

Predicting risk of drug consumption based on personality features, demographic variables and previous consumption.

Camila Batista Rahal and Elena Maestre Soteras

Abstract

Drug consumption and misuse remain significant public health concerns, shaped by an interaction of personality traits, demographic characteristics, and historical behavior. The study employs a variety of predictive models (logistic regression, lasso, random forest, and Xgboost) to analyze the relationships between this personality factors, demographic variables, and drug consumption patterns, aiming to identify significant predictors for the risk of future drug use. The participants were classified as 'consumer' or 'non-consumer.' XGBoost baseline achieve the best performance with a recall of 84.3%, followed by Random Forest with recall of 83%. Random Forest achieves the highest AUC (0.90), making it the most effective model for distinguishing between consumers and non-consumers classes. Bagging Classifier and Optimized Random Forest closely follow with an AUC of 0.89, indicating strong performance with balanced sensitivity and specificity. This research demonstrates the potential of data-driven methods for understanding and addressing drug use behaviors, while offering a foundation for future studies that integrate psychological and demographic data.

Exploratory Data Analysis

The dataset had 1,885 participants, no missing data, and the features were quantified and standardized. The dataset shows a high prevalence of participants aged 18–24, a predominance of consumers within this group, a disproportionate number of White participants compared to other ethnicities, and a higher number of male consumers than female (Table 1). Summary statistics (Table 2) confirm key predictors like age, country, ethnicity, impulsivity, and sensation-seeking. Lower education appears linked to higher use. Target imbalance (mean = 0.552) was handled with SMOTE.

Figure 1 below visualizes the distribution of drug usage frequencies across various substances. It highlights trends such as a large number of participants reporting never having used heroin or crack, while substances like ecstasy and mushrooms show relatively higher recent use. Additionally, there is a high prevalence of users for substances like chocolate, nicotine, caffeine, alcohol, and cannabis. These substances were excluded when creating the target variable.

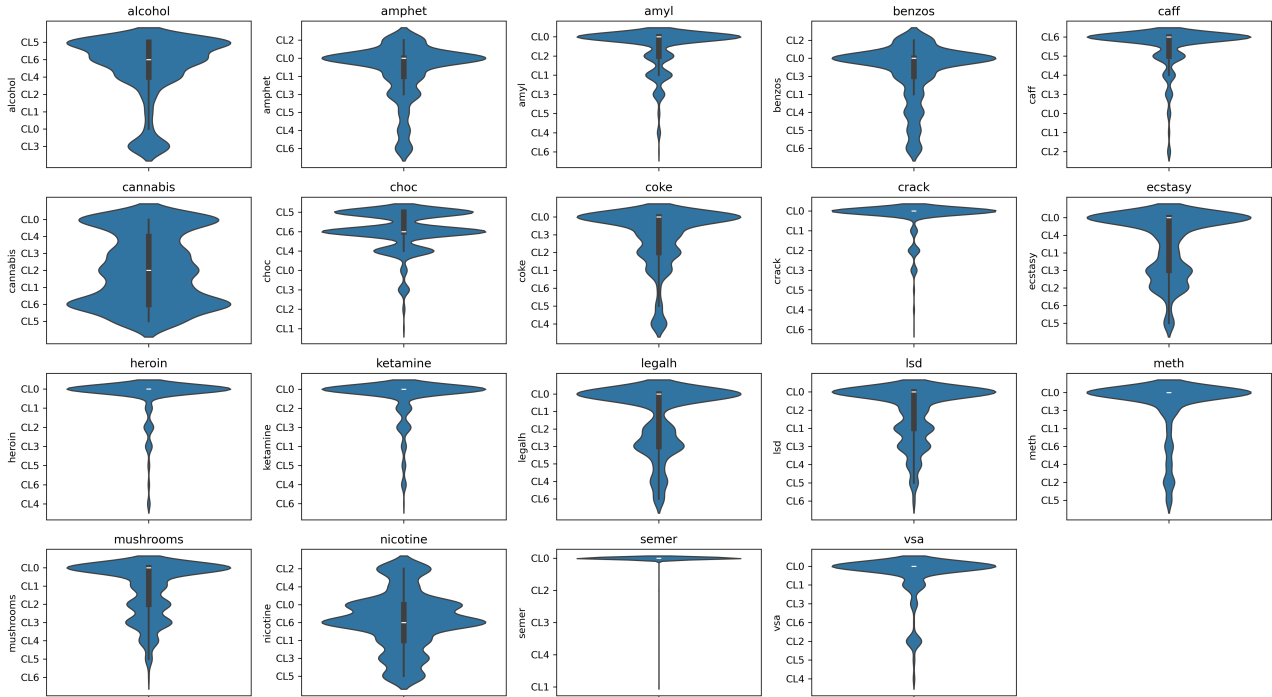


Figure 1: Violin plot of all drugs consuming behavior from CL0 (Never used) to CL6 (Used last day).

Table 1: Demographic Summary		
Target	Non-consumer	Consumer
Age category		
18-24 years	113	524
25-34 years	218	262
35-44 years	215	140
45-54 years	211	83
55-64 years	70	23
65+ years	13	5
Gender category		
Female	552	385
Male	288	652
Education category		
Doctorate degree	53	36
Left school at 16 years	53	45
Left school at 17 years	11	18
Left school at 18 years	25	74
Left school before 16 years	14	14
Master's degree	189	94
Professional certificate/diploma	149	121
Some college or university (no certificate)	85	418
University degree	261	217
Country category		
Australia	8	44
Canada	32	55
New Zealand	1	4
Other	28	90
Republic of Ireland	6	14
UK	716	328
USA	49	502
Ethnicity category		
Asian	17	8
Black	25	8
Mixed-Black/Asian	0	3
Mixed-White/Asian	8	12
Mixed-White/Black	7	12
Other	19	43
White	764	951

Table 1: Summary statistics of features and target variable

	count	mean	std	min	25%	50%	75%	max
Age	1877	0.037	0.878	-0.951	-0.951	-0.078	0.497	2.591
Gender	1877	-0.0007	0.482	-0.482	-0.482	-0.482	0.482	0.482
Education	1877	-	0.949	-2.435	-0.611	-0.059	0.454	1.984
Country	1877	0.358	0.699	-0.57	-0.57	0.96	0.96	0.96
Ethnicity	1877	-0.309	0.166	-1.107	-0.316	-0.316	-0.316	1.907
Nscore	1877	0.00055	0.998	-3.464	-0.678	0.042	0.629	3.273
Escore	1877	-0.0019	0.997	-3.273	-0.695	0.003	0.637	3.273
Oscore	1877	-0.0032	0.995	-3.273	-0.717	-0.019	0.723	2.901
Ascore	1877	0.00065	0.996	-3.464	-0.606	-0.017	0.76	3.464
Cscore	1877	0.00039	0.997	-3.464	-0.652	-0.006	0.584	3.464
Impulsive	1877	0.0052	0.954	-2.555	-0.711	-0.217	0.529	2.901
Ss	1877	-0.0074	0.962	-2.078	-0.525	0.079	0.765	1.921
Target	1877	0.552	0.497	0	0	1	1	1

Square of 7 is 49