Infosys Springboard Virtual Internship 6.0 Completion Report

Team Details

Batch Number: 4, Team A

Start date : 13-10-2025

Team Members:
Aravind Adityaa
Arijit Karmakar
Ashutosh Kumar Singh
Ezhilarasan
Ambati Sujith

Internship Duration: 8 Weeks

Project Title

CapaBot

1. Project Objective

The primary objective of CapaBot is to develop an intelligent, AI-driven system that automates the process of evaluating how well a candidate's resume aligns with a job description. The tool aims to generate an accurate match score by analyzing keywords, skills, and relevant experiences extracted from both documents using NLP techniques.

CapaBot also aims to provide a detailed breakdown of matching skills, partially matched areas, and missing skills. By identifying these gaps, the system recommends specific skills, courses, and learning paths that would help candidates improve their overall alignment with job requirements.

From the recruiter's perspective, the objective is to streamline the shortlisting process by offering a fast, consistent, and data-driven evaluation of candidate profiles. This reduces manual effort and improves decision-making accuracy. For job seekers, the objective is to offer personalized insights that highlight strengths, weaknesses, and opportunities for upskilling.

Overall, CapaBot's objective is to bridge the gap between job requirements and candidate capabilities by providing an efficient, transparent, and intelligent resume–job matching solution that enhances both recruitment efficiency and career development.

2. Project description in detail

CapaBot is an AI-powered Resume Matcher and Skill Gap Analyzer designed to compare a candidate's resume with a job description and evaluate how closely the profile aligns with the role. The system uses Natural Language Processing (NLP) techniques to extract key skills, qualifications, and relevant keywords from both documents. It then applies TF-IDF vectorization and cosine similarity to generate an objective match score that indicates the candidate's suitability for the position.

Beyond simple text matching, CapaBot categorizes skills into hard and soft skills, identifies skill gaps, and highlights the exact areas where the candidate needs improvement. It also provides personalized course recommendations for each missing skill using curated learning resources, helping users bridge knowledge gaps effectively.

To enhance user interaction, CapaBot integrates an AI-powered chatbot built on Gemini Pro. The chatbot explains analysis results, answers career-related queries, and guides users on improving their resumes. The platform includes a clean and easy-to-use web interface where users can upload resumes, paste job descriptions, and instantly view insights through visual dashboards and skill charts.

Overall, CapaBot streamlines resume evaluation, supports recruiters in shortlisting candidates, and empowers job seekers with clearer career direction. It bridges the gap between job requirements and individual capabilities, promoting skill-based growth and smarter hiring decisions.

3. Timeline Overview

Week	Activities Planned	Activities Completed
Week 1	Project introduction, understanding requirements, idea discussion	Finalized project concept, objectives, and approach
Week 2	Basic NLP learning, TF-IDF research, environment setup	Set up Flask backend, created initial data pipeline
Week 3	Resume parsing module, text extraction logic	Implemented PDF/DOCX extraction & preprocessing
Week 4	Similarity engine development (TF-IDF + Cosine Similarity)	Completed working similarity model & testing
Week 5	Skill database creation & categorization	Built full skills DB (programming, web, cloud, DS, DevOps, soft skills)
Week 6	Skill extraction logic and recommendations module	Added pattern-based extraction & course recommendation system
Week 7	Frontend UI, graph visualization, job analysis flow	Built UI, added Chart.js output, integrated backend
Week 8	AI chatbot, testing, documentation, presentation	Integrated Gemini AI, completed testing & final presentation

4. Key Milestones

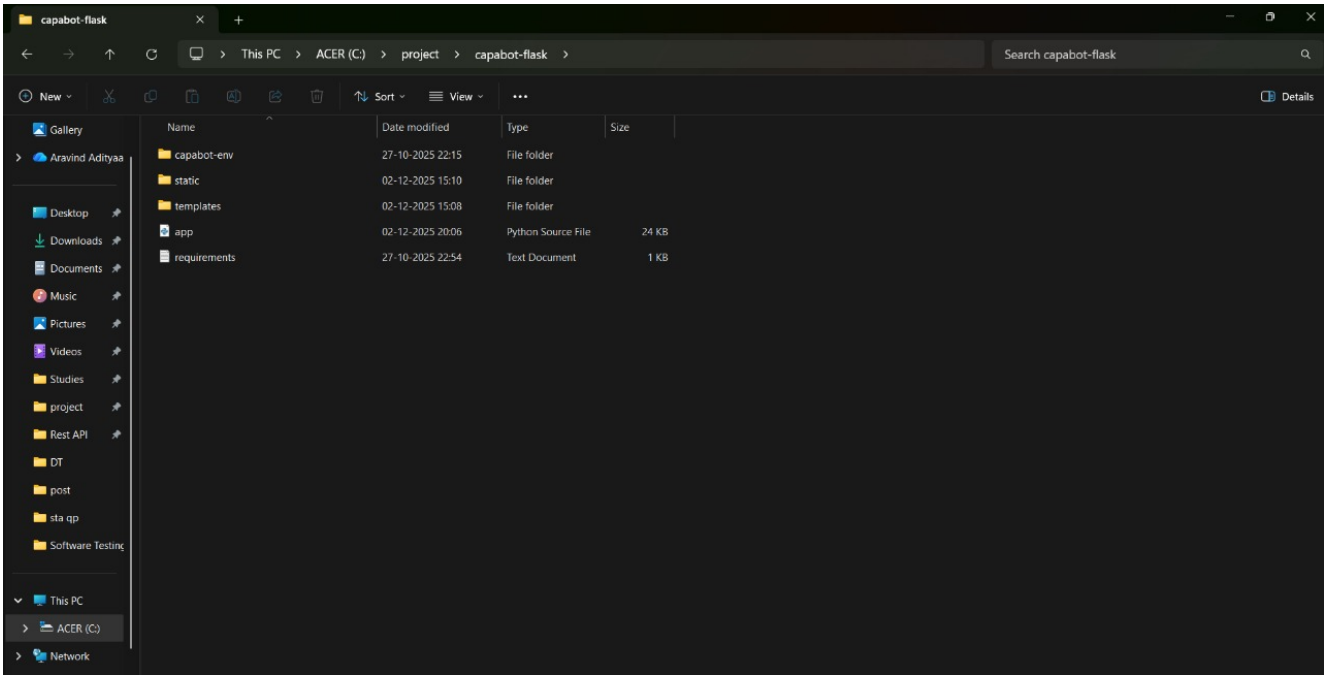
Milestone	Description	Date Achieved
Project Kickoff	Project kickoff conducted; defined objectives, scope, and team roles for CapaBot.	**07 – Nov – 2025**
Prototype / First Draft	Developed the first working model with TF-IDF matching, skill extraction, and initial UI.	**13 – Nov – 2025**
Mid-Term Review	Completed Milestone 3 review; enhanced skill categorization, recommendations, and preprocessing pipeline.	**20 – Nov – 2025**
Final Submission	Submitted the final CapaBot system with full functionality, chatbot integration, and documentation.	**02 – Dec – 2025**
Presentation	Delivered the final presentation showcasing system workflow, analysis accuracy, and end-to-end features.	**02 – Dec – 2025**

5. Project execution details

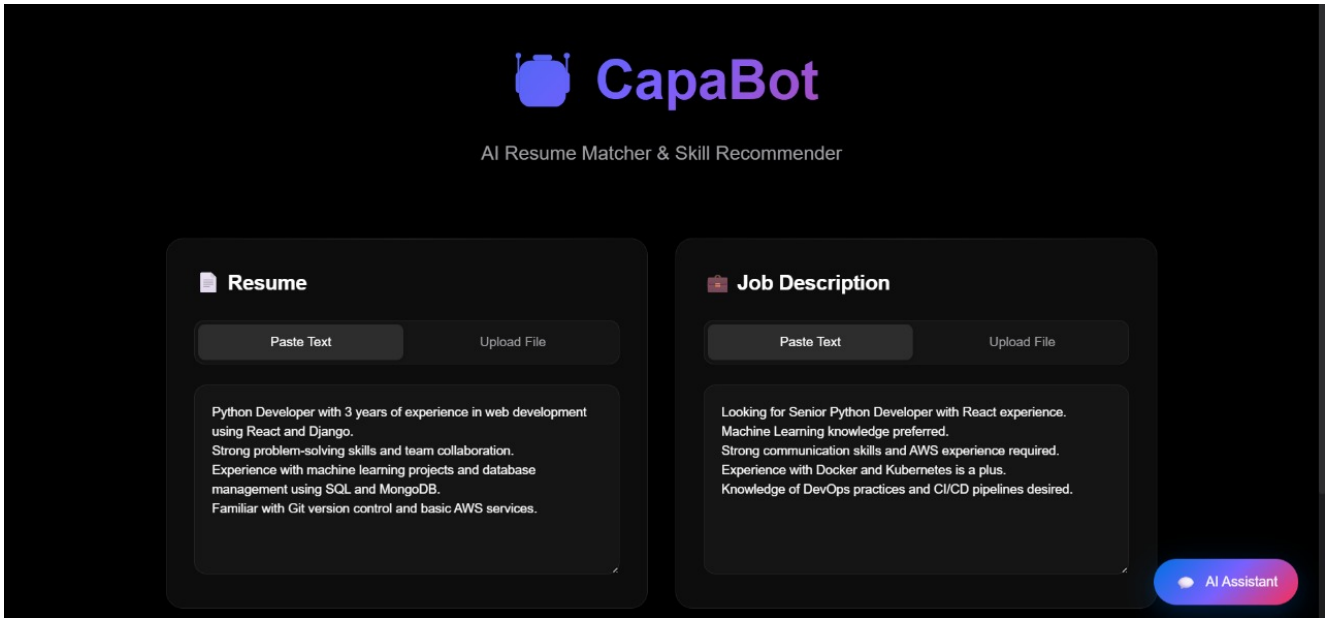
- The project began with defining the objectives, scope, and understanding the requirements for an AI-based resume–job matching system.
- Sample resumes and job descriptions were collected to analyze patterns, skill variations, and keyword structures.
- Data preprocessing pipelines were implemented using NLP techniques such as text normalization, stop-word removal, and vectorization preparation.
- A resume parsing module was developed to extract text from PDF and DOCX formats using PyPDF2 and python-docx libraries.
- Job description analysis was implemented to identify required technical and soft skills through keyword extraction and pattern-based detection.
- The matching engine was built using TF-IDF vectorization and cosine similarity to generate an objective match score between resume and job description.
- A skill categorization system was created, grouping skills into programming, web, cloud, DevOps, data science, and soft-skill domains.
- A recommendation engine was added to suggest missing skills and provide learning resources to help users improve their job alignment.
- The web interface was developed using Flask, HTML, CSS, JavaScript, and Chart.js for visual match score representation.
- Gemini Pro–based AI chatbot integration was completed to provide interactive guidance and explain analysis results to users.
- The system underwent multiple test cycles with varied resume and job description inputs to validate accuracy and reliability.
- Comprehensive project documentation and a final presentation were prepared to conclude and demonstrate the completed CapaBot system.

6. Snapshots / Screenshots

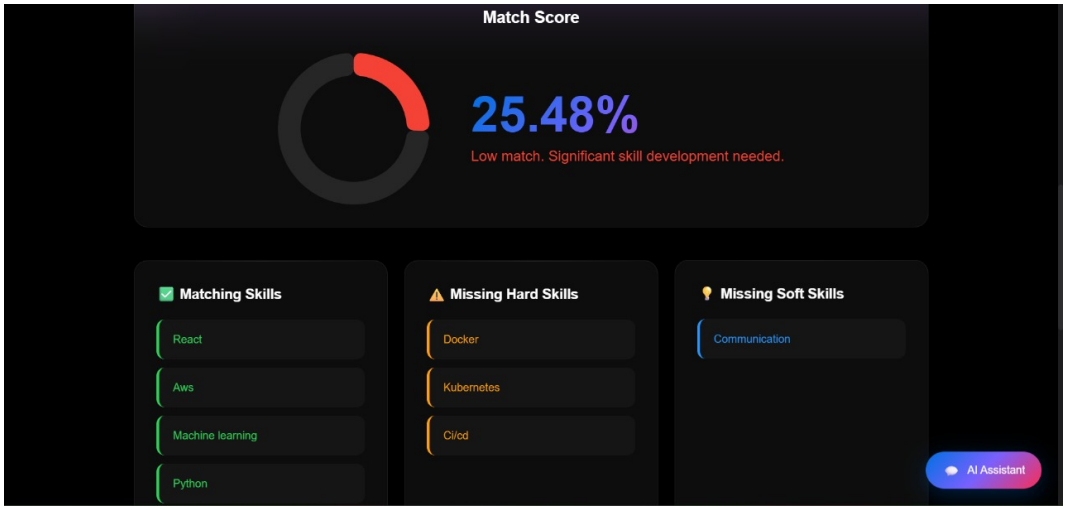
- Project Directory Structure



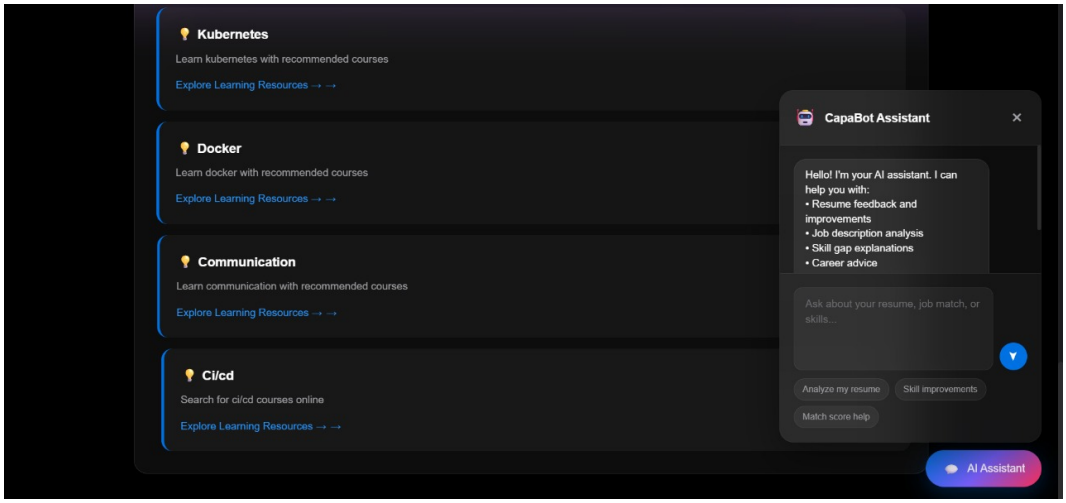
- CapaBot User Interface – Input Section



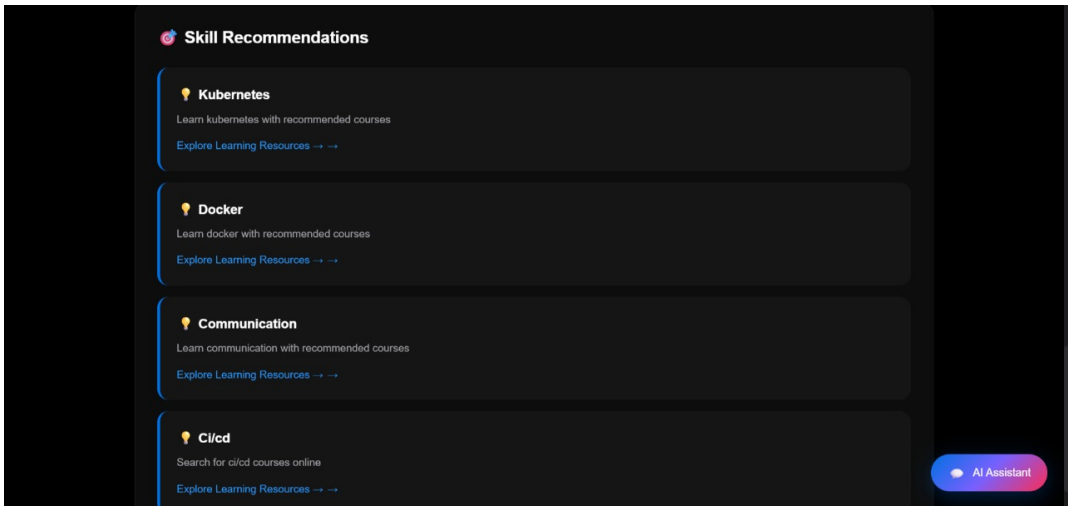
- Resume–Job Match Score Visualization



- Skill Analysis Dashboard



- Skill Recommendations & AI Assistant



7. Challenges Faced

- **Diverse Resume Formats:** Handling resumes in PDF and DOCX formats with varying layouts, fonts, and structures made text extraction and parsing challenging.
- **Inconsistent Skill Naming:** Skills appeared in different forms (e.g., *Machine Learning*, *ML*, *ML Engineer*), requiring careful normalization and keyword grouping.
- **Unstructured Job Descriptions:** Job postings often contained long paragraphs and mixed information, making it difficult to accurately extract required skills.
- **Text Preprocessing Complexity:** Removing noise while preserving key technical keywords required multiple rounds of tuning in the preprocessing pipeline.
- **Match Score Accuracy:** Achieving reliable similarity scores using TF-IDF without overfitting to keywords demanded careful feature engineering.
- **Skill Recommendation Mapping:** Ensuring that missing skills matched the job requirements meaningfully required constructing a well-structured skills database.
- **Performance Optimization:** Processing larger resumes or multiple inputs without delays required optimizing both file extraction and vectorization steps.
- **Frontend–Backend Integration:** Aligning backend outputs (JSON, skills, scores) with dynamic UI elements such as charts and tables required additional development effort.
- **Testing Across Variations:** Ensuring consistent performance with resumes of different styles, lengths, and domains was time-consuming but critical.
- **Resource Limitations:** Limited access to large datasets and heavy NLP models required us to design efficient and lightweight systems.

8. Learnings & Skills Acquired

- **Natural Language Processing (NLP):** Learned text cleaning, normalization, tokenization, and extracting meaningful information from resumes and job descriptions.
- **Machine Learning & Similarity Models:** Gained hands-on experience with TF-IDF vectorization and cosine similarity for semantic matching.
- **Resume Parsing Techniques:** Acquired skills in processing documents using PyPDF2 and python-docx, and extracting structured information.
- **Skill Categorization & Recommendations:** Learned to identify skill gaps and map them to targeted learning paths using a curated skill database.
- **Web Development & UI Integration:** Developed experience with Flask, HTML, CSS, JavaScript, and Chart.js for building intuitive dashboards.
- **AI Chatbot Integration:** Gained knowledge in using Gemini Pro API to build a contextual, interactive Q&A assistant.
- **Data Preprocessing:** Improved abilities in cleaning and normalizing unstructured text for accurate analysis.
- **Problem-Solving & Optimization:** Learned how to debug system errors, improve performance, and ensure reliability across use cases.
- **Project Management:** Strengthened planning, coordination, and documentation skills throughout the 8-week development cycle.
- **Testing & Validation:** Learned to validate the system using diverse resume–job samples for stable and repeatable accuracy.

9. Testimonials from team

Working on the CapaBot project was a valuable learning experience for all of us. As a team, we gained practical exposure to NLP, similarity modeling, and building a full-stack AI application from scratch. Collaborating on different modules—resume parsing, skill extraction, similarity scoring, UI design, and chatbot integration—helped us improve both our technical skills and our ability to work effectively as a team. This project strengthened our confidence in developing real-world AI solutions and gave us a clearer understanding of how machine learning can be applied to solve practical problems in recruitment and career development.

10. Conclusion

Working on CapaBot: AI Resume Matcher and Skill Gap Analyzer has been a highly rewarding and practical learning experience. The project provided deep exposure to Natural Language Processing, machine learning concepts, text similarity models, and full-stack development.

By developing modules for resume parsing, job analysis, similarity scoring, and skill recommendations, we were able to convert theoretical concepts into a functional real-world application. Integrating the Gemini Pro chatbot further enhanced the system's interactivity and practical relevance.

This internship strengthened our technical, analytical, and problem-solving abilities. Collaborating as a team helped us improve communication, planning, and execution skills. Overall, this project aligns strongly with our academic interests

11. Acknowledgements

We express our sincere gratitude to **Infosys Springboard** for providing us with the opportunity to work on the CapaBot project and gain practical experience in AI and Machine Learning.

We would also like to thank our mentor, **Sangeetha**, for her continuous guidance, constructive feedback, and support throughout the internship. Her insights were instrumental in helping us address challenges effectively and improve the overall quality of our work.

We appreciate the contribution of all team members for their collaboration, commitment, and consistent effort toward the successful completion of this project.

Finally, we acknowledge everyone who supported our learning and contributed, directly or indirectly, to the completion of the CapaBot project.