An AI-Based Skill Gap Analysis and Personalized Career Recommendation System

Using Natural Language Processing and Machine Learning

Prachi Surve[1], Shrikant Chauhan[2], Shraddha Mishra[3], Mohd Danish Shah[4]

[1]Assistant Professor, D. Y. Patil Deemed to be University, School of Humanities & Sciences, Belapur, Maharashtra, India

[2]Student, Department of MCA, D. Y. Patil Deemed to be University, School of Humanities & Sciences, Belapur, Maharashtra, India.

[3]Student, Department of MCA, D. Y. Patil Deemed to be University, School of Humanities & Sciences, Belapur, Maharashtra, India.

[4]Student, Department of MCA, D. Y. Patil Deemed to be University, School of Humanities & Sciences, Belapur, Maharashtra, India.

---***---

Abstract - *Students and early-career individuals often find it difficult to determine which careers match their existing skills and what they should learn to become suitable for a desired role. Existing career guidance and job-matching systems provide useful individual components such as resume analysis, job matching, personality-based guidance, skill-gap detection, and course recommendation, but these components are often developed for different purposes and are not always connected into a transparent student-centered process. This paper proposes a theoretical AI-based framework that connects these stages into one explainable pipeline. The proposed system uses established Natural Language Processing techniques for extracting and normalizing skills, structured career-skill representations for role requirements, set-based comparison for identifying skill gaps, TF-IDF and cosine similarity as an interpretable lexical baseline, and Sentence-BERT (SBERT) embeddings for semantic career matching. Identified gaps are connected to learning resources to form a personalized learning path. No new algorithm, dataset, implementation, or experimental result is claimed. The contribution is the theoretical integration of established techniques into a traceable sequence in which career recommendations connect to user skills, skill gaps connect to role requirements, and learning recommendations connect to identified gaps.*

Key Words: Natural Language Processing, Skill Gap Analysis, Career Recommendation, Semantic Similarity, SBERT, Resume Analysis, Personalized Learning

1. INTRODUCTION

Career selection is no longer simply a matter of choosing an occupation from a list. Students need to understand their existing competencies, identify suitable career roles, recognize missing skills, and determine what they should learn next. This creates a gap between traditional career guidance and the practical process of preparing for a particular career.

The reviewed literature contains systems for personalized career guidance, resume parsing, job matching, skill-gap detection, course recommendation, and reskilling. NLP is used to extract information from resumes and job descriptions, while BERT-family models can address differences in terminology during matching. However, these activities are not always connected into one clear chain from a student's current abilities to a suitable career and then to the learning required to reach it.

The proposed framework therefore follows: **Student Profile → NLP Skill Extraction → Skill Normalization → Career Requirement Modeling → Skill Matching → Skill-Gap Analysis → Career Ranking → Personalized Learning Path.** It does not introduce a new NLP model or mathematical algorithm; established methods are selected for each stage and connected into a single theoretical architecture.

1.1 PROBLEM STATEMENT

Existing career guidance systems demonstrate useful techniques, but traditional keyword matching can depend heavily on exact terminology. Similarity-based recommendation also does not necessarily explain what a student is missing for a target role. A further limitation is that a career recommendation may not be connected to a concrete learning process. The research problem is therefore: **How can established NLP, information-retrieval, semantic-similarity, and recommendation techniques be integrated into a theoretical system that identifies a student's skills, determines skill gaps, recommends suitable career roles, and produces a personalized learning path in an explainable manner?**

1.2 OBJECTIVES AND RESEARCH QUESTIONS

The objectives are to extract and normalize student skills; represent career roles through structured skill requirements; identify missing competencies; use lexical and semantic matching; rank suitable careers; connect gaps with learning resources; and maintain an explainable relationship between profile, career, gaps, and learning.

RQ1: How can NLP transform unstructured career information into a reliable skill profile?
RQ2: How can current skills and role requirements be compared to produce interpretable skill-gap reports?
RQ3: How can semantic similarity improve career matching beyond literal keyword overlap?
RQ4: How can identified gaps be translated into personalized learning paths?

2. LITERATURE REVIEW AND RESEARCH GAP

2.1 Profile-Based Career Guidance

Profile-based systems show that career guidance can be personalized using user information, aspirations, and career theories. Kulugh et al. describe an AI-powered personalized career guidance approach in which user profiles and career theories support recommendations [2]. Agarkar et al. combine NLP, knowledge representation, personality-related features, and recommendation techniques [1]. These studies establish the value of personalization, but the proposed framework places skill evidence and explicit gap analysis at the center of the process.

2.2 NLP for Resume and Skill Analysis

Resume and career information are largely unstructured text. Sujitha et al. discuss NLP-based resume analysis and skill-gap identification as a basis for personalized development [5]. Shingade et al. use semantic embeddings and supporting recommendation components for career and skill analysis [4]. These approaches motivate a pipeline in which raw profile text is first transformed into a standardized representation before matching.

2.3 Semantic Matching for Career Recommendation

Classical information-retrieval methods such as TF-IDF and cosine similarity provide an interpretable way to represent and compare textual information. Semantic approaches based on BERT and SBERT can represent contextual similarity when related experience is expressed using different terminology. Reimers and Gurevych introduced Sentence-BERT for efficient sentence-level semantic similarity [7]. Vanga et al. describe a career portal using NLP and BERT-based semantic similarity for job matching and skill-gap detection [6].

2.4 Skill Gap and Learning Recommendation Skill-gap studies

Connect missing competencies with development roadmaps and learning resources. Roy Choudri et al. discuss personalized career guidance that links capabilities with job requirements and learning resources [3]. The common practical issue is that identifying a gap is only the beginning; a useful student-facing system should also indicate what can be learned to address that gap.

2.5 Research Gap

The research gap is not the absence of algorithms. Established techniques already exist for extraction, matching, gap identification, and recommendation. The gap addressed here is their focused integration into a transparent, student-centered sequence: skill extraction normalization career matching skill-gap explanation career ranking learning recommendation. The contribution is integration and traceability rather than a new standalone algorithm.

3. PROPOSED THEORETICAL SYSTEM

The architecture is modular. Each module has a defined input, established technique, processing role, output, and connection to the next module. No module is presented as a newly invented algorithm.

3.1 User Profile Acquisition Established technique:

Structured profile representation and document processing. Input: resume, education, projects, certifications, experience, interests, and optional preferences. Processing: collect career-relevant information and exclude unnecessary attributes. Output: career-relevant user profile. Connection: pass profile to NLP preprocessing.

3.2 Text Preprocessing Established technique:

Text cleaning, tokenization, normalization, and removal of irrelevant text. Input: raw resume/profile text. Processing: reduce formatting noise and produce consistent textual units. Output: cleaned career-related text. Connection: pass text to skill extraction.

3.3 Skill Extraction Established technique:

Named Entity Recognition (NER), vocabulary matching, and NLP entity extraction. Input: cleaned resume/profile text. Processing: identify programming languages, tools, frameworks, databases, and other competencies. Output: extracted skill list. Connection: pass skills to normalization.

3.4 Skill Normalization Established technique:

Vocabulary-based normalization and synonym / abbreviation mapping. Input: extracted skills. Processing: map known variants such as JS and JavaScript to a common label when the relationship is defined. Output: standardized skill profile. Connection: compare standardized skills with career requirements.

4. CAREER REQUIREMENT AND GAP MODEL

4.1 Career Requirement Modeling

Established technique: structured knowledge representation. Input: career/job descriptions and industry requirements. Processing: represent each role using required skills, optional skills, and learning-resource mappings. Output: career-skill repository. Connection: provide reference requirements to matching.

4.2 Skill Matching and Gap Analysis Established technique:

set comparison and set difference. Let Su denote the standardized student skill set and Sr the required skill set for a target role. The shared skills can be represented by Su Sr, while missing requirements can be represented by Sr Su. An interpretable coverage value can be expressed as C = |Su Sr| / |Sr| when a role has a defined required-skill set. Output: matched skills, missing skills, coverage, and gap explanation.

4.3 Semantic Career Matching Established technique:

Sentence-BERT (SBERT) embeddings with cosine similarity. Input: student profile/resume and career/job descriptions. Processing: convert texts to dense semantic representations and compare their contextual alignment. Output: semantic similarity for each candidate career. Connection: combine semantic evidence with explicit skill evidence.

4.4 Career Ranking Established technique:

Similarity-based ranking and Top-K recommendation. Input: skill coverage, missing skills, semantic similarity, and requirements. Processing: order candidate careers by available evidence of fit while exposing missing requirements as readiness limitations. Output: ranked career list. Connection: send selected career and gaps to learning.

5. PERSONALIZED LEARNING AND EXPLAINABILITY

5.1 Personalized Learning Path Established technique:

Content-based learning-resource recommendation and prerequisite-aware sequencing. Input: selected career and missing skills. Processing: map gaps to relevant resources and order them using relevance and known prerequisites. Output: personalized learning recommendations. Connection: present the path with the selected career and explanation.

5.2 Explainability Layer Established technique:

Evidence-based recommendation explanation. Input: outputs of previous modules. Processing: show why a role was recommended, which skills support it, which requirements are missing, and which learning actions correspond to those gaps. Output: human-readable recommendation report. The explanation maintains a direct relationship between the evidence extracted from the profile and the recommended action.

5.3 Integrated Decision Logic

The framework is intentionally sequential. A career is not recommended from semantic similarity alone. Explicit skills provide interpretable evidence, semantic similarity addresses contextual variation, and gap analysis identifies requirements for which evidence is absent. The ranking stage combines these sources to generate career alternatives, after which the learning stage converts missing skills into actions. This makes the proposed system a decision-support framework rather than an opaque career classifier.

6. OVERALL SYSTEM FLOW

Student Resume/Profile Text Preprocessing NLP Skill Extraction Skill Normalization Standardized Student Skill Profile Career Requirement Repository Explicit Skill Matching + Gap Analysis SBERT Semantic Matching Career Ranking Selected Career + Missing Skills Learning Resource Recommendation Personalized Learning Path + Explanation.

The modules are connected rather than independent. The output of one stage becomes the input of the next, creating a traceable chain from current skills to career alternatives and finally to learning actions. The same structure can support multiple candidate roles because the student profile can be compared against each role in the career-skill repository.

7. THEORETICAL EVALUATION PLAN

The present study reports no experimental results because the proposed system has not been implemented or experimentally validated. Future implementation can evaluate each module using established measures. The evaluation should test the complete pipeline as well as individual components so that errors can be localized.

Module Possible evaluation
Skill extraction Precision, Recall, F1-score
Skill normalization Normalization accuracy
Career matching Precision, Recall, F1-score
Career ranking Precision@K, NDCG, Top-K evaluation
Skill-gap detection Precision, Recall, F1-score
Learning recommendation
Expert judgment and user feedback

Explainability User questionnaire / explanation assessment A future implementation should compare an interpretable lexical baseline such as TF-IDF with cosine similarity against the selected semantic approach using the same evaluation data. Skill-extraction evaluation should use manually labeled entities or another justified reference standard. Ranking evaluation should use a predefined relevance judgment rather than informal inspection. Results reported by other studies must not be reused as results of this proposed framework.

8. DISCUSSION

The central theoretical observation is that skill-gap analysis is better treated as a connected pipeline than as a single prediction task. Skill extraction creates the representation needed for later comparison. Normalization reduces terminology variation. The career repository provides an explicit reference point. Gap analysis makes the recommendation interpretable, while semantic matching addresses contextual similarity that exact word overlap may miss.

The combination of explicit and semantic evidence also addresses a limitation of using similarity alone. A semantic score can indicate that a profile is related to a role, but it cannot by itself clearly identify which competencies are absent. The gap module therefore remains an independent decision layer and supplies the evidence used for learning recommendations.

The ranking stage should not be treated as a deterministic statement about a student's future. Several roles may be reasonable for the same profile, and a ranked list can communicate alternatives together with their remaining skill requirements. This makes the framework suitable for career decision support while leaving the final choice with the user.

The learning module closes the gap between recommendation and action. Instead of stopping at a career label, the system identifies the competencies required to move toward that role and maps them to learning resources. The final explanation therefore has a traceable chain: profile evidence career requirement skill gap learning action.

9. LIMITATIONS AND ETHICAL CONSIDERATIONS

The framework depends on the quality and coverage of profile and career data. A resume may omit skills that a student possesses, while job requirements can change. Semantic models can reduce terminology problems but can also produce similarity values that are difficult to interpret. Explicit skill evidence should therefore remain part of the recommendation.

The proposed coverage comparison also depends on how required skills are defined. A role with a poorly specified requirement list can produce a misleading gap report. Likewise, a skill may have different proficiency expectations across roles. Future implementation should therefore consider proficiency levels and evidence of experience rather than treating every skill as equally strong.

Privacy is important because resumes may contain personally identifiable information. Only information needed for career analysis should be supplied to the recommendation process, and protected or unnecessary attributes should not determine career ranking. Personality or interest measures may be added later as optional inputs, but should not be the sole basis for a career decision.

The framework should be treated as decision support rather than an authority that determines a student's career. Users should be able to review extracted skills, correct inaccurate profile information, and understand why a recommendation was produced.

9.1 Scope of the Proposed Framework The framework is intended primarily for students and early-career users who need a structured view of their current competencies and possible next steps. It is not intended to replace human career counseling, recruitment decisions, or formal professional competency assessment. The quality of the final recommendation remains dependent on the completeness of the profile, the definition of role requirements, and the quality of the learning-resource repository.

9.2 Practical Implications For students, the main value of the proposed framework is the conversion of an unstructured resume into an actionable career-development view. A user can see the competencies already represented in the profile, compare them with requirements for several roles, and identify the gaps that affect readiness. For academic institutions, the same framework can support curriculum review by revealing recurring gaps between student profiles and role requirements. For career-support services, the explanation layer can make recommendations easier to inspect than a single unexplained ranking score.

9.3 Future Implementation Considerations A practical implementation should maintain versioned skill and career vocabularies because technologies and role requirements change over time. It should also retain the evidence used for each recommendation so that users can inspect the source of a detected skill or missing requirement. User corrections can be incorporated as profile updates, while evaluation should distinguish extraction errors, matching errors, ranking errors, and learning-resource errors. These considerations provide a direct path from the theoretical architecture to a testable implementation without assuming performance in advance.

10. CONCLUSION AND FUTURE WORK

This paper proposed a theoretical AI-based framework for skill-gap analysis and personalized career recommendation. The system connects established techniques into one sequence: NLP extracts skills, normalization creates comparable representations, structured career requirements provide the comparison reference, set-based analysis identifies missing competencies, SBERT and cosine similarity support semantic career matching, ranking produces career alternatives, and learning-resource recommendation converts gaps into a personalized development path.

The paper does not claim a newly invented algorithm, implemented software, new dataset, or experimental performance. Its contribution is the integration of established techniques into an explainable and traceable architecture.

The central principle is: Current Skills Suitable Careers Missing Skills Required Learning.

Future work can implement the architecture, construct a labeled dataset, compare lexical and semantic matching under identical conditions, evaluate skill extraction, test alternative ranking strategies, and conduct user studies on explanation quality and learning recommendations.

ACKNOWLEDGEMENT The author sincerely thanks the faculty and academic mentors for their guidance and support in developing the proposed research framework.

REFERENCES

[1] P. Agarkar, R. Mahant, S. Chavan, and A. Bhosale, “A Comprehensive Base Paper: AI-Powered Personalized Career Guidance System,” International Journal of Ingenious Research, Invention and Development, Vol. 5, Issue 3, 2026.

[2] V. E. Kulugh, K. P. Aondover, and A. S. Faki, “Artificial Intelligence-Powered Personalised Career Guidance System,” Dutse Journal of Pure and Applied Sciences, Vol. 11, No. 1c, pp. 92–104, 2025.

[3] S. M. Roy Choudri, S. Tejaswi, K. S. Saina Sai Kumar, D. S. Babu, and C. V. Kalyan, “Modern Career Guidance: A Personalized Approach to Success,” Journal of Nonlinear Analysis and Optimization, Vol. 16, Issue 1, 2025.

[4] S. Shingade, K. Tardalkar, A. Parab, A. Jamble, and S. Pathak, “AI-Based Career Copilot for Skill Gap Analysis and Job Recommendation Using Semantic Embeddings.”

[5] M. Sujitha, T. V. Ananthan, G. Lakshmi, A. Ramya Teja Sree, and G. Sahithi, “Skill Gap Analysis Using Machine Learning,” International Journal of Scientific Research in Engineering and Management, Vol. 9, Issue 3, 2025.

[6] L. Vanga, A. Zunje, and P. Patil, “Career Bridge AI: An Intelligent Career Portal with Bidirectional Job Matching, Resume Analysis, and Skill Gap Detection,” International Journal of Innovative Research in Engineering, Vol. 7, Issue 3, 2026.

[7] N. Reimers and I. Gurevych, “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks,” Proceedings of EMNLP-IJCNLP, 2019.