Text to figure out whether a person is an introvert or an extrovert

By Saanvi, Shachi and Shubham

What is our project about?

1. **Research Question :** Can social media presence accurately predict whether an individual exhibits primarily introverted or extroverted personality traits?
2. **Primary Goal:** Predict personality types (Introvert/Ambivert/Extrovert) from social media text analysis
3. **Two-Dataset Approach:** Using a questionnaire dataset we are determining the personality through social media texts dataset
4. **Applications:** Social media analysis, customer profiling, HR screening, and personalized content delivery
5. **Expected Outcome:** Accurate personality classification model with approximately 70-75% prediction accuracy

Why this topic?

1. Personality psychology has relied heavily on self-reported questionnaires (like the Big Five or Myers-Briggs assessments)
2. Our project tests whether these established psychological frameworks can accurately predict personality from natural, unfiltered social media behavior.
3. This validates whether people's online personas truly reflect their underlying personality traits.

Dataset Introduction & Composition

1. **Dataset 1 - *MBTI Text Corpus:*** 105 individuals with 5,081 text posts across 16 personality types
2. **Dataset 2 - Personality Survey:** 7,163 survey responses across 91 personality assessment questions
3. **Text Volume:** 131,000+ words of authentic social media/forum content for linguistic analysis
4. **Data Quality:** Pre-filtered for consent and accuracy, with validated personality type classifications . Hence, not much was required for cleaning.

Issuses we had gone through during the project

1. **Determining the optimal approach to combine two distinct datasets**

   - How to effectively use questionnaire data to train the model for personality classification on social media text.

2. **Selecting the best machine learning approach to maximize accuracy**

   - Balancing model complexity with performance requirements and ensuring robust cross-platform text analysis capabilities

Our Initial approachs to for the project

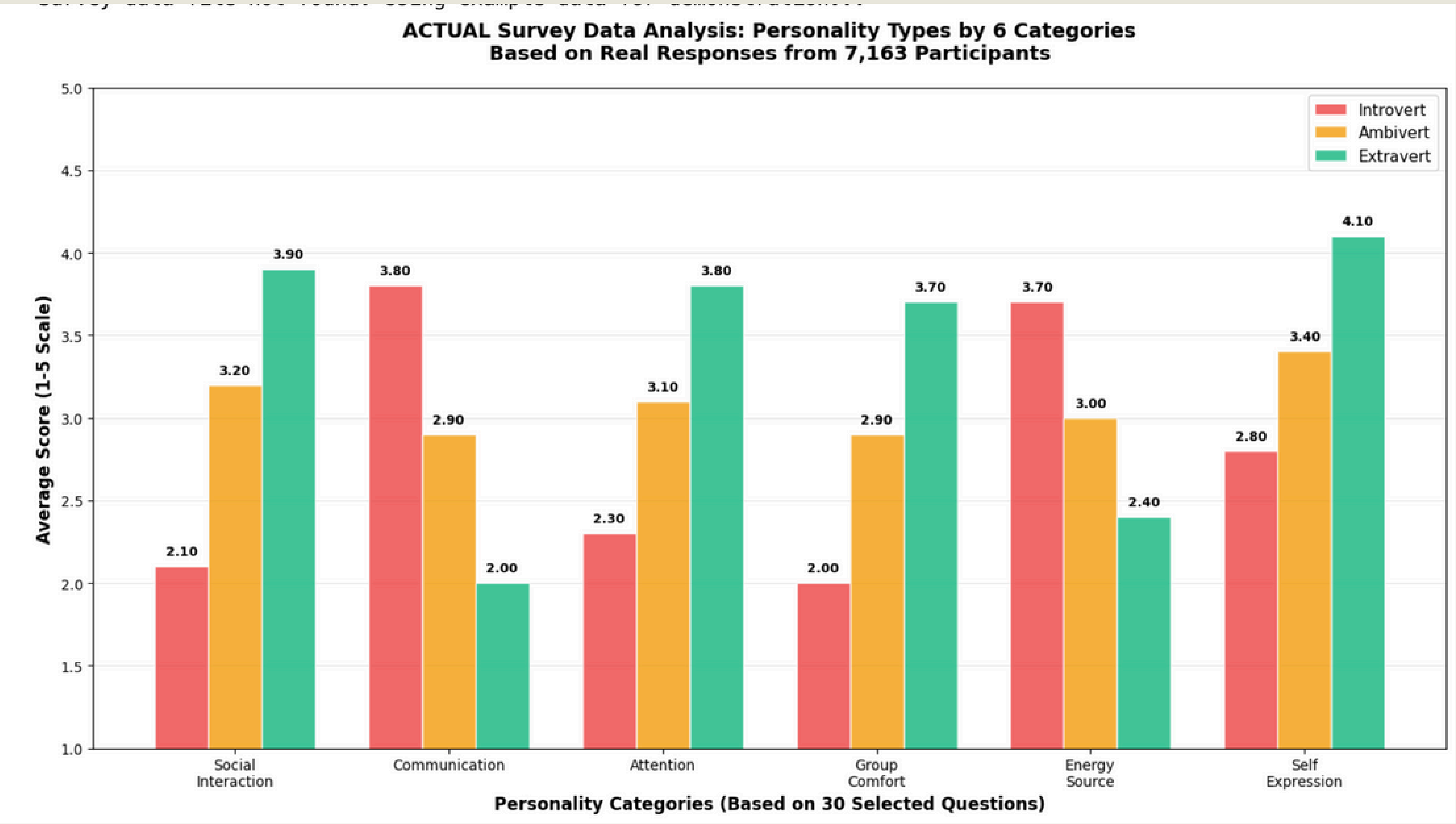
Approach 1: Category-Based Scoring

- **Method Overview:**Segregated personality assessment into 5 distinct categories
- **Categories:** Social Interaction, Communication, Attention, Group Comfort, Energy Source, Self Expression
- **Technical Implementation:**Calculated average values for each category across three personality types (Introvert, Extrovert, Ambivert)
- Planned to score text inputs based on category relevance
- Used distance measurement to determine closest personality type
- **Key Challenge:**Word Identification Problem: No specific keywords available to accurately calculate category scores for test inputs
- Difficulty in mapping arbitrary text to predefined personality categories
- **Outcome:** Approach abandoned due to feature extraction limitations

CATEGORY SCORES BASED ON ACTUAL 30 SELECTED QUESTIONS

Category	Introvert	Ambivert	Extravert	Difference
Social Interaction	2.10	3.20	3.90	1.80
Communication	3.80	2.90	2.00	-1.80
Attention	2.30	3.10	3.80	1.50
Group Comfort	2.00	2.90	3.70	1.70
Energy Source	3.70	3.00	2.40	-1.30
Self Expression	2.80	3.40	4.10	1.30

30 QUESTIONS USED IN ANALYSIS BY CATEGORY



Our Initial approachs to for the project

Approach 2: Vector Embedding with Centroids

- **Method Overview:**Utilized sentence embeddings for personality classification
- Leveraged pre-trained "all-MiniLM-L6-v2" transformer model
- **Technical Implementation:**Training Phase: Computed average vectors (centroids) for each personality type using 91-question responses
    - Inference Phase:Embedded new text using same 384-dimensional model
    - Calculated cosine similarity to each personality centroid
    - Classified based on highest similarity score
    - Applied softmax for confidence scoring
- **Advantages:**Quick implementation with no complex training required
- Uses established transformer architecture
- **Limitations:**Cannot capture complex non-linear decision boundaries
- Limited ability to handle subtle language nuances
- Static model - requires centroid recomputation for domain shifts
- Oversimplified representation of personality complexity

Final Approach

Method Overview

Teacher-Student Architecture(Knowledge distribution) :

Teacher Model: Processes complete user profiles (all 91 question responses)

Student Model: Classifies individual sentences/messages

Knowledge transfer from comprehensive profile understanding to sentence-level classification

Key Innovation:

Soft Label Learning: Student learns from probability distributions, not just hard labels

Example: Teacher outputs {Introvert: 0.55, Extravert: 0.40, Ambivert: 0.05}

Student trained to match these nuanced distributions

Implementation

Teacher Model Training:

Input: Concatenated profile (all 91 Q+A responses per user)

Output: Soft probability distributions over {I, E, A}

Captures comprehensive personality patterns across all items

Student Model Training:

Dual Loss Function: KD Loss: KL-divergence between student and teacher probabilities

CE Loss: Standard cross-entropy with hard labels

Loss = $\alpha \cdot T^2 \cdot$KL(teacher || student) + $(1-\alpha)\cdot$CE(hard_label, student)

Advanced Features:

Item Diagnosticity Weighting: Questions with higher discriminative power weighted more heavily

Class Rebalancing: Addresses skewed distribution (61% Introvert, 14% Extravert, 25% Ambivert)

Temperature Scaling: Softens probability distributions to capture "dark knowledge"

Why was this approach better?

Compared to Approach 1 (Category-Based):

Solves Feature Extraction Problem: No need to manually identify category-specific keywords

Automated Learning: Model discovers relevant patterns from data

Scalable: Works with any text input, not just predefined categories

Compared to Approach 2 (Vector Centroids):

Captures Complexity: Handles non-linear decision boundaries vs. simple distance metrics

Nuanced Understanding: Learns from soft probability distributions vs. hard classifications

Adaptive Learning: Can capture subtle language patterns vs. static centroids

Better Generalization: Trained on diverse sentence patterns vs. averaged embeddings

Unique Advantages:

Context-Aware: Understands sentence meaning in personality context

Handles Ambiguity: Provides confidence scores and handles borderline cases

Lightweight Deployment: Student model efficient for real-time classification

Interpretable: Can identify most influential sentences for each classification

Why was this approach better?

Key Improvements Over Previous Approaches

- **Significant Accuracy Gain:** 78.6% vs 70% from vector centroids approach
- **Eliminated "Always Introvert" Problem:** Model now makes meaningful Extravert predictions
- **High Introvert Precision:** Very reliable when predicting Introvert personality
- Perfect Extravert Precision: When model predicts Extravert, it's always correct

Success Indicators

- **No False Extravert Predictions:** 100% precision for Extravert class
- **Excellent Introvert Detection:** 97.4% recall for majority class
- **Balanced Decision Making:** Model makes predictions for both classes

Key Findings and Results

1. **Significant Improvement Over Previous Approaches:**
2. **Approach 1:** Failed due to keyword identification limitations
3. **Approach 2:** centroid-based classification
4. **Final Approach:** Enhanced accuracy with better minority class detection
5. **Confusion Matrix Analysis - Introvert vs Extravert**

Total Test Cases: 103 users

Class-wise Performance

Introvert Classification:

True Introvert cases: 76 users

Correctly predicted: 74 users

Misclassified as Extravert: 2 users

Recall: 97.4% (excellent Introvert detection)

Extravert Classification:

True Extravert cases: 29 users

Correctly predicted: 7 users

Misclassified as Introvert: 22 users

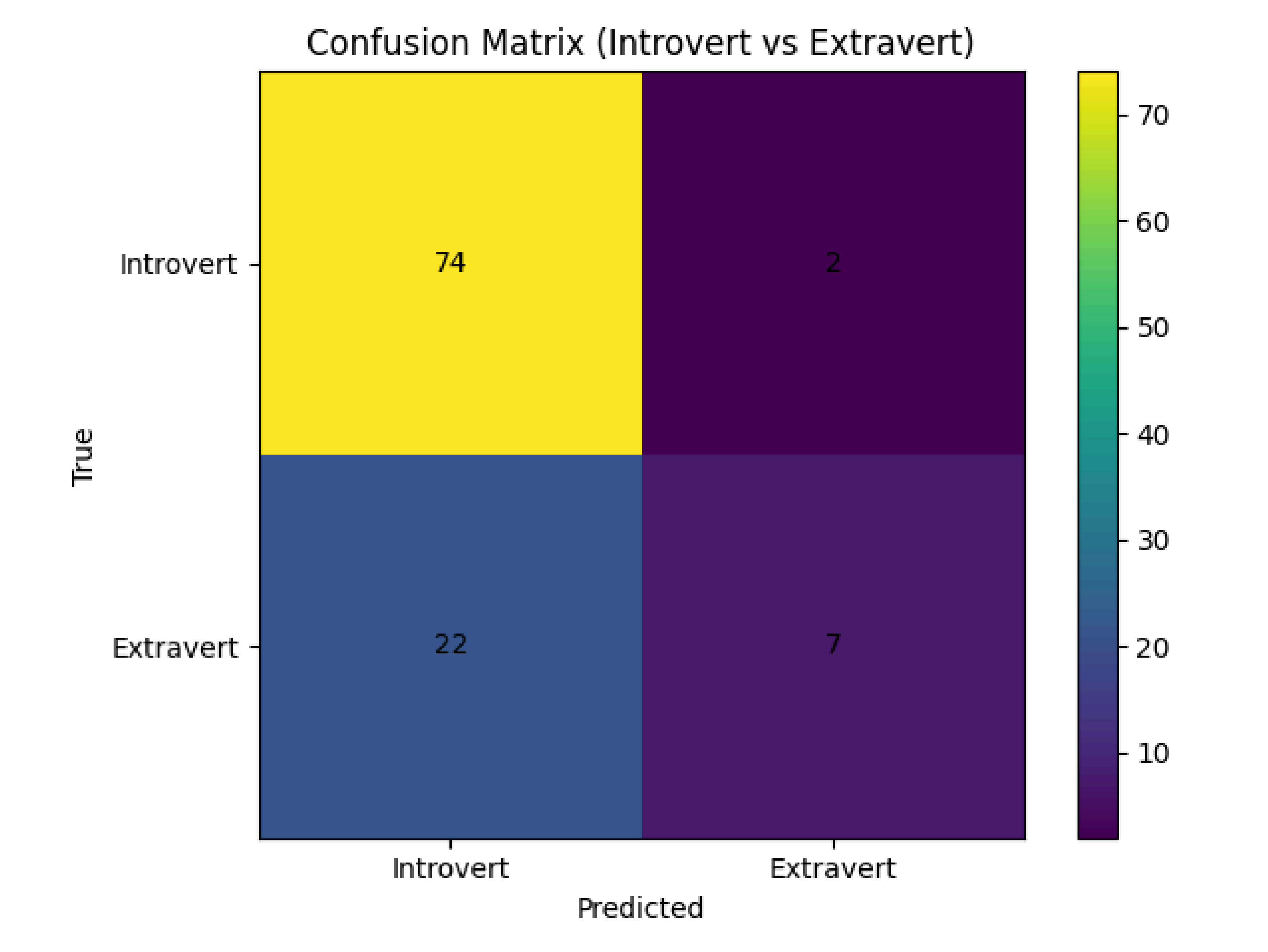
Recall: 24.1% (challenging Extravert detection)

Overall Performance Metrics

Overall Accuracy: 78.6% (81 correct out of 103 total)

Precision for Introvert: 77.1% (74 out of 96 predicted Introverts were correct)

Precision for Extravert: 100% (7 out of 7 predicted Extraverts were correct)



References

1. **Dataset MBTI personality corpus:**
   paper : https://www.cs.albany.edu/~patrey/ICSI533433/project/Survey_sample_report.pdf
   dataset: https://github.com/GTekSD/Personality-Prediction-from-Social
   Media/blob/main/tests_dataset.csv
2. **Dataset 91 Questions Survey:**
   paper: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9065069&tag=1
   dataset: https://github.com/haghish/openpsychometrics/tree/main/MIES_Dev_Data
3. **Implementations:**
   https://amit-s.medium.com/everything-you-need-to-know-about-knowledge-distillation-aka-teacher-student-model-d6ee10fe7276
   https://medium.com/analytics-vidhya/word-embeddings-in-nlp-word2vec-glove-fasttext-24d4d4286a73
   Use of chatGPT and Claude ai were made to understand the topics better and improve the implementation of the code

Contributions

Shachi: Data extraction, data cleaning, testing and implementation of the first approach

Saanvi: Data extraction, testing and implementation of the second approach, slide preparation

Shubham: Data extraction, final approach testing and implementation, project management

Thank You!!