Demographics and Products' Influence on Patient Emotional Desires

Abel Lula, Alex Clare, Amanda Pate,
Jodie Chen, Rigoberto Garcia, Ruyi Lu

TABLE OF CONTENTS

INTRODUCTION

- Stigma around cannabis and marijuana products dissipating & exponentially growing
- Pertinent to understand people's needs and how cannabis products can help.

01

Overview

- Statement of the Problem, Abstract, Codebook

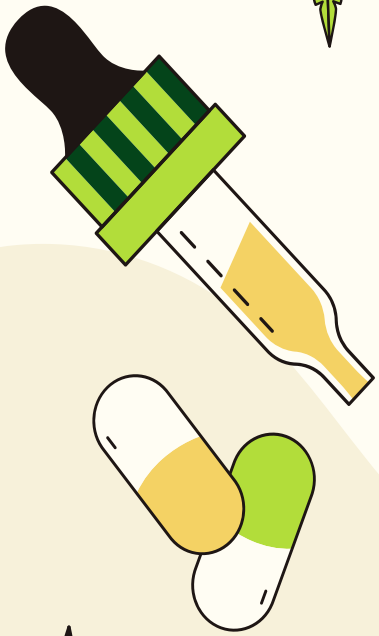
Problem Statement

➔ What demographics and, most importantly, which cannabis products are most influential in deciding how a patient wants to feel?

ABSTRACT

- Use logistic and random forest models to see what products and demographics most influence how a person wants to feel
  - Logistic models: better accuracies and demonstrated how certain cannabis products are more influential for the desired emotions
- Goal: to recommend certain products depending on what someone wants to feel!

About Our Dataset



- Dataset provided by Veriheal,
  - Patients have the option to fill out a survey on cannabis consumer demographics
  - Data used from 10/14/2023-11/13/2023
- 9192 observations in total
- 30 variables after data cleaning

Important Predictors

- **AGE:** numerical, indicates the age of each patient
- **STATE:** categorical, indicates what region in America each patient is from: Midwest, Northeast, South, or West
- **Flowers, Oils, Vape, Conc, CBD, Eds:** numerical, 1 for patient being interested in the product, 0 for not interested

Response Variables: People's Feelings

- **Pain, Happy, Sleep, Chill, Relax, Focus, Creative**
- Indicates whether or not the patient was interested in each feeling listed in the survey. 1 for interested, 0 for not interested

02

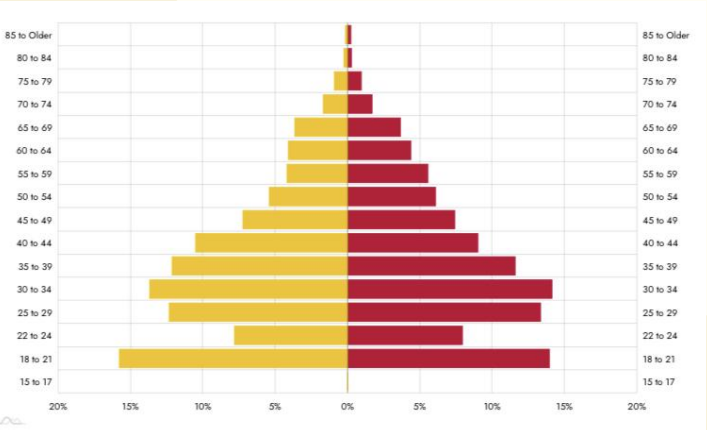
EDA

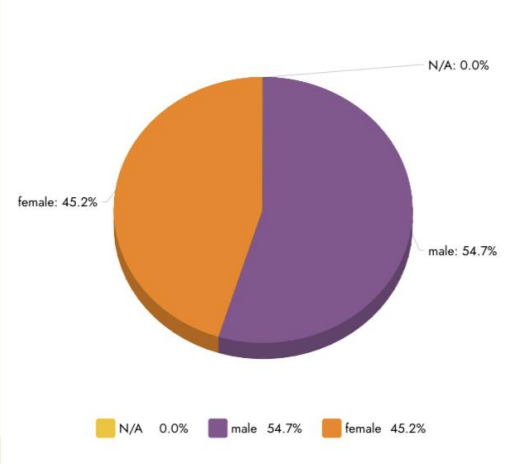
Exploring the Data

Basic Demographics

Age

Age distribution of survey participants





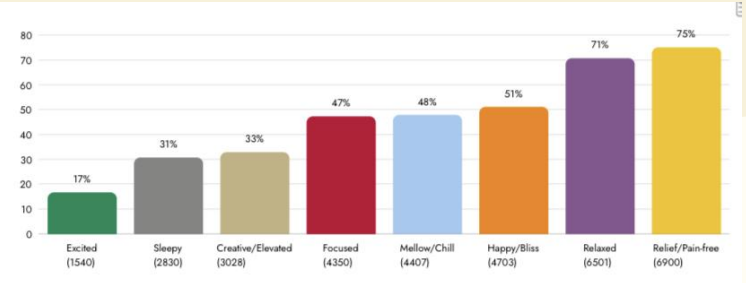
Gender

Gender distribution of survey participants

Cannabis Demographics

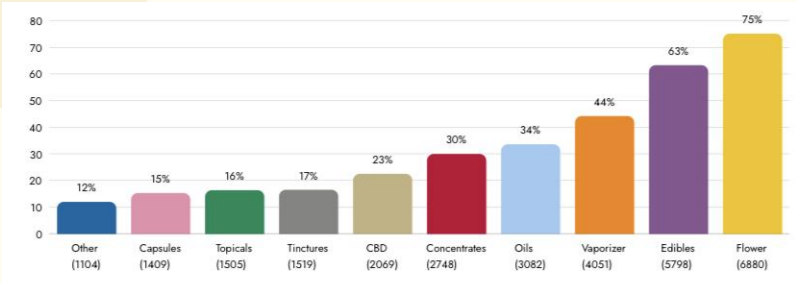
Feelings

How participants wanted to feel



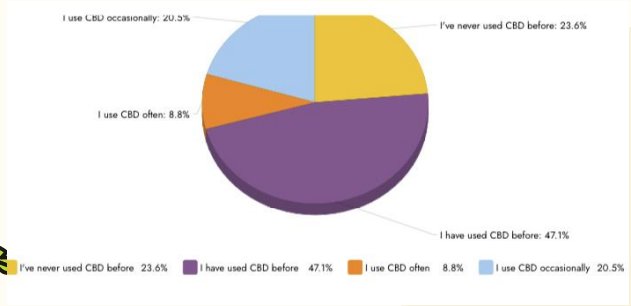
Products

What products participants were interested in



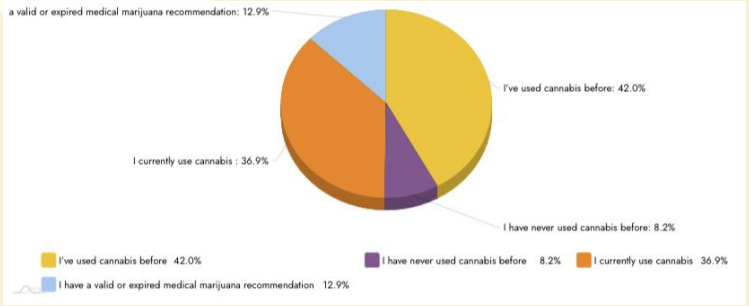
CBD Experience

Participants' experience with CBD



Cannabis Experience

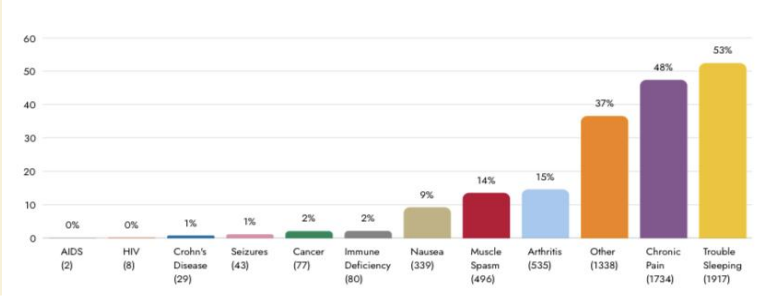
Participants' experience with cannabis



Medical History

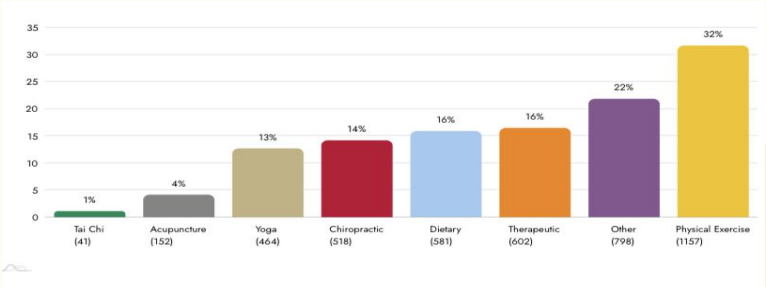
Conditions

Distribution of conditions participants have



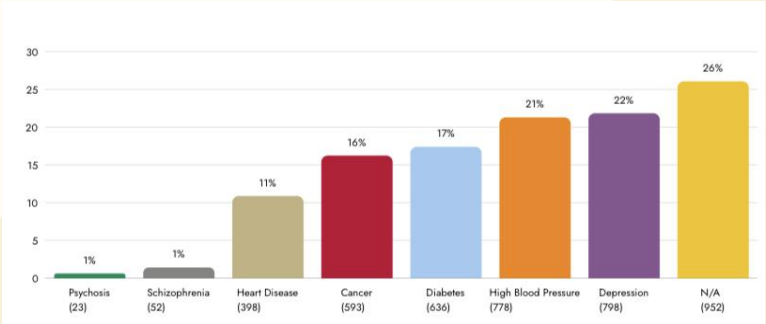
Family History

Distribution of conditions within participants' family history



Alternatives Tried

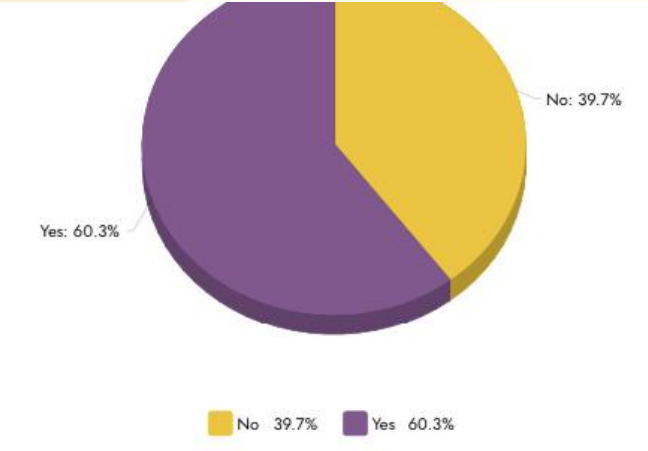
Alternative treatment methods tried for conditions

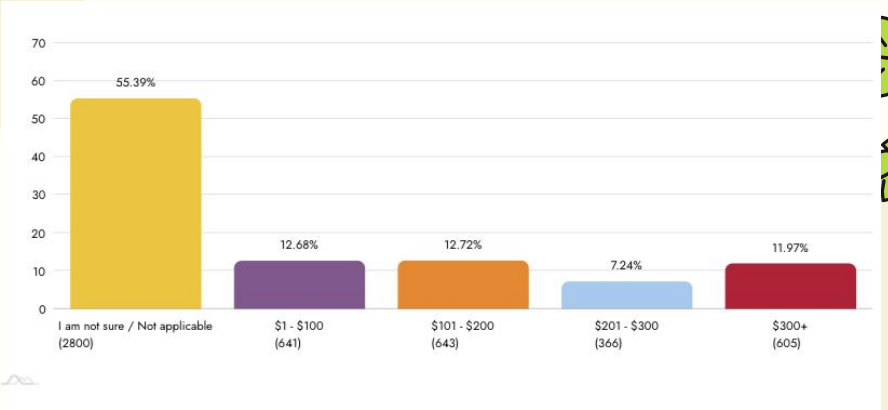


Medical History Cont.

Health Insurance

Whether or not participants have health insurance





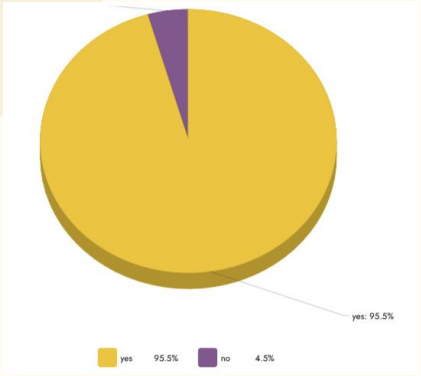
Cost

The distribution of costs for those who do

Dispensary Demographics

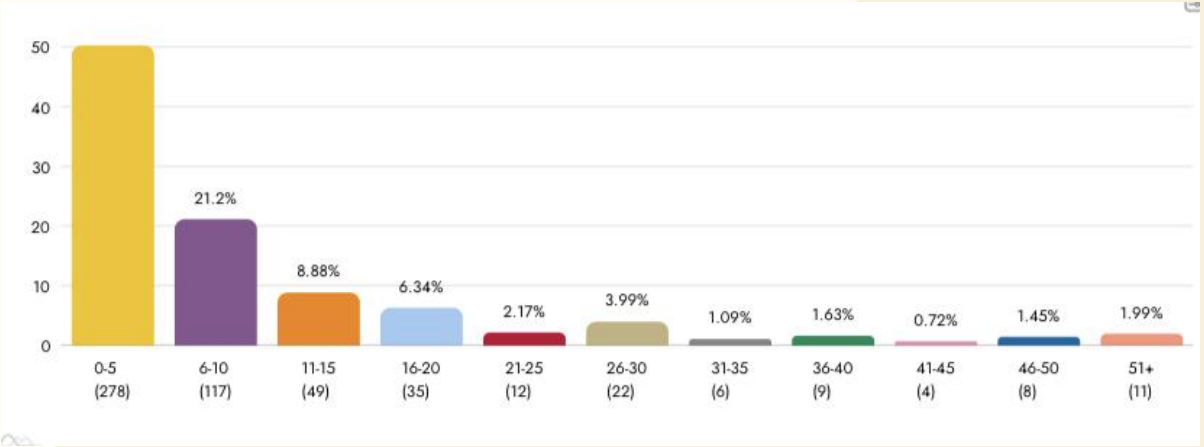
Dispo Enjoyment

Percentage of participants enjoying their dispensary



Travel

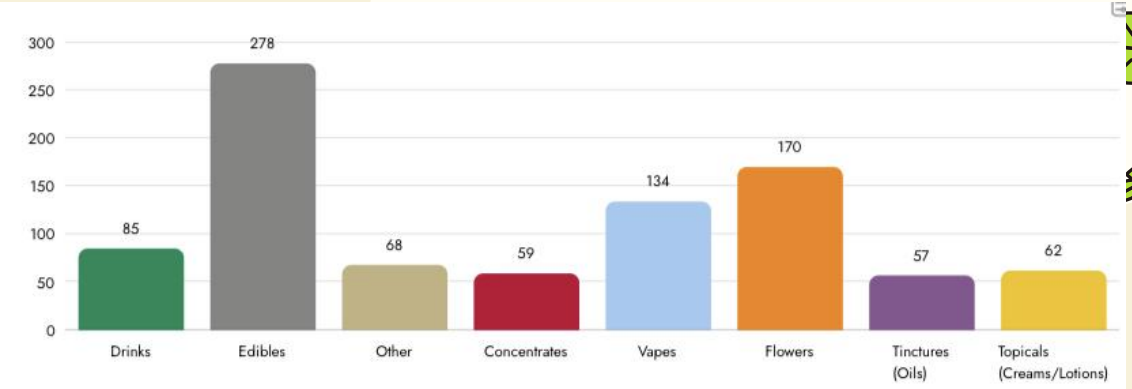
Distribution of how far participants drive to their dispensary



Dispensary Demographics Cont.

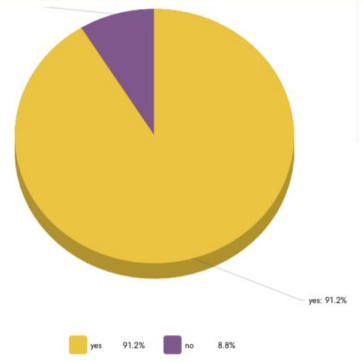
More

Products participants wish their dispensary had more of



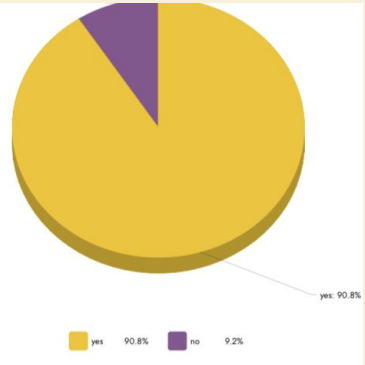
Product Selection

Percentage of participants that feel their dispensary has a wide selection of products



Strain Selection

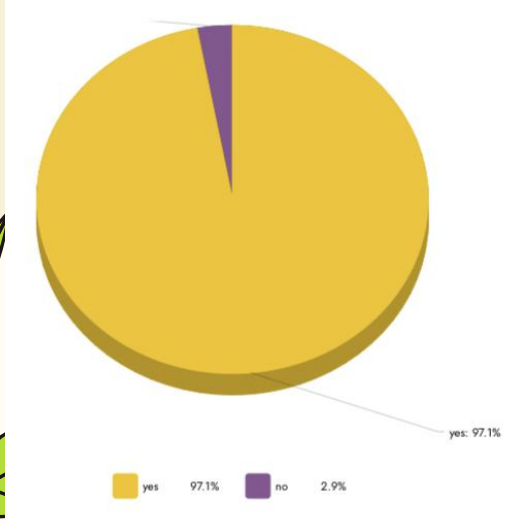
Percentage of participants that feel their dispensary has a wide selection of strains

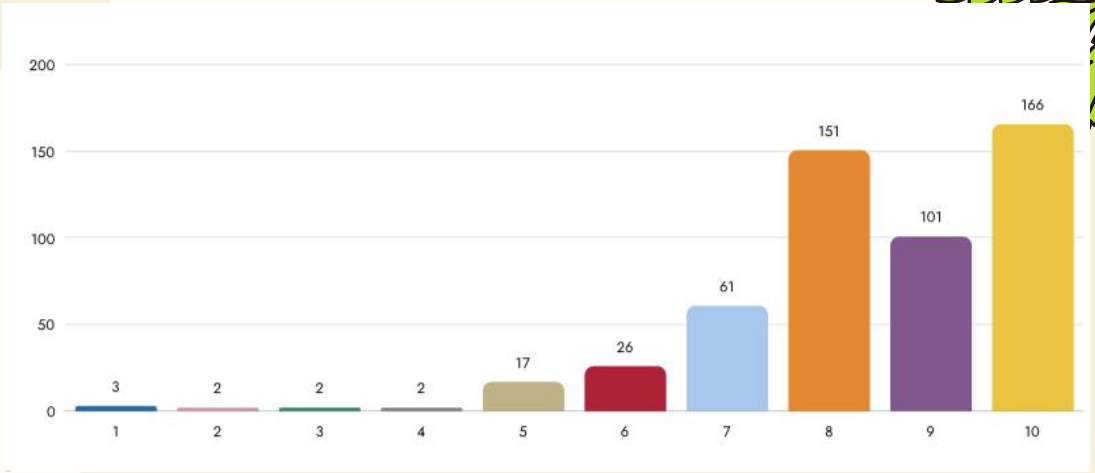


Dispensary Demographics Cont.

Products Effective?

Percentage of participants that feel their products are effective for their conditions





Effectiveness

Scale of 1-10 for effectiveness of cannabis as treatment

03

ANALYSIS

Logistic and Random Forest Models

Logistic Model: Feature Selection

- After fitting a logistic regression model, for each emotion state, performed **stepwise BIC** in both directions for each model
  - Ex: For pain-free model, lowest BIC score was achieved when gender and vape was removed → best model

Backwards BIC Logistic Model (pain-free) Output

Coeffs	Estimate	P-value	Significant?
STATEMidwest	0.627	< 2e-16	Yes
STATENortheast	-0.105	< 2e-16	Yes
STATESouth	-0.059	1.10e-06	Yes
STATEWest	-0.025	0.141	**No**
AGE	0.002	< 2e-16	Yes
Flowers	0.037	0.0004	Yes
Oils	0.05	9.57e-07	Yes
Conc	0.049	3.42e-06	Yes
CBD	0.075	1.59e-11	Yes
Eds	0.069	4.16e-13	Yes

Backwards BIC Logistic Model: Product and Region Influence

Feeling	Most Influential Product	Most Influential Region
PAIN-FREE	CBD	Northeast
HAPPY	Flowers	West
SLEEP	Conc	None
CREATIVE	Flowers	South
RELAX	Eds	Northeast
CHILL	Flowers	None
FOCUS	Flowers	None

Random Forest Models

Model Setup and Feature Ordering

1

Setup

For each model: built with the 8 predictors selected, and each emotion as response variable

2

Close Up

For each model: 3 features selected in each tree, 500 trees in total

3

Prediction

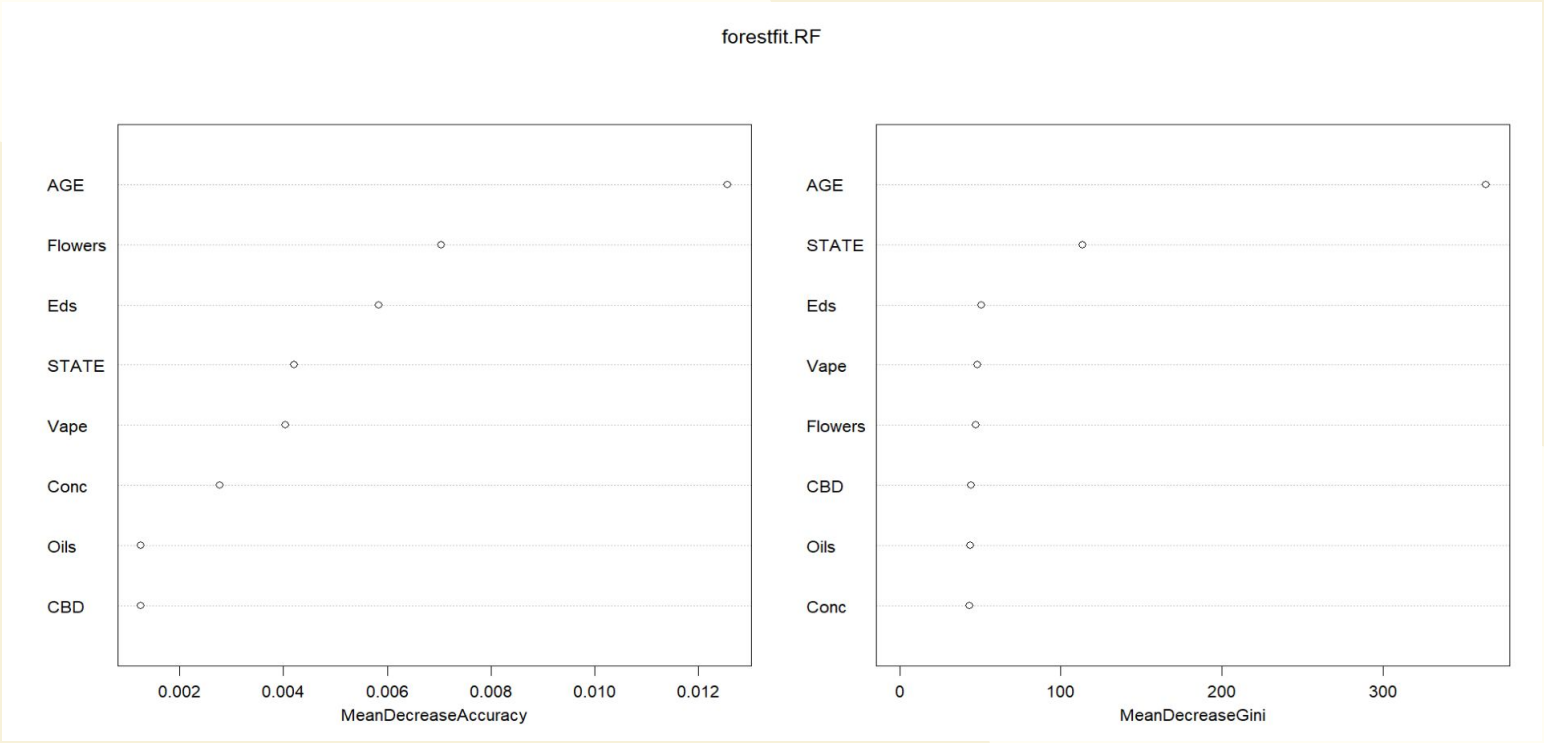
Calculate misclassification rate for each model

4

Ordering

Compare relative importance of the products and demographics

Random Forest: Feature Importance Plot for “Pain-Free”



Random Forest: Overall Importance

Feeling	Relative Importance of all Features							
PIN-FREE	Age	State	Flowers	Eds	Vape	Conc	Oils	CBD
HAPPY	Age	Flowers	Eds	Vape	Conc	State	Oils	CBD
SLEEP	Age	Conc	Vape	State	CBD	Flowers	Oils	Eds
CHILL	Age	Eds	Flowers	Vape	State	Conc	Oils	CBD
CREATIVE	Conc	Age	Oils	Flowers	Vape	State	Eds	CBD
RELAX	Age	Flowers	Eds	Vape	State	Oils	Conc	CBD
FOCUS	Oils	Vape	Age	Flowers	Eds	State	Conc	CBD

Demographics: generally, age is more important than geographic location

Random Forest: Demographics Importance

Feeling	Most Important → Least Important				
PAIN-FREE	Age	STATEMidwest	STATENortheast	STATESouth	STATEWest
HAPPY	Age	STATEWest	STATENortheast	STATESouth	STATEMidwest
SLEEP	Age	STATESouth	STATEMidwest	STATENortheast	STATEWest
CHILL	Age	STATEWest	STATENortheast	STATEMidwest	STATESouth
CREATIVE	Age	STATEWest	STATENortheast	STATESouth	STATEMidwest
RELAX	Age	STATENortheast	STATESouth	STATEWest	STATEMidwest
FOCUS	Age	STATEMidwest	STATEWest	STATESouth	STATENortheast

Demographics: for example, stronger desire for “Relax” for patients living in states in the Northeast

Random Forest: Product Importance

Feeling	Most Important → Least Important					
PAIN-FREE	Flowers	Eds	Vape	Conc	Oils	CBD
HAPPY	Flowers	Eds	Vape	Conc	Oils	CBD
SLEEP	Conc	Vape	CBD	Flowers	Oils	Eds
CHILL	Eds	Flowers	Vape	Conc	Oils	CBD
CREATIVE	Conc	Oils	Flowers	Vape	Eds	CBD
RELAX	Flowers	Eds	Vape	Oils	Conc	CBD
FOCUS	Oils	Vape	Flowers	Eds	Conc	CBD

EXAMPLE CONFUSION MATRICES

Random Forest Model for 'Sleep'

Actual \ Predicted	0	1
0	6270	2746
1	102	74

Logistic Model for 'Happy'

Actual \ Predicted	0	1
0	2607	1533
1	1901	3151

MORE EXAMPLE CONFUSION MATRICES

Random Forest Model for Pain Free

Actual \ Predicted	0	1
0	111	89
1	2220	6772

Logistic Model for Chill

Actual \ Predicted	0	1
0	3239	2030
1	1565	2358

MODEL MISCLASSIFICATION RATES

Model	PAIN-FREE	HAPPY	SLEEP	CHILL	CREATIVE	RELAX	FOCUS
Logistic	25.36%	37.36%	30.68%	39.11%	31.05%	29.53%	39.78%
Random Forest	25.12%	37.62%	30.98%	38.99%	31.23%	29.66%	39.53%
Better Model	Random Forest	Logistic	Logistic	Logistic	Logistic	Logistic	Random Forest

04

Conclusions

Model Similarities

- Most of the random forest models found age and state to be important
- A specific cannabis product always one of the top 3 most important factor
- After feature selection in logistic models, gender no longer as a significant factor

Influential and Important Products

- We can see what products are missing across two logistic models to distinguish desired products based on their feelings
- From the random forest models, we can see the relative importance of the products
    - e.g, the most influential product for ‘Pain-Free’ is Flowers

05

Beyond this Project

Future Applications

Future Studies

1

Longitudinal Studies: changes in patient preferences and outcomes over time

2

Medication Interaction: how medication interactions influence patient preferences

Potential Applications

1

Personalized Patient Recommend ations

2

Product Development and Marketing

3

Clinical Trials Design

Shortcomings

1. Lack of data
2. Quality of data
3. Modeling bias

Thank You!

We hope you learned something new

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon** and infographics & images by **Freepik**