

Projet Machine Learning

Pour ce projet, vous allez implémenter une solution pour un problème (régression ou classification) de traitement de données de Kaggle ou ailleurs. Il s'agit d'un projet **en groupe de 2**, qui compte pour 25% de la note finale du cours. Vous êtes libre du choix de projet. Vous devez rendre votre code Python (Utiliser Google Colaboratory <https://colab.research.google.com/?hl=fr>). Le notebook contiendra les parties suivantes :

Décrire brièvement le projet.

Quels sont les objectifs ? Qu'est-ce qui constituerait une bonne solution ou un bon résultat ? Comment prévoyez-vous vous y prendre, quelles idées ou intuitions avez-vous pour parvenir aux objectifs ?

Description des données

Maintenant que vous avez choisi un projet et téléchargé les fichiers de données, vous allez regarder les données avec lesquelles vous devrez travailler un peu plus proche. Discutez des données, les types d'attributs, et les propriétés statistiques des données. Quelles difficultés se présentent ? (ex. : bruit, fléau de dimensionnalité, informations manquantes, déséquilibre des classes, etc.). Implantez des algorithmes de prétraitement des données afin de corriger ces difficultés. Décrivez ces algorithmes et discutez de leurs résultats. L'objectif ici n'est pas de faire une grande liste de statistiques sur les données, mais d'en tirer des leçons pour guider la réalisation du reste du projet.

Discutez également de ce qui manque dans vos données. Quelles informations additionnelles voyez-vous comme nécessaire ou utile à avoir pour réaliser votre projet mais manquent au jeu de données. Où et comment pourriez-vous obtenir ces informations supplémentaires afin d'enrichir votre jeu de données ? Discutez également comment les différents jeux de données vont interagir ensemble dans votre système. Quelles sont les valeurs en commun pour lier les données ensemble ? Comment allez-vous utiliser les données ? Comment allez-vous structurer les données, les stocker, les charger en mémoire ? Discutez également de la procédure de tests que vous envisagez.

Traitement des données

Décrivez l'algorithme que vous avez choisi d'implanter. Décrivez, d'un point de vue technique, comment il fonctionne et les composantes clefs. Justifiez votre choix et vos décisions de design. Pourquoi avez-vous choisi cet algorithme ? Quelles alternatives avez-vous considérées, et pourquoi celle-ci est-elle émergée comme la meilleure ?

Résultat final

Présentez et analysez vos résultats. Quelles sont les performances de votre solution ? Dans quelles conditions est-ce qu'elle retourne les résultats idéaux et dans quelles conditions est-ce qu'elle échoue ? Présentez des statistiques pertinentes (taux de succès, précision, rappel, temps moyen de calcul, complexité algorithmique, etc.) et des études de cas comme exemples spécifiques. Ne vous limitez pas à un score global pour votre système, mais analysez en détails dans quels cas (classes, attributs, etc.) il fonctionne mieux ou moins bien, et qu'est-ce qui cause ces différences.

Plagiat strictement interdit : Ne partagez pas votre code source à quiconque.