Personality Prediction Using Machine Learning

Hritik Koul
P.V. G College of Engineering and Technology
Pune, India

Prof. Milind Ramesh Mahajan
Assistant Professor
P.V.G College of Engineering and Technology
Pune, India

Ishwari Pawar
P.V.G College of Engineering and Technology
Pune, India

Raviraj Kokare
P.V.G College of Engineering and Technology
Pune, India

Nikhil Chaudhari
P.V.G College of Engineering and Technology
Pune, India

***Abstract*—Human Resource Management is evidently upheld by and given more open doors by the improvement of the Job Characteristics Model (JCM) which thus depends on the idea of present-day job plan. Luckily, the advancement in present day data frameworks, computerized advances, the widespread access of electronic innovation and web prompted the tendency of the worldwide Human Resource Management improvement and make the framework more appropriate. Associations need to guarantee that they utilize the ideal individual for the right job.**

No association is indistinguishable as far as labor force, ability, climate, techniques and market type. Also subsequently, one enrollment technique can't be applied to all. The framework will assist with lessening the HR office responsibility. The framework will help the HR division to effortlessly waitlist the applicant in view of the element positioning strategy. The HR needs to include the positioning the size of 1 to 10 for the elements which are needed for a specific job position. This framework will yield the idea of the competitor in light of info given. Subsequently, the framework will empower a more viable method for shortlisting competitors from countless candidates giving a specialist labor force to the association.

Keywords—Human Resource Management, Job Characteristics Model.

I. Introduction

The most common way of matching position searchers to opportunities is a troublesome undertaking because of I) absence of adequate data of occupation searchers, ii) absence of adequate data of opening iii) trouble of the matching system. Accessible data about work searchers is regularly restricted to their Curriculum Vitae (CV) as well as introductory letter. The framework will help the HR office to effectively waitlist the competitor in light of the positioning strategy. The Human Resources (HR) necessities to include the score the size of 1 to 10 for different elements that are considered during enlistment. This situation will choose the right idea of a contender for a specific work profile in view of the post prerequisites.

Selection representatives assess and channel work searchers, positioning them on different standards. This incorporates the amount of the prerequisites are fulfilled, guaranteeing the up-and-comer is the best counterpart for the empty post. Required abilities are those abilities a task searcher should have to be considered for the gig. Human evaluators would acknowledge a task searcher who fulfills the greater part of the fundamental necessities. Not at all like human evaluators, matching calculations anticipate that this data should be introduced expressly in the opportunity information. That is, the opening should bunch abilities into required classes, and afterward, rank the abilities inside the gathering for the matching system to be as powerful. The undertaking is being created to help the Human Resource office to choose the right possibility for a specific work profile which thus gives a specialist labor force to the association. The proposed strategy matches work searchers in light of elements extricated from CV and meetings led in view of the prerequisites of the business. The proposed framework incorporates a character expectation test that incorporates a rundown of inquiries and situations that assist with anticipating the character of the candidate. The inquiries are of various character attributes.

The proposed framework can be utilized in any organization space. As of now, IT organizations have their own enlistment frameworks however all in all unique site. This framework will be a little module which can be incorporated into their authority sites. Banks and advertising organizations have frameworks wherein candidates are approached to choose and apply just for the ideal position. In the event that the position they need isn't accessible right now, the candidate can't present his resume. Here, this framework assesses the up-and-comers in light of the meeting they give. Questioner poses specific inquiries and rates the up-and-comer on the size of 1 to 10 in light of his/her response. When the rating is presented, the framework will produce the result name in view of the kind of character applicant has. Then, at that point, the organization can take the choice about tolerating the applicant and planning the possibility to their separate position.

II. LITERATURE REVIEW

Personality prediction is an extremely far reaching and fluctuating field of study. In the previous ten years, there has been a lot of work done on computerized personality discovery and evaluation of online media clients. Prior chips away at personality prediction from web-based media information generally utilized AI procedures. (Golbeck et al., 2011) proposed a technique where the personality of clients could be anticipated through the information gathered from their Facebook profiles. For this review, two AI calculations, M5'Rules, and Gaussian Processes were utilized. The outcomes showed that the exhibition and connections created by M5'Rules were more grounded when contrasted with those in the Gaussian Process where it showed no relationships. However, their review was powerful, they didn't zero in on the organization thickness between the clients, which would have contributed tremendously to the exploration in the space of personality prediction.

Around the same time, (Golbeck et al., 2011) again proposed a model to foresee the personality utilizing the information clients share on the most famous microblogging webpage, Twitter. To gather information for this examination, a personality test for 50 Twitter clients was conducted and information was gathered from their Twitter profiles of similar clients utilizing the Twitter API. LIWC (semantic Inquiry and word count) apparatus was utilized to include extraction and insights on 81 highlights of text were created. In this review, feeling examination of tweets was additionally performed and two relapse calculations, ZeroR and Gaussian Processes were utilized. The outcomes showed that the methods utilized in this model can be utilized to foresee the personality of Twitter clients from their tweets yet the model can show better outcomes on a bigger dataset. (Quercia et al., 2011) introduced a review to break down the personality of various kinds of Twitter clients (well known clients, forces to be reckoned with, and so forth) and the connection between these clients in light of how they cooperate on web-based stages. Relapse investigation with 10-overlay cross approval procedure was performed and the root mean square mistake was determined which ends up being 0.88 which was significantly low. This study saw that there are various essential likenesses notwithstanding contrasts among various Twitter clients.

The soundness of personality attributes is the most contended point among therapists and scientists. Cobb-Clark et al., 2011) introduced a review, aftereffects of which show that the personality stays stable among working grown-ups yet not steady. Different other numeric prediction procedures, LinR, REPTree, and DTable were introduced in the model proposed by (Wald et al., 2012). The information was gathered by studying 537 Facebook clients. Out of three numeric prediction strategies, the Dtable model showed better execution. Personality prediction from online information ultimately became significant in an expert climate. (Shen et al., 2013) introduced a review to anticipate the personality of clients from email messages. Since messages comprise of private information, the creator created email extraction apparatuses. This review utilized sack of-words highlights, grammatical feature labeling, and Naïve Bayes classifier. Their review was the main effort to foresee the personality of an email author. (Bai et al., 2013) proposed two displaying approaches, gradual relapse and perform multiple tasks relapse that was utilized for personality prediction of 444 Sina Weibo clients. The information was gathered by directing a study that incorporates a personality stock test and information from similar clients' profiles. The outcomes show a moderate relationship among the 5 characteristics and the normal mean outright blunder of the performing various tasks' relapse model ended up being 13.8.

(Lima and Castro, 2014) proposed a model to anticipate personality through the advanced impressions of an online media client. In this methodology, the model was prepared utilizing three AI calculations: Naïve Bayes, SVM, and a Multilayer Perceptron Neural Network. This paper presents a framework called PERSOMA, a personality prediction framework for web-based media information examination. This framework is a multi-name classifier, in light of the way that every personality aspect in the Big Five Model is isolated into one twofold classifier. The proposed framework extricates the meta-credits from the tweets posted by clients. Finally, the order calculation is utilized to precisely foresee characters through the tweets. The presentation was assessed by applying it on 41 gatherings of tweets which were made by consolidating tweets from three unique writing sources: Obama-McCain Debate, Sanders, and SemEval2013. These datasets were isolated by manual grouping. Results showed that the most troublesome quality to anticipate was transparency while the other four attributes showed high exactness and the normal precision of the framework was viewed as 83. One more intriguing approach to gathering information was presented by (Wan et al., 2014).

They proposed a model to break down the substance of a Chinese long range informal communication site Weibo. Information was gathered by creeping the situations with 131 Sina Weibo clients, after which they directed a personality test in view of the Five-factor model with similar clients. The removed substance was then examined by LIWC word reference which created 5 classes from 71 particular elements of text. In this model two AI calculations, Logistic Regression and Naïve Bayes were utilized. Credulous Bayes showed better outcomes on accuracy while the two calculations had comparative outcomes on review. (Najib and Nawab, 2015) proposed a framework to recognize age, orientation, and personality characteristics through the tweets of a client. For the examination, the PAN-2015 dataset was utilized which is an assembled corpus in which the tweets are characterized by creator and time of creator, language, orientation, and relating personality attributes. This study underlined the significance of content-based elements which can be utilized to separate among texts composed by individuals having a place with unmistakable profiles. Four AI classifiers: J48, Random Forest, SVM, and Naïve Bayes were utilized for preparing the model. The outcomes showed that the framework didn't play out that well on the testing information as it did on the preparation information.

III. PROPOSED ARCHITECTURE

Search Space definition.

To perform hyperparameter tuning, we want to characterize the hunt space, in other words which hyperparameters should be advanced and in what range. Here, for this somewhat little model, there are as of now 6 hyperparameters that can be tuned:

- the dropout rate for the three dropout layers
- the quantity of channels for the convolutional layers
- the quantity of units for the thick layer
- its initiation works

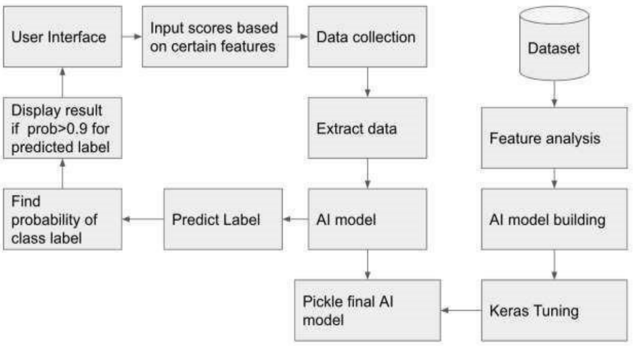

Fig. 1 Architecture

In Keras Tuner, hyperparameters have a sort (potential outcomes are Float, Int, Boolean, and Choice) and a special name. Then, at that point, a bunch of choices to assist with directing the pursuit should be set:

- a negligible, a maximal and a default an incentive for the Float and the Int types
- a bunch of potential qualities for the Choice kind
- alternatively, an examining strategy inside direct, log or switched log. Setting this boundary permits you to add earlier information you may have about the tuned boundary. We'll find in the following area how it tends to be utilized to tune the learning rate for example
- alternatively, a stage esteem, i.e the insignificant advance between two hyperparameter values.

Model Compilation.

Then, at that point, how about we move to display aggregation, where other hyperparameters are likewise present. The gathering step is the place where the streamlining agent alongside the misfortune work and the measurement are characterized. Here, we'll involve absolute entropy as a misfortune capacity and precision as a measurement. For the enhancer, various choices are accessible.

Algorithmic Steps.

1. Score for each feature can be given as an input from User Interface.
2. Input will be passed to AI model instance and in turn AI model will predict the output label for the candidate
3. On the model building front, Dataset acquired from Kaggle is used to train the AI model
4. Data visualization and Data pre-processing will be performed to make data ready and in the acceptable form
5. AI model will be trained by using deep learning algorithm named ANN
6. Tuning for the AI model will be done using advanced keras tuning techniques aiming to achieve the highest possible accuracy.
7. Once the model attains best accuracy, it is pickled and stored for future reference.
8. AI models will be integrated with UI to make future predictions.

IV. SYSTEM IMPLEMENTATION

1. Dataset Finalization: The dataset from Kaggle has been finalized for AI model training.
2. Data Visualization and Data Cleaning: Dataset is analyzed thoroughly for missing values, outliers, correlation etc.
3. Dataset sample:

	Gender	Age	openness	neuroticism	conscientiousness	agreeableness	extraversion	Personality
0	Male	17	7	4	7	3	2	extraverted
1	Male	19	4	5	4	6	6	serious
2	Female	18	7	6	4	5	5	dependable
3	Female	22	5	6	7	4	3	extraverted
4	Female	19	7	4	6	5	4	lively

Fig. 2 Sample Dataset

4. Sample of distribution of each trait within each of the personality label categories.

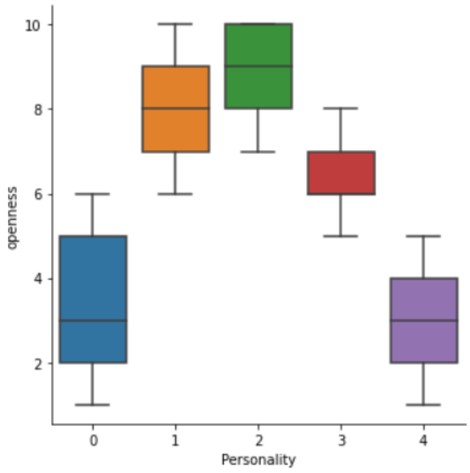

Fig. 3 Sample of distribution of openness with each personality label

5. Sample correlation matrix of dataset.

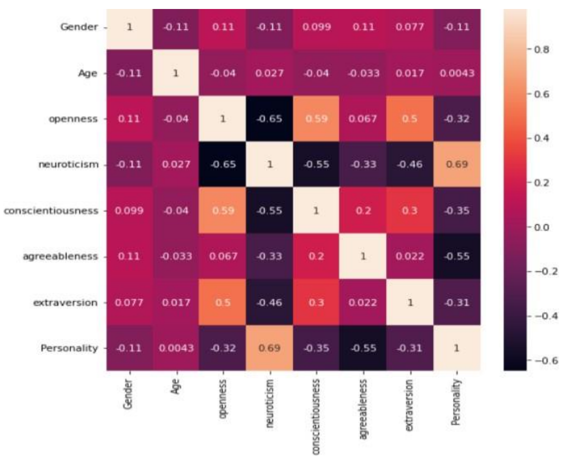

Fig. 4 Correlation matrix of dataset

V. CONCLUSION

The system will help the HR department to select the right candidate for a job. This in turn will provide expert workforce for the organization. The recruiter need not sort through thousands of resumes. The recruiter enters a score and the personality trait output is the basis of selecting or rejecting a resume. We will make use of deep learning-based ANN algorithms to train our AI model and they have high time complexity. The system can be integrated with the recruiting company's existing website. Thus, the genuineness of the employer is guaranteed as the system is part of the company's official website. The system would be used in many business sectors that will require expert candidates, thus reducing the workload on the human resource department.

.

REFERENCES

[1] Ahmad,N., & Siddique, J. (2017). *Personality Assessment using Twitter Tweets.* Procedia Computer Science, 112, 1964-1973. doi:10.1016/j.procs.2017.08.067.

[2] An, G., & Levitan, R. (2018). *Lexical and Acoustic Deep Learning Model for Personality Recognition*. Interspeech 2018. doi:10.21437/interspeech.2018-2263

[3] Arnoux, P. H., Xu, A., Boyette, N., Mahmud, J., Akkiraju, R., & Sinha, V. (2017, May). 25 tweets to know you: *A new model to predict personality with social media*. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 11, No. 1).

[4] Arroju, M., Hassan, A., & Farnadi, G. (2015). *Age, gender and personality recognition using tweets in a multilingual setting*. In 6th Conference and Labs of the Evaluation Forum (CLEF 2015): Experimental IR meets multilinguality, multimodality, and interaction (pp. 23-31).

[5] Azucar, D., Marengo, D., & Settanni, M. (2018). *Predicting the Big 5 personality traits from digital footprints on social media*: A meta-analysis. Personality and Individual Differences, 124, 150-159. doi:10.1016/j.paid.2017.12.018

[6] Bai, S., Hao, B., Li, A., Yuan, S., Gao, R., & Zhu, T. (2013). *Predicting Big Five Personality Traits of Microblog Users*. 2013.

[7] IEEE/WIC/*ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies* (IAT). doi:10.1109/wi-iat.2013.70

[8] [Balakrishnan, V., Khan, S., Fernandez, T., & Arabnia, H. R. (2019). *Cyberbullying detection on twitter using Big Five and Dark Triad features.Personality and Individual Differences*, 141, 252-257. doi:10.1016/j.paid.2019.01.024

[9] Bleidorn, W., & Hopwood, C. J. (2018). *Using Machine Learning to Advance Personality Assessment and Theory*. doi:10.31234/osf.io/ctr5g

[10] Catal,C., Song, M., Muratli, C., Kim, E. H., Tosuner, M. A., & Kayikci, Y. (2017). *Cross-Cultural Personality Prediction based on Twitter Data. Journal of Software,* 12(11), 882-891. doi:10.17706/jsw.12.11.882-891

[11] Carducci, G., Rizzo, G., Monti, D., Palumbo, E., & Morisio, M. (2018). *Twitter Personality: Computing Personality Traits from Tweets Using Word Embeddings and Supervised Learning. Information*, 9(5), 127.

[12] Golbeck, J., Robles, C., Edmondson, M., & Turner, K. (2011). Predicting Personality from Twitter. 2011 IEEE

[13] Golbeck, J., Robles, C., & Turner, K. (2011). *Predicting personality with social media. Proceedings of the 2011 Annual Conference Extended Abstracts on Human Factors in Computing Systems* - CHI EA 11. doi:10.1145/1979742.1979614

[14] Third Intl Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Intl Conference on Social Computing, 149-156. doi:10.1109/passat/socialcom.201.33