Northeastern UNIVERSITY

Personalized Recommendations Based on MBTI Classification

Yenting Kuang
Siyuan Gao
Keerthana Balswamy

IE7500 Applied Natural Language Processing in Engineering
December 4, 2024

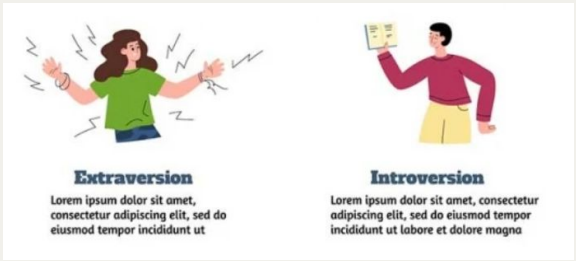
Northeastern UNIVERSITY

Catalogue:

(1) Introduction

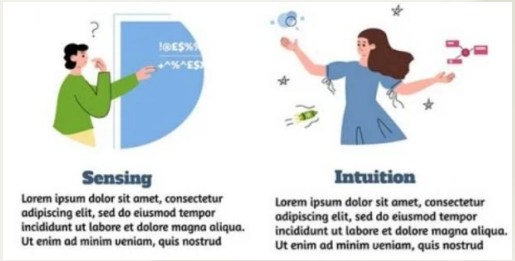
- **Background: Four dimensions of 16 MBTI personality types:**

Extraversion (E) vs. Introversion (I)



Energy Source

Sensing (S) vs. Intuition (N)



Information Processing

Feeling (F) vs. Thinking (T)

Decision-making

Judging (J) vs. Perceiving (P)

Lifestyle preference

- **Motivation**
  - Users perspective: Save users time in finding relevant products and improve user experience
  - Business perspective: Provide better guidance for product optimization and marketing strategies in anticipation of better business results in A/B testing of products.

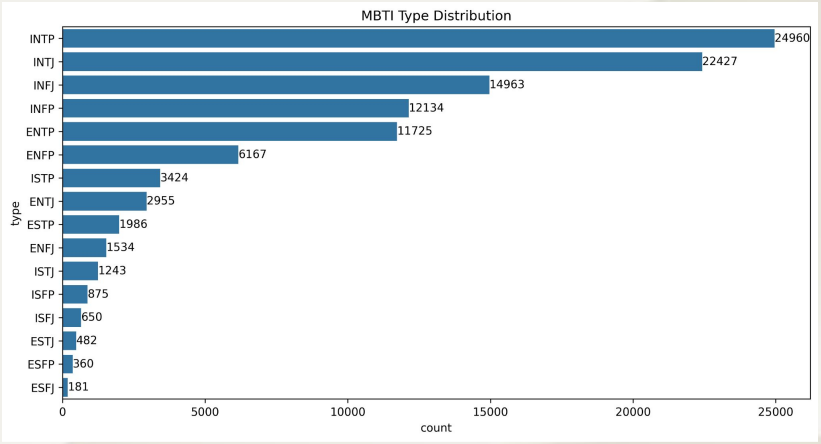
Image Source: https://www.esquirehk.com/money-investment/mbti-personality-test-16-personalities-career

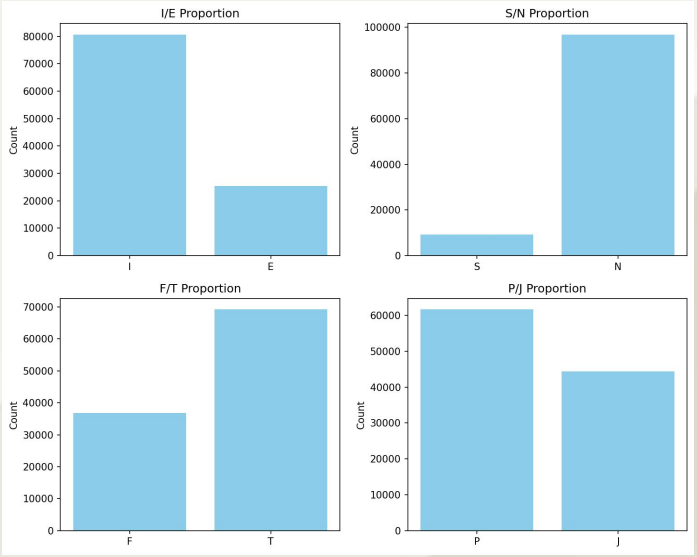
(2) Approach

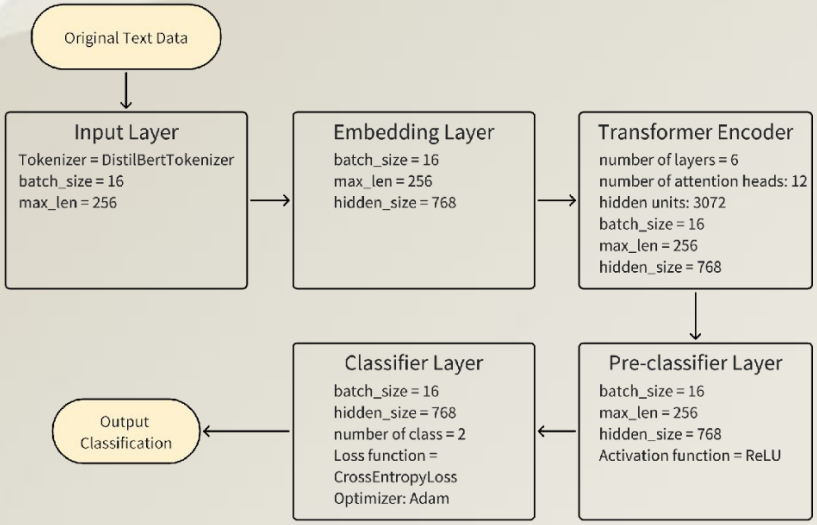
- Lemmatization
- Text vectorization
  - Word2Vec model
- Imbalanced Data Processing
  - SMOTE (Synthetic Minority Over-sampling Technique)
- Models
  - LSTM
  - GRU
  - BERT

(3) Results

- Corpus
  - “MBTI 500”: 106K+ rows of text posts
- Data preparation
- Exploratory Data Analysis (EDA)
- Models
  - LSTM (Bidirectional)
  - GRU (Bidirectional)
  - BERT







Northeastern UNIVERSITY

(3) Results and Discussion

Model	MBTI Type			
	I/E	S/N	F/T	J/P
LSTM	72.56%	77.75%	81.44%	70.63%
GRU	71.47%	78.03%	81.00%	70.16%
BERT	77.11%	90.96%	81.45%	75.64%

< 60%

- **Data Imbalance: Imbalanced data often leads models to overfit to the majority class**
  - SMOTE: LSTM and GRU
  - WeightedRandomSampler: BERT

- **P/J Classification: Relatively weak textual feature differences between P and J?**
  - Reduce learning rate: 5e-5 → 1e-5
  - Increase max_length: 256 → 512

(4) Conclusion

Model	MBTI Type			
	I/E	S/N	F/T	J/P
LSTM	72.56%	77.75%	81.44%	70.63%
GRU	71.47%	78.03%	81.00%	70.16%
BERT	77.11%	**90.96%**	81.45%	75.64%

- **Conclusion**

BERT demonstrated the best performance, achieving an accuracy of **90.96%** in the S/N dimension classification task. Across all four dimensions, the model achieved an average accuracy of **81.29%**.

- **Limitation and future work**

We only performed binary classification for each of the four MBTI dimensions instead of a comprehensive 16-type classification. If interdependencies exist among the four MBTI dimensions, a 16-type classification could yield more interpretable and meaningful results.