Model Card for Litter Bin Status Model*

Bayesian Logistic Regression for Predicting Litter Bin Status

Jiwon Choi

December 2, 2024

1 Model Details

- **Model Name:** Bayesian Logistic Regression for Litter Bin Status
- **License:** MIT
- **Framework:** R (`rstanarm` package)
- **Training Algorithm:** Bayesian logistic regression with Markov Chain Monte Carlo (MCMC) sampling
- **Prior Distributions:**
  - Coefficients: Normal(0, 5)
  - Intercept: Normal(0, 5)
- **Training Iterations:** 4000 (2000 warmup)
- **Chains:** 4

2 Intended Use

- **Purpose:** Predict whether a litter bin is "Existing" (1) or "Not Existing" (0) based on its type, servicing frequency, and neighborhood housing characteristics.
- **Domain:** Urban waste management and resource allocation.
- **Users:** Urban planners, policymakers, waste management officials.
- **Limitations:** Not designed for use outside the context of Toronto or with data not represented in the training dataset.

*Code and data are available at: LINK.

3 Data

- **Datasets Used:**
  - Demographic datasets: Age, dwelling type, and household size at the ward level.
  - Litter bin dataset: Includes attributes like bin status, type, and servicing frequency.
- **Preprocessing Steps:**
  - Standardization of numeric predictors.
  - Merging datasets on a common `Ward` key.
  - Filtering rows with missing values for the target variable (`STATUS`).
  - Encoding categorical variables (e.g., `ASSET.TYPE`).
- **Dataset Size:** 10,468 observations with 8 predictors.

4 Model Architecture

- **Formula:**

```
STATUS ~ ASSET.TYPE + DAYS.SERVICED + Low_Density_Housing +
         Medium_Density_Housing + High_Density_Housing
```

- **Response Variable:** Bin status (`STATUS`), binary: 1 (Existing) or 0 (Not Existing).
- **Predictors:**
  - `ASSET.TYPE`: Type of litter bin (categorical).
  - `DAYS.SERVICED`: Frequency of servicing (numeric).
  - `Low_Density_Housing`, `Medium_Density_Housing`, `High_Density_Housing`: Proportions of housing types (numeric).

5 Performance Metrics

- **Posterior Predictive Mean (mean_ppd):** 0.9
- **Coefficient Estimates:** Shown with credible intervals in Figure 8.
  - Key predictors: `ASSET.TYPEWR3`, `ASSET.TYPEWR4`, `DAYS.SERVICED`, housing type proportions.

6 Diagnostics

- **Posterior Predictive Check:** Figure 9 shows the alignment of observed (`y`) and replicated (`y_rep`) data, confirming model fit.
- **Trace Plots:** Figure 10 confirms convergence for all parameters.
- **Rhat Statistics:** All parameters have Rhat indicating convergence (Figure 11).

7 Ethical Considerations

- **Potential Biases:**
  - Housing type proportions may reflect socioeconomic disparities.
  - The model's predictions rely on data that might not account for seasonal or temporal variations.
- **Mitigations:**
  - Use standardized preprocessing to minimize scaling biases.
  - Ensure diverse representation in training data.

8 Limitations

- The model is designed specifically for Toronto and may not generalize to other cities.
- Predictors like `ASSET.TYPE` and `DAYS.SERVICED` assume consistent definitions across the dataset.
- Temporal trends (e.g., annual changes) are not included in the model.

9 Responsible Use

- **Guidance:** Users should ensure that input data follows the same preprocessing steps. Periodic retraining is recommended for sustained accuracy.
- **Monitoring:** Model predictions should be regularly evaluated against updated waste management data to detect drift.

10 References