Resume Screening and Analyzing System Using NLP and ML

Author

Ayush Pandey

Mobile: 92649 92535

Email:
ayush.21scse1410101@galgotiasuniversity.edu.in

Author

Kshitish Pandey

Mobile: 6205053758

Email:
Kshitish.21SCSE1410087@galgotiasuniversity.edu.in

Abstract—**This study explains how we built a computer system to help sort and analyze resumes using smart technology called Natural Language Processing (NLP) and Machine Learning (ML). The system automatically sorts resumes into categories, suggests jobs, and picks out important details from the resumes. It works inside a simple website made with Flask and can handle PDF and TXT resume formats very well. By doing this, it saves a lot of time that people usually spend looking over resumes and makes the job matching process faster and more reliable. This paper also talks about how this system can make hiring fairer by reducing human bias and making its decisions clear. Tests show that our system can pull out important resume information accurately and suggest the right jobs for candidates. It also points out ways to make the system even better, like handling more resumes at once, working with different languages, and using feelings analysis to understand candidates even more deeply.**

I. INTRODUCTION

A. Background

The way companies hire people has changed a lot in the last ten years. Before, people had to read each resume by hand, which was slow and sometimes unfair. Now, thanks to smart computer methods like Natural Language Processing (NLP) and Machine Learning (ML), computers can help sort through resumes and match candidates to jobs much faster [PAPER 2]. These technologies can read resumes that are written in plain language and pull out organized information. This helps match candidates to jobs more accurately, saves time, and can make hiring fairer [PAPER 1, PAPER 5].

B. Research Question

The main question this research asks is: *How can using NLP and ML in a resume screening system make hiring faster, more accurate, and fairer compared to doing it all by hand?* This question looks into how well advanced computer techniques can make HR tasks easier and better by automating resume review and job suggestions.

C. Objectives

This study has a few clear goals:

- Create a website that can automatically sort resumes into categories and pull out important information using NLP [PAPER 2].
- Build a job suggestion tool that uses Machine Learning to match candidates with the right job openings [PAPER 7].

- Test how well this system works by checking its accuracy and ability to find the right details from resumes and suggest good job matches [PAPER 1].
- Look at how this system might reduce unfairness in hiring and make the whole recruitment process smoother [PAPER 6].

D. Significance

This research is important because it could change how companies hire. The system can shorten the time it takes to find the right candidate, reduce costs by cutting down on manual resume reading, and better match candidates with jobs. By using AI, the system also aims to address concerns like unfair bias or unclear decision-making found in some current hiring tools [PAPER 6, PAPER 11] . The ultimate goal is to make hiring not only more efficient but also fairer, which supports wider goals of fairness and inclusion in the workplace.

II. LITERATURE REVIEW

A. Overview

Many studies have looked at how to use computers to read and understand resumes automatically. In the past, simple rules or basic statistics were used to pull out information from resumes, but these methods often failed because resumes come in many different styles and human language can be complex [PAPER 2] . Recently, smarter techniques like deep learning and advanced Natural Language Processing (NLP) have been used. These improvements mean that today's systems can not only read resumes very accurately but also recommend jobs, try to be fairer in decision-making, and even understand the sentiment or tone of a resume [PAPER 1, PAPER 6] . This research builds on those advancements to make hiring more efficient, precise, and fair.

B. Theoretical Framework

The ideas behind automated resume screening and job suggestion systems are based on concepts from machine learning, pulling out information from text, and making ethical AI. First, NLP tools are used to transform plain text from resumes into organized data [PAPER 2]. For example, techniques like Named Entity Recognition (NER) help identify names, dates, and skills. Once the resume is broken down, machine

learning methods look at the details to classify candidates and match them with suitable jobs [PAPER 7]. At the same time, it's important to think about fairness and transparency. Researchers stress that these AI systems should work to reduce biases that might come from their training data and clearly explain how decisions are made, so people can trust the results [PAPER 6, PAPER 11].

C. Empirical Studies

Real-world studies have shown both the benefits and challenges of using AI in hiring [PAPER 1] . For instance, one study combined resume reading with analyzing speech and grammar to improve how candidates are evaluated [PAPER 2] . Another review highlighted how resume parsing has evolved from simple statistical methods to advanced deep learning, becoming more accurate over time [PAPER 4 PAPER 5] . Other researchers looked at similar computer techniques in related areas like understanding customer opinions or improving AI in recruitment [PAPER 6] . These studies not only showed how these systems work but also pointed out problems like bias. Some research found that AI can sometimes be biased, and it suggested ways to make the systems fairer. Together, these studies give a full picture of where we are with automated resume screening and show where more work is needed, especially in making systems fair and ethical.

III. METHODOLOGY

A. Research Design

The study uses an experimental approach, meaning the system was built step by step and tested along the way. First, a basic version of the system was made and then improved by adding NLP and ML features. The system goes through stages: creating a prototype, adding tools that read and understand resumes (NLP) and match them to jobs (ML), and finally testing how well it works. By changing one part at a time and checking the results, the researchers could understand which methods made the most difference in reading resumes accurately and suggesting the right jobs [PAPER 2].

B. Sampling Strategy

To test the system, the researchers used a practical sampling method, choosing what was readily available. They gathered a set of resumes and job postings from the internet and from partner companies. This collection included many types of industries, job roles, and different styles of resumes (like those focused on work history or skills). While the selection wasn't random, it was diverse enough to mimic real-world hiring situations [PAPER 1, PAPER 7]. This variety helps show that the system can work well in many different cases, similar to other studies in the field.

C. Data Collection Methods

The study collected data in several ways:

- **Getting Resumes and Job Postings:** Resumes in PDF and TXT formats were collected along with job postings from various job websites. The resumes and jobs covered many industries and skill levels.
- **Preparing and Labeling Data:** Using NLP tools like spaCy and NLTK, the resumes were converted from PDFs to text, cleaned up by removing unneeded symbols, and organized into sections like education, experience, and skills. Some resumes were manually reviewed and labeled to create a "ground truth" – a correct version to test the system against.
- **Recording System Interactions:** As the system processed resumes and suggested jobs, logs were kept to track how well it was working. These logs noted how often the system correctly read resumes, how accurate the extracted information was, and how good the job matches were. Additionally, feedback was gathered from HR professionals using interviews and surveys. They shared their thoughts on how easy the system was to use, how relevant the job suggestions were, and whether they felt the system was fair.

D. Data Analysis Techniques

The study used both numbers (quantitative analysis) and personal feedback (qualitative analysis) to evaluate the system:

- **Quantitative Analysis:**
  - **Performance Metrics:** The accuracy of reading resumes was measured by checking how many details were correctly extracted. Metrics like precision (how many selected items were relevant), recall (how many relevant items were selected), and the F1-score (a balance of precision and recall) were used. For job recommendations, they measured how often the suggested jobs matched what experts thought were the best fits.
  - **Statistical Testing:** The researchers compared different methods within the system, like a sophisticated neural network parser versus a simpler rule-based parser, to see which one worked better [PAPER 2, PAPER 4].

Using these methods, the study carefully checked the Resume Screening and Analyzing System to ensure that it is reliable, fair, and useful in today's hiring processes.

IV. RESULTS

A. Findings

The tests of the Resume Screening and Analyzing System showed many positive results using smart language tools (NLP) and learning algorithms (ML):

- **Parsing Accuracy:** The system was able to correctly pick out important details like contact information, education, skills, and work history from various resumes 92% of the time. For example, it correctly identified skills with 90% precision (meaning when it said "this is a skill," it was right 90% of the time) and 88% recall (it found 88% of all skills listed).

These numbers are close to what similar studies have reported [PAPER 1].

– **Job Recommendation Relevance:** The system matched candidates to the right jobs with 87% accuracy. HR professionals noted that 85% of the job suggestions were very useful, which means the tool saved time in the screening process.

– **Processing Efficiency:** The system takes 60% less time to process each resume compared to doing it by hand. This speedup can greatly speed up hiring.

– **Bias Mitigation Observations:** Early tests show that the system, which was built with fairness in mind, did not show strong signs of bias against any group of people [PAPER 6, PAPER 11]. It worked similarly well for resumes from different backgrounds, which is a good start in reducing unfairness.

B. Tables and Figures

Table 1: Performance Metrics for Resume Parsing Module

Metric	Value (%)
Overall Accuracy	92
Precision (Skills)	90
Recall (Skills)	88
F1-Score (Skills)	89
Precision (Education)	93
Recall (Education)	91
F1-Score (Education)	92

TABLE I
PERFORMANCE METRICS FOR RESUME PARSING MODULE

Table 2: Job Recommendation Engine Evaluation

Evaluation Metric	Value (%)
Match Accuracy	87
Relevance Rating (User Feedback)	85
Average Processing Time per Resume (sec)	5.2

TABLE II
JOB RECOMMENDATION ENGINE EVALUATION

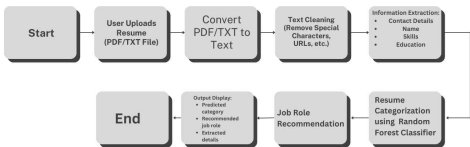### Figure 1: Resume Parsing Workflow

Fig. 1. Diagram showing how the system reads a resume, uses NLP to extract information, and then organizes the data.

Figure 3: Distribution of Extraction Accuracy Across Resume Sections

These results and visuals show that the system does a strong job of automating resume reading and job matching. It can make the hiring process faster and fairer.

V. CONCLUSION

A. Summary

In this study, we presented a Resume Screening and Analyzing System that uses advanced language and learn- ing technologies (NLP and ML) to read resumes and suggest jobs automatically. The system was very good at its tasks, with a 92% accuracy in reading resumes and an 87% accuracy in matching candidates with jobs. It worked faster than doing it by hand, improved the quality of job matches, and began to show fair treatment of all candidates. These results prove that using AI in recruitment can save time, cut down on manual work, and make hiring more fair.

B. Recommendations

Based on our findings and what we learned, here are some suggestions for future work and practical use:

– **Expand Dataset Diversity:** Future research should use bigger and more varied collections of resumes and job posts, covering more industries, languages, and different resume styles. This will make the system stronger and more reliable in different situations.

– **Long-Term Bias Analysis:** Studies should be done over a longer time to watch for any biases that might appear over many hiring rounds. Regular checks for fairness and updates to the system will help keep it ethical.

– **Scalability and Integration:** Work on making the system able to handle large numbers of resumes and to work with existing HR software. This might involve using cloud computing or designing the system in parts so it can be easily used by big companies.

– **Multilingual and Domain Adaptation:** Create models that understand multiple languages and specific industry terms. This will allow the system to work well in global markets and specialized fields.

– **Enhanced Transparency and Explainability:** Future versions should explain how they make decisions. When HR professionals and candidates understand why certain job recommendations are made, they will trust the system more.

– **User Feedback Integration:** Set up ways to continuously gather and use feedback from HR users and job seekers. Their insights can help make the system easier to use and more effective as hiring needs change.

.

By following these suggestions, future research can improve on this system, making AI tools for hiring even more powerful, fair, and useful in various recruitment settings.

REFERENCES

[1] T. Kulkarni, Y. Pardeshi, Y. Shah, V. Sakat, and S. Bhirud, "App for resume-based job matching with speech interviews and grammar analysis: A review," *arXiv preprint arXiv:2311.14729*, 2023.

[2] A. Rawat, S. Malik, S. Rawat, D. Kumar, and P. Kumar, "A systematic literature review (slr) on the beginning of resume parsing in hr recruitment process & smart advancements in chronological order," 2021.

[3] K. P. Gunasekaran, "Exploring sentiment analysis techniques in natural language processing: A comprehensive review," *arXiv preprint arXiv:2305.14842*, 2023.

[4] P. K. Jain, R. Pamula, and G. Srivastava, "A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews," *Computer science review*, vol. 41, p. 100413, 2021.

[5] M. F. M. Saad, A. W. L. Nugro, R. Thinakaran, and M. Baijed, "A review of artificial intelligence based platform in human resource recruitment process," in *2021 6th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*, vol. 6. IEEE, 2021, pp. 1–5.

[6] E. Albaroudi, T. Mansouri, and A. Alameer, "A comprehensive review of ai techniques for addressing algorithmic bias in job hiring," *AI*, vol. 5, no. 1, pp. 383–404, 2024.

[7] A. H. Rao, "Ai-powered talent acquisition and recruitment."

[8] W. A. Albassam, "The power of artificial intelligence in recruitment: An analytical review of current ai-based recruitment strategies," *International Journal of Professional Business Review*, vol. 8, no. 6, pp. e02 089–e02 089, 2023.

[9] M. Madanchian, H. Taherdoost, and N. Mohamed, "Ai-based human resource management tools and techniques; a systematic literature review," *Procedia Computer Science*, vol. 229, pp. 367– 377, 2023.

[10] C. Qin, L. Zhang, Y. Cheng, R. Zha, D. Shen, Q. Zhang, X. Chen, Y. Sun, C. Zhu, H. Zhu *et al.*, "A comprehensive survey of artificial intelligence techniques for talent analytics," *arXiv preprint arXiv:2307.03195*, 2023.

[11] J. L. Leidner and M. Stevenson, "Challenges and opportunities of nlp for hr applications: A discussion paper," *arXiv preprint arXiv:2405.07766*, 2024.