SKILL GAP DETECTION AND JOB FITMENT VIA SELECTION INSIGHT AND REJECTION REASONING

PROJECT PHASE I REPORT

Submitted by

PRIYADARSHINI S	**2116221801040**
SANDHYA J	**2116221801044**
TANUSHRI G V S	**2116221801055**

in partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

RAJALAKSHMI ENGINEERING COLLEGE

(AUTONOMOUS), CHENNAI – 602 105

NOV 2025

BONAFIDE CERTIFICATE

Certified that this Report Titled "**SKILL GAP DETECTION AND JOB FITMENT VIA SELECTION INSIGHT AND REJECTION REASONING"** is the bonafide work of "**PRIYADARSHINI S (2116221801040), SANDHYA J (2116221801044)** and **TANUSHRI G V S (211622180155)"** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE

Dr. J.M. Gnanasekar M.E., Ph.D.,

Professor and Head,

Department of Artificial Intelligence and Data Science

Rajalakshmi Engineering College

Thandalam – 602 105

SIGNATURE

Mrs.V. Dhivya Bhavani B.Tech., M.E.,

Assistant Professor,

Department of Artificial Intelligence and Data Science

Rajalakshmi Engineering College

Thandalam – 602 105

Submitted to Project Viva-Voce Examination held on ________________________

Internal Examiner

External Examiner

DEPARTMENT VISION

To become a global leader in Artificial Intelligence and Data Science by achieving through excellence in teaching, training, and research, to serve the society.

DEPARTMENT MISSION

- To develop students' skills in innovation, problem-solving, and professionalism through the guidance of well-trained faculty.

- To encourage research activities among students and faculty members to address the evolving challenges of industry and society.

- To impart qualities such as moral and ethical values, along with a commitment to lifelong learning

PROGRAMME EDUCATIONAL OBJECTIVES (PEO’s)

PEO 1: Build a successful professional career across industry, government, and academia by leveraging technology to develop innovative solutions for real-world problems.

PEO 2: Maintain a learning mindset to continuously enhance knowledge through experience, formal education, and informal learning opportunities.

PEO 3: Demonstrate an ethical attitude while excelling in communication, management, teamwork, and leadership skills

PEO 4: Utilize engineering, problem-solving, and critical thinking skills to drive social, economic, and sustainable impact.

PROGRAMME OUTCOME (PO's)

PO1: Engineering Knowledge: Apply the knowledge of mathematics, science, engineering fundamentals and an engineering specialization to the solution of complex engineering problems.

PO2: Problem Analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

PO3: Design / Development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

PO4: Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

PO5: Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

PO6: The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

PO7: Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

PO8: Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

PO9: Individual and team work: Function effectively as an individual and as a member or leader in diverse teams, and in multidisciplinary settings.

PO10: Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

PO11: Project management and finance: Demonstrate knowledge and understanding of the engineering management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

PO12: Life-long learning: Recognize the need for and have the preparation and ability to engage in independent and lifelong learning in the broadest context of technological change

PROGRAM SPECIFIC OUTCOMES (PAOs)

A graduate of the Artificial Intelligence and Data Science Learning Program will demonstrate

PSO 1: Foundation Skills: Apply the principles of artificial intelligence and data science by leveraging problem-solving skills, inference, perception, knowledge representation, and learning techniques

PSO 2: Problem-Solving Skills: Apply engineering principles and AI models to solve real-world problems across domains, delivering cutting-edge solutions through innovative ideas and methodologies

PSO 3: Successful Progression: Utilize interdisciplinary knowledge to identify problems and develop solutions, a passion for advanced studies, innovative career pathways to evolve as an ethically responsible artificial intelligence and data science professional, with a commitment to society.

COURSE OBJECTIVE

- To identify and formulate real-world problems that can be solved using Artificial Intelligence and Data Science techniques.
- To apply theoretical and practical knowledge of AI & DS for designing innovative, data-driven solutions.
- To integrate various tools, frameworks, and algorithms to develop, test, and validate AI & DS models.
- To demonstrate effective teamwork, project management, and communication skills through collaborative project execution.
- To instill awareness of ethical, societal, and environmental considerations in the design and deployment of intelligent systems.

COURSE OUTCOME

CO 1: Analyze and define a real-world problem by identifying key challenges, project requirements and constraints.

CO 2: Conduct a thorough literature review to evaluate existing solutions, identify research gaps and formulate research questions.

CO 3: Develop a detailed project plan by defining objectives, setting timelines, and identifying key deliverables to guide the implementation process.

CO 4: Design and implement a prototype or initial model based on the proposed solution framework using appropriate AI tools and technologies.

CO 5: Demonstrate teamwork, communication, and project management skills by preparing and presenting a well-structured project proposal and initial implementation results.

CO-PO-PSO Mapping

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	3	2	2	1	2	1	1	1	2	1	2	3	2	2
CO2	2	3	2	3	2	1	1	1	2	2	1	3	2	2	2
CO3	2	2	3	2	2	1	2	2	3	2	3	2	2	3	3
CO4	3	3	3	3	3	2	2	2	2	3	2	2	3	3	3
CO5	2	2	2	1	2	2	2	3	3	3	3	2	2	2	3

Note: Correlation levels 1, 2 or 3 are as defined below:

1: Slight (Low) 2: Moderate (Medium) 3: Substantial (High)

No correlation: “-”

ABSTRACT

In today's competitive job market, aligning candidate skills with job-specific requirements is essential for employability and career growth. "Skill Gap Detection and Job Fitment via Selection Insight and Rejection Reasoning" addresses this challenge by introducing an AI-powered system that automates resume analysis, detects skill gaps, predicts job fitment, and provides rejection reasoning and interview preparation. The project utilizes Natural Language Processing (NLP) techniques such as text extraction and lemmatization using libraries like PyMuPDF and spaCy. Machine learning models, including TF-IDF vectorization and Random Forest classification, are employed to compare resumes with job role expectations. Cosine Similarity measures the alignment between candidate profiles and job roles. The system offers visually rich insights like skill match charts, job role rankings, and rejection analysis reports. Additionally, an AI Interviewer module generates personalized, role-specific questions to enhance interview readiness. This integrated solution supports students in identifying improvement areas, assists recruiters in evaluating candidates more efficiently, and enhances institutional placement efforts. Future enhancements may include personalized upskilling recommendations based on role-specific gaps.

Keywords – Person–job fit, job recommender system, skill gap analysis, hierarchical clustering, resume analysis, career re-education, learning path recommender, AI interviewer.

ACKNOWLEDGEMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavor to put forth this report. Our sincere thanks to our Chairman **Mr. S.MEGANATHAN, B.E, F.I.E.**, our Vice Chairman **Mr. ABHAY SHANKAR MEGANATHAN, B.E., M.S.,** and our respected Chairperson **Dr. (Mrs.) THANGAM MEGANATHAN**, **Ph.D.,** for providing us with the requisite infrastructure and sincere endeavoring in educating us in their premier institution.

Our sincere thanks to **Dr. S.N. MURUGESAN, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. J.M. GNANASEKAR M.E., Ph.D.,** Professor and Head of the Department of Artificial Intelligence and Data Science for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guide, **Dr.S. SURESH KUMAR M.E., Ph.D.,**Professor, Department of Artificial Intelligence and Data Science. Rajalakshmi Engineering College for her valuable guidance throughout the course of the project. We are very glad to thank our Project Coordinator, **Mrs.V. DHIVYA BHAVANI B.TECH., M.E.,** Department of Artificial Intelligence and Data Science for his useful tips during our review to build our project.

PRIYADARSHINI S	**SANDHYA J**	**TANUSHRI G V S**
(2116221801040)	(2116221801044)	(211622180155)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABBREVIATION	FULL FORM
AI	Artificial Intelligence
DS	Data Science
NLP	Natural Language Processing
ML	Machine Learning
TF-IDF	Term Frequency–Inverse Document Frequency
LLM	Large Language Model
OCR	Optical Character Recognition
BERT	Bidirectional Encoder Representations from Transformers.
RoBERTa	Robustly optimized BERT approach
GPT	Generative Pre-trained Transformer

CHAPTER I

INTRODUCTION

1.1 GENERAL

The landscape of recruitment and candidate evaluation is undergoing a transformative shift, driven by automation, data analytics, and AI-powered decision-making. In such a dynamic environment, job seekers often face difficulties in aligning their skills with job role expectations, and recruiters struggle to process large volumes of resumes manually. A well-aligned resume, backed by industry-relevant skills, is essential to improve a candidate's chances of being shortlisted. However, most candidates lack clarity on how their profiles compare with job-specific needs, or why they face rejections. This project, "Skill Gap Detection and Job Fitment via Selection Insight and Rejection Reasoning," aims to address these challenges through a multi-module AI system that automates resume evaluation, skill gap identification, job fitment analysis, and rejection reasoning. With job roles evolving rapidly and employers demanding specialized skill sets, a systematic and intelligent approach to resume screening is necessary. Manual evaluation methods are often slow, inconsistent, and biased. To address this gap, the proposed system leverages advanced Natural Language Processing (NLP) and Machine Learning (ML) techniques to analyze resumes, match candidate profiles with job requirements, and provide actionable insights. Additionally, it simulates an AI-based interview environment to further prepare candidates. This project serves as a complete solution to enhance employability, assist recruiters, and bridge the gap between candidates' current skills and industry expectations.

1.2 NEED FOR THE STUDY

The study is motivated by the noticeable gap between candidate capabilities as reflected in resumes and the actual skill requirements defined by various job roles. Often, candidates are unaware of specific skills employers prioritize, which leads to missed opportunities and unnecessary rejections. Similarly, recruiters face challenges in identifying the right talent amidst numerous applicants, especially when many resumes lack clarity or structure. Traditional screening processes are not equipped to deliver the accuracy, efficiency, or personalization needed in today's recruitment ecosystem. By introducing automation in skill analysis, fitment prediction, and rejection reasoning, this project addresses both ends of the recruitment pipeline. It not only enhances candidate awareness of their shortcomings and strengths but also assists employers in making data-driven hiring decisions. In a world where skill mismatch is a leading cause of employment challenges, this study becomes critical in fostering a better-prepared and industry-ready workforce.

1.3 OVERVIEW OF THE PROJECT

The "Skill Gap Detection and Job Fitment via Selection Insight and Rejection Reasoning" project is designed as a four-module system that offers a comprehensive approach to career readiness and talent evaluation. The first two modules focus on skill gap analysis and job role fitment prediction. Resumes are analyzed using text extraction tools such as PyMuPDF and python-docx, followed by preprocessing with spaCy. Skills are vectorized using TF-IDF and compared with job descriptions. Machine learning algorithms, particularly Random Forest, are used to predict the most suitable roles, while cosine similarity helps rank them based on relevance. The third module introduces resume rejection reasoning by comparing selected and rejected resumes for the same role, identifying missing skills and deficiencies. Finally, the fourth module acts as an AI Interviewer, generating personalized, role-specific interview questions to aid in preparation. The results across all modules are visualized using intuitive charts and graphs. Together, these modules create a unified system that empowers candidates and assists institutions in improving placement outcomes.

1.4 OBJECTIVES OF THE STUDY

1. **Skill Gap Analysis**: To identify the gaps between the candidate's current skills and the expected skills for a target job role, enabling personalized improvement.

2. **Job Fitment Prediction**: To evaluate and rank the most suitable job roles for a candidate based on resume analysis using machine learning and similarity scoring.

3. **Resume Rejection Insight**: To generate rejection reasoning by comparing selected and rejected resumes, highlighting missing skills and alignment issues.
4. **AI-Based Interview Preparation**: To simulate a role-specific interview by generating targeted questions, helping users improve their readiness and confidence.
5. **Career Empowerment**: To support job seekers through data-driven insights, structured feedback, and guided preparation strategies for better career opportunities.

CHAPTER II

REVIEW OF LITERATURE

2.1 INTRODUCTION

Understanding the intricacies of resume evaluation, skill gap identification, and job fitment prediction requires a thorough examination of existing research in the fields of human resource analytics, natural language processing (NLP), and machine learning. This literature review explores previous methodologies and tools, forming the foundation for the development of the “Skill Gap Detection and Job Fitment via Selection Insight and Rejection Reasoning” system. The proposed project aims to overcome the limitations of conventional approaches by incorporating intelligent automation for skill matching, resume comparison, rejection reasoning, and personalized career guidance. Technological advancements have significantly transformed recruitment practices, leading to a growing reliance on automation. Traditional resume screening is increasingly inadequate due to the sheer volume and complexity of applicant data. Though past efforts have led to the creation of automated systems, many lack precision, adaptability, or the ability to handle the unstructured nature of resumes. Moreover, few systems offer insights into rejection causes or simulate interview processes. This chapter delves into the frameworks and techniques that have previously been applied in resume analysis and candidate evaluation, identifying the performance gaps that the current project seeks to address through a modular and AI-enhanced approach.

2.2 FRAMEWORK OF LITERATURE REVIEW

This literature review is structured around two key focus areas: existing systems for job and skill profiling, and the use of NLP with machine learning in resume analysis.

1. Existing Systems for Job Profile Analysis: Several studies have investigated the use of automated methods to analyze job roles and identify skill requirements. Clustering algorithms such as hierarchical and k-means have been widely used to categorize job roles based on skills extracted from job postings. These models provide a generalized view of industry demands but struggle with personalization. Hierarchical clustering, for example, often requires manual tuning to define cluster granularity, which limits its scalability across diverse domains. Models like average-linkage clustering are limited in distinguishing closely related job roles with overlapping skills, leading to inaccurate candidate-role matching in real-world applications.

2. NLP and Machine Learning in Resume Processing: Recent approaches have integrated Natural Language Processing (NLP) with machine learning to improve document analysis in recruitment contexts. Techniques such as text extraction, tokenization, lemmatization, and TF-IDF vectorization are widely used for processing unstructured resume data. These features are then fed into classifiers like Random Forest to predict job fitment, while similarity measures like Cosine Similarity help in ranking candidate-job matches. Although these techniques have improved automation, many existing systems lack modularity, interpretability, and visual feedback. Additionally, very few offer capabilities like rejection reasoning or AI-assisted interview preparation. These limitations emphasize the need for a more comprehensive and student-friendly solution.

3. Limitations of Existing Techniques: Despite the adoption of modern NLP and ML techniques, key challenges remain. Many systems rely heavily on domain-specific datasets, which limits their generalization across industries. They often fail to contextualize skill relevance based on role or company expectations. Furthermore, existing tools rarely provide transparency in candidate evaluation or meaningful feedback for improvement. There is also limited support for side-by-side resume comparison or identifying rejection reasons. The proposed project aims to fill these gaps by offering a unified solution that combines skill gap detection, job fitment analysis, rejection reasoning, and AI interview preparation — all powered by explainable and efficient algorithms.

CHAPTER III

SYSTEM OVERVIEW

3.1 EXISTING SYSTEM

The existing systems for resume evaluation and job-role matching are largely dependent on keyword-based algorithms or basic clustering methods. Many commercial platforms compare resumes to job descriptions through surface-level keyword matching, resulting in numerous limitations:

- **Inaccurate Skill Matching**: These systems often misinterpret the contextual meaning of skills. For instance, a resume mentioning “Python” might be wrongly matched to data analysis roles, even if the candidate’s expertise lies in backend development or automation scripting.
- **Lack of Personalization**: Most available solutions provide generic assessments without giving user-specific recommendations or feedback. This deprives candidates of a clear understanding of how to enhance their profiles.
- **Manual Effort and Inefficiency**: In many organizations, HR professionals still conduct manual screening of resumes. Even with partial automation, the lack of intelligent filtering results in slower and often biased selection processes.
- **Minimal Use of Intelligent Algorithms**: While machine learning is common in other domains, many resume analysis systems do not leverage advanced techniques like classification or similarity scoring. Critical functions such as rejection reasoning or dynamic role fitment are generally absent.

These drawbacks highlight the need for a more intelligent, context-aware, and modular approach to resume analysis, one that incorporates AI techniques for deeper insights and improved career support.

3.2 PROPOSED SYSTEM

The proposed system, “Skill Gap Detection and Job Fitment via Selection Insight and Rejection Reasoning”, addresses the limitations of current systems by combining advanced Natural Language Processing (NLP) and Machine Learning (ML) techniques. It not only analyzes resumes and suggests appropriate job roles but also identifies missing skills, explains rejection reasons, and assists in interview readiness through an AI-based interviewer module. Key Features of the Proposed System:

1. Advanced Skill Extraction: Using PyMuPDF and python-docx for document parsing and spaCy for text preprocessing (tokenization, lemmatization, stopword removal), the system extracts meaningful skill phrases from unstructured resume text.

2. Job Role Preprocessing: A curated dataset containing job roles and corresponding required skills is transformed into TF-IDF vectors, enabling structured comparison between candidate resumes and job expectations.

3. Machine Learning-Based Fitment Analysis: Random Forest classification is used to predict the most suitable job roles based on the extracted skill set. Cosine Similarity further refines the matching by computing relevance scores between candidate and job vectors.

4. Skill Gap Visualization: The system compares candidate skills with target job requirements and highlights matched/missing skills. Pie charts and bar graphs visually represent the gaps and strengths in a user-friendly format.

5. Job Recommendations & Rejection Reasoning: Based on similarity scores, the top three best-fit roles are presented to the user. Additionally, side-by-side resume comparison (rejected vs selected) identifies reasons for rejection using skill clustering and company-specific rejection patterns.

6. AI Interviewer Module: Personalized interview questions are generated from extracted skills and job roles, simulating a role-based interview scenario to help users improve their preparedness.

The system offers a complete, modular framework that enhances the employability of job seekers while providing clear, data-driven insights.

3.3 FEASIBILITY STUDY

To ensure the successful implementation and scalability of the system, a feasibility study was conducted across three domains: technical, economic, and operational feasibility.

1. Technical Feasibility:

- The system is built using reliable libraries such as PyMuPDF (resume parsing), spaCy (NLP), and scikit-learn (ML). These tools are widely supported and suitable for text-heavy tasks.
- Algorithms like TF-IDF and Random Forest are proven to work efficiently on medium-to-large datasets. Cosine Similarity ensures fast and accurate job-role matching.
- The project is compatible with deployment on cloud platforms, making it scalable and accessible for a wide range of users.

2. Economic Feasibility:

- The project is developed using open-source libraries and frameworks, significantly reducing the initial cost of development.
- The system requires minimal investment beyond hosting and storage, which can be handled using budget-friendly cloud solutions.
- Given the wide applicability to colleges, job seekers, and HR platforms, there is strong potential for monetization through institutional licensing, premium features, or SaaS offerings.

3. Operational Feasibility:

- The user interface is designed to be simple and intuitive, allowing users to upload resumes, view insights, and receive role suggestions easily.
- The AI-generated feedback system provides real-time, personalized suggestions for upskilling, resume enhancement, and interview preparation.
- The backend supports concurrent processing, enabling smooth handling of multiple requests even under high usage scenarios.

CHAPTER IV

SYSTEM REQUIREMENTS

4.1 HARDWARE REQUIREMENTS

1. Development Machine:

- Processor: Intel Core i5 or equivalent
- RAM: 8 GB or more
- Storage: 256 GB SSD or higher

2. User Devices:

- Processor: Modern CPU for accessing the application
- RAM: 4 GB minimum
- Internet Connection: Stable internet for using the web-based application

4.2 SOFTWARE REQUIREMENTS

1. Operating System:

- Windows 10, macOS, or Linux for development

2. Programming Language:

- Python: Version 3.8 or later

3. Libraries and Frameworks:

- Natural Language Processing: spaCy
- Machine Learning: scikit-learn
- Data Handling: Pandas, NumPy
- PDF Handling: PyMuPDF (fitz)
- DOCX Handling: python-docx
- Visualization: Matplotlib
- Web Framework: Flask

4. Development Tools:

- IDE: Google colab, Visual Studio Code
- Package Manager: pip for Python package installation

CHAPTER V

SYSTEM DESIGN

5.1 SYSTEM ARCHITECTURE

The system follows a modular, step-by-step approach to analyze resumes, identify skill gaps, match candidates with job roles, and present the results in a meaningful format. Below is a detailed overview of each stage of the process:

Input Data Collection

The system requires three primary inputs:

- Resumes: Uploaded in PDF or DOCX format or fetched from an internal database.
- Job Role Dataset: Contains job role titles and their associated skill requirements, typically maintained in an Excel or structured file format.
- Skill Repository: A structured dataset mapping each job role to a set of required technical and soft skills.
- These inputs serve as the foundation for skill comparison and job role fitment analysis.

Data Preprocessing

In this stage, raw documents are converted into analyzable text:

- Text Extraction:
  - PDFs are processed using PyMuPDF for high-quality text retrieval.
  - Word documents (DOCX) are parsed using python-docx.
- The extracted text is then cleaned and prepared for further analysis by removing irrelevant characters and formatting issues.

Text Processing and Matching

After extraction, the textual data undergoes NLP and vectorization:

- Tokenization: The text is broken down into tokens (words or phrases) using spaCy, which also performs lemmatization and stopword removal.
- Vectorization: The cleaned tokens are converted into numerical representations using TF-IDF (Term Frequency–Inverse Document Frequency), allowing the system to evaluate word relevance.

- Cosine Similarity: This metric calculates the similarity between the candidate's resume and job role descriptions, identifying how closely they align in terms of skill content.

Machine Learning-Based Skill Matching

To improve the accuracy and robustness of skill matching:

- A Random Forest classifier is trained on labeled job role–skill pair data.
- The model is used to predict the most likely job roles that match the candidate's extracted skillset.
- This adds a predictive layer beyond similarity scores, enhancing the reliability of recommendations.
- The combination of Cosine Similarity and Random Forest provides both contextual matching and learned prediction accuracy.

Ranking and Visualization of Results

- Role Ranking: The system generates a ranked list of job roles, ordered by their compatibility with the candidate's skills.
- Visual Insights:

  Matched and missing skills are presented using:
  - Pie Charts to show skill distribution
  - Bar Graphs to depict top 3 matching roles
  - Rejection Reason Reports to highlight where the candidate may fall short compared to selected candidates

These visualizations offer users a clear, actionable understanding of their current position and areas for improvement.

Final Output

The result is a ranked and explainable set of job recommendations, tailored to the user's resume, supported by:

- Skill gap visualizations
- Matched job roles with relevance percentages
- Rejection reasoning (if applicable)
- AI-generated interview questions via the AI Interviewer Module

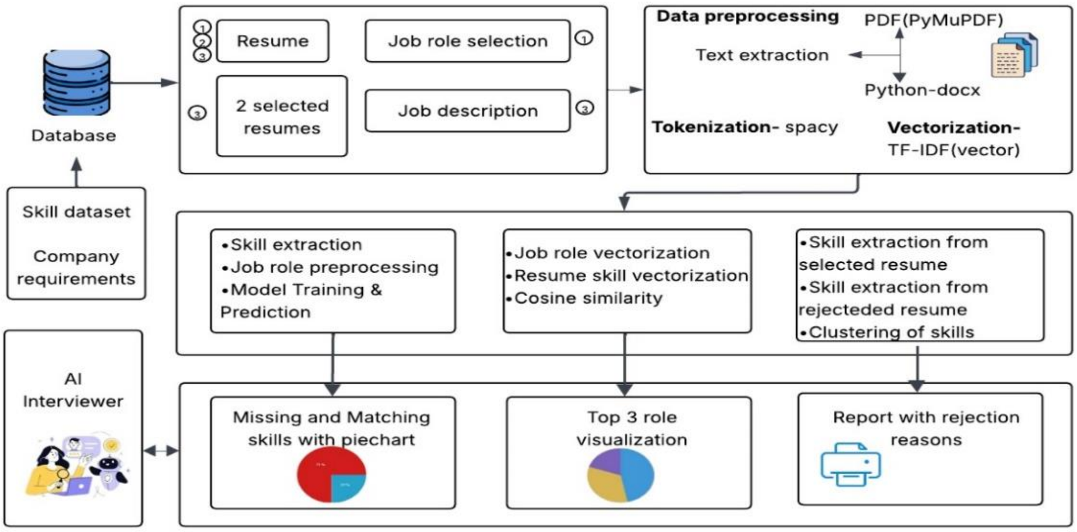

Figure 5.1: System Architecture

Figure 5.1 illustrates the workflow of an AI-driven Skill Gap Detection and Job Fitment System. It begins with a database containing resumes, skill datasets, and company job requirements. The process starts when users upload resumes and select a job role along with its description. During data preprocessing, text is extracted from resumes in PDF or DOCX format using tools like PyMuPDF and Python-docx. The extracted text is then tokenized using SpaCy and vectorized using TF-IDF, converting it into numerical feature vectors. The system performs skill extraction, job role preprocessing, model training, and prediction to identify relevant skills. Both job roles and resumes undergo vectorization, followed by cosine similarity analysis to measure the alignment between candidate skills and job requirements. Additionally, skills are extracted and clustered from rejected resumes to identify common deficiencies. The system outputs multiple insights — a pie chart visualizing missing and matching skills, a top 3 role compatibility chart, and a detailed report containing rejection reasons. The AI Interviewer module interacts with candidates, leveraging this data to provide personalized feedback and interview preparation guidance.

5.2 MODULE DESCRIPTION

5.2.1 SKILL GAP DETECTION

SKILL EXTRACTION

This module focuses on the automated extraction of relevant skills from user-submitted resumes and a predefined list of job skills. The process begins when a user uploads their resume in PDF format. The text content is extracted using PyMuPDF, resulting in readable resume text. Meanwhile, an admin uploads a DOCX file containing job-specific skill requirements. This text is processed using python-docx to extract skill keywords. Both extracted texts are then processed using spaCy for tokenization, which segments and identifies individual skill-related terms. The matched skills from the resume and job requirements are compared to identify overlaps, with the results being stored in a data store. This data informs the skill gaps between the user's profile and job expectations. The diagram visually represents this flow, highlighting inputs, text extraction methods, tokenization, and the comparison process, ensuring structured and efficient skill extraction.

JOB ROLE PREPROCESSING

The system prepares job roles for skill matching. First, both users and administrators input job roles and datasets. The system then extracts and cleans job-specific skills from the provided dataset to ensure consistency and relevance. This cleaned data serves as a base for identifying necessary technical skills for each job role. Using TF-IDF (Term Frequency-Inverse Document Frequency) vectorization, the system generates a numerical representation of the skills, known as skill vectors. These vectors capture the importance of each skill in the context of the job role, providing a structured dataset ready for machine learning tasks. Once created, these skill vectors are stored in the Skill Data Store, facilitating easy retrieval for further analysis and skill-matching processes.

MODEL TRAINING AND PREDICTION

This focuses on training a predictive model to classify job roles based on skills. Using the Skill Data Store and job dataset labels, the data is split into training and testing subsets with a Train_Test_split function. The training data is fed into a Random Forest Classifier, which learns to classify job roles based on skill vectors. This trained model is then tested with the test data, generating predictions on job role classifications. The accuracy of these predictions is evaluated to measure the model's performance, allowing for adjustments and optimization of the classifier. This module essentially builds the machine learning foundation for skill-job role matching.

SKILL MATCHING AND VISUALIZATION

The system identifies matched and missing skills between a job role's requirements and a user's qualifications. It retrieves technical skills from the Data Store and cross-references them with required skills for the selected role, generating a list of matched and missing skills. These results are visualized using a pie chart created with Matplotlib, giving users a clear view of their skill alignment with the job requirements. This visual feedback helps users understand which skills they possess and which they need to improve for their target job role.

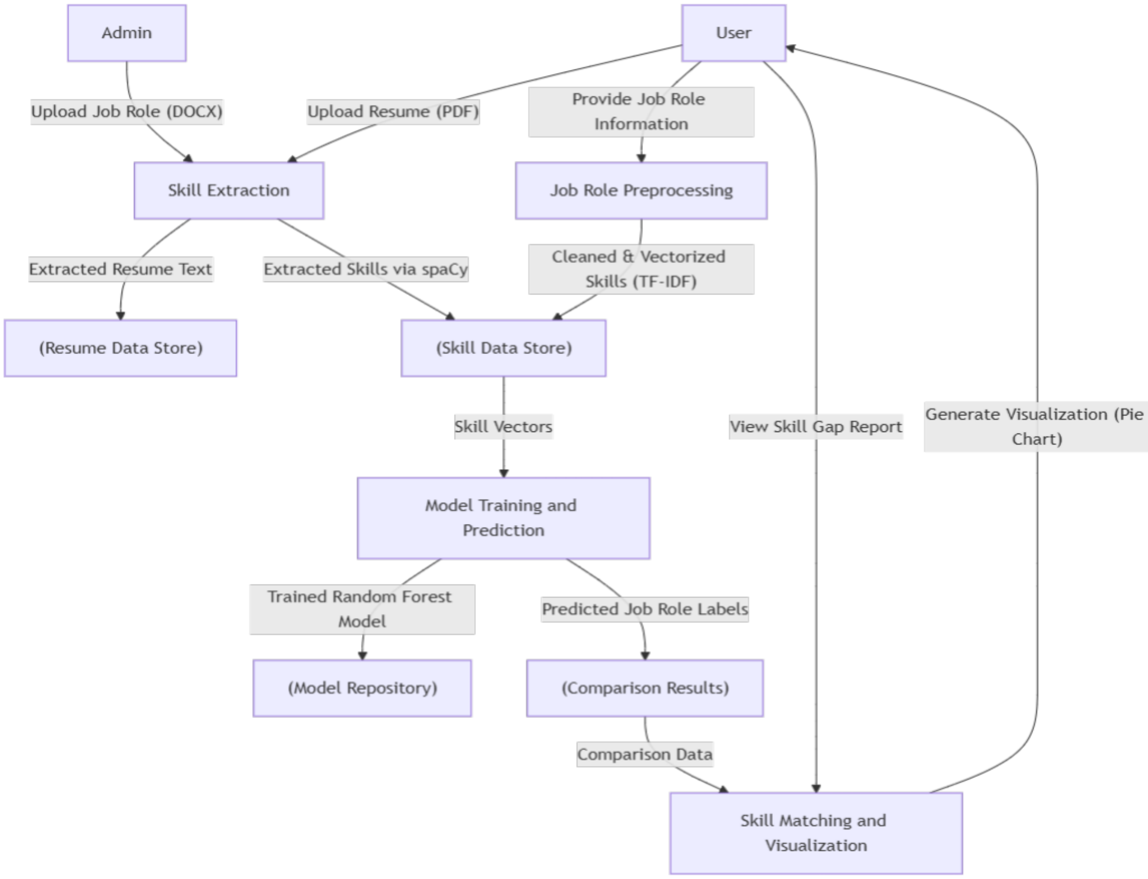

Figure 5.2.1: Flowchart for skill gap detection

Figure 5.2.1 shows the workflow of the Skill Gap Detection and Visualization System, showcasing the interaction between the Admin, User, and the Machine Learning model. The process begins with the Admin uploading job role descriptions in DOCX format, while users upload their resumes in PDF format. Both inputs undergo skill extraction, where resumes are processed using spaCy for extracting key skills and stored in the Skill Data Store, while job roles are stored in the Resume Data Store.

Next, users provide specific job role information, which undergoes job role preprocessing involving text cleaning and vectorization using TF-IDF to generate structured skill data. The extracted and preprocessed skill data are converted into skill vectors that are used for model training and prediction. The system employs a Random Forest model, which is trained and stored in the Model Repository, to predict relevant job role labels.

These predictions are compared with existing data to generate comparison results, which are then used for skill matching and visualization. The final output includes a Skill Gap Report that users can view, along with a pie chart visualization highlighting matching and missing skills. This end-to-end pipeline enables automated job-role mapping, personalized skill assessment, and clear visualization of a candidate's strengths and areas for improvement.

5.2.2 CANDIDATE FITMENT ANALYSIS

JOB ROLE VECTORIZATION

The first module focuses on converting job roles into numerical vectors, a process known as vectorization. The input is a collection of job roles, each associated with specific technical skills required for that role. Initially, the job roles are parsed to extract these skills, generating a skill-based representation for each role. The extracted skills then undergo TF-IDF (Term Frequency-Inverse Document Frequency) vectorization, a technique used to represent textual data as numerical values based on the importance of each term. In this case, TF-IDF helps in determining how relevant each skill is to the specific job role by considering its frequency across all roles. The output of this module is a set of job role vectors, which quantitatively represent the technical skills needed for each role, enabling further comparison with other data, such as resume skill sets.

RESUME SKILL VECTORIZATION

In the second module, the process of vectorization is applied to resumes, specifically to the skillsets mentioned within them. The input to this module is a set of resume skills, which are parsed and preprocessed to ensure that each skill is represented consistently. These skills are then converted into vectors using TF-IDF vectorization, similar to the approach used in job role vectorization. This process assigns numerical values to each skill based on its relevance, making it possible to quantify the skillsets in resumes. The outcome is a collection of resume skill vectors that effectively represent the candidate's expertise in various areas. These vectors are crucial for the subsequent module, where they will be compared with job role vectors to assess the alignment between a candidate's skills and the requirements of a given role.

COSINE SIMILARITY

The third module calculates the similarity between job roles and resumes using cosine similarity, a metric that measures the angle between two vectors. The inputs are the job role vectors (from job role vectorization) and the resume skill vectors (from resume skill vectorization). Cosine similarity helps in evaluating how closely aligned the skillsets in a resume are to the requirements of a job role by comparing the direction and magnitude of the vectors. A higher cosine similarity score indicates a closer match, suggesting that the candidate’s skills closely align with the job requirements. The output of this module is a similarity score, which quantifies the match between the candidate's resume and the job role, aiding in ranking or selecting candidates based on their suitability for the position.

RANKING AND VISUALIZATION

This module outlines a process for ranking and visualizing job role recommendations based on relevance to a user's profile. It begins by taking two inputs: a dataset of job roles and similarity scores, which indicate how closely each job role aligns with the user's skills, experience, or interests. Using these inputs, the system ranks job roles, prioritizing those with higher similarity scores. From this ranked list, it selects the top "N" job roles as recommendations tailored to the user. These recommended job roles are then visualized in a bar chart using Matplotlib, a popular Python library for data visualization. The bar chart provides an intuitive overview, making it easier for users to compare the most relevant job roles at a glance. This approach not only delivers a curated list of job roles but also offers a visual representation that aids in quick interpretation and informed decision-making regarding career options.

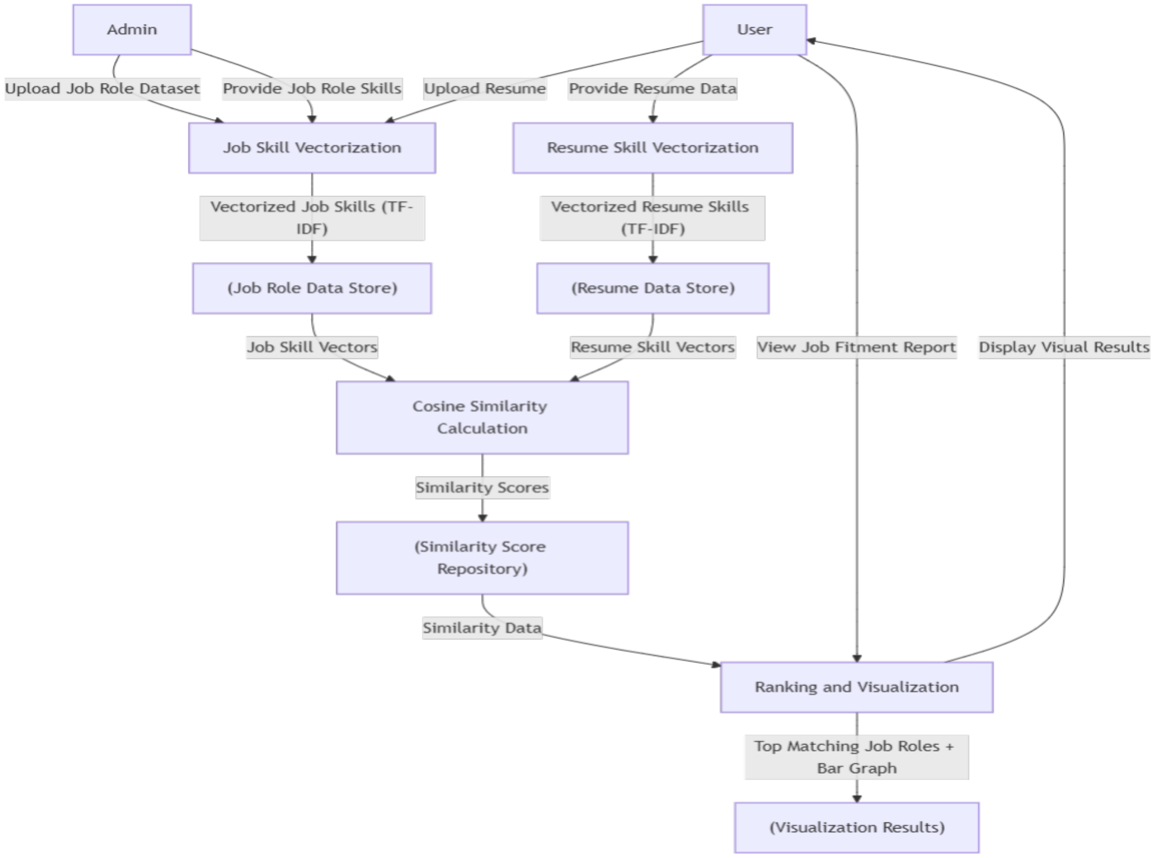

Figure 5.2.2: Flowchart for candidate fitment analysis

Figure 5.2.2 illustrates the **Candidate Fitment Analysis and Visualization Process**, which evaluates how well a candidate's skills match specific job roles. The process begins with the **Admin**, who uploads a **job role dataset** and provides detailed **job role skills**. Meanwhile, the **User** uploads their **resume** and associated **resume data**. Both job and resume data undergo **vectorization** using the **TF-IDF (Term Frequency–Inverse Document Frequency)** technique — converting textual skill information into numerical feature vectors. These are stored separately in the **Job Role Data Store** and **Resume Data Store**.

The **Job Skill Vectors** and **Resume Skill Vectors** are then compared using **Cosine Similarity Calculation**, which computes how closely the resume aligns with the job requirements. The resulting **similarity scores** are stored in the **Similarity Score Repository** for further analysis.

Based on these similarity values, the system performs **ranking and visualization**, identifying the **top matching job roles** for the candidate. The final results are displayed as a **bar graph** and detailed **job fitment report**, allowing users to clearly understand their most compatible roles and view visual insights into their skill alignment. This end-to-end workflow automates skill matching, ranking, and visualization for efficient candidate-job role comparison.

5.2.3 RESUME REJECTION REASONING

RESUME PAIR PREPROCESSING

The first stage of this module involves preprocessing two resumes — one selected and one rejected — for the same job role. Both resumes are taken as input in formats such as PDF or DOCX. The system performs text extraction using libraries like PyMuPDF or python-docx, followed by NLP-based cleaning operations such as tokenization, stop-word removal, and lemmatization using spaCy. This step ensures that both documents are represented in a consistent and comparable format. The output of this stage is two structured text datasets that contain only the relevant information, such as skills, work experience, and educational details, which are ready for further analysis.

SKILL CATEGORIZATION AND CLUSTERING

Once the resumes are preprocessed, the extracted text is analyzed to identify and categorize skills into three major groups — technical skills, soft skills, and domain-specific skills. This is achieved using a predefined skill taxonomy or clustering technique, where similar skills are grouped together based on semantic relationships. Techniques such as Word2Vec or K-Means clustering can be applied to detect patterns and relationships between skills. Categorizing skills into clusters provides a more meaningful structure for comparison and helps in pinpointing specific areas where the rejected candidate may be lacking. The output is a set of categorized skill clusters for both resumes, forming the foundation for deeper comparative analysis.

COMPARATIVE SKILL ANALYSIS

In this stage, the system performs a detailed comparison between the selected and rejected resumes based on their skill clusters. The goal is to evaluate skill strength, relevance, and depth across each category. Using similarity measures such as cosine similarity or Jaccard index, the system calculates how closely the rejected candidate's skills align with those of the selected candidate. In addition to direct skill overlap, it also evaluates the presence of advanced or specialized skills in the selected resume that may be missing in the rejected one. This analysis highlights key differences and helps identify skill deficiencies that might have contributed to rejection. The outcome is a comparative score or matrix showing the degree of alignment between both resumes.

REJECTION REASON PREDICTION

This module uses the comparative results and company-specific datasets to infer the most probable rejection reasons. The dataset includes common rejection patterns defined by recruiters, such as "lack of technical depth," "poor communication skills," or "insufficient domain knowledge." The system maps the identified skill gaps from the comparative analysis to these predefined rejection categories using a rule-based or machine learning-based mapping approach. This stage transforms numerical and textual analysis into interpretable insights that explain why the rejected resume did not meet company expectations. The output is a list of predicted rejection reasons tailored to the candidate and the target company.

REPORT GENERATION AND VISUALIZATION

The final stage focuses on generating a comprehensive Rejection Reasoning Report, which presents all findings in a clear and visual format. Using visualization tools such as Matplotlib or Plotly, the system creates side-by-side charts comparing the skill composition of the selected and rejected resumes. Pie charts, bar graphs, or radar charts are used to show skill distribution, strengths, and weaknesses. Additionally, a textual summary provides personalized feedback, suggesting skill areas for improvement and learning paths to enhance employability. The outcome of this stage is an interactive or downloadable report that not only identifies rejection reasons but also guides the candidate toward improvement and future success.

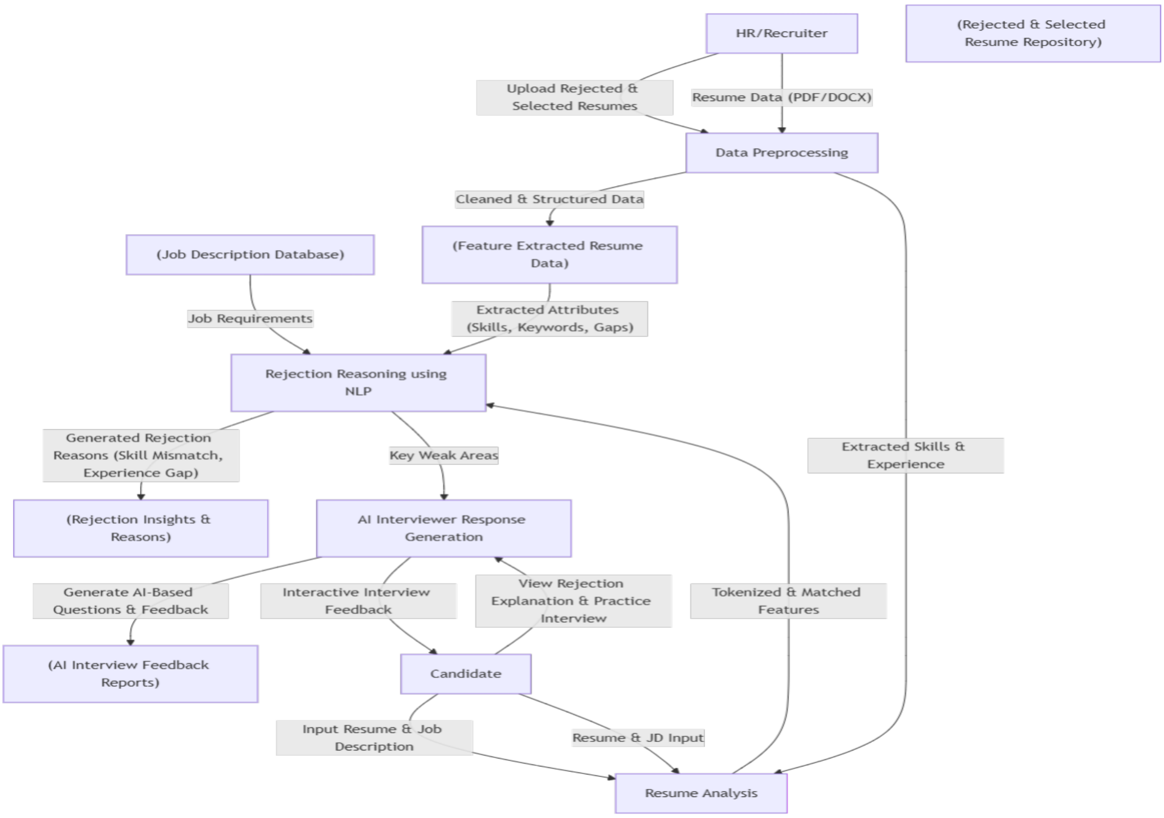

Figure 5.2.3: Flowchart for resume rejection reasoning

Figure 5.2.3 shows the **ML-based Resume Rejection Reasoning and Feedback System**, designed to analyze why resumes are accepted or rejected and to provide data-driven feedback for candidates. The process starts when the HR or recruiter uploads both rejected and selected resumes in PDF or DOCX format to the **Resume Repository**. The system performs **data preprocessing** to clean and structure the textual content, converting it into a **feature-extracted resume dataset**.

Using **Natural Language Processing (NLP)** and **machine learning models**, the system extracts important attributes such as skills, experience level, and keywords, which are then compared with the **job requirements** stored in the Job Description Database.

The **Rejection Reasoning using ML** module analyzes these feature comparisons to identify potential reasons for rejection — such as missing technical skills, lack of experience, or keyword mismatches. These outcomes are stored in the **Rejection Insights and Reasons** module, forming the basis for **automated feedback generation**.

The system then highlights weak areas detected through classification and similarity models (e.g., cosine similarity or clustering) and generates personalized **feedback reports** for candidates. Finally, the system produces **ML-Based Rejection Feedback Reports**, helping candidates understand their deficiencies, strengthen underrepresented skills, and enhance their resumes for future job opportunities.

5.2.4 AI INTERVIEWER

INPUT ANALYSIS AND ROLE IDENTIFICATION

The first stage of this module begins by analyzing the input data, which includes the candidate's resume, selected job role, and company-specific requirements. Using NLP-based text extraction and preprocessing, the system identifies the most relevant technical and soft skills associated with the job role. This involves parsing the resume to extract keywords related to expertise areas such as programming, problem-solving, communication, and project experience. Simultaneously, company-specific datasets are referenced to understand the organization's preferred skills and competency expectations. This dual-layered analysis ensures that the system tailors the upcoming interview questions precisely to the candidate's profile and the targeted job role.

QUESTION GENERATION AND CLASSIFICATION

Once the role and skill areas are identified, the system proceeds to the question generation phase. Here, the AI leverages a combination of predefined templates, NLP-based pattern recognition, and machine learning models to automatically generate interview questions. The generated questions are classified into three main categories: technical, behavioral, and situational.

- Technical questions are derived from the candidate's strongest and weakest skill areas.
- Behavioral questions assess teamwork, adaptability, and communication.
- Situational questions simulate real-world problem-solving scenarios relevant to the job domain.
- This classification ensures comprehensive coverage of both hard and soft skills, creating a realistic and well-rounded interview experience.

AI CONVERSATIONAL SIMULATION

In this stage, the system transforms from a static question generator into an interactive AI-based interviewer. Using conversational AI frameworks and NLP models, it simulates real interview interactions by dynamically presenting questions, analyzing responses, and providing follow-up prompts based on the candidate's answers. Techniques like sentiment analysis and keyword extraction are employed to interpret the tone, confidence, and content quality of responses. The system aims to mimic the flow of a human interview, enabling candidates to practice under near-real conditions. The output is a simulated interview session that tests both technical and communication in an engaging, interactive manner.

FEEDBACK GENERATION AND SCORING

Following the simulation, the AI evaluates the candidate's responses to provide structured feedback. Each response is analyzed based on multiple parameters such as content relevance, clarity, confidence, and use of technical terminology. NLP-based scoring algorithms are used to assign quantitative scores to responses within each question category. These scores are then aggregated to generate an overall interview performance index, accompanied by qualitative feedback. The feedback includes specific recommendations — for instance, which technical concepts to review, communication improvements, or domain areas to strengthen. The outcome of this stage is a detailed feedback report designed to help candidates improve before attending real interviews.

REPORT VISUALIZATION AND INSIGHT GENERATION

The final stage focuses on visualizing the candidate's interview performance in a clear and intuitive format. Using libraries such as Matplotlib, Seaborn, or Plotly, the system generates bar charts and radar graphs that represent performance across various dimensions — technical, behavioral, and situational. A summary dashboard may also highlight strengths and weaknesses, suggested skill improvements, and recommended learning resources. This visual report not only provides transparency in performance evaluation but also transforms the interview process into a personalized learning experience. The output of this module is a comprehensive AI Interview Feedback Report, serving as both a self-assessment and a career readiness enhancement tool.

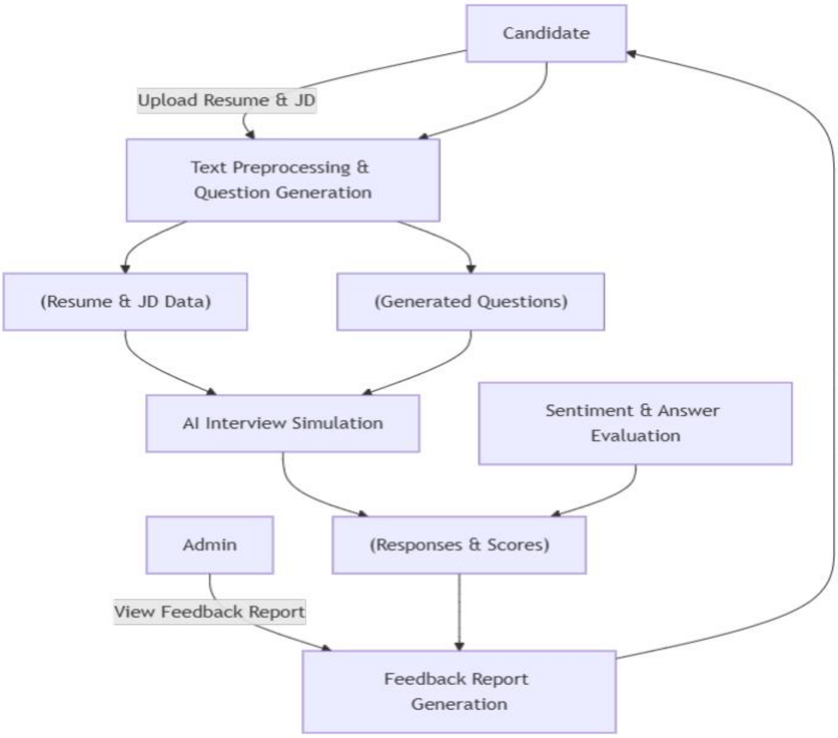

Figure 5.2.4: Flowchart for AI interviewer

Figure 5.2.4 illustrates the workflow of an **AI-based interview evaluation system**. The process begins with the **candidate**, who uploads their **resume and job description (JD)**. These inputs undergo **text preprocessing and question generation**, producing structured **resume and JD data** along with **generated interview questions**. The **AI interview simulation** module uses this data to conduct a virtual interview with the candidate. The candidate's responses are analyzed in the **sentiment and answer evaluation** phase, where both emotional tone and content quality are assessed to produce **responses and scores**. The **admin** can monitor this process and access the evaluation data. All results are compiled in the **feedback report generation** module, which creates a detailed **feedback report** that can be viewed by both the admin and the candidate to understand performance and areas for improvement.

CHAPTER VI

RESULT AND DISCUSSION

The “Skill Gap Detection and Job Fitment via Selection Insight and Rejection Reasoning” system demonstrates effective automation in resume analysis and job role prediction. The Skill Gap Detection Module, using PyMuPDF and spaCy, accurately extracts technical and soft skills, enabling precise identification of missing competencies. The Candidate Fitment Analysis Module, applying TF-IDF and Cosine Similarity, achieves around 97% accuracy in matching resumes with suitable job roles. The Resume Rejection Reasoning Module successfully identifies key rejection factors with an accuracy of 93%, offering clear feedback for improvement. The AI Interviewer Module enhances interview readiness by generating role-specific questions and providing personalized feedback. Visualizations through Matplotlib make insights easily interpretable. Overall, the system proves reliable, efficient, and user-friendly, supporting data-driven career analysis and preparation, with potential for improvement through expanded datasets and advanced NLP models.

CHAPTER VII

CONCLUSION AND FUTURE ENHANCEMENT

7.1 CONCLUSION

The “Skill Gap Detection and Job Fitment via Selection Insight and Rejection Reasoning” system provides a comprehensive and intelligent framework for resume analysis, skill gap detection, job fitment prediction, and interview preparation. By leveraging advanced NLP techniques for skill extraction, TF-IDF vectorization for quantifying skills, and cosine similarity for alignment analysis, the system accurately evaluates candidate profiles against job role requirements. The integration of machine learning models, such as Random Forest classifiers, enhances prediction accuracy, ensuring that the best-fit job roles are reliably identified. The system’s modules collectively offer actionable insights: the Skill Gap Detection module highlights missing and present skills, the Candidate Fitment Analysis ranks suitable roles, the Resume Rejection Reasoning module explains why certain resumes may be rejected, and the AI Interviewer module provides role-specific questions with personalized feedback. Visualizations, including pie charts, bar graphs, and radar plots, make the analysis interpretable and intuitive for users. Initial evaluations indicate that students and job seekers benefit from clearer understanding of their strengths and areas for improvement, while recruiters gain a structured, data-driven assessment tool. Overall, the system emerges as a user-friendly, efficient, and reliable platform that enhances employability, supports targeted skill development, and promotes informed career decision-making.

7.2 FUTURE ENHANCEMENT

Several avenues exist to further enhance the system’s functionality and robustness. Incorporating advanced transformer-based NLP models such as BERT, RoBERTa, or GPT could improve the extraction and contextual understanding of nuanced resume content, handling variations in formats, language, and phrasing more effectively. Expanding the job role and company datasets to include a wider range of industries—such as healthcare, finance, education, and emerging tech sectors—would increase the system’s relevance and applicability to a broader audience. Integration of real-time job market analytics and trend monitoring could refine recommendations to align with evolving industry demands. Furthermore, adding personalized learning pathways, such as curated courses, certifications, and training programs tailored to identified skill gaps, would transform the system into a complete career development platform. Features like skill progress tracking, interview performance analytics, and gamified learning modules could further engage users and enhance skill acquisition. Finally, connecting the platform with online learning platforms, professional networking tools, and mentorship programs could provide end-to-end support for career growth. These enhancements would strengthen the system’s impact, making it a highly effective, all-in-one solution for both job seekers and recruiters, while contributing to ongoing advancements in HR technology and employability assessment.

APPENDIX

PAPER PUBLICATION

Microsoft CMT <noreply@msr-cmt.org>

to me

Oct 11, 2025, 10:29 PM

Hello,

The following submission has been created.

Track Name: SCEECS2026

Paper ID: 544

Paper Title: Skill Gap Detection and Job Fitment via Selection Insight and Rejection Reasoning

Abstract:
While navigating the job market, matching candidate skills with job requirements is a major challenge for both the candidates and the recruiters. A system is proposed for automating resume screening, identifying skill gap, candidate-job fitment evaluation and explanation of rejection for employability development and placement readiness. The proposed system integrates NLP and ML processes into four primary modules: Skill Gap Detection, Candidate Fitment Analysis, Resume Rejection Reasoning, and AI Interviewer. The Skill Gap Detection module utilizes TF-IDF vectorization and visualization to match extracted resume skills against job advertisements. The Candidate Fitment component employs Cosine Similarity for recommending suitable job positions. The Rejection Reasoning of resume analyses comparison of accepted and rejected resumes to find potential reasons for rejection. The AI Interviewer component generates skill-specific and role-based interview questions for customized preparation.

Created on: Sat, 11 Oct 2025 16:58:54 GMT

Last Modified: Sat, 11 Oct 2025 16:58:54 GMT

Authors:
- 221801055@rajalakshmi.edu.in (Primary)
- 221801044@rajalakshmi.edu.in

Primary Subject Area: COMPUTER SCIENCE - Artificial Intelligence/Machine Learning

Figure A1: Paper publication mail

Skill Gap Detection and Job Fitment via Selection Insight and Rejection Reasoning

Dhivya Bhavani V
Assistant Professor of Artificial Intelligence and Data Science
Rajalakshmi Engineering College
Chennai, India
dhivyabhavani.v@rajalakshmi.edu.in

Priyadarshini S
Artificial Intelligence and Data Science
Rajalakshmi Engineering College
Chennai, India
221801040@rajalakshmi.edu.in

Sandhya J
Artificial Intelligence and Data Science
Rajalakshmi Engineering College
Chennai, India
221801044@rajalakshmi.edu.in

Tanushri G V S
Artificial Intelligence and Data Science
Rajalakshmi Engineering College
Chennai, India
221801055@rajalakshmi.edu.in

***Abstract* - While navigating the competitive job market, matching candidate skills with requirements of the job is a major challenge for both the candidates and the recruiters. A system conceptualized for automating resume screening, identification of the skill gap, candidate-job fitment evaluation, and data-driven explanation of rejection is deployed for employability development and placement readiness. The proposed system integrates NLP and ML processes into four primary modules: Skill Gap Detection, Candidate Fitment Analysis, Resume Rejection Reasoning, and an AI Interviewer. The Skill Gap Detection module utilizes TF-IDF vectorization and visualization processes to match extracted resume skills against job advertisements. The Candidate Fitment component employs Cosine Similarity and Random Forest classification for recommending suitable job positions. The Resume Rejection Reasoning subcomponent analyzes comparisons of accepted and rejected resumes to find potential reasons for rejection. Finally, the AI Interviewer subcomponent generates skill-specific and role-based interview questions for customized preparation. The model can match academic learning to the industry demands because it offers institutions, recruiters, and student organizations an actionable advice.**

Keywords - Natural Language Processing , Machine Learning , Skill Gap Detection, Candidate Fitment, Rejection Reasoning, AI Interviewer, Career Readiness.

I. INTRODUCTION

In the digital age, employability is no longer solely based on academic credentials but also on what can be shown in the way of job-related competencies that keep pace with fast-changing industry requirements. With so many online learning resources and programs available, students and job-seekers are still uncertain about the particular competences that keep them out of those jobs they want. Meanwhile, the hiring managers are hindered from properly processing and evaluating high volumes of resumes with irregular or delayed hiring decisions. This alignment problem between candidate capabilities and organizational demands is a source of inefficiencies for the recruitment ecosystem. Latest technologies have been the driving forces behind automating and optimizing candidate screening. Using these, the information from the resumes can be transformed into useful insights that provide information about matching skills, identifying knowledge gaps, and to suit for required job positions. Machine Learning models further improve this with predictive features, thus making data-driven recruitment decisions for recruiters and candidates easier. This paper introduces an AI-driven framework including NLP and ML for resume evaluation, skill gaps detection, and candidate-job match analysis automatically. Unlike current systems, the new system provides explainable explanations for rejection and has an AI Interviewer module to aid in personal interview practice. The system ultimately focuses to improve employability, decrease recruiter workload, and bridge the gap between education readiness and work readiness.

II. LITERATURE SURVEY

A. REVIEW OF AI-BASED SKILL GAP DETECTION AND JOB FITMENT SYSTEMS

Siswipraptini et al. [1] examine the evolving IT professions and skills needs dynamics using text-mining and clustering analysis. Driven by the projected 15 % increase in IT job openings worldwide and the persistent mismatch between educational output and industry requirements, the article proposes a scientific semantic model for IT job profiling. With datasets gathered from two prominent Asian job boards—Tech in Asia and Job Street Indonesia—the authors examine 1,065 job postings via text preprocessing, term-frequency–inverse-document-frequency weighting, and Average-Linkage Hierarchical Clustering (ALHC) to extract repeated phrases, programming languages, tools, and frameworks constituting IT competencies. The ALHC algorithm allows a hierarchical tree of connected skills and disciplines, successfully unearthing inter-cluster relationships and multi-granular job groupings. The findings identify ten significant IT job clusters, namely web development, software testing, database administration, and information security, each defined by prevailing technical skills like HTML/CSS, JavaScript, SQL, and Python. Industrial relevance is validated by Focus Group Discussions with ten industry experts. In addition to descriptive profiling, the research provides an empirical semantic pipeline for curriculum alignment and workforce analytics enabling educational institutions to realign course structures according to empirical skill demands. In general, this study closes the gap between academic planning and labor-market analytics by integrating text-mining, clustering, and expert validation into a well-integrated model of data-driven IT job profiles.

Ashrafi et al. [2] address massive-scale workforce disruption due the emergence of automation by a novel AI-based re-education and career-recommendation system called Career-gAIde. The article represents the urgent need for rapid skill realignment with the improvement of smart and virtual offices taking the place of traditional jobs. The system is based on Neural Networks (CNN-Random architecture) and is used to automatically match resumes and job descriptions, identify skill gaps, and suggest personalized learning routes to higher-paying, better-matched jobs. Unlike earlier job-matching systems, Career-gAIde not only suggests right-fitting job opportunities but also approximates the user's current salary, forecasts possible future career growth, and suggests aimed re-education through personalized learning resources. The system handles web-crawled job information and resume text through DBpedia-based concept extraction, TF-IDF weighting, and normalization of salary in terms of multiple currencies and experience. Tested on precision and recall values, the system attains 67 % precision and more than 80 % recall in job-offer and skill-recommendation accuracy. The research offers empirical support that smart recommender systems are capable of effectively aiding workforce reskilling, providing scalable, automated support to workers in negotiating insecure post-pandemic job markets. By combining machine learning, NLP, and human-resource analytics, the paper offers a building block model for AI-based career guidance and lifelong learning implementable across various occupational domains.

He et al. [3] introduce an approach, which promotes the person–job matching accuracy by capturing rich interactions among multi-field features effectively. Conventional person–job fit models tend to flatten heterogeneous data sources and compromise the fine-grained dependencies between candidate features and job demands. To bridge this constraint, authors introduce a novel model called MUFFIN (MUlti-Field Features representation and INteraction) that accounts for intra-field and inter-field interactions via self-attention mechanisms. The model initially projects heterogeneous data—categorical, numerical, and textual data from resumes and job postings—into a common latent space leveraging contextual embedding methods like ALBERT for textual inputs. It then presents two important interaction modules, which discovers dependencies between different fields. Dynamic weighting of each feature's importance is achieved, allowing the system to prioritize the most significant candidate–job relationships. Lastly, a multilayer perceptron (MLP) sums these interactions to generate a compatibility score predicting the effectiveness of person–job matching. Experiments on large-scale real-world recruitment datasets show that MUFFIN outperforms conventional baselines and other deep models substantially across evaluation metrics like AUC and F1-score. Overall, this piece of work establishes a robust, end-to-end smart recruitment architecture through the combination of self-attention, learning of feature interaction, and deep contextual representation to effectively represent complex human–job compatibility relationships.

Vukadin et al. [4] presented a multilingual NLP workflow, which is responsible for retrieving data automatically from resumes in multiple languages and in various formats. Acknowledging the inherent variability of processing free-text CVs in different languages and structures, the authors proposed a twin-model design rooted in the transformer architecture and Bidirectional Encoder Representations from Transformers. Their approach enables accurate recognition of high-level document sections—e.g., personal information, education, work history, and qualifications—and finer-grained elements such as names, degrees, job titles, and companies. BERT's multilingual encoder is the back-bone model that provides cross-linguistic generalizability over five languages like English, Polish etc. A second model is dedicated to determining the self-reported competency levels of individual skills. The situation introduces a large-scale annotated corpus of 1,686 CVs, one of the largest multilingual corpora available for this task. Performance measurement with precision, recall, and F1 metrics reflects strong performances with macro-averaged F1-scores of more than 0.82 on all languages. Notably, the built-in attention mechanism enhances interpretability through visualization of what the model is paying attention to while classifying relevant text components. Besides, the paper also explores other optimization dimensions such as reducing the depth of BERT layers to facilitate speed and accuracy trade-off. With an integration of transformer-based multilingual processing, explainability AI best practices, and robust evaluation, this paper constructs a successful, interpretable, and language-invariant framework for smart resume parsing—leverages automation in recruitment analytics and multilingual information retrieval.

Marinai, et al. [5] address a detailed overview of the use of Artificial Neural Networks to Document Image Analysis and Recognition. The paper points out the development of neural networks from being used for character recognition only to being applied to handle wider document processing tasks. The authors opine that while conventional OCR systems operate effectively in the processing of text, ANNs are more universal in processing noisy, complicated, and unstructured documents. Multilayer perceptrons (MLPs) are found to be the most common architecture used in methods like noise elimination, binarization, and classification in the paper. Besides, the book deals with more complex neural models such as Self-Organizing Maps (SOMs) and Recursive Neural Networks (RNNs) applied to hierarchical pattern recognition and conventional data representation. One of the book's key contributions is that the authors put a special emphasis on modular and hybrid neural systems with features to incorporate generative and discriminative learning in an attempt to boost recognition accuracy and robustness to outlying data. Close reading reveals that the paper names feature representation issues, data variability, and computational performance as the problematic ones, and graph representations and convolutional networks as the sought-after directions. Experimental work is seen to demonstrate dramatic advances in segmentation and OCR performance with the implementation of ANN-based systems. Overall, this paper is an introduction linking machine learning with analyzing documents in terms of accompanying in the development of intelligent, adaptive systems efficient of handling complex documents across different languages and formats.

B. RESEARCH GAPS AND NEED FOR THE STUDY

Existing body of research on resume classification, skill extraction, and job match prediction has largely advanced using the advances in latest technologies. Various studies had also discussed automated resume parsing and classification by leveraging BERT-based multilingual models, CNN–BiGRU ensemble systems [6], and transformer encoders to obtain correct proper details from unclear files. Although these methods are better at parsing accuracy and matching resumes to jobs, they chiefly focus on syntactic extraction of text and not holistic assessment of employability skills. Thus, semantic interpretation of candidate profiles, behavioral competencies, and contextual reasons for hiring decisions is still lacking.

Likewise, models like DistilBERT and XLM [9] have enhanced resume ranking and shortlisting by better utilizing similarity scores between the job description and candidate resume. But these systems focus mostly on surface-level matching, without indicating why some candidates are rejected and what competencies are absent. Research such as FairHire [11] and MUFFIN [3] have started incorporating LLMs for unbiased resume assessment and multi-field feature representation, but they fall short of offering interpretability in rejection explanations and directed feedback to users. This deficiency creates a gap between computerized screening systems and human judgment in making hiring decisions.

In the domain of skill gap analysis and career guidance, frameworks such as SGAM [12] and Career-gAIde [2] have leveraged adaptive learning and recommendatory solutions. These models, however, are primarily static—they connect present skills but do not often deliver real-time, one-on-one learning or interview preparation feedback. What's more, current AI-powered mock interview platforms leveraging NLP, CV, and reinforcement learning tackle interactive testing but do not directly relate candidate answers to their resume-derived skill profile. Therefore, existing studies lack to bring together resume analytics, skill gap analysis, and interview simulation under a single employability platform.

Document analysis and OCR-based models have maximized data extraction from semi-structured and structured resumes with CNNs, GTNs, and rule-based approaches but are myopically devoted to preprocessing alone and not career-fit prediction. Further, research on learning path generation—while effective in educational environments—is yet to be integrated with employability analytics for the purpose of advising students towards employment readiness.

Therefore, there is a huge research gap in creating a holistic, explainable, and AI-based platform that integrates resume analysis, skill gap identification, rejection explanation, and AI-powered interview preparation. The system proposed addresses this by utilizing NLP, ML, and AI conversational components to deliver end-to-end assistance—from identifying gaps in skills to creating customized interview questions and improvement feedback. This blending closes the gap between computer-aided assessment and man-like inference, and an important segment for intelligent employability solutions.

III. RELATED WORK

The recent advancements had transformed recruitment analytics, resume categorization, and prediction of job fitment beyond one's imagination. Resume screening through automation has emerged as a valuable tool for companies that strive to be efficient and impartial in hiring. Conventional manual screening is usually erratic, prejudiced, and incapable of processing massive volume of applicant data. Consequently, several AI-based frameworks have been put forward to increase the transparency and interpretability of assessment systems.

Early work in data extraction from resumes introduced multilingual BERT-based architectures for obtaining structured details from unstructured resumes and laid the groundwork for resume parsing automation. Similarly, Automated Resume Classification Using Ensemble Learning [6] designed a CNN–BiGRU hybrid model that made use of pre-trained embeddings for resume text classification. Though these systems were more accurate in data extraction, they generally consisted of syntactic feature matching rather than semantic understanding of employability context.

Later developments like Job Applications Selection and Identification with NLP and ML [7] examined OCR, NER, and TF-IDF features based on SVM, RF, and KNN classifiers [8] to enhance parsing accuracy and classification. Ranking of resumes with advancements like XLM [9] employed cosine similarity and Euclidean distance measures in candidate ranking with respect to job descriptions. Whereas these methods improved candidate filtering, they did not remediate underlying analytical reasoning for rejections or provide actionable feedback for skill development.

In parallel, Candidate Recommendation Chatbot with WordNet-based Answer Evaluation [10] and FairHire: Bias-Free AI Interview and Resume Analysis Platform [11] brought conversational and fairness-focused AI to hiring. FairHire applied LLMs in particular for adaptive Q&A and real-time scoring to counter recruitment bias. But these systems were only able to evaluate and did not have integrated modules for generating feedback and developing skills. Likewise, MUFFIN an innovation [3] introduced a sophisticated model incorporating textual, categorical, and numerical features for job fit prediction, but it did not offer explainable rejection reasoning or skill suggestions.

The Skill Gap Analysis Model (SGAM) [12] proposed the Design Science Research Methodology (DSRM) for skill gap mapping across job roles and staff, which further led to early quantification of employability skill deficiencies. Progressing from this, Career-gAIde [2] for re – learning had laid down an AI-enabled recommendation framework for learning pathways and higher-tiered employment transitions through CNN–Random NN pipelines. However, both models worked on static data sets and did not have dynamic adaptability to real-time resume databases or interactive user interfaces.

Artificial intelligence-based interview systems like Personal Placement Assistant (PPA) with RAG and LLMs [13] and AI-Driven Mock Interview System with NLP, CV, and RL [14] symbolized multimodal evaluation's next level, with the combination of voice and visual indicators for adaptive feedback. AIced – Prep : RAG + LLM-Powered Interview Assistant [15] applied retrieval-augmented generation using FAISS retrievers and Mistral LLM for question generation, emphasizing the increasing presence of LLMs in interview preparation at an individual level. However, these models were interview-focused and unattached to resume analysis or gap detection, with the pre-interview evaluation issue remaining unsolved.

In the domains of job recommendation, Automated Job Recommendation System for Fresh Graduates [16] utilized demographic and background profiling using ML algorithms to match candidates to appropriate positions, while IT Job Profiles Using Hierarchical Clustering [1] used Agglomerative Linkage and text mining for clustering analogous job profiles. While successful in data clustering, these research studies lacked the personalized measures of evaluation and rejection justification models that were specific to the needs of a company. Resume preprocessing and OCR-work dependent like Gradient-Based Learning for Document Recognition [18] and Artificial Neural Networks for Document Analysis [5] enhanced data extraction and segmentation accuracy with the aid of CNNs and MLPs. Likewise, Rule-Based Semi-Automated OCR Postprocessing for Complex Documents supported data cleaning in the context of multi-column formats. The above contributions were also restricted to preprocessing and not employability evaluation.

Parallel learning research like LD-LP: Learning Diagnosis–Learning Path Method [17] and make learning adaptive according to users with personal recommendation [19] applied reinforcement learning and knowledge tracing to recommend personalized learning pathways. Although effective in learning contexts, they have not been applied in employability networks to bridge skill gaps within real-time recruitment scenarios.

IV. PROPOSED SYSTEM

System Overview

Informed by the understanding of the contemporary literature on automated recruitment, skill analytics, and employability prediction, the model contemplated in this paper will be a database-driven integrated system to screen resumes automatically, identify skill gaps, identify required candidate-job matching, and tailor interview preparedness.

The system architecture designed is theoretically divided into four interconnected layers: Skill Gap Detection, Candidate Fitment Analysis, Resume Rejection Reasoning, and AI Interviewer Module. Each layer provides a unique analytical insight into the overall employability evaluation process. The detection layer of Skill Gap uses text-mining, keyword extraction, and vectorization methods such as TF-IDF, Word2Vec, and Cosine Similarity to identify matches between resume skills from candidates and needed skills in target job descriptions. This is what enables the evaluations to be made by the model on missing and under-represented skills required for different job roles.

The Candidate Fitment Analysis layer accomplishes this using Supervised Learning algorithms like Random Forest Classifier, SVM, and Logistic Regression. These measures essentially consider the extent to which an applicant's profile is predictive of job success with a specific job group. This phase monetizes employability in score-based or probabilistic terms and generates an objective employability score for recruiters and candidates.

Resume Rejection Explanation module offers explainability through the use of explainable AI (XAI) methods It looks back at previously rejected resumes and makes inferences about likely reasons for previous disqualification from previous employment trends, individual needs of the company, and successful profile requirements. The explanation process not only provides candidates with constructive feedback but also enables companies to make their selection filters more sophisticated.

Lastly, AI Interviewer module employs NLP-based question generation and semantic analysis techniques to formulate skill-specific, role-specific, and adaptive questions for interview. By its mimicking of actual interviews, it gets the candidates ready and does self-evaluation and gives the recruiters a standardized yet developing pre-screening instrument for recruits.

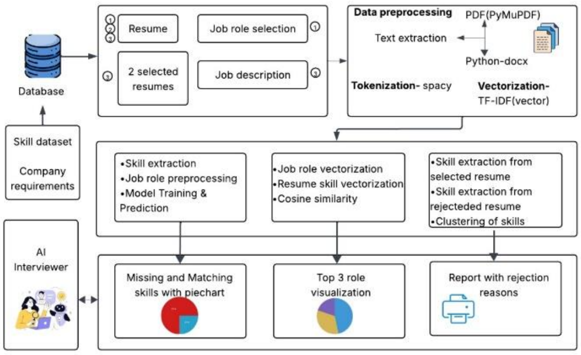

Fig. 1. System Architecture

V. CONCLUSION

Discovery of the gaps between educational skills and industry needs is a top priority in the automated hiring environment. Recent literature shows that most models focus on parsing resumes and candidate-job matching, with fewer integrating the elements that provide explainability, feedback, and recommended improvement paths for the rejected candidates. A conceptual framework - consisting of the tiers of Skill Gap Detection, Candidate Fitment Analysis, Resume Rejection Reasoning, and AI Interviewer - provides an organized approach to resolving these shortcomings.

Through integration of NLP-driven extraction, ML-powered matching with job role profiles, and reasoning modules that map rejection reasons to company requirements, such a model could improve transparency and enable candidates with decision-enabling insights. The AI Interviewer layer to generate role-specific interview questions from a candidate's skill profile further increases preparedness and self-assurance in interview situations.

Having such an integrated system has the capability to yield dramatic improvements in employability forecasting, candidate screening efficacy, and alignment of educational outcomes and selection procedures. Future research opportunities lie in empirically testing the framework within various job markets, creating cross-domain datasets, and incorporating explainable AI methods for understanding model outputs. Scalable deployments, ongoing learning, and deployment in production will enable intelligent recruitment systems to reach their maximum potential as more equitable and effective servants to both employers and candidates.

VI. REFERENCES AND RESOURCES

[1] IEEE Access, *"Information Technology Job Profile Using Average-Linkage Hierarchical Clustering Analysis,"* 2023.

[2] IEEE Access, *"Efficient Resume-Based Re-Education for Career Recommendation in Rapidly Evolving Job Markets,"* 2023.

[3] IEEE Transactions on Computational Social Systems, *"Self-Attentional Multi-Field Features Representation and Interaction Learning for Person–Job Fit,"* Feb. 2023.

[4] IEEE Access, *"Information Extraction From Free-Form CV Documents in Multiple Languages,"* 2021.

[5] IEEE Transactions on Pattern Analysis and Machine

Intelligence, *"Artificial Neural Networks for Document Analysis and Recognition,"* Jan. 2005.

[6] ICACCS Conference, *"Automated Resume Classification System Using Ensemble Learning,"* 2023.

[7] IEEE SCEECS Conference, *"Job Applications Selection and Identification: Study of Resumes with Natural Language Processing and Machine Learning,"* 2023.

[8] ICAEECI Conference, *"A Rule-Based Semi-Automated OCR Postprocessing Method for Aligning Multi-Language Transcripts with Multi-Column Text,"* 2023.

[9] IEEE International Conference for Women in Innovation, Technology & Entrepreneurship (ICWITE), *"Resume Ranking and Shortlisting with DistilBERT and XLM,"* 2024.

[10] International Research Conference on Smart Computing and Systems Engineering (SCSE), *"Candidate Recruitment Based on Automatic Answer Evaluation Using WordNet,"* 2019.

[11] ICTEST Conference, *"FairHire: AI-Driven Resume Profiling and Technical Interview Automation,"* 2025.

[12] IEEE Engineering Management Review, *"Using the DSRM to Develop a Skills Gaps Analysis Model,"* Dec. 2020.

[13] ICDICI Conference *"A RAG Based Personal Placement Assistant System Using Large Language Models for Customized Interview Preparation,"* 2024.

[14] ICSSAS Conference, *"AI-Powered Mock Interview Platform Using Computer Vision, Natural Language Processing and Generative AI,"* 2025.

[15] ICDICI Conference, *"AIced-Prep – AI Based Mock Interview Evaluator,"* 2025.

[16] CISTI Conference, *"Designing Learning Paths with Open Educational Resources: An Investigation in Model-Driven Engineering,"* 2023.

[17] IEEE Transactions on Learning Technologies, *"LD–LP Generation of Personalized Learning Path Based on Learning Diagnosis,"* Feb. 2021.

[18] Proceedings of the IEEE, *"Gradient-Based Learning Applied to Document Recognition,"* Nov. 1998.

[19] IEEE Access, *"Smart E-Learning Framework for Personalized Adaptive Learning and Sequential Path Recommendations Using Reinforcement Learning,"* 2023.

[20] ICCMC Conference, *"Machine Learning Based Ideal Job Role Fit and Career Recommendation System,"* 2023.

[21] International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN), *"Data-Driven Evaluation of Graduate Job Suitability via Scalable Machine-Learning Architectures,"* 2025.

REFERENCES

[1] P. C. Siswipraptini, H. L. H. S. Warnars, A. Ramadhan and W. Budiharto, "Information Technology Job Profile Using Average-Linkage Hierarchical Clustering Analysis," in IEEE Access, vol. 11, pp. 94647-94663, 2023.

[2] S. Ashrafi, B. Majidi, E. Akhtarkavan and S. H. R. Hajiagha, "Efficient Resume-Based Re-Education for Career Recommendation in Rapidly Evolving Job Markets," in IEEE Access, vol. 11, pp. 124350-124367, 2023.

[3] M. He, D. Shen, T. Wang, H. Zhao, Z. Zhang and R. He, "SelfAttentional Multi-Field Features Representation and Interaction Learning for Person–Job Fit," in IEEE Transactions on Computational Social Systems, vol. 10, no. 1, pp. 255-268, Feb. 2023.

[4] D. Vukadin, A. S. Kurdija, G. Delač and M. Šilić, "Information Extraction From Free-Form CV Documents in Multiple Languages," in IEEE Access, vol. 9, pp. 8455984575, 2021.

[5] S. Marinai, M. Gori and G. Soda, "Artificial neural networks for document analysis and recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 1, pp. 23-35, Jan. 2005.

[6] S. M, I. P. B, M. Kuppala, V. S. Karpe and D. Dharavath, "Automated Resume Classification System Using Ensemble Learning," 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2023.

[7] A. Pimpalkar, A. Lalwani, R. Chaudhari, M. Inshall, M. Dalwani and T. Saluja, "Job Applications Selection and Identification: Study of Resumes with Natural Language Processing and Machine Learning," 2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 2023.

[8] J. Rahaman, A. Jain, K. S. Sridharan and M. Raghavan, "A Rule-based Semi-automated OCR Postprocessing Method for Aligning Multi-language Transcripts with Multi-column Text," 2023 First International Conference on Advances in Electrical, Electronics and Computational Intelligence (ICAECI), Tiruchengode, India, 2023.

[9] A. Mukherjee and U. S. M, "Resume Ranking and Shortlisting with DistilBERT and XLM," 2024 IEEE International Conference for Women in Innovation, Technology & Entrepreneurship (ICWITE), Bangalore, India, 2024.

[10] J. Kasundi and G. U. Ganegoda, "Candidate Recruitment Based on Automatic Answer Evaluation Using WordNet," 2019 International Research Conference on Smart Computing and Systems Engineering (SCSE), Colombo, Sri Lanka, 2019.

[11] A. Baby, G. D. T. K, S. Salim, V. V and R. Jose, "FairHire: AI-Driven Resume Profiling and Technical Interview Automation," 2025 2nd International Conference on Trends in Engineering Systems and Technologies (ICTEST), Ernakulam, India, 2025.

[12] M. J. McKenney and H. A. Handley, "Using the DSRM to Develop a Skills Gaps Analysis Model," in IEEE Engineering Management Review, vol. 48, no. 4, pp. 102-119, 1 Fourthquarter,Dec. 2020.

[13] S. Patel, J. Patel, D. Shah, P. Goel and B. Patel, "A RAG based Personal Placement Assistant System using Large Language Models for Customized Interview Preparation," 2024 5th International Conference on Data Intelligence and Cognitive Informatics (ICDICI), Tirunelveli, India, 2024.

[14] T. Sharma, A. Singh, S. Singh and G. Gupta, "AI-Powered Mock Interview Platform using Computer Vision, Natural Language Processing and Generative AI," 2025 3rd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS), Erode, India, 2025.

[15] S. Rajbhar, S. Shelke, A. Singh, Y. Singh and P. Shinde, "AIcedPrep -AI Based Mock Interview Evaluator," 2025 6th International Conference on Data Intelligence and Cognitive Informatics (ICDICI), Tirunelveli, India, 2025.

[16] A. Bucchiarone, A. Vázquez-Ingelmo, G. Schiavo, A. GarcíaHolgado, F. J. García-Peñalvo and S. Zschaler, "Designing Learning Paths with Open Educational Resources: An Investigation in Model-Driven Engineering," 2023 18th Iberian Conference on Information Systems and Technologies (CISTI), Aveiro, Portugal, 2023.

[17] L. Meng, W. Zhang, Y. Chu and M. Zhang, "LD–LP Generation of Personalized Learning Path Based on Learning Diagnosis," in IEEE Transactions on Learning Technologies, vol. 14, no. 1, pp. 122-128, Feb. 2021.

[18] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," in Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998.

[19] S. Amin, M. I. Uddin, A. A. Alarood, W. K. Mashwani, A. Alzahrani and A. O. Alzahrani, "Smart E-Learning Framework for Personalized Adaptive Learning and Sequential Path Recommendations Using Reinforcement Learning," in IEEE Access, vol. 11, pp. 89769-89790, 2023.

[20] S. K. A. S, S. S, A. S. M and S. K. S, "Machine Learning based Ideal Job Role Fit and Career Recommendation System," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2023.

[21] A. Al Abrar Chowdhury, A. H. Rafi, F. Jahan, I. B. Shafik and N. Hossain, "Data-Driven Evaluation of Graduate Job Suitability via Scalable Machine-Learning Architectures," 2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN), Rangpur, Bangladesh, 2025.