

Analyse des données : XPORTER

Nom de Projet : XPORTER

Nom du Groupe : AUTOMOBILE

Membres du Groupe :

- Martin BERTHIER <martin.berthier@u-psud.fr>
- Quentin BERTRAND <quentin.bertrand@u-psud.fr>
- Yohann BLACKBURN <yohann.blackburn@u-psud.fr>
- Toufic TALHA <toufic.talha@u-psud.fr>
- Victor LEMAIRE <victor.lemaire@u-psud.fr>

URL du défi : <https://codalab.lri.fr/competitions/652>

Dossier GitHub du projet : <https://github.com/Feynallein/automobile>

→ Pourquoi XPORTER

Nous avons choisi XPORTER tout d'abord car il s'agissait du sujet de projet qui nous plaisait le plus, de plus il nous permet de mettre en application ce que l'on a vu en cours du niveau de tout ce qui touche au *training data* tout en l'appliquant à des cas assez concrets.

→ Explication des données et description du problème

Nous avons à notre disposition plusieurs jeux de données pour répondre au problème qui est le suivant : « Prédire le nombre de voitures qui passeront à une date et à une heure donnée, ainsi que des informations météorologiques supplémentaires », nous avons donc plusieurs jeux de données pour la formation d'une matrice de données X_{train} de la dimension $\text{num_training_samples} \times \text{num_features}$ et d'un tableau y_{train} d'étiquettes de la dimension $\text{num_training_samples}$.

→ Approche choisie

Pour réaliser ce projet, nous avons fait une grande sélection de modèles sur le site scikit-learn, puis nous avons effectué une grande batterie de tests sur nos jeux de données avec chacun de ces modèles afin d'obtenir les valeurs les plus optimales et nous permettre de trouver le modèle réalisant les meilleurs résultats en utilisant nos données.

→ Description des binômes

Tout d'abord, notre équipe Automobile, ne comportant que 5 membres, possède donc un monôme et donc a dû choisir un membre, ainsi qu'une partie du projet, à assigner à ce monôme. Le groupe de *preprocessing* est composé de Toufic ainsi que de Victor et a travaillé sur le fichier *preprocess.ipynb* du *repository*. Le second groupe, celui du *model* est composé de Martin et de Quentin, il a donc effectué tout leur recherche sur le fichier *model.ipynb* contenu également dans le *repository* de notre projet. Et enfin la dernière partie, *visualization*, est traitée par le monôme de l'équipe formé par Yohann, qui lui a composé sur le fichier *visualization.py* se situant, comme les précédents, dans le *repository*.

→ Premier résultats

Nous avons effectué tous ce qui était demandé lors des séances passées et avons à réussir des résultats plutôt concluants lorsque chaque groupe de l'équipe travaillait, si on peut dire, indépendamment. Toutefois, la difficulté majeure, que nous n'avons pas réussi à surmonter, est la fusion des trois parties qui depuis le début étaient indépendantes. En effet, nous n'avons jusqu'ici réussi à obtenir aucuns résultats probants avec la fusion de nos trois parties de notre projet, nous empêchant donc de proposer des graphiques à intégrer dans ce *porposal*. Nous espérons régler ce problème au plus vite pour le bon avancement de ce challenge.

Références :

- Scikit-learn
- Sklearn
- Codalab
- Charlearn