International Journal of Engineering Research and Sustainable Technologies
Volume 3, No.3, Sep 2025, P 33 – 39
ISSN: 2584-1394 (Online version)

SKILL GAP ANALYSIS USING MACHINE LEARNING

[1]Rameshbabu.V, [2]Latha R, [3]Sreenithi R
[1]Department of CSE, [2]Department of EFL, [3]Dept. of Computer Technology
[1]Dr.MGR Educational and Research Institute, Tamil Nadu, India
[2]SRM Institute of Science and Technology, TN, India
[3]MIT, Anna University, Chennai, TN, India
*** Corresponding author email address**: rameshbabu.cse@drmgrdu.ac.in

DOI: htpps://doi.org/10.63458/ijerst.v3i3.122 | ARK: https://n2t.net/ark:/61909/IJERST.v3i3.122

Abstract
The "Skill Gap Analysis Using Machine Learning" project aims to bridge the gap between user skillsets and industry requirements. It examines user resumes, finds skill gaps, and gives a path for skill improvement using Natural Language Processing (NLP) and machine learning techniques. Additionally, the system suggests resources for interview preparation based on the professional domains that the user has chosen. This platform supports firms in workforce development while providing users with practical insights for career growth.

Keywords: Skill Gap Analysis, Machine Learning, Natural Language Processing (NLP), Resume Analysis, Skill Enhancement, Industry Requirements, Career Growth, Personalized Development Roadmap, Interview Preparation Resources.

1. Introduction

The Skill Gap Analysis Using Machine Learning project aims to help people get the skills they need for job openings. It does this by using machine learning and resume analysis (NLP) to examine resumes, spot skill gaps, and offer tips for improvement. The project users also receive a plan that aims to teach them new skills and assist them in interviews for their target professions. The tool is help student's, employee's, and organization's attention to properly skill them. It enables enhanced career progression through proper guidance and resources and helps organizations build a skilled workforce for the present and coming demands. The Skill Gap Analysis Using Machine Learning project aims to help people gain the skills industries need. This system applies Natural Language Processing (NLP) and machine learning algorithms to examine resumes. It spots gaps in knowledge or expertise. To bridge these gaps, users get a custom plan with resources to learn and prepare for interviews. This method helps them acquire the skills they need to pursue their career goals with confidence. This program proves useful for students, job hunters, and professionals who want to move up in their careers. It also gives companies useful data that reveals the accepted standards in the user's field and offers personalized advice. The project also tries to help businesses find and fix skill gaps among their staff helping them plan training and create strategies to address these issues. It also serves the user's business and organization through skills and opportunities assessments, integrating it with the data concerns and phenomena of the users.

2. Literature Review

2.1 Utilizing generative AI to assist individuals in gaining IT skills.

This research considers how generative AI may assist. acquiring new IT skills. It considers means by which personalized learning plans for staff based on their skill profiles. The essay also discusses how generative AI tools can suggest training timetables to improve skills. It means the necessity to incorporate artificial intelligence into corporate training . This study is of great significance to IT firms seeking to future-proof their employees**.**

2.2 Nacelle: Conversing with AI and a Knowledge Graph Skills Gap Analysis to Enable Continuing Learning at Office This paper introduces a system called 'Nacelle based on Utilizing chat AI and knowledge graphs to determine skill gaps. It seeks to create environmentally friendly learning environments at workplaces. The research illustrates how Nacelle estimates missing skills and proposes course material. It is definite regarding a specific emphasis Knowledge graphs assist in structuring and interlinking information. This Work is a new way of creating teams using chat tools

2.3 Applying Graph Methods to Investigate Skill Deficits

This research considers the use of graph algorithms and databases to demonstrate what the workers are capable of. It discusses how graph-based Methods can help project staffing and create candidates to be trained. The research emphasizes the scalability and performance of large graph algorithms data sets. It also looks at how they impact one another existing HR systems. It is suitable for large and dynamic workforce organizations.

2.4 Machine Learning and AI Technology-Induced Skill Gaps and Opportunities for Continuous Development of Middle-Skilled Employees :

This paper discusses how machine learning and AI technology influence skill gaps in middle-skilled employees This paper examines how machine learning and AI technology impacts skill gaps among middle-skilled employees. It emphasizes the need for continuous learning to remain current in changing job functions. The research also presents challenges that employees encounter in keeping up with technological changes. It emphasizes personal and organizational development strategies. The paper is a guide to enhance preparedness for AI and ML-enabled industry profession.

2.5 Closing the Skills Gap through Learning Analytics

The research looks into means in which learning analytics can help bridge the gap between education and industry needs. It discusses how institutions can track and enhance student skill improvement through the use of analytics systems. The article also addresses the applicability between student aspirations and employment market needs. It places great importance on the role of evidence-based approaches to curriculum design. This study is useful to schools which wish to increase students' employability.

3. System Methodology

Manual Assessment: Involves human review of resumes and qualifications, usually carried out by hiring managers or HR specialists. Time-It can result in bias, and it is not suitable for working with big data. Static Online Platforms: Career websites or job portals offer users static job descriptions and generic skill suggestions. Lack personalization, since they are unable to thoroughly scan individual resumes or career objectives. Basic Skill Matching Tools :Some systems offer keyword-based matching between job descriptions and user profiles. These tools often fail to identify deeper insights such as transferable skills or potential growth. Learning Management Systems (LMS) : Many organizations use LMS platforms for employee training and skill development. However, these systems often lack AI-driven features to identify skill gaps or provide personalized learning paths. Static Reports from Surveys or Databases Traditional skill gap analyses are based on industry reports or surveys, offering insights at an aggregate level. These are not tailored to individual needs, and they quickly become outdated due to the fast-changing job market. Rule-Based Resume Screening Software: Some existing tools screen resumes based on predefined rules or keywords. These systems are limited in flexibility and often overlook nuanced insights about a candidate's potential. Limitations of Existing Systems: Lack of personalization in skill recommendations., Inability to adapt to the rapidly changing skill demands in various industries. Limited use of advanced technologies like machine learning or NLP for deep analysis. Insufficient support for interview preparation or career-specific roadmaps.

4. Proposed System

The system automatically extracts skills from resumes using Natural Language Processing (NLP). Parsing resumes with different formats, such as PDFs and DOCs, was a challenge. NLP techniques, libraries like Spacy and NLTK for text processing, and PDF parsing libraries are used. In Implementation, Text extraction from resumes was done using PyPDF2 for PDFs and python- docx for Word documents. Skills were then extracted using pre- trained models for named entity recognition (NER) to identify key skills.

The system maps user career interests to suggest the most relevant skillset for that career. Aligning diverse career interests with precise skill categories. Machine learning algorithms, including clustering techniques, to categorize careers and identify skill sets are used. From the Implementation: Career interests are mapped using supervised learning, where users input their career aspirations, and the system suggests top skills using a pre- trained classifier.

Identifying the gap between the skills listed in a resume and the skills required for a specific career. Ensuring the skill gap analysis accounts for both hard and soft skills and matches them against a comprehensive database. NLP, machine learning models (such as decision trees or SVM), and a database of required skills for various roles. Implementation: A similarity score is calculated between the skills extracted from the resume and the predefined list of required skills for a given career. This is done using cosine similarity or other text similarity measures.

Generate a roadmap for users, suggesting relevant courses, resources, and interview questions to bridge the skill gap. Suggesting accurate, up-to-date, and career- relevant resources. Web scraping for course and resource recommendations, recommendation engines are used. Implementation: Integration with online learning platforms and APIs (like Coursera, LinkedIn Learning) to fetch recommendations for skill development resources. Track users' progress in acquiring new skills and adapting to industry demands. Designing a tracking mechanism that accurately reflects user improvement. Databases for tracking progress, APIs for updates on skill certifications. Users input their learning progress, which is then recorded in a database. Periodic assessments or quizzes may be implemented to evaluate skill proficiency.

4.1 Challenges Overcome

Developing the resume parsing and skill gap analysis system involved addressing several key challenges. Handling diverse resume formats (PDF, DOCX, TXT) presented difficulties in extracting structured data due to varying layouts and encoding inconsistencies. To mitigate this, libraries like PyPDF2 and python-docx were employed to standardize text extraction processes. Another significant hurdle was ensuring data privacy and security, especially when processing sensitive user information. Implementing encryption protocols and adhering to data protection regulations were essential to safeguard user data throughout the system.Flyaps

4.2 Technologies Used

The system leverages a combination of advanced technologies to achieve its objectives. For natural language processing tasks, libraries such as spaCy and NLTK were utilized to extract and process skills from resumes . Machine learning models for classification and gap analysis were developed using frameworks like Scikit-learn and TensorFlow. To store and manage skill and career data efficiently, relational databases like MySQL or PostgreSQL were employed. Web scraping and API integrations, facilitated by tools like BeautifulSoup and Scrapy, enabled the system to fetch updated online resources and job trends . The backend was developed using Python frameworks such as Flask or Django, while the frontend was built with HTML, CSS, and JavaScript to ensure a responsive user interface. Geeks for Geeks

4.3 Implementation

The implementation of the system was carried out in several phases. The frontend was developed using HTML, CSS, and JavaScript to create an intuitive user interface, allowing users to upload resumes, input career interests, and view their skill gap analysis. The backend, developed in Python, integrated machine learning models, data processing workflows, and database management functionalities. The entire application was deployed as a web service using Flask or Django, hosted on cloud platforms like AWS or Heroku to ensure scalability and accessibility. Testing was conducted using Python's unittest module for unit testing and manual testing for user interaction scenarios to ensure the system's reliability and user-friendliness.

The proposed system follows a structured methodology to analyze resumes, identify skill gaps, and generate personalized development roadmaps. The process begins with data collection, sourcing resumes in various formats (PDF, Word, TXT), and compiling career-related data such as job roles, required skills, and online course content from platforms like Coursera and LinkedIn Learning. The collected data undergoes preprocessing, where text is extracted and cleaned using techniques like tokenization, lemmatization, and removal of noise to ensure consistency. In the feature extraction phase, Natural Language Processing (NLP) techniques such as Named Entity Recognition (NER) and pre-trained models like spaCy or BERT are employed to extract relevant technical and soft skills from the resumes. Users input their career interests, which are then mapped to skill requirements of corresponding job roles. The system then performs skill gap identification by comparing the extracted skills with job-specific skill sets using similarity measures like cosine or Jaccard similarity. Skills are further categorized to offer detailed insights. Machine learning models are trained on labeled datasets to detect gaps and predict required proficiency levels, using algorithms such as Random Forests and SVMs, and evaluated through accuracy, precision, and MSE metrics. The identified gaps are used to create a personalized development roadmap suggesting courses, certifications, and interview prep materials via a recommendation engine integrated with external platforms through APIs and scraping. An intuitive user interface facilitates resume upload and result visualization, while a feedback loop helps improve the system. The backend is developed using Python frameworks like Flask or Django, and the application is deployed on cloud platforms such as AWS or Heroku. Rigorous testing—including unit, integration, and user testing—ensures reliability and performance. Finally, the system supports continuous improvement by incorporating user feedback and updating resources to maintain relevance.

5. Result and Discussions

The system initiates the process by allowing the user to upload a resume and select a preferred domain. Once submitted, the system parses the resume to extract relevant information such as skills, qualifications, and experience. These extracted attributes are then matched against a predefined database containing industry-specific skill sets. Based on this comparison, the system identifies gaps between the user's current skillset and the required competencies for the chosen domain. It then generates a comprehensive skill gap analysis report, offering personalized recommendations for improvement. The analysis is quantified using two key metrics: the **Skill Gap Index (SGI)**, calculated using the formula

SGI = ((Required Skills−Acquired Skills) / Required Skills)) X 100 (1)

and the **Proficiency Match Rate (PMR)**, given by

PMR = (Skills Matched / Total Required Skills) X 100 (2)

These metrics help users understand their current standing and guide them in bridging the skill gaps effectively.

TF-IDF (Term Frequency-Inverse Document Frequency)

TF-IDF helps in converting textual data into numerical format and determining the importance of each word in a document relative to a corpus.

$$TF(t,d) = \frac{number\ of\ times\ t\ appears\ in\ d}{total\ number\ of\ terms\ in\ d}$$
$$IDF(t) = log\frac{N}{1+df}$$
$$TF-IDF(t,d) = TF(t,d) * IDF(t) \quad (3)$$

This will be used to assess how relevant a particular skill or keyword is within a resume or job description.

Cosine Similarity, Used to calculate the similarity between two text vectors (e.g., resume vs. job description) to identify if the skills match or if there are gaps.

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}, \quad (4)$$

Where A⃗\vec{A}A and B⃗\vec{B}B are the vector representations of two documents (resume and job description).

Clustering (K-Means)

You may use clustering to group similar resumes based on their skill sets. The K-Means algorithm works by minimizing the distance between data points and their cluster centers.

$$\mathbf{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y_i})^2. \quad (5)$$

Logistic Regression (for Skill Gap Prediction)

This can be used to predict whether there is a skill gap (binary classification: gap or no gap).

P(y = 1 ∣ X) = 1 + exp(−(w0 + w1X1 + w2X2 + ⋯ + wnXn)) (6)

Support Vector Machines (SVM) for Classification

Used for classifying whether a resume matches the required job skills.

w, bmin21 ∥ w ∥ 2 & yi(w⋅xi+b)≥1,∀I (7)

Where:

- xi\mathbf{x}_ixi are the feature vectors,
- yiy_iyi are the labels (match/no match),
- w\mathbf{w}w is the weight vector.

Natural Language Processing for Text Preprocessing

- Stopword Removal: Removing common words that do not contribute to meaning.
- Tokenization: Breaking text into words.
- Lemmatization: Converting words to their base form.

These steps are essential before applying machine learning models and are not directly represented by a formula but are key preprocessing steps.

Error Metrics (for Model Evaluation)
To evaluate the performance of your models, you'll need to use standard metrics like:

- Accuracy: Accuracy = Total Prediction / Correct Predictions
- Precision: Precision = TP + FPT

5.1 The Resume processing Work Flow

The resume processing workflow begins with the user logging into the system and uploading their resume. The system then validates the resume format to ensure it is compatible. If the format is valid, the resume is parsed to extract relevant skills. These extracted skills are stored and analyzed to identify any skill gaps. Based on this analysis, the system generates a roadmap to help the user bridge those gaps. Finally, the results, including skill gaps and recommendations, are displayed to the user, marking the completion of the process.

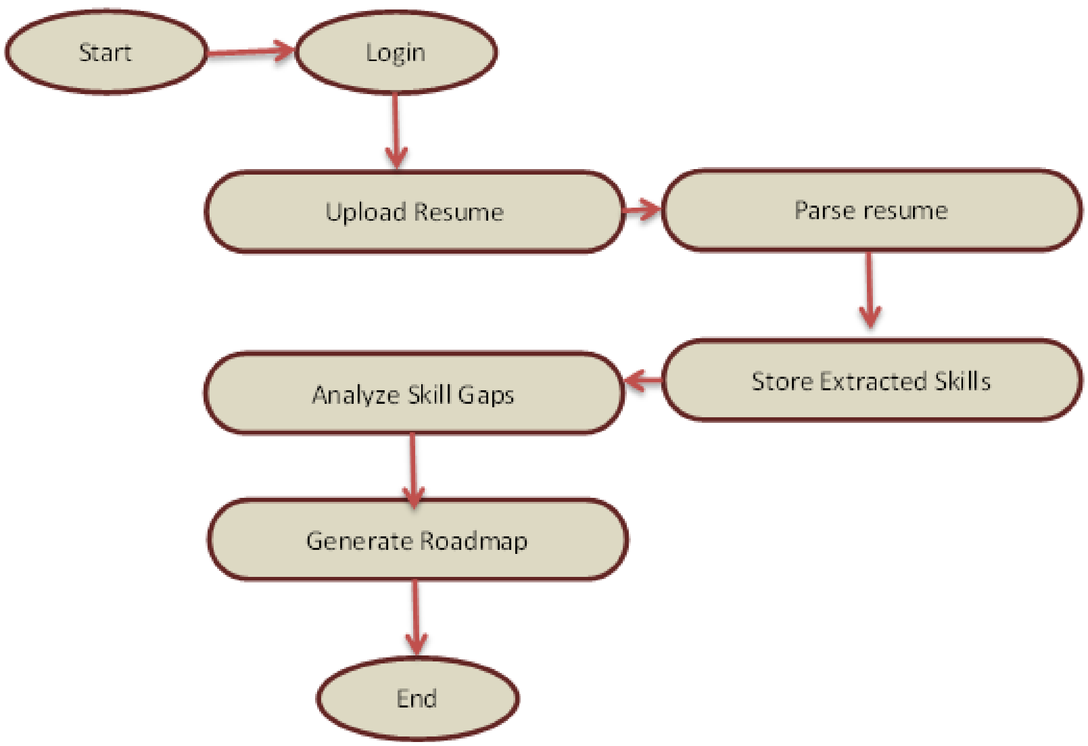

Fig:1 Resume processing Work Flow

6. Conclusion

The "Skill Gap Analysis Using Machine Learning" project offers a transformative approach to helping students upskill by identifying the gaps between their current skill set and the industry's requirements. Through the integration of Natural Language Processing (NLP) and machine learning algorithms, the project analyzes resumes, career interests, and job descriptions to assess the skills that students already possess and highlight the areas that need improvement. By doing so, it provides a comprehensive and personalized analysis of an individual's strengths and weaknesses in relation to their future career aspirations.

One of the key features of the system is its ability to offer tailored recommendations. After identifying the skills gaps, the project that can help students fill those gaps. Furthermore, it provides personalized development roadmaps, guiding students step-by-step through the process of skill enhancement. Students can prioritize their efforts effectively, focusing on the most critical skills that will make them more competitive in the job market. suggests targeted learning resources such as online courses, tutorials, articles, and practice exercises. In addition to the personalized learning resources, the project also emphasizes the importance of practical application. By offering students access to simulated interview questions, mock assessments, and real-world projects, it bridges the gap between theoretical knowledge and practical experience. This approach not only boosts the students' confidence but also equips them with the hands-on experience needed to demonstrate their expertise to potential employers. It ensures that students do not just learn new skills but also apply them in meaningful ways. Ultimately, the "Skill Gap Analysis Using Machine Learning" project serves as a comprehensive tool for students to strategically upskill, thus enhancing their employability. By focusing on the specific needs of individual students and providing a roadmap for improvement, the system fosters continuous learning and professional development.

References

1. Manning, C. D., & Schütze, H. (1999). Foundations of Statistical Natural Language Processing. MIT Press.
2. Chollet, F. (2017). Deep Learning with Python. Manning Publications.
3. Bengio, Y., et al. (2009). Learning Deep Architectures for AI. Foundations and Trends in Machine Learning.
4. Vaswani, A., et al. (2017). Attention is All You Need. Proceedings of NeurIPS.
5. Liddy, E. D. (2001). Natural Language Processing: A Survey of the State of the Art.
6. Parker, D. H., & Koller, D. (2020). Machine Learning for Skill Gap Analysis. Journal of Career Development.
7. Joulin, A., et al. (2017). Bag of Tricks for Efficient Text Classification. arXiv.
8. Joachims, T. (2002). Optimizing Search Engines Using Clickthrough Data. Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
9. Xia, M., & Wang, X. (2019). Job-Skill Matching Using Machine Learning. IEEE Transactions on Human-Machine Systems.
10. Joachims, T. (2002). Optimizing Search Engines Using Clickthrough Data. Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
11. Xia, M., & Wang, X. (2019). Job-Skill Matching Using Machine Learning. IEEE Transactions on Human-Machine Systems.
12. Bolton, M. R., & Bernhardt, D. (2018). Personalized Career Pathways and Skill Development via Machine Learning. International Journal of Career Development.
13. Liu, X., et al. (2020). An Empirical Study on Skill Gap Analysis and Prediction Using Machine Learning. International Journal of Data Science and Analytics.
14. Deng, L., & Liu, Y. (2018). Deep Learning for Natural Language Processing. Springer.
15. Ng, A. Y. (2017). Machine Learning Yearning. Deeplearning.ai.
16. Zhang, H., & Zhao, L. (2019). Machine Learning for Career Development: Applications and Trends. Journal of Career Assessment.
17. Kuhn, M., & Johnson, K. (2013). Applied Predictive Modeling. Springer.
18. Pan, Y., & Yang, Q. (2010). A Survey on Machine Learning in Job Recommendation Systems. ACM Computing Surveys.