Datasheet for the Dataset of Reported Crimes in Toronto*

Yiyi Yao

September 24, 2024

We made a dataset to analysis crimes in Toronto to figure out a fitting model.

Motivation

1. *For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.*

   - The dataset was created to enable comprehensive analysis of crime trends in Toronto. Its purpose is to fill the gap in publicly available structured data for understanding and modeling crime patterns.

2. *Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?*

   - The Toronto Police Service, on behalf of the City of Toronto, created the dataset to make crime data available for public use.

3. *Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.*

   - The funding for the creation of the dataset is provided by the City of Toronto, supporting open data initiatives.

Composition

1. *What do the instances that comprise the dataset represent (for example, documents, photos, people, countries)? Are there multiple types of instances (for example, movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.*

---

*Code and data are available at: https://github.com/Yaoee111/Toronto-crimes.

- The dataset represents crime occurrences in Toronto, including details like the year, type of crime, and geographical division where the crime occurred.

2. *How many instances are there in total (of each type, if appropriate)?*

   - There are thousands of crime records, one for each reported crime.

3. *Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (for example, geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (for example, to cover a more diverse range of instances, because instances were withheld or unavailable).*

   - The dataset is a sample, covering reported crimes within Toronto; it may not be fully representative of unreported crime.

4. *What data does each instance consist of? "Raw" data (for example, unprocessed text or images) or features? In either case, please provide a description.*

   - Each instance consists of structured features such as crime type, year, and total crime count.

5. *Is there a label or target associated with each instance? If so, please provide a description.*

   - No specific target variable is associated with the dataset, though crime counts may be used as a target for predictive models.

6. *Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (for example, because it was unavailable). This does not include intentionally removed information, but might include, for example, redacted text.*

   - Some records may lack verified locations, or include crimes later deemed unfounded.