Predicting Drug Abuse Through Decision Trees

Lucas Manker
COSC 4750
Final Project Report
May 10, 2020

Predicting Drug Abuse Through Decision Trees

Lucas Manker
University of Wyoming

COSC 4750

Abstract

Drug addiction is a serious problem facing many communities around the world. Dependence on drugs can be precipitated by a poor economy, social issues, and psychological traits. The effects of economic and social issues are well researched and generally understood, but the effect of certain psychological traits upon drug addiction are difficult to quantify. A deeper understanding of the effect that psychological traits have upon the propensity to become addicted to drugs can allow government and healthcare entities to properly combat drug addiction.

Based on a review of recent literature on the subject, there is evidence that some psychological traits impact the likelihood of drug dependence. To examine this further, a data set from the UCI Machine Learning Repository was used to gain some insight into correlations between certain psychological traits and drug use. The data was curated to be more accessible for the purposes of the examination, and a decision tree was created to assign respondents into levels of dependence upon narcotics.

The decision tree relied on psychological features to make classifier predictions for the level of use among specific illicit drugs. This model produced poor results, and was particularly inaccurate in terms of predicting heavy users. To combat this weakness in the model, the use of "gateway drugs" was added as a feature, and this allowed for a slightly more accurate prediction model. The end results still lack a reasonable amount of correlation between psychological traits and drug use, and it is clear that a larger data set or more finely tuned model is necessary to provide a deeper foundation to draw conclusions.

Introduction

Drug addiction within the United States has been a topic of concern for many decades, and the economic and societal toll has prompted many politicians to act. "The report draws attention to a problem of epidemic proportions, as indicated by the fact that more than 6 percent of the United States population has a substance abuse disorder and an estimated 1,354,000 people die prematurely every year as a consequence of drug overdose and alcohol abuse" (Aldo Badiani 2017). The War on Drugs was begun to try to stem the flow of drugs entering the United States, and make attempts at combating and preventing addiction. The Drug Enforcement Agency was founded to stop the smuggling of drugs, and the D.A.R.E. program was implemented within elementary schools across the United States to try to stop addiction before it began, but as of today it appears that drugs have won the War on Drugs.

The psychological factors which contribute to an individual's addiction to narcotics has not been fully documented, and used to the benefit of the War on Drugs. It is perhaps this key piece of knowledge which could allow for a more informed and effective approach in the fight against drug addiction. Recent literature seems to agree that psychology should play a role in understanding drug addiction. "It is necessary to understand in every case how the specific characteristics of the drug and the personality of the user interact and are modified by the social setting and its controls"(Lisa Maher 2017).

Data Mining techniques applied to large data sets combining drug use statistics and psychological groupings could provide some illumination to the question of correlation between psychological traits and drug abuse. In this study, I attempted to create a small model to classify and predict which psychological traits would lead individuals down the path towards addiction.

A similar study was conducted by Alireza Amirabadizadeh, Hossein Nezami, Michael G. Vaughn, Samaneh Nakhaee, and Omid Mehrpour in Iran to predict Heroin use within Iran. The study was more concerned with demographic information, but they also created a decision tree to classify potential Heroin users. The study found that if an individual used gateway drugs such as hash, there was a 91.23% likelihood the individual would use Heroin at some point (Alireza Amirabadizadeh 2017). Although Iran has a culture which is different from the culture of the respondents in the data set being examined, it appears that gateway drugs do contribute to future addiction to more serious narcotics.

Methods

The data set being used for this study was the UCI Drug Consumption Data Set. The set contained 1885 respondents, with 32 attributes. The attributes are as follows:

- **ID**: The respondent number associated with their responses.
- **Age**: The age of the respondent.
- **Gender**: The gender of the respondent.
- **Education**: The level of education completed by the respondent.
- **Country**: The country of residence for the respondent.

- **Ethnicity**: The ethnicity of the respondent.
- **NScore**: Values associated with the neuroticism of the respondent.
- **EScore**: Values associated with the extraversion of the respondent.
- **OScore**: Values associated with the respondent's openness to new experience.
- **AScore**: Values associated with the agreeableness of the respondent.
- **CScore**: Values associated with the conscientiousness of the respondent.
- **Impulsive**: Values associated with the impulsiveness of the respondent.
- **SS**: Values associated with the respondent impulsively seeking new sensations.

Drug use statistics were then recorded into seven classes.

- **CL0**: Never Used
- **CL1**: Used over a Decade Ago.
- **CL2**: Used in Last Decade.
- **CL3**: Used in Last Year.
- **CL4**: Used in Last Month.
- **CL5**: Used in Last Week.
- **CL6**: Used in Last Day.

The drugs respondents were queried about were as follows:

- **Alcohol**
- **Amphet**: Amphetamine consumption.
- **Amyl**: Amyl Nitrite consumption.
- **Benzos**: Benzodiazepine consumption.
- **Caff**: Caffeine consumption.
- **Cannabis**
- **Choc**: Chocolate consumption.
- **Coke**: Cocaine consumption.
- **Crack**: Crack Cocaine consumption.
- **Ecstasy**
- **Heroin**
- **Ketamine**
- **Legalh**: A class of legal highs.
- **LSD**
- **Meth**
- **Mushrooms**
- **Nicotine**
- **Semer**: A fake drug used to identify untruthful respondents.
- **VSA**: Volatile substance inhalation.

This data was then simplified and curated with basic Python functions with the goal of creating a decision tree. The classes of drug use were simplified into 3 classes for the decision tree to label.

- **Not_User**: CL0, CL1, CL2
- **User**: CL3, CL4
- **Heavy_User**: CL5, CL6

Due to the large amount of data, it was necessary to examine one drug at a time, and attempt to make a prediction model designating the level of drug consumption expected from the respondent. Alcohol, Marijuana, Cocaine, Crack Cocaine, and Heroin were chosen to be examined. The model was split in a 70% train and 30% test ratio using SKLearn, and the trees were created. The results (as explained later) were sub-par, and provided little insight into any correlation between psychological traits and drug use, so improvements were made. Marijuana has a reputation as a gateway towards heavier drug use, so it was added as a feature for the classification of Crack and Heroin users. Cocaine also has a slightly less infamous reputation as leading users towards heavier drugs, so this too was added as a feature for Crack and Heroin predictors. Confusion matrices and F1-Scores were then computed to provide an illustration of the accuracy of the predictions.

Results

Alcohol:

		Predicted		
		Not_User	User	Heavy_User
Actual	Not_User	1	1	9
	User	1	13	43
	Heavy_User	9	69	420

F1 Not_User: 0.091
F1 User: 0.186
F1 Heavy_User: 0.866
Top 3 GINI-Gains: OScore, EScore, AScore.

Marijuana:

		Predicted		
		Not_User	User	Heavy_User
Actual	Not_User	57	8	57
	User	9	9	44
	Heavy_User	58	46	278

F1 Not_User: 0.460
F1 User: 0.143
F1 Heavy_User: 0.731
Top 3 GINI-Gains: AScore, OScore, EScore.

Cocaine:

		Predicted		
		Not_User	User	Heavy_User
Actual	Not_User	190	45	71
	User	36	15	23
	Heavy_User	77	29	80

F1 Not_User: 0.624
F1 User: 0.184
F1 Heavy_User: 0.444
Top 3 GINI-Gains: OScore, AScore, EScore.

Crack Cocaine:

		Predicted		
		Not_User	User	Heavy_User
Actual	Not_User	425	12	46
	User	17	1	7
	Heavy_User	46	2	10

F1 Not_User: 0.875
F1 User: 0.05
F1 Heavy_User: 0.165
Top 3 GINI-Gains: NScore, EScore, Age. **Heroin**:

		Predicted		
		Not_User	User	Heavy_User
Actual	Not_User	404	19	49
	User	7	1	4
	Heavy_User	58	8	16

F1 Not_User: 0.859
F1 User: 0.05
F1 Heavy_User: 0.212
Top 3 GINI-Gains: EScore, Ethnicity, NScore.

Cocaine use was then added as a feature to assist in making the model more accurate. Only Crack Cocaine and Heroin use were examined after designating Cocaine use as a feature.

Crack Cocaine:

		Predicted		
		Not_User	User	Heavy_User
Actual	Not_User	429	15	39
	User	21	1	3
	Heavy_User	34	3	21

F1 Not_User: 0.887
F1 User: 0.045
F1 Heavy_User: 0.347
Top 3 GINI-Gains: EScore, Cocaine Use, NScore.

Heroin:

		Predicted		
		Not_User	User	Heavy_User
Actual	Not_User	412	23	37
	User	7	1	4
	Heavy_User	50	9	23

F1 Not_User: 0.876
F1 User: 0.03
F1 Heavy_User: 0.32
Top 3 GINI-Gains: Ethnicity, Cocaine Use, NScore.

Discussion

The accuracy of the models certainly leaves a lot to be desired, but there is some useful analysis that can be conducted. In the case of alcohol, the model was reasonably accurate in determining if a respondent could be classified as a heavy user. As alcohol is legal in all of the countries surveyed, it stands to reason that many of the respondents regularly use the drug. The most important features also correlate effectively to what is known about alcohol. Alcohol has a reputation as a social drug, and the psychological traits of extraversion, agreeableness, and openness to new experience allow the model to accurately predict whether a person uses alcohol. Marijuana falls in the same category as alcohol as being a drug tied to socialization, and mirrors the same important features in that regard.

Cocaine use is where the model predictions for heavy users starts to decline, and the F1 Score for the prediction of such falls to less than that of a coin flip. The features most important to the prediction are interesting, and also echo that of Alcohol and Marijuana. Cocaine is also generally seen as a party drug, and the features most important to the model are the OScore, Ascore, and Escore. Crack Cocaine's use in Western society is different from the previous three drugs, and it is seen as drug only used by addicts and not associated with socialization. This correlates to the three features which provided the largest GINI gains: NScore, Escore, and Age. Heroin is also not classified as a party drug, or a drug used while socializing. The GINI gains from the features mirror this, as the NScore is also an important metric when determining a resondent's Heroin use.

Heroin and Crack Cocaine use seem to correlate with the respondent's level of neuroticism. Did this neuroticism develop after the drug use, or before? Most would categorize addicts as having some level of neuroticism, as their lower quality of life can mold personalities into something generally neurotic. Neurotic individuals are anxious, and impatient; they may turn to drugs to quickly satisfy some need. "Psychoactive drugs such as opium or cocaine seem to act as a shortcut in the pleasure mechanism" (Thiago Perez Bernades de Moraes 2015). Further research could identify respondents involved in the early stages of addiction and gauge their neuroticism, and then re-survey the respondent some years later for a comparison.

A curious piece of the model was that impulsiveness was not categorized as a good predictor for drug abuse. Drug addicts are often perceived as being impulsive, and slaves to their own desires. "Individuals with drug addiction experience difficulties in the accurate evaluation of emotional cues and in appreciating the consequences of their actions to themselves and to other people" (Georgia S. Aslanidou 2018). This seems contrary to what has been seen in the model, and should be researched more fully.

The addition of Cocaine use as a feature improved the model somewhat, especially in terms of predicting heavy users of Heroin and Crack Cocaine. In the revised model Cocaine use became a solid predictor for the use of Heroin and Crack Cocaine, and was the feature with the second largest GINI gain in both trees. This correlation could point to Cocaine use as a predictor for drugs more associated with extreme addiction.

It's difficult to make any solid assessments of the data as the accuracy of the model is so low. Future predictions could use a more finely-tuned model, with even more gateway drugs as features, and perhaps a pruning of unimportant features which simply add noise.

Conclusion

Drug use has been a scourge upon many communities within many countries. The War on Drugs has been largely ineffective at preventing and alleviating drug abuse. This study lacks a level of accuracy necessary to provide any conclusions and aid to those who seek to combat drug abuse, but it shows some promise. Demographic and income metrics are traditionally used to provide a foundation upon which the War on Drugs focuses its efforts, but this study provides some insight into the fact that psychological factors might play an important role in prevention and treatment.

This study has shown that in most cases, the features of income and demographics matter little when making predictions on drug abuse. Education level and ethnicity were shown to be sub-par features in the decision tree model, and psychological traits such as extraversion, neuroticism, and agreeableness were much more impactful in making predictions. Future studies should rely on a larger data set across many years to correlate psychological traits with drug addiction. These studies could provide valuable information towards winning the War on Drugs.

Although my work is extremely basic compared to most of what gets submitted to AAAI, sharing my work at this conference would be an end goal. Decision trees are a popular model at AAAI, and the unique topic could be of interest to some people. This would of course, require months and even years of further research to get this study up to AAAI standards but it would be an interesting topic to present.

References

Aldo Badiani, Kent C. Berridge, M. H. 2017. Addiction research and theory: a commentary on the surgeon general's report on alcohol, drugs, and health.

Alireza Amirabadizadeh, Hossein Nezami, M. G. V. 2017. Identifying risk factors for drug use in an iranian treatment sample: A prediction approach using decision trees.

Georgia S. Aslanidou, K.V. Petrides, A. S. 2018. Trait emotional intelligence profiles of parents with drug addiction and of their offspring.

Lisa Maher, G. D. 2017. Qualitative research.

Thiago Perez Bernades de Moraes, G. L. S. d. T. 2015. Addiction to drugs, evolution and society: a study on the addiction starting from the evolutionary psychology.