Sway of Natural Language Processing And Sentiment Analysis on Automated Smart Hiring System

Asmita Vishalkumar Patil
All India Shree Shivaji Memorial Society
Shivajinagar, Pune.

Dr.Deepali Ujalambkar
(Assistant Professor)
All India Shree Shivaji Memorial Society
Shivajinagar, Pune.

1. ABSTRACT

Smart Hire research project suggests automating prior assessment rounds in recruiting procedures to save time, expenses and resources. The suggested approach employs AI evaluation techniques such as NLP and sentiment analysis to screen candidates based on the hard and soft abilities necessary for a given job. The goal of this research is to automate the recruiting process using machine learning approaches for personality prediction. The technology uses video and tone analysis to evaluate candidates' confidence and other characteristics. In addition, a Chatbot that replicates the functionality of resume builder is also created, which provides several templates to pick from and simplifies the process of making resumes. The suggested approach has the potential to dramatically decrease the time and expense involved with traditional manual hiring recruitment operations. The system is meant to automatically send emails to selected candidates with a single click. The new approach also intends to speed up recruiting for a wider number of candidates.

Keywords: AI, ML, NLP, d-RSA, OCEAN, STT, HR, TF-IDF, GOCI, MOTP, MOTA, ATA

INTRODUCTION

The recruiting process is an important component of every organization's performance, and employing the proper personnel is key for meeting corporate objectives. Traditional recruiting techniques, on the other hand, are typically time-consuming and resource-intensive, and they can result in biases and mistakes in applicant selection. According to survey, 118 people try to fill a job opening every hour, with an average cost per hiring of $4,120 . These data show the need for more efficient and cost-effective recruitment systems that can provide accurate applicant selection in a shorter period of time [1].

In today's competitive business environment, recruiting is an essential activity for every corporation. Significant time and resource commitments are required throughout the recruitment process. Companies frequently struggle to hire the appropriate people, and interviews may be time-consuming and tiring. In recent years, there has been an increased interest in using artificial intelligence (AI) and machine learning (ML) to expedite and enhance the recruiting process [2]. AI-powered systems may select candidates based on their talents, personality traits, and other relevant criteria, saving time and effort during pre-screening rounds. Furthermore, AI-powered solutions can assist reduce bias in the recruiting process, resulting in more accurate applicant selection.

Consequently, the use of AI and machine learning in hiring processes has become more widespread across industries. An innovative method for automating the pre-screening phases of the hiring process is presented in this research study [2]. The recommended method uses artificial intelligence (AI) assessment algorithms to filter applicants according to the hard and soft skills required for a particular position. Moreover, this method uses machine learning and natural language processing (ML) to forecast candidates' personalities and uses video and tone analysis to evaluate applicants' confidence and other traits [3].

The recommended method also provides quick insights, a synopsis of the candidate's biography, and faster hiring in larger quantities. Two algorithms are presented in this work: the answer relevance algorithm, which assesses the importance of candidate replies, and the personality insights algorithm, which gauges OCEAN skills [3].

Making resumes is made easier using Smart Hire's resume builder, which offers a variety of templates. The proposed method has the potential to significantly reduce recruitment expenses and time while increasing the accuracy of candidate selection. The overall goal of the Smart Hire research project is to use artificial intelligence (AI) to drive automation and efficiency in the hiring process. The design, implementation, and evaluation of the recommended system are described in this study project, demonstrating its usefulness and worth to HR/Recruiting teams in businesses. It provides brief insights and a synopsis of the candidate's background. Both candidates and HR/Recruiting teams (employers and workers) will find this program to be straightforward to use and intuitive [4].

2. LITERATURE SURVEY

The login system is crucial for identifying and authenticating users, and RSA and One Time Pad (OTP) are cryptographic algorithms used to protect user information. However, the strong login system can be slow due to the high computation time required for RSA decryption. A new modified login system, d-RSA [5], is proposed to speed up the login process with the same security level as RSA and OTP. This system uses a new private key with lower Hamming weight, making it faster than the existing system. Experimental results show that the new modified system can speed up the login process, but may be slower during the registration process.

In [6], Automating pre-screening rounds can save companies time and resources by ensuring candidates possess the necessary hard and soft skills for a specific job role. Two algorithms, personality insights and answer relevancy, are proposed for assessing these skills. The personality insights algorithm helps in assessing OCEAN skills, while the answer relevancy algorithm evaluates the significance of answered questions. A website was implemented to simulate the recruitment process using these algorithms. This approach can significantly reduce the time and resources required for interviews.

Spatial temporal video-text detection technique with two principal steps as potential text region detection and a filtering process. The first step involves dynamically dividing consecutive video frames into sub blocks to detect changes. A significant difference between homologous blocks indicates the appearance of a text region. The filtering process uses temporal redundancy to form an effective text region. Experiments show the approach's effectiveness, with a precision rate of 89.39% and recall of 90.19 [7].

In [8], an end-to-end scene text recognition method is highlighted for videos based on multi frame tracking, utilizing temporal information to improve performance. The method uses a unified deep neural network to detect and recognize text in each frame of the input video, followed by multi frame text tracking through associations of texts in the current frame and several previous frames. Experiments on ICDAR datasets show that the proposed method outperforms state-of-the-art methods in end-to-end video text recognition.

In [9], Human Resource Management depends heavily on competency assessment, but traditional methods can be time-consuming and costly. This study presents an interview bot that automates the interview process using artificial intelligence technology. The bot uses Behavioral Event Interview method to identify competency levels based on past behavioral experiences. The bot uses data training and testing to determine competency levels, predicting competence levels with acceptable accuracy. This reliable tool supports the assessment center process, especially in physical and social distancing constraints. It provides flexibility in time and place, is delivered in Indonesian, and is more cost-efficient than traditional interviews.

Utilization of ChnSentiCorp and ChnFoodReviews for sentiment information-based network model (SINM) to analyze Chinese text sentiment. The model uses Transfomer encoder and LSTM components, utilizing a Chinese emotional dictionary to automatically find sentiment knowledge in Chinese text. SINM uses a hybrid task learning method to learn valuable emotional expressions and predict sentiment tendencies. Experiments on ChnSentiCorp and ChnFoodReviews datasets have shown SINM to achieve better performance and generalization ability than most existing methods, enhancing the overall sentiment knowledge in Chinese text [10].

3. METHODOLOGY

4.1.1 System Architecture: -

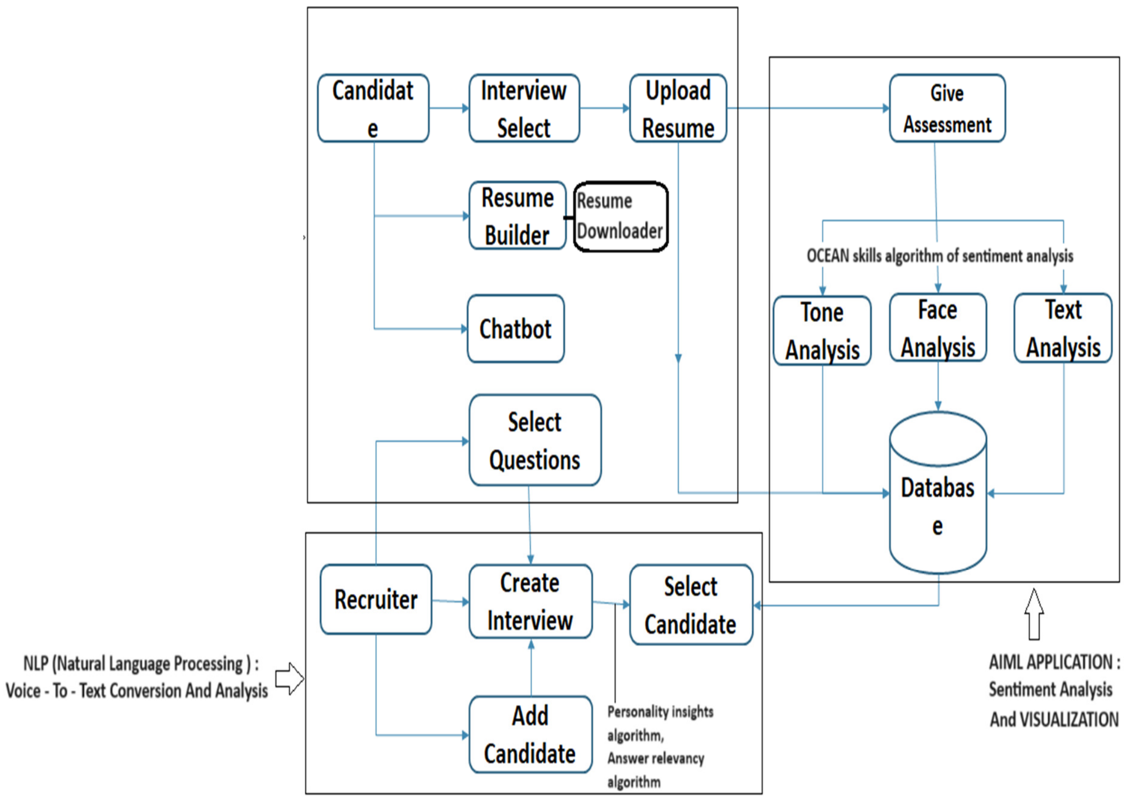


Fig. 1. System Architecture

Fig. 1 describes three important Blocks of Smart Hiring System Architecture [11]: -

- Resume Build-Up: This involves gathering candidate information, creating resumes, allowing candidates to upload their resumes, and providing options for downloading resumes.
- Interview Process and Selection: This stage involves selecting candidates, scheduling interviews, ensuring interview questions and answers are appropriate, conducting interviews via video

calls with speech-to-text (STT) conversion, and ultimately selecting candidates. The Personality insights algorithm and the answer relevancy algorithm are used to assess personality traits and the relevance of answers.

- Sentiment Analysis and Visualization: This aspect focuses on analyzing candidate sentiments through tone, facial expressions, and text analysis. The results are visualized using graphs. Algorithms are used to evaluate candidates' skills based on the OCEAN model.

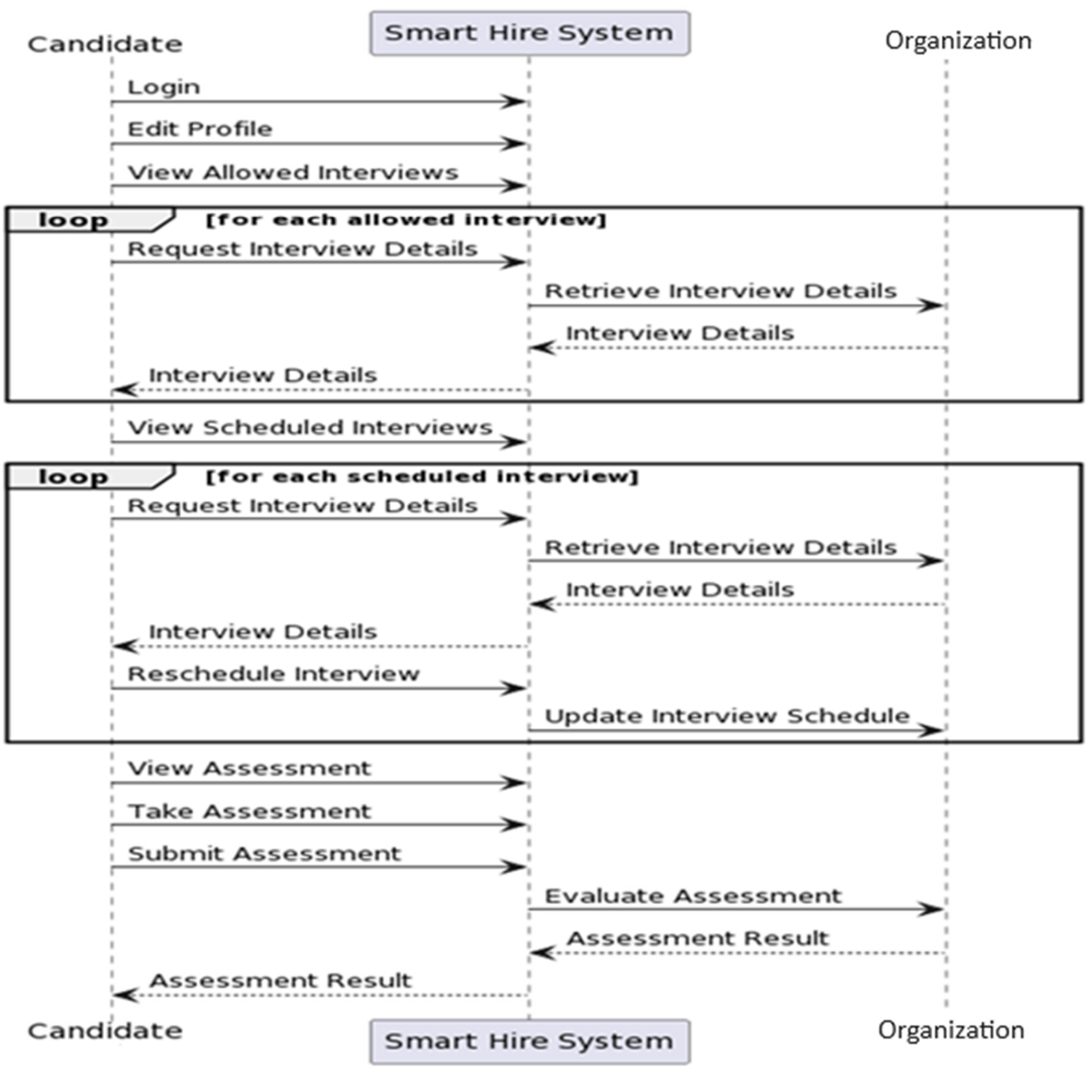


Fig. 2. Module Transitions

Fig. 2 shows Smart Hire System has 2 major modules as:

1. Organization (Recruiter) Module

The Recruiter Module provides several functions to the recruiter persona, allowing them to replicate the complete recruiting process on our system. The recruiter can construct a job posting as well as a list of interview questions. The recruiter can screen candidates for the job ad based on their submissions and the scores produced by our evaluation algorithms. The module has a simple interface with interactive visualizations and statistical analysis of candidate data.

2. Candidate Module

Candidate Module which allows candidates to create their profile or resume and apply for job openings that match their skills and qualifications. Candidates are required to answer video interviews for the job profile they applied for, which are processed by our system, and scores are displayed to the recruiter. Additionally, candidates can utilize our chatbot feature to resolve any queries they may have, and our Resume Builder functionality to generate a resume from multiple templates.

4.1.2 Algorithmic approach: -

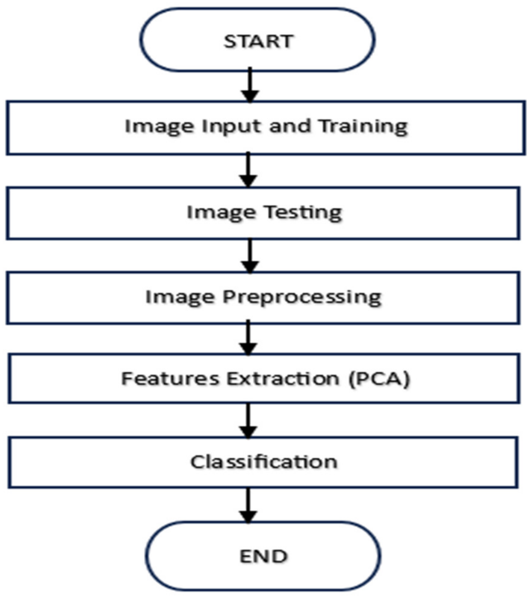
The Smart-Hire research project entails building a website that generates the whole hiring process using algorithms. During the screening process, a question bank is utilized by the recruiter to create questions that evaluate candidates' hard and soft skills. Candidates are given a time restriction in which to respond to these questions throughout the recording process. After that, we use video interviews that have been recorded to test our assessment algorithms. While the answer relevancy algorithm looks at candidates' hard abilities, our personality test algorithm analyzes their soft capabilities. To help the recruiter select the best candidate for the job, the following methods are utilized to compute the total rating characteristics for each skill.

- The personality insights algorithm

It is a psychometric instrument that draws on Jungian psychology to help you develop a tone of voice by helping you identify the fundamental traits of your audience. The

Five Factor Model (FFM) facilitates our comprehension of personality variations. It functions similarly to a tool that divides personality qualities into five categories: neuroticism, agreeableness, extraversion, conscientiousness, and openness. With the use of this model, we may gain a better understanding of a variety of outcomes, such as social interactions, mental health, and behavior in many contexts. [12].

Fig 3 describes flow of Personality Insights Algorithm through Smart-Hire: -



- The answer relevancy algorithm

Relevancy refers to the accuracy and effectiveness of search results used by a search engine to decide whether a particular Web page is relevant to a search query that has been typed by the user.

Stepwise Procedure of Answer Relevancy Algorithm in Smart Hire:

Step 1: Start.

Step 2: Receive request body containing interview ID, answer array, and interview timing

Fetch interview details from database using the interview ID.

Step 3: Check if interview details are found.

Step 4:

4a. If Yes:

Remove fillers from candidate's answer array and original answer array.

Calculate time percentage based on interview timing and duration.

Perform string comparison between filtered candidate's answer array and original answer array.

For each answer in the original answer array and the corresponding candidate answer: Calculate the similarity percentage using a custom matching function.

If:

similarity percentage is available:

Add it to the Answer_Result_Array.

Else:

Add 0 to the Answer_Result_Array.

Calculate the sum of percentages in Answer_Result_Array.

Calculate the final percentage by dividing the sum by the total number of answers multiplied by 100.

Send a success response with answer percentage list, overall percentage, and time result.

4b. If No:

Send an error response indicating failure to find interview details.

Catch any errors and log them.

Step 5: End.

This process involves pre-processing of documents in the relevant answer collection, followed by calculating document frequency (DF) and term frequency-inverse document frequency (TF-IDF) [13].

Document frequency (DF) is calculated by totaling the number of times a specific term appears in all of the documents. Finding the term frequency-inverse document frequency (TF-IDF): TF-IDF is a method for determining a word's relevance within a particular document relative to its value throughout all papers. It takes into account both the frequency (term frequency) and rarity (inverse document frequency) of a word across all of our publications. This aids in our comprehension of the words that are important in a certain text and those that are used frequently in several papers.

Pre-processing entails converting all letters to upper or lower case, eliminating apostrophes, punctuation, and common words (such as "and," "the," and "is"), converting numbers to words, and condensing words to their most basic form in order to ensure consistency across all documents. We make use of methods such as stemming(), remove_apostrophe(), remove_punctuation(), remove_stop_words(), convert_lower_case(), and so on. Following all of that, we double-check to make sure we haven't unintentionally added any more punctuation, popular terms, or pointless word ends. The end result is a cleaned-up document with all the words we want, and a bunch of these cleaned-up documents ready to use for finding relevant answers through performance of the state-of-the-art question answering system-BERT (Bidirectional Encoder Representations from Transformers) [14].

.

- OCEAN skills of the candidate's algorithm

Precise assessments of personalities are the main emphasis of the "Openness to experience, Conscientiousness, Extroversion, Agreeableness, and Neuroticism" [15] paradigm. This approach integrates the built-in functionality of the Random Forest, K-Nearest Neighbors, and logistic regression algorithms. The target's qualities are represented by the result of the personality insights algorithm and the answer relevancy algorithm when properly weighted. The end product, sentiment analysis charts using AWS, is entirely the responsibility of the OCEAN algorithm; AWS manages the input and output processing of the charts. All characteristics must be in byte format; therefore, whether our input is in text or picture format, it will be transformed to Text Byte and picture Byte, respectively. This is the only requirement.

4. RESULTS

The study focused on the practical application of the suggested approach of examine the usage and understanding of individual's utilization with convenient and precarious schemes like Web-based Single Sign-On (SSO) for evaluating job prospects for distinct profiles [16]. Each profile evaluated at least 10 individuals. Due to limited resources, human evaluations with experts in the relevant domains were utilized as a baseline to evaluate security and privacy for algorithm-generated findings.

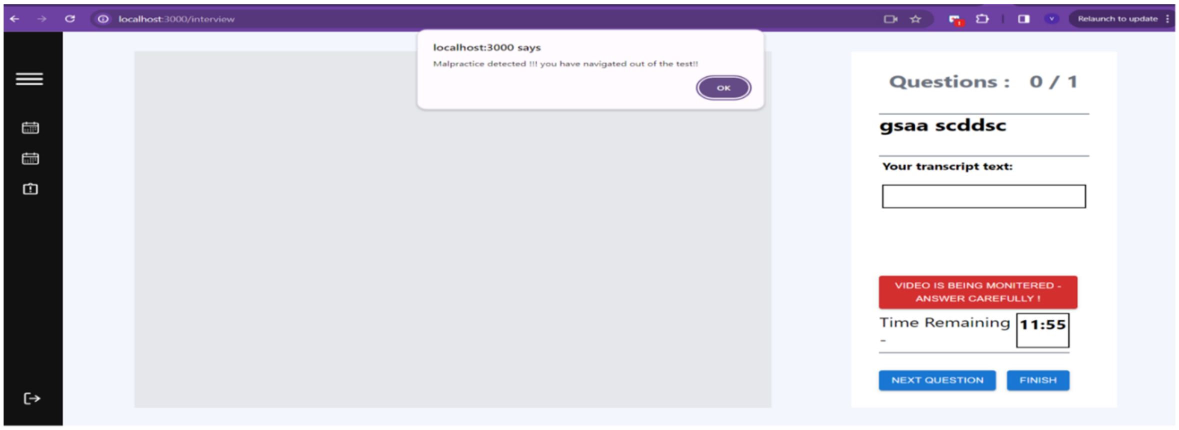


Fig. 4 Result for AI Screening

The "Result for AI Screening" in the figure shows what the AI system discovered by looking at the data or pictures. It might be a simple "yes" or "no" answer, or it could give more details depending on what the AI was programmed to look AI Screening for Hierarchical Adaptive Structure Mesh and has been applied to 4k video resolution as HASM-4K [17]. For example, if it finds any unexpected mouse or keyboard clicks, it could mean something went wrong during the interview process, and it might stop the interview abnormally.

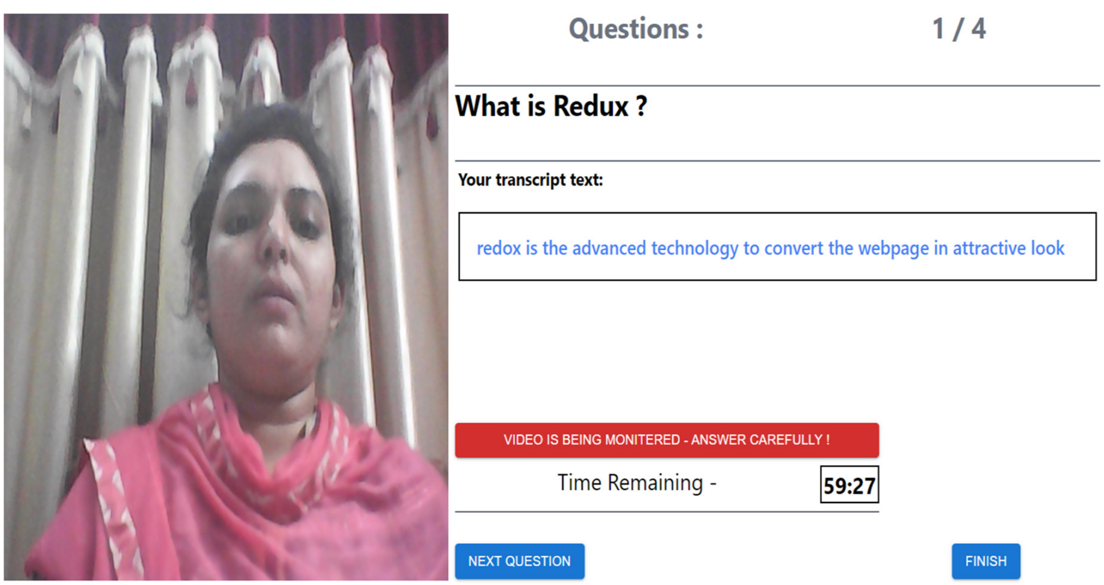


Fig. 5. Result for Natural Language Processing (Speech – To – Text Conversion STT)

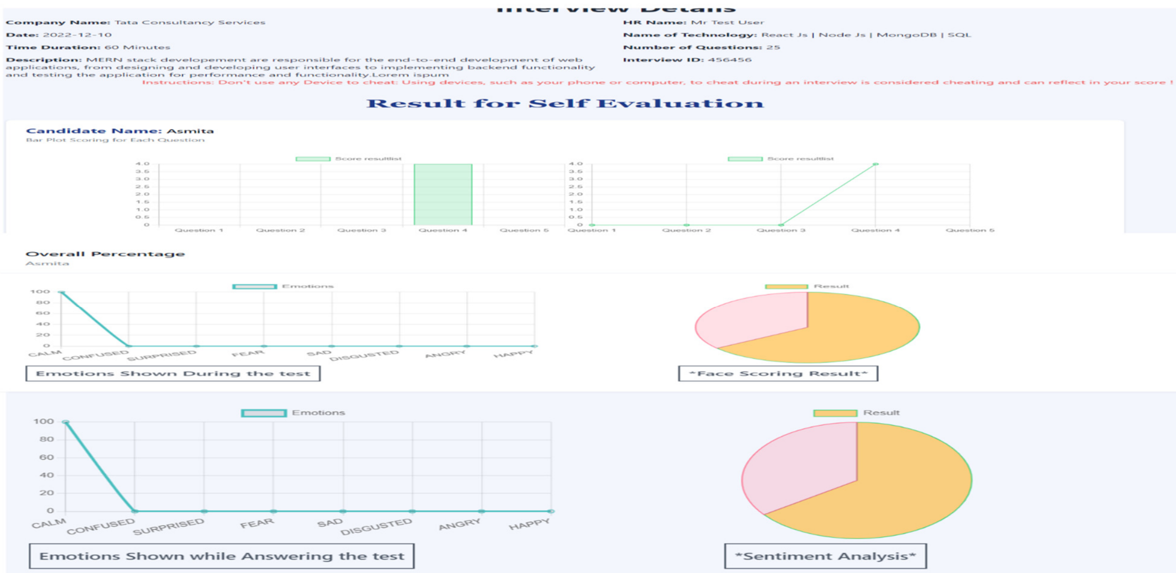


Fig. 6. Result for Sentiment Analysis

Classifier's performance is evaluated by comparing the classifier with the other previous classifiers based on the evaluation standards "Accuracy", "Precision", "Recall" [18].

- MOTP (Multiple Object Tracking Precision):

This metric measures how accurate our tracking is by measuring the average distance between where we think an object is and where it is actually. Lower values indicate higher precision in object tracking.

$$\text{MOTP} = \frac{\textit{Total distance between predicted and actual positions}}{\textit{Total number of tracked objects}}$$

- MOTA (Multiple Object Tracking Accuracy):

This metric measures the accuracy of object tracking, incorporating both false positives, false negatives, and mismatches. Higher values indicate better accuracy in object tracking.

$$\text{MOTA} = 1 - \frac{FN+FP+idsw}{GT}$$

Where:

FN is the number of false negatives (missed detections).

FP is the number of false positives (incorrect detections).

idsw is the number of identity switches (when a tracker assigns a new ID to an existing object).

GT is the total number of ground truth objects.

- ATA (Average tracking accuracy): This metric likely measures the percentage of time for which the system is average tracking objects within a minimum spectral range. Higher values indicate a higher percentage of time spent average tracking objects within the specified spectral range.

$$\text{ATA} = \frac{correct\ Associations}{Total\ number\ of\ associations}$$

Where:

The amount of accurate associations between projected object locations actual ground truth object locations is called Correct Associations.

Total number of associations is the total amount of associations between predicted actual ground truth object positions that the tracking system has created.

This formula determines the percentage of accurate connections among all associations formed by the tracking system. This gives an indication of the tracking system's average accuracy over all associated object pairings. Better tracking accuracy is indicated by a higher ATA number.

Table 1: -

	I (Existing System)	II (Proposed system)
MOTP	402 - 885	20 - 100
MOTA	25%	60%
ATA	25% with min Spectral Range	60% with min Spectral Range

Table 1 appears to compare two systems, I and II and shows how the proposed system is more accurate than the existing one based on several performance metrics as MOTP, MOTA, and ATA [18].

Result of the various tasks of smart hire as shown in Fig. 4, Fig. 5, Fig. 6 The Attributes under inspect value of every task of smart hire, it gives some result count for corresponding metrics which are enlisted in Table 1. Table 1 compare two different systems and shows how the proposed system is more accurate than the existing one

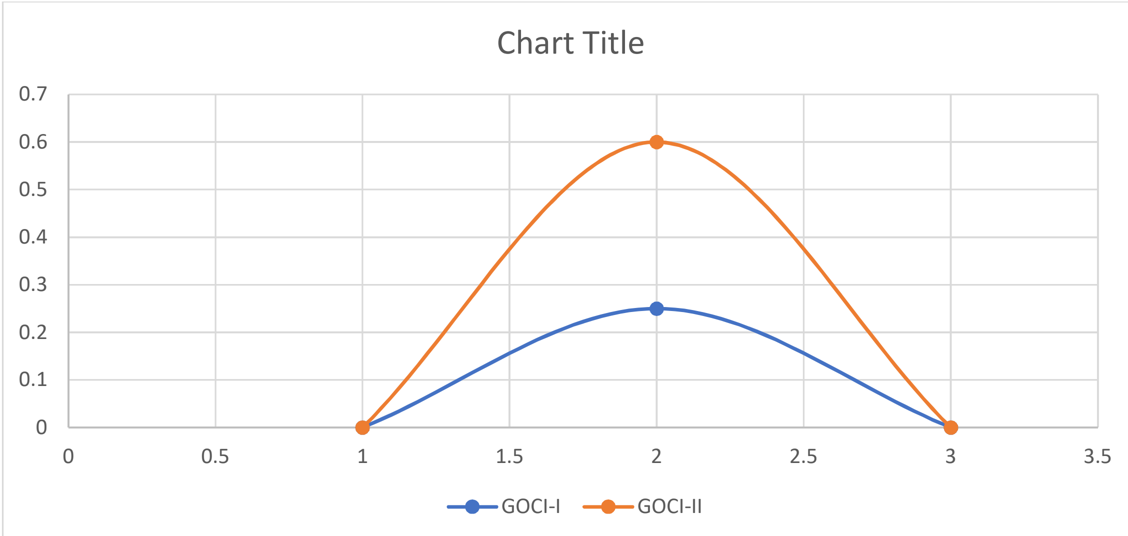


Fig. 7 X-Y Scatter Plot for Table 1

Fig. 7 depicts the increasing level of proposed algorithm. From Table 1 there is a difference between I (2,0.25) and II (2,0.60) i.e. 35. So we can say all the results of the study showed a more accuracy of 35% in finding the best-fit candidates using the proposed algorithms, which was considered a decent accuracy for algorithms running on an open dataset. However, the researchers acknowledged the need for testing the system on a larger scale with broader job profiles and thousands of candidates to obtain more accurate efficiency of the algorithm. K-means clustering algorithm with a collective hierarchical clustering algorithm for its ability to predict system behaviour with a high degree of accuracy [19].

5. CONCLUSION

The offered smart hire project makes a substantial contribution to the human resource business, namely the recruiting process. By automating time-consuming portions of the recruiting process, such as evaluating candidate qualifications and soft skills, our technology saves recruiters a significant amount of time and effort. The system includes a website and algorithms that do an AI-based analysis of each candidate's job application, making it easier for recruiters to identify the best-fit prospects. The Personality Insights

Algorithm, which offers BIG 5 personality values for soft skill-based questions, and the Answer Relevancy Algorithm, which evaluates applicant replies to topic knowledge-based questions, help recruiters better assess each candidate's ability. Our approach minimizes staffing needs and the length of the recruiting process while providing a more effective means of evaluating candidates' qualifications.

Smart Hire offers a unique method for optimizing the recruiting process, increasing efficiency, accuracy, and objectivity. This strategy has the potential to alter recruiting by making it more efficient and productive, benefiting both applicants and recruiters. More testing with a broader range of job profiles and a larger dataset is needed to improve the algorithm's accuracy and efficiency.

Smart Hire research project also recommends face recognition, which is used to recognize and match the candidate's photo on the résumé, as well as the person conducting the video interview. This prevents candidates from cheating on both the system and the recruiter.

6. FUTURE SCOPE

This research project initiative is a significant time and cost saver for everyone engaged in the job search process, it has a broad future to reach. The project may be made more durable and secure. Collaborations with a few firms may be formed to evaluate its real-world performance and promote it as a comprehensive tool for the hiring process through block chain technology [20]. Currently, the system includes algorithms for processing both hard and soft skill questions. However, we may eventually integrate the processing of logical and analytical questions, which will encompass a greater range of interview topics.

The block chain technology will primarily assist recruiters, but it may also be used to educate others. Block chain makes sure records are accurate, checks information, and puts things into action. AI helps figure things out, looks at data, and finds patterns to draw conclusions [20]. Students (before to entering the recruitment process) can do mock interviews on the system and assess themselves based on the score for their responses, and the applicant can then practice the points or abilities that he or she lacks and repeat the interview in order to become great at it. This will boost their confidence for the actual interviews. Colleges can utilize this instructional resource to train their learners. To understand these future plans, sentiment analysis charts and a full interview video can be downloaded at the end for more details.

7. REFERENCES

1. *Department of Census & Statistics, "QUARTERLY REPORT OF THE SRI LANKA LABOUR FORCE SURVEY," Ministry of Finance, Economy and Policy Development , 2019.*
2. *O. Ferret, B. Grau, M. Hurault-Plantet, G. Illouz, C. Jacquemin and L. Monceaux, "How NLP Can Improve Question Answering," KNOWLEDGE ORGANIZATION, vol. 29, no. 3-4, pp. 135-155, 2002.*
3. *P. Senarathne, M. Silva, A. Methmini, D. Kavinda and S. Thelijjagoda, "Automate Traditional Interviewing Process Using Natural Language Processing and Machine Learning," 2021 6th International Conference for Convergence in Technology (I2CT), Maharashtra, India, 2021, pp. 1-6, doi: 10.1109/I2CT51068.2021.9418115.*
4. *E. Malherbe and M. -A. Aufaure, "Bridge the terminology gap between recruiters and candidates: A multilingual skills base built from social media and linked data," 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), San Francisco, CA, USA, 2016, pp. 583-590, doi: 10.1109/ASONAM.2016.7752295.*
5. *K. Somsuk and N. Thammawongsa, "Applying d-RSA with login system to speed up decryption process in client side," 2017 IEEE 3rd International Conference on Engineering Technologies and Social*

Sciences (ICETSS), Bangkok, Thailand, 2017, pp. 1-5, doi: 10.1109/ICETSS.2017.8324150.

6. *Priyam Shah, Mihir Shah, Amit Bhujbal, Rohini Nair, Automation in the Interview Assessment Process , International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 08 Issue: 12 | Dec 2021*
7. *B. Bouaziz, T. Zlitni and W. Mahdi, "AViTExt: Automatic Video Text Extraction; A new Approach for video content indexing Application," 2008 3rd International Conference on Information and Communication Technologies: From Theory to Applications, Damascus, Syria, 2008, pp. 1-5, doi: 10.1109/ICTTA.2008.4530004.*
8. *X. Wang et al., "End-to-End Scene Text Recognition in Videos Based on Multi Frame Tracking," 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, Japan, 2017, pp. 1255-1260, doi: 10.1109/ICDAR.2017.207.*
9. *Siswanto, J., Suakanto, S., Andriani, M., Hardiyanti, M., Kusumasari, T.F., 2022. Interview Bot Development with Natural Language Processing and Machine Learning. International Journal of Technology. Volume 13(2), pp. 274-285*
10. *G. Li, Q. Zheng, L. Zhang, S. Guo and L. Niu, "Sentiment Infomation based Model For Chinese text Sentiment Analysis," 2020 IEEE 3rd International Conference on Automation, Electronics and Electrical Engineering*

(AUTEEE), Shenyang, China, 2020, pp. 366-371, doi: 10.1109/AUTEEE50969.2020.9315668.

11. *Asmita Vishalkumar Patil, "Automate Smart Hiring System Using Natural Language Processing and Sentiment Analysis",Volume 18 – Issue 1 – January 24,https://doi.org/10.7910/DVN/DVPJBS*
12. *Widiger, Thomas A. (ed.), The Oxford Handbook of the Five Factor Model, Oxford Library of Psychology (2017; online edn, Oxford Academic, 10 Sept. 2015),*
13. *Havrlant, Lukáš & Kreinovich, Vladik. (2017). A simple probabilistic explanation of term frequency-inverse document frequency (tf-idf) heuristic (and variations motivated by this explanation). International Journal of General Systems. 46. 27-36. 10.1080/03081079.2017.1291635.*
14. *B. Sigdel, G. Lin, Y. Miao and K. Ahmed, "Testing QA Systems' ability in Processing Synonym Commonsense Knowledge,"* 2020 24th International Conference Information Visualisation (IV)*, Melbourne, Australia, 2020, pp. 317-321, doi: 10.1109/IV51561.2020.00059.*
15. *H. Yang* et al.*, "Ocean Color Algorithm Development Environment for High-Speed Data Processing of GOCI-II,"* IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia, Spain, 2018, pp. 7968-7971, doi: 10.1109/IGARSS.2018.8517983.*
16. *C. Scott, D. Wynne and C. Boonthum-Denecke, "Examining the Privacy of Login Credentials Using Web-Based Single Sign-on - Are We Giving*

Up Security and Privacy for Convenience?," 2016 Cybersecurity Symposium (CYBERSEC), Coeur d'Alene, ID, USA, 2016, pp. 74-79, doi: 10.1109/CYBERSEC.2016.019.

17. *W. Badawy, "On Scalable Video Codec for 4K and High Definition Video Streaming – the Hierarchical Adaptive Structure Mesh Approach "HASM-4k"," 2020 IEEE 10th International Conference on Consumer Electronics (ICCE-Berlin), Berlin, Germany, 2020, pp. 1-5, doi: 10.1109/ICCE-Berlin50680.2020.9352175.*
18. *K. Fujihira and N. Horibe, "Multilingual Sentiment Analysis for Web Text Based on Word to Word Translation," 2020 9th International Congress on Advanced Applied Informatics (IIAI-AAI), Kitakyushu, Japan, 2020, pp. 74-79, doi: 10.1109/IIAI-AAI50415.2020.00025.*
19. *V. Venkatesh and A. S. Vindhya, "Predicting the Accuracy of Fractionation of Patron's Activities in Online Social Networks Using Novel K-Means Clustering Algorithm Comparing with Agglomerative Hierarchical Clustering Algorithm,"* 2023 6th International Conference on Contemporary Computing and Informatics (IC3I)*, Gautam Buddha Nagar, India, 2023, pp. 2563-2566, doi: 10.1109/IC3I59117.2023.10397982.*
20. *M. Rath, "A review of Artificial Intelligence Emerging technologies and challenges in Block Chain Technology,"* 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Tirunelveli, India, 2019, pp. 1031-1035, doi: 10.1109/ICSSIT46314.20*

19.8987807.