

MLOps_Immobilier_France

Ce projet vise à mettre en place un pipeline MLOps.

Introduction

L'objectif de ce projet est de développer une solution de prédiction du prix d'un appartement en France en utilisant un modèle de machine learning. Cette solution sera accessible via une API avec FastAPI, authentification des utilisateurs, une base de données SQL pour le backend, et conteneurisée avec Docker. La solution sera testée et monitorée de manière périodique, aura un pipeline CI/CD, et une interface Web Front sera créée pour permettre aux utilisateurs d'utiliser l'outil. Le déploiement sera effectué sur AWS avec un lien public pour se connecter.

Objectifs

- Développer un modèle de machine learning pour la prédiction de prix immobilier en France
- Implémenter des tests unitaires cohérents pour garantir la qualité du modèle
- Créer une API avec FastAPI pour gérer les prédictions et l'authentification des utilisateurs
- Créer une base de données SQL pour stocker les données de l'API en backend
- Conteneuriser l'API avec Docker et isoler la partie API et BDD
- Tester et monitorer de manière périodique l'évolution du projet
- Créer un pipeline CI/CD pour automatiser les tests et le déploiement
- Créer une interface Web Front pour permettre aux utilisateurs d'utiliser l'outil
- Déployer la solution clé en main sur AWS avec un lien public pour se connecter

Spécifications techniques

Modèle de machine learning

- Le modèle de machine learning sera entraîné sur un jeu de données immobilier en France
- Le modèle sera conçu pour prédire le prix immobilier en fonction des caractéristiques d'un bien immobilier (surface, nombre de pièces, quartier, etc.)
- Le modèle sera développé en Python et utilisera des bibliothèques telles que scikit-learn, pandas et numpy
- Les métriques de performance à suivre seront le coefficient R^2 , la MAE et la MSE.

API

- L'API sera développée avec FastAPI
- L'API fournira des endpoints pour la prédiction des prix immobiliers et l'authentification des utilisateurs
- L'API stockera les données dans une base de données SQL en backend
- L'API sera documentée avec Swagger pour permettre une utilisation facile

Base de données

- La base de données sera développée en SQL et stockera les données de l'API en backend
- La base de données sera hébergée sur AWS RDS

Conteneurisation

- L'API sera conteneurisée avec Docker
- Le conteneur sera isolé de la base de données pour garantir la sécurité et la stabilité du système

Tests et monitoring

- Des tests unitaires cohérents seront développés pour garantir la qualité de l'API et du modèle de machine learning
- Les tests seront automatisés et inclus dans un pipeline CI/CD
- L'évolution du projet sera monitorée de manière périodique pour garantir la stabilité du système

Interface Web Front

- Une interface Web Front sera développée pour permettre aux utilisateurs d'utiliser l'outil
- L'interface sera développée en Python avec des bibliothèques telles que Streamlit
- L'interface se connectera à l'API pour obtenir les prédictions de prix immobiliers

Déploiement sur AWS

- La solution sera déployée sur AWS EC2 pour l'API et RDS pour la base de données SQL
- Le déploiement sera automatisé avec un pipeline CI/CD pour garantir la stabilité du système
- Un lien public sera fourni pour permettre aux utilisateurs de se connecter à l'interface Web Front

Conclusion

Ce cahier des charges décrit les spécifications techniques et les fonctionnalités requises pour le développement d'une solution de prédiction de prix immobilier en France avec un modèle de machine learning et une API avec FastAPI.