

# Drug Consumption

2조

고정민 김관석 김수연

김유경 서상수

# 목차

## STEP 1

문제 및 목표 정의

이번 분석의 배경과 문제점을 파악한 후 이를 해결하기 위한 목표 정의

>>

## STEP 2

y 변수 이진 분류

각 약물별 분포 및 특성에 따라 threshold를 정해 y변수 이진 분류

>>

## STEP 3

모델 설명

성격적인 특성을 반영해 약물 중독에 대해 설명할 수 있는 모델과 예측에 초점을 둔 모델 제시

>>

## STEP 4

분석 결과

각 모델의 분석 결과에 따라 내용 정리 및 예측 모델을 통한 개인별 약물 사용 예측

# 문제 및 목표 정의

## 배경

- ✓ 2017년 트럼프의 opioid crisis 선언  
마약성 진통제 '오피오이드' 남용에 대한 공중보건 위기사태를 선포함.
- ✓ 코로나로 인한 약물중독 사망자 수 30% 증가  
방역 조치로 인해 약물 중독자를 위한 프로그램이 전혀 진행되지 못 함.  
거리두기로 인한 외로움, 불안감 등이 커진 것도 주요 요인으로 생각됨.

## 과제 목표

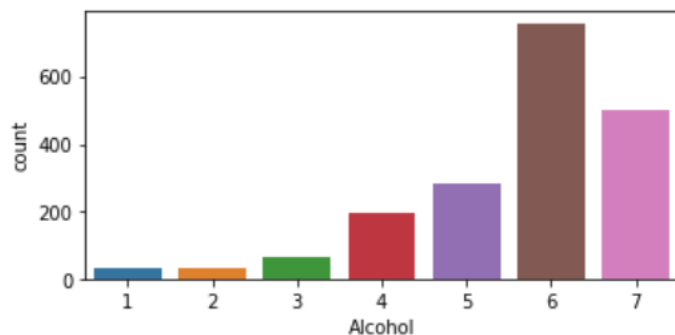
- ✓ 주어진 정보를 바탕으로 한 사람의 약물 중독 여부를 예측
- ✓ 약물의 중독성을 파악하여 위험 대비
- ✓ 나아가 성격적인 특성이 약물 중독에 얼마나 영향을 미치는지 분석



필라델피아 켄싱턴 마약 (좀비?)

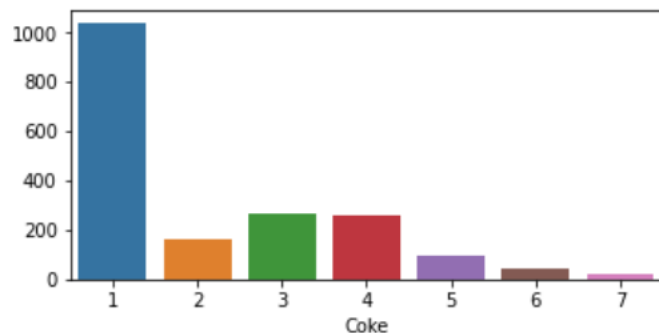
# y 변수 이진 분류

- ✓ CL0 사용 X / CL1 사용 X 10년 ↑ / CL2 10년 안에 사용 / CL3 1년 안에 사용 /
- ✓ CL4 한달 안에 사용 / CL5 1주 안에 사용 / CL6 하루 안에 사용



## 합법 약물

- ✓ Alcohol, Caff, Choc, Nicotine
- ✓ 모두 주변에서 쉽게 구할 수 있는 약물에 해당하며, 일정 수준 사용은 중독이라고 볼 수 없기 때문에 0-3 / 4-6 기준으로 Non-user / User 분류

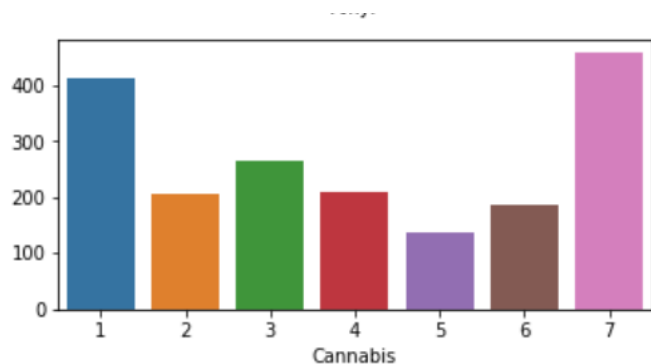


## 불법 약물 (대부분)

- ✓ Amphet, Amyl, Benzos, Coke(코카인), Crack, Ecstasy, Heroin, Ketamine, Legalh, LSD, Meth, Mushrooms(환각버섯), VSA
- ✓ 모두 주변에서 쉽게 구할 수 없는 약물에 해당하며, 사용한 사람이 거의 없기 때문에 0-1 / 2-6 기분으로 Non-user / User 분류

# y 변수 이진 분류

- ✓ CL0 사용 X / CL1 사용 X 10년 ↑ / CL2 10년 안에 사용 / CL3 1년 안에 사용 /
- ✓ CL4 한달 안에 사용 / CL5 1주 안에 사용 / CL6 하루 안에 사용



## 대마초

- ✓ Cannabis
- ✓ 지역에 따라 불법, 합법이 나뉘며 마약 중에서 가장 잘 알려져 있는 만큼 사용자가 많기에 threshold를 조정하여 0-2 / 3-6 기준으로 Non-user / User 분류

## 분석

- ✓ SS는 Impulsive와의 Correlation을 고려해 모델에 포함하지 않고 진행해 위 약물 사용자에게 대한 성격적 특성 파악!

# 모델 설명

## 설명 가능한 모델

- ✓ Logistic Regression

로지스틱 회귀를 사용해 회귀 계수를 통한 약물별 성격 특성 요인의 영향에 대해 분석

- ✓ Decision Tree

Decision Tree를 사용해 Tree 시각화 및 Tree를 통한 변수별 해석 제공

## 예측모델

- ✓ Neural Net

뉴럴 네트워크 모델을 통해 예측 성능을 향상시켜 조원들의 약물 중독 위험성을 판별

# 데이터 분석

## A. 인구통계

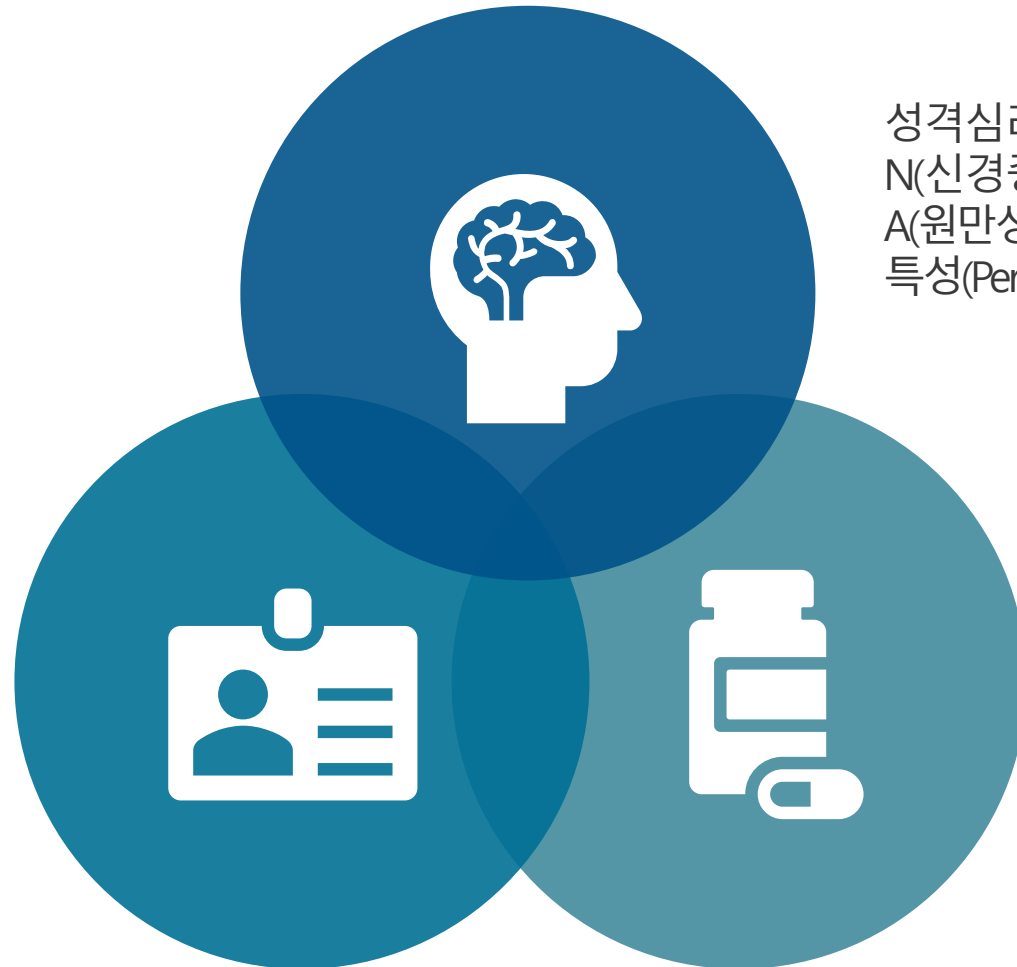
연령, 성별, 학력수준, 국가, 인종 등의 기본적인 인구통계적인 특성(Demographic traits)에 대한 분석

## B. 성격지표

성격심리학의 대표적인 지표 - N(신경증), E(외향성), O(개방성), A(원만성), C(성실성) 등 성격적인 특성(Personality traits)에 대한 분석

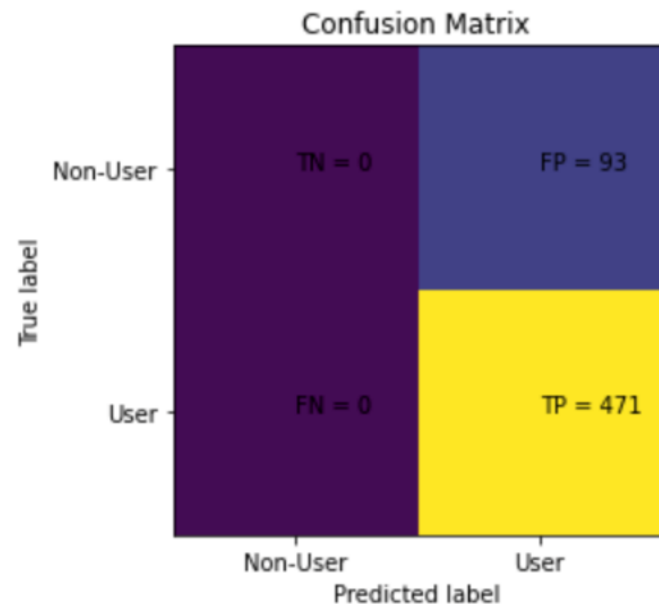
## C. 목적변수

알코올, 암페타민, 헤로인 등 총 18가지 약물 간의 관계 및 그에 대한 분석



# 분석 결과 – 로지스틱 회귀

## A. 합법 약물 – Alcohol, Caff, Choc, Nicotine



✓ 모델 학습이 제대로 되지 않아 예측된 Non-user가 아예 없는 경우도 나타남- Alcohol, Caff, Choc (좌측 confusion matrix)

✓ **Nicotine** (우측 하단 coef)

- 담배에 함유된 중독성 물질
- 남성의 경우 User일 확률 높음
- O(개방성) 높을수록 User일 확률 ↑
- C(성실성) 높을수록 Non-user일 확률 ↑
- Impulsive(충동성) 높을수록 User일 확률 ↑
- 충동성, 개방성 등의 성격 특성은 니코틴 사용에 긍정적 효과를, 성실성은 부정적 효과 보임

### Nicotine

	coef	std err	z	P> z
Gender_1	-0.4463	0.136	-3.285	0.001
Age	-0.2075	0.042	-4.963	0.000
Education	0.0846	0.028	3.040	0.002
Nscore	0.7300	0.400	1.826	0.068
Escore	-0.1384	0.488	-0.284	0.777
Oscore	2.2997	0.438	5.248	0.000
Ascore	-0.3998	0.456	-0.876	0.381
Cscore	-1.6159	0.455	-3.551	0.000
Impulsive	1.5652	0.404	3.870	0.000



# 분석 결과 – 로지스틱 회귀

## B. 불법 약물 – Amphet, Musrooms / Cannabis

### Amphet

	coef	std err	z	P> z
Gender_1	-0.8324	0.132	-6.301	0.000
Age	-0.2253	0.047	-4.808	0.000
Education	0.0936	0.028	3.339	0.001
Nscore	-0.1851	0.389	-0.476	0.634
Escore	-0.9353	0.472	-1.983	0.047
Oscore	2.3010	0.433	5.313	0.000
Ascore	-0.8856	0.442	-2.005	0.045
Cscore	-2.0215	0.444	-4.550	0.000
Impulsive	1.5308	0.398	3.844	0.000

### Fresh Cannabis



### Amphet / Mushrooms

- ✓ 암페타민은 피로와 식욕을 낮추고 기민성을 증가시키는 각성제
- ✓ ADHD, 기면증, 비만증 등의 치료제
- ✓ O(개방성)이 높은 사람일수록 약물 사용빈도 ↑ (오픈마인드?)
- ✓ 성실(C)한 사람들은 약물과 거리가 먼 듯 함.

### Cannabis

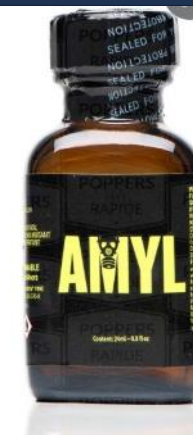
- ✓ 대마초, 마리화나와 비슷한 마약성분
- ✓ O(개방성), C(성실성), E(외향성)/A(신경증) 순으로 연관성 보임
- ✓ O(개방성)이 높은 사람일수록 약물 사용빈도 ↑
- ✓ 성실(C)한 사람들은 약물과 거리가 먼 듯 함 -> 위의 경우와 동일

# 분석 결과 – 로지스틱 회귀

## C. 불법 약물 – Amyl, Benzos, Coke,

**Amyl**: 클럽 약물로 알려져 있으며, 일부 사람들은 성관계 개선 위해 사용하기도 함

- ✓ 남성일수록 User ↑ / E(외향성) 높을수록 User ↑ - 클럽과의 관계 / C(성실성) 높을수록 Non-user ↑



**Benzos**: 신경안정제에 속하는 항정신성의약품 / 사용 중단 후 간질 발작 등 중독성 (3일분 이상 판매 x)

- ✓ N(신경증), O(개방성) 높을수록 User ↑ - 신경 안정제 역할
- ✓ E(외향성), A(원만성), C(성실성) 높을수록 Non-user ↑ - 대인관계 원만하고 사교적, 성실하면 안 먹음

**Coke**: 코카인은 쾌감, 집중력, 창의성 ↑ / 계속 복용 시 수면 장애, 인성장애, 폭력, 반사회, 정신적 장애

- ✓ O(개방성), Impulsive(충동성) 높을수록 User ↑
- ✓ A(원만성), C(성실성) 높을수록 Non-user ↑ - 대인관계 원만하고 성실하면 안 먹음



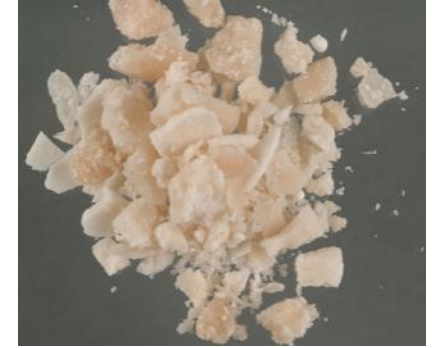
Cocaine

# 분석 결과 – 로지스틱 회귀

## D 불법 약물 – Crack, Ecstasy

**Crack** : 코카인+베이킹소다 가열 -> 코카인보다 강력, 약하게 지속 / 저렴해 흑인 갱단 유행

- ✓ 남성일수록, O(개방성), Impulsive(충동성) 높을수록 User ↑
- ✓ E(외향성), A(원만성), C(성실성) 높을수록 Non-user ↑



**Ecstasy** : 암페타민보다 싸고 환각 3~4배 / 중독성 상대적으로 가볍 / 알약이라 사용 급증

- ✓ 남성일수록 사용 많이 하며, O(개방성)과의 양의 관계가 다른 약물보다 월등히 높아 쉽게 구할 수 있어서 한 번 해 볼 여지 때문.
- ✓ C(성실성) 높을수록 Non-user ↑



전체적으로 C(성실성)은 높을수록 약물 사용 안하는 것으로 나타남.

# 분석 결과 – 로지스틱 회귀

E 불법 약물 – Heroin, Ketamine, Legalh, LSD, Meth, VSA,

**Heroin, Ketamine, Legalh, LSD, Meth, VSA 공통**

- ✓ E(외향성) 높을수록 Non-user ↑ - 건강하고 활력 있는 삶
- ✓ O(개방성) 높을수록 User ↑ - 모든 약물 공통
- ✓ A(원만성), C(성실성) 높을수록 Non-user - 원만하여 타인과 잘 어울리고, 성실하기 때문
- ✓ Impulsive(충동성) 높을수록 User ↑

Heroin



Ketamine



LSD



Meth

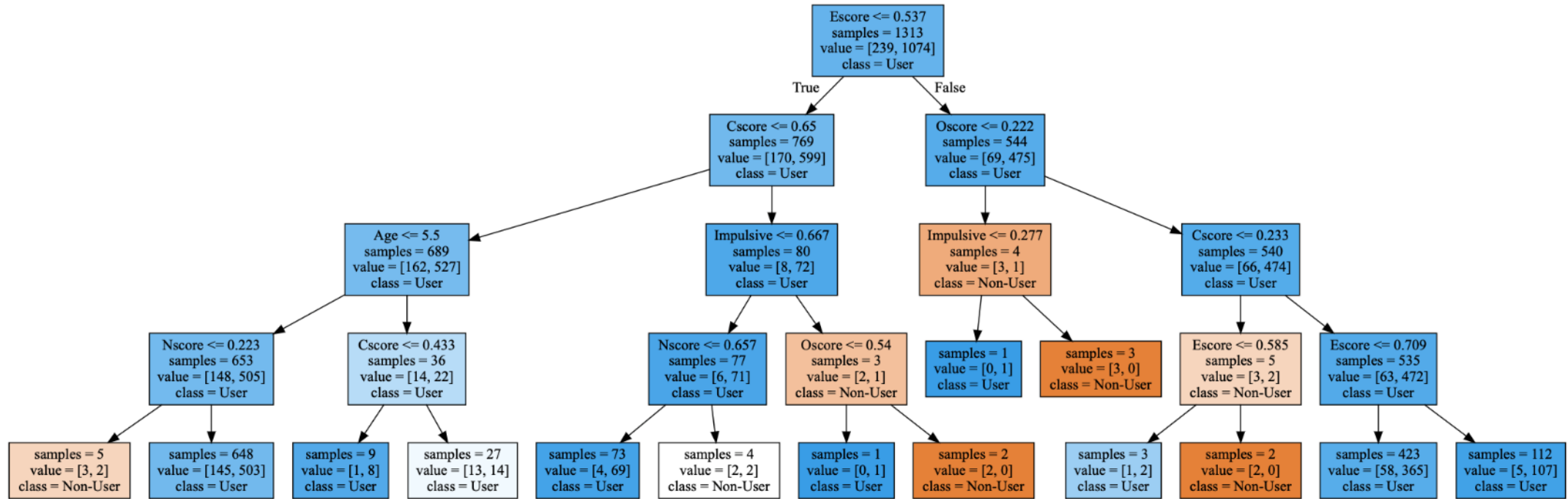


VSA



# 분석 결과 – Decision Tree

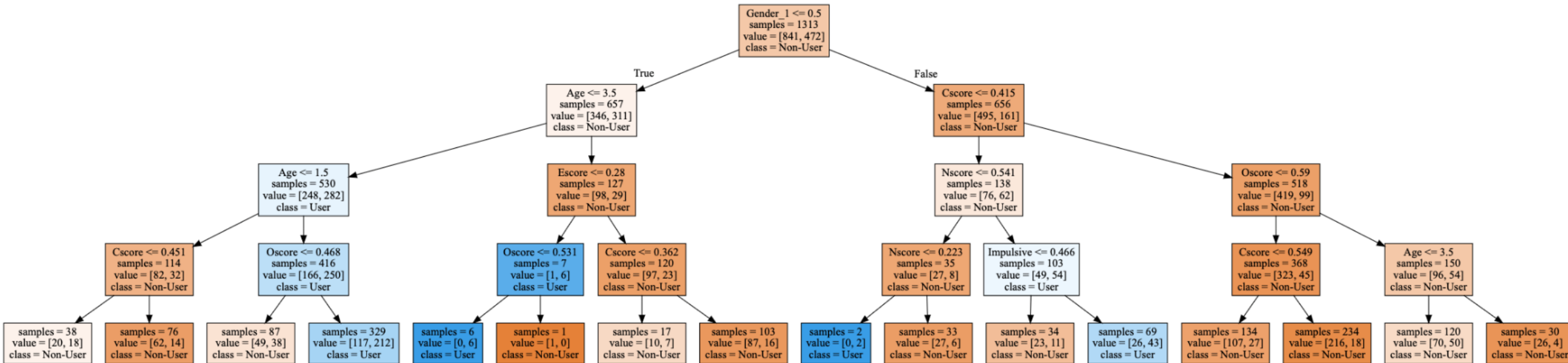
## A. 합법 약물 – Alcohol, Caff, Choc, Nicotine



- ✓ 합법 약물의 경우, Country 변수의 유무에 상관 없이 성격 지표의 영향이 더 컸다.
- ✓ 애초에 환경에 구애받지 않는, 접근성 높은 약물들이기에 인구통계학적 변수들은 영향력이 적을 수밖에 없는 것으로 보인다.

# 분석 결과 – Decision Tree

B. 불법 약물 – Amphet, Amyl, Benzos, Coke(코카인), Crack, Ecstasy, Heroin, Ketamine, Legalh, LSD, Meth, Mushrooms(환각버섯), VSA

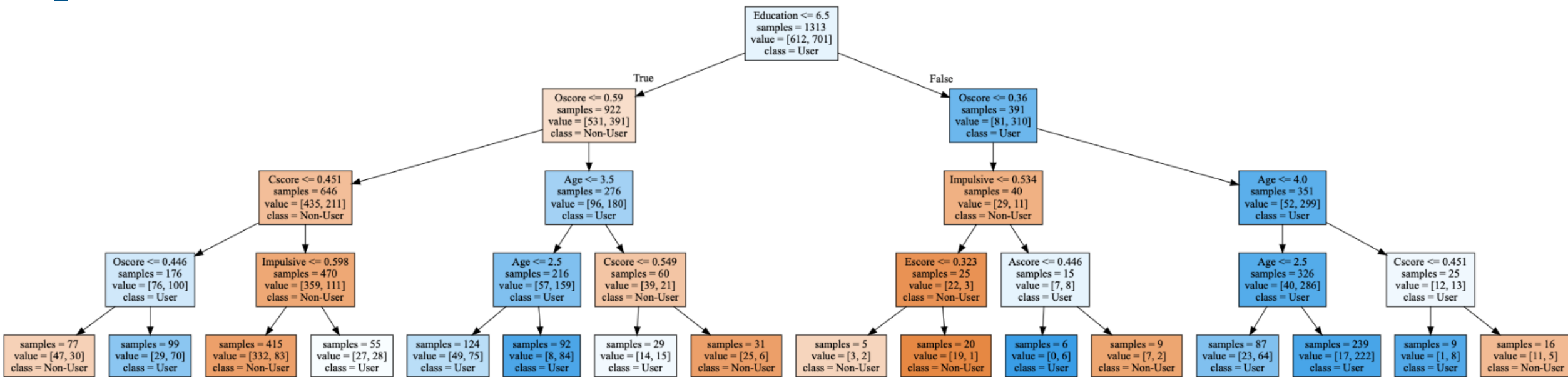


- ✓ 부분적 합법 약물 또는 불법 약물의 경우에는, 아직까지는 인구통계학적 변수가 성격 변수보다 높은 feature importance를 갖는 것으로 보인다!
- ✓ 접근성 낮고 환경의 제약이 큰 불법 약물들은 거주 국가 변수 다음으로 성별 변수가 영향을 끼친다.
- ✓ 압도적이지는 않지만 주로 남자가 더 잠재 중독자가 될 가능성을 보인다!



# 분석 결과 – Decision Tree

## C. Cannabis



- ✓ 다만, Cannabis와 같이 어느 정도 접근성이 있는 약물의 경우에는, 최상위 노드로 교육수준 변수가 위치해 있어, 대학 수료를 기점으로 분류를 하고 있으나, 결과적으로 보면 분류된 2개의 class 안의 User의 수가 비슷하기 때문에 크게 유의미한 변수라고 단정짓기는 어려울 것으로 보인다.
- ✓ 오히려, 주목해봐야 할 것은 O(개방성)인데, 이 점수가 높을수록 User로 분류되는 경우가 많다. 따라서 불법 약물임에도 불구하고, 미약하게나마 환경적 제약을 극복한(?) 이러한 약물들은 성격 지표들을 고려하지 않을 수 없겠다.

# 데이터 분석

## 정리

### 합법 약물

- ✓ 사용한 분석 모델을 통해서는 해당 약물들의 사용에 대해 제대로 설명하기 어려움.
- ✓ 애초에 환경에 구애받지 않는, 접근성 높은 약물들이기에 인구통계학적 변수들은 영향력이 적을 수밖에 없는 것으로 보임.

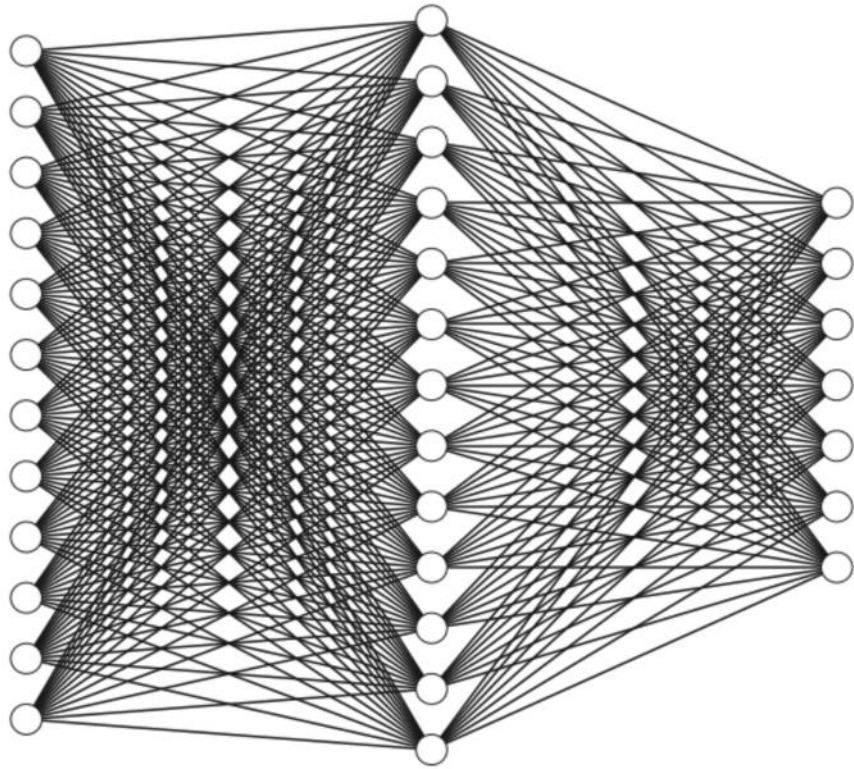
### 불법 약물

- ✓ 남성일수록, O(개방성), Impulsive(충동성) 이 높을수록 많이 사용하는 경향이 있음
- ✓ C(성실성), A(원만성), E(외향성)이 높을 수록 덜 사용하는 경향이 있음.
- ✓ 거주 국가에 따라 접근성이 달라지기 때문에 이 부분도 큰 영향을 끼침.
- ✓ 약물 특성에 따라 N(신경증)이 관여하는 경우가 있음.



# 예측 – Neural Network

## A 모델 구조



Neural Network model

- ✓ Naïve 한 형태로 Hidden Layer 1개짜리 뉴럴 네트워크 모델 구성
- ✓ 해석보다는 모델을 통한 예측에 목적을 둔 모델.
- ✓ Hidden Layer가 2개 이상인 경우 딥러닝 모델이라고 함.

# 예측 – Nerual Net

## B. 설문조사

1) I seek out the patterns of the universe.  
 Very Inaccurate ☐ ☐ ☐ ☐ ☐ Very Accurate

2) I am scientific.  
 Very Inaccurate ☐ ☐ ☐ ☐ ☐ Very Accurate

3) I follow a schedule.  
 Very Inaccurate ☐ ☐ ☐ ☐ ☐ Very Accurate

4) I prefer very structured environments.  
 Very Inaccurate ☐ ☐ ☐ ☐ ☐ Very Accurate

5) I am talkative.  
 Very Inaccurate ☐ ☐ ☐ ☐ ☐ Very Accurate

6) I put myself first.  
 Very Inaccurate ☐ ☐ ☐ ☐ ☐ Very Accurate

Extroversion		64%
Orderliness		48%
Emotional Stability		74%
Accommodation		34%
Inquisitiveness		54%

✓ <https://similarminds.com/bigfive.html>

✓ 위의 홈페이지에서 각자 설문조사 진행!

✓ 조사 결과에 따라 왼쪽과 같이 Big 5 성격 지표에 대한 결과를 얻을 수 있음.

✓ 해당 자료와 인구통계학적 특성들을 넣어 약물 사용 예측 진행.



- ✓ CL0 사용 X / CL1 사용 X 10년 ↑ / CL2 10년 안에 사용 / CL3 1년 안에 사용 /
- ✓ CL4 한달 안에 사용 / CL5 1주 안에 사용 / CL6 하루 안에 사용

[illegible]

# 예측 – Nerual Net

D.참 값 -> 틀린것 하이라이트

Drug	알콜	카페인	초코	Amp het	Amyl	Bezn os	Coke	Crac k	Ecsta sy	Hero in	Ketai me	Leag alh	LSD	Meth	Mus h	Nico tine	VSA	cann abis
정민	4	3	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2??
관석	5	6	5	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0
수연	5	6	5	0	0	0	0	0	0	0	0	0	0	0	0	0!!	0	0
유경	1	5	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

- ✓ 전체적으로 쉽게 구할 수 있는 약물에 대해서는 5나 6으로 예측했으나 사람에 따라 맞는 사람과 안 맞는 사람이 있었다. (예측 힘들것으로 예상)
- ✓ 불법 약물에 대해서는 다 0으로 예측했으며 Cannabis(대마)의 경우는 다 6으로 예측했으나 우리나라 특성 상 불가능하다. 그런데 정민 님은 2이다...?



thx