Drug and Alcohol Dependence 208 (2020) 107839

Contents lists available at ScienceDirect

Drug and Alcohol Dependence

journal homepage: www.elsevier.com/locate/drugalcdep

Full length article

Understanding Opioid Use Disorder (OUD) using tree-based classifiers

Adway S. Wadekar

Saint John's High School, 378 Main Street, Shrewsbury, MA 01545, United States

ARTICLE INFO

Keywords:
Opioid Use Disorder
Machine learning
Random forest
Marijuana

ABSTRACT

Background: Opioid Use Disorder (OUD), defined as a physical or psychological reliance on opioids, is a public health epidemic. Identifying adults likely to develop OUD can help public health officials in planning effective intervention strategies. The aim of this paper is to develop a machine learning approach to predict adults at risk for OUD and to identify interactions between various characteristics that increase this risk.

Methods: In this approach, a data set was curated using the responses from the 2016 edition of the National Survey on Drug Use and Health (NSDUH). Using this data set, tree-based classifiers (decision tree and random forest) were trained, while employing downsampling to handle class imbalance. Predictions from the tree-based classifiers were also compared to the results from a logistic regression model. The results from the three classifiers were then interpreted synergistically to highlight individual characteristics and their interplay that pose a risk for OUD.

Results: Random forest predicted adults at risk for OUD with remarkable accuracy, with the average area under the Receiver-Operating-Characteristics curve (AUC) over 0.89, even though the prevalence of OUD was only about 1 %. It showed a slight improvement over logistic regression. Logistic regression identified statistically significant characteristics, while random forest ranked the predictors in order of their contribution to OUD prediction. Early initiation of marijuana (before 18 years) emerged as the dominant predictor. Decision trees revealed that early marijuana initiation especially increased the risk if individuals: (i) were between 18–34 years of age, or (ii) had incomes less than $49,000, or (iii) were of Hispanic and White heritage, or (iv) were on probation, or (v) lived in neighborhoods with easy access to drugs.

Conclusions: Machine learning can accurately predict adults at risk for OUD, and identify interactions among the factors that pronounce this risk. Curbing early initiation of marijuana may be an effective prevention strategy against opioid addiction, especially in high risk groups.

1. Introduction

Opioid Use Disorder (OUD) is a physical or psychological reliance on opioids, a substance found in certain prescription pain medications and illegal drugs like heroin (Centers for Disease Control, n.d.). The United States is in the midst of an opioid overdose epidemic, which is now a public health crisis (Finn, 2018). OUD covers a wide range of illicit and prescribed drugs of the opioid class. These range from those that are illegal to ones commonly sold in black markets, such as heroin, to opioids used for substitution of street drugs such as methadone, to analgesics used mainly in hospital settings such as morphine to common painkillers available on prescription such as codeine and oxycontin (Daniulaityte et al., 2014). Thus, OUD covers a wide range of drugs accessed through many different sources and by people from different walks of life. Perhaps the most well-known and notorious type of OUD is Heroin Use Disorder, however, only about 20 % people with OUD list heroin as their "drug of choice". Most people with OUD use analgesic opioids, or common painkillers that are prescribed by doctors, whether it be for themselves or for someone else. A significant population of people with OUD obtain prescription drugs from some other medium as well (Finn, 2018; Tinker, 2019).

Several characteristics increase the likelihood of addiction to opioids and other substances. These include mental illnesses, disabilities, issues with personal and social relationships, proximity to other drug users, and past traumatic events (National Institute on Drug Abuse, 2017). Contemporary studies have evaluated the association of some of these characteristics with substance use disorders. Most of these studies, however, attempt to explain substance use disorders by considering each characteristic in isolation. For example, according to the Department of Health and Human Services, individuals with disabilities have a substance abuse rate two to four times of that of the non-disabled population (National Rehabilitation Information Center, 2011). Similarly, the National Bureau of Economic Research reports an association between mental illness and substance abuse (Center for

E-mail address: adway@adway.io.

https://doi.org/10.1016/j.drugalcdep.2020.107839
Received 11 November 2019; Received in revised form 14 December 2019; Accepted 22 December 2019
Available online 15 January 2020

Table 1
Summary of Predictor Variables.

Variable	Levels
Outcome	
Opioid Dependence or Abuse in the past year: Yes/No	Two
Predictor Variables (Demographic)	
1 ***Gender:*** Male/Female	Two
2 ***Age:*** 18-25 yrs, 26-34 yrs, 35-49 yrs, 50-64 yrs, 65 yrs or older	Four
3 ***Race:*** Non Hispanic White, Non Hispanic African American, Non Hispanic Native American/Alaskan Native, Non Hispanic Native Hawaiian, Non Hispanic Asian, Non Hispanic More than One Race, Hispanic	Seven
Predictor Variables (Socioeconomic)	
4 ***Income:*** < $20k, $20k-$49k, $50k-$75k, > $75k	Four
5 ***Employment:*** Employed full time, Employed part time, Unemployed, Other (including not in the labor force)	Four
6 ***Education:*** Less than high school, High school grad, Some college/Assoc degree, College Graduate	Four
7 ***Approached by someone selling drugs:*** Yes/No	Two
8 ***Heorin fairly or very easy to obtain:*** Yes/No	Two
Predictor Variables (Physical & Psychological)	
9 ***Any Disability:*** Computed based on the difficulties in one or more of: vision, hearing, walking, thinking, running errands, and dressing, Yes/No	Two
10 ***Any mental illness in the past year:*** Estimated based on psychological distress, based on symptoms such as feeling nervous, feeling hopeless, feeling restless or fidgety, feeling so sad or depressed that nothing could cheer you up, feeling that everything was an effort, and feeling down on yourself, no good, or worthless, and functional, social, relational, and work impairment, suicidal thoughts, attempts, and plans, and impairment at home, work, etc. includes depression and suicidal thoughts. Yes/No	Two
11 ***First use of alcohol before 18 years:*** Yes/No, No includes both those who use alcohol after 18 years and non-users.	Two
12 ***First use of marijuana before 18 years:*** Yes/No, No includes both those who used marijuana after 18 years and non users.	Two
13 ***Overall Health:*** Self-reported, Excellent, Very Good, Good, Fair/Poor.	Four
14 ***Parole/supervised release status in the past year:*** Yes/No	Two
15 ***Probation status in the past year:*** Yes/No	Two
16 ***Perception of great risk trying heroin once or twice:*** Yes/No	Two
17 ***Perception of great risk using heroin weekly:*** Yes/No	Two
18 ***Obesity:*** BMI, Normal (< 25.0), Overweight (25.0-29.0), Obese (> 29.0)	Three

Behavioral Health Statistics and Quality, 2016). Fiellin et al. (2013) examine the association between substance abuse in youth and substance use disorders in adult life.

Although each characteristic may individually increase the odds of substance abuse by three to four times, neither one is a determinant. Therefore, it is necessary to move beyond explanation to prediction, where intricate associations between these characteristics can be considered in an integrated manner in order to determine who is likely to develop OUD. Identifying adults at risk for OUD along with potential confounders that enhance this risk can help public health officials in planning effective intervention and prevention strategies. This paper presents a machine learning approach to predict individuals that are at risk for OUD and to understand how interactions between various demographic, socioeconomic, physical, and psychological predictors increase this risk. The approach uses a combination of logistic regression, decision trees, and random forest classifiers synergistically.

2. Materials and methods

2.1. Curating the data

The paper takes an epidemiological perspective, where OUD is an outcome, and certain characteristics increase its likelihood. To develop a predictive relationship between OUD and these characteristics, a data set is needed where individuals are classified as having OUD or not along with their other characteristics that are associated with this outcome. The 2016 edition of the National Survey on Drug Use and Health (NSDUH), conducted by Substance Abuse and Mental Health Services Administration (SAMSHA) included OUD outcome for the first time (Substance Abuse and Mental Health Services Administration, 2016). This was the 36th edition of the NSDUH survey, and it covers the civilian, noninstitutionalized population of the United States (including civilians living on military bases) older than 12 years. This data is publicly available and details about the sampling methodology can be found in the NSDUH codebook (Center for Behavioral Health Statistics and Quality, 2016).

The survey includes over 56,000 observations, with each observation comprising more than 2600 variables (Substance Abuse and Mental Health Services Administration, 2016). The raw or edited variables were directly reported from the interviews, and these raw variables served as the source for recoded or imputed variables. Recoded and imputed variables do not have any missing values, and SAMHSA recommends using these instead of the raw source variables whenever possible (Center for Behavioral Health Statistics and Quality, 2016).

A data set was curated from these survey responses. Respondents are classified according to the outcome (OUD/No-OUD), if they had a Heroin Use Disorder, Prescription Pain Reliever Use Disorder or both. These disorders consider both dependence and abuse. Therefore, respondents are classified as having pain reliever or opioid abuse if they reported a positive response to more than one of: (i) having serious problems at home or work due to abuse; (ii) using substance regularly, and then doing something that might have put them in physical danger; (iii) causing actions that repeatedly got them in trouble with the law; (iv) continuing to use substance even though it caused problems with family and friends. Adults are classified as having dependence on a substance if they responded positive to three or more of: (i) spent more than a month getting over the effects of substance abuse; (ii) unable to set limits on substance use, or used more often than intended; (iii) needed to use substance more than before to get desired effects or noticed that using the same amount had less effect than before; (iv) unable to reduce the amount used or stop using the substance every time he or she tried or wanted to; (v) continued to use substance even though it was causing problems with emotions, nerves, mental health, or physical problems; (vi) reduced or gave up participation in important activities due to substance use. (Center for Behavioral Health Statistics and Quality, 2016).

In addition to the OUD/No-OUD classification, characteristics that were believed to be germane to the prediction of OUD were included in the data used to train and test the three models. Table 1 summarizes these predictor variables along with the different levels that they can take. All of the variables, except for disability and obesity are available in the recoded form in the NSDUH survey. A binary value of any disability is computed from the raw variables on six types of disabilities listed in Table 1. Obesity takes three values based on the individual's Body Mass Index (BMI), which a, continuous variable. The curated data set consists of 42,324 observations. Data for adolescents between the

ages of 12 and 17, and those with missing values were eliminated.

2.2. Training the classifiers

In machine learning, the curated data is split into training and test sets. The training set is used to train the classifiers, and the test set is held out for validation, after the model has been trained (Tattar, 2018). Usually, the training and test sets contain 60 %–80 % and 40 % – 20 % of the observations, respectively. These standard train-test split combinations are unsuitable for OUD prediction, because the percentage of OUD to No-OUD observations is 1%–99% leading to a class imbalance problem. Imbalance makes the classification problem that this paper approaches challenging, and in effect moot if not solved, because even if the classifier predicted all observations as No-OUD, it would be incorrect only 1 % of the time, while not finding any observations with OUD.

Several techniques were considered to address this class imbalance problem. Class weights, which were considered initially, could impose a heavier cost when errors are made in the minority or rare class. Downsampling randomly removes instances in the majority class, until the two classes are balanced. In upsampling, instances are randomly replicated in the minority class (Tattar, 2018). The downsampling method was ultimately implemented by eliminating the No-OUD observations from the training set until it is balanced with an equal number of OUD and No-OUD observations.

The balanced training set was used to train decision tree and random forest classifiers. The logistic regression model was also built as a baseline for comparison. The three classifiers were then used to predict the outcomes (OUD/No-OUD) for the observations in the holdover, or test data set. For each of the three classifiers, this split-downsample-train-test process was repeated 50 times and the average performance measures and confidence intervals using a 95% significance level were computed. The analysis was conducted using the packages in the R programming language (Graham et al., 2018; Wickham et al., 2019; Breiman et al., 2018; Kuhn, 2019; Milborrow, n.d.; Therneau et al., 2019; Tobias et al., 2015; Xavier et al., 2019).

2.3. Evaluating the classifiers

The outcomes predicted by each classifier are compared against the observed outcomes, in machine learning, called the ground truth. These predictions can be organized into four groups in a confusion matrix as shown in Fig. 1. True positive refers to an observation where the classifier predicts OUD, and OUD is indeed present. False positive refers to an observation where the classifier predicts OUD, even though it is absent. In a true negative observation, the classifier does not predict OUD and it is indeed absent. In a false negative observation, the classifier does not predict OUD, but it is present. True negatives and true positives are correct classifications, where false negatives and false positives are incorrect classifications.

The performance of the classifier can be evaluated using sensitivity and specificity. Sensitivity, also known as recall, is the ability of the classifier to correctly identify those with OUD (true positive rate), while specificity is the ability of the classifier to correctly identify those without OUD (true negative rate). Sensitivity is the percentage that a person with OUD is correctly classified as such. If the classifier is highly sensitive and it classifies a person as not having OUD, then it can be fairly certain that that individual indeed does not have OUD. If the classifier is highly specific, and it predicts that the individual has OUD, then it can be fairly certain that in fact, the individualhas OUD.

		Observed (Ground Truth)	
		OUD	No-OUD
Predicted	OUD	True Positive (TP)	False Positive (FP)
	No-OUD	False Negative (FN)	True Negative (TN)

Fig. 1. Confusion Matrix.

A classifier will usually outputs a probability that a test input belongs to the OUD or the No-OUD class. This output is compared against a default threshold of 50 % to determine the ultimate class of the test input. If the probability that a test input belongs to the OUD class is greater than 50 %, it is labeled as having OUD. Likewise, if the probability that a test input belongs to the No-OUD class is greater than 50 %, it is labeled as not having OUD, or No-OUD. Now, if this decision threshold is varied, the number of true and false classifications will change, and correspondingly, the sensitivity and specificity will change. Thus, varying the decision threshold leads to a different tradeoff between sensitivity and specificity. A classifier with a high sensitivity usually has low specificity; it is therefore successful in finding actual cases of OUD but it also has a relatively high rate of false positives.

The ROC (Receiver Operating Characteristics) curve, which shows the sensitivity or the true positive rate as a function of the specificity or the false negative rate for different thresholds was used to assess this tradeoff. Each point on the ROC curve represents a sensitivity and specificity pair corresponding to a particular threshold. When the threshold is higher, the false positive fraction decreases with increased specificity but the true positive fraction and sensitivity will decrease. When the decision threshold is lower, the true positive fraction and sensitivity increases, but the false positive fraction also increases, and therefore the true negative fraction and specificity decreases. The area under the ROC curve (AUC) is a measure of how well the classifier can distinguish between OUD and No-OUD classes. Overall, the performance of a classifier is regarded as fair if the AUC falls in the range of 0.70-0.80, good if it falls in the range of 0.80-0.90, and exceptional if it is greater than 0.90.

3. Results and discussion

The average values and 95 % confidence intervals for sensitivity, specificity, and AUC were computed for the three classifiers over 50 runs. The average performance measures as well as the 95 % confidence intervals are shown in Table 2 for the three. The random forest classifier can predict adults at risk for OUD with remarkable accuracy, with average AUC of over 0.89, an improvement over prior AUC of 0.86 (Wadekar, 2019). Thus, incorporating accessibility of drugs within one's community, criminal involvement, and perception of risk produces a more comprehensive approach with better prediction accuracy. The performance measures of the random forest classifier are slightly better compared to the logistic regression classifier. Moreover, the confidence intervals of the random forest classifier are narrower compared to the other two classifiers. Both the logistic regression and random forest classifiers outperform the decision tree classifier, but not by a very large margin.

The classifiers show higher sensitivity over specificity. That means, they correctly identify many with OUD, but flag some incorrectly. If specificity were higher than sensitivity, the classifiers would miss a higher percentage with OUD. In OUD prediction, however, it is desirable to have higher sensitivity over specificity because identifying people at risk for OUD can ultimately save lives by providing them with early support. Thus, in the interest of public health, higher sensitivity as opposed to a higher specificity is a more desirable outcome rather than the reverse scenario.

The different types of analysis results provided by the three classifiers can be used synergistically to improve our understanding of OUD. Table 3 lists the average p-values of the specific levels of the predictors from the logistic regression model. These p-values are computed using

Table 2
Average (Confidence Intervals) of Performance Measures – Evaluated on 50 Runs.

Classifier	Sensitivity	Specificity	AUC
Logistic Regression	0.8150 (± 0.0123)	0.8007 (± 0.0104)	0.8854 (± 0.0104)
Decision Tree	0.7958 (± 0.0123)	0.8037 (± 0.0120)	0.8477 (± 0.0060)
Random Forest	0.8197 (± 0.0081)	0.8103 (± 0.0026)	0.8938 (± 0.0023)

Table 3
Predictors, Levels, and p-values.

Predictors & Levels	p-values
Disability – yes	4.46e-10 (***)
Age 18–25	0.014076 (*)
Age 26–34	0.003557 (**)
Age 35–49	0.012485 (*)
Age 50–64	0.038768 (*)
Income 20k-49k	0.019390 (*)
Mental Illness Yes	< 2e-16 (***)
Education – College	0.034171 (*)
Health – Fair/Poor	2.72e-05 (***)
Health – Good	1.61e-06 (***)
Health – Very Good	0.043476 (***)
Employment – Other	0.017493 (*)
Employment – Unemployed	0.000341 (***)
First use of alcohol before 18 years	4.52e-05 (***)
First use of marijuana before 18 years	2.36e-14 (***)
Probation – Yes	0.002999 (**)
Perception of risk, weekly use of marijuana – yes	5.36e-05 (***)
Perception of risk, trying marijuana once in a lifetime – yes	0.006285 (**)
Heroin easy or very easy to obtain – yes	< 2e-16 (***)
Approached by someone selling drugs – yes	< 2e-16 (***)
Obesity – Obese	0.025378 (*)

Table 4
Percentage Contribution of Variables in Prediction of OUD.

Variable	Relative Importance
Early initiation of marijuana before 18	13.52 %
Mental Illness	11.78 %
Approached by someone selling drugs	8.52 %
Age	7.29 %
Overall health	7.16 %
Perception that drugs are easy to obtain	7.03 %
Income	6.26 %
Education	6.25 %
Employment	5.47 %
Early initiation of alcohol before 18	5.43 %
Race	5.36 %
Obesity	4.43 %
Disability	3.49 %
Gender	2.49 %
Perception of risk of trying heroin in lifetime	1.99 %
Probation	1.77 %
Perception of risk of using heroin weekly	1.27 %
Parole	0.49 %

the entire data set, without the train/test split. In the table, ***, ** and * indicate 0.001 %, 0.01 % and 0.05 % levels of significance respectively. The p-values split the characteristics into the two groups whether statistically significant or not. However, p-values alone cannot determine the relative importance of the predictors. For example, Table 3 suggests that early initiation of both alcohol and marijuana have a significant association with OUD. Thus, curbing early initiation of both alcohol and marijuana may prevent OUD, but it may be onerous to prohibit early use of alcohol compared to marijuana, given its wide used in social settings.

Intervention strategies can be cost effectively developed if the predictors are ranked in the order of their contribution to OUD prediction. This ranking can be produced by the random forest classifier as shown in Table 4. First use of marijuana before the age of 18 years has the highest contribution among all the predictors. It has a much higher contribution to OUD compared to early initiation of alcohol. Once again, prohibiting early initiation of marijuana in all demographic groups is desirable, but it is unlikely to be cost-effective. It then becomes necessary to narrow the socioeconomic and demographic groups that are more affected by early initiation of marijuana.

Decision trees reveal the interactions between early initiation of marijuana and other predictors. Early initiation of marijuana affects adults with no full-time employment, or those who have not completed high school, or with incomes under $49,000, or between the ages of 18–34 years, or of Hispanic and White heritage, or with fair/poor health, or on probation, or those who live in unsafe neighborhoods with easy access to drugs. It is important to note that some predictors such as probation are not identified as statistically significant by logistic regression. Random forest also pegs their relative contribution as very low. However, it is the interaction of these predictors with early initiation of marijuana that increases the risk for OUD, which can only be identified by decision trees.

The approach demonstrates the advantages of using a combination of several models. Logistic regression and random forest classifiers show comparable prediction accuracy. Logistic regression can reveal which predictors and their levels are statistically significant for OUD prediction, while random forest can rank the predictors according to their relative importance. Finally, although the prediction accuracy of decision tree is lowest, this model can identify the interactions between the significant risk factors identified by the other two models.

This research collectively overcomes many drawbacks of prior efforts. The study is reproducible because it uses public domain data, rather than private administrative or electronic health records. It is inclusive because it considers abuse and dependence of both prescription opioids and heroin, unlike some that focus only on prescription opioids (Li et al., 2018). It covers the general adult population, rather than a specific cohort such as Medicare recipients or Caucasians (Crosier et al., 2017). The sample size is larger compared to contemporary studies (Crosier et al., 2017). The method is parsimonious because it needs a small number, of predictors. Despite its parsimony, it is still comprehensive and considers many dimensions of a person's life, rather than only a narrow subset such as behavioral markers or personality traits (Ahn WY et al., n.d.; Ahn and Vassileva, 2016). Finally, this approach enjoys a higher prediction accuracy compared to contemporary approaches that apply machine learning to substance use disorders (Acion et al., 2017; Crosier et al., 2017).

This study can be beneficial in many ways. First, it can be used in primary care settings, to proactively identify at risk adults, by eliciting responses to only a few questions. Second, it identifies the risk factors and narrows their impact to certain socioeconomic and demographic groups, which can guide public health officials in prevention and early intervention. Third, it can inform policy makers regarding the regulation and legalization of substances. For example, this research adds another dimension to the ongoing debate about legalizing marijuana by highlighting that early initiation of marijuana may also increase susceptibility to OUD in adult life.

This is one of the first comprehensive approaches to predict adults at risk for OUD. Using three models, that have their respective advantages, synergistically on a secondary data set in the public domain, this approach improves our understanding of the interactions among the characteristics that increase an adult's risk of developing OUD. It

showcases that many large secondary data sets hold rich information, vasts amounts of information, and employ statistically sound data collection methodology. Using these secondary data can alleviate the burden of collecting, anonymizing, cleaning, and recoding primary data. However, certain limitations also arise from using secondary data, because only those characteristics that are included in the survey can be incorporated into the model. For example, in predicting OUD, the use of substances and other prescription medications by family members may be relevant. Additionally, debilitating and painful health conditions such as neuropathy may also lead to abuse of opioids. However, because these data types are not collected in the NSDUH survey, they cannot be considered in the model.

4. Conclusions

The proposed machine learning approach predicts adults at risk for OUD with remarkable accuracy. The dominant predictor of OUD is first use of marijuana before the age of 18 years. Socioeconomic and demographic groups affected by such early initiation are also identified. The machine learning models are capable of finding a "needle in a haystack", given the low number of observations with OUD. Finally, it is shown how a combination of different machine learning methods can be used to comprehensively and synergistically predict Opioid Use Disorder in adults.

Contributors

Adway S. Wadekar performed all of the methods and analyses described in this paper. He is the sole author on this paper as well. A.S. Wadekar also prepared the article alone and he has approved the final article.

Role of funding source

There is no funding from any organization for this research.

Declaration of Competing Interest

The author does not have any conflicts of interest to report.

References

Acion, L., Kelmansky, D., Laan, M. der, Sahker, E., Jones, D., Arndt, S., 2017. Use of a machine learning framework to predict substance use disorder treatment success. PLoS One 12, e0175383. https://doi.org/10.1371/journal.pone.0175383.

Ahn WY, Ramesh D, Moeller FG, Vassileva J, n.d. Utility of Machine-Learning Approaches to Identify Behavioral Markers for Substance Use Disorders: Impulsivity Dimensions as Predictors of Current … - PubMed - NCBI [WWW Document]. URL https://www.ncbi.nlm.nih.gov/pubmed/27014100 (accessed 5.9.19).

Ahn, W.-Y., Vassileva, J., 2016. Machine-learning identifies substance-specific behavioral markers for opiate and stimulant dependence. Drug Alcohol Depend. 161, 247–257. https://doi.org/10.1016/j.drugalcdep.2016.02.008.

Center for Behavioral Health Statistics and Quality, 2016. 2016 National Survey on Drug Use and Health Public Use File Codebook.

Centers for Disease Control, n.d. Module 5: Assessing and Addressing Opioid Use Disorder (OUD) [WWW Document]. URL https://www.cdc.gov/drugoverdose/training/oud/accessible/index.html (accessed 5.2.19).

Crosier, B.S., Mateu-Gelabert, P., Guarino, H., Borodovsky, J., 2017. Finding a needle in the haystack: using machine-learning to predict overdose in opioid users. Drug Alcohol Depend. 171, e49. https://doi.org/10.1016/j.drugalcdep.2016.08.146.

Daniulaityte, R., et al., 2014. Sources of pharmaceutical opioids for non-medical use among young adults. J. Psychoact. Drugs 46 (3), 198–207. https://doi.org/10.1080/02791072.2014.916833.

Fiellin, L.E., Tetrault, J.M., Becker, W.C., Fiellin, D.A., Hoff, R.A., 2013. Previous use of alcohol, cigarettes, and marijuana and subsequent abuse of prescription opioids in young adults. J. Adolesc. Health 52, 158–163. https://doi.org/10.1016/j.jadohealth.2012.06.010.

Finn, K., 2018. Why marijuana will not fix the opioid epidemic. Med. 115, 191–193.

Graham, Williams, Vere Culp, Mark, Cox, Ed, Nolan, Anthony, White, Denis, Medri, Daniele, et al., 2018. Package 'rattle' - Graphical User Interface for Data Science in R.

Wickham, H., François, R., Henry, L., Müller, K., 2019. Package 'dplyr' - a Grammar of Data Manipulation.

Breiman, Leo, Cutler, Adele, Liaw, Andy, Wiener, Matthew, 2018. Package 'randomForest'' - Breiman and Cutler's Random Forests for Classification and Regression.

Li, X., Chaovalitwongse, W.A., Curran, G., Tilford, J.M., Felix, H., Martin, B.C., 2018. Using machine learning to predict opioid overdoses among prescription opioid users. Value Health 21, S245. https://doi.org/10.1016/j.jval.2018.04.1663.

Kuhn, Max, 2019. Package 'caret' - Classification and Regression Training.

National Institute on Drug Abuse, 2017. The Science of Drug Use: Discussion Points [WWW Document]. URL (accessed 5.2.19). https://www.drugabuse.gov/related-topics/criminal-justice/science-drug-use-discussion-points.

National Rehabilitation Information Center, 2011. Substance Abuse & Individuals with Disabilities Volume 6 National Rehabilitation Information Center Number 1, January 2011 [WWW Document]. URL (accessed 5.9.19). https://naric.com/?q=en/publications/volume-6-number-1-january-2011-substance-abuse-individuals-disabilities.

Stephen Milborrow, n.d. Plot "rpart" Models: An Enhanced Version of "plot.rpart.".

Substance Abuse and Mental Health Services Administration, 2016. National Survey on Drug Use and Health, 2016 ed. [WWW Document]. URL (accessed 5.9.19). https://www.datafiles.samhsa.gov/.

Tattar, P.N., 2018. Hands-On Ensemble Learning With R.

Therneau, Terry, Atkinson, Beth, Ripley, Brian, 2019. Package 'rpart'.

Tinker, B., 2019. Opioid Epidemic: Eastern United States Most Affected - CNN. [WWW Document]. URL. https://www.cnn.com/2019/02/22/health/opioid-deaths-states-study/index.html.

Tobias, Sing, Sander, O., Beerenwinkel, Niko, Lengauer, Thomas, 2015. Package 'ROCR' - Visualizing the Performance of Scoring Classifiers.

Wadekar, A., 2019. Predicting Opioid Use Disorder (OUD) using a random Forest. In: 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC). Presented at the 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC). pp. 960–961. https://doi.org/10.1109/COMPSAC.2019.00161.

Xavier, Robin, Turck, Natacha, Hainard, Alexandre, Tiberti, Natalia, Lisacek, Frédérique, Sanchez, Jean-Charles, Müller, Markus, Siegert, Stefan, Doering, Matthias, 2019. Package "pROC" - Display and Analyze ROC Curves.