Eric Max Liu,
The Webb School of California,
Claremont, USA
E-mail: ericliu@Ibb.org

PREDICTING ILLICIT DRUG USE WITH ARTIFICIAL NEURAL NETWORK

Abstract. The purpose of this study is to find the significant predictive factors for illegal drug use using an artificial neural network. Here, illegal drugs are defined as heroin, ecstasy, crack, or cocaine. For this study, I used a public database with researcher permission that recorded more than 1885 respondents on illegal drug use and their characteristics. The respondents were separated into a training sample and testing sample. Using the training sample, I built a neural network model using programming language R in RStudio. Finally, I tested and refined the model using the testing sample and found the ROC (Receiver Operating Characteristic). Of the 1885 respondents, 1043 (55%) used illegal drugs. The artificial neural network returned that the most significant predictors were extraversion, age, education, openness, and impulsiveness (From most important to least important). Extraversion, age, and education had a negative correlation with illicit drug use while openness and impulsiveness had positive correlations. The ROC is 0.86 in the training sample and 0.77 in the testing sample. I hope that the results of this research paper will help prevent individuals who are likely to be drug users from becoming one. Educators and researchers can use this profiling framework to identify people with high risk of using illegal drugs, and find ways to help them to reduce such a risk. For example, support groups can be set up and actively reach out to the high risk people.

Keywords: Illicit Drugs, artificial neural network, factors.

1. Introduction

Illicit drugs are drugs that are illegal to consume or the abuse of medically prescribed drugs (i.e., use of a medically prescribed drugs outside of its intended purpose). Illicit drug are typically divided into several categories: stimulants, opioids, hallucinogens, and depressants/sedatives [11]. According to a national survey done by the National Institute of Drug Abuse in 2017, 49.5% of people aged 12 years and older have used illicit drugs at least once in their lifetime and 11.2% have used it in the past month [6]. Also in that year, more than 71.000 overdose deaths have been reported in the United States in the International Drug Policy Consortium report [5]. While illicit drug consumption is more popular in some places than others, people of all classes are susceptible to illicit drug use. Most notably, numerous celebrities have reported the use of illicit drugs. The use of illicit drugs can lead to addiction and permanent changes to the brain [10; 11]. Individuals who are addicted have intense cravings for drugs and even when treated they are very likely to go into relapse [11]. Illicit drugs have also prompted numerous political debates over whether certain drugs should be made legal [12]. Therefore, the issue of illicit drug use is without doubt an important one. This paper explores what the predictive factors are for illicit drug use using an artificial neural network.

1.2 Literature Review

Past studies of illicit drug use have focused on the effects of drug use and predictors of drug use in adolescents. For example, a study done by David, Laurie, and Kevin examined the dependence on drugs and alcohol of adolescents as they grow up [3]. They studied four different groups of drug users: light, moderate, and heavy drug and alcohol users, and examined the effect of familial alcoholism and personality on this dependence. They found that the group who have had heavy drug and alcohol use in adolescence have a significantly higher chance of drug or alcohol dependence in later years. Furthermore, the group with heavy drug use was more associated with familial alcoholism, negative emotionality, and low constraint. Bertrand et al have studied how socioeconomic factors affect illicit drug use among young adults [7]. Socioeconomic factors among young adults include occupation, education, employment stability, and unemployment. They has found that young adults with a low socioeconomic status have a significantly higher rates of illicit drug abuse. Another study done by David, Judith, and Chenshu examined the relationship between drug use and risk of major depressive disorder, alcohol dependence, and substance use disorders [1]. The results showed that early use of tobacco can lead to higher risk of alcohol dependence and substance use disorders and early use of alcohol and other illicit drugs can lead to extremely high risks of depressive disorders, alcohol dependence, and substance use disorders. Another study done by Fehrman et al explores the personal traits and characteristics that affect drug consumption risk [13]. In their study, they studied an individual's risk of drug consumption using models using numerous predictive models. The data they used was collected through an online survey created in England. For each type of drug, they found the model that best fit to measure its correlation with personal traits. They also found the best predictors for illicit drug use with each type of drug use. A sensitivity of over 70% was achieved for most of the drugs when using the best model. Their results have shown that the most successful model overall was the decision tree, and linear discriminant analysis and k-nearest neighbors are also successful for several of the drugs.

1.3 Objective and Hypothesis

This study aims to find factors that help predict illicit drug use using an artificial neural network.

Unlike many past studies, this study sample covers age groups beyond adolescence and explores a wide range of personal characteristics besides socioeconomic or environmental factors. For example, the five big personality traits (neuroticism, openness, extraversion, agreeableness, and conscientiousness), gender, ethnicity, education, country of residence, ImpSS, and alcoholism and are the independent variables. Furthermore, this study uses an artificial neural network, a model not yet tested to find the best predictors for illicit drug use in past studies. My study also looks at illicit drugs as a whole instead of individual drugs which could be more helpful when considering that all illicit drugs are banned due to their negative effects. Finding the risk of illicit drug consumption as a whole is more accurate than calculating the risk based off each individual drug.

I hypothesize personal traits can have a significant impact on whether an individual will use drugs.

2. Data and Methodology

2.1 Data

In this study, I used the data that Elaine Fehrman and her research team collected via a survey using the tool Survey Gizmo, over the course of a year between March 2011 and March 2012, in England [6]. In their survey, illicit drugs are defined as cocaine, crack, heroine, and ecstasy.

The database contains records for 1885 respondents. For each respondent, 12 attributes are recorded: Personality measurements which include NEO-FFI-R (neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness), BIS-11 (impulsivity), ImpSS (sensation seeking), level of education, age, gender, country of residence and ethnicity.

Participants were asked about their use of 18 categories of legal (alcohol, amphetamines, amyl nitrite, benzodiazepine, cannabis, chocolate, caffeine, ketamine, legal highs, LSD, methadone, mushrooms, nicotine and volatile substance abuse and one fictitious drug (Semeron) which was introduced to identify over-claimers) and illegal drugs (cocaine, crack, ecstasy, heroin,). For each drug, they were asked to selectone of the answers:

"Never Used"
"Used over a Decade Ago"
"Used in Last Decade"
"Used in Last Year"
"Used in Last Month"
"Used in Last Week"
"Used in Last Day" [4].

In this study, the outcome of illicit drug use is defined as if a person selected any answer than "Never Used" to any of the four illegal drugs. It is a binary variable with 1=ever used illicit drug use and 0=no use of illicit drug.

All the 12 attributes of participants are included in the model to test if they predict illicit drug use.

2.2 Methodology

An artificial neural network was created using the open source software R Studio. The artificial neural network is modeled after a biological neural network (a brain), though the human brain is much more sophisticated. An artificial neural network consists of input units, where it takes information from the data it's given, and output units, where the computer decides what the information it takes in means. Between them, there are numerous hidden units, which helps transform the data in the input units into data the output units can use. Hidden units can compose several layers depending on the complexity of the problem, and each level further processes the data in the last one. In this particular study, the input units are the predictive factors and the output unit is whether the individual uses illicit drugs. There is only one hidden layer. Lines are drawn between these units representing the Iight/ importance of each unit. In this study, black lines represent a positive correlation while grey lines represent negative correlation. When data is fed into the neural network, the network uses a process called backpropagation. The input values are fed into the network, and the network will come to a result based on what it has. The result it comes up with is then compared with the actual result and the network will adjust its weight values accordingly. As more data is fed in, the predictive model becomes more and more likely to predict the correct outcome based on the input values.

The data was split into two random samples: a training sample to build the NN model, and a test sample to test/validate the model.

All of this is done on RStudio with the package neuralnet and Neural Net Tools.

3. Results

Of the 1885 individuals in the survey, 1043 (55%) of them have used one of four illicit drugs: heroine, ecstasy, crack, and cocaine.

I first built a correlogram, a matrix of correlations between different variables (a graphical representation), to see if there are any correlations between the different predictive factors. In the graph, blue represents a positive correlation and red represents a negative correlation. The darker the hue of the colors, the greater the correlation between the two factors.

According to the correlogram, Impulse and ImpSS (Impulsive Sensation Seeking), Openness and ImpSS, Age and Country, Extraversion and Consciousness, and Country and NoDrugUse have a strong positive correlation while Neuroticism and Extraversion has a strong negative correlation.

I then proceeded to feed the data into an artificial neural network. The lines between the different variables represent the strength of the factor and whether it has a negative or positive correlation (The thicker the line, the more significant the variable is. Black lines represent a positive correlation and grey lines represent negative correlation)

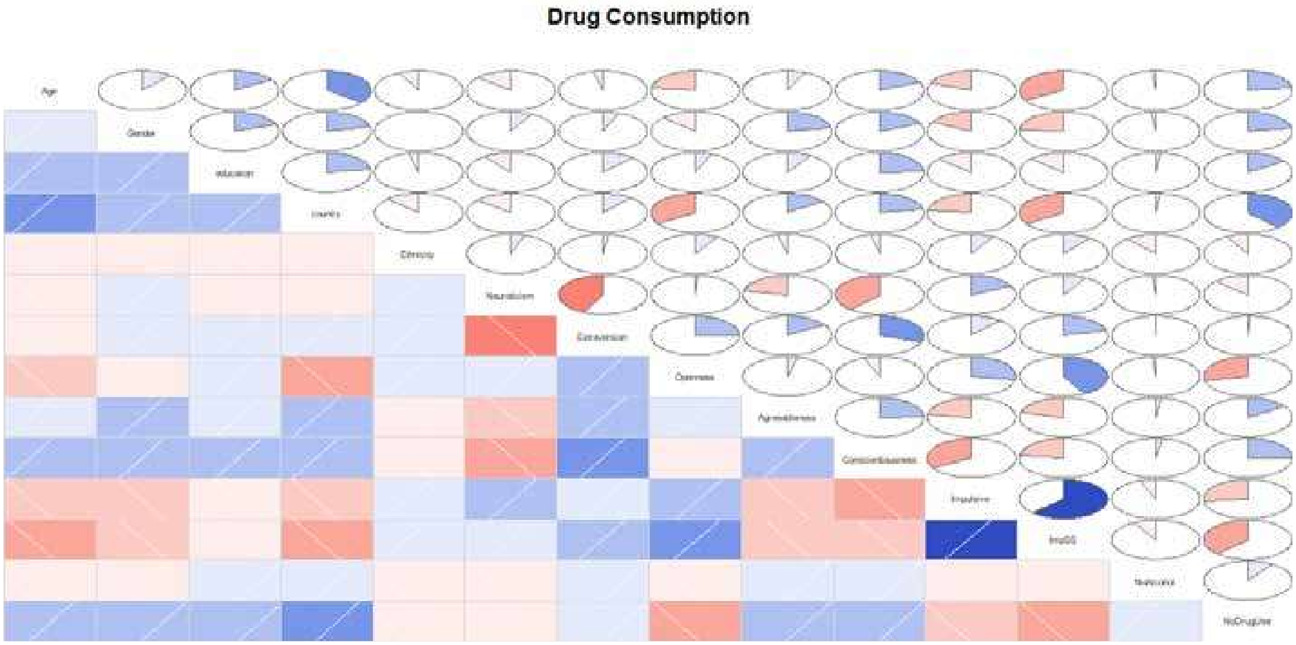

Figure 1. Correlogram of the Predictive Factors

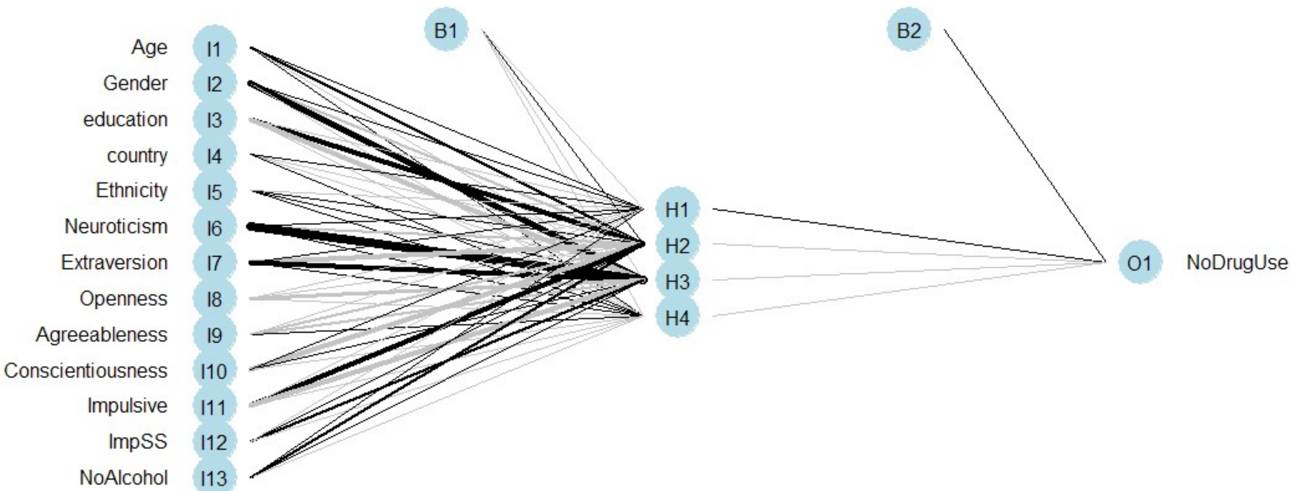

Figure 2: Artificial Neural Network (Training Sample)

As the lines seem like an incomprehensible mess, the strength of each predictor may be difficult to see so I put the data into a bar chart. H1 is the most significant of the hidden units, so only the lines connected to H1 are important. The most significant predictors of illicit drug use according to the artificial neural network are extraversion, age, education, openness, and impulsiveness. To be more specific, an introverted, young, uneducated, open, and impulsive person is more likely to use illicit drugs.

For the training sample, I found that the Receiver Operating Characteristic (ROC) 0.86 for the artificial neural network. For the testing sample, the Receiver Operating Characteristic 0.77 for the artificial neural network.

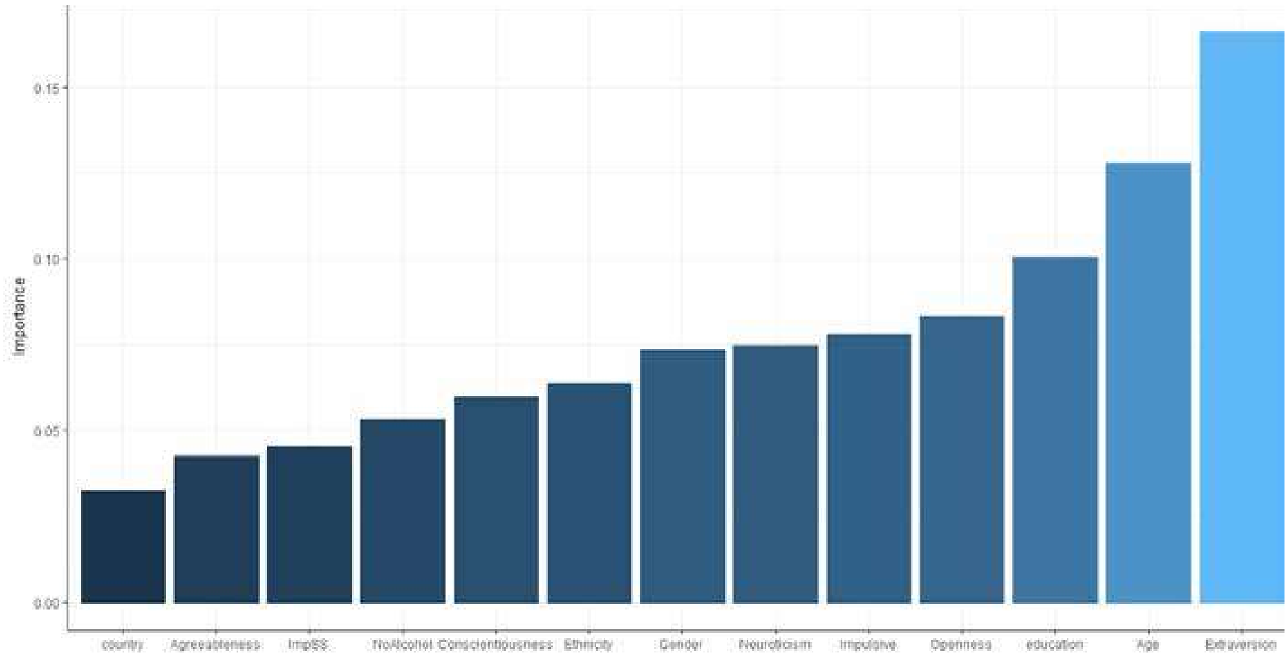

Figure 3. Bar Chart of Most Significant Predictors of Illicit Drug Use

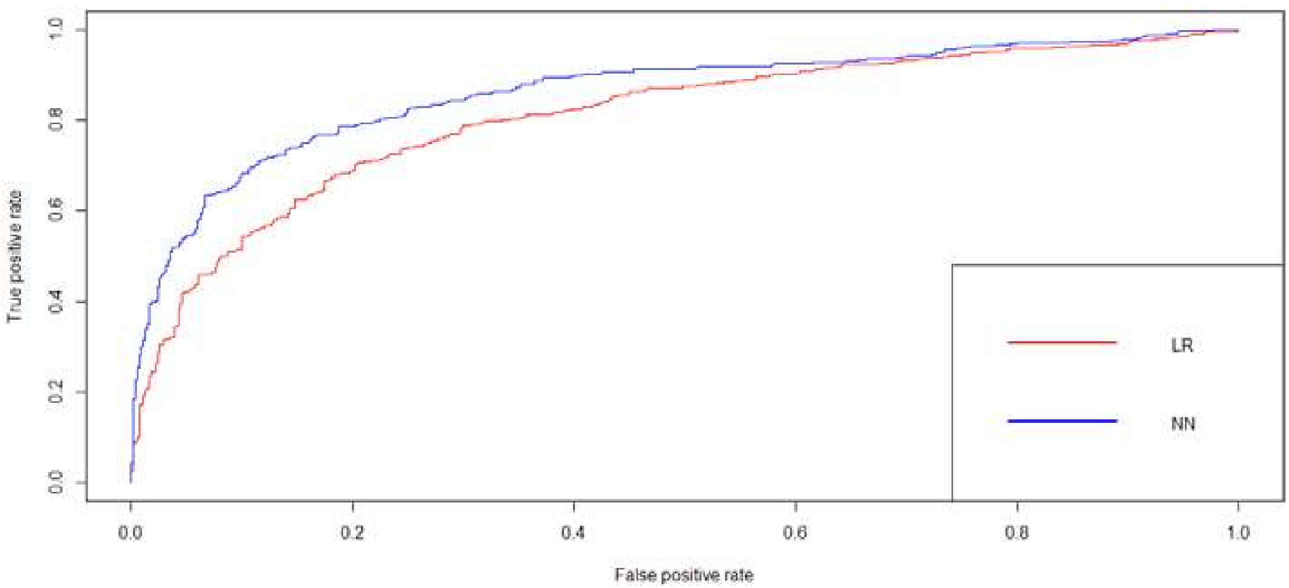

Figure 4. Receiver Operating Characteristic in Training Sample

4. Discussion

In this study, we found that the significant predictors the artificial neural network returns are extraversion, age, education, openness, and impulsiveness. The receiver operating characteristic for the artificial neural network in the training and testing samples were 0.86and 0.77, respectively, which suggest that the model is a good fit of the data. As such, this model is considered as successful at predicting whether an individual uses illicit drugs.

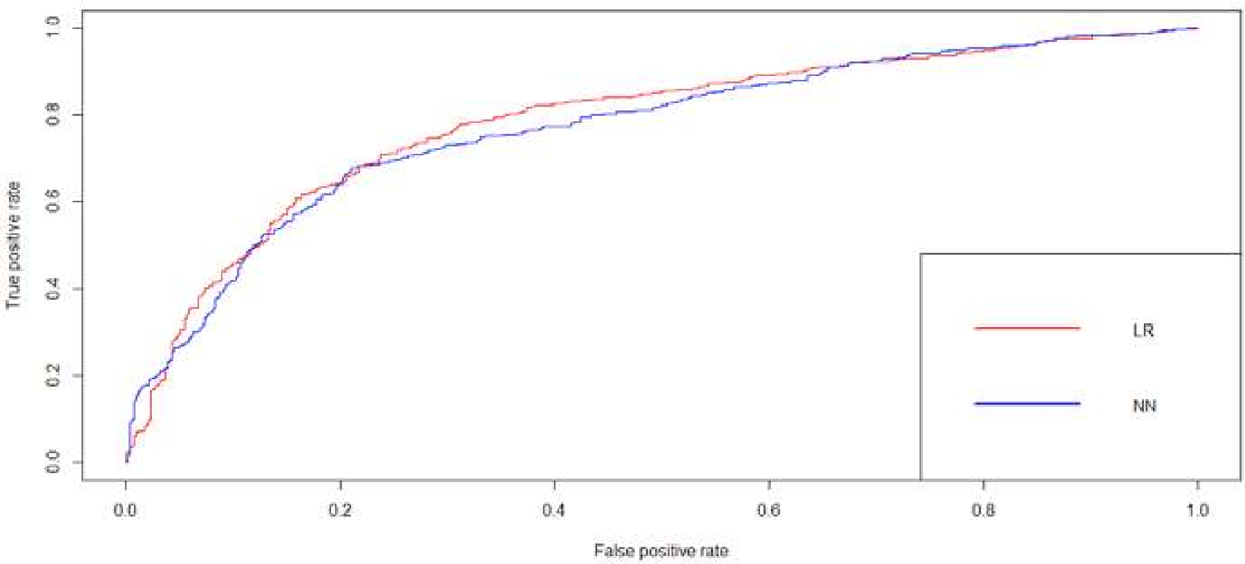

Figure 5. Receiver Operating Characteristic in Testing Sample

The knowledge of which factors are related with higher likelihood of illicit drug use can be used to better identify the high risk population, and to provide any intervention and/or treatment. For example, as extraversion is found to be an important predictor for illicit drug use according to the model, one strategy is to make extroverted individuals more aware of the harmfulness of illicit drug use. Perhaps it's because extroverted individuals are more exposed to peer pressure, and so they are more likely to use illicit drugs if their friends already do. Meanwhile, education level is found to be another important predictor. Therefore, more health education on the risk of illicit drug can be provided to people with lower levels of education.

Overall, the findings from this study is consistent with previous research. For example, the following is a table of several risk and protective factors in five domains during adolescence from the NIH website (NIH).

Table 1.

Risk Factors	Domain	Protective Factors
Early Aggressive Behavior	Individual	Self-Control
Lack of Parental Supervision	Family	Parental Monitoring
Substance Abuse	Peer	Academic Competence
Drug Availability	School	Anti-drug Use Policies
Poverty	Community	Strong Neighborhood Attachment

Although the generalizability (external validity) of the model is not tested, a split-sample validation has been conducted and showed that this model is a good fit. Other factors, such as income and familial background, can also help predict illicit drug use. However, information on these factors was not collected in the data. In future studies, more factors can be included in order to further improve the prediction of illicit drug use.

References:

1. Brook D. W., Brook J. S., Zhang C., Cohen P., & Whiteman M. Drug use and the risk of major depressive disorder, alcohol dependence, and substance use disorders. Archives of general psychiatry, 59(11), 2002.– P. 1039–1044.
2. Brook J. S., Morojele N. K., Pahl K., & Brook D. W. Predictors of drug use among South African adolescents. Journal of adolescent health, 38(1), 2006.– P. 26–34.
3. Chassin L., Flora D. B., & King K. M. Trajectories of alcohol and drug use and dependence from adolescence to adulthood: the effects of familial alcoholism and personality. Journal of abnormal psychology, 113(4), 2004.– 483 p.
4. Fehrman Elaine. (2011–2012). Drug consumption (quantified) Data Set. Survey. March. International Drug Policy Consortium. (2018). Progress Report. URL: http://fileserver.idpc.net/library/Progress%20 Report_2018.pdf.
5. "National Survey of Drug Use and Health". (2017). Survey.
6. Redonnet B., Chollet A., Fombonne E., BoIs L., & Melchior M. Tobacco, alcohol, cannabis and other illegal drug use among young adults: The socioeconomic context. Drug and alcohol dependence, 121(3), 2012.– P. 231–239.
7. Single E., Kandel D., & Faust R. Patterns of multiple drug use in high school. Journal of Health and Social Behavior, 1974.– P. 344–357.
8. Vitaro F., Brendgen M., Ladouceur R., & Tremblay R. E. Gambling, delinquency, and drug use during adolescence: Mutual influences and common risk factors. Journal of gambling studies, 17(3), 2001.– P. 171–190.
9. Illicit Drug Addiction: Symptoms and Resources for Help. (n.d.). Retrieved from URL: https://www.healthline.com/health/addiction/illicit-drugs#types-of-drugs
10. American Society of Addiction Medicine. Public Policy Statement: Definition of Addiction. 2011.
11. Race K. Pleasure consuming medicine: The queer politics of drugs. Duke University Press. 2009.
12. Fehrman E., Muhammad A. K., Mirkes E. M., Egan V., & Gorban A. N. The Five Factor Model of personality and evaluation of drug consumption risk. In Data Science 2017.– P. 231–242. Springer, Cham.