

RESUME SCREENING WITH PERSONALITY PREDICTION

R.THENMALAR
Assistant Professor
Dept of Computer Science &
Engineering
RVS College of Engineering &
Technology,
Coimbatore
thenmalarcbe@gmail.com

YUVRAJ TRIPATHI
712819104720
Dept of Computer Science &
Engineering
RVS College of Engineering &
Technology,
Coimbatore
stripyuvi@gmail.com

K. SOWMIYA
712819104027
Dept of Computer Science &
Engineering
RVS College of Engineering &
Technology,
Coimbatore
asowmiyakani@gmail.com

S. CHEZHIYAN
712819104009
Dept of Computer Science &
Engineering
RVS College of Engineering &
Technology,
Coimbatore
cheliyan.sa@gmail.com

A.SARANYA
712819104723
Dept of Computer Science &
Engineering
RVS College of Engineering &
Technology,
Coimbatore
ss.sarasuru@gmail.com

Abstract— The personality of a human plays a major role in his personal and professional life. Many organizations have also started short listing the candidates based on their personality as this increase efficiency of work because the person is working in what he is good at than what he is forced to do. The project is based on identifying the personality of an individual using machine learning algorithms and Model classify the personality. The prediction of the personality of an individual is a critical problem in both areas whether it is considered in the context of organizations or in the case of our daily lives. Prediction of personality depends on many factors and these factors may vary from one individual to another. Personality prediction is identifying the personalities of individuals through their actions in different situations and observing their behaviours in various circumstances. Five characteristics of different individuals commonly known as big five characteristics namely, openness, neuroticism, conscientiousness, agreeableness and extraversion are stored in a dataset along with gender and age of individual and used for training. Before training the model, data is preprocessed like handling missing values, data discretization, standardization etc. This preprocessed data is then used to train the model. User rates himself for different behavioural characteristics and based upon the information provided by user his/her personality is predicted using trained ML model.. Personality traits show the different characteristics of different people based on their thoughts. The accuracy of personality prediction achieved by using Logistic regression Classifier is 75.25%. we refer to as personality types. In this Project people with similar personalities are grouped together based on identifying personality model.

Keywords—Machine Learning, Personality Prediction, Resume screening, Logistic Regression.

I. INTRODUCTION

Predicting the personality of a human has always been a difficult task for every individual and when it comes to the recruitment process, it is strenuous for interviewers to determine the actual personality of the interviewee and now it deals with the scenario of online interviews and so it becomes more difficult. Two categories can define the personality of a human better and those are verbal and non-verbal. These categories contain some critical factors such that verbal includes communication skills, use of specific phrases or facial expressions, etc. Non-verbal includes the posture, speech tone, etc. of an individual. There are some more important factors for identifying the personality like the

handwriting of an individual, social media activities such as updating posts, profile picture, reaction on others' posts, and resume analysis by interviewer or HR.

We can predict the of an individual bases on this five personality traits.

Openness : To encounter is an overall admiration for craftsmanship, feeling, experience, abnormal thoughts, creative mind, interest, and assortment of involvement. Individuals who are available to encounter are mentally inquisitive, delicate to magnificence and tend to attempt new things. In addition, people with high transparency are said to seek after self-completion explicitly by searching out something serious. On the other hand, those with low openness try to acquire satisfaction through diligence and are portrayed as logical and information driven—here and there even apparent to be one sided and shut leaning. Some conflict stays about how to decipher and investigate the openness factor.

Conscientiousness : Is a propensity to show self-restraint, behave obediently. It is identified with the manner by which individuals command and organize. High conscientiousness is regularly seen as being difficult and centered. Low conscientiousness is related to adaptability and immediacy, however can likewise show up as messiness and absence of loyalty. High outcomes on conscientiousness demonstrate an inclination for arranged as opposed to unconstrained nature. The normal level of conscientiousness ascends among youthful grown-ups and afterward decreases among more established adults.

Extraversion (or extroversion) : Is portrayed by volatility, friendliness, forceful, gossipy, and massive eagerness. Those persons who are high in extraversion are friendly, and generally obtain energy in friendly circumstances. Being around others motivate them to feel energized and invigorated. Generally, persons who have low extraversion (or thoughtful) quality will be more held and have less energy to spend in friendly circumstances. Public events can feel depleting and thoughtful people need some amount of time of loneliness oftenly and calm to “re-energize”

Agreeableness: In a live interaction of a person with others, characterized by the degree of compassion and co-operation. This trait has attributes like trust, graciousness, selflessness, feeling and alternative prosocial behaviors.

Persons who have this trait as a higher value are to be additional agreeable. Generally, those having low in this quality will be extra serious and normally even calculating.

Neuroticism: An individual with an undeniable degree of pleasantness in a character test is generally amicable, careful and warm. They have a hopeful perspective on human instinct and coexist well with others. An individual who scores low on pleasantness may place their own advantages over those of others. They will generally be threatening, uncooperative and inaccessible.

II. RELATED WORK

All the correlated works that have been done that are related to the current problem are follows. [1] F. Celli, E. Bruni, and B. Lepri. Automatic Personality and Interaction Style Recognition from Facebook Profile Pictures. [2] Personality Prediction System from Facebook Users Author links open overlay panel TommyTanderaHendro, DerwinSuhartono ,RiniWongso, Yen Lina, Prasetio ,Computer Science Department, School of Computer Science, Bina Nusantara University, Jl. K. H. Syahdan No. 9 Kemanggisian, Jakarta 11480, Indonesia. [3] R. McCrae and O. P. John. An Introduction to the Five-Factor Model and its Applications. J. Pers., 60(2):175 -- 215, 1992 [4] Soto, Christopher J., Anna Kronauer, and Josephine K. Liang. "Five- Factor Model of Personality." The Encyclopedia of Adulthood and Aging (2015): 1-5. [5] Wijaya A, Febrianto N, Prasetia I, Suhartono D. SistemPrediksiKepribadian "The Big Five Traits" Dari Data Twitter. Jakarta: Bina Nusantara University, School of Computer Science; 2016. [6] D. Markovikj, S. Gievska, M. Kosinski, and D. Stillwell, "Mining facebook data for predictive personality modeling," in Proceedings of the 7th international AAAI conference on Weblogs and Social Media (ICWSM 2013)

III. PROBLEM DEFINITION

Predicting personality has many applications in real world. Use of social media is increasing day by day. Huge amount of textual data as well as images continue to explode to the web daily. Current work focuses on Linear Discriminate Analysis AdaBoost over Twitter standard dataset. Disadvantage :

- Thus, according to Existing results it is found that AdaBoost has low level accuracy
- Boosting technique learns progressively, it is important to ensure that you have quality data. AdaBoost is also extremely sensitive to Noisy data and outliers so if you do plan to use AdaBoost then it is highly recommended to eliminate them.
- AdaBoost has also been proven to be slower

IV. PROPOSED SYSTEM

This Machine Learning algorithm Logistic Regression used for the classification problems that uses the concepts of probability to do the predictive analysis. This algorithm limits the cost function which is multinomial function between 0 and 1. It overcomes the problem of Logistic Regression as it can have a value larger than 1 or smaller than 0. Sigmoid function is given as follow, 23 In this stage of the Personality Prediction, the model is producing the output as different personalities of different individuals which have been processed by the Intelligent Agent. These are the different

salient stages that are used in the Personality Prediction. In this general architecture, the input for the intelligent agent is an audio file as well as a text file which makes this a model. This modal provides better results than other models.

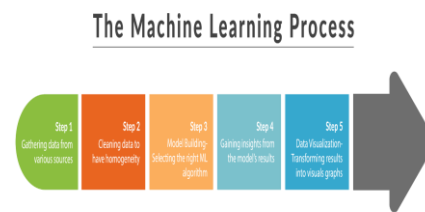


Fig 1. Machine learning process

A. Data Collection

- Dataset on various movement parameters are collected to train the model and predict outcomes. These values are arranged in a spreadsheet for better access.
- Open to Experience: It involves various dimensions, like imagination, sensitivity, attentiveness, preference to variety, and curiosity.
- Conscientiousness: This trait is used to describe the carefulness and diligence of the person. It is the quality that describes how organized and efficient a person is.
- Extraversion: It is the trait that describes how the best candidates can interact with people that is how good are his/her social skills.
- Agreeableness: It is a quality that analyses the individual behavior based on the generosity, sympathy, cooperativeness and ability to adjust with people.
- Neuroticism: This trait usually describes a person to have mood swings and has extreme expressive power.

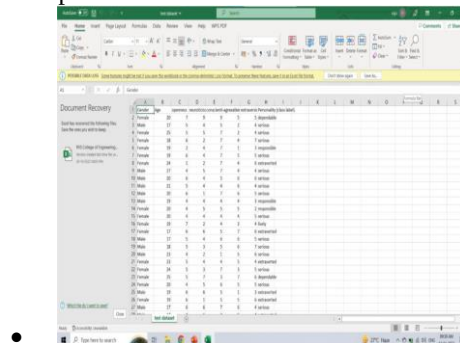


Fig 2. Data Collected in Spreadsheet

B. Data Pre Processing

Information is collected from various sources and stores in spreadsheet in raw format. The dataset might have null values, undesigned spaces, unknown formats. These variables confuse the ML model and isn't suitable for analysis. Data from spreadsheets should be extracted through a variable to the ML model for testing and training.

Preprocessing techniques are used in order to refine data and make it suitable for the algorithms to perform tasks.

Performance Evaluation is then done on this model and confusion matrix is obtained.

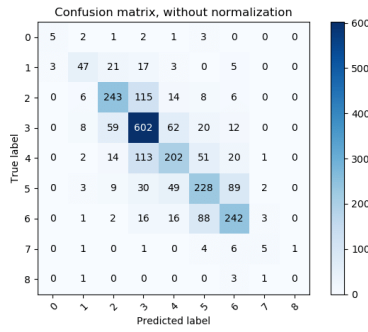


Fig 7. Confusion Matrix of Logistic regression after Performance Evaluation

C. Decision Tree Classifier

Decision tree is a supervised learning technique that can be used for both classification as well as regression. The internal nodes represent the features of a dataset and branches represent decision rules and each leaf node the outcomes. Decision trees usually act like human thinking which makes it easy to understand.

It is very useful for solving decision related problems. We fit the model to the training set by importing **DecisionTreeClassifier** from **sklearn.tree** library

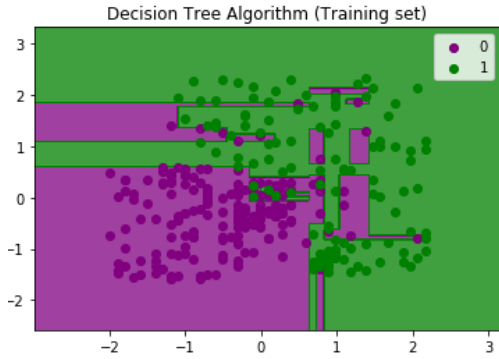


Fig 8. Visualizing Training Set Result using Decision Tree Classifier

Performance Evaluation is then done on this model and confusion matrix is obtained.

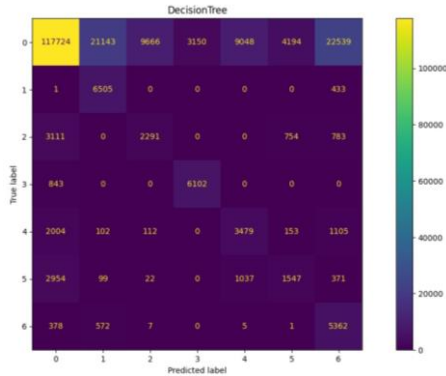


Fig 9. Confusion Matrix of Decision Tree Classifier after Performance Evaluation

VI. FINALIZED MODEL

After the completion of performance evaluation of the proposed three models, and comparing the results of performance from [Fig 3, Fig 5, Fig 7, Fig 9], it is shown that **Logistic Regression** works the best with peak performance, precision, recall and F1 scores with least number of False positives and False negatives when executed in our Dataset compared to other proposed classification models.

The Decision Tree classifier helps to think about all the possible outcomes for a problem. There is less requirement of data cleaning compared to other algorithms. Hence the finalized model is stored in a pickle file and is accessed later.

Pickel is used to serialize python object structures, which refers to converting object which is in memory to byte stream that is stored as binary file on disk. When we load it, this file is de-serialized back to python object. Serialization refers to the process of converting an object in memory to byte stream that is stored on a disk or sent over a network.

VII. SOFTWARE REQUIREMENTS

A. Python IDLE 3.11.1

IDLE is an integrated Development Environment for python. Python3.11.1 is the newest version of the Python programming language and it contains many new features and optimizations including fine-grained error locations, atomic grouping, etc.

B. Sci-Kit Learn

Sci-Kit Learn is an open source data analysis library, and the gold standard for Machine Learning in python ecosystem. Key concepts and features include algorithmic decision making methods including classification, regression, etc. It provides simple and efficient tools for predictive analysis and is built on NumPy, SciPy, Matplotlib.

F tkinter

Python tkinter is a standard GUI (Graphical User Interface) package for creating desktop applications with a graphical interface in Python. It provides a set of tools for creating graphical user interfaces and event-driven programming. With tkinter, developers can create windows, buttons, labels, and other GUI elements with just a few lines of code. It also supports different types of widgets, such as canvas, text box, radio buttons, checkboxes, and more.

VIII. OUTPUT SCREENS

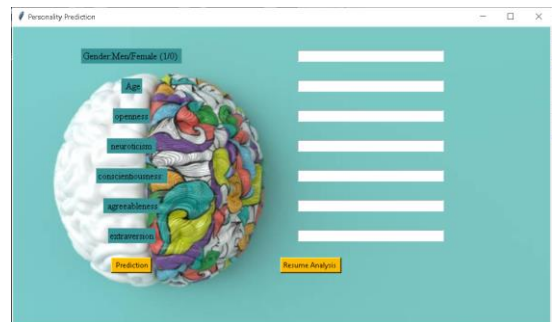


Fig 10. User Home screen

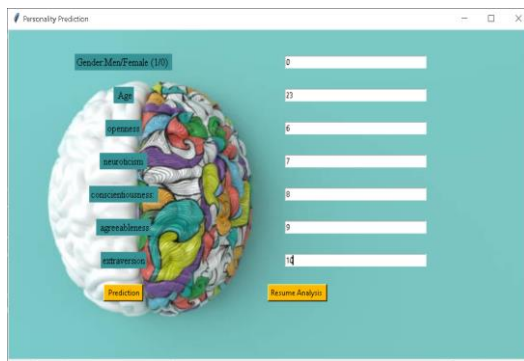


Fig 11. Inputs for prediction.

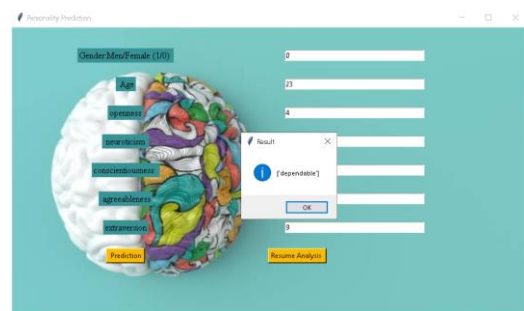


Fig 12. Personality Prediction Output

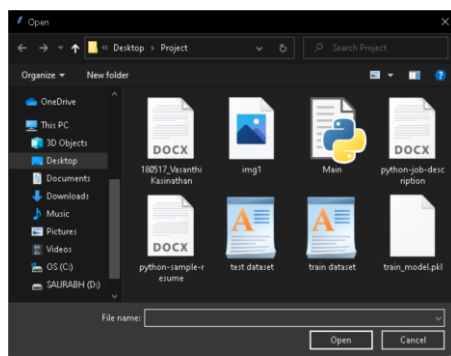


Fig 13. Data Collection For Resume screening

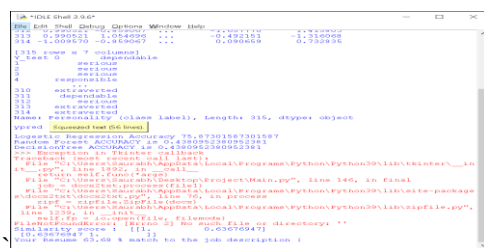


Fig 14. Output for resume screening.

REFERENCES

- [1] F. Celli, E. Bruni, and B. Lepri. Automatic Personality and Interaction Style Recognition from Facebook Profile Pictures. In Proc. ACM Multimedia -MM '14', pages 1101--04. ACM, 2014.
- [2] Personality Prediction System from Facebook Users Author links open overlay panel TommyTanderaHendro, DerwinSuhartono ,RiniWongso, Yen Lina, Prasetyo ,Computer Science Department, School of Computer Science, Bina Nusantara University, Jl. K. H. Syahdan No. 9 Kemanggisian, Jakarta 11480, Indonesia
- [3] R. McCrae and O. P. John. An Introduction to the Five-Factor Model and its Applications. J. Pers., 60(2):175 -- 215, 1992 [4] Soto, Christopher J., Anna Kronauer, and Josephine K. Liang. "Five- Factor Model of Personality." The Encyclopedia of Adulthood and Aging (2015): 1-5.
- [4] Wijaya A, Febrianto N, Prasetya I, Suhartono D. SistemPrediksiKepribadian "The Big Five Traits" Dari Data Twitter. Jakarta: Bina Nusantara University, School of Computer Science; 2016
- [5] D. Markovikj, S. Gievska, M. Kosinski, and D. Stillwell, "Mining facebook data for predictive personality modeling," in Proceedings of the 7th international AAAI conference on Weblogs and Social Media (ICWSM 2013), Boston, MA, USA, 2013..
- [6] Hassanein, M., Hussein, W., Rady, S. and Gharib, T.F., 2018, December. Predicting personality traits from social media using text semantics. In 2018 13th International Conference on Computer Engineering and Systems (ICCES) (pp. 184-189). IEEE
- [7] Sumner, Chris, et al. "Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets." 2012 11th international conference on machine learning and applications. Vol. 2. IEEE, 2012.
- [8] Qaiser, Shahzad, and Ramsha Ali. "Text mining: use of TF-IDF to examine the relevance of words to documents." International Journal of Computer Applications 181.1 (2018): 25-29.
- [9] Kunte, Aditi, and SujaPanicker. "Personality Prediction of Social Network Users Using Ensemble and XGBoost." Progress in Computing, Analytics and Networking. Springer, Singapore, 2020. 133-140.
- [10] Aydin, Berkay, et al. "Automatic personality prediction from audiovisual data using random forest regression." 2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, 2016
- [11] Youyou W, Kosinski M, Stillwell D. Computer-based personality judgments are more accurate than those made by humans. In National Academy of Sciences; 2015. p. 1036-1040.