





Personality Prediction using MBTI Dataset

Name Harshit Patel
Ashish Srivastava
Divyanshu Chaudhary
Harsh

Personality Prediction Using the MBTI Dataset

Predict personality traits from text using machine learning and NLP techniques.

Leveraging the widely used Myers-Briggs Type Indicator for practical insights.





Introduction to MBTI

MBTI Overview

Self-report tool identifying psychological preferences across four dichotomies.

16 Personality Types
Examples include ISTJ,
ENFP, each representing
unique traits.

Applications

Used in career counseling, team building, development, and 80% of Fortune 500 companies.

The MBTI Dataset

Source

Kaggle dataset: 8,600+ users, 50 recent posts each, from PersonalityCafe forum.

Preprocessing

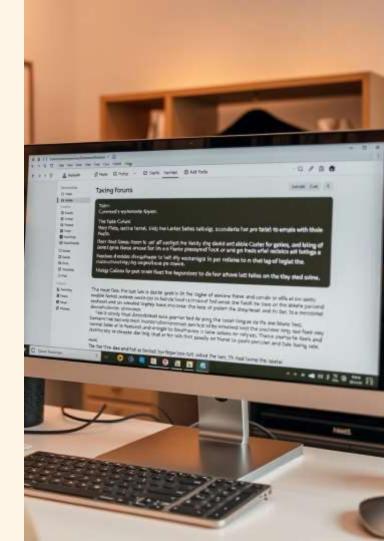
URLs and mentions removed for clean text suitable for analysis.

Imbalance

Uneven representation across types; INFP notably abundant.

Example

"I'm an INFP and I love rainy days..." demonstrating personal style.



Methodology: Feature Extraction

Text Preprocessing

Tokenization, stemming, and stop word removal to refine input.

Vectorization

Using TF-IDF and word embeddings for rich text representation.

Context & Sentiment

N-grams capture word sequences; sentiment analysis reveals emotional tone.

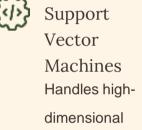
Methodology: Machine Learning Models



Naive Bayes

Fast and effective

baseline for text classification.



data with strong

accuracy.



Random

Forest

Reduces

overfitting using

ensemble

learning.



Neural Networks

Captures

complex

patterns via

deep learning.

Machine learning







Graded Network



Figet nedent forest



Districting Purchase







Neural Networ

Performance Metrics

Accuracy	Overall prediction correctness (e.g., 70%)
Precision	Avoids false positives (e.g., 80% for INFP)
Recall	Identifies all instances (e.g., 65% for ENTP)
F1-score	Balance of precision and
Confusion Matrix	recall Visualizes classification errors and overlaps



Results and Analysis

Model Performance

SVM model led with 72% accuracy among tested algorithms.

Feature Insights

Key words like "creative" help identify specific personality types.

Error Patterns

Mistakes often occur between similar types like INTJ and INFJ.

Addressing

Imbalance rechniques like oversampling improve fairness and model bias.



Applications and Conclusion

Personalized

Raiso Pareer, datation, and relationship advice by

personality.

Targeted Marketing

Customize messaging based on predicted personality types.

Team Building

Create balanced groups for optimal collaboration and productivity.

Ethics & Future

Avoid stereotypes, incorporate multimodal data like audio and images.



/