Motor Vehicle Crashes

You will need to read the attached article before starting this week's graded discussion.

Iranitalab, & Khattak. (2017). Comparison of four statistical and machine learning methods for crash severity prediction. Accident Analysis and Prevention, 108, 27-36.

[M2\_CSE575\_Comparison-of-four-statistical-and-machine-learning-methods-for-crash-severity-prediction.pdf](https://asu.instructure.com/courses/45793/files/11859682/download?wrap=1)

[Actions](https://asu.instructure.com/courses/45793/discussion_topics/951428)

Meet the requirements of this graded assignment by completing two steps:

* Write and post a 100- to 150-word response to the following question: **Based on the methods used, the results achieved, and the conclusions drawn in this use case, what question or questions does this journal article answer or raise for you? Why?**
* Read through other students' responses and reply to at least two posts to explore the factors your peers considered and the reasoning they employed. Reply posts must be at least 50 words.

The primary purpose of this article was to determine crash severity prediction utilizing not only machine learning methods, such as nearest-neighbor classification, support vector machines and random forests but also contemporary statistical methods. The goal of the study was to optimize resource expenditure from three classes of stakeholders- transportation safety planners, hospitals and medical agencies, and insurance companies to be better prepared for when accidents do happen.

The results of the paper demonstrate that the final comparison results are contingent on the type of stakeholder that is going to use them. For instance, transportation safety planners are more inclined to use a combination of Nearest-Neighbor-Classification (NC) and K-Means-Clustering (KC), while an insurance company focuses on the statistical method of Overall Prediction Error (OPE) and Specific Prediction Error (SPE). This led me to question whether a general-purpose methodology framework could be constructed based on more inherently quantitative or qualitative factors.

---------------------------------

Hi Arpit. The paper didn’t just exclusively use machine-learning methodologies but also relied on more contemporary statistical calculations, such as Overall Prediction Error (OPE) and Specific Prediction Error (SPE) in order to mitigate losses

Typically within statistics you want a good sample size that can accurately reflect the principles of a larger population, which is an inherent constraint in statistics. You are right about the need for a more holistic and diverse data set that would mirror a larger set of conditions under the state space, but perhaps Nebraska was simply a starting point for their study

------------------------------

Hi Gnanavi. The authors did make extensive use of clustering-based methods within the study, but it was important to note that near the end of the report (specifically within the conclusion) that the model had a trend to underestimate the crash costs. In fact, some prediction models misclassified a significant portion of the more severe crashes as less severe. Hence, it was considered less costly, which led to broad underestimation of the overall costs of crashes, by the best method (which was inherently clustering-based): Nearest Neighbors Classification (NNC)

NNC.

* \_
* \_
* \_