Skill Gap Analysis Using Machine Learning

prof. M.Sujitha
Asst.Professor CSE
DR.M.G.R Educational
and Reasearch Institute
Chennai,India

Dr.T.V.Ananthan
HOD-CSE-DS(AI)
DR.M.G.R Educational
and Reasearch Institute
Chennai,India

G.Lakshmi
Student-CSE-DS(AI)
DR.M.G.R Educational
and Reasearch Institute
Chennai,India

A.Ramya teja sree
Student-CSE-DS(AI)
DR.M.G.R Educational
and Reasearch Institute
Chennai,India

G.Sahithi
Student-CSE-DS(AI)
DR.M.G.R Educational
and Reasearch Institute
Chennai,India

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,

Career Domains, Workforce Development, Skill Recommendations, Employee Development, Data- Driven Insights, Job Market Alignment, User-Centric Design, Actionable Insights, Predictive Analytics, Skill Matching, Resource Optimization.

I. 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.

II. RELATED WORK

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. **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

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.

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.

4. **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.

III. EXISTING SYSTEM

1. **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.

2. **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..

3. **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

3. **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.

4. **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.

5. **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.

IV. PROPOSED SYSTEM

1. **Resume Analysis and Skill Extraction:**

- **Feature:** The system automatically extracts skills from resumes using Natural Language Processing (NLP).
- **Overcome:** Parsing resumes with different formats, such as PDFs and DOCs, was a challenge.
- **Technologies Used:** NLP techniques, libraries like Spacy and NLTK for text processing, and PDF parsing libraries.
- **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.

2. **Career Interest Mapping:**

- **Feature:** The system maps user career interests to suggest the most relevant skillset for that career.
- **Overcome:** Aligning diverse career interests with precise skill categories.
- **Technologies Used:** Machine learning algorithms, including clustering techniques, to categorize careers and identify skill sets.
- **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.

3. **Skill Gap Identification:**

- **Feature:** Identifying the gap between the skills listed in a resume and the skills required for a specific career.
- **Overcome:** Ensuring the skill gap analysis accounts for both hard and soft skills and matches them against a comprehensive database.
- **Technologies Used:** 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.

4. **Personalized Development Roadmap:**

- **Feature:** Generate a roadmap for users, suggesting relevant courses, resources, and interview questions to bridge the skill gap.
- **Overcome:** Suggesting accurate, up-to-date, and career-relevant resources.
- **Technologies Used:** Web scraping for course and resource recommendations, recommendation engines.
- **Implementation:** Integration with online learning platforms and APIs (like Coursera, LinkedIn Learning) to fetch recommendations for skill development resources.

5. Skill Development Tracking:

- **Feature:** Track users' progress in acquiring new skills and adapting to industry demands.
- **Overcome:** Designing a tracking mechanism that accurately reflects user improvement.
- **Technologies Used:** Databases for tracking progress, APIs for updates on skill certifications.
- **Implementation:** Users input their learning progress, which is then recorded in a database. Periodic assessments or quizzes may be implemented to evaluate skill proficiency.

Challenges Overcome:

- Data Inconsistencies**:** Handling varied resume formats (PDF, Word, HTML) and inconsistent data structure across resumes.
- Data Privacy: Ensuring that user data (e.g., resumes and personal details) is securely handled.

Technologies Used:

1. NLP Libraries**:** Spacy, NLTK, transformers (for extracting skills from resumes).
2. Machine Learning Models**:** Scikit-learn, TensorFlow, for classification and gap analysis.
3. Database: MySQL or PostgreSQL for storing skill and career data.
4. Web Scraping**:** BeautifulSoup, Scrapy, and APIs for fetching online resources and job trends.
5. Python Libraries: PyPDF2, python-docx for text extraction from resumes, Flask or Django for web interface.

Implementation:

- Frontend: Developed using HTML, CSS, and JavaScript to create an intuitive user interface where users can upload resumes, input career interests, and view their skill gap analysis.
- Backend: Python was used for backend development, integrating machine learning models, data processing, and database management.
- Deployment**:** Deployed as a web application using Flask or Django, hosted on platforms like AWS or Heroku.
- Testing: Unit testing with Python's unittest module and manual testing for user interaction scenarios.

V. METHODOLOGY

1. Data Collection:

Data Sources:

- Resumes: Collect resumes from users in different formats (PDF, Word, TXT).
- Career Data: Gather information about various job roles, skills required for those roles, and industry demands.
- Skill and Course Data: Collect data from platforms like Coursera, LinkedIn Learning, etc., for relevant courses and learning materials.

Data Preprocessing:

- Standardize resume data by extracting text content from PDFs or DOCX files.
- Clean and preprocess the text data by removing unnecessary formatting, special characters, and stop words.
- Use techniques like stemming, lemmatization, and tokenization to extract meaningful features from the text.

2. Feature Extraction (Resume Analysis):

Skills Extraction Using NLP:

- Implement NLP techniques like Named Entity Recognition (NER) to identify skills mentioned in resumes.
- Use pre-trained models (e.g., spaCy, BERT) to extract technical and soft skills listed in the resumes.

Career Interest Mapping:

- Users provide their career interests or job titles.
- The system maps these interests to a database of required skills for those careers.

3. Skill Gap Identification:

Comparison of Extracted Skills with Job Requirements:

- Compare the skills extracted from resumes with the required skills for specific job roles or career aspirations using similarity algorithms like cosine similarity or Jaccard similarity.
- Identify missing skills by calculating the difference between the skills present in the resume and those required by the user's chosen career.

Skills Categorization:

- Skills are categorized into hard and soft skills, technical vs. non-technical, and domain-specific skills based on the user's career interest.

4. Machine Learning Models for Gap Analysis: Model

Training:

- Train machine learning models using labeled data (e.g., resumes with known skill gaps) to predict skill gaps.
- Use classification models (e.g., Random Forest, Support Vector Machine, Decision Trees) to classify whether a particular skill gap exists in a user's profile.
- Implement regression models if predicting the level of proficiency or expertise required for each skill is needed.

Evaluation Metrics:

- Evaluate the performance of machine learning models using metrics like accuracy, precision, recall, and F1-score for classification tasks.
- Use mean squared error (MSE) for regression models that predict proficiency levels.

5. Personalized Development Roadmap Generation:

Roadmap Creation:

- Based on the identified skill gaps, the system generates a personalized learning plan that includes:
  - Suggested courses, books, and certifications.
  - Recommended practice problems, tutorials, and other resources.
  - Interview questions related to the missing skills.

Recommendation Engine:

- Use a recommendation engine to suggest resources from online platforms like Coursera, LinkedIn Learning, or edX.
- Web scraping and API calls are used to fetch updated data on available courses and learning paths.

6. User Interaction and Feedback:

User Interface:

- Design a user-friendly interface for users to upload resumes, input career interests, and view skill gap analysis results.
- Use front-end technologies like HTML, CSS, and JavaScript to create a responsive web interface.

Feedback Mechanism:

- Allow users to provide feedback on the accuracy of the gap analysis and suggest improvements to the roadmap.
- Use the feedback to refine the recommendation engine and improve the system's performance over time.

7. System Deployment:

Backend Development:

- Develop the backend using Python (Flask or Django) to process user inputs, handle database interactions, and manage machine learning models.

Web Deployment:

- Deploy the system on a cloud platform (e.g., AWS, Heroku, or Azure) to make it accessible to users.
- Implement APIs to integrate with third-party platforms for fetching course recommendations and job market data.

8. Testing and Evaluation:

- Unit Testing**:** Test individual components such as resume parsing, skill extraction, and gap analysis using unit tests.
- Integration Testing**:** Test the full system workflow, from resume upload to roadmap generation.
- User Testing: Conduct user testing with a sample group of students or job seekers to ensure the system is user-friendly and meets their needs.
- Performance Testing: Ensure the system handles large datasets and provides accurate results within an acceptable timeframe.

9. Continuous Improvement:

- Based on user feedback and system performance, make iterative improvements.
- Continuously update the skill databases and integrate new learning resources to keep the system relevant.

VI.ARCHITECTURE

1. Data Flow Diagram

1. User uploads resume and selects a domain.

2. The system parses resumes to extract skills and qualifications.

3. Extracted data is compared with a predefined database of industry-specific skills.

4. The system generates a skill gap analysis report with recommendations.

KEY FORMULAS:

1. Skill Gap Index (SGI):

SGI = frac{text{Required Skills - Acquired Skills}}{text{Required Skills}} times 100

(Equation 1)

2. Proficiency Match Rate (PMR):

PMR =frac{text{Skills Matched}}{text{Total Required Skills}} times 100

(Equation 2)

1. 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.

Formula:

$$TF(t,d) = \frac{number\ of\ times\ t\ appears\ in\ d}{total\ number\ of\ terms\ in\ d}$$

$$IDF(t) = log\frac{N}{1+df}$$

$$TF-IDF(t,d) = TF(t,d) * IDF(t)$$

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

2. 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.

Formula:

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

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

3. 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.

Formula:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2.$$

4. Logistic Regression (for Skill Gap Prediction)

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

Formula:

$$P(y = 1 \mid X) = 1 + \exp(-(w0 + w1X1 + w2X2 + \cdots + wnXn))\,1$$

5. Support Vector Machines (SVM) for Classification

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

Formula:

$$w, bmin21 \parallel w \parallel 2 \&$$

$$yi(w \cdot xi + b) \geq 1, \forall i$$

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.

6. 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.

7. Error Metrics (for Model Evaluation)

To evaluate the performance of your models, you'll need to use standard metrics like:

- **Accuracy:**
  Accuracy = Total Predictions/ Correct Predictions
- **Precision:**

$$\text{Precision} = TP + FPT$$

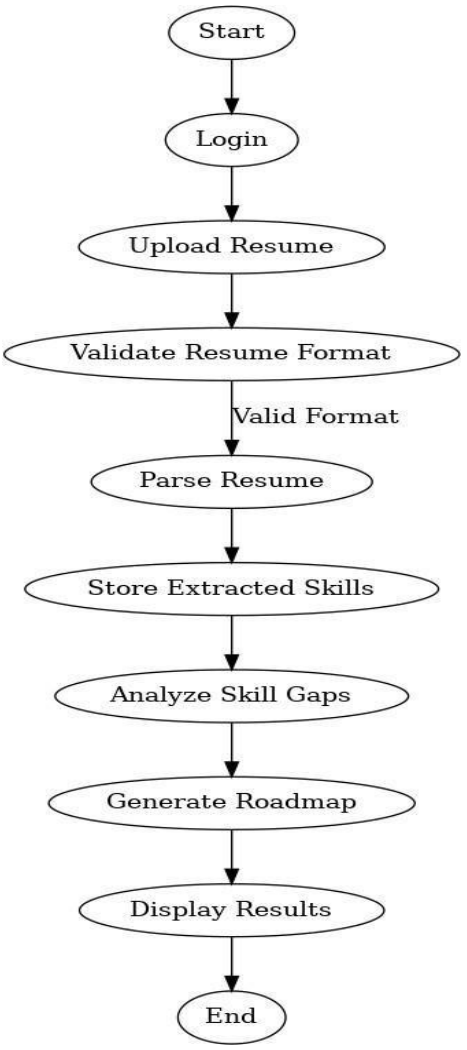

Fig:1

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 analysed 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.

VII.CONCLUSION

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.

IX.REFERENCES

1. **Manning, C. D., & Schütze, H.** (1999). Foundations of Statistical Natural Language Processing. MIT Press.
- Provides foundational concepts and techniques in NLP that are crucial for your project's text analysis.

2. **Chollet, F**. (2017). Deep Learning with Python. Manning Publications.
- A resource for understanding deep learning, which may be useful for advanced NLP models used in skill gap analysis.

3. **Bengio, Y.,** et al. (2009). Learning Deep Architectures for AI. Foundations and Trends in Machine Learning.
- Discusses deep learning, which could enhance your NLP models for resume and job description analysis.

4. **Vaswani, A.,** et al. (2017). Attention is All You Need. Proceedings of NeurIPS.
- Introduces the Transformer architecture, which is central to modern NLP techniques.

5. **Liddy, E. D**. (2001). Natural Language Processing: A Survey of the State of the Art.
- Offers a survey of NLP methods, which can guide your text extraction and processing methods.

6. **Parker, D. H., & Koller, D**. (2020). Machine Learning for Skill Gap Analysis. Journal of Career Development.
- Focuses on the application of machine learning for skill gap identification in various industries.

7. **Joulin, A., et** al. (2017). Bag of Tricks for Efficient Text Classification. arXiv.
- Relevant for implementing text classification techniques, which will help classify skills in resumes and job descriptions.

8. **Joachims, T**. (2002). Optimizing Search Engines Using Clickthrough Data. Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

• Covers algorithms for feature extraction, helpful in determining which skills are relevant from resumes and job descriptions.

9. **Xia, M., & Wang, X**. (2019). Job-Skill Matching Using Machine Learning. IEEE Transactions on Human-Machine Systems.

• Discusses the application of machine learning in aligning job descriptions with candidate skills, which is directly applicable to your project.

10. **Joachims, T**. (2002). Optimizing Search Engines Using Clickthrough Data. Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

• Covers algorithms for feature extraction, helpful in determining which skills are relevant from resumes and job descriptions.

11. **Xia, M., & Wang, X**. (2019). Job-Skill Matching Using Machine Learning. IEEE Transactions on Human-Machine Systems.

• Discusses the application of machine learning in aligning job descriptions with candidate skills, which is directly applicable to your project.

12. **Bolton, M. R., & Bernhardt, D**. (2018). Personalized Career Pathways and Skill Development via Machine Learning. International Journal of Career Development.

• Explores how personalized skill development pathways can be suggested through machine learning, aligned with your project's goal.

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.

• Provides practical examples and case studies on how machine learning can be used for skill gap prediction.

14. **Deng, L., & Liu, Y**. (2018). Deep Learning for Natural Language Processing. Springer.

o Deep learning approaches for NLP that could improve the accuracy of skill detection in resumes.

15. **Ng, A. Y**. (2017). Machine Learning Yearning. Deeplearning.ai.

o A comprehensive guide to building AI systems, including applications for analyzing skill gaps in various fields.

16. **Zhang, H., & Zhao, L**. (2019). Machine Learning for Career Development: Applications and Trends. Journal of Career
Assessment.

o Focuses on machine learning applications in career assessment, which aligns with your project on analyzing students' skill sets.

17. **Kuhn, M., & Johnson, K**. (2013). Applied Predictive
Modeling. Springer.

o Provides essential techniques and practices in predictive modeling, relevant for predicting skill gaps in students' profiles.

18. **Pan, Y., & Yang, Q**. (2010). A Survey on Machine Learning in Job Recommendation Systems. ACM Computing Surveys.

o Discusses algorithms for job matching, which are central to identifying gaps in skills based on job requirements.

These references will help you understand the theoretical and technical aspects of skill gap analysis using machine learning.