

Rapport de projet - Groupe Autobus 2

Rapport de projet

Team name/X-porters

Group name: Autobus2

members : Marion pobelle, Aimane Benammar, Garance Roux, Romain soliman, Elie Raspaud, lucas foissey

(J'ai travaillé seul sur cette partie de rédaction du rapport, en effet j'ai essayé de contacter le reste du groupe mais ils n'ont pas répondu en effet j'ai travaillé avec Marion pobelle sur la partie du modèle)

Introduction :(fait par Aimane et Marion)

Notre groupe fait parti du projet X-porters. Le prof nous a désigné le groupe et le sujet sur lequel on travaille .

L'objectif de ce projet est de pouvoir réaliser et faire une régression et une classification des données . . Ce projet nous a appris le fonctionnement et l'importance d'une régression . et bien évidemment *la visualisation* des données qui est une technique d'exploration et d'analyse des données numériques à l'aide de graphiques.

Les informations qu'on dispose pour ce projet sont des fichiers . Pour explorer ces données nous avons à notre disposition un code de base que nous avons modifié pour répondre aux problèmes abordés . Pour le faire preuve . Le prof nous a séparés en 3 groupes et que chaque groupe a travaillé sur une tâche précise que nous allons expliciter par la suite .

Pre-processing : (?????????)

Model :(Aimane et Marion)

l'objectif était de trouver les hyper-paramètres et les modèles qui donnent un résultat optimal sur un ensemble de tests afin de trouver les prédictions qui sont les plus fiables .

Pour le faire preuve . il fallait tester plusieurs modèles de régression, trouver les hyper-paramètres les plus optimaux des deux meilleurs et expliquer chaque modèle

Avant de commencer cette partie . on a fait des recherches sur internet pour comprendre le concept

et nous avons également regardé des vidéos sur YouTube et on a lu certains livres qui parlent de cette partie afin de mieux comprendre le principe de la régression.

on a commencé par hyper-parameters du TP2 pour chercher le meilleur modèle qui convient. Puis on a essayé de trouver le meilleur hyper-parameters pour améliorer la performance du programme. Et pour déterminer lequel est le meilleur et le plus performant nous avons examiné les performances obtenues sur l'ensemble de données de validation mais également sur l'ensemble de données d'apprentissage pour éviter un sur-ajustement nous avons utilisé donc la fonction `train_test_split` de `scikit-learn`. Et on a gardé les deux modèles qui ont le meilleur score sur l'ensemble de données valide

les hyper-paramètres les plus importants d'un arbre de décision sont `max_depth`, `min_samples_split`, `min_samples_leaf` et `max_features`.

Dans le cas de `max_depth`, plus l'arbre est profond, plus il y a d'informations sur les données à prendre en considération. Cependant, si la valeur de `max_depth` est trop grande, l'arbre de décision adaptera les données.

`min_samples_split` spécifie le nombre minimum d'échantillons requis pour diviser un nœud. Cet hyper-paramètre empêche également une situation de sur-ajustement et réduit le temps d'entraînement.

`min_samples_split` et `min_samples_leaf` se complètent et travaillent ensemble pour limiter le nombre de divisions et de feuilles dans l'arbre.

Enfin, `max_features` représente le nombre de fonctionnalités à considérer lors de la recherche du meilleur fractionnement. Cette hyper-paramètre peut être utilisé pour contrôler et éviter un sur-ajustement car moins de fonctions sont prises en compte

Nous avons utilisé `RandomizedSearchCV` pour essayer de trouver les meilleurs hyper-paramètres valeurs de notre modèle sur une grille de valeurs hyper-paramètres choisies en regardant les informations données par `DataManager`. Nous avons pas utilisé l'exact valeurs fournies par les informations pour empêcher notre modèle de situations de sur-ajustement.

Visualization : (???????)

Conclusion :(fait par Aimane et Marion)

Ce projet nous a permis d'apprendre à analyser, à organiser et à interpréter les données. Il nous a permis également d'utiliser des insights pour des questions et des problèmes du monde réel. En effet Chaque groupe d'entre nous à bien avancé dans sa propre partie . en attendant que chacun entre nous puisse améliorer ce qu'on a pu faire jusqu'au aujourd'hui..

References:

- [1] <https://www.youtube.com/watch?v=NUXdtN1W1FE>
- [2] <https://www.youtube.com/watch?v=a9UrKTVEeZA>
- [3] <https://www.youtube.com/watch?v=erfZsVZbGJI>
- [4] https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html