2025 IEEE International Conference on Computer, Electronics, Electrical Engineering & their Applications (IC2E3) | 979-8-3315-2439-5/25/$31.00 ©2025 IEEE | DOI: 10.1109/IC2E365635.2025.11167576

AI-Driven Recruitment: Resume Screening and Skill Matching with NLP

Barath G V
Department of Computer Science and Engineering,
Amrita School of Computing,
Amrita Vishwa Vidyapeetham,
Chennai, India.
barathsegvm23@gmail.com

Sanjev R
Department of Computer Science and Engineering,
Amrita School of Computing,
Amrita Vishwa Vidyapeetham,
Chennai, India.
sanjevrv@gmail.com

V Shenbaga Priya
Department of Computer Applications,
School of Computer Information
and Mathematical Sciences,
B.S.Abdur Rahman Crescent
Institute of Science and Technology,
Vandalur, Chennai, India.
shenbagapriya@crescent.education

Pandiyavathi Thenral Manoharan
Department of Computer Applications,
School of Computer Information and Mathematical Sciences,
B.S.Abdur Rahman Crescent Institute of Science and Technology,
Vandalur, Chennai, India.
pandiyavathi@crescent.education

Saranya S
Department of Computer Science and Engineering,
Vels Institute of Science, Technologies and Advanced Studies,
Pallavaram,Chennai,India.
saranyas.se@vistas.ac.in

Fizza Ghulam Nabi
Department of Industrial Engineering and Management,
University of the Punjab,
Lahore 54000, Pakistan.
engr.fizza@yahoo.com

Sindhu Ravindran*
Department of Computer Science and Engineering,
Amrita School of Computing,
Amrita Vishwa Vidyapeetham,
Chennai, India.
r_sindhu@ch.amrita.edu*

***Abstract*—The employment process is greatly enhanced and effective by the application of artificial intelligence (AI) in recruitment process. This work presents a data-driven approach to improve resume screening and job matching. This paper presents a data-driven approach to improve resume screening and job matching using Natural Language Processing (NLP) techniques, specifically spaCy and Word2Vec, along with AI-powered technologies. By focusing on skill extraction through keyword matching and vectorization, the system efficiently processes resumes and job descriptions.In order to help understand the process, The study also explores how these AI technologies promote more inclusive and equitable recruitment practices, streamlining the hiring process, improving candidate-job alignment, and offering a scalable, ethical solution for modern HR systems..This research provide a diverse filed of AI in HR for scalability, ethical and equitable recruitment systems.**

I. Introduction

The sector of human resources is undergoing fast change due to the integration of Artificial Intelligence (AI) in recruiting,which brings with it both enormous benefits and problems[12]. In the past, hiring practices included time-consuming procedures including hand-screened resumes, employee recommendations, and newspaper advertising.[6] Despite being simple to use, these approaches frequently produced less-than-ideal results because of ingrained human biases and the enormous volume of applications that needed to be processed[5]. While some of these procedures were streamlined by the move to digital hiring in the late 1990s, recruiting saw a complete transformation with the introduction of artificial intelligence[17].

A. 1.1 The Impact of AI on Recruitment

AI has revolutionized the recruitment procedure by facilitating quicker and more objective decision making[13]. Artificial Intelligence (AI) has significantly reduced the time and resources required to locate and employ top talent by automating processes like resume screening and utilizing chatbots for first

candidate interactions[8]. With natural language processing, modern AI systems can assess soft skills, forecast job fit, and even perform preliminary interviews[9]. The developments in AI have an ability to reduce the biases associated with human decision-making in addition which increases recruiting efficiency[7].

B. 1.2 Importance of Responsible AI

However artificial intelligence (AI) has many advantages applicable in hiring presents serious difficulties, especially when it comes to propagation of prejudices, If AI systems are not carefully created[16]. The AI algorithms design is opaque decision-making process can provide results that are challenging in interpretations[10]. In addition to that,creating Responsible AI for hiring, which points up the responsibility, transparency, and justice in AI design and implementation[11].

C. 1.3 Problem Statement and Study Objectives

Even with advancements AI, there are still existing problems with its flexibility, efficiency, and justice in the present recruiting processes[7]. Recruiters invest a significant amount of time recurrently in CV matching with job specifications[4]. Moreover, individuals continually apply for the jobs which does not match the required qualifications, which makes shortlisting even more difficult as an individual[14]. An intelligent resume rating system is needed for a recruiter, one that can match with the job descriptions automatically and selects the most eligible applicants within a short periodic interval and accurately[9]. The main objective of this research is to create an effective system that accurately do resume screening process which utilizes machine learning approaches, specifically through learning-to-rank resumes[3]. By automating hiring process, the motive of study hopes to save expenses of organizations, increase productivity, and also guarantee that the best applicants can be paired with open positions[1]. Candidates gain from this strategy as it places them in positions best suit their qualifications and makes recruiting teams function more smoothly[2].

D. 1.4 The Potential of AI-Enhanced Recruitment

AI's has capacity to learn, analyze and predict with vast amounts of data enables recruiters to pinpoint repective candidates using Job description in extensive applicant pools[8]. This implementation particularly important in the worldwide talent market, who's ability to effectively connect candidates with job openings in various regions is crucial[6].

However, AI can be viable in decreasing the administrative tasks of HR professionals can concentrate on more strategic aspects rather than matching resume with job specification[15].

II. Literature survey

In recent years, cutting-edge technologies are transforming many different ways businesses and job seekers can interact with recruiting in the constantly evolving market[12]. Inefficiencies in traditional hiring practices often lead to mismatches between candidates and job roles[5]. To speed up recruiting and mainly enhance the job matching, this review of literature explores a various method that utilizes data gathering techniques, machine learning algorithms, and natural language processing (NLP)[3]. The motive of this research is to examine the various modules that including automatic resume summarization and job categorization with Word embeddings techniques[1]. Both employers and job seekers finally see the advantages from the research findings, aiming to improve candidate matching, streamlines hiring process, and boost the recruitment strategies[2].

I. The method combines skills of the applicants into a multimodal network, LSTM networks for text analysis and facial analysis using ResNet-50. It tests the impact of the biases regarding gender and ethnicity on CV evaluation through three major scenarios: neutral, prejudiced, and agnostic. The SensitiveNets was utilized to remove the identifiable traits like gender and race, which helps in enhancing privacy in the learning domain while maintaining performance. This model offers a unique perspective on data analysis to reduce biases in recruitment procedures[1].

II. BERT, an advanced language model which helps the system to analysing text features from resumes and calculates a score using the combination of factors like location, education, experience, and skills. Using these factor for ratings, machine learning algorithms are utilized to predict the candidate's success rate. This model can be customized to achieve different recruiting objectives by recasting each component. further the technology provide a similarity measurement to candidates with job requirements, aiding recruiters to make decisions using data[8].

III. The dataset containing 14,906 candidate resumes (CRs) and eight job descriptions (JDs) from Kaggle which is prepared for analysis. CRs include the qualifications and skills of respective candidates based on resume, while JDs lists job requirements, company details, responsibilities, and preferred skills. AI techniques like Jaccard similarity and Natural Language Processing which has been used to evaluate compatibility between CRs and JDs. To evaluate the compatibility, this researchers examines descriptors, modifiers, and fundamental and additional skills and also It also categorizes candidates based on Most Suitable (MOS), Moderately Suitable (MDS), and Not Suitable (NTS) techniques[13].

IV. The collection of 2,452 resumes on different fields such as engineering and medicine. Information such as name, education, skills. The work history are also encompassed. The model architecture bring into play data pre-processing steps such as segmentation, tokenization, stemming, lemmatization, part-of-speech tagging, and entity annotation with the Spacy tool. Named Entity Recognition (NER) tool take advantage off identification and categorizing entities. After training the model with an 80-20 train-test split, the resumes will be compared to job descriptions using cosine similarity to finalise selections[4].

V. The method is using machine learning and deep learning to automate the process of evaluating and selecting resumes,aims to simplify the recruitment process. The system is able to evaluate resumes of many file formats (PDF, DOC, DOCX), utilize NER to extract relevant details regarding a

resume's skills, background, and qualifications, and assess resumes alignment with job criteria. Cosine similarity and vectorization techniques like TF-IDF and BERT used to assess the alignment between resumes and a job. The Linear SVC classifier enhances recruiting effectively by sorting resumes based on job categories[8].

VI. The system uses Large Language Models (LLMs) and Natural Language Processing (NLP) to automate the process of analysis of resumes. NLTK is utilized for text preprocessing and BERT is for summarization, and spaCy is for Named Entity Recognition (NER) to extract key elements from the text. Their proposed model utilizes web scraping for data collection, the T5 model used for generating relevant questions and BERT helps in delivering the answers. The dataset enhances adaptability in real world for job descriptions and resumes[2].

VII. The recommended method involves the organizing and tidying up job listings, and then utilizing web scraping to extract relevant information from LinkedIn. Job Categorization system employs Word2Vec embeddings to precisely categorize job descriptions. The Resume Summarization compares candidates with positions by utilizing qualifications on respective resume. Questions in interviews are created on candidate's resume by the mentioned skills and abilities. The Word2Vec model demonstrates the potential in recognizing the job titles using word embedding with an efficient accuracy[11].

III. Methodology

A. Data Extraction from PDF Resumes

The process is extracting information from resumes begins with loading the necessary libraries such as spaCy and pdfplumber, where spacy is crucial for natural language processing and The initial step involves extracting text from PDF resumes using pdfplumber. This text extraction is a vital component of a broader data extraction and analysis each of the resumes, here the contents from each PDF page is read and combined together as a string. This string captures all relevant information, including personal details, education, work experience, and skills of all resumes separately in a dataframe usinf pandas. Once the text is extracted, it undergoes process of cleaning and removal of unnecessary whitespace and irrelevant artifacts, preparing the data for further analysis. This meticulous extraction is an essential quality of the text which directly influences the accuracy of subsequent processes of skill matching. To identify relevant skills from the resumes and job descriptions , a list of skill keywords is defined, encompassing all the skills required from resume.By intersecting the extracted words from resumes to skill keywords, the relevant skills are matched, enhances the ability to align candidate profiles with job requirements effectively.

B. Text Cleaning and Normalization

The process of extracting information from resumes starts with extracting text from PDF documents is essential for broader data extraction and analysis tools designed for handling PDFs and natural language processing. Once the raw text

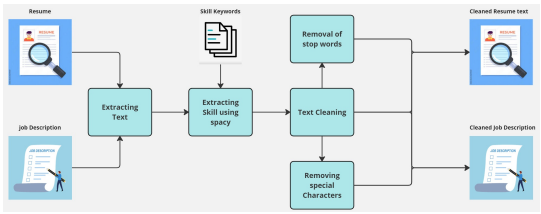

Fig. 1.

is obtained, the next step is to clean and normalize the data, as raw text can be messy and inconsistent it may also often containing irrelevant information such as formatting artifacts, special characters and typos. Text cleaning is crucial during this stage; all text is converted to lowercase to maintain uniformity, ensuring that "Python" and "python" are treated as same entries. Punctuation and unwanted characters, such as URLs and email addresses, are removed, as they do not contribute to the analysing resumes. The contractions are expanded (e.g., changing "don't" to "do not") to standardize language usage. The goal of this meticulous cleaning process is to create a uniform and streamlined dataset, allowing the focus to be on relevant information that contributes to skill extraction. Clean text is vital for accurate analysis, enhancing the model's ability to recognize skills and qualifications effectively. This stage is pivotal in the workflow which sets a foundation for successful extraction and analysis.

C. Skill Extraction through Keyword Matching

The focus shifts to skill extraction where cleaned and normalized texts are ready then crucial step in evaluating a candidate's suitability for specific roles in a competitive job market. To facilitate this, a comprehensive list of keywords representing various skills relevant to different job sectors are compiled, including programming languages, frameworks, tools, and methodologies. The extraction process involves systematically searching the cleaned text for these keywords. Whenever a keyword is found, it is marked and recorded. This keyword matching approach is effective spacy technique, allowing for the creation of a clear inventory of the skills possessed by each candidate from pool of resume. It directly influences the model's ability to comprehend a candidate's qualifications and match them with the potential job demands on basis of job description and respective skills. Natural language processing is utilized to identify skills from the text, processing the content to extract entities which is related to skills. This extraction is supplemented with keyword matching against the data of predefined list. Additionally, resumes are loaded from a specified folder then extracting text and skills while handling both PDF and TXT formats. This method ensures that both common skills are captured that providing a comprehensive view of each candidate's abilities. By employing this targeted approach to skill identification estimated analysis becomes more efficient and adaptable, ensuring a thorough evaluation of each candidate's qualifications.

D. Vectorization of Skills and Text Data

The step converts both skills and text data into a numerical format for analysis and train the word2vec model for an effective model. This transformation, known as vectorization, creates a mathematical representation of the text, enabling calculations and comparisons. Vectorization uses advanced techniques to turn text data into vectors with a series of numbers that represent specific words from resumes and job description. This process allows the creation of a vector space model to analyze common relationships between different pieces of information. Vectorization is significant for comparing resumes and job descriptions by representing text data numerically. Therfore, it allows for calculating similarities and differences in candidates' skills and qualifications. This capability is essential for deriving meaningful insights and making informed decisions in the job matching process. The enhanced understanding is the key how well a candidate's skills align with job requirements. In this approach, a model is trained using extracted skills from resumes and job descriptions. Skills from both sources are compiled into sentences for training. The resulting output is in numerical representations that improve the recruitment process by analyzing the relevance of these skills to job roles.

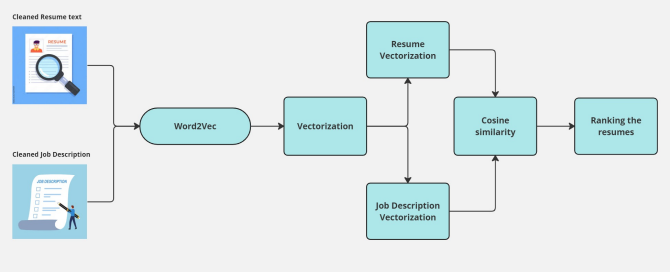

Fig. 2.

E. Cosine Similarity Calculation for Job Matching

Cosine similarity is a mathematical technique used to assess the compatibility between two sets of data, such as resumes and job descriptions which vectorized and output of numerical representation. This scoring system identifies candidates who not only meet basic qualifications but also those who exceed them, offering a nuanced view of suitability. Cosine similarity transforms the analysis from a binary "fit" or "not fit" decision into a more detailed assessment of compatibility. It evaluates the cosine of the angle between two vectors, providing a measure of their alignment. Smaller angle indicates that greater similarity whereas the larger angle has significant differences. Then by calculating cosine similarity scores between two vectors of Job description and Resume, it becomes possible rank to match skills specific job requirements. This scoring mechanism is essential in recruitment process offers a data-driven insights that supported to decision-making by emphasizing skills and competencies. This enables the identification of most suitable candidates which ultimately streamlining recruitment and improving outcomes.

$$\text{Cosine Similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|}$$

$$\frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

The cosine similarity score helps to determine two pieces of information how well align, such as skills in a resume and the requirements of a JD. The score ranges from -1 to 1:

- 1: Perfect match – the two sets are completely aligned means the content is identical in direction and intent.
- 0: No match – – the two sets are unrelated and showing no meaningful connection between them.
- -1: Complete mismatch – the content is entirely opposite though this is rare in text-based comparisons.

The higher cosine similarity scores indicate the candidate's skills and qualifications align more closely with the job requirements and it makes easier to shortlist resumes that are well-suited for a particular role which improving the job matching process.

F. Shortlisting Resumes for Job Matching

The process of shortlisting resumes for a specific job description involves several key steps. Initially, skills are extracted from the job description, which is then cleaned and prepared for analysis. The Utilizing techniques such as natural language processing, the relevant skills are identified and converted into a vector representation and predict the scores from job description skills. Each resume is processed to extract and vectorize its skills, allowing for a comparison between the resume and the job description. Next calculating score using cosine similarity between the job skills vector and each resume vector, a similarity score is generated. This score reflects how closely the skills of a candidate align with the requirements of the job. The results, which include the resume identifiers and their corresponding similarity scores, are compiled into a structured format. These results can be sorted based on similarity scores to identify the top matches. The final step involves in saving the shortlisted candidates to a CSV file for further review or to take an action. And also the method not only improvise the recruitment process but also provides the data-driven approach to identifying the best-fit candidates for specific job roles.

IV. Result and Analysis

In this study, we explored various word embedding techniques such asWord2Vec Tf-Idf and Transformer based models like Bert, Roberta, Distilbert to effectively match JD with candidate resumes. We carefully processed each resume to extract relevant skills and qualifications to transforming them into vectors with high-dimensional space. The cosine similarity measures how closely the embeddings of the resumes aligned with those of the job descriptions, allowing us to gauge

each candidate fit the job requirements. Based on the cosine similarity scores we ranked the candidates to giving us a clear picture of who the most suitable can didates were for each position. This ranking helped streamline the selection process, making it easier to identify top matches for the roles.

Resume ID	Score	Job Title
13264796.pdf	0.969244	SDE
17926546.pdf	0.968187	SDE
13149176.pdf	0.968109	SDE
61579998.pdf	0.964718	SDE
32081266.pdf	0.964615	SDE

V. CONCLUSION

This research successfully illustrates the effectiveness of word embedding techniques including Word2Vec and Tf-Idf alongside transformer-based models such as BERT, RoBERTa, and DistilBERT which helps in improving the matching process between job descriptions and candidate resumes. By converting resumes and job descriptions into high-dimensional vector representations. We utilized cosine similarity to measure the candidates's qualifications aligned with job requirements.

The findings reveals a significant enhancement to identify suitable candidates with higher cosine similarity scores to JD indicating stronger matches. Our comparative analysis showed that Word2Vec consistently provided the most accurate rankings, highlighting its valuable role in recruitment processes. By embracing these advanced technique organizations can streamline their hiring efforts and to enhance candidate selection accuracy which ultimately improve their overall recruitment efficiency.

VI. KEY CONSIDERATION

While the results are promising, it's important to acknowledge some key considerations that this approach. One major concern is the reliance on the quality and comprehensiveness of the training data. If the model is trained on a limited or biased dataset, it may fail to capture the full spectrum of skills or qualifications, potentially leading to inaccurate matches. Additionally, word embeddings might struggle to account for nuanced language, such as contextual meanings or emerging skills that are not adequately represented in the training data.

Moreover cosine similarity serves a valuable metric for matching, it overlooks other critical factors like cultural fit and soft skills are the essentials for successful hiring. Addressing these considerations will be crucial for optimizing automated job matching processes and resume screening process in the future research ensuring that organizations find the right candidates for their teams.

VII. FUTURE WORK

In Future the research could be enhancing the candidate matching process by integrating diverse data sources across various profiles and with performance reviews to provide a more comprehensive view of candidates. Fine tuning transformer-based models like BERT or RoBERTa on specialized datasets could be further improve the understanding of complex relationships between skills and job requirements . Additionally, developing hybrid models that combine traditional keyword matching with advanced embedding techniques may ensure crucial skills that are not overlooked. Implementing feedback loops from hiring managers would also facilitate continuous model refinement with aligning predictions with real-world hiring practices and ultimately improving candidate selection.

VIII. REFERENCE

[1] Julian, A., Haripriya, K. (2024, March). NLP based Resume Analysis and Adaptive Skill Assessment System. In 2024 3rd International Conference for Innovation in Technology (INOCON) (pp. 1-5). IEEE.

[2] Varalakshmi, P., Bugatha, N. M. K. (2024, March). AI-Powered Resume Based QA Tailoring for Success in Interviews. In 2024 Third International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS) (pp. 1-6). IEEE.

[3] Jaiswal, G., Uttam, A., Dubey, D. D., Mall, P. K. (2024, March). Resume Analyser and Job RecommendationSystem Based on NLP. In 2024 2nd International Conference on Disruptive Technologies (ICDT) (pp. 1584-1587). IEEE.

[4] Ambareesh, S., Thakur, N. K., Bhattarai, U., Yadav, S. K., Thakur, J. N., Mahato, A. K. (2024, March). Resume Shortlisting Using NLP. In 2024 4th International Conference on Data Engineering and Communication Systems (ICDECS) (pp. 1-5). IEEE.

[5] Rigotti, C., Fosch-Villaronga, E. (2024). Fairness, AI recruitment. Computer Law Security Review, 53, 105966.

[6] Ajayi, F. A., Udeh, C. A. (2024). Innovative recruitment strategies in the IT sector: A review of successes and failures. Magna Scientia Advanced Research and Reviews, 10(2), 150-164.

[7] Albaroudi, E., Mansouri, T., Alameer, A. (2024). A Comprehensive Review of AI Techniques for Addressing Algorithmic Bias in Job Hiring. AI, 5(1), 383-404.

[8] Jacob, P. M., Jacob, S., Cheriyan, J., Nair, L. S. (2023, December). ResumAI: Revolutionizing Automated Resume Analysis and Recommendation with Multi-Model Intelligence. In 2023 Global Conference on Information Technologies and Communications (GCITC) (pp. 1-7). IEEE.

[9] Tanberk, S., Helli, S. S., Kesim, E., Cavsak, S. N. (2023, September). Resume Matching Framework via Ranking and Sorting Using NLP and Deep Learning. In 2023 8th International Conference on Computer Science and Engineering (UBMK) (pp. 453-458). IEEE.

[10] Varsha, P. S. (2023). How can we manage biases in artificial intelligence systems–A systematic literature review. International Journal of Information Management Data Insights, 3(1), 100165.

[11] Vivek, R. (2023). Enhancing diversity and reducing bias in recruitment through AI: a review of strategies and

challenges. . . /Informatics. Economics. Management, 2(4), 0101-0118.

[12] Albassam, W. A. (2023). The power of artificial intelligence in recruitment: An analytical review of current AI-based recruitment strategies. International Journal of Professional Business Review, 8(6), e02089-e02089.

[13] Sridevi, G. M., Suganthi, S. K. (2022). AI based suitability measurement and prediction between job description and job seeker profiles. International Journal of Information Management Data Insights, 2(2), 100109.

[14] Delecraz, S., Eltarr, L., Becuwe, M., Bouxin, H., Boutin, N., Oullier, O. (2022, May). Making recruitment more inclusive: Unfairness monitoring with a job matching machine-learning algorithm. In Proceedings of the 2nd International Workshop on Equitable Data and Technology (pp. 34-41).

[15] Koumoutsos, A., Bakas, G. (2022). Artificial Intelligence tools, Recruiting process Biases.

[16] Pena, A., Serna, I., Morales, A., Fierrez, J. (2020). Bias in multimodal AI: Testbed for fair automatic recruitment. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 28-29).

[17] Ochmann, J., Michels, L., Tiefenbeck, V., Maier, C., Laumer, S. (2024). Perceived algorithmic fairness: An empirical study of transparency and anthropomorphism in algorithmic recruiting. Information Systems Journal, 34(2), 384-414.

[18] Pang, R. (2024, February). Research on the Application of Word2Vec-Based Job-Person Matching Methods in Corporate Recruitment. In Proceedings of the 2024 16th International Conference on Machine Learning and Computing (pp. 506-510).