

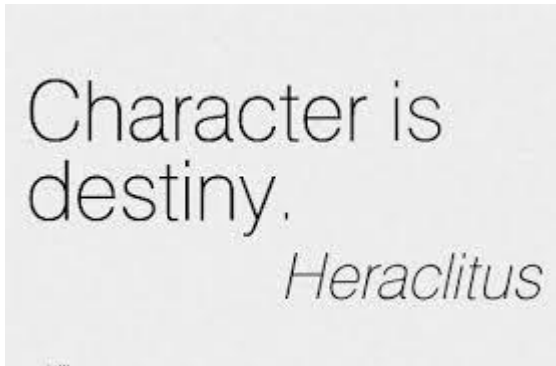
MBTI Personality Prediction

Determine your type by one simple click

Skye Sheng

June 2020

Motivation



Long and annoying process

- 15 + minutes
- 100 + questions



Simple and fast way

- One user input (social media post)
- One click
- How they say not what they say

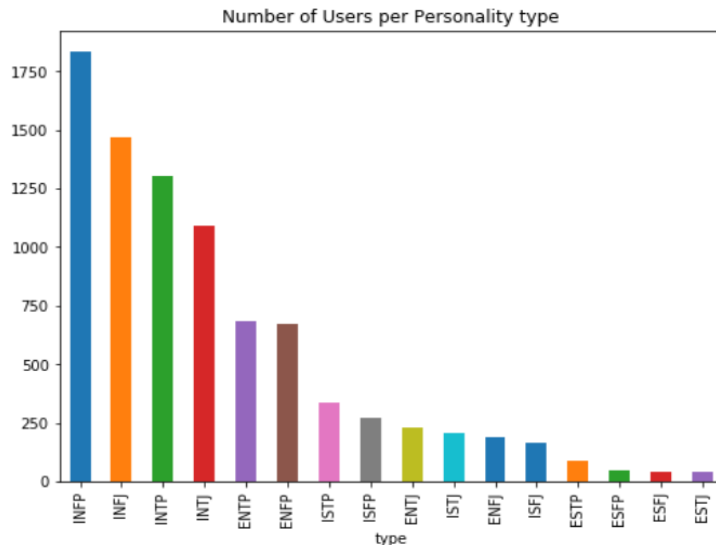
Demo

DATA

- Data Source: <https://www.kaggle.com/datasnaek/mbti-type>
- Over 8600 rows of data
- Row: the most recent 50 posts a user has posted on PersonalityCafe forum
- Features generated (per post): number of videos, music, images, words, question, exclamation, ellipsis

RDS used to

- Query and display data entry in the database
- Store new user input and corresponding predictions



Type	Raw post
INFJ	<code>http://www.youtube.com/watch?v=IRcrv41hgZ4 Two baby deer on left and right munching on a beetle in the middle! Using their own blood, two cavemen diary today's latest happenings on their designated cave diary wall. I see it as... </code>

Algorithm

Modeling

- Build 4 binary classifiers
 - Introverted vs Extroverted
 - Intuition vs sensing
 - Feeling vs thinking
 - Perception vs judgment
- Split data: 60-40 train test split
- Models: logistic regression, random forest, decision tree

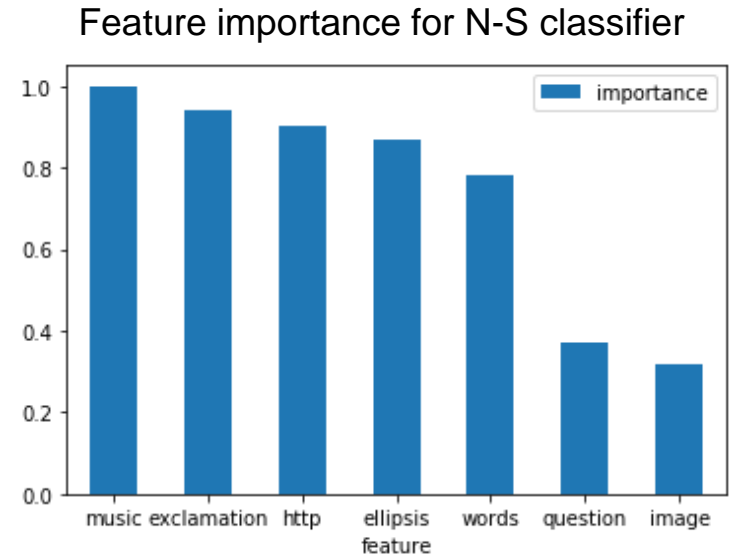
Evaluation

- Logistic regression stands out for all classifiers
- Success: accuracy above 0.7
- Good at distinguishing intuition and sensing

Model	Classifier	Accuracy
Logistic regression	I-E	0.78
	N-S	0.92
	F-T	0.68
	P-J	0.70

Insights

- Music, exclamation marks, http, ellipsis and words per post are identified as important factor in distinguishing intuition and sensing
- An intuition inclined person tends to share music, refer to a web link, use exclamation, ellipsis more often than a sensing-dominated person.
- Training data were from a specific forum, so it might not generalize well to posts on other social platform
- Consider other feature engineering: NLP



Thank You!

Developer: Skye Sheng

Email: skyesheng2020@u.northwestern.edu

QA: Manish Kumar