COLLEGE ED MATHEMATIQUES
DOCTORAL STIC
PAYS DE LA LOIRE

Nantes Université

THÈSE DE DOCTORAT DE

NANTES UNIVERSITÉ

ÉCOLE DOCTORALE N° 641
Mathématiques et Sciences et Technologies du numérique de l'Information et de la Communication
Spécialité : *Informatique*

Par
Gaëlle JOUIS

Explicabilité des modèles profonds et méthodologie pour son évaluation

Application aux données textuelles de Pôle emploi

En vue de la soutenance de Thèse à Nantes Université, le 14 Février 2023
Unité de recherche : umr6004 – LS2N
Thèse N° : 2019-0989

Rapporteurs avant soutenance :

Céline HUDELOT Professeure des universités - CentraleSupélec
Philippe LENCA Professeur des universités - IMT Atlantique

Composition du Jury :

Examinateurs :	Richard DUFOUR	Professeur des Universités - Nantes Université
	Gilles VENTURINI	Professeur des Universités - Ecole Polytechnique de l'Université de Tours
Dir. de thèse :	Harold MOUCHERE	Professeur des Universités - Nantes Université
Co-dir. de thèse :	Fabien PICAROUGNE	Maitre de conférences - Nantes Université

Invité(s) :

Alexandre HARDOUIN Scientifique des données - Pôle Emploi

COLLEGE ED MATHEMATIQUES
DOCTORAL STIC
PAYS DE LA LOIRE

Nantes Université

Titre : Explicabilité des modèles profonds et méthodologie pour son évaluation : application aux données textuelles de Pôle emploi

Mot clés : Apprentissage profond, Explicabilité, Réseaux de Neurones, Intelligence Artificielle

Résumé : L'intelligence Artificielle fait partie de notre quotidien. Les modèles développés sont de plus en plus complexes. Les régulations telles que la Loi Pour une République Numérique orientent les développements logiciels vers plus d'éthique et d'explicabilité. Comprendre le fonctionnement des modèles profonds a un intérêt technique et humain. Les solution proposées par la communauté sont nombreuses, et il n'y a pas de méthode miracle répondant à toutes les problématiques. Nous abordons la question suivante : comment intégrer l'explicabilité dans un projet d'IA basé sur des techniques d'apprentissage profond ?

Après un état de l'art présentant la richesse de la littérature du domaine, nous présentons le contexte et les prérequis de nos travaux. Ensuite nous présentons un protocole d'évaluation d'explications locales et une méthodologie modulaire de caractérisation globale du modèle. Enfin, nous montrons que nos travaux sont intégrés à leur environnement industriel.

Ces travaux résultent en l'obtention d'outils concrets permettant au lecteur d'appréhender la richesse des outils d'explicabilité à sa disposition.

Title: Explainability of deep models and methodology for its evaluation: application to textual data from Pôle emploi

Keywords: Deep learning, Explainability, Neural Networks, Artificial Intelligence

Abstract: Artificial intelligence is part of our daily life. The models developed are more and more complex. Regulations such as the French Law for a Digital Republic (Loi Pour une République Numérique) are directing software development towards more ethics and explainability. Understanding the functioning of deep models is of technical and human interest. The solutions proposed by the community are numerous, and there is no miracle method that answers all the problems. We address the following question: how to integrate explainability in an AI project based on deep learning techniques?

After a state of the art presenting the richness of the literature in the field, we present the context and prerequisites for our work. Then we present a protocol for evaluating local explanations and a modular methodology for global model characterization. Finally, we show that our work is integrated into its industrial environment.

This work results in concrete tools allowing the reader to apprehend the richness of the explicability tools at their disposal.