

Career Lens: An ML-Based Personalized Career Guidance System for Resume Skill Gap Analysis and Learning Path Recommendation

Mrs. S. Sandhya¹, K. Harini², P. Saanvi³, K. Sai Charitha⁴

Associate Professor, CSE, GNITS, Hyderabad, India

Student, CSE, GNITS, Hyderabad, India

Student, CSE, GNITS, Hyderabad, India

Student, CSE, GNITS, Hyderabad, India

Abstract - Industries are being reshaped continuously by new technologies, and the skill sets that employers expect from candidates change just as quickly. For students, recent graduates, and other job seekers, this creates a genuine problem: it is often unclear whether the skills they already have line up with what a particular career path actually demands. Career counselling has traditionally relied on manual evaluation, broad generic advice, or resume screening built around keyword matching approaches that rarely reveal exactly where an individual's skill gaps lie.

This paper presents Career Lens, a machine-learning-based system built to give job seekers a personalized view of their career readiness. The system reads a candidate's resume, pulls out relevant skills using Natural Language Processing (NLP), and checks them against the skill requirements tied to a specific job role. Rather than producing a single match-or-no-match verdict, Career Lens scores the similarity between what the candidate already knows and what the role requires, and sorts each skill into one of three bands Strong, Moderate, or Missing so users can see exactly where they stand.

Once the gaps are identified, Career Lens goes a step further by recommending certification courses, learning resources, and hands-on project ideas tailored to those specific gaps. The overall goal is to make career planning explainable and centred on the individual rather than generic, connecting a candidate's actual skill profile to what industry is really asking for. In doing so, the proposed framework is intended to help students and job seekers make better-informed decisions about which skills to build next, narrowing the gap between what is taught in the classroom and what employers expect on day one.

Keywords - Machine Learning, Natural Language Processing, Resume Analysis, Skill Extraction, Skill Gap Analysis, Career Guidance System, Recommendation System, Personalized Learning

1. INTRODUCTION

Today's job market rarely stays still. Organizations keep revising what they expect from candidates as technology advances and business needs shift, and this steady stream of new, specialized roles means candidates have to keep evaluating and updating their own knowledge just to stay relevant. Many students and job seekers, however, simply do not have a clear picture of what skills their target roles actually require.

One of the biggest obstacles candidates run into is figuring out the gap between what they can already do and what industry expects. Job portals and professional networking sites do offer job recommendations, but these tend to focus on matching a profile to open roles rather than pointing out what to improve. Traditional career counselling, meanwhile, depends on human expertise, which makes it hard to scale or personalize for every individual.

Most existing resume-analysis systems are built for recruitment screening applicants and ranking candidates rather than for helping candidates improve. They typically lean on keyword matching, which struggles to capture the semantic relationships between skills. Two different tools or technologies can represent essentially the same competency, but a simple keyword search often misses that connection entirely.

Career Lens was built to address these gaps by combining Machine Learning and NLP into a single career-guidance framework. It takes a resume and a target job role as input, extracts the meaningful technical information from the resume, and compares it against the role-specific requirements stored in a skill knowledge base.

Rather than simply telling a candidate whether they match a role, the proposed approach digs into individual skills and sorts them into Strong, Moderate, and Missing categories, giving users a clear sense of where their strengths lie and where they still need work.

Beyond identifying the gaps, Career Lens builds personalized learning pathways recommending certification courses, technical resources, and project ideas suited to each gap so users have a structured way to close them rather than guessing at what to learn next. The central aim of this work is an explainable, personalized career-support system that helps people make data-driven decisions about their own skill development. By connecting individual skill profiles to industry expectations, the system is meant to help narrow the mismatch between what students learn academically and what employers actually require.

2. LITERATURE REVIEW

Skill Identification from Online Job Advertisements

A number of studies have looked at extracting skills directly from online job postings using NLP techniques. This line of work is mainly concerned with identifying which skills are most frequently demanded and building a picture of industry requirements useful for constructing skill knowledge bases, though it looks at employer-side data rather than individual candidates.

Resume Analysis and Classification Systems

On the candidate side, machine-learning-based resume analysis systems have been built to automate screening and classification, generally relying on resume parsing, feature extraction, and classifiers such as Support Vector Machines, Naive Bayes, and Logistic Regression. The focus of most of this work, though, is recruitment decisions rather than helping candidates understand or improve their own skill sets.

2.1 Comparison of Existing Approaches

System	Approach	Advantages	Limitations
Resume Screening Systems	Keyword matching and classification	Automated resume filtering	No detailed skill gap analysis
Job Recommendation Platforms	Profile-based recommendation	Suggests suitable jobs	No personalized improvement path

System	Approach	Advantages	Limitations
Career Counselling Systems	Manual assessment	Expert guidance	Time consuming and less scalable
Online Learning Platforms	Course recommendation	Provides learning resources	Not connected with candidate skills
Career Lens	NLP + ML skill analysis	Personalized career guidance	Requires continuous dataset updates

Job Recommendation Systems

Job recommendation systems are widely used across employment platforms, typically relying on collaborative filtering or content-based techniques to suggest suitable openings based on a user's profile. These systems do improve job discovery, but they generally stop short of analysing missing competencies or recommending a structured path for learning them.

NLP-Based Skill Extraction Approaches

NLP methods tokenization, named entity recognition, keyword extraction, semantic similarity have all been applied to pull meaningful information out of unstructured resume text. These techniques go beyond simple keyword matching and enable more intelligent comparison between a candidate's skills and a role's requirements.

Personalized Learning Recommendation Systems

Separately, research on personalized learning systems has looked at recommending educational resources based on a user's interests and abilities. Much of this work, however, operates independently of resume analysis and job-role requirements, and building a single system that connects skill assessment directly to learning recommendations remains an open problem.

3. RESEARCH GAP

Looking across this body of work, a few gaps stand out:

- Existing resume analysis systems are built primarily for recruitment and candidate ranking, not for helping users improve their own skills.

- Most job recommendation systems suggest suitable roles but do not explain why a candidate is or is not a good fit.
- Keyword-matching approaches fail to capture the semantic relationships between skills and job requirements.
- Learning-recommendation platforms tend to suggest courses without factoring in a user's existing competencies or career objectives.
- There is a clear need for one integrated framework that ties together resume analysis, skill gap identification, and personalized learning recommendations.

3.1 Proposed Solution Over Existing Limitations

Career Lens is designed to close these gaps by offering an end-to-end career guidance pipeline. Where most existing systems stop at resume screening or job matching, Career Lens goes further comparing candidate skills directly against specific job requirements, sorting skill proficiency into clear categories, and generating actionable recommendations for whatever is missing.

4. EXISTING SYSTEM VS PROPOSED SYSTEM

Feature	Existing Systems	Career Lens
Resume Understanding	Keyword-based analysis	NLP-based skill extraction
Skill Gap Detection	Mostly unavailable	Strong, Moderate and Missing classification
Role Specific Analysis	Limited	Target job role-based comparison
Recommendations	Generic suggestions	Personalized courses and projects
Explainability	Limited	Provides skill improvement insights

5. PROPOSED SYSTEM METHODOLOGY

Career Lens combines Machine Learning and NLP techniques to analyse candidate profiles and generate personalized skill-improvement recommendations. It takes a resume and a target job role as input, pulls out the relevant technical skills from the resume, and compares them with what the selected role requires. From that comparison, the

system identifies skill gaps and builds a customized learning pathway.

The overall methodology unfolds across several stages resume preprocessing, skill extraction, job-role requirement analysis, skill similarity computation, skill gap classification, and recommendation generation each contributing to an explainable system that helps users understand both their strengths and the areas where they still need to grow.

5.1 Resume Processing Module

The pipeline begins with the uploaded resume. Since resumes typically arrive as unstructured documents most often PDFs the system first extracts the raw text and converts it into a structured representation. This extracted content then goes through preprocessing: cleaning the text, stripping out unnecessary symbols, and normalizing the formatting. This step matters because it reduces noise and keeps the resume content consistent with how job-requirement descriptions are represented, which in turn improves the quality of everything downstream.

5.2 NLP-Based Skill Extraction

With clean text in hand, NLP techniques are used to pull out the meaningful skills programming languages, frameworks, tools, and other domain-specific competencies from the resume content. This happens in a few steps:

- Tokenizing the resume text into meaningful units.
- Normalizing and preprocessing that text further.
- Identifying technical skills through skill-matching techniques.
- Mapping the extracted skills into standardized categories.

5.3 Job Role Requirement Analysis

Each target job role is analysed against a predefined skill-requirement dataset, which lists the expected skills, relevant certifications, and learning resources for that role. This dataset effectively serves as the reference knowledge base against which a candidate's current skill profile is evaluated.

5.4 Weighted Skill Gap Analysis

At the core of Career Lens is its weighted skill gap analysis. Extracted resume skills are compared against job-specific requirements through similarity-based analysis, with each skill given a relevance score based on how closely it matches what is expected.

Cosine similarity is used to measure this relationship between candidate skills and required skills:

$$\text{Similarity}(A,B) = (A \cdot B) / (|A| |B|)$$

Based on the resulting similarity score, each skill falls into one of three bands:

Similarity Score	Skill Category
75% - 100%	Strong Skill
50% - 75%	Moderate Skill
Below 50%	Missing Skill

5.5 Recommendation Generation Module

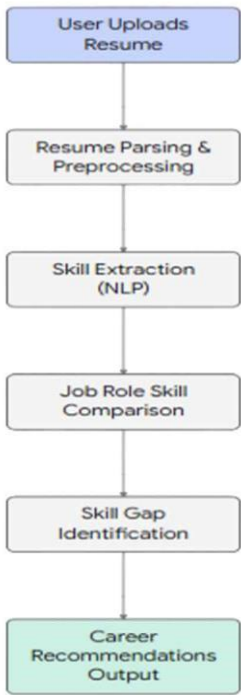
Once missing and moderate skills have been identified, the recommendation module takes over, mapping each gap to relevant courses, certifications, and hands-on projects. This gives users a structured improvement path rather than leaving them to search for learning resources on their own.

For example, a candidate applying for a Machine Learning Engineer role who lacks Deep Learning skills would be pointed toward relevant deep learning courses along with project ideas such as image classification or other neural-network-based applications.

6. SYSTEM ARCHITECTURE

Career Lens is organized into frontend, backend, NLP processing, machine learning, database, and recommendation layers. The frontend handles user interaction resume upload and result visualization while the backend coordinates communication between the other modules.

System Architecture Flow:



7. ALGORITHM

Algorithm: Personalized Skill Gap Analysis

Input: Resume R and Target Job Role J

Step 1: Extract text from the uploaded resume.

Step 2: Perform NLP preprocessing on the extracted text.

Step 3: Identify candidate skills from the resume content.

Step 4: Retrieve the required skills for the selected job role.

Step 5: Calculate similarity between candidate skills and required skills.

Step 6: Classify skills into Strong, Moderate, and Missing categories.

Step 7: Generate personalized course, certification, and project recommendations.

Output: Skill gap report and personalized learning path.

8. TECHNOLOGY STACK

Component	Technology
Programming Language	Python
Backend Framework	Flask
Frontend	HTML, CSS, JavaScript
Machine Learning	Scikit-learn
NLP Libraries	NLTK / spaCy
Data Processing	Pandas, NumPy
Database	MySQL / MongoDB
Development Tools	VS Code, Git

9. DATASET DESCRIPTION

How well a career guidance system performs depends heavily on the quality of the skill and job-role data it is comparing against. Career Lens relies on a structured knowledge base containing job roles, the technical skills tied to each one, how important each skill is, and recommended learning resources this dataset is what allows the system to spot missing competencies and generate improvement suggestions that make sense.

The dataset spans multiple job categories Software Developer, Data Scientist, Machine Learning Engineer, Web Developer, and other technical roles each mapped to a set of required skills covering programming languages,

frameworks, databases, tools, and domain-specific technologies. It also includes mappings from individual skills to relevant courses, certifications, and project recommendations.

Attribute	Description	Example
Job Role	Target career position	Machine Learning Engineer
Required Skills	Skills expected for role	Python, TensorFlow, SQL
Skill Weight	Importance of skill	High/Medium/Low
Learning Resources	Recommended improvement resources	Courses and Projects

10. EXPERIMENTAL SETUP

The system was tested by feeding it resumes with different technical backgrounds and pairing each one with multiple target job roles. For each resume, the system extracts the available skills, compares them against role-specific requirements, and produces a skill analysis report.

The experimental setup itself is built on Python-based machine learning libraries alongside standard NLP tools, using resume parsing, skill-matching algorithms, and recommendation mapping to evaluate how well the personalized guidance actually works.

10.1 Evaluation Metrics

Performance of the skill extraction and recommendation modules can be assessed using standard machine learning metrics:

- Accuracy: measures the correctness of skill identification.
- Precision: evaluates the relevance of the extracted skills.
- Recall: measures the ability to identify required skills.
- F1-score: provides a balance between precision and recall.

11. RESULTS AND DISCUSSION

In practice, Career Lens successfully analyses resumes and produces role-specific skill insights identified skills, missing skills, proficiency categories, and recommended learning paths. Sorting skills into Strong, Moderate, and Missing gives users a concrete sense of where they stand in terms of career readiness, and because the recommendations

are tied to the specific job role selected, they are noticeably more useful than the generic suggestions typical of older systems.

The recommendation module adds further value by turning identified gaps into an actionable learning plan, so users can work on specific certifications and projects that map directly to their career goals rather than picking courses at random.

11.1 Output Analysis

- Resume-based skill extraction gives automated, consistent candidate profiling.
- Role-specific comparison surfaces the actual technical competency gaps.
- Personalized recommendations support a structured approach to skill improvement.
- Because the analysis is explainable, users understand why a given recommendation was made rather than just receiving it.

12. ADVANTAGES AND APPLICATIONS

Advantages:

- Delivers personalized career guidance rather than one-size-fits-all advice.
- Cuts down on the need for manual skill assessment.
- Raises awareness of exactly which skills industry is looking for.
- Encourages continuous learning and ongoing professional development.

Applications:

- Students preparing for campus placements.
- Fresh graduates mapping out their career paths.
- Academic institutions offering career support services.

13. CONCLUSION AND FUTURE SCOPE

Career Lens is an ML-based, personalized career guidance framework aimed at closing the gap between a candidate's actual skills and what industry expects. By bringing together NLP-based resume analysis, skill gap identification, and recommendation generation in one system, it offers a more intelligent approach to career planning and skill development than the generic tools currently available.

Future work could extend the system by integrating real-time job market APIs, applying more advanced language models for deeper resume understanding, building adaptive recommendations that respond to user feedback over time, and expanding the framework to cover additional

professional domains beyond the technical roles it currently supports.

REFERENCES

- [1] J. Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," 2019.
- [2] T. Joachims, "Text Categorization with Support Vector Machines," Machine Learning: ECML, 1998.
- [3] C. Manning, P. Raghavan, and H. Schütze, "Introduction to Information Retrieval," Cambridge University Press.
- [4] F. Sebastiani, "Machine Learning in Automated Text Categorization," ACM Computing Surveys, 2002.
- [5] Research articles related to NLP-based resume analysis and recommendation systems.
- [6] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proc. NAACL-HLT, pp. 4171–4186, 2019.
- [7] T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," Proc. European Conference on Machine Learning (ECML), pp. 137–142, 1998.
- [8] C. D. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval. Cambridge, U.K.: Cambridge University Press, 2008.
- [9] F. Sebastiani, "Machine Learning in Automated Text Categorization," ACM Computing Surveys, vol. 34, no. 1, pp. 1–47, 2002.
- [10] A. Malherbe, S. Aufaure, and M. Hacid, "A Survey on Skill Identification From Online Job Advertisements," IEEE Access, vol. 9, pp. 118134–118153, 2021.
- [11] S. Ashrafi, B. Majidi, E. Akhtarkavan, and S. H. Razavi Hajiagha, "Efficient Resume-Based Re-Education for Career Recommendation in Rapidly Evolving Job Markets," IEEE Access, vol. 11, pp. 124350–124367, 2023.
- [12] J. Leskovec, A. Rajaraman, and J. D. Ullman, Mining of Massive Datasets, 3rd ed. Cambridge, U.K.: Cambridge University Press, 2020.
- [13] C. C. Aggarwal, Machine Learning for Text. Cham, Switzerland: Springer, 2018.
- [14] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.
- [15] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," Proc. International Conference on Learning Representations (ICLR), 2013.