## Projet de 2e année

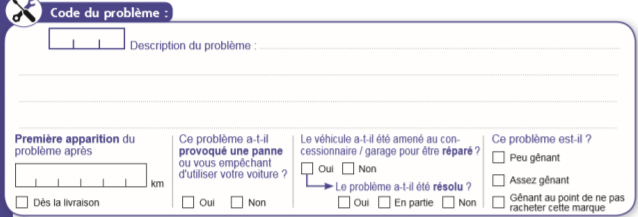
***Sujet – Détection de points d’impact dans des courbes***

Nom du tuteur : Steven Golovkine

Entreprise ou organisme : Renault SA – Direction Qualité et Satisfaction Client – Pôle Data science

1. Contexte

L’enquête NEVQS (New European Vehicle Quality Survey) est une enquête trimestrielle réalisée dans 6 pays (France, UK, Allemagne, Espagne, Italie, Russie) auprès des clients des principaux constructeurs automobile (Renault, Nissan, Toyota, Ford, VW, PSA, Opel, VAZ). L’enquête NEVQS évalue la satisfaction des clients après trois mois et douze mois d'usage du véhicule, durée à partir de laquelle le client peut exprimer un retour d'expérience significatif concernant tous les aspects de son véhicule. Les résultats permettent aux ingénieurs de mieux comprendre les motifs d’insatisfaction client et d'établir des plans de progrès du véhicule. L’enquête recueille notamment, sous forme de texte libre (ie. verbatim), la description des plaintes exprimés par la personne interviewée. Il s’agit de textes « courts », maximum 4 lignes (cf extrait ci-dessous).



*Extrait du questionnaire NEVQS*

1. Sujet proposé

Pour adresser les problèmes aux équipes projets compétentes, les plaintes client doivent au préalable être catégorisés selon la prestation du véhicule concernée : qualité perçue, confort acoustique, consommation de carburant. **Il existe 20 classes de prestations différentes**. Actuellement, la catégorisation est faite « manuellement » via une lecture humaine. Le processus est fastidieux et couteux. L’objet de ce projet est de construire un modèle de classification automatique des plaintes client. Renault fournit les données suivantes :

* un échantillon d’apprentissage d’environ 100 000 verbatim avec la catégorie de prestation associée (1 catégorie par verbatim). Les verbatim sont tous en langues anglaise. Le fichier est au format \*.csv et comprendra les champs suivants : ID, Texte, Label
* un échantillon de test d’environ 20 000 verbatim

1. Méthodologie envisagée

Méthodes d’apprentissage supervisée (éventuellement de type deep learning) développées en langage Python 3.x. Renault fournira un accompagnement méthodologique pour appréhender les spécificités liées à l’analyse des données de type textuelles (pré-processing, embeddings, etc ...).

1. Résultats attendus

Un partage du code sous forme de notebook Jupyter ou script python est attendu. Le rapport du projet devra comporter une comparaison des résultats des différentes méthodes testées.

1. Confidentialité

Du fait de la nature des données, une clause de confidentialité sera signée par le groupe de travail.

1. Bibliographie

* Ramsay and Silverman, Functional Data Analysis, 2005, Springer
* Ramsay, Hooker and Graves, Functional Data Analysis with R and MATLAB, 2009, Springer
* Gijbels, Hall and Kneip, On the Estimation of Jump Points in Smooth Curves, Ann. Inst. Statis. Math., Vol. 51, No. 2, 231-251 (1999)
* Kneip, Poss and Sarda, Functional Linear Regression with Points of Impact, The Annals of Statistics, 20616, Vol. 44, No. 1, 1-30