

Meyer Briggs Type Indicator (MBTI) Personality Prediction

Rachel Lim DSI 16

Agenda

- Problem Statement
- Background
- Approach
- Exploratory Data Analysis (EDA)
- Modelling Process
- Performance
- Putting the Model to Test
- Limitation
- Conclusion



Problem Statement



Hiring is an art



**Skill fit
Culture fit
Relationship**



Social Media

About MBTI

- One of the most most well-known and common personality assessments
- Based on Jung's theory of psychological types
- Measures a person's **preference** across 4 different traits
- Understand someone's natural tendencies

Img:
<https://medium.com/towards-artificial-intelligence/training-ai-to-predict-myers-briggs-personality-types-from-texts-d4e3d7baf17>

Energy

I - Introversion

Introverts get their energy from being alone or in small groups. Remember: being introverted does not mean you are shy. Some employees may be introverted but also very outspoken and confident.

E - Extroversion

Extroverts get their energy from other people. Introverts often find extroverts exhausting because they like to work on multiple things at a quick pace.

Information

N - Intuition

Intuitive people focus on possibilities. They see the big picture and how everything connects. They also enjoy ideas and concepts for their own sake.

S - Sensing

Sensors are realists. They use their five senses to assess how things are and pay attention to concrete facts and details.

Decisions

T - Thinking

Thinkers are logical and make decisions analytically. They value fairness and enjoy finding the flaws in an argument. Usually they are level-headed.

F - Feeling

Feelers base decisions on personal values and how their decisions may affect others. They are usually described as empathetic and warm.

Organization

J - Judging

Judgers like rules, deadlines, and structure. They prefer to have clear, detailed instructions and want to make plans and stick to them.

P - Perceiving

Perceivers are spontaneous and flexible. They see deadlines as negotiable and like to improvise.

Uses of MBTI:

Where:

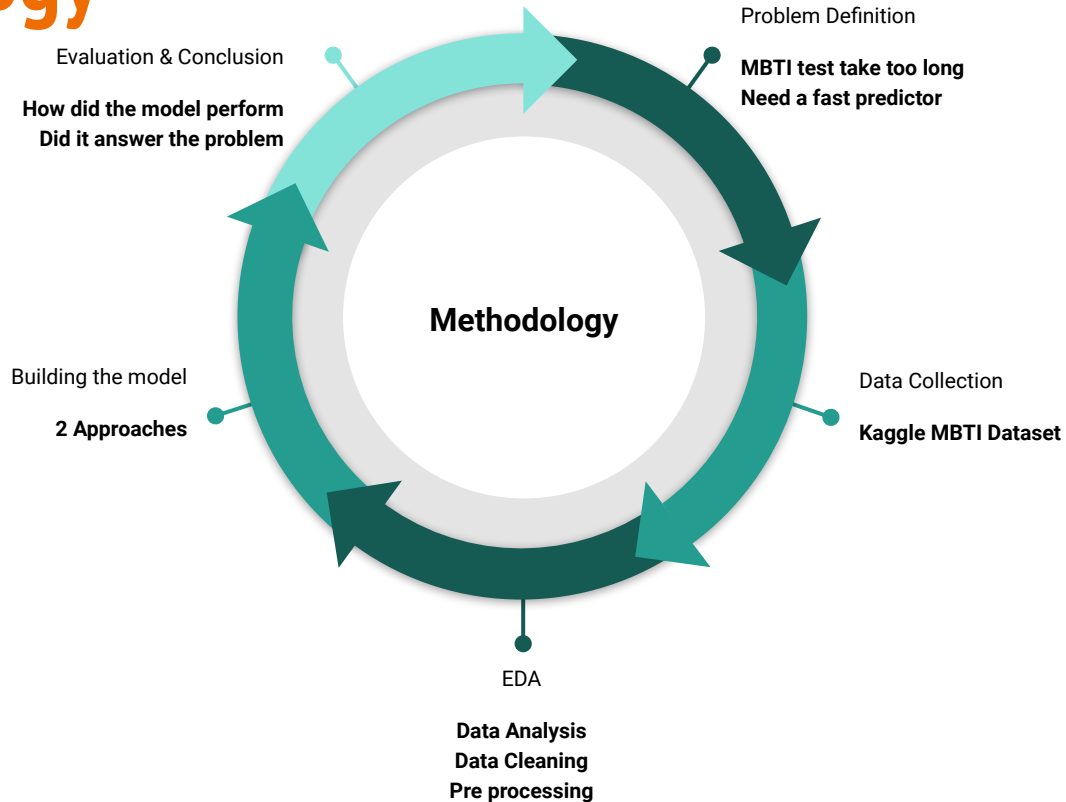
- Career Advice
- Team Cohesion

How:

- Understand how each type react
- Better relationships
- Engagement Style



Methodology



Exploratory Data Analysis (EDA)

- Data Source
 - <https://www.kaggle.com/datasnaek/mbti-type>
- Data distribution
- Length of posts, words per comment, POS per comment
- Prominent Words

Data

This dataset contains 8675 rows of data, on each row is a person's:

- Type (This person's 4 letter MBTI code/type)
- A section of each of the last 50 things they have posted in PersonalityCafe (Each entry separated by "|||" (3 pipe characters))

```
df.shape
```

```
(8675, 2)
```

type	posts
INFJ	'http://www.youtube.com/watch?v=qsXHcwe3krw http://41.media.tumblr.com/tumblr_lfouy03PMA1qa1rooo1_500.jpg enfp and intj moments https://www.youtube.com/watch?v=iz7IE1g4XM4 sportscenter not top ten plays https://www.youtube.com/watch?v=uCdfze1etec pranks What has been the most life-chan...
ENTP	'I'm finding the lack of me in these posts very alarming. Sex can be boring if it's in the same position often. For example me and my girlfriend are currently in an environment where we have to creatively use cowgirl and missionary. There isn't enough... Giving new meaning to 'Game' theory. ...
INTP	'Good one _____ https://www.youtube.com/watch?v=fHiGbolFFGw Of course, to which I say I know; that's my blessing and my curse. Does being absolutely positive that you and your best friend could be an amazing couple count? If so, than yes. Or it's more I could be madly in love in case I ...

Data - Sample post

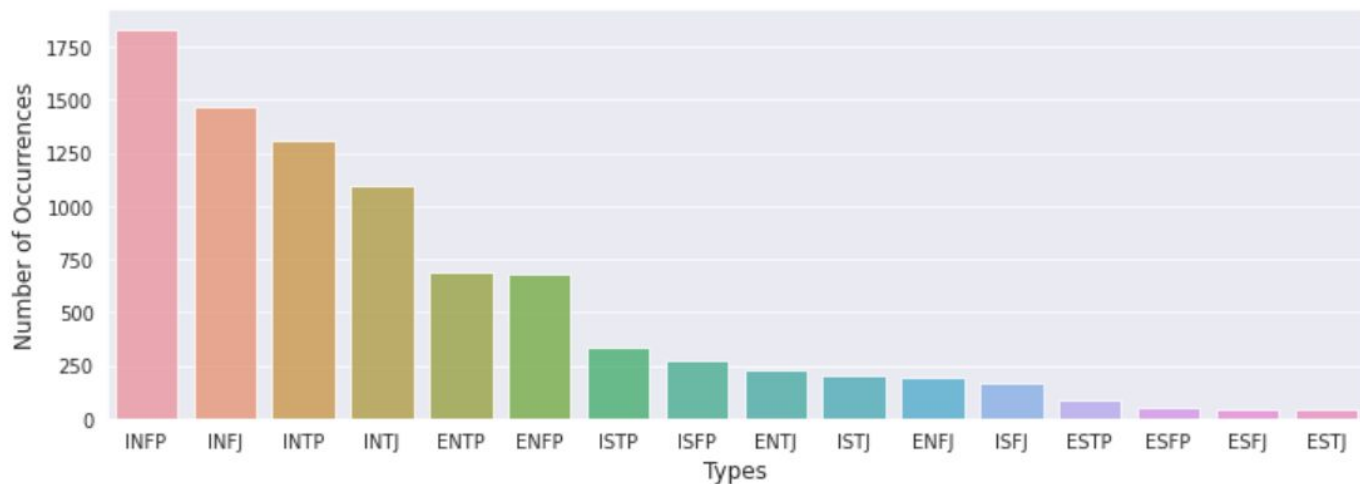
```
df.loc[8612:8612]['posts'].values[0]
```

"You tell him! Christ. I do wanna thank everyone for their input so far.|||Um we're lesbians...|||She just quit her canvassing job and I'm proud of her for standing up for herself. She's DEFINITELY a gifted writer, but she's more of a poet. I'm not sure how you incorporate that in a job...|||I showed her the posts and I think she's going to look into the libraries...she IS an avid reader, after all. Any other ideas in case this doesn't work out?|||Thanks! I'm glad I could make someone laugh!|||Yes, you bet your ass I can. I swear I'm sober. Just FYI.|||What, I can't talk about what pretty much everyone does during lonely nights? I AM A SEXUAL 8. And my girlfriend LOVES it.|||INFPs are adorable. I don't know what I'd do without mine...oh wait. Yes I do. I'd be masturbating a lot more and leaving the house recreationally less.|||TootsieBear267 Thank you! It comes naturally. XD|||What would you like to read?|||Hello all, Yeah, yeah, an ENTJ posting in your thread. I have a (very serious) girlfriend who is your type, and she's been rather upset lately about job choices. She's just out of high school...|||I want minions because then they'd be mini-mes.|||Sorry to disappoint you. :P|||Thank you guys! I like to think I'm fascinating.|||Hello everyone! I'm Yuugi, and I'm an ENTJ 8w7 sx/sp/so. I enjoy long, romantic walks from my house to the bar (hopefully returning with an INF P in my arms). My hobbies include world domination...|||When you panic whenever you lose something because that's imperfect, dammit!|||Welcome! I'm new as well. You mentioned anime...which ones? Sent from my SAMSUNG-SGH-I337 using Tapatalk|||I'd go in to teaching, to keep it realistic, but it depends on how long you want to be on school. Sent from my SAMSUNG-SGH-I337 using Tapatalk|||I'm definitely an N but I like literal things. I'm a pretty black-and-white thinker. I have Nonverbal Learning Disorder, which impairs my ability to process nonverbal cues. But at the same time,..."

Data

```
df.isnull().sum()
```

```
type      0  
posts     0  
dtype: int64
```



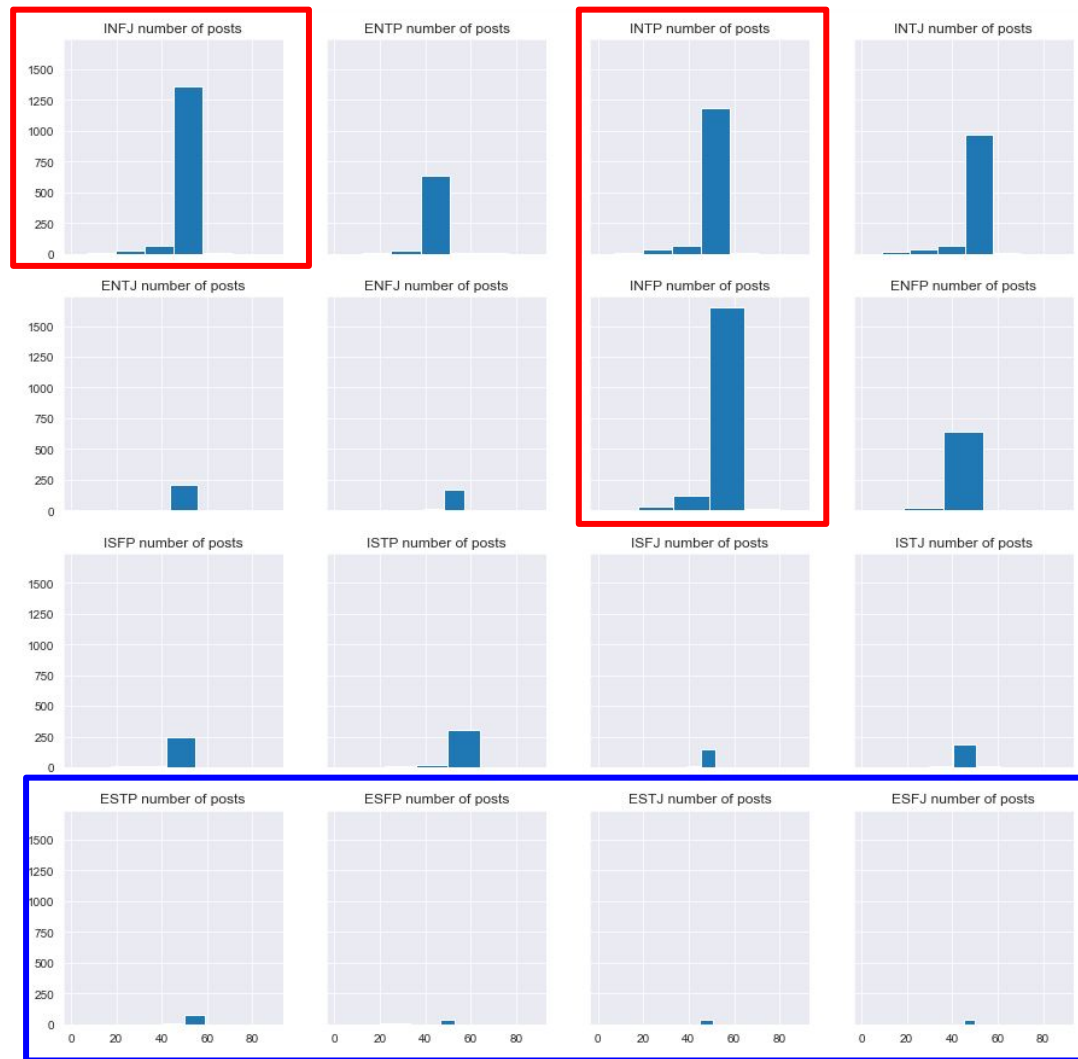
Baseline:
Majority Class: 0.211

	count	%_spread
INFP	1832	0.211
INFJ	1470	0.169
INTP	1304	0.150
INTJ	1091	0.126
ENTP	685	0.079
ENFP	675	0.078
ISTP	337	0.039
ISFP	271	0.031
ENTJ	231	0.027
ISTJ	205	0.024
ENFJ	190	0.022
ISFJ	166	0.019
ESTP	89	0.010
ESFP	48	0.006
ESFJ	42	0.005
ESTJ	39	0.004

Number of Posts

There are certain types that contain posts that doesn't have the full 50 'subposts', (e.g. INFJ and INTP types) while some types have more than 50 subposts (INFP).

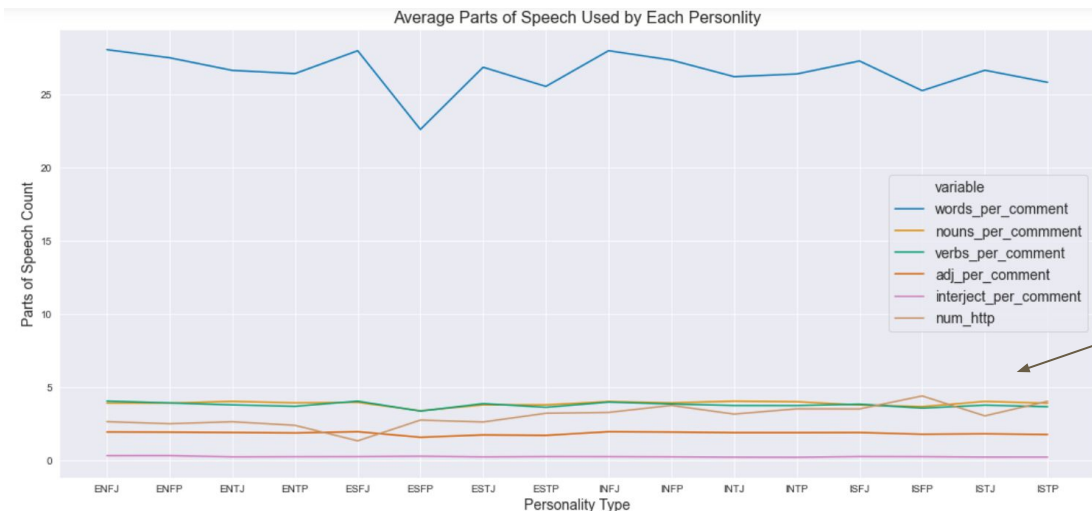
From these graphs, we also see that there are certain types (ES types) have very little posts and subposts



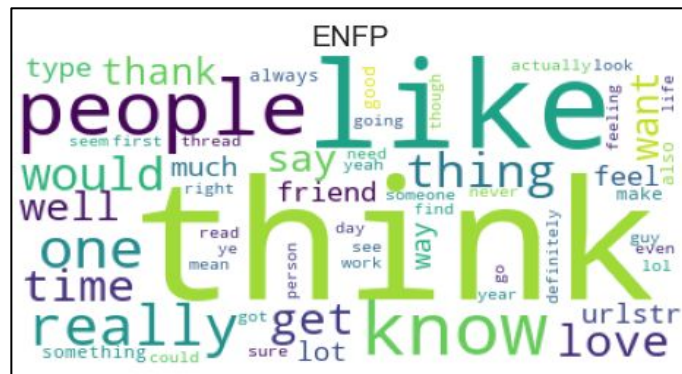
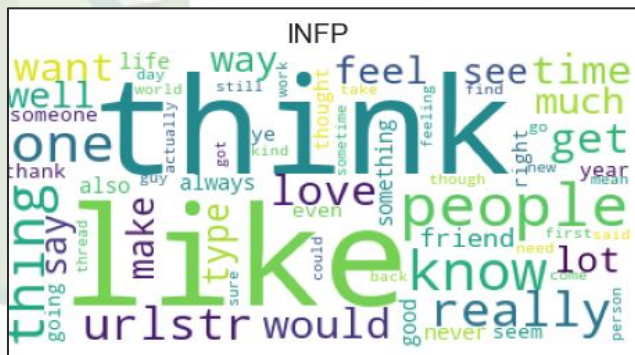
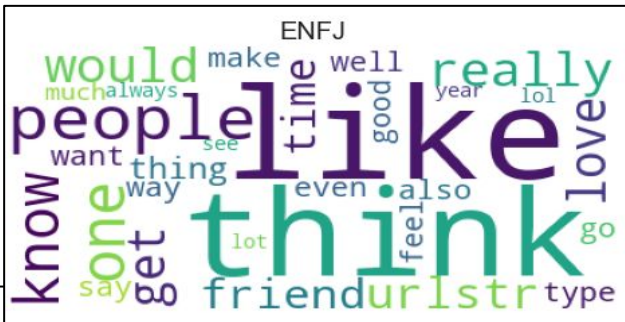
Parts of Speech

Part of Speech Tagging involves identifying the part of speech for every word. Using SpaCy library to tokenize and tagging the posts, we try to determine if a certain personality types uses more of a POS, particularly:

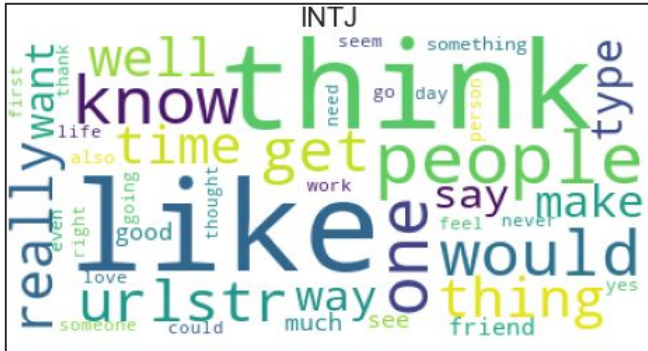
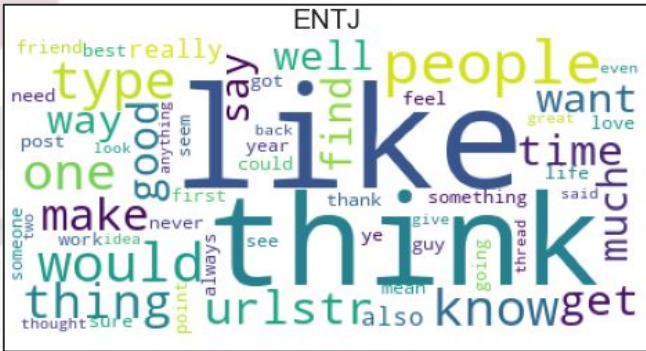
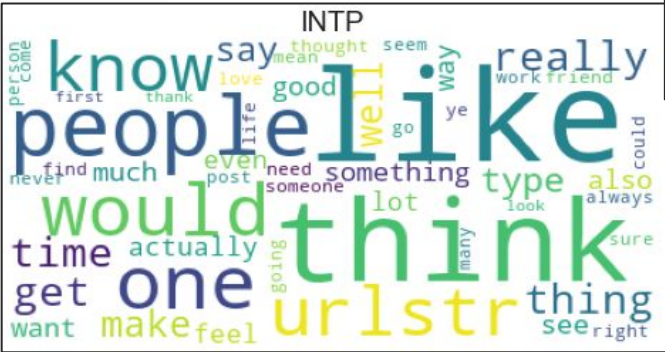
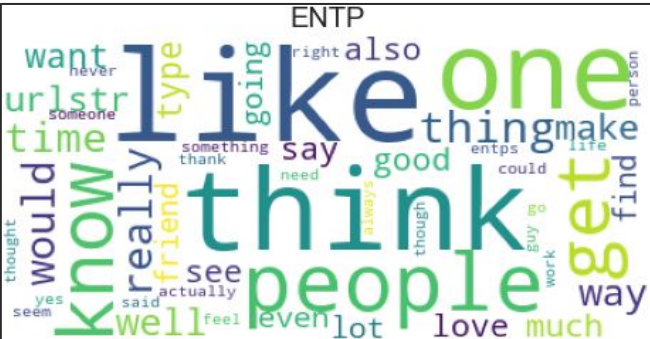
- nouns ("NOUN")- a word that functions as the name of a specific object or set of objects
- verbs ("VERB") - action words in a sentence that describe what the subject is doing
- adjectives ("ADJ") - modifies a noun or noun phrase or describes its referent
- Interjections ("INTJ") - demonstrates the emotion or feeling



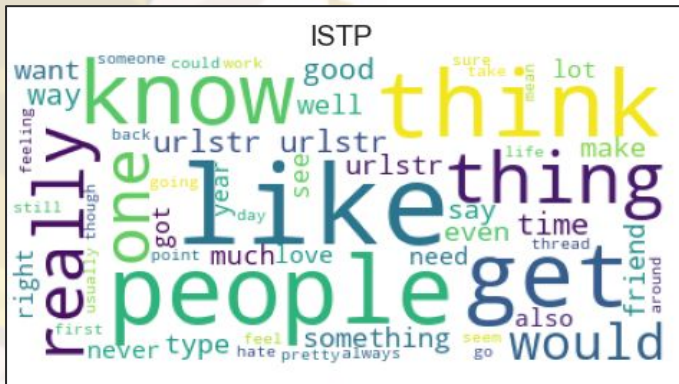
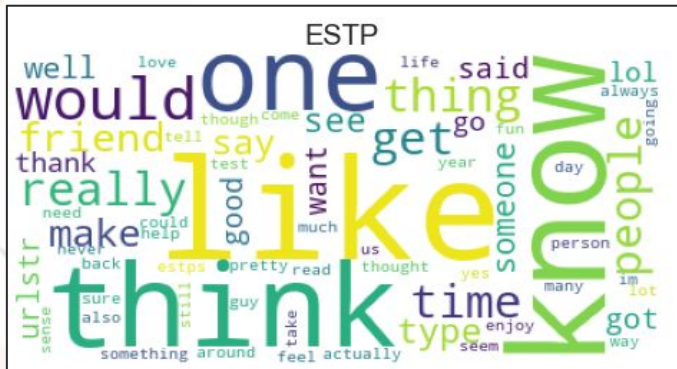
There is no significant variance in the POS/comment for each type

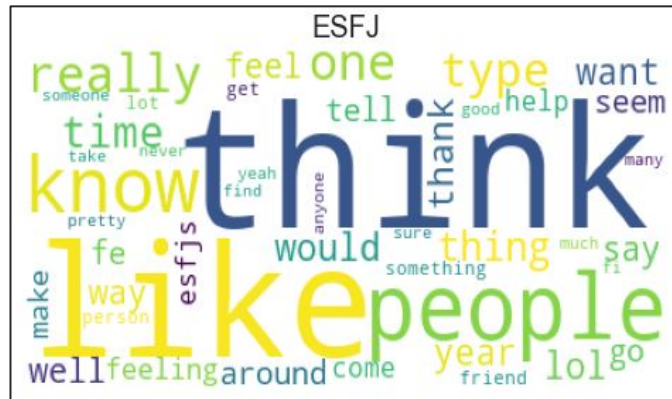


NT Types - Intro



SP Types - Adv





Prominent Words

INFJ	ENTP	INTP	INTJ	ENTJ	ENFJ	INFP	ENFP	ISFP	ISTP	ISFJ	ISTJ	ESTP	ESFP	ESTJ	ESFJ
like	like	like	like	like	like	like	like	like	like	like	like	like	like	like	like
think	think	think	people	think	think	think	think	think	people	think	think	think	think	think	think
people	people	people	think	people	people	people	people	people	think	people	people	people	people	people	people
know	one	would	would	would	know	really	know	really	get	really	one	get	know	would	know
one	know	one	one	one	really	know	really	know	know	know	know	know	really	one	one
really	would	know	know	know	one	one	one	one	would	would	would	one	would	know	really
would	really	really	really	get	would	would	would	get	one	one	really	really	one	get	would
get	get	get	get	really	get	feel	get	urlstr	really	time	get	would	get	really	get
time	time	urlstr	time	time	feel	get	love	would	urlstr	get	time	well	well	well	time
feel	well	time	urlstr	well	love	urlstr	time	time	time	feel	urlstr	type	urlstr	time	type

words	count
like	16
people	16
know	16
one	16
would	16
really	16
get	16
time	16
well	16
much	16
think	16
good	15
say	15
things	14
urlstr	14

These common words will be removed since they appear in all types and will not be of an impact

Data Preparation

- Word Vectorizing using **Term Frequency-Inverse Document Frequency (TF-IDF)** Vectorizer
 - Max Features = 10,000 (total Word size = 78899 , after removing stopwords)
- **SMOTETomek** to handle imbalance classes
 - Combine over- and under-sampling using SMOTE (Synthetic Minority Oversampling Technique) and Tomek links to create a more balanced dataset

Modelling Approach

Approach 1:

- Multiclass Classifier (across all 16 types)
- 5 folds cross validation using GridSearch
- Models: Logistic Regression, Naive Bayes, Random Forest, XGBoost
- Metrics: F1-Score, Matthews Correlation Coefficient (MCC) Score

Approach 2:

- Binary Classifiers for each MBTI axis (I vs E, N vs S, F vs T, J vs P)
- 5 folds cross validation using GridSearch
- Models: Logistic Regression, XGBoost
- Metrics: AUC ROC score

Approach 1: Multiclass Modelling

Baseline:

Majority Class: 0.211

	Logistic Reg	Naive Bayes	Random Forest	XGBoost
Accuracy	0.351	0.349	0.228	0.281
Misclassification Rate	0.649	0.651	0.772	0.719
F1-Score Macro	0.232	0.225	0.126	0.156
F1-Score Weighted	0.364	0.352	0.185	0.242
Matthews Correlation	0.271	0.258	0.090	0.148

Multiclass Modelling Logistic Regression

Best CV score : 0.33706095725488866
 Training score : 0.643840330350998
 Validation score : 0.35103038770520434
 Accuracy : 0.35103038770520434
 MCC Score : 0.2709083104972687

True \ Pred																
	ENFJ	ENFP	ENTJ	ENTP	ESFJ	ESFP	ESTJ	ESTP	INFJ	INFP	INTJ	INTP	ISFJ	ISFP	ISTJ	ISTP
ENFJ	17	12	2	3	2	1	0	1	4	6	2	1	2	4	4	2
ENFP	10	88	8	19	5	3	1	2	23	31	9	7	2	6	2	7
ENTJ	2	1	13	14	0	0	0	5	2	6	12	5	5	3	4	4
ENTP	6	15	7	71	1	1	3	5	15	9	26	39	5	5	6	12
ESFJ	4	1	0	1	0	0	1	0	2	2	1	0	0	2	0	0
ESFP	1	4	2	1	0	0	1	0	1	1	3	1	0	1	0	0
ESTJ	0	1	1	3	0	0	1	1	2	0	0	0	0	2	1	1
ESTP	1	3	2	3	0	1	0	2	2	1	2	2	0	2	0	8
INFJ	18	27	15	28	8	3	1	6	171	66	41	30	12	26	13	20
INFP	17	67	11	14	2	4	6	9	70	250	27	45	10	34	16	23
INTJ	4	8	21	30	3	6	3	4	31	22	134	54	6	9	10	15
INTP	8	11	15	51	1	4	4	7	16	30	75	160	9	11	6	22
ISFJ	0	6	1	1	2	1	0	0	7	4	3	8	14	5	3	0
ISFP	4	12	2	6	1	0	1	0	5	11	6	3	5	27	2	4
ISTJ	3	0	5	3	0	1	0	1	4	7	8	6	6	1	17	6
ISTP	2	7	11	8	1	0	1	2	1	9	8	11	3	4	3	40

Approach 2: Binary Modelling

==== I vs E Value Counts =====

I 0.770

E 0.230

Name: ie, dtype: float64

==== N vs S Value Counts =====

N 0.862

S 0.138

Name: ns, dtype: float64

==== T vs F Value Counts =====

F 0.541

T 0.459

Name: tf, dtype: float64

==== J vs P Value Counts =====

P 0.604

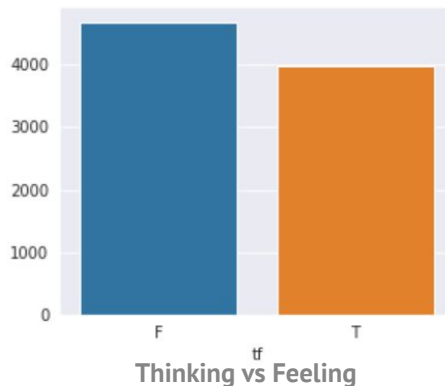
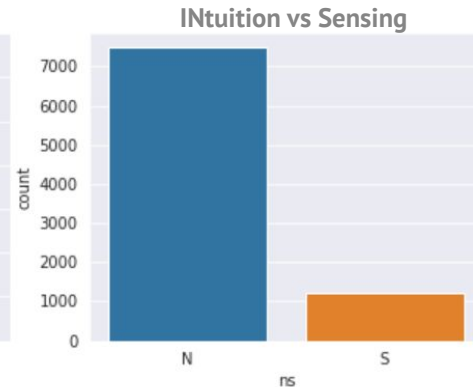
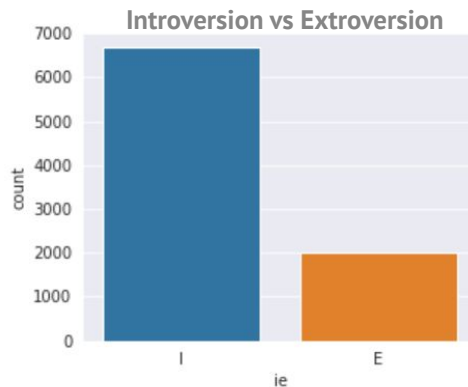
J 0.396

Name: jp, dtype: float64

Encoding

Convert sets into binary labels:

- 0 for I-types vs 1 for E-types
- 0 for N-types vs 1 for S-types
- 0 for T-types vs 1 for F-types
- 0 for J-types vs 1 for P-types



Binary Modelling Performance

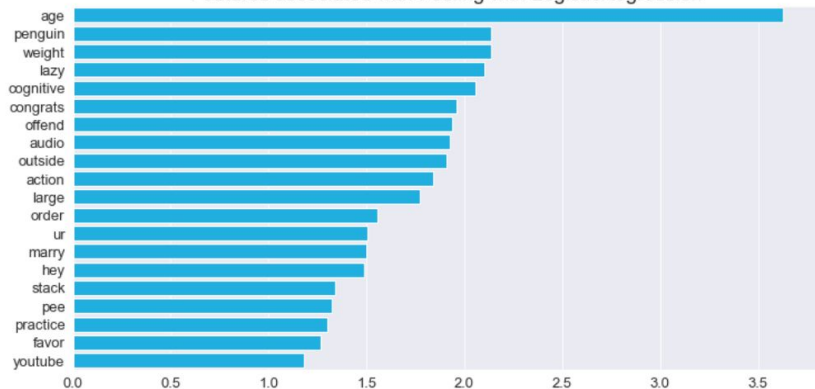
	IE LogReg	IE XGBoost	NS LogReg	NS XGBoost	TF LogReg	TF XGBoost	JP LogReg	JP XGBoost
Accuracy	0.624	0.740	0.686	0.836	0.502	0.530	0.656	0.640
Misclassification Rate	0.376	0.260	0.314	0.164	0.498	0.470	0.344	0.360
Sensitivity (Recall)	0.347	0.033	0.251	0.043	0.532	0.840	0.710	0.889
Specificity	0.707	0.952	0.756	0.963	0.467	0.165	0.575	0.259
Precision	0.262	0.172	0.141	0.156	0.540	0.543	0.718	0.647
True Positive	229.000	22.000	99.000	17.000	824.000	1,301.000	1,228.000	1,538.000
False Positive	645.000	106.000	602.000	92.000	701.000	1,097.000	482.000	839.000
False Negative	431.000	638.000	296.000	378.000	725.000	248.000	502.000	192.000
True Negative	1,558.000	2,097.000	1,866.000	2,376.000	613.000	217.000	651.000	294.000
F1-Score	0.299	0.056	0.181	0.067	0.536	0.659	0.714	0.749
AUC Score	0.529	0.494	0.508	0.503	0.506	0.502	0.701	0.651
Matthews Correlation	0.050	-0.030	0.005	0.010	-0.002	0.007	0.284	0.193

Prominent Words

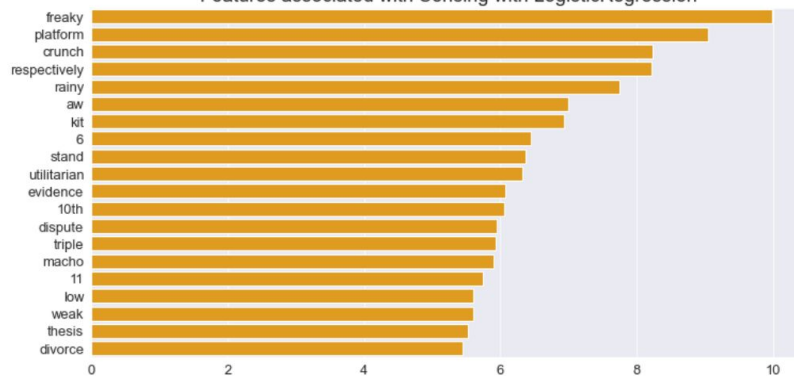
Features associated with Extrovert with LogisticRegression



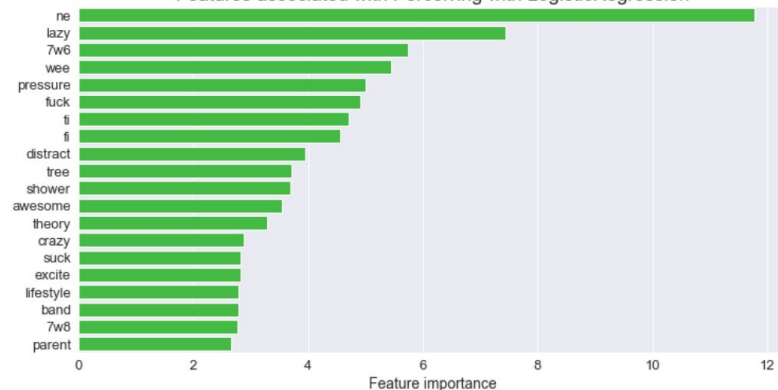
Features associated with Feeling with LogisticRegression



Features associated with Sensing with LogisticRegression



Features associated with Perceiving with LogisticRegression



Putting the Model to Test

Famous Quotes:

"I will build a great wall--and nobody builds walls better than me, believe me--and I'll build them very inexpensively. I will build a great, great wall on our southern border, and I will make Mexico pay for that wall. Mark my words."

"We have a 5 billion dollar website. I have so many websites. I have them all over the place... I hire people... it costs me three dollars."

"The line of 'Make America great again,' the phrase, that was mine, I came up with it about a year ago, and I kept using it, and everybody's using it, they are all loving it. I don't know, I guess I should copyright it, maybe I have copyrighted it."

Putting the Model to Test

```
predictMBTI(X_test)
```

```
===== Introversion vs Extroversion =====  
[[0.08689449 0.91310551]]
```

```
===== INTuition vs Sensing =====  
[[0.82423829 0.17576171]]
```

```
===== Thinking vs Feeling =====  
[[0.48061794 0.51938206]]
```

```
===== Judging vs Perceiving =====  
[[0.43385374 0.56614626]]
```

```
Results: ENFP
```

```
'ENFP'
```



ESTPs aren't policy people. They don't want to talk about things in fine details. Instead, they use their energy to uncover solutions. Spontaneous by their nature, ESTPs are also very perceptive. If you need someone to be the face of a cause or to rally people behind an unpopular initiative, choose an ESTP.

Putting the Model to Test

Famous Quotes:

Everyone can rise above their circumstances and achieve success if they are dedicated to and passionate about what they do. Do not judge me by my successes, judge me by how many times I fell down and got back up again.

"Justice and I became friends, though we were opposites in many ways – he was extroverted, I was introverted; he was lighthearted, I was serious.

I learned that courage was not the absence of fear, but the triumph over it. The brave man is not he who does not feel afraid, but he who conquers that fear. No one is born hating another person because of the color of his skin, or his background, or his religion.

Putting the Model to Test

```
predictMBTI(X_test2)
```

```
===== Introversion vs Extroversion =====  
[[0.65223171 0.34776829]]
```

```
===== INTuition vs Sensing =====  
[[0.82881884 0.17118116]]
```

```
===== Thinking vs Feeling =====  
[[0.58140514 0.41859486]]
```

```
===== Judging vs Perceiving =====  
[[0.54958012 0.45041988]]
```

```
Results: INTJ
```

```
'INTJ'
```



**INFJ
Personality
Type**

No other personality type is better suited to create a movement to right a wrong, no matter how big or small,” and INFJs “have very strong opinions and will fight tirelessly for an idea they believe in

Limitations

- Models are better at predicting the majority types
- Dataset is not balanced (we need more E, S types)
- Posts scrapped are mainly word posts and no image. Certain personality may be more expressive using images or media other than words
- The posts available is not comprehensive enough to determine the various types

Conclusion

- Certain personality traits may not be defined well in words or language.
- Personality Trait is a far more complex area and more than just words, and language expression.
- More research needs to be done to determine if in future we are able to predict the personality just by text alone.
- Final Note: MBTI does not predict Success. It assists people in becoming more aware of why people we work with and deal with on a daily basis seem to react, organise and behave in the way they do