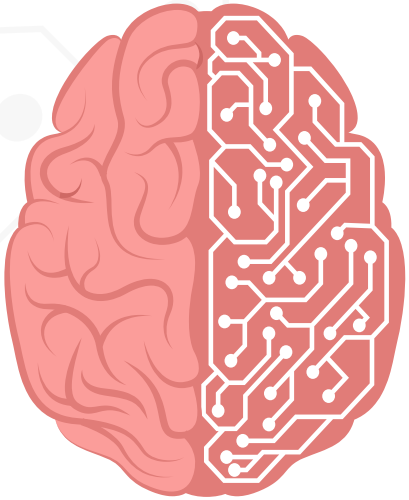
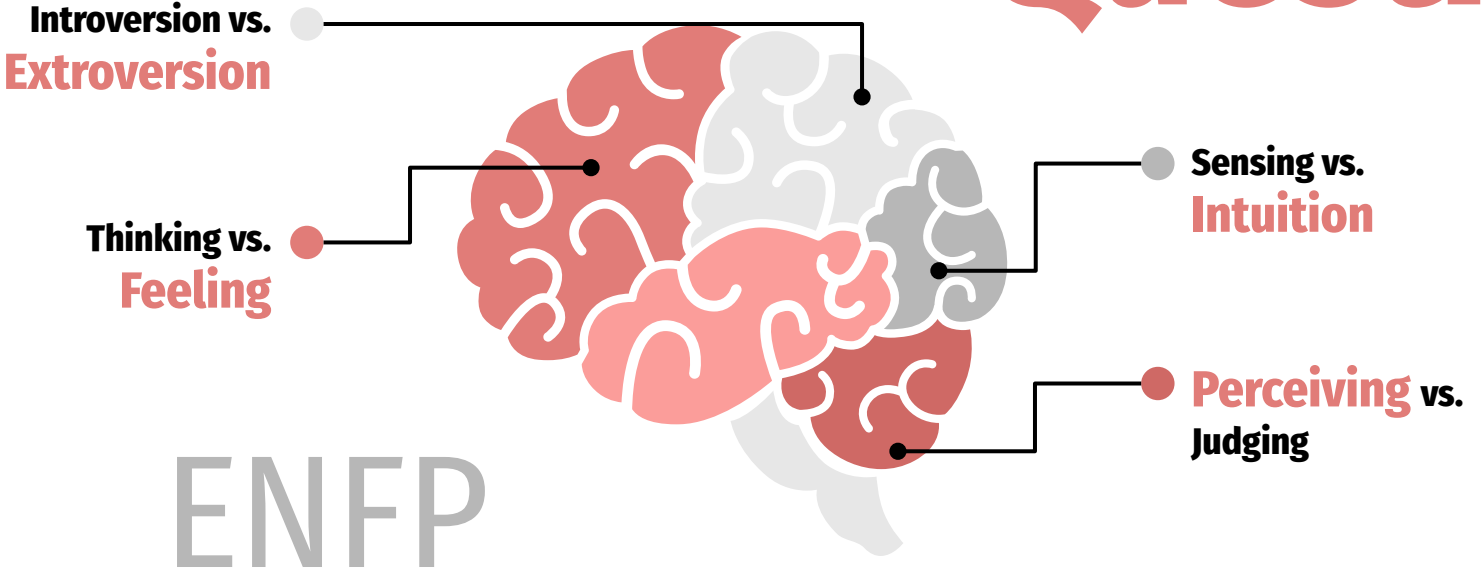


READING YOUR PERSONALITY

Brian Tam

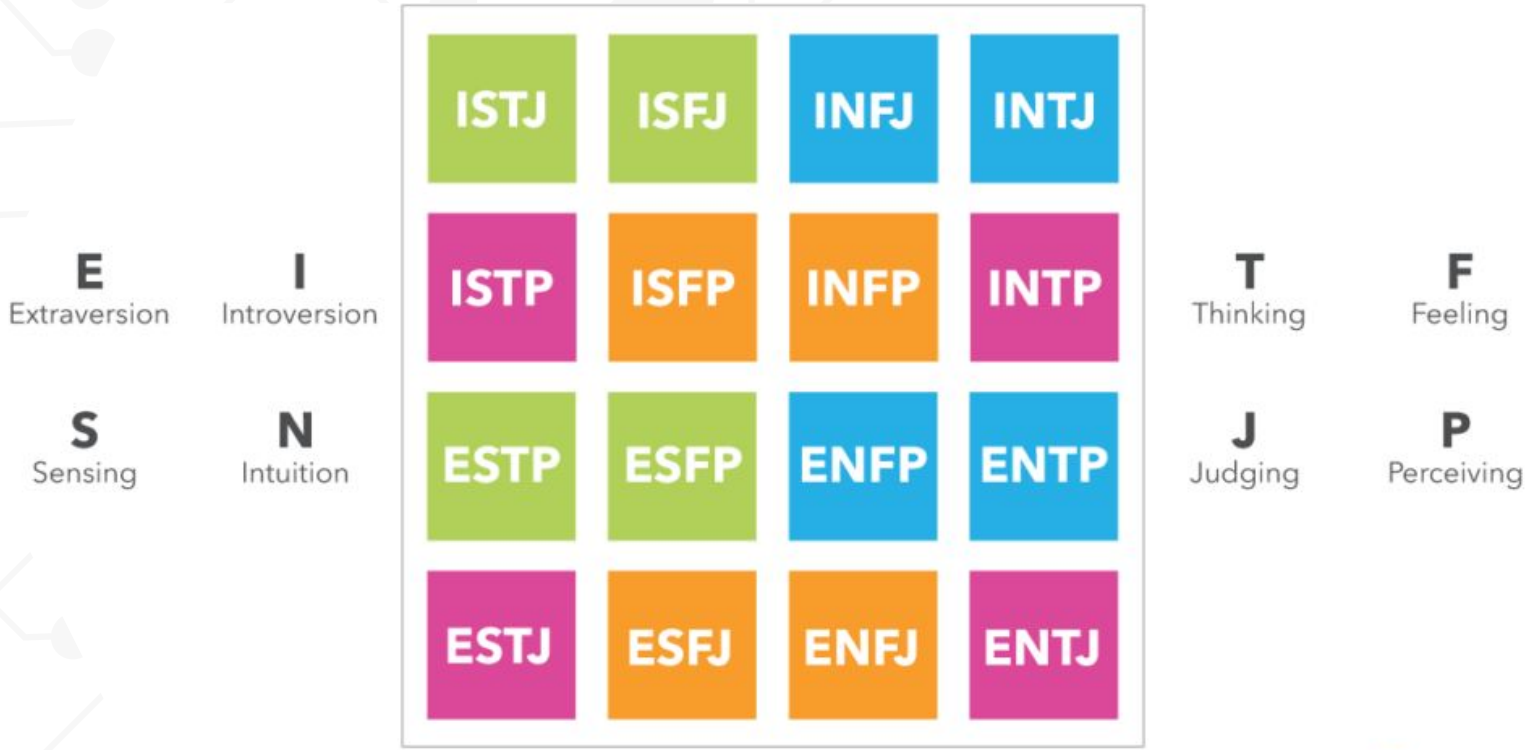
Present Solution.

93 Questions



ENFP

16 Personalities



Can We

Can We

Do

Can We Do Better?

Try :

Classifying using Natural Language Processing.

01. Overview

02. Features

03. Models & Scores

04. Future Work

01.

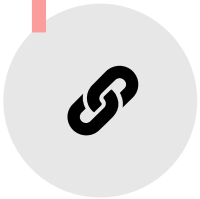
MBTI Dataset

- Labeled posts taken from the “personality cafe” online forum
- 8000+ people
- > 50 posts/person

personality cafe

Preprocessing

1

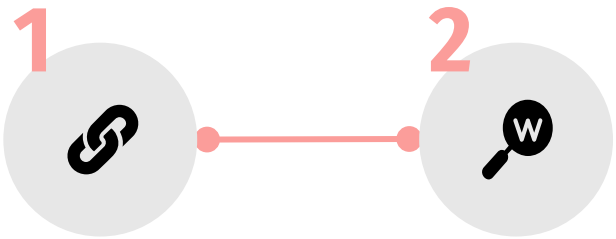


Removed links

Videos and images are hard to interpret.

~~http://www.youtube...~~

Preprocessing



Removed links

Videos and images are hard to interpret.

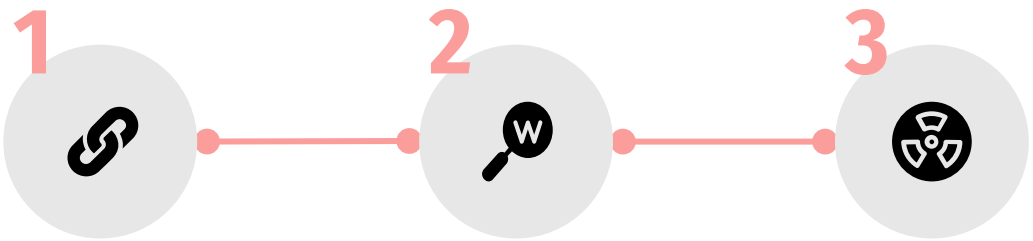
~~http://www.youtube...~~

Spacy

Tokenize
Lemmatize
Remove stop-words

I'm running!! → run

Preprocessing



Removed links

Videos and images are hard to interpret.

~~http://www.youtube...~~

Spacy

Tokenize
Lemmatize
Remove stop-words

I'm running!! → run

Regex

Remove numbers
And punctuation.

Born @ 2012 → born

Preprocessing

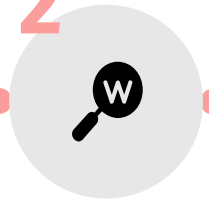
1

Removed links

Videos and images are hard to interpret.

~~http://www.youtube...~~

2

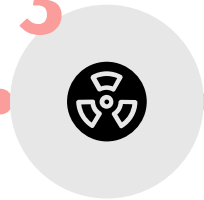


Spacy

Tokenize
Lemmatize
Remove stop-words

I'm running!! → run

3

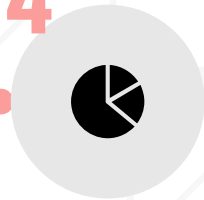


Regex

Remove numbers
And punctuation.

Born @ 2012 → born

4



TF-IDF

Vectorizer

	ability	able
Text1	0.0	0.43
Text2	0.12	0.1

INTP VS ENFP


like
people
thing
find
type
want
feel
come
mean
good
work
use
look
love
time
way
need
friend
try
person
actually
lot
post
life
read
talk
sure
tell
usually
understand


people
like
love
want
thing
time
find
friend
feel
work
good
type
come
try
way
lot
mean
actually
need
thank
life
look
guy
right
sure
infp
talk
person
maybe
lol

02.

Topic Modeling Features

Non-negative

Non-negative matrix

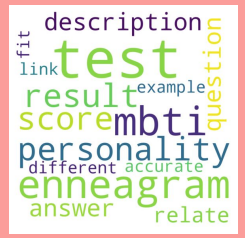
Non-negative matrix factorization

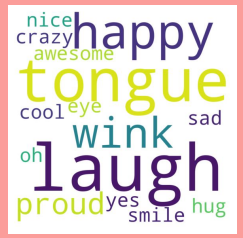
Dating

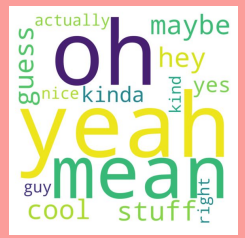
VS

Entertainment

40 Topics







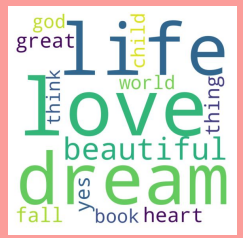
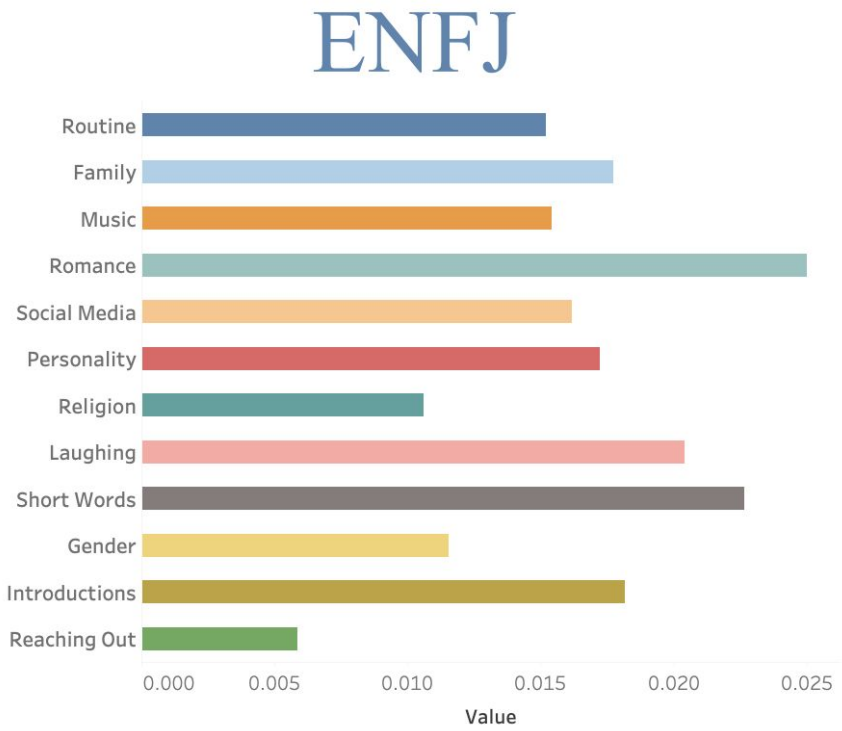


Tableau Viz:

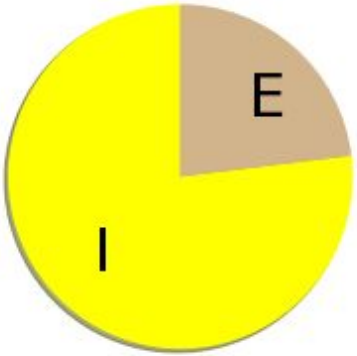
https://public.tableau.com/profile/bgood2me#!/vizhome/MyerBriggsTopics/Dashboard1?publish=yes



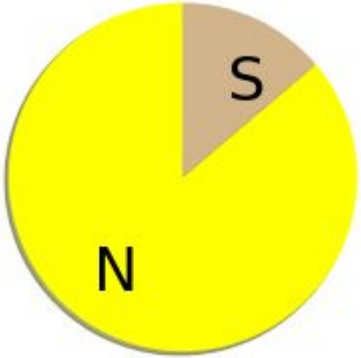
03.

Models & scores

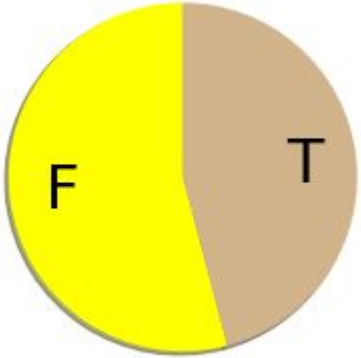
Oversample for Class imbalance



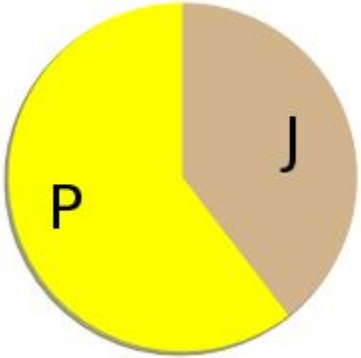
Extrovert Vs. Introvert



Sensing Vs. Intuition



Feeling Vs. Thinking



Judging Vs. Perceiving

Each personality attribute required a different model.

% Accuracy:

	E vs I	N vs S	T vs F	P vs J
My Models	85% Logistic Reg	90% Random Forest	84% SVC	77% Logistic Reg

Each personality attribute required a different model.

% Accuracy:

	E vs I	N vs S	T vs F	P vs J
My Models	85% Logistic Reg	90% Random Forest	84% SVC	77% Multinomial NB
Majority Guess	77% I	86% S	54% F	60% J

Each personality attribute required a different model.

% Accuracy:

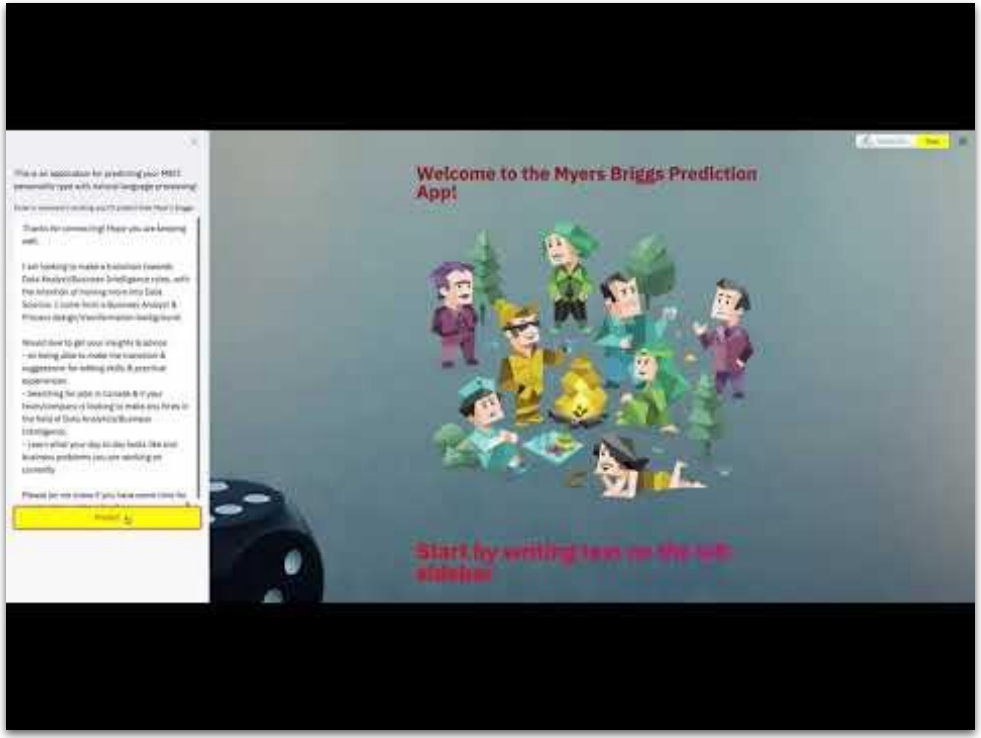
	E vs I	N vs S	T vs F	P vs J	OVERALL
My Models	85% Logistic Reg	90% Random Forest	84% SVC	77% Multinomial NB	50%
Majority Guess	77% I	86% S	54% F	60% J	21%

Each personality attribute required a different model.

% Accuracy:

	E vs I	N vs S	T vs F	P vs J	OVERALL
My Models	85% Logistic Reg	90% Random Forest	84% SVC	77% Multinomial NB	50%
Majority Guess	77% I	86% S	54% F	60% J	21%
Random Guess	50%	50%	50%	50%	6%

App Demo.



Can We

Can We

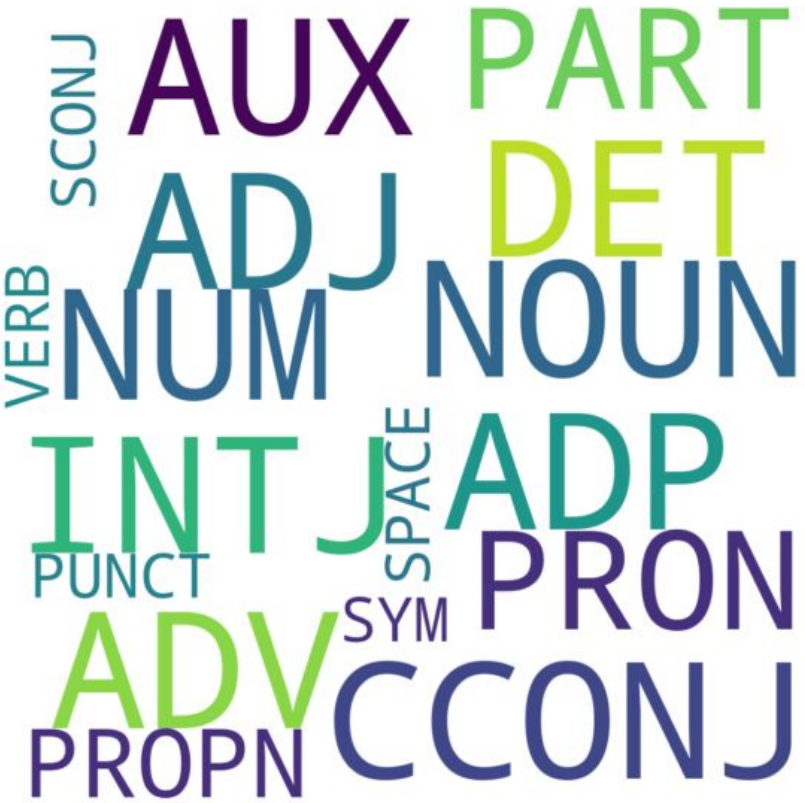
Do

Can We Do Better?

Future Work (implemented)

Parts of Speech

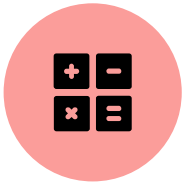
Provides the usage of parts of speech of your sentence



Future Work (implemented)

Parts of Speech

Provides the usage of parts of speech of your sentence



Vader Analysis

Sentiment Analysis

Future Work (implemented)

Parts of Speech

Provides the usage of parts of speech of your sentence

Vader Analysis

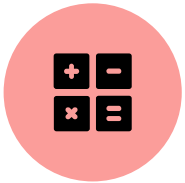
Sentiment Analysis

48% Accuracy

Future Work.

Parts of Speech

Provides the usage of parts of speech of your sentence



Vader Analysis

Provides a score for the text's emotions

Neural Net

Patterns Recognition

Document Upload

Upload long document attachments to analyze

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

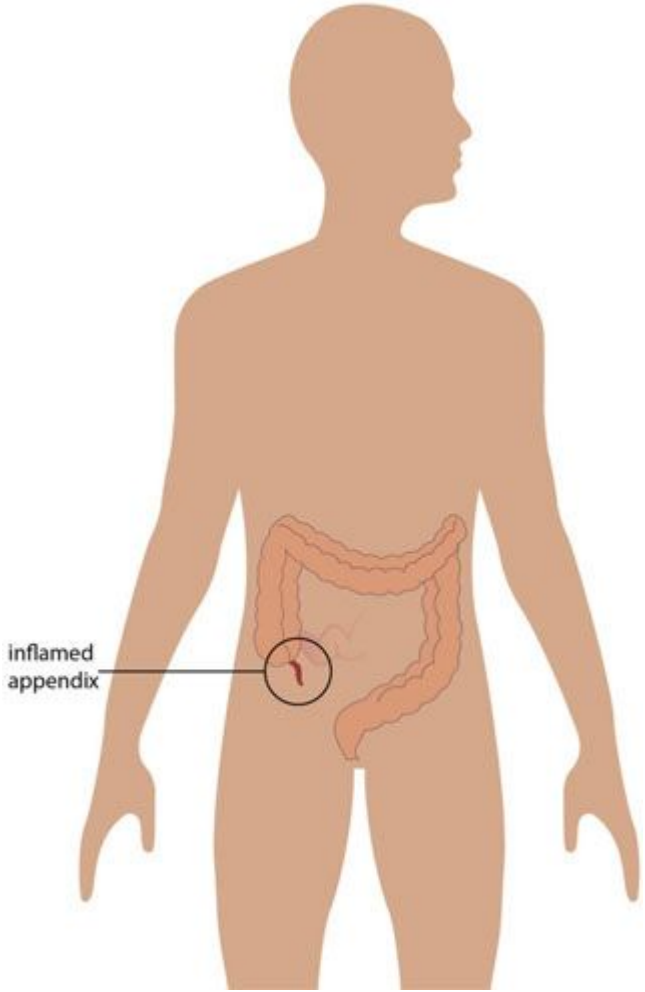
Thanks!

Do you have any questions?

btgood2me@gmail.com
+312-493-5666
github.com/brianhtam

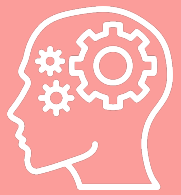
Deployed App:
https://rocky-island-30400.herokuapp.com/

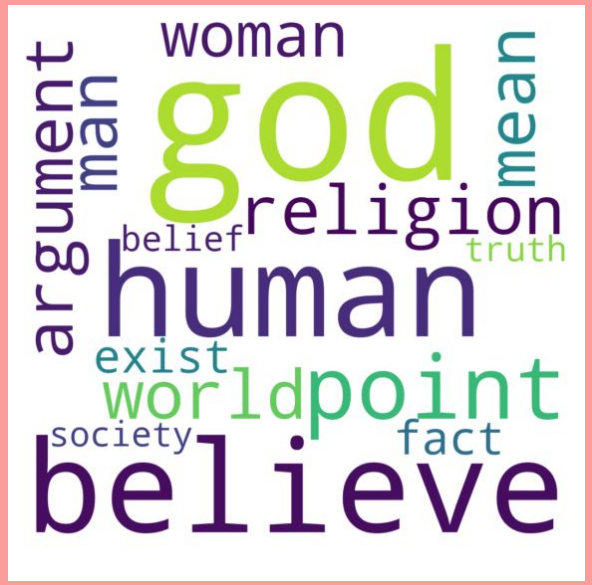
Tableau Viz:
https://public.tableau.com/profile/bgood2me#!/vizhome/MyerBriggsTopics/Dashboard1?publish=yes

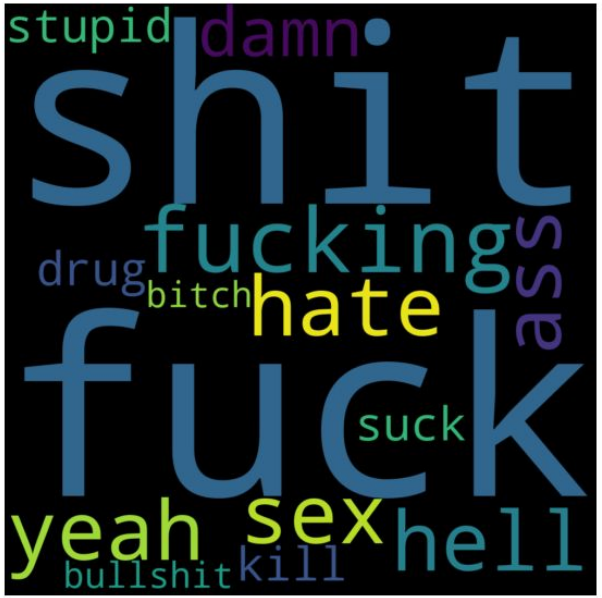


APPENDIX

Vadar : sentiment analysis



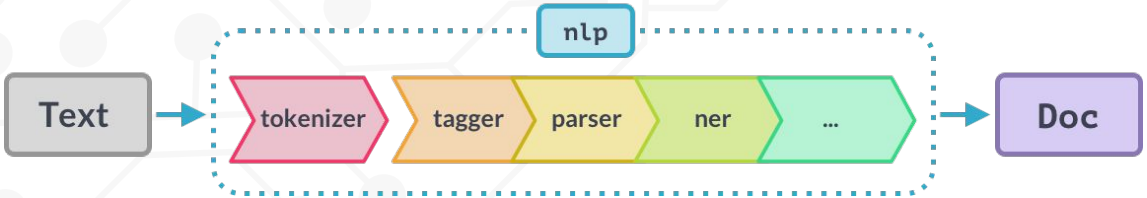


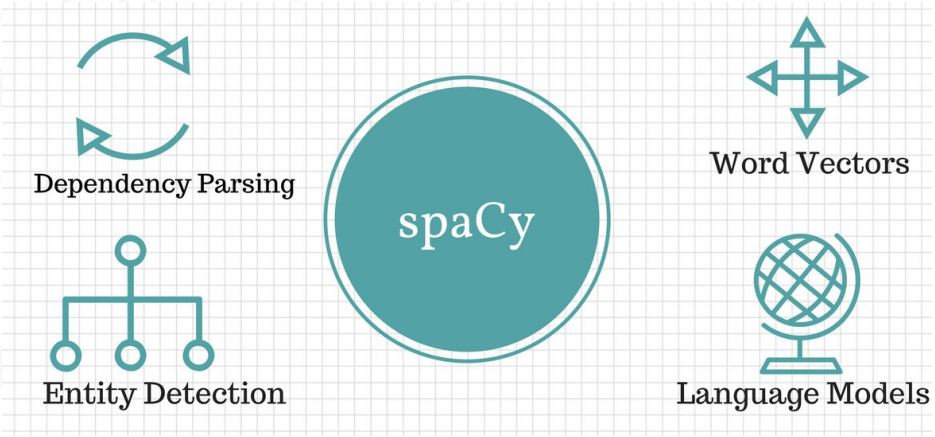


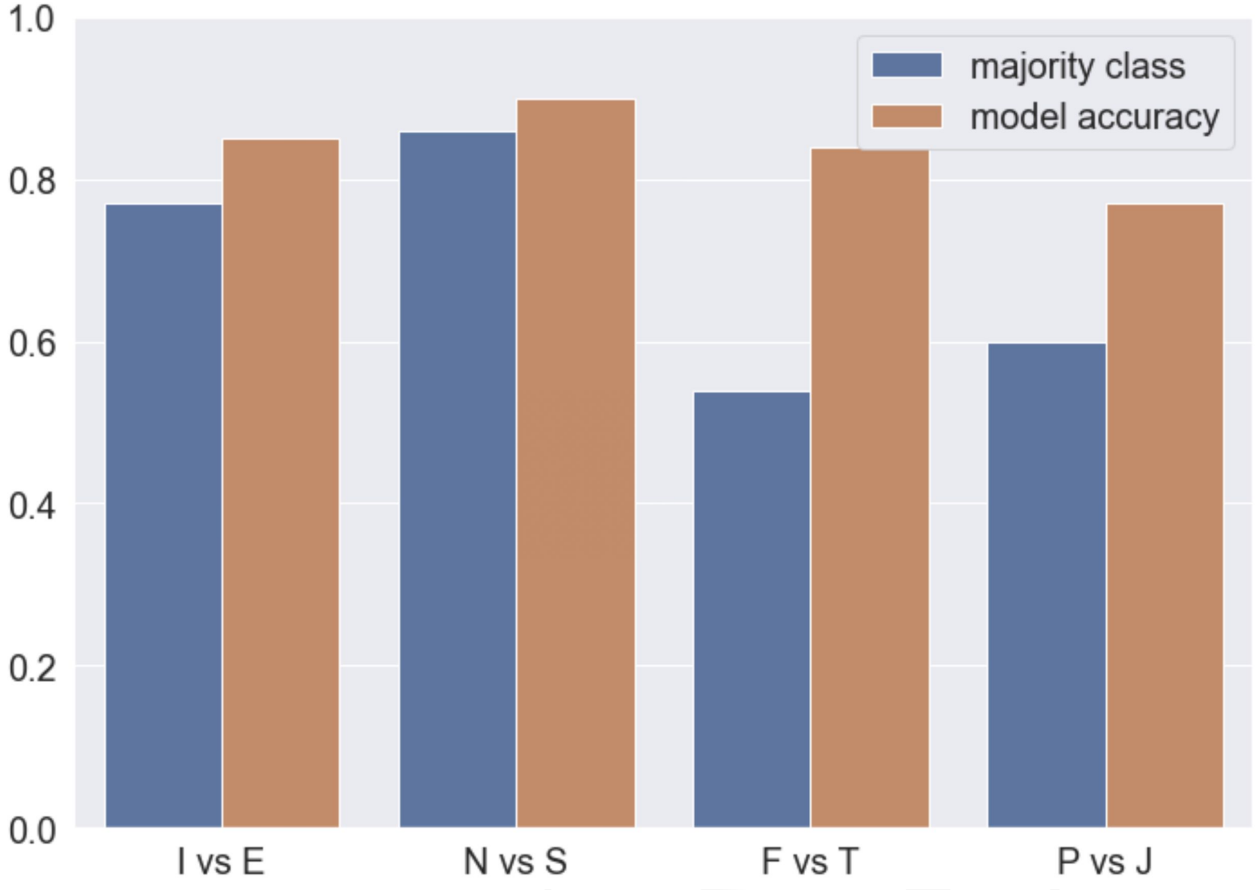
Scoring metrics . . .

F1 scores	90% I 69% E	48% N 94% S	83% F 85% T	71% P 80% J
	vs.	vs.	vs.	vs.
	85% accuracy Logistic Reg	90% accuracy Random Forest	84% accuracy SVC	77% accuracy Multinomial NB

Preprocessing.






1.0
0.8
0.6
0.4
0.2
0.0
majority class
model accuracy
I vs E
N vs S
F vs T
P vs J

```
briantam=# \connect myers_briggs
You are now connected to database "myers_briggs"
myers_briggs=# \dt
                List of relations
 Schema |       Name        | Type  |  Owner
--------+-------------------+-------+----------
 public | benchmark         | table | briantam
 public | bgg               | table | briantam
 public | cleaned_posts     | table | briantam
 public | polarity_scores   | table | briantam
 public | pos               | table | briantam
 public | sentiment         | table | briantam
 public | twitter           | table | briantam
 public | twitter_origin    | table | briantam
(8 rows)
```

Show my scores.

Problem:

Class Imbalance

Solution:

Oversampling

EACH person is different