

Scraping, Analyse et Restitution des Offres de Tesla Modèle 3 en France

GAVAZZI Adrien - 2024

[Voir website](#)

Introduction.....	2
Contexte et Objectifs.....	2
Nécessité du Scraping Automatisé.....	2
Stockage et Gestion des Données.....	3
Analyse et Visualisation des Données.....	3
Développement d'un Modèle de Classification d'Image.....	3
Interface Utilisateur Intuitive.....	3
Technologies et Architecture du Projet.....	3
Processus Détaillé.....	4
1. Scraping des données.....	4
2. Automatisation du scraping.....	4
3. Analyse des données.....	4
4. IA - Classification d'image.....	4
5. Développement de l'interface web.....	5
Expansion Future du Projet: Intégration d'une Fonctionnalité de Publication Automatisée sur Twitter...6	6
Contexte et Objectif.....	6
Description Technique de l'Extension.....	6
Calcul de l'Indicateur de Fiabilité:.....	6
Création et Publication Automatique de Cartes d'Offre:.....	6
Automatisation et Intégration.....	6
Conclusion et Perspectives.....	7
Synthèse du Projet.....	7
Apport Technologique et Professionnel.....	7
Invitation à Collaborer.....	7
Prochaines Étapes.....	7

Introduction

Contexte et Objectifs

Le marché des véhicules électriques connaît une croissance exponentielle, et Tesla, en tant que leader incontesté dans ce secteur, attire particulièrement l'attention tant des consommateurs que des analystes. Mon projet vise à exploiter cette dynamique en fournissant une analyse approfondie des offres de Tesla modèle 3 disponibles en France, en se basant sur les données extraites du site leparking.fr. L'objectif principal est de développer une application capable de collecter, stocker, analyser et présenter ces données de manière automatisée et intuitive. Ce projet permettra non seulement d'identifier les tendances du marché, mais aussi de fournir une plateforme utile pour les acheteurs potentiels et les passionnés de Tesla.

Nécessité du Scraping Automatisé

La première étape du projet implique le scraping de données, une technique essentielle pour extraire les informations pertinentes des pages web. Utiliser Python couplé à BeautifulSoup pour ce processus permet d'obtenir les détails nécessaires tels que les prix, les kilométrages, les années de modèle, et les lieux des annonces. Cette automatisation est cruciale, car elle permet de recueillir des données à grande échelle et de manière régulière, garantissant ainsi que l'application dispose toujours des informations les plus récentes et les plus précises. L'utilisation de Docker et Make.com pour orchestrer ces scripts de scraping plusieurs fois par jour aide à maintenir la base de données à jour sans intervention manuelle.

Stockage et Gestion des Données

Les données collectées sont stockées dans MongoDB, une base de données NoSQL qui offre flexibilité et évolutivité, permettant de gérer facilement les grands volumes de données générés par le scraping quotidien. Ce choix technologique assure une récupération rapide des données et supporte des opérations de recherche complexes, ce qui est indispensable pour l'analyse et la visualisation des données dans les étapes ultérieures.

Analyse et Visualisation des Données

L'analyse des données est réalisée à l'aide de Jupyter Notebook et Plotly, des outils qui facilitent la visualisation interactive des données. Cette partie du projet est essentielle pour transformer les données brutes en informations compréhensibles et exploitables. Les graphiques et les analyses statistiques aident à déceler des motifs et des tendances, telles que les variations saisonnières des prix ou les modèles les plus vendus, fournissant ainsi des insights précieux pour les décideurs et les consommateurs.

Développement d'un Modèle de Classification d'Image

Le projet intègre également une dimension de machine learning avec le développement d'un modèle de classification d'images pour identifier les modèles de Tesla à partir des photos des annonces. Cette fonctionnalité augmente la valeur de l'application en automatisant la catégorisation des annonces, ce qui améliore l'exactitude des filtres et des recherches utilisateurs sur la plateforme.

Interface Utilisateur Intuitive

Enfin, la présentation des données se fait à travers une interface web développée en React avec Material-UI, offrant une expérience utilisateur riche et réactive. Cette interface permet aux utilisateurs de filtrer et de trier les offres basées sur divers critères, rendant l'application non seulement un outil d'analyse mais aussi une plateforme pratique pour les acheteurs potentiels et les amateurs de Tesla.

Technologies et Architecture du Projet

Python & BeautifulSoup: Utilisés pour le scraping de données.

MongoDB: Base de données NoSQL choisie pour le stockage des données scrapées.

Docker & Make.com: Utilisés pour automatiser et isoler l'exécution du script de scraping.

React & Material-UI: Framework et bibliothèque utilisés pour le développement de l'interface utilisateur.

Jupyter Notebook & Plotly: Outils d'analyse de données pour visualiser les insights extraits des données.

Machine Learning (Python): Utilisation de techniques de classification d'image pour identifier les modèles de voiture.

Processus Détaillé

1. Scraping des données

Script Python:

- Extraction de données: Le script utilise BeautifulSoup pour naviguer et parser le HTML du site leparking.fr. Il extrait les éléments pertinents tels que le modèle, le prix, le kilométrage, l'emplacement des annonces et des informations sur les vendeurs.
- Normalisation des données: Les données brutes sont nettoyées et normalisées, par exemple, conversion des formats de dates et de prix pour assurer la cohérence dans la base de données.
- Stockage des données: Les données nettoyées sont stockées dans MongoDB, ce qui permet un accès facile et rapide pour les analyses ultérieures.

2. Automatisation du scraping

Make.com & Docker:

- Le processus de scraping est automatisé via Make.com qui planifie l'exécution du script trois fois par jour dans un conteneur Docker. Cela garantit que le scraping est effectué dans un environnement contrôlé et isolé, minimisant les risques d'erreurs dues à des environnements d'exécution inconsistants.
- Notifications par email: Après chaque session de scraping, un résumé des logs est envoyé par email pour surveiller l'état et le résultat du scraping.

3. Analyse des données

Jupyter Notebook & Plotly:

- Visualisation des données: Des graphiques interactifs sont créés en utilisant Plotly pour visualiser différentes métriques telles que la distribution des prix, le kilométrage, et la fréquence des annonces.

- Insights statistiques: Le notebook analyse les tendances sur le marché des Tesla modèle 3, offrant des insights comme les variations saisonnières des prix ou des volumes de vente.

4. IA - Classification d'image

Notebook:

- Prétraitement des images: Les images des annonces sont prétraitées pour la classification, incluant des opérations telles que la normalisation, le redimensionnement et l'augmentation de données.
- Single Shot Détection: Utilisation d'un modèle de détection d'objet afin d'ajouter sur l'image l'emplacement de la voiture avant le traitement par le modèle CNN.
- Entraînement du modèle: Un modèle de classification d'images est entraîné pour reconnaître différents modèles de Tesla modèle 3. Le modèle utilise un détecteur à shot unique pour localiser et classer les véhicules dans les images.
- Validation et performance: Le modèle atteint une précision de 0.93, ce qui démontre son efficacité à classer correctement les modèles de voiture à partir des images scrapées.

5. Développement de l'interface web ([Voir website](#))

React & Material-UI:

- Structure de l'application: L'application est structurée autour de React, utilisant Material-UI pour les composants de l'interface utilisateur, offrant une expérience utilisateur riche et réactive.
- Fonctionnalités interactives: L'application permet aux utilisateurs de filtrer et de trier les annonces de Tesla modèle 3 selon différents critères tels que le prix, le kilométrage, l'année et la localisation.
- Visualisation des détails des annonces: Un composant modal détaillé montre des informations plus approfondies sur chaque annonce, incluant des graphiques de prix comparatifs basés sur des modèles de régression stockés et calculés en arrière-plan.

Expansion Future du Projet: Intégration d'une Fonctionnalité de Publication Automatisée sur Twitter

Contexte et Objectif

Dans le prolongement de mon projet initial sur l'analyse des offres de Tesla modèle 3, je prévois d'étendre les capacités de mon application en intégrant une fonction de publication automatique sur Twitter. L'objectif est de partager régulièrement des offres jugées intéressantes basées sur des critères analytiques précis, augmentant ainsi la visibilité et l'utilité de mon application pour un public plus large. Cette extension automatise le flux de travail complet, de la collecte des données à la communication au consommateur, en fournissant des mises à jour en temps réel sur les meilleures offres disponibles.

Description Technique de l'Extension

Développement d'un Script de Sélection Automatisée:

Le nouveau script s'intégrera dans le pipeline existant en examinant quotidiennement les nouvelles offres collectées dans la base de données MongoDB. Il utilisera des fonctions analytiques pour évaluer chaque annonce selon sa 'classe de prix', déterminée par la comparaison du prix de l'annonce à un prix de référence calculé à partir des données historiques des véhicules similaires. Cette évaluation s'appuiera sur des modèles de régression et des analyses statistiques déjà implémentés dans l'interface React du projet.

Calcul de l'Indicateur de Fiabilité:

En plus de la classe de prix, un indicateur de fiabilité sera calculé pour chaque offre. Cet indicateur prendra en compte le volume de données disponibles pour le modèle et l'année spécifiques de la voiture, ainsi que la comparaison des prix des véhicules avec un kilométrage similaire (à plus ou moins 10 000 km). Cela permettra d'assurer que les offres promues ne sont pas seulement intéressantes en termes de prix, mais également soutenues par des données robustes, augmentant ainsi la confiance des utilisateurs dans les recommandations publiées.

Création et Publication Automatique de Cartes d'Offre:

Pour chaque offre sélectionnée comme étant intéressante, une carte d'offre sera générée automatiquement en utilisant Jinja pour le templating en Python. Cette carte inclura des informations clés telles que le modèle, le prix, le kilométrage, une image de la voiture, et un lien vers l'annonce. Les cartes d'offres seront ensuite publiées automatiquement sur un compte Twitter dédié, permettant une diffusion rapide et large des meilleures offres.

Automatisation et Intégration

Le processus entier, de la sélection des offres à leur publication sur Twitter, sera entièrement automatisé. Des tests réguliers et une surveillance du système seront mis en place pour s'assurer que le pipeline fonctionne sans interruption et avec précision. Cette automatisation garantit une efficacité maximale tout en réduisant la nécessité d'une intervention manuelle, ce qui est crucial pour maintenir l'actualité et la pertinence des données publiées.

En intégrant cette nouvelle fonctionnalité de publication sur Twitter, le projet ne se contente pas de fournir des analyses internes, mais devient également une plateforme active de communication et de marketing. Cela non seulement augmente la visibilité des données collectées, mais aussi ajoute une valeur considérable à l'application, la transformant en un outil dynamique et interactif pour les utilisateurs intéressés par l'achat de véhicules Tesla. L'extension envisagée met en lumière l'engagement du projet à innover et à évoluer constamment, répondant aux besoins changeants des utilisateurs et exploitant les technologies pour automatiser et améliorer l'expérience utilisateur.

Conclusion et Perspectives

Synthèse du Projet

Ce projet représente une application complète et intégrée, démontrant une maîtrise avancée des technologies de pointe dans les domaines du scraping web, de l'analyse de données, du machine learning, et du développement web. En rassemblant des données précises et actualisées sur les Tesla Modèle 3 disponibles en France, ce projet ne se contente pas de fournir une plateforme analytique, mais offre également une ressource précieuse pour les acheteurs potentiels et les passionnés de Tesla. Chaque composant du projet, du backend au frontend, a été conçu avec soin pour assurer robustesse, efficacité et facilité d'utilisation, rendant les informations non seulement accessibles, mais aussi engageantes.

Apport Technologique et Professionnel

Grâce à une architecture bien pensée et une mise en œuvre méticuleuse, ce projet illustre mon aptitude à concevoir et à déployer des solutions logicielles complexes qui répondent à des besoins réels du marché. La capacité à intégrer des analyses de données avancées et des fonctionnalités de machine learning dans une application web conviviale montre un équilibre compétent entre compétences techniques et sensibilité utilisateur. Ce projet sert également de preuve de mon engagement envers l'excellence professionnelle et l'innovation technologique, qualités que je suis impatient de mettre au service de nouveaux défis.

Invitation à Collaborer

Je suis passionné par l'utilisation de la technologie pour résoudre des problèmes pratiques et créer de la valeur ajoutée pour les entreprises et les consommateurs. Si vous êtes intéressé par les détails techniques du projet, souhaitez discuter des possibilités d'application dans votre propre contexte d'affaires, ou envisagez d'ajouter une expertise en développement de logiciels et en analyse de données à votre équipe, je vous invite chaleureusement à me contacter. Ensemble, nous pouvons explorer comment des approches similaires pourraient profiter à vos opérations et projets, et comment je pourrais contribuer à réaliser vos objectifs avec autant de passion et de rigueur que j'ai mis dans ce projet.

Prochaines Étapes

Je suis ouvert à des opportunités de collaboration, de consultation, ou d'emploi où je pourrais apporter non seulement mon expertise technique, mais aussi ma capacité à travailler de manière transversale entre différentes équipes et domaines de compétence. Engager une conversation sur ce projet ou sur d'autres idées pourrait être le début d'un partenariat fructueux et innovant.

Je suis impatient de partager avec vous l'enthousiasme et l'expertise que ce projet incarne. N'hésitez pas à me contacter pour discuter de la manière dont nous pouvons travailler ensemble pour transformer les défis technologiques en opportunités.