1.0 Background

Traffic collisions remain a significant public safety concern in Toronto. According to a report produced by the Toronto Police Service, from the year 2022 to 2023, 600 people were killed or seriously injured as a result of motor vehicle collisions [1]. Studies highlight that timely emergency medical services (EMS) response plays a critical role in mitigating the severity of collision outcomes and reducing fatalities [2]. Given this, identifying the contributing factors to MVCs and exploring effective mitigation strategies is crucial for improving road safety. This project will analyze traffic and EMS data to identify factors increasing traffic collision risk and determine how optimizing EMS station locations can reduce response time and minimize accident severity. By gaining these insights, our team seeks to develop evidence-based solutions to enhance road safety in Toronto.

2.0 Problem Statement

This project addresses two key objectives: understanding the risk of collisions through predictive modeling of contributing factors and minimizing collision impact by optimizing EMS response times. Considering this scope, the aim of this project is to explore the following questions: (1) Based on the features available and the data present, how can we predict the risk of collisions? (2) Which features have the highest importance and relevance when determining the risk of a collision? (3) How can we optimize current infrastructure (i.e. EMS placement) to minimize the distance to high-risk collision areas, thereby decreasing response times?

3.0 Datasets and Sources

This project will leverage multiple datasets, combining data from the City of Toronto, Environment Canada and various reputable traffic sources. GIS data from Wiki Open Street Map and historical collision data from Toronto Open Data Catalog (TODC) for traffic collisions will provide details on each incident such as accident location (latitude, longitude, ward), time of day, and road factors (speed limits, traffic signals, lighting, and road type) [3]. This allows for the identification of patterns in relation to time, geography and road conditions and assess the effects of congestion and its relationship to accident frequency. TODC will be used for data on existing EMS station placement [4], and available properties for new infrastructure [5]. Weather data from Environment Canada (i.e. temperature, precipitation, and visibility), will be integrated to assess the impact of weather conditions on accident occurrences [6].

4.0 Methodology

A predictive model will estimate traffic accident risk, and an optimization model will determine the optimal placement of ambulance stations based on accident hotspots. Logistic regression will be the baseline for predicting accident probability. Classification models such as Random Forest Classifiers will be applied to capture nonlinear relationships between features and the target. The K-means clustering algorithm [7] and/or DBSCAN [8] will be used to group accident-prone areas for hotspot analysis [9], and the cluster centroids and high-density areas will represent the centers of high-risk zones. A linear programming problem will be formulated to optimize the placement of ambulance stations, by minimizing the distance to these centroids. The number of new stations to place will be determined by the City of Toronto's budget constraints, and possible locations will be determined based on existing station locations and available property. Optimization tools such as CVXPY or Gurobi will be used to solve this problem. The results will: (1) guide the optimal placement of new ambulance stations, and (2) allow prediction of future accident risk based on input factors.

5.0 References

[1] “Total KSI,” Toronto Police Service Public Safety Data Portal, https://data.torontopolice.on.ca/pages/total-ksi (accessed Oct. 3, 2024).

[2] J. P. Byrne et al., “Association between emergency medical service response time and motor vehicle crash mortality in the United States,” JAMA Surgery, vol. 154, no. 4, p. 286, Apr. 2019. doi:10.1001/jamasurg.2018.5097

[3] Transport Canada, “Canadian Motor Vehicle Traffic Collision Statistics: 2022,” Transport Canada, May 02, 2024. https://tc.canada.ca/en/road-transportation/statistics-data/canadian-motor-vehicle-traffic-collision-statistics-2022

[4] City of Toronto, Nov. 29, 2022, “Ambulance Station Locations”, Open Data Dataset. [Online]. Available: https://open.toronto.ca/dataset/ambulance-station-locations/

[5] “Open Data Dataset.” https://open.toronto.ca/dataset/toronto-signature-sites/

[6] E. and C. C. Canada, “Daily Data Report for October 2022 - Climate - Environment and Climate Change Canada,” climate.weather.gc.ca, Oct. 31, 2011. https://climate.weather.gc.ca/climate_data/daily_data_e.html?StationID=51459

[7] Scikit. https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html. (Accessed Oct. 3, 2024).

[8] Scikit. https://scikit-learn.org/dev/auto_examples/cluster/plot_dbscan.html. (accessed Oct. 3, 2024).

[9] H. Elle, “Unveiling patterns in geospatial data: A clustering approach,” Medium, https://medium.com/@huntresselle/unveiling-patterns-in-geospatial-data-a-clustering-approach-9a5fde1b5bf5 (accessed Oct. 3, 2024).