Addictive Behaviors Reports 8 (2018) 128–135

Contents lists available at ScienceDirect

Addictive Behaviors Reports

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

Data mining techniques for drug use research

Rafael Jiménez, Joella Anupol, Berta Cajal, Elena Gervilla*

University of the Balearic Islands, Ctra. Valldemossa, km 7.5, 07122 Palma de Mallorca, Spain

ARTICLE INFO

Keywords:
Substance use
Adolescence
Data mining
Motives
Cannabis
Tobacco
Alcohol
Cocaine

ABSTRACT

Drug use motives are relevant to understand substance use amongst students. Data mining techniques present some advantages that can help to improve our understanding of drug use issue. The aim of this paper is to explore, through data mining techniques, the reasons why students use drugs.

A random cluster sampling of schools was conducted in the island of Mallorca. Participants were 9300 students (52.9% girls) aged between 14 and 18 years old (M = 15.59, SD = 1.17). They answered an anonymous questionnaire about the frequency and type of drug used, as well as the motives.

Five classifiers techniques are compared; all of them have much better performance (% of correct classifications) than the simplest classifier (more repeated category: drug use/never drug use) in all the compared drugs (alcohol, tobacco, cannabis, cocaine). Nevertheless, alcohol and tobacco have the lower percentage of correct classifications concerning the drug use motives, whereas these use motives have better classification performance when predicts cannabis and cocaine use. When we analyse the specific motives that better predicts the category classification (drug use/never drug use), the following reasons are highlighted in all of them: "pleasant activity" (most frequent among drug users), and "friends consume" and "addiction" (both of them most frequent among never drug users). These results relate to the social dimension of drug use and agree with the statement that environmental context influences adolescent's involvement in risk behaviours. Implications of these results are discussed.

1. Introduction

The global drug consumption among adolescents is not declining, even though its consequences are well known (Broman, 2009; Lee, Brook, Finch, & Brook, 2015; Moss, Chen, & Yi, 2014). In this sense, underage drinking endangers teens in many ways. It can lead to physical and psychological problems (Heron et al., 2013; Mota et al., 2013; Popovici & French, 2013; Risher et al., 2015), later addiction, and other drug-related problems that can last throughout their lives (Dawson, Li, & Grant, 2008; Scaglione et al., 2015).

In the European Union, an estimated of 17.2 million and 2.3 million young people (15–34 years old) have consumed cannabis and cocaine in the last year respectively, representing the most used illicit drugs (European Monitoring Centre for Drugs and Drug Addition, 2018a). For instance, in young adults, Spain has one of the highest prevalence in both illegal substances, given the last year cannabis use (17.1%) and cocaine use (3%); regarding legal substance use, alcohol (79.2%) and tobacco (40.8%) use is predominant (European Monitoring Centre for Drugs and Drug Addition, 2018b; Observatorio Español de las Drogas y las Adicciones [Spanish Drug Monitoring Agency], 2018a).

In addition, the European School Survey Project on Alcohol and Other Drugs (ESPAD Group, 2016) that collects data on substance use among 15 to 16-year-old students, states that in the last 30 days, an average of 21% European adolescents have smoked, 48% consumed alcohol, and 46% reported to have been intoxicated at least once; on the other hand, the lifetime use of cannabis is an average of 16% and of cocaine is 2%.

More specifically in Spain, the last National Survey on Drug Use in Secondary School Students points out that 27.3% of 14–18-year student population have smoked tobacco and 67% have consumed alcohol in the last month, being the number of binge drinking episodes (31.7% in the past 30 days) an increasing problem in this country; furthermore, 18.3% of Spanish students (14–18 years old) have used cannabis and 1% have consumed cocaine in the last month (Observatorio Español de las Drogas y las Adicciones [Spanish Drug Monitoring Agency], 2018b).

Drinking motives (including normative beliefs, expectancies and social motives) are important constructs to take into consideration to understand alcohol use among students (Armeli, Conner, Cullum, & Tennen, 2010; Bekman et al., 2011; Crutzen, Kuntsche, & Schelleman-Offersman, 2013; Hasking, Lyvers, & Carlopio, 2011; Lyvers, Hasking, Hani, Rhodes, & Trew, 2010; Maddock & Glanz, 2005). On this subject, social motives are crucial: teenagers assume that drinking makes parties

* Corresponding author.
E-mail addresses: rafa.jimenez@uib.es (R. Jiménez), joella.anupol@uib.es (J. Anupol), berta.cajal@uib.es (B. Cajal), elena.gervilla@uib.es (E. Gervilla).

https://doi.org/10.1016/j.abrep.2018.09.005
Received 3 May 2018; Received in revised form 10 September 2018; Accepted 16 September 2018
Available online 20 September 2018

more enjoyable, helps them to approach others and share their feelings and experiences (Kuntsche, Knibbe, Gmel, & Engels, 2005). Cannabis is widely consumed as a stress-coping strategy (Fox, Towe, Stephens, Walter, & Roffman, 2011; Hyman & Sinha, 2009). And little research has been done on tobacco use motives (Vinci, McVay, Copeland, & Carrigan, 2012).

Boys, Marsden, and Strang (2001) found in a sample of polydrug users that the most popular reasons for cannabis use were to relax (96.8%), to become intoxicated (90.7%), to enhance activity (72.8%), to decrease boredom (70.1%), to sleep (69.6%) and to feel better (69.0%). In the case of cocaine, the main reasons were to help keep going (84.5%) and to help stay awake (69.0%). The most common reason for alcohol use was to get intoxicated (89.1%), to relax (82.7%), to enjoy company (74.0%), to increase confidence (70.2%) and to feel better (69.0%).

Substance use motives can also be mediators of the association between certain risk factors and the use of a specific substance. For example, it seems that drinking motives intervene in the association between alcohol expectancies and risky drinking behaviour (Van Tyne, Zamboanga, Ham, Olthuis, & Pole, 2012) or between personality and alcohol use (Littlefield, Sher, & Wood, 2010; Willem, Bijttebier, Claes, & Uytterhaegen, 2012).

Most of the cited research conducted classical statistical techniques. In contrast, few studies use data mining tools, which allow finding new ways to analyse and represent data (Larose, 2006; Palmer, Jiménez, & Gervilla, 2011; Witten, Frank, & Hall, 2011). Data Mining is the process of discovering “interesting, unexpected or valuable structures in large databases” (Hand, 2007). Data mining techniques add new ways of analysing data and representing results, as described in the development and resolution of the first objective of the study.

The paper has two main objectives. Firstly, to compare five (classical and modern) data mining techniques which are barely used in the drug related context; we compare their performance to correctly classify the participants in drug user/never drug user, according to the knowledge of the drug use motives they have reported about themselves or about others. Secondly, we want to examine the frequent reasons why high school students use drugs and if they differ from the type of substance used. In line with previous research, we hypothesize that social and pleasant motives would emerge as the most important variables. Both objectives analyse, from a complementary point of view, the motives of use. With the first objective we aim to explore the relative weight of motives of use (and its possible interaction) to predict drug users and never drug users through data mining techniques; so we are interested in knowing which of the motives can predict better the use of a certain drug or the fact of not consuming. However, in the second objective the focus is to know the ranking of the more frequent motives of use between drug users.

2. Material and methods

2.1. Participants

A random cluster sampling of schools was conducted in the island of Mallorca, and 22 schools out of 47 were chosen. A total of 9300 students, aged between 14 and 18 years old, provided responses on the analysed variables. After eliminating the unreliable answers of some adolescents, the final sample included 9284 adolescents (47.1% boys and 52.9% girls) with an average age of 15.59 years (SD = 1.17).

It is worth noting that the final sample size represented 41.16% of the population size it was extracted from (N = 22,593).

2.2. Procedure

Participation in the study was voluntary and written informed consent was given by all participants' parents or legal guardians. Moreover, the study protocol was approved by the research ethical committee. The adolescents anonymously answered a questionnaire which asked about the frequency of use of different addictive substances as well as the motives of using drugs.

We analysed drug use motives through a series of classification techniques included in Data Mining: two classical machine learning techniques, Decision Trees (DT) and Artificial Neural Networks (ANN); two modern statistical techniques, k-Nearest Neighbours (K-NN) and Naïve Bayes (NB); and a classical statistical technique, Logistic Regression (LogR).We also analyse the differences (%) in which reasons are chosen by users of a specific drug and never drug users, in order to know the effect size (Phi: Φ) of a specific motive; the higher the Φ value, the more likely that the motive is selected by the predictive model.

ANN are data processing systems whose structure and functioning are inspired by biological networks and their fundamental characteristics are parallel processing, distributed memory and adaptability to the surroundings. In this work, we used the backpropagation algorithm to analyse data.

DT create sequential partitions of a set of data that maximise the differences of a response variable, and can easily be converted to classification rules.

K-NN constructs a classification method without making assumptions concerning the shape of the function that relates the dependent variable with the independent variables; this way, k similar (neighbouring) observations are used to classify each of this in a specific category.

NB is a classification technique based on Bayes' theorem; it can predict the probability of a given case belonging to a certain class. Euclidean distance was chosen to search for the ‘neighbours’.

Finally, LogR is a classical statistical technique and may be used to classify a new observation, whose group is unknown, in one of the groups, based on the values of the predictor variable. For an extended description of these tools, see Palmer et al. (2011).

To implement these techniques, we used the freely distributed platform Weka (*Waikato Environment for Knowledge Analysis*, version 3.8.1) (Witten et al., 2011) and R (version 3.5.1) (R Core Team, 2013).

3. Results

52.7% of the students in the sample drink alcohol, 25% smoke tobacco, 18.6% use cannabis and 1.6% use cocaine.

Table 1 informs about the motives of addictive substance use by drug type. Grey colour highlights the greatest differences between drug users/never drug users. “Pleasant activity” is a central motive (boxed values) in all the compared drugs, with the highest value difference in cannabis users (Φ = 0.404): 61.2% of cannabis users chose this reason, in front of 21.5% of never drug users. “Relaxing” is also a central motive with the highest differences in cannabis (Φ = 0.349) and tobacco users (Φ = 0.271). “They are not so dangerous” is not a central motive, but it has been chosen by 36.5% of cocaine users (in front of 14.5% by never drug users, Φ = 0.152). “Friends consume” has the greatest differences in all the compared drugs, although the discriminant differences are due to the greater choice of this motive in never drug users (76%), in front of cannabis (26.8%, Φ = 0.492), tobacco (35.9%, Φ = 0.401), alcohol (40.8%, Φ = 0.318), and cocaine (26.3%, Φ = 0.282) users. We observe the same pattern of differences in the “Addiction” motive, that has been chosen by 53.9% by never drug users, in front of cannabis (27.3%, Φ = 0.269), alcohol (28.3%, Φ = 0.240), tobacco (32.9%, Φ = 0.212), and cocaine (30.7%, Φ = 0.118) users.

Moreover, Table 1 highlights (boxed values) the more frequent reasons in drug users/never drug users. Regarding drug users, most of them said that they use substances to forget problems, to find new sensations or because they find it pleasant. To relax is also a highly frequent reason, except in alcohol users. In cocaine users, they also consume to intensify dance and music (54.7%) and to last longer

Table 1
Motives adolescents give to use addictive substances by the substance they use.

Never drug use: n=1949 (21.3%) **Alcohol use:** n=4773 (52.7%) **Tobacco use:** n=2232 (25%) **Cannabis use:** n=1651 (18.6%) **Cocaine use:** n=139 (1.6%)	**Never drug use**	**Alcohol use**		**Tobacco use**		**Cannabis use**		**Cocaine use**	
	%	%	Φ	%	Φ	%	Φ	%	Φ
1. Improving relations	51.8	36.6	.139	30.2	.220	27.9	.243	29.2	.114
2.To forget problems	68.4	55.1	.122	58.7	.100	57.6	.112	59.9	.046
3. Pleasant activity	21.5	42.0	.192	52.0	.313	61.2	.404	66.4	.264
4. Better with yourself	23.6	17.2	.073	20.7	.029	22.1	.018	32.4	.052
5. To intensify dance and music	31.0	41.1	.093	38.7	.080	39.5	.089	54.7	.128
6. Improve sexual relations	15.8	9.7	.088	10.8	.075	11.9	.057	24.1	.056
7. Last longer	46.0	29.9	.153	28.2	.185	25.2	.216	56.2	.052
8. To lose inhibition	19.4	23.7	.046	21.7	.029	22.0	.032	23.4	.025
9. Friends consume	76.0	40.8	.318	35.9	.401	26.8	.492	26.3	.282
10. Addiction	53.9	28.3	.240	32.9	.212	27.3	.269	30.7	.118
11. New sensations	60.2	56.5	.034	59.3	.009	58.6	.016	63.5	.017
12. Against established	27.7	19.6	.088	20.3	.086	19.2	.100	26.3	.008
13. They are not so dangerous	14.5	17.5	.036	22.6	.103	27.2	.157	36.5	.152
14. Relaxing	25.6	35.6	.095	52.3	.271	60.1	.349	58.4	.186
15. Creativity	13.6	10.3	.048	13.7	.001	14.6	.014	20.4	.049

Grey colour highlights the greatest differences between drug users/never drug users.
Boxed values highlight the more frequent reasons in drug users/never drug users.

Table 2
Motives adolescents give to use addictive substances by the substance they use for one single substance users.

Never drug use: n=1949 (21.3%) **Only alcohol use:** n=2264 (25.3%) **Only tobacco use:** n=145 (1.6%) **Only cannabis use:** n=52 (0.6%)	**Never drug use**	**Only alcohol use**		**Only tobacco use**		**Only cannabis use**	
	%	%	Φ	%	Φ	%	Φ
1. Improving relations	51.8	44.0	.078	38.1	.070	23.1	.094
2.To forget problems	68.4	54.1	.146	60.0	.046	42.3	.091
3. Pleasant activity	21.5	30.4	.101	30.2	.053	46.2	.097
4. Better with yourself	23.6	13.7	.128	16.5	.043	11.5	.046
5. To intensify dance and music	31.0	41.5	.108	19.6	.063	11.5	.069
6. Improve sexual relations	15.8	8.6	.112	5.8	.071	11.5	.019
7. Last longer	46.0	32.4	.139	22.9	.119	11.5	.113
8. To lose inhibition	19.4	25.5	.073	19.6	.001	13.5	.024
9. Friends consume	76.0	48.7	.279	55.4	.120	26.9	.183
10. Addiction	53.9	27.6	.268	42.1	.060	19.2	.113
11. New sensations	60.2	54.3	.059	57.9	.012	40.4	.066
12. Against established	27.7	19.5	.096	20.3	.042	19.2	.031
13. They are not so dangerous	14.5	12.7	.025	11.6	.021	31.4	.077
14. Relaxing	25.6	19.7	.072	45.1	.107	56.1	.111
15. Creativity	13.6	7.4	.102	5.9	.055	14.6	.005

Grey colour highlights the greatest differences between drug users/never drug users.
Boxed values highlight the more frequent reasons in drug users/never drug users.

(56.2%). Regarding never drug users, most of them said that the others use substances because friends consume (76%), to forget problems (68.4%) and to feel new sensations (60.2%); to be addicted (53.9%) and to improve relations (51.8%) are also frequent reasons.

Table 2 shows the motives that adolescents give to use addictive substances by the substance they use (for adolescents that only consume alcohol or tobacco or cannabis).

According to Table 2, we can observe that the most frequent reasons (boxed values) of the adolescents that only use alcohol are: to forget problems, to find new sensations and because their friends also drink. These are also the same most frequent reasons why teenagers who smoke say they use substances.

It is also relevant to highlight that 44% of only alcohol users say they use drugs to improve relationships, and that adolescent smokers say that addiction (42.1%) and relaxation (45.1%) are additional frequent motives.

Regarding cannabis users, the most frequent reason is to relax, and they also frequently mention that they use it to forget problems, because it is a pleasant activity or to find new sensations (see Table 2). 31.4% of cannabis users point out that this substance is not so harmful.

To assess the predictive power of motives for drug use, we implemented data mining predictive models. We ran Decision Trees (DT), K-Nearest Neighbours (KNN), Logistic Regression (LogR), Naïve Bayes (NB) and Artificial Neural Networks (ANN). To be able to predict drug use and abstinence of the analysed substances, in the selection of the subsamples we controlled that there was a balance between consumers (subjects that used the analysed substance) and never-consumers (adolescents that never use any substance). To generate the models and to estimate their classification accuracy (predictive power), we have used on every classifier the *k-fold cross-validation* technique: the original sample is randomly partitioned into k equal sized subsamples. Of the k subsamples, a single subsample is retained as the validation data to test

Table 3
Data mining classification tools performance, against a ZeroR classifier.

	Correct classifications (%) & Training elapsed time-seconds (TS): Mean(SD) from 100 models					
Classifiers[a]	ZeroR	DT	K-NN	LogR	NB	ANN
	%	%	%	%	%	%
Alcohol (n = 3898)	49.97(0.05)	70.25(2.44)	71.29(2.11)	71.91(2.01)	70.81(2.29)	70.19(2.23)
Tobacco (n = 3898)	49.97(0.05)	73.43(1.90)	75.00(2.15)	74.58(2.06)	74.12(2.13)	74.31(2.01)
Cannabis (n = 3302)	49.97(0.06)	79.50(2.18)	80.05(1.85)	78.18(1.94)	79.20(2.09)	79.44(2.18)
Cocaine (n = 278)	49.63(0.74)	77.77(7.57)	80.47(8.03)	80.43(7.71)	83.13(7.00)	77.29(7.85)
	Correct classifications (%) & Training elapsed time-seconds (TS): Mean(SD) from 100 models					
	TS	TS	TS	TS	TS	TS
Alcohol (n = 3898)	0.00(0.01)	0.04(0.03)	0.00(0.01)	0.33(0.08)	0.00(0.01)	15.64(2.34)
Tobacco (n = 3898)	0.00(0.00)	0.05(0.01)	0.00(0.00)	0.31(0.07)	0.00(0.00)	15.91(1.58)
Cannabis (n = 3302)	0.00(0.00)	0.03(0.01)	0.00(0.00)	0.24(0.08)	0.00(0.01)	12.31(1.90)
Cocaine (n = 278)	0.00(0.00)	0.00(0.01)	0.00(0.00)	0.02(0.01)	0.00(0.00)	1.06(0.25)

Note: Number of cases (n) for each substance have been balanced for the training process (50% of cases for each classification category: drug use/never drug use).
[a] ZeroR: Simplest classifier (predicts the most repeated classification value); DT: Decision Tree (C4.5 algorithm); K-NN: K Nearest Neighbor; LogR: Logistic Regression; NB: Naïve Bayes; ANN: Artificial Neuronal Network (Multilayer Perceptron).

the model, and the remaining k-1 subsamples are used as training data. The cross-validation process is then repeated k times (the folds), with each of the k subsamples used exactly once as the validation data. Specifically, we have used k = 10 folds on every classifier, and we have repeated 10 times each k-fold cross-validation, to finally obtain 100 models for every substance.

Table 3 presents the model performance for each substance use (mean, and standard deviation, of the correct classifications and training elapsed time from 100 models). We also show the comparison of these classifiers against the ZeroR classifier (the simplest classifier), that predicts the mode (the most repeated value) for the classification variable (in this case, both categories have been balanced). In other words, the ZeroR classifier do not use any variable (predictor) to estimate the predicted category (class) for a specific case, just use the mode, whereas the other classifiers consider the information of the potential predictors. In this regard, it should be noted that if a motive of use is frequent among drug users (reason about yourself), and it is also frequent among never drug users (reason reported about others), this reason cannot discriminate between both groups and it would not be selected by the predictive model.

3.1. Alcohol

All the compared data mining techniques offer a much better performance in terms of correct classifications (difference greater than 20%) than ZeroR models, with similar values in all of them (see Table 3): LogR have the highest value of correct classifications (M = 71.91%, SD = 2.01) and ANN the lowest value (M = 70.19%, SD = 2.23). On the other hand, we observe that ANN have the highest training elapsed time-seconds (TS), with a mean of 15.64 s (SD = 2.34) to estimate a model, whereas TS is lower than a second in the other techniques. The main advantage of DT technique refers to the graphical representation of the relationship between the predictors variables (motives of use) and the predicted classification, and it shows why the model classifies a case as a drug user or never drug user. DT for alcohol use highlights the following reasons as better predictors of alcohol use/never drug use: “friends consume”, “pleasant activity” and “addiction” (Fig. 1). We can observe that these motives (selected by the model), have the highest Φ value in Table 1. Additionally, we can describe a rule of prediction related to a concrete node (subsample). Node 3 (n = 1023) shows that 63.4% of alcohol users have not chosen “friends consume” and neither “pleasant activity”, whereas in node 4 (n = 609) 88% of alcohol users have not chosen “friends consume”, but they have chosen “pleasant activity”. On the other hand, node 9 (n = 897) shows that 75.6% of never drug users have chosen “friends consume” and “addiction”, but they haven't chosen “pleasant activity”.

3.2. Tobacco

The five data mining techniques present a similar performance (correct classifications), and much better performance (difference greater than 23%) than ZeroR models. K-NN is the one that offers the better performance in classifying the adolescents (M = 75% correct classifications, SD = 2.15) and DT the lowest value (M = 73.43%, SD = 1.90) (see Table 3). ANN have the highest TS (M = 15.91 s, SD = 1.58), whereas TS is lower than a second in the other techniques. DT for tobacco use highlights the following reasons as better predictors: “friends consume”, “pleasant activity”, “relaxing” and “addiction” (Fig. 2) (motives with a high Φ value in Table 1). Node 4 (n = 734) shows that 90.1% of tobacco users have not chosen “friends consume”, but they have chosen “pleasant activity” (the same rule and similar probability than in alcohol users). On the other side, node 11 (n = 660) shows that 83.8% of never drug users have chosen “friends consume” and “addiction”, but they haven't chosen “pleasant activity” and “relaxing” (similar rule and higher probability than node 9 of alcohol model).

3.3. Cannabis

The data mining techniques present a similar performance (correct classifications), and much better performance (difference greater than 28%) than ZeroR models. K-NN offers the better performance in classifying the adolescents (M = 80.05% correct classifications, SD = 1.85) and LogR the lowest value (M = 78.18%, SD = 1.94) (see Table 3). ANN have the highest TS (M = 12.31 s, SD = 1.90), whereas TS is lower than a second in the other techniques. DT for cannabis use highlights the same predictors than the tobacco model: “friends consume”, “pleasant activity”, “relaxing” and “addiction” (Fig. 3) (motives with the high Φ value in Table 1). Node 3 (n = 782) shows that 92.7% of cannabis users have not chosen “friends consume”, but they have

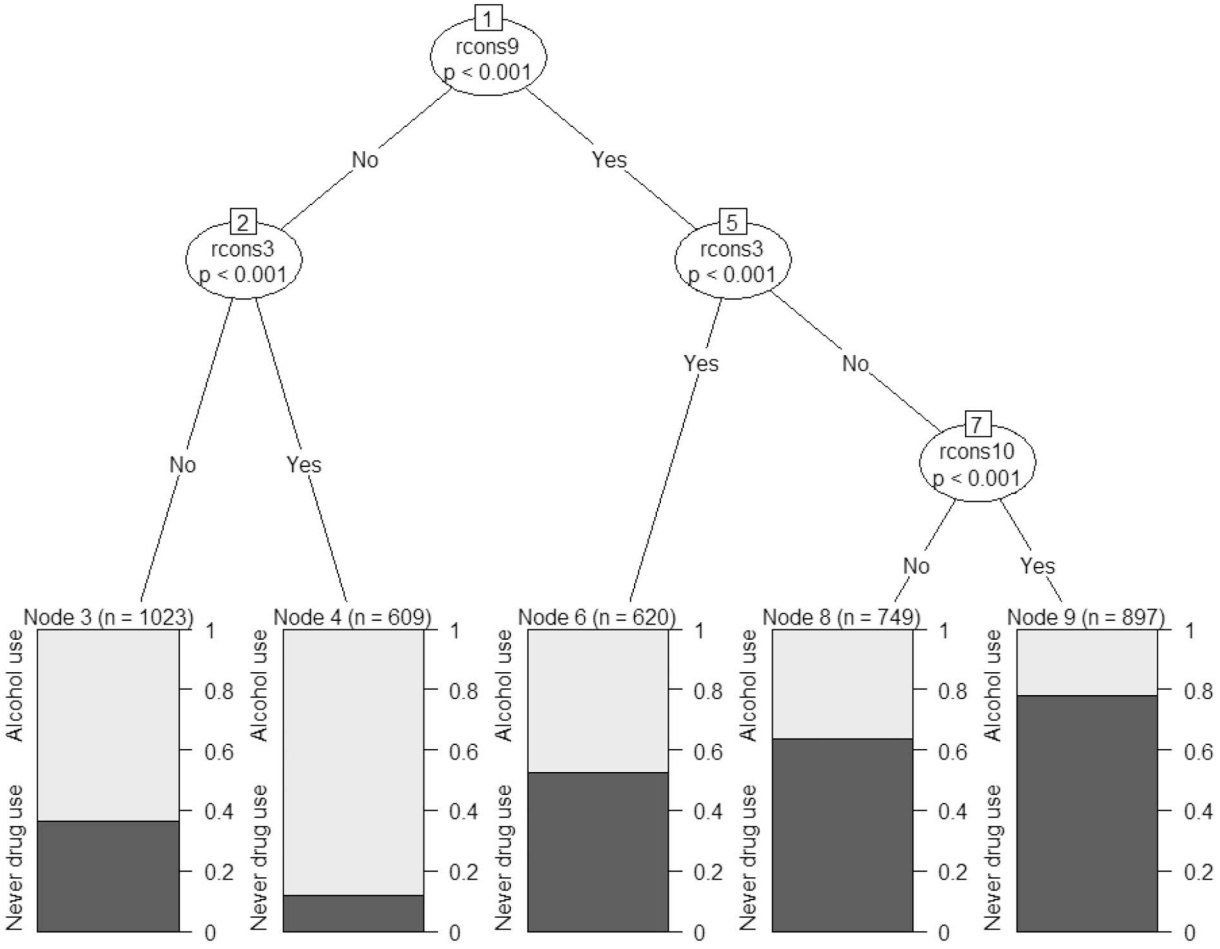

Fig. 1. Alcohol use classification pruned tree (rcons9: friends consume; rcons3: pleasant activity; rcons10: addiction).

chosen “pleasant activity” (the same rule and similar probability than in tobacco users). On the other side, node 10 (n = 458) shows that 95.3% of never drug users have chosen “friends consume” and “addiction”, but they haven't chosen “pleasant activity” and “relaxing” (same rule and higher probability than node 11 of tobacco model).

3.4. Cocaine

All the data mining techniques present a much better correct classification (difference greater than 27%) than ZeroR models. Nevertheless, because of the small sample size of balanced data for cocaine users/never drug users (n = 278), in cocaine models there are greater differences in performance (correct classifications) between data mining techniques than the others substance models, as well as greater variance in the performance mean (over 100 models). NB offers the better performance in classifying the adolescents (M = 83.13% correct classifications, SD = 7.00) and ANN the lowest value (M = 77.29%, SD = 7.85) (see Table 3). ANN have again the highest TS (M = 1.06 s, SD = 0.25), but it is obviously much lower (small sample size) than the other substances (big sample size). DT for cocaine use highlights the following reasons as better predictors: “friends consume”, “pleasant activity”, and “they are not so dangerous” (Fig. 4). Node 3 (n = 75) shows that 92% of cocaine users have not chosen “friends consume”, but they have chosen “pleasant activity” (same rule and similar probability in cannabis users). On the other side, node 7 (n = 107) shows that 85.7% of never drug users have chosen “friends consume”, but they haven't chosen “they are not so dangerous”; otherwise, node 6 (n = 34) shows that 61.8% of cocaine users have chosen “they are not so dangerous” and “friends consume”.

4. Discussion

This paper aimed to compare classical and modern data mining techniques which are barely used in the drug related context and to examine the frequent reasons why high school students use drugs and if the reasons differ from the type of substance used.

We have compared five classification techniques from an exploratory point of view, in order to explore if any pattern could be discovered in relation to the discrimination of adolescent drug users and never drug users. Results show that in all the analysed techniques alcohol and tobacco have the lower percentage of correct classifications concerning the better predictors (drug use motives), whereas these motives of use have better classification performance in the prediction of cannabis and cocaine use. Concerning the time needed to train the models, artificial neural networks need more computational resources than the other techniques, and therefore, it is the one that requires significantly more time. On the other hand, we have also showed the descriptive ability of decision trees, because they allow to graphically represent the model predictive rules.

Therefore, what is the best technique for classification? There is no general answer that can help us to know prior to data analysis which technique or algorithm we should apply to obtain the best classificatory model. In this sense, Nisbet, Elder, and Miner (2009) indicate that if different classificatory algorithms are used, we will discover that the

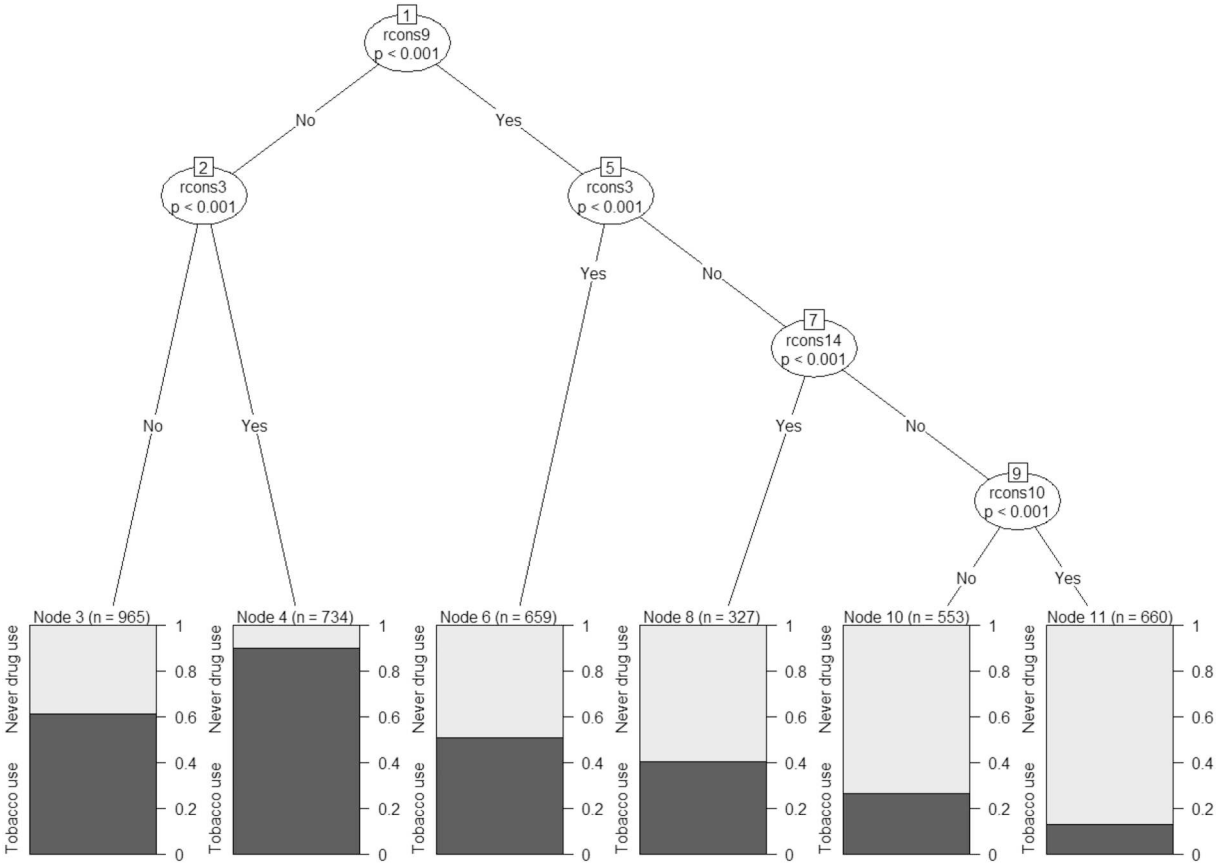

Fig. 2. Tobacco use classification pruned tree (rcons9: friends consume; rcons3: pleasant activity; rcons14: relaxing; rcons10: addiction).

best algorithm for classifying a set of data may not work well in another set of data; in other words, different techniques or algorithms have a better functioning in different data sets, and in this sense, they claim that to use a diversity of algorithms is the best option.

Regarding the second goal, we hypothesized that social and pleasant motives would emerge as the most important variables. Results confirm our hypothesis and show that the most frequent reasons to use substances are to feel new sensations, to forget problems, because friends consume and because it is a pleasant activity. However, it is important to highlight that some relevant differences emerge between motives chosen by drug users and those chosen by never drug users; in this regard, never drug users perceive that drug users consume because of the influence of friends and the effect of addiction as well. However, these two motives are not chosen with the same frequency by drug users. Anderson, Grunwald, Bekman, Brown, and Grant (2011) stated that it would be useful to analyse non-consumers motives. We have offered data on this issue comparing drug users and never drug users.

When we analyse the specific motives that better predicts the category classification (drug use/never drug use), the following reasons are highlighted in all of them: “pleasant activity” (most frequent among drug users), and “friends consume” and “addiction” (both of them most frequent among never drug users).

These motives of use are in line with the popular reasons offered by Boys et al. (2001), relate to the social dimension of drug use and agree with the findings that link environmental context with adolescents' involvement in risk behaviours such as drug use (Hakkarainen, Karjalainen, Raitasalo, & Sorvala, 2015; Schellerman-Offermans, Kuntsche, & Knibbe, 2011; Trucco, Colder, Wieczorek, Lengua, & Hawk, 2014).

Research in this area indicates that prevention interventions for young people should target norms and perceptions of normality. Thus, social agents like friends, mass media but also fathers and teachers have an active role in monitoring adolescents' substance use risky behaviour. Davis and Spillman (2011) studied the reasons why some individuals seem to have more resilience when faced with drugs use than others. The most cited reasons for not consuming were: fear of the physical damage, parental disapproval of drug use and a belief that drugs would interfere with personal goals. In addition, the last European drug report (European Monitoring Centre for Drugs and Drug Addition, 2018a) showed that students in countries with fewer users perceive cannabis use as riskier. And these lead, again, to the social perspective and the rule perception on drug use, both shaped by the information offered in the social context (i.e. friends).

The interest of this study lies in having a big sample and to apply data mining techniques that allow extracting associations in big matrices. Crutzenand and Giabbanelli (2014) have also implemented these techniques to analyse the association between drinking motives and binge drinking. Our study has extended this analysis to tobacco, cocaine and cannabis.

The results of this study should be interpreted in a transversal self-reported perspective.

Future research should focus on identifying the reasons why teenagers start using drugs. Some researchers have found that those who had their first drink in a party, had a higher likelihood of future risky drinking (Kuntsche & Muller, 2012) and they studied if this social motive depends on gender and age at this stage of development (Kuntsche et al., 2015).

Finally, it is necessary to quantify if the number of drinking peers present in the drinking event makes adolescents drink more. Thrul and Kuntsche (2015) have found this effect in young adults. Furthermore, the role of the normative perception regarding substance use must be studied to compare it to the real prevalence of consumption in a particular context (i.e. a group of friends) (Stock et al., 2014).

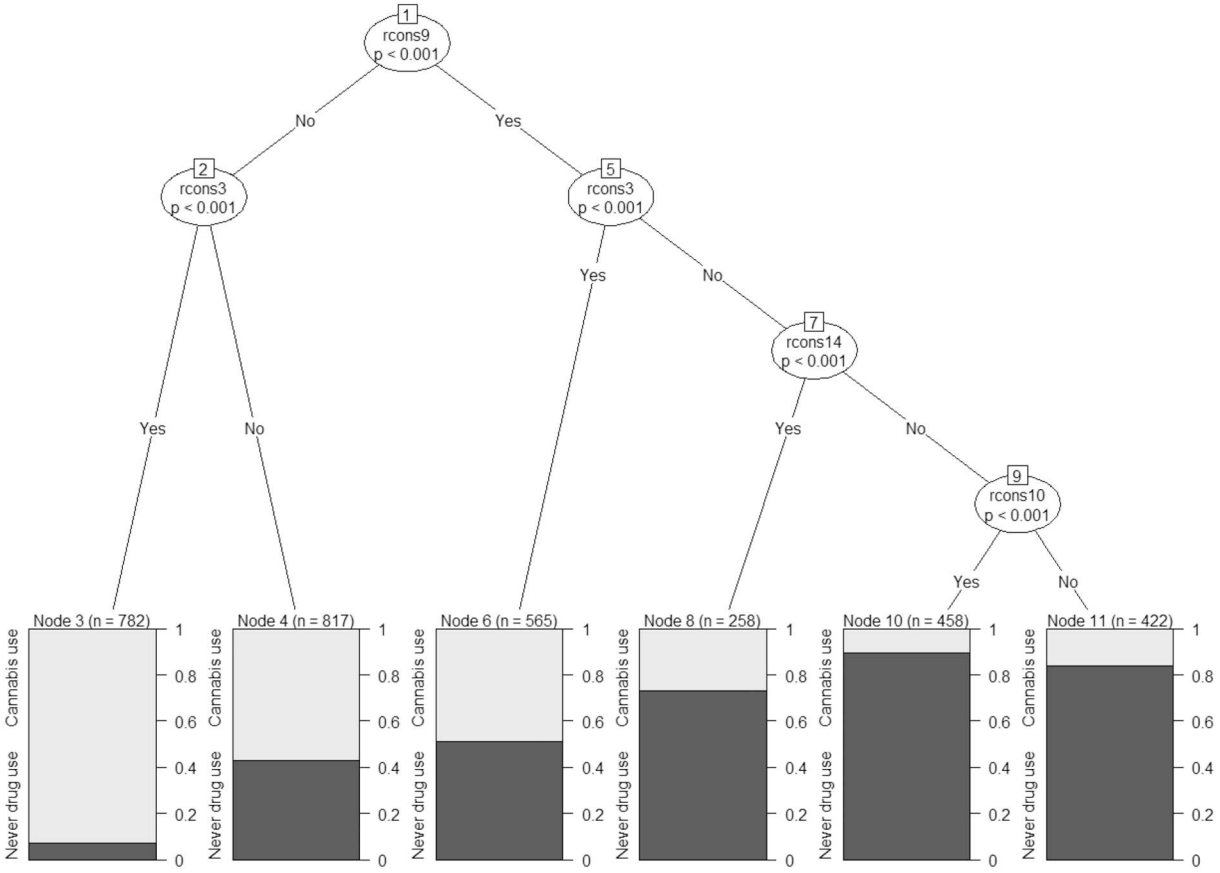

Fig. 3. Cannabis use classification pruned tree (rcons9: friends consume; rcons3: pleasant activity; rcons14: relaxing; rcons10: addiction).

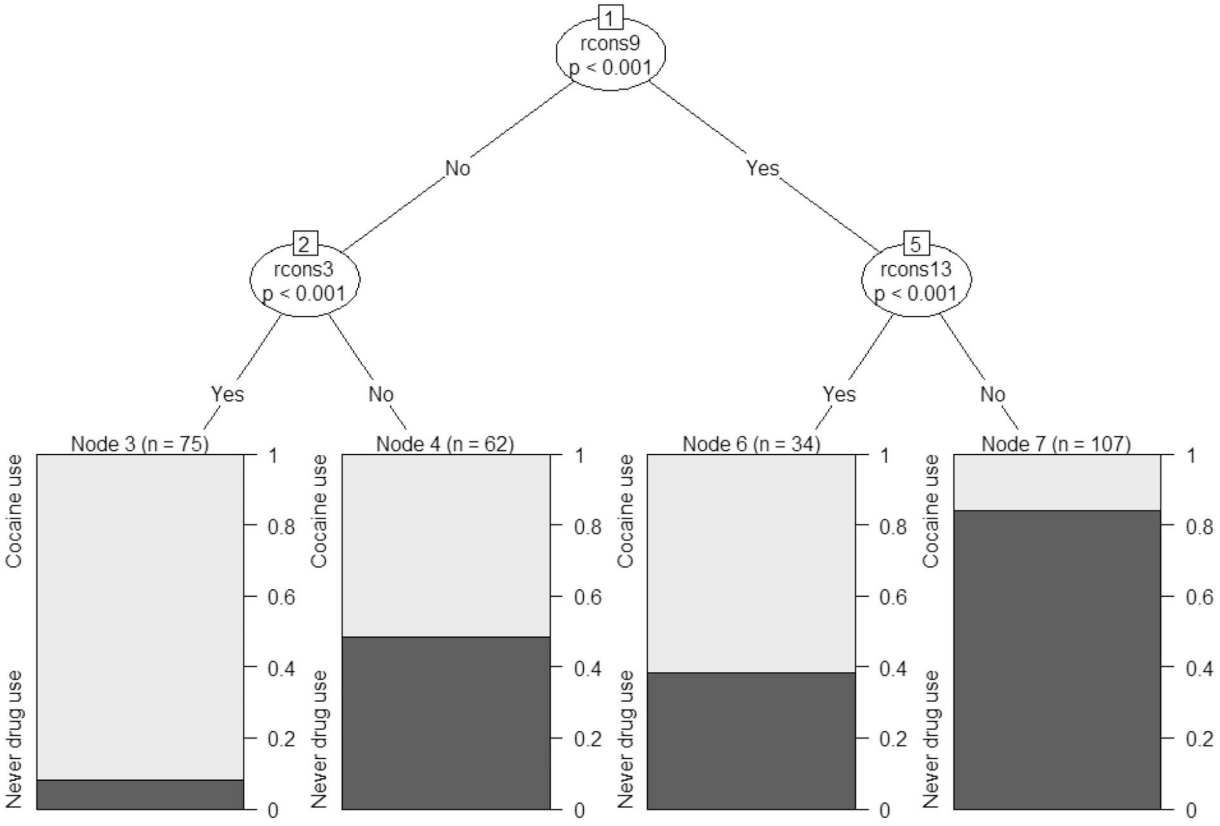

Fig. 4. Cocaine use classification pruned tree (rcons9: friends consume; rcons3: pleasant activity; rcons13: they are not so dangerous).

5. Conclusions

Never drug users perceive that adolescents use drugs because friends consume, to forget problems and to feel new sensations. Half of the adolescents who drink alcohol believe that people use drugs to feel new sensations, to forget problems and because friends consume. Adolescents who smoke mainly think that people use drugs to find new sensations, to forget problems, to relax, and because it is a pleasant activity. And cannabis and cocaine users highlight as the more frequent motive of use that it is a pleasant activity, to relax and to find new sensations.

Data mining techniques are useful techniques to analyse substance use risk and protective factors. Friends' use and pleasant activity are the main motives in order to distinguish between adolescent substance users and not users.

References

Anderson, K. G., Grunwald, I., Bekman, N., Brown, S., & Grant, A. (2011). To drink or not to drink: Motives and expectancies for use and nonuse in adolescence. *Addictive Behaviors, 36*, 972–979.

Armeli, S., Conner, T. S., Cullum, J., & Tennen, H. (2010). A longitudinal analysis of drinking motives moderating the negative affect-drinking association among college students. *Psychology of Addictive Behaviors, 24*(1), 38–47.

Bekman, N. M., Anderson, K. G., Trim, R. S., Metrik, J., Diulio, A. R., Myers, M. G., & Brown, S. A. (2011). Thinking and drinking: Alcohol-related cognitions across stages of adolescent alcohol involvement. *Psychology of Addictive Behaviors, 25*(3), 415–425.

Boys, A., Marsden, J., & Strang, J. (2001). Understanding reasons for drug use amongst young people: A functional perspective. *Health Education Research, 16*(4), 457–469.

Broman, C. L. (2009). The longitudinal impact of adolescent drug use on socio economic outcomes in young adulthood. *Journal of Child & Adolescent Substance Abuse, 18*, 131–143.

Crutzen, R., Kuntsche, E., & Schelleman-Offersman, K. (2013). Drinking motives and drinking behaviour over time: A full cross-lagged panel study among adults. *Psychology of Addictive Behaviors, 27*(1), 197–201.

Crutzenand, R., & Giabbanelli, P. (2014). Using classifiers to identify binge drinkers based on drinking motives. *Substance Use & Misuse, 49*(1–2), 110–115.

Davis, S. J., & Spillman, S. (2011). Reasons for drug abstention: A study of drug use and resilience. *Journal of Psychoactive Drugs, 43*(1), 14–19.

Dawson, D. A., Li, T.-K., & Grant, B. F. (2008). A prospective study of risk drinking: At risk of what? *Drug and Alcohol Dependence, 95*(1–2), 62–72.

ESPAD Group (2016). *ESPAD report 2015: Results from the European school survey project on alcohol and other drugs.* Luxembourg: Publications Office of the European Unionhttps://doi.org/10.2810/289970.

European Monitoring Centre for Drugs and Drug Addition (2018a). European drug report: Trends and developments. Retrieved from http://www.emcdda.europa.eu/system/files/publications/8585/20181816_TDAT18001ENN_PDF.pdf.

European Monitoring Centre for Drugs and Drug Addition (2018b). Spain: Country Drug Report 2018. Retrieved from http://www.emcdda.europa.eu/countries/drug-reports/2018/spain_en.

Fox, C. L., Towe, S. L., Stephens, R. S., Walter, D. D., & Roffman, R. A. (2011). Motives for cannabis use in high-risk adolescent users. *Psychology of Addictive Behaviors, 25*, 492–500.

Hakkarainen, P., Karjalainen, K., Raitasalo, K., & Sorvala, V. M. (2015). School's in! Predicting teen cannabis use by conventionality, cultural disposition and social context. *Drugs: Education, Prevention and Policy*. https://doi.org/10.3109/09687637.2015.1024611.

Hand, D. J. (2007). Principles of data mining. *Drug Safety, 30*(7), 621–622.

Hasking, P., Lyvers, M., & Carlopio, C. (2011). The relationship between coping strategies, alcohol expectancies, drinking motives and drinking behaviour. *Addictive Behaviors, 36*, 479–487.

Heron, J., Maughan, B., Dick, D. M., Kendler, K. S., Lewis, G., Mcleod, J., ... Hickman, M. (2013). Conduct problem trajectories and alcohol use and misuse in mid to late adolescence. *Drug and Alcohol Dependence, 133*(1), 100–107.

Hyman, S. M., & Sinha, R. (2009). Stress-related factors in cannabis use and misuse: Implications for prevention and treatment. *Journal of Substance Abuse Treatment, 36*, 400–413.

Kuntsche, E., Knibbe, R., Gmel, G., & Engels, R. C. M. E. (2005). Why do young people drink? A review of drinking motives. *Clinical Psychology Review, 25*, 841–861.

Kuntsche, E., & Muller, S. (2012). Why do young people start drinking? Motives for first-time alcohol consumption and links to risky drinking in early adolescence. *European Addiction Research, 18*(1), 34–39.

Kuntsche, E., Wicki, M., Windllin, B., Roberts, C., Gabhain, S. N., van der Sluijs, W., ... Demetrovics, Z. (2015). Drinking motives mediate cultural differences but not gender differences in adolescent alcohol use. *Journal of Adolescent Health, 56*(1), https://doi.org/10.1016/j.jadohealth.2014.10.267.

Larose, D. T. (2006). *Data mining methods and models.* Hoboken, NJ: Wiley.

Lee, J. Y., Brook, J. S., Finch, S. J., & Brook, D. W. (2015). Trajectories of marijuana use from adolescence to adulthood predicting unemployement in the mid 30s: Marijuana trajectories predicting unemployement. *The American Journal on Addictions*. https://doi.org/10.1111/ajad.12240.

Littlefield, A. K., Sher, K. J., & Wood, P. K. (2010). Do changes in drinking motives mediate the relation between personality change and "maturing out" of problem drinking? *Journal of Abnormal Psychology, 119*, 93–105.

Lyvers, M., Hasking, P., Hani, R., Rhodes, M., & Trew, E. (2010). Drinking motives, drinking restraint and drinking behaviour among young adults. *Addictive Behaviors, 35*, 116–122.

Maddock, J., & Glanz, K. (2005). The relationship of proximal normative beliefs and global subjective norms to college students' alcohol consumption. *Addictive Behaviors, 30*, 315–323.

Moss, H. B., Chen, C. M., & Yi, H. (2014). Early adolescent patterns of alcohol, cigarettes, and marijuana polysubstance use and young adult substance use outcomes in a nationally representative sample. *Drug and Alcohol Dependence, 136*, 51–62.

Mota, N., Parada, M., Crego, A., Doallo, S., Caamaño-Isorna, F., Rodríguez, S., ... Corral, M. (2013). Binge drinking trajectory and neuropsychological functioning among university students: A longitudinal study. *Drug and Alcohol Dependence, 133*(1), 108–114.

Nisbet, R., Elder, J., & Miner, G. (2009). *Handbook of statistical analysis & data mining applications.* San Diego, CA: Academic Press.

Observatorio Español de las Drogas y las Adicciones [Spanish Drug Monitoring Agency] (2018a). Encuesta domiciliaria sobre Alcohol y otras Drogas en España (EDADES) [household survey on alcohol and drugs in Spain]. Retrieved from http://www.pnsd.msssi.gob.es/profesionales/sistemasInformacion/sistemaInformacion/pdf/2017_Estadisticas_EDADES.pdf.

Observatorio Español de las Drogas y las Adicciones [Spanish Drug Monitoring Agency] (2018b). Encuesta estatal sobre uso de drogas en Estudiantes de Enseñanzas Secundarias (ESTUDES) [National Survey on Drug Use among Secondary School Students]. Retrieved from http://www.pnsd.msssi.gob.es/profesionales/sistemasInformacion/sistemaInformacion/pdf/ESTUDES_2016_Informe.pdf.

Palmer, A., Jiménez, R., & Gervilla, E. (2011). Data mining: Machine learning and statistical techniques. In K. Funatsu, & K. Hasegawa (Eds.). *Knowledge-oriented applications in data mining* (pp. 373–396). Viena: InTech. Open Access Publisher.

Popovici, I., & French, M. T. (2013). Binge drinking and sleep problems among young adults. *Drug and Alcohol Dependence, 132*(1), 207–215.

R Core Team (2013). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. (ISBN 3-900051-07-0, URL) http://www.R-project.org.

Risher, M. L., Fleming, R. L., Risher, C., Miller, K. M., Klein, R., & Wills, T. (2015). Adolescent intermittent alcohol exposure: Persistence of structural and functional hippocampal abnormalities into adulthood. *Alcoholism, Clinical and Experimental Research, 39*(6), https://doi.org/10.1111/acer.12725.

Scaglione, N., Mallett, K., Turrisi, R., Reavy, R., Cleveland, M. J., & Ackerman, S. (2015). Who will experience the most alcohol problems in college? The roles of middle and high school drinking tendencies. *Alcoholism, Clinical and Experimental Research, 39*, 2039–2046.

Schellerman-Offermans, K., Kuntsche, E., & Knibbe, R. A. (2011). Associations between drinking motives and changes in adolescents alcohol consumption: A full cross-lagged panel study. *Addiction, 106*(7), 1270–1278.

Stock, C., Mcalaney, J., Pischke, C. R., Vriesacker, B., Van Hal, G., & Akvardar, Y. (2014). Student estimations of peer alcohol consumption: Links between the social norms approach and the health promoting university concept. *Scandinavian Journal of Public Health, 42*, 52–59.

Thrul, J., & Kuntsche, E. (2015). The impact of friends on young adults' drinking over the course of the evening-an event-level analysis: Impact of friends on young adults' drinking. *Addiction, 110*(4), https://doi.org/10.1111/add.12862.

Trucco, E. M., Colder, C. R., Wieczorek, W. F., Lengua, L., & Hawk, L. W. (2014). Early adolescent alcohol use in context: How neighborhoods, parents, and peers impact youth. *Development and Psychopathology, 26*(2), 1–12.

Van Tyne, K., Zamboanga, B. L., Ham, L. S., Olthuis, J. V., & Pole, N. (2012). Drinking motives as mediators of the associations between alcohol expectancies and risky drinking behaviours among high school students. *Cognitive Therapy and Research, 36*(6), 756–767.

Vinci, C., McVay, M. M., Copeland, A. L., & Carrigan, M. H. (2012). The relationship between depression level and smoking motives in college smokers. *Psychology of Addictive Behaviors, 26*(1), 162–165.

Willem, L., Bijttebier, P., Claes, L., & Uytterhaegen, A. (2012). Temperament and problematic alcohol use in adolescence: An examination of drinking motives as mediators. *Journal of Psychopatholy and Behavioral Assessment, 34*(2), 282–292.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques* (Third Edition). San Francisco, CA: Morgan Kaufmann Publishers.