LET'S GET PLACED-NLP APPROACH TO PREDICT PERSONALITY TRAITS

Mrs. Sunita Suralkar, Muskan Kewalramani, Chandni Panjabi, Sanjana Narang, Simran Galani

Assistant Professor, Student, Student, Student, Student

Computer Engineering

Vivekanand Education Society's Institute of Technology Mumbai, India

Abstract: Introducing a web app where students can learn key placement by opening a website and students can offer skills tests and test their skills or they can learn about the company's recruitment process to be defined by job role or they can test confidence and various personality traits. While we are preparing for this whole process of humorous interviews, we suggest that A candidate's feelings are removed from his discourse using divisiveness, using deep CNN. Its first layers process the text in chronological order. Each word is represented by an input as a vector of the default length using Word2vec, and the sentences are represented as a variable number of word vectors. Finally, the 5 main personality traits: Openness, Awareness, Acceptance, Consensus, Neuroticism (OCEAN) commonly used by many employers, is compiled using a Convoluted Neural Networks (CNN) and a personalized graph, which gives a complete overview of the individual's personality which will be written and approved by the company. The results are finally presented in the form of a detailed review.

I. INTRODUCTION

A. Overview:

Let's Get Placed is a placement practice kind of website for the students who nearly want to sit for placement or in general want to go for interviews. According to personal, professional, academic and project details, the system will provide various companies eligibility criteria to candidate the list of companies in which he/she can apply in future. The system will provide practice aptitude tests and also users have to voice record the answer and have to upload that recording and Machine Learning algorithms will be used to convert speech to emotional quotient to know the nature or confidence of the student. The system will provide the probability of the selection of candidates.

In our website there are total modules viz.

   (1) E-learning Module

   (2) Aptitude Practice Test

   (3) A module which will consider all possible factors related to behaviour, user's emotions and qualities needs to be implemented so that any nervousness and loopholes in the real interview can be reduced.

B. Problem Definition:

In today's world with increasing competition and with the new advances in technology and communication, oral on-line interviewing, which simulates the actual physical interview, is now possible. Interviewers can now see how candidates interact with their questions and study their body language Our website named LET'S GET PLACED will have various functionalities and features that use the latest technologies that will very accurately analyse the information and produce an error free output. This will help students in preparing themselves to appear for an interview.

C. Current problems:

In today's age, the entire process is man-powered. Humans need to analyse huge amounts of video submissions. Going through all of them is a tedious job and can cause missing a few files. The judge has his own inherent biases when it comes to the interviewing process. Deceptive IM is difficult to detect even through computers and data science. It is not possible to be recognised by a human. It has to be done with a lot of data or information. Interviews which happen in person need people to meet at a centralised location. Videos can be taken from anywhere and at any time.

II. METHODOLOGY

Details on personality types were collected (MBTI and five majors) for more details, Data status checked. There are too many details of the MBTI available that are highly scientifically based, but there are very few details about the BIG FIVE symptoms. Machine learning algorithms thrive on data so a method was developed to integrate MBTI and BIG Five data. Data from three different sources was converted into alternative form and advanced with the requirements of ML algorithms. Features from the text were extracted to inject vectorize data with word bags and GloVe method Several separate reading algorithms was used and trained to predict unknown future text. Rescue results were checked.

III. PROPOSED SOLUTION

The proposed model, which works on audio provided by the candidates, can be divided into five major parts viz. Various Assessments of aptitude process and e learning module. The NLP method of obtaining Deceptive Impression Management using text-to-text-based text from input is using specific concepts. ANN using MFCC as a major the variation itself helps us to understand the nature of candidates. Happy, Sad, Calm, Anger, Fearful, Confidence. The final output will be in the form of a graph which includes the result of personality traits prediction and aptitude final result.

The entire problem has been broken down into three subsections to process the data better. The user is allowed to use the website after entering details and can test his/her practice skills in aptitude tests or interviews or can choose to learn about the topics asked in interviews. The input from the third section(interview) is divided into audio and text. The candidate is made to answer 3-4 questions for a duration of 90 seconds wherein he can choose to speak as long as he wants. Thinking time of 30 seconds is given. It is these answers that are then analysed further using Natural Language Processing and Audio processing to get a complete overview of the interviewee.
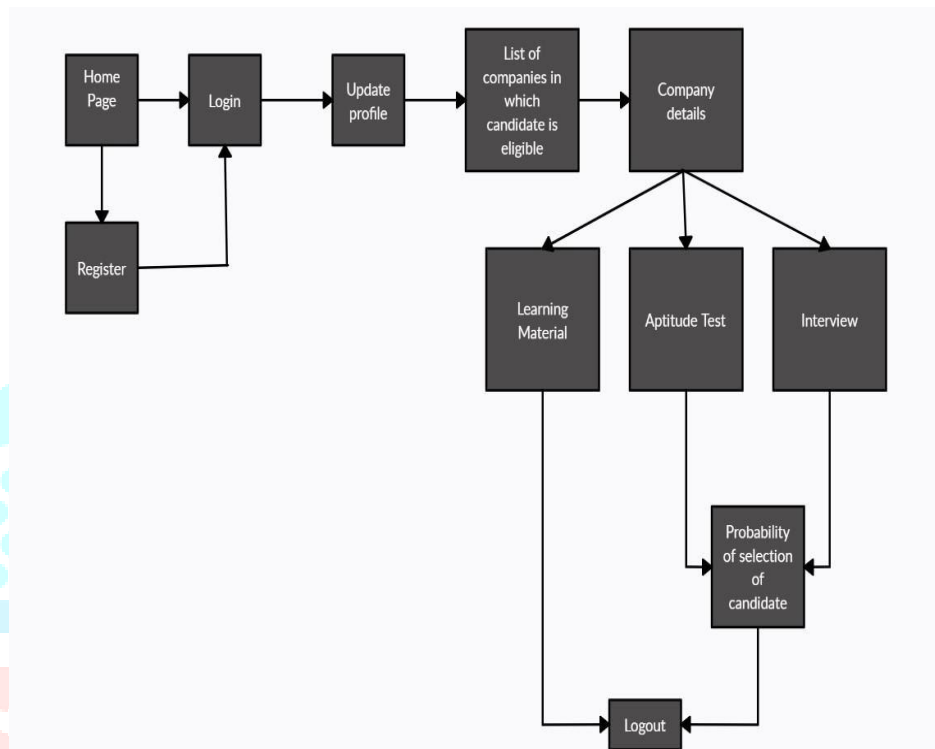

fig.1: block diagram

IV. ALGORITHM

A. DATASET

The Big Five personality traits, also known as the five-factor model (FFM) and the OCEAN model, is a taxonomy, or grouping, for personality traits. The five factors are:

- Openness to experience (inventive/curious vs. consistent/cautious)
- Conscientiousness (efficient/organized vs. easy-going/careless)
- Extraversion (outgoing/energetic vs. solitary/reserved)
- Agreeableness (friendly/compassionate vs. challenging/detached)
- Neuroticism (sensitive/nervous vs. secure/confident)

Myers - Briggs Type Indicator (MBTI) is a self-explanatory questionnaire that reveals different psychological preferences in how people view the world and make decisions.
The combination of the four preferred pairs leads to 16 possible combinations. The 16 genres are usually called a four-letter abbreviation - the first characters of each of their four favorite characters (except in the intuition mode, which uses the abbreviation "N" to separate it from the introduction). For example:
ESTJ: extraversion (E), hearing (S), thinking (T), judgment (J) INFP: introversion (I), intuition (N), hearing (F), understanding (P).

B. ALGORITHM USED

Our method includes processing of input and filtering data, feature extraction, and segmentation. We use two types of features: a set number of document-level style elements, and the semantic word elements combined to represent different lengths of input text. This representation of various lengths is embedded in CNN, where it is processed sequentially by combining words into n-gram, n-gram sentences, and sentences into complete document. The values obtained are combined with style elements at the document level to form the document representation used for the final partition. Specifically, our approach involves the following steps

• Promotion. This includes sentence segregation and data purification and integration, such as lowercase lettering.

• Document level feature release. We have used the first element of the Maresis base, which includes international features such as word count and average sentence length.

• Filtering. Some of the essay sentences may not have human traits. Such sentences can be overlooked in the removal of the semantic element for two reasons: first, they represent sound that reduces the performance of the editor, and secondly, the removal of those sentences greatly reduces the input size, and thus training time, without affecting results. Therefore, we remove such sentences before the next step.

• Word-level feature release. We represent individual words by embedding words in a continuous vector space; in particular, we tried word2vec embedding.5 This provides the feature of the document's variable length: the document is represented as a number of dynamic sentences, represented as a variable number of vector elements of fixed words.

• Separation. Divided, we use deep CNN. Its first layers process the text in chronological order. Each word is represented in the input as a vector feature of a limited length using Word2vec, and the sentences are represented as a variable number of voice vectors. In another layer, this variable vector length is reduced to the vector of the average length of each sentence, which is a type of sentence inserted in a continuous vector space. At that level, the documents are shown as a variable number of embedded fixed lengths. Finally, on a deeper layer, this document vector of various lengths is reduced to a document document vector of limited length. This vector of the measured length feature is also combined with the document level-level features that provide the long-document document vector, which is used for the final partition.
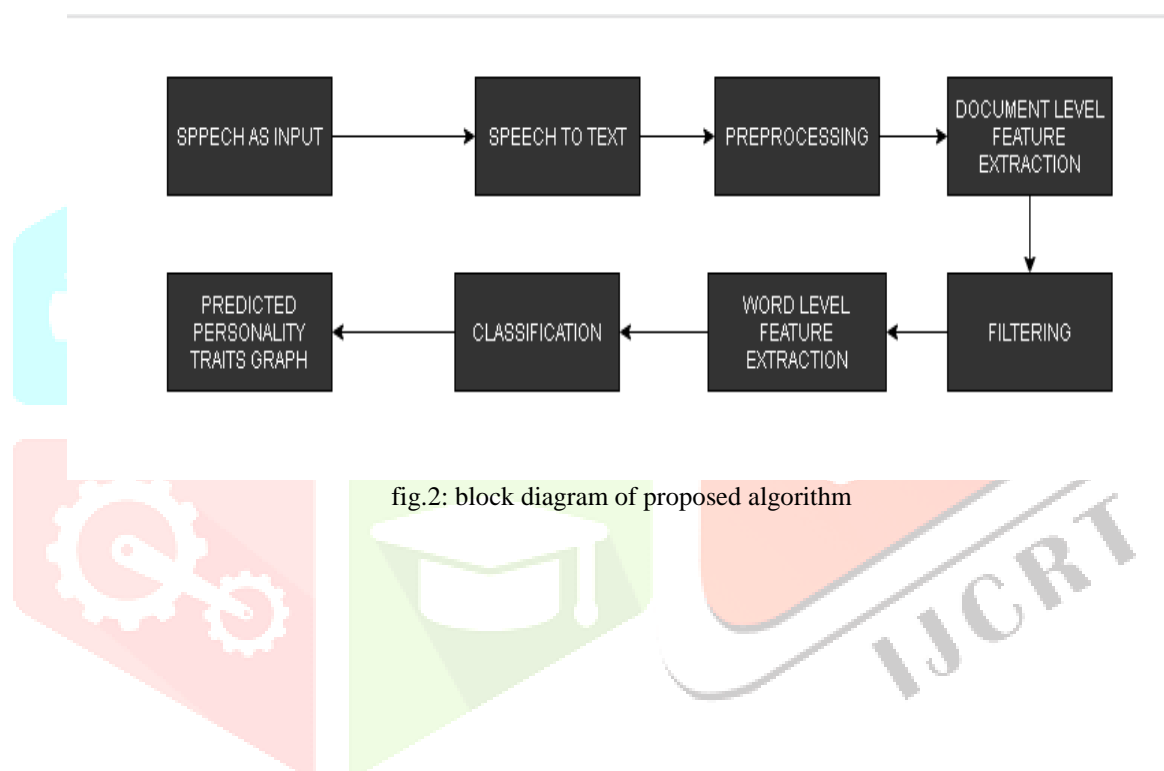

fig.2: block diagram of proposed algorithm

Document vector d	Filter	Classifier	Convolution filter	Personality traits				
				EXT	NEU	AGR	CON	OPN
N/A	N/A	Majority	N/A	51.72	50.02	53.10	50.79	51.52
Word n-grams	Not used	SVM	N/A	51.72	50.26	53.10	50.79	51.52
Mairesse[12]	N/A	SVM	N/A	55.13	58.09	55.35	55.28	59.57
Mairesse (our experiments)	N/A	SVM	N/A	55.82	58.74	55.70	55.25	60.40
Published state of the art per trait[12]	N/A	N/A	N/A	56.45	58.33	56.03	**56.73**	60.68
CNN	N/A	MLP	1, 2, 3	55.43	55.08	54.51	54.28	61.03
CNN	N/A	MLP	2, 3, 4	55.73	55.80	55.36	55.69	61.73
CNN	N/A	SVM	2, 3, 4	54.42	55.47	55.13	54.60	59.15
CNN + Mairesse	N/A	MLP	1, 2, 3	54.15	57.58	54.64	55.73	61.79
CNN + Mairesse	N/A	SVM	1, 2, 3	55.06	56.74	53.56	56.05	59.51
CNN + Mairesse	N/A	sMLP/FC	1, 2, 3	54.61	57.81	55.84	**57.30**	62.13
CNN + Mairesse	Used	sMLP/MP	1, 2, 3	**58.09**	57.33	**56.71**	56.71	61.13
CNN + Mairesse	Used	MLP	1, 2, 3	55.54	58.42	55.40	56.30	**62.68**
CNN + Mairesse	Used	SVM	1, 2, 3	55.65	55.57	52.40	55.05	58.92
CNN + Mairesse	Used	MLP	2, 3, 4	55.07	**59.38**	55.08	55.14	60.51
CNN + Mairesse	Used	SVM	2, 3, 4	56.41	55.61	54.79	55.69	61.52
CNN + Mairesse	Used	MLP	3, 4, 5	55.38	58.04	55.39	56.49	61.14
CNN + Mairesse	Used	SVM	3, 4, 5	56.06	55.96	54.16	55.47	60.67

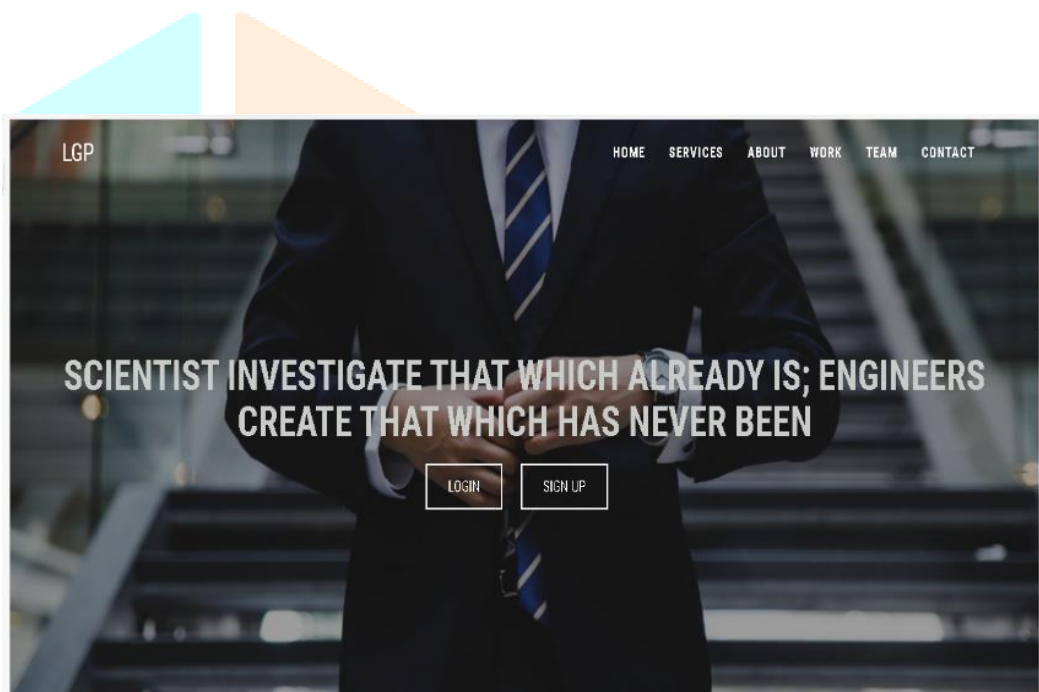fig.3: Accuracy obtained with different configurations.

V. RESULTS

fig.4: first screen on opening the website for user

After successful login, the candidate is required to update his profile by entering all the required details for placement activities. The candidate can improve his/her knowledge by e-learning module. Also, he/she can have practice sessions to be confident about placement activities.
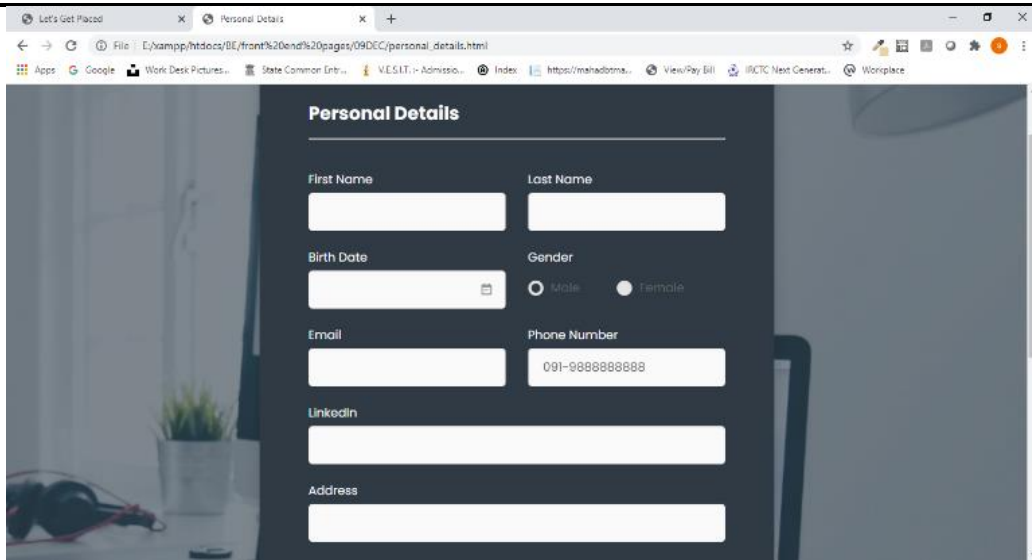
fig.5: update profile

Once the candidate has login, he/she has to update their profile by entering all the details regarding personal details, academic details, skills, achievements, project details, languages known, etc.
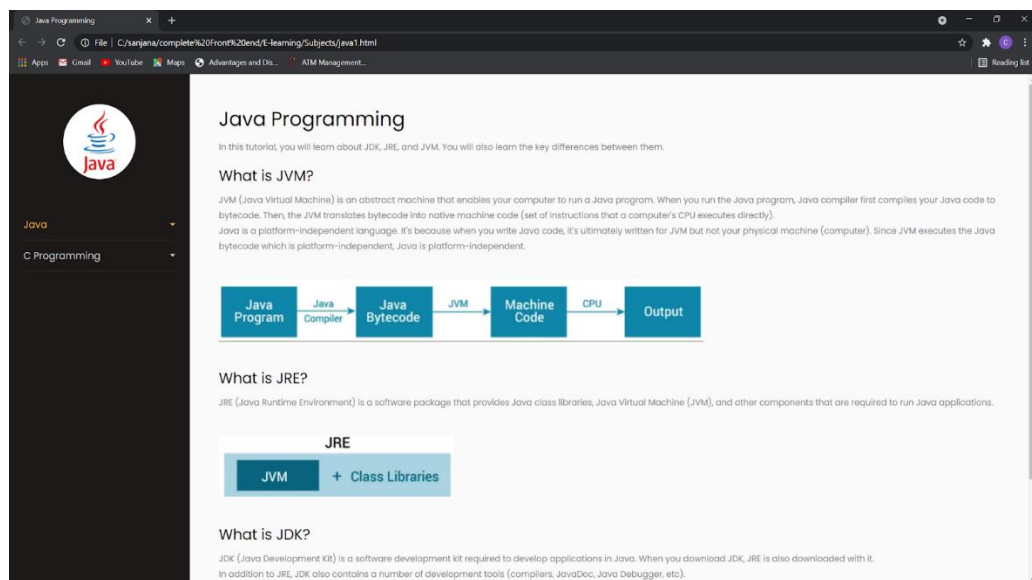
fig.6: e-learning

E-learning and Practice modules are built using JavaScript, jQuery and php for server-side scripting. They can learn new technology, languages for their technical interviews.
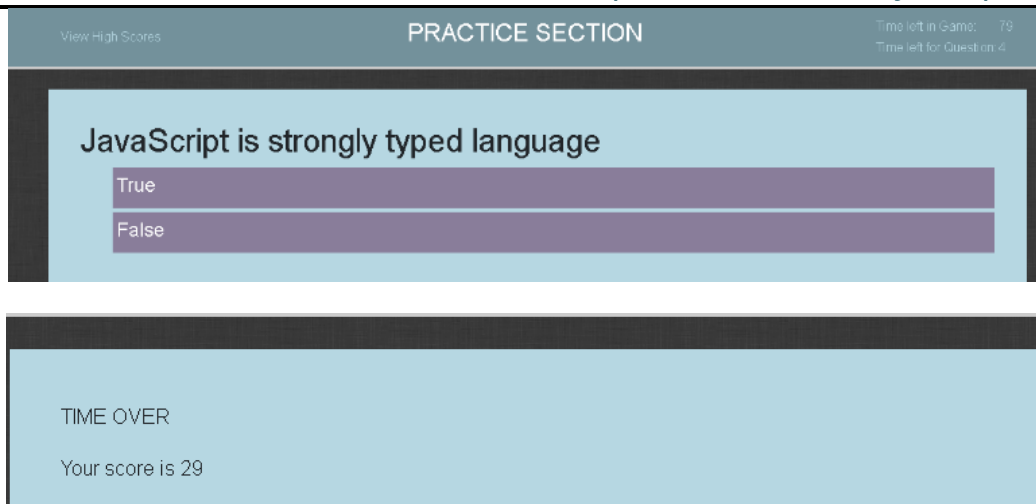
fig.7: practice session

Practice sessions will help the candidate to clear the aptitude round of placement activities. These modules will enhance the knowledge of candidates and boost their confidence.
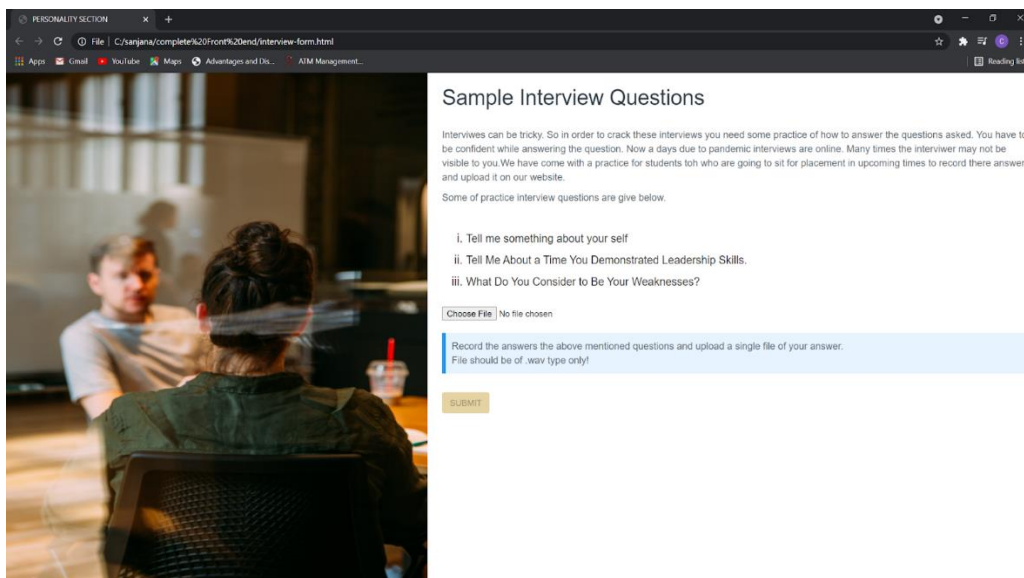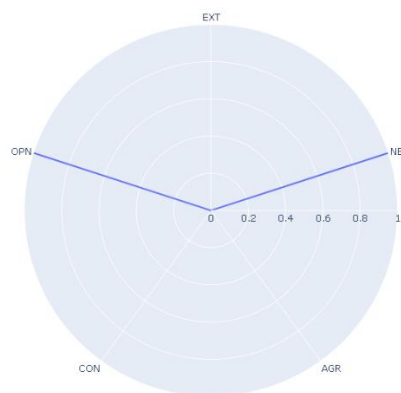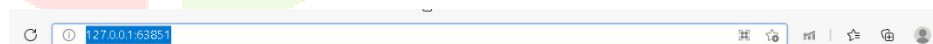

Fig.8: practice interview

fig.9: personality traits prediction

VI. CONCLUSION

Online Recruitment process has become the most effective means of recruiting, screening, hiring, and retaining highly skilled professionals. The purpose of any appointment is to find a candidate who is eligible. The means of recruitment is also equally important. Keeping the current trend in perspective, emphasizing E employment is important and good for effective and efficient hiring. Let's get placed will produce comprehensive reports for all candidates to help them understand where they stand and give them a hint how an actual recruitment process will be. The emotion of the speaker as he answers a question is noted to find his natural response to them the students are able to practice and attempt the aptitude test. Students will be also notified about when and how the procedure of the selected company will take place i.e. the details of the recruitment process of the company.

REFERENCES

[1] A. B. Kmail, M. Maree, M. Belkhatir and S. M. Alhashmi, "An Automatic Online Recruitment System Based on Exploiting Multiple Semantic Resources and Concept-Relatedness Measures," 2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI), Vietri sul Mare, 2015, pp. 620-627, doi: 10.1109/ICTAI.2015.95.

[2] Kevin Tromba, Joel Dumoulin, Elena Mugellini,Omar Abou Khaled and Salah Hawila, Stress Detection using Speech Analysis, HumanTech Institute, 2018, Stress Detection using Audio Analysis.

[3] Society of Human Resources, A Guide to Conducting Behavioral Interviews with Early Career Job Candidates, (2016) Behavioral Interview Guide: Early Career Job Candidates Society for human resource management.

[4] Speech Recognition. In Wikipedia [Web]. Wikimedia Foundation inc.. Retrieved May 15, 2007 https://en.wikipedia.org/wiki/Speech_recognition

[5] Urmila Shrawankar and Vilas Thakare, Noise Estimation and Noise Removal Techniques for Speech Recognition in Adverse Environment, (2010), https://link.springer.com/content/pdf/10.1007/978-3-6 42-16327-2_40.pdf

[6] Navonil Majumdar,Soujania Poria, Alexander Gelbukh,Erik Cambria ,Deep learning based Document-Modelling for Personality detection from text,Nanyang Technological University,(2016),Personality Detection with CNN

[7] Stephanie, Z-test: Definition & two proportion z-test, 2014