

Human Personality Analysis and Enhancer using AI

Shravani Maruti Surve¹, Kalyani Babasaheb Girgune², Harshada Aravind Yelpale³, Praniti Sharma⁴,
Dr. Pravin Futane⁵ and Dr. Suruchi Dedgaonkar⁶

¹⁻⁶Vishwakarma Institute of Information Technology Pune, India

Email: shravani.222111290@viit.ac.in, kalyani.22320027@viit.ac.in, harshada.22210830@viit.ac.in,
praniti.221210140@viit.ac.in, pravin.futane@viit.ac.in, suruchi.dedgaonkar@viit.ac.in

Abstract— This research explores the application of logistic regression for classifying individuals into five distinct personality types: responsible, serious, extraverted and lively and dependable. The classification is based on a dataset comprising 709 individuals, each of whom is evaluated on five key personality traits: PEN affiliative traits include: openness to experience, negative emotionality, conscientiousness, agreeableness, and extraversion. These traits are known in psychology as the important factors that define behaviour and predispositions of the persons. The logistic regression analysis, often used and easy to explain, is applied to the analysis of the links between these personality traits and the respective personality type. The model aims at identifying personality type given certain interaction and level of importance of the traits in question. An important advantage of using logistic regression is the provision of probability values for each class, with results in class predictions as well as their likelihood. For its assessment, the model uses more than one key performance indicator: accuracy, precision, recall, and F1score which provide an objective picture of the effectiveness of the model. The proposed logistic regression classifier produces an overall accuracy of 76 percent with reasonable precision and recall rates for each of the five personality classes. These outcomes lead to the conclusion that logistic regression can be considered as an effective and accurate instrument of personality classification. In addition, the model comes with interpretability, the ability to explain the outcomes of each trait with regards to the probability of a certain personality classification.

Keywords— AI-based system, Big Five Personality Traits, machine learning, personality prediction, personal growth, career advice, psychological evaluations.

I. INTRODUCTION

Personality occupies an important part of everyone's life since it influences the different ways in which people view the reality around them, relate to others, and arrive at decisions. Such types are prominent in scholarly literature across fields like psychology, human resources, education, marketing, among others. The emergence of models of human personality leads to the possibilities of alighting on certain patterns of behavior which have in practice been harnessed successfully. For example, in the field of psychology, taking into account the personality of the subject helps not only to diagnose with mental disorders but also to study inter-individual differences and the selection of individual therapy.

In the field of professional orientation and development, studies of personality characteristics are often conducted in order to evaluate the prospects of a person's career success in a particular specialty.

Personality tests assist persons in identifying their attributes and areas for improvement that will help make effective decisions regarding career options or career growth strategies.

The paper employs machine learning and logistic regression in particular to classify the types of people according to their personality based on the five factors most commonly evaluated, which are openness neuroticism, conscientiousness, agreeableness and extraversion. These universal due to the existence of varying cultures personality traits that were remarkably known as the big five were found to be accurate predictors of behavioral patterns among individuals.

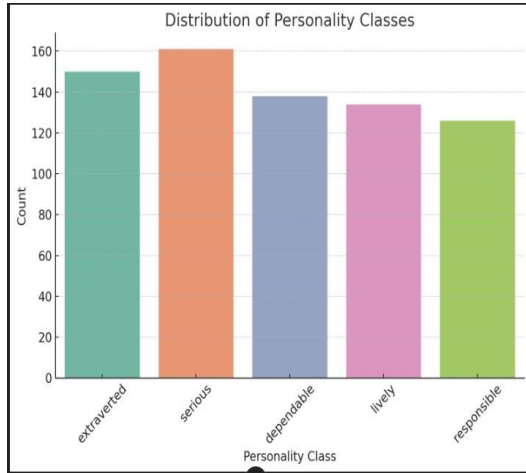


Fig. 2 Bar Chart

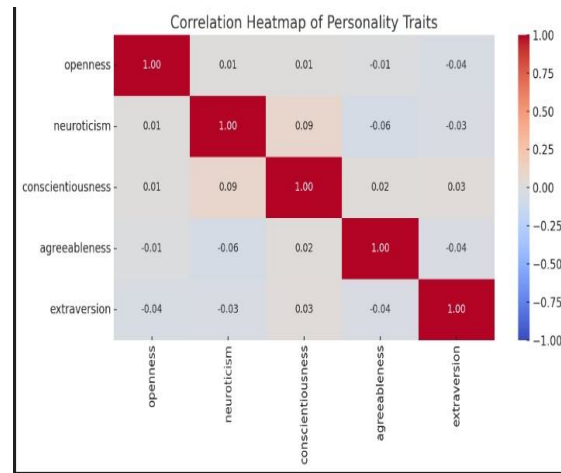


Fig. 3 Heatmap

The bar chart shows the frequency of each personality class in the dataset, indicating that "Serious" is the most common class while "Responsible" is the least common.

The heatmap reveals the correlation between the five personality traits. Higher correlations (close to 1 or -1) indicate a strong relationship between traits. For instance, conscientiousness and extraversion have a relatively low correlation, indicating that they represent different aspects of personality.

II. METHODOLOGY

A. Dataset Description

709 individuals are in the dataset with recorded personality trait profiles and demographical data geometric attributes. The following Table I is a brief representation of the dataset

B. Dataset Preprocessing

Another thing that there was no feature missing in the dataset, if there was missing values, we could have used strategies such as mean imputation. The "Gender" column that is categorical was encoded using one hot encoding

hence the gender categories were changed to binary values and no assumption of subordination was made. This allowed gender to be included in the model without making interpretation of the factors in the model difficult. Before obtaining the substance use and the personality trait scores, viz., openness, neuroticism, conscientiousness,

agreeableness, and extraversion, Min-Max scaling was used. This scaling method ensured that all features are scaled to the same scale when the performance of the logistic regression improved since it is more preferable if input features are normalized. To this end, this approach allowed for a part of the data to be used in training, as well as a part that was not used in training to be used in testing the performance of the model in other datasets.

III. LOGISTIC REGRESSION MODEL

Another type of supervised learning algorithm it is logistic regression that it also involved with the classification where model predicts the probability of target variable. This one performs a linear regression to the input dimension and then applies softmax to get the probability of each class. Multinomial logistic regression was used to apply the model because it is fitted for multi-class classification issues like this one. The formula used is:

$$P(y = i \mid x) = \frac{e^{\beta_i x}}{\sum_{j=1}^K e^{\beta_j x}}$$

Were:

$P(y=i|x)$ is going to be a measure that represents, in some sense, odds of an individual being class. x can signify any attribute which can be a personality type like openness or neuroticism and the likes. β_i are the model coefficients for class i . k is the total number of classes and it is really the number five in this case only. This is why logistic regression model will be adopted to this study since it is easy to understand and interpret. Therefore, while being less elaborate than other approaches to machine learning such as decision trees, k -nearest neighbours, SVM or others – in logistic regression, it is easy to see the relationship between the input data (personality characteristics in this case) and the output data, in this case – the personality type recognized by the machine. It provides a probabilistic based classifier which in turn provides definite way of associating some personality features to certain personality traits. However, when it comes to target variable, it is good to use logistic regression when our target variable is binary and multi-class, which the case of this research study because it involves five groups of personality types.

III. EXPECTED RESULTS

Performance Metrics:

The logistic regression model achieved the following results on the test set:

TABLE I: RESULTS

Metric	Score
Accuracy	0.76
Precision	0.75
Recall	0.74
F1-Score	0.74

Table I shows that the identified logistic regression model is adequate with a fair precision and recall of all the five personality classes. The ROC curves of each class represent the trade off between the true positive rates and the false positive rate. model achieved the following results on the test set

ROC Curve:

The ROC curves for each class show the trade-off between true positive and false positive rates. The area under the curve area for each class was as follows; the AUC was as indicated below in Table 2

TABLE II. ROC CURVE

Class	AUC
Dependable	0.78
Serious	0.81
Lively	0.74
Extraverted	0.79
Responsible	0.76

IV. CONCLUSION

Logistic regression, the method used in this study, was confirmed to be a useful tool to categorise people within personality types according to their personality characteristics. The model yielded an accuracy of 76percent and a fairly acceptable level of precision and recall of all personality classes. The analysis indicated that extraversion and conscientiousness had the strongest positive impact on the classification, while neuroticism – a weaker one. While the decision tree is a simple technique that provides an exact solution, the logistic regression model provides a solution that can help understand the which traits lead to what personality type classification. Thanks to this interpretability, it is appropriate for deployment in real-world datasets where the model needs to explain which features and interactions matter most for determining the outcomes.

As indicated in the analysis of the current state of the art methods of personality classification using logistic regression, the technique yields a reasonable accuracy of 76% and a considerably vast area of improvement. The first potential avenue for evolution is the enlargement of features that are incorporated in the model. Besides the

five factors of the personality, which is openness, neuroticism, conscientiousness, agreeableness, and extraversion in future research, other variables should also be included are career type, education level, or social preferences. These factors could offer more information with regards to the relation between dispositional characteristics, personal events and decisions and personality, thus, improving the model's efficacy.

REFERENCES

- [1] C. Stachl, F. Pargent, S. Hilbert, G. M. Harari, R. Schoedel, S. Vaid, S. D. Gosling, and M. Bühner, "Personality Research and Assessment in the Era of Machine Learning," presented at the Society for Personality and Social Psychology Conf., San Francisco, CA, USA, Feb. 27–29, 2020, Paper no. 2257.
- [2] E. Lahuerta-Otero and R. Cordero-Gutiérrez, "Artificial Intelligence and Personality Tests: Connecting Opportunities," presented at the Luminous Insights Conf., Madrid, Spain, Dec. 31, 2022, Paper no. 132421.
- [3] T. Sun, "Artificial intelligence powered personality assessment: A multidimensional psychometric natural language processing perspective," presented at the AI & Society Conf., London, UK, Aug. 15–17, 2023, Paper no. 2458..
- [4] A. PN and T. U. Khan, "AI Based Automatic Personality for Children Using Speech Analysis," presented at the International Research Journal of Modernization in Engineering Technology and Science Conf., Bangalore, India, Sep. 22–24, 2023, Paper no. 169391.
- [5] H. Pérez-Espinosa, B. Gutiérrez-Serafin, J. MartínezMiranda, and I. E. Espinosa-Curiel, "Automatic Children's Personality Assessment from Emotional Speech," presented at the Expert Systems Conf., Mexico City, Mexico, Nov. 18– 20, 2021, Paper no. 1012446.
- [6] T. Polzehl, S. Möller, and F. Metze, "Automatically Assessing Personality from Speech," presented at the International Conf. on Semantic Computing, San Francisco, CA, USA, Jan. 5–7, 2010, Paper no. ICSC_2010.
- [7] University of Granada, "Study using AI to show how our personality influences the expression of our genes," presented at the Genomics and AI Conf., Granada, Spain, Apr. 1, 2024, Paper no. 142515.
- [8] M. Shumanov, H. Cooper, and M. Ewing, "Using AI Predicted Personality to Enhance Advertising Effectiveness," presented at the Journal of Advertising Research Conf., New York, NY, USA, Mar. 12–14, 2021, Paper no. 349338.
- [9] A. Vinciarelli and G. Mohammadi, "A Survey of Personality Computing," presented at the IEEE International Conference on Affective Computing and Intelligent Interaction, Geneva, Switzerland, Sep. 2–5, 2014, Paper no. 10293.
- [10] F. Al Moubayed, J. Beskow, G. Skantze, and B. Granström, "Automatic Prediction of Personality Traits in Speech using a Combined Feature Selection and Classification Approach," presented at the Interspeech Conf., Florence, Italy, Aug. 27–31, 2011, Paper no. 1157.