

IA de Confiance

Durée : 2 heures

Remarque importante : Ce sujet est assez ouvert vous devez imaginer des solutions et ne pas restreindre votre imagination.

Contexte : Une entreprise financière utilise un réseau de neurones profonds pour automatiser le processus d'octroi de prêts. Cependant, l'opacité de ce modèle pose des défis en termes d'explicabilité, car les décisions de prêt doivent être transparentes et compréhensibles pour les clients ainsi que les régulateurs financiers. Vous êtes chargé(e) d'évaluer et d'améliorer la transparence de ce modèle en utilisant des techniques d'explicabilité telles que SHAP, LIME, Génération d'un modèle de substitution (surrogate model), et une analyse contrefactuelle avec l'outil What-If Tool (WIT).

Enoncé du Problème : L'entreprise constate une augmentation du nombre de demandes de prêts rejetées sans explication claire pour les clients. Certains clients se plaignent également de biais dans les décisions de prêt. Pour résoudre ces problèmes, l'entreprise souhaite améliorer la transparence et l'explicabilité de son modèle d'octroi de prêts basé sur un réseau de neurones profonds.

Questions :

1. **Analyse des Caractéristiques les Plus Importantes :** Quelle méthode utiliser pour identifier les caractéristiques les plus importantes utilisées par le modèle dans la prise de décision d'octroi de prêts. En considérant des caractéristiques (features) de votre choix (âge, niveau de revenu, etc.), esquissez un dessin qui montre la visualisation des caractéristiques et leurs effets sur la prise de décision (positives ou négatives).
2. **Interprétation des Prédictions Individuelles :** Montrer avec des exemples de votre choix comment peut-on utiliser une méthode comme LIME pour réaliser des explications individuelles (locales) du modèle. Comment ces explications peuvent-elles aider les clients à comprendre les raisons derrière les décisions de prêt qui leur sont spécifiques ?
3. **Détection et Correction des Biais :** Montrer comment utiliser une analyse contrefactuelle avec l'outil WIT pour détecter les biais potentiels dans le modèle d'octroi de prêts. Comment pouvez-vous ajuster les caractéristiques des demandes de prêt pour minimiser ces biais tout en maintenant la précision du modèle ?
4. **Stratégies d'Amélioration Continue :** Élaborez des stratégies pour intégrer la transparence et l'explicabilité dans le processus de développement continu du modèle d'octroi de prêts. Comment pouvez-vous garantir que le modèle reste transparent et équitable à mesure que de nouvelles données sont collectées et que le modèle est mis à jour ? Cette question est générale. Elle ne dépend pas d'une méthode particulière mais représente une sorte de stratégie d'entreprise pour assurer la transparence et l'explicabilité des modèles d'IA. Cela concerne aussi bien le choix des données de départ, les processus de test des modèles qui intègrent l'explicabilité, le type de restitutions des techniques d'explicabilité à destination des utilisateurs, etc.