



# Camping vs Ultralight Backpacking

Categorizing potential  
customers for targeted  
advertising



**Background**

**01**



**Model Development**

**04**



**Data Collection**

**02**



**Results**

**05**



**Data Cleaning &  
Exploratory Analysis**

**03**



**Limitations & Next  
Steps**

**06**



# Ultralight Backpacking vs Car Camping



- Defining the groups
- Gear Choices
- Relevance for advertising

# Data Collection

- Ultralight Subreddit: "r/Ultralight is the largest online Ultralight Backcountry Backpacking community! This sub is about overnight backcountry backpacking, with a focus on moving efficiently, packing light, generally aiming at a sub 10 pound base weight, and following LNT principles. Join us and ask yourself the question: Do I really need that?"
- Camping Subreddit: "A subreddit for campers concerned more about the act of camping and less concerned about hiking long distances or light gear. Primarily for tent/hammock camping. No RV camping here."

# Data Cleaning Decisions

- Combining Datasets
- Dropping Missing Values
- Dropping any short posts
- Dropping duplicate posts

# Exploratory Data Analysis

- The average length of posts in r/ultralight is 4x longer than the average length of posts in r/camping
- Surprisingly, on average, the posts in r/ultralight are more positive than the posts in r/camping

## Ultralight

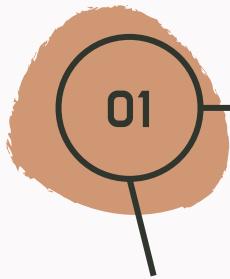
Got Hike Bag  
Don Use  
Miles Tent  
www Lighterpack  
**Hiking** Water  
Pack Weight  
Day Just  
Trail Com  
https

## Camping

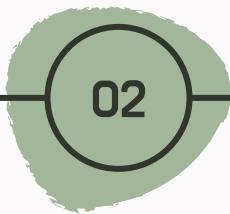
Tent Trip Gear Time  
Looking Really Like Use  
Make New **Sleeping** Car  
Need People Thanks  
Good Want Night  
People Know  
Want Don Going  
Camp Just  
**Camping**

# **Model Development Work Flow**

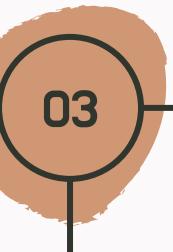
## **Training the Model**



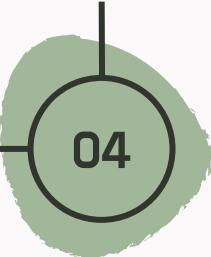
**Pre-processing  
the data:  
TF-IDF Vectorizer**



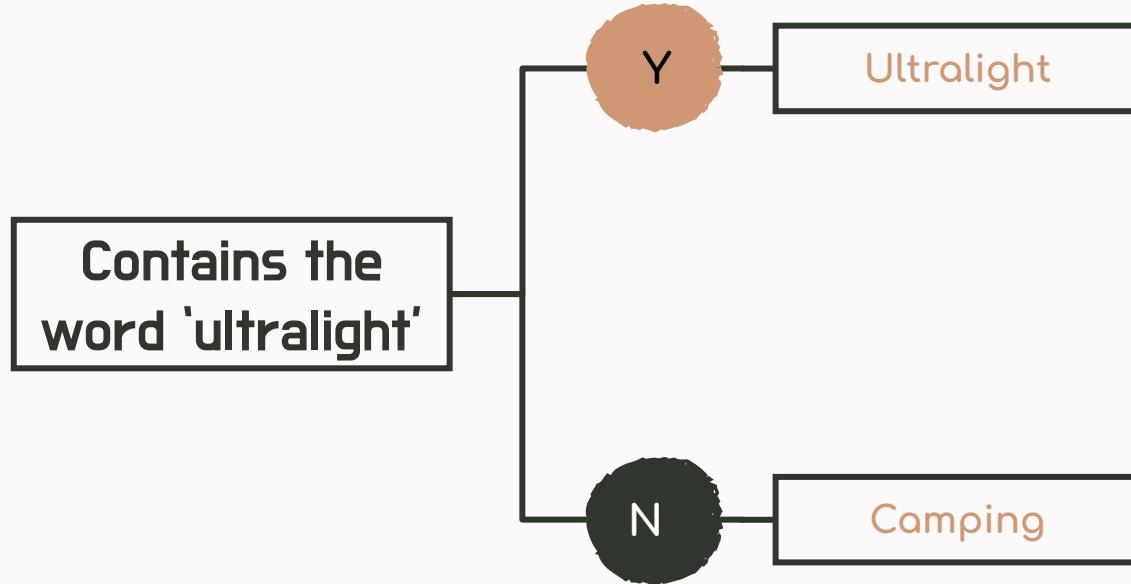
## **Evaluating Performance**



**Testing the  
model**



# Decision Tree

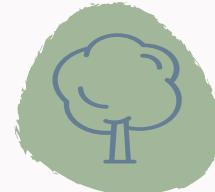


# The Winning Model!



## Voting Classifier

- Voting Mechanism
- Weights the results of other 2 models
- 99.6% accuracy



## AdaBoost Classifier

- Combination of Decision Trees
- 'Boosts' weak model
- 99.4% accuracy



## Extra Trees

- Combination of Decision trees
- Averages results
- 99.3% accuracy

# Results

- Final model accuracy: 99.6%!
- Only 9 errors out of 3,186 posts
- $\frac{1}{3}$  misclassifications: acceptable/ human error
- $\frac{1}{3}$  misclassifications: spam/ unrelated content
- $\frac{1}{3}$  misclassifications: true errors



# Next Steps

# THANKS!

Heavily utilized libraries:  
Pandas, Numpy, and MatPlot Lib  
Sci-Kit Learn

Special Thanks to Tim Book and Rowan Schaefer

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



Final Note: I may be biased

