

Big-Five Personality Prediction

A1: DHANUSHA, MCA DAYANANDA SAGAR COLLEGE OF ENGINEERING

CO-A2: PROF THILOTHAMME, ASST PROF OF MCA, DEPT OF MCA, DAYANANDA SAGAR COLLEGE OF ENGINEERING

ABSTRACT

Desires provide more dedicated and personalized services that fit into individual's need, reckoning on the inner personalities of the user. Personality defines characters which makes unique and important in psychology. Manual efforts to access one's personality intakes huge cost and can't be tired real time. This research aims to live a person's experience in field and also their personality. To predict Personality of person, a psychological analysis called Big-Five Personality is used where the user attempts a collection of predefined questions and these questions answered by the user is used for predicting personality using logistic regression model.

Although personality prediction is crucial for explaining the user's outer form from their inner personality. Big-Five Personality approach aims to predict the person's behaviour from their assessment. The Machine Algorithm used here analyses the personalities supported five different personality traits namely Openness, Extroversion, Agreeableness, Conscientiousness, Neuroticism (OCEAN).

1 INTRODUCTION

Personality is that the particular combination of emotional, attitudinal, and behavioural response patterns of a private in psychological definition (Wikipedia, 2012). Big Five Personality traits theory, personality may be divided into five different dimensions which are openness, conscientiousness, extraversion, agreeableness and neuroticism. Openness is expounded to imagination, creativity, curiosity, tolerance, political liberalism. Conscientiousness for a spontaneous one. Extroversion for tendency to hunt stimulation within the external world and to specific positive emotions. they're outgoing, friendly, and socially active. Agreeableness focuses for maintaining positive relations, compassionate and cooperative.

Neuroticism refer to degree of emotional stability, impulse control, and anxiety.

Above mentioned traits are the commonly used model of personality in psychology. The test consists of 20 questions that the user rates on how true they're about them on a 3-point scale where Agree, Neural and Disagree values 0, 1, and 2. Personality is predicted with the obtained marks by the user. For the prediction purpose, a training dataset including marks and an adjective for a personality is fed to the model for training. Here, we use Logistic Regression classifier model.

2 LITERATURE SURVEY

In 2014 associate Integrated E-Recruitment System for machine-control Personality Mining Ranking was planned by Faliagka, an automatic candidate ranking was implemented by this method. It had been supported objective criteria that the candidate's details would be extracted from the candidate's LinkedIn profile. The persons temperament traits were mechanically extracted from their social presence using linguistic analysis. The candidate's rank was derived from individual choice criteria victimisation Process (AHP), whereas their weight was controlled by the recruiter (admin). the constraints of the system were that senior positions that needed experience and sure qualifications were screened inconsistently. Liden et al. revealed the overall issue of Personality: The interrelations among the large 5 temperament factors (Openness, Conscientiousness, sociability, Agreeableness, and Neuroticism) were analyzed during this paper to check.

3 METHODS

LOGISTIC REGRESSION:

Logistic Regression is used when the dependent variable (target) is categorical. Logistic regression transforms its output using the logistic sigmoid function to return a probability value which can then be mapped to two or more discrete classes. [5] In view of the quantity of classifications, Logistic relapse can be delegated:

(A) Binomial: target variable can have just 2 potential sorts: "0" or "1". (B) Multinomial: target variable can have at least 3

potential sorts which are not ordered. (C) Ordinal: it manages target factors with requested classes. Here, every class can be given a score like 0, 1, 2, 3.

3.1 DATA COLLECTION

The assessment consists of questions based on each Big-Five traits from each category. Questions are framed to according to psychological analysis. The user rates on 3 point scale about how true they're about them where Agree, Neural and Disagree values 0, 1. Personality is predicted consistent with the obtained marks of the user which is evaluated and converted into percentage for each Big-Five Personality Traits. And a class label for personality is evaluated and given. These label can be serious, dependable, lively, extraverted, responsible. For the prediction purpose, a training dataset including marks and an adjective for a personality is fed to the model for training.

We also use the Pickled File saving method to save and reload machine learning model. The logistic regression classifier saves the data to a file, and then loaded the data into the model, so that the model doesn't need to be trained each time when the model is loaded. The loaded model is then used to calculate the accuracy score and predict outcomes on new unseen (test) data. This saves and restores machine learning models quick with just two lines of code. And it optimizes the model's parameters on the training data.

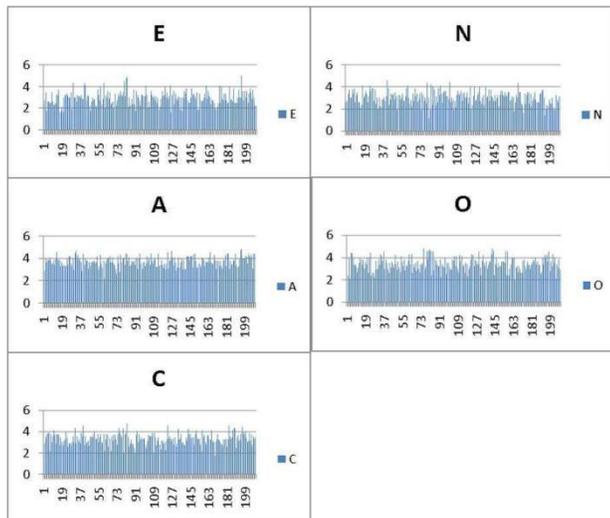


Figure 1: Personality Score Distribution

In order to train our system using classification methods in machine learning, we need to do the discretization on the initial scores and use the discrete values as data labels. [6] The discretization functions we used are shown below:

$$\alpha = E(x) - \sigma(x), \beta = E(x) + \sigma(x),$$

For each dimensions, it means that we separate the label scores into three classes, low-score group 1 to α , middle-score group α to β and high-score group β to 5, where $E(x)$ is the mean value of personality score for dimension x , $\sigma(x)$ is the Standard variation of dimension x and x is chosen from O, C, E, A, N the five dimensions for personality. [7]

4 MODEL TRAINING

Logistic regression model used here gives accurate prediction of a person. The data is classified into test and train data set. Train data set train with model and test data set is used for testing and predicting the personality. Foremost CSV format specified data is assigned to train data set and the values are extracted into array. This array is loaded into data-frame with values and classifier variable is assigned with logistic regression with multi nominal target variable. Classifier is opened as pickled file and prediction is obtained.

Traits	Result
Openness	87.5 %
Extroversion	75.0 %
Agreeableness	75.0 %
Conscientiousness	87.5 %
Neuroticism	75.0 %

Table 1: Prediction values

5 DISCUSSION

From the above values, it is easy to predict that different attribution has different weight and different dimension has different high-weight attribution. The difference of these five personality with their behaviour performances which are evident that they are correlation of behaviour and personality. The personality of this person is dependable, they mainly focus what other people may think of them and agreeable.

5.1 RESULTS ANALYSIS

From the table, we can draw a meaningful conclusions. Traits as agreeableness refers to being helpful and sympathetic towards others. People with high scores in agreeableness tend to be dependable on others. Therefore, their self-comment proportion is relatively high. Also in order to get the attention of others.

Conscientiousness people are disciplined, organized, and achievement-oriented. People having a high score in this dimension tend to be helpful for others and responsible they account for.

Extraversion is displays high degree of sociability, assertiveness, and talkativeness. An extroverted person, may tend to portray their character with emotional perception.

Neuroticism refer to degree of emotional stability, impulse controllability. People in this dimension tend to be easily angry for other things. Therefore, their seriousness is shown towards others and they may be relatively show angry proportions towards people.

Openness is strong intellectual curiosity and a preference for novelty and variety. People that are curious with others tend to make many new friends. They tend to be lively as possible, maintaining a healthy co-relation with others.

5.2 CONCLUSION

Personality predicting opens a dimension in acknowledging new outer perspective of a person and depicting psychology. Although many machine-learning models may use, logistic regression model shows effective performance since the target variable is known and their values can be easily used in evaluation.

5.3 FUTURE ENHANCEMENT

Further research in this personality prediction may involve an approach with multi-target variable and also unlabelled data as training data. With deep machine learning technique, we can analyse various other possible factors that helps in predicting the personality. Targeting multiple variable enhances in clear prediction. This opens a window in computer science and psychology also. We recommend resources for extending the understanding of the person based on their conversions. With objective data, trait's selection strategy is given to psychological experiments will increase their quality levels. With upgradation, system will be accessed online to treat any kind of acknowledgements. For extroverted patients, it helps in understanding their weak areas. By considering

the correlation between these five personality dimensions it makes them orthogonal. To fix this problem, with the help of multi-target variable learning techniques to fix our training algorithm.

REFERENCES

- [1] <https://www.kaggle.com/prmohanty/python-how-to-save-and-load-ml-models>
- [2] Big 5 personality test: <https://openpsychometrics.org/tests/IPIP-BFFM/>
- [3] <https://partheniumprojects.com/personality-prediction-system-through-cv-analysis/>
- [4] <https://github.com/jcl132/personality-prediction-from-text>
- [5] <https://www.quora.com/How-is-predictive-modeling-used-in-logistic-regression>
- [6] <https://stackabuse.com/scikit-learn-save-and-restore-models/>
- [7] <https://arxiv.org/pdf/1204.4809>
- [8] <https://openpsychometrics.org/tests/IPIP-BFFM/>
- [9] <https://www.geeksforgeeks.org/saving-a-machine-learning-model/>
- [10] <https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc?gi=bfe01db01ede>
- [11] The Big Five Personality Test: from personality-testing.info courtesy [ipip.ori.org](https://openpsychometrics.org/printable/big-five-personality-test.pdf)