



An Overview of Personality Recognition through Machine Learning For E-Recruitment

¹Priyanka Kamble, ²Umesh Kulkarni, ³Ankit Sanghavi

¹Student, ²Professor, ³Professor

¹Department of Computer Engineering,

¹ARMIET College of Engineering and Technology, Shahapur, India

Abstract : In our daily lives it is crucial to identify the personality of a human based on his behavior, therefore the prediction of personality has a large scope in the field like recruitment of organizations or industries. The personality can be positive or negative that can be identified by their actions in different situations and behaviors in different conditions. In the research study it is noticed that, personality traits are based on the Big Five Model i.e. OCEAN model. In the previous study, many investigations done in the field of personality recognition where they have used different techniques and different algorithms to predict the personality of different people by using verbal and non-verbal cues. Some have used segments from the resume, resume classification, text based approach to predict the personality of the different candidates. Some of the case studies used facial expressions where they have used CNN features. One study reviewed on the prediction of personality by using various deep learning algorithms. Apart from this there are certain limitations on the single model as discussed above. As per research terminology, there is no other system which can be used by the interviewee and interviewer on the same platform for personality prediction which helps candidates to get selected in the organization by checking his ability from the resume and interviewer to select the perfect candidate for the organization by analyzing the candidates during the interview. As the system discussed above can be a better solution for the candidates and HR of the organizations.

IndexTerms - Big five personality traits, Personality Recognition, Machine learning, deep learning, CNN.

I. INTRODUCTION

It is difficult for every individual to predict the personality of humans. At the recruitment process it becomes more crucial as the interviewer has to recruit the best suitable candidate for the organization. As the interviewer has responsibility to determine the actual personality of the interviewee when it comes to online interviews hence it becomes difficult.

The personality of the human can be determined by two categories such as verbal and non-verbal where verbal personality can be determined by some of the factors which includes communication skills, use of specific phrases or facial expression, etc. The non-verbal personality factors include the posture, speech tone, etc. of an individual human. The personality can be detected by some other factors too such as handwriting of the person, social media activities like updating posts, profile pictures, reactions or comments on other posts, resume analysis by interviewer or HR and psychometric test.

In recent years, online or intelligent recruitment systems have played an important role in the recruitment process with advanced text processing techniques. The job recruiters posted most of the job requirements on various online platforms in recent years. The online recruitment systems gather a large number of resumes from different sources. In the recruitment process the resume or CV is the golden key for the job seekers. As the recruitment companies spend a lot of time and energy to filter and pull out the data from the resumes according to the job requirements. Therefore many recruitment tools are developed for the retrieval of the information from the resume.

CV/ Resume analysis becoming beneficial for the organization in prediction of personality of candidates for human resource management. Also the CV analysis can be beneficial for the candidate to select the appropriate job title for the organization. Some organizations conduct the online aptitude test and personality assessment tests for evaluating the personality of candidates. Nowadays, multimodal systems are popularly used to determine the personality traits of speakers from seminar videos, YouTube videos and vlogs.

The job seekers can postal their personality by using a resume which helps to express themselves in a good manner. Humans with great communication skills are more likely to have an impressive personality than those who face difficulties in communicating with others. Great communication can build strong bonds among people. The posture and hand gestures of individuals can help them to present their personalities effectively.

In recent years, the recruitment process inserted psychometric assessment which helps to assess whether the candidate is capable of growing vigorously in a specific professional role. Psychometric testing is a calculation of an individual's logical skills and personality traits which can help to understand aspects of mental ability and behavioral style that organizations are unable to measure during conversation and interviews. The psychometric assessments are astronomically used in recruitment and selection process, performance evaluation, employee engagement, employee training and development, career guidance, high-potential identification and succession planning. The psychometric test in recruitment is a strong correlation between psychometric test score and job performance. For example, the high scorer candidate will show the high performance on the job. Hence the psychometric

testing or assessment becoming the testing part of the interviewee in the recruitment process. The psychometric test is easy to administer and scale to match the hiring requirement.

Similarly, every person has unique behavior as the posture, facial expressions and hand gestures of individuals helps candidates to present their personalities effectively during the recruitment process.

The human and computer are linked by Natural Language Processing (NLP) that focuses on how computers can be utilized to understand original language text/speech to perform some effective tasks.

II. RELATED WORK

A lot of work has been already done in prediction of personality which is always been a crucial task and mainly in recruitment process as follows:

In the existing e-recruitment process, they worked on personality prediction through personality assessment tests and cognitive skills tests. In existing systems, interviewers used to make decisions by filtering interviewees on the basis of their CVs and behavior during online interviews sessions. They simply mark the personality on the basis of online interviews and scan the submitted CVs and shortlist the candidates within the suggested system.

Hung-Yue Suen, et al. [1] experiments on the recognition of individual personality traits through automatic analysis of video interviews. The working interviewing system developed using asynchronous video interview (AVI) processing with tensor flow-based AI engine to perform automatic personality recognition (APR) based on the feature extracted from AVIs. The true personality scores are evaluated by the facial expression and self-reported questionnaires of 120 real job applicants [1].

Octavia Arriaga, et al. [2] proposes the general CNN building framework for designing real time CNNs. They validate the model by creating a real time vision system to accomplish the tasks of face detection, gender classification and emotion classification. The IMBD gender dataset and FER-2013 emotion dataset is used by author to develop the model. The author successfully reduces the gap between the slow performances and real time architecture [2].

Dina Al-Hammadi, et al. [3] compares four different machine learning techniques with their performances to recognize personality traits from non-verbal cues. As the author reported the result by using speaker personality corpus provided by the Interspeech 2012 challenge. The proposed work gives an increase in performance. The non-verbal things are focused in this research such as speech attributes, prosody, gestures, facial expressions and body movement. For implementation purpose the proposed system used Support Vector Machine which gives the good result as 60% for all traits whereas naïve Bayes, SVM with linear kernel and SVM with Radial Basis Function (RBF) are also used by researcher which increase the result high for extraversion, conscientiousness, and neuroticism [3].

Shivam Gupta [4] states a fully automatic recognition of facial emotions using the computer vision and machine learning algorithms which classifies eight different emotions. The support vector machine algorithm is best suited for the classification process with accuracy 94.1% according to the author. The fully automatic real time coding facial expressions in the continuous video stream is achieved by author at least for applications in frontal views can be assumed using webcam [4].

Bernhard Kratzwalda, et al. [5] investigates how emotion recognition is done by using neural networks and deep learning. The emotions are affecting human decision making which can be the goal of affective computing. The proposed scheme is a tailored form of transfer learning for affective computing which has past work as the traditional machine learning used as utilized methods with advanced deep learning and recurrent neural networks and transfer learning. The proposed system summarizes the baselines from traditional machine learning and deep learning and uses the inherent nature of affective computing with multiple innovations concerning the network architecture.

Jaejin Cho, et al. [6] improves the acoustic information and conversation transcripts are used to improve the emotion recognition. The system had researched in the field of deep learning networks to generate vectors at frame level from raw spectral representation and merge the features over time into utterance level feature vectors by classifying it with a softmax layer. The proposed system is compared with SVM and DNN training. The system is combined with a transcript based system to improve the acoustic system. The acoustic system in this work is based on the LSTM network and the multi resolution CNN is used for prediction of emotion from ASR transcripts and human annotated transcripts.

Gunaseelan B [7] extracts various features from the text and built a classifier to predict whether a text line in the resume is heading or not. They researched that XG boost outperforms the other classifier in predicting headings. They have extracted the skillsets information from the resume pdf and word files format.

Ricardo Stegh Camati [8] et al. investigates the new TB-APR approach using projective tests to build a corpus. The method developed is labeling of textual corpus using z-test projective instrument conducive to mitigate the limitation of inventories which is very sensitive and offers the possibility of collective applications. The proposed work used a corpus of 363 individuals labeled by the z-test. The proposed model used bag-of-words techniques with some state of art machine learning inductors.

Berkay Aydin et al. [9] presents the multimodal method for predicting personality from the short video clips. According to the big five personality traits to predict personality, the proposed work used audiovisual features and random decision forest regression. This paper is presenting the framework which uses audio features, global motion energy features and facial landmark features to perceive automatic personality prediction. In this proposed work the visual features extracted from the videos, the overall visual activity of the person is computed by using Weighted Motion energy Images (WMEI). The proposed system is developed for the Chalearn First Impressions Challenge where the system is presented with an in depth analysis of the usage of non-verbal audio visual features, energy image based features and facial landmarks in automatic personality perception. The overall observation of all the features results in obtaining accuracies in the range of 85% to 89%.

Mr. Ramraj S et al. [10] predicts the domain adaptation, used for the sensitive nature of resume data. Also the classifier is trained on job description snippets as a large dataset, used to classify the resume data. The paper aims to compare the results gained by all the various state of art algorithms generated by the same data so the efficiency of each algorithm can be evaluated. As per author's research, it is evident that character level CNN gives the better F1 score as compared to other models. As a result, CNN can also be used to classify the text rather than images and exceptional results as compared to text classification algorithms. In future the dataset size can be increased by using Generative Pre-Trained Transformer (GPT-2) algorithm which will help to train the model to learn better with more information.

Made Artha Agastya et al. [11] reviews on personality trait recognition has an essential role in the job screening process which can perform the analysis based on the survey, the handwriting of the participant, or conduct the interview which takes a lot of time

and money. Hence, there should be a system that can help in the screening process. This paper is the study to classify the latest deep learning algorithms in personality trait classification. The author collected 25 key papers from Scopus, IEEE, science direct, Emerald Insight and ACM on the basis of Preferred Reporting Items for Systematic reviews and Meta-Analyses Method (PRISMA).

Christopher Pramerdorfer et al. [12] overviews the state of art in image based facial expression recognition using CNNs and highlights algorithmic differences and their performance. In this paper the previous bottlenecks and consequently for advanced research directions are identified. This paper is demonstrated to overcome the previous bottleneck which outperform without requiring auxiliary training data or face registration with deep CNNs which obtains FER2013 test accuracy of 75.2%. The FER2013 dataset is used to study the FER performance which is the most common dataset in CNNs based FER.

Allan Robey, Kaushik Shukla, et al. [13] directs to examine the organization oriented enrollment that they will hire a passionate and right candidate for a particular job profile. In this paper [13] they have used NLP to structure unstructured data sources but NLP doesn't have a user interface to communicate with the framework. NLP needs to include the permit clients in the organization.

Hung - Yue Suen et al. [14] suggests that there is not only a considerable calculation instrument for approximating interpersonal abilities in organized but also they made decisions about their non-verbal correspondence. They have used the Novel Deep Learning model for great simultaneous permissibility and accuracy to anticipate an interviewee's relational abilities and characteristics.

MICHAEL M. TADESSE et al. [15] proposes the relationship between client personalities with their nature in social networks. The online interactions are monitored by the SNA which gives the accuracy better than LIWC, SPLICE and the combination of SNA+LIWC+SPLICE. For implementation purposes they have used the combination of SNA, LIWC and SPLICE so the performance is low.

Chen-Lin Zhang, et al. [16] refers YouTube videos to predict personality by the author in this paper for this they have developed the DBR model for improving visual regression performance. They have used traditional CNN for development. They divided the video into images and audio features. For images they have used deep regression and for audio features they have used regression. The 10000 videos with time duration 15 seconds from YouTube dataset is used for development by the author.

Dr. V. Suma [17] extends the work of analysis of the Indian e-commerce market using data mining approach for prediction of demand of renovated electronics. The real world factors and real world datasets are used from the three random e-commerce websites which are considered for analyzing the work. Based on the result of the author, the highly accurate prediction can be made with the proposed approach in spite of the impact of varying customer's behavior and market factors. In the CLSC method, the item cost and demand are a few questionable parameters that are to be measured absolutely by implementing advanced forecast strategies.

Dr. Abraham Chandy [18] administers the savvy asset in cloud computing that is used for an enormous information processing system that is fulfilled in this paper influencing machine learning. The work stack request is associated with asset assignment where the proposed work predicts the work stack utilizing the arbitrary timberland and degrades the assets utilizing the genetic calculations.

Heysem Kaya and Albert Ali Salah suggests that proposed work is used for the interactive media work applicant screening task. The multimodal combination is used for proposed work that is followed by the choice tree (DT). The DT is used to create text based clarification of its choices. The DT is better accuracy wise as per research work. The job candidate is analyzed by the short video to produce apparent personality scores and a prediction about whether the candidate will be selected for the job interview or not. The system gives the explanation of its decision that provides a visual and textual explanation [19].

Matthew J.W. McLarnona, et al. [20] studies within person, two wave data that investigates faking on a conscientiousness and faking conditions using latent transition analysis (LTA) that identify different types of faker. The author examined whether counterproductive workplace behavior (CWB) differed across the faking types. The result supported three class solutions in honest responding and faking conditions. As the respondent can be classified as honest respondent, slight fakers and extreme fakers. The high agreeableness and low neuroticism as predictive of stable response patterns are supported by the system.

Mihai Gavrilescu and Nicolae Vizireanu propagate the psychological qualities that are examined consequently by the facial features of people to assess the diversity. The Facial Expression Recognition (FER) framework is used to predict the sentimental state of the individual. The emotions and facial movements are analyzed with proposed architecture with 16PF qualities [21].

Table 1: Comprehensive Analysis of Personality Prediction Based On the Previous Study

| Sr. No. | Years | Methodology | Datasets /Achievements | Limitations |
|---------|-------|--|--|--|
| 1. | 2019 | Convolutional Neural Network (CNN), Tensorflow, Big five | Datasets are R, R2, MSE with above 90% Accuracy | The system used only 120 job applicants to predict their personality. Restricted or limited population can be produced |
| 2. | 2021 | CNN | Used dataset of IMDB gender with 96% accuracy and FER-2013 with 66% accuracy | more parameters in naïve implementation leads to less robust features. The CNNs for gender classification found biased towards western facial features. |

| | | | | |
|-----|------|--|--|---|
| 3. | 2021 | Sampling technique, OCEAN Model | Used Dataset of Speaker Personality Corpus (SPC) with higher result using SVM | The smaller datasets are considered for personality recognition. |
| 4. | 2018 | SVM, noise-tip method, | Cohn-kanade (CK) database is used. Used images, facial landmark datasets with accuracy 93.7% | Complexity level increases with real time emotion recognition. |
| 5. | 2018 | Deep Learning, neural networking, transfer learning, lexicon based methods | Holistic setting across 6 benchmark dataset, ISEAR, semEval-2007 & SemEval-2018 with improvement of performance up to 23.2% in F1 Score and 11.6 % in regression task. | Some of the algorithms like naive network struggling with task of emotion recognition. Decreases performance with the larger dataset |
| 6. | 2018 | Deep neural networks, LSTM network | Datasets used USC-IEMOCAP with accuracy 24% & US English telephone speech with accuracy 3.4% | The degree of system can be improve by the trained transcript performance |
| 7. | 2020 | Random forest, LSTM, N-gram RNN models, NLP, sampling techniques | Test dataset used from the recruitment process with 85% accuracy | Fails to extract the skillset from the resume with all type of file formats |
| 8. | 2020 | Bag-of word techniques, | Datasets are used corpus z-test, AUC-ROC with 0.85 average | Provides the low performance |
| 9. | 2016 | Random forest regression, Big five model, | Dataset used from first impression challenge, video clips from the youtube with 10 fold cross validation accuracy of 0.90197 and test set accuracy of 0.898 | Increase in data can trouble the system |
| 10. | 2020 | CNN, SVM, ML, subsampling technique | Dataset used from Kaggle with 159571 labeled samples, images dataset with 0.886 F1 score. Dataset used PASCAL VOC | Poor performance on the small dataset with CNN |
| 11. | 2020 | Deep Learning, | Dataset used MyPersonality with accuracy 67% and 61% on CxD dataset | Lower accuracy in personality trait recognition |
| 12. | 2016 | CNN based FER method | Used dataset FER-2013 with 75.2% test accuracy | Decreases in performance due to FER dataset bias in real world |
| 13. | 2019 | NLP | .csv Dataset is used with 75.4 % accuracy | Only depends on CV and aptitude tests The candidate's interview does not matter |
| 14. | 2020 | NLP, data collection, AVI-AI technique | Dataset are R, R2, MSE with 95% accuracy in communication skills and 98.8% accuracy in personality traits | Fails to explain candidate's probability which is going to be scrupulous and extraverted |

| | | | | |
|-----|------|---|--|--|
| 15. | 2018 | OCEAN Model., LIWC, SPLICE linguistic features, social network Analysis (SNA) | XG boost classifier with 74.2% accuracy and SNA with 78.6% accuracy | Lower performance |
| 16. | 2019 | Deep bimodal regression (DBR), CNN | APA competition dataset with best regression mean accuracy (0.9130) | Does Not work on discriminative audio for personality prediction |
| 17. | 2018 | Closed loop supply chain (CLSC) method | Real world e-commerce website datasets with higher conclusions trade analytics within the field of restored items. | Could be achieve the better result with the machine learning calculations to analyze the design within the huge information for the more precise expectation of future advertise requests for the restored items |
| 18. | 2019 | Cloud computing, random forest algorithm | Achieves high performance in time, cost, memory utilization | Issued the security problems that has to be tended to amid the assignment of the comes about |
| 19. | 2018 | Visual feature extraction, experimental result using regression, experimental result using classification | Dataset used from youtube with 67.3% accuracy | Fails to work on both the interviewers and interviewees during interaction |
| 20. | 2019 | Counterproductive workplace behavior (CWB), latent transition analysis (LTA) | Results highly supported three class solution in honest responding and faking condition | Increases in CWBs might have negative financial implications Limited result by single source data |
| 21. | 2017 | CNN features | Datasets used MMI with accuracy 75% and JAFFE with 70% accuracy | Lower expectation precision for few 16PF characteristics |

III. CONCLUSION

The objective of this paper is to study various methods to extract and analyze verbal and non-verbal data to predict the personality of an individual. The research methodology discussed the OCEAN model and how it is used to predict personality. Various techniques are developed to predict personality, some of the techniques providing lesser accuracy and some of them are providing high accuracy. Among all of the higher accuracy providing algorithms worked on limited people / dataset or used on verbal or non-verbal features. As per the previous study, the emotion feature's performance is very low but can be achieved with the help of advanced methodologies. Basically, advanced convolutional neural network (CNN), VGG neural network, Support vector machine (SVM), Naïve Bayes, Logical regression, deep learning algorithms are used by the methods presented in previous study. Multimodal are effective according to case studies and to make test automatic. The intelligent multimodal agent can identify the personality traits better based on verbal and non-verbal features. So, we conclude that several methodologies have been done in the field of personality recognition but none of the system is completely perfect and there is still scope to improve such systems.

REFERENCES

- [1] Hung- Yue Suen, Kau-En Hung and Chien-Liang Lin, "TensorFlow-based Automatic Personality Recognition Used in Asynchronous Video Interviews" IEEE Access vol. 7, pp. 61018-61023, 2019,
- [2] Octavio Arriaga, Paul G. Ploger and Matias Valdenegro, "Real Time Convolutional Neural Networks for Emotion and Gender Classification", 2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC), Jan 2021
- [3] Dina Al-Hammadi, Prof. Roger K. Moore, "Using Sampling Techniques and Machine Learning Algorithms to Improve Big Five Personality Traits Recognition from Non-Verbal Cues", 2021 National Computing Colleges Conference (NCCC), March 2021.
- [4] Shivam Gupta, "Facial emotion recognition in real time and static images", 2nd International Conference on Inventive Systems and Control (ICISC), June 2018.
- [5] Bernhard Kratzwald, Suzana Ilic, Mathias Kraus, Stefan Feuerriegel, helmut Prendinger, "Deep Learning for affective computing: text based emotion recognition in decision support", Decision Support System pp. 1-28, Sept 2018.
- [6] Jaejin Cho, Raghvendra Pappagari, et. al., "Deep Neural Networks for Emotion Recognition Combining Audio and Transcripts", Interspeech, Computer science, Sept 2018.

- [7] Gunaseelan B, Supriya Mandal and Rajagopalan V, "Automatic Extraction of Segments from Resumes using Machine Learning", IEEE, 2020.
- [8] Ricardo Stegh Camati and Fabrício Enembreck, "Text-Based Automatic Personality Recognition: a Projective Approach", 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC) October 11-14, 2020. Toronto, Canada, 2020.
- [9] Berkay Aydin, Ahmet Alp Kindiroglu, Oya Aran and Lale Akarun, "Automatic Personality Prediction from Audiovisual Data using Random Forest Regression", 2016 23rd International Conference on Pattern Recognition (ICPR) Cancún Center, Cancún, México, December 4-8, 2016.
- [10] Mr. Ramraj S, Dr.V. Sivakumar and Kaushik Ramnath G, "Real-Time Resume Classification System Using LinkedIn Profile Descriptions", IEEE International Conference on Computational Intelligence for Smart Power System and Sustainable Energy (CISPSSE-2020),2020
- [11] Made Artha Agastya, Dini Oktarina Dwi Handayani and Teddy Mantoro, "A Systematic Literature Review of Deep Learning Algorithms for Personality Trait Recognition", IEEE 2020.
- [12] Christopher Pramerdorfer and Martin Kampel, "Facial Expression Recognition using Convolutional Neural Networks: State of the Art", arXiv organization, Dec 2016.
- [13] Allan Robey, Kaushik Shukla, Kashish Agarwal, Keval Joshi and Professor Shalmali Joshi, "Personality Prediction System through CV Analysis", International Research Journal of Engineering and Technology (IRJET), Feb 2019.
- [14] Hung- Yue Suen 1 , Kuo- En Hung 1 and Chien- Liang Lin 2, "Intelligent video interview agent used to predict communication skill and perceived personality traits", 2020.
- [15] Michael Mesfin Tadesse, Hongfei Lin and V, "Personality Predictions Based on User Behavior on the Facebook Social Media Platform",October 2018.
- [16] Zhang, Chen-Lin, et al. "Deep bimodal regression for apparent personality analysis". European conference on computer vision. Springer, Cham, 2016.
- [17] Suma, V., and Shavige Malleshwara Hills. "Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics." Journal of Soft Computing Paradigm (JSCP) 2, no. 02 (2020): 101-110.
- [18] Chandy, Abraham. "Smart resource usage prediction using cloud computing for massive data processing systems.", J Inf Technol 1, no. 02 (2019): 108-118.
- [19] Heysem Kaya,Albert Ali Salah,"Multimodal Personality Trait Analysis for Explainable Modeling of Job Interview Decisions", 2018 (pp. 255- 275).
- [20] McLarnon MJW, DeLongchamp AC, Schneider TJ (2019) Faking it! "Individual differences in types and degrees of faking behavior", Pers Individ Differ 138:88–95.
- [21] Mihai Gavrilescu, Nicolae Vizireanu, "Predicting the Sixteen Personality Factors(16PF) of an individual by analyzing facial features", 30 August 2017.

