

Projet - Réalisation d'une interface d'analyse de données par apprentissage supervisé

Pegdwendé Nicolas Sawadogo
pegdwende.sawadogo@univ-lyon2.fr

13 octobre 2021

1 Objectifs et cahier des charges

Il s'agit dans ce projet de réaliser une interface graphique permettant d'appliquer des algorithmes de prédiction sur un jeu de données. Les utilisateurs de votre outil sont supposés ne rien connaître en programmation python. Le cahier des charges est le suivant :

1. L'outil prendra en entrée un jeu de données annoté au format csv (avec une virgule pour séparateur). Le jeu de données comportera des noms de colonnes à la première ligne (des exemples de jeux de données vous sont fournis).
2. L'application permettra à l'utilisateur de définir les variables prédictives, et la variable cible.
3. En fonction du type de la variable cible (catégorielle ou numérique), l'application proposera au moins 3 algorithmes de classification, ou 3 algorithmes de régression que l'utilisateur peut appliquer.
4. L'utilisateur pourra choisir un ou plusieurs de ces algorithmes à appliquer aux données. Pour chaque algorithme à appliquer, les valeurs optimale des principaux hyper-paramètres pourront être soit définies par l'utilisateur, soit identifiées automatiquement.
5. Pour chacun des modèles à appliquer, on utilisera une validation croisée et on fournira en sortie les métriques d'évaluation , le temps de calcul, et surtout une/des figure(s) permettant de percevoir de façon synthétique la différence entre les prédictions et la réalité (Vous pouvez par exemple vous servir par exemple d'une représentation factorielle dans le cas d'une classification).
6. Vous devrez utiliser l'une des deux bibliothèques « Dash » ou « Bokeh » pour la conception de l'interface graphique et y intégrer les visualisations à l'aide de « Plotly »(pour les rendre interactives).

2 Livrables et calendrier

En guise de livrables, vous devrez fournir un rapport d'une part et le code source (commenté) d'autre part. Le rapport devrait tenir sur 5 à 8 pages et *pourra* présenter la répartition du travail, l'architecture de l'application, un mini guide d'utilisation, ainsi que tout autre point que vous jugerez utile.

L'ensemble des livrables est attendu pour le dimanche **28 novembre 2020** au plus tard. Votre travail se conclura par une soutenance à faire courant décembre 2020 (date à préciser ultérieurement).

3 Critères d'évaluation

Vous serez évalués principalement sur :

- Respect du cahier des charges
- Fonctionnement, fluidité et facilité de prise en main de l'outil
- Capacité à identifier les meilleurs modèles (et paramétrages)
- Pertinence et utilité des sorties
- Présentation du rapport et de l'application (interface et code source)

4 Quelques ressources

- <https://dash-gallery.plotly.host/Portal/>
- <https://docs.bokeh.org/en/latest/docs/gallery.htmlserver-app-examples>