## 1. Introduction

This project applies machine learning to predict and classify demographic indicators, analyze correlations, and identify regional trends. It focuses on forecasting literacy rates, child sex ratios, and population growth, as well as classifying districts based on socio-economic factors. The goal is to provide actionable insights for policy-making, resource allocation, and urban planning.

## 2. Methodology

- Predictive Modelling: regression models (Linear Regression, Random Forest Regression) will be used to predict future demographic indicators.

- Correlation Analysis: Correlation matrices and \*Principal Component Analysis (PCA)\* will identify key relationships between demographic

## 3. Model Training and Evaluation

Models was trained and evaluated using metrics such as Mean Absolute Error (MAE) and R-squared.

|  |  |  |
| --- | --- | --- |
| Metric | Multiple Regression | Decision Tree Regression |
| R^2 | 0.9987941262817102 | 0.6965521605453784 |
| MSE | 0.0296565780394756 | 7.46282500000002 |
| RMSE | 0.17221085343112263 | 2.7318171607924313 |

# 4. Conclusion

Conclusion, the dataset works well with multiple linear regression