JOURNAL AFRICAIN DES SCIENCES
VOL. 3, **2** (2026) 46-57

https://jafrisci.org

ISSN (online) : 3107-0949 ; ISSN (print) : 3107-0930

OPEN ACCESS

ARTICLE DE RECHERCHE

Article Info.:
Reçu : le 19/12/2025
Accepté: le 03/02/2026
Publié : le 16/02/2026

APPROCHE HYBRIDE DATA SCIENCE-IA POUR LA REDUCTION DES ECARTS ENTRE CONCEPTION ET EXECUTION DES PROJETS D'INFRASTRUCTURES ROUTIERES : CAS DE LA RDC

Kipulu Ngimbi Serge [1,*], **Mabela Matendo Rostin** [1], **Nkulani Mansiamina Joseph** [1], **Kisasa Matwi Patrick** [1]

https://doi.org/10.70237/jafrisci.2026.v3.i2.03

Resumé

Les projets d'infrastructures routières jouent un rôle essentiel dans le développement économique et territorial de la République Démocratique du Congo. Toutefois, leur mise en œuvre est fréquemment marquée par des écarts significatifs entre les prévisions établies lors de la conception et les résultats observés à l'exécution, notamment en termes de coûts, de délais et de qualité. Ces dérives mettent en évidence les limites des approches classiques de gestion de projets, souvent statiques et peu adaptées à la complexité des environnements opérationnels. Cet article propose une approche hybride combinant les techniques de Data Science et les modèles d'Intelligence Artificielle afin d'analyser, d'anticiper et de réduire ces écarts. La méthodologie repose sur l'exploitation de données historiques de projets routiers publics, intégrant une analyse exploratoire, un clustering non supervisé et une modélisation prédictive. Les résultats révèlent l'existence de profils de projets à niveaux de risque différenciés et démontrent la capacité des modèles à anticiper les dérives avant les phases critiques. Cette approche renforce l'aide à la décision et contribue à améliorer la gouvernance des projets routiers en contextes contraints.

Mots clés : Data Science ; Intelligence Artificielle ; Projets d'infrastructures routières ; Écarts conception–réalisation ; Clustering ; Modélisation prédictive ; Gouvernance des projets en RDC

Abstract

Road infrastructure projects play a crucial role in the economic and territorial development of the Democratic Republic of Congo, yet their implementation is frequently affected by significant gaps between design forecasts and actual execution outcomes, particularly in terms of cost, schedule, and construction quality. These recurrent deviations reveal the limitations of traditional project management approaches, which rely on static monitoring tools and offer limited analytical capacity in complex project environments.
This paper proposes a hybrid approach combining Data Science techniques and Artificial Intelligence models to analyze, anticipate, and reduce design–execution gaps in road infrastructure projects. The methodology exploits historical data from public road projects through exploratory data analysis, unsupervised clustering, and predictive modeling of observed deviations. The results identify distinct project risk profiles and demonstrate the ability of predictive models to anticipate potential cost and schedule overruns before critical execution phases. This approach strengthens decision support and improves project governance in constrained institutional contexts, while offering a transferable framework for other developing countries facing similar infrastructure challenges.

Key words : Data Science; Artificial Intelligence; Road Infrastructure Projects; Design–Execution Gaps; Clustering; Predictive Modeling; Project Governance in the Democratic Republic of Congo

1. INTRODUCTION

1.1. Contexte et justification de l'étude

Les infrastructures routières constituent un pilier fondamental du développement socio-économique des États, en assurant la mobilité des personnes et des biens, l'intégration des marchés et l'accès équitable aux services essentiels. Dans les pays en développement, et plus particulièrement en Afrique subsaharienne, les investissements dans les réseaux routiers sont considérés comme un levier stratégique de croissance, de réduction de la pauvreté et de cohésion territoriale.[1] La performance des projets routiers conditionne ainsi non

[1] World Bank. (2019). *Roads for Development: Improving Access, Mobility and Connectivity*. Washington, DC: World Bank Publications.

Correspondance : *sergekipulu@gmail.com* ***(K. N. Serge)***

Copyright : © The Author(s) Published under a Creative Commons Attribution 4.0 International Licence (CC BY 4.0)

[1] *Apprenant au DEA à l'Université de Kinshasa (UNIKIN), Faculté des Sciences, Département de Mathématiques et Informatique,Option ; Data science et iA, République Démocratique du Congo*

seulement l'efficacité des politiques publiques, mais égalementla crédibilité des institutions en charge de la planification et de l'exécution des infrastructures.

En République Démocratique du Congo, pays à forte étendue territoriale et doté d'un important potentiel économique, les infrastructures routières jouent un rôle central dans le désenclavement des zones rurales, la fluidité des échanges interprovinciaux et l'intégration régionale[2]. Malgré les efforts consentis par l'État et ses partenaires techniques et financiers, le secteur routier congolais demeure confronté à de multiples contraintes structurelles, notamment la faiblesse des capacités institutionnelles, l'insuffisance des mécanismes de suivi des projets, la complexité des environnements opérationnels et la variabilité des conditions géographiques et climatiques[3].

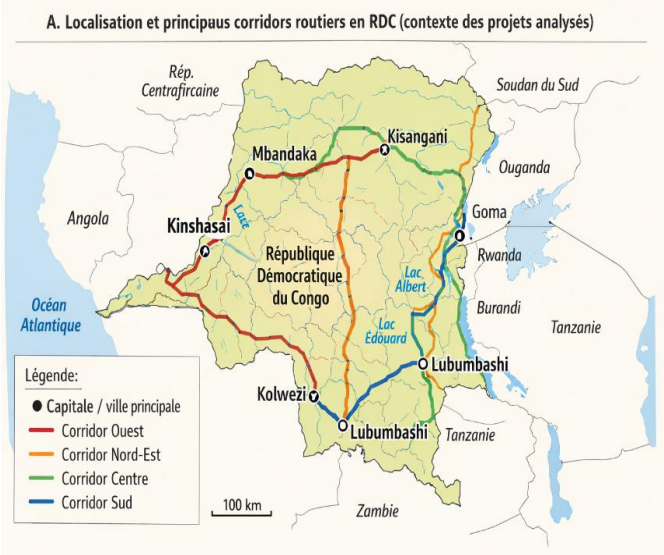


Figure 1.1. *Localisation et principaux corridors routiers en RDC (contexte des projets analysés)*

Ces contraintes rendent la gestion des projets d'infrastructures particulièrement délicate et exposée à des risques élevés de dérives.

L'un des constats récurrents dans la mise en œuvre des projets routiers en RDC réside dans l'existence d'écarts significatifs entre les hypothèses formulées lors de la phase de conception et les résultats observés lors de l'exécution. Ces écarts se traduisent principalement par des dépassements de coûts, des retards dans les délais de livraison et, dans certains cas, une dégradation de la qualité des ouvrages réalisés[4]. De telles dérives compromettent l'atteinte des objectifs initiaux des projets et limitent leur impact socio-économique à moyen et long terme.

Les approches traditionnelles de gestion et de suivi des projets routiers, généralement basées sur des outils descriptifs, des indicateurs statiques et des rapports périodiques, montrent leurs limites face à la complexité croissante des projets et à la diversité des données produites tout au long de leur cycle de vie[5]. Ces méthodes peinent à exploiter efficacement les volumes importants de données historiques disponibles et à anticiper, de manière proactive, les risques susceptibles de générer des écarts entre conception et réalisation.

Dans ce contexte, l'intégration des techniques avancées de Data Science et d'Intelligence Artificielle apparaît comme une opportunité majeure pour améliorer l'analyse, la compréhension et la maîtrise des performances des projets d'infrastructures routières.

La justification de cette étude repose ainsi sur la nécessité de proposer une approche analytique innovante, fondée sur l'exploitation des données et l'apprentissage automatique, capable de dépasser les limites des méthodes classiques. Une telle approche vise à identifier des profils de projets à risque,à prédire les dérives potentielles avant les phases critiques d'exécution et à renforcer les mécanismes d'aide à la décision au sein des institutions publiques en charge des infrastructures routières en République Démocratique du Congo. Une figure de synthèse illustrant le positionnement de l'approche proposée par rapport aux méthodes traditionnelles peut être insérééa la fin de cette section afin de mettre en évidence l'apport méthodologique de l'étude.

1.2. Problématique de recherche

La gestion des projets d'infrastructures routières repose traditionnellement sur des méthodes de planification et de contrôle fondées sur des outils normatifs tels que les calendriers prévisionnels, les estimations budgétaires initiales et les mécanismes de suivi périodiques.

Bien que ces approches aient contribué à structurer les pratiques de gestion de projets, elles montrent aujourd'hui des limites significatives lorsqu'elles sont confrontées à la complexité croissante des projets publics et à la multiplicité des facteurs susceptibles d'influencer leur performance [6]. En particulier, ces méthodes se caractérisent par une capacité limitée à intégrer des données hétérogènes issues de différentes phases du cycle de vie des projets, ainsi que par une faible aptitude à anticiper les risques de dérives avant leur matérialisation.

Dans le contexte des projets routiers en République Démocratique du Congo, ces insuffisances sont amplifiées par des contrainte institutionnelles, organisationnelles et environnementales spécifiques. Les mécanismes de suivi traditionnels reposent principalement sur des analyses rétrospectives, intervenant souvent après l'apparition des écarts entre conception et réalisation, ce qui réduit considérablement les marges de manœuvre pour des actions correctives efficaces[7]. De plus, la fragmentation des sources d'information et l'absence de systèmes analytiques intégrés limitent la

[2] Banque Africaine de Développement. (2020). *Stratégie sectorielle des transports en République Démocratique du Congo*. Abidjan: BAD.

[3] Ministère des Infrastructures et Travaux Publics (MITP). (2021). *Rapport sur l'état des infrastructures routières en RDC*. Kinshasa, document institutionnel.

[4] Flyvbjerg, B., Holm, M. K. S., & Buhl, S. (2002). *Underestimating costs in public works projects: Error or lie?* Journal of the American Planning Association, 68(3), 279–295.

[5] Kerzner, H. (2017). *Project Management: A Systems Approach to Planning, Scheduling, and Controlling*. 12th ed., Hoboken, NJ: Wiley.

[6] PMI. (2021). *A Guide to the Project Management Body of Knowledge (PMBOK® Guide)*. 7th ed., Newtown Square, PA: Project Management Institute.

[7] Müller, R., & Lecoeuvre, L. (2014). *Operationalizing governance categories of projects*. International Journal of Project Management, 32(8), 1346–1357.

capacité des décideurs à disposer d'une vision globale et dynamique de la performance des projets.

Face à ces limites, l'exploitation des données massives produites par les projets d'infrastructures routières apparaît comme une opportunité encore largement sous-exploitée. Les avancées récentes en Data Science et en Intelligence Artificielle offrent des perspectives nouvelles pour analyser des volumes importants de données historiques, identifier des schémas récurrents et modéliser des relations complexes entre les variables influençant la performance des projets[8]. Ces approches permettent non seulement de dépasser les analyses descriptives classiques, mais également de développer des modèles prédictifs capables d'anticiper les écarts potentiels entre conception et exécution.

La problématique centrale de cette recherche réside ainsi dans la nécessité de concevoir une approche méthodologique innovante, fondée sur les données et l'apprentissage automatique, permettant d'améliorer la compréhension des facteurs générateurs d'écarts et de renforcer la capacité des acteurs publics à piloter les projets routiers de manière proactive. Une figure conceptuelle illustrant le passage d'une gestion traditionnelle réactive vers une gestion predictive et orientée données peut être propose à l'issue de cette section afin de mettre en évidence l'évolution paradigmatique induite par l'intégration de la Data Science et de l'Intelligence Artificielle.

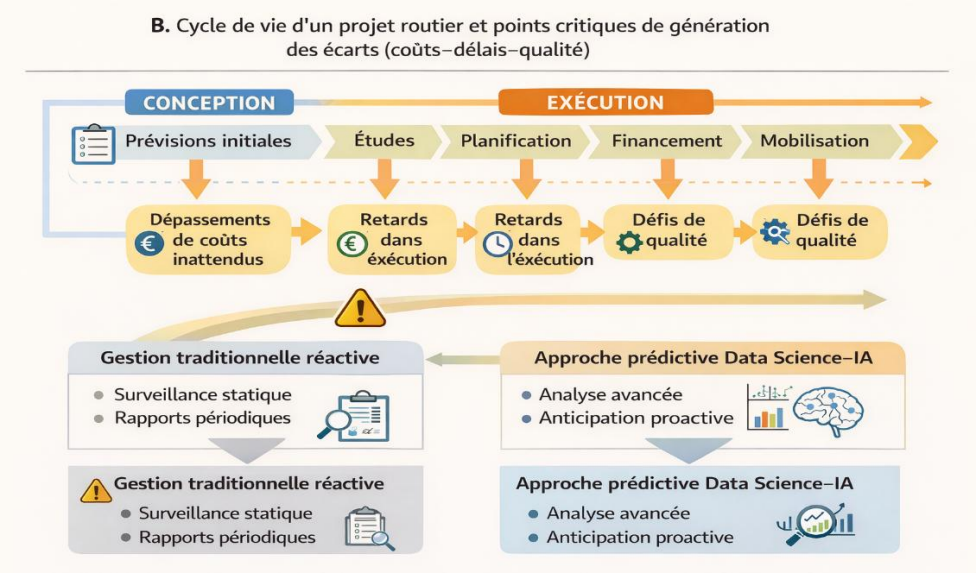


Figure 1.2. *Cycle de vie d'un projet routier et points critiques de génération des écarts (coûts–délais–qualité)*

1.3. Objectifs de l'article

L'objectif général de cet articleest de proposer et de valider une approche hybride fondée sur les techniques de Data Science et d'Intelligence Artificielle afin d'analyser, d'anticiper et de contribuerà la réduction des écarts entre la conception et la réalisation des projets d'infrastructures routières en République Démocratique du Congo.

Cette approche vise à dépasser les limites des méthodes traditionnelles de gestion de projets en exploitantde manière systématique les données historiques disponibles pour améliorer la qualité des décisions tout au long du cycle de vie des projets [9].

Dans cette perspective, l'article poursuit plusieurs objectifs spécifiques. Il s'agit, en premier lieu, d'identifier et de structurer les variables clés caractérisant les projets routiers susceptibles d'influencer l'apparition des écarts de coûts, de délaiset de qualité. En second lieu, l'étude vise à regrouper les projets selon des profils homogènes à l'aide de techniques de clustering non supervisé, afin de mettre en évidence des catégories de projets présentant des niveaux de risque différenciés. L'article cherche également à développer et à évaluer des modèles prédictifs basés sur des algorithmes d'Intelligence Artificielle capables d'anticiper les dérives potentielles avant les phases critiques d'exécution. Enfin, l'objectif est de démontrer l'apport opérationnel de cette approche hybride en matière d'aide à la décisionet de gouvernance des projets routiers, en proposant un cadre analytique réplicable et adaptable à d'autres contextes de pays en développement[10].

Une représentation synthétique des objectifs et de leur articulation au sein de la démarche méthodologique peut être proposée sous forme de schéma conceptuel, facilitant la compréhension du lien entre les objectifs scientifiques et les résultats attendus.

1.4. Contributions scientifiques et pratiques

Sur le plan scientifique, cet article se distingue par son originalité méthodologique, fondée sur l'intégration conjointe des techniques de Data Science et des modèles d'Intelligence Artificielle dans l'analyse des projets d'infrastructures

[8] Bakhshi, N., Ireland, V., & Gorod, A. (2016). *Systems thinking perspectives on digital transformation in project management*. International Journal of Information Management, 36(6), 1265–1275.

[9] Shmueli, G., Bruce, P. C., Gedeck, P., & Patel, N. R. (2020). *Data Mining for Business Analytics: Concepts, Techniques, and Applications*. Hoboken, NJ: Wiley.

[10] Flyvbjerg, B. (2021). *Top Ten Behavioral and Political Biases in Project Management: An Overview*. Project Management Journal, 52(6), 531–546.

routières. Contrairement aux travaux antérieurs qui mobilisent généralementdes approches isolées, descriptivesou exclusivement prédictives, l'étude propose un cadre hybride combinantune analyse exploratoire des données, un regroupement non supervisé des projets et une modélisation prédictive des écarts entre conception et réalisation[11]. Cette articulation méthodologique permet de capturer à la fois la structure latente des données et les relations complexes entre les variables influençant la performancedes projets, contribuant ainsi à enrichir les travaux existants en gestion de projets et en ingénierie des infrastructures.

L'originalité de la démarche réside également dans son application à un contexte africain encore peu exploré dans la littérature scientifique internationale.
En mobilisant des données issues de projets routiers publics en République Démocratique du Congo, l'article apporte des éléments empiriques nouveaux sur les dynamiques de dérives de coûts et de délais dans des environnements caractériséspar des contraintes institutionnelles, organisationnelles et opérationnelles spécifiques.

Une figure méthodologique illustrant l'articulation entre les différentes composantes de l'approche hybride peut être insérée afin de mettre en évidence la valeur ajoutée scientifique de la démarche proposée.

Sur le plan pratique, les contributionsde cette recherche se traduisent par des apports concrets pour la gouvernance publique des projets d'infrastructures routières. L'approche développée offre aux décideurs et aux gestionnaires de projets des outils analytiques permettant d'identifier précocement les projets à risque, d'anticiper les dérives potentielles et de renforcer les mécanismes de pilotage et de contrôle[12].En favorisant une prise de décision fondée sur les données, cette contribution participe à l'amélioration de la transparence,de l'efficacité et de la redevabilitédans la gestion des investissements publics, tout en ouvrant des perspectivespour l'institutionnalisation de pratiques de gestion prédictive au sein des administrations en chargedes infrastructures routières.

2. Revue de littérature

2.1. Projets d'infrastructures routières et causes des écarts

Les projets d'infrastructures routières figurent parmi les investissements publics les plus complexes en raison de leur envergure financière, de leur durée d'exécution et de la multiplicité des acteurs impliqués.

La littérature scientifique meten évidence que ces projets sont particulièrement exposés à des écarts entre les prévisions établies lors de la phasede conception et les résultats observés lors de l'exécution, quels que soient le contexte géographique ou le niveau de développement économique[13].
Ces écarts constituent un enjeu majeur pour la performance des politiques publiqueset la soutenabilité des investissements en infrastructures.
Les dépassements de coûts représentent l'une des formes les plus documentéesde ces écarts. Plusieurs études soulignent que les budgets initiaux des projets routiers sont fréquemment sous-estimés, en raison de facteurs tels que l'insuffisance des études préalables, l'évolution des prix des matériaux, les modificationsdu périmètre des travaux et les aléas institutionnels ou politiques .

Dans les pays en développement, ces dépassements sont souvent accentués par des faiblesses dans les mécanismesde contrôle financier et par un faible capacité d'anticipation des risques.

Tableau 1. Synthèse des taux moyens de dépassement de coûts dans les projets d'infrastructures routières selon les régions du monde

Région du monde	Type dominant de projets routiers	Taux moyen de dépassement de coûts (%)	Intervalle observé (%)	Principales causes identifiées
Europe occidentale	Routes nationales, autoroutes	15 – 25	10 – 40	Ajustements techniques, inflation modérée, modifications réglementaires
Amérique du Nord	Autoroutes, infrastructures urbaines	20 – 30	15 – 50	Changements de périmètre, contraintes environnementales, procédures administratives
Asie	Routes interurbaines, projets rapides	25 – 40	20 – 60	Sous-estimation initiale, complexité des chantiers, pression sur les délais
Afrique	Routes nationales, routes rurales	30 – 60	25 – 80	Faiblesses institutionnelles, aléas financiers, contraintes logistiques
Amérique latine	Routes interurbaines, corridors économiques	25 – 45	20 – 65	Instabilité économique, renégociations contractuelles
Pays en développement (moyenne)	Projets routiers publics	35 – 55	25 – 75	Gouvernance des projets, risques exogènes, insuffisance des études préalables

[11] Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed., New York: Springer.

[12] OECD. (2020). *Infrastructure Governance: A Framework for Policy Makers*. Paris: Organisation for Economic Co-operation and Development.

[13] Cantarelli, C. C., Flyvbjerg, B., Molin, E. J. E., & van Wee, B. (2012). *Cost overruns in large-scale transportation infrastructure projects: Explanations and their theoretical embeddedness*. European Journal of Transport and Infrastructure Research, 12(1), 5–18.

Les retards d'exécution constituent une autre cause majeure des écarts entre conception et réalisation des projets routiers. La littérature identifie une diversité de facteurs explicatifs, incluant les contraintes de financement, les défaillances dans la planification des activités, les problèmes de coordination entre les parties prenantes et les conditions environnementales défavorables[14].

Les retards prolongent la durée des chantiers, augmentent les coûts indirects et retardent la mise à disposition des infrastructures au bénéfice des populations. Un graphique illustrant la relation entre durée prévue et durée réelle des projets routiers peut être proposé pour mettre en évidence l'ampleur des dérives temporelles.

Les défaillances de qualité constituent une troisième catégorie d'écarts fréquemment évoquée dans les travaux consacrés aux projets d'infrastructures routières. Elles se traduisent par des ouvrages ne respectant pas les normes techniques initialement définies, une durabilité insuffisante des infrastructures ou la nécessité de travaux correctifs prématurés[15].
Ces défaillances sont souvent liées à des choix techniques inadéquats, à une supervision insuffisante des chantiers ou à des compromis réalisés pour compenser des contraintes budgétaires et temporelles.

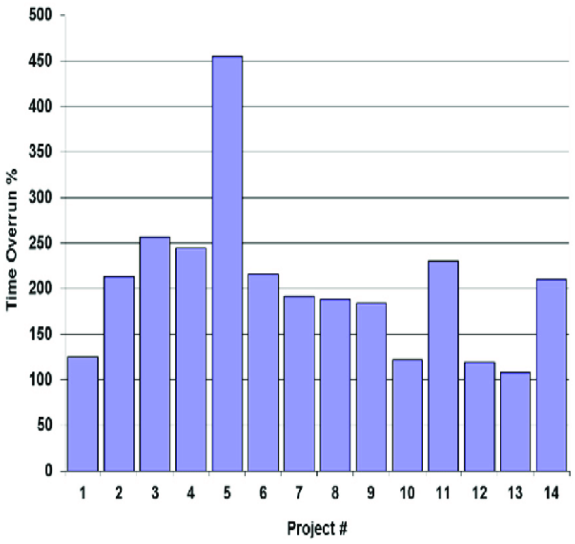


Figure 1. *Comparaison entre durées prévues et durées réelles des projets d'infrastructures routières*

L'analyse de la littérature met ainsi en évidence que les écarts observés dans les projets routiers résultent rarement d'un facteur isolé, mais plutôtd'une combinaison de causes interdépendantes. Cette complexité justifie le recours à des approches analytiques capables d'exploiter simultanémentun grand nombre de variables et de capturer des relations non linéaires, ouvrant la voie à l'intégration des techniques de Data Science et d'Intelligence Artificielle dans l'analyse et la gestion des projets d'infrastructures routières.

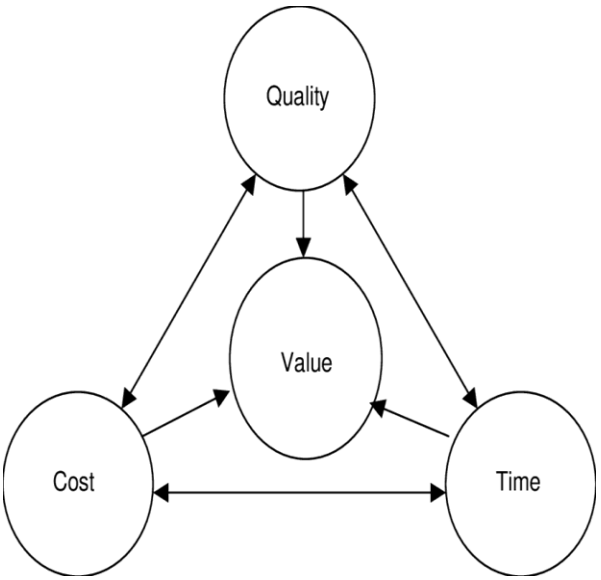


Figure 2. *Interaction systémique entre dépassements de coûts, retards d'exécution et défaillances de qualité dans les projets routiers*

2.2. Apports de la Data Science dans la gestion des projets

L'émergence de la Data Science a profondément renouvelé les approches d'analyse et de pilotage des projets complexes, en particulier dans le domaine des infrastructures publiques. Contrairement aux méthodes traditionnelles centrées sur des indicateurs isolés et des analyses statiques, la Data Science permet une analyse multidimensionnelle intégrant simultanément des variables techniques, financières, temporelles et contextuelles issues de l'ensemble du cycle de vie des projets[16]. Cette capacité d'intégration favorise une compréhension plus fine des mécanismes générateurs de performance ou de dérives et ouvre la voie à des analyses exploratoires capables de révéler des structures latentes au sein des données.

Dans le cadre des projets d'infrastructures routières, l'analyse multidimensionnelle permet de dépasser les lectures univariées des écarts pour appréhender leurs interactions complexes. Les techniques de Data Science facilitent l'identification de corrélations non évidentes entre les caractéristiques des projets, telles que la localisation géographique, le mode de financement, le type d'entreprise exécutante ou encore les conditions environnementales, et les résultats observés en termes de coûts, de délais et de qualité[17].

[14] Doloi, H., Sawhney, A., Iyer, K. C., & Rentala, S. (2012). *Analysing factors affecting delays in Indian construction projects*. International Journal of Project Management, 30(4), 479–489.

[15] Love, P. E. D., Edwards, D. J., & Irani, Z. (2012). *Moving beyond optimism bias and strategic misrepresentation: An explanation for social infrastructure project cost overruns*. IEEE Transactions on Engineering Management, 59(4), 560–571.

[16] Provost, F., & Fawcett, T. (2013). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. Sebastopol, CA: O'Reilly Media.

[17] Batini, C., & Scannapieco, M. (2016). *Data Quality: Concepts, Methodologies and Techniques*. Berlin: Springer.



Figure 3. *Représentation multidimensionnelle des variables de performance des projets d'infrastructures routières*

L'exploitation efficace de ces analyses repose toutefois sur la disponibilité de données structurées, cohérentes et historisées. À cet égard, les entrepôts de données constituent un socle fondamental pour l'application de la Data Science à la gestion des projets d'infrastructures.
En centralisant des données hétérogènes provenant de multiples sources opérationnelles, les Data Warehouses permettent de consolider l'information, d'assurer sa qualité et de faciliter des analyses longitudinales sur de larges ensembles de projets[18]. Cette structuration favorise la production d'indicateurs de performance fiables et comparables dans le temps et entre projets.

Dimension de performance	Indicateur	Description analytique	Utilité décisionnelle
Coûts	Taux de dépassement budgétaire	Écart relatif entre le coût réel et le coût prévu du projet	Identification des projets à risque financier
Coûts	Coût unitaire par kilomètre	Rapport entre le coût total du projet et la longueur de la route réalisée	Comparaison de l'efficience entre projets
Délais	Taux de retard d'exécution	Écart relatif entre la durée réelle et la durée planifiée	Détection des dérives temporelles
Délais	Durée moyenne des phases critiques	Temps cumulé des phases clés (mobilisation, travaux, réception)	Optimisation de la planification
Qualité	Taux de non-conformité technique	Proportion des ouvrages ne respectant pas les normes définies	Amélioration du contrôle qualité
Qualité	Fréquence des travaux correctifs	Nombre d'interventions correctives après réception	Évaluation de la durabilité des ouvrages
Contexte	Indice de complexité du projet	Score synthétique intégrant localisation, environnement et contraintes	Segmentation des projets par niveau de risque
Gouvernance	Indice de performance globale	Indicateur composite coûts–délais–qualité	Aide à la priorisation des actions correctives

Les indicateurs de performance issus des entrepôts de données ne se limitent plus à des mesures descriptives, mais s'inscrivent dans une logique analytique orientée vers l'aide à la décision. La Data Science permet d'enrichir ces indicateurs par des analyses croisées, des segmentations de projets et des visualisations avancées, offrant aux décideurs une vision synthétique et dynamique de la performance des portefeuilles de projets[19]. Un tableau récapitulatif des principaux indicateurs de performance utilisés dans la littérature, structurés selon les dimensions coûts, délais et qualité, peut être inséré afin de clarifierle cadre analytique mobilisé.
L'ensemble de ces apports positionne la Data Science comme un levier stratégique pour moderniser la gestion des projets d'infrastructures routières.

En favorisant une approche intégrée, fondée sur les données et orientée vers l'anticipation des risques, elle préparele terrain pour l'intégration des modèles d'Intelligence Artificielle, qui prolongent ces analyses par des capacités prédictives et décisionnelles avancées.

2.3. Intelligence Artificielle et prédiction des risques projets

L'Intelligence Artificielle constitueun prolongement naturel des approchesde Data Science pour l'analyseet l'anticipation des risques dans les projets d'infrastructures routières. En exploitant les données historiques, elle permet de dépasser les analyses descriptives traditionnelles et d'introduire des capacités prédictives orientées vers l'aideà la décision[20].
Les techniques de clustering non supervisé sont largement mobilisées pour identifier des profils homogènes de projets à partir de leurs caractéristiques financières, temporelles,

[18] Inmon, W. H. (2015). *Building the Data Warehouse*. 4th ed., Hoboken, NJ: Wiley.
[19] Müller, R., & Jugdev, K. (2012). *Critical success factors in projects*. International Journal of Managing Projects in Business, 5(4), 757–775.
[20] Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach*. 4th ed., Harlow: Pearson.

techniques et contextuelles. Ces méthodes permettent de révélerdes catégories de projets présentant des niveaux de risque différenciés, facilitant ainsi la segmentation des portefeuilleset l'analyse exploratoire des dérives potentielles[21].

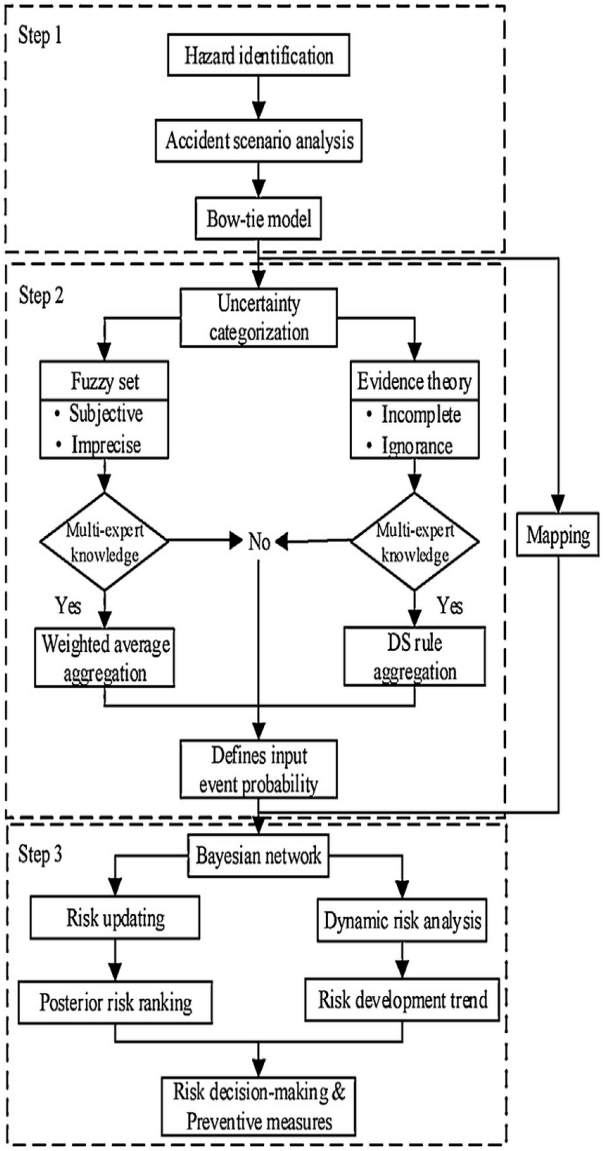


Figure 4. *Segmentation des projets d'infrastructures routières en clusters de risque par Intelligence Artificielle*

Les modèles prédictifs supervisés complètent cette segmentationen établissant des relations entreles caractéristiques initiales des projets et les écarts observés lors de l'exécution.

Des algorithmes tels que les forêts aléatoires ou les machines à vecteursde support sont couramment utilisés pour prédire les risques de dépassement de coûts ou de retards, grâce à leur capacitéà modéliser des relations complexes[22].

La combinaison de ces approches non supervisées et supervisées permetde développer des cadres analytiques hybrides efficaces pour la gestion prédictive des projets routiers et constituele fondement méthodologique de l'approche proposée dans cet article[23].

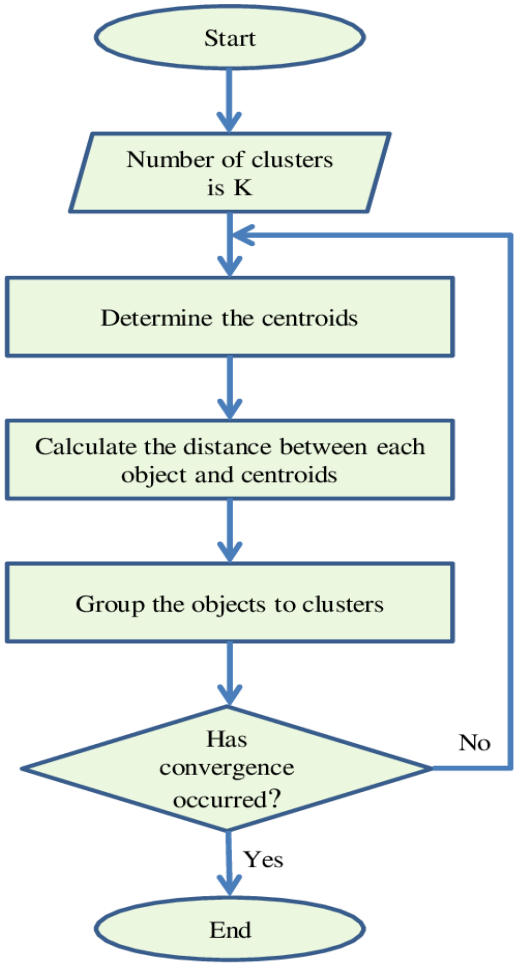


Figure 5. *Articulation entre clustering et modèles prédictifs pour l'anticipation des risques projets*

2.4. Synthèse critique et positionnement de l'étude

L'analyse de la littérature consacrée aux projets d'infrastructures routières met en évidence une abondance de travaux portant sur les causes des écarts entre conception et réalisation, notamment en matière de coûts, de délais et de qualité. Toutefois, une part importante de ces études demeure centrée sur des approches descriptives ou explicatives, reposant sur des analyses statistiques classiques et des indicateurs agrégés[24]. Bien que pertinentes pour identifier des tendances générales, ces approches présentent des limites notables lorsqu'il s'agit d'appréhender la complexité des interactions entre les multiples facteurs influençant la performance des projets.

Par ailleurs, les travaux mobilisant la Data Science ou l'Intelligence Artificielle se concentrent souvent sur une seule

[21] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666.

[22] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York: Springer.

[23] Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2007). Supervised machine learning: A review of classification techniques. *Informatica*, 31(3), 249–268.

[24] Love, P. E. D., Sing, C. P., Wang, X., & Irani, Z. (2013). Overruns in transportation infrastructure projects. *Journal of Management in Engineering*, 29(4), 312–325.

catégorie de méthodes, soit exploratoires, soit prédictives, sans proposer de cadre intégré permettant de combiner ces deux dimensions[25]. Cette fragmentation méthodologique limite la capacité à exploiter pleinement le potentiel des données historiques, en particulier pour la segmentation des projets à risque et l'anticipation des dérives futures.

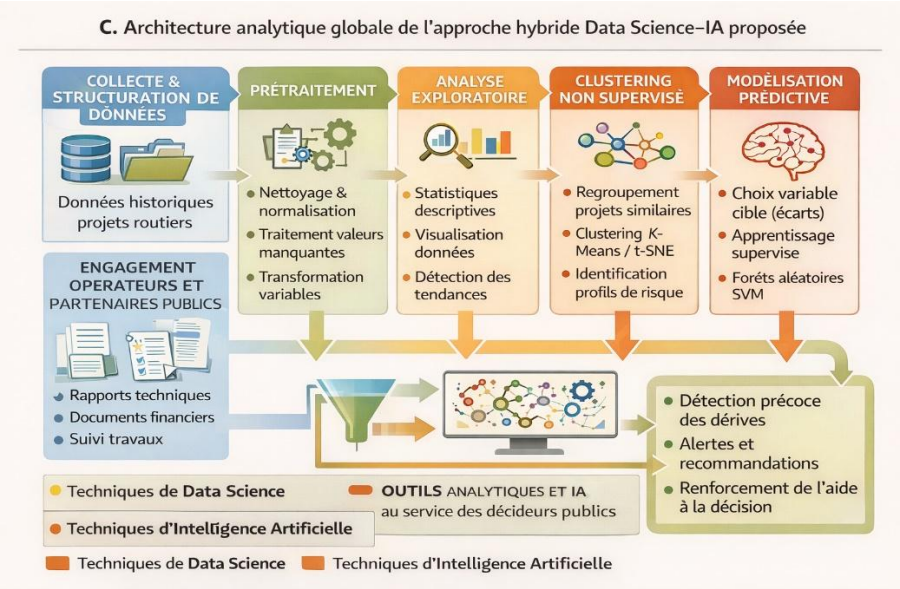
De plus, la majorité des études empiriques existantes portent sur des contextes de pays développés, laissant encore insuffisamment explorées les réalités spécifiques des pays en développement, et notamment celles des projets routiers en Afrique subsaharienne[26].
Dans ce contexte, la présente étudese positionne comme une contribution visant à combler ces lacunes à traversla proposition d'une approche hybride combinant le clustering non supervisé et les modèles prédictifs supervisés.
Cette articulation méthodologique permet, d'une part, d'identifier des profils homogènes de projets routiers caractérisés par des niveaux de risque différenciés et, d'autre part, de développer des modèles capables d'anticiper les écarts entre conception et exécution avant les phases critiques de réalisation[27]. En intégrant ces deux dimensions au sein d'un même cadre analytique, l'étude dépasse les approches isolées et offre une lecture plus complète et opérationnelle des risques projets.

Le positionnement de cette recherche repose également sur son ancrage dans le contexte de la République Démocratique du Congo, caractérisé par des contraintes institutionnelles, organisationnelles et opérationnelles spécifiques. En mobilisant des données issues de projets routiers publics congolais, l'étude apporte un éclairage empirique original et propose un cadre méthodologique transférable à d'autres contextes similaires.
Ce positionnement prépare ainsi la transition vers la section méthodologique, dans laquelle l'approche hybride proposée est formalisée et mise en œuvre de manière opérationnelle.

3. Méthodologie de recherche

3.1. Cadre général de l'étude

Cette étude adopte une démarche empirique fondée sur l'analyse de données historiques relatives aux projets d'infrastructures routières. L'approche méthodologique repose sur l'intégration progressive des techniques de Data Science et d'Intelligence Artificielle, depuis l'exploration des données jusqu'à la modélisation prédictive des écarts entre conception et réalisation[28]. L'objectif est de construire un cadre analytique cohérent, reproductible et orienté vers l'aide à la décision.
L'architecture globale du système analytique s'articule autour de cinq étapes principales : collecte et structurationdes données, prétraitement, analyse exploratoire, clustering non supervisédes projets et modélisation prédictive des écarts. Cette chaîne de traitement vise à assurer la cohérence méthodologique et la robustesse des résultats [29].



3.2. Données utilisées

Les données exploitées proviennent de projets routiers publics et sont issues de documents institutionnels, de rapports techniques et financiers ainsi que de dossiers de suivi des travaux[30]. Elles couvrent l'ensemble du cycle de vie des projets, de la conception à l'exécution.

Les variables considérées se répartissent en trois catégories principales. Les variables techniques décrivent les caractéristiques physiques et contextuelles des projets.
Les variables financières portent surles coûts prévisionnels et réels ainsi queles modalités de financement. Les variables temporelles concernent les durées planifiéeset effectives des différentes phasesdu projet. Cette structuration permet une

[25] Marzouk, M., & Enaba, M. (2019). Predicting cost overruns in construction projects using machine learning. *Journal of Construction Engineering and Management*, 145(1), 04018100.
[26] Osei-Kyei, R., & Chan, A. P. C. (2017). Developing transport infrastructure in developing countries. *Transport Reviews*, 37(3), 282–305.
[27] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260.
[28] Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage.
[29] Shmueli, G., Bruce, P. C., Gedeck, P., & Patel, N. R. (2020). *Data Mining for Business Analytics*. Wiley.
[30] World Bank. (2018). *Infrastructure Project Appraisal and Monitoring Guidelines*. Washington, DC.

analyse intégrée des performances projets[31].

3.3. Prétraitement et analyse exploratoire

Un prétraitement rigoureux des donnéesa été réalisé afin de garantir leur qualité analytique. Il a consisté au traitement des valeurs manquantes, à la correction des incohérences et à la normalisationdes variables quantitatives[32]. L'analyse exploratoire a permis d'examinerles distributions, d'identifier les tendances générales et de détecter d'éventuelles anomalies, constituant ainsi une base solide pour les étapes de modélisation[33].

3.4. Clustering des projets routiers

Le regroupement des projets a été effectué à l'aide de l'algorithme K-Means, choisi pour son efficacité et son usage éprouvé dans l'analyse de données multidimensionnelles[34].
Le nombre optimal de clusters a été déterminé à partir de critères analytiques tels que la méthode du coude, permettant d'obtenir une partition équilibréeet interprétable[35].
Les clusters identifiés ont été interprétés comme des profils de projets caractérisés par des niveaux de risque différenciés.

3.5. Modélisation prédictive des écarts

La prédiction des écarts a été réalisée à l'aide de modèles supervisés, notamment la forêt aléatoire et la machine à vecteurs de support, sélectionnés pour leur capacité à modéliser des relations non linéaires complexes[36]. La variable cible correspond aux catégories d'écarts observés entre les valeurs prévues et réelles des coûts et des délais. Les performances des modèles ont été évaluées à l'aide de métriques standards incluant la précision, le rappel et le score F1[37].
Cette démarche méthodologique condensée constitue le socle empirique de l'approche hybride proposée et prépare la présentation des résultats et de leur analyse.

4. Résultats et analyses

4.1. Résultats du clustering

L'application de l'algorithme de clustering non supervisé a permis de segmenter les projets d'infrastructures routières en groupes distincts présentantdes caractéristiques homogènes.
Le nombre de clusters retenu résulte d'un compromis entre la cohérence interne des groupes et leur interprétabilité, tel que recommandé dans la littérature méthodologique[38]. Chaque cluster obtenu reflète un profil spécifique de projets, différencié principalement par les niveaux de coûts, de délais et de complexité technique.

L'analyse des caractéristiques moyennes des clusters met en évidence l'existence d'un groupe de projets présentant des dépassements de coûts et des retards significativement supérieurs à la moyenne. Ces projets à risque élevé se distinguent par des hypothèses initiales optimistes, une complexité contextuelle accrueet des conditions d'exécution défavorables, confirmant les constats rapportésdans les travaux antérieurs sur les projets d'infrastructures publiques[39].

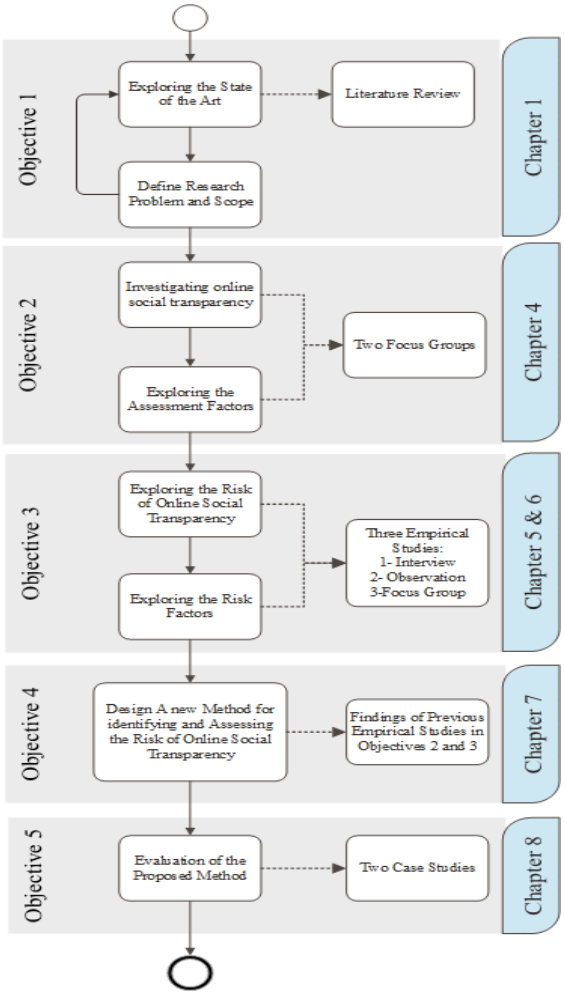


Figure 6. *Distribution des projets routiers selon les clusters de risque identifiés.*

Cette segmentation permet ainsi d'identifier, dès les phases amont, les catégories de projets les plus exposées aux dérives, offrant une première base analytique pour un pilotage différencié des projets.

4.2. Résultats de la prédiction

Les modèles prédictifs supervisés ont été mobilisés afin d'anticiper les écarts entre conception et réalisation à partir des caractéristiques initiales des projets. Les résultats montrent que

[31] Kerzner, H. (2017). *Project Management: A Systems Approach*. Wiley.
[32] Batini, C., & Scannapieco, M. (2016). *Data Quality: Concepts, Methodologies and Techniques*. Springer.
[33] Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley.
[34] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666.
[35] Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis*. Wiley.
[36] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
[37] Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC. *Journal of Machine Learning Technologies*, 2(1), 37–63.
[38] Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis*. Wiley.
[39] Flyvbjerg, B. (2014). What You Should Know About Megaprojects and Why. *Project Management Journal*, 45(2), 6–19.

les deux algorithmes testés présentent des performances satisfaisantes, avec une meilleure stabilité globale observée pour la forêt aléatoire, notamment en termes de précisionet de capacité de généralisation[40]. Ces résultats confirment la pertinence des approches d'apprentissage automatique pour la prédiction des risques projets.

L'analyse de l'importance des variables met en évidence le rôle prépondérant des facteurs financiers initiaux, des durées planifiées des phases critiques et de certains indicateurs de complexité contextuelle dans la prédiction des dérives[41]. Ces éléments soulignent l'importance stratégique des décisions prises lors de la phasede conception et justifient l'intégration de mécanismes analytiques avancésdès les premières étapes des projets.

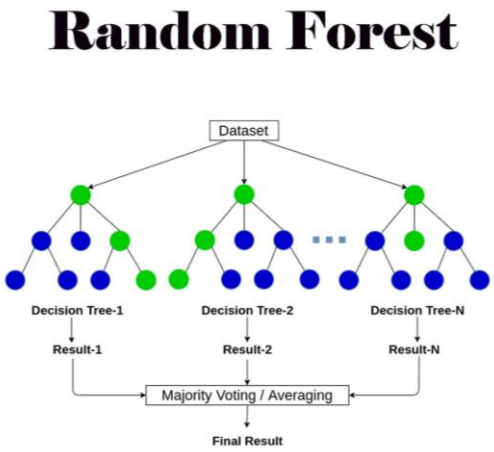


Figure 7. *Importance relative des variables dans la prédiction des écarts projets*

4.3. Validation de l'approche hybride

La confrontation des résultats issus du clustering et de la modélisation prédictive met en évidence une forte complémentarité entre les deux approches. Les projets appartenant aux clusters identifiés comme les plus risqués sont également ceux pour lesquels les modèles prédictifs estiment une probabilité élevée de dépassements, ce qui confirme la cohérence interne de l'approche hybride proposée[42]. Cette convergence renforce la crédibilité analytique du cadre méthodologique développé.
La robustesse de l'approche a été évaluéeà travers des tests de sensibilité et des validations croisées, montrant une stabilité globale des résultats face à des variations modérées des paramètres et des jeuxde données[43].

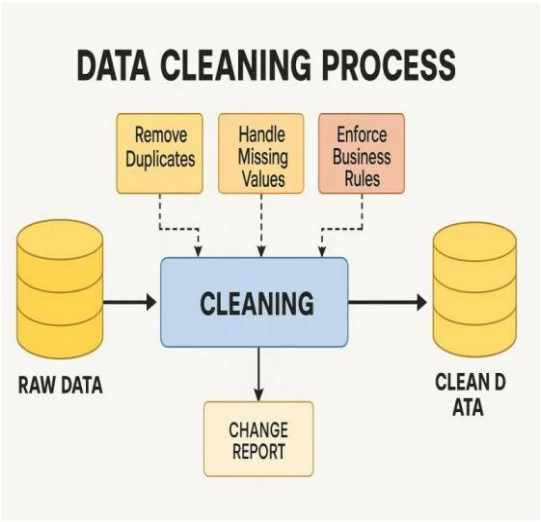


Figure 8. *Articulation entre clustering et prédiction dans la validation de l'approche hybride*

Dans l'ensemble, ces résultats démontrent que la combinaison du clustering non supervisé et des modèles prédictifs supervisés permet d'améliorer significativement l'analyse et l'anticipation des écarts entre conception et réalisation. L'approche hybride proposée constitue ainsi un outil pertinent pour renforcer l'aide à la décision et la gouvernance des projets d'infrastructures routières, préparant la transition vers la discussion des implications scientifiques et pratiques de l'étude.

5. Discussion

Les résultats de cette étude confirmentla pertinence de l'approche hybride combinant le clustering non superviséet la modélisation prédictive pour l'analyse des écarts entre conception et réalisation des projets d'infrastructures routières. La segmentation des projets en profils homogènes met en évidence que les dérives de coûts et de délais s'inscrivent dans des configurations récurrentes, liéesà des caractéristiques techniques, financières et contextuelles spécifiques, et non à des événements isolés ou aléatoires[44].

Les performances des modèles prédictifs montrent qu'il est possible d'anticiper, avec un niveau de fiabilité satisfaisant, les projets susceptibles de présenter des écarts significatifs. L'importance relative des variables financières initiales et des durées planifiées souligne le rôle déterminant des décisions prises en phase de conception, confirmant que de nombreuses dérives trouvent leur origine en amont du cycle de vie des projets[45]. Ces résultats sont cohérents avec les constats de la littérature, tout en apportant une valeur ajoutée méthodologique par l'intégration conjointe d'analyses exploratoires et prédictivesau sein d'un cadre unifié[46].

Sur le plan pratique, l'approche proposée offre des perspectives concrètes pour la gouvernance des projets routiers.Pourles

[40] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
[41] Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The Elements of Statistical Learning*. Springer.
[42] Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 25(3), 289–310.
[43] Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer.
[44] Cantarelli, C. C., Flyvbjerg, B., Molin, E. J. E., & van Wee, B. (2012). Cost overruns in large-scale transportation infrastructure projects. *European Journal of Transport and Infrastructure Research*, 12(1), 5–18
[45] Flyvbjerg, B. (2014). What You Should Know About Megaprojects and Why. *Project Management Journal*, 45(2), 6–19.
[46] Marzouk, M., & Enaba, M. (2019). Predicting cost overruns in construction projects using machine learning. *Journal of Construction Engineering and Management*, 145(1), 04018100.

décideurs publics, elle constitue un outil d'aide à la décision permettantde prioriser les projets et de renforcer le pilotage des investissements publics.

Pour les gestionnaires de projets, elle favorise une gestion plus proactive des risques grâce à l'identification précoce des projets à forte probabilité de dérive. Enfin, pour les ingénieurs et les planificateurs, elle souligne l'intérêt d'intégrer systématiquement les analyses de données et les retours d'expérience des projets antérieurs afin d'améliorer la qualité des estimations et la robustesse des choix techniques[47].

6. Conclusion et perspectives

6.1. Conclusion générale

Cette étude avait pour objectif de proposer une approche analytique innovante pour analyser et réduire les écarts entre conception et réalisation des projets d'infrastructures routières, en mobilisant conjointement les techniques de Data Science et d'Intelligence Artificielle. Les résultats obtenus montrent que la combinaison du clustering non supervisé et de la modélisation prédictive permet d'identifier des profils distincts de projets à risque et d'anticiper, avec une fiabilité satisfaisante, les dérives potentielles de coûts et de délais[48].
L'apport principal de cette recherche réside dans la formalisation d'un cadre hybride intégrant à la fois une lecture exploratoire des données et une capacité prédictive orientée vers l'aide à la décision.

L'Intelligence Artificielle apparaît ainsi comme un levier pertinent pour renforcerla gouvernance des projets routiers, en permettant une identification précocedes risques et une meilleure allocationdes ressources dès les phases amont du cycle de vie des projets[49].

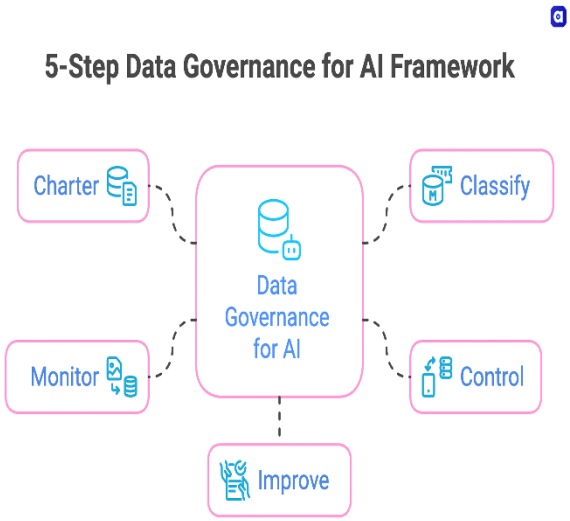


Figure 9. *Synthèse des apports de l'approche hybride IA–Data Science dans la gestion des projets routiers*

6.2. Limites de l'étude

Malgré ces contributions, certaines limites doivent être soulignées. La première concerne la qualité et la disponibilité des données utilisées, qui dépendent fortement des dispositifs institutionnels de collecte et de suivi des projets. Des lacunes ou des incohérences dans les données historiques peuvent influencer la précision des analyseset des prédictions.

La seconde limite tient à la généralisation des résultats, ceux-ci étant principalement fondés sur des projets routiers réalisés dans un contexte institutionnel et opérationnel spécifique.
Bien que le cadre méthodologique soit transférable, les résultats empiriques doivent être interprétés avec prudence lorsqu'ils sont appliqués à d'autres contextes[50].

6.3. Perspectives de recherche

Les perspectives ouvertes par cette recherche sont multiples. L'intégration de données en temps réel issues de capteurs, de systèmes de suivi numérique ou d'images satellitaires pourrait permettre un pilotage dynamique des projetset une détection encore plus précocedes dérives. Par ailleurs, l'exploration d'approches de Deep Learning, capablesde modéliser des relations plus complexes et non linéaires, constitue une piste prometteuse pour améliorer les performances prédictives[51].

7. Évolution future de l'approche : données historiques vers données temps réel et Deep Learning

Enfin, l'extension de cette approche à d'autres pays africains offrirait l'opportunité de comparer les profils de risques projets à l'échelle régionale et de contribuer à la construction de cadres analytiques adaptés aux réalités des pays en développement. Ces perspectives confirment le potentiel de la Data Science et de l'Intelligence Artificielle comme outils structurants pour la transformation durable de la gouvernancedes infrastructures en Afrique.

Références

[1] Abadi, D. J. (2020). Consistency trade-offs in modern distributed systems. *Communications of the ACM, 63*(2), 100–109. https://doi.org/10.1145/3365438
[2] Banque mondiale. (2018). *Infrastructure project appraisal and monitoring guidelines*. World Bank Publications.
[3] Banque mondiale. (2022). *Projet d'appui à la gestion des infrastructures routières en République démocratique du Congo* (Document de projet). World Bank Group.
[4] Batini, C., & Scannapieco, M. (2016). *Data quality: Concepts, methodologies and techniques*. Springer. https://doi.org/10.1007/978-3-319-24106-7
[5] Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
[6] Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324
[7] Bureau Central d'Études du Gouvernement (BCECO). (2023). *Rapport annuel sur l'état des infrastructures en*

[47] OECD. (2020). *Infrastructure Governance: A Framework for Policy Makers*. Paris: OECD Publishing.
[48] Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 25(3), 289–310.
[49] Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach*. 4th ed., Pearson.
[50] Osei-Kyei, R., & Chan, A. P. C. (2017). Developing transport infrastructure in developing countries. *Transport Reviews*, 37(3), 282–305.
[51] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Cambridge, MA: MIT Press.

République démocratique du Congo. Kinshasa, RDC.

[8] Cantarelli, C. C., Flyvbjerg, B., Molin, E. J. E., & van Wee, B. (2012). Cost overruns in large-scale transportation infrastructure projects: Explanations and their theoretical embeddedness. *European Journal of Transport and Infrastructure Research, 12*(1), 5–18.

[9] Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Sage Publications.

[10] Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). Wiley.

[11] Flyvbjerg, B. (2014). What you should know about megaprojects and why: An overview. *Project Management Journal, 45*(2), 6–19. https://doi.org/10.1002/pmj.21409

[12] Fonds National d'Entretien Routier (FONER). (2023). *Rapport d'activités et d'exécution des programmes d'entretien routier*. Kinshasa, RDC.

[13] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

[14] Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

[15] Inmon, W. H. (2015). *Building the data warehouse* (4th ed.). Wiley.

[16] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters, 31*(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

[17] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

[18] Kerzner, H. (2017). *Project management: A systems approach to planning, scheduling, and controlling* (12th ed.). Wiley.

[19] Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2007). Supervised machine learning: A review of classification techniques. *Informatica, 31*(3), 249–268.

[20] Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer.

[21] Love, P. E. D., Sing, C. P., Wang, X., & Irani, Z. (2013). Overruns in transportation infrastructure projects. *Journal of Management in Engineering, 29*(4), 312–325. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000162

[23] Marzouk, M., & Enaba, M. (2019). Predicting cost overruns in construction projects using machine learning algorithms. *Journal of Construction Engineering and Management, 145*(1), 04018100. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001580

[24] Ministère des Infrastructures, Travaux Publics et Reconstruction (ITPR). (2023a). *Stratégie nationale de développement des infrastructures routières 2023–2030*. Kinshasa, RDC.

[25] Ministère des Infrastructures, Travaux Publics et Reconstruction (ITPR). (2023b). *Cadre institutionnel de suivi-évaluation des projets d'infrastructures publiques*. Kinshasa, RDC.

[26] Organisation de Coopération et de Développement Économiques (OCDE). (2020). *Infrastructure governance: A framework for policy makers*. OECD Publishing. https://doi.org/10.1787/6c38c9f5-en

[27] Osei-Kyei, R., & Chan, A. P. C. (2017). Developing transport infrastructure in developing countries. *Transport Reviews, 37*(3), 282–305. https://doi.org/10.1080/01441647.2016.1232918

[28] Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC. *Journal of Machine Learning Technologies, 2*(1), 37–63.

[29] Programme des Nations Unies pour le Développement (PNUD). (2021). *Diagnostic de la gouvernance des infrastructures en République démocratique du Congo*. Kinshasa, RDC.

[30] Provost, F., & Fawcett, T. (2013). *Data science for business*. O'Reilly Media.

[31] Russell, S., & Norvig, P. (2021). *Artificial intelligence : A modern approach* (4th ed.). Pearson.

[32] Shmueli, G. (2010). To explain or to predict? *Statistical Science, 25*(3), 289–310. https://doi.org/10.1214/10-STS330

[33] Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.

[34] World Bank. (2018). *Infrastructure project appraisal and monitoring guidelines*. World Bank Publications.