Model Evaluation Report

Multi-Class Classification of 16 MBTI Personality Types

1. Evaluation Metrics Used

Metric	Purpose
Accuracy	Overall percentage of correct predictions
Top-K Accuracy	True label within top K predicted classes
Macro Precision	Average precision across all 16 classes
Macro Recall	Average recall across all 16 classes
Macro F1-Score	Harmonic mean of precision & recall (primary metric)
Weighted F1	F1-score weighted by class support

2. Model Performance Comparison

Model	Test Accuracy	Top-3 Acc	Macro F1	Train-Test Gap
XGBoost	98.22%	99.65%	0.9822	1.73%
Random Forest	97.57%	99.41%	0.9757	2.42%
Logistic Regression	91.90%	97.81%	0.9189	0.14%
LDA	90.56%	97.43%	0.9055	0.16%

Baseline Comparison: XGBoost achieves 98.22% vs 6.25% random guess — **15.7× improvement**

3. Strengths and Weaknesses

Model	Strengths	Weaknesses
XGBoost	Highest accuracy, best Top-K, handles interactions	Slight overfitting, complex, longer training
Random Forest	Strong accuracy, feature importance, robust	Highest overfitting, more memory
Logistic Reg.	Best generalization, interpretable, fast	Lower accuracy, assumes linearity
LDA	Fastest training, dimensionality reduction	Lowest accuracy, assumes Gaussian

4. Handling Class Imbalance

Strategy: Stratified sampling maintained class proportions across train/validation/test splits. This ensures all 16 personality types are proportionally represented. **Macro F1-score** was used as the primary metric to treat all classes equally, regardless of frequency. Result: High macro F1 (0.90-0.98) indicates good performance across all personality types.

5. Practical Implications

Finding	Practical Implication
98% accuracy achievable	ML can reliably predict personality from surveys
Top-3 accuracy ~99.5%	True type almost always in top 3 suggestions
Ensemble methods best	Use XGBoost/RF for production systems
Linear models useful	Use Logistic Regression for explainability

Recommended Deployment:

- **Primary:** XGBoost for maximum accuracy
- **Explainability:** Logistic Regression coefficients for feature interpretation
- **User Experience:** Show top-3 predictions with confidence probabilities

Limitations: Models trained on self-reported surveys; 16 discrete types simplify continuous personality traits; high accuracy may reflect survey structure.