



9 Juin 2023

DTNum - Pôle données

Séminaire IA

*L'IA explicable: des techniques pour des décisions
transparentes*

13h00
13h10
13h25
14h25
14h40
15h45
16h00
17h00
17h30
18h00

Ouverture – Su Yang *Directeur du pôle données* – 10 min

Introduction – Jérôme Fournel *Directeur Général* – 15 min



Arthur Maillart (*ex Université de Lyon - Detralytics*) et **Christian Yann Robert** (*ENSAE*)

Éclairer l'IA : l'explicabilité des modèles de machine learning pour sonder la boîte noire et applications en assurance



PAUSE

Astrid Bertrand (*Télécom Paris – Autorité de contrôle prudentiel et de résolution/Banque de France*)

Concevoir des outils d'explicabilité centrés sur les utilisateurs (une perspective IHM Interaction Homme Machine sur l'XAI)



PAUSE

Table ronde : Su Yang (DTNUM) / Céline Castets-Renard (Université d'Ottawa) / Gilles Clabecq (Service de la sécurité juridique et du contrôle fiscal) / Elodie Quezel (Signaux Faibles) / Arthur Maillart (*ex Université de Lyon – Detralytics*)



Mot de conclusion – Su Yang *Directeur du pôle données* – 15 min

Éclairer l'IA : l'explicabilité des modèles de machine learning pour sonder la boîte noire et applications en assurance

Les bonnes performances des modèles de machine learning ont favorisé leur utilisation dans l'industrie. Cependant, ce gain en performance s'est fait au détriment de l'interprétabilité. Aujourd'hui, l'explicabilité des modèles de machine learning est devenue un sujet clé. Les raisons sont multiples : améliorer nos connaissances, détecter d'éventuels biais dans les données, respecter une réglementation ou encore accroître notre confiance dans des modèles qui prennent des décisions automatiques. Cette présentation a pour objectif de clarifier le concept d'explicabilité des modèles de machine learning. Nous nous concentrons sur les méthodes basées sur des ensembles d'arbres, c'est-à-dire des méthodes type random forest ou gradient boosting avec des arbres en weak learners. Pour cela, nous définissons les différents termes utilisés dans ce domaine, puis nous proposons une taxonomie des différentes approches. Enfin, nous illustrons l'importance de l'explicabilité en examinant des exemples d'application dans des exemples tirés du monde de l'assurance.

Christian Robert



Christian est professeur de statistiques actuarielles et de mathématiques à l'Institut Polytechnique de Paris (ENSAE). Il a une expérience en tant que directeur de recherche du Laboratoire de Finance et d'Actuariat à l'ISFA (Lyon) et en tant que Responsable scientifique de la Chaire d'excellence Data Analytics and Models for Insurance (DAMI). Christian est membre agrégé de l'Institut des Actuaire (Paris). Il a publié plus de 60 articles scientifiques.

Arthur Maillart



Arthur est actuaire de formation. Il a effectué sa thèse CIFRE sous la direction de Christian Robert. Ses travaux portaient sur l'interprétabilité des modèles de machine learning. Il s'est en particulier intéressé aux ensembles d'arbres (p.e. random forest, gradient boosting). Pour mettre en œuvre ces méthodes, il s'est concentré sur des cas concrets tels que la tarification auto et l'estimation de la fonction d'indice de queue avec de modèles de Machine Learning. Arthur est par ailleurs consultant en actuariat et machine learning et continue ses travaux sur l'interprétabilité des modèles au sein de son cabinet.

Concevoir des outils d'explicabilité centrés sur les utilisateurs (une perspective IHM Interaction Homme Machine sur l'XAI)

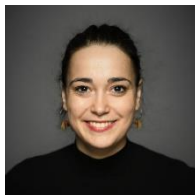
Cette présentation concerne la conception centrée sur l'utilisateur et l'explicabilité dans le domaine de l'IA. Elle aborde les différents aspects de l'explicabilité, notamment pour qui elle est nécessaire, les différents degrés d'explicabilité et les "coûts" associés à l'interprétabilité de l'IA. Celle-ci explore également les perspectives légales en matière d'explicabilité, notamment la contrainte de "human oversight" dans l'IA Act. La présentation décrit les différentes interfaces homme-machine (IHM) qui permettent de rendre l'IA interprétable, ainsi que les différents types d'interaction et les méthodes d'évaluation de l'explicabilité par les utilisateurs. Elle met en évidence les biais cognitifs qui peuvent affecter les prises de décision lors de l'utilisation de l'IA. Enfin, la présentation présente une application concrète de l'explicabilité dans le domaine des robots-conseillers en assurance-vie

Astrid Bertrand



Astrid Bertrand est doctorante à Télécom Paris – Institut Polytechnique de Paris, et à l'Autorité de contrôle prudentiel et de résolution (ACPR/Banque de France), sous la direction de David Bounie (Télécom Paris) et Winston Maxwell (Télécom Paris). Elle travaille sur la mesure de l'efficacité des systèmes actuels de lutte contre le blanchiment d'argent et sur les différentes approches de l'explicabilité de l'utilisation de l'IA dans la lutte contre le blanchiment d'argent. Elle est titulaire d'un diplôme d'ingénieur de Centrale Lyon et d'un MSc en durabilité et innovation sociale de HEC.

Elodie Quézel



Directrice opérationnelle de « Signaux Faibles » - un service numérique interministériel ayant pour objectif de soutenir les entreprises en difficulté. Le service repose sur un algorithme qui cible les entreprises afin de mettre en place des actions d'accompagnement le plus rapidement possible. Elodie Quézel est intrapreneuse publique. Elle a participé à l'expérimentation alors défi d'intérêt général et contribué le passage à l'échelle nationale. Diplômée de Sciences Po, passionnée par le service public et l'entrepreneuriat, après l'avoir observé à l'œuvre en Europe et en Asie, elle a construit une expertise sur les transformations par le numérique, en contribuant lors d'une 1er expérience à la préfiguration et mise en place opérationnelle du GIP Grande Ecole du Numérique.



Céline Castets-Renard

Céline Castets-Renard est titulaire de la Chaire de recherche de l'Université sur l'intelligence artificielle responsable à l'échelle mondiale, chercheure régulière au Centre de recherche en droit, technologie et société de l'Université d'Ottawa, et professeure titulaire à la Section de droit civil de la Faculté de droit.

Ses travaux portent sur le droit et la régulation du numérique dans des domaines variés de droit privé allant de l'impact de la technologie sur les contrats et la responsabilité civile à la propriété intellectuelle, à la protection des données personnelles, au commerce électroniques, en passant par les questions éthiques liées à la régulation des voitures autonomes et la cybersécurité.

Gilles CLABECQ



De l'audit des systèmes d'information des entreprises, aux travaux de prévision des recettes fiscales ou d'analyse de l'impact budgétaire des mesures fiscales présentées en loi de finances, Gilles Clabecq a construit un parcours professionnel à la DGFIP orienté sur la compréhension, le traitement, l'analyse et la valorisation de la donnée. Chef du bureau SJCF-1D (ex Mission requêtes et Valorisation) depuis avril 2022, il pilote la mise en œuvre du data-mining dans les opérations de ciblage des opérations de contrôle fiscal à la DGFIP.

Arthur Maillart



Arthur est actuaire de formation. Il a effectué sa thèse CIFRE sous la direction de Christian Robert. Ses travaux portaient sur l'interprétabilité des modèles de machine learning. Il s'est en particulier intéressé aux ensembles d'arbres (p.e. random forest, gradient boosting). Pour mettre en œuvre ces méthodes, il s'est concentré sur des cas concrets tels que la tarification auto et l'estimation de la fonction d'indice de queue avec de modèles de Machine Learning. Arthur est par ailleurs consultant en actuariat et machine learning et continue ses travaux sur l'interprétabilité des modèles au sein de son cabinet.