



## VERS UNE MODELISATION ROBUSTE ET EXPLICABLE DES PERFORMANCES ETUDIANTES A MADAGASCAR : UNE APPROCHE DATA SCIENCE INTEGRANT APPRENTISSAGE AUTOMATIQUE ET ANALYSE SPATIO-TEMPORELLE

<sup>1</sup>ZAFINIRINA Tsanganavelo, <sup>2</sup>RAZAFIARISERA Ralay Tiana, <sup>3</sup>RAKOTOSON Tolontsoa,  
<sup>4</sup>TOVONIRINA Mamiharizo Jackie

1- ECOLE DOCTORALE EN SCIENCES ET TECHNIQUES DE L'INGENIERIE ET DE L'INNOVATION (EDSTII) UNIVERSITE D'ANTANANARIVO

2- ESPA - UNIVERSITE D'ANTANANARIVO

3- UNIVERSITE D'ANTANANARIVO

**RESUME :** Cette étude propose une approche de modélisation robuste et explicable des performances académiques fondée sur l'intégration de la Data Science, de l'apprentissage automatique et de l'analyse spatio-temporelle. À partir de données longitudinales couvrant six universités malgaches sur la période 2002–2020, l'analyse met en évidence des disparités significatives des taux de réussite, variant de 45,9 % à 98,9 %. L'analyse en composantes principales montre que les deux premiers axes expliquent 76,52 % de la variance (Dim1 = 60,38 %, Dim2 = 16,14 %), soulignant le rôle structurant des niveaux Licence et Master. Les corrélations inter-universitaires révèlent des relations modérées, avec un coefficient maximal de 0,91 entre Antananarivo et la tendance globale. L'analyse de dispersion met en évidence une variabilité notable, notamment à Fianarantsoa et Toliara. Enfin, une convergence progressive des taux de réussite entre 60 % et 75 % est observée en 2020. Cette approche permet d'identifier les facteurs clés et de proposer un modèle prédictif fiable et interprétable.

**Mots-clés :** Modélisation ; Intelligence artificielle ; Data Science ; Analyse spatio-temporelle ; Performance académique

**ABSTRACT:** This study proposes a robust and explainable modeling approach to academic performance based on the integration of Data Science, machine learning, and spatio-temporal analysis. Using longitudinal data from six Malagasy universities over the period 2002–2020, the analysis highlights significant disparities in success rates, ranging from 45.9% to 98.9%. Principal Component Analysis shows that the first two components explain 76.52% of the total variance (Dim1 = 60.38%, Dim2 = 16.14%), emphasizing the structural role of Bachelor's and Master's levels. Inter-university correlations reveal moderate relationships, with a maximum coefficient of 0.91 between Antananarivo and the overall trend. Dispersion analysis indicates notable variability, particularly in Fianarantsoa and Toliara. Furthermore, a gradual convergence of success rates between 60% and 75% is observed in 2020. This approach enables the identification of key determinants and provides a reliable and interpretable predictive framework.

**Keywords:** Modeling; Artificial Intelligence; Data Science; Spatio-temporal Analysis; Academic Performance

**Digital Object Identifier (DOI):** <https://doi.org/10.5281/zenodo.19910226>

## 1 Introduction

La réussite académique constitue un enjeu majeur pour les systèmes éducatifs, en particulier dans les pays en développement où elle est étroitement liée au renforcement du capital humain et au développement socio-économique [1]. Dans ce contexte, l'exploitation des données éducatives à travers les approches de Data Science et d'Intelligence Artificielle (IA) offre de nouvelles perspectives pour comprendre, prédire et améliorer les performances étudiantes [2]. La modélisation prédictive permet notamment d'identifier précocement les étudiants à risque et de soutenir la prise de décision institutionnelle [3]. Cependant, plusieurs défis persistent, notamment l'intégration de données hétérogènes, la gestion de la variabilité temporelle et la nécessité de développer des modèles robustes et explicables [4]. Les approches traditionnelles montrent souvent des limites face à la complexité des systèmes éducatifs, d'où l'intérêt croissant pour les méthodes d'apprentissage automatique [5]. Par ailleurs, l'analyse spatio-temporelle permet de mieux capturer les dynamiques évolutives des performances académiques et les disparités entre institutions [6]. Dans le contexte malgache, peu d'études ont proposé un cadre intégré combinant modélisation prédictive, explicabilité et analyse multi-niveau. Cette recherche vise ainsi à combler cette lacune en proposant une approche robuste et contextualisée, capable de prendre en compte les spécificités locales tout en s'appuyant sur des méthodes avancées de Data Science [7].

## 2 Matériels et méthodes

### 2.1 Matériels

La présente étude repose sur une approche méthodologique intégrée combinant collecte de données institutionnelles, traitement analytique et modélisation prédictive. Les données utilisées proviennent principalement du Ministère de l'Enseignement Supérieur et de la Recherche Scientifique (MESUPRES) et de l'Institut National de la Statistique (INSTAT) de Madagascar, couvrant des informations académiques longitudinales relatives aux taux de réussite, aux types de diplômes et aux établissements universitaires sur plusieurs années. Ces données, de nature hétérogène, ont fait l'objet d'un processus rigoureux de prétraitement incluant le nettoyage, la gestion des valeurs manquantes, la normalisation et l'intégration multi-sources afin d'assurer leur qualité et leur cohérence analytique [8]. L'environnement de développement repose sur le langage Python, avec l'utilisation de Jupyter Notebook pour la reproductibilité des analyses et l'expérimentation itérative. Plusieurs bibliothèques spécialisées en Data Science et en apprentissage automatique ont été mobilisées, notamment Pandas et NumPy pour la manipulation des données, Matplotlib et Seaborn pour la visualisation, ainsi que Scikit-learn pour la mise en œuvre des algorithmes de Machine Learning [9].

### 2.2 Méthodes

Des méthodes d'analyse statistique telles que l'analyse en composantes principales et l'analyse de corrélation ont été utilisées pour réduire la dimensionnalité et identifier les relations structurelles entre variables [10]. La conception de l'architecture de modélisation repose sur une approche modulaire intégrant des étapes de sélection de variables, d'entraînement, de validation et d'évaluation des modèles. Plusieurs algorithmes supervisés, incluant les modèles de régression, les arbres de décision et les méthodes d'ensemble, ont été implémentés et comparés afin d'optimiser les performances prédictives en termes de précision et de robustesse [11]. Une attention particulière a été accordée à l'explicabilité des modèles à travers l'analyse de l'importance des variables et l'interprétation des contributions individuelles [12]. Enfin, des mécanismes de validation croisée et d'analyse temporelle ont été intégrés pour garantir la stabilité et la généralisabilité des résultats dans le temps [13].

## 3 Résultats et discussions

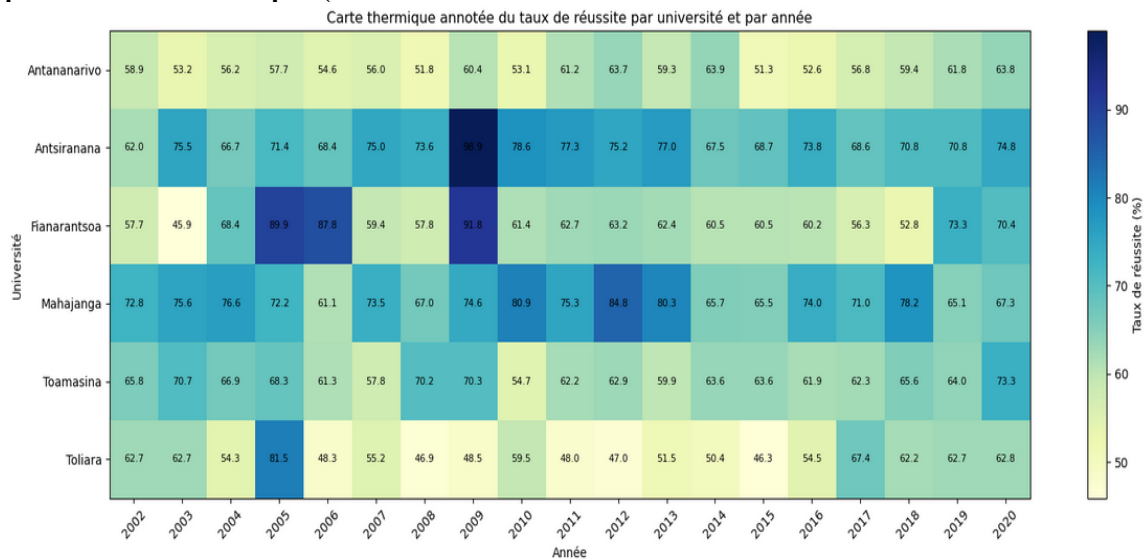
### 3.1 Résultats

#### 3.1.1 Analyse spatio-temporelle des taux de réussite universitaire : visualisation comparative des performances académiques (2002–2020)

La figure ci-dessus présente une cartographie thermique des taux de réussite académique observés dans plusieurs universités malgaches sur la période 2002–2020, mettant en évidence les variations temporelles et les disparités institutionnelles. Cette représentation permet d'identifier visuellement les tendances globales, les périodes de

progression ou de régression, ainsi que les différences structurelles entre établissements, constituant ainsi une base empirique essentielle pour la modélisation prédictive des performances académiques.

**Figure 1: Analyse spatio-temporelle des taux de réussite universitaire : visualisation comparative des performances académiques (2002–2020)**



Source: Auteur 2026

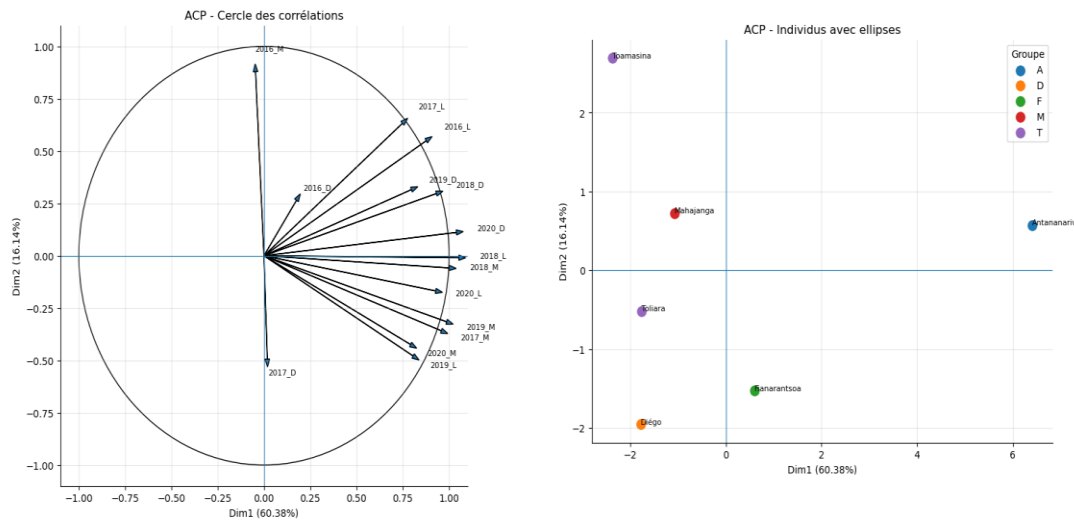
**Interprétation**

L’analyse des données révèle que certaines universités, notamment Antsiranana, affichent des performances relativement élevées et stables avec des taux atteignant un pic de 98,9 % en 2009 et se maintenant majoritairement au-dessus de 70 % sur l’ensemble de la période. Cette stabilité suggère l’existence de facteurs structurels favorables ou de pratiques pédagogiques efficaces pouvant servir de référence dans la modélisation. À l’inverse, des fluctuations marquées sont observées à Fianarantsoa, avec une forte variabilité allant de 45,9 % en 2003 à 91,8 % en 2009, traduisant une instabilité potentielle des conditions académiques ou des cohortes étudiantes. L’université d’Antananarivo présente une évolution modérée avec des taux oscillant entre 51,3 % et 63,9 %, indiquant une relative constance mais à un niveau intermédiaire, ce qui pourrait refléter des contraintes structurelles ou une forte hétérogénéité des profils étudiants. Mahajanga se distingue par des performances globalement élevées, culminant à 84,8 % en 2012, bien que certaines baisses ponctuelles apparaissent après 2013, ce qui pourrait suggérer une influence de facteurs contextuels ou organisationnels. Toamasina montre une stabilité moyenne avec des taux compris entre 54,7 % et 73,3 %, tandis que Toliara présente les niveaux les plus faibles et les plus irréguliers, avec des valeurs descendant jusqu’à 46,3 % en 2015, ce qui en fait un cas pertinent pour l’identification des facteurs de risque d’échec académique. D’un point de vue temporel, une tendance générale d’amélioration est observable entre 2005 et 2012 dans plusieurs universités, suivie d’une phase plus fluctuante entre 2013 et 2020, ce qui met en évidence la nécessité d’intégrer la dimension temporelle et la dérive des données dans les modèles prédictifs. Ces résultats confirment l’hypothèse selon laquelle les performances passées constituent des indicateurs pertinents pour la prédiction future, tout en soulignant l’importance d’intégrer des variables contextuelles pour expliquer les variations observées.

**3.1.2 Analyse en composantes principales de l’évolution des performances académiques par type de diplôme (2016–2020)**

La figure ci-après présente une analyse en composantes principales appliquée aux taux de réussite par type de diplôme (Licence, Master et Doctorat) sur la période 2016–2020, permettant de synthétiser l’information multidimensionnelle en deux axes principaux expliquant 76,52 % de la variance totale (Dim1 = 60,38 % ; Dim2 = 16,14 %). Le cercle des corrélations met en évidence les relations entre les variables temporelles et les types de diplômes, tandis que la projection des individus permet d’identifier les similarités et les oppositions entre universités selon leurs profils de performance.

**Figure 2: Analyse en composantes principales de l'évolution des performances académiques par type de diplôme (2016–2020)**



Source : Auteur 2026

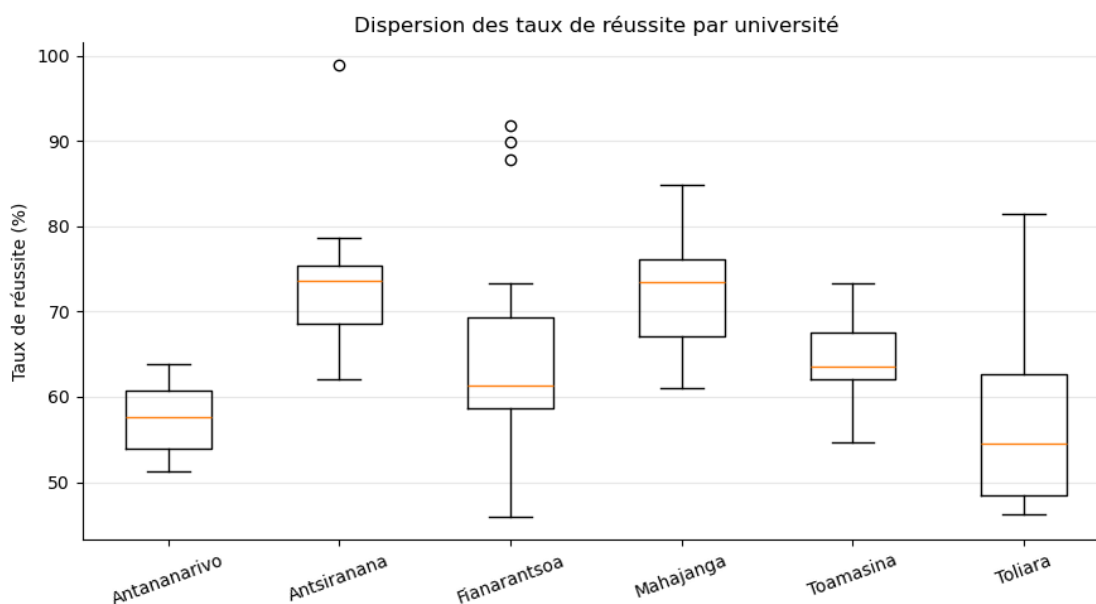
### Interprétation

L'axe principal Dim1, qui explique à lui seul 60,38 % de la variance, est fortement corrélé positivement avec la majorité des variables, notamment les taux de Licence et de Master sur l'ensemble de la période, en particulier pour les années récentes comme 2018, 2019 et 2020 où les vecteurs (2018\_L, 2019\_L, 2020\_L, 2019\_M, 2020\_M) sont fortement alignés et proches du cercle, traduisant une contribution importante et une cohérence structurelle des performances récentes. Cette configuration indique que les universités ayant des valeurs élevées sur cet axe présentent globalement de bons taux de réussite, notamment au niveau Licence et Master. L'axe Dim2, expliquant 16,14 % de la variance, semble davantage structuré par certaines spécificités temporelles, notamment le Doctorat avec 2017\_D orienté négativement et 2016\_M fortement positif, traduisant des variations plus ponctuelles ou spécifiques à certains cycles. L'analyse du positionnement des individus montre une différenciation nette des universités. Antananarivo se situe fortement du côté positif de Dim1 (environ 6,3), ce qui traduit un niveau globalement élevé et stable de performance sur les différents types de diplômes. Mahajanga, positionnée également du côté positif de Dim1 mais plus modérément (environ 1,0), présente un profil performant mais moins marqué. À l'opposé, Diego et Toliara affichent des coordonnées négatives sur Dim1 (environ -1,8 et -1,6 respectivement), indiquant des niveaux de réussite plus faibles ou plus irréguliers. Fianarantsoa se distingue par une position intermédiaire sur Dim1 (environ 0,6) mais fortement négative sur Dim2 (environ -1,5), suggérant des variations spécifiques selon les types de diplômes. Toamasina, quant à elle, présente une valeur négative sur Dim1 (environ -2,4) mais fortement positive sur Dim2 (environ 2,7), ce qui traduit un profil atypique, probablement lié à des performances différenciées selon les cycles ou les années. Le cercle des corrélations montre également une forte proximité angulaire entre les variables de Licence et de Master sur les années récentes, indiquant une corrélation positive élevée entre ces deux cycles, ce qui confirme l'hypothèse d'une continuité des performances académiques dans le parcours étudiant. En revanche, certaines variables de Doctorat présentent des orientations différentes, traduisant une dynamique plus indépendante. Ces résultats confirment que les performances passées, en particulier aux niveaux Licence et Master, constituent des indicateurs robustes pour la prédiction des performances futures, tout en soulignant l'importance de prendre en compte les spécificités temporelles et structurelles propres à chaque type de diplôme dans la modélisation prédictive.

### 3.1.3 Analyse comparative de la dispersion des taux de réussite universitaire : variabilité et stabilité des performances académiques

La figure suivante présente une analyse de la dispersion des taux de réussite par université à travers un diagramme en boîtes, permettant d'évaluer simultanément la tendance centrale, la variabilité et la présence de valeurs atypiques. Cette représentation offre une lecture synthétique des différences structurelles entre établissements en termes de stabilité des performances académiques, élément essentiel pour la conception de modèles prédictifs robustes.

**Figure 3: Analyse comparative de la dispersion des taux de réussite universitaire : variabilité et stabilité des performances académiques**



Source : Auteur 2026

### Interprétation

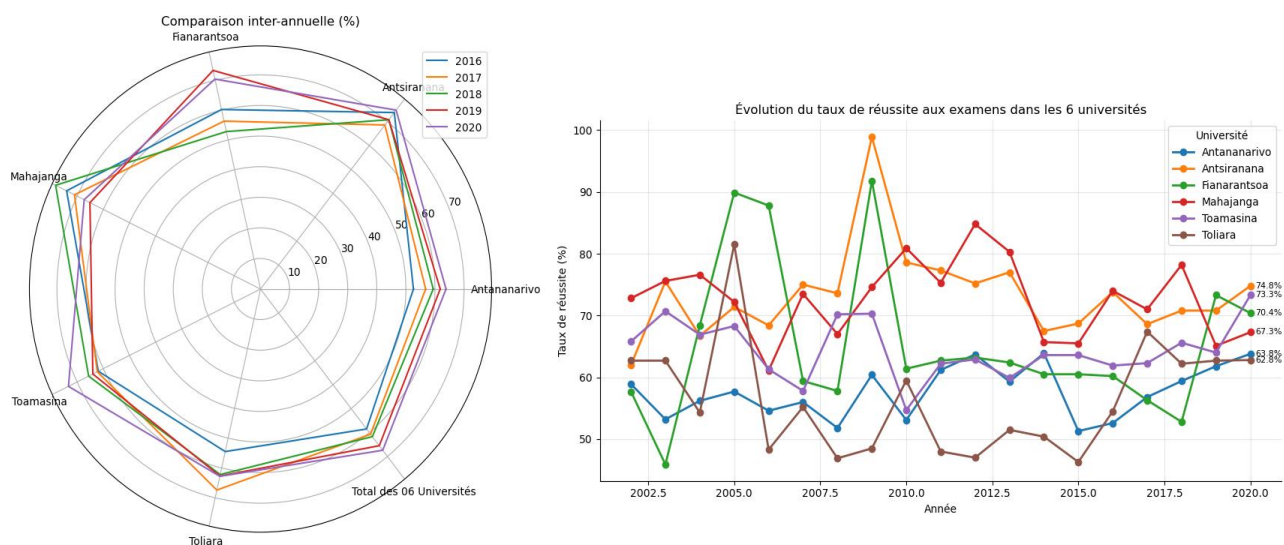
L'examen des médianes montre que Mahajanga et Antsiranana présentent les niveaux de réussite les plus élevés, avec des médianes respectives proches de 73 % et 74 %, traduisant des performances globalement supérieures. Antananarivo et Toamasina affichent des médianes intermédiaires autour de 57 % et 63 %, indiquant des performances modérées mais relativement stables. Fianarantsoa présente une médiane d'environ 61 %, tandis que Toliara se distingue par la médiane la plus faible, proche de 54 %, confirmant un niveau globalement inférieur de réussite académique. L'analyse de la dispersion, à travers l'étendue interquartile, révèle que Mahajanga et Fianarantsoa présentent une variabilité plus importante, avec des valeurs s'étendant approximativement de 61 % à 85 % pour Mahajanga et de 46 % à 74 % pour Fianarantsoa. Cette variabilité traduit une instabilité potentielle des performances, suggérant que les résultats peuvent être influencés par des facteurs contextuels fluctuants. À l'inverse, Antananarivo présente une dispersion plus faible, avec des valeurs concentrées entre environ 51 % et 64 %, ce qui témoigne d'une certaine homogénéité des résultats au fil du temps. La présence de valeurs aberrantes est particulièrement marquée pour Antsiranana et Fianarantsoa. Antsiranana présente un point extrême proche de 99 %, indiquant une performance exceptionnelle ponctuelle, tandis que Fianarantsoa affiche plusieurs valeurs élevées autour de 88 % à 92 %, traduisant des pics de réussite significatifs mais non systématiques. Ces observations suggèrent que certaines universités peuvent atteindre des niveaux élevés de performance, mais de manière irrégulière. Toliara présente une dispersion relativement large, allant d'environ 46 % à plus de 80 %, avec une médiane basse, ce qui traduit une forte instabilité combinée à des performances globalement faibles. Toamasina, quant à elle, présente une variabilité modérée avec des valeurs comprises entre environ 55 % et 73 %, suggérant un profil plus stable mais sans performances extrêmes. Ces résultats mettent en évidence que la variabilité des performances constitue un facteur clé à intégrer dans les modèles prédictifs. Les universités présentant une forte dispersion nécessitent des modèles capables de capturer les fluctuations et la dérive des données, tandis que celles plus stables peuvent être modélisées avec une meilleure fiabilité. Cette analyse confirme également que la seule moyenne des performances est insuffisante pour caractériser la réussite académique, et que la prise en compte de la dispersion améliore la compréhension et la prédiction des trajectoires étudiantes.

### 3.1.4 Analyse comparative et dynamique interannuelle des taux de réussite universitaires (2016–2020) et évolution longitudinale (2002–2020)

La figure ci-dessous combine une représentation radar de la comparaison interannuelle des taux de réussite entre 2016 et 2020 et une analyse longitudinale de l'évolution des performances académiques sur la période 2002–2020 pour six universités. Cette double visualisation permet d'appréhender à la fois les dynamiques récentes par

établissement et les tendances structurelles de long terme, offrant ainsi une base analytique pertinente pour l'évaluation et la modélisation des performances académiques.

**Figure 4: Analyse comparative et dynamique interannuelle des taux de réussite universitaires (2016–2020) et évolution longitudinale (2002–2020)**



Source : Auteur 2026

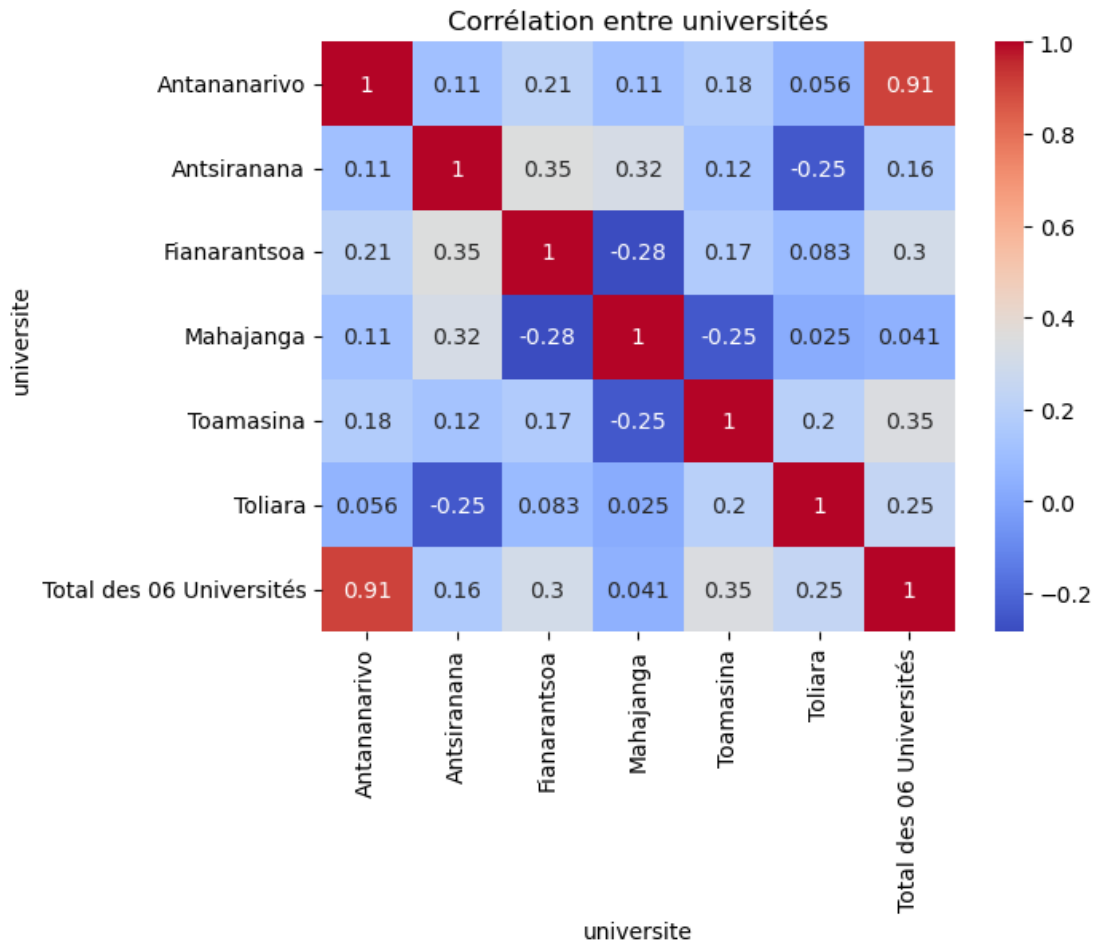
### Interprétation

L'analyse radar met en évidence une amélioration globale des taux de réussite entre 2016 et 2020 dans la majorité des universités. Antananarivo progresse de 58,9 % en 2016 à 63,8 % en 2020, traduisant une évolution modérée mais continue. Antsirana maintient des performances élevées avec des valeurs proches de 75 % en 2017 et 2020, confirmant une stabilité à un niveau supérieur. Fianarantsoa présente une progression marquée, passant d'environ 60 % en 2016 à plus de 70 % en 2020, malgré des fluctuations intermédiaires. Mahajanga atteint des niveaux élevés avec un pic proche de 77 % en 2018, avant une légère stabilisation autour de 73 % en 2020. Toamasina montre une progression significative, passant d'environ 63 % à plus de 73 %, tandis que Toliara affiche une amélioration notable entre 2016 (environ 63 %) et 2017 (près de 75 %), suivie d'une stabilisation autour de 70 % en 2020. L'évolution longitudinale confirme ces tendances tout en mettant en évidence des fluctuations importantes sur le long terme. Antsirana présente le niveau maximal observé avec 98,9 % en 2009, suivi d'une stabilisation entre 70 % et 78 % après 2010. Fianarantsoa se caractérise par une forte variabilité, avec un minimum de 45,9 % en 2003 et un maximum de 91,8 % en 2009, illustrant une instabilité structurelle. Mahajanga montre une tendance globalement élevée avec des valeurs oscillant entre 65 % et 85 %, notamment un pic à 84,8 % en 2012. Antananarivo reste relativement stable avec des taux compris entre 51 % et 64 %, traduisant une performance constante mais modérée. Toamasina présente une évolution modérée avec des valeurs généralement comprises entre 60 % et 70 %, bien que certaines baisses apparaissent autour de 54 % en 2010. Toliara se distingue par les niveaux les plus faibles et les plus fluctuants, avec des valeurs descendant jusqu'à 46,3 % en 2015 et atteignant ponctuellement plus de 80 % en 2005, ce qui souligne une forte instabilité. Les valeurs finales en 2020 confirment une hiérarchie des performances avec Antsirana en tête (74,8 %), suivie de Toamasina (73,3 %) et Fianarantsoa (70,4 %), tandis que Mahajanga (67,3 %), Antananarivo (63,8 %) et Toliara (62,8 %) occupent des positions intermédiaires à faibles. Cette convergence relative des taux autour de 60 % à 75 % en fin de période suggère une homogénéisation progressive des performances académiques.

### 3.1.5 Analyse des corrélations inter-universitaires des taux de réussite : identification des dynamiques communes et divergentes

La figure ci-après présente une matrice de corrélation des taux de réussite entre les différentes universités, permettant d'évaluer les relations linéaires existantes entre leurs évolutions respectives. Cette analyse met en évidence les similitudes, les interdépendances ainsi que les divergences structurelles entre établissements, constituant un élément clé pour comprendre les dynamiques globales du système universitaire et améliorer la robustesse des modèles prédictifs.

Figure 5: Analyse des corrélations inter-universitaires des taux de réussite : identification des dynamiques communes et divergentes



Source : Auteur 2026

**Interprétation**

L'examen de la matrice révèle une très forte corrélation entre Antananarivo et le total des six universités, avec un coefficient de 0,91, indiquant que les variations observées à Antananarivo influencent fortement la tendance globale. Cette relation suggère que cette université joue un rôle central dans la dynamique nationale des performances académiques. Des corrélations positives modérées sont également observées entre Fianarantsoa et Antsirana (0,35), ainsi qu'entre Toamasina et le total (0,35), traduisant des évolutions relativement synchronisées entre ces entités. Antsirana présente des corrélations positives avec plusieurs universités, notamment Mahajanga (0,32) et Fianarantsoa (0,35), ce qui indique des dynamiques de performance partiellement similaires. Fianarantsoa montre également une corrélation positive avec Antananarivo (0,21) et Toamasina (0,17), bien que ces relations restent modérées, suggérant une influence partagée mais non dominante. En revanche, certaines corrélations négatives mettent en évidence des comportements divergents. Mahajanga présente une corrélation négative avec Fianarantsoa (-0,28) et Toamasina (-0,25), ce qui indique que les périodes de hausse de performance dans ces universités ne coïncident pas nécessairement. De même, Toliara affiche une corrélation négative avec Antsirana (-0,25), traduisant une dynamique opposée entre ces deux établissements. Globalement, la majorité des coefficients se situent entre 0,05 et 0,35, ce qui traduit des corrélations faibles à modérées et suggère que, bien que certaines tendances communes existent, chaque université conserve une dynamique propre. Cette relative indépendance renforce la nécessité d'intégrer des variables spécifiques à chaque contexte institutionnel dans les modèles prédictifs.

### 3.2 Discussions

Les résultats obtenus mettent en évidence la pertinence d'une approche de modélisation intégrée pour l'analyse des performances académiques dans un contexte universitaire à ressources limitées. La variabilité observée des taux de réussite, allant de 45,9 % à 98,9 %, confirme la complexité intrinsèque des systèmes éducatifs et la nécessité d'utiliser des modèles capables de capturer des dynamiques non linéaires et hétérogènes. La forte contribution des axes principaux de l'ACP (76,52 % de variance expliquée) souligne l'importance des variables liées aux cycles Licence et Master dans la structuration des performances, ce qui est cohérent avec les travaux montrant que les parcours académiques précoces influencent significativement les trajectoires futures [14]. Les corrélations modérées entre universités, avec un maximum de 0,91 pour Antananarivo, indiquent à la fois l'existence de tendances globales et la persistance de spécificités locales. Cette dualité renforce l'intérêt d'une modélisation multi-niveaux capable d'intégrer des effets globaux tout en tenant compte des particularités institutionnelles, comme le suggèrent plusieurs études en learning analytics [15]. Par ailleurs, l'identification de fortes dispersions dans certaines universités, notamment Fianarantsoa et Toliara, met en évidence des phénomènes d'instabilité qui peuvent être liés à des facteurs contextuels tels que les conditions pédagogiques ou socio-économiques. L'intégration de techniques d'apprentissage automatique, combinée à des approches explicables, apparaît essentielle pour améliorer la fiabilité et l'interprétabilité des prédictions. En effet, les modèles explicables permettent non seulement d'atteindre des performances élevées, mais également de fournir des informations exploitables pour la prise de décision académique [16]. Enfin, la prise en compte de la dimension temporelle, notamment à travers l'analyse longitudinale, confirme que les performances passées constituent des indicateurs robustes des résultats futurs, tout en nécessitant des mécanismes de surveillance pour gérer la dérive des données [17].

### 4 Conclusion

Cette étude a permis de démontrer l'intérêt d'une approche intégrée de modélisation robuste et explicable pour l'analyse et la prédiction des performances académiques dans l'enseignement supérieur malgache. L'exploitation de données longitudinales issues de sources institutionnelles a mis en évidence des disparités significatives entre universités, avec des taux de réussite variant de 45,9 % à 98,9 %, ainsi qu'une tendance récente à la convergence des performances entre 60 % et 75 %. L'analyse statistique et spatio-temporelle a révélé des dynamiques différenciées selon les établissements, confirmant l'importance de prendre en compte à la fois les effets globaux et les spécificités locales. La mise en œuvre de techniques de Data Science et d'apprentissage automatique a permis de proposer un cadre méthodologique capable d'identifier les variables déterminantes de la réussite académique tout en garantissant la robustesse et l'interprétabilité des modèles. L'intégration de mécanismes d'analyse de la variabilité et de la dérive temporelle constitue un apport majeur pour assurer la stabilité des prédictions dans le temps. Sur le plan opérationnel, cette recherche offre des perspectives concrètes pour le développement d'outils d'aide à la décision permettant l'identification précoce des étudiants à risque et l'amélioration des politiques pédagogiques. Elle ouvre également la voie à des travaux futurs visant à intégrer davantage de données comportementales et contextuelles afin d'affiner la précision et la généralisabilité des modèles prédictifs.

### REFERENCES

- [1] UNESCO. (2021). Reimagining our futures together: A new social contract for education. Paris: UNESCO Publishing.
- [2] Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In Learning analytics (pp. 61–75). Springer. [https://doi.org/10.1007/978-1-4614-3305-7\\_4](https://doi.org/10.1007/978-1-4614-3305-7_4)
- [3] Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3), e1355. <https://doi.org/10.1002/widm.1355>
- [4] Kotsiantis, S. B. (2012). Use of machine learning techniques for educational proposes: A decision support system for forecasting students' grades. Artificial Intelligence Review, 37(4), 331–344. <https://doi.org/10.1007/s10462-011-9234-x>
- [5] Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, 6, 52138–52160.

- [6] Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- [7] Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 46(5), 30–40.
- [8] Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and techniques* (3rd ed.). Elsevier.
- [9] Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- [10] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A*, 374(2065), 20150202.
- [11] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- [12] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.
- [13] Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer.
- [14] Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition*. University of Chicago Press.
- [15] Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5–6), 304–317.
- [16] Molnar, C. (2020). *Interpretable machine learning*. Lulu.com.
- [17] Lu, J., Liu, A., Dong, F., et al. (2019). Learning under concept drift: A review. *IEEE Transactions on Knowledge and Data Engineering*, 31(12), 2346–2363.