# Ottawa Traffic Collisions – Describing and Predicting Injuries with Supervised Learning

Ben Nikkel, Anja Wu

Carleton University

## Introduction

- In their most recent report, the World Health Organization estimated that global deaths due to road traffic was over **1.3 million people per year**.
- Countries have **reduced this burden** by implementing road, vehicle and user safety standards strategies.
- Many traffic collision **features can influence an injury outcome** – people, vehicle, environment, infrastructure, fuel price and climate change elements are relevant.
- **Supervised machine learning** is suitable for relating diverse predictors to traffic collision injuries; previous work have used Decision Tree (**DT**) and k-Nearest Neighbour (**kNN**) methods.

## Research Questions

1. **How can Ottawa collision injuries be described** – top locations, severity distribution per mode of transportation, occurrence correlation, severity correlation, and traffic control implications?
2. Can we build a model to **predict the likelihood of injury** using Decision Tree and k-Nearest Neighbour methods?

## Methods

- **Data Source**: https://open.ottawa.ca/ – accessed in September 2023.
- **Preprocessing**
  - Missing integer values were changed to zeros.
  - Latitude/Longitude values were truncated to 4 decimal points.
  - A binary injury variable was created; 0 = no injury, 1 = at least one injury.
  - Ordinal features were converted to integer scales.
  - Nominal features were converted to one-hot encoding.
- **Descriptive Analysis**
  - Correlation heat maps used **Pearson correlation coefficients**.
- **Modeling**
  - Packages (Python): SciKit-Learn, imblearn.over_sampling
  - 80-20 Train-Test split with nested 70-30 Train-Validate split of outer Train set.
  - **kNN**
    - Hyperparameters: nearest neighbours = 6
    - Class imbalance was addressed using Random OverSampling and SMOTE.
  - **DT**
    - Hyperparameters: entropy criterion, max depth = 6
    - Class imbalance was addressed using Random OverSampling.
    - Random Forest (n_estimators = 100)
- **Model Evaluation**
  - DT was evaluated against a variety of techniques using PyCaret [Python].

## Acknowledgements

Thank you to **City of Ottawa** for making the collision data publicly available and for **Mahmud Hasan** for guidance during this project.

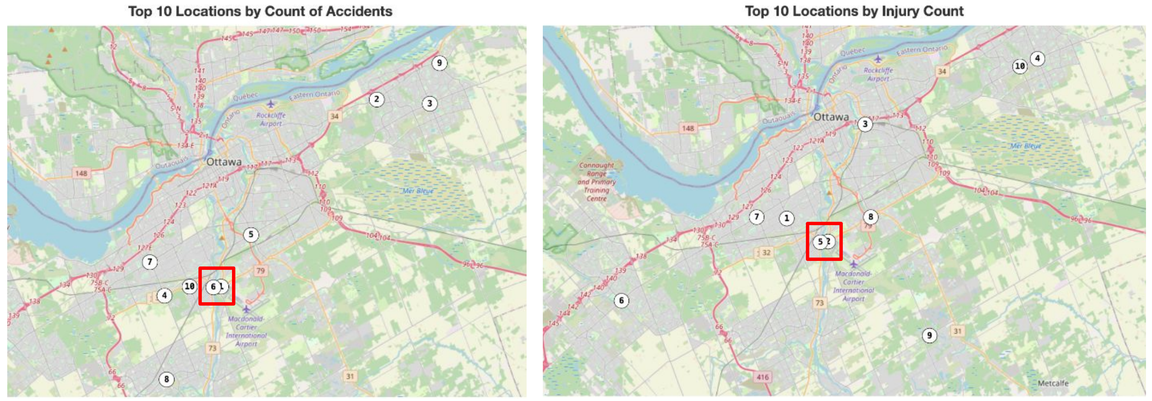


**Figure 1**. Top 10 Locations in Ottawa, Ontario with the most collisions and the most collisions involving at least one injury. **Top**: geographic maps marked by the locations with the most collisions (**left**) and most collisions with injuries (**right**). **Bottom**: Frequency tables of locations with the most collisions (**left**) and the most collisions with injuries (**right**).

| Collisions | | | Collisions with an Injury | | |
|---|---|---|---|---|---|
| **Rank** | **Location** | **n** | **Rank** | **Location** | **n** |
| **1** | Hunt Club Rd @ Riverside Dr | 232 | **1** | Meadowlands Dr @ Merivale Rd | 33 |
| **2** | St. Joseph Blvd @ Jeanne D'arc Blvd | 218 | **2** | Hunt Club Rd @ Riverside Dr | 32 |
| **3** | Innes Rd @ Tenth Line Rd | 168 | **3** | Riverside Dr @ Tremblay Rd/Hwy417 Ic117 Ramp52 | 32 |
| **4** | West Hunt Club Rd @ Woodroffe Ave | 157 | **4** | Innes Rd @ Tenth Line Rd | 32 |
| **5** | Bank St @ Walkley Rd | 157 | **5** | Prince Of Wales Dr @ West Hunt Club Rd | 29 |
| **6** | Prince Of Wales Dr @ West Hunt Club Rd | 155 | **6** | Hazeldean Rd @ Terry Fox Dr | 28 |
| **7** | Baseline Rd @ Woodroffe Ave | 155 | **7** | Baseline Rd @ Woodroffe Ave | 28 |
| **8** | Greenbank Rd @ Strandherd Dr | 147 | **8** | Hunt Club Rd @ Bridle Path Dr/Daze St | 28 |
| **9** | St. Joseph Blvd/Old Montreal Rd @ Trim Rd | 147 | **9** | Albion Rd @ Mitch Owens Rd | 27 |
| **10** | Merivale Rd @ West Hunt Club Rd | 146 | **10** | Innes Rd @ Jeanne D'arc Blvd/Mer Bleue Rd | 27 |

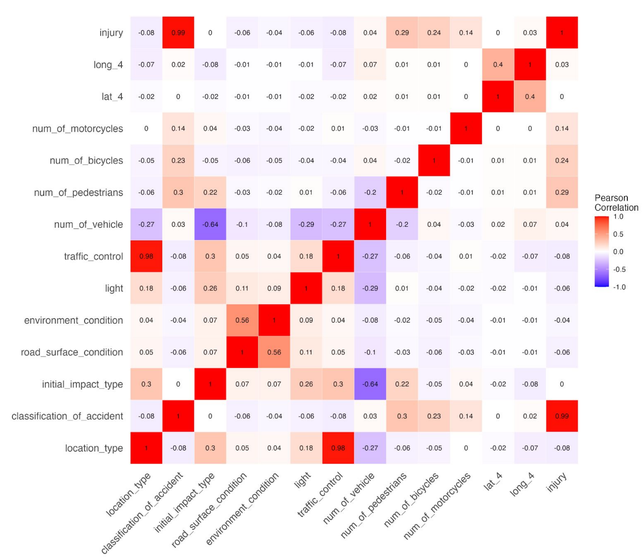


**Figure 2.** Correlation matrix of all predictor variables and injury occurrence.

| Rank | Location |
|---|---|
| Precision | 0.88 |
| Recall | 0.24 |
| F1 Score | 0.37 |
| Accuracy | 0.85 |

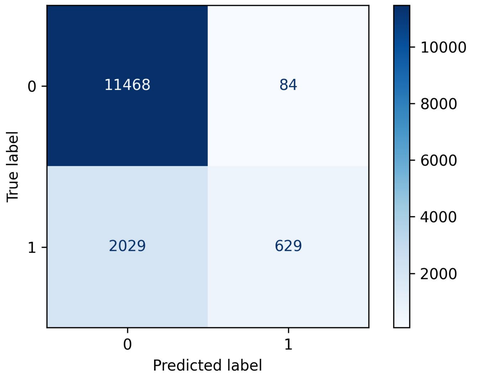


**Figure 3**. Decision Tree model test measures (**left**) and confusion matrix (**right**).

## Results

1.The "**Hunt Club Bridge**" is one of the most collision/injury prone locations (Figure 1).

2.Six locations had 3 **major** injuries. Three locations had 2 **fatal** injuries. Both outcomes involved **8th Line Rd @ Parkway Rd**.

3.**Mode of transportation** features (**motorcycles, pedestrians, bicycles**) correlated with an injury occurring (Figure 2).

4.**Environment** features did not correlate with injury severity.

5.Most collisions involved only **Vehicles** (95.28%) and of these most were without injury (80%). **Pedestrian** collisions almost always involved injury (> 99%). 3.33% of **Motorcycle** injuries were ***fatal*** while 2.43% of **Pedestrian** injuries were ***fatal***.

6.Combined, collisions usually occurred when there was a **traffic control** (signals, stop signs, etc.). Separated, the most common type of traffic control was no control.

7.Compared to the best benchmark model (Ridge Classifier), **Decision Tree** model (Figure 3) had the highest **accuracy** (0.85 > 0.84) and **precision** (0.88 > 0.69) but not **recall** (0.24 < 0.26).

8.**kNN** performed worse than chance. **Random Forest** performed worse than the final decision tree model.

## Implications

- **First responders** can better anticipate the degree of injury with this model.
- High collision/injury locations may deserve targeted **interventions**.
- **Environmental** features may not correlate to traffic collision injuries when analyzed **categorically**.

## Future Work

- **Time** analysis – does time of day, day of week, seasonality, COVID affect injury incidence?
- Incorporate **traffic volume** data for more comprehensive model construction.
- Intervention analysis – did traffic **camera installation** result in fewer collisions and injuries?

## Conclusion

- Multiple Ottawa locations have high collision frequency, with the "**Hunt Club Bridge**" being most prevalent.
- **8th Line Rd @ Parkway Rd.** is a **dangerous** traffic location.
- **Mode of transport** affects the occurrence and type of injury.
- **Decision Tree** is a useful method for predicting traffic collision injuries with this dataset.

## References

Abdi A, Seyedabrishami S, O'Hern S. (2023). A Two-Stage Sequential Framework for Traffic Accident Post-Impact Prediction Utilizing Real-Time Traffic, Weather, and Accident Data. J Adv Transp 2023.

Nogayeva S, Gooch J, Frascione N. (2021). Science & Justice The forensic investigation of vehicle – pedestrian collisions : A review. Sci Justice 61:112–118.

World Health Organization. (2018). Global status report on road safety 2018: summary. WHO/NMH/NVI/18.20.

Yuan Y, Wang S, Liu Z, Cui G, Wang Y. (2022). Influencing factors analysis of side right-angle collisions severity at intersections based on decision tree. Int J Crashworthiness 27:59–69.