



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

Avis de Soutenance

THESE DE DOCTORAT

Présentée par

Monsieur EL KHATTABI MOHAMMED ZAKARIAE

Discipline : Informatique
Spécialité : Informatique

Sujet de la thèse

Geometry-Inference Based-Clustering: Nouvelle Métrique pour la Sélection Optimale du Nombre de Clusters sous K-means

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

Thèse présentée et soutenue le **samedi 20 avril 2024 à 10h** à l'Ecole Nationale des Sciences Appliquées de Fès, devant le jury composé de :

NOM ET PRÉNOM	TITRE	ETABLISSEMENT	
Abdellatif EZZOUHAIRI	PES	Ecole Nationale des Sciences Appliquées de Fès	Président
Said AGOUJIL	PES	Ecole Nationale du Commerce et de Gestion de Meknès	Rapporteur
Brahim AKSSASSE	PES	Faculté des Sciences de Meknès	Rapporteur
Ghizlane KHAISSIDI	PES	Ecole Nationale des Sciences Appliquées de Fès	Rapporteur
Khalid HADDOUCH	PH	Ecole Nationale des Sciences Appliquées de Fès	Examineur
Abdelhak BOULAALAM	PH	Ecole Nationale des Sciences Appliquées de Fès	Examineur
Adil KENZI	PES	Ecole Nationale des Sciences Appliquées de Fès	Examineur
Lahcen OUGHDIR	PES	Ecole Nationale des Sciences Appliquées de Fès	Directeur de Thèse

Laboratoire de recherche : Ingénierie, Systèmes et Applications
Etablissement : Ecole Nationale des Sciences Appliquées de Fès



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

Résumé de la thèse

Cette thèse approfondit le domaine du clustering en se concentrant sur l'algorithme K-means et ses métriques associées. Une analyse approfondie de l'état actuel des métriques pour la détection optimale du nombre de clusters est présentée. L'importance cruciale de la normalisation des données est examinée, en mettant en évidence les normes Min-Max et standard. Une étude comparative détaillée évalue quatre métriques de clustering sur différents ensembles de données, mettant en évidence l'impact significatif de la normalisation sur les résultats. Les conclusions soulignent la sensibilité de la précision du clustering à l'interaction complexe entre la distribution des données, la forme de normalisation et les métriques utilisées. La thèse propose également une nouvelle métrique/heuristique basée sur l'inférence géométrique pour la sélection automatisée du nombre optimal de clusters. Cette approche novatrice a montré son efficacité sur des ensembles de données gaussiens, bien que des travaux en cours visent à l'améliorer pour les données non gaussiennes, démontrant ainsi un progrès significatif dans le domaine du clustering.

Mots clés : Clustering, K-means, Métriques de clustering, Normalisation des données, Normes Min-Max, Norme standard (zéro-moyenne, unité d'écart type), Détection optimale du nombre de clusters, Étude comparative, Ensembles de données gaussiens et non gaussiens, Précision du clustering, Inférence géométrique (GIC), Heuristique, Sélection automatisée du nombre optimal de clusters