

Personality Prediction Using Sklearn

Sumair Qadir¹, and Monika Mehra²

¹ M.Tech Scholar, Department of Electronics and Communication Engineering, RIMT University Mandi Gobingrah, Punjab India

² Professor, Department of Electronics and Communication Engineering, RIMT University Mandi Gobingrah, Punjab India

Correspondence should be addressed to Sumair Qadir; sumairzargar1702@gmail.com

Copyright © 2022 Made Sumair Qadir et al. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT- Personality is one characteristic that influences how people interact with the outside environment. A person's personality can be regarded as an essential component of their behaviour. People's personalities are determined by how they connect with others. This article discusses Automated Personality Classification, which is a system that uses Data Mining Algorithms to analyse a user's personality based on specific parameters. In this study, we present a technique for analysing an applicant's personality. This technique will be useful for businesses and other organisations who want to hire people based on their personalities rather than their technical skills. The Big Five Personality qualities are used to predict personality traits, and the categorization is done using the Nave Bayes Algorithm and Support Vector Machine.

KEYWORDS- Nave Bayes Algorithm, Support Vector Machine, Automated Personality Classification, Data Mining

I. INTRODUCTION

In recent years, personality categorization has been one of the most explored issues. Personality is the sum of an individual's actions and qualities that affects how he or she behaves in various situations. Personality may impact a person's choices in different areas such as literature, clothing, music, and movies [1]. His or her personality can also influence how he or she interacts with the outside world and his or her surroundings. Personality may also be utilised as a differentiator during the hiring process, career counselling, and health counselling, among other things. Analyzing a person's behaviour to predict their personality is an old practise. This manual approach of personality prediction took a long time and a lot of effort. Analyzing personality based on one's nature was a time-consuming activity that needed a great deal of human work. This ancient approach of determining personality took a long time and had a small sample size Furthermore, when analysing a user's personality analysis did not produce reliable findings. Because people are prone to biases and preconceptions, the accuracy of the results was harmed

because the analysis was done manually. As a result, data mining techniques are used to examine and analyse data in order to uncover any hidden patterns or information in a huge data collection. These methods are employed to extract user attributes and then train the model to predict the personality of future users we may analyse the personality of a candidate looking for a job in a company that values behaviour and personality above technical competence using these strategies. In addition, the candidate learns which personality qualities he or she possesses and which traits are lacking As a result, we may help him/her develop those attributes or reinforce the others as needed The primary goal of this paper is to provide an overview of the data mining methods that are used to forecast a user's personality. In this study, we focus on an online exam that the candidate would take, after which his or her personality would be predicted using the Big Five Personality Traits. This allows us to eliminate people who are seeking for a specific position inside the company. As a result, the organization's resources would be saved, and they would only interview those people who were the best fit for the job.

II. ABOUT SKLEARN

Sklearn, sometimes known as scikit learn, is a Python package that includes a variety of unsupervised and supervised learning algorithms. It's utilised extensively throughout the bank for categorization, predictive analytics, and a variety of other machine learning applications. In Python, Scikit-learn (Sklearn) is the most usable and robust machine learning package. It uses a Python consistency interface to give a set of fast tools for machine learning and statistical modelling, such as classification, regression, clustering, and dimensionality reduction. NumPy, SciPy, and Matplotlib are the foundations of this package, which is mostly written in Python.

III. A QUICK REVIEW OF THE LITERATURE

New techniques to automated personality classification: Concepts and Prospects. This publication proposes a

number of fresh studies on the idea of Automated Personality Classification is a method of classifying people based on their personalities. To begin with, all options are considered explored and what improvements to the current challenges of Automated Personality Classification may be achieved. Then there's the Dynamic APC, which is an extension of the Automated Personality Classification [APC] issue, taken into account, as well as how to eliminate inconsistencies in textual data This study was conducted entirely in the setting of social networks and data mining technologies [2].

A. Psychopathy Prediction Using Twitter Content

A large number of people use Twitter, a microblogging service, to share their experiences and ideas about their daily lives. Despite the fact that academics have frequently dismissed the approach of predicting personality by analysing tweets because they believe it includes too little evidence to forecast meaningful information, these tweets may be integrated to construct a bigger image of the user who is sending them. Select RUSBoost [3], a novel type of ensemble learning that combines four classification learners and four feature selection approaches, has been applied to predict psychopathy via Twitter.

B. B. Play an educational game (Detecting personality of players in an educational game)

Education Data Mining [4] tries to provide methods for predicting student behaviour using data from chat interactions, class discussions, and other sources. Individuals behaviour and personality were utilised for analysis in approaches like Intelligent Tutoring System (ITS) and Educational Data Mining (EDM). As a result, a user-adaptable system is created that uses approaches such as Intelligent Tutoring System and Educational Data Mining to analyse student behaviour during interactions.

C. Detection System for Personality and Happiness

This [5] study offered a technique for predicting personality and personality trait a user's happiness The authors created a method based on Eysenck's theory of human personality, in which they collected text messages from social media sites like Facebook and Twitter and then classified them into distinct personality types. Although a direct relationship between personality traits and happiness has yet to be demonstrated, there is a correlation that may be discovered in the near future.

D. Using Social Media to Predict Personality

Social media [6] is a platform where individuals show themselves to the outside world by sharing personal information and allowing others into their life. Personality has a vital role in a variety of human interactions it may be used to predict work satisfaction, professional achievement, and the success of love relationships. Until now, taking a personality test was the only way to reliably estimate a user's personality. However, gathering data from social media sites made this unfeasible, making accurate personality analysis difficult.

E. A data-mining analysis of the efficacy of online learning

Because of technical advancements that allow for conversations even from a distance, online learning [7] has become quite popular. The majority of research show how effective online learning has aided students in improving their learning ability while also measuring the learning process. It's feasible because to data mining, which allows us to evaluate students' online learning experiences based on their log files. However, the created predictive model suggests that students must put in more effort to be good silent learners.

IV. WORK TO BE DONE

To address the shortcomings of the current approach, we suggested an Automatic Personality Classification system that classifies applicants' personalities using Data Mining techniques. To mine user attributes data and learn from trends, the system employs techniques such as Nave Bayes, Support Vector Machine, and Big Five Model, as well as sophisticated data mining. Based on previous classifications, this learning may now be used to classify/predict user personality. The system examines a wide range of user features and behaviours and creates its own user characteristics patterns in the database based on the patterns discovered. The technology now predicts a new user's personality based on personality data stored from prior user data categorization. This technique is beneficial for predicting candidates' personalities. applications vying for various positions in a company We created an Automatic Personality Classification system that assigns each applicant a unique user name and password. After logging in with his or her user name and password, each applicant completes the survey. The survey has 30 questions, each of which determines one of the Big Five Personality Traits. After the applicant completes the 30-question survey, he or she will be able to see the results of his or her personality. The survey examines the applicant's personality using the Big Five Personality Traits. In the registration page, we also presented a graph depending on the various qualifications supplied by the applicants. The graph then forecasts which candidates are all qualified for a certain job position in a company. The traits of applicants from various streams are grouped using algorithms such as Nave Bayes and Support Vector Machine, and a graph depicting which set of candidates would be best fit for a given role in the business is displayed. This sort of graph will aid in the screening of candidates while also conserving the organization's resources. When the organization sees the graph, it can only call those candidates from a particular stream whose scores are higher than those from other streams As a result, this form of For a company, the method would be both cost-effective and precise in terms of selecting applicants for a certain job.

A. *The Bayes Algorithm (sometimes known as the Nave Bayes Algorithm)*

The Nave Bayes Algorithm, which is a sort of inductive learning algorithm, is extensively used in data mining and is regarded to be one of the most efficient and effective algorithms. Because the conditional independence assumption on which the whole method is based is rarely fulfilled in real-world applications, the performance of the Nave Bayes Algorithm in categorising data is fairly accurate. The Nave Bayes Algorithm[8] is based on the application of Bayes theorem. Multinomial Nave Bayes Algorithm is a version of the Nave Bayes Algorithm that is also used for classification. The Multinomial Nave Bayes Algorithm employs a multinomial distribution in which the number of times a given word appears or the weight of that word is used as a classification feature. The Nave Bayes method combines efficiency (i.e., optimal time performance) with reasonable accuracy (i.e., reasonable accuracy). The Nave Bayes Method's sole theoretical flaw is that it implies conditional independence within the language characteristics. Despite the fact that the Nave Bayes Algorithm provides an oversimplified model, the classification results are remarkably accurate. The findings are saved in our system datasets as either correct or inaccurate. As a result, the system's whole operation is dependent on the associated probability of being correct or incorrect. The steps of the Nave Bayes Algorithm are as follows

- A frequency table is created from the dataset.
- The probability table is constructed by calculating the probabilities of each question.
- The posterior probability of each class is then calculated using naive Bayesian formulae. The prediction's result is the class with the highest probability.

B. *Machine to Support Vectors*

A supervised learning technique called a Support Vector Machine analyses data and detects patterns for classification purposes[8]. To forecast whether a test document is a member of an existing class or not, a set of training data is obtained and then tagged as a component of category. The data set is represented by the Support Vector Machine models as a point in space separated by a line or hyperplane. . The support vector machine approach is based on the premise that if a classifier performs well in the most difficult comparisons, it will almost certainly do much better in the easiest comparisons. Support Vector Machine, a nonlinear classifier, frequently outperforms other classifier approaches in classification. The Support Vector Machine works by non-linearly mapping input data into a high-dimensional space where the data is separated linearly, resulting in accurate classification results.

The following are the stages involved in using a Support Vector Machine:

- Make vectors out of the answers to the questions you've been given.
- Next, determine the vectors' weights.

- Find the highest-valued vectors and the value of personality.
- Finally, determine your personality type.

C. *The Big Five Personality Characteristics*

The Big Five Personality Traits are five personality dimensions or domains that may be used to analyse or forecast a user's personality [9].

The following are the Big Five Factors:

- Imaginative capacity or openness to new experiences
- Agreeableness
- Extraversion
- Emotional Stability vs. Neuroticism
- Conscientiousness

For estimating a user's personality, the Big Five Personality Model is the most frequently acknowledged and researched model. The Big Five Personality Traits may be found in people of all ages, backgrounds, and countries. The Big Five Personality scores are extremely accurate and may accurately determine a user's genuine personality to a great extent.

V. ANALYSIS OF THE RESULTS

A curve is created for candidates with various qualifications based on the survey performed by the applicants. The graph assists the business in weeding out applicants for a certain job position inside the company. This allows the business to screen applicants based on their personality, saving time and money because applicants who are not appropriate for the position are merely filtered out in the first round. This enables the company to conduct interviews with only those candidates who have passed the personality test. The company will save time and money by not having to interview the remaining applicants. The graph shows us which candidates are most likely to be qualified for a certain job position. Each candidate will take the personality test, and applicants with comparable credentials will be placed together in one class, with the graph drawn individually for each class of applicants' qualifications. This graph will then assist the business in classifying the applicants with the highest scores, who will subsequently be interviewed for the position. This Personality Prediction System will assess candidates' overall personalities and will be valuable for companies that hire people based on their personalities rather than their technical skills. The system will assist all firms in hiring for management and sales positions, resulting in a gain in income because the resources used in the recruitment process will be greatly reduced. In addition, in a separate area of the system, the candidates would come to know their own personalities. They will get to realise where they lack and what all personality attributes they have to acquire. It will also allow them to see where they are strong and where they need to improve. As a result, the Personality Prediction System will benefit both the company and the candidates by identifying their strong and weak areas.

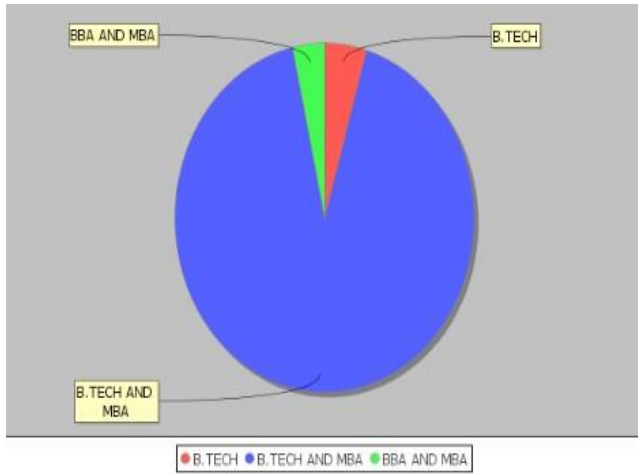


Figure 1: Graph: comparison B.TECH, BBA, MCA

VI. CONCLUSION & FUTURE DIRECTION

In this part, we describe and discuss how the classification algorithms fared for predicting the personality of the user. With an average accuracy of roughly 60%, the Nave Bayes Algorithm is the most accurate of the two algorithms examined. Due to the difficulty in separating a class of a word as the dataset was not entirely correct, the Support Vector Machine algorithm performed somewhat lower than Nave Bayes. Personality analysis and prediction have been increasingly popular in recent years[10]. The present system's ability to extract a user's personality is extremely useful in a variety of disciplines, including the recruiting process, medical counselling, and so on. The term "personality detection" refers to the process of extracting the user's behaviour traits from a survey. . This work emphasises on presenting a state-of-art review of an emrging filed i.e. personality identification from survey. The state-of-the-art approaches for personality identification and prediction are also discussed in this work. Apart from the work done thus far, future work will primarily focus on the following goals: We hope to integrate more personality qualities in future work so that we can present a more full personality to the user, as well as forecast personality using textual data and sentiment analysis. There might be a module that gives the user career advice and counselling sessions that are tailored to his personality.

REFERENCES

- [1] I. Cantandir, I. Fernandez-Tobiaz, A. Belllogin, I. Cantandir, I. Fernandez-Tobiaz, I. Fernandez-Tobiaz, I. Fernand (2013).
- [2] Aleksandar Kartelj, Veljko Milutinovi, Vladimir Filipovi (2012). Proceedings of the 35th International Convention MIPRO.IEEE. Novel techniques to automated personality classification: Ideas and their potentials.
- [3] Randall Wald, Taghi M. Khoshgoftaar, Amri Napolitano, and Chris Sumner are three of the authors (2012). IEEE 11th International Conference on Machine Learning and Applications, Using Twitter Content to Predict Psychopathy.

- [4] Arthur C Graesser, Fazel Keshtkar, Candice Burkett, Haiying Li, and Fazel Keshtkar (2014). Springer International Publishing, Using Data Mining Techniques to Detect the Personality of Players in an Educational Game.
- [5] A Learning Game Pedro Isasi, Yago Saez, Carlos Navarro, Asuncion Mochon (2014). A personality and happiness detecting system International Journal of Interactive Multimedia and Artificial Intelligence is a publication dedicated to the study of interactive multimedia and artificial intelligence.
- [6] J. Golbeck, C. Robles, and K. Turner [6] (2011). CHI'11 Extended Abstracts on Human Factors in Computing Systems, pp. 253-262, Predicting personality with social media.
- [7] Henny Vander Meijden, Nurbih A Shukora, Zaidatun Tasira, Nurbih A Shukora (2015). 555-562 in Science Direct - Procedia - Social and Behavioral Sciences, An Examination of Online Learning Effectiveness Using Data Mining.
- [8] C.D. Manning, P. Raghavan, H. Schutze, C.D. Manning, P. Raghavan, H. Schutze, C.D. Manning, P. Ragma (2008). The Basics of Information Retrieval. ISBN: 978-0-521-86571-5, Cambridge University Press
- [9] R.R. McCrae, P.T. Costa (1992). Psychological Assessment Resources' revised NEO personality inventory (NEO-PI-R) and NEO five-factor inventory (NEO- FFI).
- [10] Sebast'ian Ventura, Crist'obal Romeroan (2010). IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)VOL. 40, NO. 6. Educational Data Mining: A Review of the State of the Art, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)VOL. 40, NO. 6.