurs et très variables par rapport aux appartements et studios standards, ce qui induit un biais dans l'apprentissage du modèle et dégrade sa capacité à faire des prédictions précises sur la majorité du marché.

*Amélioration envisagée* : Dans les perspectives futures, une approche consisterait à segmenter le modèle par type de bien, ou à entraîner des modèles distincts pour les villas et pour les autres logements. Une autre possibilité serait d'exclure temporairement les villas de l'entraînement initial afin de stabiliser les métriques globales, puis d'intégrer ces catégories via des modèles spécialisés ou une phase de post-traitement.

## Évolutivité limitée du micro-service ML :

Actuellement, juste deux modèles (un pour le marché marocain et l'autre pour le marché français) sont utilisée en production. Il n'est pas possible de comparer plusieurs modèles ou de versionner les prédictions.

*Amélioration envisagée* : Implémenter un gestionnaire de modèles (ex : MLflow) pour tester plusieurs modèles, moniter leur performances, et activer des stratégies A/B.

## Visualisation réservée à l'administrateur :

L'utilisateur standard ne bénéficie pas d'une vision agrégée du marché ou de son historique d'interactions.

*Amélioration envisagée* : Intégrer un mini-dashboard côté utilisateur (ex : moyenne des prix dans sa région) et une fonctionnalité « historique de prédictions » dans MongoDB.

## Absence de support multilingue :

L'application fonctionne uniquement en français, ce qui peut limiter son accessibilité à un public plus large.

*Amélioration envisagée* : Intégrer un système de traduction via Angular i18n pour supporter au moins deux langues : français et anglais.

# Conclusion Générale

## 1 Récapitulation

Ce projet a consisté à concevoir, développer et déployer une plateforme web intelligente de prédition des prix immobiliers, en s'appuyant sur l'analyse automatisée de données issues de plusieurs sources francophones et marocaines. Il a couvert l'ensemble du cycle de vie d'un projet data : de la collecte web (scraping) à la modélisation prédictive, en passant par le traitement, la visualisation et la mise en production via une architecture MEAN.

L'application permet aujourd'hui à un utilisateur de saisir les caractéristiques de son bien immobilier et d'obtenir instantanément une estimation du prix, grâce à des modèles de machine learning entraînés sur des données réelles. L'administrateur peut quant à lui suivre les tendances du marché via un tableau de bord Power BI connecté à la base MySQL.

## 2 Points clés

Plusieurs aspects fondamentaux ont marqué la réussite de ce projet :

- La mise en place d'un pipeline de données automatisé, robuste et orchestré avec Airflow.
- L'utilisation conjointe de plusieurs bibliothèques de scraping (Playwright, Selenium, BeautifulSoup4) pour extraire efficacement des données complexes.
- Le développement de modèles performants (Random Forest, XGBoost) avec des métriques satisfaisantes ( $R^2 > 0.9$  sur les données test).
- La création d'une plateforme MEAN complète, intégrée à un micro-service Python et sécurisée via JWT.
- Une interface utilisateur claire et intuitive, adaptée aux différents profils d'usage (utilisateur / administrateur).
- Un travail collaboratif bien organisé, avec une répartition claire des responsabilités et une méthode agile (Scrum) appliquée sur toute la durée du projet.

### 3 Retour d'expérience

Ce projet a été une expérience formatrice sur les plans technique, méthodologique et humain. Il nous a permis d'approfondir nos compétences dans des domaines variés : traitement de données massives, intelligence artificielle, développement full-stack, intégration de services, déploiement avec Docker, et collaboration agile.

Travailler sur un projet aussi transversal nous a confrontés à des défis concrets : dépendance aux sources tierces, gestion d'erreurs dans le scraping, orchestration des processus, compatibilité inter-services, ou encore gestion de la sécurité et des performances. Ces obstacles ont été surmontés grâce à une communication fluide, une répartition équilibrée des tâches et une veille technologique constante.

### 4 Conclusion

Ce travail démontre la faisabilité et la pertinence d'une solution logicielle permettant de démocratiser l'accès à l'estimation des prix immobiliers grâce à la donnée ouverte et aux technologies modernes. Il représente une base solide sur laquelle il est possible d'itérer, en intégrant de nouvelles sources, en affinant les modèles ou en enrichissant l'interface utilisateur.

Les perspectives d'évolution sont nombreuses : support multilingue, géolocalisation avancée, version mobile, intégration de photos comme input au modèle, ou encore ouverture de l'API à des partenaires tiers.

Ce projet illustre l'intérêt des synergies entre data science, ingénierie logicielle et visualisation intelligente, dans un secteur à fort impact socio-économique tel que l'immobilier.

# Bibliographie

- [1] Bosvieux, Jean. « L’immobilier, poids lourd de l’économie ». *Constructif*, vol. 49, n°1, 2018, pp. 10–14. Fédération Française du Bâtiment.
- [2] Le Bon Coin. « Site officiel de petites annonces immobilières en France ». Disponible sur : <https://www.leboncoin.fr/> (consulté en 2025).
- [3] Le Figaro Immobilier. « Portail d’annonces immobilières en France ». Disponible sur : <https://www.lefigaro.fr/> (consulté en 2025).
- [4] Avito.ma. « Annonces immobilières au Maroc ». Disponible sur : <https://www.avito.ma/> (consulté en 2025).
- [5] Mubawab.ma. « Portail immobilier marocain ». Disponible sur : <https://www.mubawab.ma/> (consulté en 2025).
- [6] Yu, Ying, Wang, Xin, Zhong, Ray Y. et Huang, George Q. « E-commerce logistics in supply chain management: Implementations and future perspective in furniture industry ». *Industrial Management & Data Systems*, vol. 117, n°10, 2017, pp. 2263–2286. Emerald Publishing Limited.
- [7] Goy, Anna, Ardissono, Liliana et Petrone, Giovanna. « Personalization in e-commerce applications ». In : *The Adaptive Web: Methods and Strategies of Web Personalization*, Springer, 2007, pp. 485–520.
- [8] Kaggle. « House Prices - Advanced Regression Techniques ». Disponible sur : <https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques> (consulté en 2025).
- [9] Yu, Ying, Wang, Xin, Zhong, Ray Y. et Huang, George Q. « E-commerce logistics in supply chain management: Implementations and future perspective in furniture industry ». *Industrial Management & Data Systems*, vol. 117, n°10, 2017, pp. 2263–2286. Emerald Publishing Limited.