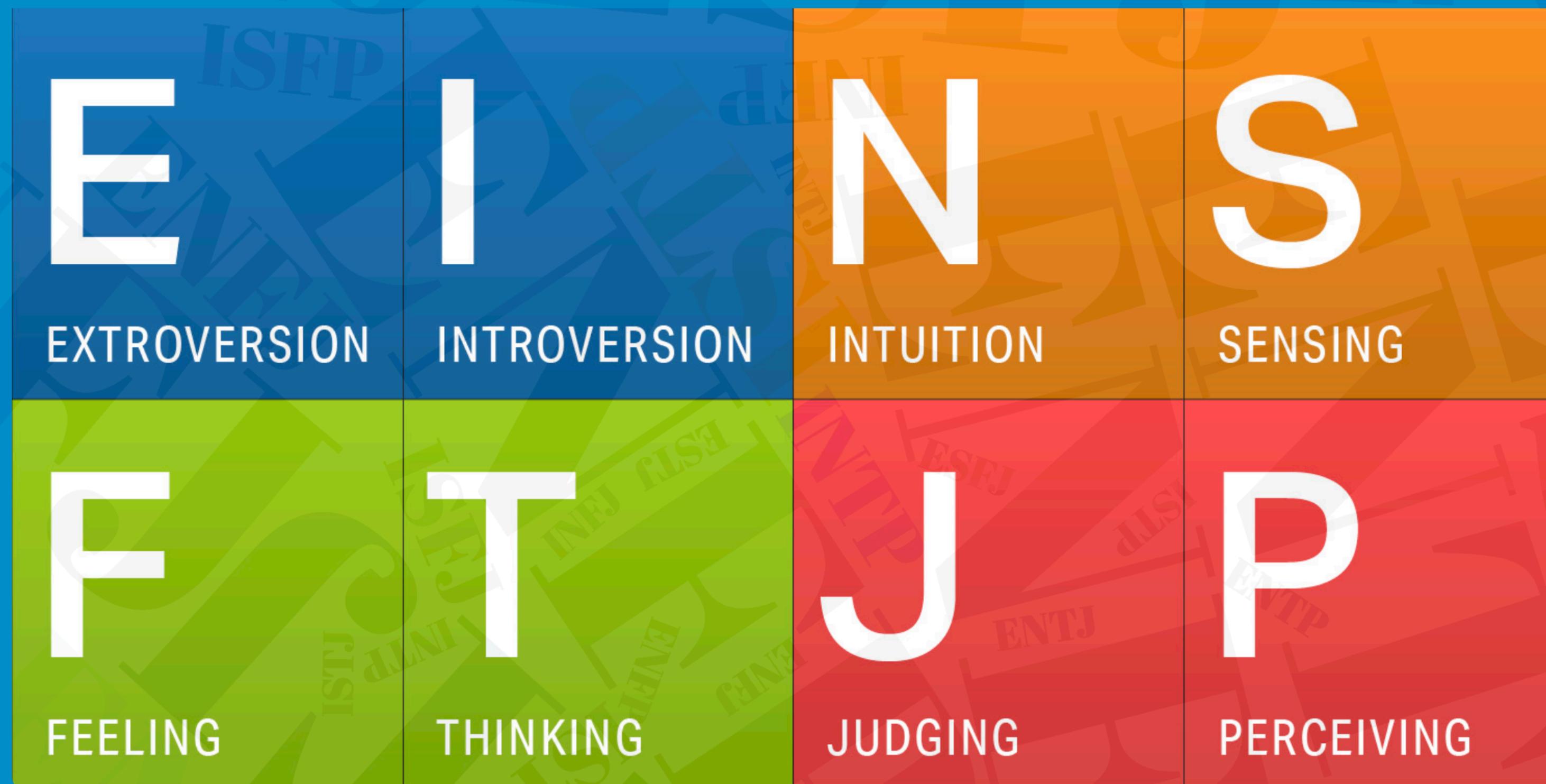
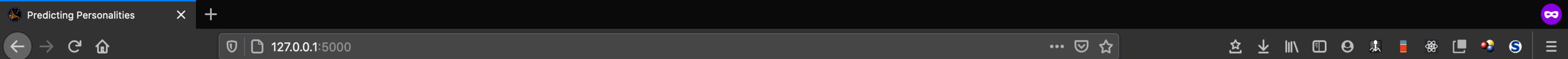




What is the Myers-Briggs Type Indicator?





127.0.0.1:5000

Getting Started Most Visited Monsterblader.github...

What's missing? Identifying the personalities.



ENFP: Barney

Type: ENFP
Characterization: The Advocate
Personality:
ENFPs are introspective, values-oriented, inspiring, social and extremely expressive. They actively send their thoughts and ideas out into the world as a way to bring attention to what they feel to be important, which often has to do with ethics and current events. ENFPs are natural advocates, attracting people to themselves and their cause with excellent people skills, warmth, energy and positivity. ENFPs are described as creative, resourceful, assertive, spontaneous, life-loving, charismatic, passionate and experimental.

ISTP: Sideshow Bob

Type: ISTP
Characterization: The Craftsmen
Personality:
ISTPs are adventurous and independent. They like to figure out how things work. They have great mechanical and technical skills. They live in the "here and now" which makes them adaptable and spontaneous. They like to keep busy and are action-oriented. ISTPs thrive on new and exciting situations.

INFP: Ralph Wiggum

Type: INFP
Characterization: The Dreamer
Personality:
INFPs are introspective, private, creative and highly idealistic individuals that have a constant desire to be on a meaningful path. They are driven by their values and seek peace. Empathetic and compassionate, they want to help others and humanity as a whole. INFPs are imaginative, artistic and often have a talent for language and writing. They can also be described as easygoing, selfless, guarded, adaptable, patient and loyal.

A Tool to Predict One's Myers-Briggs Type



Hi, everyone! I'm a San Francisco native who attended Caltech in Pasadena and has spent time all over the country. My favorite cities are San Francisco, Boston, Raleigh, and Denver. I am a bootcamp veteran, having acquired a skill set in web development, and where I, amazingly, met Josh Shaman who now works for Metis. I bike, play piano, and dance in my spare time.



I N T P



Predicting Personalities × +

127.0.0.1:5000

Getting Started Most Visited Monsterblader.github...

Finding an MBTI dataset

The screenshot shows a web browser window with a search result for an MBTI dataset on Kaggle. The title bar says "Predicting Personalities". The address bar shows "127.0.0.1:5000". The search results page has a header with "the place to discover yourself" and "personality cafe". The main content area shows a "Myers Briggs Forum" section with a "Follow Forum" and "Start Discussion" button. On the left, there's a sidebar for "kaggle" with links to Home, Compete, Data, Notebooks, Discuss, and Courses. The main content area displays a dataset titled "(MBTI) Myers-Briggs Personality Type Dataset" by Mitchell J, updated 3 years ago (Version 1). It includes a preview image of a brain composed of gears and text stating "Includes a large number of people's MBTI type and content written by them". Below the preview are buttons for "Data", "Tasks", "Notebooks (50)", "Discussion (4)", "Activity", "Metadata", "Download (60 MB)", and "New Notebook". A horizontal ellipsis at the bottom indicates more results.

Predicting Personalities

Getting Started Most Visited Monsterblader.github...

127.0.0.1:5000

Natural Language Processing



Text Classification
POS Identification
Entity Extraction

Text Summarization
Text Generation
Sentiment Analysis

spaCy MENU Search docs

Industrial-Strength Natural Language Processing IN PYTHON

NLTK 3.5

Natural Language Toolkit

.....

Working with spaCy

Classifier

- LinearSVC
- SVC
- Multinomial Naive Bayes
- Decision Tree
- Random Forest

Accuracy

0.27
not enough computer
0.27
0.20
0.26

Processing the data for NLTK

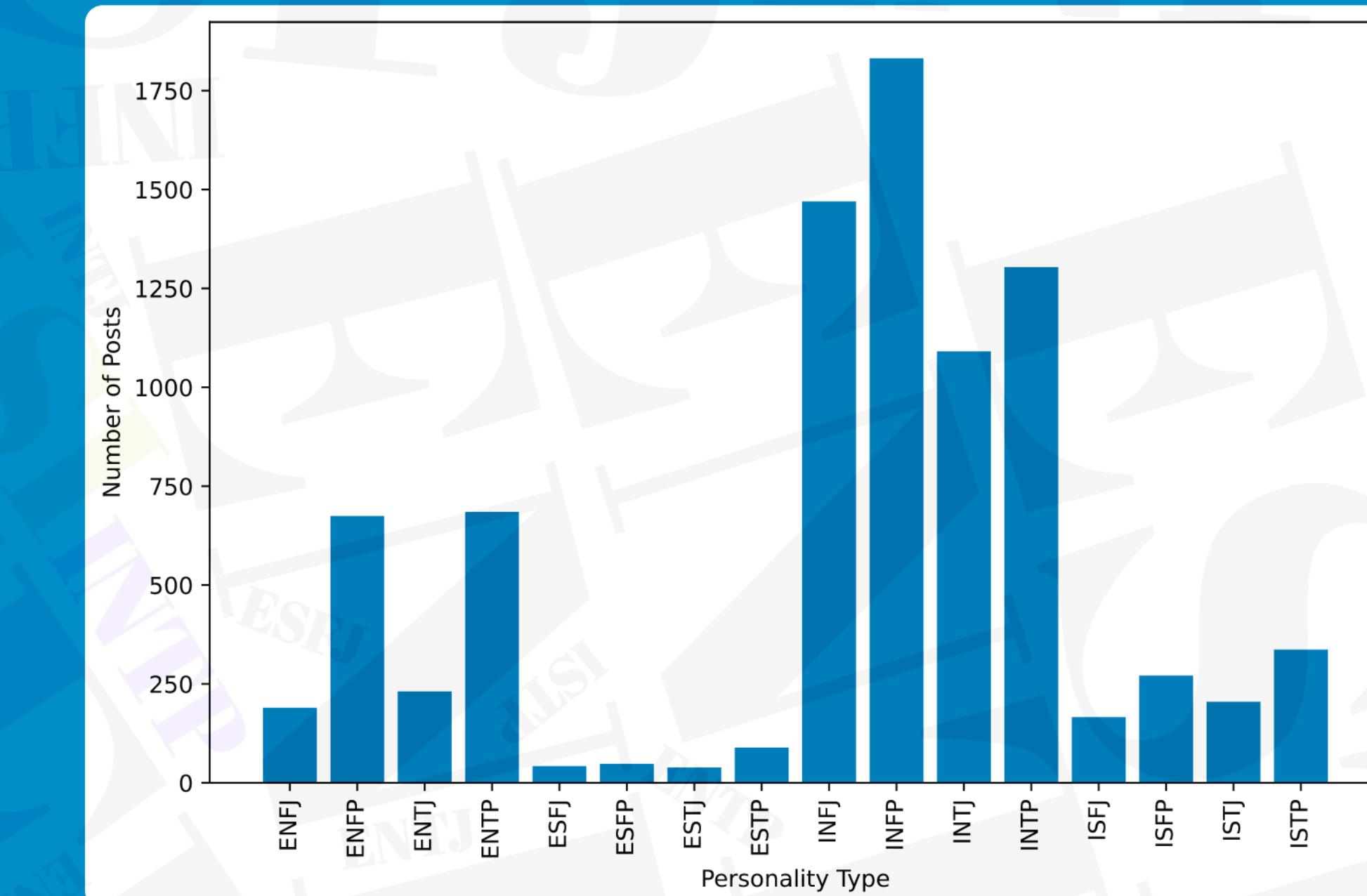
1. Remove special characters
 2. Remove single-letter words
 3. Convert to lower case
 4. Lemmatize
 5. Stop words
 6. Term frequency-inverse document frequency

Create a baseline.

```
In [21]: print(classification_report(y_test,y_pred))
print(accuracy_score(y_test, y_pred))
```

	precision	recall	f1-score	support
ENFJ	0.00	0.00	0.00	26
ENFP	0.91	0.10	0.18	102
ENTJ	1.00	0.03	0.06	31
ENTP	0.65	0.17	0.26	103
ESFJ	0.00	0.00	0.00	6
ESFP	0.00	0.00	0.00	9
ESTJ	0.00	0.00	0.00	3
ESTP	0.00	0.00	0.00	18
INFJ	0.40	0.55	0.46	207
INFP	0.36	0.85	0.51	281
INTJ	0.63	0.33	0.43	173
INTP	0.50	0.57	0.53	195
ISFJ	0.00	0.00	0.00	19
ISFP	0.00	0.00	0.00	38
ISTJ	0.00	0.00	0.00	39
ISTP	1.00	0.02	0.04	52
accuracy			0.42	1302
macro avg	0.34	0.16	0.15	1302
weighted avg	0.49	0.42	0.36	1302

0.42242703533026116



Predicting Personalities X +

Getting Started Most Visited Monsterblader.github...

127.0.0.1:5000

Undersample or Oversample?

Sets the maximum number of data points per class to 500

```
In [20]: print(classification_report(y_test,y_pred))
print(accuracy_score(y_test, y_pred))
```

	precision	recall	f1-score	support
ENFJ	0.00	0.00	0.00	36
ENFP	0.28	0.77	0.41	98
ENTJ	1.00	0.04	0.07	26
ENTP	0.34	0.77	0.47	117
ESFJ	0.00	0.00	0.00	3
ESFP	0.00	0.00	0.00	6
ESTJ	0.00	0.00	0.00	7
ESTP	0.00	0.00	0.00	12
INFJ	0.44	0.26	0.32	66
INFP	0.34	0.36	0.35	84
INTJ	0.64	0.30	0.41	70
INTP	0.61	0.35	0.44	78
ISFJ	0.00	0.00	0.00	28
ISFP	0.00	0.00	0.00	36
ISTJ	0.00	0.00	0.00	22
ISTP	0.91	0.17	0.29	58
accuracy			0.36	747
macro avg	0.29	0.19	0.17	747
weighted avg	0.40	0.36	0.31	747

0.3627844712182062

Predicting Personalities

Getting Started Most Visited Monsterblader.github...

In [28]:

```
print(classification_report(y_test,y_pred))
print(accuracy_score(y_test, y_pred))
```

	precision	recall	f1-score	support
ENFJ	1.00	0.06	0.12	81
ENFP	0.67	0.10	0.17	105
ENTJ	0.84	0.26	0.40	81
ENTP	0.64	0.15	0.24	95
ESFJ	0.84	0.66	0.74	73
ESFP	0.67	0.45	0.54	78
ESTJ	1.00	0.62	0.76	68
ESTP	0.93	0.53	0.68	75
INFJ	0.39	0.52	0.44	233
INFP	0.26	0.87	0.40	268
INTJ	0.58	0.33	0.42	155
INTP	0.47	0.55	0.51	194
ISFJ	1.00	0.24	0.39	67
ISFP	1.00	0.06	0.11	88
ISTJ	0.90	0.12	0.21	76
ISTP	0.89	0.11	0.20	72
accuracy			0.42	1809
macro avg	0.76	0.35	0.39	1809
weighted avg	0.64	0.42	0.40	1809

0.4223327805417358

In [40]:

```
print(classification_report(y_test,y_pred))
print(accuracy_score(y_test, y_pred))
```

	precision	recall	f1-score	support
ENFJ	0.82	0.58	0.68	130
ENFP	1.00	0.07	0.12	106
ENTJ	0.62	0.72	0.67	138
ENTP	0.75	0.11	0.20	105
ESFJ	0.82	0.92	0.87	154
ESFP	0.78	0.87	0.82	153
ESTJ	0.99	0.83	0.90	151
ESTP	0.90	0.72	0.80	156
INFJ	0.47	0.49	0.48	230
INFP	0.29	0.87	0.43	269
INTJ	0.76	0.34	0.47	170
INTP	0.57	0.58	0.58	212
ISFJ	0.96	0.56	0.71	161
ISFP	0.74	0.44	0.55	147
ISTJ	0.88	0.53	0.66	122
ISTP	0.87	0.57	0.69	155
accuracy			0.60	2559
macro avg	0.76	0.58	0.60	2559
weighted avg	0.73	0.60	0.60	2559

0.602188354826104

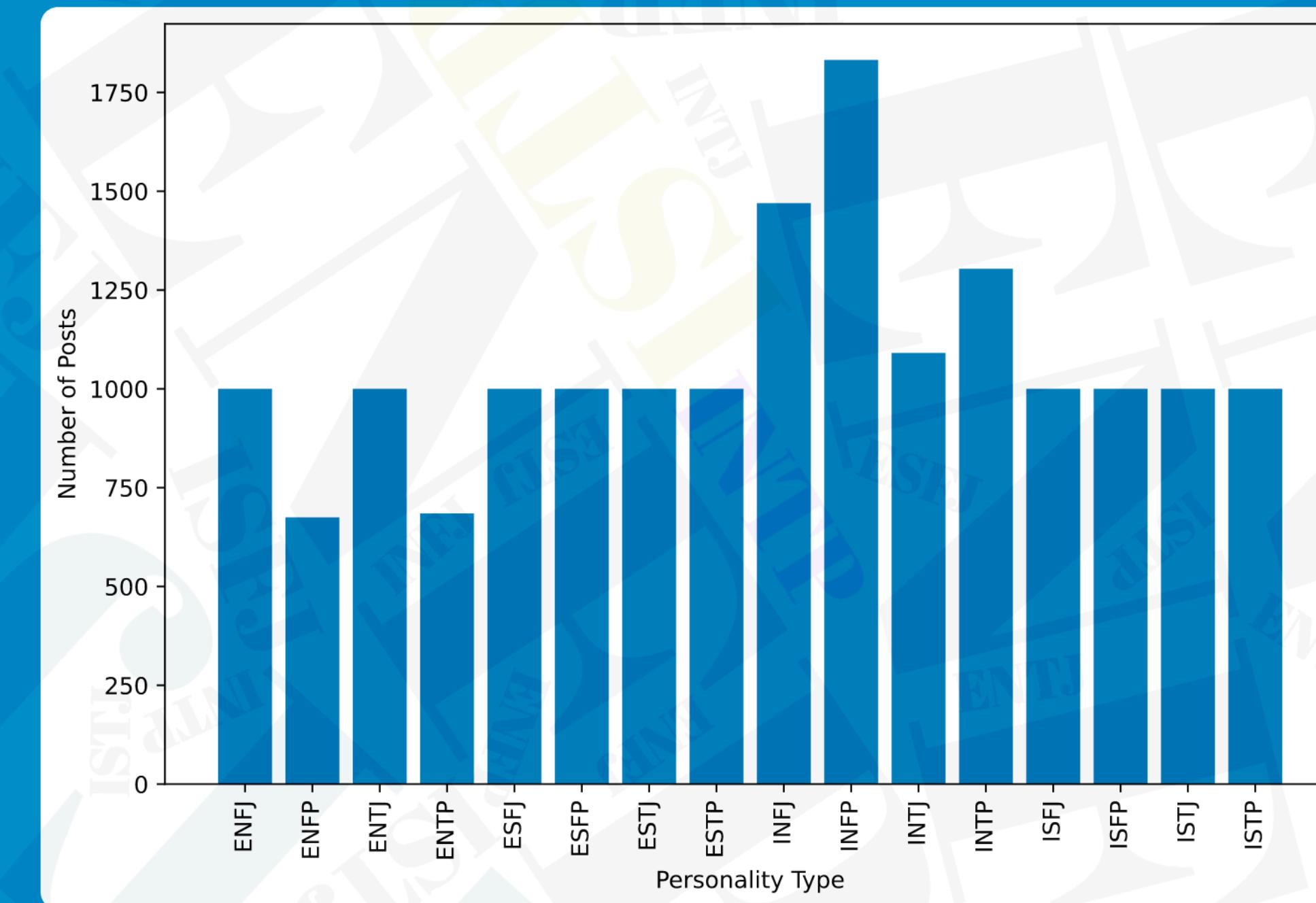
In [13]:

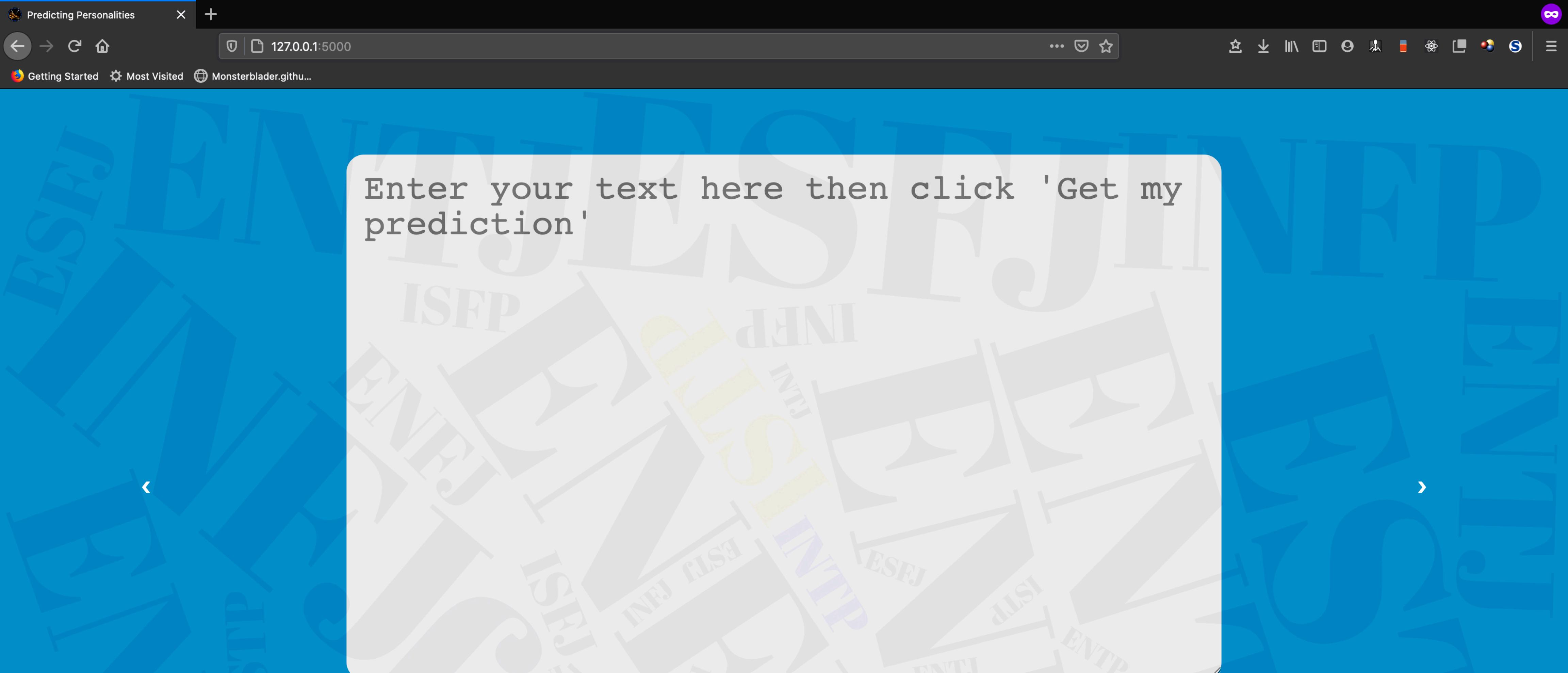
```
print(classification_report(y_test,y_pred))
print(accuracy_score(y_test, y_pred))
```

	precision	recall	f1-score	support
ENFJ	0.68	0.75	0.71	224
ENFP	1.00	0.05	0.10	113
ENTJ	0.61	0.79	0.69	228
ENTP	0.62	0.05	0.09	103
ESFJ	0.90	0.95	0.93	229
ESFP	0.76	1.00	0.86	215
ESTJ	1.00	0.93	0.96	231
ESTP	0.84	0.84	0.84	238
INFJ	0.60	0.39	0.47	212
INFP	0.37	0.81	0.51	286
INTJ	0.80	0.25	0.38	174
INTP	0.65	0.51	0.57	204
ISFJ	0.88	0.76	0.81	229
ISFP	0.68	0.64	0.66	224
ISTJ	0.77	0.74	0.76	190
ISTP	0.68	0.71	0.69	209
accuracy			0.69	3309
macro avg	0.74	0.63	0.63	3309
weighted avg	0.73	0.69	0.67	3309

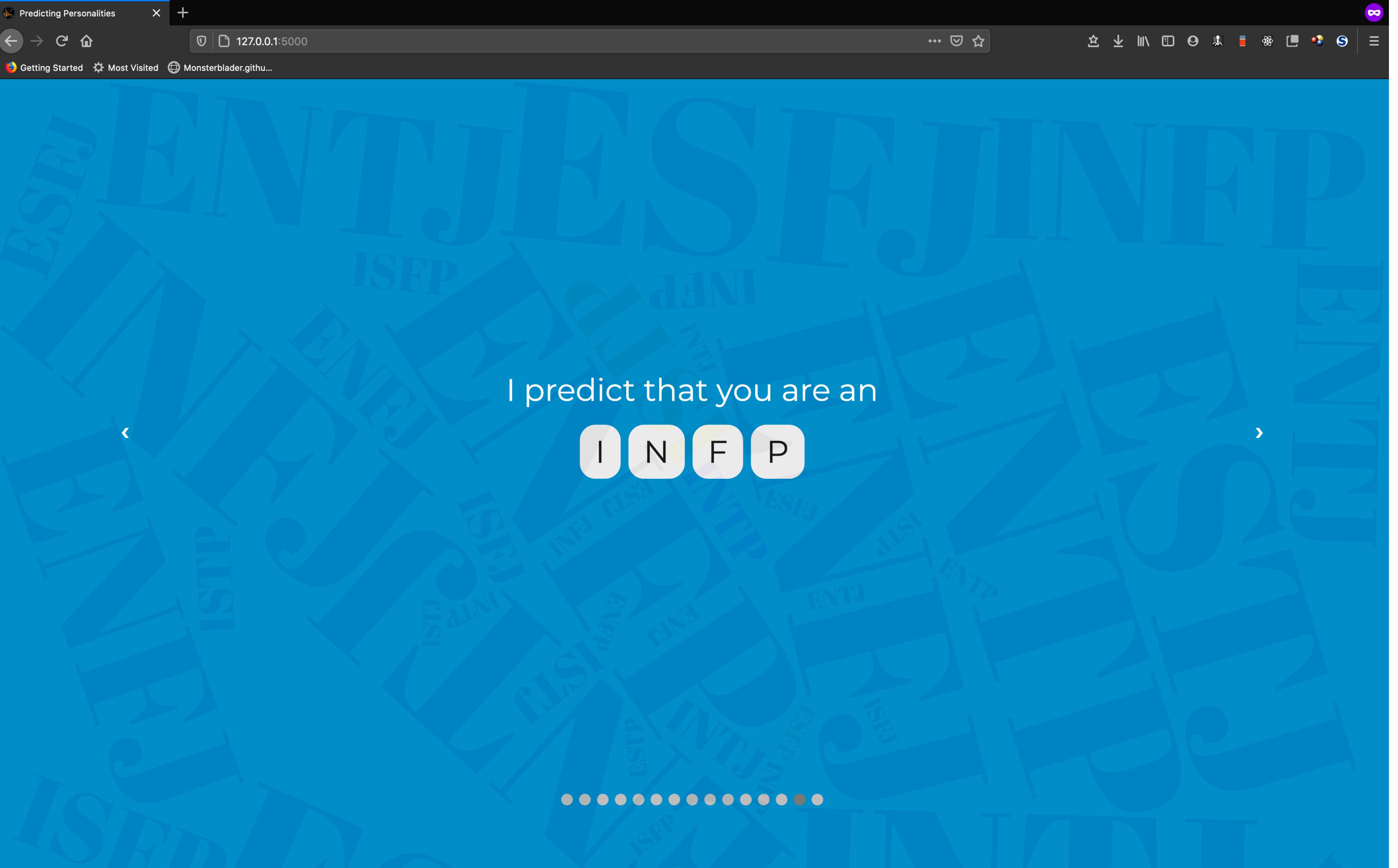
0.6863100634632819

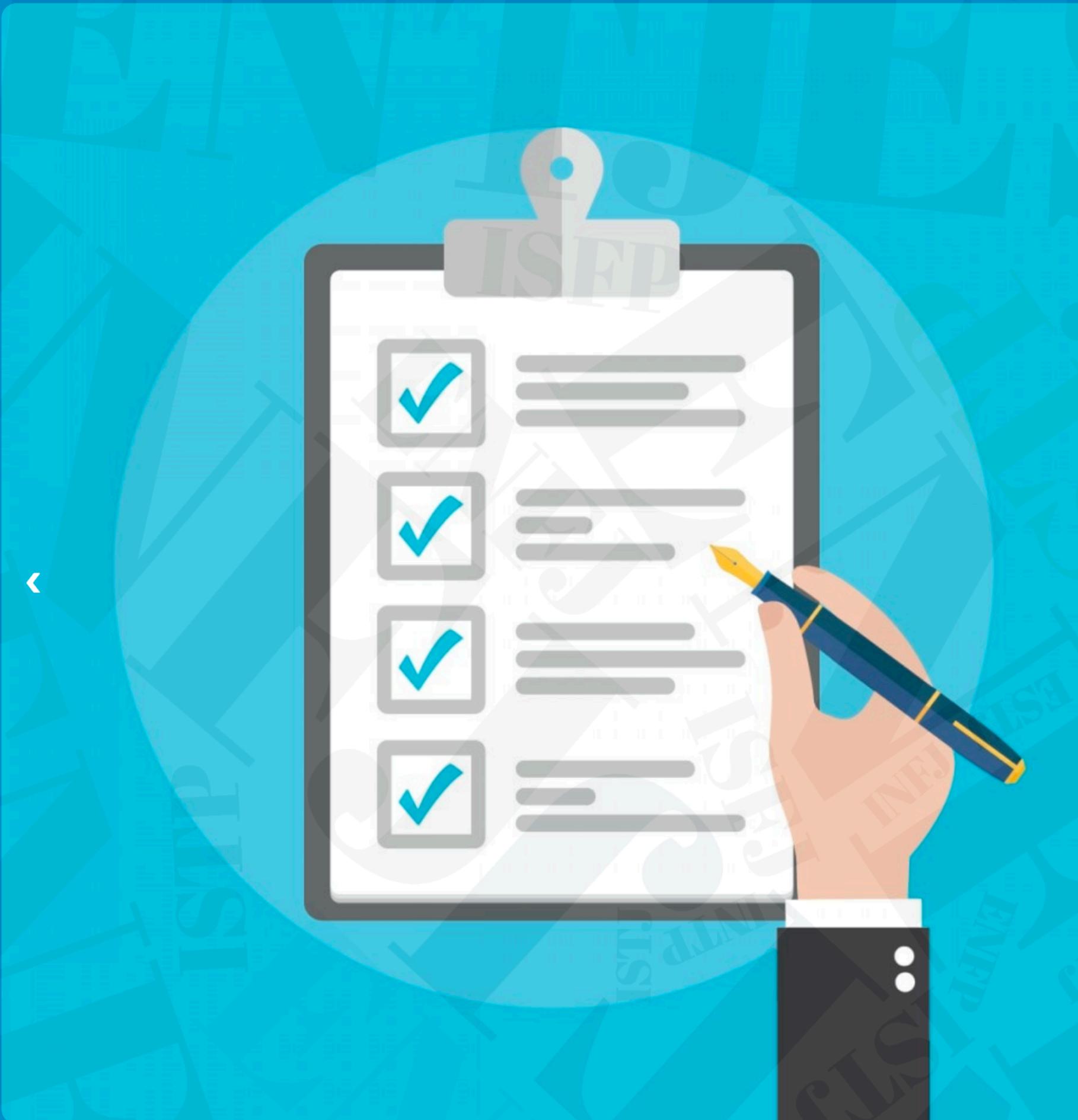
New distribution plot for final settings





Get my prediction!





Take aways:

- Text processing models work better with larger bodies of text.
 - Introverts either like to post more than extroverts, they are more curious about themselves, or they would rather connect with others anonymously.

Next steps:

- Topic modeling
 - spaCy