

Skill Extractor in Resumes using Machine Learning and Streamlit

Varun Sai Ganga
Dept. of Information Technology
Velagapudi Ramakrishna Siddhartha
Engineering College, Kanuru
Vijayawada, India,
gangavarunsai@gmail.com

Sri Navya Deepika Regalla
Dept. of Information Technology
Velagapudi Ramakrishna Siddhartha
Engineering College, Kanuru
Vijayawada, India,
navyaregalla007@gmail.com

S Kranthi
Dept. of Information Technology
Velagapudi Ramakrishna Siddhartha
Engineering College, Kanuru
Vijayawada, India,
kranthisri41@gmail.com

Lakshmi Lokeshwar Vissakodeti
Dept. of Information Technology
Velagapudi Ramakrishna Siddhartha
Engineering College, Kanuru
Vijayawada, India,
lokeshwarvk@gmail.com

Abstract— Finding a job can be difficult as the market is constantly changing and job searchers are never sure if their talents will fit the demands of the company or can they follow their passion. Users can receive tailored insights by submitting their resumes. These insights include skill recommendations and matching of important skills and experience against a large job database, and the discovery of positions that are suitable based on market trends and credentials. The Skill Extractor represents a dramatic change in career exploration with its market-driven skill ranking, tailored job recommendations, and broad format compatibility. It empowers people to confidently navigate the job market and make well-informed decisions about their professional journey. This system is compatible with several formats (PDF, Doc, etc.). Job searchers frequently feel lost, not knowing whether their skill set is relevant or which route would lead to their ideal position. The Skill Extractor empowers people to make knowledgeable decisions about their career path by giving them the resources they need to successfully navigate the job market. Because it may be used in a variety of formats, accessibility is guaranteed, meeting the demands of a wide range of job searchers, many of whom feel lost and unsure of the value of their experience and the best way to advance into the roles they want.

Keywords— NLP, resume analysis, skill recommendation, job matching, career guidance

I. INTRODUCTION

A more complex and automated solution is required due to increase in the job applications and the requirement for prompt and accurate assessments. But it might be difficult to write a CV that fairly highlights one's abilities, background, and credentials. Additionally, resumes that fail to pass the screening process of job interviews often result in job seekers being eliminated from consideration, regardless of their actual qualifications. One of the most challenging tasks of this type of job matching is that there is typically a bulk of information to coordinate against. Furthermore, these data usually submitted in a free form, as each individual has their own preference to prepare the data[2]. This interface of proposed method and also offers benefits for job seekers, as they can receive recommendations based on the job role predicted. By utilizing cutting-edge technologies like ML and NLP

techniques, it surpasses the conventional resume assessments. With Skill Extractor, job seekers can ensure that their resumes are optimised for the positions they are applying for, increasing their chances of landing their dream job. Thus, the Skill Extractor interface of proposed method has the potential to revolutionize the job application process, making it more efficient and effective for job seekers[3]. Job seekers can increase their chances of finding their ideal job by optimizing their resumes for the job roles they are suitable. The features including as evaluating a candidate's soft skills, accounting for changing company expectations as market trends and requirements change, and analysing the veracity of a candidate's resume, cannot yet be automated [8]. Hiring was made easy and help of cutting-edge technology of the Skill Extractor. By using Skill Extractor to tailor their resumes to the positions they are looking for, job seekers can find their suitable job position analysis using their resume data.

II. REVIEW OF LITERATURE

In [1] Bhushan Kinge and team suggested a technique for detecting, identifying, and classifying different resumes using machine learning and neural network models, such as SVM, KNN, Word2Vec, Cosine similarity, etc. These techniques take a student's resume as input and extract information about the student's qualifications, skills, and links to their GitHub and LinkedIn profiles, among other things.

In [2] Rodriguez and team suggested a system that uses a clustering algorithm to match job searchers' profiles to the specifications of positions that potential employers had posted.

In [3] Sruthi and team done streamlines resume uploads (PDF/Word) using a Streamlit-based approach. Pyresparser displays data after parsing resumes. Positions are suggested by skill analysis, and learning for gaps is advised. Resumes are evaluated for skill-based assessments by NLP and Spacy.

In [5] Smiti Singhal and team concentrates on taking information out of resumes and analyzing it in order to make it into something recruiters can use.

In [7] Vamsi and team suggested an online interface to filter resumes and determine the kind of job that corresponds with the resume; MLP works better than other methods like KNN, SVM, and Logistic Regression

In [8] Tijare and team designed a resume screening algorithm with the ability to extract important data from CVs and resumes. Their goal is to streamline and expedite the hiring process for businesses and individuals alike.

In [9] Nirmithi and team suggested a combined method for resume parsing; the knowledge bases for talents, employers, and designations were created and they used a deep learning model to reach an accuracy of 90–92%.

In [10] Khan and team suggested system will recruit a company in processing several resumes at once by extracting the standardized and condensed resume from the database and providing the result examination.

III. PROPOSED METHODOLOGY

To extract proper information from resumes, we have employed Keyword Matching Algorithm, which finds key terms in resumes related to job needs, skills, and qualifications, and Natural Language Processing (NLP), which recognizes and classifies entities like names, skills, and job titles. Two different approaches are used by our system to analyze resumes efficiently. To ensure accurate extraction, the Keyword Matching Algorithm finds essential terms associated with job requirements, abilities, and certifications. Natural Language Processing (NLP) simultaneously contributes to a more detailed knowledge of individuals' professional backgrounds by recognizing and classifying items such as names, talents, and job titles. By combining these two methods, resume parsing becomes more accurate and thorough, facilitating the focused extraction and contextual interpretation of crucial resume content.

In addition to an applicant's skills, the adapted ranking algorithm considers their experience, talents, and contextual fit. This all-encompassing strategy acknowledges that a good match entails more than just technical skills; it also entails learning about the candidate's background, special skills, and degree to which they mesh with the hiring organization's requirements and culture.

A. Steps Involved

1) Collect Information

The first stage of the recruiting procedure is gathering resumes from several sources because job seekers frequently send their resumes via email attachments, online job portals, or other recruitment channels.

Gather resumes and take textual data out of them. Job seekers frequently send resumes through email attachments, online job portals, or other recruiting channels. Recruiters are able to get a wide range of resumes from candidates thanks to this multipronged strategy, which guarantees a thorough picture of the talent that is out there.

2) Handle the Information Gathered

By dividing the text data into smaller pieces like words or phrases, you can tokenize it. To find and classify entities in resumes, such as names, companies, abilities, and locations, use Named Entity Recognition

(NER). This aids in comprehending the background and obtaining important data. Tokenization and Named Entity Recognition are the methods used to extracting important details from resumes, including a person's history, experiences, and abilities. When used for resume analysis, NER is extremely helpful in sorting through large volumes of text to identify particular entities. It can correctly identify, for example, a person's identity, the businesses they have worked for, their skills, and even the places specified in the document. This procedure not only helps to comprehend a candidate's professional path, but it also makes it easier to extract important information that is necessary to make wise hiring selections.

3) Classify using NLP and Keyword Matching Algorithms

Use NLP methods to examine the data that has been tokenized and enriched by NER. Utilize keyword matching algorithms to find pertinent terms and phrases associated with job specifications.

It is possible to develop an extensive technique for examining and extracting relevant data from job specifications by integrating matching techniques, tokenization, NER, and keyword extraction. Improved decision-making about job matching and recruitment is made possible by this methodology, which increases the processing efficiency of enormous amounts of textual data.

This procedure is further refined by keyword matching algorithms, which find relevant terms and phrases related to job demands. These algorithms align the extracted items and keywords with the particular requirements listed in job descriptions by utilizing sophisticated pattern recognition techniques. Keyword matching is integrated into the Skill Extractor to guarantee that recommendations are closely aligned with potential employers' expectations, while also streamlining the discovery of relevant skills and experiences.

4) Determine the Suitable Job Roles

Use the enhanced information obtained by tokenization and NER to help you decide on the best job responsibilities. Candidates are matched with specific employment vacancies based on contextual information, skills, and recognized entities. A comprehensive approach is developed to match applicants with job openings by leveraging enriched data obtained via Named Entity Recognition (NER) and tokenization. While skill-matching algorithms assess applicant ability, semantic analysis highlights key tasks. With the help of dynamic matching that adapts to shifting job requirements, a personalized ranking system takes experience, talents, and contextual fit into account.

Incorporating ethical considerations helps to reduce biases and promotes a data-driven, effective, and inclusive hiring process. According to the recognized entities and skills acquired by NER, skill-matching algorithms are essential in assessing an applicant's talents. This procedure makes sure that candidates are

precisely matched with the nuanced skill sets and experiences specified in the job specifications, rather than just matching them based on general job requirements. The Skill Extractor provides precise evaluations of candidates' fit for particular positions by utilizing the comprehensive information gleaned from resumes.

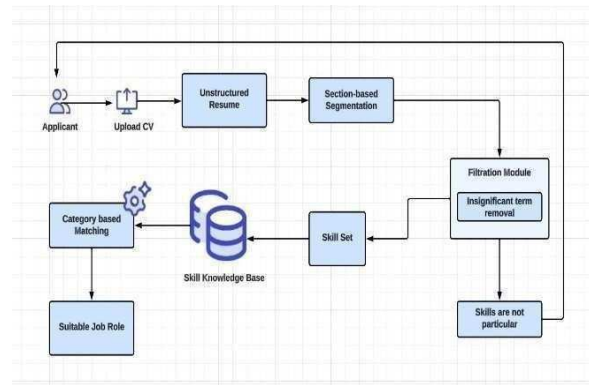


Fig. 1. System Architecture

Applicants start the workflow by uploading their resumes in the recruitment process shown in the block diagram. The CVs are then filtered using a specific module that uses a "category-based removal" procedure. Resumes that don't match certain requirements, including the job title or industry, are eliminated by this process. After filtering, the remaining resumes move on to a matching module using a "skill set" matching method. This module methodically matches each applicant's qualifications with the precise specifications listed in the job description. The hiring manager receives the applicant's résumé if their profile shows them to be a strong fit for the position. The resume is labelled as "not suitable" if, on the other hand, the applicant's skills do not match the job requirements. Both section-based and unstructured resume formats are accepted in the hiring process. A section-based resume is structured with sections like a summary, work experience, and education, but an unstructured resume has no set format. The procedure has a feedback loop that runs from the hiring manager to the filtration module, which is an important feature. With time, a more responsive and sophisticated filtering process will be made possible by the hiring manager's ability to express preferences and share information about the kinds of resumes they are looking for through this feedback mechanism. The filtration criteria are continuously optimized with the help of this iterative feedback loop.

B. Algorithms

1) Keyword Matching Algorithm

Redefining the job search process is mostly due to the keyword matching technology built within the Skill Extractor. When applicants submit their resumes, our advanced algorithm carefully reads through the text to find and extract important terms that represent certain abilities, credentials, and experiences. A thorough

evaluation of these keywords' applicability to the demands of the market is subsequently made possible by an advanced matching method that compares them to a sizable job database. The method improves the accuracy of skill and career recommendations by combining contextual and contextual correlations with explicit keyword matching. With this method, job seekers are equipped with customized insights that guarantee their special skills match the varied and continually shifting needs of businesses. A key component of the Skill Extractor, the keyword matching algorithm helps to refine the job search process and gives users a tactical edge in navigating constantly shifting job market.

Steps to Implement:

1. Make two lists: one for keywords and the other for keywords that match.
2. Tokenize the resume text into individual words or phrases.
3. To divide the text into a list of words, utilize Python's `split()` function. Go over the list of keywords again.
4. Verify whether each keyword appears in the list of tokenized terms from the resume text.
5. Print the keywords that matched.

2) Tokenization

This tokenization-driven strategy improves the system's overall performance and provides users with a thorough and in-depth review of their resumes. Tokenization essentially serves as a key method for empowering job searchers by making sure that their information is properly interpreted and utilized by the Skill Extractor to deliver customized insights in the competitive employment opportunity marketplace. The system is able to classify and evaluate each piece of information independently thanks to this methodical division, which also makes the textual data less difficult. The algorithm obtains a thorough grasp of the user's professional profile by dissecting the information into these fine-grained components. With the content divided into these easily understood sections, the Skill Extractor is able to obtain a more detailed picture of the user's professional profile, which facilitates more precise matching of the user's skills to job requirements.

Steps to implement:

1. Import the `nltk` library.
2. Define the text you want to tokenize.
3. The text can be divided into list of terms using the `nltk.word_tokenize()` method.
4. Print the text that has been tokenized.

3) Named Entity Recognition

One of the primary features of the Skill Extractor is Name Entity Recognition (NER), which transforms the way resumes are viewed and utilized for career exploration. The NER capability carefully recognizes and classifies textual elements, including names, companies, locations, dates, and particular abilities, when users submit their resumes. This simplifies the

process of extracting important data and offers an organized comprehension of the user's work history. Through entity discrimination, the system learns important details about specialized talents, school backgrounds, and job experiences, which helps to provide more precise and individualized skill recommendations. NER goes beyond simple keyword recognition by enumerating terms' contextual relevance and cultivating a sophisticated comprehension of the user's domain knowledge.

As a result of the Skill Extractor's integration of NER, users may now navigate the job market with greater accuracy and sophistication thanks to an in-depth and contextually rich interpretation of their professional story.

Steps to implement:

1. Create a variable with the text message you want to analyze.
2. Create a list of entities.
3. Tokenize the message into individual words.
4. Verify if every word appears in the list of predefined entities.
5. Print the entities that are matched.

IV. RESULTS AND ANALYSIS

The Streamlit app for talent extraction and recommendation boosts user engagement by providing a dynamic and interactive platform for skill assessment. It also delivers individualized suggestions based on current market trends, assisting users in aligning their abilities with industry expectations. The incorporation of job role recommendations enhances the usefulness of the tool by directing users toward appropriate career pathways according to their skills. The app's ability to give information fast and precisely is one of its main advantages. The app's speed and accuracy combine to make it a dependable resource for skill evaluation, career counseling, and recommendations—all of which improve user experience and usefulness.



Fig. 2. Uploading Resume

To upload a resume, click the "Upload Resume" button or choose "Browse". Find the computer folder where your resume is stored, select the file, and press "Open."

After the file has been selected, it can begin uploading automatically, and you might see a progress bar showing the upload status.

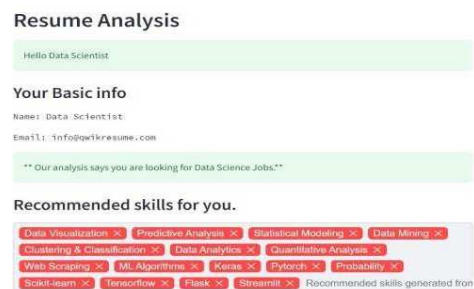


Fig. 3. Recommended Skills

Using advanced algorithms, the "Skill Extractor" offers customers specific skill recommendations based on their uploaded resumes. Using the NLTK library's capabilities, the program finds important terms and patterns that indicate the user's area of expertise. With the help of the NLTK library, the application finds key terms and patterns that correspond to the user's field of expertise. By streamlining the resume analysis process and providing individualized insights, this cutting-edge approach guarantees a more accurate and effective match between candidates' talents and job needs.

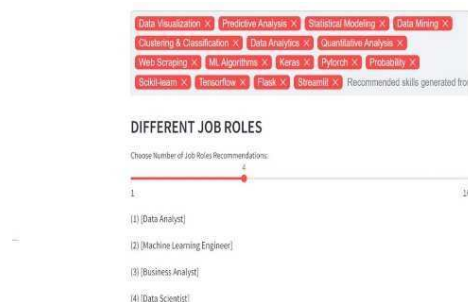


Fig. 4. Suitable Job Roles

Users receive strategic advice on career pathways that effectively leverage their abilities through a methodical comparison of the extracted talents with preset criteria for a range of professional roles. The technology helps users by offering strategic career guidance by methodically comparing the skills collected from resumes with predetermined criteria across a range of professional roles. Users can obtain insightful knowledge on the best job paths that best utilize their special talents thanks to this customized study. The technology enables people to make well-informed decisions by matching the extracted abilities with a wide range of criteria for various professions, which promotes a more intentional and focused approach to career development.

V. CONCLUSION

The Skill Extractor skillfully evaluates resumes that have been uploaded and provides customized skill recommendations that are in line with emerging technologies. In addition, it streamlines the job search process by offering tailored recommendations for suitable career roles. Frequent updates ensure that the tool remains relevant and effective in the ever-changing employment market by keeping it in line with industry trends. With its skillful assessment of submitted resumes, the Skill Extractor provides tailored skill recommendations that are in line with emerging technology. Beyond its ability to evaluate skills, the application enhances the job search process by providing personalized recommendations for appropriate career possibilities. Through segmenting resumes into manageable chunks and recognizing identified entities such as experiences, education, and talents, the analyzer is able to obtain a thorough grasp of each applicant's professional profile. It provides recommendations based on skills. In addition to helping people assess their skills, the Skill Extractor proves to be a useful tool for matching people with positions that suit their experience in the quickly evolving employment sector. By providing customized career position recommendations, the tool helps users get the most out of their job search experience. Through an awareness of each person's distinct set of abilities, experiences, and educational background, the analyzer recommends roles that fit the user's professional profile and provides appropriate career pathways. Job searchers benefit from this individualized strategy in two ways: first, it saves time; second, it increases the probability of discovering positions that not only fit the candidate on paper but also complement their career goals.

With its cutting-edge approach to keeping users informed, competitive, and successful in their career initiatives, the Skill Extractor continues to lead the way as the employment market changes. It is a dependable ally for people navigating the complexities of the contemporary job market because of its dedication to providing frequent updates and remaining up to speed with industry trends.

REFERENCES

- [1] Bhushan Kinge, Shrinivas Mandhare, Pranali Chavan, S. M. Chaware, "Resume Screening using Machine Learning and NLP: A proposed system", *International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT)*, ISSN : 2456-3307, Volume 8 Issue 2, pp. 253-258, March-April 2022.
- [2] Rodriguez, Leah G., and Enrico P. Chavez. "Feature selection for job matching application using profile matching model." 2019 IEEE 4th International Conference on computer and communication systems (ICCCS). IEEE, 2019.
- [3] Sruthi, Patlolla, et al. "Skill Extractor: A Case Study using RNN-based Keyword Extraction." *E3S Web of conference* . Vol. 430. EDP Sciences, 2023.
- [4] Mittal, Vrinda, et al. "Methodology for resume parsing and job domain prediction." *Journal of Statistics and Management Systems* 23.7 (2020): 1265-1274.
- [5] Sharma, Anushka, Smiti Singhal, and Dhara Ajudia. "Intelligent Recruitment System Using NLP." 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV). IEEE, 2021.
- [6] Green, Nathan, Michelle Liu, and Diane Murphy. "Using an Electronic Resume Analyzer Portal (eRAP) to Improve College Graduates Employability." *Information Systems Education Journal* 18.3 (2020): 28-37.
- [7] Vamsi Naidu, P., et al. "Resume Screening Using Machine Learning." *Second International Conference on Image Processing and Capsule Network ICIPCN 2021* 2. Springer International Publishing, 2022.
- [8] Tijare, Poonam, et al. "a smart resume screening tool for customized shortlisting." *ITM Web of Conferences*. Vol. 56. EDP Sciences, 2023.
- [9] Bhoir, Nirmiti, et al. "Resume Parser using hybrid approach to enhance the efficiency of Automated Recruitment Processes." (2023).
- [10] Khan, Er Farzana, et al. "Resume Parser and Summarizer."
- [11] Das, Papiya, Manjusha Pandey, and Siddharth Swarup Rautaray. "A CV parser model using entity extraction process and big data tools." *International Journal of Information Technology and Computer Science* 9.2 (2018): 21-31.
- [12] Daryani, Chirag, et al. "An automated resume screening system using natural language processing and similarity." *ETHICS AND INFORMATION TECHNOLOGY* [Internet]. VOLKSON PRESS (2020): 99-103.
- [13] "NLP Based Resume Parser Analysis", *International Journal of Emerging Technologies and Innovative Research* (www.jetir.org | UGC and issn Approved), ISSN:2349-5162, Vol.10, Issue 5, page no. pp219-221, May-2023.
- [14] Bhaliya, Nirali, Jay Gandhi, and Dheeraj Kumar Singh. "NLP based extraction of relevant resume using machine learning." (2020).
- [15] Sanyal Satyaki, et al. "Resume parser with natural language processing." *International Journal of Engineering Science* 4484 (2017).