



Avis de Soutenance

Madame Meryem YANKOL

Sciences Economiques

Soutiendra publiquement ses travaux de thèse intitulés

Nouvelles approches du Machine Learning pour la détection des fraudes financières et l'analyse de la survie dans les secteurs de l'assurance

dirigés par Monsieur Sessi Noude TOKPAVI

Ecole doctorale : Sciences de la Société : Territoires, Economie, Droit - SSTED

Unité de recherche : LÉO - Laboratoire d'Economie d'Orléans

Soutenance prévue le **mercredi 06 juillet 2022** à 13h30

Lieu : UFR DEG Université d'Orléans Rue de Blois, 45067 Orléans Cedex

Salle : des thèses

Composition du jury proposé

M. Sessi Noude TOKPAVI	Université d'Orléans	Directeur de thèse
M. Olivier DARNE	Université de Nantes	Rapporteur
M. Arthur CHARPENTIER	Université du Québec à Montréal	Rapporteur
Mme Denisa BANULESCU RADU	Université d'Orléans	Co-encadrante de thèse
M. Christophe BOUCHER	Université Paris Nanterre	Examineur
M. Sylvain BENOIT	Université Paris Dauphine	Examineur

Mots-clés : Détection de la fraude, Assurance, Analyse de survie, Apprentissage automatique, SHAP, LIME

Résumé :

Les secteurs de la banque et de l'assurance sont mis au défi de traiter de grandes quantités de données pour répondre à des problématiques commerciales et stratégiques. Diverses études ont déjà porté sur le développement d'outils d'aide à la décision pour aider les compagnies d'assurance à être mieux équipées pour lutter contre la fraude à l'assurance. Par ailleurs, dans un secteur de plus en plus concurrentiel, la connaissance du client est devenue un élément crucial pour les entreprises afin de conserver leur clientèle. L'évolution constante du marché et le grand nombre d'offres rendent également plus facile la résiliation des polices par les assurés. Par conséquent, cette thèse propose quatre articles de recherche qui contribuent à la littérature en étudiant les avantages des nouvelles approches d'apprentissage automatique pour la détection de la fraude à l'assurance et l'analyse de survie dans le secteur de l'assurance et de la banque. La contribution principale est d'utiliser des données réelles et d'appliquer de nouveaux modèles présentant de bonnes performances prédictives permettant d'identifier les facteurs de risque et les comportements. Des modèles économétriques traditionnels ainsi que de nouveaux algorithmes d'apprentissage automatique présentant de bonnes performances prédictives et une grande efficacité opérationnelle sont testés. Chaque méthode est évaluée sur la base de critères de performance appropriés et la question des données non cylindrées est également prise en compte. Les résultats sont interprétés à l'aide de la méthode SHapley Additive exPlanations, d'explications de modèles localement interprétables et de diagrammes de dépendance partielle pour les techniques d'apprentissage automatique. La méthode Natural Language Processing est utilisée pour des variables textuelles et les résultats confirment que la performance des modèles est améliorée lorsque ces dernières sont incluses. Les techniques d'apprentissage automatique, basées sur des algorithmes de Boosting, sont la méthode la plus performante pour la détection des fraudes et l'analyse de survie.