

# Leveraging Machine Learning Tools and Techniques to Determine the Association and Correlation Between Students' Drug Addict Behavioural Parameters

**Bahisht Samar**  
*Amity University, Noida*

---

## ABSTRACT

Students are the future leaders and assets of a country. But in this era students are losing their education level for drug addiction. So, a nation must be concerned about these students and this social problem. And we need to solve this problem as soon as possible. In this study, we have tried to find out the most significant causes of drug addiction for a student. We have also tried to find out the correlation among the factors. Those may help students and guardians to know about the main causes of drug addiction which are primary steps of drug addiction. For performing this study, we have collected 130 instances data from various universities from Bangladesh where each has 13 unique attributes. For processing dataset to select the significant attributes, we have used Machine learning techniques. And also, for finding out the most correlated features of dataset we have used Association Rule Mining techniques. Apriori algorithm is used for implementing Rule Mining and we found 8 rules.

## INTRODUCTION

Education is the backbone of a nation. The students are playing an important role in a country. But drug addiction is a great problem for dropping students' education level. In Bangladesh there are nearly 25 lakhs who are drug addicted, 80 percent are young men who are 15 years to 30 years old [1]. Addiction exerts a long and potent influence on the encephalic that manifests in three distinct ways: craving for the object of addiction, loss of control over its use, and perpetuating involution with it despite adverse consequences [2]. From the 1970s on, rates of alcohol consumption and binge imbibing have remained fairly constant. College students have always represented an astronomically immense portion of the population abusing drugs and alcohol on a customary substratum [3]. The abuse of drugs is increasing day by day among college students. Categorically, 32.4 percent of college students noted they had binge imbibed in the past a fortnight. In integration, 40.8 percent of college students reported they had been intoxicated in the past month [4]. By analysing student's activities, we can predict one's drug addict behaviour. For this reason, we collected 130 data by using Google form. Those data are put to excel sheets for calculated P-value. Basically, the rule of mining technique of Machine Learning is used by us. The prediction system is too important for us. By this system we can give alarm and we should take care of these students.

## METHODOLOGY

In this analysis different approaches have been used. The approaches have been given below:

- A. Data collection
- B. Data preprocessing
- C. Applying Correlation
- D. Applying Association
- E. Tools and techniques.

### A) Data collection

For this research purpose we collected data from North Western University, Khulna, Bangladesh and from other institutes by Google form. The data set contains 130 instances where have 10 attributes. In 'Table 1' the features list has been shown.

Table 1. Features List

Features	DATA DISTRIBUTION	
	Subcategory	Count
Class	Undergraduate	111
	Postgraduate	10
	11-12	6
	<5	2
	Other	1
Married_Status	Yes	3
	No	127
Age	21-30	70
	31-40	1
	11-20	58
	41-50	1
Gender	Male	81
	Female	49
Ever_Tried_To_Take_Drug	Yes	16
	No	114
Drug_Availability	Yes	45
	No	85
Drug_Abuse_Increasing_Among_Your_Friends	Yes	67
	No	63
Smoking_Preferred_Over_Alcohol	Yes	368
	No	628
Main_Cause_For_taking_Substance	Environment	120
	Genetics	10
Addicted_With_Environmental_issues	Personal Problem	50
	Problems at School	3
	Family Pressure	7
	Passing Time	16
	Economic Problems	5
	Neglected - Appearances	13
	Educational Pressure	2
	Health issues	1
	Others	33

## B) Data preprocessing

Missing value of the database has been handled by Trimmed- mean function. After that all variables have been categorized in nominal value. To figure 1, there has been shown the system architecture.

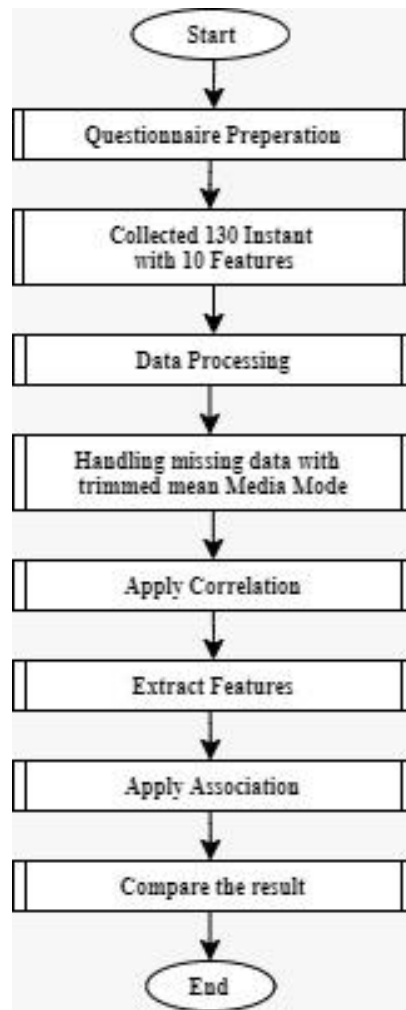


Figure 1: System Architecture

### C) Applying Correlation

Correlation is a statistical measure that denotes the extent to which two or more variables fluctuate together [9]. A Chi-square test is designed to analyse categorical data. That signifies that the data has been counted and divided into categories. The Chi Square statistics is utilized for testing relationships between categorical variables. When there is no relationship subsisting on the categorical variables is the null hypothesis [10]. Where cross tabulation presents the distributions of two categorical variables simultaneously [11]. System architecture for Correlation Chi-Square test P-value work is given in Figure 2.

### D) Applying Association

Rule Mining is a technique of identifying the correlation among the attributes of a dataset [12]. In this study we utilize Apriori algorithm. The algorithm can be branched into two components. The self-join rule in order to find the periodic groups with  $k+1$  item, the periodic  $K$ -itemset helps to use [13]. The process has been illustrated in Figure 3.

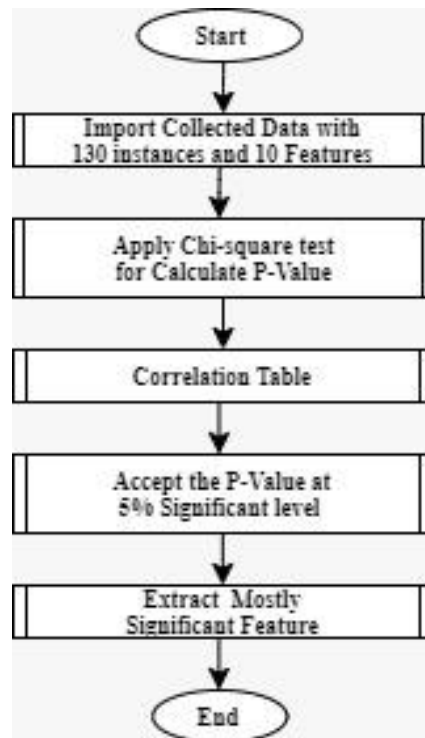


Figure 2: System architecture for Correlation Chi-Square test P-Value

Apriori algorithm is a classical algorithm in data mining. It is utilized for mining frequent itemsets and germane sodality rules [14]. Moreover, to discover all rules it's required which have greater support than the threshold support as well as greater confidence than the threshold confidence.

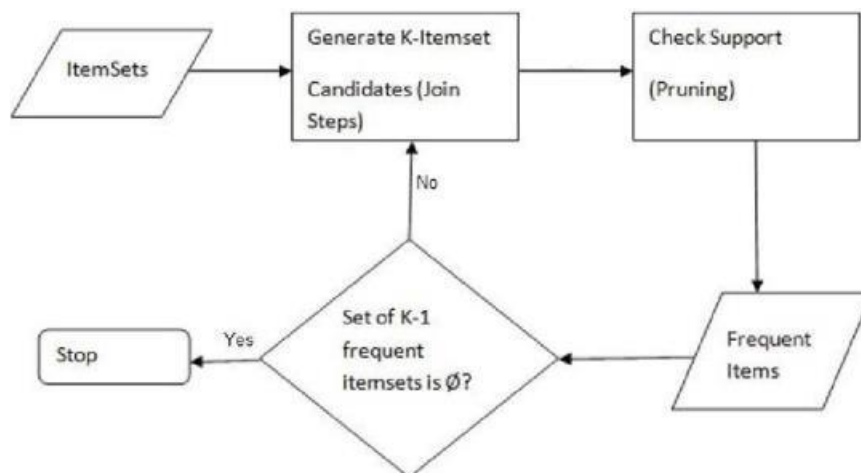


Figure 3: Apriori Algorithm Process

E) Tools and techniques

For our work R programming language has been used for statistical analysis, graphics representation and reporting. Some important functions have been used. Which is given below:

subset () : After satisfying a particular condition, the subset functions returns vector, data frame etc.

mean () : For using the trimmed-mean.

chisq.test() : For the chi-Square test it has been used.

Apriori(): It provides the rule of association mining.

plot(): A generic object to plot the objects of R.

inspect(): It represents all the summary of option, plot and statistics.

## OUTCOME

Association rule is known as frequent if-then associations of Association rule mining. Two segments 'if' and 'then' where 'if' known as antecedent and 'then' known as consequent. Mainly, within the data an antecedent is identified, whereas consequent discovered in blend with the antecedent. The time shows up in the information by support and the occasions the if-then statements are discovered true that showed by confidence. There is a third measurement that is called lift that can be utilized to distinguish confidence with anticipated confidence. The outcomes have been shown in Table-2. We have found a total of 4 rules. Where,

Support = 45%

Confidence = 95%

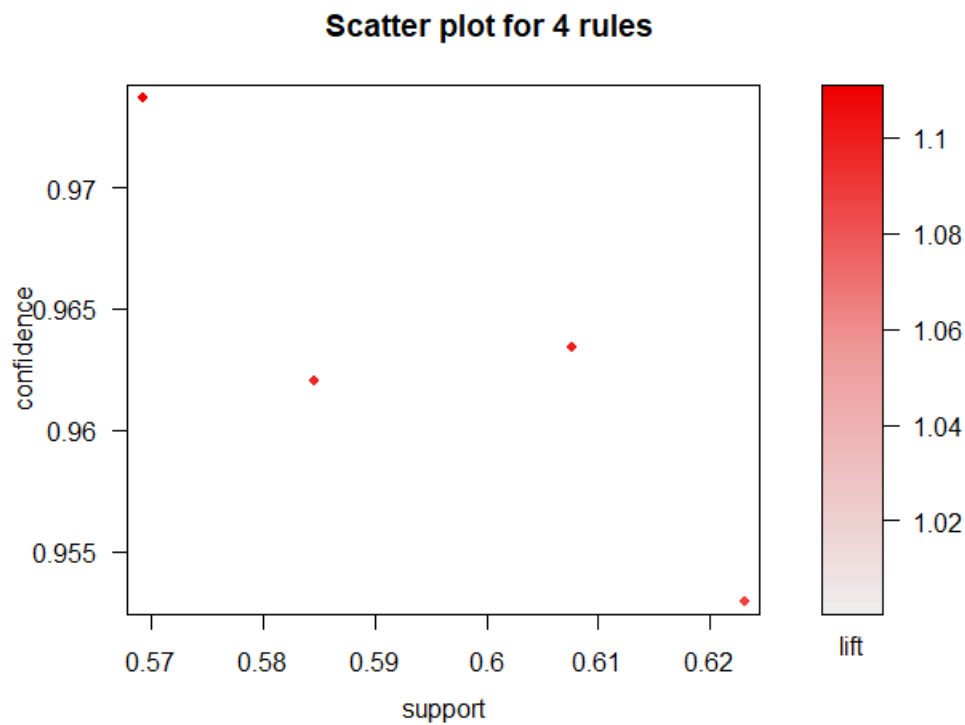


Figure 4: Scatter plot for 4 rules

Table 2: 4 rules descriptions.

Rules	LHS	RHS	Support	Confidence	Lift
[1]	{married_Status=No, drugs_availability=No, drug_abuse_increasing=Yes}	{tried_taking=No}	0.56 92308	0.973 6842	1.11 0342
[2]	{married_Status=No, drugs_availability=No, drug_abuse_increasing=Yes, main_cause_for_taking_substance=Environment}	{tried_taking=No}	0.53 07692	0.971 8310	1.10 8228
[3]	{class=Under_Graduate, married_Status=No, drugs_availability=No, drug_abuse_increasing=Yes}}	{tried_taking=No}	0.51 53846	0.971 0145	1.10 7297
[4]	{class=Under_Graduate, married_Status=No, drugs_availability=No, drug_abuse_increasing=Yes,	{tried_taking=No}	0.48 46154	0.969 2308	1.10 5263

The scatter plot is shown in Figure 4. Basically, scatter plot shows the amount one variable is influenced by another variable. Scatter plot to line graphs in that they utilized horizontal and vertical axes to information focus. Correlation is the connection between the two factors. It's called Correlation. To show a positive pattern it goes left to right. This defines an uphill connection among X & Y. When the X-value increment (move right), the Y-values will in general increment (move up).

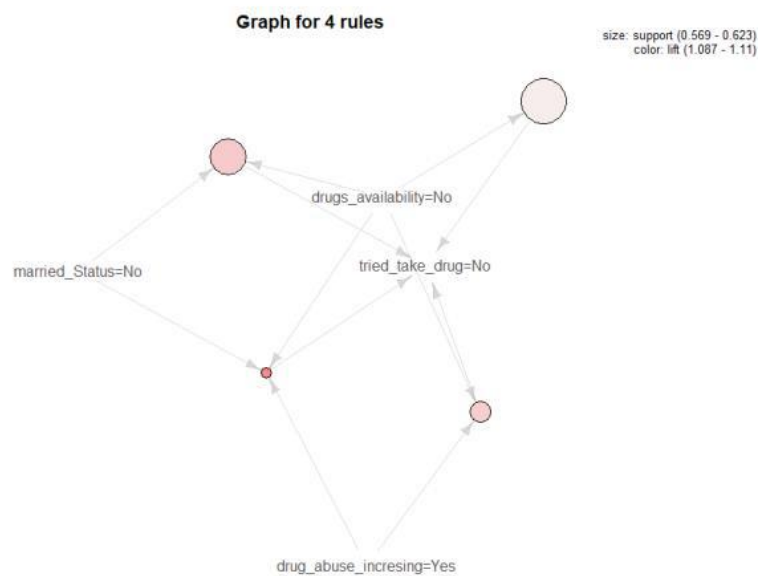


Figure 5: Apriori Algorithm Process

The most significant association rules are shown in Figure 5. This introduces the graph visualization of those rules. Item sets are introduced as vertices though rules are introduced to as coordinated edges between item sets, it is shown to the system chart. To the parallel coordinates, it's very own axis where each factor is given and there is all the axes are set in parallel to one another. Every axis can have an alternate scale, as every factor works off an alternate unit of estimation, or all the axes can be standardized to keep all the scales uniform. Which associated over all axes the value is plotted as progression of lines. This illustrates that each line is a group of points positioned on each axis, that have all been associated together.

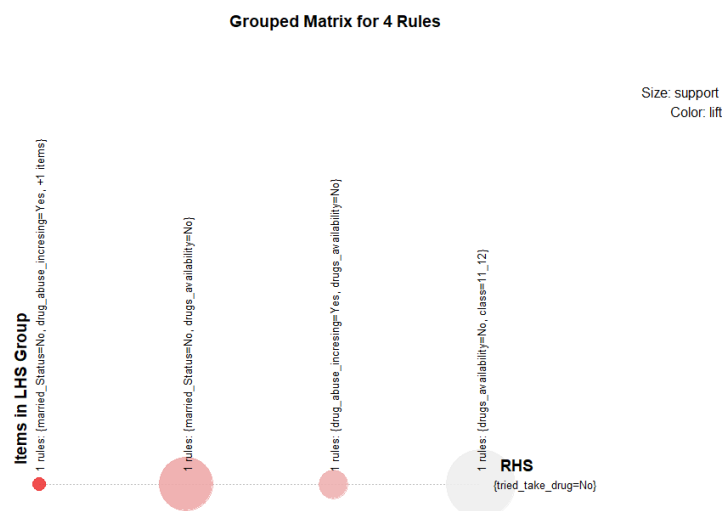


Figure 6: Parallel coordinates plot for 4 rules

**CONCLUSION**

This study is mainly arranged with students. For this reason, we collected data from various institutions. But this is the main matter we know that drug addiction among students is increasing day by day. Then Apriori was applied to our data where support 0.45 and there is confidence is 0.90. From there we determined 4 association rules. Where another the main cause is married status. Who married status is single, he is passing time on drugs. Another reason we found that environmental issue. For this environmental issue students are taking substance. In there the biggest factor is personal problems. Students are taking substance for their personal problems. The universities

and the college students are more addicted where the age is 21 to 30, they are taking substance and they are unable to get success in their educational life. Although this research used several correlation factors and obtained results are accordance with the background data and feedback and it's a fact that machine learning methods can produce an effective rule. In future we will do this research with a more advanced and intelligent system with explored information. We hope an advanced intelligent system will help to detect the harmful factor of drug addicted students.

## REFERENCES

1. M. Shazzad, S. Abdal, M. Momen Majumder, J. Alam Sohel, S. Monowar Ali and S. Ahmed, "Drug Addiction in Bangladesh and its Effect", Bangladesh Journals Online, 2013. [Online]. Available: <https://www.banglajol.info/index.php/MEDTODAY/article/view/17927>. [Accessed: 10- Feb- 2020].
2. "How addiction hijacks the brain", How addiction hijacks the brain - Harvard Health, 2020. [Online]. Available: [https://www.health.harvard.edu/newsletter\\_article/how-addiction-hijacks-the-brain](https://www.health.harvard.edu/newsletter_article/how-addiction-hijacks-the-brain). [Accessed: 14- Feb- 2020].
3. A. M. Arria, K. M. Caldeira, H. K. Allen, B. A. Bugbee, K. B. Vincent and K. E. O'Gradyb, "Prevalence and incidence of drug use among college students: An 8-year longitudinal analysis", NCBI, 2017. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5638668/>. [Accessed 18 February 2020].
4. "Facts and Statistics of College Drug Abuse", Addiction Center, 2019. [Online]. Available: <https://www.addictioncenter.com/college/facts-statistics-college-drug-abuse/>. [Accessed: 15- Feb- 2020].
5. G. J, "Correlation between drug use by teenagers and drug use by older family members.", National Center for Biotechnology Information, vol. 13, 1987, no. 1-2, pp. 95-108, 2009. Available: <https://www.ncbi.nlm.nih.gov/pubmed/3687887>. [Accessed 17 March 2020].
6. Ahad, A., Chowdhury, M., Kundu, I., Zahan Tanny, N. and Rahman, M., 2017. Causes of Drug Addiction among Youth in Sylhet City of Bangladesh. IOSR Journal Of Humanities And Social Science (IOSR-JHSS), [online] Volume 22(Issue 5, Ver. 7 (May. 2017), pp.27-31. Available at: <<http://www.iosrjournals.org/iosr-jhss/papers/Vol.%2022%20Issue5/Version-7/E2205072731.pdf>> [Accessed 20 April 2020].
7. Singh, J., Kaur, M. and Singh, G., 2018. EVALUATION OF DRUG ADDICTION CAUSES IN PUNJAB USING MACHINE LEARNING. International Journal of Advanced Research in Computer Science, [online] Volume 9(March-April 2018), pp.730-733. Available at: <<http://ijarcs.info/index.php/Ijarcs/article/view/5832/4813>> [Accessed 22 May 2020].
8. Acion, L., Kelmansky, D., der Laan, M., Sahker, E., Jones, D. and Arndt, S., 2017. Use of a machine learning framework to predict substance use disorder treatment success. PLOS ONE, [online] Available at: <<https://journals.plos.org/plosone/article/authors?id=10.1371/journal.pone.0175383>> [Accessed 26 May 2020].
9. M. Rouse, "What is correlation?" WhatIs.com, 2016. [Online]. Available: <https://whatis.techtarget.com/definition/correlation>. [Accessed: 16- Feb- 2020].
10. "Tutorial: Pearson's Chi-square Test for Independence", Department of Linguistics, 2018. [Online]. Available: <https://www.ling.upenn.edu/~clight/chisquared.htm>. [Accessed: 16- Feb- 2020].
11. "Using Chi-Square Statistics in Research", Statistics Solutions, 2020. [Online]. Available: <https://www.statisticssolutions.com/using-chi-square-statistic-in-research/>. [Accessed: 17- Feb- 2020].
12. A. Rai, "Association Rule Mining", An Overview and its Applications, 2019. [Online]. Available: <https://www.upgrad.com/blog/association-rule-mining-an-overview-and-its-applications/>. [Accessed: 17- Feb- 2020].
13. R: The R Project for Statistical Computing: 2019. <https://www.rproject.org/>. [Accessed: 17- Feb- 2020].
14. "A beginner's tutorial on the apriori algorithm in data mining with R implementation", HackerEarth, 2020. [Online]. Available: <https://www.hackerearth.com/blog/developers/beginners-tutorial-apriori-algorithm-data-mining-r-implementation>. [Accessed: 16- Feb- 2020]