Big Five Personality Prediction Using Machine Learning Algorithms

Prof. Dhanraj Jadhav [#1], Tejas Bakade [#2], Siddhant Deshpande [#3], Omkar Pethe [#4], Shreyas Kale [#5], Yashwant Madane [#6]
JSPM's Rajarshi Shahu College of Engineering

Article Info
Page Number: 1128 – 1133
Publication Issue:
Vol. 71 No. 3 (2022)

Abstract

With the tremendous increase in use of Machine Learning and Data Science, it becomes easier to collect and process the data related to human personality and behavior. Personality prediction systems and recommendation models allow for better discovery of human moods and interests. The personality trait is an internal attribute that may be derived from a pattern of actions, attitudes, feelings, and habits in a person. Extraversion, Agreeableness, Openness, Conscientiousness, and Neuroticism are the five broad personality qualities outlined by theory. Earlier research demonstrated an increasing interest in personality prediction among persons working in industries such as career development as well as counselling, and health care, among others. Completing surveys, which take an inordinate amount of time and are seldom, is a modern technique of determining the Big-Five Personality type. This research includes a survey on utilizing machine learning to recognize for big five personality traits, model based on the responses to a series of personality questions.

Keywords: Machine Learning, Personality Prediction, Big Five Model

Article History
Article Received: *12 January 2022*
Revised: *25 February 2022*
Accepted: *20 April 2022*
Publication: *09 June 2022*

Introduction

Machine Learning plays a crucial role in Personality Prediction Systems. Algorithms from ML boosts the process of evaluating the survey that reduces the time required for predicting the personality trait. Recent progress in machine learning reduces the human effort to evaluate the data in the survey and increase the efficiency of human personality prediction. Considering the experimentation view, researchers are exploring various domains and trying to create the blueprints to predict the human personality trait. In this survey paper, we are focused on predicting the personality trait from the answers given by each person in a question set of 50 questions based on human behavior. The set of questions processed the Big Five Personality Factor Model was used to assess the personality attribute.

Big Five Personality Traits

Related Work

For Woodworth Psychoneurotic Inventory, sometimes regarded during World War I, it was created as the first personality test. It was employed by the US military to screen army recruits. Post-Traumatic Stress Disorder (PTSD) was the focus of the screening (PTSD). It was common to employ communication Model for Processes (PCM). Taibi Kahler developed PCM, which was used to pick astronauts.[1].

Text, aural, and visual signals are the most basic approaches for determining personality. When working with textual data, data preparation is a critical step that has a direct impact on the outcome. Textual characteristics are often taken from Support Vector Machines (SVMs), as well as Nave Bayes Classifier (NBC), and other Machine Learning models are used to process raw data. Vectors are used to express word embeddings (Word2Vec, Glo Ve, and so on), then further approaches are employed to produce results.

The stream-of-consciousness as well as MyPersonality datasets were utilized by K. C. Pramodh et al. in 2016 [2]. My-Personality dataset is made up of roughly 10,000 Facebook Status Updates from 250 people. For their model and their research, they employed Natural Language Toolkit. The F1-scores for traits OPN are 0.665, CON is 0.632, EXT is 0.625, AGR is 0.624 and NEU is 0.637.

Using Convolutional Neural Networks, Action Unit Recognition Systems, and face detection, Carles Venture's experiment classified personality traits an according to Big Five Personality Model. In their study, they utilized dataset initial Impressions., which contains over 10,000 distinct films.

In 2018, Gokul K. and college used a Bayesian network to categories with a classifier personality for extroverts as well as introverts. As a dataset, they applied self-recorded audio samples. Using Voice Activity Detection and Attribute Selection to reduce dimensionality, they utilized the Auditory Nerve Modeling algorithm and 88.3 percent accuracy was attained.

B. Simoski developed his Social Contagion Model by using twenty-five people's replies to the Big Five Questionnaire to define personality on Big Five Model.

Student Life is a dataset that entails data from a group of 48 students. Abir Abyaa, in 2018, employed supervised learning algorithms to categories people's personalities types based on the basis of Big Five model Random Forests (RF), Logistic Regression (LR), Support Vector Machines (SVM), and c4.5 Decision Trees are all available in this mode, a K-nearest Neighbors were used.

Based on the Big Five Model, Willy has implemented C4.5 Decision Trees in 2019. A dataset of approximately 110 million tweets per day was obtained by using the Twitter API. A 64.30% accuracy rate was achieved.

In 2019, Tao Hong conducted a study on detecting personality traits and classifying sentiment. They applied MDSTC dataset speech, facial emotions, as well as Skin Response are all part of this to classify personality using Deep Neural Networks. They outperformed state of art models for terms in performance metrics.

Big Five Model was used by Imanuel Buhapoda Drexel to detect personality based on Gaussian NAB classifier. For cross validation, they used Fasttext word embeddings, Word2Vec embeddings, as well as k fold cross conformation k=5 is all available.

Using a Multiview Learning Approach, Songcheng Gao in 2020 has developed a mobile application that acquires Student Life datasets and a MultiTask approach. Data was collected

from 183 students from two distinct academic institutions. For data gathering methods are done through a questionnaire completed through a mobile application. The performance measure is root mean absolute error as well as mean squared error.

The parts that follow are describe in detail in methodology we used for create an model based question set for the Big Five Personality Factor Model.

Summary Paper:

Name, Year & Reference	Dataset	Summary	Gaps/Challenges Future
In the Age of Machine Learning, Personality examination and evaluation August 2019 C. Stachl, Gabriella M. H. R. Schoede	Phone study mobile sensing dataset	A number of key methodological issues have been considered. We examined the challenges and There are certain possible drawbacks to using machine learning model that must be considered.	Challenges: To improve performance accuracy Applying precious algorithm to model
Behavior and Personality Analysis in a non-social context Dataset June 2018 Dario Dotti Mirela Popa Stylianos Asteriadis	Novel Dataset	Using a dataset to find a machine learning model for behavior interpretation and personality recognition in an unconstrained indoor scenario in a smart home setting.	Challenges: Providing Findings on the relationship between nonverbal behavior and personality types.
Recent developments in personality-based deep learning prediction, 2020, Y. Mehta, N. Majumder, A. Gelbukh, E. Cambria		Personality characteristics are estimated using an in-depth learning algorithm based on the Big-Five factor model and MBIT data.	

Observation On Survey:

Using different ML techniques and algorithms it is observed that prediction of personality is possible. Considering the current scenario, automated personality detection has a lot of practical potential. In the next few years, this paper describes many of the in-depth learning models discussed in this study are expected to be extensively used in industrial settings. A large set of publicly available datasets will be needed for training these models. Deep learning models for automated personality detection have seen a tremendous increase in popularity. Deep neural networks have been found to perform best in most modalities. Deep learning

models will continue to improve Perhaps in the long term as smart and more efficient algorithms follow.

Techniques:

In order to achieve a certain goal or result, various approaches and technologies are used for effective treatment. Survey shows that these techniques include some ML Decision trees, neural networks, and k-means clustering etc. In many cases, a few deep learning techniques are often used in order to achieve increased precision such as neural networks. These techniques really play well in terms of contributing to a good project.

1. **Decision Tree:**

Through regressing the decision tree's categorization rules to a list of un-ruled and unordered variables, examples, the decision tree (DT) could be utilized as a form of instancebased learning. Using a top-down recursive method, it compares property values at each interior node of the decision tree, makes judgments based on the numerous attributes' values, grows branches from each interior node, and arrives at a decision at each as =>Figure 1.1.

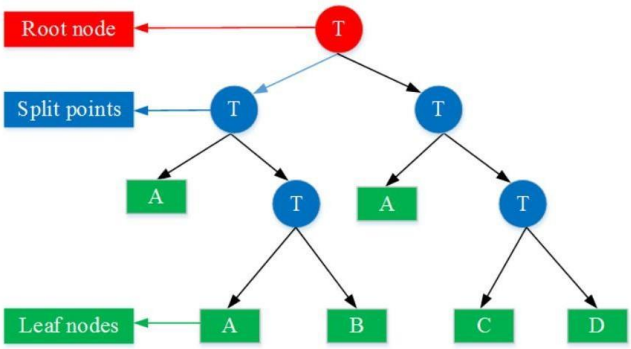

Figure 1.1: General form of decision tree

An expression rule of conjunction corresponds to the path between a root node and a leaf node, and a collection of rules for discontinuous expressions is homologous to the decision tree. One of the hugest benefits of decision trees that is they require little prior knowledge of the user in order to learn, but they also have a number of drawbacks.

2. **Neural Networks:**

Neural Networks is a technique in which the algorithm tries to detect the underlying relations between the data with a process that is made to act like a human brain.

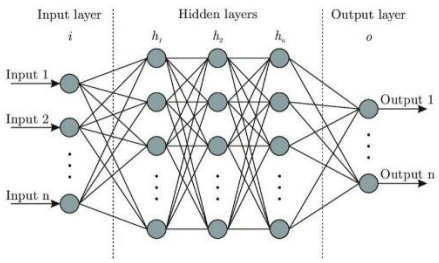

Figure 1.2: General form of Neural networks

=>Figure 1.2: The number of input neurons you use to build your neural network. One input neuron is required for each feature in the input vector. The number of characteristics in your dataset that are connected are the number of tabular features. To avoid overfitting, closely choose these variables, excluding any that involve patterns that will not generalize outside the training sample. In the case of MNIST, this is the dimension of your image (28*28=784). You can specify how many predictions you want to make in the output neurons. Regression [13]: For tasks of regression, it is a one possible value for example house price. One neuron per predicted value is used in multivariate regression (e.g., a task that has three output variables will require a neural network output layer with three nodes in the output layer, each with the linear activation function).

Classification [14]: We use one output neuron per positive classification (spam versus non-spam), where the probability of the affirmative class is represented by the output. On the output layer, we apply a soft-max activation function to verify that the probabilities add up to one when a class is composed of more than one output neuron (For example, in object identification); an instance may be categorized as a vehicle, an animal, or a building, among other things.

Conclusion:

This paper pointed out that automated personality detection has a wide range of applications It might be put to use in the workplace; thus, this field is very hot trending and upcoming the data used to train a machine learning model determines its effectiveness, though. For many applications in this field, it is not possible to collect enough labelled data to train a neural network. For the detection of personalities, it is critical to have vast, accurate, and diverse datasets. There are currently very few studies examining other personality measures like the MBTI or PEN. The majority of studies employ Big-Five personality models. Generally, personality is assessed by completing a survey with numerous questions. Usually, questions are asked in surveys to assess personality. But even if everyone taking the survey answers truthfully, its reliability to accurately label a person's personality remains questionable. We need to look into more precise and efficient methods of determining personality types.

References:

[1] N. Majumder, Y. Mehta, E. Cambria, and A. Gelbukh “Recent trends in deep learning-based personality detection,”
[2] Vijayalata and K.C.Pramodh “Automatic personality recognition of authors using big five factor model,”
[3] Gokul and Lalitha, “Personality Identification Using Auditory Nerve Modelling of Human Speech,” 2018
[4] Pawan Kumar Tiwari, P. S. . (2022). Numerical Simulation of Optimized Placement of Distibuted Generators in Standard Radial Distribution System Using Improved Computations. International Journal on Recent Technologies in Mechanical and Electrical Engineering, 9(5), 10–17. https://doi.org/10.17762/ijrmee.v9i5.369
[5] Simoski, and Klein, “Applying machine learning algorithms for deriving personality traits in social network”.

[6] A. Chawla, “Phishing website analysis and detection using Machine Learning”, Int J Intell Syst Appl Eng, vol. 10, no. 1, pp. 10–16, Mar. 2022.

[7] Abyaa, Khalidi Idrissi, and. Bennani, “Predicting the learner’s personality from educational data using supervised learning,” 2018.

[8] Malla, S., M. J. . Meena, O. . Reddy. R, V. . Mahalakshmi, and A. . Balobaid. “A Study on Fish Classification Techniques Using Convolutional Neural Networks on Highly Challenged Underwater Images”. International Journal on Recent and Innovation Trends in Computing and Communication, vol. 10, no. 4, Apr. 2022, pp. 01-09, doi:10.17762/ijritcc.v10i4.5524.

[9] Willy, Setiawan, and Nugraha, “Implementation of Decision Tree C4.5 for Big Five Personality Predictions with TFRF and TF-CHI2 on Social Media Twitter,” 2019

[10] T. Hong, X. Sun, F. Tian, and F. Ren, “Sentiment Classification and Personality Detection Via Galvanic Skin Response Based on Deep Learning Models,” 2019.

[11] I. B. Drexel, “Feature Engineering and Word Embedding Impacts for Automatic Personality Detection on Instant Message,” 2019.

[12] S. Gao, W. Li, L. J. Song, X. Zhang, M. Lin, and S. Lu, “Personality Sensing: A Multi-View Multi-Task Learning Approach for Personality Detection based on Smartphone Usage” 2020.

[13] Ghazaly, N. M. . (2022). Data Catalogue Approaches, Implementation and Adoption: A Study of Purpose of Data Catalogue. International Journal on Future Revolution in Computer Science &Amp; Communication Engineering, 8(1), 01–04. https://doi.org/10.17762/ijfrcsce.v8i1.2063

[14] Artificial Intelligence, Wikipedia

[15] Medical expert systems are born from the necessity of assisting physicians and patients in the diagnosis and treatment of diseases, even as a means through which to make alerts and notify physicians and patients. Science Direct

Authors and Affiliations

Prof. Dhanraj Jadhav
Professor at JSPM’s Rajarshi Shahu College of Engineering
Tejas Bakade
Pursuing Bachelor of Technology at JSPM’s Rajarshi Shahu College of Engineering
Siddhant Deshpande
Pursuing Bachelor of Technology at JSPM’s Rajarshi Shahu College of Engineering
Omkar Pethe
Pursuing Bachelor of Technology at JSPM’s Rajarshi Shahu College of Engineering
Shreyas Kale
Pursuing Bachelor of Technology at JSPM’s Rajarshi Shahu College of Engineering
Yashwant Madane
Pursuing Bachelor of Technology at JSPM’s Rajarshi Shahu College of Engineering