Classifying Drug Consumption Patterns with Machine Learning

Sandip Patel

Project Overview

Understanding Drug Consumption Patterns based on demographic, Personality Scores, Psychological traits.

➜ **Challenges**
  - ◆ Data Quality and Availability
  - ◆ Privacy and Ethical Concerns
  - ◆ Complexity of Drug Use Behavior
  - ◆ Intervention Effectiveness
 - ◆ Policy and Regulatory Barriers

➜ **Opportunities**
  - ◆ Predictive Analytics for Early Intervention
  - ◆ Personalized Treatment Plans
  - ◆ Enhanced Public Health Surveillance
  - ◆ Community-Based Programs
  - ◆ Educational and Awareness Campaigns

Overview of the Proposed Vision for Tackling Drug Consumption Classification Using Data Science

Vision Statement: Utilize advanced data science techniques to develop a comprehensive understanding and classification system for drug consumption patterns, leveraging demographic, psychological, and behavioral data to inform targeted public health interventions and policy-making.
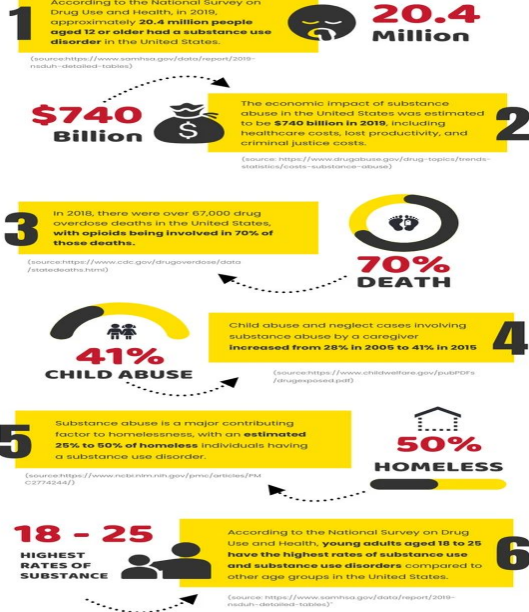

➤ **Data Exploration and Preprocessing**
➤ **Exploratory Data Analysis (EDA)**
➤ **Model Development**
➤ **Model Evaluation**
➤ **Insights and Interpretability**
➤ **Deployment and Application**
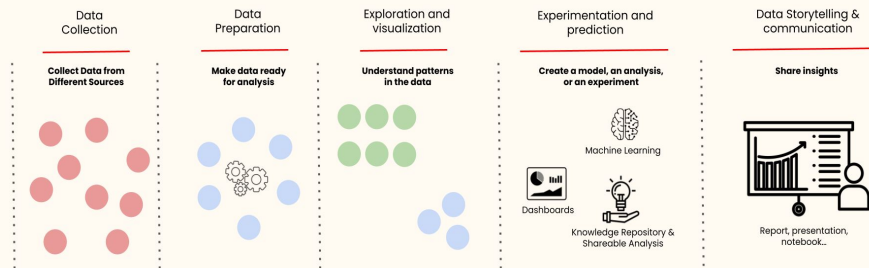
Long-term Vision: By creating a reliable and interpretable classification system, this project aims to contribute to a data-driven approach in tackling drug misuse. The ultimate goal is to reduce the prevalence of drug consumption through informed public health strategies and to support individuals at risk with targeted interventions based on comprehensive data analysis.

The Multi-Faceted Impact of Addressing Drug Abuse

The Impact of Substance Abuse on Society

1. According to the National Survey on Drug Use and Health, in 2019, approximately **20.4 million people aged 12 or older had a substance use disorder** in the United States.
(source:https://www.samhsa.gov/data/report/2019-nsduh-detailed-tables)
20.4 Million

2. The economic impact of substance abuse in the United States was estimated to be **$740 billion in 2019**, including healthcare costs, lost productivity, and criminal justice costs.
(source: https://www.drugabuse.gov/drug-topics/trends-statistics/costs-substance-abuse)
$740 Billion

3. In 2018, there were over 67,000 drug overdose deaths in the United States, **with opioids being involved in 70% of those deaths.**
(source:https://www.cdc.gov/drugoverdose/data/statedeaths.html)
70% DEATH

4. Child abuse and neglect cases involving substance abuse by a caregiver **increased from 28% in 2005 to 41% in 2015**
(source:https://www.childwelfare.gov/pubPDFs/drugexposed.pdf)
41% CHILD ABUSE

5. Substance abuse is a major contributing factor to homelessness, with an **estimated 25% to 50% of homeless** individuals having a substance use disorder.
(source:https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2774244/)
50% HOMELESS

6. According to the National Survey on Drug Use and Health, **young adults aged 18 to 25 have the highest rates of substance use and substance use disorders** compared to other age groups in the United States.
(source: https://www.samhsa.gov/data/report/2019-nsduh-detailed-tables)
18 – 25 HIGHEST RATES OF SUBSTANCE

❖ **Improved Public Health:**
  ➢ Early intervention prevents drug abuse escalation, reducing healthcare burdens and deaths.
  ➢ Nearly 92,000 drug overdose deaths occurred in the U.S. in 2020.

❖ **Enhanced Quality of Life:**
  ➢ Personalized treatment boosts recovery rates, benefiting individuals, families, and communities.
  ➢ Effective interventions can significantly lower the $740 billion annual cost of drug-related expenses in the U.S.

❖ **Workplace Safety and Productivity:**
  ➢ Addressing drug abuse reduces absenteeism, improves productivity, and lowers healthcare costs.
  ➢ Supporting interventions boosts overall workplace efficiency.

❖ **Corporate Social Responsibility and Innovation:**
  ➢ Companies enhance reputation by engaging in public health initiatives.
  ➢ Collaborations with healthcare providers position businesses as leaders in corporate social responsibility and innovation.

An Introduction to the Dataset

Database records information on 1885 respondents, including 12 attributes: personality traits (NEO-FFI-R), impulsivity (BIS-11), sensation seeking (ImpSS), education, age, gender, country, and ethnicity. Attributes are initially categorical but quantified later. Respondents reported usage of 18 drugs (legal and illegal) and one fictitious drug (Semeron). For each drug, usage frequency is categorized into seven classes. The database comprises 18 classification problems, each with seven usage classes.

	ID	Age	Gender	Education	Country	Ethnicity	Nscore	Escore	Oscore	Ascore	...	Ecstasy	Heroin	Ketamine	Legalh	LSD	Meth	Mushroom
0	1	0.49788	0.48246	-0.05921	0.96082	0.12600	0.31287	-0.57545	-0.58331	-0.91699	...	CL0	CL0	CL0	CL0	CL0	CL0	Cl
1	2	-0.07854	-0.48246	1.98437	0.96082	-0.31685	-0.67825	1.93886	1.43533	0.76096	...	CL4	CL0	CL2	CL0	CL2	CL3	Cl
2	3	0.49788	-0.48246	-0.05921	0.96082	-0.31685	-0.46725	0.80523	-0.84732	-1.62090	...	CL0	CL0	CL0	CL0	CL0	CL0	Cl
3	4	-0.95197	0.48246	1.16365	0.96082	-0.31685	-0.14882	-0.80615	-0.01928	0.59042	...	CL0	CL0	CL2	CL0	CL0	CL0	Cl
4	5	0.49788	0.48246	1.98437	0.96082	-0.31685	0.73545	-1.63340	-0.45174	-0.30172	...	CL1	CL0	CL0	CL1	CL0	CL0	Cl

5 rows × 32 columns

Source: https://www.kaggle.com/datasets/mexwell/drug-consumption-classification
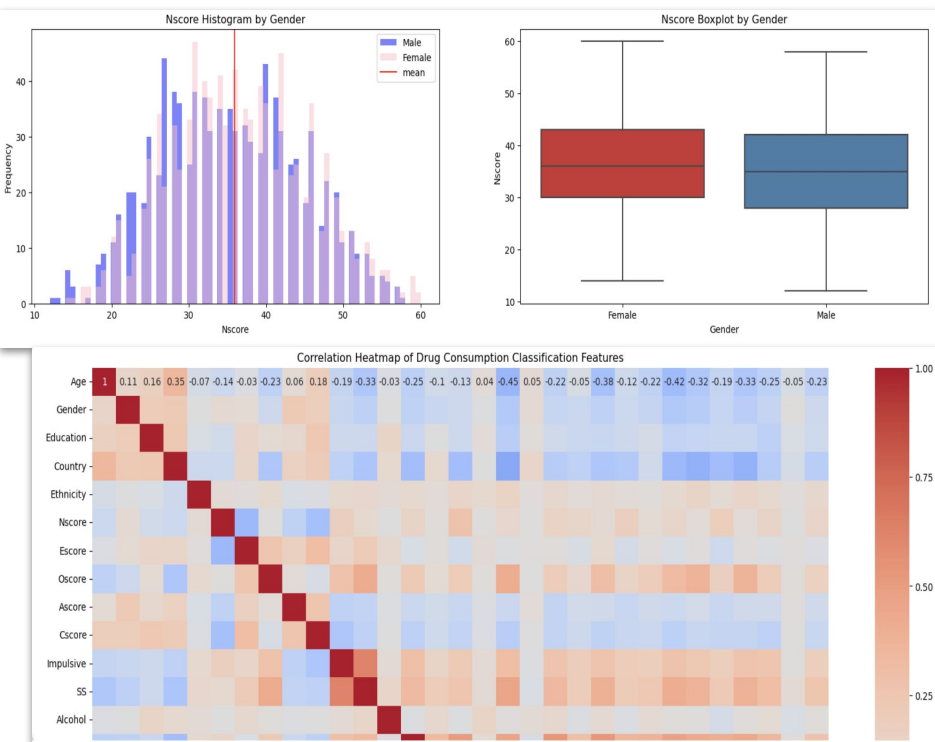
An Introduction to the Dataset

❖ **Data Quality Concerns:**

➢ Data is almost clean, except 19 null values in Cscore.

➢ The original data used numeric codes for all features, which were mapped to human-readable values.

➢ Self-Reporting Bias

➢ Over-Claiming

➢ Data Imbalance

➢ Outliers

➢ Inconsistent Data

➢ Temporal Relevance

➢ Ethical and Privacy Concerns

➢ Quantification Errors

Preliminary EDA findings

Nscore Histogram by Gender

Nscore Boxplot by Gender

Correlation Heatmap of Drug Consumption Classification Features

Age shows significant negative correlations with many substance uses, suggesting younger individuals are more likely to engage in these activities.

Gender influences substance use patterns, with males possibly more inclined towards certain substances.

Personality traits like Neuroticism, Extraversion, and Sensation Seeking are strongly related to substance use behaviors.

Country and Education also play a role in substance use tendencies, indicating potential cultural and socio-economic influences.

Next Steps
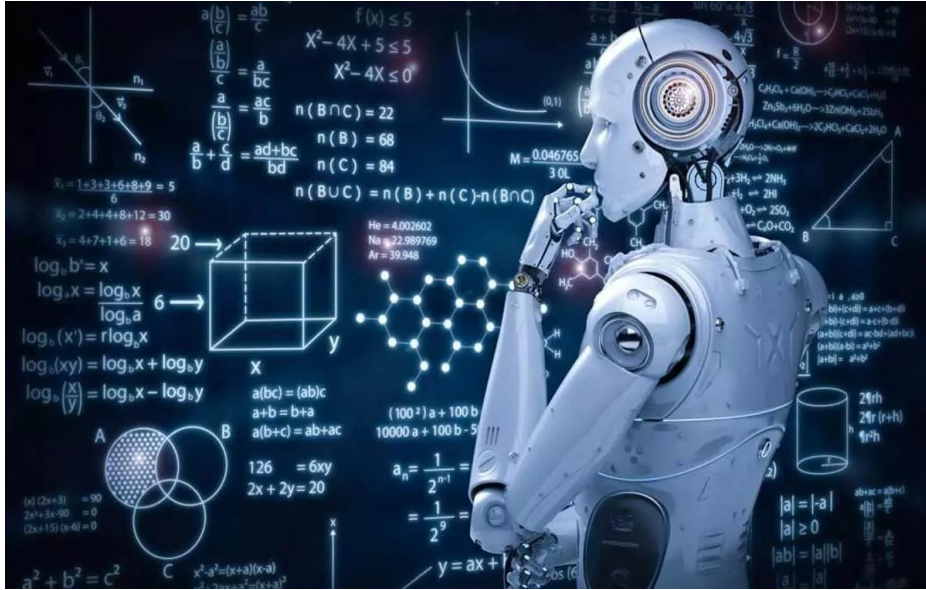
- ❖ Statistical Analysis
- ❖ Feature Engineering
- ❖ Dealing with Data Imbalance
- ❖ Data Splitting
- ❖ Model Selection
- ❖ Model Training and Hyperparameter Tuning
- ❖ Model Evaluation
- ❖ Model Interpretation and Explainability