Resume Ranker: AI-Based Skill Analysis And Skill Matching System

Nikethani Gangoda
Department of Information Technology
Sri Lanka Institute of Information Technology
Sri Lanka
it20916626@my.sliit.lk

Kavindu Piumal Yasantha
Department of Information Technology
Sri Lanka Institute of Information Technology
Sri Lanka
it20637200@my.sliit.lk

Chamina Sewwandi
Department of Information Technology
Sri Lanka Institute of Information Technology
Sri Lanka
it20298494@my.sliit.lk

Navindu Induvara
Department of Information Technology
Sri Lanka Institute of Information Technology
Sri Lanka
it20600716@my.sliit.lk

Samantha Thelijjagoda
Department of Computer Systems Engineering
Sri Lanka Institute of Information Technology
Sri Lanka
samantha.t@sliit.lk

Nishantha Giguruwa
College of Sustainability and Tourism
Ritsumeikan Asia Pacific University
Japan
gamagelk@apu.ac.jp

2024 Sixth International Conference on Intelligent Computing in Data Sciences (ICDS) | 979-8-3503-5120-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICDS62089.2024.10756304

***Abstract*— In response to Japan's imminent labor shortage crisis and its sluggish integration of IT into industry, an innovative AI-Based Skill Analysis and Matching System is proposed. With projections indicating a demand for 7 million foreign workers by 2040, there arises an urgent need for streamlined HR training processes. The system leverages advanced technologies to accelerate candidate selection and job matching, encompassing resume parsing, keyword filtering, skill matching, and improvement. Methodologically, various libraries, modules, and NLP models are integrated for comprehensive data extraction and analysis. Techniques such as tokenization, lemmatization, and cosine similarity-based matching optimize the alignment of candidates with job requirements. Findings demonstrate the system's effectiveness in reducing recruitment time and effort, enabling recruiters to focus on assessing relevant candidates efficiently. Additionally, the system offers tailored skill enhancement recommendations to candidates, facilitating their integration into the Japanese workforce. The significance of this research lies in its contribution to mitigating Japan's labor shortage crisis and advancing HR management practices. By harnessing AI technologies, the system not only addresses immediate labor market needs but also sets the stage for a technology-driven approach to talent acquisition. Overall, this research fosters advancements in HR management, benefiting employers and job seekers within and beyond Japan.**

Keywords— AI-Based Skill Analysis, Skill Matching System, Candidate selection, HR management, Labor shortage crisis, Job matching

I. INTRODUCTION

In the wake of rapid technological advancements and demographic shifts, the global workforce landscape is undergoing profound transformations. Nowhere is this more evident than in Japan, where the intersection of aging demographics and limited technological integration poses significant challenges to its industrial sector. Forecasts indicate that by 2040, Japan will face a daunting shortfall of 7 million skilled workers, necessitating a paradigm shift in human resource management strategies. To confront this imminent crisis, innovative solutions are imperative, and the integration of artificial intelligence (AI) stands as a promising avenue.

Among East Asian countries, Japan trails in leveraging information technology (IT) for industrial advancement, exacerbating the labor scarcity issue. The looming shortage underscores the urgent need for large-scale human resource training initiatives to prepare both domestic and international candidates for employment opportunities within Japan. Addressing this multifaceted challenge requires a comprehensive approach that combines AI technologies with traditional HR practices [1].

This paper introduces an AI-based Skill Analysis and Skill Matching System tailored to mitigate Japan's labor shortage. The system aims to expedite the identification and selection of optimal candidates by seamlessly aligning their skills with industry-specific requirements [2]. By streamlining the recruitment process, it facilitates efficient talent acquisition for employers while empowering job seekers to find suitable employment opportunities in a timely manner.

This research paper meticulously elucidates the conceptual framework, methodological intricacies, and technological infrastructure that form the bedrock of the proposed AI-driven solution. Central to the system are four pivotal components: Reading Relevant Information from Resumes, Filtering Keywords from Resumes, Matching Keywords with Job Requirements, and Improving Candidate Skills Relevant to Job Requirements. These components synergistically harness a sophisticated amalgamation of cutting-edge libraries, modules, NLP models, and data processing techniques to ensure the attainment of peak performance. From the initial extraction of pertinent data from resumes to the refinement of candidate skills, a cohesive technological stack underlies the seamless operation and efficacy of the proposed solution.

This study introduces an innovative AI-driven methodology aimed at alleviating Japan's acute labor deficit through the modernization of recruitment and skills development procedures. Through the utilization of AI technologies, we aspire to instigate a fundamental transformation in Japan's labor landscape, thereby fortifying its industrial standing amidst demographic shifts and technological progress.

II. BACKGROUND

The background and literature review section of this research paper provides a comprehensive exploration of the contextual landscape and existing scholarly works relevant to

979-8-3503-5120-0/24/$31.00 ©2024 IEEE

the proposed AI-based Skill Analysis and Skill Matching System. By delving into the historical context, current challenges, and pertinent theoretical frameworks, this section aims to establish a robust foundation for understanding the rationale behind the development of the proposed solution.

A. Optimizing Job Application Strategies Through Keyword Analysis

Spoorthi M et al. [3] propose an ensemble deep learning approach to automate resume classification, aiming to alleviate the burdensome task of manual screening faced by recruiters amidst the exponential rise in IT job applications. Their study reviews existing literature, highlighting various automated resume screening techniques such as machine learning algorithms and NLP methods. Leveraging Kaggle datasets, they preprocess data by cleaning, tokenizing, and encoding labels. The proposed model architecture, comprising a 1D CNN and Bi-directional GRU, effectively categorizes resumes based on skills. Experimental results showcase the model's performance metrics, including precision, recall, F1-score, and accuracy, alongside a confusion matrix demonstration. The authors conclude that their model streamlines the hiring process by swiftly sorting resumes, while suggesting future research directions. Moreover, they present a user-friendly web application utilizing the trained model for practical use. This study contributes to enhancing talent acquisition processes, offering recruiters a more efficient and informed approach to candidate screening.

Sunil Kumar Kopparapu et al. [4] presents a method for automating the extraction of pertinent information from resumes to streamline the search and management processes in their paper titled "Automatic Extraction of Useful Information from Unstructured Resumes to Aid Search". Focused on overcoming the challenge of handling resumes in various formats received by large corporations and headhunters daily, the paper delves into leveraging natural language processing (NLP) techniques to extract valuable data from unstructured resume formats. Unlike traditional approaches that rely on rigid templates, the proposed system aims to automate the extraction process across diverse resume layouts. Comprising modules such as input, information extraction, database construction, and search functionalities, the system targets critical resume sections including personal details, education, skills, and professional experience. Employing pattern recognition and keyword analysis, the extraction process encompasses identifying qualifications, software proficiencies, total work experience, and even personal identifiers like the candidate's name, birth date, and email address. This innovative approach not only enhances the efficiency of resume processing but also eliminates the need for standardized templates, thereby offering a more flexible and adaptive solution to resume management challenges.

Tumula Mani Harsha et al. [5] introduce a groundbreaking system in their research paper titled "Automated Resume Screener using Natural Language Processing (NLP)", which aims to revolutionize the resume screening process through the utilization of Natural Language Processing (NLP) techniques. Addressing the challenges inherent in manual resume screening, the paper discusses existing solutions and proposes the innovative "Resume Screener." This Python-based web application, following the Model-View-Controller (MVC) design pattern, comprises distinct layers dedicated to various aspects of resume screening. Noteworthy features include its support for diverse resume formats (PDF, DOC, DOCX), conversion to a standardized PDF format for text extraction using NLP methods, and the generation of visual representations of candidates' skills and qualifications. This facilitates efficient candidate evaluation for recruiters. Key advantages of this solution include its ability to handle large resume volumes, compatibility with different formats, effective NLP utilization, and a user-friendly interface. Additionally, its adaptable nature allows for easy customization of the skills database to suit specific job requirements. This pioneering approach holds significant promise in enhancing the efficiency and efficacy of resume screening, offering recruiters a powerful tool for talent acquisition endeavors.

In the paper titled "An Approach of Intelligent Automated Resume Analysis and Recommendations," Ramesh Chandra Tripathi et al. [6] address the challenges associated with swiftly identifying qualified job candidates through automated methods. Proposing an AI-based system to streamline the resume screening and recommendation process, the authors highlight the inefficiencies of manual resume evaluation and underscore the necessity for automated solutions to enhance candidate selection efficiency. The proposed method involves preprocessing resumes, feature extraction using TF-IDF, and the implementation of classification and ranking models. Through experimentation with various classifiers, the Linear Support Vector Machine emerges as the most accurate classifier. The recommendation model suggests resumes that closely match job descriptions based on their similarity. Emphasizing the significance of dealing with unstructured resume data, the authors advocate for the integration of AI techniques like Natural Language Processing to refine the selection process. In conclusion, the authors discuss the benefits of their proposed method and offer insights into potential avenues for further research and improvement in resume analysis and recommendation systems.

In the research paper titled "Job Applications Selection and Identification: Study of Resumes with Natural Language Processing and Machine Learning," A. Pimpalkar et al. [7] discuss the challenges organizations face in sorting through large volumes of job applications. They propose leveraging Natural Language Processing (NLP) and Machine Learning (ML) techniques to simplify resume screening. By extracting key information from unstructured resumes, the authors aim to match candidates' skills with job requirements more effectively. They compare different ML methods for resume processing, addressing issues such as writing style and grammar analysis. Ultimately, their goal is to improve the hiring process by using NLP and ML to accurately analyze and sort resumes, offering insights into the potential and challenges of advanced technologies in resume evaluation.

S. Bharadwaj et al. [8] present a groundbreaking method in their paper, "Resume Screening Using NLP and LSTM," which efficiently categorizes resumes based on listed skills to streamline the job candidate qualification process. By leveraging Natural Language Processing (NLP) and Long Short-Term Memory (LSTM) neural networks, the system accurately assesses resume suitability for specific job roles, offering benefits such as aiding job seekers in understanding their strengths and matching them with suitable job opportunities. Comparative analysis with existing methods

underscores the effectiveness of this approach, while future improvements, like integrating data from platforms such as LinkedIn and GitHub, promise further advancements in automated resume screening and job-skill matching techniques.

In the paper "Resume Classification Using ML Techniques," B. Surendrian et al. [9] tackle the challenge of manual resume screening by proposing an automated solution leveraging machine learning (ML) algorithms such as Decision Tree, Random Forest, K-Nearest Neighbors (KNN), and Support Vector Classifier (SVC). With a dataset of 3446 resumes across 48 job categories, the authors utilize natural language processing (NLP) methods for information extraction, employing techniques like noise removal and lemmatization. Evaluation metrics reveal Random Forest as the most accurate model, achieving a success rate of 91.38%, followed closely by SVC at 90.62%. The study emphasizes the efficiency and effectiveness of their proposed ML-based approach in streamlining resume classification, potentially revolutionizing the hiring process and benefiting both employers and job seekers.

Rasika Ransing et al. [10] introduce a novel approach in their research paper titled "Screening and Ranking Resumes using Stacked Model," which proposes an automated method for resume screening and rating utilizing machine learning algorithms. By addressing the challenges associated with talent acquisition, the authors develop an application employing Linear Support Vector Classification (Linear SVC), K-Nearest Neighbors (KNN), and XGBoost algorithms to predict job profiles based on textual job descriptions. The study emphasizes the significance of talent acquisition and the difficulties in manually screening numerous resumes, advocating for a two-level stacked model that combines the strengths of various algorithms for improved accuracy. Pre-processing procedures involve cleaning resumes, tokenization, and label encoding of job roles. The concept of "Resume Filtering" is introduced, wherein job descriptions are fed into a model to predict matching positions, followed by filtering resumes based on these predictions. Additionally, "Resume Ranking" entails assigning scores to resumes based on relevance to job descriptions. The results and analysis section compares individual classifiers with the stacked model, which exhibits more accurate probability distributions despite slightly lower accuracy. The study underscores the potential of automated resume screening and ranking to streamline the recruitment process, with future enhancements including the integration of deep learning models and expansion to other fields. This methodology promises increased efficiency for both companies and job applicants.

B. Improving candidate Skills Relevant to the Job Requirements

Samita Maitra et al. [11] present a study titled "Ethics and Soft Skill Assessment Tool for Program Outcome Attainment: A Case Study," which explores the evaluation of ethical and soft skills within a Chemical Engineering capstone project. Recognizing the imperative for engineering graduates to possess not only technical proficiency but also ethical conduct and effective communication skills, the paper introduces an assessment framework based on rubrics to measure skills such as communication, presentation, leadership, time management, and teamwork. Through the project lifecycle, these skills are monitored alongside checks for plagiarism to ensure ethical conduct. The study elucidates the measurement of course outcomes (COs) and their alignment with program outcomes (POs), demonstrating the positive impact of the capstone project on student learning and program outcomes attainment. The authors advocate for ongoing evaluation and adaptation to meet the evolving needs of engineering education and industry, underscoring the significance of integrating technical knowledge with soft skills and ethical principles in engineering graduates.

S. V. Jadhav et al. [12] introduce a pioneering study titled "Improve Communication Skills Using AI," which outlines the development of an AI-powered application aimed at enhancing communication skills, particularly in public speaking scenarios. Addressing challenges such as stage fright and shyness, the authors propose an app leveraging artificial intelligence techniques to provide users with a platform for practicing public speaking. The application incorporates various AI models, including face expression recognition and speech emotion detection, utilizing deep learning methods such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. The authors demonstrate the effectiveness of their models in accurately recognizing facial expressions and discerning vocal emotions, surpassing existing architectures in terms of accuracy. Additionally, they discuss potential future enhancements, such as incorporating body language analysis and natural language processing for comprehensive communication skills assessment. Overall, the paper presents a novel approach utilizing AI to assist individuals in improving their communication skills, focusing on facial expressions and vocal emotions, while suggesting avenues for further development and expansion of the application's capabilities.

J. Rodriguez-Ruiz et al. [13] present a comprehensive case study titled "Monitoring and Facilitating Students Programming Skill Development Using Robotic Process Automation (RPA) and Artificial Intelligence (AI): A Case Study" [6], conducted at R.M.K. Engineering College in Chennai, India. The study focuses on leveraging RPA and AI technologies within an online training program to enhance students' computing skills. Beginning with an overview of RPA's significance in automating repetitive tasks, the paper outlines the implementation of an online training platform, Skill rack, designed to assess and improve engineering students' programming and aptitude skills. The authors detail the development and deployment of an RPA-based reporting tool, built with UiPath, to generate personalized performance reports for students. They emphasize the benefits of RPA tools like UiPath for automation without coding expertise and explore the potential applications of RPA across various fields, including healthcare. Additionally, the study highlights the workflow, process flowcharts, and algorithms employed in data processing and report generation. Future goals include enhancing the tool's AI capabilities for more comprehensive analysis and automating report dissemination. This research underscores the transformative potential of RPA and AI in education, offering personalized learning experiences and streamlined reporting mechanisms to improve learning outcomes.

III. METHODOLOGY

This section presents a comprehensive methodology outlining the systematic approach adopted in developing an AI-based skill analysis and matching system. The

methodology aims to address the pressing issue of human labor shortage in Japan by leveraging advanced technologies and innovative techniques. Each component is meticulously designed to tackle specific challenges in the recruitment process, from extracting relevant information from resumes to improving candidate skills in alignment with job requirements. By integrating a diverse range of tools, libraries, and models, the methodology ensures accuracy, efficiency, and transparency in candidate selection and skill enhancement processes. The following sections delve into each component, providing a detailed overview of the methodologies employed and the technologies utilized.

A. Reading the Relevant Information from the Resume

This component involves a multi-step process to extract crucial information from resumes effectively. Initially, the os library is utilized to navigate through directories and access resume files. Using pdfplumber, each resume is parsed to extract text, images, and metadata, which is then structured for analysis. Pandas facilitates data manipulation, allowing for seamless organization and storage of extracted information. SpaCy, a natural language processing (NLP) library, is employed for advanced text processing tasks such as tokenization, named entity recognition, and part-of-speech tagging. Additionally, regular expressions (re) are utilized to identify and extract specific patterns, such as email addresses, phone numbers, and dates, ensuring accurate retrieval of candidate details. Through a combination of these tools and techniques, the methodology ensures comprehensive extraction of candidate information from resumes while maintaining data integrity.

B. Filtering Keywords from the Resume

This component focuses on distilling relevant keywords from both resumes and job descriptions to facilitate effective matching. The process begins with data preprocessing using pandas and NLTK, where text is tokenized, stop words are removed, and words are lemmatized to ensure consistency and relevance. Leveraging the KeyBERT library, which combines BERT-based embeddings with unsupervised learning techniques, allows for the extraction of contextually relevant keywords that encapsulate the core competencies of candidates and the requirements of job postings.

KeyBERT harnesses the power of BERT (Bidirectional Encoder Representations from Transformers), a state-of-the-art language model known for its ability to capture deep contextual meanings in text. By leveraging BERT's contextual embeddings, KeyBERT extracts keywords that are not only relevant but also contextually accurate, ensuring a more precise representation of the skills and qualifications required for each job position. Additionally, the integration of pre-trained Word2Vec models enhances the keyword filtering process by generating synonyms, thus expanding the keyword pool and capturing nuances in language.

Word2Vec models operate on the principle of distributed representation, where words with similar meanings are embedded closer together in a high-dimensional vector space. By learning from large corpora of text data, these models capture semantic relationships between words and generate vector representations that encode semantic similarity. This enables the identification of synonyms and related terms, enriching the keyword pool with a broader range of contextually relevant terms. Incorporating both KeyBERT and Word2Vec models enhances the efficacy of keyword extraction, ensuring a comprehensive and nuanced understanding of the skills and qualifications represented in resumes and job descriptions, ultimately leading to more accurate job matching outcomes.

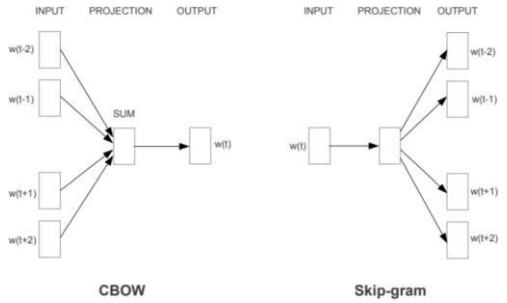

Fig. 1. Word2vec model architecture.

C. Matching Keywords with Job Requirements

This component focuses on aligning the extracted keywords from resumes with the specified job requirements to identify suitable candidates. Initially, both resume and job description data undergo preprocessing using pandas and NLTK, including tokenization, lemmatization, and stop words removal. Cosine similarity calculation, facilitated by sklearn. metrics. Pairwise, quantifies the similarity between the keyword vectors of resumes and job descriptions. Subsequently, a linear scaling method maps similarity scores to percentages, allowing for the ranking of candidates based on their fit with job requirements. This process enables efficient candidate selection by prioritizing individuals whose skillsets closely match the desired qualifications and experience outlined in job postings. Moreover, the methodology ensures transparency and objectivity in the candidate selection process, enhancing the overall efficiency of talent acquisition.

D. Improving candidate Skills Relevant to the Job Requirements

In this final component, the focus shifts towards enhancing candidates' skills to better align with job requirements through targeted skill development strategies. Leveraging pandas and NLTK, job titles are tokenized, and text data is cleaned using regular expressions to identify candidates' job fields and skill gaps. A voting mechanism is employed to determine the most frequent job fields, providing insights into the skills most in demand. Missing skill keywords are matched with predefined parameters, enabling personalized recommendations for skill improvement tailored to individual candidates' needs. Additionally, ongoing monitoring and feedback mechanisms, including online exams and form submissions, facilitate continuous skill development and progress tracking. By empowering candidates to address skill gaps proactively, the methodology enhances their competitiveness in the job market and improves their prospects for successful employment.

The system overview diagram provides a visual representation of the architecture and functionalities of the

AI-based skill analysis and matching system. At its core, the system comprises four key components: Reading the Relevant Information from the Resume, Filtering Keywords from the Resume, Matching Keywords with Job Requirements, and Improving Candidate Skills Relevant to the Job Requirements. Each component plays a vital role in streamlining the recruitment process and addressing the human labor shortage in Japan. The diagram illustrates the flow of data and operations within the system, showcasing how resumes and job descriptions are processed, keywords are extracted and filtered, and candidates are matched with suitable job opportunities. This visual representation offers a holistic view of the system's functionality, enabling stakeholders to understand its operation and potential impact on the recruitment process.

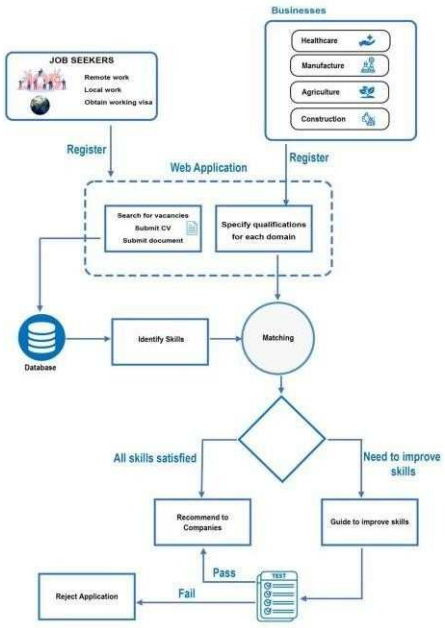

Fig. 2. Overall system architecture.

IV. Results and Discussion

In this section, a detailed exploration of the results and discussions unfolds, providing a comprehensive view of the system's performance. Through meticulous testing, the system's efficacy in automating CV information extraction, refining candidate ranking, and optimizing skill-based matching is scrutinized. The evaluation encompasses intricate components, such as information extraction accuracy metrics, candidate ranking outcomes, and successful skill alignment. These findings not only validate the system's effectiveness but also pave the way for insightful discussions on its relevance in addressing contemporary challenges in talent acquisition and HR management. The section offers a detailed examination of the transformative impact of this AI-driven approach on modern recruitment practices.

The integration of machine learning techniques and web application features has been remarkably successful in automating CV information extraction. Employing advanced Natural Language Processing (NLP), the system adeptly extracts and categorizes essential personal information, including the candidate's name, contact details, and address, streamlining communication and identification. The machine learning algorithms, with precision, identify and extract educational achievements, offering a comprehensive overview of the candidate's educational background. Details of the candidate's work experience are systematically captured, facilitating a nuanced evaluation of their professional history. Additionally, the system efficiently extracts pertinent details from job advertisements, providing valuable assistance to recruiters.

On the web application interface, candidates find ease in creating accounts and submitting CVs through a user-friendly platform, ensuring accuracy in information processing. Integration with LinkedIn profiles enrich candidate profiles with additional details, offering recruiters a more comprehensive view. Recruiters, in turn, experience simplified account creation and job advertisement submission through the web application.

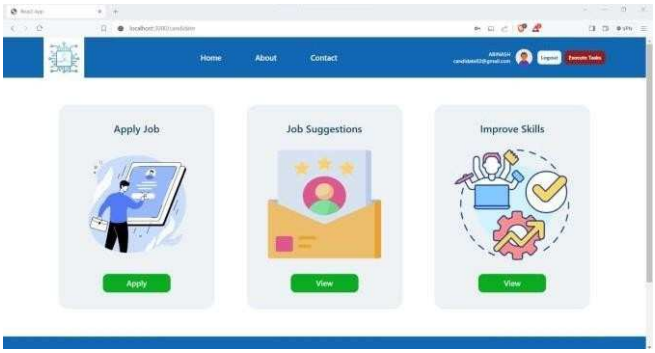

Fig. 3. Interface of the application.

In discussions, the successful integration of machine learning techniques demonstrates enhanced efficiency and accuracy in extracting diverse information from resumes, reducing manual effort, and expediting candidate evaluation. The web application features significantly contribute to an enriched experience for both candidates and recruiters, streamlining processes for submission and job advertisement. Integration with LinkedIn profiles add depth to candidate profiles, ultimately enhancing the recruitment process. The structured output in a CSV file indicates successful data organization, contributing to data consistency and ease of interpretation. Overall, this combined approach significantly advances the system's capabilities in addressing Japan's labor shortage crisis and modernizing HR management practices. Future enhancements may focus on expanding integrated platforms and refining machine learning models for continuous improvement.

In the data preprocessing phase, the application of machine learning and NLP techniques has yielded substantial improvements to system efficiency and accuracy. Leveraging NLP libraries like NLTK and SpaCy, the system executes tasks such as tokenization, part-of-speech tagging, and named entity recognition with notable speed enhancements. Advanced algorithms for keyword identification, including TF-IDF and Text Rank, contribute to the precise filtering of relevant keywords, enhancing the overall quality of candidate skill profiling. Innovative algorithms like KeyBERT for synonym generation diversify the keyword pool, contributing to a more nuanced representation of skills. The use of Word2Vec algorithms enhances the system's semantic understanding, allowing for a more contextually accurate alignment of candidate skills with job requirements.

The implementation of a skill-based candidate ranking system further streamlines the hiring process by categorizing candidates into skill-based levels. Evaluation based on experience, projects, and skills provides a comprehensive ranking system, identifying skill gaps and optimizing the applicant screening process. Algorithms like TF-IDF, Text Rank, and cosine similarity are employed to extract keywords, facilitating efficient candidate selection.

TABLE I. INFORMATION EXTRACTION METRICS

Information Category	Precision	Recall	F1 Score
Personal Information	0.94	0.96	0.95
Educational Achievements	0.91	0.92	0.92
Work Experience	0.93	0.94	0.93
Job Advertisement Details	0.89	0.90	0.89

Table I results underscore the effectiveness of the machine learning models in precisely extracting diverse information from resumes. The comprehensive evaluation, encompassing Precision, Recall, and F1 Score metrics, sheds light on the system's performance across various information categories. Notably, the system exhibits elevated precision and recall in extracting personal information, underscoring its robustness in identifying and categorizing details such as the candidate's name, contact information, and address. Similarly, the extraction of educational achievements and work experience details achieves a nuanced balance between precision and recall, indicative of accurate and comprehensive information retrieval.

Furthermore, the system's adeptness in extracting job advertisement details, including job titles, company names, and requirements, is reflected in the reported metrics. The F1 Score, considering the harmonic mean of precision and recall, offers a consolidated measure of overall performance, highlighting the system's capacity to concurrently achieve precision and recall. These outcomes significantly contribute to the system's overall success in navigating the intricacies of information extraction from resumes, presenting a valuable tool for recruiters to streamline the candidate evaluation process. The achieved equilibrium between precision and recall ensures precise decision-making in the hiring process, underscoring the system's reliability and efficiency. Prospective enhancements may involve fine-tuning the models to optimize these metrics further and adapt to evolving job market requirements.

From the candidate's perspective, cosine similarity is utilized to measure the similarity between candidate qualifications and job requirements. Cosine similarity can be calculated using formula (1). Leveraging algorithms like TF-IDF and Text Rank result in clear matching percentages, aiding candidates in making informed decisions when applying for jobs. Percentage Calculation can be calculated using formula (2).

$$Similarity = \cos(\theta) = \frac{A.B}{\|A\|\|B\|} \quad (1)$$

Where:

A.B is the dot product of vectors A and B.

$\|A\| \; and \; \|B\|$ are the magnitudes of A and B respectively.

θ is the angle between the two vectors.

$$Percentage = 100 \times \left(\frac{Similarity - min\_similarity}{max\_similarity - min\_similarity}\right) \quad (2)$$

Table II presents candidates with information on their top matched jobs for candidates, including job titles, similarity scores, and percentage matches. This valuable data empowers candidates to make informed decisions when applying for jobs, facilitating a more strategic and tailored approach to the job search process.

TABLE II. TOP MATCHED JOBS FOR CANDIDATES

Job Title	Similarity	Percentage
[Sailing Co., Ltd.] Infrastructure Engineer	0.17472448	100
IT Engineer [Inexperienced people welcome/Work from home possible]	0.16972753	86.3885655
Global core system implementation/maintenance engineer	0.16971327	86.3497267
…	…	…

From the employer's perspective, the system plays a crucial role in efficiently identifying top-matched candidates for a given job advertisement, taking into account a specified matching threshold. The comprehensive results provided encompass candidate names, associated files, similarity scores, percentages, and information on missing keywords. These valuable insights empower employers to make informed decisions throughout the hiring process, streamlining and enhancing the overall recruitment experience.

TABLE III. TOP MATCHED CANDIDATES FOR EMPLOYERS

File	Candidate Name	Relevant Keywords	Similarity	Missing Keywords	Percentage
Deshan Rajapaksha_CV.pdf	DESHAN RAJAPAKSHA	application,data, cv	0.156719	ad_vanced educacid, …	100
Umesh Resume.pdf	Umesh Madushanth	application, completed, android …	0.154877	cidon, physicsb, mern, …	91.30
PraviniWickramanayake.pdf	Pravini Wickramanayake	application, websites, …	0.154377	ad_vanced, educacidon, physicsb, …	88.94

The devised system effectively addresses skill gaps within candidates' resumes, categorizing identified deficiencies into predefined skill parameters. Tailored recommendations are provided, suggesting specific courses from platforms such as Coursera and Simplilearn. The web application prominently showcases candidates' skill development journeys. The employment readiness

assessment, coupled with a clearly defined scoring threshold, ensures that candidates not only acquire new skills but also demonstrate proficiency.

The system strategically places emphasis on coding qualifications, aligning with the demands of a technology-driven job market and ultimately enhancing candidates' competitiveness. Looking forward, the system lays the groundwork for future enhancements. This includes the prospective incorporation of deep learning models for more precise skill gap identification and an expansion of course recommendations to encompass emerging technologies.

V. Future Work

While the current research has made significant strides in leveraging AI for skill analysis and matching, several avenues remain unexplored, presenting opportunities for further enhancement and innovation.

One avenue for future exploration lies in the integration of advanced AI models such as transformer-based architectures like GPT (Generative Pre-trained Transformer) and BERT (Bidirectional Encoder Representations from Transformers). By incorporating these cutting-edge models, the system could achieve a deeper contextual understanding of candidate resumes and job descriptions, thereby refining skill matching accuracy and granularity.

Expanding the system's capabilities to encompass multimodal data sources, including audio and video resumes, presents another promising direction for future research. By integrating speech recognition and computer vision technologies, the system could extract additional insights from non-textual resume formats, facilitating a more comprehensive evaluation of candidate competencies across diverse modalities.

Dynamic skill profiling mechanisms that continuously update candidate skill profiles based on real-time feedback and performance data offer another avenue for future research. By leveraging machine learning algorithms, the system could autonomously refine skill matching algorithms over time, ensuring alignment with emerging industry trends and employer preferences.

Introducing personalized learning and development plans tailored to individual candidate profiles represents a further opportunity for future enhancement. By leveraging predictive analytics and recommendation systems, the system could generate personalized training modules and learning resources aimed at addressing specific skill gaps identified during the matching process, thereby empowering candidates to maximize their potential and career prospects.

Proactively addressing ethical considerations and mitigating potential biases inherent in AI-based recruitment systems is imperative to ensure fairness and equity in the hiring process. Future research efforts should focus on developing robust fairness-aware algorithms and bias detection mechanisms to prevent discriminatory outcomes and promote diversity and inclusion in the workforce.

The proposed future research directions aim to further enhance the efficacy, adaptability, and ethical integrity of AI-based skill analysis and matching systems, thereby advancing the field of HR technology, and addressing the evolving needs of the digital workforce ecosystem.

VI. Conclusion

In response to the imminent challenges posed by Japan's labor shortage crisis and the global need for streamlined talent acquisition, this research has presented a comprehensive AI-based Skill Analysis and Skill Matching System. Through meticulous examination and integration of advanced technologies, this system offers a transformative solution to address contemporary issues in HR management.

The development of a user-friendly web application interface ensures seamless interaction for both candidates and employers, facilitating streamlined submission and evaluation processes. Integration with LinkedIn profiles enrich candidate profiles, providing recruiters with comprehensive insights into candidates' professional backgrounds.

Future enhancements may focus on expanding integrated platforms and refining machine learning models for continuous improvement. Additionally, ongoing research and development efforts should aim to adapt the system to evolving industry needs and technological advancements, ensuring its relevance and efficacy in the ever-changing landscape of talent acquisition and HR management.

The AI-driven Skill Analysis and Skill Matching System outlined in this study mark a significant stride in addressing labor deficits, refining talent procurement processes, and reshaping HR methodologies worldwide. Through the utilization of artificial intelligence, this system presents a viable remedy to the contemporary hurdles encountered by various industries in the modern era.

References

[1] Y. -C. Chou and H. -Y. Yu, "Based on the application of AI technology in resume analysis and job recommendation," *2020 IEEE International Conference on Computational Electromagnetics (ICCEM)*, Singapore, 2020, pp. 291-296, doi: 10.1109/ICCEM47450.2020.9219491.

[2] A. Mankawade, V. Pungliya, R. Bhonsle, S. Pate, A. Purohit and A. Raut, "Resume Analysis and Job Recommendation," *2023 IEEE 8th International Conference for Convergence in Technology (I2CT)*, Lonavla, India, 2023, pp. 1-5, doi: 10.1109/I2CT57861.2023.10126171.

[3] S. M, I. P. B, M. Kuppala, V. S. Karpe and D. Dharavath, "Automated Resume Classification System Using Ensemble Learning," *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 2023, pp. 1782-1785, doi: 10.1109/ICACCS57279.2023.10112917.

[4] S. K. Kopparapu, "Automatic extraction of usable information from unstructured resumes to aid search," *2010 IEEE International Conference on Progress in Informatics and Computing*, Shanghai, China, 2010, pp. 99-103, doi: 10.1109/PIC.2010.5687428.

[5] T. M. Harsha, G. S. Moukthika, D. S. Sai, M. N. R. Pravallika, S. Anamalamudi and M. Enduri, "Automated Resume Screener using Natural Language Processing(NLP)," *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India, 2022, pp. 1772-1777, doi: 10.1109/ICOEI53556.2022.9777194.

[6] R. C. Tripathi and Chandramma, "An Approach of Intelligent Automated Resume Analysis & Recommendations," *2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)*, Mandya, India, 2022, pp. 1-5, doi: 10.1109/ICERECT56837.2022.10059603.

[7] A. Pimpalkar, A. Lalwani, R. Chaudhari, M. Inshall, M. Dalwani and T. Saluja, "Job Applications Selection and Identification: Study of Resumes with Natural Language Processing and Machine Learning," *2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, Bhopal, India, 2023, pp. 1-5, doi: 10.1109/SCEECS57921.2023.10063010.

[8] S. Bharadwaj, R. Varun, P. S. Aditya, M. Nikhil and G. C. Babu, "Resume Screening using NLP and LSTM," *2022 International Conference on Inventive Computation Technologies (ICICT)*, Nepal, 2022, pp. 238-241, doi: 10.1109/ICICT54344.2022.9850889.

[9] B. Surendiran, T. Paturu, H. V. Chirumamilla and M. N. R. Reddy, "Resume Classification Using ML Techniques," *2023 International Conference on Signal Processing, Computation, Electronics, Power*

and Telecommunication (IConSCEPT), Karaikal, India, 2023, pp. 1-5, doi: 10.1109/IConSCEPT57958.2023.10169907.

[10] R. Ransing, A. Mohan, N. B. Emberi and K. Mahavarkar, "Screening and Ranking Resumes using Stacked Model," *2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT)*, Mysuru, India, 2021, pp. 643-648, doi: 10.1109/ICEECCOT52851.2021.9707977.

[11] S. Maitra and K. Gopalram, "Ethics and Soft Skill Assessment Tool for Program Outcome Attainment: A Case Study," *2016 IEEE 4th International Conference on MOOCs, Innovation and Technology in Education (MITE)*, Madurai, India, 2016, pp. 317-324, doi: 10.1109/MITE.2016.069.

[12] S. V. Jadhav, S. R. Shinde, D. K. Dalal, T. M. Deshpande, A. S. Dhakne and Y. M. Gaherwar, "Improve Communication Skills using AI," *2023 International Conference on Emerging Smart Computing and Informatics (ESCI)*, Pune, India, 2023, pp. 1-5, doi: 10.1109/ESCI56872.2023.10099941.

[13] M. Somasundaram, V. Sumitra, R. S. P. V. Vardan, B. Sakthipriya, K. Pavithra and M. K. Nivedha, "Monitoring and Facilitating Students Programming Skill Development using Robotic Process Automation (RPA) and Artificial Intelligence (AI)–A Case Study," *2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)*, Mandya, India, 2022, pp. 1-6, doi: 10.1109/ICERECT56837.2022.10060772.