Relationship between Demographic Factors and the Misuse of Recreational Drugs among the Students of the Faculty of Science of the University of Kelaniya

Thanuja Dewanmini (1st Author)
Department of Statistics and Computer Science, Faculty of Science
University of Kelaniya
Paliyagoda, Sri Lanka

Harshana Munasinghe (2nd Author)
Department of Statistics and Computer Science
Faculty of Science
University of Kelaniya
Paliyagoda, Sri Lanka
harshanam@kln.ac.lk

Abstract — **Drug abuse has been identified as a global health issue. It can be identified as one of the root causes of destroying peoples' lives and souls. This has a huge impact on their academic performance as well as the productivity of their lives. This study has explored the factors affecting the misuse of recreational drugs among university students. The main objective of the study was to identify the relationship between demographic factors and the misuse of recreational drugs among the students of the Faculty of Science of the University of Kelaniya. Data were collected through an online questionnaire survey. Snowball sampling was used as the sampling technique. A descriptive analysis was performed and a binary logistic regression, Support Vector Machine (SVM), and Probabilistic Neural Network (PNN) were used to evaluate the best statistical model to predict drug usage. Participants included 220 students from the Faculty of Science of the University of Kelaniya. The descriptive analysis showed that male drug usage was higher than female drug usage. Also, the drug consumption of the third-year students was higher than the other students. Students whose parents were illiterate showed a higher value of drug consumption than the other students. Also, the drug usage among the students who lived in the hostel was significantly higher than the others. It was also revealed that the most commonly used drug is alcohol. Among the fitted models, SVM Non-Random spilt model showed the highest accuracy (93.1818%) in predicting drug usage. Based on the results, gender, religion, year of study, involvement of a part-time job, participation in the sports activities, financial support from bursary or mahapola, mother's education level, father's occupation, ethnicity, marital status of the parents were identified as the associated factors of the drug usage among the students.**

Keywords — Alcohol, Drug abuse, Drug usage

I. INTRODUCTION

Human beings have had a deep connection to their social environment since past time, and have therefore maintained an intimate, cooperative, and long-term relationship with one another. As can be seen throughout history, people adapted and coordinated themselves to achieve their common goals, and this generally resulted in positive and desired outcomes. Globalization has fundamentally altered the essence of traditional society, transforming it into a tangle of contradictions unlike anything seen before in any conventional society. These changes and complexities have resulted in, and continue to result in, a variety of social issues among citizens. Teenage pregnancy, drug addiction, poverty, domestic violence, child abuse, trafficking, traffic accidents, divorce, and crime are only a few examples of today's social issues, many of which have negative implications for individuals and communities. Drug abuse has recently emerged as one of the world's most significant social issues. This problem hasn't an age limit. In every age limit, people are addicted to this matter. Substance abuse among youth is a huge problem. University students are facing this problem at that moment. The majority of drug users will argue that using drugs is a personal matter and that their actions affect only them and not others. Everyone knows, however, that it is not as easy as the drug addicts say. It is a well-known fact that drug addiction influences not only drug users, but also their families, culture, and world. Drug abuse is not a straightforward issue that just affects a few people. It is a multi-faceted, all-encompassing social phenomenon that affects almost every country in the world. Drug misuse and addiction are not new occurrences. Millions of people all over the world have lost control of their lives as a result of drug addiction. In particular, illegal drug use has increased dramatically across the world in recent decades. Many people have begun to use drugs without realizing the harmful effects that these substances can have on their own lives and the lives of their dependents. In the origin, People only used drugs on rare occasions at first, and they were capable of handling themselves in unusual circumstances. However, the scene has gradually changed as drugs have taken hold of people's

lives and made them believe they could not survive without them. Drug abuse has enslaved a large portion of the population in many countries around the world today, which is tragic. Drug addiction has now been added to the list of social problems, and it is particularly noticeable among the poor, and working population of Sri Lanka.[1]

This research has explored the Relationship between Demographic Factors and Misuse of Recreational Drugs among the Students of the Faculty of Science of the University of Kelaniya. The demographic factors of the students were considered for the analysis. Namely, Gender, Ethnicity, Religion, Accommodation place, Marital status of parents, Life status of parents, Father's Occupation, Father's Education Level, Mother's Education Level, Scholarship status, Engagement of the part-time jobs, Engagement in sports activities, Accommodation place and year of study at university. These factors were identified by studying the relevant research papers. These variables were used to find the determinants of drug usage and to develop a statistical model to predict drug usage

II. OBJECTIVES

The main objective of this study was to identify the relationship between demographic factors and misuse of recreational drugs among university students. Also, this study helped to identify the present situation of drug addiction among the Students of the Faculty of Science of the University of Kelaniya, Which year university students more use drugs, Investigate the drug abuse of students who participates in sports, Investigate the drug abuse of students who do a part time job, Develop a statistical model to predict the drug usage, Identify the types of drug currently being used by university students, their involvement in multiple drug misuse and drugs combinations, To determine the frequency of drug abuse in a sample of a university student. And also, the demographic variables were used to find the determinants of the drug usage and to develop a statistical model to predict the drug usage.

III. METHODOLOGY

In this study gender, ethnicity, religion, marital status of parents, life status of parents, father education, mother education, father occupation, accommodation, year, the status of the scholarships (Mahapola/bursary), engagement in a part-time job, and engagement in sports were used to find out the influence on the drug usage of the university students. All the outputs of the research are obtained with the use of Minitab, SPSS, and MATLAB software. A sample of 220 students who are studying at the Faculty of Science of the University of Kelaniya was used in this research. The sample was obtained with the use of the Snowball Sampling procedure. A descriptive analysis was carried out to check the composition of the sample. Correlation analysis was conducted using the Chi-square test in order to find out the most influential factors of drug usage. After checking the association between the independent variables and the dependent variable, the following variables were identified as they were having a

significant association with the response variable. Gender, Religion, Father's Occupation, Year of study, Engagement in a part-time job, Engagement in sports activities. These variables were used to fit the binary logistic regression model in order to predict drug usage. Since the VIF for all independent variable were less than 10 multicollinearity was identified among the independent variables. Support Vector Machine (SVM) random split and nonrandom split data divide method and Probabilistic Neural Network (PNN) random split and nonrandom split data divide methods were used to evaluate the best statistical model to predict drug usage. The polynomial kernel function has been chosen as the optimal kernel function in the SVM. 0.1 was the optimal spread parameter in the PNN. Model diagnostic analysis was carried out to check the accuracy of all the fitted models.

IV. RESULTS AND DISCUSSION

According to the fitted binary logistic regression model, male drug usage is 2.008 higher than the females. For students who are doing a part-time job their drug consumption value is higher at 1.564 than the students who are not doing a part-time job. Drug consumption of the students who are engaged in Sports is 0.493 times higher than a student who is not doing sports. Third-year students' drug consumption and second-year students' drug composition are higher than the fourth-year students' drug composition. As a value, it is 1.289 times and 0.066 times respectively. Third-year students' drug consumption is 1.494 times higher than first-year students. Drug usage of Hindu students is 2.946 times higher than the drug usage of Muslim students and 1.411 higher than Buddhists. Christian students' drug consumption is 2.079 times higher than muslin students' drug consumption. The students whose father works in the private sector have used drugs 1.328 times higher than students whose father is Unemployed (base category).

According to the Table 1, we can conclude that the Gender, Part-time job variable (Do you do any part-time job?), Year 3, Hinduism (Religion), variables are significant in the fitted model. Since they are positively correlated, they have a positive association with students' drug usage after controlling other variables.

Based on the results of binary logistic analysis Gender, part-time job, Year 3, Hinduism (Religion) variables are the only influenced factor for drug usage from the final model with the accuracy of 79.1% correctly classified.

Table 1: Variable in the Equation Table

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1ᵃ	G1	2.008	.463	18.808	1	.000	7.451
	R1	1.535	.851	3.251	1	.071	4.641
	R2	2.079	1.126	3.410	1	.065	7.996
	R3	2.946	1.152	6.537	1	.011	19.031
	P1	1.564	.476	10.812	1	.001	4.780
	S1	.493	.387	1.625	1	.202	1.637
	F1	-.271	.945	.082	1	.775	.763
	F2	1.328	.934	2.020	1	.155	3.772
	F3	1.095	.941	1.353	1	.245	2.988
	F4	.323	.926	.122	1	.727	1.381
	Y1	-.205	.609	.114	1	.736	.814
	Y2	.066	.587	.013	1	.911	1.068
	Y3	1.289	.581	4.926	1	.026	3.630
	Constant	-4.922	1.411	12.162	1	.000	.007

a. Variable(s) entered on step 1: G1, R1, R2, R3, P1, S1, F1, F2, F3, F4, Y1, Y2, Y3.

Then fitted the support vector machine to the data set. First, we divide the data set-in two-part training set and testing set. In the analysis below the two sets are divided as 80% training set and 20% testing set. Artificial Neural network modeling has been carried out using MATLAB mathematical software. To measure the performance using a classification matrix. First, we divide the whole data set nonrandom split. This data set has fourteen variables. Firstly, label the demographic variable Gender(G), Ethnicity(E), Religion(R), Marital Status of Parents (MS), Life Status of Parents (LS), Father Education (FE), Mother Education(ME), Father Occupation (FO), Accommodation(A), Bursary/Mahapola (BM), Part-time job(P), Sports(S), Year(Y). Then remove one by one variable and checked the accuracy. Then checked the most accurate input combination by removing one by one variable out of fourteen variables. Father Occupation, Part time job, Mother Education, Religion, Marital Status, Sport, Year, Gender, Ethnicity and student have a bursary or mahapola variables were included most accuracy input combination model. Then did the parameter adjustment of this model. There are many Kernel Functions that can be used in support vector machine. Polynomial, Gaussian, Rbf (Radial Basis Function), Linear kernel are the most commonly used Kernel Function. When the Kernel Function change while all other factors are Keep Constant. Gaussian Kernel Function and Rbf Kernel Function have the almost same performance in this SVM. All the other kernel function performance was lesser than the polynomial kernel function performance. Therefore, the polynomial Kernel function has been chosen as the kernel function in the SVM. After changing the kernel function in the SVM final model has included the polynomial kernel function and Father Occupation, Part time job, Mother Education, Religion, Marital Status, Sport, Year, Gender, Ethnicity and student have a bursary or mahapola input variables.

Table 2. Overall performance of SVM Non-Random Split model

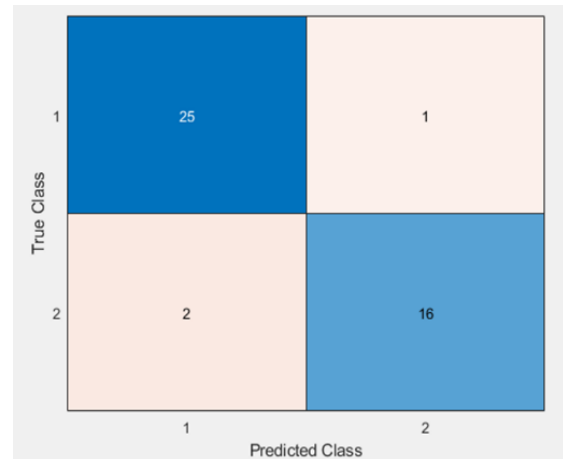| Accuracy	Misclassification	Precision	Recall	F-measure
93.1818	6.82	0.89	0.94	0.91

Fig. 1. Classification Chart

1-Not Use Drug, 2- Drug Use

This is the classification matrix of the optimum model Out of the training data set, the model predicts the 25 students are not using the drug. They are actually not drug use students. And also, the model predicts that 16 students are used drug. They are actually drug used students. The model predicts the two students are not used drugs. But actually, they have used drugs. And also, the model predicts the one student who used a drug. But actually, this student not used the drug

Then checked the SVM model by using the random split data dividen method. This method firstly identified the most accurate input combination. Removed one by one variable in the full model and checked the accuracy. Then we identified the less accurate variable in the full model. Accommodation(A), Life Status of Parents (LS), Ethnicity(E), Father Occupation (FO), Marital Status Parents (MS) variable have the lowest association in this SVM random split data dividen model. Then removed the lowest association variables combination. Checked the highest accuracy model. Father Occupation (FO), Marital Status of Parents (MS), Religion(R), Mother Education (ME), Sport(S), Student who have Bursary or Mahapola (BM), Part-time job(P), Gender(G), Year(Y) Variables were included highest accuracy random split support vector machine model. Then did the parameter adjustment of this model. There are many Kernel Functions that can be used in support vector machine. Polynomial, Gaussian, Rbf (Radial Basis Function), Linear kernel are the most commonly used Kernel Function. When the Kernel Function change while all other factors are Keep Constant. Gaussian Kernel Function and Rbf Kernel Function have the almost same performance in this SVM. All the other kernel function performances are lesser than the polynomial kernel function performance. Therefore, the polynomial Kernel function has been chosen as the kernel function in the SVM. After changing the kernel function in the SVM final model has included the polynomial kernel function and Father Occupation, Part time job, Mother Education, Religion, Marital Status, Sport, Year, Gender

and student have a bursary or mahapola input variables. The below table shows the performance statistic of the fitted model.

Table 3. Overall performance of SVM Random Split model

Accuracy	Misclassification	Precision	Recall	F-measure
77.95	22.05	0.82	0.82	0.82

Then fitted the Probabilistic Neural Network model to the data set. First, we divide the data set-in two-part training set and testing set. In the analysis below the two sets are divided as 80% training set and 20% testing set. Artificial Neural network modeling has been carried out using MATLAB mathematical software. To measure the performance using a classification matrix.

This data set have fourteen variables. In this section to identify what are optimum input combination. Firstly, label the demographic variable Gender(G), Ethnicity(E), Religion(R), Marital Status of Parents (MS), Life Status of Parents (LS), Father Education (FE), Mother Education (ME), Father Occupation (FO), Accommodation(A), Bursary/Mahapola (BM), Part-time job(P), Sports(S), Year(Y)these inputs were presented to the network in different combinations and the optimum combination of inputs. Performance measures were used to identify the best input combination. All the other factors and parameters in the network were kept constant during the training process. Marital Status of Parents, Sports, Gender, Religion, Life Status of Parents, Student have Bursary or Mahapola, Year these inputs variables were identified highest accuracy model. PNN has only one parameter adjustment. It is the spread parameter. After getting the optimum input combination model then we changed the spread parameter. The spread parameter should be near zero. Because The network performs the function of the closest neighbor classifier. As the spread widens, the designed network considers many nearby design vectors.

Table 4. Performance of different Spread Parameter in PNN For Non-Random Split

Spread	Accuracy	Misclassification	Precision	Recall	F-measure
0.1	90.91	9.09	0.93	0.82	0.88
0.2	90.91	9.09	0.88	0.88	0.88
0.3	90.91	9.09	0.88	0.88	0.88
0.4	90.91	9.09	0.88	0.88	0.88
0.5	90.91	9.09	0.88	0.88	0.88

According to the above table, we can conclude that the optimal model is the first model. This model spread parameter 0.1 and it has the highest Precision than the other model. All the models have the same accuracy and misclassification rate. After increasing the spread parameter to 0.1 all the performance values are the same.

Changing the spread parameter in the PNN final model has included the polynomial 0.1 spread parameter and Mother Education, Religion, Life Status of Parents, Sport, Year, Gender and student have a bursary or mahapola input variables. The below table shows the performance statistic of the fitted model.

Table 5. Overall performance of PNN Non-Random Split model

Accuracy	Misclassification	Precision	Recall	F-measure
90.9091	9.09	0.93	0.82	0.88

Fig. 2. Architecture of the PNN for Random Split data set

Fig.3. Confusion Chart
1-Not Use Drug, 2- Drug Use

This is the classification matrix of the optimum model Out of the training data set, the model predicts the 26 students are not using the drugs. They are remove actually not drug use students. And also, the model predicts that 14 students are used drugs. They are actually drug used students. The model predicts the three students are not used drug. But actually, they have used drugs. And also, the model predicts the one student who used a drugs. But actually, this student not used the drug.

Then fitted the PNN model by using the random data split method. This method firstly identified the most accurate input combination. Removed one by one variable

in the full model and checked the accuracy. Accommodation(A), Part time job(P), Year(Y), Sport(S), Religion(R), Marital Status of Parents (MS), Gender(G) and Year(Y) Variables were included highest accuracy model. This model includes only eight variables out of fourteen variables.

Table 6. Performance of different Spread Parameter in PNN For Random Split

Spread	AA	AM	AP	AR	AF
0.1	76.82	23.18	0.82	0.80	0.81
0.2	76.82	23.18	0.83	0.79	0.81
0.3	73.18	26.82	0.81	0.76	0.78
0.4	73.18	26.82	0.80	0.78	0.78

According to the above table, we can conclude that the optimum spread parameter 0.1. This model has high Average Accuracy (AA), Average Misclassification (AM), Average Precision (AP), Average Recall (AR)and Average F-measure (AF).

Changing the spread parameter in the PNN final model has included the polynomial 0.1 spread parameter and Accommodation, Religion, Marital Status of Parents, Sport, Year, Gender, Part-time job and Farther Education input variables. The below table shows the performance statistic of the fitted model.

Table 7. Overall performance of PNN Random Split model

Accuracy	Misclassification	Precision	Recall	F-measure
76.82	23.18	0.82	0.80	0.81

According to the above four models, we can conclude that the Non-Random split method model accuracy is higher than the Random split method. support vector machine non-random split method accuracy (93.1818%) is higher than the SVM random split (77.9545%), probabilistic neural network random split (90.9091%) and non-random split method (76.8182%). According to the SVM random and non-random split method, both two models include the same dependent variable there is father occupation, part-time job, mother education, religion, marital status, sports, year, gender and student have a bursary or mahapola. But SVM non-random split has included one extra variable it is student ethnicity. And also, according to the probabilistic neural network random split and non-random split method, both models have the same dependent variable there are religion, sport, year and gender. Then we consider all four models have the same dependent variable there are religion, sport, year, gender variables. Finally, we conclude that the most suitable and accurate model support vector machine (SVM) non-random split data dividend method. It has high accuracy and a low misclassification rate. It is correctly classified than the other model, a student who have used drug and a student who has not used.

Table 8. Model Comparison

Model	Accuracy
Binary Logistic Model	79.1%
Support Vector Machine Random Split Model	77.95%
Support Vector Machine Non-Random Split Model	93.18%
Probabilistic Neural Network Random Split	76.83%
Probabilistic Neural Network Non-Random Split	90.91%

According to the Binary Logistic, SVM and PNN non-random split statistical model, SVM Non-Random spilt model is the highest accuracy (93.18%) model to classify the drug usage. Based on the result, gender, religion, year, does the student do any part-time job? Do the student participants in the sports activities? does the student have a bursary or Mahapola? Mother's Education Level, Father's Occupation, Ethnicity, Marital Status of Parents were in influenced factors for student drug usage.

V. CONCLUSION

In this research, I have studied the Relationship between Demographic Factors and Misuse of Recreational Drugs among the Students of the Faculty of Science of the University of Kelaniya. The demographic factors of the students were considered for the analysis. Namely, gender, ethnicity, religion, accommodation place, marital status of parents, life status of parents, father's Occupation, Father's Education Level, Mother's Education Level, Does the student has a bursary or Mahapola? Does the student do any part-time job? Does the student participants in the sports activities? Accommodation place and year of study at university. This is a survey-based study. Data were collected by using a questionnaire. The Snowball sampling technique was used to collect the data.220 students were respondents to this questionnaire. Out of 88-drug use respondents, 13.6 % female respondents and 55.4 % male respondents, according to that result, male drug usage is more than female drug usage. Burger students have used more drug than the other students and according to their religion. Hinduism students have used drug higher than the other students. The student whose parents are Divorced has a higher value of drug consumption than the other students and the students whose parents are Both Dead have a higher value of drug consumption than the other students. According to these two results, we can conclude that parent life status and marital status highly impact student drug consumption. furthermore, Students whose parents are Illiterate have a higher value of drug consumption than other students. we can conclude that parents' education status is an important factor in student drug consumption. The students whose father works in the private sector have used drugs higher than the other students. The students who live in the hostel have used drugs higher than boarding and home and students who have bursary used drugs higher

than those who don't have bursary and Mahapola. And also, the students who are doing a part-time job and students who have participated in sports used drugs higher those who are not doing a part-time job and have not participated in a sport. Third-year students' drug consumption is higher than the other year student. The majority of the students do not use any drug (75.9%). And most commonly used drug is alcohol.

Binary Logistic regression fitted model male drug usage is 2.008 higher than the females. For students who are doing a part-time job their drug consumption value is higher at 1.564 than the students who are not doing a part-time job. Drug consumption of the students who are engaged in Sports is 0.493 times higher than a student who is not doing sports. Third-year students' drug consumption and second-year students' drug composition are higher than the fourth-year students' drug composition. As a value, it is 1.289 times and 0.066 times respectively. Third-year students' drug consumption is 1.494 times higher than first-year students.

According to the fitted Binary Logistic regression model, Support Vector Machine and Probabilistic Neural Network non-random split and random split statistical model, SVM Non-Radom spilt model is the highest accuracy (93.18%) model to correctly classified the drug usage. In this model use the polynomial kernel function. Out of the training data set, the model predicts the 25 students are not using the drug. They are remove actually not drug use students. And also, the model predicts the 16 students are using the drug. They are actually drug used students. The model predicts the two students are not used drugs. But actually, they have used drugs. And also, the model predicts the one student who used a drug. But actually, this student not used the drug. Based on the result, gender, religion, year, student engage part-time job, student participants in the sports activities, student have a bursary or Mahapola, Mother's Education Level, Father's Occupation, Ethnicity, Marital Status of Parents were in influenced factors for student drug usage.

This study is a sensitive case study because of that, I have faced a huge problem in the data collection part. Lots of time was spent on the data collecting part. This research is limited to the Faculty of Science, University of Kelaniya. This study is does not apply to students outside of this designation. This research can be extended to the whole University system to draw more accurate results.

REFERENCES

[1] T. H. Bennett and K. R. Holloway, "Drug use among college and university students: Findings from a national survey," *J. Subst. Use*, vol. 20, no. 1, pp. 50–55, 2015, doi: 10.3109/14659891.2013.878762.

[2] B. Kasundu, M. Mutua Mutiso, P. Sutter Chebet, and P. Wanjiku Mwirigi, "Factors Contributing To Drug Abuse among the Youth in Kenya: A Case of Bamburi Location," *Elixir Soc. Sci.*, vol. 46, no. 16, pp. 8259–8267, 2012, [Online]. Available: http://www.elixirpublishers.com/articles/1350378627_46 (2012) 8259-8267.pdf.

[3] S. Senanayake *et al.*, "Smoking, Alcohol Consumption, and Illegal Substance Abuse among Adolescents in Sri Lanka: Results from Sri Lankan Global School-Based Health Survey 2016," *Adv. Public Heal.*, vol. 2018, pp. 1–7, 2018, doi: 10.1155/2018/9724176.

[4] N. Geramian *et al.*, "Development of a questionnaire to assess drug abuse among high school students of Isfahan province, Iran: An action research," *Int. J. Prev. Med.*, vol. 5, no. 14, p. 146, 2014, doi: 10.4103/2008-7802.157678.

[5] K. Holloway and T. Bennett, "Characteristics and correlates of drug use and misuse among university students in Wales: a survey of seven universities," *Addict. Res. Theory*, vol. 26, no. 1, pp. 11–19, 2018, doi: 10.1080/16066359.2017.1309031.

[6] M. Meray, F. Ahmed, and A. Rania, "Study of Drug Abuse among University Students in Sohag, Egypt," *Journal of Forensic,Toxicology and Medicolegal Analysis*, vol. 1, no. 2, pp. 19–23, 2016.

[7] O. Rezahosseini, A. Roohbakhsh, V. Tavakolian, and S. Assar, "Drug Abuse among University Students of Rafsanjan, Iran.," *Iran. J. psychiatry Behav. Sci.*, vol. 8, no. 2, pp. 81–5, 2014, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/25053961%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4105608.

[8] M. Zaman *et al.*, "Drug abuse among the students," *Pakistan J. Pharm. Res.*, vol. 1, no. 1, p. 41, 2015, doi: 10.22200/pjpr.2015141-47.

[9] T. Evgeniou and M. Pontil, "Support vector machines: Theory and applications," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2049 LNAI, no. May, pp. 249–257, 2001, doi: 10.1007/3-540-44673-7_12.

[10] T. H. misuse among university students in the U. I. for preventiongh Bennett and K. R. Holloway, "Drug misuse among university students in the UK: Implications for prevention," *Subst. Use Misuse*, vol. 49, no. 4, pp. 448–455, 2014, doi: 10.3109/10826084.2013.846378.

[11] F. Seguel, G. Santander, and M. L. Ramos, "Consumo de drogas y factores demográficos asociados en estudiantes de primer año de una universidad Chilena," *Acta Paul Enferm*, vol. 25, no. 2, pp. 33–39, 2012.

[12] T. Osman *et al.*, "Epidemiology of Substance Use among University Students in Sudan," *J. Addict.*, vol. 2016, pp. 1–8, 2016, doi: 10.1155/2016/2476164.

[13] I. L. M. Mahir and T. M. F. Wazeema, "Social Aspects of Drug Addiction in Sri Lanka," *J. Polit. Law*, vol. 13, no. 2, p. 54, 2020, doi: 10.5539/jpl.v13n2p54.

[14] R. Atkinson, F. Flint, and N. Gilbert, "Snowball Research Strategies," *Dep. Sociol. Univ. Surrey Guildf. GU2 5XH United Kingdom.*, no. 33, p. null-null, 2001.

[15] T. M. Franke, T. Ho, and C. A. Christie, "The Chi-Square Test: Often Used and More Often Misinterpreted," *Am. J. Eval.*, vol. 33, no. 3, pp. 448–458, 2012, doi: 10.1177/1098214011426594.

[16] K. A. Pituch and J. P. Stevens, "Binary Logistic Regression," *Appl. Multivar. Stat. Soc. Sci.*, pp. 454–490, 2020, doi: 10.4324/9781315814919-16.

[17] F. Ancona, A. M. Colla, S. Rovetta, and R. Zunino, "Implementing probabilistic neural networks," *Neural Comput. Appl.*, vol. 5, no. 3, pp. 152–159, 1997, doi: 10.1007/BF01413860.

[18] J. Sujatha and S. P. Rajagopalan, "Performance evaluation of machine learning algorithms in the classification of parkinson disease using voice attributes," *Int. J. Appl. Eng. Res.*, vol. 12, no. 21, pp. 10669–10675, 2017.

[19] D. P. Ranjan, N. Namita, and R. M. Chaturvedi, "A study of socio-demographic factors contributing to the habit of drug abuse in the urban slum community of Mumbai," *Biomed. Res.*, vol. 21, no. 3, pp. 277–284, 2010.