



# Vivekanand Education Society's Institute of Technology

(Autonomous Institute Affiliated to University of Mumbai, Approved by AICTE & Recognised by Govt. of Maharashtra)  
*NAAC accredited with 'A' grade*

## Semester: VI Review: 6.2

Title of the Project: MBTI personality analysis project

Domain: AIML, Data Science, Psychology

Group Members:

Member 1: Sadneya Samant (46)

Member 2: Anshi Tiwari (56)

Guide: Dr. Ravita Mishra

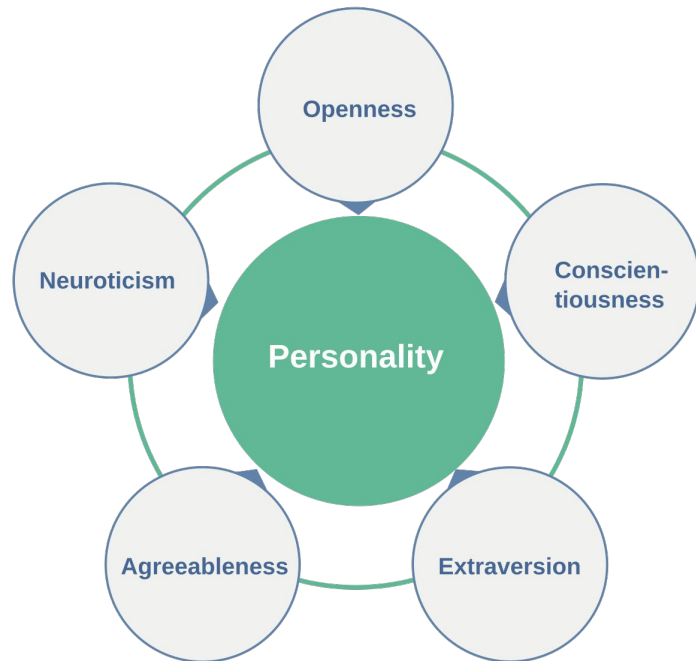


# Content

- Introduction
- Problem Statement
- Requirements
- Literature Survey
- Dataset Overview
- Feature Engineering
- Distribution Analysis
- Implementation (Clustering)
- Visualization
- Results and Analysis
- References

# Introduction to Project

- Project explores the relationship between Big Five personality test data and MBTI personality types
- Focus on classifying individuals into four main MBTI quadrants (IT, IF, ET, EF)
- Goal: Create a data-driven approach to personality assessment using clustering and decision trees





# Problem Statement

1. Traditional personality assessments often rely on **subjective self-reporting**
2. **Challenges in current personality assessments:**
  - **Inconsistency** between different personality frameworks (Big Five vs. MBTI)
  - **Lack of data-driven validation** for personality classifications
  - **Binary categorization** in a continuous personality spectrum
3. **Research Question:** Can we use machine learning to objectively classify individuals into MBTI quadrants based on Big Five personality test responses?



# Requirements

1. **Data sources:** Personality test responses, behavioral data, career performance metrics.
2. **Software Requirements: Programming Language:** Python  
**Libraries & Frameworks:**
  - a. **Data Processing & Analysis:** Pandas, NumPy
  - b. **Machine Learning:** Scikit-learn
  - c. **Data Visualization:** Matplotlib
3. **Hardware Requirements:**
  - a. High-performance computing resources for efficient model training and processing, including:
  - b. Multi-core CPUs/GPU for parallel computations.
  - c. Sufficient RAM and storage to handle large datasets.



# Literature Survey

Title	Published Date	Description
MBTI in Recruitment and Team Building	Proceedings of the 2022 7th International Conference on Social Sciences and Economic Development 2022	<ul style="list-style-type: none"><li>● MBTI is applied in HR for recruitment and team building.</li><li>● It affects job choice, behavior, decision-making, and teamwork.</li><li>● Personality alignment boosts job satisfaction, leadership, and work style.</li><li>● Helps with better hiring and team efficiency.</li><li>● Limitations: lacks strong evidence, ignores external job factors.</li><li>● Calls for more research to validate its workplace use.</li></ul>



# Literature Survey

Title	Published Date	Description
A Study of the Effects of an Individual's Personality and Characteristics on Job Behavior Using the Myers-Briggs Type Indicator	June 2023	<ul style="list-style-type: none"><li>• Study explores how MBTI traits affect job behavior, satisfaction, and workplace relationships.</li><li>• Analyzes the four MBTI dimensions and their role in decision-making, performance, leadership, and social interaction.</li><li>• Finds that while personality is stable, work experiences influence behavioral adaptation.</li><li>• Limitations: Doesn't account for factors like salary, job variety, or experience.</li><li>• Recommends future research across varied roles and organizations.</li></ul>



# Literature Survey

Title	Published Date	Description
Is Big Five better than MBTI? A Personality Computing Challenge using Twitter Data	2023	<ul style="list-style-type: none"><li>• Study compares MBTI and Big Five using Twitter data.</li><li>• Analyzed tweets in multiple languages with machine learning.</li><li>• Tested features like text patterns and linguistic categories.</li><li>• Found MBTI traits easier to predict than Big Five.</li><li>• Reason: Better class balance and more available data for MBTI.</li><li>• Aims to improve automated personality prediction.</li></ul>





# Dataset Overview

- **Source:** Big Five Personality Test dataset from Kaggle
- **Features:**
  - **EXT1-10:** Extroversion measures
  - **AGR1-10:** Agreeableness measures
  - **OPN1-5:** Openness measures
- **Successfully preprocessed data by:**
  - Removing missing values
  - Converting to numeric format
  - Creating composite personality scores

5 rows x 110 columns

Column Data Types:

EXT1	int64
EXT2	int64
EXT3	int64
EXT4	int64
EXT5	int64
...	
endelapse	int64
IPC	int64
country	object
lat_appx_lots_of_err	object
long_appx_lots_of_err	object
Length: 110, dtype: object	



# Feature Engineering

Created two key composite scores:

- **Extroversion Score:** Average of EXT1-10 and OPN1-5 measures
- **Agreeableness Score:** Average of AGR1-10 measures

Binary classifications:

- Introvert ( $<3.0$ ) vs. Extrovert ( $\geq 3.0$ )
- Thinker ( $<3.0$ ) vs. Feeler ( $\geq 3.0$ )

```
df = df.dropna()
df = df.apply(pd.to_numeric, errors='coerce')

# Feature Engineering: Calculate average scores for Extroversion & Agreeableness
extroversion_related = [f'EXT{i}' for i in range(1, 11)] + [f'OPN{i}' for i in range(1, 6)]
agreeableness_related = [f'AGR{i}' for i in range(1, 11)]

# Calculate new personality scores based on multiple factors
df['Extroversion_Score'] = df[extroversion_related].mean(axis=1)
df['Agreeableness_Score'] = df[agreeableness_related].mean(axis=1)

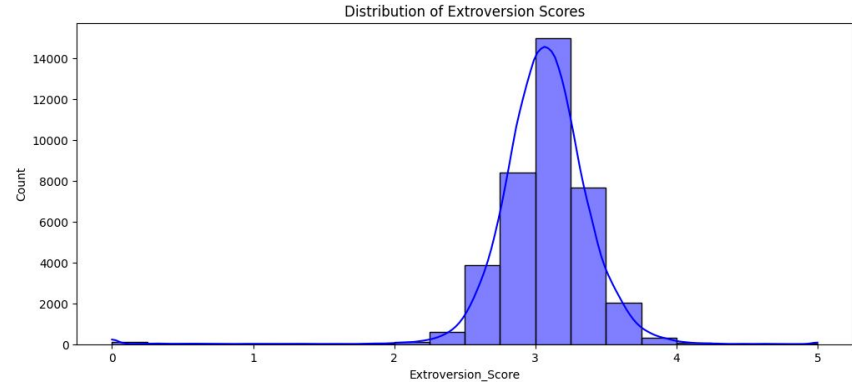
# Refined classification
df['Personality_IE'] = df['Extroversion_Score'].apply(lambda x: 'Introvert' if x < 3.0 else 'Extrovert')
df['Personality_TF'] = df['Agreeableness_Score'].apply(lambda x: 'Thinker' if x < 3.0 else 'Feeler')
```



# Distribution Analysis

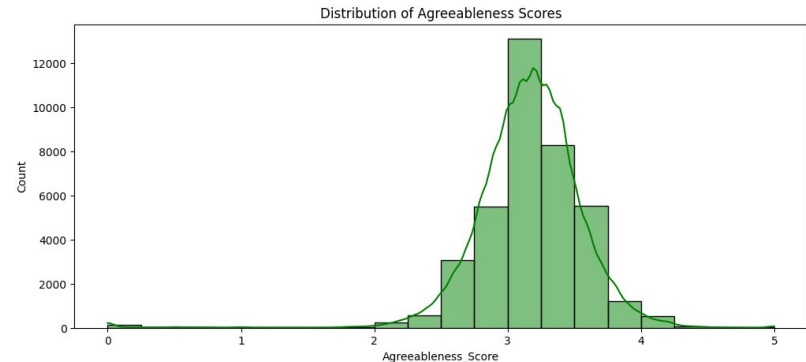
## Extroversion Distribution:

- Bell-shaped distribution with slight positive skew
- More balanced between introverts and extroverts



## Agreeableness Distribution:

- Well-balanced distribution between thinkers and feelers
- Correlation analysis showed strong relationships between related traits





# Clustering Implementation

## Applied two clustering methods:

- K-Means
- Gaussian Mixture Model (GMM)

## Process:

- Selected key features (AGR3-5, EXT3-4)
- Standardized data
- Set n\_clusters=4 to match MBTI quadrants (IT, IF, ET, EF)
- Evaluated using silhouette and Davies-Bouldin scores

## Used PCA to reduce dimensions for visualization

### K-Means Results:

- Four distinct clusters visible
- Clear separation between personality types

### GMM Results:

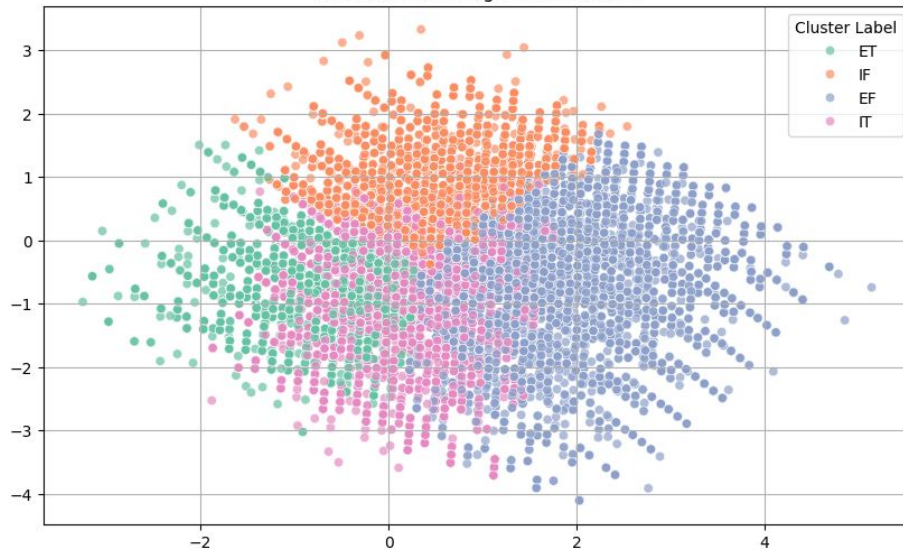
- More flexible boundaries between clusters
- Comparable performance metrics to K-Means

Both methods effectively grouped similar personality patterns

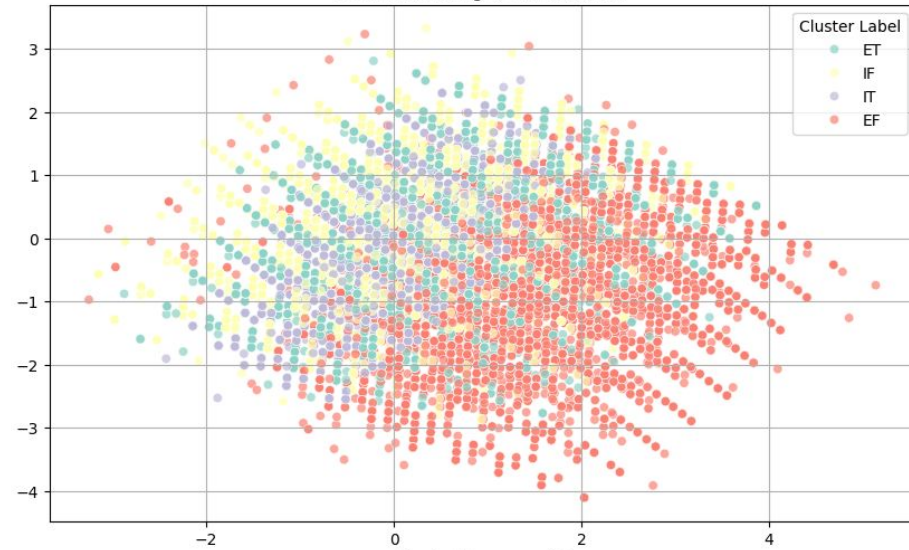


# Clustering Visualization

K-Means Clustering (PCA Reduced)



GMM Clustering (PCA Reduced)





# Result and analysis

Successfully classified individuals into IT, IF, ET, EF quadrants

- Certain questions have higher predictive power for personality traits.
- Distinct clustering patterns emerged despite personality being continuous.
- Balanced distribution among personality types.
- K-means got a higher silhouette score and a lower David-Bouldin score proving to be the better algorithm comparatively.

K-Means Clustering Scores:

Silhouette Score: 0.19975401595848513

Davies-Bouldin Score: 1.4449750215499377

GMM Clustering Scores:

Silhouette Score: 0.07887408681780998

Davies-Bouldin Score: 3.10714602781144



# Top Predictive Questions (Feature Importance)

## Most Influential Questions for IE:

1. EXT1 – *"I am the life of the party."*
2. EXT7 – *"I don't mind being the center of attention."*
3. EXT9 – *"I like to draw attention to myself."*
4. EXT2 – *"I talk a lot."*
5. EXT4 – *"I start conversations."*

## Most Influential Questions for TF:

1. AGR9 – *"I make decisions with my head more than my heart."*
2. AGR6 – *"I value logic over feelings when making decisions."*
3. AGR3 – *"I prefer objective analysis."*
4. AGR10 – *"I prioritize logic in my interactions."*
5. AGR4 – *"I consider myself an analytical thinker."*



# Interactive Form and Personality Scoring

## User Flow:

- Users respond to structured questions (EXT + TF style)
- Responses are scored on a scale of 1 to 5
- Average scores determine:
  - **Extraversion Score**
  - **Thinking Score**

## Classification Logic:

- If Extraversion Score > 3 → **E**, else **I**
- If Thinking Score > 3 → **T**, else **F**

TF9: I prioritize logic in my interactions.

Strongly Disagree    Disagree    Neutral    Agree    Strongly Agree

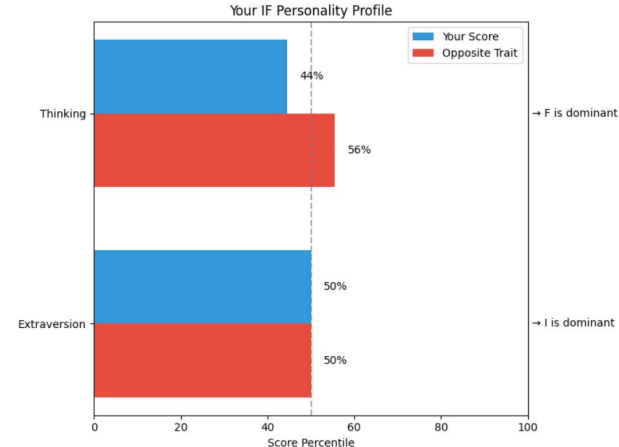
Disagree    Agree

1 2 3 4 5

Submit Quiz

Processing your responses...

🌸 Your Personality Type: IF 🌸





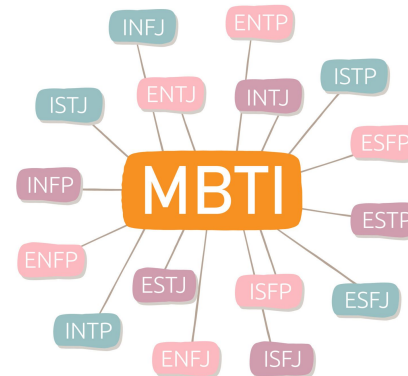
# Future Scope

- To implement this application fully, we need to make sure that the application takes in user input and determines the MBTI type based on the clusters.
- Thus the user answers questions based on the parameters and gets assigned a cluster which determines their personality type.
- A further expansion would have a classification system for all 16 personality types based on all 4 parameters and advanced machine learning algorithms that have a very high accuracy rate.

Match the questions to the type of openness.

4 points

	open	bipolar	highly closed	moderately closed
"What is your birthday?"	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
"Did you enjoy the movie?"	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
"What did you eat for dinner?"	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
"What do you imagine married life will be like?"	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>





# References

- A Study of the Effects of an Individual's Personality and Characteristics on Job Behavior Using the Myers-Briggs Type Indicator - [link](#)
- Research on the Application of MBTI in Organizations - [link](#)