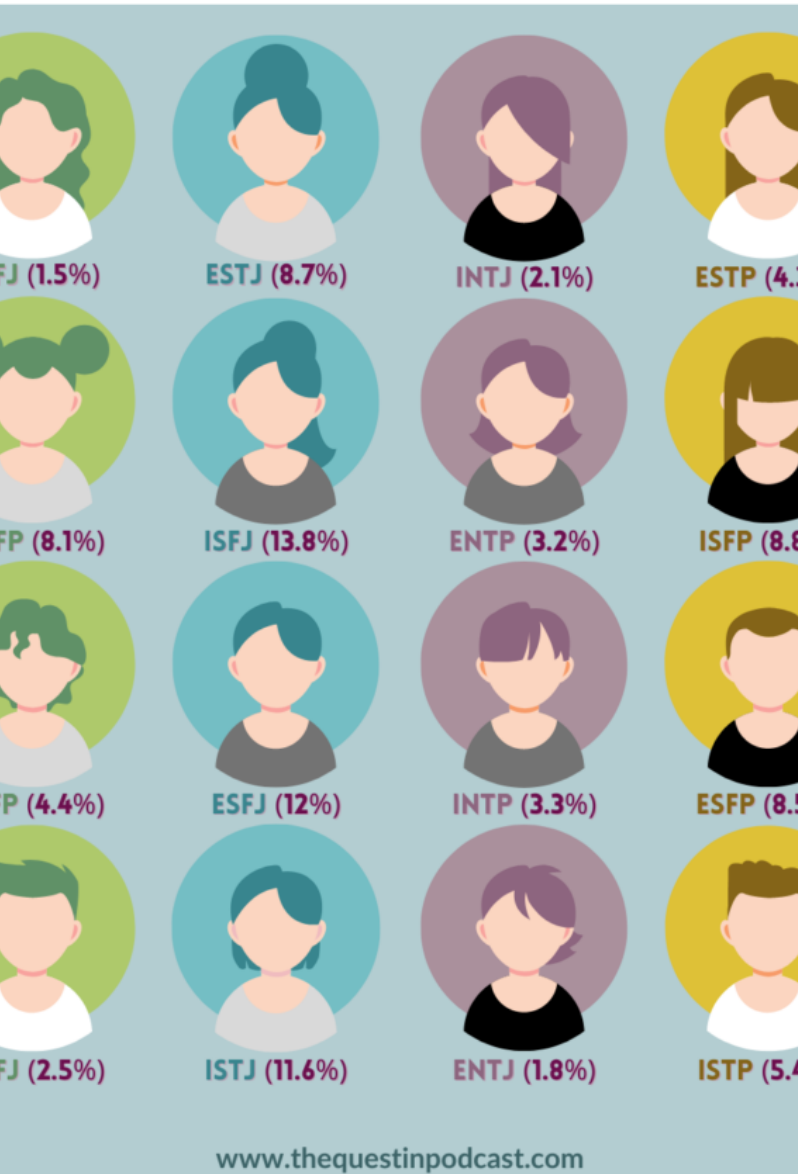


MBTI STATISTICS

POPULATION PERCENTAGES



MBTI Classification with NLP

Predicting people's MBTI types through online posts

Nathan Torento

Kaggle Dataset

(MBTI) Myers-Briggs Personality Type Dataset

8600 rows containing “type” describing a person’s MBTI type, as well as “posts” containing the last 50 posts a user posted on forums in a website called “PersonalityCafe”

Screenshot Sample of Dataset Below

	type	posts
0	INFJ	'http://www.youtube.com/watch?v=qsXHcwe3krw ...
1	ENTP	'I'm finding the lack of me in these posts ver...
2	INTP	'Good one _____ https://www.youtube.com/wat...
3	INTJ	'Dear INTP, I enjoyed our conversation the o...
4	ENTJ	'You're fired. That's another silly misconce...

What's Your Personality Type?

Use the questions on the outside of the chart to determine the four letters of your Myers-Briggs type.
For each pair of letters, choose the side that seems most natural to you, even if you don't agree with every description.

1. Are you outwardly or inwardly focused? If you:

- Could be described as talkative, outgoing
- Like to be in a fast-paced environment
- Tend to work out ideas with others, think out loud
- Enjoy being the center of attention

then you prefer

E
Extraversion

- Could be described as reserved, private
- Prefer a slower pace with time for contemplation
- Tend to think things through inside your head
- Would rather observe than be the center of attention

then you prefer

I
Introversion

2. How do you prefer to take in information? If you:

- Focus on the reality of how things are
- Pay attention to concrete facts and details
- Prefer ideas that have practical applications
- Like to describe things in a specific, literal way

then you prefer

S
Sensing

- Imagine the possibilities of how things could be
- Notice the big picture, see how everything connects
- Enjoy ideas and concepts for their own sake
- Like to describe things in a figurative, poetic way

then you prefer

N
Intuition

ISTJ

Responsible, sincere, analytical, reserved, realistic, systematic. Hardworking and trustworthy with sound practical judgment.

ISFJ

Warm, considerate, gentle, responsible, pragmatic, thorough. Devoted caretakers who enjoy being helpful to others.

INFJ

Idealistic, organized, insightful, dependable, compassionate, gentle. Seek harmony and cooperation, enjoy intellectual stimulation.

INTJ

Innovative, independent, strategic, logical, reserved, insightful. Driven by their own original ideas to achieve improvements.

ISTP

Action-oriented, logical, analytical, spontaneous, reserved, independent. Enjoy adventure, skilled at understanding how mechanical things work.

ISFP

Gentle, sensitive, nurturing, helpful, flexible, realistic. Seek to create a personal environment that is both beautiful and practical.

INFP

Sensitive, creative, idealistic, perceptive, caring, loyal. Value inner harmony and personal growth, focus on dreams and possibilities.

INTP

Intellectual, logical, precise, reserved, flexible, imaginative. Original thinkers who enjoy speculation and creative problem solving.

3. How do you prefer to make decisions? If you:

- Make decisions in an impersonal way, using logical reasoning
- Value justice, fairness
- Enjoy finding the flaws in an argument
- Could be described as reasonable, level-headed

then you prefer

T
Thinking

- Base your decisions on personal values and how your actions affect others
- Value harmony, forgiveness
- Like to please others and point out the best in people
- Could be described as warm, empathetic

then you prefer

F
Feeling

4. How do you prefer to live your outer life? If you:

- Prefer to have matters settled
- Think rules and deadlines should be respected
- Prefer to have detailed, step-by-step instructions
- Make plans, want to know what you're getting into

then you prefer

J
Judging

- Prefer to leave your options open
- See rules and deadlines as flexible
- Like to improvise and make things up as you go
- Are spontaneous, enjoy surprises and new situations

then you prefer

P
Perceiving

ESTP

Outgoing, realistic, action-oriented, curious, versatile, spontaneous. Pragmatic problem solvers and skillful negotiators.

ESFP

Playful, enthusiastic, friendly, spontaneous, tactful, flexible. Have strong common sense, enjoy helping people in tangible ways.

ENFP

Enthusiastic, creative, spontaneous, optimistic, supportive, playful. Value inspiration, enjoy starting new projects, see potential in others.

ENTP

Inventive, enthusiastic, strategic, enterprising, inquisitive, versatile. Enjoy new ideas and challenges, value inspiration.

ESTJ

Efficient, outgoing, analytical, systematic, dependable, realistic. Like to run the show and get things done in an orderly fashion.

ESFJ

Friendly, outgoing, reliable, conscientious, organized, practical. Seek to be helpful and please others, enjoy being active and productive.

ENFJ

Caring, enthusiastic, idealistic, organized, diplomatic, responsible. Skilled communicators who value connection with people.

ENTJ

Strategic, logical, efficient, outgoing, ambitious, independent. Effective organizers of people and long-range planners.

Objectives

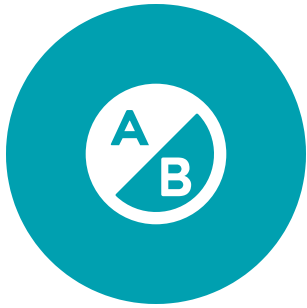
1

Creating a text classification model to determine a user's MBTI type based on their posts.

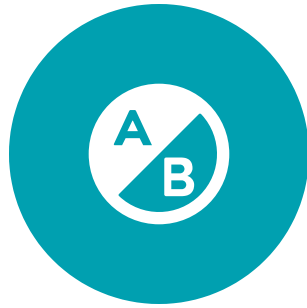
2

Assessing the difficulty of this task by evaluating the maximum possible accuracy.

Targets



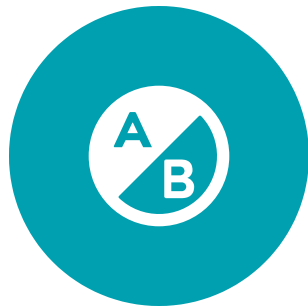
E/I



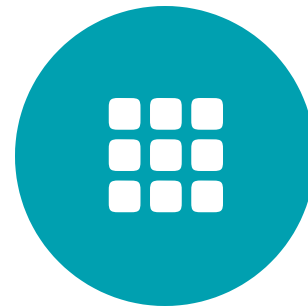
S/N



T/F



J/P



MBTI Type

Data Wrangling

- Preprocessing

Splitting Posts

Converting Emojis to Text

Removing @ Mentions

Removing Non-Alphanumeric Characters

Standardizing MBTI Mentions

Lowercasing

Tokenization and Lemmatization

Removing Stopwords

Reassembling Tokens into Strings

- Feature Extraction

Creating Dummy Variables for MBTI Classifications

Measuring Post Length

Readability Assessment

Capitalization Analysis

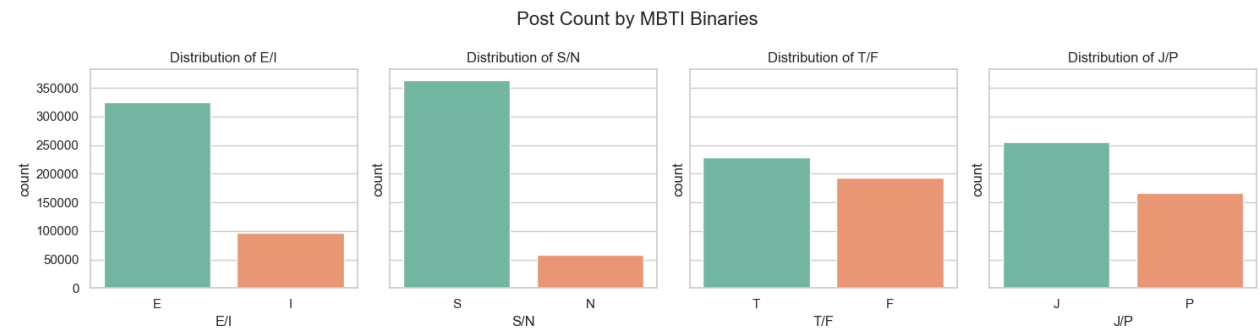
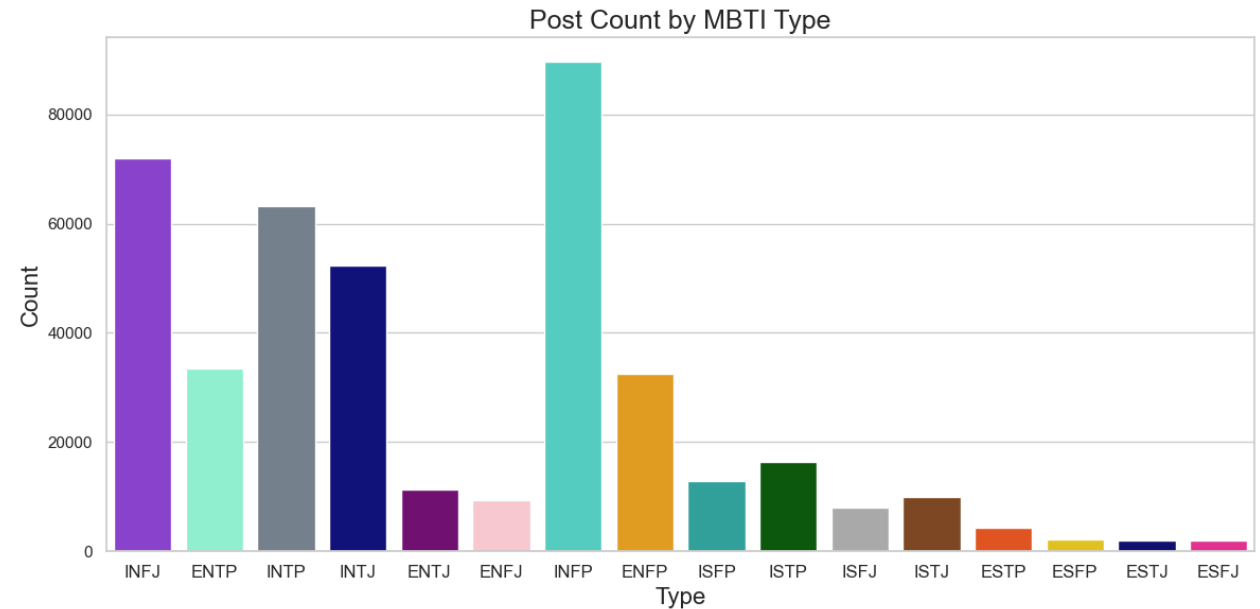
URL Analysis

Tracking MBTI Mentions

Emotion Probability Vector

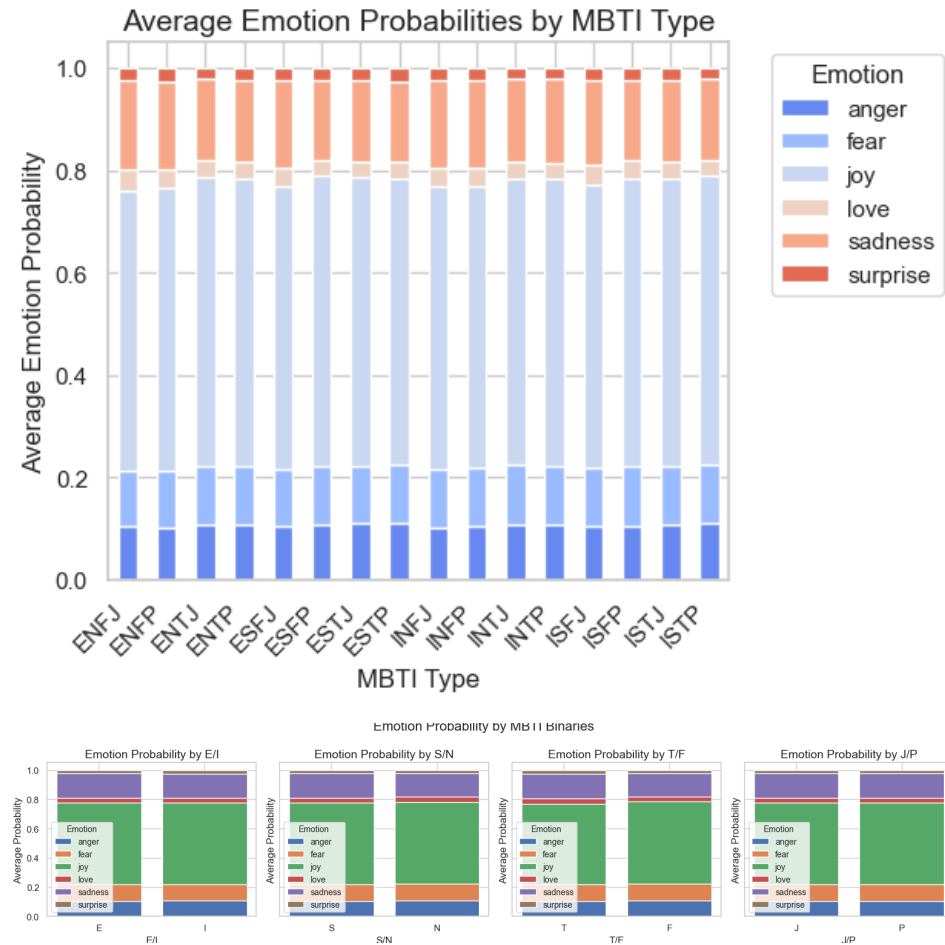
Exploratory Data Analysis

- post distribution is highly uneven
- requires us to draw stratified random samples across our targets for later



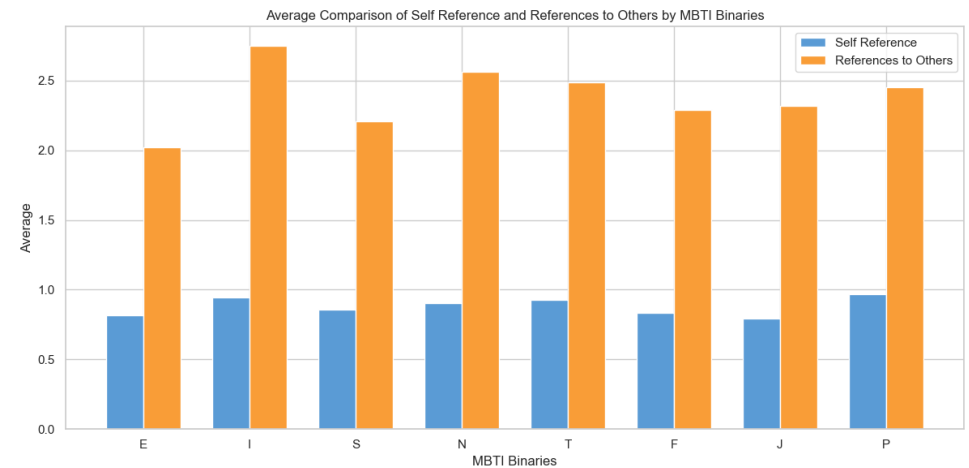
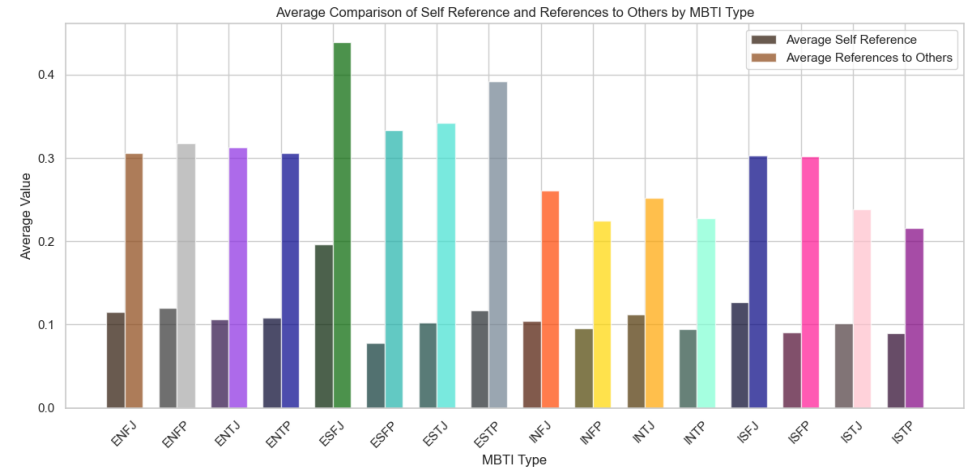
Exploratory Data Analysis (cont.)

- surprising uniformity in emotional expression
- **weakens predictive power: MBTI types may not significantly differ in their emotional articulation in text**



Exploratory Data Analysis (cont.)

- minor variations, especially with references to others
- **strengthens predictive power: non-negligible difference across all targets**



Analysis

1 Model Creation

Supervised Models

Logistic Regression
RandomForestClassifier
LinearSVC

2 Model Optimization

Hyperparameter Optimization

Grid Search

3 Model Evaluation

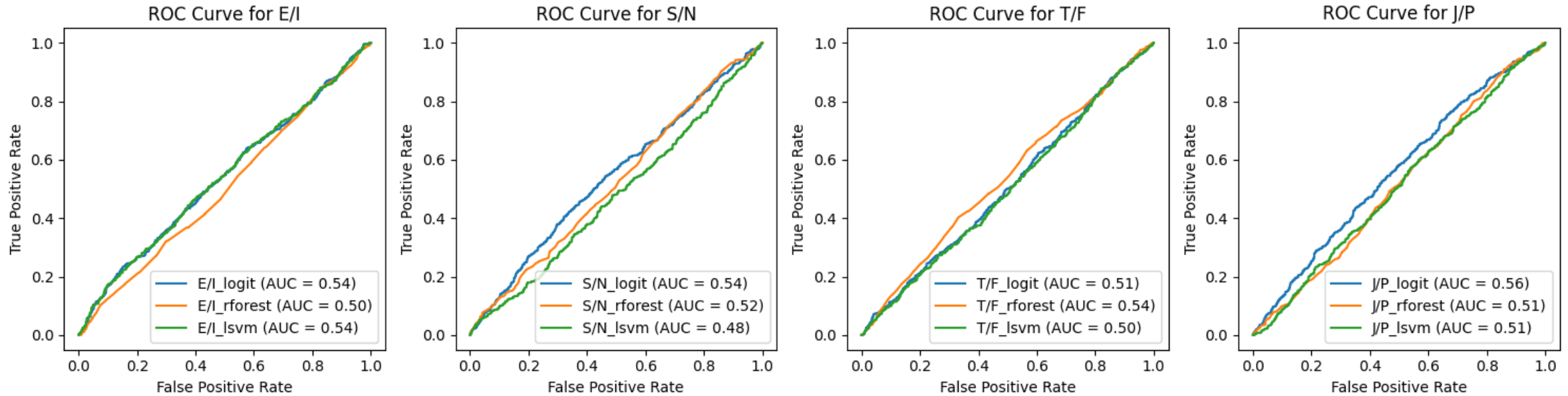
Model Evaluation Techniques

ROC Curve
Accuracy Score

MBTI Binary Classification Models

Binary Target	Best Model	Accuracy Score
0 (E/I)	E/I_lsvm	0.7
1 (S/N)	S/N_logit	0.5
2 (T/F)	T/F_rforest	0.3
3 (J/P)	J/P_logit	0.4

MBTI Binary Classification Models (cont.)



MBTI Type Model

Classification Report

	precision	recall	f1-score	support
ENFJ	0.04	0.1	0.06	0.06
ENFP	0.03	0.01	0.01	0.01
ENTJ	0.06	0.05	0.06	0.06
ENTP	0.06	0.02	0.03	0.03
ESFJ	0.1	0.15	0.12	0.12
ESFP	0.12	0.2	0.15	0.15
ESTJ	0.11	0.12	0.12	0.12
ESTP	0.08	0.15	0.11	0.11
INFJ	0.07	0.27	0.11	0.11
INFP	0.09	0.03	0.04	0.04
INTJ	0	0	0	0
INTP	0.09	0.02	0.04	0.04
ISFJ	0	0	0	0
ISFP	0.16	0.08	0.11	0.11
ISTJ	0.2	0.01	0.02	0.02
ISTP	0.06	0.12	0.08	0.08
accuracy	0.08	1.616		
macro avg	0.08	0.08	0.07	
weighted avg	0.08	0.08	0.06	

BEST MULTICLASS MODEL FOR 'TYPE' BASED ON ACCURACY:

MODEL NAME: TYPE_LOGIT

ACCURACY: 0.08

Conclusion

certain features have greater fluctuation and reveal greater insight

models demonstrated some capability of predicting MBTI binaries more than MBTI types

but for now, MBTI cannot be accurately and strongly predicted by one's post online in a forum