Predicting School Attendance Rate in Sub-National African Countries (Using Educational Infrastructure)

Saturday, 11th November, 2023

_____

TEAM NEURAL CRAFT

Our Team

Aberejo Habeeblah
Presenter 1

Caleb Balogun
Presenter 2

Project Lead:
- Tochukwu Collins

Assistant Project Lead:
- Qudus Bello

Query Analyst:
- Chukwuegbo Love

Other Active Members

Caleb Balogun	Model Development
Onyebuchi Mkpuluma	Visualization
Aberejo Habeeblah	Presenter

Problem Statement

- In subnational African countries, forecasting school attendance rates is essential for effective education planning and resource allocation.

- This project hope to address this issue by developing machine learning models that predict attendance rates based on a comprehensive analysis of educational and socio-economic factors.

- If this problem is solved, improving attendance predictions will empower education policymakers to enhance resource allocation and educational outcomes in these regions.

Aim

- The aim of this project is to develop accurate machine learning models for predicting school attendance rates in subnational African countries, utilizing educational and socio-economic factors as key predictors.

Our Approach

- **Feature Selection:** Identify the most relevant features that impact school attendance rates within the selected regions from the data provided.
- **Model Development:** Create machine learning models that leverage the chosen features to predict school attendance rates with a high degree of accuracy.
- **Model Evaluation:** Assess the performance of the developed models through rigorous testing and validation using historical attendance data.
- **Policy Recommendations:** Provide insights and recommendations to education policymakers on resource allocation and interventions to improve attendance rates and educational outcomes.

Dataset Description

- The dataset, compiled by Climate Change and African Political Stability (CCAPS), whose focus is analyzing how climate change, conflict, governance, and aid intersect to impact African and international security.
- The dataset features provides data on literacy rates, primary and secondary school attendance rates, access to improved water and sanitation, household access to electricity, and household ownership of radio and television at the subnational level, specifically the first administrative district level.
- The dataset contains values between the year 2003 to 2011 in regions within Africa.

Dataset Description

● Below is a flowcharts diagram explain how the data wrangling, processing and exploration is done.

● **Identification:** Getting our data
● **Cleansing:** Identifying data quality issues like null values and getting rid of them.
● **Transformation:** Making a new copy of our dataset into a usable format.
● **Integration:** Combining the dataset into a unified schema.
● **Loading:** Moving the cleaned unified data into algorithms for model building.

NOTE: Codes addressing all of this can be found in the respectve cells of the colab file

Modeling

- **Machine Learning Models: We utilized ML models like**
  - **Support Vector Regression**
  - **K-Nearest Neighbors Regression**
  - **XGBoost Regression.**

- **Evaluation Metrics: Model performance was evaluated using evaluation metrics like**
  - **Mean Squared Error (MSE)**
  - **Mean Absolute Error (MAE)**
  - **R-squared (R2_Score)**

Conclusion

- **Key Findings:**
  XGBoost Regression is recommended for attendance rate prediction in educational institutions.

- **Conclusions:**
  The project successfully predict attendance rate with good accuracy.

Summary

- **Training the model after thorough data preprocessing techniques, the model was evaluated with evaluation metrics which include, R2Score, MAE, MSE which gave general comment on the model and the model is can be considered ready for Real-world deployment of solution.**