Drug Use Decision Tree

Lucas Manker

Data

- Data downloaded from UCI ML Repository

- 1885 Respondents from 7 countries

- 32 Features

Features

1. ID
2. Age
3. Gender
4. Education Level
5. Country of Current Residence
6. Ethnicity
7. NScore (Neuroticism - anxiety, moodiness, worry, fear, anger, frustration)
8. EScore (Extraversion - enjoy human interaction, talkative, assertive)
9. OScore (Openness - openness to new experiences)
10. AScore (Agreeableness - easily persuaded to try new things)

Features cont.

11. CScore (Conscientiousness - empathy for others)
12. Impulsiveness (Tending to act on a whim)
13. ImpSS (Impulsively seeking new sensations)
14. Alcohol Use
15. Amphetamine Use
16. Amyl Nitrate Use (Inhalants popular in the EU, "poppers")
17. Benzos
18. Caffeine
19. Marijuana
20. Chocolate

Features cont.

21. Cocaine
22. Crack Cocaine
23. Ecstasy
24. Heroin
25. Ketamine
26. Legal Drugs (legal inhalants, etc)
27. LSD
28. Meth
29. Mushrooms
30. Nicotine

Features cont.

31. Semer (fake drug to weed out people not taking survey seriously)
32. VSA (Glue/Gasoline/Air Duster inhalants)

Drug Use Classes

- 6 Classes of drug use

- Never used, Used a decade ago, Used last decade, Used last year, Used last month, Used last week, Used yesterday

Data Manipulation

- First 20 features read from file and put into a matrix
- Picked 5 drugs to examine and predict based on first 20 labels
- Drug use was simplified into 3 classes:
  - Never used, Used over a decade ago, Used last decade : Not_User
  - Used last year, Used month : User
  - Used last week, Used last day : Heavy_User

Decision Tree

A decision tree using SKLearn was implemented with 70% training data and 30% testing data. A confusion matrix, along with precision, recall, and F1 scores were then computed.

Alcohol

F1-Score for NOT_USER: 0.091

F1-Score for USER: 0.186

F1-Score for HEAVY_USER: 0.866

		PREDICTED		
A C T U A L		NOT_USER	USER	HEAVY_USER
	NOT_USER	1	1	9
	USER	1	13	43
	HEAVY_USER	9	69	420

Three Most Important Features:

1. Openness to new experiences : **0.0201**
2. Extraversion: **0.0198**
3. Agreeableness: **0.0193**

Marijuana

F1-Score for NOT_USER: 0.460

F1-Score for USER: 0.143

F1-Score for HEAVY_USER: 0.731

		PREDICTED		
A C T U A L		NOT_USER	USER	HEAVY_USER
	NOT_USER	57	8	57
	USER	9	9	44
	HEAVY_USER	58	46	278

Three Most Important Features:

1. Agreeableness : **0.0444**
2. Openness to new experiences: **0.0444**
3. Extraversion: **0.0430**

Cocaine

F1-Score for NOT_USER: 0.624

F1-Score for USER: 0.184

F1-Score for HEAVY_USER: 0.444

		PREDICTED		
A C T U A L		NOT_USER	USER	HEAVY_USER
	NOT_USER	190	45	71
	USER	36	15	23
	HEAVY_USER	77	29	80

Three Most Important Features:

1. Openness to new experiences : **0.0558**
2. Agreeableness: **0.0491**
3. Extraversion: **0.0410**

Crack

F1-Score for NOT_USER: 0.875

F1-Score for USER: 0.05

F1-Score for HEAVY_USER: 0.165

		PREDICTED		
A C T U A L		NOT_USER	USER	HEAVY_USER
	NOT_USER	425	12	46
	USER	17	1	7
	HEAVY_USER	46	2	10

Three Most Important Features:

1. Neuroticism : **0.0211**
2. Extraversion: **0.0191**
3. Age: **0.0170**

Heroin

F1-Score for NOT_USER: 0.859

F1-Score for USER: 0.05

F1-Score for HEAVY_USER: 0.212

		PREDICTED		
A C T U A L		NOT_USER	USER	HEAVY_USER
	NOT_USER	404	19	49
	USER	7	1	4
	HEAVY_USER	58	8	16

Three Most Important Features:

1. Extraversion : **0.0234**
2. Ethnicity: **0.0204**
3. Neuroticism: **0.0197**

Improvements

- Marijuana use was then added as a feature to help predict harder drug use, but this actually made the model even more imprecise.

- Cocaine use was then used in a similar manner to predict Crack and Heroin use. This change increased the accuracy of predicting heavy users.

Crack v2

F1-Score for NOT_USER: 0.887 **(old score 0.875)**

F1-Score for USER: 0.045 **(old score 0.5)**

F1-Score for HEAVY_USER: 0.347 **(old score 0.165)**

1. Extraversion: **0.0242**
2. Cocaine Use: **0.0227**
3. Neuroticism: **0.0184**

		PREDICTED		
A C T U A L		NOT_USER	USER	HEAVY_USER
	NOT_USER	429	15	39
	USER	21	1	3
	HEAVY_USER	34	3	21

Three Most Important Features:

Heroin v2

F1-Score for NOT_USER: 0.876 **(old score 0.859)**

F1-Score for USER: 0.03 **(old score 0.05)**

F1-Score for HEAVY_USER: 0.32 **(old score 0.212)**

1. Ethnicity : **0.0238**
2. Cocaine Use: **0.0220**
3. Neuroticism: **0.0205**

		PREDICTED		
A C T U A L		NOT_USER	USER	HEAVY_USER
	NOT_USER	412	23	37
	USER	7	1	4
	HEAVY_USER	50	9	23

Three Most Important Features:

Questions!