Letty Wu

Problem

- Reddit is one of the top 10 the most visited websites in the US
- 2019 Ad revenue – $100 million
- Twitter generated more than 2.99 billion in advertising service revenues in 2019

Join Google AdSense!

- Big data with machine learning algorithm
- Using machine learning model to predict between subreddits, then use distinguished words to target advertising better.

Sample Data

A subreddit for the discussion of nutrition science.

A place for current and future nutrition professionals to discuss all aspects of the profession.

Workflow

Data Collection

Used Pushshift API to pull posts from subreddits

1

Data Cleaning & EDA

Removed Outliers, used sklearn to preprocess text data

2

Modeling

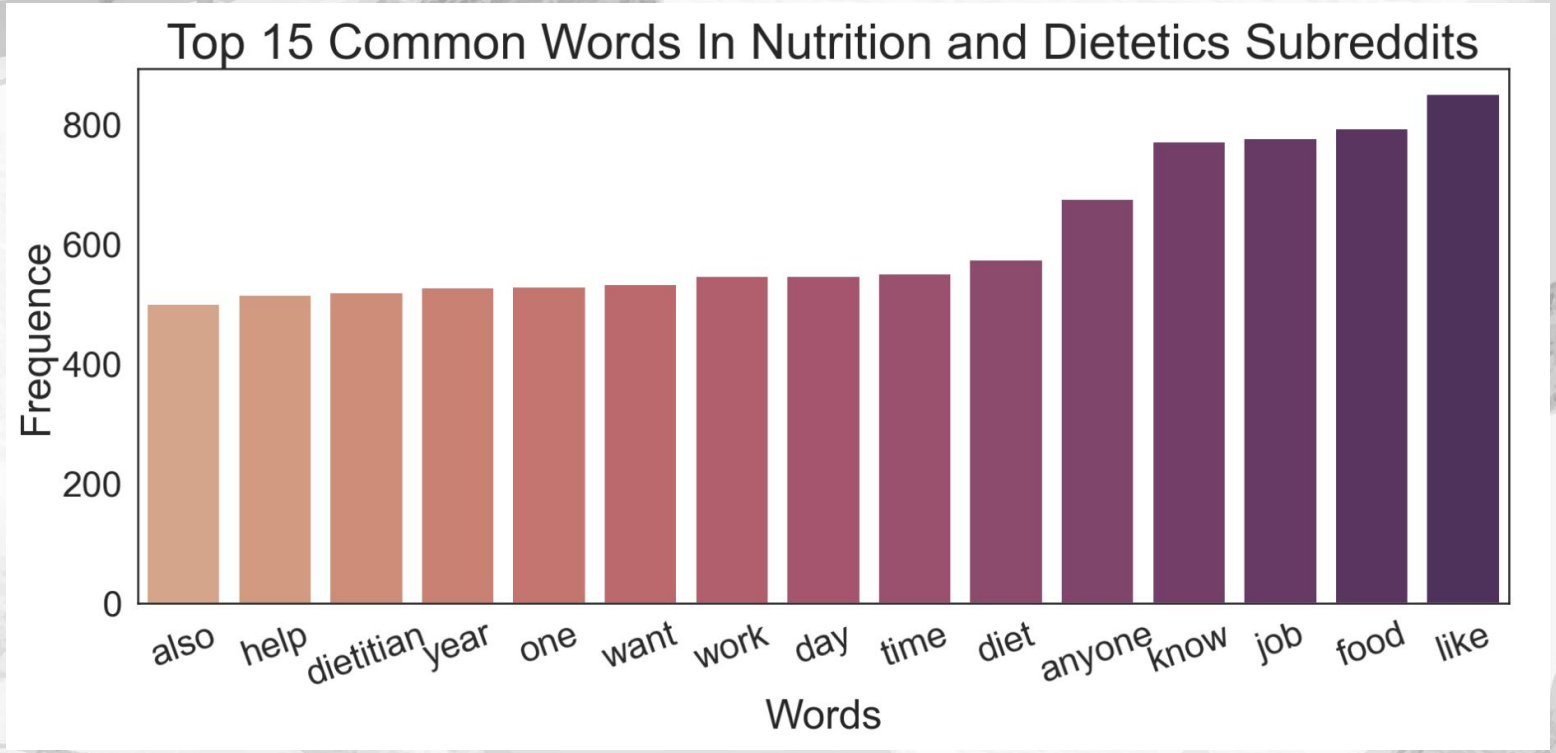
Tried 13 different classifiers, and did Gridsearch on 3 classifiers

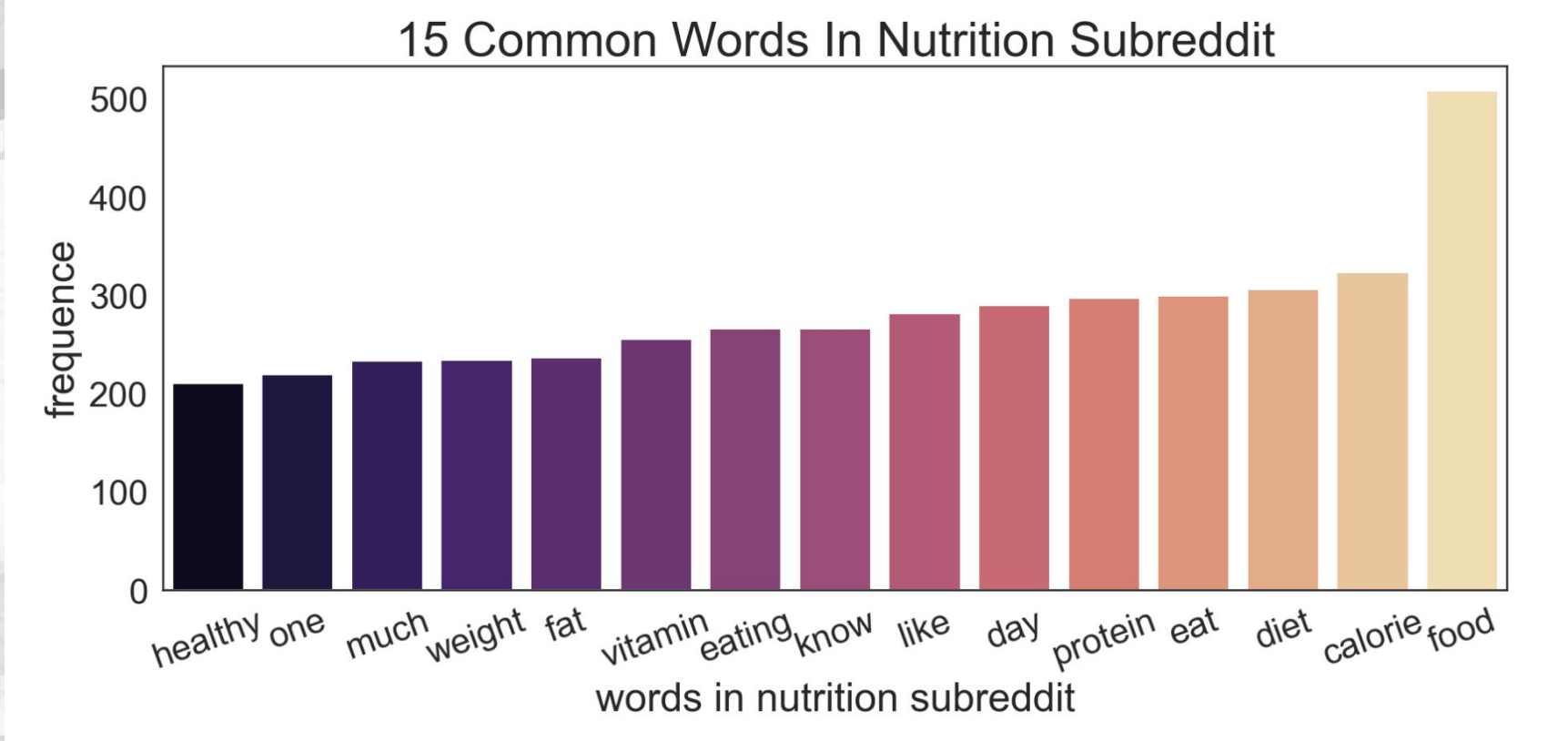
3

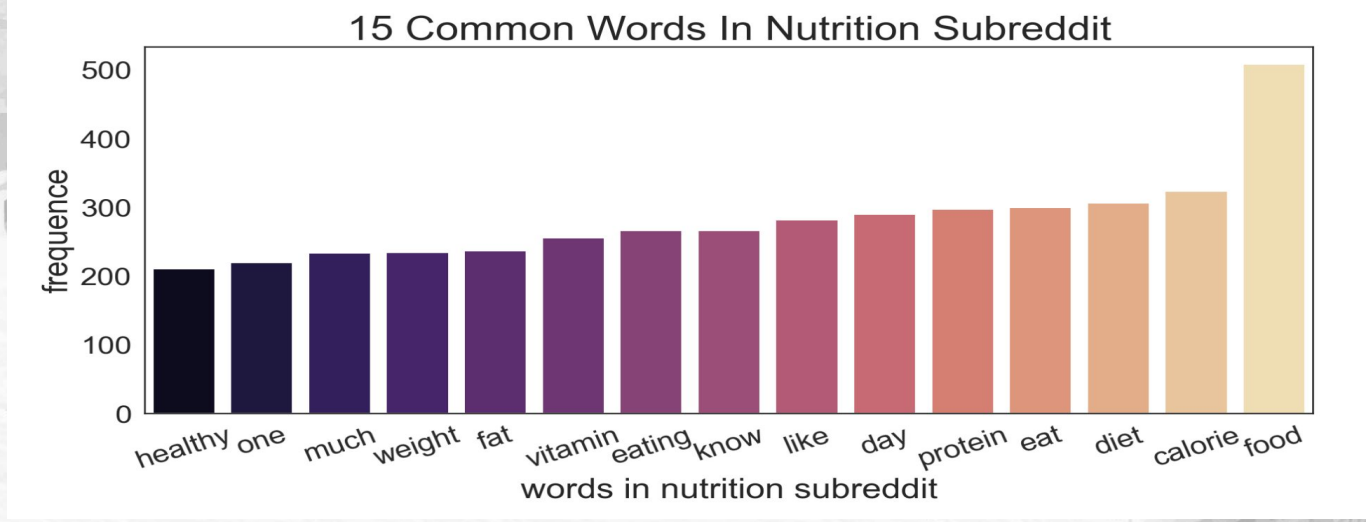
Evaluation

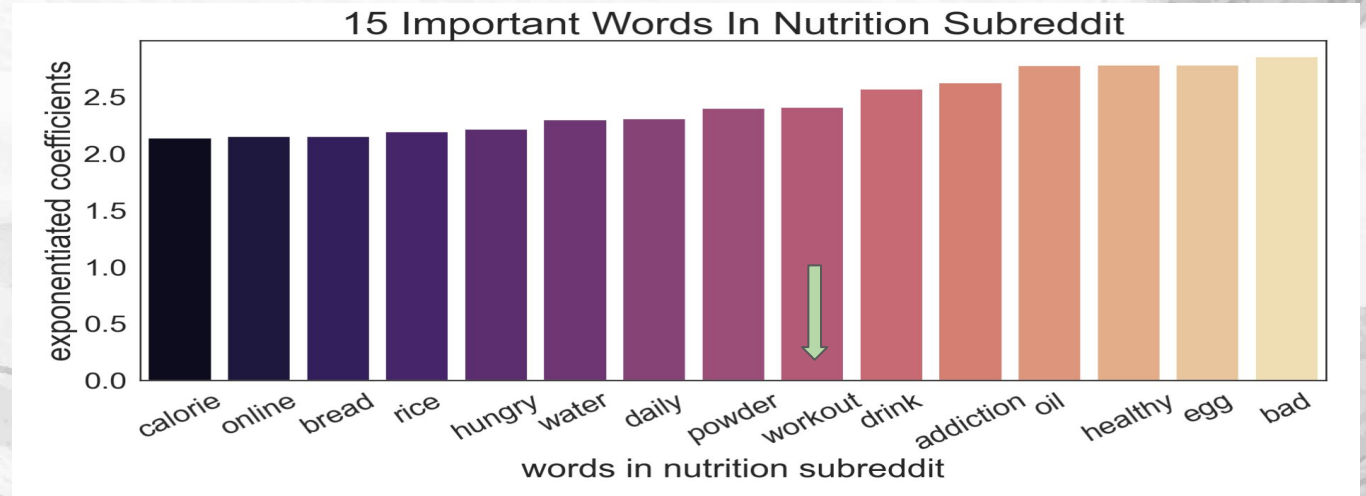
Used accuracy as the main metric

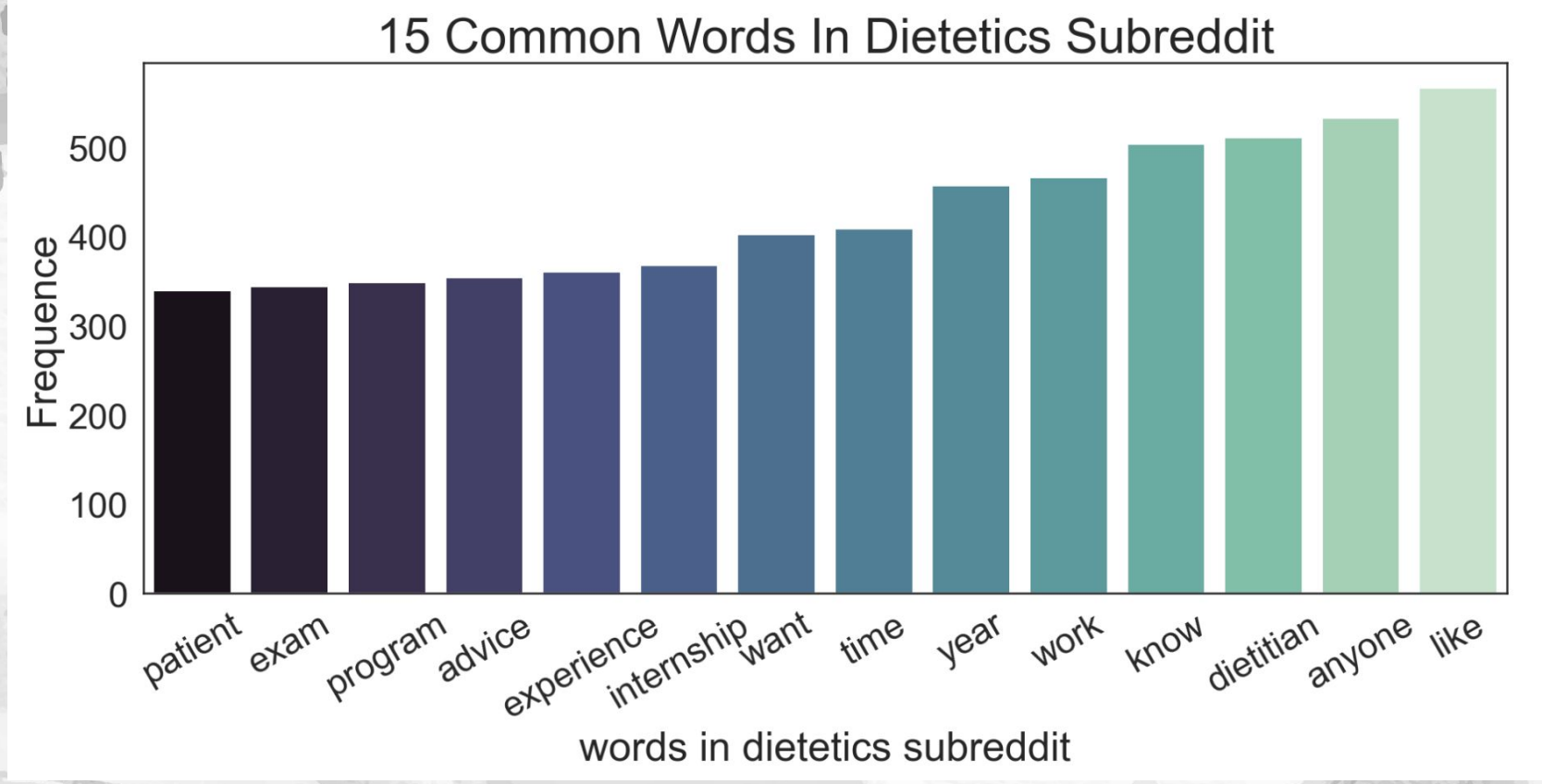
4

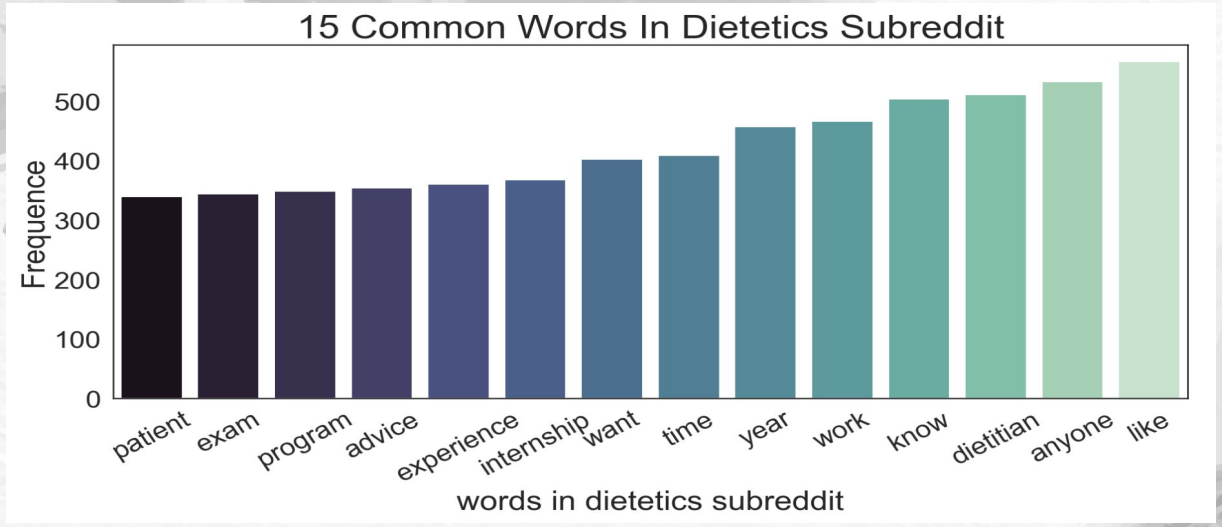

Top 15 Common Words In Nutrition and Dietetics Subreddits
Frequence
0
200
400
600
800
also
help
dietitian
year
one
want
work
day
time
diet
anyone
know
job
food
like
Words

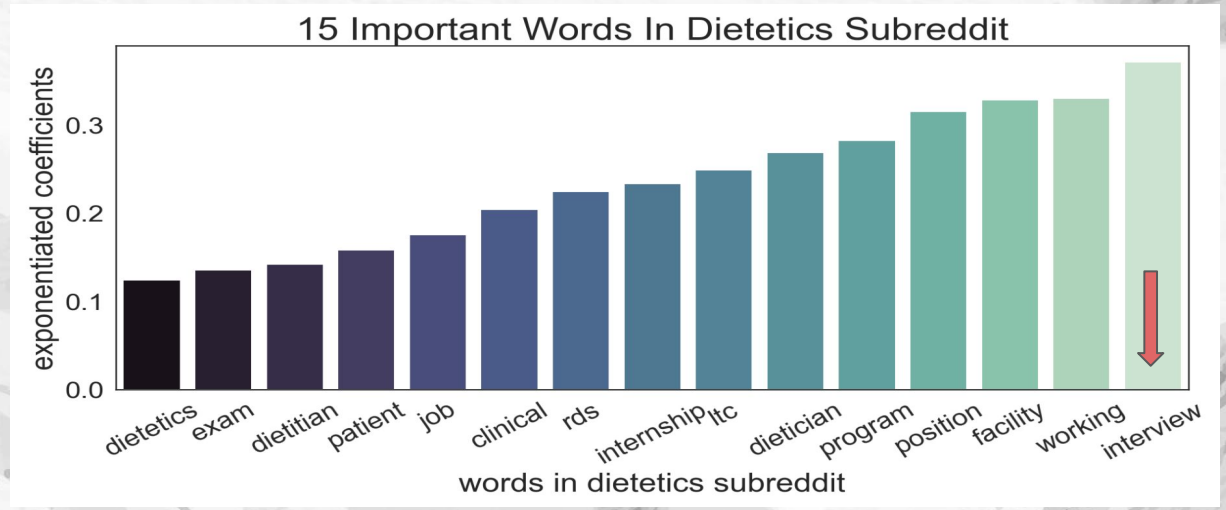

15 Common Words In Nutrition Subreddit
frequence
500
400
300
200
100
0
healthy
one
much
weight
fat
vitamin
eating
know
like
day
protein
eat
diet
calorie
food
words in nutrition subreddit

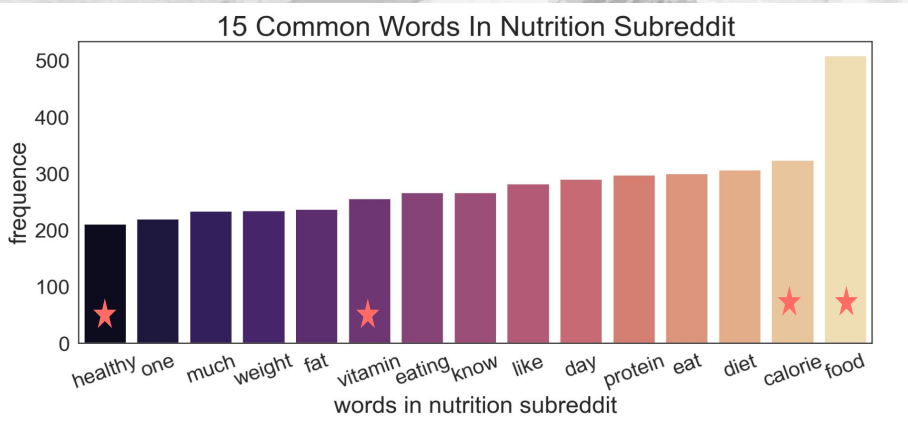

15 Common Words In Nutrition Subreddit
frequence
500
400
300
200
100
0
healthy
one
much
weight
fat
vitamin
eating
know
like
day
protein
eat
diet
calorie
food
words in nutrition subreddit

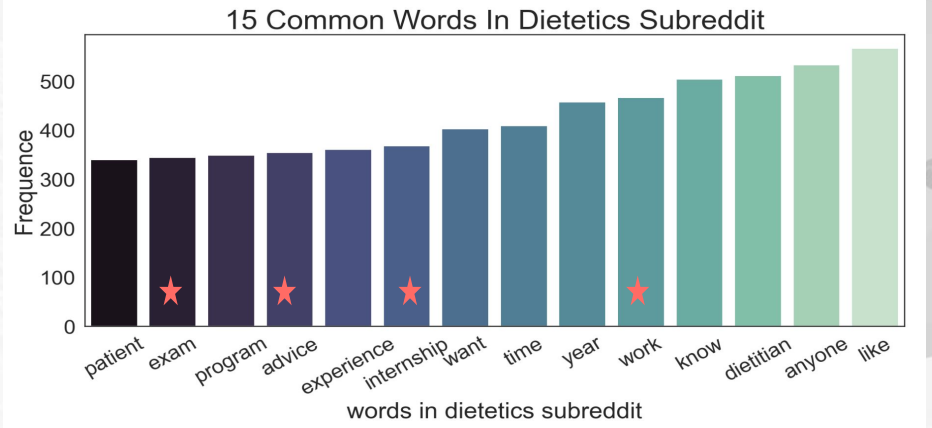

15 Important Words In Nutrition Subreddit
exponentiated coefficients
2.5
2.0
1.5
1.0
0.5
0.0
calorie
online
bread
rice
hungry
water
daily
powder
workout
drink
addiction
oil
healthy
egg
bad
words in nutrition subreddit

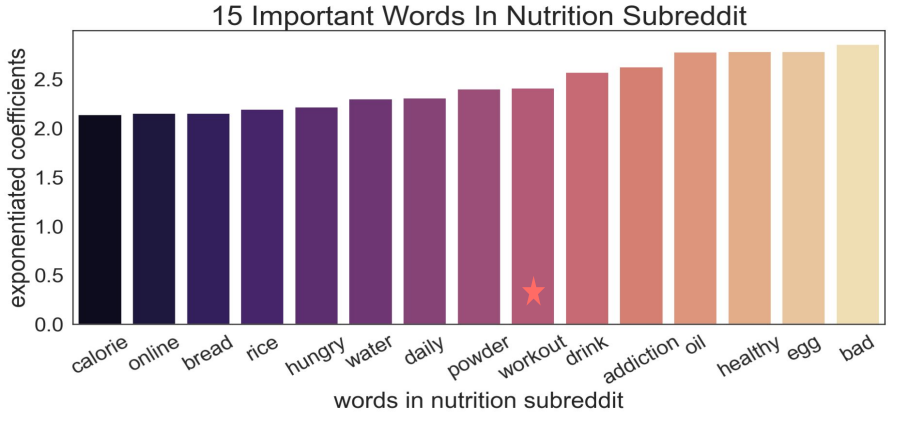

15 Common Words In Dietetics Subreddit
Frequence
0
100
200
300
400
500
patient
exam
program
advice
experience
internship
want
time
year
work
know
dietitian
anyone
like
words in dietetics subreddit

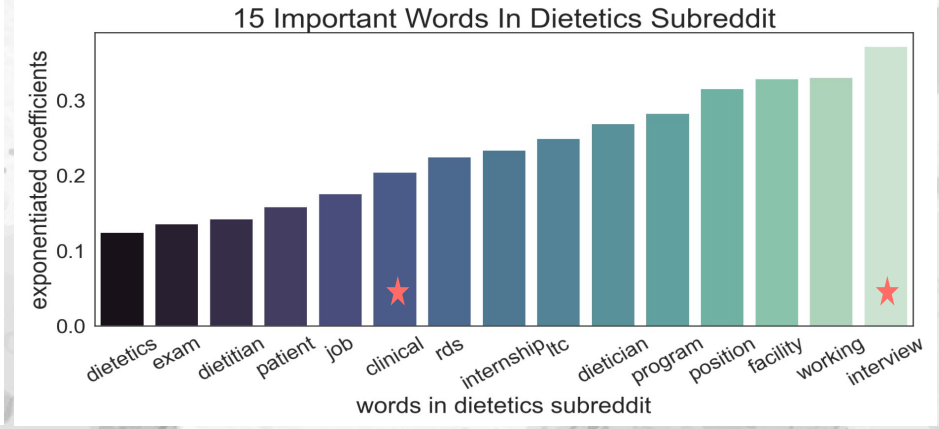

15 Common Words In Dietetics Subreddit
Frequence
0
100
200
300
400
500
patient
exam
program
advice
experience
internship
want
time
year
work
know
dietitian
anyone
like
words in dietetics subreddit


15 Important Words In Dietetics Subreddit
exponentiated coefficients
0.0
0.1
0.2
0.3
dietetics
exam
dietitian
patient
job
clinical
rds
internship
ltc
dietician
program
position
facility
working
interview
words in dietetics subreddit


15 Common Words In Nutrition Subreddit
frequence
0
100
200
300
400
500
healthy
one
much
weight
fat
vitamin
eating
know
like
day
protein
eat
diet
calorie
food
words in nutrition subreddit


15 Common Words In Dietetics Subreddit
Frequence
0
100
200
300
400
500
patient
exam
program
advice
experience
internship
want
time
year
work
know
dietitian
anyone
like
words in dietetics subreddit


15 Important Words In Nutrition Subreddit
exponentiated coefficients
0.0
0.5
1.0
1.5
2.0
2.5
calorie
online
bread
rice
hungry
water
daily
powder
workout
drink
addiction
oil
healthy
egg
bad
words in nutrition subreddit

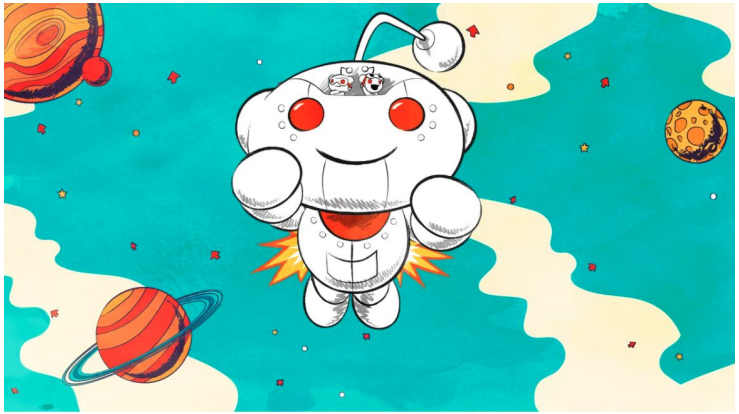

15 Important Words In Dietetics Subreddit
exponentiated coefficients
0.0
0.1
0.2
0.3
dietetics
exam
dietitian
patient
job
clinical
rds
internship
ltc
dietician
program
position
facility
working
interview
words in dietetics subreddit

Model Accuracy

Multinomial Naive Bayes	90.08%
Voting Classifier	89.13%
LogisticRegression	88.22%

Conclusion

- Google AdSense has variety of advertisers
- Automate the process and customize the strategy for placing ads
- **Join Google AdSense!**



Thank you!