

Machine Learning Lab

Aayushi Goel 2241103

The best model : Decision Trees

Dataset Description: Handwriting Features for Personality Classification:

This dataset is designed to analyze handwriting traits and determine whether a person is an **introvert (0) or extrovert (1)**. The dataset includes **four numerical handwriting attributes**, each providing insights into writing style and personality. The features are as follows:

- **Variance:** Represents fluctuations in writing pressure. A higher variance suggests more emotional or spontaneous behavior, while a lower variance may indicate steadiness and control.
- **Skewness:** Measures the tilt of the handwriting. A forward tilt (positive skewness) may be linked to extroversion, whereas a backward tilt (negative skewness) can be associated with introversion.
- **Kurtosis:** Determines how sharp or flat the curves in the handwriting are. High kurtosis implies attention to detail and structure, while low kurtosis may indicate a more relaxed and free-flowing writing style.
- **Entropy:** Captures the complexity of handwriting strokes. Higher entropy suggests randomness or creativity, while lower entropy reflects more structured and precise handwriting.

The **Class** column serves as the target variable:

- **0** = Introvert
- **1** = Extrovert

This dataset provides a **quantitative approach** to analyzing handwriting patterns and their correlation with personality traits. The extracted features allow for effective **machine learning-based classification** of individuals based on their writing styles.

Why Decision Trees Are the Best Model for Handwriting-Based Personality Prediction

The **Decision Tree model** proved to be the most effective in classifying individuals as **introverts or extroverts** based on their handwriting characteristics. Here's why it outperformed other models:

1. Adapts to Non-Linear Patterns

Unlike linear models, Decision Trees can **detect intricate patterns** between handwriting traits and personality.

- For instance, **stroke pressure variance** may not impact personality in a straightforward way. Instead, a sudden shift in variance might serve as a decisive factor in distinguishing introverts from extroverts.
- The model efficiently **segments data at meaningful points** (e.g., "If Variance > X, then Extrovert"), making it more suitable than Linear Regression for this type of classification.

2. Handles Diverse Data Distributions

Each handwriting trait (Variance, Skewness, Kurtosis, and Entropy) follows a distinct distribution:

- **Variance and Entropy** may be normally or skewedly distributed.
- **Kurtosis and Skewness** could exhibit long-tail distributions.
- Unlike **Linear Regression**, which assumes a uniform distribution, Decision Trees adaptively **split the data where necessary**, making them robust across different data types.

3. Identifies Key Features Clearly

Decision Trees provide insights into **which handwriting attributes** play the biggest role in determining personality.

- If **Skewness** emerges as the most critical factor, it suggests that **tilt direction** is a strong indicator of extroversion.
- If **Entropy** ranks higher, it highlights the importance of **stroke randomness** in

predicting personality.

Since Decision Trees **outline clear decision paths**, they are more transparent than **Neural Networks**, which function as a "black box."

4. Performs Well on Small & Medium Datasets

This model is particularly suited for scenarios where **data collection is limited**.

- Unlike **Neural Networks**, which need large datasets to generalize well, Decision Trees perform effectively even with fewer samples.
- To further improve accuracy, **ensemble methods** like Random Forest can be applied, reducing overfitting.

5. Fast & Computationally Efficient

Unlike **KNN**, which requires searching the entire dataset for every prediction, Decision Trees:

- Train quickly
- Provide instant predictions
- Require minimal computational power

This makes them ideal for **real-time handwriting analysis applications**.

Conclusion: Why Decision Trees Stand Out

- Effectively captures complex, non-linear relationships
- Works well with different types of data without extensive preprocessing
- Clearly identifies key handwriting features influencing personality
- Performs reliably on smaller datasets without overfitting
- Efficient, interpretable, and practical for real-world applications

For **handwriting-based personality classification**, Decision Trees offer the best balance of **accuracy, interpretability, and efficiency**, making them the top choice for this task.