WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Personality Prediction Ocean Model Using Machine Learning

Dr. C. Sunitha [a], Abirami N [b]

[a] *Associate Professor and Head, Department of Software Systems, Sri Krishna Arts and Science College, Coimbatore 641008, India*
[b] *Student, Department of Software Systems, Sri Krishna Arts and Science College, Coimbatore 641008, India*
DOI: https://doi.org/10.55248/gengpi.4.1023.102623

A B S T R A C T

Numerous instances of machine learning applications are evident in our daily routines one particularly noteworthy utilization involves the categorization of individuals based on their distinct personal characteristics. A significant focal point of this application pertains to sorting individuals by the big five personality traits. This categorization framework is recognized as the five-factor model (FFM) or the OCEAN model. This operates as a methodology for arranging diverse personality traits. The OCEAN model encompasses the dimensions of "Openness to experience, Conscientiousness, Extroversion, Agreeableness, and Neuroticism". To achieve precise evaluations of personalities this system attempts to employ three different algorithms Logistic regression, K-Nearest Neighbors Algorithm, and Random Forest algorithm". In line with the results, "Logistic Regression achieved better accuracy when compared to other algorithms."

Keywords: OCEAN Model, FFM, KNN, Logistic Regression, Random Forest, Personality Traits, Machine learning

Introduction

Numerous contemporary personality psychologists argue that personality can be broken down into five fundamental dimensions, commonly known as the "Big 5" personality traits. The objective of this project is to predict individual personalities using the personality traits defined by the OCEAN Model or The Five-Factor Model. The acronym "OCEAN" stands for "Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism". Openness encompasses creativity and a sense of curiosity, while conscientiousness is associated with thoughtfulness. Extraversion primarily relates to sociability, while agreeableness is linked with kindness. Neuroticism often involves feelings of sadness or emotional instability. Gaining an understanding of each personality trait and comprehending the implications of scoring high or low in these traits can offer valuable insights into your own personality without the need for a dedicated personality traits test. Additionally, this awareness can enhance your comprehension of others, enabling you to gauge their position on the spectrum for each of these personality traits.

The analysis delved into the influence factors affecting the correlation among the Five-Factor Model of personality traits. The dataset encompassed these significant traits, encompassing individual's gender and age. As a preliminary step to model training, the data was subjected to preprocessing, which involved addressing missing values, discretizing, and standardizing the information. The overarching goal of the project was to establish a system that could enhance the accuracy of predicting individual personalities. This was to be achieved by amalgamating outcomes from three separate machine learning techniques: k-Nearest Neighbours, random forest, and logistic regression.

The central challenge is to identify an individual's personality based on the big five traits or the five-factor model, using an 8-point Likert scale for measurement. The scale ranges from 1 (indicating lower values) to 8 (indicating higher values). For example, an extroversion score of 1 indicates pure introversion, while a score of 8 suggests high sociability. Neuroticism values can span from 1 (frequent mood changes) to 8 (consistent mood stability). Agreeableness scores are indicative of cooperation and adaptability in teamwork, crucial for effective decision-making. A high agreeableness score signifies strong cooperative and adaptable tendencies. Furthermore, the project showcases one of its practical applications by assessing compatibility for marriage. This is achieved by comparing the personalities of both the groom and the bride. If the personality traits are found to be well-matched, the project declares them as compatible couples.

Indeed, identifying one's own personality can be valuable for students when it comes to making informed career choices. Understanding your personality traits can help you align your career path with your strengths and preferences. For example, someone with a highly extroverted personality might thrive in a career that involves a lot of social interaction, such as sales or marketing, while a person with a more introverted personality might excel in roles that require deep analytical thinking, like data analysis or programming. Recognizing your personality can provide valuable insights into the types of roles and work environments that suit you best, ultimately leading to more satisfying and fulfilling career decisions.

1.1 Ocean Model

- **Openness to Experience**: Openness is a fundamental personality trait defined by an individual's receptiveness to new experiences, ideas, and concepts. Individuals with high levels of openness typically display curiosity, imagination, and an open-minded approach to life. They are often eager to explore different viewpoints, embrace novel experiences, and frequently engage in creative activities. Conversely, individuals with low levels of openness may lean towards conventionality, favor routine, and exhibit resistance to change or new ideas.
- **Conscientiousness**: Conscientiousness represents a personality dimension distinguished by robust impulse control, a propensity for strategic thinking, and a commitment to goal-oriented behaviors. Individuals who score high in conscientiousness typically exhibit traits such as organization, meticulous attention to detail, forward planning, and a conscientious awareness of how their actions affect others. This personality trait often corresponds to a strong sense of responsibility and a disciplined approach to tasks and responsibilities.
- **Extraversion**: Extraversion is characterized by traits such as enthusiasm, sociability, assertiveness, and the expression of emotions. People with high extraversion scores tend to thrive in social environments, deriving energy and happiness from social interactions. They are typically outgoing and enjoy being around others. In contrast, introverts may find social situations tiring and may need time alone to rejuvenate. They often exhibit more reserved and introspective tendencies, and social interactions can be draining for them.
- **Agreeableness**: Agreeableness, one of the key personality traits, encompasses attributes like trust, kindness, friendliness, and empathy. Individuals with high levels of agreeableness are often characterized by their cooperative, compassionate, and considerate nature. They tend to work collaboratively, nurture positive relationships, and prioritize harmonious interactions. Conversely, those with lower scores in agreeableness may exhibit traits such as competitiveness and occasionally engage in manipulative behavior. Cooperation and fostering harmony in their relationships might be of lesser importance to individuals with lower agreeableness scores.
- **Neuroticism**: Neuroticism is a personality trait characterized by features such as moodiness and emotional instability. Individuals with low scores on neuroticism tend to have stable and even-tempered emotional states. They experience fewer mood swings and tend to worry less irrationally. They are generally patient and composed. Conversely, individuals with high neuroticism scores often exhibit emotional volatility. They may experience frequent mood swings, irrational worries, and impatience. However, it's worth noting that high neuroticism doesn't imply any inherent problems; it simply indicates a propensity for more emotional fluctuations and anxiety.

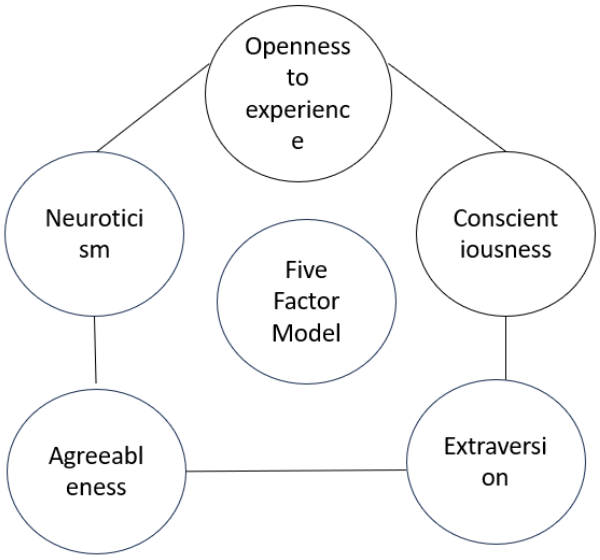

Fig. 1 - Ocean Model.

2. Literature Survey

In Feb 2019, Prediction System through CV Analysis is proposed by Allan Robey, Kashish Agarwal, Keval Joshi, Shalimali Joshi with an objective of simplifying the process of hiring a candidate for an organization. This model tries to design a plan to integrate job characteristics model to E_HR system to search for a model of efficient operation on human resource management in the Internet age.

In April 2022, Personality Prediction using Machine learning is proposed by Devesh Agarwal, Mr. M. Karthikeyan. This project helps to write the personality test and check the personality of the person. From the personality classification, the person can view the type of personality and can improve the personality based upon the results.

In Sep 2021, Personality Prediction Via CV Analysis using Machine Learning is proposed by Atharva Kulkarni, Tanuj Shankarwar, Siddharth Thorat. This model attempts to examine different machine learning algorithm approaches for efficiently predicting personality through CV analysis using NLP techniques.

3. Proposed System

3.1 Problem Objective

The primary objective is to identify a person's strong skill set based on their predicted personality traits. This system empowers users to readily discern their personality traits. In this project, a distinctive approach has been adopted to predict an individual's personality using Machine Learning algorithms, specifically Logistic Regression, K-Nearest Neighbours algorithm, and Random Forest.

By employing these three algorithms, it becomes possible to compare the accuracy of various models and accurately forecast human personality traits. Moreover, the project demonstrates a practical application by evaluating marriage compatibility. It accomplishes this by comparing the personalities of both the groom and the bride. When the personality traits align well, the project concludes that they are a compatible couple. This illustrates how the prediction of personality traits can be applied to real-life situations, aiding individuals in making significant life decisions such as choosing a life partner.

3.2 Advantages of proposed system

- **Personalized Usage**: This project enables individuals to predict and understand their own personality traits independently by considering logical values that span from 1 to 8. This self-assessment feature enhances user engagement, leading to higher satisfaction and loyalty.
- **Cost-Effective**: Once the model is developed, individuals can ascertain their personality traits without the need to consult others for minor reasons. This time-saving aspect eliminates the necessity of consulting professionals for personality identification.

  For example, young individuals can leverage this tool to make informed career choices.
- **24/7 Availability**: The model's availability round-the-clock ensures that users receive immediate support, leading to improved customer satisfaction and reduced response times.
- **Scalability**: The current model focuses on the Big Five traits or the Five-Factor Model, known as the OCEAN model. If new features emerge in the future, this system can be extended to accommodate the additional personality traits.
- **Improved Insights**: A personality prediction web application can gather user interaction data, facilitating the identification of their personality traits and offering valuable insights into user preferences and behavior.

Overall, a well-designed prediction system provides numerous benefits to both businesses and users, including tailored interactions, cost savings, 24/7 accessibility, scalability, and enhanced insights.

3.3 Input Dataset

As part of this project's functionality, we engage users through an interface resembling a structured form. The information inputted by users encompasses the significant big five personality traits, in addition to their gender and age. This comprehensive set of input values ensures that the predictive model is equipped with essential data for precise personality trait prediction.

Table 1 – Input data set.

S. No	Attribute	Data Type	Range of values
1	Gender	Nominal	Male/Female
2	Age	Numeric	17-28(preferred)
3	Openness	Numeric	1-8
4	Conscientiousness	Numeric	1-8
5	Extraversion	Numeric	1-8
6	Agreeableness	Numeric	1-8
7	Neuroticism	Numeric	1-8

3.4 Targeted features or output features

The ocean model primarily aims to predict the following target variable

- **Responsible**: This personality trait characterizes individuals who take ownership of their responsibilities and obligations, consistently fulfilling them with reliability and ethical conduct. Responsible individuals demonstrate a strong sense of accountability for their actions and decisions. They are trustworthy in managing tasks, making informed choices, and acting in the best interests of their organization or team. Their presence contributes to establishing order, dependability, and a sense of reliability in both personal and professional contexts.
- **Dependable**: A dependable person is someone who consistently delivers on their commitments and promises. They can be relied upon to fulfil their responsibilities and meet deadlines. Dependable individuals demonstrate a strong work ethic and reliability in both personal and professional contexts. Their consistent and trustworthy behaviour fosters trust and confidence in their abilities.
- **Serious**: A serious person is someone who approaches tasks and situations with a focused and earnest demeanour. They prioritize their responsibilities and commitments, often displaying a strong sense of dedication and diligence. Serious individuals tend to take matters seriously, demonstrating a no-nonsense attitude and a commitment to achieving their goals. Their seriousness is often associated with a sense of responsibility and a dedication to excellence.
- **Lively**: A lively person is characterized by their vibrant and energetic demeanour, often bringing enthusiasm and positivity to their interactions and surroundings. They exude a zest for life and a contagious sense of joy. Lively individuals are known for their active engagement in social activities and their ability to uplift the mood of those around them. Their spirited nature contributes to a lively and enjoyable atmosphere in various social and communal settings.
- **Extroverted**: An extroverted person is characterized by their outgoing and sociable nature, thriving in social interactions and group settings. They often seek opportunities to engage with others, enjoy being the centre of attention, and are energized by social activities. Extroverts tend to be expressive and comfortable in social situations, readily forming connections and friendships. Their openness and sociability contribute to their ability to build strong networks and navigate social environments with ease.

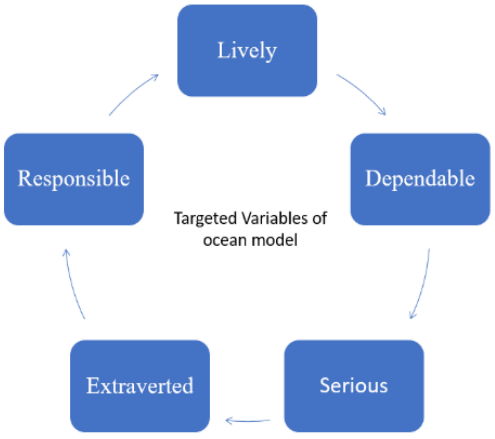

Fig. 2 – Output feature variables

4. Process flow

- **Data Collection**: Data collection is a systematic process that involves the methodical gathering of information about specific variables. This procedure is instrumental in addressing questions, testing hypotheses, and evaluating the results of a study or analysis. The primary aim is to collect and quantify data related to specific factors within a predefined framework, which in turn allows for the exploration of pertinent inquiries and the assessment of outcomes. In the context of this model, data collection is accomplished by sourcing data from Kaggle, a popular platform that hosts a variety of datasets for diverse analytical purposes.
- **Data Pre-Processing**: Data pre-processing is a crucial step that entails the transformation of raw data before feeding it into the machine learning algorithm. The objective here is to clean and structure the data effectively for analysis. Raw data gathered from diverse sources often needs formatting to align with the analytical requirements and suit specific machine learning models. For instance, certain algorithms may not be compatible with null or missing values, necessitating the implementation of strategies to handle such data gaps. Furthermore, the data should be prepared in a manner that facilitates its utilization with various machine learning algorithms, enabling the selection of the most optimal model for the task at hand.

- **Training Data and Test Data**: In order to construct and assess the model's performance, the dataset is divided into two subsets: the training set and the test set. As a common practice, this division is often done in a 3:1 ratio, with 70% of the data allocated for training purposes and the remaining 30% reserved for testing and evaluation. This partitioning allows the model to learn from the training data and subsequently evaluate its predictive abilities on unseen data, helping to gauge its generalization performance.
- **Model Creation**: Model creation encompasses the construction of a machine learning model using the prepared dataset. This procedure entails situating machine learning within the organizational context, exploring and selecting suitable algorithms, preparing and refining the dataset, executing cross-validation to evaluate the model, enhancing the performance of the machine learning model through optimization techniques, and, finally, deploying the model for practical application.
- **Model Prediction**: Model Prediction: Predictive modelling is a data-driven approach that employs statistical techniques, machine learning, and data mining to forecast and project future events or outcomes based on historical and existing data. This process involves the analysis of both present and past data, utilizing the insights gained to construct a model capable of making predictions regarding probable future scenarios. During the training phase with three distinct models, the system independently evaluates each algorithm. This methodology finds applications in various domains, including personality assessment and marriage compatibility evaluation. Once the result is obtained, users gain insight into their personality traits and can apply this knowledge for their specific purposes.

4.1. System flow diagram

A system flow diagram is a visual representation that illustrates the entire system's functionality. It outlines the journey of information or processes from their origin to their destination. The diagram is typically segmented into various stages, each providing insights into different aspects of the system's operation. System flowcharts can vary in complexity, ranging from basic, hand-drawn outlines of processes to comprehensive diagrams that delve deeply into data handling procedures. They serve multiple purposes, including the analysis of existing systems or the modelling of new ones.

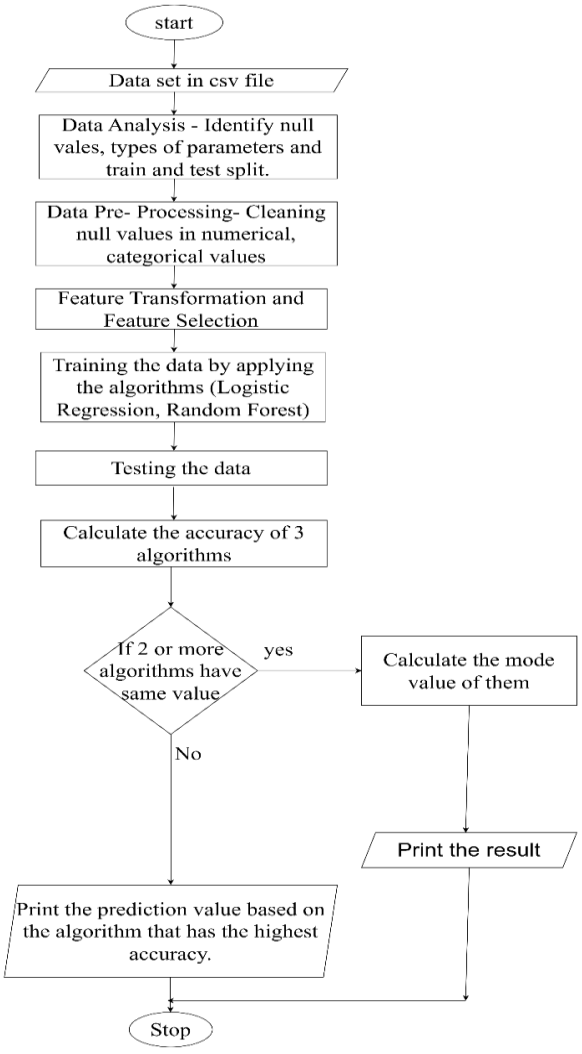

Fig. 3 – System flow diagram

5. Algorithms used

5.1. Logistic regression

Logistic Regression serves as a statistical analysis technique frequently applied in predictive analytics and modelling, finding extended utility within the realm of machine learning. It finds application in statistical software to establish connections between the dependent variable and one or more independent variables by estimating probabilities through a logistic regression equation.

Given that there were more than two classes involved which includes age, gender, Openness, Neuroticism, Conscientiousness, Agreeableness, Extroversion the implementation involves multinomial logistic regression. The chosen solver, "newton-cg," is particularly well-suited for larger datasets, enhancing the model's efficacy when dealing with substantial data volumes.

5.2. K-Nearest Neighbours Algorithm

K-Nearest Neighbours (KNN) stands out as one of the foundational yet crucial classification algorithms in the field of Machine Learning. It falls within the supervised learning category and is extensively employed in tasks involving pattern recognition, data mining, and intrusion detection.

What distinguishes KNN is its practical applicability across real-world scenarios. Its non-parametric nature allows it to thrive without any underlying assumptions about data distribution, a contrast to algorithms like Gaussian Mixture Models (GMM) that presuppose a Gaussian distribution for the provided data. In KNN, prior data—referred to as training data—is furnished, and it classifies points into groups defined by a specific attribute. This attribute-based classification allows KNN to make predictions based on proximity to known data points.

5.2. Random forest

Random Forest stands as a widely recognized machine learning algorithm that falls under the umbrella of supervised learning techniques. This versatile algorithm finds application in both Classification and Regression tasks within the domain of Machine Learning. Its foundation rests upon the concept of ensemble learning, a strategy that amalgamates multiple classifiers to address intricate problems and enhance model performance. Contrasting the conventional reliance on a single decision tree, the essence of random forest lies in its ability to draw predictions from numerous trees. The algorithm then aggregates these individual tree predictions through a majority vote mechanism, culminating in the final output prediction. The effectiveness of random forest is amplified as the number of constituent trees increases, contributing to elevated accuracy while also acting as a safeguard against the peril of overfitting, wherein a model becomes excessively fine-tuned to the training data and performs poorly on new, unseen data.

5. Results and Discussions

This project is centered around creating a system with the capability to predict an individual's personality traits with elevated accuracy and also demonstrated its one of the applications for checking marriage compatibility by using the same model. The approach involves amalgamating the outcomes derived from three distinct machine learning techniques: the k-nearest neighbor algorithm, the random forest algorithm, and the logistic regression algorithm. The prediction mechanism entails comparing the outputs generated by these three algorithms, with the final prediction being determined by the mode of these three outputs. In instances where the outputs diverge, the system prioritizes the result produced by the initial algorithm, namely the logistic regression. This choice is founded on the logistic regression algorithm's consistent display of higher accuracy compared to other algorithms in the mix.

6. Conclusion

In conclusion, predicting personality traits based on the Ocean Model using machine learning is a fascinating area of research with promising applications across various domains. This endeavor presents significant challenges, including the multifaceted nature of personality, the need for high-quality and diverse datasets, and the translation of abstract personality traits into quantifiable features. However, it offers valuable insights and potential benefits.

Machine learning models for personality prediction have practical use cases in fields such as human resources, marketing, and mental health. For instance, in HR, these models can aid in candidate selection, team composition, and employee development. In marketing, personalized advertising and product recommendations can be tailored to individual personalities. Moreover, in the mental health domain, early detection of personality-related risk factors may assist in providing timely interventions and support. Ethical considerations are paramount in the development and deployment of such models. Data privacy, informed consent, and the responsible use of predictions are critical aspects that must be addressed. Continuous improvement and validation of these models are also essential to ensure their accuracy and relevance. Overall, while predicting personality traits using machine learning remains a challenging endeavor, its potential benefits make it an exciting avenue for future research and practical applications.

References

Allan Robey, Kashish Agarwal, Keval Joshi, Shalimali Joshi (2019).Personality Prediction System through CV Analysis, IRJET vol 06 issue 02, e-ISSN-2395-0056, p-ISSN-2395-0072

Devesh Agarwal, Mr. M. Karthikeyan(2022). PERSONALITY PREDICTION USING MACHINE LEARNING, IRJMETS Vol:04 /Issue:04/ April-2022, e-ISSN-2582-5208

Atharva Kulkarni, Tanuj Shankarwar, Siddharth Thorat (2021). Personality Prediction Via CV Analysis using Machine Learning, IJERT Sep 2021 Vol 10 Issue 09, ISSN – 2278-0181

Priyanka Kamble, Umesh Kulkarni, Ankit Sanghavi (2022). An Overview of Personality Recognition through Machine Learning for E- Recruitment, JETIR Aug 2022 Vol 9 Issue 8, ISSN-2349-5162

Binisha Mohan, Dinju Vattavayil Joseph, Bharat Plavelil Subhash, "Personality Detection of Applicants and Employees Using K-mode Algorithm And Ocean Model, https://arxiv.org/submit/4652175

G. Sudha, Sasipriya K K, Sri Janani S, Nivethitha D, Saranya S, Karthick Thyagesh G (2021). Personality Prediction Through CV Analysis using machine Learning Algorithms for Automated E-Recruitment Process, IEEE, 978-1-6654-1447-0/21/

Dr. Nidhi Roy Choudhury, Vaibhavi A. Shet (2022). Categorical Classification of Personality Using Ocean Model Based on Alcohol Consumpton, The International Journal of Indian Psychology, ISSN 2348-5286

https://www.enjoyalgorithms.com/blog/personality-prediction-using-ml

www.javatpoint.com

https://www.geeksforgeeks.org/

https://i2.wp.com/farthertogo.com/wp-content/uploads/2018/01/5-Factor-Model-graphic_001.jpg