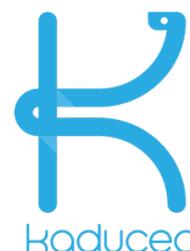




Institut de Recherche
en Informatique de Toulouse
CNRS - INP - UT3 - UT1 - UT2J



Stage de fin d'études - Profil Master 2

Entreprise : Kaduceo [1]

Laboratoire : Institut de Recherche en Informatique de Toulouse (IRIT) – UMR 5505 – Equipe SIG (Systèmes d'Informations Généralisées) [2]

Lieu du stage : Kaduceo (96 avenue Jules Julien)

Responsables du stage : Matthieu Ortala (Kaduceo), Noemie Salaun-Penquer (Kaduceo), Julien Aligon (IRIT), Chantal Soulé-Dupuy (IRIT)

Contexte du stage : La mission s'inscrit dans le cadre d'une collaboration entre l'entreprise Kaduceo et l'équipe SIG de l'IRIT. Le cœur de métier de Kaduceo porte sur l'aide à la décision pour les données de santé, en particulier sur les hospitalisations et consultations. Ces données ont, en particulier, besoin d'être catégorisées automatiquement afin de mieux prendre en charge les patients et d'améliorer les conditions de travail des professionnels de santé (par exemple dans les Centres Hospitaliers). A cette fin, l'utilisation de modèles prédictifs peut être une solution mais ils sont généralement difficiles à appréhender. En effet, le manque d'explications sur ces modèles mène à un effet « black box », entraînant une possible perte de confiance de l'utilisateur. Afin d'éviter ce problème, quelques stratégies d'explications de prédictions ont été proposées dans la littérature [3, 4, 5].

Objectifs du stage : L'objectif du stage porte sur la conception et la réalisation d'une nouvelle méthode d'explication pour les modèles prédictifs. Le but sera d'identifier, dans le dataset, le sous ensemble d'instances (le plus réduit possible) expliquant au mieux le modèle. Ces instances pourraient être vues comme des exemples caractéristiques du modèle. Ainsi, l'utilisateur n'aurait plus qu'à analyser ces instances afin de comprendre et valider son modèle.

Déroulement du stage : Les principales étapes du stage sont les suivantes :

- Analyser et préparer les données du cas d'usage
- Appliquer des algorithmes de prédictions
- Concevoir et implémenter l'algorithme de sélection d'instances pour l'explication

Il est à noter que le bon déroulement du stage pourra mener à une possible thèse CIFRE par la suite.

Compétences souhaitées : Le stagiaire devra avoir un intérêt certain pour l'analyse de données et le machine learning. De fortes compétences en Python sont attendus. Le stagiaire devra également disposer d'une capacité à s'exprimer auprès de non-spécialistes en informatique.

Rémunération : Ce stage est rémunéré pour une durée de 6 mois, aux alentours de 900€ par mois.

Candidature : les candidat(e)s sont priés d'envoyer au plus tôt un CV, leurs relevés de notes et une lettre de motivation relative au sujet. Ces documents sont à envoyer au format PDF par email à

matthieu.ortala@kaduceo.com, noemie.salaunpenquer@kaduceo.com, julien.aligon@irit.fr, et chantal.soule-dupuy@irit.fr

Références :

[1] <https://kaduceo.com>

[2] <https://www.irit.fr/-Equipe-SIG->

[3] Ribeiro, Marco Tulio and Singh, Sameer and Guestrin, Carlos, "Why Should I Trust You?": Explaining the Predictions of Any Classifier, SIGKDD 2016

[4] Gabriel Ferrettini, Julien Aligon and Chantal Soulé-Dupuy, "Explaining single predictions : a faster method", SOFSEM 2020 (To Appear)

[5] Gabriel Ferrettini, Julien Aligon et Chantal Soulé-Dupuy, « Un cadre d'aide à l'exploitation des résultats de prédictions », à destination d'experts de domaine, Inforsid 2019