



Avis de Soutenance

Monsieur Steve BELLART

Informatique et applications

Soutiendra publiquement ses travaux de thèse intitulés

Calcul d'explications formelles pour les modèles d'apprentissage automatique à ensemble d'arbres

dirigés par Monsieur Gilles AUDEMARD et Monsieur Jean-Marie LAGNIEZ

Soutenance prévue le **lundi 18 décembre 2023** à 10h00

Lieu : UFR des Sciences Jean Perrin Rue Jean Souvraz SP 18, 62300 Lens

Salle : des thèses

Composition du jury proposé

M. Gilles AUDEMARD	Université d'Artois	Directeur de thèse
Mme Wassila OUERDANE	Centrale Supélec, Université Paris-Saclay	Rapporteuse
M. Nicolas MAUDET	Sorbonne Université	Rapporteur
M. Jean-Marie LAGNIEZ	Université d'Artois	Co-directeur de thèse
Mme Marie-Jeanne LESOT	Sorbonne Université	Examinatrice
M. Pierre MARQUIS	Université d'Artois	Examineur

Résumé :

L'intelligence artificielle (IA) permet aux machines de prendre des décisions autonomes. Au fil du temps, l'accent a été mis sur la capacité des machines à apprendre et à prendre des décisions par elles-mêmes, en utilisant des méthodes d'apprentissage automatique. Cependant, un défi majeur posé par les systèmes obtenus est leur nature de boîte noire, rendant difficile la compréhension de leur processus de prise de décision. Cette opacité devient particulièrement problématique dans des domaines tels que la médecine, la finance, le recrutement et le système judiciaire, où des décisions inexplicables peuvent avoir des implications majeures. Reconnaissant l'importance de la transparence, l'union européenne (UE) a mis en œuvre des réglementations telles que le règlement général sur la protection des données (RGPD) et a proposé de nouvelles directives en 2021 pour garantir une utilisation éthique et transparente de l'IA. La demande de clarté favorise non seulement la confiance, mais s'aligne également sur les exigences réglementaires. Dans cette thèse, nous examinons l'explication de prédictions issues de modèles à ensembles d'arbres, en particulier les forêts aléatoires et les arbres boostés. Bien que ces modèles soient basés sur des arbres de décision, considérés comme naturellement interprétables, leurs prédictions peuvent être difficiles à expliquer. Dans le paysage actuel de l'IA, la confiance, l'éthique, l'interprétabilité et l'explicabilité sont primordiales. Nous proposons des solutions adaptées pour expliquer ces deux types de modèles. Pour les forêts aléatoires, nous avons défini une nouvelle notion d'explication formelle qui est à la fois concise et calculable en temps polynomial. Pour les modèles d'arbres boostés, tels que ceux appris avec des bibliothèques comme LGBM et XGBoost, nous proposons une méthodologie permettant d'expliquer les résultats obtenus dans les problèmes de régression, en cherchant à expliquer pourquoi une prédiction se situe dans un certain intervalle. Enfin, nous introduisons des notions d'explications personnalisées exploitant les bases de connaissances des utilisateurs et leurs préférences.