A Predictive Model using Personality Traits: A Survey

Rutuja Tarale[1], Prof. Pramila Chawan[2]

[1]M. Tech Student, Dept of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India
[2]Associate Professor, Dept of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India
---***---

Abstract - *Choosing the correct career is a crucial undertaking, but with the abundance of new career alternatives and opportunities that arise every day, it can be challenging. The CSIR estimates that roughly 40% of students make the wrong profession choice as a result of their confusion. As a result, pupils' productivity declines as a result of poor selection. In order to avoid negative outcomes brought on by making the wrong option, it is crucial to select the right vocation at the right age. A crucial responsibility for the generation of today, as the world becomes increasingly technological, is choosing a career. This problem primarily affects pupils who are interested in different things. Everyone wants their children to become engineers or doctors, but nobody ever asks the child what he or she is interested in. Moreover, parents are often concerned about their children's futures. So, this software benefits both the parents and the students.*

Key Words: Language Models, Personality Prediction, Psycholinguistic Features, NLP, MBTI.

1.INTRODUCTION

With the increase of study and inquiry in various fields, there are numerous new job opportunities in each subject. This further complicates the task of selecting a career for students concentrating in tenth or twelfth school. The causes of this confusion may include self-incapacity and self-character traits ignorance, ignorance of the various options available, comparable premiums in many professions, lack of presentation, advertise bombardment, anticipated public activities, peer pressure, etc. Due to these issues, the understudy may decide on a job that is not a good fit for them, which could lead to work disappointment, subpar performance, stress, and other negative outcomes including social negligence. So, there needs to be proper guidance for the understudy's brain science, interest, and capacity to work in a particular area.

One of the key factors on which our future depends is our career. Modernization has led to the emergence of several new job options, but on the flip side, an ever-growing number of graduates are graduating each year, boosting competitiveness. In order to provide the finest profession option that matches their personality and talents, it is also necessary to be aware of their hobbies, strengths, and weaknesses. Since we are in the information age, information is constantly growing in many areas, and much beneficial knowledge is collected from this data and applied to various changes.

1.1 Personality Traits

The term "personality traits" is used to describe a set of relatively consistent thought, feeling, and behavior patterns that have been linked to a variety of significant life decisions and consequences. Particularly, personality characteristics have frequently been linked to personal outcomes (such as happiness, psychopathology), interpersonal outcomes (such as relationship satisfaction), and social institutional outcomes (such as career choices, work success). As a result, there is growing interest in creating models that can automatically predict people's levels of personality traits using internet data about human behavior and preferences (i.e., digital footprints).

Need of Personality Traits?

1) Recommender systems

2) Product and service personalization

3) Career Analysis

4) Job screenings

5) Social network analysis

6) Sentiment analysis

1.2 Career Analysis

The choice of the best university or college no longer marks the conclusion of career planning, which was once a one-time event. It continues until we find the ideal position and a fulfilling work profile. One often chooses a stream based on their prior greatest achievement after finishing their education. Students that receive 90 percentiles in science desire to be engineers or doctors, and the same is true for commerce and all other subjects. More often than not, it turns out to be a hasty or poor decision. The criteria by which we assess our career decision are still unjustified.

Most Common Factors that Influence the Career Choice Among Students:

Parents' Desire/Parental

1) Pressure Peer Persuasion

2) Past Performance

3) Potential

4) Placements

5) Personality Driven

6) Affordability

1.3 MBTI

The Myers-Briggs Type Indicator (MBTI), used in personality typology, is an introspective self-report questionnaire that identifies various psychological preferences in how people view the world and make decisions. Although being extensively accepted by the scientific community, it is widely viewed as pseudoscience. Introversion or extraversion, sensing or intuition, thinking or feeling, and evaluating or perceiving are the four categories that the exam tries to rate. To create four-letter test results like "INTJ" or "ESFP," one letter is selected from each group.

Fig-1: MBTI types

2. LITERATURE SURVEY

Ashutosh Shankhdhar Et al. [1], this study sought to enhance the recommendation process by helping readers choose a career path based on their personality traits, areas of interest, and readiness to enroll in a course. It also recommended the best universities for them depending on their location and tuition costs. Academic performance was evaluated in order to determine whether the student possessed the necessary skills to follow a specific predicted vocation. This also includes details about the student's gender, nationality, age, how many days they attend class, whether they participate in extracurricular activities, whether they pay attention in class, and how many online courses they are capable of. In the disciplines of data, techniques like KNN, Neural Networks, K-means clustering, D-Tree, and many other cutting-edge algorithms are used to

calculate some data that is useful to predict the proper careers. This system functions as a career counselling system in which students submit their academic records, take an aptitude test, and fill out information about their interests, grades, and other factors.

Vignesh S Et al. [2], in order to forecast the best department for a person based on their skills as determined by an objective test, this article attempted to design a career system. One will automatically select the proper course if they complete the online evaluation we have provided in our system, which will help lower the failure rate due to selecting the incorrect career route. Dataset variety, appropriate career analysis, and improved visualization. Recent datasets and better algorithms could be employed, however there are no personality traits. Techniques including HTML, CSS, Flask-API, Neural Networks, and K-means clustering were employed. For department recommendation purposes, the success rate in each of the clusters is calculated and utilized to determine which clusters will have a higher success rate and a lower failure rate. In this project, a web-based application for career guidance has been conceived and developed after extensive research on the topic.

Yash Mehta Et al. [3], The authors of this research suggested a brand-new deep learning-based model for predicting language-based personality traits. This model includes employed language model embeddings and conventional psycholinguistic features as features. They have also examined how specific psycholinguistic characteristics affect a personality trait's ultimate prediction. According to the findings, psycholinguistic characteristics that model language typically outperform those that do not. Overall, the BERT-base + MLP model outperformed BERT-large + MLP in terms of predicting the Big Five personality traits and the MBTI dimensions. Our models' predicted performances on the Essays dataset and the Kaggle dataset outperformed the existing state-of-the-art by 1.6% and 1%, respectively. Also, results from our interpretable machine learning study partially concur with earlier psychological studies.

Kartikey Joshi Et al. [4], They have developed a system for recommending careers that will assist today's youngsters in deciding which professional route is best for their future and will yield the best outcomes if they take that path. This will enhance the student's performance and pique their attention, allowing them to concentrate on their desired job. This system is built on a test that students must take, and depending on their responses, it will produce a summarized result. This system's primary goal is to give a general summary of the artificial intelligence methods we employed to forecast the student's success. This system will also concentrate on how we are identifying attributes in student data by means of prediction algorithms. Using this technique turned out to be advantageous for educators, educational

institutions, and students alike. The system will proclaim the final conclusion depending on the knowledge we placed into it and the artificial intelligence that underpins it. Absence of a personality test and accuracy dependent on particular features are both flaws in the approach. SVM, Decision Tree, Artificial Intelligence, and Student Career Prediction are some of the terms that were employed in the creation of this research.

Md. Yeasin Arafath Et al. [5], They used classification to examine survey data on successful alumni who are now employed, and they used that data to forecast the careers of final-year students based on certain desirable characteristics. We primarily examined various academic, technical, and interpersonal aspects of the alumni's undergraduate careers as well as their current line of work. Their current line of work is regarded as a class label, and the quality traits are regarded as features. This data are used to train the models, which then use the replies of the running students who have finished their third year to forecast their careers based on the same quality criteria as test sets. Several classification methods are available. So, we used various classification algorithms and conducted a comparison of the classifiers' performances. The accuracy, precision, recall, and f-measure of a model are only a few of the metrics used to assess its performance. Several datasets are utilized to increase diversity. It might be better if more other ML methods were used. techniques for data mining like ID3, Random Forest, and Cart.

Chanchal Suman Et al. [6], They have put up a paradigm for resolving the issue of guessing a user's personality from videos. background, facial and from the user's video are taken the audio features. The most often used measurements in the literature are the big-five personality traits. Extraversion, Neuroticism, Agreeability, Conscientiousness, and Openness are these traits. Many practical uses exist for automatic personality characteristic prediction, including forensics, recommender systems, personalized services, etc. In this work, we have put forth a framework for tackling the issue of predicting a user's personality characteristics from videos. From the user's video, ambient, face, and audio elements are retrieved. The dataset made public in Chalearn-16 is used to assess the system's performance. According to the results of the experiments, employing a small number of the video's photos can yield better results than using the entire collection. The Fac-model visual's architecture is the same as the Ambvisual's model architecture for extracting features.

Madhura Jayaratne al. [7], In order to estimate HEXACO trait values from textual material, the authors created a regression model using natural language processing (NLP) and machine learning techniques. With an average correlation of r=0.39, they discovered that term frequency-inverse document frequency (TF-IDF) with Latent Dirichlet

Allocation (LDA) themes performed the best when compared to the other five text representation techniques. In contrast, IBM's Personality Insights service, developed using text data from Twitter, claims an average correlation of r=0.31 while big research using Facebook messages-based inference of Big 5 personality showed an average correlation of r=0.35. Their research demonstrates that personality traits can be accurately inferred from the text of answers to common interview questions about past conduct and situational judgement. In order to self-rate people's personalities, we used information from over 46,000 job applicants who participated in an online chat interview that also contained a personality questionnaire based on the six-factor HEXACO personality model.

Zhanming Guan Et al. [8], The unique model Personality2vec, which is based on network representation learning, fully utilises the semantic, personality, and structural data in users' messages. The model trains on the network using a novel biased walk algorithm and a better skip-gram method, and it eventually produces a personality vector for each user. The personality vectors used by the author are to forecast users' Big Five personality ratings using regression techniques. Using three personality datasets, experimental results demonstrate that personality2vec beats seven widely used approaches. First off, the majority of techniques primarily concentrate on the grammatical and semantic content of users' writing. Earlier techniques used psychological dictionaries like LIWC and MRC, which were created based on psychological and statistical expertise, to extract linguistic properties. Some techniques attempt to employ NLP approaches to extract semantic features and combine those features with linguistic features for personality prediction in recent years, with the emergence of deep learning and the maturation of NLP technology.

Z. Mushtaq al. [9], They suggested a system that could be erroneous because it asks consumers to fill out questionnaires to obtain their personality insights. and make an effort to fill out the survey carefully. Nevertheless, when it comes to social media, individuals do not give their ideas any thought before posting them. Hence, the information gleaned from social media could be invaluable in identifying the personality type of the user. In this article, we suggest a method for analyzing user data from social media sites by fusing two already-existing machine learning methods, such as K-Means Clustering and Gradient Boosting, in order to identify user personality type. Also, this study contributes to the analysis of the actual relationship between a user's personality and the data they post on social media. The Myers-Briggs Type Indicator (MBTI), developed by Swiss psychiatrist Carl Jung, was used in this essay. A well-known personality assessment tool used to determine a person's personality type, areas of strength, and personal preferences is the Myers-Briggs Personality Indicator.

Hans Christian Et al. [10], They have created a novel prediction that combines features from numerous pre-trained language models, such as BERT, RoBERTa, and XLNet, with a multi-model deep learning architecture. Bidirectional context feature combined with extraction approach creates a prediction model for each attribute utilizing numerous social media data sources, including Facebook and Twitter. Findings have a maximum accuracy of 86.2% and a 0.912 f1 measure score on the Facebook dataset, and an accuracy of 88.5% and a 0.882 f1 measure score on the Twitter dataset. The tools used for model creation include BERT, RoBERTa, XLNet, NLP Features, and the Big Five personality model. A well-known personality model, known as the major fve personality traits, has frequently been embraced in the literature as the de facto standard for personality evaluation due to its simplicity and shown effectiveness. Many techniques, some of which are based on ensembled models and deep learning, can be used to extract embedded contextualized words from textual data for personality prediction systems.

3. PROPOSED SYSTEM

3.1 Problem Statement

To implement a Predictive Model for the Career Analysis using MBTI Trait and psycholinguistic features.

3.2 Problem Elaboration

The accuracy of prediction actually lies with the set of relevant skill parameters, interpersonal and academic factors. This system works as a career counselling system in which student give academic record and give the test of aptitude and also fill details such as hobbies, marks, etc and then the system will recommend a suitable career for that student. This system is also helpful to determine the internal traits of the student which will also be considered to find an appropriate career.

3.3 Proposed Methodology

The system will ask the user or student to create a paragraph-style essay in order to anticipate their personality and then offer a job that fits their as well as taking into account particular decisions or preferences made. One of the most recent embedding techniques or algorithms for low-dimensional vector space is called long short term memory (LSTM), which makes the proposed system perform better because it is simple for the system to grasp. The accuracy of machine learning models such as Long Short-Term Memory (LSTM), Decision Tree, KNN, and Logistic Regression will be compared. The most accurate and efficient model among these will be taken into consideration. We will create corresponding analyses for corresponding domains utilizing the 16 personalities from the MBTI Kaggle dataset. The model will next be trained (80%) and tested (20%) using the

aforementioned procedures. The dataset is then classified using the aforementioned techniques. The logic for making recommendations is then developed. We will choose the best model after evaluating the accuracy of the three models, and we will then develop a model for a recommendation system for each area. Hence, machine learning has assisted us in enhancing system performance and developing a superior recommendation system.
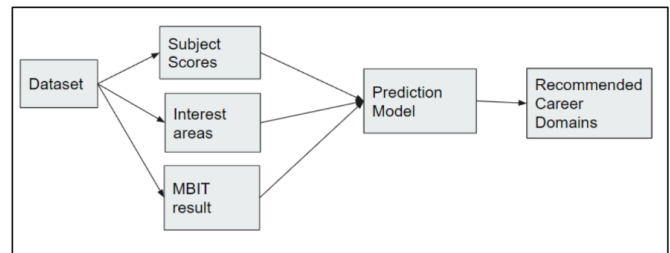

Fig-2: Proposed system Block Diagram

Data Collection:

Data from the Kaggle MBTI dataset is used for all career and personality categories. This information was gathered from the PersonalityCafe forum and offers a wide variety of persons engaging in casual online social interaction. The last 50 items a user wrote on the website are listed in 8675 entries together with their MBTI binary personality type in this dataset. It is about 8.7GB in size. Both JSON and picture formats are supported. The JSON datasets will be transformed into CSV files. The appropriate dataset will contain a variety of domain-related parameters.

Training and Testing the dataset:

We will first do preprocessing on the datasets to eliminate duplicate, invalid, and null values from the dataset before starting the model training procedure. The relevant datasets will then be split into a training set and a testing set. Moreover, the training to test set ratio will be 80:20. Three models will be used to classify the data. The predictive model will be used to build recommendation prediction logic in our suggested system. A test set will be used to test the model. Recall, precision, support, F1 measure, accuracy, and other parameters will be used to evaluate the system's performance.

4. CONCLUSIONS

When a student takes a test and provides information about their grades and interests, the proposed career system will process the data and match them with the ideal career option based on their abilities. It also takes into account their personality traits, aptitude test results, and field of interest. This website served as a career counsellor, predicting careers based on tests and records. Not only would this save

time and effort, but it will also give a more accurate picture of one's personality and job prospects.

REFERENCES

[1] A. Shankhdhar, A. Agrawal, D. Sharma, S. Chaturvedi and M. Pushkarna, "Intelligent Decision Support System Using Decision Tree Method for Student Career," 2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC), Mathura, India, 2020, pp. 140-142, doi: 10.1109/PARC49193.2020.246974.M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[2] S. Vignesh, C. Shivani Priyanka, H. Shree Manju and K. Mythili, "An Intelligent Career Guidance System using Machine Learning," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2021, pp.987-990, doi:10.1109/ICACCS51430.2021.9441978.

[3] Y. Mehta, S. Fatehi, A. Kazameini, C. Stachl, E. Cambria and S. Eetemadi, "Bottom-Up and Top-Down: Predicting Personality with Psycholinguistic and Language Model Features," 2020 IEEE International Conference on Data Mining (ICDM), Sorrento, Italy, 2020, pp. 1184-1189, doi: 10.1109/ICDM50108.2020.00146.

[4] K. Joshi, A. K. Goel and T. Kumar, "Online Career Counsellor System based on Artificial Intelligence: An approach," 2020 7th International Conference on Smart Structures and Systems (ICSSS), Chennai, India, 2020, pp. 1-4, doi: 10.1109/ICSSS49621.2020.9202024.

[5] M. Y. Arafath, M. Saifuzzaman, S. Ahmed and S. A. Hossain, "Predicting Career Using Data Mining," 2018 International Conference on Computing, Power and Communication Technologies (GUCON), Greater Noida, India, 2018, pp. 889-894, doi: 10.1109/GUCON.2018.8674995.

[6] Chanchal Suman, Sriparna Saha, Aditya Gupta, Saurabh Kumar Pandey, Pushpak Bhattacharyya, A multi-modal personality prediction system, Knowledge-Based Systems, Volume 236, 2022, 107715, ISSN 0950-7051, https://doi.org/10.1016/j.knosys.2021.107715.

[7] M. Jayaratne and B. Jayatilleke, "Predicting Personality Using Answers to Open-Ended Interview Questions," in IEEE Access, vol. 8, pp. 115345-115355, 2020, doi: 10.1109/ACCESS.2020.3004002.

[8] Z. Guan, B. Wu, B. Wang and H. Liu, "Personality2vec: Network Representation Learning for Personality," 2020 IEEE Fifth International Conference on Data Science in Cyberspace (DSC), Hong Kong, China, 2020, pp. 30-37, doi: 10.1109/DSC50466.2020.00013.

[9] Z. Mushtaq, S. Ashraf and N. Sabahat, "Predicting MBTI Personality type with K-means Clustering and Gradient Boosting," 2020 IEEE 23rd International Multitopic Conference (INMIC), Bahawalpur, Pakistan, 2020, pp. 1-5, doi: 10.1109/INMIC50486.2020.9318078.

[10] Christian, H., Suhartono, D., Chowanda, A. et al. Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging. J Big Data 8, 68 (2021). https://doi.org/10.1186/s40537-021-00459-1

[11] M. T. Apple, J. Falout and G. Hill, "The Relationship Between Future Career Self Images and English Achievement Test Scores of Japanese STEM Students," in IEEE Transactions on Professional Communication, vol. 63, no. 4, pp. 372- 385, Dec. 2020, doi: 10.1109/TPC.2020.3029662.

[12] A. V. Kunte and S. Panicker, "Using textual data for Personality Prediction:A Machine Learning Approach," 2019 4th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2019, pp. 529-533, doi: 10.1109/ISCON47742.2019.9036220.

BIOGRAPHIES

Rutuja Tarale is currently pursuing M. Tech from VJTI COE, Mumbai. She has done her B.E. (Computer Engineering) from Xavier Institute of Engineering, Mumbai.

Prof. Pramila M. Chawan, is working as an Associate Professor in the Computer Engineering Department of VJTI, Mumbai. She has done her B.E.(Computer Engineering) and M.E.(Computer Engineering) from VJTI College of Engineering, Mumbai University. She has 32 years of teaching experience and has guided 80+ M. Tech. projects and 100+ B. Tech. projects. She has published 140+ papers in the International Journals, 20+ papers in the National/International Conferences/ Symposiums. She has worked as an Organizing Committee member for 21

International Conferences and 5 AICTE/MHRD sponsored Workshops/STTPs/FDPs. She has participated in 14 National/International Conferences. She has worked as NBA Coordinator of the Computer Engineering Department of VJTI for 5 years. She had written a proposal under TEQIP-I in June 2004 for 'Creating Central Computing Facility at VJTI'. Rs. Eight Crore were sanctioned by the World Bank under TEQIP-I on this proposal. Central Computing Facility was set up at VJTI through this fund which has played a key role in improving the teaching learning process at VJTI.