Original Article

# Data Classification Technique for Assessing Drug Use in Adolescents in Secondary Education

**Samuel-Soma M. Ajibade[1*], Oluwadare Joshua Oyebode[2], Johnry P. Dayupay[3], Nathaniel G. Gido[4], Almighty C. Tabuena[5], Osias Kit T. Kilag[6]**

[1]Department of Computer Engineering, Istanbul Ticaret Universitesi, Istanbul, Turkey. Email: asamuel@ticaret.edu.tr

[2]Department of Civil and Environmental Engineering, Afe Babalola University, Ado Ekiti, Ekiti State, Nigeria. oyebodedare@yahoo.com

[3]Cebu Technological University, Moalboal, Cebu, Philippines. Email: johnrydayupay@gmail.com

[4]Dept of Education, Madridejos Community College, Madridejos Cebu, Philippines. Email: maritimeresearchunit@gmail.com

[5]Philippine Normal University, Manila, Philippines. Email: tabuena.ac@pnu.edu.ph

[6]PAU Excellencia Global Academy Foundation, Inc. Cebu, Philippines. Email: pice.edu.ph@gmail.com

## Abstract

The reasons why students abuse drugs are crucial information. Knowledge of the difficulties associated with drug use can be improved by employing data mining techniques, which have many advantages. The focus of this study is to examine the causes of drug abuse among Lagos's high school students using data mining methods. In February of 2021, a cross-sectional study was conducted. Four hundred teenagers and young adults were present. They were given a questionnaire to fill out about their drug use habits, the types of drugs they take, and why they take them. We found that 59.1% of students drank alcohol, 23.6 % smoked cigarettes, 15.4 % used cannabis, and 3.1% used cocaine. In addition, the performance of 5 classifiers is compared in terms of correctly classified instances (CCI), with all of them performing better than the simplest classifier (more frequent category: used drug/never used drugs) in terms of the percentage of correctly classified instances. KNN yielded the highest CCI across the board when various drugs were compared (alcohol: 82.40 percent, tobacco: 66.22 percent, cannabis: 91.16 percent, and cocaine: 94.24). Use motives obtained a higher classifier performance when it came to alcohol and tobacco use, but the opposite was true for cannabis and cocaine. Peer pressure and the community in which a teen lives are two major factors that we found to have a significant impact on that teen's drug use.

**Keywords:** Drug use, Data mining method, alcohol, tobacco, cannabis, cocaine, secondary education

**DOI: 10.47750/pnr.2022.13.S04.114**

## INTRODUCTION

Adolescent drug use continues to rise despite widespread awareness of the risks associated with it [1]. Underage drinking places teens in multiple dangers. It's been linked to "both psychological and physical problems" [2] and "future addiction and other issues related to drug use" [3] that can last a lifetime. "About 20 million European Union teenagers and young adults between the ages of 14 and 36 consumed illicit drugs like cocaine and cannabis in the previous year" [4]. This includes 17.2 million people between the ages of 15 and 24, and 2.3 million between the ages of 25 and 34. For example, among young adults, Spain has one of the highest rates of illegal drug use, with 17.1% of them having used cannabis and 3.0% having used cocaine in the past year; the most commonly used legal drugs in Spain are tobacco (40.8%) and alcohol (79.2%), respectively [5].

In addition, "an average of 21% of European teenagers have been involved in smoking, while 48% have drunk alcohol, and about 46% have been high on drugs at least once in the last 30 days," as reported by the European School Survey Project on Alcohol and Other Drugs, which collected data on substance usage among students aged 15 to 16. Additionally, "2% of the population has ever tried cocaine and 16% have tried cannabis in their lifetimes." [6]. Recent data from Spain's National Survey on Drug Use in Secondary School Students shows that 27.3% of students aged 14–18 have smoked tobacco at some point, while 67.3% have consumed alcohol in the last month, with binge drinking being a common occurrence among this group (32 percent in the previous 30 days). And "1% of Spanish students aged 14-18 have used cocaine in the past month," while 18.3% of Spanish student's overall report using cannabis. [7]. "When trying to make sense of student alcohol use, it is crucial to take into account drinking motives (which include normative beliefs, expectations, and social motives)" [8].

"Social motives are important here: teenagers believe that drinking at parties makes it more interesting and pleasurable, and that it facilitates them walking up to other people and expressing their feelings and insights" [9]. Many people turn to cannabis when they need a stress reliever [10]. Furthermore, very little study has been conducted on why people smoke [11]. Almost all of the cited research relied on conventional statistical methods. However, data mining techniques, which allow researchers to uncover novel ways to evaluate and visualize data, are rarely used in studies [8]. "interesting, unexpected, or valuable structures in large databases" [12] are what data miners seek out when they set out on their quest. D The use of data mining techniques expands the available options for studying data and displaying the findings.

Numerous studies in Europe, the United States, and elsewhere have looked into the common causes of adolescent drug use, but no such research has been done in Nigeria. To better predict who will use drugs and who won't, we'll employ data mining methods to assess the significance of various factors related to drug use and the nature of the relationships among them. We anticipate that factors like socialization and hedonism will play significant roles. In general, there are two things you want to achieve. First, we'll look at the effects of five data mining methods that are rarely used in the context of drug use, and see how well they can classify individuals as drug users or non-users based on their knowledge of drug use motives. Next, we determine whether or not there are differences between the drugs when analyzing the most common reasons why high school students abuse them. Both of these objectives are geared toward understanding the underlying causes of drug use among adolescents.

## Research Methodology

### Study design

A cross-sectional research was carried out to investigate the common reasons for drug use among secondary school students and adolescents.

### Study setting and population

The largest country in Africa, Nigeria, can be found in West Africa. Nigeria shares borders with Niger, Chad, Cameroon, and Benin, starting in the north. About 216 million people call its 923,769 square kilometers of land their home (356,669 square miles). There are six distinct geopolitical regions in Nigeria: the North Central, North East, North West, South East, South South, and South West. Specifically, the city of Lagos, which is in the South West region of Nigeria, was chosen for this research. As of 2018, Lagos was the most populous metropolis in Nigeria and the second most populous city in Africa, with a population of 23.5 million. The pilot phase of this study ran from January 11-15, 2021, and data collection took place from February 1st to April 15th, 2021, at a number of secondary schools in Lagos. Fifteen secondary schools were selected through a cluster-based random sampling strategy. There are a total of 15 institutions included here, 10 public and 5 private. Students and adolescents ranging in age from 13 to 19 years old, from classes in Junior Secondary Schools (JSS) all the way up to classes in Senior Secondary Schools (SSS) make up the 400 participants (SSS). Anyone enrolled in one of the participating secondary schools was welcome to participate.

### Data sources and management

Students responded anonymously to a survey about their use of addictive substances and the factors that led them to try them. Student demographics such as gender, age, socioeconomic status, parental status (both parents/single parent), family size, and household income were collected. The data was gathered via a questionnaire. The researchers gathered information by giving the participants an online version of a pre-designed, structured questionnaire. Students were given tablets by the data collection team with which to fill out the survey. The information was collected, entered into a database, and then processed for analysis. Fifteen secondary schools were selected through a cluster random sampling procedure. Four hundred students, ages 13 to 19, participated in the research. The final sample included 395 students (55.6% male and 44.4% female) with an average age of 15.93 after eliminating students who provided unreliable responses. Artificial Neural Networks (ANN), k-Nearest Neighbors (K-NN), Decision Trees (DT), Nave Bayes (NB), and Logistic Regression were just some of the data mining classification techniques used to assess drug users' motivations (LogR). The effect size (Phi:) of a given motivation is calculated by comparing the proportion of people who engage in that motivation with the proportion of people who do not; if the proportion is larger, the prediction model is more likely to choose that motivation. With the help of DT, classification algorithms can easily incorporate differences in a response variable. This is because DT breaks down data into sequences. KNN uses KNN observations rather than assumptions about the underlying structure of a component that connects the variables (dependent and independent) in a study into categories. NB uses Bayes' theorem to assess the likelihood that an event fits into one of several categories. 'Nearby' locations were determined using Euclidean distance. These methods were implemented using Weka (Waikato Environment for Knowledge Analysis). [13].

## Experimental Results

57% of college students use alcohol, 23% tobacco smoking, 15% use cannabis, and 3% use cocaine. Table 1 provides a breakdown of the reasons people use addictive substances. Drugs that have been previously used and those that haven't been used are distinguished by their yellow tint. "Pleasurable activity" is a central motive (bold values) in all of the compared drugs, with the greatest value difference in alcohol users (Φ = 0.497): 79.3% of alcohol users ended up selecting this reason, compared to 28.4% of never used drugs. "Relaxing" is a central motive as well, with the greatest differences among alcohol (Φ = 0.365) and tobacco users (Φ = 0.271). "To get over issues" is another common motivation, which has largest differences among alcohol (Φ = 0.386) and tobacco users (Φ = 0.266). However, 42.3 percent of cocaine users (in front of 19.7 percent of never-used drugs, Φ = 0.220) chose "they are not so harmful" as their key motivation. Whereas the discriminant differences seem to be because of the greater preference of this motive in never used drug (85 percent), in front of cannabis (31.2 percent, Φ = 0.422), tobacco (39.6 percent, Φ = 0.438), alcohol (52.6 percent, Φ = 0.480), and cocaine (23.1 percent, Φ = 0.254) users, "friends take drugs" does have the biggest discrepancies in all the compared drugs. A similar pattern of inconsistencies is evident in the "Addiction" reason, which is selected by 60.8% of never-used substances in compared to cannabis (37.6%, Φ = 0.299), alcohol (34.5 percent, Φ = 0.252), cigarette (30.6 percent, Φ = 0.212) and cocaine (27.1 percent, Φ = 0.181) users.

**Table 1: Motivation for using addictive substances by drug type**

| Never used drugs: n=98 (24.7%) Alcohol use: n=229 (57.9%) Tobacco use: n=93 (23.6%) Cannabis use: n=61 (15.4%) Cocaine use: n=12 (3.1%) | Never drug use | Use of Alcohol | | Use of Tobacco | | Use of Cannabis | | Use of Cocaine | |
|---|---|---|---|---|---|---|---|---|---|
| | % | % | Φ | % | Φ | % | Φ | % | Φ |
| 1. Improving relations | 56.4 | 39.2 | 0.149 | 32.1 | 0.255 | 25.4 | 0.234 | 28.6 | 0.144 |
| 2.To get over issues | 64.6 | 72.3 | 0.386 | 67.5 | 0.266 | 54.1 | 0.167 | 62.0 | 0.101 |
| 3.Pleasurable activity | 28.4 | 79.3 | 0.497 | 58.8 | 0.299 | 64.4 | 0.444 | 69.5 | 0.295 |
| 4.Better with yourself | 32.6 | 19.8 | 0.086 | 25.5 | 0.065 | 22.8 | 0.014 | 36.7 | 0.087 |
| 5.To intensify dance and music | 39.8 | 46.1 | 0.069 | 35.6 | 0.087 | 41.2 | 0.100 | 62.5 | 0.164 |
| 6.Improve sexual relations | 16.7 | 19.4 | 0.077 | 12.6 | 0.073 | 15.4 | 0.045 | 26.7 | 0.069 |
| 7. Last longer | 48.9 | 30.8 | 0.163 | 29.3 | 0.195 | 24.6 | 0.229 | 52.7 | 0.102 |
| 8.To lose inhibition | 20.4 | 25.6 | 0.065 | 22.8 | 0.040 | 33.1 | 0.043 | 33.5 | 0.036 |
| 9.Friends take drugs | 85.0 | 52.6 | 0.480 | 39.6 | 0.438 | 31.2 | 0.422 | 23.1 | 0.254 |
| 10.Addiction | 60.8 | 34.5 | 0.252 | 30.6 | 0.212 | 37.6 | 0.299 | 27.1 | 0.181 |
| 11.New feelings | 61.6 | 65.6 | 0.045 | 69.4 | 0.020 | 68.7 | 0.027 | 74.6 | 0.028 |
| 12.Against established | 38.9 | 29.7 | 0.101 | 42.5 | 0.109 | 41.4 | 0.122 | 48.5 | 0.030 |
| 13.They are not so dangerous | 19.7 | 28.6 | 0.047 | 33.7 | 0.124 | 38.3 | 0.168 | 42.3 | 0.202 |
| 14.Relaxing | 26.7 | 78.6 | 0.365 | 76.6 | 0.324 | 71.2 | 0.360 | 58.4 | 0.197 |
| 15.Creativity | 24.7 | 20.4 | 0.059 | 14.8 | 0.012 | 25.5 | 0.025 | 31.6 | 0.061 |

Having a balance of consumers and non-consumers in our sample sizes was critical to our ability to forecast drug use and abstinence in teenagers (teenagers who had never used the substance). A cross-validation approach known as the k-fold cross-validation was used to generate the models and evaluate their predictive potential. Randomly dividing the initial sample into k equal-sized sub-samples is done. Validation of the model is carried out using a sample from the k sample population; the remaining k-1 data samples were utilized to train the model. Each of the k samples is used exactly once in the cross-validation process, which is repeated k times. For each substance, we generated 100 models using k = 10 folds of cross-validation, with each k-fold cross-validation run ten times.

## Outcome Data

Table 1 also emphasizes (bold values) the most common reasons for using/never using drugs. The majority of drug users stated that they take drugs to get over issues, discover new feelings, or simply because it is pleasant. Except in the case of alcoholics, another common reason is to relax. Cocaine users also use it to make dance and music more intense (62.5 percent) and to make it last longer (52.7 percent). The majority of never drug users said that others use drugs due to friends that take drugs (85%), to get over issues (64.6%), to perceive new feelings (61.6%); being addicted (60.8%) and improving relationships (56.4%) are also common reasons.

The motivation that teenagers give to use addictive drugs broken down by substance (for teenagers that take only alcohol, tobacco or cannabis) is displayed in Table 2. Table 2 shows that the most common motives (bold values) for teenagers/students who take alcohol only are "to get over issues", "to experience new feelings", and "because their friends also drink". These are also the most common motives given by teenagers who smoke who claim to use drugs. It's also worth noting that 54.1 percent of only the alcohol consumers claim that they usually take drugs in order to improve their relationships, and that 64.3 percent of teenage smokers say addiction (64.3 percent) and relaxation (56.2 percent) are other common motivations. The most common reason given by cannabis users is to have some relaxation, and another commonly noted reason is that it is used to get over issues, as it is a pleasurable activity, or to discover new feelings. 53.6 percent of cannabis users believe the drug is not as harmful as it appears. We used data mining measures to assess the predictive ability of drug use motives. We used Decision Trees (DT), K-Nearest Neighbors (KNN) and Nave Bayes (NB) to solve our problems.

**Table 2: Motivation Teenagers give to use addictive drugs by substance used for a single substance user**

| Never used drugs: n=98 (24.7%) Alcohol use: n=148 (37.4%) Tobacco use: n=27 (6.7%) Cannabis use: n=15 (2.8%) | Never drug use | Use of Alcohol | | Use of Tobacco | | Use of Cannabis | |
|---|---|---|---|---|---|---|---|
| | % | % | Φ | % | Φ | % | Φ |
| 1. Improving relations | 62.9 | 54.1 | 0.089 | 49.2 | 0.081 | 34.2 | 0.105 |
| 2.To get over issues | 79.5 | 65.2 | 0.157 | 71.1 | 0.057 | 53.4 | 0.110 |
| 3.Pleasurable activity | 32.6 | 41.5 | 0.112 | 41.3 | 0.064 | 57.3 | 0.108 |
| 4.Better with yourself | 34.7 | 24.8 | 0.139 | 27.6 | 0.054 | 22.6 | 0.057 |
| 5.To intensify dance and music | 42.1 | 52.6 | 0.129 | 30.7 | 0.074 | 22.6 | 0.080 |
| 6.Improve sexual relations | 26.9 | 19.7 | 0.123 | 6.9 | 0.082 | 22.6 | 0.031 |
| 7. Last longer | 57.1 | 43.5 | 0.151 | 33.1 | 0.131 | 22.6 | 0.122 |
| 8.To lose inhibition | 30.0 | 36.6 | 0.084 | 30.7 | 0.012 | 24.6 | 0.035 |
| 9.Friends take drugs | 87.1 | 59.8 | 0.291 | 66.5 | 0.131 | 37.0 | 0.194 |
| 10.Addiction | 64.0 | 38.7 | 0.279 | 64.3 | 0.082 | 30.3 | 0.124 |
| 11.New feelings | 71.3 | 65.4 | 0.071 | 68.0 | 0.023 | 51.5 | 0.077 |
| 12.Against established | 38.8 | 30.6 | 0.107 | 31.3 | 0.053 | 30.3 | 0.042 |
| 13.They are not so dangerous | 25.6 | 23.8 | 0.036 | 22.7 | 0.032 | 53.6 | 0.088 |
| 14.Relaxing | 36.7 | 30.8 | 0.083 | 56.2 | 0.118 | 67.2 | 0.122 |
| 15.Creativity | 24.7 | 18.5 | 0.123 | 16.0 | 0.066 | 25.7 | 0.016 |

## Main Results

In Table 3, we can see the model's results for each substance, including the mean and standard deviation (SD) of correctly classified cases, as well as the time required to train 100 models. The classifiers were evaluated against ZeroR, the simplest classifier available, which predicts the mode (the most common value for the classification variable). While other classifiers take into account information from potential predictors, the ZeroR classifier uses only the mode to measure the projected category for a given instance.

## Alcohol

In the aspects of correctly classified instances, the compared data mining methods outperform ZeroR models significantly as seen in Table 3. The highest percentage of correctly classified instances belongs to KNN (M = 82.40 percent, SD = 3.22), while the lowest percentage belongs to DT (M = 81.36 percent, SD = 3.55). DT, on the other hand, has the largest training elapsed time-seconds (ET), with a mean of 16.75 s and SD = 3.45) for estimating a model, while the ET is lower than 1 sec in other techniques.

### Tobacco

The data mining approaches perform closely similar to ZeroR models in terms of correctly classified instances. As shown in Table 3, KNN has the best performance in classifying teenagers (M = 86.22 percent correctly classified instances, SD = 3.26) and DT has the worst performance (M = 84.54 percent, SD = 2.98) DT, on the other hand, has the longest training elapsed time-seconds (ET), with a mean of 17.02s and SD = 2.69) for estimated model, while other methods have ET of less than 1 second.

### Cannabis

Data mining approaches perform similarly to ZeroR models in terms of correctly classified instances. KNN has the best classification performance (M = 91.16 percent correct classifications, SD = 2.96) and NB has the worst (M = 90.31 percent, SD = 3.20). DT, on the other hand, has the highest training elapsed time-seconds (ET), with a mean of (13.42 s and SD = 2.98) for estimating a model, whereas other techniques have ETs of less than 1 second.

### Cocaine

In aspects of correctly classified instances, the compared data mining methods outperform ZeroR models significantly as seen in Table 3. The highest percentage of correctly classified instances belongs to NB (M = 94.24 percent, SD = 8.22), while the lowest percentage belongs to DT (M = 88.89 percent, SD = 8.68). DT, on the other hand, has the largest training elapsed time-seconds (ET), with a mean of 3.17s and SD = 1.36) for estimating a model, while the ET is lower than 1 sec in other techniques.

**Table 3: Performance of data mining classification techniques compared to a ZeroR classifier**

**Table 3a: Mean & SD from 100 models for Correctly Classified Instances (%) & Training elapsed time-seconds (ET)**

| Classifiers | ZeroR | DT | KNN | NB |
|---|---|---|---|---|
| Alcohol | 60.86 (0.09) | 81.36 (3.55) | 82.40(3.22) | 81.92(3.40) |
| Tobacco | 60.86 (0.09) | 84.54(2.98) | 86.22(3.26) | 85.23(3.24) |
| Cannabis | 59.92 (0.20) | 90.61(2.18) | 91.16(2.96) | 90.31(3.20) |
| Cocaine | 47.65 (0.24) | 90.29(8.68) | 91.58(9.14) | 94.24(8.22) |

**Table 3b: Correctly Classified Instances (%) & Training elapsed time-seconds (ET): Mean, SD from 100 models**

| Classifiers | ET | ET | ET | ET |
|---|---|---|---|---|
| Alcohol | 0.01 (0.02) | 16.75(3.45) | 0.01(0.02) | 0.00(0.00) |
| Tobacco | 0.00 (0.00) | 17.02(2.69) | 0.03(0.01) | 0.23(0.06) |
| Cannabis | 0.01 (0.02) | 13.42(2.98) | 0.05(0.04) | 0.20(0.05) |
| Cocaine | 0.01 (0.02) | 3.17(1.36) | 0.05(0.04) | 0.39(0.12) |

## Discussion

The authors of this paper set out to answer the questions, "What are the most common reasons why secondary school students take drugs? Do these reasons vary depending on the substance used?" and "What are the differences between contemporary and traditional data mining approaches that are rarely utilized in scenario related to use of drug?".

With an eye toward uncovering underlying patterns, we compared three different classification strategies for identifying subgroups, with a focus on comparing drug users and nonusers amongst adolescents. Classification accuracy was highest when using these motivations of use to predict the use of cannabis and cocaine, while it was lowest when applying these motivations of use to predict the use of alcohol and tobacco. Compared to other methods, DT's lengthy model training process is a result of its high computational demand.

What then, would you say is the best way to categorize data? Unfortunately, it does not appear that there is a simple answer that can tell us the method or algorithm to use to get the best classification model before we begin data analysis. In this regard, Nisbet, Elder, and Miner (2009) argue that the technique with the best performance used for the classification of one set of data may not be useful for other datasets if multiple classification methods are employed. Therefore, it appears that using a combination of algorithms is the best course of action, as different approaches or algorithms perform better on different datasets.

In terms of the secondary objective, we anticipated that metrics related to a desire to help others and spread joy would be the most telling. Our research backs up our hypothesis that the most common reasons people experiment with drugs are to try something new, cope with stress, socialize with others, and indulge in a pleasurable activity. It's important to note that drug users and people who have never tried drugs chose different motivations for using; people who have never tried drugs believe that drug users consume due to peer pressure and the impacts of drug abuse. Contrarily, people who use drugs don't make the same tradeoffs between these two factors. Several authors, including Anderson, Grunwald, Bekman, Brown, and Grant (2011), have argued that it is worthwhile to investigate the reasons why some people choose not to purchase consumer goods. In this article, we present data on the subject, drawing parallels between drug users and those who don't partake in this activity.

## Limitations

Fifteen secondary schools in the state of Lagos are participating in this study. This precludes generalizing the findings to the other states in Nigeria. However, the findings of this study are useful because they reveal data on the prevalence of drug use amongst young students and adolescents in Lagos state, data that can be extrapolated to situations with similar characteristics. In addition, the COVID-19 pandemic occurred while this study was being conducted, which may have made it more challenging to obtain a representative sample. The response rate that was received, on the other hand, is adequate for descriptive study.

## Conclusion

People who don't engage in drug use often assume that young people turn to drugs as a means of coping with difficult emotions. Half of alcoholics believe that drug use is driven by curiosity about the human experience, the desire to find solace from problems, and peer pressure. Young drug users think that people use substances to experiment with their emotions, cope with problems, unwind, and have fun. Many people who partake in drug use like cannabis and cocaine do so because they find it to be a pleasurable way to unwind and explore their emotions. In order to assess the pros and cons of drug use, data mining techniques are useful.

### Competing interests

The authors declare no competing interest.

## References

1. Lee Jung Yeon, Judith Brook, Wonkuk Kim. "Triple trajectories of alcohol use, tobacco use, and depressive symptoms as predictors of cannabis use disorders among urban adults." Psychology of Addictive Behaviors. 2018; 32(4): 466.
2. David Walker, Cynthia Kuhn, Leon Risher. The effects of peri-adolescent alcohol use on the developing hippocampus. In International review of neurobiology. 2021; 160: 251-280. Academic Press.
3. Hazart Juliette, Marie Blanquet, Anne Debost-Legrand, Anne Perreve, Stéphanie Léger, Vincent Martoia, Sylvie Maurice, Georges Brousse, Laurent Gerbaud. "A screening focusing on aftereffects of alcohol consumption in a student population. A National cross-sectional survey." Journal of preventive medicine and hygiene. 2018; 59(1): E48.
4. European Monitoring Centre for Drugs and Drug Addition. European drug report: Trends and developments. 2018a. Retrieved from http://www.emcdda.europa.eu/system/ files/publications/8585/20181816_TDAT18001ENN_PDF.pdf
5. European Monitoring Centre for Drugs and Drug Addition. Spain: Country Drug Report 2018. Retrieved from http://www.emcdda.europa.eu/countries/drug reports/2018/spain_en.
6. Kraus, Ludwig, Alojz Nociar. ESPAD report: results from the European school survey project on alcohol and other drugs. European Monitoring Centre for Drugs and Drug Addiction, 2016.
7. Brime, García, León, Llorens, López, Molina, Sánchez. Observatorio Español de las Drogas y las Adicciones (2021).
8. Jiménez Rafael, Joella Anupol, Berta Cajal, Elena Gervilla. "Data mining techniques for drug use research." Addictive Behaviors Reports 8. 2018: 128-135.
9. Ajibade, S. S. M., Dayupay, J., Ngo-Hoang, D. L., Oyebode, O. J., & Sasan, J. M. (2022). Utilization of Ensemble Techniques for Prediction of the Academic Performance of Students. Journal of Optoelectronics Laser, 41(6), 48-54.
10. Smit Koen, Carmen Voogt, Roy Otten, Marloes Kleinjan, Emmanuel Kuntsche. "Why adolescents engage in early alcohol use: A study of drinking motives." Experimental and clinical psychopharmacology (2020).

11. Genrich Gregor, Céline Zeller, Hans Jörg Znoj. "Interactions of protective behavioral strategies and cannabis use motives: An online survey among past-month users." PloS one. 2021; 16 (3): e0247387.
12. Gazibara Tatjana, Marija Milic, Milan Parlic, Jasmina Stevanovic, Nebojsa Mitic, Gorica Maric, Darija Kisic Tepavcevic, Tatjana Pekmezovic. "What differs former, light and heavy smokers? Evidence from a post-conflict setting." African Health Sciences. 2021; 21(1): 112-22.
13. Cody Slattery. Data Mining/Data Privacy and the Collection/Misuse of Our Private Data. 2021; No. 5823.
14. Wahab Lukuman, Haobin Jiang. "A comparative study on machine learning based algorithms for prediction of motorcycle crash severity." PLoS one. 2019; 14(4): e0214966.
15. Ajibade, S. S. M., Ahmad, N. B., & Shamsuddin, S. M. (2018, December). A data mining approach to predict academic performance of students using ensemble techniques. In International Conference on Intelligent Systems Design and Applications (pp. 749-760). Springer, Cham.