

Personality Assessment Using Social Media for Hiring Candidates

Roshal Moraes
Computer Engineering
St. Francis Institute of Technology
Mumbai, India
roshal.moraes@student.sfit.ac.in

Larissa Lancelot Pinto
Computer Engineering
St. Francis Institute of Technology
Mumbai, India
larissapinto98@student.sfit.ac.in

Mrunal Pilankar
Computer Engineering
St. Francis Institute of Technology
Mumbai, India
mrunalpilankar1298@student.sfit.ac.in

Pradnya Rane
Computer Department
St. Francis Institute of Technology
Mumbai, India
pradnyarane@sfit.ac.in

Abstract—Personality analysis is an intuitive ability of humans, carried out every day with multiple people, and for innumerable reasons. Personality profiling, specifically, has several real-life applications, such as mental health screening tests, screening during human resource interviews, etc. Hiring candidates is one of the most crucial and time-consuming processes for any company. Personality Analysis using social media is an actively developing field. It will help in the job hiring process to get an overall idea of the candidate's ideas and ideals, providing a consistent and fair ranking policy. This system aims to automate the personality assessment process by analyzing the text extracted from his/her tweets, and suggests general job roles suitable for the candidates. The models used for personality analysis are Myers Briggs Type Indicator (MBTI) personality analysis and Big Five personality analysis. The employer from the hiring company will be shown a metric of the applied candidates' personality analysis.

Keywords—Big Five, Job Role Prediction, MBTI, Personality Assessment, Social media, Twitter Analysis

I. INTRODUCTION

Personality is defined as the traits that predict a person's behavior. Personality Assessment is expertise in psychology that includes scoring techniques and interpretation of scoring measures to determine personality traits and styles to help in diagnoses, grouping, and counseling. These assessments help to improve the accuracy in measuring the dominant personality traits by using objective instead of subjective methods. Personality assessment of a person helps us to understand how he/she will handle real-time situations and which environments are most suitable for them to progress. It helps the company to make better use of their skills. This assessment also helps to make hiring decisions easier for the recruiting team.

The traditional hiring process faces a lot of challenges and issues like attracting the right quality candidates, less time required for the hiring process, ensuring a good candidate experience and so on. Interested applicants apply to the company and the HR managers have to sort out the candidates after continuous interview processes. This becomes a time-consuming process that is human error-prone.

Personality Assessment using Social Media for Hiring Candidates System aims to extract tweets from candidate's account and perform personality assessment to help the recruiting team with the interviewing process. The results obtained must be used to shortlist candidates along with other factors. The system plans to overcome some of the inconsistencies in the system and make the selection process fair.

The system tries to incorporate the personality of the candidate as a measure to shortlist candidates by using quantitative psychological personality assessment models used widely. This helps in saving time for the hiring process as the company will have an overall idea of the candidate's personality and the job role he/she is suitable for.

II. LITERATURE REVIEW

A strong interconnection between a user's personality and their online behavior on social networks has been observed in various studies and experiments. Since social media posts made by users are reflective of their personality traits, they can be analyzed to predict the personality of a person. Personality traits of an individual are not specific but generalized. The personality analysis models considered for the project are Myers Briggs Type Indicator (MBTI) and Big Five Indicator.

MBTI is a commonly used Personality Model. It works on binary classes. MBTI describes human identity in four principal measurements i.e., Introversion or Extroversion, Sensing or Intuitiveness, Thinkers or Feelers and Judgers or Perceivers. Big Five indicators include openness, conscientiousness, extraversion, agreeableness and neuroticism. [1]

Text classification is used for personality assessment. The performance of Naïve Bayes, K Nearest Neighbor (KNN) and Support Vector Machine (SVM) for text classification is compared. The steps undertaken to preprocess the data are tokenization, stemming, filtering and weighting. In tokenization each entry in the corpus will be converted into a set of words. Stemming returns a word into a basic form

(root word) by eliminating affixes attached to it. Unlike stemming, lemmatization usually refers to finding the canonical form of a word with the use of a vocabulary and morphological analysis of words. Parts of Speech (POS) Tagging attaches a POS tag with each word obtained. Stop words removal is a process that removes words with little or no meaning. Multinomial Naive Bayes (MNB) is a modified version of the Naive Bayes algorithm based on the Naive Bayes theorem designed for classification of text documents. MNB uses multinomial distribution with the weight or count of the word as a classification feature. K-Nearest Neighbors (KNN) classification algorithm uses the K nearest neighbors of the data to predict its class. It uses a distance function between the training and test samples and the number of nearest neighbors to determine the classification results. Support Vector Machine (SVM) is based on identifying patterns in the training dataset to predict the class. SVM is a binary supervised learning algorithm. Classification on Twitter text is analyzed using the True Positive Rate (TPR) and True Negative Rate (TNR). Twitter accounts having greater number of tweets (100-1000) gave the best results. The online personality analysis method has 65% accuracy compared to Questionnaire based test Naive Bayes performed slightly better than others [2].

Decision tree, Naïve Bayes, K Nearest Neighbor (KNN) and Support Vector Machine (SVM) are compared. Decision tree used information gain as a parameter. Performance comparison of different algorithms is given in Table I.

The different feature selection methods applied for each algorithm, input type best suited and performance measures like computational demand, memory requirement, error rate are estimated and given.

KNN required a larger dataset and consumed more memory and time in the testing phase. Naïve Bayes' principle assumes all events to be independent of each other and thus showed low performance in classification.[3]

III. CHALLENGES IDENTIFIED

The traditional hiring process uses interviews to measure a candidate's personality traits subjectively which may not determine the best candidate and also uses significant resources. That's because the hiring process is flawed due to human bias and inaccurate interpretations.

Challenges include:

A. Human Error

Traditional candidate hiring systems depend on HR interview alone for personality assessment. This process is prone to human error and a lack of proper judgment. The candidate may not be able to showcase his actual personality due to nervousness or inhibitions.

B. Time Consuming

Usually, the number of applicants for a job is very high and interviewing each one of them can be time-consuming. This affects the quality of the interviews and makes it more error-prone. Time spent per candidate can be around 15-20 minutes which is not enough for proper analysis.

C. Cost

Companies invest a lot of money in the process of hiring as they need the best employees. The cost increases as the process and the time increases.

IV. PROBLEM DEFINITION

Hiring the correct candidate is essential for a company's progress and reputation. Hence, companies put in a lot of effort into hiring the right candidates. Despite the resources allotted to the HR department, it is tough for the interviewer to assess the candidate's personality in one interview. Identifying the right candidate includes judging the person's personality along with academics. Problems like human error, lack of time and resources in the hiring process occur in personality identification. These issues can be minimized using social media for personality prediction digitally. Personality prediction system helps to simplify the interview process. Personality analysis is done by giving microblogs (Tweets) written by the user to the system. Text from the resume is used to improve accuracy.

TABLE I. COMPARISON OF VARIOUS SUPERVISED MACHINE LEARNING ALGORITHMS [3]

Parameter	Decision Tree	Naïve Bayes	KNN	SVM
<i>Feature Selection</i>	Implicit selection	Explicit, Multiclass	Explicit, Multiclass	Default binary. Can be multiclass
<i>Input Type</i>	Categorical, Continuous values can be grouped into categories	Categorical, Continuous (Gaussian Naïve Bayes)	Continuous	Continuous
<i>Error rate</i>	High due to over fitting or under fitting	Error occurs if events are dependent	Depends on choice of 'K'	High when filtering of data is not done correctly
<i>Algorithm</i>	ID3	Multinomial/Bernoulli Naïve Bayes	KNN with TF-IDF	Linear SVM
<i>Best</i>	Best for larger samples	Best for smaller samples	Best for smaller sample size	Suitable for small and medium datasets
<i>Computational Demand and Memory Requirements</i>	Data classified with minimum calculation	Better results when events are independent	High memory and time requirements	Memory efficient

V. THEORETICAL MODEL

The application is based on mapping MBTI and Big Five personality traits to job roles to be suggested to recruiters.

The personality traits are obtained by using the tweets of the candidate as opposed to the traditional questionnaire designed to measure the traits as these questionnaires can be faked by the candidate. Filling the questionnaires repeatedly to measure personality changes over time is also not practical. The application uses the most recent tweets to obtain the most recent personality shift. The mapping of personality traits to job roles is summarized in Table II and Table III.

VI. PROPOSED SYSTEM METHODOLOGY

The method used to assess the personality of an individual is

shown in Fig. 1.

1. **Data Preparation:** Datasets for MBTI and Big Five are selected and are further divided into training and testing sets.

Cleaning and pre-processing data: Includes tokenization, Parts of Speech (POS) tagging, stop words removal, and lemmatization.

Tokenization: Each entry in the corpus will be converted into a set of words.

Stemming: It returns a word into a basic form (root word) by eliminating existing additive.

Lemmatization: Unlike stemming, this usually refers to finding the canonical form of a word with the use of a vocabulary and morphological analysis of words.

TABLE II. MAPPING OF MBTI PERSONALITY TRAITS TO JOB ROLES

Category	MBTI Type	Traits	Job Roles
Analyst	INTJ – The Architect	Perfectionist , Non- Social , Theoretical thinking	Investment Banker , Financial Banker , Software Developer , Economist , Executive
	INTP – The Logician	Creative Problem Solver , Independent , Focus on Theory and Precision	Computer Programmer , Software Designer , Analyst , Architect , Professor
	ENTJ –The Commander	Natural leader, logical, analytical, strategic planner	Authoritarian roles, Consultant, Research analyst, Venture Capitalist
	ENTP – The Debater	Creative, likes challenges, persistent, non-conformist	Entrepreneur, Real Estate, Marketing Director, Political roles
Diplomats	INFJ –The Advocate	Thoughtful, creative, firm principles and integrity	Behind the scene roles like Therapist, Counselor, Social Worker, HR roles, Customer Relations manager
	INFP –The Mediator	Sensitive Idealists, deeper personal values, compassionate and adaptable	Graphics designer , Psychologist , Therapist , Writer/ Editor
	ENFJ –The Protagonist	People lovers, energetic, articulate, diplomatic	Expressive and logical job roles like Advertising, Public Relations Specialist, Corporate Coach, Sales Manager, HR roles
	ENFP –The Campaigner	Curious, Confident, Creative, Alert and Communicative	Expressive Roles like Journalist, Consultant, Restaurateur, Event Planner
Sentinels	ISTJ –The Logistician	Hard-working, responsible, committed, reliable	Behind the scenes roles like CFOs, Web Development, Accountant, Auditor
	ISFJ –The Defender	Modest, Determined, Helping Nature, provide service without position of authority	Dentist, School teacher, Librarian, Customer Service
	ESTJ –The Executive	Realists, practical decision makers , Leaders	Social roles like Project Manager, Judge, Lawyer , Insurance Sales Agent
	ESFJ – The Consultant	Gregarious traditionalists, help others	Social and well-being roles like Nurse, Healthcare Worker , Social Worker, Public Relations
Explorers	ISTP – The Virtuoso	Straightforward, honest, prefer deeds to words	Utilitarian roles like civil engineer , Pilot, Economist , Data Communication Analyst
	ISFP – The Adventurer	Warm, Sensitive, like to help people, sympathetic and attentive	Fashion Designer, Storekeeper , Landscape Architect , Physical / Massage therapist
	ESTP – The Entrepreneur	Pragmatist, Love excitement, excel in crisis	High stake roles like Detective, Banker, Investor , Entertainment and Sports
	ESFP – The Entertainer	Lively, playful, value common sense, expressive, like interaction	Child Welfare Counselor, Interior Designer , Actor, Environment Scientist

TABLE III. MAPPING OF BIG FIVE PERSONALITY TRAITS TO JOB ROLES

Type	Trait	Job Roles
<i>Openness</i>	Prefer new, exciting situations, value knowledge, curious and intellectual	Guides, Pilot, knowledge seeking roles
<i>Conscientious</i>	Self- disciplined, prefer ordered and planned activities, strong focus	Accountant, Manager
<i>Extroversion</i>	Social, Action- oriented, like interactions	Event Planner, Trainers, Actors
<i>Agreeableness</i>	Considerate, kind, sympathetic, help others	School teacher, HR specialist, social workers
<i>Neuroticism</i>	High score indicates anxiety, pessimism, emotional instability	Emotional stability (indicated by low score) helps in job roles like nursing, social work and psychology

Parts of Speech (POS) Tagging: Attach a POS tag with each word obtained.

Stop words removal: Stop word is a common word that has little or no meaning.

2. Encoding: Encoding converts label associated with each data element to 0/1 model.

Tf-Idf vectorization: Tf-Idf (term-frequency inverse document frequency) is performed on the text data. Tf-Idf assigns a weight to evaluate how important a word is to a document in a collection or corpus. For a word 't' in document 'd',

$$TfIdf_t = f_{t,d} \times \log(N / d_{ft}) \quad (1)$$

Where,

$TfIdf_t$ is weight of term t

$f_{t,d}$ is occurrences of term t in document d

N is the total number of documents

d_{ft} is the number of documents containing t

Here, a document is an entry of the dataset (Social media text). This phase will convert a word into a numerical vector representation where words with similar meaning have similar representation.

3. Classification model generation: Different classification models are used to train the MBTI and Big Five datasets. The different classification models used are SVM, Decision Tree, KNN and Naïve Bayes. These models will be used to predict the classes of processed tweets of the user.
4. Testing : In the testing phase , the model is tested on the test dataset

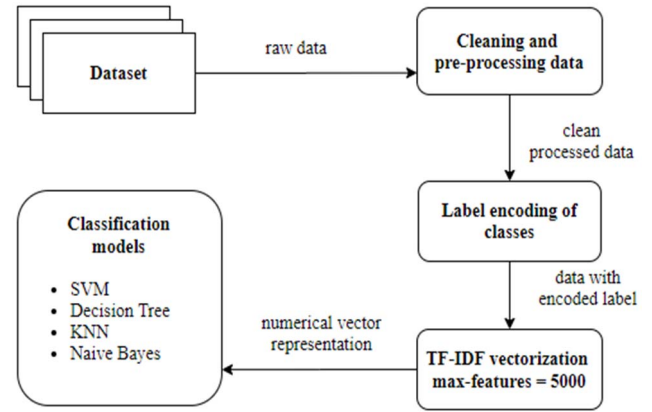


Fig. 1. Process used to implement Personality Analysis

VII. PERFORMANCE EVALUATION PARAMETER

Following parameters will be used to measure the performance of the personality prediction model:

1. Classification Accuracy

Classification accuracy is ratio of correctly predicted test cases to the total number of test cases.

Though commonly used, only accuracy is not sufficient to measure the efficiency of the model. It will only give suitable results when there are an equal number of observations in each class and when all predictions and the errors are equally important, which is not commonly seen Precision:

The Precision of a classification is given by (2)

$$\text{Precision} = TP / (TP + FP) \quad (2)$$

Recall:

The Recall of a classification is given by (3)

$$\text{Recall} = TP / (TP + FN) \quad (3)$$

Where,

TP is True Positive rate i.e. Number of elements actually true and predicted true.

FP is False Positive rate i.e. Number of elements actually false and predicted true.

FN is False Negative rate i.e. Number of elements actually true and predicted false.

VIII. EXPERIMENTAL SET UP

Hardware Requirements: The project must be built on a desktop/laptop with minimum 8GB RAM (16 GB optimal) and storage space of minimum 128 GB.

Software Requirements: Latest browser (Google Chrome, Firefox, Edge) for display and input testing needed. Python3 installation needed with scikit and Natural Language Processing (NLP) packages. SQLite database will store personality analysis and candidate details.

Big Five uses the essay dataset. Some of the earlier work done on the subject of automated personality detection using plain text was done by James Pennebaker and Laura King.[7] who compiled the essay dataset that we used in our experiment. For this, they collected stream-of-consciousness essays written by volunteers in a controlled environment and then asked the authors of the essays to define their own Big Five personality traits. This dataset consists of 2468 essays of approximately 670-700 words, which are already labeled into Big Five personality dimensions. Original dataset is then slightly modified by considering only the required columns. We access tweets of the users and feed it to the model as a text of 670-700 words. MBTI Dataset was collected from the Kaggle platform. According to Kaggle, this dataset was collected through the PersonalityCafe forum to train the model. The dataset has over 8600 rows of data, where each row has the individual's MBTI personality and his/her last 50 tweets appended as a single string. This data provides information of a large variety of people and their MBTI personality type, as well as their tweets. Each MBTI pair is considered individually for training the model. Introversion/Extroversion and Intuition/Sensing have been under-sampled to balance the data.[1].

IX. PRIMARY RESULT AND DISCUSSION

Training dataset and testing dataset is obtained by dividing available data into two parts by ratio 70:30, that is 70% used for training and 30% for testing. TF-IDF was weighting scheme used to generate the word vectors which is word frequency in document and in the entire corpus.

Table IV shows the accuracy, precision and recall of each of the four binary MBTI models. The SVM model provides best average performance for all traits in terms of all performance parameters.

The Big Five personality traits models are trained and tested on Big Five essays dataset. The results of Big Five models for all algorithms are as shown in Table V. Decision tree gives best average performance in terms of all parameters.

Thus, MBTI and Big Five results of the user are obtained using machine learning algorithms on user's tweets. These results are mapped using the theoretical model to obtain desirable job traits and roles of the user. These results will help recruiters select suitable candidate for the job position offered.

X. CONCLUSION

The online personality assessment is a simple cost effective, embeddable system which can be modified easily to be used in different application areas. MBTI and Big Five models are used to identify user's personality. We have trained text classification algorithms to predict users' personality traits and have mapped roles suitable to them. Further improvements in the results can be achieved using datasets native to the user's region and understanding their local linguistics. Also, a semantic approach can be tried to get better results.

TABLE IV. ACCURACY, PRECISION AND RECALL OF ALL ALGORITHMS FOR MBTI MODEL

	SVM			Decision Tree			Naïve Bayes			KNN		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall
<i>Introvert/Extrovert</i>	79.70%	88.09%	78.35%	67.85%	72.09%	70.96%	71.85%	92.33%	68.43%	70.52%	73.28%	73.86%
<i>Sensing/Intuition</i>	84.40%	61.97%	81.80%	89.77%	88.07%	82.08%	70.69%	26.95%	100.00%	78.86%	54.54%	74.24%
<i>Thinking/Feeling</i>	84.13%	83.40%	82.25%	79.38%	79.22%	77.14%	78.10%	65.74%	82.98%	70.90%	48.21%	81.88%
<i>Judging/Perceiving</i>	78.33%	87.64%	78.40%	67.96%	72.75%	73.12%	64.12%	98.89%	62.49%	67.73%	87.25%	67.70%
<i>Average</i>	81.64%	80.28%	80.20%	76.24%	78.03%	75.83%	71.19%	70.98%	78.48%	72.00%	65.82%	74.42%

TABLE V. ACCURACY, PRECISION AND RECALL OF ALL ALGORITHMS FOR BIG FIVE MODEL

	SVM			Decision Tree			Naïve Bayes			KNN		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall
<i>Openness</i>	77.18%	77.62%	78.03%	86.76%	86.91%	87.36%	74.68%	69.63%	78.81%	66.77%	42.40%	86.17%
<i>Conscientiousness</i>	76.09%	75.23%	77.67%	85.68%	87.42%	85.18%	74.88%	83.48%	72.14%	67.25%	50.06%	78.61%
<i>Extrovert</i>	73.53%	76.72%	73.47%	86.22%	89.86%	84.58%	68.26%	83.87%	65.08%	70.69%	54.22%	83.56%
<i>Agreeableness</i>	74.68%	78.21%	75.06%	85.95%	84.97%	88.11%	67.59%	89.55%	63.85%	68.06%	50.19%	82.77%
<i>Neuroticism</i>	75.96%	74.42%	76.70%	84.47%	85.11%	83.98%	72.04%	77.13%	69.94%	69.75%	56.16%	76.99%
<i>Average</i>	75.49%	76.44%	76.19%	85.82%	86.85%	85.84%	71.49%	80.73%	69.96%	68.50%	50.61%	81.62%

The results will help recruiters in the hiring process. Dominant personality traits and job roles suitable to the user can help the candidate to select a suitable career path as well. Target oriented personality and soft skill training for candidates, Psychological profiling for job satisfaction, career progression, and guidance can benefit from the psychological assessment. Other fields that can use the assessment are Recommendation Systems, Online Marketing, targeted advertising, determining behavior aspects for medical purposes and so on.

- [9] (MBTI) Myers-Briggs Personality Type Dataset – Kaggle, [online] Available: <https://www.kaggle.com/datasnaek/mbti-type>
- [10] J. Golbeck, C. Robles, K. Turner, "Predicting personality with social media," In CHI'11 Extended Abstracts on Human Factors in Computing Systems, pp. 253-262, 2011
- [11] Rayne Hernandez, Ian Scott Knight, "Predicting Myers-Briggs Type Indicator with Text Classification", 31st Conference on Neural Information Processing Systems, NIPS, 2017.
- [12] N. Majumder, S. Poria, A. Gelbukh and E. Cambria, "Deep Learning-Based Document Modeling for Personality Detection from Text", *IEEE Intell. Systems*, vol.32, no.2, Mar. 2017, doi: 10.1109/MIS.2017.23

REFERENCES

- [1] P.S. Dandannavar, S.R. Mangalwade, P.M. Kulkarni, "Social Media Text – A Source for Personality Prediction", International Conference on Computational Techniques, Electronics and Mechanical Systems, IEEE, 2018.
- [2] Bayu Yudha Pratama, Riyanarto Sarno, "Personality Classification Based on Twitter Text Using Naïve Bayes, KNN and SVM", International Conference on Data and Software Engineering, IEEE, 2015.
- [3] A.Helen Victoria, M. Vijayalakshmi, "An outcome based Comparative study of different text classification Algorithms", International Journal of Pure and Applied Sciences, Mathematics, ijpam.eu, 2018.
- [4] Paul D. Teiger, Barbara Barron and Kelly Teiger, "Do What You Are: Discover the Perfect Career for You Through the Secrets of Personality Type", 5th ed., Little Brown Spark, 2014.
- [5] "The best jobs for your personality type using the Myers-Briggs scale", Mar. 04, 2017. Accessed on: Dec. 28, 2019 [Online]. Available: <https://careers.workopolis.com/advice/the-best-jobs-for-your-personality-type-using-the-myers-briggs-scale/>
- [6] "Big Five Personality Traits: Finding the Right Jobs for You". Accessed on: Dec. 28, 2019 [Online]. Available: <https://www.indeed.com/career-advice/finding-a-job/big-five-personality-traits>
- [7] J.W. Pennebaker and L.A. King, "Linguistic Styles: Language Use as an Individual Difference," J. Personality and Social Psychology, vol. 77, no. 6, 1999, pp. 1296–1312.
- [8] "Hands-On Introduction To Scikit-learn (sklearn)", Mar. 04, 2017. Accessed on: Dec. 28, 2019 [Online]. Available: <https://towardsdatascience.com/hands-on-introduction-to-scikit-learn-sklearn-f3df652ff8f2>