

EVALUATING THE MBTI PERSONALITY CONSTRUCT USING TEXT DATA

BEN OLSEN

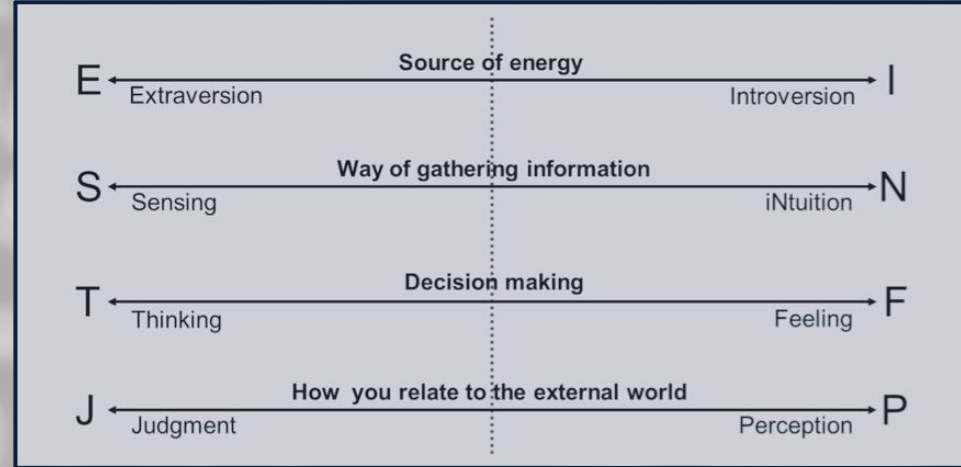
PURPOSE

**To determine the usefulness
of the MBTI as a construct of
personality using text data.**

MYERS-BRIGGS TYPE INDICATOR (MBTI)

BACKGROUND

"...introspective self-report questionnaire with the purpose of indicating different psychological preferences in how people perceive the world around them and make decisions"



MBTI RELIABILITY

BACKGROUND

Poor Retest:

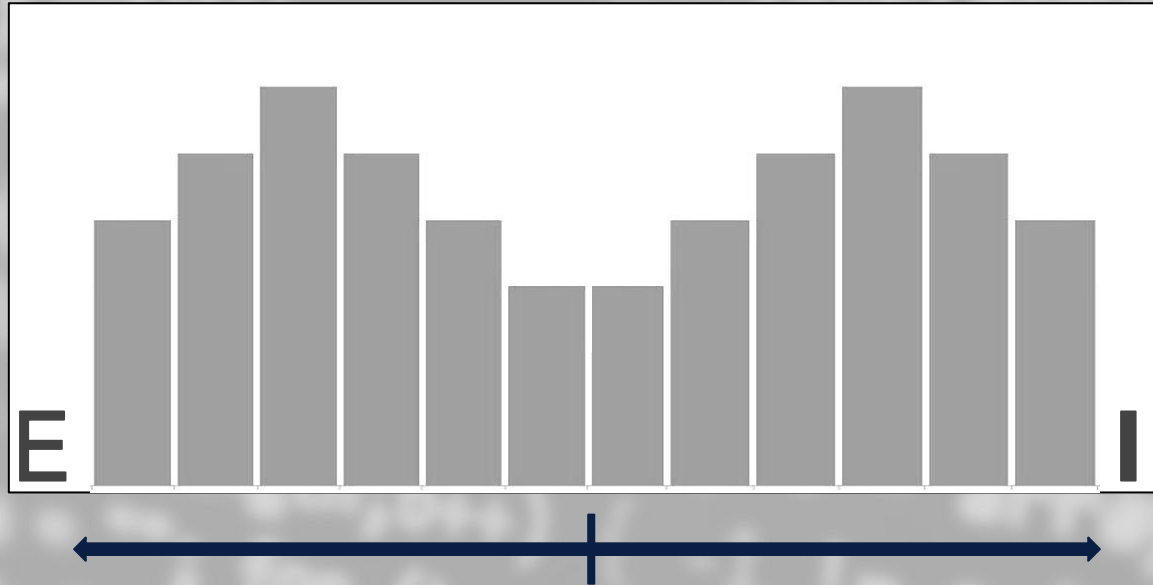
Normal distribution of test scores



MBTI RELIABILITY

BACKGROUND

Bimodal distribution

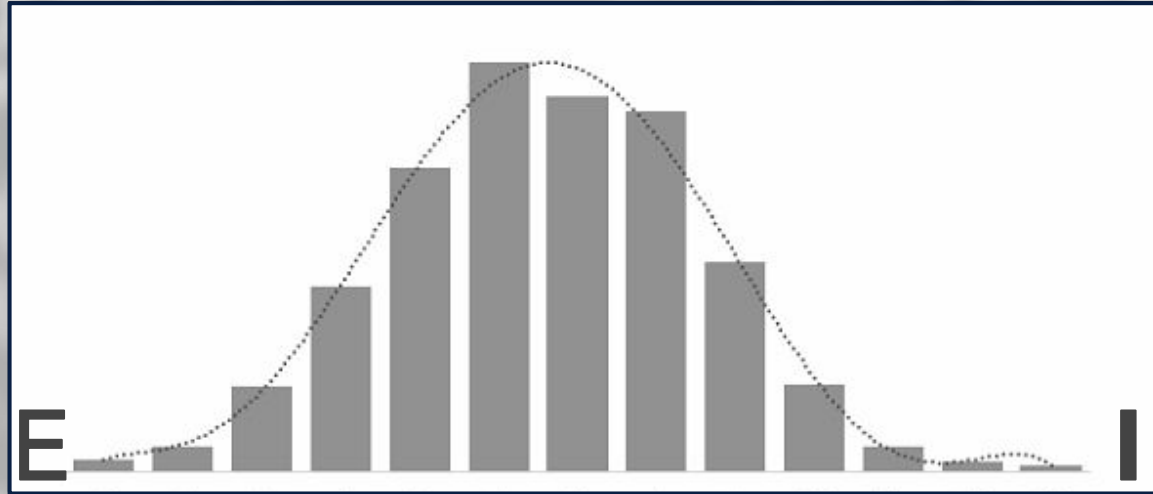


Two groups in population = good

MBTI RELIABILITY

BACKGROUND

Normal distribution



normal population = bad

HYPOTHESIS

Text can be categorized by the MBTI personality type of the person writing it (Bimodal Distribution).

DATA

MATERIALS

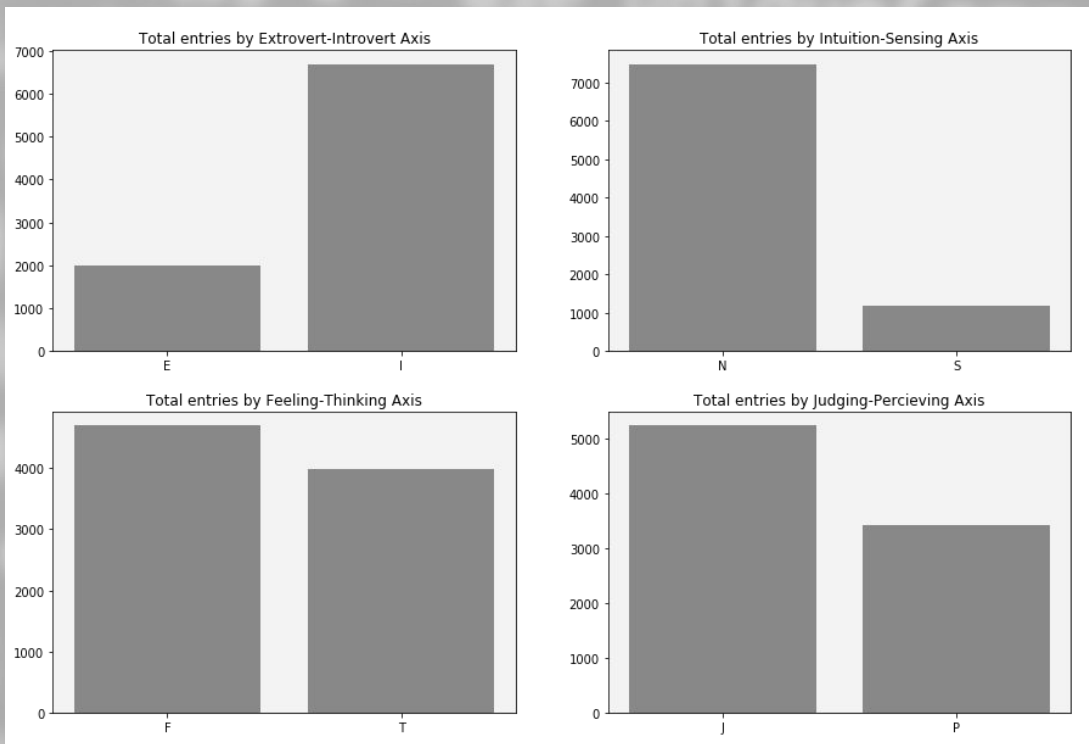
- **PersonalityCafe**
- **8600 rows**
 - **Type (4 letters).**
 - **The last 50 things posted**

	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...

DATA

MATERIALS

Regroup data by 4 axis



LIBRARIES

MATERIALS

Wrangling

- **pandas**
- **nltk**
- **numpy**
- **html**
- **re**

Analysis

- **matplotlib**
- **scipy**
- **statsmodels**
- **wordcloud**

ML

- **SGDClassifier**

SAMPLES

EXPERIMENTAL DESIGN

Sample 2: “Short-Fat”

- Not split
- More features,
less entries
(8,600)

Sample 3 “No-Names”

- Class references
removed
- Simulate unbiased
data

DATA WRANGLING

EXPERIMENTAL DESIGN

- Remove Escaping HTML Characters
- Remove Hyperlinks
- Expand Contractions
- Remove Digits
- Remove Punctuation
- Remove Stopwords

MACHINE LEARNING

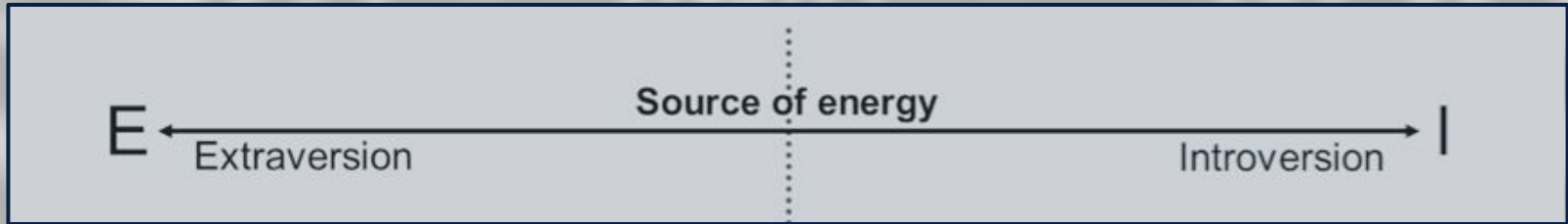
EXPERIMENTAL DESIGN

- **Feature Extraction - Bag of words**
 - Text vectorizing
 - Transforming by Inverse document frequency
- **SGDClassifier - SVM with stochastic gradient descent**
 - One classifier per axis

SIMULATING THE MBTI SCORING METHOD

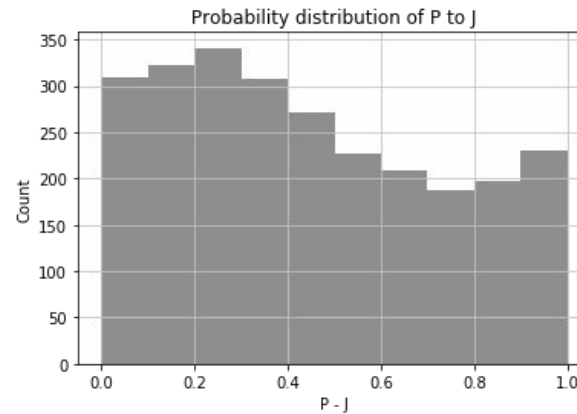
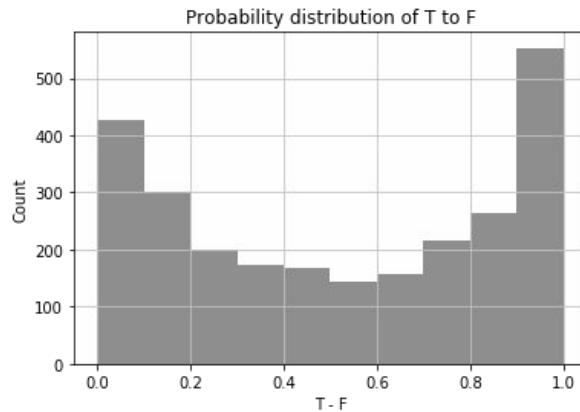
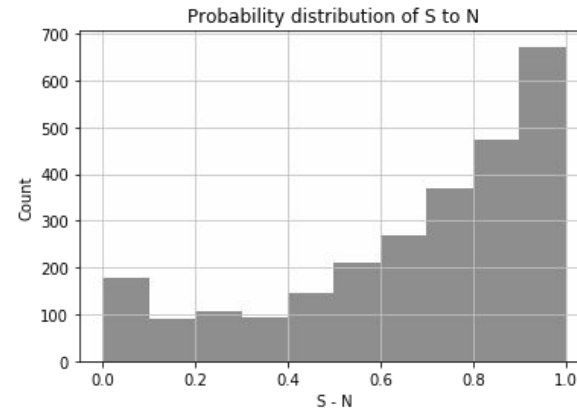
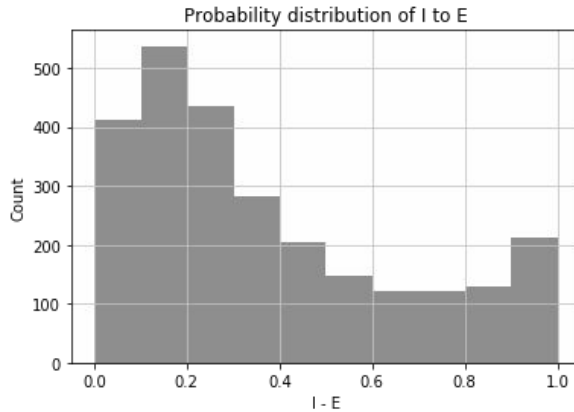
EXPERIMENTAL DESIGN

- Logarithmic loss function - probabilities
- Probability = Score
- Distribution of probabilities - bimodal vs normal



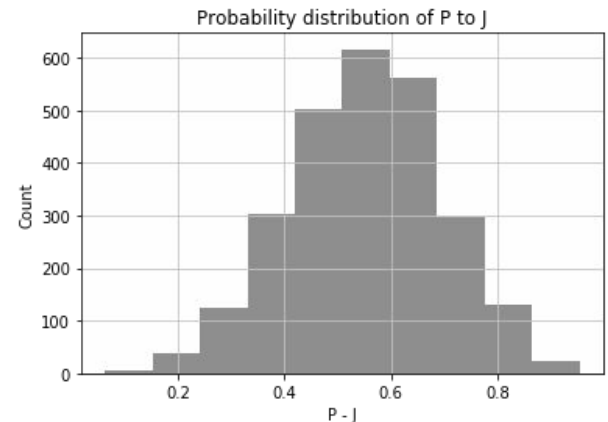
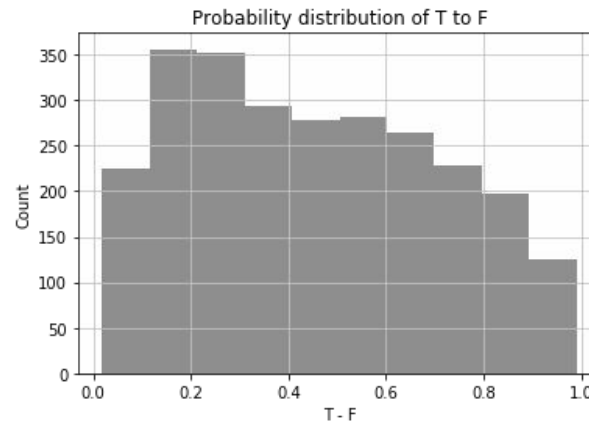
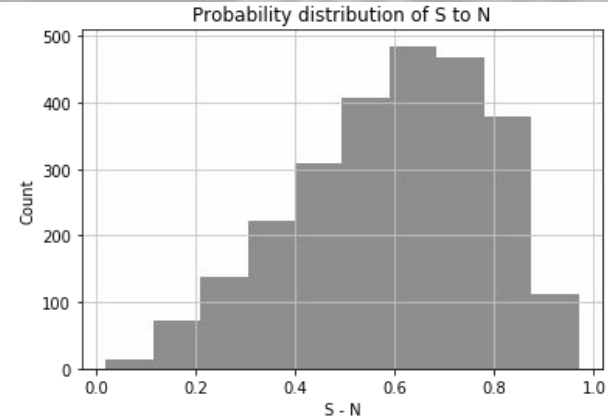
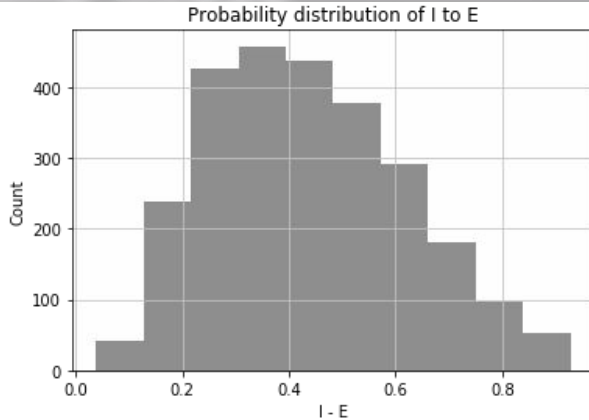
SAMPLE 2

RESULTS



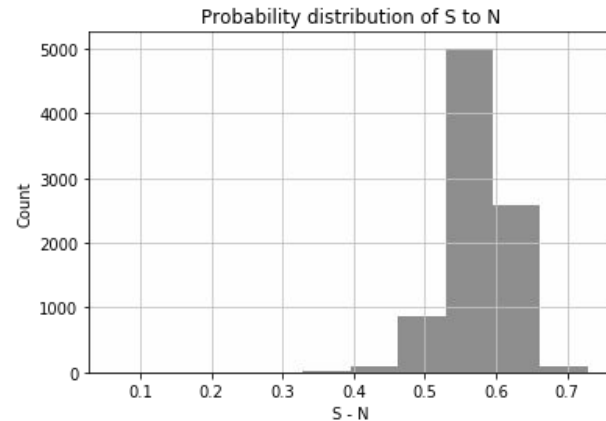
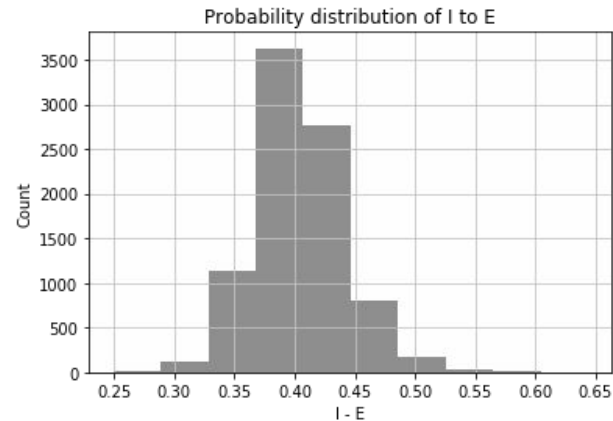
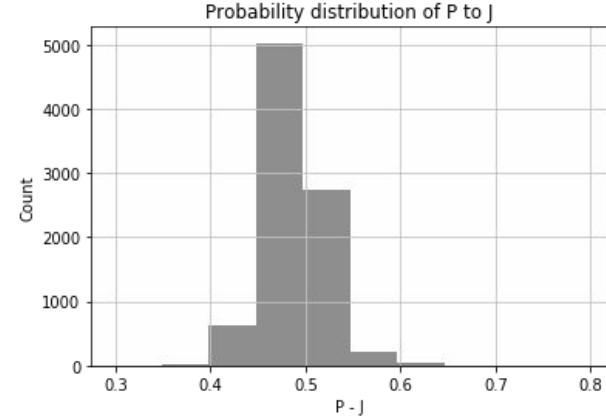
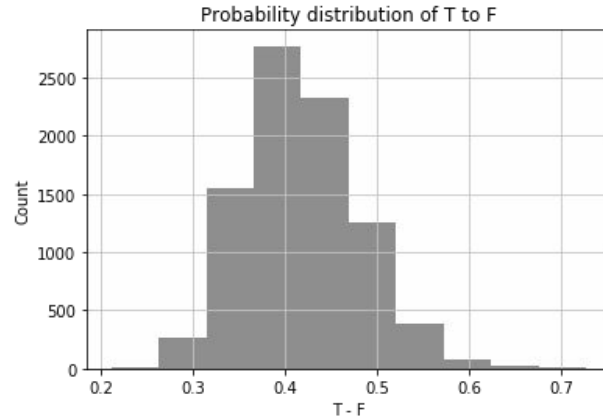
SAMPLE 3

RESULTS



SAMPLE 2 CLASSIFIER WITH SAMPLE 3

RESULTS



EVALUATING THE MBTI PERSONALITY CONSTRUCT USING TEXT DATA

- CONCLUSIONS**
- Implications
 - Applications

IMPLICATIONS

CONCLUSIONS

- **Accept Null Hypothesis**
- **Class names = strongest feature**
- **Belief influenced behavior**
- **Potential usefulness of text data**

APPLICATIONS

CONCLUSIONS

- **Targeted advertising - mildly useful**
 - Individuals who strongly identify with MBTI
- **Content delivery - very useful**
 - Scoring for audiences based on language content
- **Applicant screening - not useful**
 - Social media screening
 - Resume screening - potential