

Exploitation des données de maintenance pour la construction de profils utilisateurs

Application aux produits de moto, vélo et motoculture

Localisation : Site du Futuroscope, France

Date de début : début 2025 (flexible)

Durée : 6 mois

Encadrants : Mickaël BARON, Brice CHARDIN et Stéphane JEAN (*l'encadrant dont le nom est souligné est l'encadrant référent : baron@ensma.fr*)

Mots clés : analyse de données, qualité des données, apprentissage automatique, clustering, similarité sémantique via des LLM

Contexte du stage

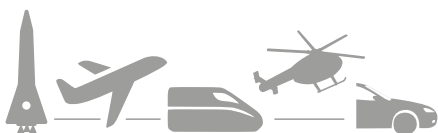
Futurosoft est un éditeur de logiciel du groupe Orisha qui conçoit et développe une solution de gestion (approvisionnement, facturation, ordres de réparation, statistiques, relation client, connecteurs avec des sites marchands, etc.) pour des concessions deux/trois roues, moto, vélo et motoculture de plaisance. Les données issues des logiciels de gestion sont stockées localement par chaque concession, et celles-ci ne sont pas inter-connectées. Par conséquent, les services proposés aux concessions ou aux groupements de concessions sur leurs données ne permettent pas d'effectuer des analyses globales. Des travaux sur la consolidation des données sont en cours au sein de la société pour les centraliser et ainsi fournir de nouvelles perspectives d'analyse.

Objectifs du projet

L'objectif du projet est de s'intéresser aux problématiques d'analyse de données, sur la base de données centralisées, et plus spécifiquement sur l'historique d'un produit (moto, vélo ou motoculture) identifié par un numéro de série. Plusieurs pistes de travaux sont à envisager.

La première piste est relative au **respect des forfaits d'entretien**. En se basant sur les forfaits d'entretien fournis par les constructeurs, il convient d'identifier clairement les opérations d'entretien qui ont été ajoutées ou omises dans les concessions. Cette piste permettra de détecter les écarts, de créer des statistiques précises sur les fréquences des ajouts et omissions d'opérations d'entretien, d'analyser les tendances et motifs communs dans les écarts observés, de mieux comprendre le respect des entretiens des produits par l'ensemble des concessions.

La deuxième piste est relative à la **construction de forfaits depuis les données**. Les forfaits constructeurs sont disponibles uniquement aux concessions qui représentent un constructeur et pour les produits récents. Cette piste s'intéresse donc à reconstruire les forfaits depuis les



données en s'appuyant sur le kilométrage et la durée par exemple afin de fournir des forfaits d'entretien au plus proche du comportement d'entretien des propriétaires des produits.

La troisième piste est relative à la **construction de profils d'entretien** du point de vue des produits ou du client des produits. Cette piste aidera à offrir des services d'entretien personnalisés et adaptés à chaque profil.

Ces trois pistes permettront de proposer à la solution fournie par Futurosoft de nouveaux services d'analyses autour de forfaits explicites (transmis par les constructeurs) et de forfaits implicites (déduits des données) tout en permettant de construire des profils de produits et des clients.

Réalisation attendue

Les objectifs du stage en adéquation avec les livrables sont les suivants.

- Études bibliographiques sur l'usage de techniques d'apprentissage automatique (par exemple du clustering ou l'utilisation de LLM pour réaliser des analyses de similarité sémantique) pour répondre aux trois pistes détaillées précédemment.
- Conception et implémentation de prototypes logiciels pour mettre en œuvre de techniques d'apprentissage automatique sur les données fournies par la société Futurosoft.
- Un rapport de synthèse incluant le rapport de stage décrivant l'étude, l'état de l'art, la conception et l'implémentation du prototype.

À noter que les prototypes développés dans le cadre de ce stage seront des démonstrateurs. Ils ne visent pas être développés à des fins de mise en production. Les données (≈ 100 Go) et les besoins de l'étude seront fournies par Futurosoft.

Profil du candidat

Le candidat doit être en Master 2 en Informatique ou en dernière année de préparation d'un diplôme d'ingénieur spécialité Informatique. Une bonne connaissance du langage de programmation Python et des bibliothèques usuelles d'apprentissage automatique est requise. Un bon niveau en français est fondamental.

Documents à fournir

- Curriculum Vitae;
- Lettre de motivation;
- Notes de Master ou équivalent;
- Tout autre document jugé utile pour enrichir le dossier de candidature (e.g. rapports de projet, dépôt Github).

