**ABSTRACT**

The problem of evaluating an individual’s risk of drug consumption and misuse is highly important and novel. An online survey methodology was employed to collect data including personality traits, sensation seeking, and demographic information. The data set contained information on the consumption of 18 central nervous system psychoactive drugs. Correlation analysis using a relative information gain model demonstrates the existence of a group of drugs (amphetamines, cannabis, cocaine, ecstasy, legal highs, LSD, and magic mushrooms) with strongly correlated consumption. An exhaustive search was performed to select the most effective subset of input features and data mining methods to classify users and non-users for each drug. A number of classification methods were employed (decision tree, random forest, k-nearest neighbours, linear discriminant analysis, Gaussian mixture, probability density function estimation, logistic regression, and naïve Bayes) and the most effective method selected for each drug. The quality of classification was surprisingly high. The best results with sensitivity and specificity being greater than 75% were achieved for cannabis, crack, ecstasy, legal highs, LSD, and volatile substance abuse. Sensitivity and specificity greater than 70% were achieved for amphetamines, amyl nitrite, benzodiazepines, chocolate, caffeine, heroin, ketamine, methadone, and nicotine. The poorest result was obtained for prediction of alcohol consumption.

Batch-17

160030878 – M. SAI HARSHITHA

160030890 – M. MOUNIKA

160031387 – T. JASWANTH