Title: Comparative Analysis of Machine Learning Techniques for Multi-Class Personality Type Classification

Step 5: Final Presentation
Group Name: Inverters
Group Members: Adrian Vaflor & Leian Carl Dela Cruz

Main Findings and Conclusions

Model	Test Accuracy	Top-3 Acc.	Macro F1	Key Insight
XGBoost	98.16%	99.20%	0.9815	Best overall performer; highest precision
Random Forest	97.57%	99.08%	0.9757	Strong alternative; highly robust.
Logistic Regression	91.90%	98.62%	0.9189	Best linear baseline; highly interpretable
LDA	90.56%	98.14%	0.9053	Fastest training; least overfitting.

• High Accuracy Achieved: 98.16% accuracy predicting 16 personality types
• Ensemble Methods Excel: XGBoost/RF outperform linear methods by 6-8%
• All Types Predictable: Per-class F1 scores: 0.9757-0.9815

Answering Research Objectives

Objective	Status	Result
Develop Classification Model	✓	XGBoost achieved 98.16% accuracy, significantly outperforming the 6.25% baseline (15.7x improvement).
Identify Key Traits	✓	Top Predictor identified: "You usually prefer just doing what you feel like at any given moment..." was the most influential feature across models.
Compare Algorithms	✓	Ranked: XGBoost > Random Forest > Logistic Regression > LDA. Ensemble methods outperformed linear models by ~6-8% .
Hard-to-predict types?	✓	None found. All 16 types were predicted with high consistency (Per-class F1 scores range: 0.975 - 0.982), proving no specific personality type was "hard" to classify .
Find the Top n Questions yielding similar Accuracy	✓	Top 35 questions is identified as the sweet spot to achieve almost the same accuracy as answering all 60 questions.

Limitations Encountered

Limitation	Impact
Synthetic/Survey Data	May not generalize to real assessments
Self-Reported Responses	Subject to response bias
16 Discrete Categories	Simplifies continuous personality traits
Single Dataset	External validity unknown

Potential Applications

1. Application Description Automated Assessment
    - Instant personality prediction from surveys.
2. HR & Team Building
    - Match team compositions by personality diversity.
3. Career Guidance
    - Recommend paths aligned with personality.
4. Personalization
    - Tailor content/recommendations to type
5. Psychology
    - Improve the mbti test questionnaires.

Recommended Future Work

• Short-term: Hyperparameter tuning, k-fold cross-validation, SHAP explainability
• Research: Real-world validation, neural networks, longitudinal studies
• Production: API deployment, user interface, model monitoring

Final Summary

This project demonstrated that ML can accurately predict 16 MBTI personality types with 98.22% accuracy. XGBoost significantly outperformed linear baselines. Models are reproducible (random_state=42) and ready for HR, education, and personalization applications.

Key Takeaway: Ensemble methods are highly effective for psychometric classification, achieving near-perfect accuracy with interpretable feature importance.