

# ANALYSIS OF MORTALITY DATA FOR VARIOUS MANNERS OF DEATH USING MACHINE LEARNING ALGORITHMS

*Ben Kim, Seattle University, 901 12<sup>th</sup> Avenue, Seattle, WA 98072, 206-296-2806, [bkim@seattleu.edu](mailto:bkim@seattleu.edu)  
Trang Ho, Seattle University, 901 12<sup>th</sup> Avenue, Seattle, WA 98072, 206-296-2806, [hot3@seattleu.edu](mailto:hot3@seattleu.edu)*

## ABSTRACT

We analyzed the 2015 mortality data from the CDC (Center for Disease Control) to estimate and predict three types of death – accidents, suicides, and homicides based on people's demographic data. This dataset contains 116 attributes and more than 2.7 million records. Also, we identified the critical attributes for each type of death. For this analysis, we used three machine learning (ML) algorithms – Bayesian Classification, Decision Trees, and Artificial Neural Networks. To determine which ML algorithms work better for each type of death, we built confusion matrices and calculated the ratios, including accuracy, error, precision, sensitivity, specificity, and F-score. Since these types of death are relatively rare compared with the total number of deaths, we used various oversampling schemes to address the class imbalance problem. In conclusion, we provide the interpretations of the results for each ML model and implications for preventive measures and social policies.

**Keywords:** Mortality, Decision Trees, Bayesian Classification, Artificial Neural Networks