



Centre d'Etudes Doctorales : Sciences et Techniques et Sciences Médicales

Avis de Soutenance

THESE DE DOCTORAT

Présentée par

Madame BRIDGE-NDUWIMANA CHARLENE BEATRICE

Discipline : Informatique

Spécialité : Informatique et Systèmes intelligents

Sujet de la thèse

Nouvelles méthodes d'imputation basées sur l'intelligence artificielle pour la préparation et l'analyse des données massives

Formation Doctorale " Sciences de l'Ingénieur, Sciences Physiques, Mathématiques et Informatique"

Thèse présentée et soutenue le **samedi 21 novembre 2026 à 10h** à l'Ecole Supérieure de Technologie de Fès, devant le jury composé de :

NOM ET PRÉNOM	TITRE	ETABLISSEMENT	
Abdelaziz AHAITOUF	PES	Faculté des Sciences Appliquées de Taza	Président
Ismail AKHARRAZ	PES	Faculté des Sciences d'Agadir	Rapporteur
Mounia ZAIM	PES	Ecole Supérieure de Technologie de Salé	Rapporteur
Anass EL AFFAR	PES	Faculté des Sciences Appliquées de Taza	Rapporteur
Faiq GMIRA	MCH	Ecole Supérieure de Technologie de Fès	Examineur
Souad ALAOUI	MCH	Faculté des Sciences Appliquées de Taza	Examineur
Aziza EL OUAZIZI	PES	Faculté des Sciences Appliquées de Taza	Directeur de thèse
Majid BEN YAKHLEF	PES	Faculté des Sciences Appliquées de Taza	Co-Directeur de thèse

Laboratoire de recherche : Laboratoire Technologies Innovantes et d'Informatique

Etablissement : Ecole Supérieure de Technologie de Fès



Centre d'Etudes Doctorales : Sciences et Techniques et Sciences Médicales

Résumé de la thèse

La qualité des données constitue un prérequis fondamental à la fiabilité des systèmes d'apprentissage automatique. Dans cette dynamique, la gestion des valeurs manquantes constitue l'un des défis les plus persistants et les plus structurants : elle conditionne la validité statistique des analyses, la stabilité des modèles prédictifs et, in fine, la confiance que l'on peut accorder aux décisions algorithmiques. En dépit d'une littérature abondante, les approches existantes demeurent fragmentées, souvent confinées à des hypothèses restrictives sur les mécanismes de manque ou à des domaines d'application étroits, laissant la question d'une stratégie d'imputation unifiée, précise et opérationnellement viable sans réponse satisfaisante.

Cette thèse s'inscrit dans cette problématique et propose une investigation systématique, depuis l'état de l'art jusqu'au développement de nouvelles méthodes d'imputation fondées sur l'intelligence artificielle pour la préparation et l'analyse des données massives. Trois axes méthodologiques complémentaires sont explorés : l'exploitation des estimateurs basés sur le gradient dans des schémas itératifs de sélection de caractéristiques et d'imputation, l'intégration de la logique floue intuitionniste et du partitionnement pour modéliser l'incertitude inhérente aux valeurs manquantes, et enfin, la conception d'une architecture d'apprentissage profond hybride convolutionnel-Transformer opérant dans l'espace latent pour réaliser une imputation multiple statistiquement fondée. Ces contributions sont évaluées de manière rigoureuse à partir des référentiels de jeu de données reconnus, en présence de mécanismes de manque et de taux de manque variés, à l'aide de protocoles statistiques non paramétriques garantissant la fiabilité des comparaisons. Les résultats établissent la pertinence des approches hybrides et itératives par rapport aux méthodes classiques et aux architectures d'apprentissage profond conventionnelles, tout en mettant en évidence le rôle structurant de la sélection de caractéristiques en amont de l'imputation. L'architecture Conv1D-Transformer proposée démontre une compétitivité manifeste par rapport aux méthodes de référence les plus récentes, avec l'avantage supplémentaire de quantifier explicitement l'incertitude associée à chaque valeur reconstruite et de corriger implicitement les effets de batch sans nécessiter d'identifiants de provenance.

Mots clés :

Valeurs manquantes, Imputation des données, Apprentissage automatique, Logique floue intuitionniste, Partitionnement flou, Estimateurs par gradient, Apprentissage profond, Autoencodeur convolutionnel-Transformer, Imputation multiple, Quantification de l'incertitude, Correction des effets de batch, Préparation des données, Mécanismes MCAR/MAR/MNAR.