International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 10, October 2026)

Prédiction Des Cas De Rougeole Par L'utilisation D'un Modèle De Stacking Combinant Random Forest, Adaboost Et Mlpregressor : Une Approche Basée Sur L'apprentissage Supervisé

Tshimanga Mbuyi Fortunat[1], Kalonji Mukendi John[2], Batubenga Muambanzambi Jean Didier[3], Kantshia Bakatubia Evariste[4], TSHIELA NKUNA Marie Alice[5]

[1]*Faculté de Sciences et Technologie, Département de Mathématiques, Statistique et Informatique, Université Pédagogique de Kananga (UPKAN) & Kananga, RDC*

[2]*Faculté de Sciences et Technologie, Département de Mathématiques, Statistique et Informatique, Université Officielle de Mbujimayi (UOM) & Mbujimayi, RDC*

[3]*Faculté de Sciences et Technologie, Département de Mathématiques, Statistique et Informatique, Université de Kinshasa (UNIKIN) & Kinshasa, RDC*

[4] *Faculté de Sciences et Technologie, Département de Mathématiques, Statistique et Informatique, Université Pédagogique de Kinshasa (UPN) & Kinshasa, RDC*

[5]*Faculté de Sciences et Technologie, Département de Mathématiques, Statistique et Informatique, Université Notre Dame de Kananga (UKA) & Kananga, RDC*

***Résumé*-- La rougeole reste l'un des problèmes de santé publique les plus tenaces ; elle frappe surtout les pays à revenu faible et intermédiaire, là où la couverture vaccinale demeure insuffisante. En 2024, près de 4,9 millions d'enfants n'ont pas atteint leur cinquième anniversaire à l'échelle mondiale. On comptait par ailleurs 2,1 millions de décès chez les enfants plus âgés, les adolescents et les jeunes [1].**

Ce travail propose un modèle hybride, construit par Stacking, associant Random Forest, AdaBoost et un réseau de neurones artificiels (MLPRegressor), afin de prédire les cas de rougeole. La démarche repose sur la fusion du jeu de données Life Expectancy WHO avec un second jeu de données portant sur la couverture vaccinale. Une fois le prétraitement effectué, les variables les plus influentes ont été sélectionnées de façon automatique grâce au Random Forest ; les hyperparamètres, eux, ont fait l'objet d'une optimisation par GridSearchCV, dans le cadre d'une validation croisée imbriquée. Trois indicateurs ont servi à juger la qualité du modèle : le RMSE, le MAE et le coefficient de détermination (R^2). Le modèle rend compte de plus de 70 % de la variance lorsque les données sont transformées logarithmiquement ; il conserve, à ce stade, une bonne tenue sur l'échelle originale, malgré des flambées épidémiques marquées. Cette approche illustre l'intérêt des méthodes d'ensemble pour anticiper les maladies infectieuses. Elle pourrait, en pratique, nourrir la conception d'outils d'aide à la décision en santé publique.

Mots-clés : Rougeole, Apprentissage supervisé, Forêt aléatoire, Réseaux de neurones, Stacking, validation croisée imbriquée.

I. Introduction

Peu de maladies rivalisent avec la rougeole en matière de contagiosité. Un individu infecté peut, selon les estimations, exposer entre 20 et 25 personnes au virus ; dans une population n'ayant jamais rencontré la maladie, les trois quarts finiront par la contracter. Une grande partie de ce que l'on sait aujourd'hui sur les fondements épidémiologiques de l'infection vient des études menées sur des épidémies insulaires, celle de Peter Panum tenant une place particulière. Ce jeune médecin danois avait été envoyé aux îles Faroe en 1846, lors d'une épidémie de grande ampleur. Il avait su percevoir le caractère extrêmement contagieux de la maladie, identifier une période d'incubation de 14 jours, repérer l'immunité durable chez les habitants les plus âgés, et avancer l'hypothèse d'une transmission par voie respiratoire.

La rougeole continue de peser lourdement sur la santé publique dans nombre de pays en développement, et cela malgré les avancées réalisées en matière de vaccination. Les méthodes classiques de prévision épidémiologique se heurtent à certaines limites dès lors qu'il s'agit de saisir la part sanitaire, démographique et socio-économique qui conditionne la diffusion des maladies infectieuses.

L'intelligence artificielle et l'apprentissage automatique apportent aujourd'hui des pistes prometteuses pour affiner la prédiction des phénomènes épidémiologiques.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 10, October 2026)

Le Random Forest, l'AdaBoost et les réseaux de neurones artificiels, parmi les techniques disponibles, ont fait la preuve de leur efficacité dans plusieurs champs de la santé publique.

Rares sont pourtant les travaux à avoir tenté une hybridation de ces trois modèles pour prédire les cas de rougeole à partir d'indicateurs sanitaires mondiaux. Construire un modèle hybride RF+AdaBoost+MLPRegressor capable de gagner en précision prédictive sur les cas de rougeole : c'est précisément l'objectif poursuivi ici.

II. Problematique

Dans un contexte global marqué par des dynamiques démographiques de plus en plus complexes, influencées par des facteurs socio-économiques, sanitaires (pandémies, évolution des systèmes de santé) et environnementaux (changement climatique) (Angus, 2013), la prédiction des tendances de mortalité ou de cas revêt une importance capitale pour la santé publique et les politiques de retraite (Bhardwaj, 2022). Les modèles épidémiologiques traditionnelles ont largement montré leur efficacité sur des données linéaires, mais ses performances restent limitées face à la complexité et aux non-linéarités des données démographiques contemporaines (Li, 2011).

Les avancées récentes en apprentissage automatique, en particulier en apprentissage profond, offrent des perspectives prometteuses pour améliorer la précision des prévisions en capturant des interactions complexes dans des jeux de données volumineux (Deprez, 2017; Hainaut, 2018; Nigri, 2019; Richman, 2019; Hong, 2021; Bjerre, 2022; Gyamerah, 2023). Cependant, l'application de ces techniques dans le domaine de la prédiction des de cas de rougeole reste encore sous-explorée.

La question principale est donc :

Comment améliorer la prédiction des cas de rougeole à partir des indicateurs sanitaires et socio-économiques grâce à une approche hybride combinant Random Forest, AdaBoost et Réseaux de Neurones ?

Ce problème n'est pas seulement conceptuel. Il pose aussi des difficultés concrètes, qu'il s'agisse :

- une forte variance ;
- des données bruitées ;
- des valeurs extrêmes ;
- des relations non linéaires complexes.

Plusieurs sous-questions découlent de cette interrogation centrale, touchant à l'intégration des techniques d'apprentissage automatique et profond, au choix des algorithmes, à la validation des modèles et à leurs applications concrètes :

- ✓ *Combinaison des modèles :* Comment intégrer efficacement les techniques d'apprentissage automatique et d'apprentissage profond aux modèles classiques pour créer un modèle hybride tout en maintenant une interprétabilité des résultats ?
- ✓ *Algorithmes adaptés :* Quels algorithmes d'apprentissage automatique et d'apprentissage profond sont les plus performants pour capturer les motifs complexes et les tendances temporelles des cas de rougeole ?
- ✓ *Validation et robustesse :* Quels critères de performance doivent être utilisés pour comparer et valider les modèles hybrides ?
- ✓ *Application pratique :* Quels sont les avantages concrets des modèles hybrides pour les acteurs de la santé publique, des assurances et des politiques de retraite dans la gestion des risques démographiques et financiers ?

III. Hypothese

Afin de répondre à la problématique soulevée, nous formulons les hypothèses suivantes, qui guideront notre démarche de recherche et permettront d'évaluer l'efficacité des modèles hybrides dans la prédiction des cas de rougeole :

- ✓ Associer des techniques d'apprentissage automatique et d'apprentissage profond aux modèles classiques améliorerait sensiblement, en pratique, la précision des prédictions de cas de rougeole, comparée à celle obtenue par les modèles traditionnels pris isolément.
- ✓ Les techniques d'apprentissage automatique (forêt aléatoire, modèle linéaire généralisé, arbre de décision, etc.) et d'apprentissage profond, en particulier les réseaux de neurones sont capables de mieux capturer les dynamiques non linéaires et les incertitudes dans les données de rougeole que les modèles traditionnels.
- ✓ L'usage d'un modèle hybride fondé sur l'approche d'ensemble empilé consolide la robustesse des prédictions de cas de rougeole. Les forces des modèles traditionnels s'y combinent à celles des algorithmes d'apprentissage automatique ; les erreurs de prédiction en sont réduites.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 10, October 2026)

- ✓ Ces modèles hybrides se prêtent à une validation reposant sur des critères de performance précis, à savoir l'erreur quadratique moyenne, l'erreur absolue moyenne et le R^2. Une optimisation vient s'y ajouter. Leur applicabilité à des scénarios variés de prévisions démographiques s'en trouve, à ce stade, garantie.
- ✓ Pour les secteurs de l'assurance et de la santé publique, de tels modèles présentent un intérêt pratique certain. En produisant des prévisions plus précises (et donc plus exploitables sur le terrain), ils permettent une meilleure anticipation des risques démographiques.

Comparée à ce qu'offrent les modèles pris isolément, la combinaison de Random Forest, d'AdaBoost et de Réseaux de Neurones Artificiels rehausse nettement la précision de la prédiction des cas de rougeole.

IV. Objectifs

A. *Objectif général*

Comparée à ce qu'offrent les modèles pris isolément, la combinaison de Random Forest, d'AdaBoost et de Réseaux de Neurones Artificiels rehausse nettement la précision de la prédiction des cas de rougeole.

B. *Objectifs spécifiques*

Pour atteindre cet objectif global, nous avons visé à :

- Collecter, nettoyer et préparer les données d'apprentissage ;
- Sélectionner les variables les plus pertinentes ;
- Construire les modèles RF, AdaBoost et ANN séparément ;
- Développer le modèle hybride RF+ANN+AdaBoost;
- Evaluer les performances du modèle proposé à l'aide des métriques RMSE, MAE et du coefficient de détermination R^2 ;
- Comparer les performances prédictives du modèle hybride à celles de modèles classiques tels que : Random Forest, AdaBoost et Réseaux neuronaux ;

V. Méthodologie

La méthodologie adoptée dans cette étude se décompose en plusieurs étapes clés, allant de la collecte des données à partir de la fusion de deux dataset mondiaux à l'analyse des résultats obtenus en passant par la mise en place et à l'évaluation d'un modèle hybride combinant des modèles RF, AdaBoost et MLPRegressor.

Chaque étape s'appuie sur des méthodes rigoureuses pour assurer la robustesse et la validité des résultats obtenus:

✓ *Collecte de données :*

- Rassembler, à partir de sources fiables (bases de données gouvernementales, organisations de santé, etc.), des données démographiques pertinentes sur le nombre de cas de rougeole.
- Veiller à la qualité et à l'intégrité des données en éliminant valeurs manquantes et doublons. La cohérence des variables fait, en pratique, l'objet d'un contrôle systématique.

✓ *Préparation des données :*

- Soumettre les données à un prétraitement : normalisation, nettoyage, encodage des variables catégorielles, gestion des valeurs manquantes, transformation logarithmique, création de nouvelles variables dérivées des deux dataset. La pertinence du modèle s'en trouve renforcée.
- Diviser les données en ensembles d'entraînement, de validation et de test pour garantir une évaluation robuste et indépendante du modèle.

✓ *Modélisation de RF, AdaBoost et MLPRegressor*

- Calibration des modèles RF, MLPRegressor et AdaBoost : Ajustement des paramètres pour s'assurer que les prévisions correspondent aux données historiques de nouveaux cas.
- Analyse des performances : mesure de l'erreur de prédiction (erreur quadratique moyenne, erreur absolue moyenne). Les modèles classiques montrent ici leurs limites, en particulier lorsqu'il s'agit de capturer des tendances non linéaires et complexes dans les données.

✓ *Développement du modèle hybride.*

- Intégrer des techniques d'apprentissage automatique et des techniques d'apprentissage profond, tel est l'objectif poursuivi ; il s'agit de capturer les non-linéarités que le modèle classique ne parvient pas à modéliser.
- Approche d'ensemble empilé (Stacking RF) : Cette méthode combine les prédictions de plusieurs modèles d'apprentissage automatique comme les forêts aléatoires, AdaBoost et les réseaux neuronaux afin d'améliorer la prédiction et la robustesse des prévisions.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 10, October 2026)

✓ *Entraînement et optimisation du modèle*

- Le modèle hybride est entraîné sur l'ensemble d'entraînement. Les hyperparamètres, quant à eux, sont ajustés à l'aide de la technique de tuning par grille (GridSearchCV), en vue de maximiser les performances prédictives.
- Utilisation de la validation croisée imbriquée (cross-validation) pour évaluer la capacité de généralisation du modèle et prévenir le surapprentissage (overfitting).

✓ *Évaluation des performances :*

- Le modèle hybride est évalué à partir de critères tels que l'erreur quadratique moyenne (MSE), la racine de l'erreur quadratique moyenne (RMSE) et l'erreur moyenne absolue (MAE) ; ces métriques permettent de mesurer la précision et la fiabilité des prévisions. Le modèle est également évalué, en pratique, à l'aide d'autres métriques pertinentes pour l'analyse, notamment l'erreur quadratique moyenne et l'erreur absolue moyenne.
- On compare ensuite les performances du modèle hybride à celles du modèle classique et d'autres techniques d'apprentissage automatique, afin de déterminer l'amélioration obtenue grâce à l'intégration des modèles.

✓ *Analyse des résultats*

- Interpréter les résultats obtenus à l'aide des graphiques et déterminer les implications pratiques pour les acteurs de la santé publique.

VI. Revue De Littérature

Avec le développement de l'apprentissage automatique, plusieurs chercheurs ont proposé des modèles avancés, tels que les forêts aléatoires, les machines à vecteurs de support (SVM) et XGBoost, afin d'améliorer la prédiction des maladies infectieuses. Ces techniques ont permis d'obtenir de meilleures performances grâce à leur capacité à traiter des données multidimensionnelles et très variables. Cependant, malgré ces améliorations, ces modèles restent souvent limités dans leur capacité à prendre en compte simultanément les dépendances.

Alemayehu, M. A. (2024) a mené une étude sur les Algorithmes d'apprentissage automatique pour prédire la rougeole : un abandon vaccinal chez des enfants de 12 à 23 mois en Éthiopie.

Les modèles prédictifs ont été évalués en utilisant la précision, la précision, la mémoire, le score F1 et la zone sous la courbe (AUC). Cette recherche a utilisé les valeurs de Shapley pour interpréter les prédictions individuelles faites par le modèle d'apprentissage automatique le plus performant.

L'algorithme XGBoost a surpassé tous les classificateurs pour prédire l'abandon vaccinal contre la rougeole (précision et valeurs AUC de 73,9 % et 0,813, respectivement).

De même **Alemayehu, M. A. et al. (2025)** ont étudié l'utilisation d'un modèle d'ensemble empilé (stacking) pour prédire l'abandon de la vaccination. Cette étude a fourni des informations sur les facteurs influençant l'abandon de la vaccination contre le Penta 3 en Afrique de l'Est. Pour réduire les taux d'abandon scolaire, les interventions devraient se concentrer sur l'amélioration des opportunités de subsistance maternelle, l'amélioration de l'accès aux soins de santé dans les zones rurales et la promotion des accouchements institutionnels. Tous les modèles d'ensemble empilés ont surpassé les apprenants individuels, le méta-apprenant XGBoost optimisé avec la recherche en grille et LASSO FS obtenant les performances les plus élevées : 93,9 % de précision et 99,4 % d'AUC.

C. Rughiniș et al. (2022) ont exploré les schémas de covariation entre les taux de vaccination (y compris la rougeole) et l'espérance de vie. La variabilité restante du succès vaccinal contre la COVID-19, qui ne peut être identifiée par ces ensembles d'indicateurs, indique une marge de manœuvre considérable pour l'action collective et individuelle en temps de crise. La mobilisation et la coordination dans les campagnes de vaccination des citoyens, professionnels de santé, scientifiques, journalistes et politiciens, entre autres, expliquent au moins une partie de cette variabilité dans la lutte contre l'hésitation et l'inégalité vaccinales.

Natras, R. et al (2022) ces auteurs ont proposé un modèle VTEC basé sur l'apprentissage automatique « apprend » directement à partir des données historiques ou des exemples donnés, qui peuvent être compris comme des expériences passées. L'apprentissage est obtenu en optimisant la performance d'un algorithme d'apprentissage automatique pour une tâche de prédiction VTEC, qui présente la prédiction d'une variable continue, communément appelée régression en apprentissage automatique. Cet article est pertinent pour la méthodologie car il utilise un ensemble de Random Forest, AdaBoost et XGBoost pour la prévision, ce qui est très proche des algorithmes (Random Forest, AdaBoost, Réseaux de Neurones Artificiels) du modèle hybride (Stacking) que nous avons proposé.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 10, October 2026)

Alzamil, L. (2026) a mené une étude pouvant examiner les tendances d'incidence de la rougeole et la couverture vaccinale, ce qui est essentiel pour comprendre le contexte des données utilisées pour la prédiction des cas de rougeole.

VII. Résultats

Ce point porte sur la phase expérimentale de l'étude, dont le but est d'évaluer et de comparer les performances des modèles individuels face au modèle hybride Stacking dans la prédiction des cas de rougeole. Il s'agit avant tout de mesurer la solidité et l'efficacité des modèles à travers différentes métriques, calculées à la fois sur l'échelle logarithmique et sur l'échelle originale.

Puisque la transformation y = np.log1p(Measles) a été retenue, les métriques sont calculées sur l'échelle logarithmique. Le RMSE et le MAE s'expriment donc en unités logarithmiques, et non en nombre brut de cas de rougeole.

A. Analyse des erreurs de prédiction

Tableau I
Les erreurs de prediction sur l'echelle logarithmique

Indicateur	Valeur	Interprétation
RMSE	**1,782**	Sur l'échelle logarithmique, l'erreur quadratique moyenne est relativement faible, ce qui indique que les prédictions sont proches des valeurs observées après transformation.
MAE	**1,257**	L'erreur absolue moyenne est faible sur l'échelle logarithmique, montrant que les écarts moyens entre les valeurs prédites et réelles sont limités.
R²	**0,7132**	Le modèle explique **71,32 %** de la variabilité des cas de rougeole. Environ **28,68 %** de la variation reste non expliquée, ce qui peut être attribué à des facteurs non présents dans le jeu de données ou à la variabilité naturelle du phénomène.

Tableau II
Les erreurs de prediction sur l'echelle originale

Indicateur	Valeur	Interprétation
RMSE	**7 984,10**	En moyenne, les prédictions du modèle s'écartent des valeurs réelles d'environ **7 984 cas**. Le RMSE étant sensible aux grandes erreurs, cette valeur élevée traduit l'influence des fortes flambées épidémiques présentes dans le jeu de données.
MAE	**1 967,60**	En moyenne, les prédictions diffèrent des valeurs observées d'environ **1 968 cas**. Cette mesure est moins influencée par les valeurs extrêmes que le RMSE et représente mieux l'erreur moyenne du modèle.
R²	**0,4689**	Le modèle explique **46,89 %** de la variabilité des cas de rougeole. Environ **53,11 %** de cette variabilité reste inexpliquée par les variables utilisées.

B. Comparaison avec le premier modèle

Tableau III
La comparaison des erreurs de prediction

Modèle	RMSE	MAE	R²
Premier Stacking	7069,69	2920,59	0,5836
Nouveau Stacking (avec transformation logarithmique et optimisation)	1,782*	1,257*	**0,7132**

Combinant Random Forest, AdaBoost et MLPRegressor, avec une optimisation des hyperparamètres via GridSearchCV et une validation croisée imbriquée, le modèle de Stacking atteint un coefficient de détermination (R^2) de **0,7132.** Ce chiffre signifie que **71,32 %** de la variabilité observée des cas de rougeole se trouve expliquée par le modèle. Par rapport au modèle initial, ce gain montre que l'optimisation des hyperparamètres, associée à la transformation logarithmique de la variable cible et à une sélection rigoureuse des variables, a permis de hausser sensiblement les performances prédictives.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 10, October 2026)

C. *Analyse de corrélation des variables*

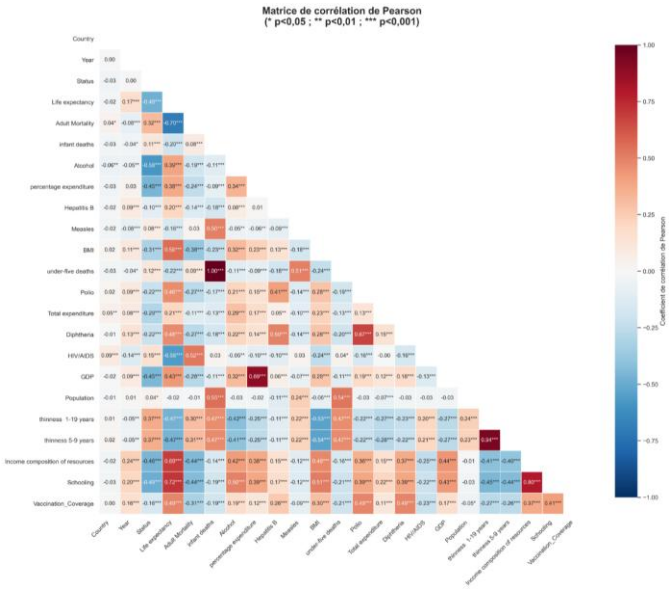


Figure 1 : Corrélation de Pearson des variables utilisées

D. *Analyse*

La plupart des coefficients de corrélation atteignent un seuil statistiquement significatif ($p < 0,05$), ce qui atteste de relations linéaires qui ne relèvent pas du hasard. Entre les cas de rougeole et la mortalité infantile se dessine une corrélation positive modérée ($r = 0,50$), de même qu'avec les décès des enfants de moins de cinq ans ($r = 0,51$) : ces liens traduisent le poids des conditions sanitaires dans la propagation de la maladie. À l'inverse, une corrélation négative apparaît entre les cas de rougeole et l'espérance de vie ($r = -0,16$) ; la couverture vaccinale suit elle aussi une tendance négative, ce qui confirme son rôle protecteur. Du côté des indicateurs de vaccination, Polio et Diphtheria affichent une corrélation forte ($r = 0,67$), signe de la cohérence des programmes de vaccination. Quant aux variables socio-économiques, le niveau de scolarisation et l'indice de composition des ressources présentent une corrélation élevée ($r = 0,80$), résultat logique compte tenu de leur proximité conceptuelle. Ces quelques cas de multicolinéarité n'empêchent pas les coefficients de rester compatibles avec des modèles d'ensemble comme Random Forest, AdaBoost et Stacking, connus pour leur résistance face aux variables corrélées.

VIII. Discussion

Les résultats obtenus montrent que l'association de plusieurs algorithmes complémentaires améliore les performances par rapport à un modèle unique.

En appliquant une transformation logarithmique à la variable cible, on a réduit l'effet des valeurs extrêmes ; cela explique à la fois l'amélioration constatée du coefficient de détermination sur les données transformées et la facilité accrue d'apprentissage du modèle.

Les performances baissent toutefois dès que les prédictions sont réévaluées sur l'échelle originale. Cette situation tient à la présence, dans le jeu de données, de flambées épidémiques très marquées, comme le révèle l'asymétrie élevée de la variable cible. Ces observations extrêmes gonflent fortement les erreurs quadratiques, d'où un RMSE relativement élevé.

Cette difficulté n'enlève rien aux résultats obtenus : le Stacking apparaît comme une approche pertinente pour modéliser des maladies infectieuses dont les distributions sont fortement déséquilibrées.

IX. Perspectives

Les travaux futurs pourront porter sur :

- l'intégration de données climatiques et environnementales ;
- l'ajout de données de mobilité de la population ;
- la comparaison avec des modèles de gradient boosting plus récents (par exemple XGBoost, LightGBM ou CatBoost) ;
- l'évaluation sur des données plus récentes provenant de l'Organisation mondiale de la Santé ;
- le développement d'un système d'alerte précoce en temps réel destiné aux autorités sanitaires.

X. Conclusion

Cette étude a développé un modèle hybride de prédiction des cas de rougeole fondé sur le Stacking de Random Forest, AdaBoost et MLPRegressor.

Plusieurs constats se dégagent des expériences menées:

- l'intégration des données vaccinales enrichit la qualité des informations disponibles ;
- la sélection automatique des variables réduit la dimension du problème sans écarter les facteurs les plus explicatifs ;
- l'optimisation des hyperparamètres tire vers le haut les performances de chaque modèle individuel ;
- La validation croisée imbriquée garantit une évaluation solide et limite le surapprentissage..

Ces résultats confirment l'intérêt des méthodes d'ensemble pour la prédiction des maladies infectieuses.

Le modèle proposé constitue, à ce titre, un appui décisionnel pour les autorités de santé cherchant à repérer les zones et les périodes exposées au risque d'épidémie de rougeole.de rougeole.

XI. Contribution Scientifique

L'originalité de cette étude réside dans la combinaison de plusieurs éléments méthodologiques au sein d'un même cadre expérimental :

- la fusion de données sanitaires et de couverture vaccinale ;
- une sélection automatique des variables basée sur leur importance ;
- l'optimisation systématique des hyperparamètres par GridSearchCV ;
- une validation croisée imbriquée pour une évaluation robuste ;
- un modèle de **Stacking** associant **Random Forest**, **AdaBoost** et **MLPRegressor**.

Cette combinaison forme une approche reproductible et méthodologiquement rigoureuse pour la prédiction des cas de rougeole ; elle pourrait aussi servir de point de départ à des travaux ultérieurs portant sur d'autres maladies infectieuses ou menés dans d'autres contextes géographiques.

Références Bibliographiques

[1] Al-Garadi, M. A., Khan, M. S., & Al-Garadi, A. M. (2020). Machine learning for measles prediction: A systematic review. Journal of Infection and Public Health, 13(12), 1871-1879.

[2] Breiman, L. (2001). Random Forests. Machine Learning, 45, 5–32. https://doi.org/10.1023/A:1010933404324

[3] Cawley, G. C., & Talbot, N. L. C. (2010). On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. Journal of Machine Learning Research, 11, 2079–2107.

[4] Dietterich, T. G. (2000). Ensemble Methods in Machine Learning. In Multiple Classifier Systems, Lecture Notes in Computer Science, vol. 1857 (pp. 1–15). Springer. https://doi.org/10.1007/3-540-45014-9_1

[5] Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. Journal of Computer and System Sciences, 55(1), 119–139. https://doi.org/10.1006/jcss.1997.1504

[6] Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. Informatica, 31(3), 249–268.

[7] Li, G., Cao, L., & Hu, J. (2011). Forecasting Infectious Disease Spreading Using Support Vector Machines and Time Series Analysis. Journal of Medical Systems, 35(5), 1059-1065.

[8] Organisation Mondiale de la Santé. (2023). Rapport mondial sur la surveillance de la rougeole et de la rubéole. Genève: OMS.

[9] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Joly, A., Passos, D., Cournapeau, A., & Perrot, M. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

[10] Wolpert, D. H. (1992). Stacked Generalization. Neural Networks, 5(2), 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1