Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
February 2024 5(2), 1–8

HUMAN BEHAVIOUR RECOMMENDATION SYSTEM USING MACHINE LEARNING

P. A. Kharade [1], Rajendra B. Mohite [1], Sandip S Kanase [2]

[1] Department of Electronics and Telecommunication Engineering, Bharati Vidyapeeth College of Engineering, Navi Mumbai, India
[2] Department of Mechanical Engineering Bharati Vidyapeeth College of Engineering, Navi Mumbai, India

Corresponding Author

P. A. Kharade, kharadepa@yahoo.co.in

DOI
10.29121/shodhkosh.v5.i2.2024.1613

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

During early adolescence, most mental disorders emerge, which contribute significantly to the global mental health burden, including India. Early identification of mental health problems is a major challenge in India. This study aimed to evaluate the effectiveness of mental health interventions among adolescents in India, and investigate personality patterns and psycho-social functioning among them using Machine Learning techniques. After conducting a literature review on Pub Med, Research gate and government websites, it was concluded that longitudinal study designs can be more efficient, less costly, and more robust to model selection, and they can have increased statistical power. The k means clustering technique was used to analyze personality patterns based on the Big 5 personality test. Psychological distress can be linked to personality patterns or traits, which refer to an individual's enduring patterns of thoughts, feelings, and behaviors that shape their characteristic ways of responding to and coping with the world around them.

Keywords: Data Science, Artificial Intelligence, Machine Learning, Personality Traits

1. INTRODUCTION

Mental health is a critical component of overall health and well-being, with the potential to impact social and economic outcomes throughout the lifespan. The adolescent years are especially crucial for establishing a foundation for healthy development and good mental health. However, there is growing concern globally about the increasing burden of mental health problems among this population. Most mental disorders emerge before the age of 25, with a higher prevalence between 11-18 years. In India, home to the world's largest number of adolescents, the burden of psychiatric disorders is significant, with a reported 6.5% of the community and 23.3% of school children and adolescents affected. Despite this, there is a considerable treatment gap for mental health disorders, with the lack of mental health programs for adolescents in schools and out-of-school adolescents being a major challenge. This study aims to address this issue by analyzing the personality patterns in adolescents using machine learning techniques and studying their

psycho-social functioning. The results suggest that psychological distress can be associated with personality patterns or traits, with some personality traits more closely associated with psychological distress than others.

An individual's particular set of traits, behaviors and cognitive processes define their personality, which in turn defines how they interact with the outside world. It is a recurring pattern of ideas, emotions, and actions that determines how someone sees, engages with, and reacts to their surroundings. A variety of genetic, environmental, and cultural factors influences a person's personality. As a result, it can greatly affect a person's relationships, job, and general well-being. Trait theory, psychoanalytic theory, humanistic theory, social cognitive theory, and other techniques can all be used to study personality.

An individual's personality is referred to as their nature, traits, or qualities. It includes the sum of a person's beliefs, attitudes, feelings, anxieties, objectives, and aspirations. The personality of a person is the deciding factor in how they will interact with their family and friends, choose a career, perform at work, and interact with others in the world. Knowing one's personality thoroughly allows one to be aware of both their strengths and weaknesses, which is beneficial for self-improvement.

Machine learning has become a potent tool for detecting personality traits, which is an important area of research in psychology. The Big Five Model, Myers-Briggs Type Indicator, Three-Trait Personality Model, and Eysenck Personality Questionnaire are a few popular theories for categorizing personality traits. A popular technique for determining a person's personality traits is the Big Five Personality Test. The test evaluates a person's degree of extroversion, agreeableness, conscientiousness, openness, and neuroticism. These traits are regarded as the fundamental building blocks of personality.

In this paper we will categorize personalities using the Big Five system. The most popular personality classification theory adopted by psychologists and researchers worldwide, this classification remains largely constant over the course of a person's lifetime.

We are employing the K-Means clustering algorithm, an unsupervised learning method, divides the unlabeled datasets into various clusters. Starting with the input unlabeled datasets, the algorithm divides them into k clusters before repeating the process until it is unable to identify the simplest cluster.

The Big Five model categorizes people into five broad categories:

Figure 1

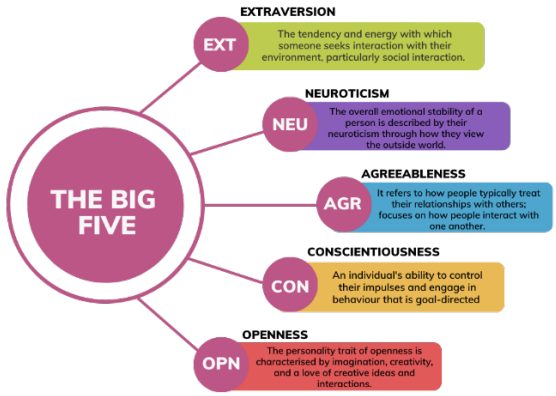


Figure 1 The Big Five Model

2. PROBLEM STATEMENT

The increasing burden of mental health problems among adolescents is a growing concern globally, and India is home to the largest number of adolescents in the world. However, there is a lack of mental health programs for adolescents in schools and out-of-school adolescents in India due to various challenges. To address this issue, this study aims to analyze the personality patterns in adolescents using machine learning techniques and study their psycho social functioning to better understand the association between psychological distress and personality traits. The findings of this study can inform the development of effective mental health programs for adolescents in India.

3. SCOPE OF THE PROJECT

Personality Assessment: Personality assessment will help people better understand their personality traits and how they relate to various career options or interpersonal interactions.

Recruitment and Selection: This can also be used to analyse the personality traits of job candidates and determine whether they are a good fit for the position's requirements and organisational culture.

Marketing and Advertising: By figuring out the personality traits of their customers, marketers and advertisers can develop more efficient and audience-specific campaigns.

Healthcare: When it comes to identifying and treating mental health conditions and enhancing patient outcomes, personality traits can also be a valuable source of knowledge for healthcare professionals.

Personal Development: It can be used as a tool for personal development, assisting people in recognising their assets and areas for improvement.

4. RELATED WORKS

Woodworth's Personal Data Sheet, released in 1917, was the first objective personality test. In World War I (WWI), that assessment was created to help soldiers who might experience nervous breakdowns due to enemy bombardment. The development of numerous competitive personality tests for use in industry followed shortly. In order to weed out candidates who would cause disruptions at work, many of these tests, like Woodworth's, focused on the concept of employee maladjustment.

Text, audio, and visual cues are the standard techniques for identifying personalities. Data pre-processing is an essential step that directly affects the outcomes when using textual data. The raw data is typically processed to extract the textual features, which are then fed into machine learning models like Support Vector Machines (SVMs), Naive Bayes Classifier, etc. [1]

In 2016, Kasula Chaithanya Pramodh et al. [2] used the MyPersonality dataset and stream-of-consciousness data. Around 10,000 Facebook Status Updates from 250 users make up the MyPersonality dataset. Their F1-scores are 0.665, 0.632, 0.625, 0.624, and 0.637 for the OPN, CON, EXT, AGR, and NEU traits, respectively. They used the Natural Language Toolkit for their model.

Waiel Tinwala, Shristi Rauniya in 2021 [3] In this study, the Big Five personality model is used to determine a person's personality. The deep convolution network

has been fed the processed data following the feature extraction at the document level. Personality traits have been categorized using a binary classifier. With the help of the Word2Vec Word embeddings library from Google, word-level feature extraction was carried out, and uniform distribution was used to handle any unidentified words that were not present in Word2Vec.

Devesh Agarwal, Mr. M. Karthikeyan in 2022 [4] This study employs k-Means clustering and Logistic regression model and feeds the model data obtained from a personality questionnaire. The dataset was divided into training data and testing data and was scaled by using sklearn Library. The training part is 70% and testing part is 30%.

Gokul K et al. [5] in 2018 have used the Bayes-Net classifier to determine whether a person is an extrovert or an introvert. Self-recorded audio samples were used as their dataset. They used attribute selection for dimensionality reduction along with voice activity detection for auditory nerve modelling, and they were successful in achieving an accuracy of 88.3%.

Yasmine Hernandez [6] The models used in this study suggest direct approaches like DISC personality questionnaires and supervised learning for text classification. Numerous fields, including marketing, education, and human resource management, benefit from personality.

Tajul Rosli Razak [7] This study suggests a fuzzy logic-based career recommendation system. The use of fuzzy logic assists students by providing recommendations for careers based on career tests.

Hiren Patel, Jash Sanghavi [8] A portal has been created to collect applicant input and conduct aptitude, interest, and personality tests. The recommended career is then produced by the machine learning model using these inputs as test data. Additionally, it has been noted that the accuracy predicted by various methods varies.

Jia Lu, Minh Nguyen, Wei Qi Yan [9] This study employs deep learning techniques to investigate the issue of human behavior recognition. Multiscale information was acquired to implement the recognition in this project in order to achieve sufficient recognition.

Abir Abyaa et al. in 2018 [10] have used the 48-students StudentLife dataset, which contains data. In order to categorize personalities according to the Big Five model, they used supervised learning algorithms. Support vector machines, random forests, logistic regression, C4.5 Decision Trees, and the k-nearest neighbor algorithm were all used.

Marwa S. Salem et al. [11] the Egyptian Twitter Users Dataset using Multinomial Naive Bayes, K-Nearest Neighbor, Support Vector Machine, and Decision Trees to categorize personality. Their analysis found that Decision Tree performed best for Conscientiousness, while K-Nearest Neighbour was the best algorithm for Openness, Extraversion, Agreeableness, and Neuroticism.

5. METHODOLOGY

The collection of input datasets for the algorithm will be one of the project's biggest challenges. We are employing the K - Mean Clustering Algorithm to carry out the test. The participant provides the dataset that will be used to test the algorithm. Giving a personality classification questionnaire accomplishes this. The K-Mean Clustering Algorithm, a personality classification algorithm, is then fed the information that has been gathered. The algorithm then assesses the data based on the big five personality traits and outputs the outcome. Here is a diagrammatic

representation of the entire procedure we will use to carry out the project to give you an idea

1) K- Means Clustering Algorithm

It is an iterative approach that separates the unlabeled dataset into k distinct clusters, each of which contains just one dataset and shares a set of characteristics.

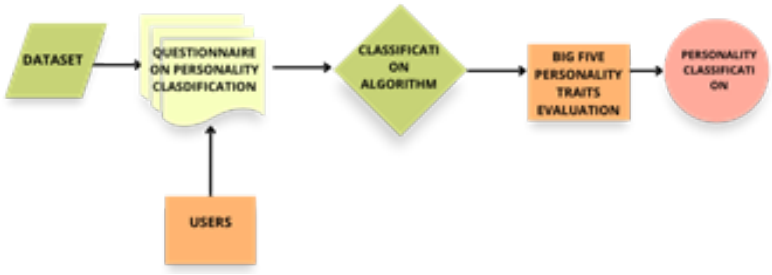
It provides a convenient method for categorizing the groups in the unlabeled dataset on its own, without the need for any training. It also enables us to cluster the data into various groups. Each cluster is connected to a centroid in this centroid-based technique.

This algorithm's major goal is to reduce the total distances between the data points and the clusters to which they belong. The algorithm starts with an unlabeled dataset and separates it into k clusters. It then continues the procedure until it is unable to discover the optimal clusters. In this algorithm, k should have a known value.

The two major functions of the k-means clustering algorithm are:

The most suitable value for K centre points or centroids is determined through an iterative method.

The closest k-center is given to each data point. The data points that are close to a certain k-center group together to form a cluster.

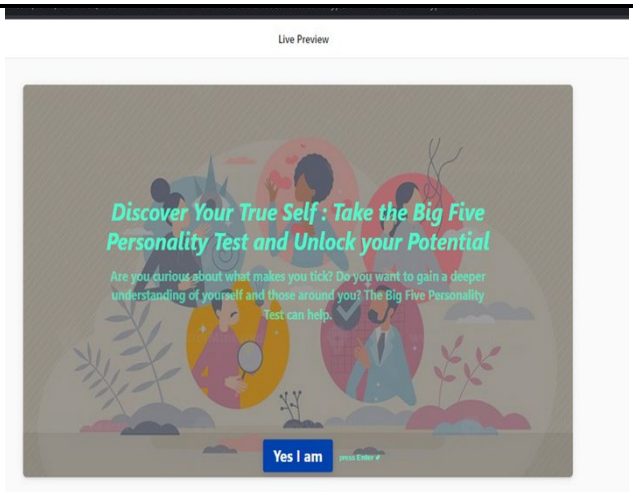


We are dividing people into 10 personality types using a datasets of 1 million responses from the Big Five personality test. The datasets consists of 250 personality questions, and the researchers feed this data into their machine learning algorithm. The data is divided into small batches of 100, resulting in 10,000 batches of data, which are fed into the algorithm in one iteration.

After the algorithm runs, we assign the resulting 10 clusters into 10 variables, and they inspect the results for Cluster 1. If someone has responded in a way that aligns with Cluster 1, they can be classified as having a personality type that corresponds to Cluster 1.

To provide more insights, we determine which personality traits each question relates to, such as extroversion, agreeableness, and so on. Then, we calculate the overall personality trait scores for each personality type. For each personality trait, they add or subtract the relevant answers to calculate the overall score.

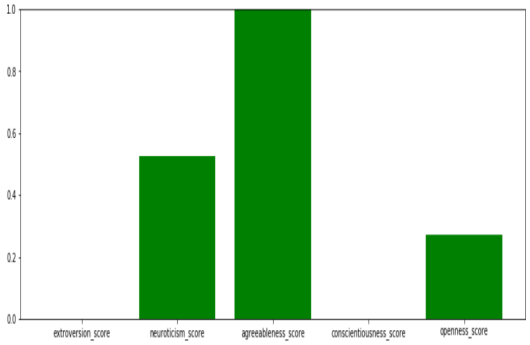
Finally, we normalize their data using the formula $z_i=x_i-\min(x)/\max(x)-\min(x)$. This process allows us to standardize the data and make it easier to compare across different variables. By analyzing the resulting clusters and personality trait scores, we can gain valuable insights into how different personality types are related to various outcomes and behaviors.

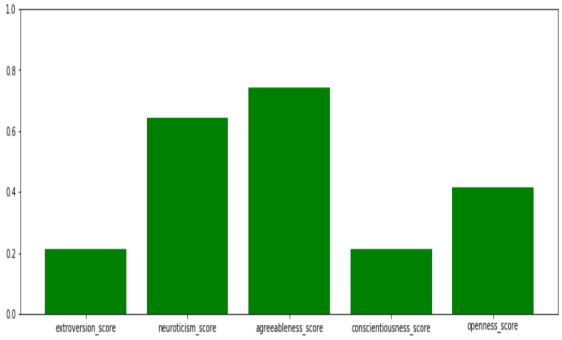


Adolescents with the resilient pattern have lower levels of neuroticism and relatively higher levels of other qualities, making them more mature than adolescents with other patterns. Over-controlled youth experienced relatively high levels of agreeableness over the course of the four years, in contrast to the resilient pattern. In other words, this group of teenagers may be extremely sensitive and overly aware of their environment.

Over the course of four years, the vulnerable pattern consistently scored lower on the other four qualities and higher on neuroticism.

Due to its large proportion of manipulative and low proportion of adaptive personality traits, it can be a personality pattern to avoid in terms of psycho social functioning. Average scores across all personality traits over the course of four years defined the moderate pattern. Adolescents in this pattern exhibit a moderate level of self-consciousness regarding all personality features





The present findings provide valuable implications for personality development in adolescence. First, this study identified patterns of personality development in adolescents., the present findings revealed that adolescents with vulnerable and moderate patterns tend to move toward maturity.findings revealed which pattern of personality development were psycho socially healthy or unhealthy in adolescents. Resilient adolescents are psycho socially healthy, over-controlled adolescents have internalizing problems, vulnerable adolescents are psycho-socially unhealthy, and moderate adolescents have externalizing problems. Furthermore, these personality patterns did not change from middle to late adolescence. Vulnerable adolescents may be particularly psycho socially unhealthy for four years. These findings provide longitudinal evidence for the relationship between persistent personality development and psycho social problems.

6. CONCLUSION

The study's findings imply that applying machine learning techniques could enhance programmes for adolescent mental health in India. The study suggests using longitudinal study designs, which have the potential to produce data with higher levels of statistical power and reliability. The Big 5 personality test can be used to discover personality patterns, which can aid in understanding the relationship between personality traits and psychological suffering. Overall, this study shows how machine learning methods could enhance mental health interventions and provide a better knowledge of the psychological aspects influencing teenagers in India.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

We would like to express our sincere gratitude to our guide Dr. Rajendra Mohite for their invaluable guidance and support throughout this study. Their expertise, encouragement, and constructive criticism were essential to the success of this research.

REFERENCES

Kai Hatano , Shogo Hihara , Kazumi Sugimura , Tetsuya Kawamoto, “Patterns of Personality Development and Psychosocial Functioning in Japanese Adolescents: A Four-Wave Longitudinal Study” Issue: 21 Jan 2023

Waiel Tinwala, Shristi Rauniyar, “Big Five Personality Detection Using Deep Convolutional Neural Networks”,

Devesh Agarwal, Mr. M. Karthikeyan “PERSONALITY PREDICTION USING MACHINE LEARNING” International Research Journal of Modernization in Engineering Technology and Science, Volume:04, Issue:April-2022

Yasmín Hernández , Alicia Martínez, Hugo Estrada, Javier Ortiz and Carlos Acevedo “Machine Learning Approach for Personality Recognition in Spanish Texts” published: 15 March 2022

Tajul Rosli Razak, Muhamad Arif Hashim, Iman Hazwan Abd Halim, Nur Fatin Farihin Shamsul “Career Path Recommendation System for UiTM Perlis Students Using Fuzzy Logic” International conference on Intelligent and Advanced Systems (ICIAS) Issue: June 2014

Hiren Patel, Jash Sanghavi, Saurin Shah, Sachet Thacker, Sandeep Mishra, " Career Guidance System Using Machine Learning" International Journal of Creative Research Thoughts (IJCRT), Volume: 09,
e:5 May 2021
Jia Lu, Minh Nguyen, Wei Qi Yan, "Deep Learning Methods for Human Behavior Recognition" 35th International Conference on Image and Vision Computing New Zealand (IVCNZ), Issue: November 2020
Marwa S. Salem, Sally Saad Ismail, Mostafa Aref , "Personality Traits for Egyptian Twitter Dataset" 8th International Conference, Issue: April 202
atalano, R. F., Fagan, A. A., Gavin, L. E., Greenberg, M., Irwin, C. E., Jr., Ross, D. A., & Shek, D. T. L. (2012). Worldwide application of prevention science in adolescent health. The Lancet, 379, 1653–1664.
Behavioral Patterns Of Students Under An Individualized Learning Strategy Tamar Levin, Zipora Libman And Rivka Amiad School Of Education. Tel Aviv University, Israel.
Donovan, J. (2005). Problem behavior theory. In C. B. Fisher & R. M. Lerner (Eds.), Encyclopedia of applied developmental science (Vol. 2, pp. 872–877). Thousand Oaks: Sage.
Donovan, J., & Jessor, R. (1985). Structure of problem behavior in adolescence and young adulthood. Journal of Consulting & Clinical Psychology, 53, 890–904.
Farrell, A. D., Kung, E. M., White, K. S., & Valois, R. (2000). The structure of self-reported aggression, drug use, and delinquent behaviors during early adolescence. Journal of Clinical Child Psychology, 29, 282–292.
Jaccard, J. (2016). The prevention of problem behaviors in adolescents and young adults: Perspectives on theory and practice. Journal of the Society for Social Work and Research, 7, 585–613. https://doi.org/10.1086/689354.
Jessor, R., & Turbin, M. (2014). Parsing protection and risk for problem behavior versus pro-social behavior among US and Chinese adolescents. Journal of Youth and Adolescence, 43, 1037–1051. https://doi.org/10.1007/s10964-014-0130-y.
Jessor, R., Turbin, M., Costa, F. M., Dong, Q., Zhang, H., & Wang, Z. (2003). Adolescent problem behavior in China and the United States: A cross-national study of psychosocial protective factors. Journal of Research on Adolescence, 13, 329–360.
What predicts adolescent delinquent behavior in Hong Kong? A longitudinal study of personal and family factors. Social Indicators Research, 129, 1291–1318. https://doi.org/10.1007/s11205-015-1170-8.
Oetting ER, Beauvais F. Peer cluster theory, socialization characteristics and adolescent drug use: A path analysis. J Counsel Psychol 1987; 34:205-213.