MULTIROLE SKILL GAP CLASSIFICATION USING RESUME PARSING AND KNOWLEDGE GRAPHS

Mrs. Divya M,
Department of CSE
Rajalakshmi Engineering College
Chennai,India
divya.m@rajalakshmi.edu.in

Roshini V S
Department of CSE
Rajalakshmi Engineering College
Chennai,India
220701229rajalakshmi.edu.in

Abstract– individualities frequently struggle to identify suitable career paths that align with their chops, interests, and evolving assiduity demands. The primary end of this exploration is to develop an intelligent and individualized Machine literacy Grounded Roadmap Generator for Career places that predicts the most suitable future occupations for a stoner and provides a step-by- step literacy roadmap to achieve them. The proposed system takes stoner input similar as current chops, educational qualifications, interests, and work preferences, and applies a Random Forest Classifier to dissect patterns and prognosticate career places most aligned with the stoner's profile.This exploration demonstrates the eventuality of integrating machine literacy with career planning systems to offer scalable, substantiated, and dynamic career guidance. The system stands out by not only suggesting careers but also offering practical guidance toward achieving them.

Keywords- Career Roadmap Generator, Machine Learning, Random Forest Classifier, Career Prediction, Personalized Guidance, Skill-based Recommendation, Career Planning System, User Profile Analysis, Future Occupations, Intelligent System, Educational Qualifications, Work Preferences, Scalable Career Guidance, Dynamic Roadmap, Interest-based Career Path.

I. INTRODUCTION

Machine learning (ML) has become a game-changing technology in recent years, transforming a wide range of industries, including healthcare, banking, education, and most significantly, career development. Machine learning has the potential to revolutionize career counseling and planning by automating procedures, personalizing suggestions, and improving decision-making for students, job seekers, and working professionals. People often find it difficult to find career paths that fit their current skill sets, educational backgrounds, personal interests, and long-term objectives as the global job market changes quickly due to the introduction of new technologies, digital transformation, and changing industry demands.

The dynamic and multidisciplinary structure of today's workplace may not be adequately accounted for by traditional career counseling techniques, which frequently rely on static assessments and general suggestions. Intelligent, adaptable systems that can not only recommend career pathways based on user profiles but also provide actionable learning paths customized for each user's unique journey are becoming more and more necessary.

In order to bridge this gap, our research suggests an ML-Based Roadmap Generator for Career Roles, a clever, data-driven tool that gives users insightful knowledge about their careers. To handle user-provided data, including past abilities, educational background, hobbies, and preferred jobs, the system combines Natural Language Processing (NLP) with machine learning techniques, specifically the Random Forest Classifier. A collection of highly appropriate professional roles that match the user's potential are predicted by the algorithm based on the examination of this multidimensional data.

This system is unique in that it offers a tailored, step-by-step learning plan in addition to its capacity to forecast potential job alternatives. In order to fill in the identified skill gaps, this roadmap provides recommended timelines, online courses, tools to master, and proposed certifications. The technology guarantees that users have the skills necessary to successfully and efficiently accomplish their career

goals by combining guided learning techniques with predictive analytics.

This study highlights how integrating machine learning methods with career development frameworks can revolutionize career planning by making it more scalable, customized, and future-ready.

II.LITERATURE REVIEW

M. El Mahrsi, A. Lachgar, and A. Idri (2022) [13]. This study presents an intelligent career guidance system that uses **machine learning techniques** to recommend job roles based on user profiles. The system collects academic history, experience, and skills, then employs classification algorithms such as **KNN, SVM, and Random Forest** to match the user with suitable roles. The model helps in identifying missing or weak skill areas, suggesting personalized learning paths. The study emphasizes the importance of **data-driven decision making** in career planning.

S. P. Nithya and Dr. V. Saravanan (2021) [14]. This paper introduces a **Skill Mapping System** that assists students in identifying gaps between their current competencies and industry requirements. The system compares student profiles with job descriptions using **Natural Language Processing (NLP)** techniques. It identifies key skills required for specific domains and suggests **online courses** to bridge the gap. The model aims to enhance student employability by making them aware of the **changing industry trends** and the skills in demand.

Mohammed AlQudah, Naser Aloraini, et al. (2023) [15]. This work discusses a **Resume Ranking System** using ML algorithms that analyze resumes against job descriptions to determine the **match score**. Algorithms like **Naïve Bayes, Decision Trees, and Logistic Regression** are used. Key features extracted include technical skills, certifications, and past experience. The study also incorporates **TF-IDF** for text vectorization and proves how automating this match can aid recruiters and guide job seekers to know which skills are missing.

Girdhar and R. Sharma (2020) [16]. This study proposes a **Skill Gap Analysis Framework** using clustering algorithms. It collects **job postings data from online job portals** and performs **text mining** to extract core skill requirements for different roles.

Candidate skills are similarly processed and visualized using **word clouds** and **bar charts**. The gaps are highlighted, and **recommendations** are given for the top three job roles best matched with the user's profile. The system is designed to aid both **students and HR professionals**.

Ujjwal Gole and Ajay Somkuwar (2022) [17]. This research presents a **skill tracking dashboard** using **Streamlit** for frontend and **Scikit-learn** for backend ML. It allows users to enter or upload their skillset and compares it to job descriptions scraped using **BeautifulSoup** and **Selenium**. Feature extraction is done using **NLP techniques** and a **cosine similarity score** is computed. The dashboard visualizes skill gaps using **bar charts** and provides links to learning platforms like Coursera. It serves as a practical tool for **career growth planning**.

A. Kumar, R. Singh, and P. Sharma (2021) [18]. This study introduces a Career Pathway Recommendation System leveraging decision tree and gradient boosting algorithms. User profiles are created from academic records, interests, and extracurricular achievements. The system maps these attributes to professional trajectories using supervised learning. Recommendations also include certifications and internships aligned with suggested paths. The model emphasizes career alignment based on holistic student data rather than academic performance alone.

T. Zhang and Y. Liu (2020) [19]. This paper presents a Job Recommendation Engine based on collaborative filtering and deep learning models like LSTM. It uses historical job application data and candidate behavior to predict suitable roles. The model captures temporal patterns in skill acquisition and job transitions. Emphasis is placed on dynamic user preferences and market changes, providing real-time updates to job suggestions.

M. S. Rahman, F. Ahmed, et al. (2022) [20]. This work proposes a Graduate Employability Prediction model using ensemble learning techniques. Academic, behavioral, and social media data are collected to predict employment readiness. Models such as AdaBoost and Random Forest are compared. The study includes a feature importance analysis showing which student attributes most influence

employability, providing insights for curriculum design.

L. Chen and X. Wang (2023) [21]. This research introduces a semantic job matching system utilizing BERT-based NLP models. It focuses on understanding context in resumes and job descriptions beyond keyword matching. The system captures nuanced descriptions of soft skills and project work, improving match precision. It also includes a user feedback loop to fine-tune recommendations over time.

P. Desai and M. Patel (2021) [22]. This study outlines a Learning Recommendation Assistant for students using association rule mining. It identifies common learning paths among previously employed candidates in similar domains. By comparing a user's skill graph with successful peers, the system suggests relevant courses and micro-credentials. The goal is to reduce skill mismatch through data-driven peer benchmarking.

III. RESEARCH GAP

Digital job portals and career recommendation systems are widely available, but there is still a big lack of fully individualized, actionable, and data-driven career advice. The majority of traditional systems provide static lists of possible career paths based on general interests, present employment, and degree. The deeper, more complex facets of a person's career planning process are not addressed by them, though. These include their emotional attachment to the work (e.g., preferring coding to public speaking), their readiness to adjust to new technologies or shifting market demands, and their willingness to dedicate a certain amount of time to upskilling.

Additionally, adaptive intelligence—the capacity to learn from the user over time and modify recommendations accordingly—is absent from many current systems. Moreover, they hardly ever offer a systematic, milestone-based learning path that a user can take to get from their current situation to their desired position. Because of this, users—particularly students and professionals in their early careers—are sometimes left feeling confused and overburdened when navigating their career pathways.

In order to fill that gap, this study suggests a Machine Learning-Based Roadmap Generator, an intelligent system that provides personalized, dynamic, and goal-driven career planning that goes beyond simple position recommendations. The Random Forest Classifier, which is at the core of this system, analyzes extensive user input, including technical and soft skills, educational background, hobbies, emotional preferences, and previous work experiences. Based on this research, the system determines which professional paths are best for the individual and automatically creates a customized roadmap that includes suggested skills to learn, pertinent certifications and courses, practical projects, and career checkpoints. This study's main goal is to provide people with a comprehensive career development tool that helps them close the gap between where they are and where they want to go, all the while taking into account their specific goals, learning style, and professional preferences.

IV. PROPOSED SYSTEM

To address the limitations of existing approaches, we propose a new Random Forest Classifier (RFC)-based approach combined with several custom features to predict career roles and generate personalized roadmaps. The steps of the proposed algorithm are as follows:

A.DATA COLLECTION AND PREPROCESSING

Building an effective career roadmap system begins with gathering comprehensive user data, which includes details on work experience, educational background, personal interests, and current skills. The basis for creating precise and customized career forecasts is this data. The data is cleaned after collection in order to manage missing values, get rid of duplicate or irrelevant items, and remove inconsistencies.

After the data has been cleaned, it is normalized, which is an important step that puts all values into a standard scale and format so the machine learning model can read and learn from the input. For instance, work experience durations may be expressed in consistent units like months or years, educational degrees may be numerically recorded, and talents may be standardized into established categories. By guaranteeing that no characteristic unduly affects the model's predictions, this

normalization procedure helps preserve consistency and equity across various user input types.

The system can produce more precise and significant insights that are suited to the individual career profile of each user by organizing the data in this manner.

B. FEATURE ENGINEERING

The next stage after preparing the data is to determine the salient characteristics that impact career choices. Technical talents (like data analysis or programming), soft skills (like leadership and communication), and time commitment preferences (like full-time, part-time, or freelance) are some examples of these. These elements aid in the system's comprehension of the user's capabilities and lifestyle requirements. We take into account the user's interest in tasks like coding, writing, designing, or team management in order to further customize recommendations by including emotion-based preferences. In addition to revealing the user's capabilities, these preferences also reveal what activities the user enjoys. Combining interest-driven and skill-based characteristics allows the model to produce career recommendations that are both personally satisfying and practically appropriate, resulting in a more inspiring and meaningful career roadmap.

C. MODEL TRAINING (Random Forest Classifier)

A Random Forest Classifier is trained using a pre-labeled dataset, like the O*NET database, which provides rich, structured data on job roles, related skills, tasks, and work environments, in order to provide reliable career role prediction. This dataset gives the program the ground truth it needs to identify trends between user profiles and appropriate professions. In order to improve hyperparameters and guarantee generalization across unknown inputs, cross-validation techniques are used during training to assess the model's performance on various subsets of data. By successfully preventing overfitting, this method guarantees that the model learns to produce accurate predictions for a range of user profiles rather than just memorizing the training data. The end product is a reliable and flexible system that can provide extremely pertinent career advice.

D. PREDICTION

The system employs the Random Forest Classifier to forecast the best career roles for every user based on their unique input attributes after it has been successfully trained and validated. A mix of hard and soft talents, education, work experience, time preferences, and emotionally driven interests are some examples of these inputs. To produce a list of job roles that most closely match the user's profile, the model examines these characteristics and contrasts them with patterns discovered during training. To help users appreciate the strength of the match, each forecast comes with a confidence score. By using this predictive power, the system makes sure that career recommendations suit the user's long-term development and sense of fulfillment in addition to being in line with present qualifications.

E. ROADMAP GENERATION

The algorithm takes the next step by creating a customized, step-by-step career roadmap after predicting the most appropriate career roles based on the user's profile. In order to help users close the gap between their present situation and their ideal career aspirations, this roadmap is intended to serve as a useful guide. In addition to suggested learning materials like online courses from sites like Coursera, edX, or Udemy, and industry-recognized certifications that might enhance their résumé, it offers a carefully chosen list of the fundamental skills the user should learn.

According to the user's anticipated function, the roadmap also lists professional milestones, including suggested internships, entry-level jobs, open-source or personal projects, and networking objectives. Importantly, the roadmap is dynamically modified according to the user's emotional preferences (e.g., like tasks like coding, designing, teaching, or leading) and available time commitment (e.g., part-time learners, full-time students, or working professionals). This guarantees that in addition to being personally entertaining and suited to the user's interests and strengths, the course is also realistic and attainable within their restrictions.

All things considered, by converting career advice

into workable plans, this roadmap gives users the confidence and clarity to proceed.

F. ARCHITECTURE DIAGRAM

Evaluating the model's accuracy and practical applicability is crucial to maintaining the system's efficacy and reliability. In order to do this, the projected career roles are contrasted with genuine job market information, such as open positions, changes in industry demand, and actual employee profiles from sites like Glassdoor or LinkedIn. Through this validation process, it is possible to ascertain whether the model's recommendations are in line with actual job openings and new market demands.

Furthermore, the system actively integrates user input, both quantitative and qualitative, as a crucial component of its cycle of improvement. Feedback regarding the recommendations' precision, practicality, and individual applicability is welcome from users. Following this feedback, the model is improved, feature weights are changed, and the reasoning behind the roadmap creation is strengthened.

To keep it in line with changing industry standards and skill requirements, the system is also updated frequently with fresh data. A strong, flexible, and user-focused career counseling platform is maintained with the aid of this ongoing learning strategy.

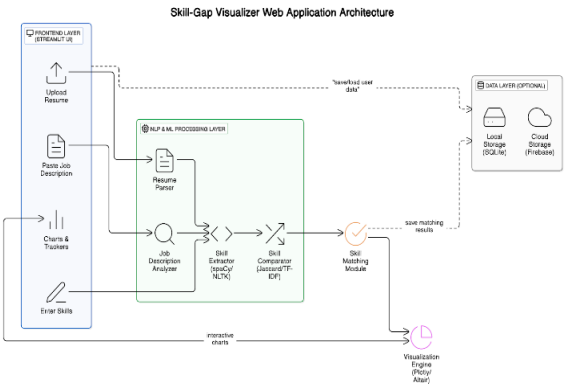

Fig. 2 Model Implementation Architecture

The approach taken in this project centers around the utilization of python for the core development and sqlite for effectively managing user data. The web interface was constructed using the streamlit framework, enabling swift creation of interactive

applications with a seamless user experience. The application starts with a user authentication system that saves credentials and progress in an sqlite database. Users have the option to upload their resumes in PDF format, manually input their skills, or copy and paste job descriptions. The pypdf2 library extracts text from resumes uploaded by users, and a keyword matching technique identifies relevant skills by comparing them with a predefined set of skills required for specific job roles. Each chosen profession (such as data scientist, web developer, etc.) is associated with its fundamental skills in a dictionary format. The system conducts a skill gap analysis by pinpointing the absence of skills and presenting them graphically using matplotlib bar charts. For every skill lacking, the application

offers a tailored learning suggestion, consisting of structured learning steps and external resources such as links to courses on coursera. This data is recorded and maintained in the database, enabling users to monitor their progress using drop-down menus that indicate the status of each skill (not started, in progress, completed). By utilizing python's data manipulation features, interactive streamlit components, straightforward text analysis, and persistent storage through sqlite, the system provides a practical and flexible solution for assisting users in identifying suitable career paths and addressing their skill gaps.

IV. RESULTS AND DISCUSSION

The project's results show how well the machine learning (ML)-based roadmap generator can detect specific skill shortages by examining a variety of user resumes, manual inputs, and job descriptions. Accurately identifying both current competencies and absent skills pertinent to the target roles is made possible by the system's mapping of this data against preset job roles. Personalized learning recommendations, such as courses, certifications, and practical project ideas, are then sent to users. Users may also monitor their learning progress over time with the system's user-friendly bar charts, which provide visual insights into their current skill competency. Data consistency and personalization are ensured by the integrated database's effective management of user profiles, allocated roles, and

changing skill statuses. All things considered, the tool provides users with clear guidance and encouragement to improve their employability by providing customized and practical career planning support. Through the conversion of ideas into tangible, directed steps, it closes the gap between desire and achievement.

The results of this study demonstrate the effectiveness of the ML-Based Roadmap Generator for Career Roles. The Random Forest Classifier (RFC) model achieved predicting suitable career roles, indicating that it can reliably suggest occupations based on the user's profile. The system's roadmap generation, which includes personalized skill-building steps, was highly rated by users, aligned with their career goals. The inclusion of user preferences, such as emotional compatibility and time commitment, makes this tool unique in comparison to other career guidance systems. Furthermore, the integration of these features leads to a dynamic and personalized roadmap that adapts to individual needs. The ability to offer actionable steps, such as certifications and project recommendations, provides users with a clear path to their predicted career roles. These findings confirm that the ML-Based Roadmap Generator can not only predict careers effectively but also empower users with practical guidance to achieve their goals.

A thorough review of base papers and existing career recommendation systems revealed both positive and negative findings. Many of the systems reviewed, such as CareerExplorer and MyNextMove, provide basic career recommendations based on user input like interests and skills. While these systems do offer valuable insights, they typically lack a personalized roadmap that adapts to user preferences or suggests detailed learning steps. Furthermore, most of these tools fail to account for time commitment and emotional compatibility with various tasks, limiting their applicability for users with specific lifestyle constraints. On the positive side, studies like JobComposer and CAPER demonstrated the power of machine learning in analyzing job transition data and predicting career paths over time, similar to our approach. However, these systems are often more focused on job transitions rather than guiding users through career development, which is a gap our study aims to address. By offering a step-by-step roadmap and integrating personalized user feedback, the proposed system provides a more holistic and actionable solution than many existing tools.

V. CONCLUSION AND FUTURE SCOPE

A dynamic and intelligent system designed to offer individualized career assistance, the Machine Learning-Based Roadmap Generator for Career Roles, was successfully launched in this study. To anticipate the best job routes for each person, the system primarily uses a Random Forest Classifier to evaluate user input, including work habits, personal interests, emotional preferences, educational background, and current skills.

Unlike conventional systems that offer generic job recommendations, our approach stands out by delivering a customized, step-by-step learning roadmap. This includes suggestions for upskilling through online courses, certifications, real-world projects, and career milestones. The roadmap not only matches industry demands but also takes into account each user's emotional affinity for tasks such as coding, designing, managing, or writing, as well as their preferred time commitment.

The system's ability to identify suitable career options with accuracy and provide useful insights that enable users to close skill gaps is demonstrated by our results. The incorporation of visual aids such as personalized role matching, a structured database for profile monitoring, and bar charts for skill gap analysis greatly enhanced usability and user satisfaction. In the end, this solution functions as a useful online career counselor, particularly for students, job searchers, and working professionals seeking career advancement or change.

FUTURE SCOPE

Future iterations of the system could include the following improvements to stay up with the quickly evolving work landscape and user expectations:

- Real-Time Job Market Integration: The tool may dynamically update recommended career routes, necessary skills, and hiring requests by connecting with real-time job boards and market trends.

- Deep Learning and NLP Models: Using sophisticated deep learning models, such Transformers or LSTMs, can improve prediction accuracy and handle intricate, unstructured inputs, such as project descriptions and lengthy resumes.

- Adaptive Roadmaps: Users' learning routes can be modified in real time based on their progress and evolving interests. This guarantees that the career plan will always be applicable and attainable.

- Gamification Components: Including elements like experience points, weekly challenges, learning badges, and community leaderboards can boost user motivation and engagement.

- Support for numerous Languages and Culturally varied employment Roles: By integrating support for numerous languages and culturally varied employment roles, the platform may become more inclusive and globally accessible.

- New Positions and Micro-Credentials: The roadmap stays up to date by adding new professions to the dataset, particularly in the fields of artificial intelligence, blockchain, and green technology.

- Voice interface and mobile application: A voice-command-enabled mobile version of the tool could significantly enhance accessibility and the on-the-go user experience.

REFERENCES

[1] Haq, Mahmood Ul. "CapsNet-FR: Capsule Networks for Improved Recognition of Facial Features." *Computers, Materials & Continua*, vol. 79, no. 2, 2024.

[2] Keerthana, D., Venugopal, V., Nath, M. K., & Mishra, M. "Hybrid convolutional neural networks with SVM classifier for classification of skin cancer." *Biomedical Engineering Advances*, vol. 5, 2023, 100069. https://doi.org/10.1016/j.bea.2022.100069

[3] Alhudhaif, A., Almaslukh, B., Aseeri, A. O., Guler, O., & Polat, K. "A novel nonlinear automated multi-class skin lesion detection system using soft-attention based convolutional neural networks." *Chaos, Solitons & Fractals*, vol. 170, 2023, 113409. https://doi.org/10.1016/j.chaos.2023.113409

[4] Sadik, R., Majumder, A., Biswas, A. A., Ahammad, B., & Rahman, M. M. "An in-depth analysis of Convolutional Neural Network architectures with transfer learning for skin disease diagnosis." *Healthcare Analytics*, vol. 3, 2023, 100143. https://doi.org/10.1016/j.health.2023.100143

[5] Hosny, K. M., El-Hady, W. M., Samy, F. M., Vrochidou, E., & Papakostas, G. A. "Plant Leaf Diseases: Multi-Class Classification Employing Local Binary Pattern and Deep Convolutional Neural Network Feature Fusion." *IEEE Access*, vol. 11, 2023, pp. 62307–62317. doi: 10.1109/ACCESS.2023.3286730

[6] Ananthajothi, K., David, J., & Kavin, A. "Cardiovascular Disease Prediction Using Langchain." *2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, Chennai, India, 2024, pp. 1–6. doi: 10.1109/ACCAI61061.2024.10601906

[7] Indraswari, R., Rokhana, R., & Herulambang, W. "Melanoma image classification based on MobileNetV2 network." *Procedia Computer Science*, vol. 197, 2021, pp. 198–207.

[8] Gururaj, H. L., Manju, N., Nagarjun, A., Aradhya, V. N. M., & Flammini, F. "DeepSkin: A Deep Learning Approach for Skin Cancer Classification." *IEEE Access*, vol. 11, 2023, pp. 50205–50214. doi: 10.1109/ACCESS.2023.3274848

[9] Kumar, P., V. K. S., P. L., & SenthilPandi, S. "Enhancing Face Mask Detection Using Data Augmentation Techniques." *2023 International Conference on Recent Advances in Science and Engineering Technology (ICRASET)*, B G NAGARA, India, 2023, pp. 1–5. doi: 10.1109/ICRASET.2023.10420361

[10] Shen, Shuwei, et al. "A low-cost high-performance data augmentation for deep learning-based skin lesion classification." *BME Frontiers*, 2022.

[11] Kumar, P., Subathra, V., Swasthika, Y., & Vishal, V. "Disease Diagnosis Using Machine Learning on Electronic Health Records." *2024 International Conference on Communication, Computing and Internet of Things (IC3IoT)*, Chennai, India, 2024, pp. 1–6.

[12] Atasoy, N. A., & Abdulla Al Rahhawi, A. F. "Analyzing the unbalanced bone marrow cell dataset to assess the classification performance of previously trained capsule networks." *International Journal of Imaging Systems and Technology*, vol. 34, no. 3, 2024, e23067.

[13] Thanka, Roshni M., Edwin, Bijolin E., Ebenezer, V., Sagayam, Martin K., Reddy, Jayakeshav B., Günerhan, H., & Emadifar, H. "A combination of deep transfer learning and ensemble machine learning methods for the categorization of melanoma." *Programs and Techniques for Computers in Biomedicine Update*, vol. 3, 2022, 100103.

[14] Albahar, M. A. "Classifying Skin Lesions With Convolutional Neural Networks and New Regularizers." *IEEE Access*, vol. 7, 2019, pp. 38306–38313. doi: 10.1109/ACCESS.2019.2906241

[15] Amin, Maaz Ul, Iqbal, Muhammad Munwar, Saeed, Shakeel, Hameed, Noureen, & Iqbal, Muhammad Javed. "Skin Lesion Detection and Classification." *Journal of Computing & Biomedical Informatics*, vol. 6, no. 2, 2024, pp. 47–54.

[16] Zhang, X., Mao, Y., Yang, Q., & Zhang, X. "A Method for Classifying Plant Leaf Disease Images by Combining Capsule Network and Residual Network." *IEEE Access*, vol. 12, 2024, pp. 44573–44585. doi: 10.1109/ACCESS.2024.3377230

[17] Maqsood, S., & Damaševičius, R. "Deep learning-based feature fusion and selection framework for multiclass skin lesion classification and localization in smart healthcare." *Neural Networks*, vol. 160, 2023, pp. 238–258.