Personality Types Analysis

Rüveyda Şahin

Data Scientist

Purpose of the Report

The purpose of this report is to present the findings of the analysis on personality types and to explain the methods used in the analysis process. The report aims to analyze certain personality traits using data science techniques and to reveal the relationships between these traits. The findings can contribute to a better understanding of the behavioral tendencies of personality types.

How did I do analysis?

I pulled the analysis data from the Kaggle. I looked at what data headers were used and what the data contained. Then I input it into the python program. And I started the analysis. The main libraries I used are; Pandas, Numpy, Matplotlib, Seaborn and Scikitlearn.

What are the Personality Types?

Personality types are classifications that help describe patterns in individual behaviors, attitudes, and emotions. These classifications aim to offer a framework for understanding individual differences and predicting how people are likely to respond in different situations. One of the most widely used personality frameworks is the **Myers-Briggs Type Indicator (MBTI)**, which categorizes people into 16 personality types based on four dichotomies:

- **Extraversion (E) – Introversion (I)**: Describes where people draw their energy from. Extraverts gain energy from social interactions, while introverts find energy in solitude.
- **Sensing (S) – Intuition (N)**: Indicates how people take in information. Sensors rely on observable facts, while intuitives look for underlying patterns.
- **Thinking (T) – Feeling (F)**: Shows how people make decisions. Thinkers prioritize objective criteria, while feelers weigh personal considerations.

- **Judging (J) – Perceiving (P)**: Describes how people approach the outside world. Judgers prefer structure, while perceivers are more open and flexible.

Analysis Results

1. **Personality Type Distribution:**

The distribution of personality types in the dataset was examined and it was observed that certain personality types were more dominant. This shows that some types may be included in the analysis more than others and therefore the model may be more successful in some types.

1. **Effect of Traits on Personality Type:** The effect of Introversion, Perception, Thinking and Judgment scores on personality types was visualized. It was observed that these traits were effective in distinguishing personality types. According to the correlation analysis, it is shown that there are intermediate performances between some traits, which can be decisive details between personality traits.
2. **Personality Type Distribution:** The personality types in the dataset were examined and it was observed that certain personality types were more dominant. However, it is seen that some types may be more prevalent than others in the analysis and therefore the models may be more successful in some types.
3. **Effect of Education and Gender on Personality:** The distribution of education level and gender on personality types was examined. This analysis showed that education level and gender may have different effects on certain personality traits.
3. **Model Performance:** Random Forest Model: For the best performance of the model, the optimal result was obtained with 150 trees using the "gini" or "entropy" criterion. The accuracy rate obtained in the training set showed how successful the model was on this data set. AdaBoost Model: The performance of the model was optimized with a flexible learning rate and tree numbers. Although a certain success was achieved in the AdaBoost model, different results were shown when compared to Random Forest.
4. **Most Influential Traits (Mutual Information):** The traits that have the greatest impact on personality type have been identified. This analysis provides access

to the model and findings that show which traits are more dispersed in the predictions.

RESULTS

1. **Performance and Success of the Model:**

During the study process, Random Forest and AdaBoost models were tested among the two models used for personality type prediction.

The Random Forest Model showed a more successful performance in predicting personality types by providing the highest accuracy rate in the training data. The success of this model stems from its ability to better process the relationship and effect of the features.

Although the AdaBoost Model achieved a certain level of accuracy, it showed lower performance than the Random Forest Model with limited accuracy. However, this model also helped in understanding the personality features in the data set.

2.

The Effect of Features on Personality Type:

Introversion, Sensing, Thinking and Judging scores emerged as the main features effective in determining personality types. These four features stand out as important factors in distinguishing personality types.

Mutual Information analysis shows the determining effect of these features on personality type. In particular, the fact that Thinking and Judging scores have higher mutual information values indicates that these features play an important role in personality estimation.

3. **Effect of Demographic Factors:**

Education Level and Gender are demographic factors that are observed to have different effects on personality traits. According to the findings of the study, while some personality types are more common at certain levels of education, it has been observed that gender also causes some changes in personality distribution.

This situation reveals that demographic factors can affect personality types and therefore these factors should be taken into consideration in personality analyses.

4. **General Evaluation and Future Studies:**

This analysis process has produced effective results with the use of different

models and techniques for personality estimation. The success of the Random Forest model can be considered as a potential tool for personality estimation. In the future, it may be possible to improve model performance and make estimations more reliable by using larger data sets and more demographic factors.

Source

https://www.kaggle.com/datasets/stealthtechnologies/predict-people-personality-types

https://scikit-learn.org/stable/

https://matplotlib.org/

https://seaborn.pydata.org/

https://pandas.pydata.org/

https://www.16personalities.com/personality-types

https://www.myersbriggs.org/my-mbti-personality-type/the-16-mbti-personality-types/