

**Title: Find Reason of Drug Addiction through Machine Learning Techniques.**

Name of student: (1) Md Azad Hossain (2) Bilkiss Ara Sweet

ID: (1) 13334351 (2) 161412020

Present Address: (1) Mockbul Hossian Hall(2) Mona Hossain Hall, City University, Dhaka.

Name of the department: Dept. of Computer Science & Engineering.

Name of Program: B. Sc. in CSE.

Principle-supervisor: Md. Rakib Uddin, Lecturer & Coordinator , Dept. of CSE.

**Introduction:**

Bangladesh is a developing country and about 18 million people live here. Most of the people live under poverty and also they are not so educated. Here in Bangladesh most of people don’t have enough knowledge about what changes can happen due to drug addiction in our human body and in psychology. Even most of our people don’t have knowledge of what sort of instances they can face in the future for drug addiction people, that why in our country drug-addicted people are increasing day by day. They have been living in such a kind of darkness because most people don’t know why people are addicted to drugs. In our country, an addict is counted as a wicket man but no one doesn’t try to understand why he is addicted. Therefore we have tried to find out some main reasons for being an addicted person. To do this work we collected data and we have applied machine learning techniques. Because machine learning techniques are used to analyze the data and we have got a better result from this. The accuracy of different techniques is so good.

**Related Works:**

We had to do some study about drug addiction to accomplish our analyses. We had read some papers about abuse of drugs. We found that no previous research have done about why people being drug-addicted using machine learning technique, We found a paper, in that paper they investigated the causes of drug addiction among the youth in Sylhet city and identified the problems faced by young drug addicts as consequences of drug addiction. In that study they collected data from 42 drug-addicted people. Data was collected by using structured questionnaire, interview schedule and observation. That study depicted that curiosity was first and foremost reason and temptation by friends was the second reason behind drug addiction among youth. This study also revealed that highest 76.19% of drug addicts changed taking of drugs one after another to get adventure by tasting different drugs. About 88.09% drug addicts feel being neglected by their relatives due to drug addiction. About 95.24% felt that they did not have any discipline in their daily life and 50% have experienced some sexual problems due to their involvement in drug addiction.

We have found another paper where they investigated about Influencing Factors of ‘Yaba’ Addiction among the Youth of Bangladesh and Its Effect in Cox’s Bazar district. They Conducted depth interviews with the yaba addicted. They found 60% of respondents revealed that due to the influence of peer group they started using yaba. There have been other reasons for taking yaba, which were curiosity, extra power & entertainment, besides parent's negligence.

In another study was conducted to determine the prevalence of tobacco smoking and factors associated with the initiation of smoking among university students in Dhaka, Bangladesh. In their study the found influence of friend is the most significant reason for initiating tobacco smoking. They were analyzed data with logistic regression.

**Proposed Method:**

* Collecting real field data from Rehab.
* data Preprocessing.
* Separating training and testing data.
* Apply Machine Learning Techniques

-Decision Trees

-SVM

-KNN

-Logistic Regression

- Random Forest

* Find predicting accuracy

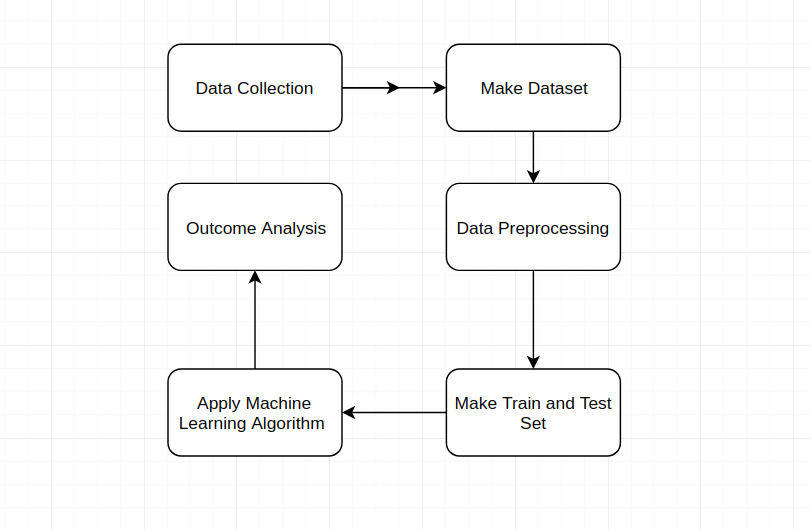


Fig: Proposed method to predictive accuracy

**Machine Learning:**

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

**Supervised machine learning algorithm:** It can apply what has been learned in the past to new data using labeled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.

Machine learning (ML) is a category of [algorithm](https://whatis.techtarget.com/definition/algorithm) that allows software applications to become more accurate in predicting outcomes without being explicitly programmed. The basic premise of machine learning is to build algorithms that can receive input data and use [statistical analysis](https://whatis.techtarget.com/definition/statistical-analysis) to predict an output while updating outputs as new data becomes available.

**Unsupervised machine learning algorithm:** It is used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system doesn’t figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data.

**Semi-supervised machine learning algorithm:** It fall somewhere in between supervised and unsupervised learning, since they use both labeled and unlabeled data for training – typically a small amount of labeled data and a large amount of unlabeled data. The systems that use this method are able to considerably improve learning accuracy. Usually, semi-supervised learning is chosen when the acquired labeled data requires skilled and relevant resources in order to train it / learn from it. Otherwise, acquiring unlabeled data generally doesn’t require additional resources.

**Logistic Regression:** Logistic regression is the appropriate regression analysis to conduct when the dependent variable is dichotomous (binary). Like all regression analyses, the logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.

**Reinforcement machine learning algorithm:** It is a learning method that interacts with its environment by producing actions and discovers errors or rewards. Trial and error search and delayed reward are the most relevant characteristics of reinforcement learning. This method allows machines and software agents to automatically determine the ideal behavior within a specific context in order to maximize its performance. Simple reward feedback is required for the agent to learn which action is best; this is known as the reinforcement signal.

### SVM: SVM is a supervised machine learning algorithm which can be used for classification or regression problems. It uses a technique called the kernel trick to transform your data and then based on these transformations it finds an optimal boundary between the possible outputs.

### 

**Decision Trees:** Decision Trees are a type of Supervised Machine Learning (that is you explain what the input is and what the corresponding output is in the training data) where the data is continuously split according to a certain parameter. The tree can be explained by two entities, namely decision nodes and leaves. The leaves are the decisions or the final outcomes. And the decision nodes are where the data is split.

**K-Nearest Neighbors (KNN):** K-Nearest Neighbors (KNN) is one of the simplest algorithms used in [Machine Learning for regression](https://quantra.quantinsti.com/course/trading-with-machine-learning-regression) and classification problem. KNN algorithms use a data and classify new data points based on a similarity measures (e.g. distance function). Classification is done by a majority vote to its neighbors. The data is assigned to the class which has the most nearest neighbors. As you increase the number of nearest neighbors, the value of k, accuracy might increase.

**Discussion:**

Drug addiction is one of the most common problems in Bangladesh and this problem increasing day by day. Although many researchers and governments are trying to find out how to diminish the increasing rate of drug addiction. Which reason is the main culprit behind it? So we tried here to find out those reasons for that reason people being addicted. To analyses cause of drug addiction we collected row data and we preprocess data. After that, we used some supervised machine learning algorithms to find out the accuracy. We have applied Logistic Regression, Decision Trees, KNN, SVM, and Random Forest algorithms. We have got the highest 97 % accuracy with the SVM algorithm.

**Conclusion and Future Work:**

In this paper, we have tried to demonstrate those reasons for that a person get addict. For this research, we have collected 327 data and we have found 73% addicted . We have applied the machine learning techniques in Jupyter Notebook to find out the accuracy and showed the comparison among different kinds of algorithms. The accuracy is between 82% to 97% among all the classifier.

In the future, we will try to apply more algorithms to find out the best accuracy and also we will try to work with many more data.

**References:**

1. Md. Abdul Ahad 1, Dr. Mitu Chowdhury , Dr. Indrajit Kundu , Nishith Zahan Tanny, Dr. M. Wakilur Rahman “Causes of Drug Addiction among Youth in Sylhet City of Bangladesh”, Volume 22, Issue 5, Ver. 7 (May. 2017) PP 27-31 e-ISSN: 2279-0837.
2. Minhazur Rahman Rezvi “Influencing Factors of ‘Yaba’ Addiction among the Youth of Bangladesh and Its Effect: A Qualitative Study”, 1-7, 2019; Article no.ARJASS.47877 ,ISSN: 2456-4761.
3. Sahadat Hossain, Shakhaoat Hossain, Fahad Ahmed, Rabiul Islam, Tajuddin Sikder, Abdur Rahman “Prevalence of Tobacco Smoking and Factors Associated with the Initiation of Smoking among University Students in Dhaka, Bangladesh”, Vol. 6, No. 1 (2017) | ISSN 2166-7403.