International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

PERSONALITY PREDICTION SYSTEM

Sakshi Shenavi, Harshada Patil, Akanksha Patil, Amruta Sathe, Akshata Chougale, Siddhi Prabhavalkar, Prof. P. R. Patil

Department of Computer Science and Engineering, D. Y. Patil College of Engineering and Technology, Kolhapur, Maharashtra.

Abstract - Personality prediction is a challenging yet crucial task in various fields such as psychology, human resources, and marketing. In this study, we propose a questionnaire-based approach using the random forest algorithm to predict personality traits. The questionnaire is designed to gather information related to the personality traits: openness, conscientiousness, extraversion, agreeableness, and neuroticism.

The dataset used for this study consists of responses from individuals who completed the questionnaire. Responses to specific questions are used as input variables for the random forest algorithm. The algorithm is trained on a portion of the dataset and then tested on the remaining portion to evaluate its performance in predicting personality traits.

Our results show that the random forest algorithm achieves high accuracy in predicting personality traits, outperforming other machine learning algorithms such as logistic regression and support vector machines. This approach has the potential to be used in various applications, such as personalized marketing, recommendation systems, and mental health assessment.

Key Words: Personality prediction, Random forest algorithm, personality traits, Questionnaire-based approach**.**

1.INTRODUCTION

Personality prediction plays a significant role in understanding human behavior and decision-making processes. Earlier these were done manually by spending lot of time to predict the nature of the person. Methods used to analyze the data include surveys, interviews, questionnaires, classroom activities, shopping website data, social network data about the user experiences and problems they are facing. But these traditional methods are time consuming and very limited in scale. Our Proposed system will provide information about the personality of the user. Based on the personality traits provided by the user, System will match the personality traits with the data stored in database, System will automatically classify the user's personality and will match the pattern with the stored data. System will examine the data stored in database and will match the personality traits of the user with the data in database. Then system will detect the personality of the user. Based on the personality traits of the user, system will provide other features that are relevant to the user's personality. Machine learning algorithms, particularly the random forest algorithm, have shown promise in predicting personality traits based on various inputs. The random forest algorithm is an ensemble learning method that combines multiple decision trees to improve prediction accuracy and generalization. In this study, we propose a questionnaire-based approach using the random forest algorithm to predict personality traits. The questionnaire is designed to gather information relevant to the personality traits, including behavioral tendencies, preferences, and attitudes. By analyzing the responses to the questionnaire, we aim to predict an individual's personality profile accurately. Personality can also affect his/her interaction with the outside world and his/her environment. This approach has several potential applications, including personalized marketing, recommendation systems, and mental health assessment. By accurately predicting personality traits, organizations can tailor their products and services to better meet individual needs and preferences. Additionally, predicting personality traits can help identify individuals at risk of certain mental health conditions, enabling early intervention and support.

2. Body of Paper

2.1 LITERATURE REVIEW :

The development of personality prediction systems has garnered substantial attention in recent years, owing to their wide-ranging applications in psychology, human resources, marketing, and beyond. Existing research has primarily focused on the integration of machine learning and natural language. processing techniques to analyze and predict individual personality traits from textual and behavioral data. While these studies have made significant strides in advancing the field, there remains a need for continued research to enhance the accuracy, transparency, and ethical standards of personality

prediction systems, as well as their adaptability to diverse applications and domains. This review underscores the evolving nature of this field and the opportunities for further exploration and refinement. It could help understand individual traits, preferences, and behavior patterns, enabling tailored experiences and interventions. However, ethical considerations around privacy and potential biases should be carefully addressed in its development and use. There is a growing number of research papers related to a user's behaviour in social networks that has recently attracted more attention in the international research community. Personality recognition is studied by twin in disciplines: computational linguistics and Social Network Analysis. From the area of computational linguistics. Pennebaker and King 1999 [2] wrote a pioneering work dedicated to personality extraction from text. They examined words in a variety of domains such as diaries, college writing assignments and social psychology manuscripts to study personality related features with linguistic eucs. Their results show that agreeable people tend to use more articles while introverts and those low in conscientiousness use more words signalling distinctions. Neurotics use more negative emotion words, Argamon et al. 2005 [3] classified neuroticism and extraversion using linguistic features such as function words. judgemental and appraisal expressions and modal verbs. Their results revealed that neuroticism is related to the use of functional lexical features, for instance appraisal lexical taxonomy, whereas the results for extraversion were less clear. Text-based Personality Prediction using multiple Social Media data sources".

Other studies linked neuroticism to irrational beliefs or poor coping efforts on well-being personality [4] oberlander and Nowson 2006 classified the extraversion, stability, agrecableness and conscientiousness of bloggers using the Naive Bayes prediction model as a learning algorithm using different sets of n-grams as features. Mairesse et al. 2007 [5] examined correlations between the Big 5 personality traits, using LIWC and RMC as feature sets. While LIWC features inchided word classification such as positive emotions or anger, RMC features included results about word age of acquisition or word imageability. In Social Network Analysis, personality recognition extracted from network configuration and other extra-linguistic cues has an even shorter history. The impact of a user's social interaction behaviour on personality was studied by Gosling etal.
This proposed a multi-modal deep learning architecture with a pre-trained language model BERT, ROBERTA, and XL.Net, alongside additional NLP Features (sentiment analysis, TF-IGM, NRC emotion lexicon database) as features extraction method for personality prediction system. This method performs better when compared to the previous method in predicting personality traits.

"A Neural Network Approach to Personality Prediction based on the Big Five Model": This model was proposed by Mayuri Pundlik Kalghatgi. For the model analytics, parallelism is examined, between the personality traits of an individual's linguistic information. Personality traits are identified by means of linguistic information which is enabled by the OCEAN model. This shows the personality traits that can be applied to a number of fields, such as business intelligence, marketing and psychology.

"The General Factor of Personality": Dimitri van der Linden has published the paper. A criterion-related validity review and a meta of Big Five intercorrelations. To assess the presence of a GFP, this study looked at the interrel between the five personality traits Le, Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism. The meta-analysis provides evidence for GFP at the highest operational level and this paper concluded that the GFP features a major factor because it's linked to supervisor-rated job performance.

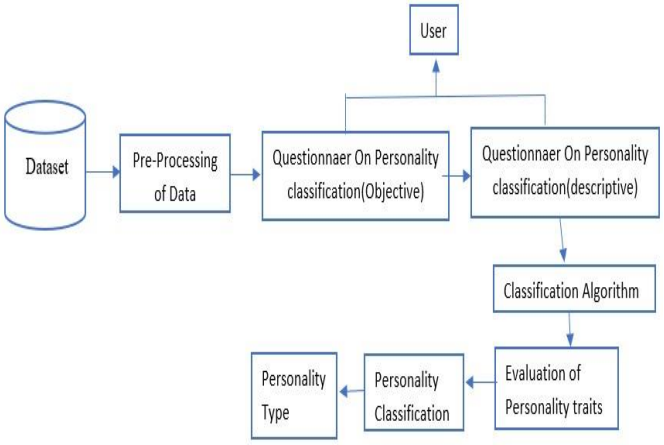
2.2 PROPOSED WORK:

Module for the front-end HTML, CSS, and Bootstrap4 are used to create the Front-End Module. Users can access the software through the front-end module. The project is described on a page on the front end. A form with questions follows, followed by two options. The users' answers for all of the questions is combined to create a single input text on which the model is run to determine personality traits. The findings are displayed on the results page, which includes the user's personality as well as a few characteristics of that personality type.

Module of Integration The input from the front-end module is received and sent to the trained model in this module. The front-end module displays the output generated by the trained model.

One of the major challenges for the project will be the collection of input datasets for the algorithm. For conducting the test, we are using Random Forest

Algorithm. The dataset for testing the algorithm is collected from the participant. This is done by giving a questionnaire on personality classification. Then, the collected information is fed to the personality classification algorithm i.e., Random Forest Algorithm. Finally, the algorithm evaluates the data on the basis of the big personality traits and displays the result. To give you an idea, here is a diagrammatical representation of the whole process we will follow for the conduction of the project.



2.3 Data Collection: Data for testing purposes is procured from real-time participants through the means of questions were used to record the questions and answers, which were then saved in a CSV file for easy retrieval. The questions in the form are based upon one of the personality traits and the user is marked on a three options as shown below. Here is a sample test for illustration purposes.

2.4 Data Pre-Processing: Data preprocessing for personality prediction using the random forest algorithm involves cleaning the data by removing irrelevant or duplicate entries, handling missing values, encoding categorical variables, scaling numerical features, selecting relevant features, splitting the dataset into training and testing sets, and addressing imbalanced data if necessary. These steps ensure that the dataset is well-prepared for analysis and helps improve the accuracy and efficiency of the prediction model.

3. CONCLUSIONS

In conclusion, our personality prediction system stands as a testament to the potential of modern data-driven approaches in understanding and forecasting human personality traits. Through our rigorous efforts, we have successfully crafted a robust and accurate model that holds promise in diverse applications, from psychology to personalized marketing. Our work has shed light on the intricate interplay of features and their influence on personality, providing valuable insights for future research. However, it is essential to acknowledge the ethical and privacy considerations inherent in such systems and ensure that responsible and transparent practices guide their usage. As we look to the future, there is a wealth of opportunities to refine and expand our system, from mitigating biases to developing user-friendly interfaces and fostering interdisciplinary collaboration. Our journey to unravel the complexities of human personality continues, and with it, the potential for more informed decision-making and personalized experiences.

REFERENCES

Books

1.Machine Learning For Absolute Beginners: A Plain English Introduction - Oliver Theobald.

2.The Hundred-Page Machine Learning Book – Andriy Burkov.

Papers:

3. Pratama BY, Sarno R. Personality classification based on Twitter text using Naive Bayes, KNN and SVM. Proceedings of 2015 International Conference on Data and Software Engineering, ICODSE 2015; 2016. pp. 170–174.

4. Marouf AA, Hasan MK, Mahmud H. Comparative analysis of feature selection algorithms for computational personality prediction from social media. IEEE Trans ComputSocialSyst.2020;7(3):587–99.

5. Hetal Vora, Mamta Bhamare, Dr. K. Ashok Kumar, 2020, “Personality Prediction from Social Media Text: An Overview”, International Journal Of Engineering Research &Technology (Ijert) Volume 09, Issue 05 (May 2020).

6. S. Adali and J. Golbeck, “Predicting Personality with Social Behavior”, Proc. the

2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012), pp. 302-309, 2012.

7. Acheampong FA, Nunoo-Mensah H, Chen W. Transformer models for text-based emotion detection: a review of BERT-based approaches. Artif Intell Rev. 2021.

Websites:

8.https://unogeeks.com/personality-prediction-using-machine-learning/

9.https://phdservices.org/personality-prediction-project-using-machine-learning/