Técnicas de Machine Learning aplicadas al consumo de sustancias psicoactivas ilícitas: Un mapeo sistémico.

Machine Learning techniques applied to the consumption of illegal psychoactive substances: A systematic mapping.

Jefferson Eduardo Campo Yule iD
Universidad del Cauca. Popayán, (Colombia)
jeffersoncy@unicauca.edu.co

Danny alberto Díaz Mage iD
Universidad del Cauca. Popayán, (Colombia)
dannyad@unicauca.edu.co

Hugo Armando Ordoñez-Eraso iD
Universidad del Cauca - Research group GTI. Popayán, (Colombia)
hugoordonez@unicauca.edu.

Resu**men**

El consumo de sustancias psicoactivas ilícitas es una problemática que se vive a diario, donde personas de diferentes edades se han visto implicadas, resaltando que muchas de estas sustancias generan trastornos tales como, por ejemplo: la Marihuana o cannabis: su consumo afecta la función cerebral de manera directa, y particularmente las partes del cerebro responsables de la memoria, el aprendizaje, la atención, la toma de decisiones. El Bazuco: es una sustancia tóxica, cuyos principales riesgos de consumirla se reflejan en el deterioro neurológico y en el organismo, y es muy rápida su disolución en el torrente sanguíneo, aspecto que hace que sea muy adictiva. La Cocaína: su consumo afecta directamente el sistema nervioso y el resto del organismo de forma inmediata, en estas afectaciones se encuentran vasoconstricción, midriasis, hipertermia, taquicardia e hipertensión. La Heroína: es una sustancia altamente adictiva, inicialmente, sus efectos son muy placenteros, lo que propicia una conducta de consumo continuada y repetitiva, además, produce sensaciones de sequedad en la boca, enrojecimiento y acaloramiento de la piel, pesadez en brazos y piernas, náuseas y vómitos, comezón intensa y enturbiamiento de las facultades mentales. Por otra parte, a nivel mundial y en regiones de América Latina esta problemática es algo que resalta mucho y de gran impacto en los jóvenes de acuerdo al contexto en el que se encuentren ya que hoy en dia hay mucha facilidad para obtener este tipo de sustancias, por ende, se han planteado una serie de trabajos que abordan desde la inteligencia artificial esa problemática, en este sentido, el presente estudio se realiza una revisión de 50 publicaciones relacionadas con el uso de métodos y técnicas de ML aplicadas al consumo de sustancias psicoactivas ilícitas. De las publicaciones incluidas se hallaron temáticas en común por lo que se hace un resumen de los artículos seleccionados por cada temática y se describen brevemente los métodos adoptados, así como también una comparativa entre ellos, anotando los métodos usados, sus resultados y demás factores importantes de la aplicación o modelo en distintas áreas y se concluye con una serie de propuestas sobre las líneas que a futuro podrían encaminar la investigación en este campo.

Abstract

The consumption of illicit psychoactive substances is a problem experienced every day, by people of different ages who have been involved in it, highlighting that many of these substances generate disorders such as, for example: Marijuana or cannabis: its consumption affects brain function directly, and particularly the parts of the brain responsible for memory, learning, attention, decision

making. Bazuco: it is a toxic substance, which main risks of consumption are reflected in the neurological deterioration and in the organism, and its dissolution in the bloodstream is very fast, an aspect that makes it very addictive. Cocaine: its consumption, directly affects the nervous system and the rest of the organism immediately, these affectations include vasoconstriction, mydriasis, hyperthermia, tachycardia and hypertension. Heroin: is a highly addictive substance, initially, its effects are very pleasant, which leads to a continuous and repetitive consumption behavior, in addition, it produces sensations of dry mouth, reddening and heating of the skin, heaviness in arms and legs, nausea and vomiting, intense itching and clouding of the mental faculties. On the other hand, in Latin American regions and all over the world, this problem is something that stands out a lot and has a great impact on young people according to the context they are in, since nowadays it is very easy to obtain this type of substances, therefore, a series of works have been proposed that address this problem from artificial intelligence, in this way, the current study is a review of 50 publications related to the use of ML methods and techniques applied to the consumption of illicit psychoactive substances. From the publications included, common themes were found, so a summary is made of the articles selected for each theme and the methods adopted are briefly described, as well as a comparison between them, noting the methods used, their results and other important factors of the application or model in different areas, and concluding with a series of proposals on the lines that could guide future research in this field.

Key Words

Machine Learning, illegal psychoactive substances, illegal drugs, treatment, predictive models, new illegal psychoactive substances.

Nomenclature	Meaning
ML	Machine Learning
IPS	Illegal Psychoactive Substances
NIPS	New Illegal Psychoactive Substances
RF	Random Forest
LR	Logistic Regression
ROC	Receiver operating characteristic curve
AUC	Area under the receiver operating characteristic curve
SVM	Support Vector Machine
DL	Deep Learning
VAR	Modelo autorregresivo vectorial
OLS	Mínimos cuadrados ordinarios
KNN	K-nearest neighbor
DT	Decision Tree
GAIN	Generative Adversarial Imputation Network
SUD	Substance use disorder

Table I. Nomenclature

I. INTRODUCTION

Illicit psychoactive substances (IPS) directly affect the nervous system which implies alterations in several functions that control aspects such as: emotions, thoughts and behavior, the composition of these IPS is based on natural or synthetic components. Therefore, the consumption of IPS is one of the problems commonly present around the world, and externalizes short-term consequences to people who consume them, among these consequences are: intoxication, aggressiveness, depression, anxiety, insomnia, anorexia, decreased libido, violent behavior, cerebrovascular accidents, auditory and visual hallucinations, psychosis and the worst of all addiction to these substances, which creates dependence, thus generating a chronic disorder that is characterized by the need to repeatedly consume the substance and the loss of the ability to control consumption, in the worst cases affecting their health, leading the consumer to a street situation (street dweller) and even resulting in death[1].

Based on the above, the problem of IPS consumption has spread worldwide, affecting countries such as the United States, China, Bangladesh, Latin American countries, among others, which can be evidenced in the articles found. In addition, the United Nations report against drugs and crime showed that a total of 284 million people worldwide between 15 and 64 years of age have consumed some type of drugs, a significant increase of 26% over the previous decade. The highest consumption occurs in young people, increasing the levels of consumption for each specific country. Therefore, this becomes a global challenge that requires each country to adopt prevention measures for both young people and people at risk, in addition to providing treatment and education to people who are already in this consumption problem [2].

On the other hand, in the social sphere there are other risk factors, both personal (failure, low self-esteem, not knowing how to manage free time, not having social skills, etc.) and in the consumer's environment (contact with people or friends who have IPS consumption habits, ease of obtaining substances, etc.) that influence the initiation of IPS consumption. Therefore, it is important to find prevention or treatment strategies for this social problem, such as the Comprehensive Policy for the Prevention and Care of IPS Consumption, which is proposed in [3].

As an advance in technology, Artificial Intelligence has become a key tool that can be used to predict and inform about the consequences both to people who are at risk of initiating IPS use, as well as to those who are already in this problem.

Over the last decades, several techniques and methods have been developed in response to this problem, which have led to rapid progress in achieving the objectives in this field. More recently, the adoption of artificial intelligence techniques such as machine learning has made it possible to improve the results in terms of accuracy and prediction to extract common patterns and behaviors in order to help make sophisticated decisions in medical applications such as computer-aided diagnosis[4]. In this sense, it is important to know the current and future status of the techniques most commonly used to address this type of problems, therefore, the aim of this article is to present a systemic mapping of artificial intelligence and ML methods applied to the consumption, abuse and addiction of IPS. For this reason, section 2 presents the methodology that will guide the review and selection of the articles and/or studies analyzed in the systemic mapping. Section 3 includes the most relevant publications for the topic that have been synthesized and subsequently analyzes important factors to consider and also answered the research questions that helped guide the systemic mapping. Finally, the paper concludes with a discussion of future challenges and possible research approaches, which can be found in section 4.

II.   RELATED WORK

Predictive ML models for IPS addiction prediction.

A variety of models have been proposed to predict the addiction of IPS by users who consistently use different types of IPS (marijuana, cocaine, heroin and bazuco). These models make use of ML to develop and improve the results. In [7], a classification of drug dependent users into three types of groups (mild, moderate and severe) is proposed. The input data collected are used to design a test algorithm, measuring the degree of addiction based on ML techniques and convolutional neural networks. In [8] a study is conducted with data related to IPS consumption and addiction of Bangladeshi university students. With the data, ML techniques were used to discover the most correlated features and behaviors from the dataset in order to determine the main causes of addiction and obtain an accurate prediction. In [9], a prediction approach for juvenile delinquency in Bangladesh due to IPS addiction is proposed using ML techniques. Such an approach studies the causes of IPS addiction in adolescents by pointing out behavioral disorders and finally predicting their tendency to delinquency. In [10] the authors have performed a predictive model using ML that consists of predicting the risk of becoming addicted to IPS, taking into account significant factors and variables from data collected from both addicted and healthy individuals. According to the methods in terms of performance, logistic regression outperformed algorithms such as KNN, SVM, Naïbe bayes, classification, Random Forest RF, regression trees, multilayer perception, adaptive boosting and GBM with 97.91% accuracy. In [11] proposes a predictive ML model in which psychological factors of major impact on the desire and cravings for IPS consumption in individuals are considered. The model aims to help develop recovery plans for people addicted to IPS according to these factors. Finally, in [12] a framework based on ML techniques is proposed to identify individual vulnerability to IPS abuse. For this purpose, socioeconomic variables are taken into account to assess the causes that are commonly found to be involved in patients suffering from IPS abuse and healthy individuals.

Trends in the use of ML techniques, methods and algorithms for predicting IPS use and abuse

It is important to review the state of the art in terms of ML techniques, methods and algorithms that provide the best accuracy percentages to build a robust predictive model. For this reason, the following articles are presented:

In [13] the authors provide a summary of how each effective machine learning method should be applied in IPS abuse addiction studies. The authors conclude that the most common method with the best predictive accuracy is RF, furthermore it is efficient with a large corpus of data regardless of the presence of null data. Two categories of computational models addressing drug use and IPS addiction are described in [14]. These models adopt predictive coding

approaches, focusing on Bayesian prediction error with a focus on the effects of IPS use. In [15], prediction models for IPS intoxication mortality were developed. Each implemented method was trained, validated and tested through data collected from samples from centers for disease control and prevention in Korea to compare their performance using performance measures (Brier score, calibration slope and calibration-in-the-large) in order to identify the appropriate technique for the medical field. In [16] provides a systematic review of the state of the literature on the applications of ML methods in IPS addiction research. For this purpose, 23 articles were evaluated of which 5 of them employed the following methods, multiple algorithms, six of them used regression techniques and two focused on classification. Two studies were found that applied reinforcement learning using direct methods. Finally, this indicates that such techniques are increasingly being used in the field of IPS addiction psychiatry for medical decision making. In [17] they propose a systematic review on the state of the literature on predictive models using ML to examine psychological disorders produced by IPS abuse in digital health care. In their results, a comparison is made and their performance, limitations and challenges of techniques such as Random Forest RF, Decision Trees DT, Support Vector Machine SVM and DL-based Neural networks are measured. It was determined that RF was the most used in terms of improving performance. Finally, in [18] provides an overview of the usage trends in methods, techniques and algorithms applied in ML models in the field of IPS abuse. The selected models predict current substance use, future risk of substance use, and treatment success. The most common methods are Naive Bayes, K-nearest Neighbors KNN, Decision Trees DT, Support Vector Machines SVM.

Predictive ML models focused on the detection of patterns and behaviors in IPS consumption.

These studies focus on identifying IPS users using predictive ML models in order to give a timely diagnosis and determine what type of IPS these users use. In [19]proposes an expert system with ML that diagnoses and identifies drug users and the type of drugs they use. The methods used (Forward Chaining and Certainty Factor) obtain an accuracy rate of 80% in their results. In [20] a predictive ML model is proposed that allows finding and correlating patterns that contribute to opioid overdose or abuse among a variety of patients. The model when classifying patients achieved an F1 score of 94.45%. This model is intended to be an efficient tool to help uncover fraudulent practices in the issuance of prescriptions. In [21] they propose a study that trains and validates a ML/DL model to identify patients with IPS abuse, in order to better understand the problems of consumption of such drugs. It is concluded that the GAIN method is best suited to predict AUDIT and DAST test scores. In [22] ML models were developed to predict the risk of developing SUD using different types of algorithms in order to compare and evaluate the accuracy of the predictions. It is claimed that the RF algorithm optimally detects psychological, health, environmental, and social behavioral characteristics that predict substance use disorders compared to others. In [23], the applicability of ML methodology to detect the trajectory of high and low severity IPS use based on the harmfulness of the substances used is demonstrated, spanning ages 10 to 30 years. The algorithm employed identified a set of psychological and health characteristics that predict the trajectory with high probabilities of ending in SUD. Finally, in [24] they propose an approach that classifies in a binary manner to predict the current vulnerability of any individual towards the use and abuse of IPS taking into account the socioeconomic environment. This approach used ML techniques and a logistic regression classifier to identify patterns and variables of key characteristics influential in IPS abuse.

Predictive ML models of effective prevention and treatment in IPS consumption for the health sector.

In this section, a review of the state of the art on proposals related to the treatment of psychoactive substance use was conducted. These proposals use ML methods, which allow health entities, physicians and professionals in health-related areas to use them as a tool to help in the treatment of psychoactive substance use and addiction.

In [25], a model using a hybrid ML technique combining PCA and K-Means++ is proposed to cluster drug addicts according to the relationship found in data collected using virtual reality in patients with IPS use. In [26], a system to aid in clinical decision making based on ML is proposed to predict the increased risk of discontinuation in patients who have initiated treatment for opioid use disorder. This model is of great importance in clinical settings as the prediction of treatment discontinuation can help develop personalized support systems to improve long-term retention of patients in treatment. A predictive model that estimates cocaine treatment success in hospitalized patients is presented in [27], which could represent a new step in treatment management thus avoiding treatment dropout rates. The model is based on a heterogeneous set of high-dimensional features, which allows to generate and learn new associations between different features. The RF method achieved 82% accuracy compared to others. In [28] ML models were developed using demographic and psychometric assessment data of patients in alcohol addiction treatment. Different predictions were

obtained according to the model and the data set in order to compare and define the most accurate model, which in this case was the algorithm (Fuzzy Unordered Rule Induction Algorithm) with 74% accuracy. In [29] a method is proposed to build a prediction model that uncovers interaction effects (length of stay, frequency of substance use) that might be ignored by other traditional hypothesis generation approaches affecting treatment completion. The extreme gradient boosting method showed better predictive performance compared to others. In [30] a state of the art study is conducted on predictive modeling approaches with ML to generate individual level predictions of complex addiction outcomes and provide neurobiological insight into the brain basis of such behaviors to help as a diagnostic and treatment management tool. As a result, limitations were found in the approaches, as many of them do not consider treatment-specific factors, among others. Finally, in [31] predictive models are proposed and evaluated using ML for the personalization of rehabilitation therapies in users with drug addictions. As input, data considering sociodemographic, pharmacological, mental, cognitive and personality information, among others, are taken into account. The best predictive model is the RF algorithm with 82.12% accuracy.

Detection and identification of new illegal psychoactive substances (NIPS)

The new illegal psychoactive substances (NIPS) are drugs that contain a chemical structure and pharmacological actions similar to the illegal drugs that currently exist, these NIPS are not being legally controlled by the UNODC [32][33][34], highlighting that in recent years the NIPS have been increasing as well as their consumption [35][34], 35][34], therefore the detection and identification of these NIPIs is becoming necessary for their regulation and control[33][35], and although these substances harm people's health, they are also being used as alternatives to evade the legal control that currently exists on drugs[32][33], and even more so taking into account the large number of people who are addicted to drugs, where most people do not belong to a population with a particular age, but people of all ages[36], therefore, with technological progress and advances in artificial intelligence, in [32] they propose a quantitative structure-retention relationship model (QSRR) as a machine learning system in order to predict the retention time of mass gas chromatography for NIPS using AutoML, clarifying that the data used for the research were obtained in previous studies, similar case to [36] where they propose a complete database called AddictedChem, which contains information of controlled substances, in addition, they developed 29 predictive models using 5 learning algorithms and 7 molecular descriptors for the detection of NIPS, In [35] they also developed learning models to identify NIPSs by using GC-MS (gas chromatography mass spectrometry) data to train and test them, on the other hand, in [33] they conducted a study of the chemical diversity of NIPSs in [33], which contains information on controlled substances, in [33], a study of the new applications of high-resolution mass spectrometry (HRMS) techniques is carried out, where the study of wastewater from a given population is taken into account, in order to obtain consumption trends, as well as to examine the effects produced inside the organism by ISP abuse, finally, as a complement in [34] they create an ML algorithm in order to predict the similarity that may exist between current drugs and NIPS, where the data used to train this developed model were obtained from the brains of mice using monoamine neurotransmitters and steroid hormones.

ML applied to ISP consumption and social networks

With the current increase in technology there has also been a high growth of social networks, as well as an exponential increase of people using social networks such as Instagram, Twitter, among others, however, these social networks are being used as a channel for marketing both controlled drugs and illegal drugs[36][37][38], where the publications made in these social networks (Twitter, Instagram) allow to know what they offer or market, in addition, with the legalization of marijuana in the United States, and more specifically in 29 states[39], it is common to see Twitter accounts associated with the marketing and information of marijuana, but in other cases, such as opioids, one of the drugs that has caused the most deaths in the United States [40] and which is also being marketed on social networks such as Twitter, this has become a problem that must be addressed, thanks to artificial intelligence, as well as the different branches that make up AI such as ML, in [41] they make use of an unsupervised machine learning approach and the biterm(BTM) model in order to thematically summarize Instagram comments which are then used to detect comments of buying and selling drugs and illegal drugs, similar case to [37] in which they developed a machine learning approach using web scraping and deep learning to detect Instagram posts focused on illegal drug trafficking, such approach was able to detect illegal drug sellers with high accuracy, in addition, it should be noted that much of the information that exists in Instagram profiles serves as data for the prediction of people who are at risk of consumption, in the same way, in [38] develop a deep learning method to classify the risk of consumption of alcohol, tobacco and other drugs in people, taking into account the content of the profiles of each of the participants, on the other hand, in [39] they present an approach using Compositional Multiview Embedding(MCE) for the classification of accounts related to Marijuana, either for information of laws for the legalization

of the same (Marijuana), marketing as a medicine, finally in [40] which is related to Twitter, they developed a methodology using machine learning to detect the marketing and sale of opioids on Twitter, they also developed wireframe as a prototype to detect, classify and report illicit tweets from pharmacies that sell illegal substances.

ML applied in heroin and opioid detection and consumption.

Opioid abuse according to the Substance Abuse and Mental Health Services Administration, is defined as prohibited use of opioid analgesics or use of heroin which is a commonly used synthetic opioid[42] highlighting that in China and more specifically in Yunnan, heroin has been the most commonly seized along with methamphetamine, They account for 80% of all illicit drugs seized[43], and opioid can be defined as any natural or synthetic substance that binds to specific opioid receptors in the human body[44], however, opioid abuse is one of the major causes of overdose deaths in the United States, For example, in the year 2020 there were more than three million deaths involving opioids and yet overdose deaths continue to increase, tripling in the last 10 years, where a percentage of deaths are due to prescription opioids, which makes this problem a crisis or epidemic for public health[44][42][45] and even more so because young people and adults are at-risk populations that have relapsed after treatment[46], thus with the different problems presented above, in [46] they make use of two ML approaches to evaluate different factors such as: demographic, psychosocial, psychological comorbidity and environmental factors to predict those at risk of relapse after treatment for opioid use. similar to [44] where they developed ML models with which they conducted screening and reuse assessments of drug candidates for the treatment of opioid use disorder, and also used the models developed to reassess the side effects of drugs currently available for the treatment of opioid use, however, in [42] they are more focused on opioid misuse, with ase to that, they developed three prediction models for the prediction of opioid misuse in adolescents, likewise, with the three models they managed to find that the prediction performance was similar in all the developed models, with the results obtained it can be said that ML techniques are promising for the prediction of opioid use, as a complement in [45] as well as the previous study are focused on the misuse of opioids and heroin use (OM), is so then, they developed a method of natural language processing using the logistic regression algorithm to improve the identification of OM from paramedical documentation, where they manage to conclude that the approach raised helps in public health as prevention to opioid use. In the same way, taking into account the problem presented for the seizure of heroin and methamphetamine in Yunnan province in [43] presented a method for recognition of the sources of heroin drugs quickly and non-destructive with the use of an infrared and ML algorithm, such an approach developed facilitates research in the criminalistic area to be easy to use and low cost.

ML applied in the detection and prediction of cocaine use.

When talking about cocaine it is necessary to mention that it comes from the coca leaf, therefore there are organizations and programs that seek to reduce or eliminate coca plantations, as is the case of the Eradication Program of the Special Project for the Control and Reduction of Illegal Crops in Alto Huallaga or CORAH located in Peru, However, it is necessary to clarify that in Peru the coca leaf is a sacred leaf that people like peasants chew to obtain energy and thus continue working, without forgetting that, being an illegal product, its price is quite high[47], in addition, it is worth noting that cocaine is one of the most traded and consumed drugs in the world, Although research has been conducted to control cocaine addiction, a drug capable of controlling addiction and lack of dependence on this drug (cocaine) has not yet been found[48][49] and one of the possible reasons for cocaine consumption may be related to depression[50], It is worth mentioning that with the advance of the drug trade, it has become a challenge for the authorities to detect them in cases of crime[51], therefore, with all the above mentioned and with the problems presented, In [48], ML and DL models are presented in order to predict the potential drugs for cocaine addiction control, as well as to know the side effects they may cause when consumed, similar to [49] in which they developed a ML/DL-based platform trained by a proteome dataset to discover lead compounds against cocaine which will help to discover drugs for cocaine addiction, and in [50] they developed models to predict the possible depressive symptoms that may be related to cocaine consumption, highlighting that the data of the people taken into account for this study were obtained from electronic records of the Health Sciences Center of the University of Texas from 2006 to 2017, on the other hand, taking into account the problem related to coca plantation in [47] they use a technique based on ML in order to select the most relevant variables to analyze the determinants of coca plantations and also to evaluate the variables selected by the ML technique they use the OLS and VAR method, concluding that eradication policies do not have a greater effect and it is necessary to seek new alternatives to reduce coca plantations. Finally, to solve or attack the problem of drug detection in a case or crime, in [51] they propose to use near infrared spectroscopy (NIR) to safely detect cocaine in samples of criminal cases or drug seizures,

noting that the use of NIR is complemented by the use of ML algorithms to predict the presence of cocaine and also to know the concentration and composition of the sample.

ML applied in the detection and prediction of marijuana use.

Marijuana is one of the most popular and most consumed IPS in the world, as people of different ages are involved[52], taking into account that one of the possible causes of the increase in marijuana consumption is due to the legalization that has been established in 15 states of the United States for recreational and medical purposes[53], however, marijuana consumption is not only seen as a cigarette but it is also being consumed in vapes, which leads to include mostly young people who normally use this type of objects[54], It should be noted that marijuana use is not only seen in ordinary people but also in people who belong to entities such as the army, where studies have been conducted and have found positive results of drug use and especially marijuana use[55], however, the causes of marijuana use may be induced by people's past, because the consumption of IPS may have started due to the abuse that the person may have received in childhood[56], in addition, studies have shown that marijuana consumption is closely related to depression and can cause suicide in adulthood[52]. Therefore, with the different problems presented regarding marijuana consumption, in [53] it is proposed to develop and evaluate ML models in order to predict the daily consumption of people, as well as to identify factors associated with this consumption, in which RF had the best performance, in addition, they managed to identify the factors that are the possible cause for the daily consumption of marijuana, on the other hand, for the case of young people or people who consume marijuana through the vaper or electronic cigarette, in [54] they propose an ML approach with the purpose of predicting the initiation of marijuana vaping, with which they managed to obtain that the approach applied using ML can be promising for the prediction of risk behavior, on the other hand, for the problems presented regarding marijuana consumption in entities such as the army, In [55] they model public data on drug use, demographic data, among others, using logistic regression, decision trees and neural network models in order to predict the risk of marijuana use through personality traits where young people were more involved than adults, similar to [56] where they use ML to examine the shared and non-shared risk factors for marijuana use in child welfare and non-child welfare youths, finally, taking into account the relationship that can occur between marijuana use, depression and suicide, in [52] they make use of ML algorithms such as the logistic regression algorithm, RF and KNN to build prediction models applied to the risks of suicide and depression, where they found that people who have consumed marijuana are at risk of becoming depressed or suicidal.

III.   METHODOLOGY

In this systemic mapping, a methodology was used (see Figure 1) that allows the collection of a significant amount of information in accordance with the proposed objective, using the methodology proposed in [5], with an adaptation of the proposed processes.
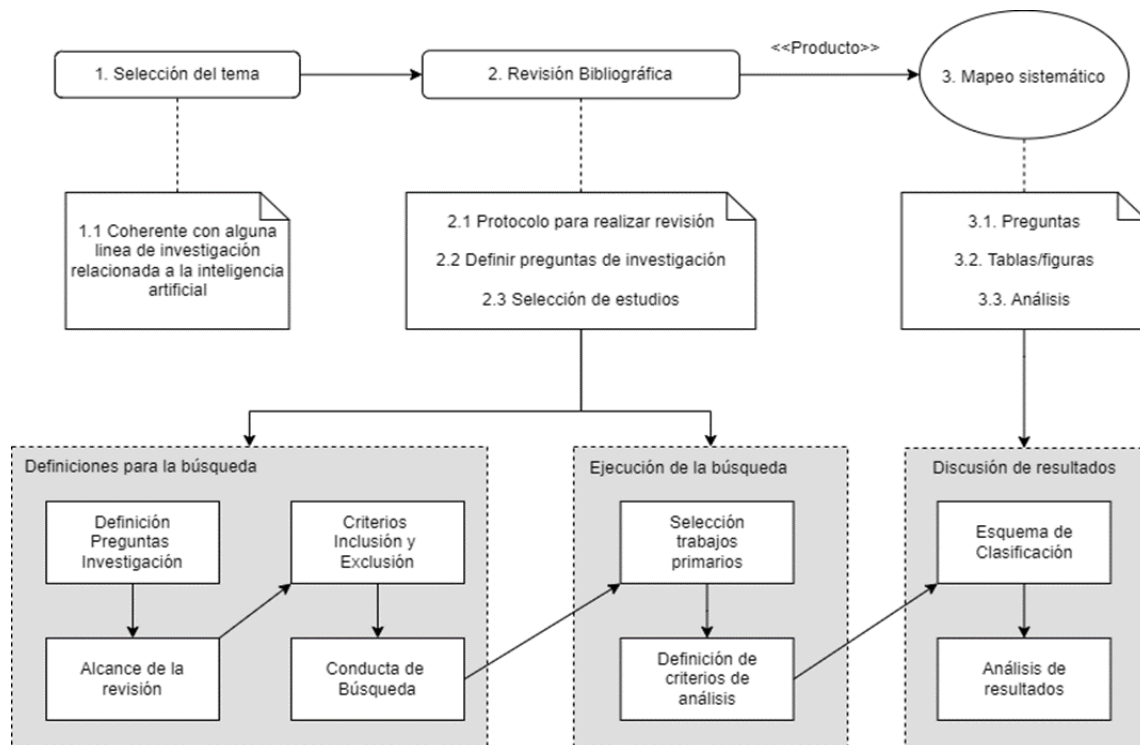
Figure. 1 Systemic mapping processes.

The new processes are: Selection of the topic related to artificial intelligence, bibliographic review that in synthesis considers the three important steps as the protocol, definition and execution of the search, finally the systemic mapping as a product in which the answers to the questions posed, data analysis, and graphs are obtained.

In addition, the processes already described in [5], as well as in Figure 1, were taken into account to perform the systemic mapping. First, the research questions and the scope of the study were formulated. Then, a search for articles was conducted in the databases described in the bibliographic review section. Subsequently, the data needed to answer the questions formulated were extracted, and both the analysis and conclusions of the mapping were made.

A. *Theme selection*

For the process of selection and delimitation of the research topic (ML techniques applied to IPS consumption), some factors and characteristics are considered, including:

a. Previous knowledge and experience in the development of predictive models using ML.
b. The potential impact that IPS consumption can have in Colombia, as this is a critical problem that impacts different public health sectors and society, among others.
c. Availability of resources, since there is a wide variety of articles focused on applications of predictive models with ML in the consumption, addiction and treatments of IPS.

B. *Bibliographic review*

In the bibliographic review, one of the steps to take into account is to formulate the research questions to guide and focus the objectives of this study (see Table 2), for which a review of the state of the art on predictive models with ML applied to IPS consumption was carried out to determine which works are relevant to the selected topic.

Figure 2 presents the questions to answer the why, how and where in terms of the formulation of the research questions.
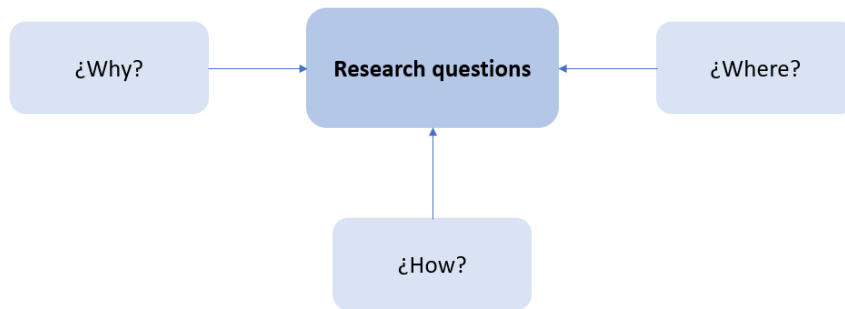
8

Figure. 2 Guide to formulate research questions.

a. Why: The questions helped to focus and delimit the scope of this study and provided a guide for the analysis and selection of studies.
b. How: Here it was important to focus the study on the selected topic; therefore, it was necessary to identify the key concepts, the topic and the different issues analyzed by the studies, which allowed defining the research questions in a coherent manner with the proposed topic.
c. Where: The research questions took into account all the studies found that were related to the selected topic.

The next step was to establish inclusion and exclusion criteria related to the subject matter and other factors that made it possible to determine which studies were relevant for the systemic mapping (see Table 3).

As for the information sources that allow searching and selecting relevant studies from scientific journals, articles, bibliographic databases such as Scopus, ScienceDirect, IEEExplore, Google Schoolar, Springer Link, web of Science were included.

The main terms that allow to focus and delimit the search for articles related to ML applied to the detection, prevention and treatment for the consumption of IPS as well as for the specific search of IPS can be seen in Table 4. The search strings used follow the steps recommended by kitchenham[6] , which consider the logical operators AND and OR for the construction of the search strings. Table 4 describes the terms used in this protocol.

It is important to clarify that the steps mentioned above define the scope of the systemic mapping and ensure that the review is refined, rigorous and exhaustive.

Question indicator	Research question	Motivation
General question	¿What is the current state of the art regarding ML techniques applied to IPS consumption?	To understand the current state of the selected topic, taking into account the different applications of ML in it (the selected topic).
PI 1	¿What is the accuracy of ML predictive models for predicting the use of different types of psychoactive substances?	Entender y comprender la eficacia del uso de los modelos de ML para ayudar a predecir el consumo de IPS en las personas
PI 2	¿What specifically are ML techniques used for in the context of treatment for illicit psychoactive substance use?	To understand and comprehend the effectiveness of using ML models to help predict IPS consumption in individuals.
PI 3	¿What can ML models, techniques and algorithms be used for to identify individuals at risk of developing addiction to illicit psychoactive substances?	Identifying the techniques that perform best in identifying individual at risk of consuming IPS

Table II. Research questions

Inclusion criteria	Exclusion criteria
Articles published in scientific journals.	Articles that are not published in scientific journals and peer-reviewed.

Articles focusing on the identification, treatment, addiction, and prevention of psychoactive substance consumption.	Articles focusing on substances that are not considered psychoactive.
Articles that use ML techniques for the analysis of patterns and behaviors related to psychoactive substance consumption.	Articles in which the use of ML techniques to analyze data related to psychoactive substance consumption is not evident.
Types of selected articles (Books, articles).	Articles that are not available or accessible, or that do not have a complete format..
Articles published in English or Spanish languages.	
Articles published between the period from 2018 to 2023.	

Table III. Inclusion and exclusion criteria

Main terms	Search queries.
ML in psychoactive substances	((((predictive AND models) AND (machine AND learning) AND (psychoactive AND substances) OR (illegal AND drugs)))) ((((predictive AND models) AND (machine AND learning) AND (cocaine OR marihuana OR heroin OR bazuco))))
Predictive models techniques in psychoactive substances	((((predictive AND models) AND (machine AND learning) AND (techniques OR methods OR algorithmics) AND (psychoactive AND substances) OR (illegal AND drugs))))
Predictive models in psychoactive substances with Drugs addiction, treatment, use, risk factors.	((Predictive models) AND (Machine AND Learning) AND (psychoactive AND substances) AND ((drugs use) OR (SUD))) ((Predictive models) AND (Machine AND Learning) AND (psychoactive AND substances) AND (drugs AND addiction)) ((Predictive models) AND (Machine AND Learning) AND (psychoactive AND substances) AND ((Treatment) OR (Healthcare) OR (diagnosis))) ((Predictive models) AND (Machine AND Learning) AND (psychoactive AND substances) AND (Risk Factors))

Table IV. Search query.

C. *Study selection*

Figure 3 shows the steps carried out for the selection of studies that were taken into account in the systemic mapping, for example, in the identification section, first a search was performed in the different databases used for the study, the results obtained allowed a global visualization of the breadth of the subject matter, as a next step, duplicate citations were eliminated, and the combination of terms and inclusion and exclusion criteria were applied, reducing the number of results (n = 121).

However, when searching in google scholar these results contained a very large number of articles due to the fact that the search engine does not allow filters to be applied, so a manual selection of each one of them was made, taking into account the inclusion and exclusion criteria and reviewing the most relevant articles (n = 37).

The next step was to read the title and summary of the studies selected for both database and manual selection, in which (n = 80) were excluded via database and (n = 21) via manual selection, some of the excluded studies used other types of techniques, or focused on addictions other than IPS. We also excluded studies that had incomplete information, were poorly structured, or were excluded because of their language (n = 5) via database and (n = 2) via manual selection. Finally, the included studies (n=36) via database and (n=14) via manual selection for a total of (n=50), met the search strategies related to IPS consumption to be considered in the systemic mapping. In addition, studies focusing on the same line were grouped. Some (n = 6) of them focused on the prediction of IPS addiction; in other studies (n = 6) a review of the most applied techniques in IPS use and abuse was carried out in order to compare the best of them; in others (n = 7) they focused specifically on the prediction of IPS use; other articles (n = 6) are focused on the prevention and treatment

of IPS use in the health sector; others that although they are focused on new substances, these substances are derived from the IPS in which they were found (n = 5); other studies (n = 5) focus on the relationship between social networks and IPS use; finally, articles that focus on a particular psychoactive substance such as marijuana (5), cocaine (5), heroin (5).
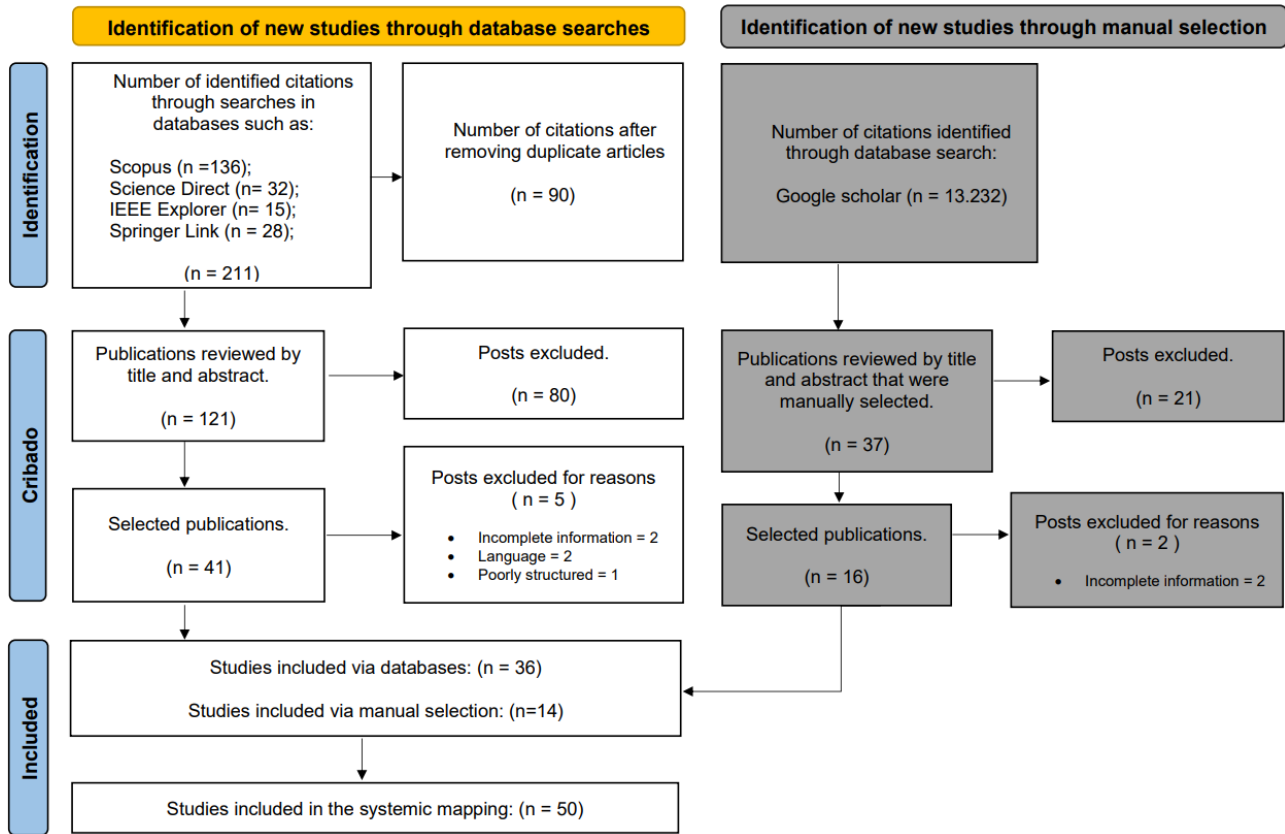
Figure. 3 PRISMA diagram

IV.   RESULTS

Table 5 shows the number of articles obtained in each step of the search process according to the main terms and their derivatives, in addition, the steps followed in Figure 3 show the process for the selection of articles.

Search query	Number of articles found	Review of title, abstract, and conclusions	Selected
1	3.109	23	8
2	4.641	27	16
3	3.573	12	6
4	402	9	4
5	507	11	4
6	319	16	5
7	519	17	4
8	313	11	3

Table V. Number of results found

Based on the articles found, we proceeded to analyze each one of them, identifying and grouping them by topics of common interest in the related works section.

A.   *Answers to research questions.*

PI1 What is the accuracy of ML predictive models for predicting the use of different types of psychoactive substances?

Considering the articles selected for this systemic mapping, Table 6 presents the methods, techniques and algorithms used in each article.

Reference	Application context	Study population	ML techniques, methods and algorithms	Model results
[53]	Daily marijuana use	Persons aged 18 or over 18	•Logistic regression •Decision tree •Random Forest •Naïve Bayes	Taking into account the area under the curve (AUC), RF had an AUC of 0.97, while DT had an AUC of 0.95, both of which outperform LR with an AUC of 0.73 and Naïve Bayes with an AUC of 0.67. Therefore, they conclude that RF is the best model, surpassing the other models.
[24]	Consumption and abuse of psychoactive substances, unauthorized drugs, and alcohol.	Substance abuse patients and healthy individuals from the rehabilitation center in Dhaka city.	•Random forest •KNearest Neighbors •Decision Tree •Linear SVC •Gaussian Naive Bayes •Logistic Regression	According to the study, the most effective model for accurately distinguishing between healthy and addicted classes is the LR classifier with a precision score of 96.72% and an AUC of 0.98.
[19]	Drug consumption	30 people selected from a survey conducted in Maratan, Indonesia.	•Expert system based on KNN	The expert system demonstrates up to an 80% accuracy rate based on various tests conducted with the data as input.
[20]	Consumption and abuse of IPS derived from opioids.	46.520 patients from the Beth Israel Deaconess Medical Center in Boston, Massachusetts..	•Logistic Regression with L2 regularization •Extreme Gradient Boosting •K-Means	The improved model, which uses LR and Extreme Gradient Boosting, to classify patients susceptible to opioid abuse, has an F1 score of 94.45% (precision 94.35%). The unsupervised algorithm K-Means was used to explore pharmacological interactions that could be of concern.
[21]	Alcohol use and substance use disorder	6,978 adults, these data are obtained from three public medical facilities.	•KNN •Multiple Imputations by Chained Equations (MICE)	The GAIN method for data imputation along with mixed-effects prediction models are the most suitable for predicting AUDIT

Reference	Problem	Population	Techniques, models and algorithm	Application
			•Multiple Imputation with Denoising Autoencoders (MIDAS)•DataWig•Generative Adversarial Imputation Networks (GAIN)	scores, with a prediction accuracy index of 0.94 and an F1 score of 0.99. For DAST, the prediction accuracy index is 0.93 and the F1 score is 0.99.
[22]	Substance use disorder	People between 12 and 14 years of age.People of 16, 19 and 22 years of age.	•Random Forest•Logistic Regression•Adaptive Boosting•Naïve Bayes•Support Vector Machine•KNN•Deep Neural Network	According to the proposed algorithms, RF identified thirty psychological, health, environmental, and social behavioral features that predict SUD in each of the five assessments conducted across a variable age range..
[23]	Substance use and its severity	Individuals from childhood (12-14) to adulthood (30) years of age.	•Random Forest•Support Vector Machine•Naïve Bayes•Adaptive Boost•Nearest Neighbor•Artificial Neural Network	The ML model for predicting both low and high trajectories leading to SUD demonstrates an accuracy of 71% between ages 10 and 12, increasing its precision to 93% by age 22. These results could inform primary and secondary prevention efforts.It is worth mentioning that among the proposed algorithms, RF and Naïve Bayes stood out as having the best results.

Table VI. ML techniques, models, and algorithms used for prevention and treatment programs for IPS consumption.

PI2 What specifically are ML techniques used for in the context of treatment for illicit psychoactive substance use?

According to the articles selected for this systemic mapping, it became evident that certain studies deal with the problem related to the treatment of IPS consumption, where different techniques, models or algorithms are applied in order to create models to present a solution; Table 7 shows the articles related to this problem.

Reference	Techniques, methods and algorithm	Application of the technique, model or algorithm
[26]	•Logistic regression•Decision tree•Random Forest•Extreme gradient boosting•Support vector machine•Artificial neural network	In this study, the mentioned algorithms were used to predict treatment discontinuation due to IPS consumption and also to create interpretable and actionable rules to assist in clinical decision-making.

Reference	Techniques, methods and algorithm	Application of the technique, model or algorithm
[27]	•Random Forest •Logistic regression •Multilayer perceptron network •Support vector machine	For this study, the mentioned algorithms are used to predict treatment outcomes for drug consumption and also to predict psychiatric outcomes.
[29]	•Artificial neural network •Random Forest •XGB	For this case, the algorithms are used to uncover interaction effects for treatment completion.
[25]	•K-Means++ Clustering Algorithm •PCA	For this case, the virtual reality model is first used to capture data from individuals watching videos of IPS treatment. Then, the data is analyzed to group drug addicts accordingly.
[28]	•Furia •ADTree •SMO •REPTree •SPegasos •Decision Table •RBFN •BayesNet •DTNB •BFTree	This article compares the accuracy of predicting treatment outcomes for individuals with alcohol dependence between the predictions made by clinical personnel and the predictions generated by the ML model developed in this study.

Table VII. ML techniques, models, and algorithms used for prevention and treatment programs for IPS consumption.

PI3 What can ML models, techniques and algorithms be used for to identify individuals at risk of developing addiction to illicit psychoactive substances?

Regarding the articles found, some take into account the problem related to people who are at risk of consuming or becoming addicted to IPS, however, each study takes into account different possible causes for those people who are at risk of consuming or becoming addicted to IPS and also each study proposes different ML techniques with their respective application to address this problem, Table 8 shows the techniques, algorithms and models used by each study.

Reference	Techniques, methods and algorithm	Application of the technique, model or algorithm
[38]	•ResNet18 •Word2Vec	ResNet18 is used for image processing, while Word2Vec is used for natural language processing in this study. They are applied to predict the risk of tobacco, alcohol, and IPS consumption among Instagram users.
[56]	•Logistic regression •Least absolute shrinkage and selection operator (Lasso) •Support vector machine (SVM)	In this case, these three techniques are employed to predict individuals at risk of marijuana consumption, taking into account those who have had involvement with Child Welfare Services and those who have not.
[54]	•Least absolute shrinkage and selection operator(Lasso) •Classification and Regression Tree(CART)	LASSO is used in this study to predict variables for marijuana consumption prediction through e-cigarettes, while CART is utilized to predict individuals at risk of marijuana consumption through e-cigarettes or vapes.

[55]	•Logistic regression •Decision tree •Red Neuronal	In this study, the various techniques or models presented are used to predict individuals at risk of marijuana consumption, taking into account their personality traits.
[46]	•Random Forest •Regularized Cox regression	In this case, the techniques, models, or algorithms are used to predict the relapse risk of individuals after undergoing opioid consumption treatment.
[12]	•Random Forest •Decision Tree •Logistic Regression •Linear Support Vector Machine •KNearest Neighbors •Gaussian Naive Bayes	The mentioned algorithms in this study are used to create an approach for identifying individual vulnerability to IPS consumption. This approach aids in preventing individuals who are consuming IPS but are not yet addicted to these substances.

Table VIII. ML techniques, models and algorithms applied in the prediction of people at risk of consuming SPI

B. Analysis

To evaluate the quality of the selected articles, the quality evaluation criteria proposed by [57] was established as a basis, which consists of different scores that are given in a range from -1 to 1 being applied to each of the selected articles (see Table 9).

Criterion	Description	Score		
		-1	**0**	**1**
C1	The study provides a detailed description of the algorithms and methods used.	No	Partially	Yes
C2	The study presents the results in a clear and detailed manner.	No	Partially	Yes
C3	The study presents the results in a clear and detailed manner.	No	1 to 10	+10
C4	The study presents the results in a clear and detailed manner.	No	Q4, Q3	Q2, Q1

Table IX. Quality evaluation criteria for the selected articles.

Using the criteria established above, Table 10 presents the results obtained for each of the articles selected in this systemic mapping.

Reference	C13	C2	C3	C4	Total
[7]	1	1	0	1	3
[8]	0	1	0	-1	0
[9]	1	1	0	1	3
[10]	1	1	0	0	2
[11]	1	0	0	1	2
[12]	1	1	0	1	3

15

[13]	1	1	0	-1	1
[14]	1	1	1	1	4
[15]	1	1	1	1	4
[16]	1	1	1	1	4
[17]	1	1	0	1	3
[18]	1	0	1	1	3
[19]	1	1	0	0	2
[20]	1	1	0	1	3
[21]	1	1	0	-1	1
[22]	1	0	1	1	3
[23]	1	1	1	1	4
[24]	1	1	0	-1	1
[25]	1	0	1	-1	1
[26]	1	1	0	1	3
[27]	1	1	0	1	3
[28]	1	1	1	1	4
[29]	0	1	0	1	2
[30]	0	1	1	1	3
[31]	1	1	0	-1	1
[32]	1	1	-1	-1	0
[33]	1	0	0	1	2
[34]	1	1	1	1	4
[35]	1	1	0	1	3
[36]	1	1	0	1	3
[37]	1	1	1	1	4
[38]	1	1	1	1	4
[39]	1	1	1	-1	2
[40]	1	1	1	1	4
[41]	0	0	0	1	1
[42]	1	1	1	1	4
[43]	1	1	-1	-1	0
[44]	1	1	-1	1	2
[45]	1	1	1	1	4
[46]	1	1	0	1	3
[47]	1	1	-1	0	1
[48]	1	1	0	1	3
[49]	1	0	0	-1	0
[50]	1	1	0	1	3
[51]	1	1	1	1	4
[52]	1	1	0	1	3
[53]	0	1	0	1	2
[54]	0	1	0	1	2
[55]	1	1	0	1	3
[56]	1	1	0	1	3

Table X. Quality results of the articles taking into account the quality criteria.

With the results obtained by applying the quality criteria for each of the selected articles, it can be seen in Figure 4 that 88% of the articles present in detail the algorithms and methods used according to criterion 1 (C1); 86% of the articles present clearly and in detail the results, obtained from criterion 2 (C2); 34% of the articles have been cited more than 10 times by authors, however a higher percentage of 58% of articles have been cited between 1 to 10 times considering criterion 3 (C3); finally 74% of articles have been published in relevant journals and conferences with a quartile of Q1 and Q2 considering criterion 4.
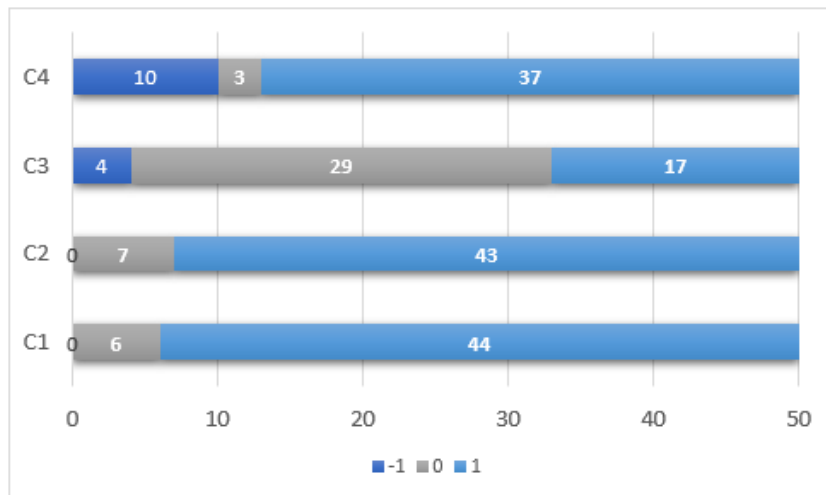

Figure. 4 Evaluation result according to the quality criteria applied to the articles.

In Figure 5 it is necessary to clarify that the larger the bubble, the greater the number of articles that exist between the subject and the year of publication, where the number inside the bubble is equivalent to the number of studies found, therefore, if there is no bubble at an intersection, it will mean that no articles were found and/or selected.

On the other hand, when analyzing Figure 5, it can be observed that for articles with a publication year between 2018 to 2022 most of the studies were found, which helps us to infer that ML applied to IPS consumption is relevant as a study for the last years, however, it is expected that by 2023 the studies related to this subject will increase.

In addition, it should be taken into account that the topics were established according to the issues addressed by each of the selected studies and, when analyzing Figure 5, it can be observed that there is a similarity between the number of articles in each topic, which suggests that ML in IPS has diverse applications for different problems presented.
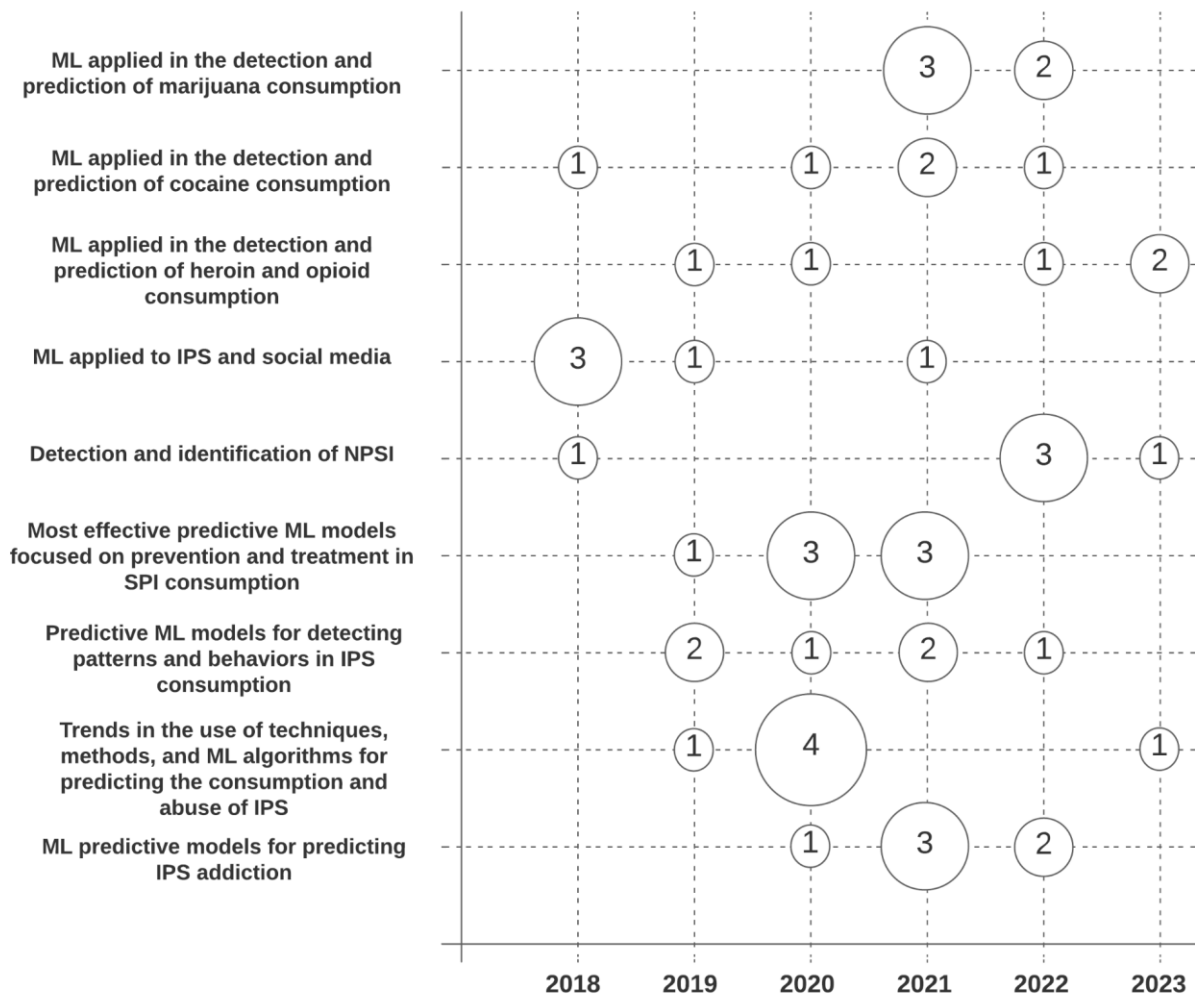
Figure. 5 Bubble chart of the articles found depicting the number of publications by topics and year of publication.

Figure 6 shows a cloud diagram of keywords that were most frequently repeated in the articles collected. Each letter size indicates the number of repetitions, the small size being a minimum number of repetitions and the large size the opposite case. The main ones (highest repetitions) are "Machine Learning", "Predictive models", "Substance abuse", "Substance Use Disorder", "Addiction", among others.

This diagram allows emphasizing the keywords that are common in the articles collected and that are widely used in the subject of consumption of illicit psychoactive substances. Certain topics of interest can also be found within this domain such as "addiction", "treatment outcome", drug and substance abuse in general, specifically "Marijuana", "Cocaine", "Opioid use disorder". Another topic of interest is the algorithms highlighted in the keywords, one of them being "Random Forest".

Figure. 6 Keyword word cloud

D. *LIMITATIONS AND CHALLENGES*

Table 11 presents the different limitations and challenges encountered in the studies selected for this systemic mapping, keeping in mind that each of the challenges and limitations presented are general for the different ML applications.

Reference	Limitations and challenges
[47]	It is necessary to create and industrialize different types of plantations other than coffee plantations, which can serve as alternatives for farmers and reduce illegal plantations.
[50]	Research using machine learning algorithms may be limited by overfitting and generalization.
[41]	Tools could be developed to increase the speed of analysis, considering ML approaches such as text classification, image analysis, and other metadata.
[38]	Class imbalance is one of the most common problems in ML, and further research is needed on this topic as it affects the effectiveness of ML models.
[42]	Adding applicability to ML models can improve results and adoption possibilities.
[23]	When focusing on a specific substance in a study, the change in preference for other substances is not taken into account. For example, switching from marijuana to a more harmful substance is common. Therefore, a study covering a wide range of substances with the quantification of associated harm for each is proposed.

[22]	In addiction studies, previous research works could be divided based on the type of addiction, differentiating between cigarette smoking, alcohol, cocaine, opioids, and multiple substance use.

Table XI. Limitations and challenges in ML applied to the treatment and consumption of psychoactive substances.

V. CONCLUSIONS

In this systemic mapping, approximately 50 articles related to the use of ML in the consumption of psychoactive substances were analyzed, in each of the studies different solutions were presented to problems such as prediction of SP consumption, prediction for treatment for SP consumption, prediction of people who are at risk of consuming IPS, prediction of NIPS, among others, where different models, techniques, algorithms and neural networks are used to develop predictive models based on ML; It should be noted that each of the solutions presented help to understand the current state of the art, in order to identify the best techniques.

In addition, many of the studies analyzed demonstrate the effectiveness and success of the ML models, techniques and algorithms used by each(articles), helping to understand the boom that AI is currently having, since AI as such is being used or applied in various problems, in order to support and help professionals such as doctors, engineers, among others.

It is noteworthy that although there are limitations and challenges to the application of ML models, techniques and algorithms, the advancement of technology will help to eliminate each of the limitations that currently exist, however, for now it is necessary to understand and find the best techniques, models and algorithms by understanding their respective application in order to find where the techniques, models and algorithms perform best.

With the above-mentioned and taking into account the results obtained in this systemic mapping, it should be clarified that the different applications of ML in the consumption of IPS may vary according to the place, where factors such as culture, the habits of the country where the problem (consumption of IPS) occurs and different environmental factors cause the results to change.

Finally, it became evident that there are various applications of ML in the consumption of IPS, such as to predict whether people need treatment for consumption, to evaluate and identify those people who may be at risk of consuming IPS, among others, where demographic data, socioeconomic data, data on consumption, cultural data, among other data are taken into account, among other data, which help and contribute to the training and improvement of the models that have been and will be developed, thus providing a useful tool for both health and government entities, in order to provide support to create prevention and treatment programs for people who already consume or who are at risk of consuming IPS.

REFERENCES

[1] Organización Panamericana de Salud, "Abuso de sustancias - OPS/OMS | Organización Panamericana de la Salud." https://www.paho.org/es/temas/abuso-sustancias (accessed May 02, 2023).

[2] ONU, "El Informe Mundial sobre las Drogas 2022 de la UNODC destaca las tendencias del cannabis posteriores a su legalización, el impacto ambiental de las drogas ilícitas y el consumo de drogas entre las mujeres y las personas jóvenes," 2022. https://www.unodc.org/unodc/es/press/releases/2022/June/unodc-world-drug-report-2022-highlights-trends-on-cannabis-post-legalization--environmental-impacts-of-illicit-drugs--and-drug-use-among-women-and-youth.html (accessed Jun. 06, 2023).

[3] Ministerio de Salud, "Política Integral para la Prevención y Atención del Consumo de Sustancias Psicoactivas," 2019.

[4] H. Zhou, J. Tang, and H. Zheng, "Machine Learning for Medical Applications," Scientific World Journal, vol. 2015. Hindawi Publishing Corporation, 2015. doi: 10.1155/2015/825267.

[5] D. Carrizo and J. Rojas, "Classification of requirements elicitation practices in agile developments: a systematic mapping," 2016.

[6] B. Kitchenham, "Procedures for Performing Systematic Reviews," 2004.

[7]   X. Gu, B. Yang, S. Gao, L. F. Yan, D. Xu, and W. Wang, "Application of bi-modal signal in the classification and recognition of drug addiction degree based on machine learning," Mathematical Biosciences and Engineering, vol. 18, no. 5, pp. 6926–6940, 2021, doi: 10.3934/MBE.2021344.

[8]   M. M. Hassan, J. H. Angon, Z. J. Peya, A. I. Keya, S. Zaman, and A. U. Dulla, "A Machine Learning Approach to Identify the Correlation and Association among the Students' Drug Addict Behavior," 2020.

[9]   M. Nesa, T. R. Shaha, and Y. Yoon, "Prediction of juvenile crime in Bangladesh due to drug addiction using machine learning and explainable AI techniques," J Comput Soc Sci, vol. 5, no. 2, pp. 1467–1487, Nov. 2022, doi: 10.1007/s42001-022-00175-7.

[10]   M. A. I. Arif, S. I. Sany, F. Sharmin, M. S. Rahman, and M. T. Habib, "Prediction of addiction to drugs and alcohol using machine learning: A case study on Bangladeshi population," International Journal of Electrical and Computer Engineering, vol. 11, no. 5, pp. 4471–4480, Oct. 2021, doi: 10.11591/ijece.v11i5.pp4471-4480.

[11]   H. Gong, C. Xie, C. Yu, N. Sun, H. Lu, and Y. Xie, "Psychosocial factors predict the level of substance craving of people with drug addiction: A machine learning approach," Int J Environ Res Public Health, vol. 18, no. 22, Nov. 2021, doi: 10.3390/ijerph182212175.

[12]   U. I. Islam, E. Haque, D. Alsalman, M. N. Islam, M. A. Moni, and I. H. Sarker, "A Machine Learning Model for Predicting Individual Substance Abuse with Associated Risk-Factors," Annals of Data Science, 2022, doi: 10.1007/s40745-022-00381-0.

[13]   N. F. Zulkifli, Z. C. Cob, A. A. Latif, and S. M. Drus, "A Systematic Review of Machine Learning in Substance Addiction," in 2020 8th International Conference on Information Technology and Multimedia, ICIMU 2020, Institute of Electrical and Electronics Engineers Inc., Aug. 2020, pp. 103–107. doi: 10.1109/ICIMU49871.2020.9243581.

[14]   J. A. Mollick and H. Kober, "Computational models of drug use and addiction: A review," J Abnorm Psychol, vol. 129, no. 6, pp. 544–555, Aug. 2020, doi: 10.1037/abn0000503.

[15]   Y. Boo and Y. Choi, "Comparing logistic regression models with alternative machine learning methods to predict the risk of drug intoxication mortality," Int J Environ Res Public Health, vol. 17, no. 3, Feb. 2020, doi: 10.3390/ijerph17030897.

[16]   K. K. Mak, K. Lee, and C. Park, "Applications of machine learning in addiction studies: A systematic review," Psychiatry Research, vol. 275. Elsevier Ireland Ltd, pp. 53–60, May 01, 2019. doi: 10.1016/j.psychres.2019.03.001.

[17]   B. Chhetri, L. M. Goyal, and M. Mittal, "How machine learning is used to study addiction in digital healthcare: A systematic review," International Journal of Information Management Data Insights, vol. 3, no. 2. Elsevier B.V., Nov. 01, 2023. doi: 10.1016/j.jjimei.2023.100175.

[18]   E. Barenholtz, N. D. Fitzgerald, and W. E. Hahn, "Machine-learning approaches to substance-abuse research: emerging trends and their implications," Current opinion in psychiatry, vol. 33, no. 4. NLM (Medline), pp. 334–342, Jul. 01, 2020. doi: 10.1097/YCO.0000000000000611.

[19]   A. Anggrawan, N. Gusti Ayu Dasriani, C. Satria, and C. Ku Nuraini, "Machine Learning for Diagnosing Drug Users and Types of Drugs Used." [Online]. Available: www.ijacsa.thesai.org

[20]   R. Vunikili, B. S. Glicksberg, K. W. Johnson, J. T. Dudley, L. Subramanian, and K. Shameer, "Predictive Modelling of Susceptibility to Substance Abuse, Mortality and Drug-Drug Interactions in Opioid Patients," Front Artif Intell, vol. 4, Dec. 2021, doi: 10.3389/frai.2021.742723.

[21]   B. Rekabdar, D. Albright, J. T. McDaniel, S. Talafha, and H. Jeong, "From machine learning to deep learning: A comprehensive study of alcohol and drug use disorder," Healthcare Analytics, vol. 2, Nov. 2022, doi: 10.1016/j.health.2022.100104.

[22]   Y. Jing et al., "Title: Analysis of substance use and its outcomes by machine learning I. Childhood Evaluation of Liability to Substance Use Disorder," 2019. [Online]. Available: http://dx.doi.org1://www.elsevier.com/open-access/userlicense/1.0/VersionofRecord:https://www.sciencedirect.com/science/article/pii/S0376871619303825

[23]   Z. Hu et al., "Analysis of substance use and its outcomes by machine learning: II. Derivation and prediction of the trajectory of substance use severity," Drug Alcohol Depend, vol. 206, Jan. 2020, doi: 10.1016/j.drugalcdep.2019.107604.

[24]   U. I. Islam, I. H. Sarker, E. Haque, and M. M. Hoque, "Predicting Individual Substance Abuse Vulnerability using Machine Learning Techniques," Dec. 2020, [Online]. Available: http://arxiv.org/abs/2101.03184

[25]   Y. Yuan, J. Huang, and K. Yan, "Virtual reality therapy and machine learning techniques in drug addiction treatment," in Proceedings - 10th International Conference on Information Technology in Medicine and

Education, ITME 2019, Institute of Electrical and Electronics Engineers Inc., Aug. 2019, pp. 241–245. doi: 10.1109/ITME.2019.00062.

[26] M. M. Hasan et al., "A machine learning based two-stage clinical decision support system for predicting patients' discontinuation from opioid use disorder treatment: retrospective observational study," BMC Med Inform Decis Mak, vol. 21, no. 1, Dec. 2021, doi: 10.1186/s12911-021-01692-7.

[27] J. Tapia-Galisteo et al., "Prediction of Cocaine Inpatient Treatment Success Using Machine Learning on High-Dimensional Heterogeneous Data," IEEE Access, vol. 8, pp. 218936–218953, 2020, doi: 10.1109/ACCESS.2020.3041895.

[28] M. Symons, G. F. X. Feeney, M. R. Gallagher, R. M. D. Young, and J. P. Connor, "Machine learning vs addiction therapists: A pilot study predicting alcohol dependence treatment outcome from patient data in behavior therapy with adjunctive medication," J Subst Abuse Treat, vol. 99, pp. 156–162, Apr. 2019, doi: 10.1016/j.jsat.2019.01.020.

[29] M. Nasir, N. S. Summerfield, A. Oztekin, M. Knight, L. K. Ackerson, and S. Carreiro, "Machine learning-based outcome prediction and novel hypotheses generation for substance use disorder treatment," Journal of the American Medical Informatics Association, vol. 28, no. 6, pp. 1216–1224, Jun. 2021, doi: 10.1093/jamia/ocaa350.

[30] S. W. Yip, B. Kiluk, and D. Scheinost, "Toward Addiction Prediction: An Overview of Cross-Validated Predictive Modeling Findings and Considerations for Future Neuroimaging Research," Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, vol. 5, no. 8. Elsevier Inc, pp. 748–758, Aug. 01, 2020. doi: 10.1016/j.bpsc.2019.11.001.

[31] Jose Tapia Galisteo, "Propuesta de modelos predictivos en salud mental para la personalización de terapias de rehabilitación en pacientes con adicciones," 2021.

[32] Y. Kobayashi and K. Yoshida, "Automated retention time prediction of new psychoactive substances in gas chromatography," in Procedia Computer Science, Elsevier B.V., 2022, pp. 654–663. doi: 10.1016/j.procs.2022.09.120.

[33] J. Klingberg, B. Keen, A. Cawley, D. Pasin, and S. Fu, "Developments in high-resolution mass spectrometric analyses of new psychoactive substances," Archives of Toxicology, vol. 96, no. 4. Springer Science and Business Media Deutschland GmbH, pp. 949–967, Apr. 01, 2022. doi: 10.1007/s00204-022-03224-2.

[34] E. Olesti et al., "Metabolomics predicts the pharmacological profile of new psychoactive substances," Journal of Psychopharmacology, vol. 33, no. 3, pp. 347–354, Mar. 2019, doi: 10.1177/0269881118812103.

[35] S. L. Wong, L. T. Ng, J. Tan, and J. Pan, "Screening unknown novel psychoactive substances using GC–MS based machine learning," Forensic Chemistry, vol. 34, Jul. 2023, doi: 10.1016/j.forc.2023.100499.

[36] M. Han et al., "AddictedChem: A Data-Driven Integrated Platform for New Psychoactive Substance Identification," Molecules, vol. 27, no. 12, Jun. 2022, doi: 10.3390/molecules27123931.

[37] J. Li, Q. Xu, N. Shah, and T. K. Mackey, "A machine learning approach for the detection and characterization of illicit drug dealers on instagram: Model evaluation study," J Med Internet Res, vol. 21, no. 6, Jun. 2019, doi: 10.2196/13803.

[38] S. Hassanpour, N. Tomita, T. DeLise, B. Crosier, and L. A. Marsch, "Identifying substance use risk based on deep neural networks and Instagram social media data," Neuropsychopharmacology, vol. 44, no. 3, pp. 487–494, Feb. 2019, doi: 10.1038/s41386-018-0247-x.

[39] U. Kursuncu et al., "What's ur Type? Contextualized Classification of User Types in Marijuana-Related Communications Using Compositional Multiview Embedding," in Proceedings - 2018 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2018, Institute of Electrical and Electronics Engineers Inc., Jan. 2019, pp. 474–479. doi: 10.1109/WI.2018.00-50.

[40] T. Mackey, J. Kalyanam, J. Klugman, E. Kuzmenko, and R. Gupta, "Solution to detect, classify, and report illicit online marketing and sales of controlled substances via twitter: Using machine learning and web forensics to combat digital opioid access," J Med Internet Res, vol. 20, no. 4, Apr. 2018, doi: 10.2196/10029.

[41] N. Shah, J. Li, and T. K. Mackey, "An unsupervised machine learning approach for the detection and characterization of illicit drug-dealing comments and interactions on Instagram," Subst Abus, vol. 43, no. 1, pp. 273–277, 2022, doi: 10.1080/08897077.2021.1941508.

[42] D. H. Han, S. Lee, and D. C. Seo, "Using machine learning to predict opioid misuse among U.S. adolescents," Prev Med (Baltim), vol. 130, Jan. 2020, doi: 10.1016/j.ypmed.2019.105886.

[43] Z. Jianqiang, N. Chunming, P. Chengyun, H. Junxun, G. Sheng, and C. Jin, "Rapid Recognition of Different Sources of Heroin Drugs by Using a Hand-Held Near-Infrared Spectrometer Based on a Multi-Layer Extreme Learning Machine Algorithm," J Braz Chem Soc, 2023, doi: 10.21577/0103-5053.20220120.

[44] H. Feng, R. Elladki, J. Jiang, and G.-W. Wei, "Machine-learning Analysis of Opioid Use Disorder Informed by MOR, DOR, KOR, NOR and ZOR-Based Interactome Networks," Jan. 2023, [Online]. Available: http://arxiv.org/abs/2301.04815

[45] J. T. Prieto et al., "The detection of opioid misuse and heroin use from paramedic response documentation: Machine learning for improved surveillance," J Med Internet Res, vol. 22, no. 1, 2020, doi: 10.2196/15645.

[46] J. P. Davis et al., "Identifying individual and environmental predictors of opioid and psychostimulant use among adolescents and young adults following outpatient treatment," Drug Alcohol Depend, vol. 233, Apr. 2022, doi: 10.1016/j.drugalcdep.2022.109359.

[47] D. Belén et al., "A machine learning approach to find the determinants of Peruvian coca illegal crops," Decision Science Letters, vol. 11, pp. 127–136, 2022, doi: 10.5267/dsl.2021.12.003.

[48] H. Feng, K. Gao, D. Chen, A. J. Robison, E. Ellsworth, and G.-W. Wei, "Machine learning analysis of cocaine addiction informed by DAT, SERT, and NET-based interactome networks," Dec. 2021, [Online]. Available: http://arxiv.org/abs/2201.00114

[49] K. Gao, D. Chen, A. J. Robison, and G. W. Wei, "Proteome-Informed Machine Learning Studies of Cocaine Addiction," Journal of Physical Chemistry Letters, vol. 12, no. 45, pp. 11122–11134, Nov. 2021, doi: 10.1021/acs.jpclett.1c03133.

[50] R. Suchting, J. N. Vincent, S. D. Lane, C. E. Green, J. M. Schmitz, and M. C. Wardle, "Using a data science approach to predict cocaine use frequency from depressive symptoms," Drug Alcohol Depend, vol. 194, pp. 310–317, Jan. 2019, doi: 10.1016/j.drugalcdep.2018.10.029.

[51] R. F. Kranenburg et al., "Rapid and robust on-scene detection of cocaine in street samples using a handheld near-infrared spectrometer and machine learning algorithms," Drug Test Anal, vol. 12, no. 10, pp. 1404–1418, Oct. 2020, doi: 10.1002/dta.2895.

[52] J. Choi, J. Chung, and J. Choi, "Exploring impact of marijuana (Cannabis) abuse on adults using machine learning," Int J Environ Res Public Health, vol. 18, no. 19, Oct. 2021, doi: 10.3390/ijerph181910357.

[53] T. Parekh and F. Fahim, "Building risk prediction models for daily use of marijuana using machine learning techniques," Drug Alcohol Depend, vol. 225, Aug. 2021, doi: 10.1016/j.drugalcdep.2021.108789.

[54] D. H. Han and D. C. Seo, "Identifying risk profiles for marijuana vaping among U.S. young adults by recreational marijuana legalization status: A machine learning approach," Drug Alcohol Depend, vol. 232, Mar. 2022, doi: 10.1016/j.drugalcdep.2022.109330.

[55] L. Zoboroski, T. Wagner, and B. Langhals, "Classical and neural network machine learning to determine the risk of marijuana use," Int J Environ Res Public Health, vol. 18, no. 14, Jul. 2021, doi: 10.3390/ijerph18147466.

[56] S. Negriff, B. Dilkina, L. Matai, and E. Rice, "Using machine learning to determine the shared and unique risk factors for marijuana use among child-welfare versus community adolescents," PLoS One, vol. 17, no. 9 September, Sep. 2022, doi: 10.1371/journal.pone.0274998.

[57] J. Homepage, N.-E. Quemá-Taimbud, M.-E. Mendoza-Becerra, and O.-F. Bedoya-Leyva, "Revista Facultad de Ingeniería Initialization and Local Search Methods Applied to the Set Covering Problem: A Systematic Mapping," Revista Facultad de Ingeniería (Rev. Fac. Ing, vol. 32, no. 63, 2023, doi: 10.19053/01211129.v32.n63.2023.15235.