Wellness Program
For employees

By Azin Faghihi

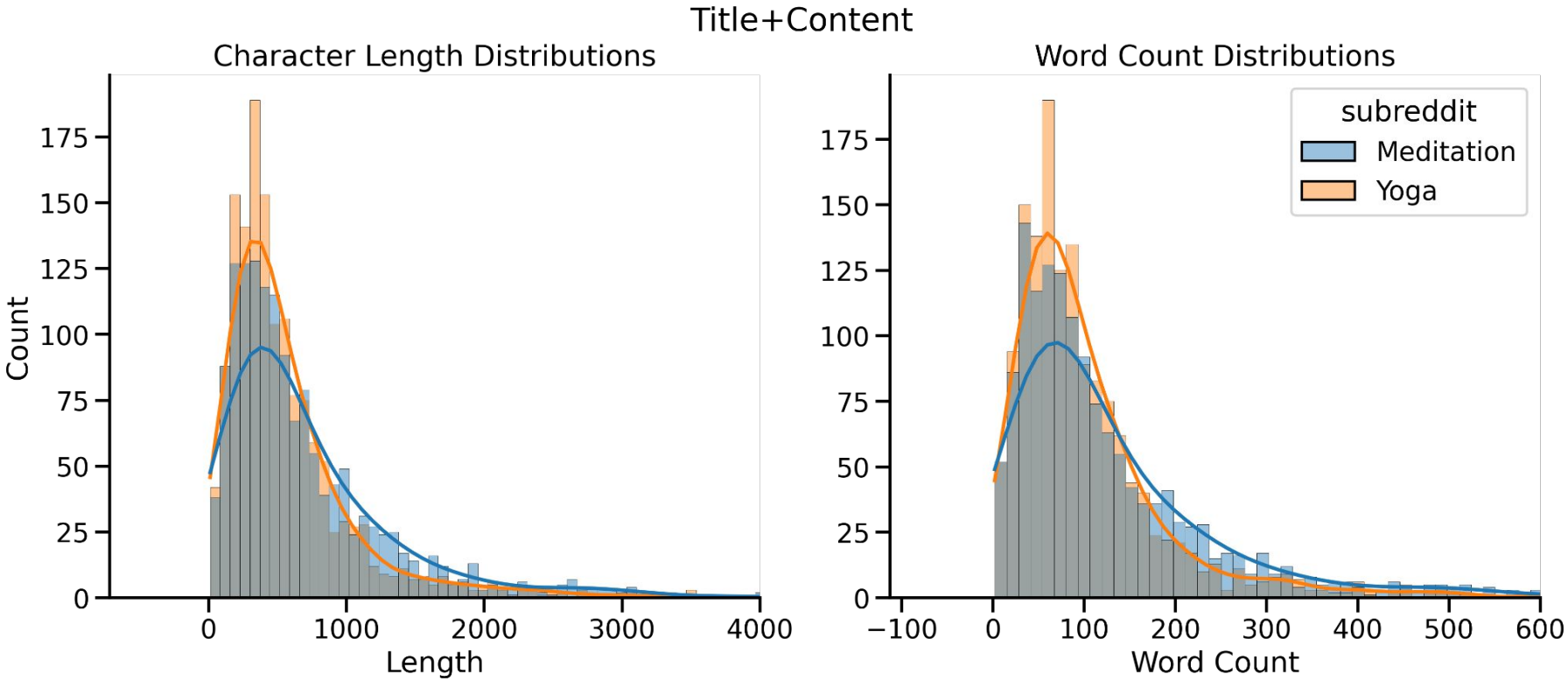
Problem statement:

How should we tailor the wellness package to each employee based on questionnaires /surveys collected from the employees.

Possible Role: Consultant at *SomeCaringStartup.com*

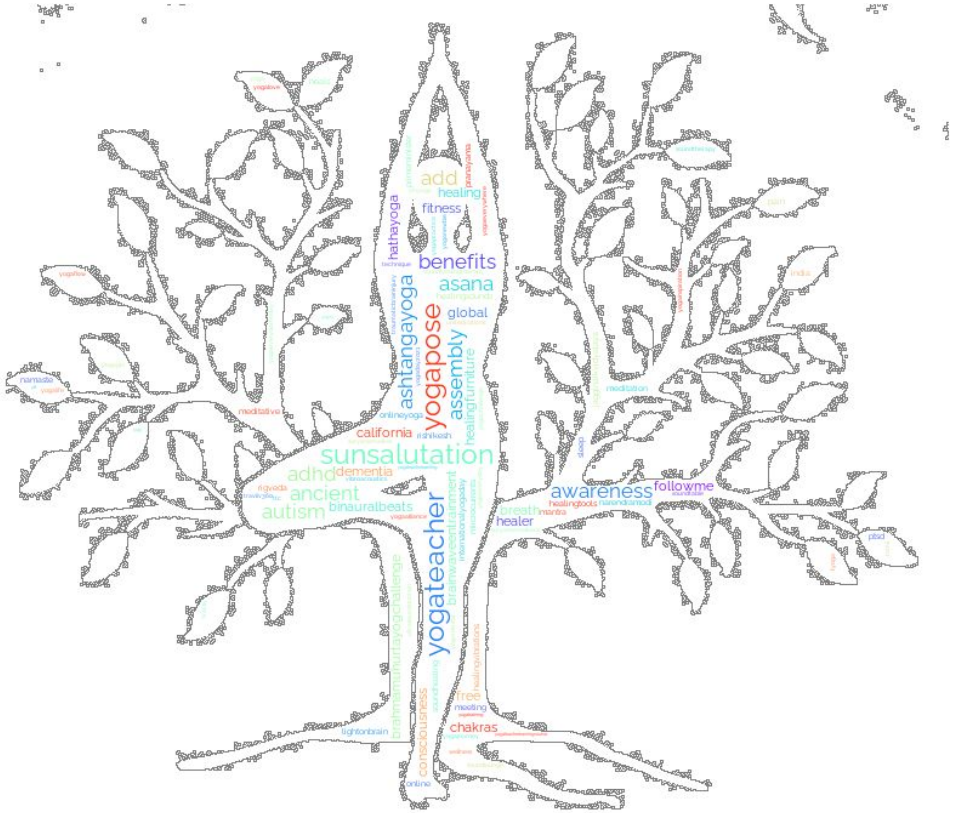
Data: reddit data r/Meditation, r/yoga

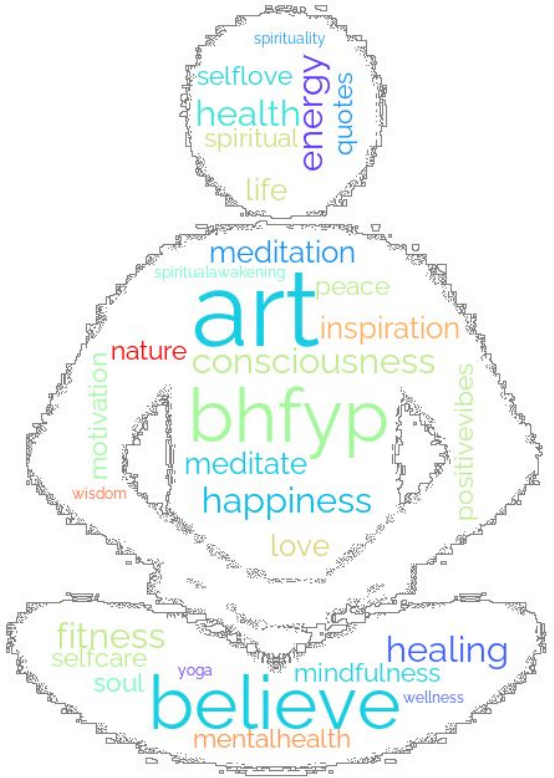
General look at the data


Title+Content
Character Length Distributions
Word Count Distributions
Count
Length
Word Count
subreddit
Meditation
Yoga
0
25
50
75
100
125
150
175
0
1000
2000
3000
4000
−100
0
100
200
300
400
500
600

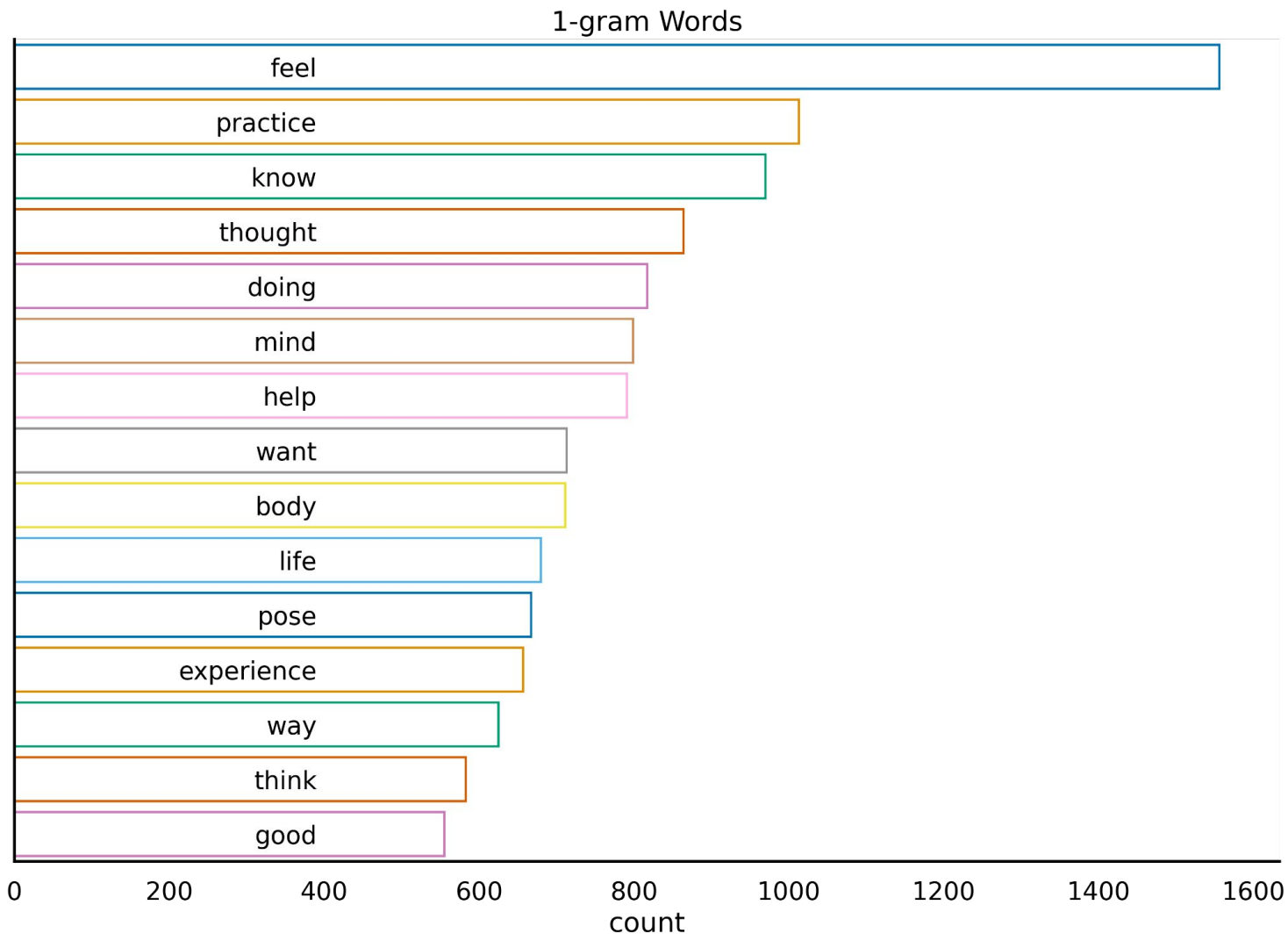
subreddit	# rows	# hashtags
Meditation	1480	33
Yoga	1480	99

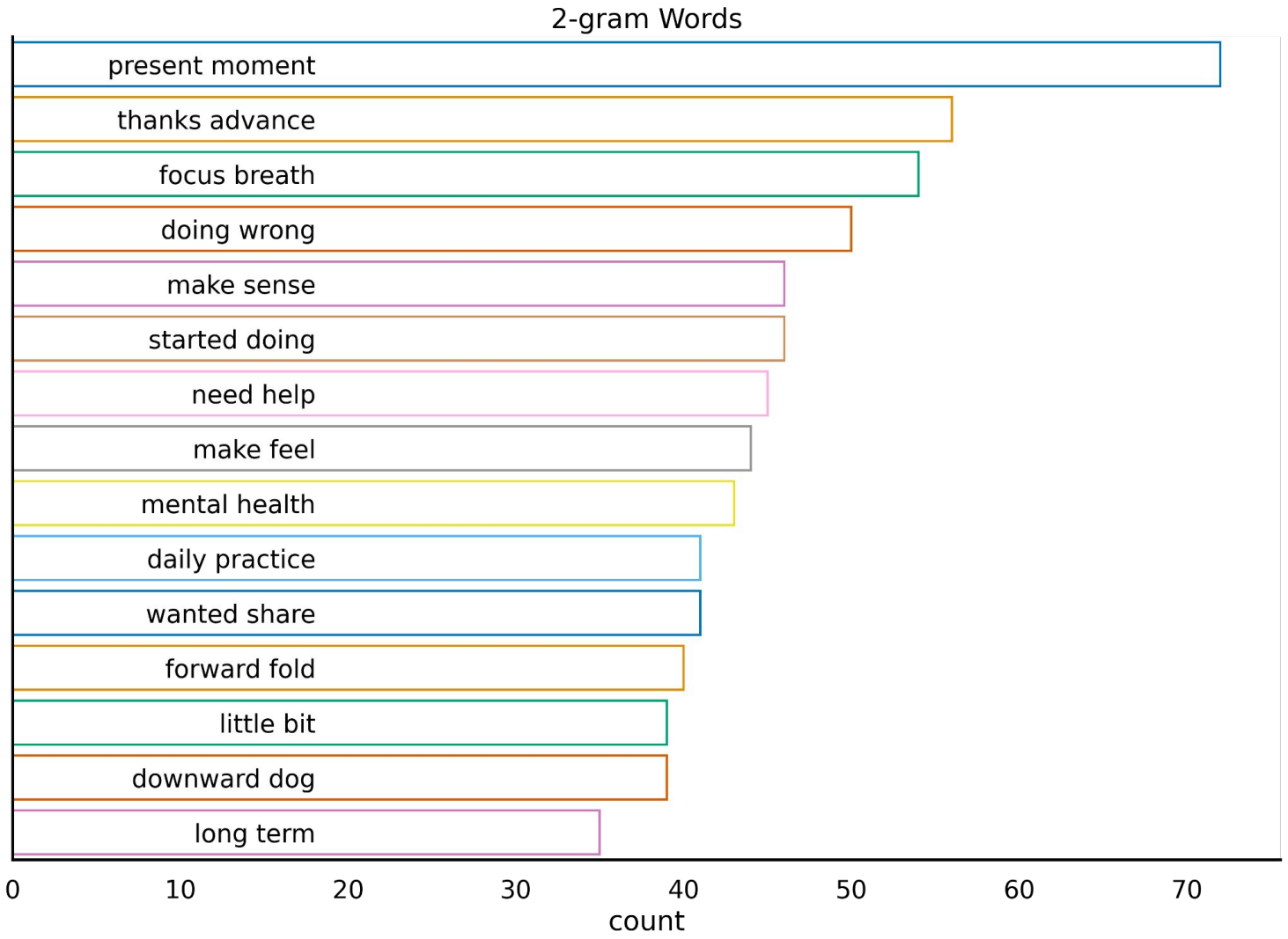
hashtags

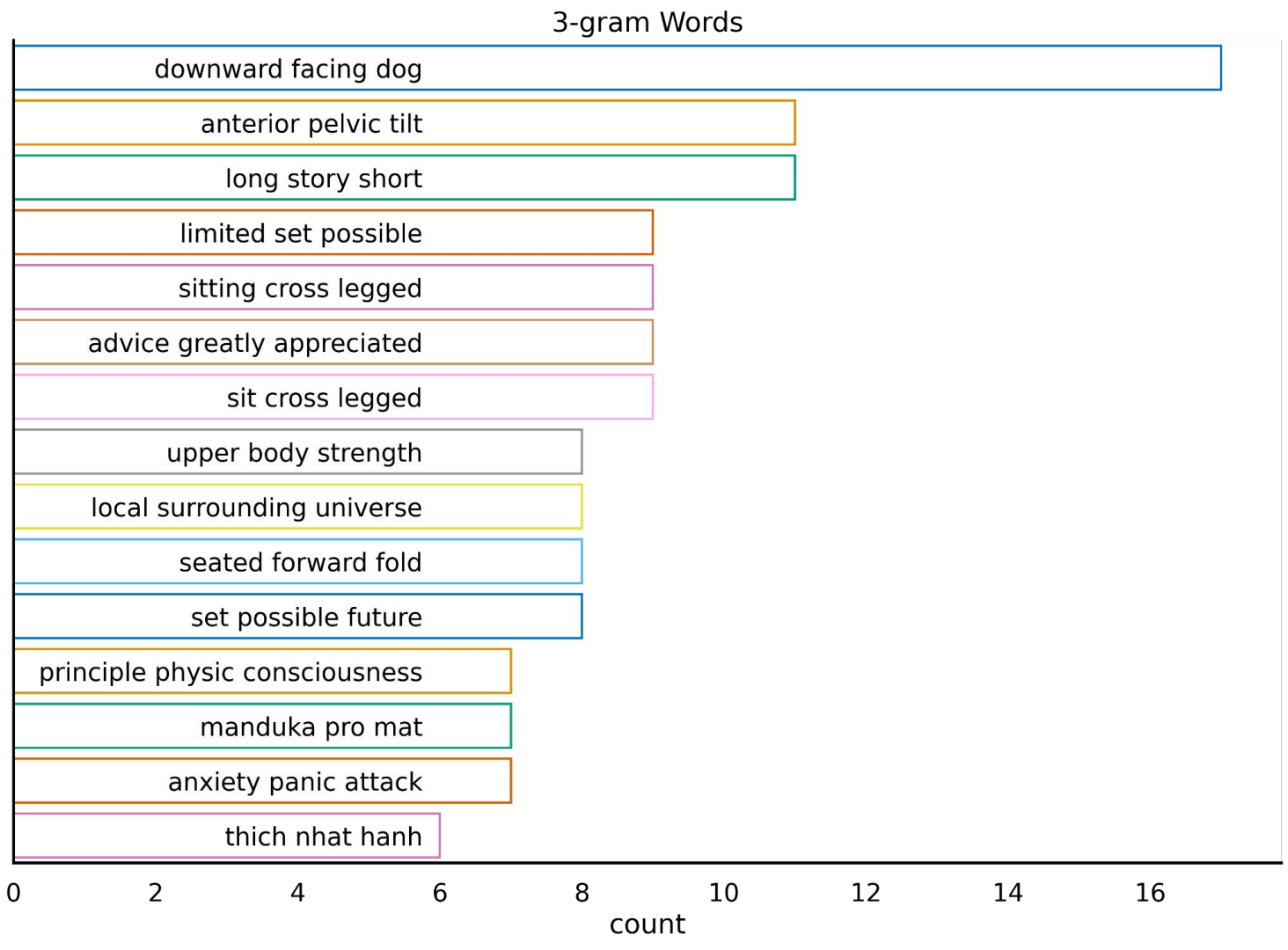

add
healing
fitness
hathayoga
benefits
asana
ashtangayoga
yogapose
global
assembly
sunsalutation
california
adhd
dementia
ancient
autism
binauralbeats
awareness
followme
healer
yogateacher
chakras


spirituality
selflove
health
energy
quotes
spiritual
life
meditation
spiritualawakening
peace
art
inspiration
nature
consciousness
motivation
bhfyp
positivevibes
meditate
wisdom
happiness
love
fitness
selfcare
healing
yoga
mindfulness
soul
believe
wellness
mentalhealth

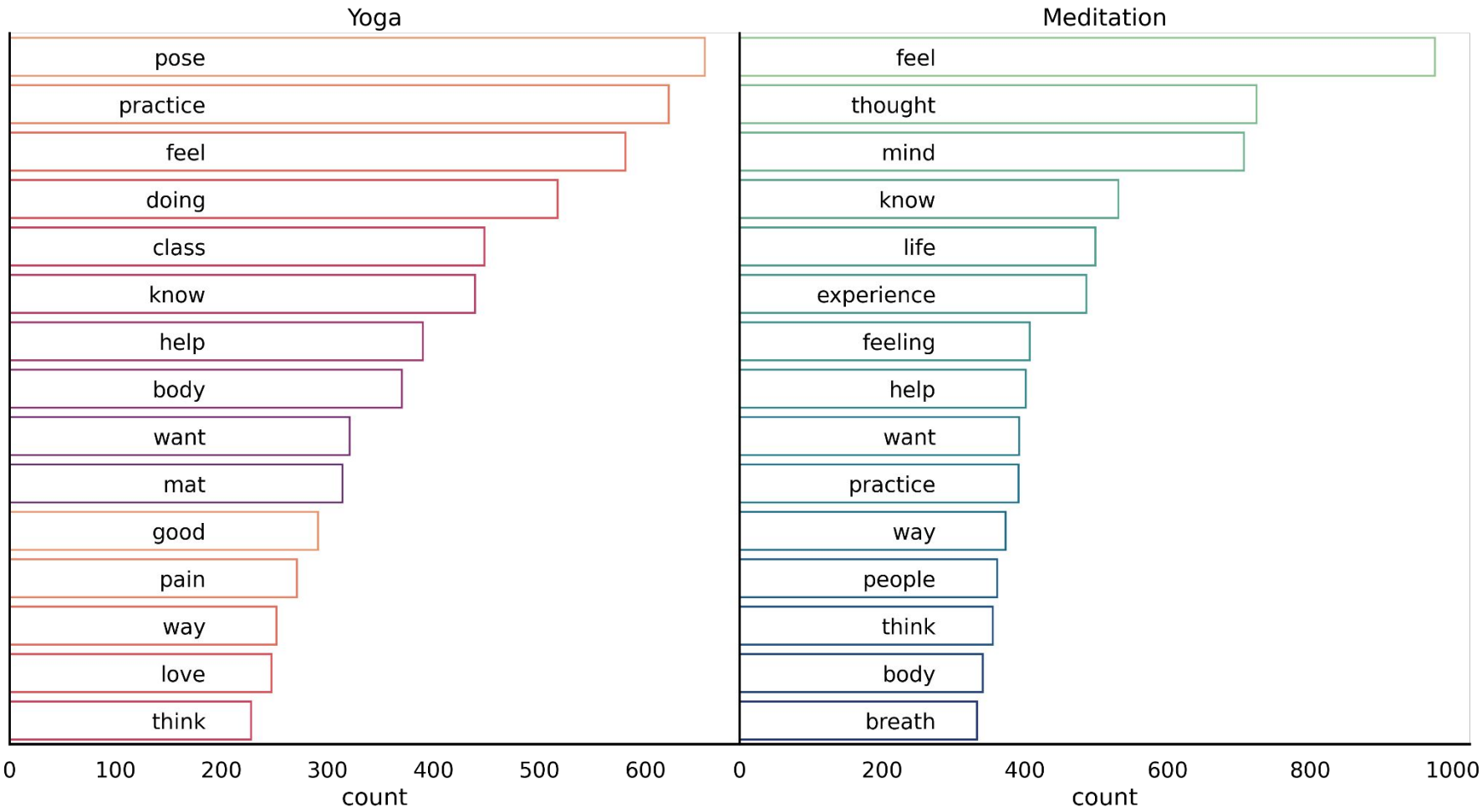
Which ***words*** appear commonly in these two subreddits?

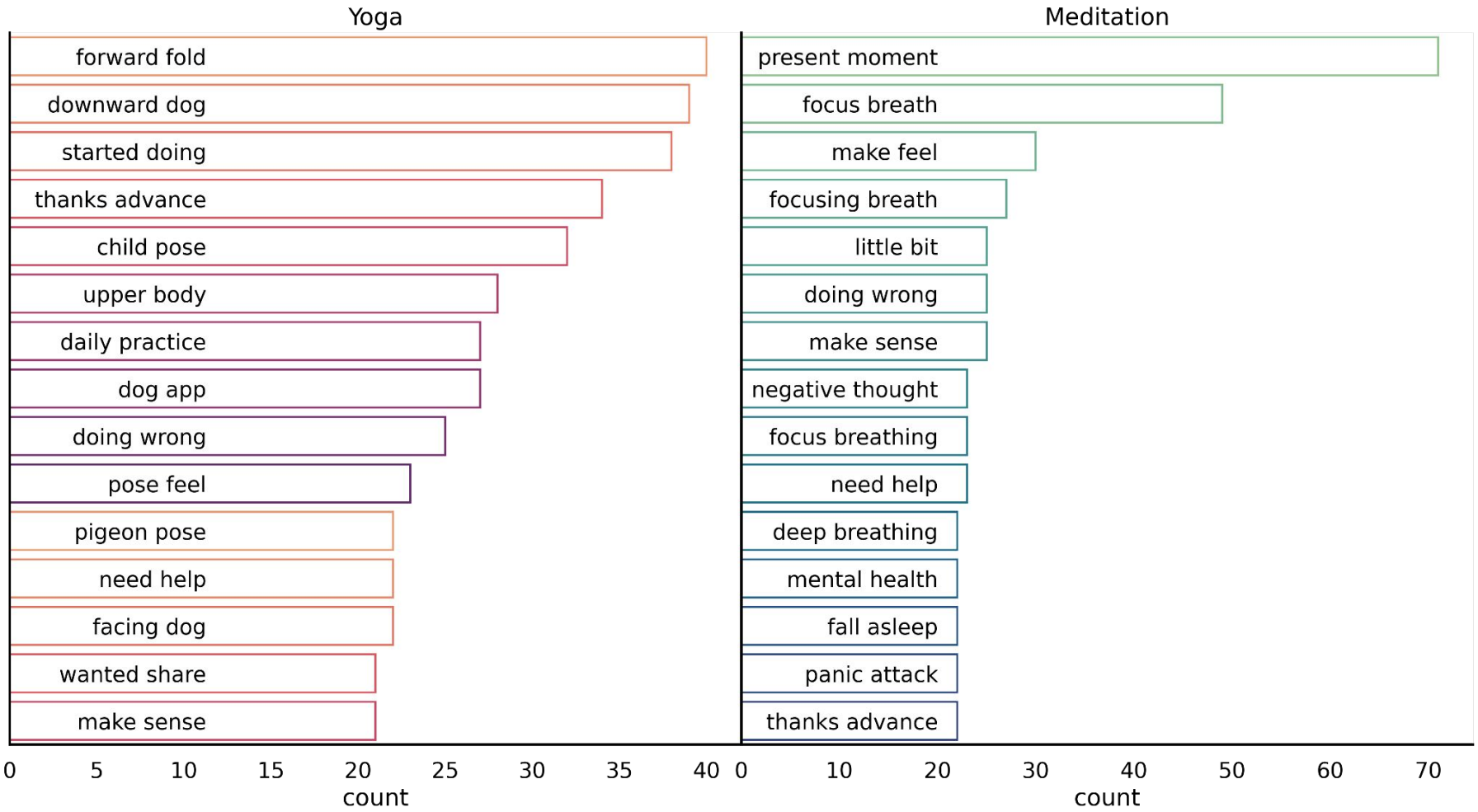

1-gram Words
feel
practice
know
thought
doing
mind
help
want
body
life
pose
experience
way
think
good
0
200
400
600
800
1000
1200
1400
1600
count

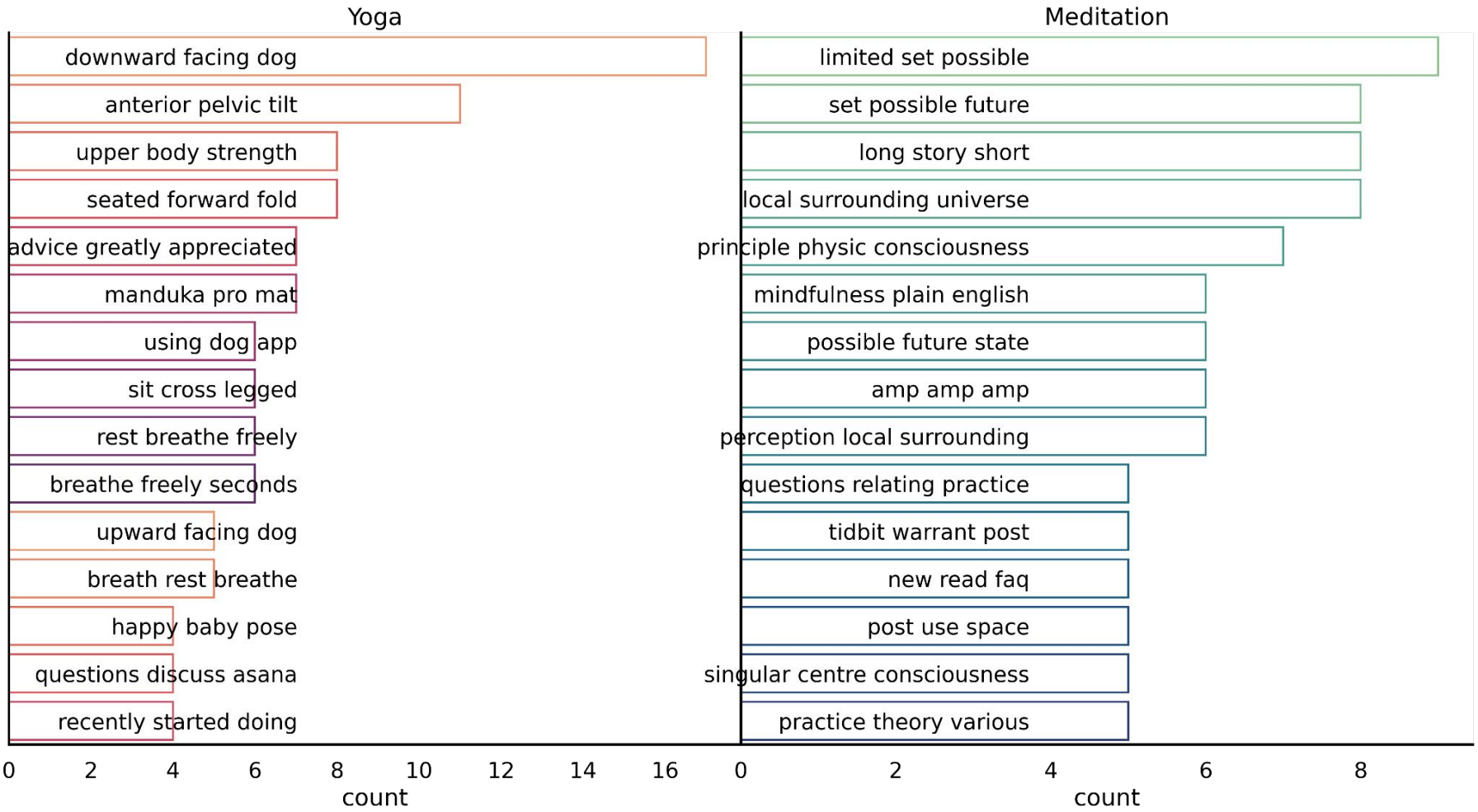

2-gram Words
present moment
thanks advance
focus breath
doing wrong
make sense
started doing
need help
make feel
mental health
daily practice
wanted share
forward fold
little bit
downward dog
long term
0
10
20
30
40
50
60
70
count


3-gram Words
downward facing dog
anterior pelvic tilt
long story short
limited set possible
sitting cross legged
advice greatly appreciated
sit cross legged
upper body strength
local surrounding universe
seated forward fold
set possible future
principle physic consciousness
manduka pro mat
anxiety panic attack
thich nhat hanh
0
2
4
6
8
10
12
14
16
count

Which ***words*** appear in *Yoga* & *Meditation* each?


1-gram Words
Yoga
pose
practice
feel
doing
class
know
help
body
want
mat
good
pain
way
love
think
0
100
200
300
400
500
600
count
Meditation
feel
thought
mind
know
life
experience
feeling
help
want
practice
way
people
think
body
breath
0
200
400
600
800
1000
count

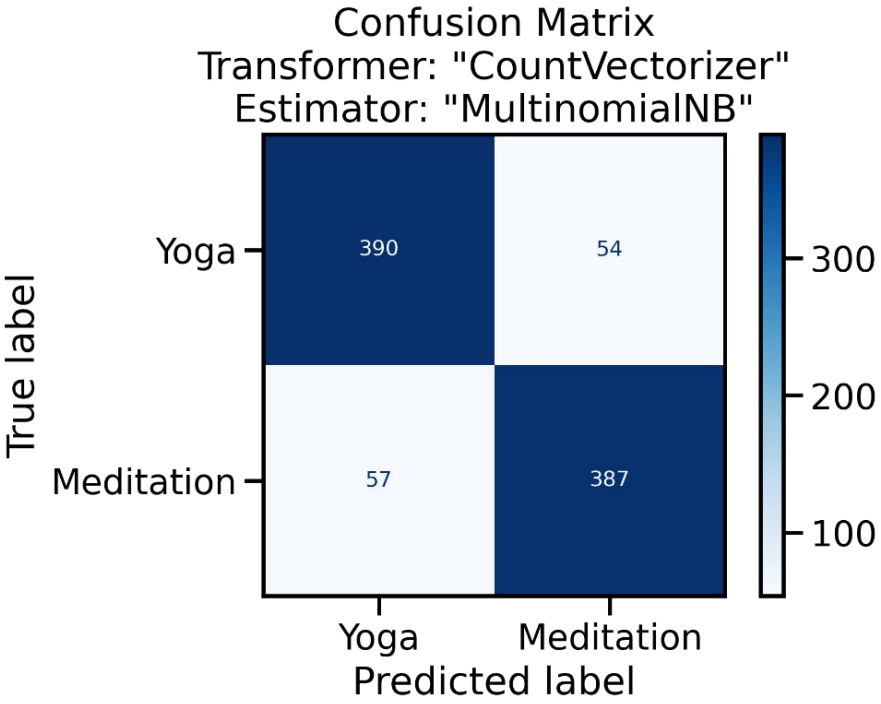

2-gram Words
Yoga
forward fold
downward dog
started doing
thanks advance
child pose
upper body
daily practice
dog app
doing wrong
pose feel
pigeon pose
need help
facing dog
wanted share
make sense
0
5
10
15
20
25
30
35
40
count
Meditation
present moment
focus breath
make feel
focusing breath
little bit
doing wrong
make sense
negative thought
focus breathing
need help
deep breathing
mental health
fall asleep
panic attack
thanks advance
0
10
20
30
40
50
60
70
count


3-gram Words
Yoga
downward facing dog
anterior pelvic tilt
upper body strength
seated forward fold
advice greatly appreciated
manduka pro mat
using dog app
sit cross legged
rest breathe freely
breathe freely seconds
upward facing dog
breath rest breathe
happy baby pose
questions discuss asana
recently started doing
0
2
4
6
8
10
12
14
16
count
Meditation
limited set possible
set possible future
long story short
local surrounding universe
principle physic consciousness
mindfulness plain english
possible future state
amp amp amp
perception local surrounding
questions relating practice
tidbit warrant post
new read faq
post use space
singular centre consciousness
practice theory various
0
2
4
6
8
count

Modeling

Estimator	Transformer	Validation Score	Train Score	Test Score	Best Score (GS)	Train Score (GS)	Test Score (GS)	F1 - score	Recall (Sensitivity)	Specificity (True Negative Rate)	Precision	Accuracy
MultinomialNB	**TfidfVectorizer**	0.881	0.991	0.884	0.881	0.941	0.883	0.885	0.905	0.860	0.866	0.883
	CountVectorizer	0.890	0.994	0.880	0.883	0.924	0.875	0.875	0.872	0.878	0.878	0.875
RandomForestClassifier	**TfidfVectorizer**	0.841	0.999	0.855	0.866	0.999	0.855	0.861	0.896	0.813	0.827	0.855
DecisionTreeClassifier	**CountVectorizer**	0.774	0.999	0.787	0.778	0.842	0.785	0.803	0.876	0.694	0.741	0.785
	TfidfVectorizer	0.752	0.999	0.757	0.760	0.820	0.762	0.788	0.883	0.642	0.711	0.762

Modeling: *Multinomial Naive Bayes*



Accuracy: 88%

Precision: 88%

F1 - score: 87%

Recall (Sensitivity): 87%

Specificity (True Negative Rate): 88%

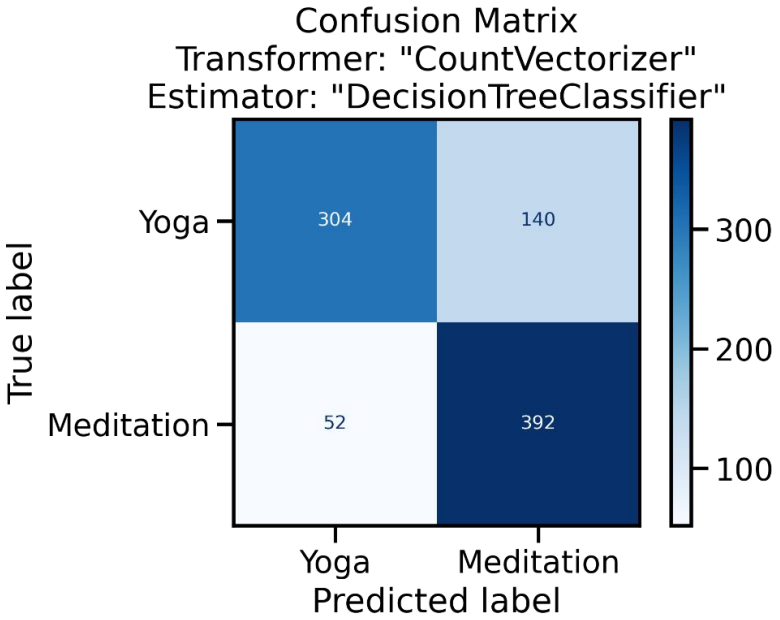
max_df: 0.9

max_features: 3000

min_df: 2

ngram_range: (1, 1)

Modeling: *Decision Tree Classifier*



Accuracy: 78%

Precision: 74%

F1 - score: 80%

Recall (Sensitivity): 88%

Specificity (True Negative Rate): 68%

max_depth: 30

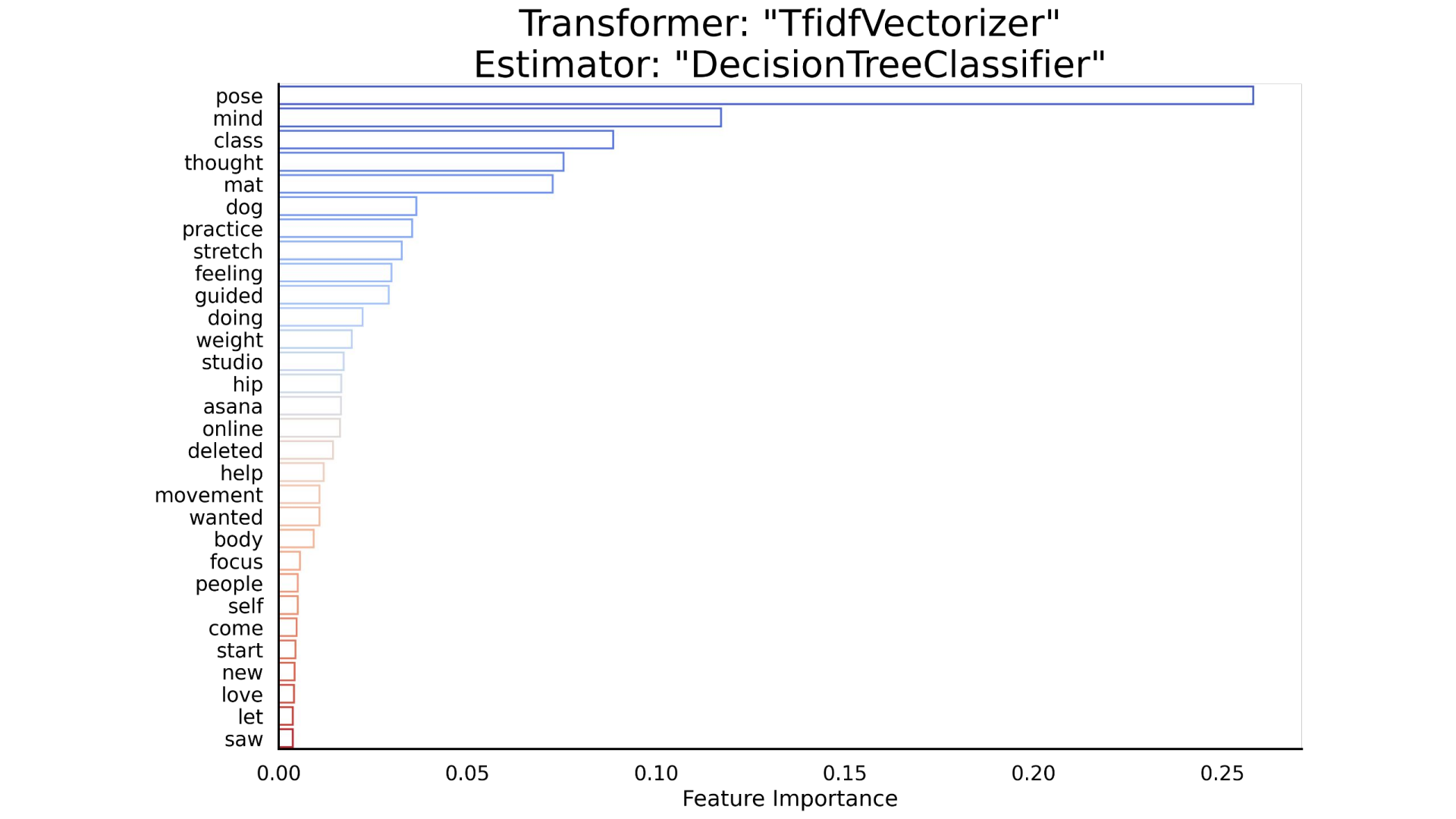
min_samples_leaf: 5

min_samples_split: 30

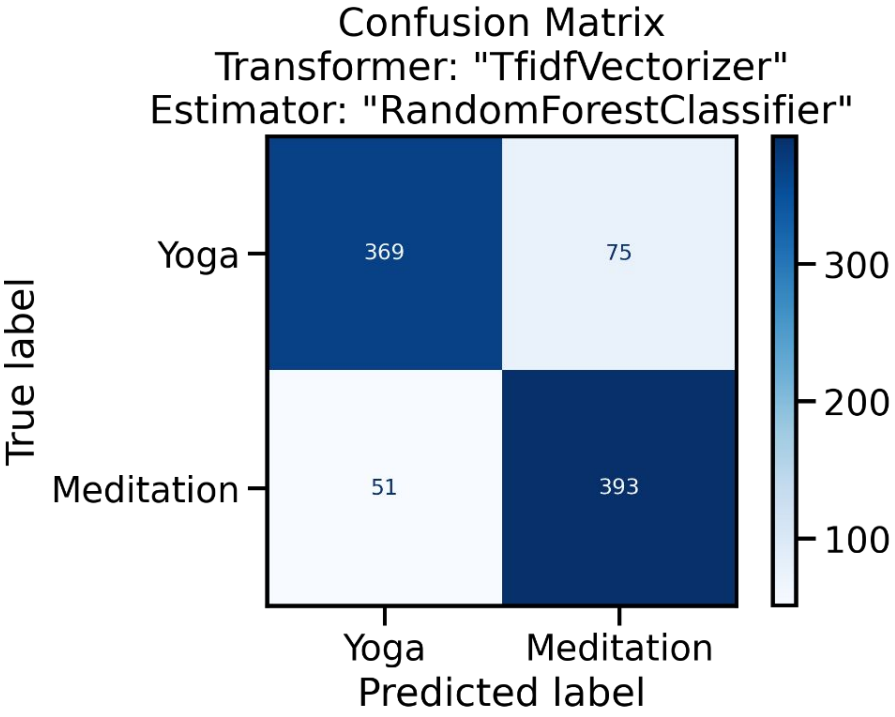
max_df: 0.9

min_df: 2

ngram_range: (1, 2)


Transformer: "TfidfVectorizer"
Estimator: "DecisionTreeClassifier"
pose
mind
class
thought
mat
dog
practice
stretch
feeling
guided
doing
weight
studio
hip
asana
online
deleted
help
movement
wanted
body
focus
people
self
come
start
new
love
let
saw
0.00
0.05
0.10
0.15
0.20
0.25
Feature Importance

Modeling: *Random Forest Classifier*



Accuracy: 86%

Precision: 84%

F1 - score: 86%

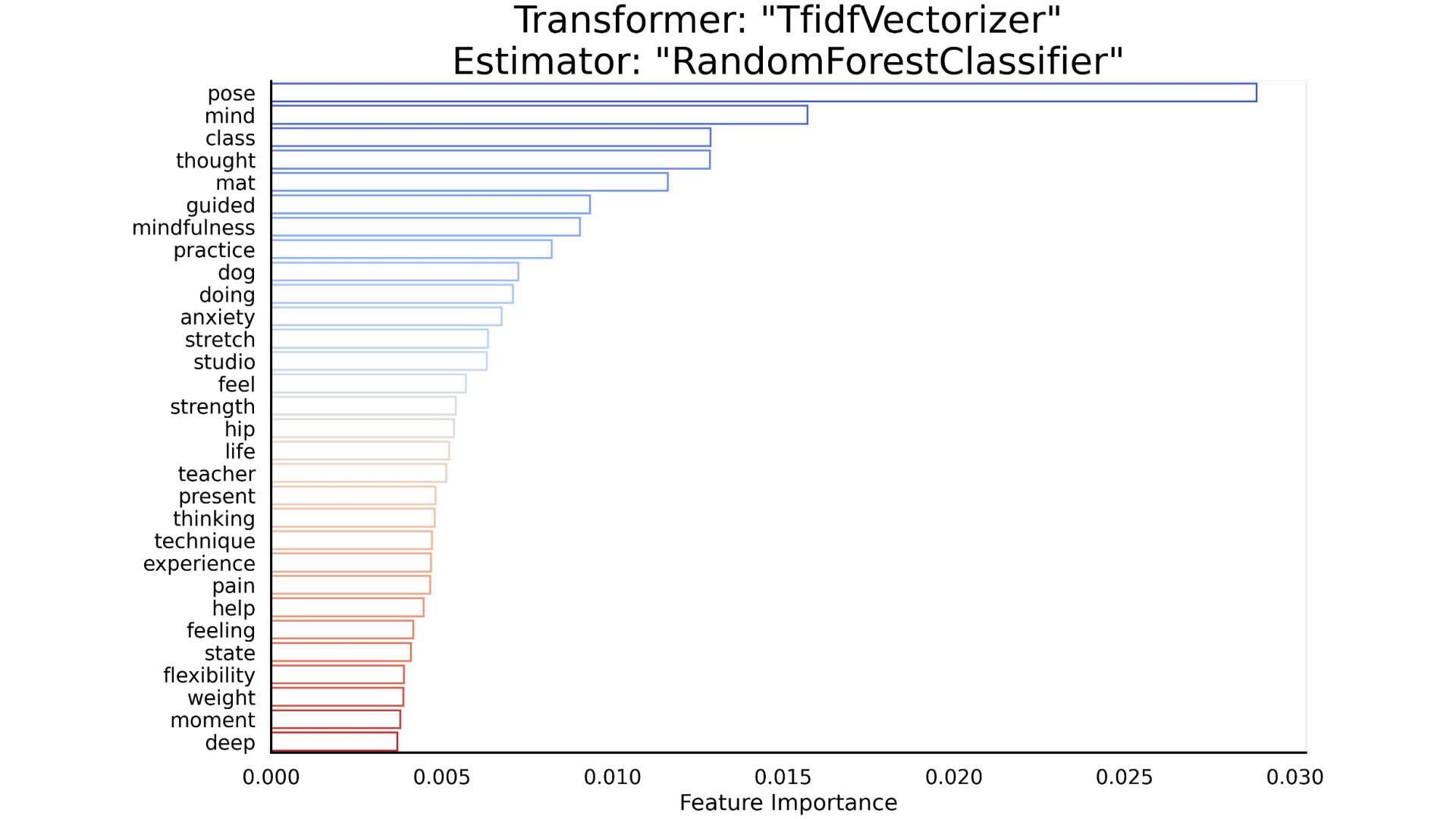
Recall (Sensitivity): 89%

Specificity (True Negative Rate): 83%

max_depth: None

n_estimators: 100

ngram_range: (1, 1)


Transformer: "TfidfVectorizer"
Estimator: "RandomForestClassifier"
pose
mind
class
thought
mat
guided
mindfulness
practice
dog
doing
anxiety
stretch
studio
feel
strength
hip
life
teacher
present
thinking
technique
experience
pain
help
feeling
state
flexibility
weight
moment
deep
0.000
0.005
0.010
0.015
0.020
0.025
0.030
Feature Importance

Modeling: *summary*

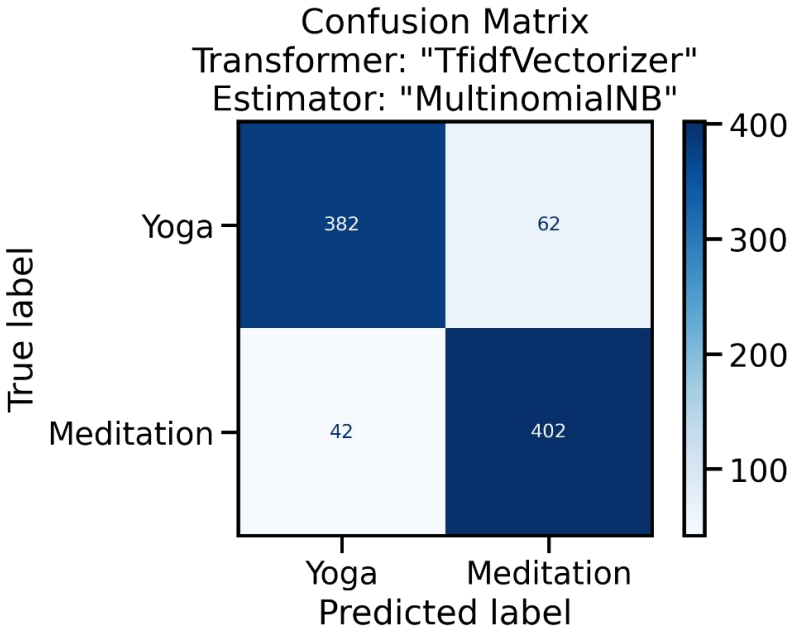
Estimator	Transformer	Validation Score	Train Score	Test Score	Best Score (GS)	Train Score (GS)	Test Score (GS)	F1 - score	Recall (Sensitivity)	Specificity (True Negative Rate)	Precision	Accuracy
MultinomialNB	TfidfVectorizer	0.881	0.991	0.884	0.881	0.941	0.883	0.885	0.905	0.860	0.866	0.883
	CountVectorizer	0.890	0.994	0.880	0.883	0.924	0.875	0.875	0.872	0.878	0.878	0.875
RandomForestClassifier	TfidfVectorizer	0.841	0.999	0.855	0.866	0.999	0.855	0.861	0.896	0.813	0.827	0.855
DecisionTreeClassifier	CountVectorizer	0.774	0.999	0.787	0.778	0.842	0.785	0.803	0.876	0.694	0.741	0.785
	TfidfVectorizer	0.752	0.999	0.757	0.760	0.820	0.762	0.788	0.883	0.642	0.711	0.762

Recommendations:

- The corpus of data collected from the employees throughout the first two years of working at the company could be used to determine if Meditation or Yoga would be a better option for the employee.

Future Work:

- Try other classification models
- Consider the posts from unique users



Accuracy: 88%

Precision: 87%

F1 - score: 89%

Recall (Sensitivity): 91%

Specificity (True Negative Rate): 86%

max_features: 4000

ngram_range: (1, 1)