AI-Powered Resume Parser for Automated Candidate Screening Using Automated Classification Logistic Regression System

Rohini M[1], Snega T[2], Chinthamani V[3], Sowmiya Shree. N.L[4], Dharani K[5]
[1]*Assistant Professor, Department of computer science and Engineering, The Kavery Engineering college (Autonomous), M. Kalipatti.*
[2,3,4,5]*UG Students, Department of computer science and Engineering, The Kavery Engineering college (Autonomous), M. Kalipatti.*

***Abstract*—Recently, the limited job opportunities and the primary reasons for updating resumes have become major contributors to the issue of unemployment. Early updates and continuous monitoring of resume automation using Machine Learning (ML) can be costly. To address the challenges associated with the Automated Classification Logistic Regression System Classification Method, early updates and automation of job applications improve the accuracy of results, and the collection of documents becomes more secure. Moreover, the Text Preprocessing Module eliminates duplicate data, minimizes unknown data, and maximizes valuable data during preprocessing. The Term Frequency-Inverse Document Frequency (TF-IDF) separates each data value type, starting with the minimum representation to reduce the unpredictable values and enhance resume update performance. Ultimately, the proposed method classifies data, tests it, and validates the predicted outcomes to assess performance. Calculation is a training process, while the testing data is multi-level in classification. Each type of data connects to its network within the classification framework. It facilitates comprehensive monitoring based on input data from video tests of the training data and checks the various recruitment models involved in the process. Finally, it is automatically updated to assist in the job application process. The proposed method offers greater reliability and achieves high performance while maintaining standard scalability. These techniques reduce time complexity, ensuring performance remains within an accurate range of 91%.**

***Index Terms*—Automated Classification Logistic Regression System, Term Frequency-Inverse Document Frequency (TF-IDF), Text Preprocessing Module, Automatic Classification.**

I. INTRODUCTION

The KNN enhances the identification of suitable candidates. Its societal impact includes streamlining hiring processes, improving efficiency and accessibility, and maintaining stability [1]. The strengths of the DL techniques are calculated from the resumes and ranked from maximum to minimum match. Text classification involves assigning data and labels of potential in the resume for use, and it automatically updates the resume used for Natural Language Processing (NLP). Boosting is a method used in machine learning to reduce errors in predictive data analysis on labeled data, allowing for guesses about unlabeled data. It's one of the ML techniques used to assess performance [2]. Bidirectional Encoder Representations from Transformers (S-BERT) can revolutionize processes. For HR professionals facing overwhelming numbers of resumes, the manual screening process is time-consuming and prone to errors.

The resume update is a job application for use in deep learning, analyzing real-time data, and verifying the process. To determine the skills, communication of the process, and analysis of the data utilized for the Random Forest (RF), i.e., reliability of the process, maintaining stability of the process, to resume updating the performance [3] generally. The strengths of the DL techniques are calculated from the resumes and ranked from maximum to minimum match. Text classification involves assigning data and labels of potential in the resume for use, and it automatically updates the resume being used.

The process lacks transparency in assessing performance, has high computational costs, and sometimes misclassifies the data. It is data-dependent and poses risks of overfitting and a lack of interpretability. Additionally, it is a more expensive tool that exhibits greater interoperability in performance; however, it requires lower efficiency in performance and consumes more power [4]. The process is limited in its contextual understanding of the techniques, which are high-cost and have a low accuracy range. It is affected by unwanted data regarding the performance of the techniques and is impersonal due to screening and the absence of a human touch. Moreover, it incurs high computational costs and sometimes misclassifies the data, while being a more expensive tool that demonstrates greater interoperability in performance [5]. Overall, the process lacks transparency in assessing performance, is data-dependent, and poses risks of overfitting and a lack of interpretability.

The main contribution ensures that each data type is normalized for performance and addresses the data variables of the techniques. All kinds of data are separated for testing and training, updating the data variables in the skills. It checks the different types of recruitment models involved in the process. Finally, it is automatically updated to assist in the job application process.

II. LITERATURE SURVEY

According to the job application, the resume is updated using machine learning (ML). It aims for a high-salary position, but the developed skills are applied to the convolutional neural network (CNN) to enhance performance [6]. The system is responsible for matching the resumes to the corresponding job descriptions; however, the process lacks transparency in assessing performance.

The text classification assigns the data and labels of the potential in the resume to be used, and it automatically updates the resume used for the Natural Language Processing (NLP). However, the process has low accuracy in performance [7]. The proposed method is deep learning (DL), which allows real-time data analysis of the resume for the skin and updates the process. However, the process has a high computational cost, and sometimes the data is misclassified.

The advanced deep learning techniques enhance resume ranking in web development, improving the recruitment process [8]. Using the embedding model, the text extraction hybrid matching approach offers highly accurate and relevant resume rankings. However, the process is data-dependent and poses risks of overfitting and a lack of interpretability.

The hybrid model compares the resume model with verification and testing of the skills used for Artificial Intelligence [9]. It focuses on achieving a more accurate level of various performance metrics and reducing the misclassification of data to predict the unknown performance data. However, the process is a more expensive tool that is utilized and exhibits greater interoperability in performance.

Boosting is a method used in machine learning to reduce errors in predictive data analysis on labeled data to make guesses about unlabeled data. It's one of the ML techniques used to assess performance, but the process of testing and verifying resume performance is slow [10]. The proposed method is a DL compared to all kinds of resumes, and proper performance updating is necessary. However, the processes require lower performance efficiency and more power consumption of the techniques.

The Resume Parsing and Shortlisting System is designed to transform recruitment through an innovative machine focused on this task [11]. It offers a refined resume parsing tool that identifies the most qualified candidates according to the recruiter's criteria. Natural Language Processing (NLP) techniques are employed, and strengths are calculated from the resumes and ranked from maximum to minimum match. However, the process is limited in its context understanding of the techniques.

Bidirectional Encoder Representations from Transformers (S-BERT) can revolutionize processes. For HR professionals facing an overwhelming number of resumes, the manual screening process is time-consuming and prone to errors [12]. The proposed method involves an NLP calculation and ranking based on scores. In our evaluation of a dataset, it demonstrates the ability to identify relevant resumes effectively. However, the process is high-cost, and the techniques have a low accuracy range.

Identifying module titles and grades involves using supervised ML models, with the Support Vector Machine (SVM) model demonstrating the highest accuracy, albeit with greater time complexity [14].

The proposed method enhances the identification of suitable candidates. Its societal impact includes streamlining hiring processes, improving efficiency and accessibility, and fostering a more productive and equitable job market. However, the process is affected by the unwanted data regarding the performance of the techniques.

They identify the most suitable candidates for their Job Description (JD), while ensuring that no talent acquisition is overlooked due to human mistakes or low performance reliability [15]. The proposed methods of Applicant Tracking Systems (ATS) have taken over the human process of resume screening, enabling the assessment of thousands of resumes in seconds. However, the process is impersonal due to screening and the lack of the human touch of the techniques.

The decision support system (DSS) assigns the data and labels of the potential in the resume to be used and automatically updates the utilized resume, and the data processing. However, this process has low accuracy in performance [16]. The proposed method, NLP, allows real-time data analysis of the resume for skin and updates the process. Nonetheless, the process incurs a high computational cost, and sometimes there is a misclassification of the data.

The hybrid model compares the resume model with verification and testing of the skills used for Artificial Intelligence [17]. DL focuses on achieving a more accurate level of various performance metrics and reducing misclassification data to predict unknown performance data. However, this process is a more expensive tool that exhibits greater interoperability in performance.

The multi-keyword scoring techniques enhance resume ranking in web development, improving recruitment [18]. Using the embedding model, the K-Nearest Neighbor (KNN) hybrid matching approach offers highly accurate and relevant resume rankings. However, the process is data-dependent and poses risks of overfitting and a lack of interpretability.

The resume used for ML is for a high-salary position, but the developed skills are applied to the Recurrent Neural Network (RNN) to enhance performance [19]. The system is responsible for matching the resumes to the corresponding job descriptions; however, the process lacks transparency in assessing performance.

The KNN enhances the identification of suitable candidates. Its societal impact includes streamlining hiring processes, improving efficiency and accessibility, and maintaining stability [20]. The strengths of the DL techniques are calculated from the resumes and ranked from maximum to minimum match. However, the process is limited in its contextual understanding of the techniques.

III. PROPOSED METHODOLOGY

The section predicts the value and collects data regarding the performance verification and testing of data variables. It predicts the value of the data in the skills testing process and training for the resume update of performance. The process automatically updates; it measures the performance's recall, precision, and reliability. It calculates classification, accurately reflecting the process to identify the appropriate skill based on recent performance.

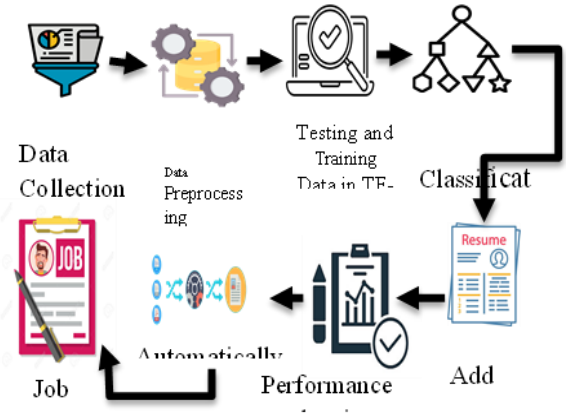


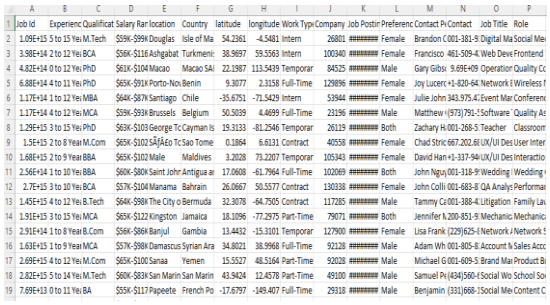
Fig. 1 Resume Updating using Automated

Classification Logistic Regression System

Figure 1 measures recall, precision, and reliability of performance. It calculates classification, accurately reflecting the process of identifying the appropriate skill based on recent performance.

A. Dataset Description

The section of the dataset measures the range of skills, quantitative performance, communication skills, extra knowledge, building in the fraction of information, and general aspects of the performance of the process to update the job application in a resume model evaluation of the performance.



Job Id	Experienc	Qualificat	Salary Ran	location	Country	latitude	longitude	Work Typ	Company	Job Postin	Preferenc	Contact Pe	Contact	Job Title	Role
1.09E+15	5 to 15 Yea	M.Tech	$59K-$99K	Douglas	Isle of Ma	54.2361	-4.5481	Intern	26801	########	Female	Brandon C	001-381-9	Digital Ma	Social Me
3.98E+14	2 to 12 Yea	BCA	$56K-$116	Ashgabat	Turkmenis	38.9697	59.5563	Intern	100340	########	Female	Francisco	461-509-4	Web Deve	Frontend
4.82E+14	0 to 12 Yea	PhD	$61K-$104	Macao	Macao SAI	22.1987	113.5439	Temporar	84525	########	Male	Gary Gibs	9.69E+09	Operation	Quality Co
6.88E+14	4 to 11 Yea	PhD	$65K-$91K	Porto-Nov	Benin	9.3077	2.3158	Full-Time	129896	########	Female	Joy Lucero	+1-820-64	Network E	Wireless I
1.17E+14	1 to 12 Yea	MBA	$64K-$87K	Santiago	Chile	-35.6751	-71.5429	Intern	53944	########	Female	Julie John	343.975.4	Event Mar	Conferenc
1.17E+14	4 to 12 Yea	MCA	$59K-$93K	Brussels	Belgium	50.5039	4.4699	Full-Time	23196	########	Male	Matthew H	(573)791-1	Software '	Quality As
1.29E+15	3 to 15 Yea	PhD	$63K-$103	George Tc	Cayman Is	19.3133	-81.2546	Temporar	26119	########	Both	Zachary H	001-268-5	Teacher	Classroom
1.5E+15	2 to 8 Year	M.Com	$65K-$102	SÃ£o Tc	Sao Tome	0.1864	6.6131	Contract	40558	########	Female	Chad Stric	667.202.6	UX/UI Des	User Inter
1.68E+15	2 to 9 Year	BBA	$65K-$102	Male	Maldives	3.2028	73.2207	Temporar	105343	########	Female	David Han	+1-337-94	UX/UI Des	Interactio
2.56E+14	1 to 10 Yea	BBA	$60K-$80K	Saint John	Antigua ar	17.0608	-61.7964	Full-Time	102069	########	Both	John Nguy	001-318-9	Wedding I	Wedding
2.7E+15	3 to 10 Yea	BCA	$57K-$104	Manama	Bahrain	26.0667	50.5577	Contract	130338	########	Female	John Colli	001-683-8	QA Analys	Performar
1.45E+15	4 to 12 Yea	B.Tech	$64K-$98K	The City o	Bermuda	32.3078	-64.7505	Contract	117285	########	Male	Tammy Ca	001-388-4	Litigation	Family Lav
1.91E+15	3 to 15 Yea	MCA	$65K-$122	Kingston	Jamaica	18.1096	-77.2975	Part-Time	79071	########	Both	Jennifer N	200-851-9	Mechanic	Mechanic
2.91E+14	1 to 8 Year	B.Com	$56K-$86K	Banjul	Gambia	13.4432	-15.3101	Temporar	127900	########	Female	Lisa Frank	(229)625-4	Network A	Network S
1.63E+15	1 to 9 Year	MCA	$57K-$98K	Damascus	Syrian Ara	34.8021	38.9968	Full-Time	92128	########	Male	Adam Wh	001-805-8	Account M	Sales Accc
2.69E+15	4 to 12 Yea	M.Com	$65K-$100	Sanaa	Yemen	15.5527	48.5164	Part-Time	92028	########	Male	Michael G	001-609-5	Brand Mar	Product Bi
2.82E+15	5 to 14 Yea	M.Tech	$60K-$83K	San Marin	San Marin	43.9424	12.4578	Part-Time	49100	########	Male	Samuel Pe	(434)560-4	Social Wo	School Soc
7.69E+13	0 to 11 Yea	BA	$55K-$117	Papeete	French Po	-17.6797	-149.407	Full-Time	29318	########	Male	Benjamin	(331)668-1	Social Me	Content C

Fig. 2 Job Application Based on the Dataset

Figure 2 measures the data in various skills and resume models of performance. Its automatic updating dataset is composed of the performed.

B. Text Preprocessing Module

The section is a module evaluation resulting from the job applications' collection of more data. It aims to check corrections to a verification process, minimize unwanted data, and maximize proper data quality checks. Ultimately, it ensures that each data type is normalized for performance and addresses the data variables of the techniques.

Equation 1 is a collection of data in the job application that identifies unknown data and unwanted skills of performance and predicts value in the data.

$$s = (s_1, s_2, s_3, \ldots . s_n) \quad (1)$$

Equation 2 predicts the value and collects the data regarding the performance verification and testing of the data variables.

$$s(x|y_k) = \frac{(\sum_i x_i)}{x_i} s_{k_i}^{x_i} \quad (2)$$

Equation 3 is for all types of data, normalized for performance and measures of the respective data variable, validating the data's veracity.

$$S = -\sum_{i=1}^{z} \frac{|s_j|}{|s_i|} \quad (3)$$

C. Term frequency – inverse document frequency (TF-IDF)

The section is a numerical statistic that measures the importance of a word in a document within a collection of documents. It's commonly used in information retrieval, and all kinds of data are separated for testing and training, updating the data variables in the skills.

Equation 4 separates each kind of data value, first detecting the maximum representation in the performance to reduce unknown data.

$$x_{ij} = pq_{ij} sq_{ij} = pq_{ij} \times log_2 \left(\frac{k}{sq_i}\right) \quad (4)$$

Equation 5 separates each kind of data value, first training the minimum representation in the performance to reduce the unpredictable values of the data.

$$x_{ij} = pq_{ij} sq_{ij} = \frac{q_{ij}}{\sqrt{\sum_{k=0}^{k} (pq_{ij} sq_{ij})}} \times log_2 \left(\frac{k}{sq_i}\right) \quad (5)$$

Equation 6 predicts the value of the resume data in the skills testing process and resume training, updating performance.

$$x(p,q) = \sqrt{\sum_{k=0}^{k} (s_{rm} - s_{rn})^2} \quad (6)$$

D. Automated Classification Logistic Regression System

The section is an automatically updated record of the various skills in the recruitment process and compares another resume model, the implementation of the resume portfolio, and checks the different types of recruitment models involved in the process. Finally, it is automatically updated to assist in the job application process.

Equation 7 tests the validity of the data and is classified as part of the process to predict misclassification of the performance.

$$\sum_{j=1}^{i} \frac{\frac{pq_j + ps_j}{pq_i + zn_i + zl_i + ps_i}}{I} \quad (7)$$

Equation 8 calculates the classification, accurately reflecting the process to identify the appropriate skill based on recent performance.

$$\sum_{j=1}^{i} \frac{\frac{pq_j + ps_j}{pq_i + zn_i + zl_i + ps_i}}{J} \quad (8)$$

Equation 9 is an automatic update of the process; it measures the performance's recall, precision, and reliability.

$$\frac{\frac{\sum_{j=0}^{n} pq_i}{\sum_{j=0}^{n} pq_j + ps_j}}{J} \quad (9)$$

IV. RESULTS AND DISCUSSION

This section evaluates the precision, recall, accuracy, FN, and time complexity scores across various parameters and approaches. Furthermore, the proposed method can detect data updating in job applications using Technology and data points in the attack dataset.

Table 1. Simulation Parameter

Simulation Process	Parameter Name
Dataset Name	Job Dataset
No of Dataset	1048576
Training Dataset	100000
Testing Dataset	48576
Language	Python

As illustrated in Table 1, the simulation parameters were evaluated using 1048576 data points collected in the feature selection process. 100000 is a training dataset, and 48576 is a testing dataset.

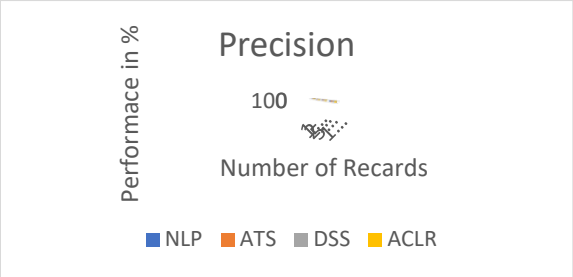


Fig. 3 Analysis of Precision

Figure 3 illustrates using precision analysis for secure health data exchange through Artificial Intelligence technology. This review assesses previous methods, including NLP, ATS, and DSS contrasts them with the proposed ACLR method. The precision levels of the performance ratings for these methods are 85.6, 89.9, and 99.2, respectively, for the various performance levels in data protection.

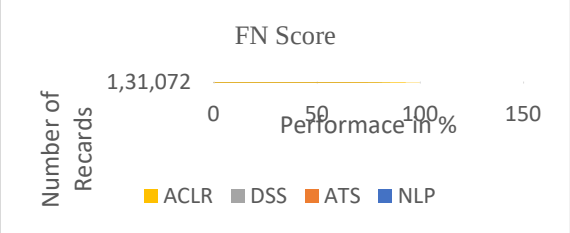


Fig. 4 Analysis of FN Score

Figure 4 illustrates using FN Score analysis for secure health data exchange through Artificial Intelligence technology. This review assesses previous methods, including NLP, ATS, and DSS contrasts them with the proposed ACLR method. The precision levels of the performance ratings for these methods are 85.6, 89.9, and 99.2, respectively, for the various performance levels in data protection.

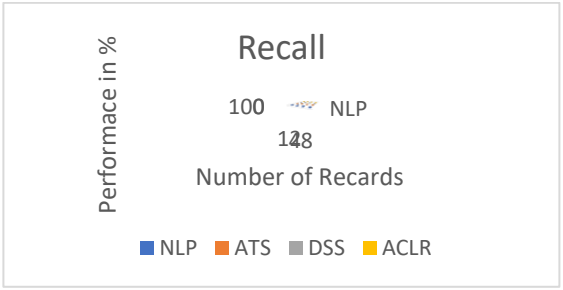


Fig. 5 Analysis of Recall

Figure 5 illustrates using recall analysis for secure health data exchange through Artificial Intelligence technology. This review assesses previous methods, including NLP, ATS, and DSS contrasts them with the proposed ACLR method. The precision levels of the performance ratings for these methods are 85.6, 89.9, and 99.2, respectively, for the various performance levels in data protection.

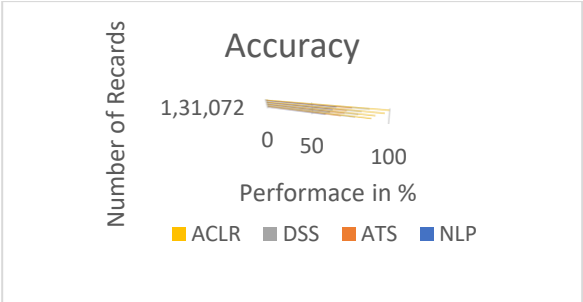


Fig. 6 Analysis of Accuracy

Figure 6 illustrates using accuracy analysis for secure health data exchange through Artificial Intelligence technology. This review assesses previous methods, including NLP, ATS, and DSS contrasts them with the proposed ACLR method. The precision levels of the performance ratings for these methods are 85.6, 89.9, and 99.2, respectively, for the various performance levels in data protection.

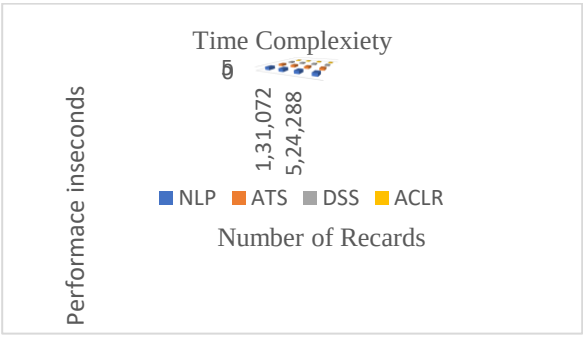


Fig. 7 Analysis of Time Complexity

Figure 7 illustrates using time complexity analysis for secure health data exchange through Artificial Intelligence technology. This review assesses previous methods, including NLP, ATS, and DSS contrasts them with the proposed ACLR method. The precision levels of the performance ratings for these methods are 5.56, 3.45, and 1.50, respectively, for the various performance levels in data protection.

V. CONCLUSION

This study explores the Automated Classification Logistic Regression of Technology in enhancing detection, prevention, and response in automatically updated systems to assist in the job application. The proposed method offers greater reliability and achieves high performance while maintaining standard scalability. Performance measurement is improved using the F1 score, time complexity, recall, precision, and accuracy based on commonly employed real-time data for result comparison. The capacity to identify unknown data is bolstered by compressing vast amounts of information into manageable archives and transforming the data. This process involves further classification and reduction of the data. It is trained and tested on labeled data to ascertain the relationship between features and classes, with the class having the highest calculated probability as the prediction. Each type of data connects to its network within the classification framework. It facilitates comprehensive monitoring based on input data from video tests of the training data and checks the various recruitment models involved in the process. The process is managed efficiently, completed within a defined time frame, and performed with multitasking efficiency, ensuring reliability. The system has achieved a 91% accuracy rate in automated job application updates.

REFERENCE

[1] Sarveshwaran, R., et al. "NLP-Based AI-Driven Resume Screening Solution for Efficient Candidate Selection." International Congress on Information and Communication Technology. Singapore: Springer Nature Singapore, 2024.

[2] Sarveshwaran, R., et al. "NLP-Based AI-Driven Resume Screening." Proceedings of Ninth International Congress on Information and Communication Technology: ICICT 2024, London, Volume 2. Vol. 2. Springer Nature, 2024.

[3] Gudipati, Praneeth Kumar. AI-Based Resume Parser and Job Matcher. Diss. CALIFORNIA STATE UNIVERSITY, NORTHRIDGE, 2025.

[4] Priyanka, J. Himabindu, and Nikhat Parveen. "DeepSkillNER: an automatic screening and ranking of resumes using hybrid deep learning and enhanced spectral clustering approach." Multimedia Tools and Applications 83.16 (2024): 47503-47530.

[5] Deshmukh, Asmita, and Anjali Raut. "Applying bert-based nlp for automated resume screening and candidate ranking." Annals of Data Science (2024): 1-13.

[6] Najjar, Arwa, Belal Amro, and Mário Macedo. "An intelligent decision support system for recruitment: resumes screening and applicants ranking." Informatica 45.4 (2021).

[7] Kothari, Param, et al. "InterviewEase: AI-powered interview assistance." (2024).

[8] Yazıcı, Mehmet Batuhan, Damla Sabaz, and Wisam Elmasry. "AI-based Multimodal Resume Ranking Web Application for Large Scale Job Recruitment." 2024 8th International Artificial Intelligence and Data Processing Symposium (IDAP). IEEE, 2024.

[9] Ujlayan, Anshul, Sanjay Bhattacharya, and Sonakshi. "A Machine Learning-Based AI Framework to Optimize the Recruitment Screening Process." International Journal of Global Business and Competitiveness 18.Suppl 1 (2023): 38-53.

[10] Bolla, Soham, et al. "AIRecruitHUB: An AI-Based Recruitment Web Framework for Bridging Talent and Opportunity." Doctoral Symposium on Computational Intelligence. Singapore: Springer Nature Singapore, 2024.

[11] Vijaykumar, Nandini, et al. "A NLP based model on resume parsing and shortlisting system for smart recruitment." AIP Conference Proceedings. Vol. 3279. No. 1. AIP Publishing LLC, 2025.

[12] Deshmukh, Asmita, and Anjali Raut. "Enhanced Resume Screening for Smart Hiring Using Sentence-Bidirectional Encoder Representations from Transformers (S-BERT)." International Journal of Advanced Computer Science & Applications 15.8 (2024).

[13] Abeywardhana, Lakmini, et al. "Optimizing Candidate Selection in it Industry: An Approach for Cv-Job Description Matching and Academic Transcript Analysis." Available at SSRN 4676402.

[14] Gulati, Vansh, et al. "Resume Analyzer Using Natural Language Processing (NLP)." (2024).

[15] Manish, Vazzula, et al. "Optimizing Resume Parsing Processes by Leveraging Large Language

Models." 2024 IEEE Region 10 Symposium (TENSYMP). IEEE, 2024.

[16] Amro, Belal, Arwa Najjar, and Mario Macido. "An Intelligent Decision Support System For Recruitment: Resumes Screening and Applicants Ranking." (2022).

[17] Shaikh, Mohaseen, Dnyaneshwar Ghanawajeer, and Mahendra Sawane. "Design and Development of an Intelligent Decision Support System for HR Recruitment." Asian Journal of Computer Science and Technology 12.2 (2023): 24-30.

[18] Nguyen, Hien, et al. "Intelligent search system for resume and labor law." PeerJ Computer Science 10 (2024): e1786.

[19] Sreekari, A. Lakshmi, and T. Devi. "Accurate resume shortlisting using multi keyword scoring technique in comparison with K-nearest neighbor algorithm." Applications of Mathematics in Science and Technology. CRC Press 516-521.

[20] Satheesh, K., et al. "Resume ranking based on job description using SpaCy NER model." International Research Journal of Engineering and Technology 7.05 (2020): 74-77.