Undetected Suicide: Classification of Undetermined Drug-Related Deaths Using Machine Learning Techniques

Liu, Daphne (School: West High School)
Yu, Mia (School: West High School)

States and the CDC monitor accidental and intentional violent deaths to answer questions that are critical for the development of effective violence prevention strategies. The National Violent Death Reporting System (NVDRS) is a major innovation in violent death surveillance linking data for individual cases from multiple sources. However, suicide underreporting is common and primarily from drug-related deaths. Therefore, the purpose of this study was to use innovative machine learning techniques to minimize suicide underreporting. Drug-related violent death data were extracted from the 2012-2015 NVDRS in Utah. The cases that were already classified as suicides or unintentional in the data were used to train and test our machine learning models. Three classification algorithms, including Random Forest Classifier (RFC), Support Vector Machines (SVM), and Artificial Neural Networks (ANN), were used to develop the classification models. The results show that the overall classification accuracy rates of all three models were 94%. Among the three models, the SVM model achieved the best F score and was thus selected to classify the undetermined drug-related deaths into suicide or unintentional deaths. According to the findings from the SVM model, drug-related suicide rates were underreported by 34% averaged across the 2012-2015 four years, ranging from 31% in 2013 to 38% in 2015. This research identified a cost-effective method to substantially reduce suicide underreporting, which is replicable in other NVDRS states. Increased accuracy in suicide reporting can directly enhance suicide surveillance quality, provide useful data for suicide etiology research, and support effective public health interventions to reduce suicide.

Awards Won:
National Institute on Drug Abuse, National Institutes of Health &amp the Friends of NIDA: First Award of $2,500
American Statistical Association: Certificate of Honorable Mention