Resume Flow-Streamlined Resume Parsing for Hiring Success

M. Raja Ramesh[1], G. Rohitha[2], B. Sri Harsha[3], E. Muralidhar Reddy[4], D. Krishna Varma[5]

[1]Associate Professor, Department of Information Technology
[2,3,4,5]Student, Department of Information Technology,
Vishnu Institute of Technology, Bhimavaram, Andhra Pradesh, India-534202

***Abstract**—Businesses have the difficulty of adjusting to the shifting dynamics of candidate sourcing in the constantly changing field of recruitment. Traditional hiring methods are seeing declining benefits as online recruiting becomes more and more common. The traditional methods of creating interview shortlists, carefully reviewing resumes, and going through a large number of applications by hand have become labo - intensive tasks. To make matters worse, today's employment market is dominated by tech-savvy people who have taken a more methodical and approachable approach to job searching. Due to this change, there is an abundance of resumes, many of which are formatted incorrectly or do not adequately highlight the candidate's qualifications. Despite the introduction of automated resume screening technologies to expedite the recruiting process, a sizable segment of businesses and individuals continue to rely on them.*

Keywords— *Candidate ranking, Machine learning, Nltk, Resume screening, SpaCy and Web application.*

I. INTRODUCTION

The $200 billion (about $620 per person) recruitment sector in the US is choosing the most qualified applicants from a large pool for available positions. Recruiters' first responsibility is to sift through the voluminous resumes that arrive in their inboxes. This is a difficult procedure because of the volume and difficulty of finding relevant talents. Roughly 75% of resumes that are submitted frequently lack the required abilities. Although there are more than 50,000 online job boards, many of them use labor-intensive methods to match each CV to each job advertisement, which results in a high time complexity. This research presents a novel approach that uses Natural Language Processing for section-based segmentation and machine learning for dataset training. The web tool exclusively matches resumes with job positions where people exhibit interest, in an effort to maximize time efficiency. Only the results are visible.

Among the most sophisticated and popular NLP libraries, spacy is well-known for its ease of usage, quickness, and precision. Utilizing state-of-the-art methods for dependency parsing, entity recognition, and part-of-speech tagging, spaCy provides powerful tools for examining unstructured text data, including job descriptions and resumes. Furthermore, a basic metric for determining how similar two vectors are to each other in a multidimensional space is cosine similarity. By comparing candidate resumes with job descriptions, cosine similarity facilitates resume processing. This allows for the selection of the most suitable candidates based on how closely their skill sets match the job criteria.

II. RELATED WORK

Several studies have attempted to use artificial intelligence (AI) to improve the efficacy of hiring and performance reviews for employees. A study by Faliagka et al. investigated the use of AI to evaluate the authenticity of carefully written resumes, especially in the setting of online ranking that is consistent in e-recruitment platforms. The study explores AI's possible effects on worker performance in addition to assessing its adaptability in hiring.

H. Fu, Z. Niu, and J. Chen developed a novel two-step method for obtaining data from resumes. Their method identifies discrete resume blocks from unprocessed text by using a text classifier to enable efficient information extraction. This creative framework addresses a critical component of the hiring process by offering a fresh approach to enhance the processing and analysis of resume data.

In their work, Schmitt, Caillou, and Sebag aim to address the concerns of recruiters as well as candidates. The project's goal is to help seniors in college locate employment that fits their company preferences and skill sets. In order to improve the job and resume matching process and guarantee compatibility between candidates and company profiles, deep collaborative filtering is investigated as a technique.

Kessler et al. focused on creating a web application for efficient resume screening in job ads in a 2009 publication. The principal aim was to enhance the ranking of candidates during the job offer management procedure, hence facilitating a more optimized and effective recruitment process. An early attempt to incorporate technology into the employment process for better results is reflected in the report.

Moreover, Yu, Guan, and Zhou's contribution from 2005 used an incremental hybrid model to address the problem of information extraction continuity. A key element in the changing environment of AI-driven recruitment tools, the paper, which was presented at the 43rd annual conference of the Communication Association, offers insights into the ongoing difficulties and developments in information extraction. When taken as a whole, these studies show the various approaches taken to utilize AI and technology to optimize various aspects of the hiring process, such as information extraction, job matching, and honesty assessment.

III. METHODOLOGY

Resume parsing is the process of extracting relevant information from resumes or CVs to facilitate efficient hiring

processes. Using spaCy and cosine similarity for this task involves natural language processing (NLP) techniques and similarity measures. Here is a methodology for implementing resume parsing using spaCy and cosine similarity:

a) Data Collection:

The first step in implementing our resume parsing system involves the collection of a standardized dataset, typically in formats like DOCX, which is then meticulously annotated with key information labels such as names, contact details, education, work experience, and skills. Subsequently, data preprocessing is undertaken to convert resumes into a uniform text format, eliminating irrelevant information, special characters, and formatting inconsistencies. With spaCy, a powerful NLP library, installed and loaded, the system performs entity recognition, identifying crucial entities like names, organizations, and skills, extracting relevant details for each category. Following this, the textual data undergoes vectorization using techniques like TFIDF or word embeddings, creating numerical vectors that represent the content of each resume.

b) Cosine Similarity Calculation:

Cosine similarity is then calculated between each resume and a job description, producing scores that signify the degree of similarity. These scores are pivotal in the subsequent step of ranking and selection, where resumes are ordered based on their cosine similarity scores, and a threshold is set to filter out less relevant ones. The parsed and ranked resumes are seamlessly integrated into the hiring system, with relevant information stored in a structured format for easy retrieval and analysis. The system undergoes thorough evaluation using a validation dataset, and finetuning is conducted based on feedback to enhance entity recognition and vectorization accuracy.

$$\cos(\theta) = \frac{A \cdot B}{||A|| \times ||B||} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_t^2} \cdot \sqrt{\sum_{i=1}^{n} B_t^2}}$$

Where A_i and B_i are the *i*th components of vectors A and B respectively.

Finally, after rigorous evaluation and refinement, the resume parsing system is ready for deployment. It can be seamlessly integrated into the existing hiring platform or utilized as a standalone tool, presenting a robust solution for the efficient and objective screening of resumes. This comprehensive process, encompassing data collection, preprocessing, entity recognition, vectorization, similarity calculation, ranking, integration, evaluation, and deployment, ensures a holistic and sophisticated approach to modernizing traditional resume sorting, promising enhanced efficiency and accuracy in the recruitment process.

c) System Architecture

Our system architecture leverages spaCy for efficient recursive parsing and cosine matching for similar data. The process begins by using spaCy's ability to process messages to extract data without rebooting. Then, using cosine similarity, we compare the parsed function text with the description function to check the effect. This architecture optimizes the

sorting process, increasing the accuracy and speed of matching candidates to jobs. It also seamlessly connects to other tools and systems to make recruiting easier.
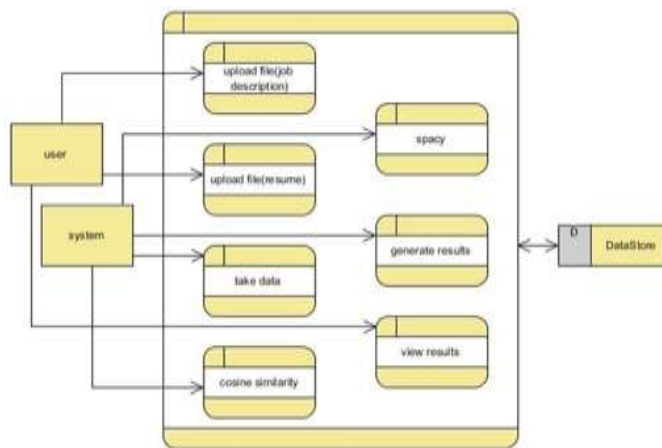

Fig. 1. System architecture of proposed system.

IV. RESULTS AND DISCUSSION

Recruiters find this technique straightforward to use as it rates resumes based on how well they match the job description, displaying a list of resumes in order of relevancy to the position. This would enable the recruiter to rapidly identify the CVs that most closely fit the job description by classifying the resumes in accordance with the job requirements. With the ability to review a large number of resumes quickly and with the right fit—something a human could not accomplish in almost real time—the method would help the recruiter shortlist profiles more quickly while also guaranteeing the authenticity of the shortlisting process. This would contribute to increasing the effectiveness and success of the hiring process in terms of choosing.

TABLE 1. Accurate results of different resumes.

Types of Resumes	Accuracy
Python developer	59.0078%
Java developer	67.8765%
Web developer	87.3456%
Cloud developer	90.13%
Database designer	45.876%

V. OUTPUT

VI. LIMITATIONS

There are acknowledged shortcomings with the current system. First off, there could be accuracy issues with the process of classifying data into entities. For this analysis to be effective, it is imperative that the segmentation be done precisely. Furthermore, if an applicant solely states Java competency yet the job description lists "Spring" as a Java extension, there may be a mismatch. Careful thought must be given to aligning these discrepancies during the data processing phase. Additionally, the system encounters a limitation concerning the retention of resumes as separate documents. Currently, there is no support for changing an existing resume document in the database to a newer version. This restriction makes it more difficult for the system to accommodate applicants who could submit revised resumes during the application process. To resolve this issue and keep the system up to date with the most recent candidate data, a way to seamlessly replace or append updated resume documents in the database will need to be implemented. These shortcomings draw attention to possible areas for system improvement, highlighting the necessity of precise data segmentation and adaptability when managing new candidate information in the database.

VII. CONCLUSION

This study emphasizes how employing automated machine learning to update the hiring process's conventional resume sorting could have a revolutionary effect. We created a state-ofthe-art automated system that uses machine learning to improve the hiring process' efficiency in order to address this problem. Companies can increase candidate selection accuracy and recruitment speed by smoothly incorporating this method into the hiring process. Including specialists is essential for improving the algorithm because of their insightful criticism, particularly HR specialists. Crucially, we can modify and tweak our approach to fit various industry needs, offering a unique solution. In In short, sorting through resumes takes a lot of time and effort, and managing the flood of job applications has become a major burden for businesses. The automated machine learning algorithm's ability to learn and improve over time is a crucial feature. The algorithm modifies its selection criteria as it reviews more resumes and gets input from hiring specialists and HR specialists. The algorithm becomes increasingly skilled at seeing patterns, comprehending industry-specific criteria, and correctly determining the appropriateness of candidates as a result of this iterative learning process.

VIII. FUTURE SCOPE

The future of biographical parsing using SpaCy and cosine similarity is promising with improved performance and accuracy. Leveraging natural language processing technologies such as SpaCy a long with cosine similarity for semantic analysis provides opportunities for improved candidates election, automatic skills extraction, and advanced resume ranking algorithms. Additional improvements may include integration with machine learning models to enable personalized recommendations, expansion to multi-language parsing, and site authentication optimization of high quality content. This technology has the potential to revolutionize the talent acquisition process , improving hiring and facilitating a better match between candidates and job opportunities.

IX. REFERENCES

[1] E Faliagka, L Iliadis, I Karydis, M Rigou, S Sioutas, A Tsakalidis, and G Tzimas, "On-line consistent ranking on e-recruitment: seeking the truth behind a well-formed CV," *The Artificial Intelligence Review*, 42(3), 515, 2014.

[2] J Chen, Z Niu, H Fu, "A Novel Knowledge Extraction Framework for Resumes Based on Text Classifier," *Proceedings of the International Conference on Web-Age Information Management*. Springer International Publishing, pp. 540-543, 2015.

[3] T Schmitt, P Caillou, M Sebag, "Matching Jobs and Resumes: A Deep Collaborative Filtering Task," *Proc. of the 2nd Global Conf. on Artificial Intelligence*, pp.1-14, 2016.

[4] R Kessler, N Béchet, JM Torres-Moreno, M Roche, and M. El-Bèze, "Job Offer Management: How Improve the Ranking of Candidates", in *Foundations of Intelligent Systems*, J. Rauch, et al., Editors. Springer Berlin Heidelberg, pp. 431-441, 2009.

[5] K Yu, G Guan, and M Zhou, "Resume information extraction with cascaded hybrid model." Proceedings *of the 43rd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, pp. 499–506, 2005.

[6] F Javed, Q Luo, M McNair, F Jacob, M. Zhao, and TS. Kang, "Carotene: A Job Title Classification System for the Online Recruitment Domain," *Proceedings of the IEEE First International Conference on Big Data Computing Service and Applications (Bigdata Service)*, pp. 286-293, 2015.

[7] S Clyde, J Zhang, and CC Yao, "An object-oriented implementation of an adaptive classification of job openings," *Proceedings of the 11th Conference on Artificial Intelligence for Applications, IEEE*, pp. 9-16, 1995.

[8] A. Zaroor, M. Maree, and M. Sabha, "A Hybrid Approach to Conceptual Classification and Ranking of Resumes and Their Corresponding Job Posts," In: Czarnowski I., Howlett R., Jain L. (eds) Intelligent Decision Technologies 2017. IDT 2017. Smart Innovation, Systems and Technologies, vol 72. Springer, Cham.