

**ASSESSMENT OF TORONTO CRIME DATA THROUGH EXPLORATORY DATA ANALYSIS,
CLUSTERING, AND CLASSIFICATION METHODS**

Katherine Ault
501092397
CIND 820: Big Data Analytics Project

Course Instructor: Dr. T. Abdou

May 16th, 2022

While the concept of crime forecasting can be traced back a century, it was through the adoption of geographic information systems (GIS) to map crime data during the 1990s which led researchers to recognize the potential for predictive analytics to forecast crime (Hvistendahl, 2016). The use of machine learning methods in the field of crime analysis to identify crime patterns and predict criminal activity is of considerable interest to law enforcement agencies with the results used to support evidence-based decisions in addition to informing choices regarding resource allocation, deployment, divisional staffing, and patrol plans (Lau, 2020).

Through exploration and analysis of the recently released Toronto Major Crime Indicators (MCI) dataset (Toronto Police Service, 2022), several research questions will be investigated through the application of various analytical tools and machine learning methods. Research questions include, but are not limited to:

- Can crime type(s) be predicted based on neighbourhood attributes (e.g., population density, unemployment rate, average income, average education level)?
- Which neighborhoods are the most violent and which the least violent?
- Which neighborhood has the highest incidence of crime and which neighbourhood the lowest?
- Applying the crime severity index weights (StatCan, 2021) to incidents, which neighbourhood has the highest overall crime weighting and which the lowest?
- What are the general crime trends within the City of Toronto?
- Are there recognizable temporal trends?
- Are specific crime types concentrated within certain geographical areas?

The data used for this project includes the Toronto MCI dataset noted above, in addition to Crime Severity Index weights for Canada (Statcan, 2021), geographic feature files (shapefiles) of Toronto

Police patrol zones and Toronto neighbourhoods, and Toronto neighbourhood profiles (all from open.Toronto.ca, 2022).

The techniques employed will be data cleaning and exploratory data analysis of the Toronto MCI and Toronto neighbourhood datasets, merging of datasets, identification of relevant factors associated with crime, K-means clustering for crime pattern recognition and k-NN classification for crime prediction; all methods to be conducted using R.

****feedback: 3 algorithms for classification then compare methods.**

REFERENCES

Hvistendahl, M. (2016, Sept 28). *Can 'predictive policing' prevent crime before it happens?*. Science. <https://www.science.org/content/article/can-predictive-policing-prevent-crime-it-happens>.

Lau, T. (2020, April 1). *Predictive Policing Explained*. Brennan Center for Justice. <https://www.brennancenter.org/our-work/research-reports/predictive-policing-explained>.

Social Development, Finance & Administration (2022). *Neighbourhoods* [Shapefile]. Open Data – City of Toronto. <https://open.toronto.ca/dataset/neighbourhoods/>.

Social Development, Finance & Administration. (2022). *Neighborhood Profiles* [Data set]. Open Data – City of Toronto. <https://open.toronto.ca/dataset/neighbourhood-profiles/>.

Statcan. (2021). *Crime Severity Index Weights*. [Data file]. Received via email request to Canadian Centre for Justice and Community Safety Statistics (statcan.ccjcss-ccsjsc.statcan@statcan.gc.ca).

Toronto Police Services. (2022). *Major Crime Indicators* [Data set]. Toronto Police Service Public Safety Data Portal. <https://data.torontopolice.on.ca/search?q=crime>.

Toronto Police Services. (2022). *Patrol Zones* [Shapefile]. Open Data – City of Toronto. <https://open.toronto.ca/dataset/patrol-zones/>.