

Project Documentation: Waste Recycling Prediction

Problem Statement

As Indian cities grow, they face a huge waste problem driven by inefficient recycling. This project tackles that head-on by using machine learning to predict a city's recycling potential. We analyzed key factors like waste generation and how trash is managed to give cities the data they need to build smarter, more effective waste strategies.

Summary of Work

- **Development:** We built a pipeline to prepare the waste data and then trained several machine learning models—Random Forest, XGBoost, and Gradient Boosting—to predict recycling rates. The goal was to figure out which factors had the biggest impact on a city's recycling success.
- **Implementation:** We integrated all our scripts using `app.py`, which now generates a clean dataset (`cleaned_data.csv`), saves our trained models (`xgboost_tuned_model.pkl`), and outputs the final predictions in a separate file (`predictions.csv`).
- **Documentation:** We created a `README.md` file that includes detailed instructions on setting up and using the project, along with a list of credits.

Challenges

- **Real-World Data Gaps:** The available data doesn't include important factors that influence recycling, like public behavior or how well waste is sorted. This was a big challenge and limited how accurate our models could be.
- **Limited Dataset:** The `train.csv` file we used had too few samples, which made it difficult to build a truly robust and reliable model.
- **Need for Relevant Data:** For better predictions in the future, we'll need to collect more specific data, especially on things like citizen engagement, public awareness, and the technology available for recycling.