# Running Head

Anthony Escalona
Seidenberg School of CSIS
New York, Ny 10038
Email: ae50483p@pace.edu

***Abstract*—Roadway traffic safety is a significant concern for transportation governing agencies as well as ordinary citizens. To provide advice for safe driving, careful analysis of road traffic information is important to identify variables closely related to fatal accidents. In this paper, I apply statistical analysis and data mining algorithms on the NYC Open Data portal dataset as an attempt to address these problems. The relationship between fatal rate and other attributes, including collision manner, weather, surface condition, light condition, and noise complaints, were investigated.**

## I. Introduction

It is estimated that around 4,000 New Yorkers are seriously injured in New York and more than 250 people are killed in traffic accidents every year. The automobile is the leading cause of injury-related death for children under 14 years of age and the second leading cause for seniors. On average, every two hours, vehicles severely injure or kill a New Yorker. The cost of these deaths and injuries impacts the city's social and economic growth greatly. New York City should no longer consider traffic crashes as mere "accidents," but as preventable incidents that can be addressed systematically. No degree of fatality is unavoidable or appropriate on the streets of the city. New York City's Vision Zero Action Plan[1] is the foundation to reduce traffic deaths and injuries. City of New York will use every available tool to enhance the safety of our streets. With this action plan, it is making a bold new commitment to improving street safety in every neighborhood and district – with increased enforcement of dangerous moving violations such as speed and failure to yield to pedestrians, new street designs and configurations to improve safety, widespread public access and communications, and a comprehensive legislative agenda to increase penalties.

Data mining is a major step in knowledge discovery. It is the process of extraction of non-trivial, valid and potentially useful information from huge databases. Some of the important data mining techniques are classification, association rule mining, segmentation, and clustering.

Predicting where and when road incidents will occur is complicated. It is possible to analyze traffic injury statistics and identify a correlation between variables based on historical traffic event data. On the other hand, visualization of data from traffic accidents provides detailed insights into how it changes over time. This paper focuses on practical issues related to the project to prevent road accidents. Analysis and visualization of data help observe the occurrence of traffic accidents and take appropriate action to enhance safety.

## II. Literature Review

A number of studies have been conducted to determine the factors leading to serious road accidents and to reduce the number and severity of injuries by removing or regulating these factors. As the traffic accident is large and heterogeneous, most scientists adopted data mining methods to carry out their studies.

Chong, Abraham, and Paprzycki [2] have applied artificial neural networks and decision trees to a specific data collection from the National Automotive Sampling Program and General Estimates Systems including traffic incident information from 1995 to 2000. The collection of data was limited only to head-on collisions. The findings revealed that neural networks were outperformed by the decision tree approach. The findings found that seat belt use, highway lighting condition, and driver alcohol use were the most important variables in fatal injuries.

Authors in [3] have applied their work using multivariate logistic regression to determine the independent contribution of driver, crash, and vehicle characteristics to driver' fatality. The result showed that increased use of seatbelts, reduced vehicle speed, and reduced number of and severity of drivers' side-impact might prevent deaths.

The purpose of paper [4] was to employ logistic regression models to develop crash-related injury prediction models. They analyzed traffic crash data in Kentucky during 2001 using logistic regressions. They concluded that the occupant's risk factors for the high level of injury severity were age, gender, and non-use of restraints. In [5], the authors used the same model to quantify the association of driver's age with traffic injury severity. Wisconsin crash data from 2000 to 2004 was used to study 602,694 drivers of a car or truck who were involved in a motor vehicle crash. It was discovered that the oldest drivers, especially those older than 85 drivers 85 years and older, had the highest risks for serious injury or fatality.

### A. Ojbectives

## III. Motivation

to do

## IV. Design

to do

## V. Implementation

to doe.

## VI. Evaluation

to do

## VII. Conclusion

The conclusion goes here.


## Acknowledgment

The authors would like to thank...


## References


[1] NYC, "Vision zero," Online, 2019, https://www1.nyc.gov/site/visionzero/index.page, last access November 2019.

[2] M. M. Chong, A. Abraham, and M. Paprzycki, "Traffic accident analysis using decision trees and neural networks," 2004.

[3] M. Bedard, G. H. Guyatt, M. J. Stones, and J. P. Hirdes, "The independent contribution of driver, crash, and vehicle characteristics to driver fatalities," *Accident Analysis Prevention*, vol. 34, no. 6, p. 717, 2002.

[4] M. Singleton, H. Qin, and J. Luan, "Factors associated with higher levels of injury severity in occupants of motor vehicles that were severely damaged in traffic crashes in kentucky, 2000-2001," *Traffic Injury Prevention*, vol. 5, no. 2, pp. 144–150, 2004.

[5] R. B. Hanrahan, P. M. Layde, S. Zhu, C. E. Guse, and S. W. Hargarten, "The association of driver age with traffic injury severity in wisconsin." *Traffic Injury Prevention*, vol. 10, no. 4, pp. 361 – 367, 2009.