



STAT 5243 - Project 5

Group member:

Zhengfei Chen

Zichen Zhao

Jingqi Wang

Topic

Analyzing the most frequently used words in social media posts, categorized by the 16 MBTI types using NLP techniques.



Introduction

MBTI:

A personality test. It is a personality type system that divides everyone into 16 distinct personality types across 4 axis:

Introversion (I) – Extroversion (E)

Intuition (N) – Sensing (S)

Thinking (T) – Feeling (F)

Judging (J) – Perceiving (P)

MBTI Personality Test



Dataset

- Contains over 8,600 rows of data
- Each row is a person's:
 - MBTI Type
 - Last 50 things they have posted



Models

	Logistic Regression	Random Forest	XGBoost	SVM
Accuracy	0.69	0.48	0.65	0.67

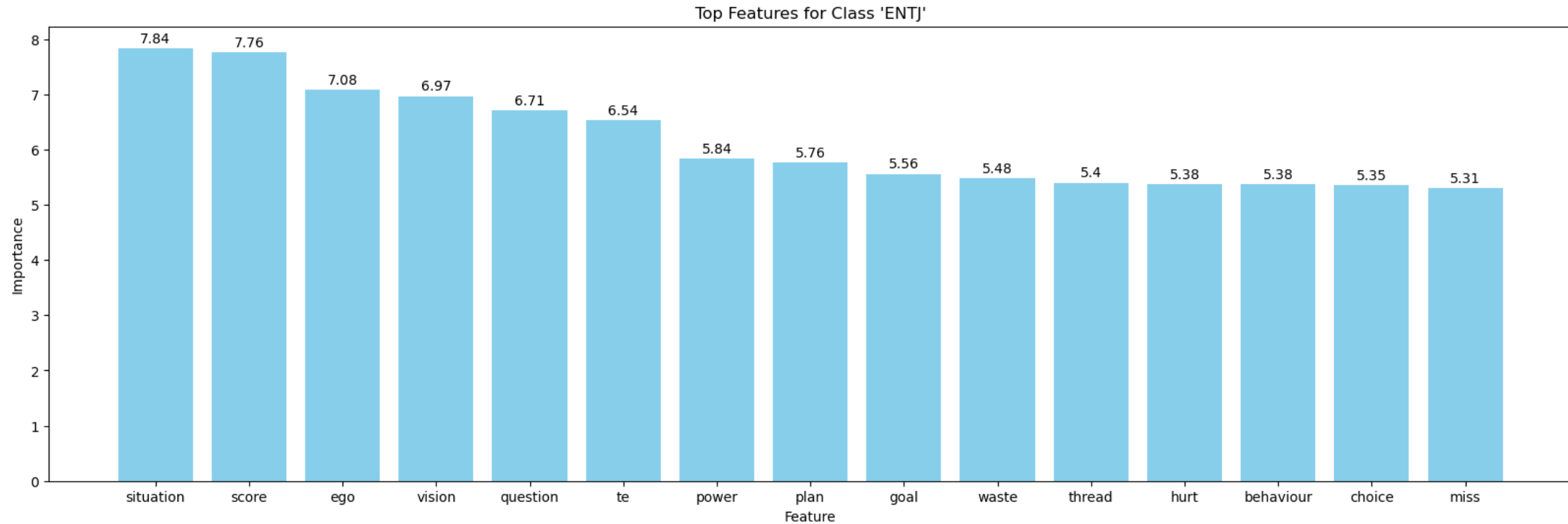


Hyperparameter Tuning

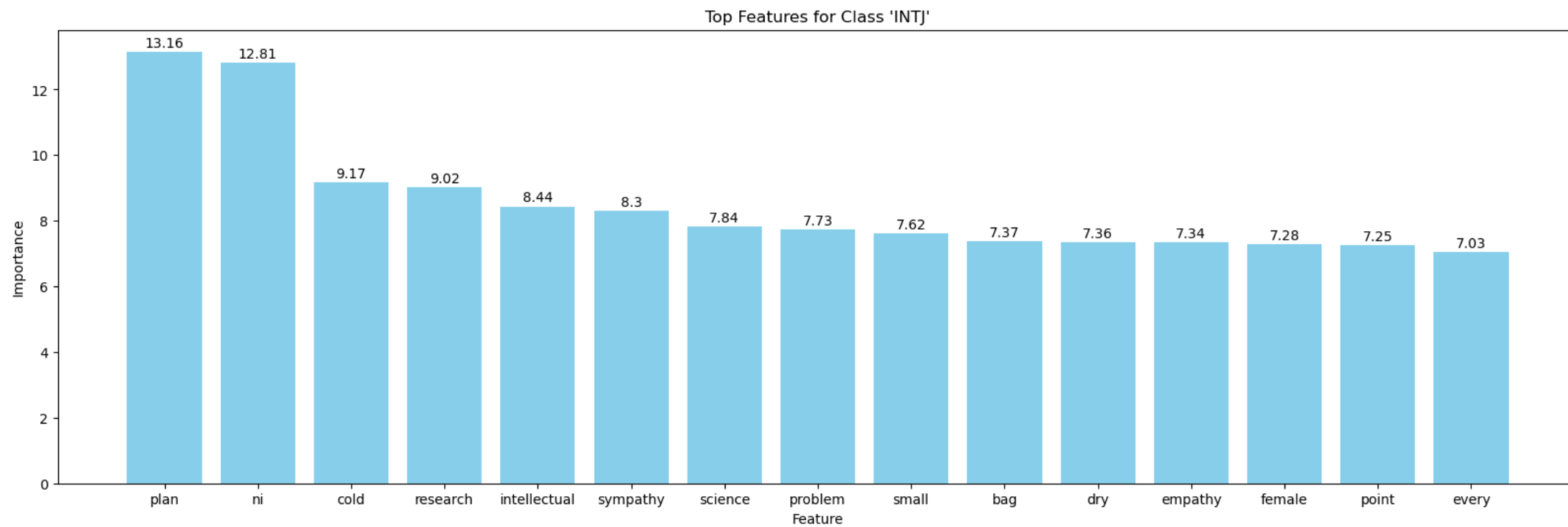
- Logistic Regression (highest accuracy)
- Using RandomsearchCV
- Performance improved



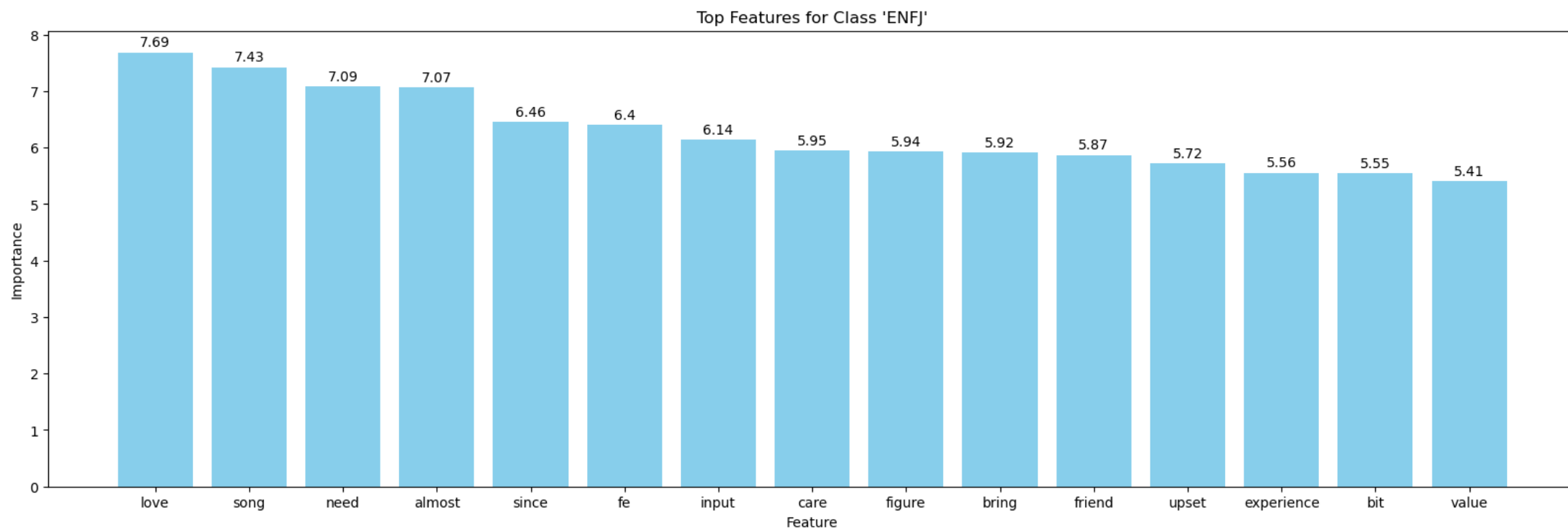
Results - ENTJ



Results - INTJ



Results - ENFJ



Applications

01

Enhanced Personalization

02

Improved Communication Strategies

03

Human Resource Development

04

Mental Health Support



Thank You!

