

Vente de voitures d'occasion au Royaume Uni

Projet CSI méthodes – Algos pour l'IA M1 MIAGE - 2022-23

Contexte

Vous êtes sollicités pour réaliser un système d'information permettant d'aider des garages à vendre des voitures d'occasion. Ce système va permettre de gérer les dépôts des offres de ventes et des offres d'achat. Et il utilisera des algorithmes d'IA pour aider le garagiste à organiser les voitures en catégories et à définir un prix de vente. Et pour parfaire votre anglais, ce système devra s'adresser aux garages et acheteurs du Royaume Uni.

Fonctionnement souhaité de l'application

L'application devra permettre à des garages d'ajouter des offres de vente de voitures d'occasion, apportant un certain nombre de précisions sur le véhicule vendu qui seront accessibles aux clients. D'autres informations ne seront accessibles que par le garage, comme le prix de vente prédit par l'algorithme d'IA, le prix de vente final accepté par un client, et un éventuel commentaire sur la différence entre les deux. Un garage pourra bien sûr modifier une offre sur un véhicule tant que celui-ci n'est pas vendu. Et une offre ne sera plus proposée aux clients passé un certain nombre de jours après sa publication, délai défini par le garage.

Les clients peuvent consulter les offres sur l'application, en faisant des recherches selon divers critères. Et ils peuvent déposer une proposition d'achat pour un véhicule, au prix affiché ou à un prix inférieur. Le garage peut décider à tout moment d'étudier les offres effectuées, et éventuellement d'en accepter une. En cas d'acceptation, toutes les autres offres sont automatiquement refusées.

Un garage peut consulter un bilan sur une période de temps choisie, visualisant pour ses offres, par exemple le nombre de dépôts, le nombre de ventes, les différences entre prix prédit, demandé et final, raffinés éventuellement par différents critères (marque, modèle,...). Les groupes d'alternants ne proposeront qu'un bilan comparant prix prédit et prix final de vente.

L'administrateur de l'application peut également obtenir des bilans, mais pour tous les garages, ou comparant les garages (à faire par les groupes de non-alternants).

Fonctionnalités utilisant l'IA

Un fichier de données vous sera fourni, recensant les ventes d'années précédentes, incluant les caractéristiques des voitures vendues et leur prix de vente. Sur la base de ces données, vous intégrerez dans l'application :

1- Une fonction permettant de suggérer une catégorie pour une voiture mise en vente.

Pour cela, les groupes d'alternants devront implémenter l'algorithme k-means, l'appliquer sur les données fournies et interpréter les résultats pour définir des catégories. Ces résultats seront ensuite utilisés pour automatiquement identifier la catégorie d'une voiture entrée par le garage lors de la mise en vente.

Les groupes de non-alternants devront implémenter l'algorithme k-means, l'appliquer sur les données fournies et appliquer d'autres méthodes de partitionnement issues de bibliothèques existantes sur ces mêmes données. Ils devront ensuite comparer les résultats, choisir la méthode la plus adaptée et interpréter ses résultats pour définir les catégories. Ces résultats seront ensuite utilisés pour automatiquement identifier la catégorie d'une voiture entrée par le garage lors de la mise en vente.

2- Une fonction permettant de prédire les prix de vente d'une voiture.

Pour cela, les groupes d'alternants devront implémenter l'algorithme de construction d'arbres de décision, l'appliquer sur les données fournies et évaluer les résultats. Le modèle construit sera ensuite utilisé pour prédire le prix de vente d'une voiture entrée par le garage lors de la mise en vente.

Les groupes de non-alternants devront implémenter l'algorithme de construction d'arbres de décision, l'appliquer sur les données fournies et appliquer d'autres méthodes de d'apprentissage supervisé issues de bibliothèques existantes sur ces mêmes données. Ils devront ensuite comparer les résultats pour choisir la méthode la plus adaptée. Le modèle construit sera ensuite utilisé pour prédire le prix de vente d'une voiture entrée par le garage lors de la mise en vente.

Caractéristiques techniques

- On vous demande d'implémenter votre base de données sur PostgreSQL (serveur de l'IDMC, n'hésitez pas à utiliser la connexion VPN).
- L'application peut être un site Web, développé dans un langage de votre choix. Ne perdez pas trop de temps sur l'interface graphique, ce n'est pas le but du projet.

Recommandations pour le bon déroulement du projet

- Les durées entre deux rendus sont parfois courtes, ne vous mettez pas dans chaque partie du projet à la dernière minute, et répartissez-vous bien le travail au sein du groupe tout en communiquant sur les choix et avancées. Les parties modélisation et IA peuvent être faites en parallèle.
- Attention, les horaires limites de dépôt sont stricts ! **Aucun délai ne sera accordé.**
- Pensez à l'ensemble des parties et contraintes du projet lorsque vous faites des choix.
- N'attendez pas la dernière semaine pour vous lancer dans l'implémentation.

Évaluation du projet

L'évaluation du projet se fera selon trois axes :

- La démarche globale de conception du SI basée sur une modélisation UML telle que vue en cours de CSI, et le respect des modèles dans l'implémentation.
- La qualité de l'implémentation et de l'application des algorithmes d'IA et (pour les groupes de non-alternants) la justification des choix des algorithmes utilisés.
- L'intégration de l'application des algorithmes d'IA dans l'application.
- Le respect de la modélisation dans l'implémentation.

Échéancier

- **Mercredi 5/10, 8h30** : présentation du projet en cours ; réflexion préalable nécessaire sur la composition des groupes (4 étudiants par groupe), sans mélanger alternants et non-alternants. Interview publique des clients (les enseignants).
- **Jeudi 6/10, 9h** : remise de la composition définitive des groupes par mél.
- **Non-alternants : mardi 8/11, 22h** **Alternants : vendredi 18/11, 22h** : dépôt d'un rapport avec
 1. le compte-rendu des interviews, la liste des acteurs, la liste des principales fonctionnalités par acteur et celles automatiques, et une première version du modèle de données accompagné du dictionnaire des données (utiliser les diagrammes UML appropriés) ;
 2. le compte rendu de l'implémentation, de l'application et des résultats de la méthode de partitionnement.
- **Non-alternants : mardi 22/11, 22h** **Alternants : vendredi 2/12, 22h** : reprendre le rapport précédent, le corriger et le compléter (éventuellement le réorganiser) avec
 1. des modèles permettant de décrire le repérage du domaine, les données à stocker, au moins un cycle de vie d'un objet d'une classe, et le workflow général de l'application ;
 2. le compte-rendu de l'implémentation, de l'évaluation et de l'application de la méthode de prédiction.
- **Non-alternants : mardi 6/12, 22h** : reprendre le rapport précédent, le corriger, et y ajouter la modélisation de quelques fonctionnalités principales, ainsi que la description des tests supplémentaires réalisés avec des implémentations d'algorithmes de partitionnement et d'apprentissage supervisé de bibliothèques existantes.
- **Non-alternants : vendredi 6/1, 22h** **Alternants : vendredi 20/1, 22h** : Dépôt du dossier complet et final reprenant le rapport précédent, avec 3 parties distinctes :
 1. Modélisation de la conception (cf. dépôts précédents, avec modélisation des fonctionnalités principales et des cycles de vie des objets pour les classes appropriées).
 2. Modélisation de l'implémentation, description de l'application obtenue et des choix effectués lors de cette implémentation (en fonction des modélisations, de contraintes, etc.) ; joindre le script SQL de création de la base.
 3. Résumés des résultats des algorithmes d'IA et description de leur intégration dans l'application.
- **Non-alternants : jeudi 12/1, 8h30-13h** **Alternants : jeudi 26/1, 13h-18h** : Soutenance du projet (40 minutes par groupe : 15 min d'exposé + 10 min de démonstration + 15 min de questions et réponses). La présence de tous les membres du groupe est bien sûr obligatoire, et la désignation de qui fait l'exposé, la démo ou répond aux questions sera faite aléatoirement juste avant la soutenance.