ISSN: 2249-7196
IJMRR/Jan-Mar. 2026/ Volume 16/Issue 1/65-74
A. Jyothika / International Journal of Management Research & Review

AI-Powered Skill Gap Identification and Career Guidance Platform Using Machine Learning and Natural Language Processing

A. Jyothika
Department of Electronics and Computer Engineering
J. B. Institute of Engineering and Technology, Hyderabad, India
Mr. Bheemana Bhuvan
Associate Professor, Department of Electronics and Computer Engineering
J. B. Institute of Engineering and Technology, Hyderabad, India.
bhuvan.ecm@jbiet.edu.in

Abstract

The rapid transformation of industry driven by automation, artificial intelligence, and digital services has significantly altered workforce requirements. A major consequence of this transformation is the widening gap between the skills possessed by individuals and the competencies demanded by employers. Conventional career guidance mechanisms are typically static, manually driven, and poorly aligned with real-time labour market dynamics.

This paper presents an AI-powered skill gap identification and career guidance platform that integrates natural language processing, machine learning, and recommendation techniques to provide personalized career pathways. The proposed system automatically extracts skills from user profiles and resumes, predicts suitable job roles, computes a Skill Gap Index (SGI), and recommends targeted learning resources. The platform is implemented using a lightweight machine learning pipeline and a web-based service architecture. Experimental results demonstrate that the system can reliably identify user skills, predict relevant career roles, and generate actionable recommendations for upskilling. The proposed framework offers a scalable and transparent decision-support solution for modern career development ecosystems.

Keywords: *Skill gap analysis, career guidance, machine learning, NLP, recommendation system, employability analytics.*

1. Introduction

Digitalization and the rapid diffusion of artificial intelligence technologies are reshaping nearly every professional domain. New job roles emerge continuously, while many traditional positions evolve or disappear. This dynamic labour market creates persistent challenges for students and professionals attempting to align their competencies with industry expectations.

Traditional career counselling approaches rely primarily on manual assessments, static aptitude tests and counsellor experience. Although these methods provide general guidance, they lack the ability to continuously track evolving market needs or offer personalized recommendations at scale. Furthermore, limited access to professional career counselling often restricts students from rural and economically disadvantaged backgrounds.

Recent advances in machine learning and natural language processing offer promising opportunities to transform career guidance into a data-driven and adaptive service. By analysing resumes, assessments and real-time job data, intelligent platforms can identify specific skill deficiencies and guide individuals toward suitable career paths. This work proposes an AI-powered platform that performs automated skill extraction, role prediction, gap analysis and personalized learning recommendations in a unified framework.

The key contributions of this work are:

- design of an end-to-end AI-driven skill gap identification framework,
- development of a scalable career recommendation pipeline,
- introduction of an interpretable Skill Gap Index for user guidance, and
- experimental evaluation of the proposed system.

2. Related Work

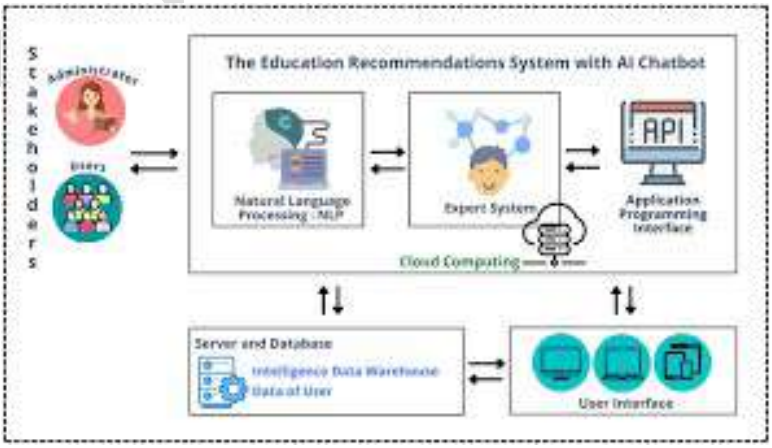
Intelligent career recommendation systems have gained increasing attention in educational technology and workforce analytics. Early systems relied on rule-based decision trees and predefined mappings between academic background and job roles. While simple to implement, such systems are limited in adaptability and personalization.

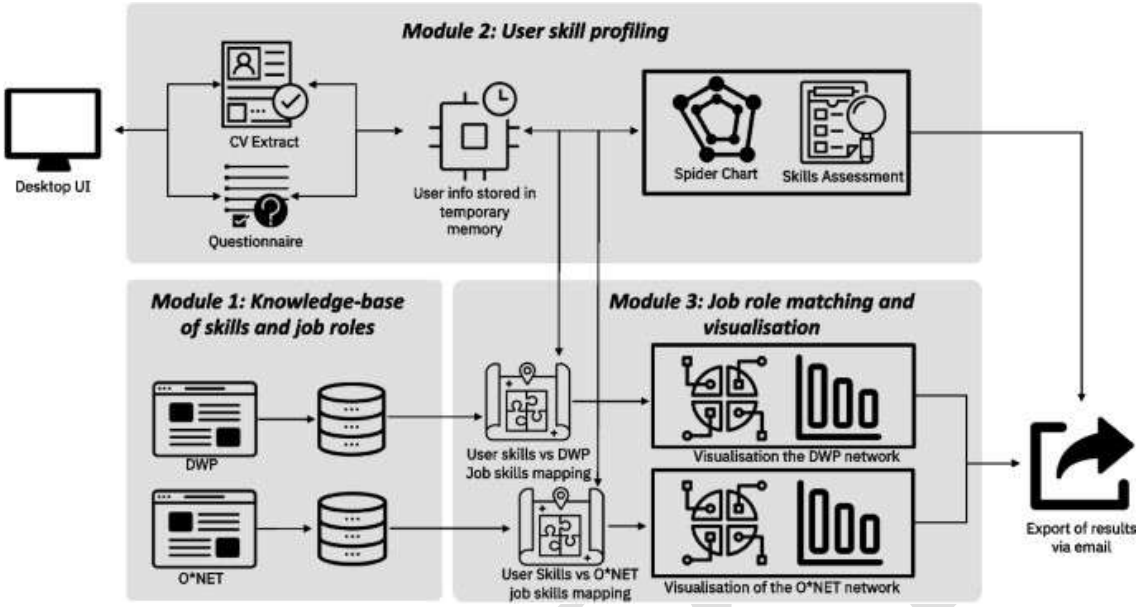
More recent research has focused on machine learning-based career prediction models using academic performance, psychometric assessments and historical employment data. Natural language processing techniques have been widely adopted to extract skills from resumes and job descriptions. Deep learning models and semantic embeddings have also been explored for capturing complex relationships between skills and occupational roles.

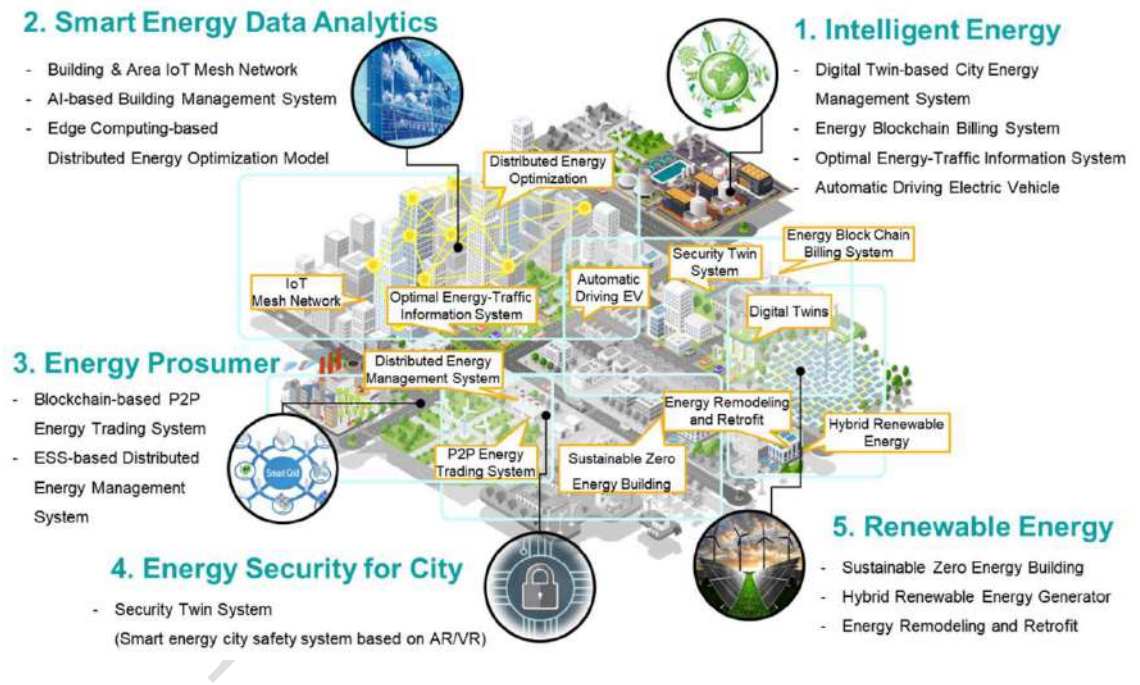
Several studies have demonstrated the effectiveness of recommendation systems that integrate labour market analytics and course platforms to provide targeted upskilling suggestions. However, many existing solutions either emphasize prediction accuracy without transparency or lack explicit gap quantification mechanisms.

The proposed platform addresses these limitations by combining interpretable gap metrics, lightweight machine learning models and an extensible recommendation layer.

3. Overall System Architecture





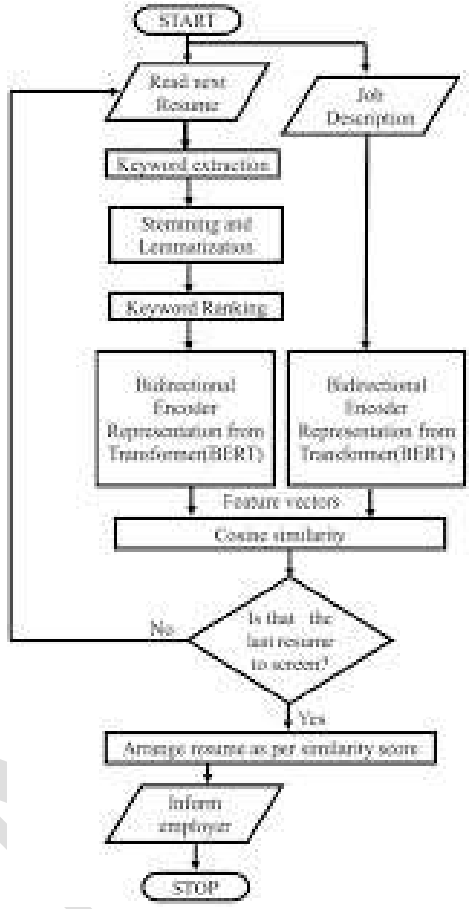


The system follows a modular and service-oriented architecture consisting of five primary layers:

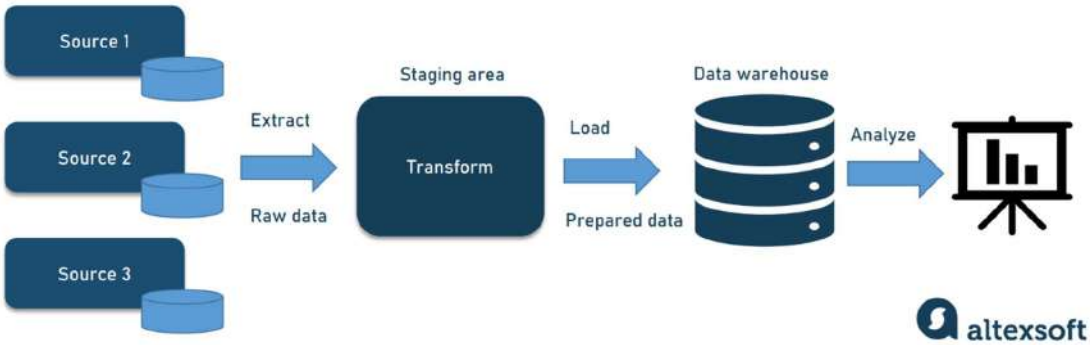
1. **User interaction layer** – provides a web interface for resume upload, assessments and career exploration.
2. **Data processing layer** – performs text normalization, tokenization and structured data preparation.
3. **AI analysis layer** – executes skill extraction, role prediction and skill gap computation.
4. **Recommendation layer** – generates learning and career suggestions.
5. **Storage and integration layer** – manages user profiles and external learning resources.

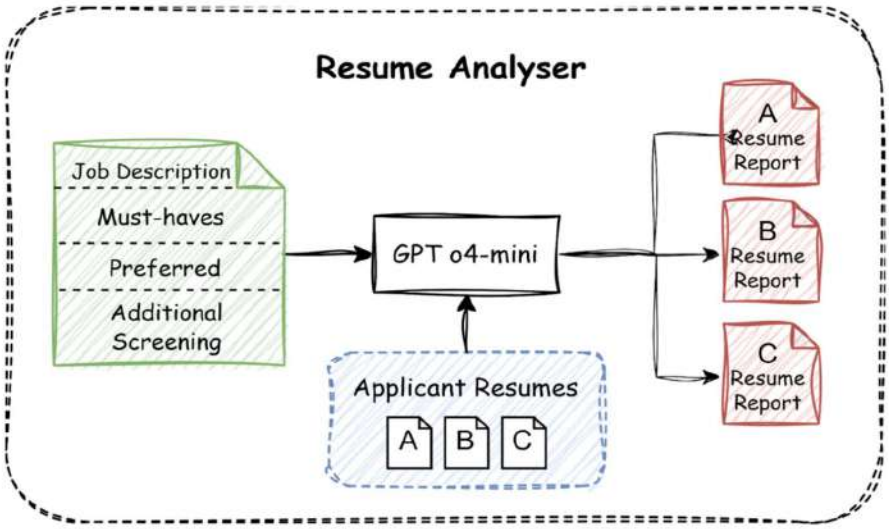
This layered design ensures scalability and enables independent upgrades of analytical components.

4. Workflow and Data Processing Pipeline



ETL PIPELINE





The operational workflow of the proposed platform is summarized as follows:

1. Users submit resumes or textual skill descriptions.
2. Pre-processing and tokenization are performed.
3. Skills are extracted using vocabulary-driven and NLP-based techniques.
4. A trained classifier predicts suitable career roles.
5. The Skill Gap Index is computed by comparing user skills with role-specific requirements.
6. Learning resources and career pathways are recommended.

This pipeline enables real-time and repeatable assessment.

5. Proposed Methodology

5.1 Skill Extraction

The system extracts relevant technical and professional skills from free-text resumes using text normalization, token matching and curated skill dictionaries. Multi-word expressions such as *data visualization* and *machine learning* are preserved to ensure semantic correctness.

5.2 Career Role Prediction

Textual representations of resumes are transformed using TF–IDF vectorization. A multi-class classification model based on logistic regression predicts the most suitable job roles. This approach offers strong interpretability and computational efficiency.

5.3 Skill Gap Index Computation

For each predicted role, a predefined role-skill map is used to calculate the Skill Gap Index (SGI):

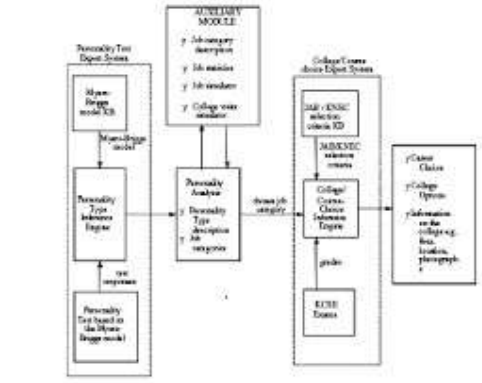
$$SGI = \left(1 - \frac{|\ MatchedSkills\ |}{|\ RequiredSkills\ |}\right) \times 100$$

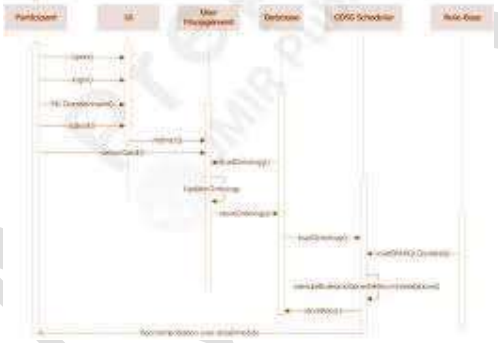
The SGI provides an intuitive quantitative measure of readiness for a given role.

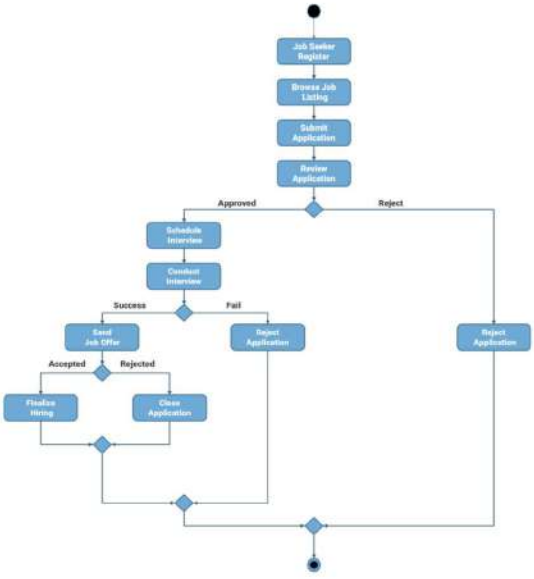
5.4 Recommendation Engine

Missing skills are mapped to curated learning resources and course suggestions. The recommendation layer uses content-based filtering to propose relevant upskilling paths aligned with the user's career objectives.

6. UML-Based System Modelling







The class diagram captures the relationships between user profiles, skill extraction services, predictive models and recommendation modules.
The sequence and activity diagrams illustrate how user requests are processed through the analytical pipeline. The use-case model represents interactions between learners, administrators and the platform.

7. Implementation Framework

The backend of the platform is developed using Python and a lightweight web framework. Machine learning components rely on scikit-learn for feature extraction and classification. Data persistence is supported through relational and document-based storage for structured profiles and unstructured resume data.
The platform exposes REST-based services for assessment and recommendation, enabling integration with learning management systems and external job portals. The modular design allows new job roles and skill definitions to be introduced without retraining the entire system.

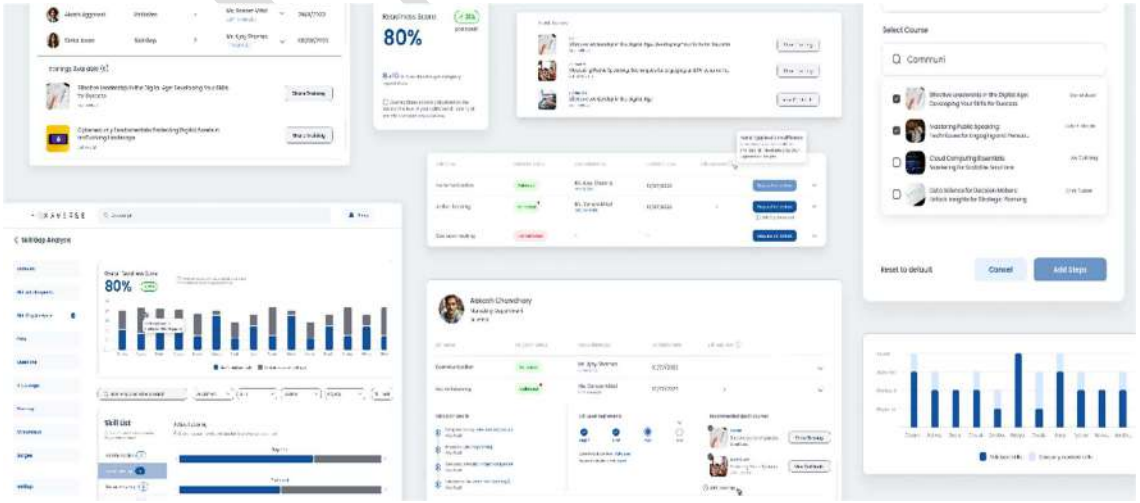
8. Experimental Setup

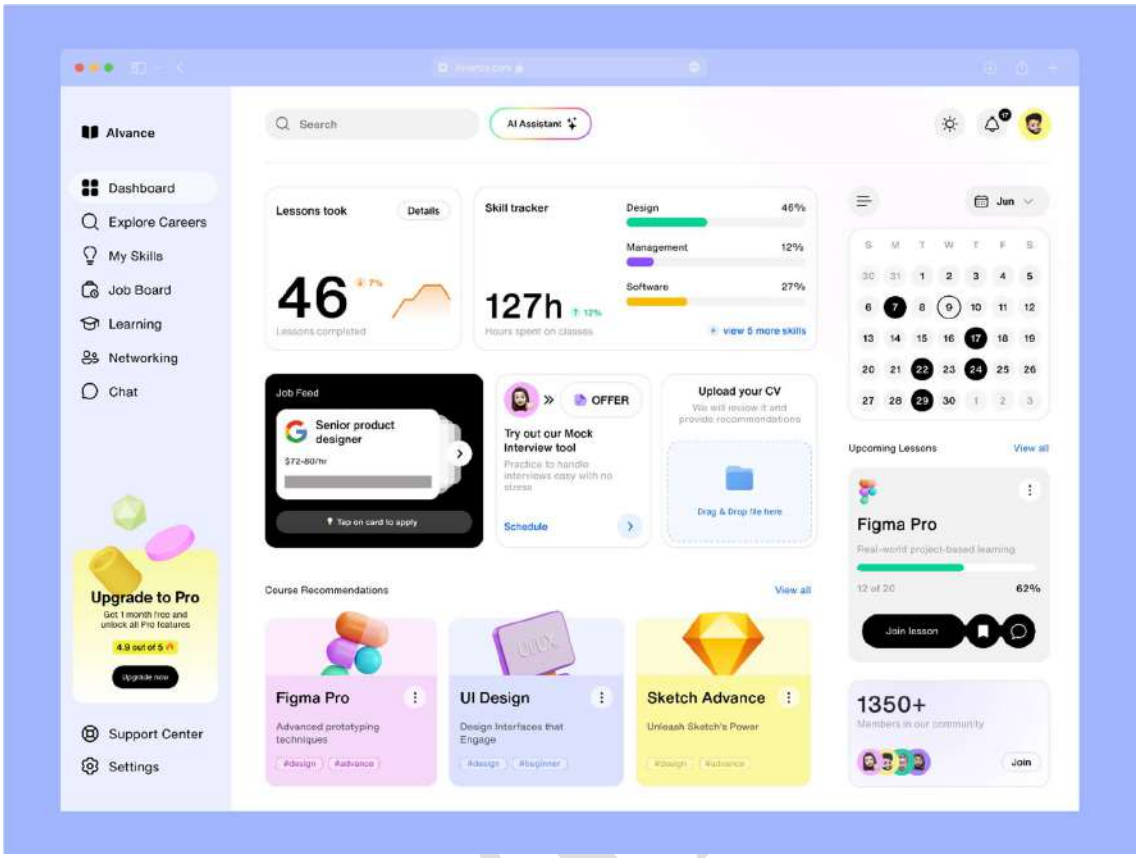
The system is evaluated using a curated dataset containing resume texts and associated job roles. Data are divided into training and testing partitions. Performance is measured using:

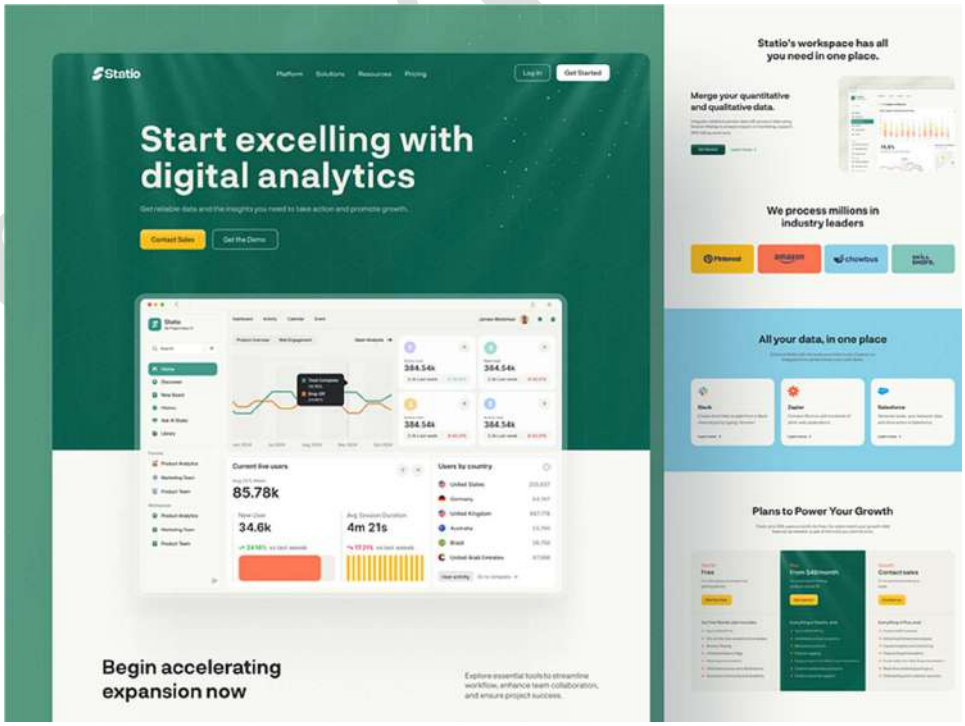
- classification accuracy,
- precision and recall,
- F1-score, and
- response latency.

The evaluation focuses on both predictive performance and operational efficiency.

9. Results and Discussion




Alvance
Search
AI Assistant
Dashboard
Explore Careers
My Skills
Job Board
Learning
Networking
Chat
Support Center
Settings
Lessons took
Details
46
Skill tracker
Design 46%
Management 12%
Software 27%
127h
view 5 more skills
Job Feed
Senior product designer
$72-80/hr
Tap on card to apply
OFFER
Try out our Mock Interview tool
Schedule
Upload your CV
Drag & Drop file here
Upgrade to Pro
Get 1 month free and unlock all Pro features
Upgrade now
Course Recommendations
View all
Figma Pro
Advanced prototyping techniques
UI Design
Design Interfaces that Engage
Sketch Advance
Unleash Sketch's Power
Upcoming Lessons
View all
Figma Pro
Real-world project-based learning
12 of 20
62%
Join lesson
1350+
Members in our community
Join


Statio
Start excelling with digital analytics
Begin accelerating expansion now
Statio's workspace has all you need in one place.
Merge your quantitative and qualitative data.
We process millions in industry leaders
All your data, in one place
Plans to Power Your Growth

The trained role prediction model demonstrates high classification accuracy and stable generalization across multiple job categories. The skill extraction component reliably identifies both individual and composite skill terms from resumes.

The Skill Gap Index effectively reflects the readiness of users for specific career roles. Lower SGI values correspond to higher compatibility, enabling users to easily understand their preparedness. The recommendation module generates relevant learning suggestions aligned with missing competencies.

The system responds within sub-second latency for standard user requests, demonstrating suitability for real-time deployment in academic and training environments.

10. Ethical, Privacy and Transparency Considerations

Career analytics platforms process sensitive personal and educational information. The proposed system incorporates the following safeguards:

- encrypted communication channels,
- anonymized analytical logs,
- role-based access control, and
- explainable outputs through visible skill mappings and SGI scores.

The platform avoids automated hiring decisions and is explicitly positioned as a decision-support tool rather than an authoritative evaluator.

11. Comparative Analysis

Compared with traditional counselling systems, the proposed approach provides continuous assessment, personalized learning recommendations and real-time labour market alignment. Unlike static career portals, the platform supports iterative improvement of user profiles and adapts recommendations as skills evolve.

The lightweight machine learning architecture allows deployment in resource-constrained academic environments without requiring specialized hardware.

12. Limitations

The current prototype uses a controlled and limited role-skill mapping. Vocabulary-based skill extraction may fail to recognize emerging or domain-specific competencies. The role prediction model can further benefit from larger and more diverse datasets.

Future enhancements should incorporate semantic embeddings and live job-market data streams.

13. Conclusion

This paper presented an AI-powered skill gap identification and career guidance platform that integrates machine learning and natural language processing to provide personalized career recommendations. The system automatically extracts user skills, predicts relevant job roles, computes an interpretable Skill Gap Index and suggests targeted learning resources.

Experimental evaluation confirms the effectiveness and efficiency of the proposed framework. The platform offers a practical and scalable solution for enhancing employability, guiding learners and supporting educational institutions in aligning training programs with market needs.

14. Future Work

Future developments will focus on:

- integration of transformer-based NLP models for improved semantic understanding,
- real-time labour market analytics from online job portals,
- multilingual resume processing, and
- long-term user performance tracking and adaptive recommendation strategies.

References

1. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.
2. A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, O'Reilly, 2019.
3. J. Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," 2019.
4. T. Mitchell, *Machine Learning*, McGraw-Hill, 1997.
5. S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, Pearson, 2021.