

# Personality prediction using machine learning algorithm for Recruitment

Darshan Shaniyar Naik<sup>1</sup>, Dr. Preethi N Patil<sup>2</sup>

<sup>1</sup>PG Student, Master of Computer Applications, RV College of Engineering, Karnataka, India

<sup>2</sup>Assistant Professor, Master of Computer Applications, RV College of Engineering, Karnataka, India

## ABSTRACT

Personality can be characterized as a lot of qualities which makes an individual one of a kind. The investigation of personality is of focal significance in brain research. There are different traditional methods of surveying one's personality which either costs a lot of manual endeavours or is impossible progressively. To take care of these issues, this exploration intends to gauge the Big-Five personality from a lot of inquiries. The client is solicited to answer a set from scarcely any inquiries and as per the inquiries replied by the client the personality of the client is anticipated utilizing logistic regression model.

**Keywords:** Big-Five personality, logistic regression, term frequency.

changed the manners by which individuals pass on their assessments and sentiments. The examination of online networking information or a makes it conceivable to distinguish significant personality attributes, that is, attributes or characteristics which depicts his/her personality. This personality uncovers just what an individual's needs others to know. In this way, it is essential to know the genuine personality of an individual. Consequently, we utilize Big five personality test. The Big five personality test incorporates five diverse personality attributes specifically, Openness, Neuroticism, Conscientiousness, Extroversion and Agreeableness. Logistic Regression, a machine learning algorithm is used in this project for the prediction of personality of a person.

## INTRODUCTION

In today's world, the personality of a person is of more importance. For example, for hiring any employee in a company it is important to know the personality of a person according to the job role. They use different strategies to know one's personality. The lifestyles have drastically

In this paper the Authors states that social media use has been on a regularly expanding exponential ascent. Use of online networking locales, for example, Twitter and Facebook, for social association has likewise become a mainstream pattern. It is assessed that on a normal, around 6,000 tweets are tweeted on Twitter each second. With individuals spending

on a normal 35 minutes on Facebook every day, it is likewise assessed that there are around 317,000 announcements on Facebook every moment. These huge volumes of information have incredible data bolted inside them. This information can be dissected and a few purposes. The utilization of such web-based social networking information for anticipating client personality is normal. Prediction models have been effectively fabricated that can foresee a few client qualities - age, sex, personality attributes, occupation, political direction and so forth. Measures in personality models, for example, the Big Five model, Disk and the Myers-Briggs Type Indicator have been the reason for all such personality prediction. A client's web-based social networking information would thus be able to be utilized to foresee his/her personality. The principle goal of this work is to survey the work completed for personality prediction utilizing social network life data[1]. The paper states the point of programmed personality discernment is to foresee the personality of the speaker saw by the audience from nonverbal conduct. Extroversion, Conscientiousness, Agreeableness, Neuroticism and Openness are the speaker qualities utilized for personality assessment. In this work, a speaker characteristic prediction approach for programmed personality appraisal is proposed to display the connection between discourse sign and personality attributes utilizing frequency area straight prediction (FDLP) method. Among a few component extraction procedures, FDLP highlights render expanded execution. SSPNet Speaker Personality Corpus is utilized for investigations and assessment. The proposed technique predicts the speaker attributes with 90-99% characterization accuracy[2]. The authors in the paper suggests Albeit programmed personality

investigation has been concentrated broadly lately, it has not yet been received for continuous applications and genuine practices. As far as we could possibly know, this exhibit is a first endeavor at anticipating the broadly utilized Big Five personality measurements and various social measurements from nonverbal conduct prompts continuously. The proposed framework means to dissect the nonverbal conduct of the individual that cooperates with a little humanoid robot through a live gushing camera. And conveys the anticipated personality and social measurements on the fly[3]. This paper suggests personality is a mental develop planned for clarifying the wide assortment of human practices in terms of a couple, steady and quantifiable individual attributes. In this regard, any innovation including getting, prediction and union of human conduct is probably going to profit by Personality Computing draws near, for example from innovations equipped for managing human personality. This paper is a study of such advancements and it targets giving not just a strong information base about the best in class, yet additionally a reasonable model basic the three fundamental issues tended to in the writing, to be specific Automatic Personality Recognition (surmising of the genuine personality of a person from social proof), Automatic Personality Perception (induction of personality others ascribe to an individual dependent on her recognizable conduct) and Automatic Personality Synthesis (age of counterfeit personalities by means of exemplified specialists). Besides, the article features the issues despite everything open in the field and recognizes potential application areas[4]. In this paper, authors present a structure for anticipating the personality characteristics by breaking down tweets written in Turkish. The

prediction model is built with a bunching based methodology. Since the model depends on etymological highlights, it is language explicit. The prediction model uses highlights material to Turkish language and identified with composing style of Turkish Twitter clients. Their methodology utilizes unknown BIG5 survey scores of volunteer members as the ground truth so as to create personality model from Twitter posts. Trial results show that built model can anticipate personality characteristics of Turkish Twitter clients with moderately little errors[5]. This paper explains that with the fast development of social media, clients are engaging in virtual socialism, producing a tremendous volume of literary and picture substance. Considering the substance, for example, announcements/tweets and shared posts/retweets, preferring different posts is mirroring the online conduct of the clients. Foreseeing personality of a client from these advanced impressions has become a computationally testing issue. In a profile-based methodology, using the client created printed substance could be helpful to mirror the personality in social media. Utilizing enormous number of highlights of various classifications, for example, conventional etymological highlights (character-level, word-level, basic), psycholinguistic highlights (enthusiastic effects, recognitions, social connections) or social network highlights (network size, betweenness) could be helpful to foresee personality characteristics from social media. As indicated by a generally mainstream personality model, specifically, big-five-factor model (BFFM), the five components are openness-to-encounter, conscientiousness, extraversion, agreeableness, and neuroticism. Foreseeing personality is re-imagined as anticipating every one of these

qualities independently from the removed highlights. Generally, it takes colossal number of highlights to show signs of improvement precision on any prediction task in spite of the fact that applying highlight determination calculations may improve the presentation of the model. In this article, we have looked at the presentation of five component choice calculations, in particular the Pearson connection coefficient (PCC), data gain (IG), symmetric uncertainly (SU) evaluator, and chi-squared (CHI) strategy. The exhibition is assessed utilizing the exemplary measurements, specifically, exactness, review, f-measure, and precision as assessment matrices[6]. In this paper mentioned that "In the event that you need to find the genuine character of an individual, you have just to see what they are energetic about". Programmed assessment of personality characteristics from the posts posted on different social media platforms is an intriguing field of research. The interests of an individual are reflected in the substance of pictures which clients post to their social media accounts. Famous personality prediction frameworks characterize the personality of an individual based on five expansive personality characteristics Openness, Conscientiousness, Extraversion, Agreeableness and neuroticism (otherwise called the Big Five Personality Traits). This paper proposes a novel methodology of foreseeing an individual's personality reflected by his/her advantages. The task recognizes huge highlights in each picture posted by the client in his/her social media profile and afterward maps the highlights distinguished in the picture to a predefined class which depends on the regular intrigue spaces typically saw in social media. For instance, an individual slanted towards experience may post a

higher number of photographs identified with bold exercises like trekking, surfing, and so on. The classes, for example, can be; creature sweetheart, foodie, gutsy and so on. The noteworthy highlights in a picture posted by, for instance, a creature sweetheart may be a pooch, a feline, or some other creature. Typically, an individual is bound to warm up to other individual on the off chance that they share comparable interests. This task assists with interfacing a person to others having comparable interests with him/her and make social media progressively "recognizable" to the user[7]. With the improvement of social networks, an enormous assortment of approaches has been created to characterize clients' personalities dependent on their social exercises and language use propensities. Specific methodologies vary concerning distinctive machine learning calculations, information sources, and capabilities. The objective of this paper is to research the consistency of the personality characteristics of Facebook clients dependent on various highlights and proportions of the Big 5 model. They investigate and look at four machine learning models and play out the relationship between every one of the capabilities and personality characteristics. The outcomes for the prediction precision show that regardless of whether tried under similar informational collection, the personality prediction framework based on the XGBoost classifier beats the normal standard for all the capabilities, with a most elevated prediction exactness of 74.2%. The best prediction execution was gone after the additional variant attribute by utilizing the individual social network investigation highlights set, which accomplished a higher personality prediction exactness of 78.6%[8]. In this paper,

computerized, reasoning based investigation which expects to consequently recognize personality characteristics from video by utilizing machine learning and picture handling methods is clarified. To determine the character and feeling of individuals by examining their conduct is now a significant undertaking for particularly HR. On the off chance that an individual's non-verbal communication is sufficiently broke down, much data can be acquired about that individual's personality characteristics. The fundamental undertaking of this investigation is to foresee the five factor model personality measurements from video pictures by utilizing machine learning methods and counterfeit neural networks. In this investigation, video pictures and the passionate conditions of the individual got from recordings were used and a man-made brainpower based framework was created to have the option to anticipate consequently the personality characteristics of an individual from recordings with 0.8951 exactness by utilizing fake neural networks, machine learning and picture preparing methods. Inside the extent of the investigation, prediction of 5 significant personality qualities, extraversion, agreeableness, conscientiousness, neuroticism and openness were emphasized[9]. These days, foreseeing personality from social media end up being a drifting exploration zone in computational phonetic. This paper suggests personality is mix of trademark and conduct of individual contrasts in individuals. Customarily, personality appraisal through meeting or self-report inventories performed by brain research specialists are costly, tedious, troublesome and work concentrated. With the advancement of Internet utilization, individuals can communicate their exercises, feeling,

contemplations, and sentiments through social media. Posts, remarks and notices made by clients of social media can uncover personal data. This paper gives an outline on the personality prediction dependent on language highlights. This paper may go about as a help material for the individuals who wish to think about personality prediction from the Facebook notice content[10].

## PROPOSED SYSTEM

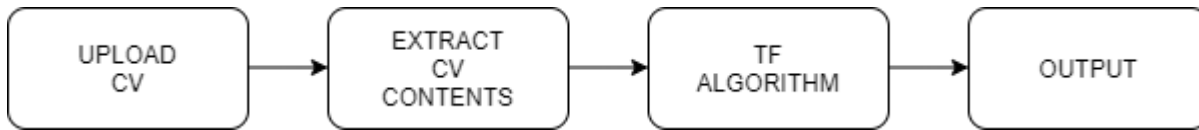
Different procedures are utilized with the end goal of personality prediction. Ordinary personality appraisal relies upon self-report stock, which costs a great deal to gather data. Various models make use of Big five personality test which uses various machine learning technologies or techniques. In our system we expect to predict the personality of any person using logistic regression as well as K-Nearest Neighbor algorithm.

This system uses the big five personality traits, namely, Openness, Agreeableness, Neuroticism,

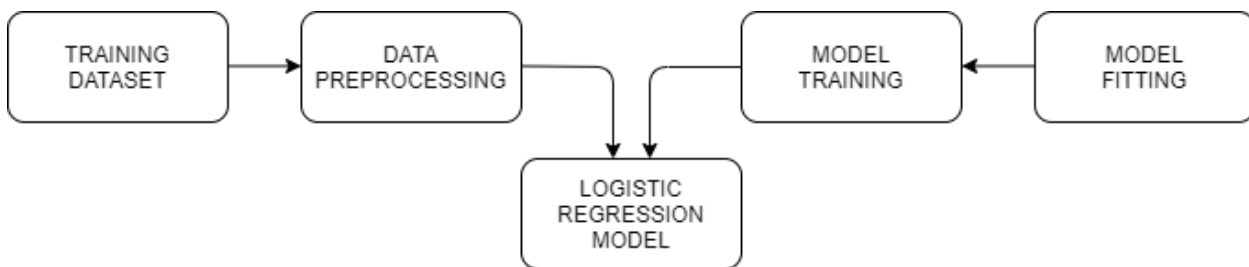
Extraversion and Conscientiousness for prediction of personality of a person. These characteristics are the best acknowledged and most generally utilized model of personality in scholastic brain research. The test consists of 20 questions that the user must rate on how true they are about them on a 3-point scale where Agree, Neural and Disagree values 0, 1, and 2 accordingly. According to the obtained marks of the user the personality is predicted. 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 KNN and Logistic Regression classifiers.

This system also includes TF algorithm which calculates the term frequency in a document or a file. Here, we extract the table of contents from a document file uploaded by the user, and then calculate the term frequency of each word in the document using TF algorithm. The aim of this algorithm in this project is to check the word frequency in the user's technical skills and work history and hence decide on what is he good at.

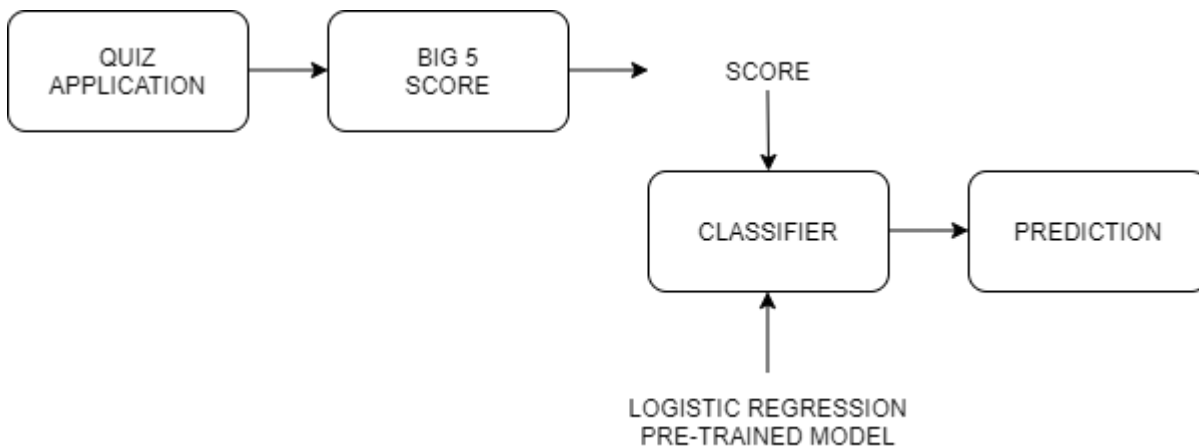
## SYSTEM ARCHITECHTURE



Fig(a): Term Frequency Algorithm



Fig(b): Model training



Fig(c): Block diagram of Personality prediction



## TF-IDF Algorithm

TF-IDF represents Term frequency-inverse document frequency. The tf-idf weight is a weight regularly utilized in data recovery and content mining. Varieties of the tf-idf weighting plan are frequently utilized via web crawlers in scoring and positioning a document's pertinence given an inquiry. This weight is a factual measure used to assess how significant a word is to a document in an assortment or corpus. The significance expands relatively to the occasions a word shows up in the document yet is counterbalanced by the frequency of the word in the corpus.

## Personality and The Big Five Model

The term personality is gotten from the Latin word *persona*, which implies the cover utilized by entertainers in a theater. A lot of characteristics that portray an individual and includes feelings, conduct, personality and the brain characterizes a personality. Because of the assorted variety of credits, it is vital to check personality as it doesn't give any authoritative structure through which individuals can be ordered and analyzed.

The arrangement of human feelings is immense, because of which a comparable issue happens when one attempts to distinguish the opinion installed in a message, in this manner making it try to pick the essential feelings for an order. Along these lines, so as to mechanize estimation investigation, for example, numerous analysts acknowledge a streamlined portrayal of assumptions by methods for their extremity. So also, for determining personality. Personality can change contingent upon various circumstances.

In investigation of the personality structure, meaning of the Big Five Model or Five Factor Model came into utilization. The Big Five characteristics can be depicted as follows:

Openness is identified with creative mind, innovativeness, interest, resistance, political radicalism, and thankfulness for culture. Individuals scoring high on openness like change, acknowledge new and irregular thoughts, and have a decent feeling of style.

Conscientiousness gauges the inclination for a composed way to deal with life as opposed to an unconstrained one. Scrupulous individuals are bound to be efficient, solid, and predictable. They appreciate arranging, look for accomplishments, and seek after long haul objectives. Non upright people are commonly progressively nice, unconstrained, and innovative. They will in general be increasingly lenient and less limited by rules and plans.

Extroversion quantifies an inclination to look for incitement in the outside world, the organization of others, and to communicate positive feelings. Outgoing people will in general be all the more cordial, benevolent, and socially dynamic. They are normally vivacious and chatty; they wouldn't fret being at the focal point of consideration and make new companions all the more without any problem. Contemplative people are bound to be singular or saved and look for situations portrayed by lower levels of outside Stimulation

Agreeableness identifies with an attention on keeping up positive social relations, being neighborly, sympathetic, and helpful. Pleasant individuals will in general trust others and adjust to their requirements. Unpalatable individuals

are increasingly centered around themselves, less inclined to bargain, and might be less naïve. They additionally will in general be less limited by social desires and shows and are increasingly emphatic

## LOGISTIC REGRESSION

Logistic regression turns into an arrangement procedure just when a choice limit is brought into the image. The setting of the limit esteem is a significant part of Logistic regression and is reliant on the arrangement issue itself.

The choice for the estimation of the limit esteem is significantly influenced by the estimations of exactness and review. In a perfect world, we need both exactness and review to be 1, yet this only from time to time is the situation. If there should be an occurrence of a Precision-Recall tradeoff we utilize the accompanying contentions to settle on the edge.

1.Low Precision/High Recall: In applications where we need to diminish the quantity of bogus negatives without essentially lessening the number bogus positives, we pick a choice worth which has a low estimation of Precision or high estimation of Recall. For instance, in a malignant growth analysis application, we don't need any influenced patient to be delegated not influenced without giving a lot of regard to if the patient is in effect unjustly determined to have disease. This is on the grounds that, the nonappearance of malignant growth can be recognized by further clinical ailments however the nearness of the malady can't be identified in a previously dismissed applicant.

2. High Precision/Low Recall: In applications where we need to diminish the quantity of bogus positives without essentially lessening the

number bogus negatives, we pick a choice worth which has a high estimation of Precision or low estimation of Recall. For instance, on the off chance that we are grouping clients whether they will respond decidedly or contrarily to a personalized promotion, we need to be certain beyond a shadow of a doubt that the client will respond emphatically to the notice in light of the fact that something else, an adverse response can cause misfortune potential deals from the client.

## RESULTS

The result of this project includes several functionalities. Initially a web page will be appeared where a candidate will be able to fill the details like name, age etc. Then the candidate will have option to upload his resume. After uploading resume the candidate need to submit the online form. After submitting the candidate personal details from, the candidate will be able to see a test screen consists of several questions. The questions are categorized into 5 types, each relates to personality traits of big five model. After answering all the question by the candidate, the software will calculate the score of a candidate based on the selection made in the test. In the next page, A candidate's personality will be measured in percentages with respect to Big5 model such as Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism and shown on the web page. The web page shows the technology in which the candidate is good at by applying term frequency algorithm on the data extracted from the resume file.Describes a candidate based on the test input given by applying logistic regression.



## CONCLUSION

This project utilizes term frequency algorithm for calculating the frequency of words utilized in the resume of an individual. Term frequency algorithm is applied on technical skills and work history of a candidate to identify the best technical skill of the candidate. logistic regression on the Big five personality test marks dataset. This results in providing a word describing the personality of the candidate according to the marks scored by that person. This system gives an accuracy of around 90%.

## REFERENCES

- [1] P. S. Dandannavar, S. R. Mangalwede, P. M. Kulkarni, "Social Media Text - A Source for Personality Prediction", 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)
- [2] S. Jothilakshmi, R. Brindha, "Speaker trait prediction for automatic personality perception using frequency domain linear prediction features", 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)
- [3] OyaCeliktutan, EvangelosSariyanidi, HaticeGunes, "Let me tell you about your personality!<sup>†</sup>: Real-time personality prediction from nonverbal behavioural cues", 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)
- [4] Alessandro Vinciarelli, GelarehMohammadi, "A Survey of Personality Computing", IEEE Transactions on Affective Computing
- [5] EsenTutaysalgir, Pinar Karagoz, Ismail H. Toroslu, "Clustering based Personality Prediction on Turkish Tweets", 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)
- [6] Ahmed Al Marouf, Md. Kamrul Hasan, Hasan Mahmud, "Comparative Analysis of Feature Selection Algorithms for ComputationalClustering based Personality Prediction From Social Media", IEEE Transactions on Computational Social Systems

- [7] YukteeSahu. Yash Ramani, Viral Parekh, Nishit Maru, “Clustering based Personality Prediction from Social Media Images: A Content Driven Approach”, 2019 6th International Conference on Computing for Sustainable Global Development (INDIACom)
- [8] Michael M. Tadesse, Hongfei Lin, Bo Xu, Liang Yang, “Clustering based Personality PredictionBased on User Behavior on the Facebook Social Media Platform”, IEEE Access
- [9] ErtanTaş, Mustafa ErselKamaşak, “Prediction Of Personality Traits From Videos By Using Machine Learning Algorithms”, 2019 4th International Conference on Computer Science and Engineering (UBMK)
- [10] Zan Mo Mo Aung, Phyu HninnMyint, “Clustering based Personality PredictionBased on Content of Facebook Users:A Literature Review”, 2019 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)