

Introduction

- Monitoring toxic chemical releases is vital for public health and sustainable urban development
- Existing tools like EnviroMapper lack advanced data visualization and analysis features
- This limits stakeholders' ability to track trends, assess regulatory impact, and make informed decisions
- The project aims to create an interactive platform to improve accessibility and usability of TRI data
- Enhanced tools will empower users to track pollution trends and support environmental risk mitigation

What's New


- Dynamic Data Interactivity
- Advanced Clustering Analysis
- Geospatial Visualization

Section 2


Data

- Source: From the Toxics Release Inventory (TRI) provided by the Environmental Protection Agency (EPA)
- Size: 3 million records.
- Timeframe: Spans from 1987 to 2023
- Content: Includes chemical release quantities, facility information, and location data.


Methodology

**Data Cleaning & Preparation**


- Removed incomplete entries
- Dropped irrelevant fields
- Standardized Zip Code format

**Exploratory Data Analysis**


- Statistical summaries
- Feature importance analysis
- Correlation matrix

**Model Training**

- Algorithm: *HistGradientBoostingRegressor*
- Hyperparameter tuning
- Performance metrics evaluated (e.g., MAE, RMSE)

**Clustering**

- PCA for dimensionality reduction
- K-Means to cluster facilities
- Grouped by chemical release profiles

**User Interface**

- Interactive Visualizations
- Embedded dashboard for data exploration

Evaluation

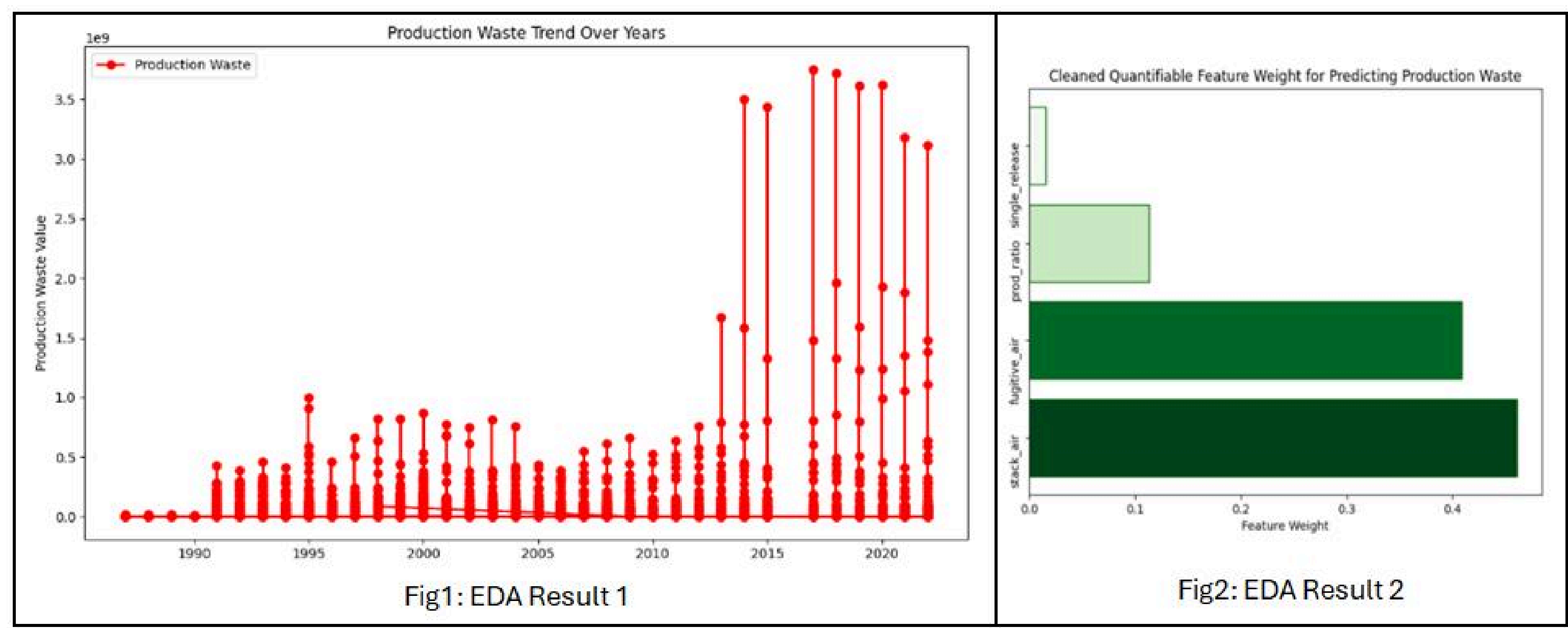
- Usability:
User surveys compare Tox-E-Mapper and EnviroMapper in terms of task efficiency, satisfaction, and ease of use.
- Accuracy:
 - Optimal cluster count identified using the elbow method (inertia plot).
 - Prediction accuracy assessed using MAE and RMSE metrics.

Results

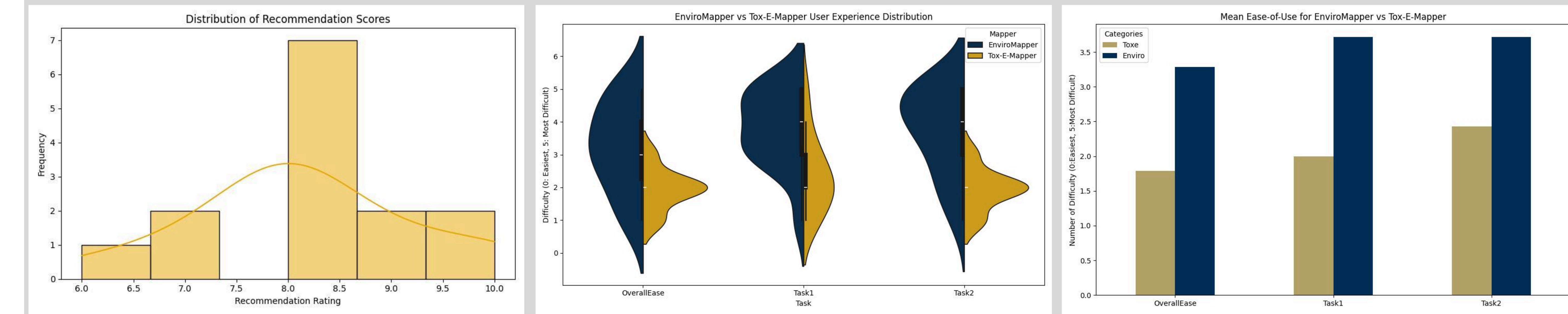
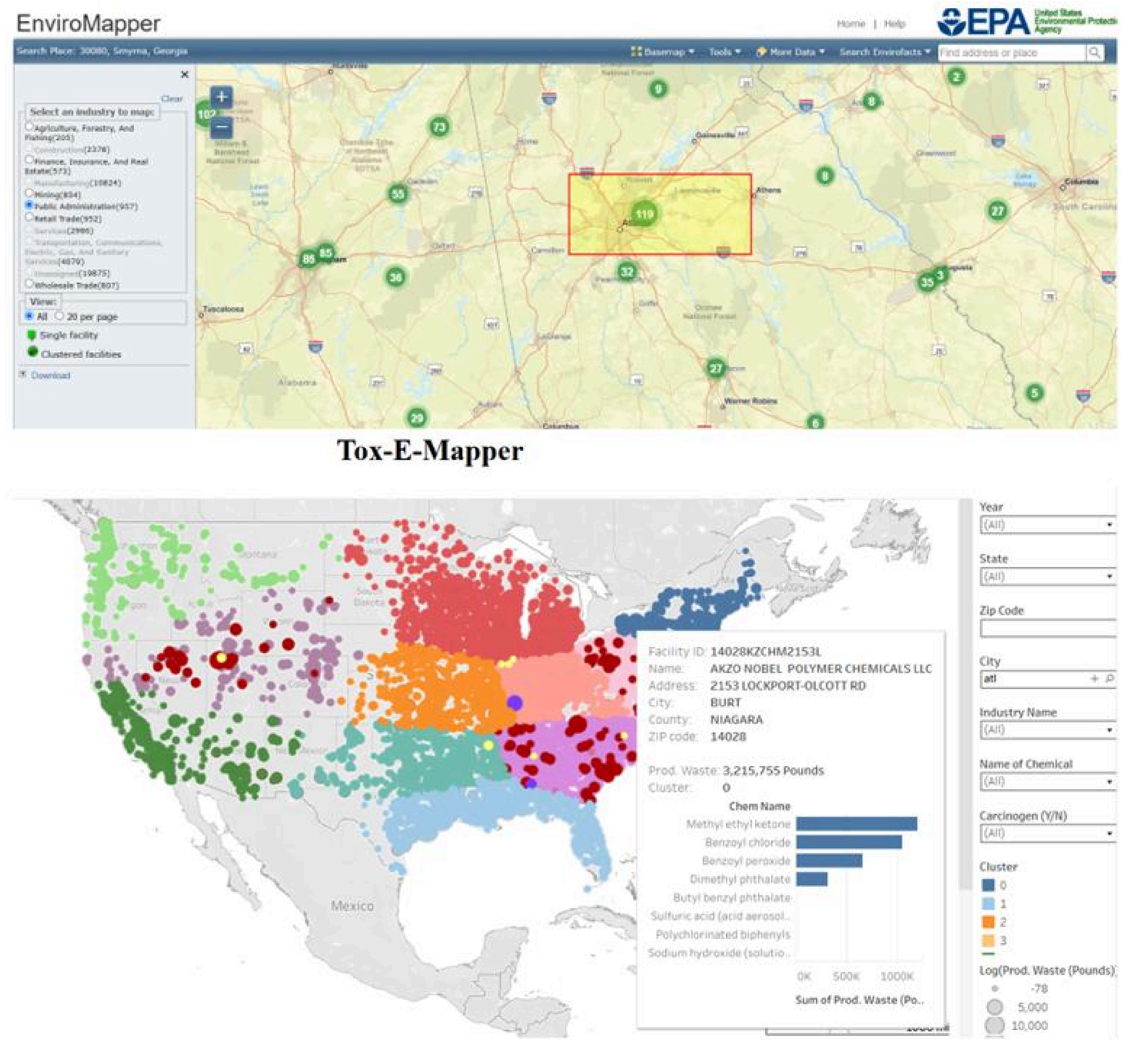
1. Sample Metrics from an experiment training model for forecasting with Histogram Gradient Boosting Regressor

Scoring Function	MAPE	MAE	MSE	RMSE
neg_mean_absolute_percentage_error	0.26	118481.36	54316723669698.88	7369988.04
neg_mean_squared_error	1.37e+20	143174.61	53606213011906.93	7321626.39
neg_root_mean_squared_error	0.26	118481.36	54316723669698.88	7369988.04
neg_mean_absolute_error	0.26	118481.36	54316723669698.88	7369988.04

2. User Survey results



Comparison : Enviromapper Vs Tox-e-mapper



Limitations

- Incomplete Dataset: Only includes government-reported industries.
- Prediction Issues: Model often predicted zeros, limiting insights.
- Clustering Time: Clustering process was computationally intensive and time-consuming.

FutureScope

- Explore alternatives to forecasting, such as classification or anomaly detection, if data sparsity continues
- Purchase a domain to publish the platform for public access.