

Rapport de projet : AfricaEduVision

1. Introduction

Le projet AfricaEduVision a pour objectif de développer une application interactive permettant d'analyser, de comparer et de prédire l'évolution de l'alphabétisation et des facteurs socio-économiques en Afrique. Ce projet vise à fournir un outil d'aide à la décision pour les chercheurs, décideurs politiques et ONG, afin de mieux comprendre les liens entre éducation, développement et société.

2. Méthodologie

La méthodologie suivie pour ce projet peut être résumée en plusieurs étapes :

- Collecte des données : jeu de données sur 30 pays africains concernant l'éducation, la fécondité, les mariages précoces, l'urbanisation et le PIB par habitant.
- Nettoyage des données : imputation des valeurs manquantes, suppression des doublons, homogénéisation des formats (années, pourcentages).
- Analyse exploratoire (EDA) : visualisation de l'évolution des indicateurs et exploration des corrélations.
- Modélisation prédictive : mise en place de trois approches — Prophet (séries temporelles), Random Forest (machine learning) et LSTM (deep learning).
- Comparaison multi-pays : classements, distributions et visualisations animées pour comparer les trajectoires.
- Documentation et mise en forme : intégration d'une interface utilisateur claire avec Streamlit.

3. Technologies utilisées

- Python (langage principal)
- Pandas et NumPy (traitement et nettoyage des données)
- Plotly (visualisations interactives)
- Streamlit (développement de l'application web interactive)
- Prophet (prévisions de séries temporelles)
- Scikit-learn (modélisation Random Forest et métriques d'évaluation)
- TensorFlow / Keras (réseaux de neurones LSTM)
- Git & GitHub (gestion de version et collaboration — prévu)

4. Difficultés rencontrées et solutions

- Gestion des chemins d'accès aux fichiers : utilisation des chemins relatifs et absolus pour garantir la portabilité.
- Affichage du logo dans Streamlit : conversion en base64 pour un rendu universel.
- Nettoyage des données : certaines colonnes présentaient des valeurs manquantes, traitées par imputation ou médiane.
- Modélisation avec Random Forest : difficulté à générer des projections futures résolue par la création de scénarios socio-économiques simulés.
- Modélisation LSTM : nécessité de normaliser les données et de générer des séquences pour permettre un apprentissage cohérent.
- Limitation des données : certains pays ne disposaient pas de séries temporelles suffisamment longues pour entraîner un modèle robuste.

5. Résultats obtenus

L'application AfricaEduVision permet désormais :

- Une analyse exploratoire interactive par pays et par période.
 - La mise en évidence de corrélations entre alphabétisation, fécondité, mariages précoces et PIB.
 - Des prévisions à l'horizon 2030 basées sur différents modèles (Prophet, Random Forest, LSTM).
 - Une comparaison multi-pays via des classements, distributions et animations dynamiques.
- Ces résultats offrent une base solide pour analyser les liens entre éducation et développement.

6. Perspectives

- Améliorer la précision des modèles en intégrant davantage de variables exogènes (santé, politique, infrastructures scolaires).
- Ajouter un module d'explicabilité (SHAP, LIME) pour mieux comprendre les décisions des modèles ML/DL.
- Développer une API REST pour interagir avec d'autres systèmes.
- Étendre la base de données à plus de pays et plus d'années.

7. Conclusion

Le projet AfricaEduVision constitue une première étape vers la création d'un outil d'analyse et de prévision de l'alphabétisation en Afrique. Malgré les défis rencontrés, l'application offre déjà une valeur ajoutée en combinant visualisations interactives et modèles prédictifs. Elle peut servir de support aux décideurs et acteurs du développement dans la mise en place de politiques éducatives adaptées.