

UCLA ASA DataFest 2021 Write-Up

Team K-Bandits: *Eustina Kim (kimjee8955@gmail.com), Tiffany Feng (tiffanyfeng456@gmail.com), Diana Pham (dianapham005@gmail.com), and Kienna Qin (kqin@g.ucla.edu)*

Our team explored which characteristics of each occupation (students, health professionals, and veterans) determined the use of benzodiazepines (BZDs), a type of drug that produces sedation and lowers anxiety levels, in the sample of participants in the United States during 2018. BZDs are the most commonly prescribed medication in the U.S., and, unfortunately, are frequently abused and taken with other drugs. For our project, we created three multinomial logistic regression models to predict use/misuse of BZDs for each occupation.

For data preparation, we conducted feature engineering by first creating a new response variable that categorized BZD use into no use, medical use, and non-medical use. We also created two new predictor variables that classified whether or not the participant has used drugs other than BZD both medically and non-medically in their lifetime, respectively. Then, we separated the data into three datasets according to occupation. For variable selection, we ran random forest models on our full datasets and used the variable importance outputs to select variables with mean decreases in accuracy above the third quartile for each occupation.

During our model selection process, we used both a validation set approach and 10-fold cross-validation to evaluate models using random forest, bagging, multinomial logistic regression, and support vector machines. After comparing our models, we selected multinomial logistic regression as the final model since it performed well and produced results that were interpretable. Our final models had cross-validation accuracies from 68-80% and validation set accuracies from 70-80% for the three occupations, so we were able to predict BZD use with high accuracy for all three occupations.

To better understand what specific categories contributed most to BZD use in each occupation in our sample, we categorized the predictors in our dataset into those related to mental health, opioid prescriptions for pain, drug use / abuse, and demographics. By looking at each category of predictors' mean percent decrease in accuracy for each occupation, we found that students' BZD use was influenced strongly by the drug and pain categories, veterans were influenced most by the mental health and pain categories, and health professionals had similar influence across all categories. We gained further insight on how these categories influenced BZD use in each occupation in our sample by creating barcharts and boxplots on a selection of the most important predictors in each category. First, the history of drug use had a large significance on BZD use and misuse, especially students, and those who have used other drugs besides BZD non-medically had higher instances of using BZD non-medically. In addition, participants with mental health histories had higher instances of BZD use. Lastly, veterans had the strongest relationship between age and BZD use, with lower ages corresponding to higher instances of non-medical BZD use.

Overall, our findings indicate that, in our sample, having a history of drug use, mental illness, and prescriptions for opioids for treating pain are important in predicting the use of BZD for all three occupations, but the largest impacts for each occupation were a history of drug use for students, mental illness and opioid prescriptions for pain for veterans, and there were similar influences for each category for health professionals. Although our data was not a representative sample, this could still demonstrate the importance of investigating the unique differences in determining BZD use for each occupation, which could lead into interventions such as targeted programs to reduce BZD misuse in each occupation. Hopefully with these insights, we can make a positive impact on the lives of those affected by the misuse of BZDs in the United States.