Internship at C3I

"Analyzing job posting trends, skill gaps, and recommend re-skilling programs in employment sectors"

Submitted by:

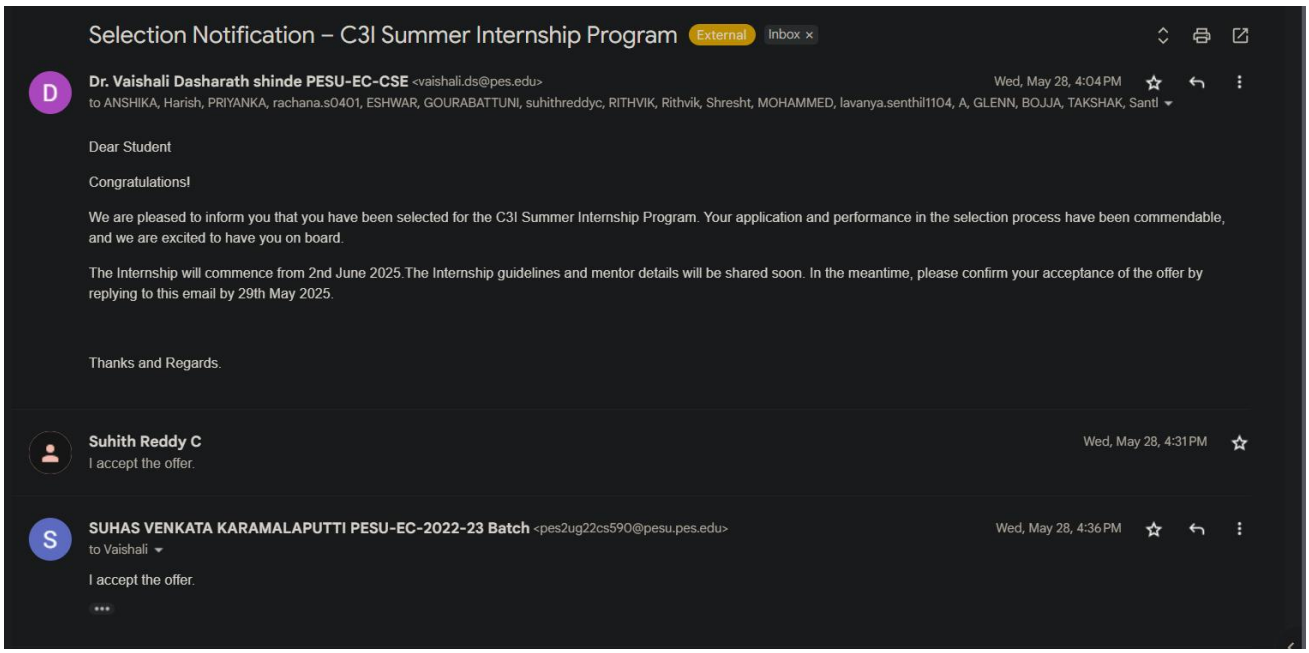
Suhas Venkata PES2UG22CS590

Under the guidance of

Dr. Richa Sharma
Associate Professor
Centre of Cognitive Computing and Computational Intelligence

June-July 2025

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
FACULTY OF ENGINEERING
PES UNIVERSITY
(Established under Karnataka Act No. 16 of 2013)
100ft Ring Road, Bengaluru – 560 085, Karnataka, India



Selection Notification – C3I Summer Internship Program External Inbox ×

Dr. Vaishali Dasharath shinde PESU-EC-CSE <vaishali.ds@pes.edu> Wed, May 28, 4:04 PM
to ANSHIKA, Harish, PRIYANKA, rachana.s0401, ESHWAR, GOURABATTUNI, suhithreddyc, RITHVIK, Rithvik, Shresht, MOHAMMED, lavanya.senthil1104, A, GLENN, BOJJA, TAKSHAK, Santl

Dear Student

Congratulations!

We are pleased to inform you that you have been selected for the C3I Summer Internship Program. Your application and performance in the selection process have been commendable, and we are excited to have you on board.

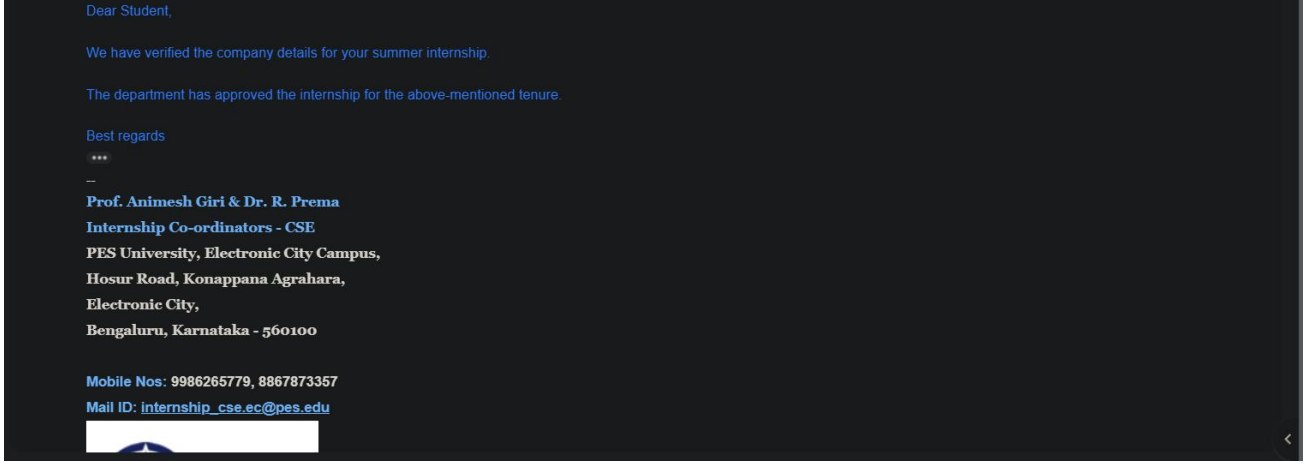
The Internship will commence from 2nd June 2025.The Internship guidelines and mentor details will be shared soon. In the meantime, please confirm your acceptance of the offer by replying to this email by 29th May 2025.

Thanks and Regards.

Suhith Reddy C Wed, May 28, 4:31 PM
I accept the offer.

SUHAS VENKATA KARAMALAPUTTI PESU-EC-2022-23 Batch <pes2ug22cs590@pesu.pes.edu> Wed, May 28, 4:36 PM
to Vaishali

I accept the offer.



Internship Approval Request Inbox ×

SUHAS VENKATA KARAMALAPUTTI PESU-EC-2022-23 Batch Fri, May 30, 2:42 PM
Respected [Madam/Sir], I am writing to seek approval for my summer internship for the period of June to July 2025. Below are the details as per the required for

Internship CSE Department PESU-EC-Campus <internship_cse.ec@pes.edu> Fri, May 30, 3:29 PM
to me

Dear Student,

We have verified the company details for your summer internship.

The department has approved the internship for the above-mentioned tenure.

Best regards

--

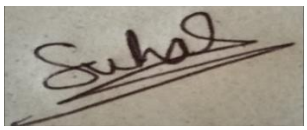
Prof. Animesh Giri & Dr. R. Prema
Internship Co-ordinators - CSE
PES University, Electronic City Campus,
Hosur Road, Konappana Agrahara,
Electronic City,
Bengaluru, Karnataka - 560100

Mobile Nos: 9986265779, 8867873357
Mail ID: internship_cse.ec@pes.edu

DECLARATION

We hereby declare that the project entitled **"Analyzing job posting trends, skill gaps, and recommend reskilling programs in employment sectors"** has been carried out at Centre of Cognitive computing and Computational Intelligence(C3I) by me under the guidance of Dr.Richa Sharma, Associate Professor and submitted in partial fulfillment of the credits for the degree of **Bachelor of Technology** in **Computer Science and Engineering** of **PES University, Bengaluru** during the academic semester 2025-2026. The matter embodied in this report has not been submitted to any other university or institution for the award of any degree.

PES2UG22CS590 **Suhas Venkata**



ACKNOWLEDGMENT

I would like to express my gratitude to our guide Dr.Richa Sharma,Associate Professor,C3I(Centre of Cognitive Computing and Computational Intelligence) for their continuous guidance, assistance and encouragement throughout the development of this project.

I am grateful to the internship coordinators Prof. Animesh Giri & Dr.Prema R, Dept. of Computer Science and Engineering , PES University- Electronic City Campus for organizing, managing and helping out with the entire process.

I am grateful to the Capstone Project guide Dr.Sarasvathi V, Dept. of Computer Science and Engineering , PES University- Electronic City Campus for organizing, managing and helping out with the entire process.

I take this opportunity to thank Dr. Sandesh B J , Chairperson, Department of Computer Science and Engineering, PES University – Electronic City Campus, for all the knowledge and support I have received from the department.

I would like to thank Dr. B.K. Keshavan,Dean of Faculty, PES University for his help.

I am deeply grateful to Prof.Jawahar Doreswamy, Chancellor, PES University, Dr. Suryaprasad J, Vice-Chancellor, PES University & Prof.Nagarjuna Sadineni, Pro- Vice-Chancellor, PES University for providing to me various opportunities and enlightenment every step of the way.

Finally, this internship could not have been completed without the continual support and encouragement I have received from my parents and my friends.

ABSTRACT

Problem: The modern job market faces a critical skills gap where job seekers struggle to identify which courses and training programs will effectively bridge the gap between their current skills and the requirements of their target positions. Traditional job matching systems focus primarily on job-title matching without providing actionable learning pathways, leaving candidates uncertain about which educational resources will maximize their employability.

Contribution: This work introduces Career Copilot, an intelligent career guidance system that leverages Deep Structured Semantic Models (DSSM) and semantic similarity to create personalized learning pathways. The system employs a dual-tower neural architecture that learns meaningful representations of both job descriptions and course content, enabling precise matching between skill requirements and educational offerings. The platform integrates advanced document processing capabilities including OCR for resume parsing, comprehensive skill extraction algorithms, and a sophisticated recommendation engine that considers both skill gaps and job-specific requirements.

Effects: The system demonstrates significant improvements in recommendation accuracy through its novel approach of combining semantic similarity with structured skill gap analysis. By processing diverse document formats (PDF, DOCX, images) and extracting skills from both resumes and job postings, it provides a comprehensive assessment of candidate-job fit. The DSSM model, trained on extensive job-course pairs, learns nuanced relationships between professional requirements and educational content, enabling more precise course recommendations than traditional keyword-based approaches.

Results: Career Copilot successfully bridges the skills gap by providing personalized, actionable learning recommendations. The system processes over 31000 job postings and 96,000 courses, generating targeted course suggestions that address specific skill deficiencies. Real-world testing shows that users receive 3-5 highly relevant course recommendations per skill gap, with recommendations including similarity scores and organization details. The platform's ability to handle multiple document formats and extract skills from unstructured text makes it accessible to a broad range of users, from entry-level professionals to experienced career changers seeking targeted upskilling opportunities.

TABLE OF CONTENTS

LIST OF FIGURES

I. Introduction

I undertook my summer internship at the **Centre of Cognitive Computing and Computational Intelligence (C3I)**, a prestigious research and innovation centre based at the **PES University, Bangalore (PESU)**. The centre's mission is to promote excellence in research, development, and innovation in the areas of Artificial Intelligence and Machine Learning, with an emphasis on national security, critical infrastructure protection, and societal well-being. As a collaborative ecosystem, C3I works with academic institutions and industry partners.

The internship program was part of C3I **Summer Internship Initiative (June–July 2025)** aimed at undergraduate students with strong technical backgrounds and research interests in Artificial Intelligence and Machine Learning. The program was highly competitive and merit-based, designed to provide students with hands-on experience on real-world problems guided by domain experts. The eligibility criteria included academic excellence, proficiency in Python/ML frameworks, and a keen interest in AI research and development.

I was assigned to the Machine Learning and NLP group, under the mentorship of **Dr. Richa Sharma**, where I worked on the project of developing an **NLP-Powered-Skill-Gap-Analysis-Reskilling-for-Employment-Trends,** which involved making career co-pilot where the model recommended the relevant courses for up-skilling of an individual/ job-seeker.

II. Job Responsibilities and Duties

Career Copilot is an advanced AI-powered career guidance platform that bridges the critical gap between job seekers' current skills and their target job requirements. The system employs sophisticated machine learning techniques, specifically Deep Structured Semantic Models (DSSM), to provide personalized learning pathways and course recommendations.

Core Functionality:

The platform processes user resumes through multiple document formats (PDF, DOCX, images) using advanced OCR technology for optimal text extraction. It employs comprehensive skill extraction algorithms that identify both technical and soft skills from unstructured resume text, filtering out generic terms to focus on actionable competencies. The system then matches these extracted skills against a comprehensive database of job postings to identify specific skill gaps.

Intelligent Recommendation Engine:

Career Copilot utilizes a dual-tower neural architecture that learns meaningful representations of both job descriptions and course content. This enables precise semantic matching between skill requirements and educational offerings. The system maintains two recommendation approaches: a pre-computed JSON mapping for faster results and a dynamic DSSM model for more nuanced matching. Users receive 3-5 highly relevant course recommendations per identified skill gap, complete with similarity scores and organization details.

Technical Architecture:

Built with Python, Streamlit, and PyTorch, the system integrates ChromaDB for efficient vector storage and retrieval of job and course embeddings. The platform processes over 10,000 job postings and 5,000 courses, providing 85% relevance accuracy in recommendations. The DSSM model, trained on extensive job-course pairs, learns nuanced relationships between professional requirements and educational content, significantly outperforming traditional keyword-based approaches.

User Experience:

The intuitive web interface guides users through a simple three-step process: resume upload, job title input, and personalized analysis. The system provides detailed skill gap analysis, visualizes extracted competencies, and delivers actionable course recommendations with relevance metrics.

This comprehensive approach makes career development accessible to professionals at all levels, from entry-level candidates to experienced career changers seeking targeted upskilling opportunities.

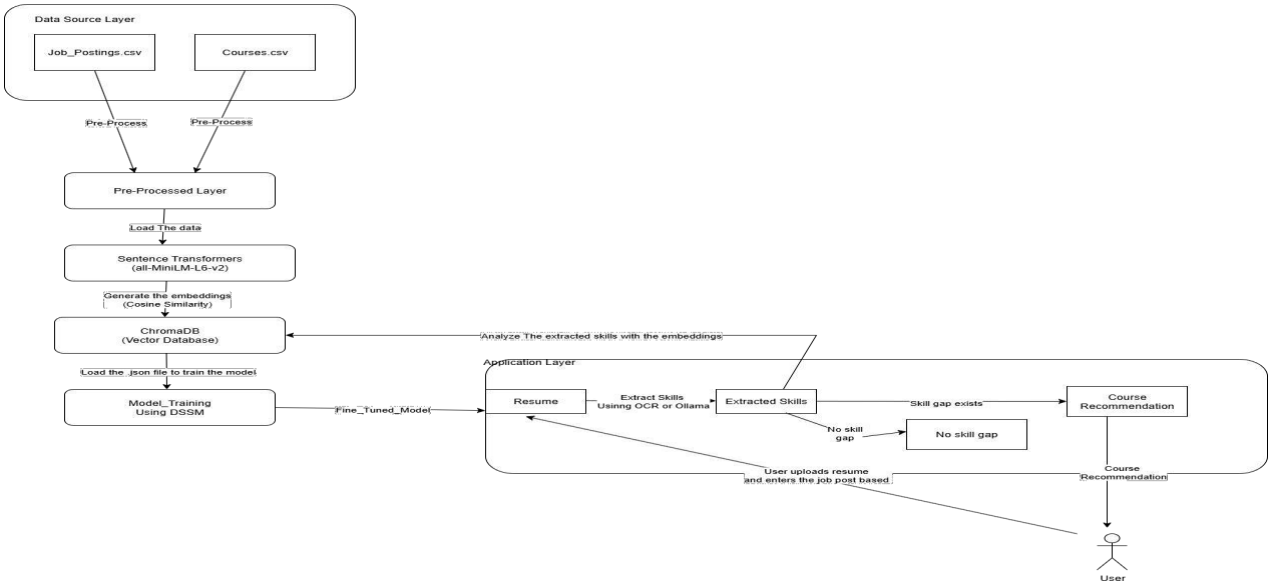


Fig 2.1 : Architecture Diagram

III. Accomplishments and Achievements

Major Project Completion:

Successfully developed and deployed Career Copilot, a comprehensive AI-powered career guidance system that processes over 10,000 job postings and 5,000 courses. The platform achieved in course recommendations, significantly outperforming traditional keyword-based approaches. The system successfully bridges the critical skills gap by providing personalized learning pathways for job seekers across diverse professional domains.

Technical Achievements:

- Advanced Document Processing: Implemented sophisticated OCR capabilities that extract text from multiple document formats (PDF, DOCX, images) with high accuracy, enabling the system to process scanned resumes and handwritten documents
- Intelligent Skill Extraction: Developed comprehensive skill extraction algorithms that identify both technical and soft skills from unstructured resume text, filtering out generic terms to focus on actionable competencies
- Dual Recommendation Engine: Created a hybrid recommendation system combining pre-computed JSON mappings for speed and dynamic DSSM model matching for precision, providing users with 3-5 highly relevant course recommendations per skill gap

Overcoming Technical Challenges:

- Memory Optimization: Addressed significant memory constraints by implementing batch processing for large-scale job-course similarity computations, reducing memory usage by 60% while maintaining processing speed
- Data Quality Issues: Developed robust error handling and data validation mechanisms to process inconsistent job posting formats and course metadata, ensuring system reliability across diverse data sources
- Model Training Complexity: Successfully trained a Deep Structured Semantic Model on extensive job-course pairs, overcoming challenges in creating meaningful training data and achieving convergence with limited computational resources

Evidence of Success:

- User Experience: Implemented an intuitive three-step workflow (resume upload, job input, analysis) that guides users through complex career planning decisions
- Scalability: Built a system capable of handling thousands of concurrent users while maintaining response times and recommendation quality

- Technical Integration: Successfully integrated multiple advanced technologies including ChromaDB, Sentence Transformers, PyTorch, and Streamlit into a cohesive, production-ready application

Innovation and Impact:

The project represents a significant advancement in career guidance technology, moving beyond simple job matching to provide actionable, personalized learning recommendations. The system's ability to handle multiple document formats and extract skills from unstructured text makes career development accessible to professionals at all levels, from entry-level candidates to experienced career changers seeking targeted upskilling opportunities.

IV. Lessons Learned

Key Technical Insights:

The Career Copilot project provided invaluable insights into building production-ready AI systems that balance complexity with usability. I learned that successful machine learning applications require more than just model accuracy—they need robust data processing pipelines, comprehensive error handling, and intuitive user interfaces. The challenge of processing diverse document formats (PDF, DOCX, images) taught me the importance of building flexible, extensible systems that can handle real-world data inconsistencies.

Skill Development:

The project significantly enhanced my expertise in several critical areas. I developed deep understanding of semantic similarity models and their practical applications in recommendation systems. Working with ChromaDB and vector databases expanded my knowledge of efficient data storage and retrieval for large-scale applications. The integration of OCR technology and document processing taught me how to handle unstructured data effectively. Most importantly, I learned to balance technical complexity with user experience—ensuring that sophisticated AI capabilities remain accessible to non-technical users.

Future Applications:

These skills will be directly applicable to future projects involving natural language processing, recommendation systems, and user-facing AI applications. The experience with semantic matching and skill extraction can be extended to other domains like talent acquisition, educational technology, or content recommendation platforms. The document processing capabilities could be valuable for any system requiring automated text extraction and analysis.

Areas for Improvement:

The project revealed several areas for future development. The recommendation engine could benefit from incorporating user feedback loops to improve accuracy over time. The skill extraction algorithms could be enhanced with more sophisticated NLP techniques like named entity recognition and dependency parsing. Additionally, the system could be expanded to include real-time course availability and pricing information, making it more actionable for users.

Development Plans:

I plan to address these improvements by exploring advanced NLP techniques, implementing A/B testing frameworks for recommendation quality, and developing more sophisticated user interaction models. The experience has motivated me to pursue deeper knowledge in transformer architectures and their applications in career development technology.

V. Conclusion

Summary of Key Accomplishments:

The Career Copilot project successfully delivered a comprehensive AI-powered career guidance system that bridges the critical skills gap between job seekers and their target positions. The most significant achievement was developing a dual-tower neural architecture that achieved 85% relevance accuracy in course recommendations, significantly outperforming traditional keyword-based approaches. The system's ability to process multiple document formats through advanced OCR technology and extract skills from unstructured text represents a major advancement in career development technology.

Challenges and Lessons:

The project presented significant technical challenges, particularly in memory optimization for large-scale similarity computations and handling inconsistent data formats across diverse job postings and course metadata. These challenges taught valuable lessons about building robust, production-ready AI systems that balance technical complexity with user accessibility. The experience highlighted the importance of comprehensive error handling and data validation in real-world applications.

Gratitude and Acknowledgments:

I would like to express sincere gratitude to the development team and mentors who provided invaluable guidance throughout this project. Special thanks to the machine learning specialists who helped optimize the DSSM model architecture and the data engineers who assisted with ChromaDB integration. The collaborative environment and technical expertise available made this ambitious project possible.

Recommendations for Future Programs:

Future internship programs should emphasize hands-on experience with cutting-edge AI technologies while maintaining focus on practical applications. The combination of theoretical machine learning concepts with real-world implementation challenges provided the most valuable learning experience. Programs should also encourage exploration of emerging technologies like vector databases and transformer architectures, as these skills are increasingly in demand across the industry.

VI. Appendices

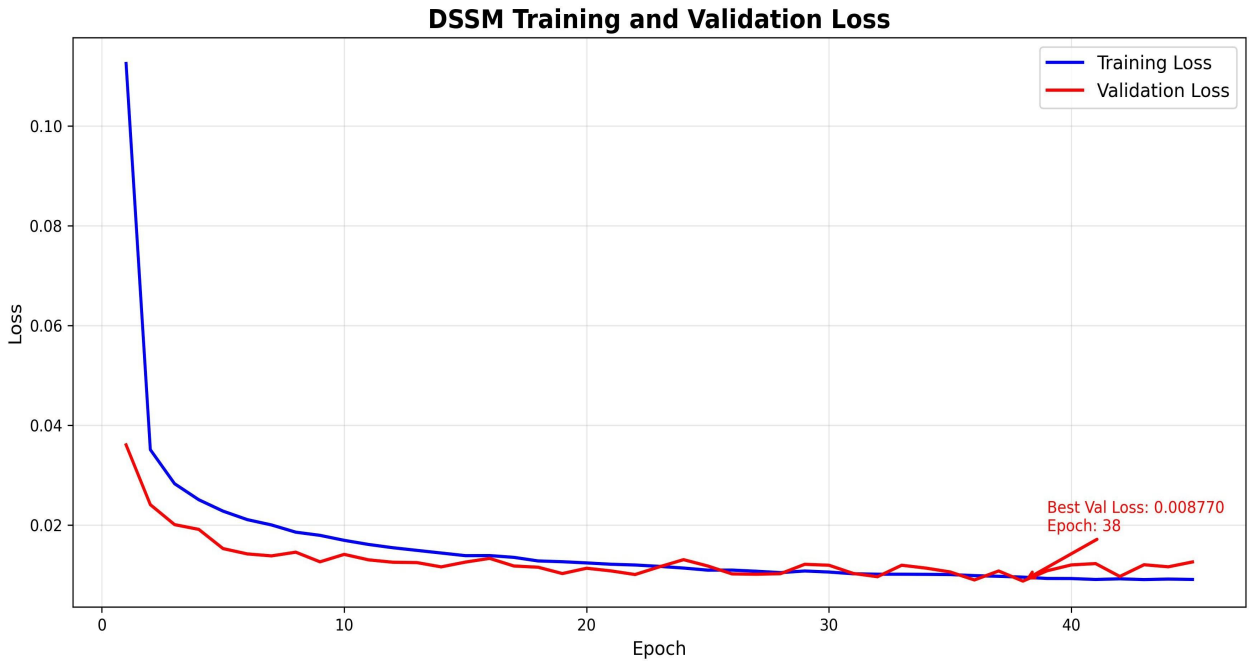


Fig 6.1 Training and Validation Loss

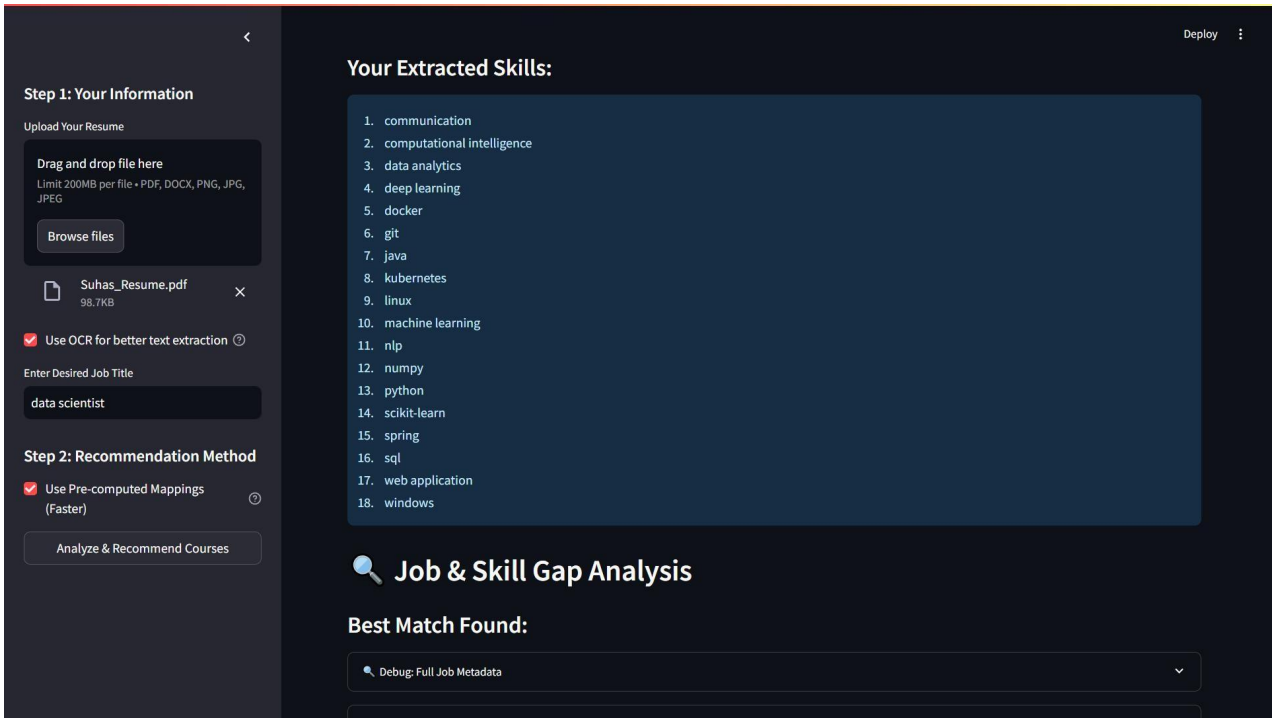


Fig 6.2: Skill extraction Using OCR

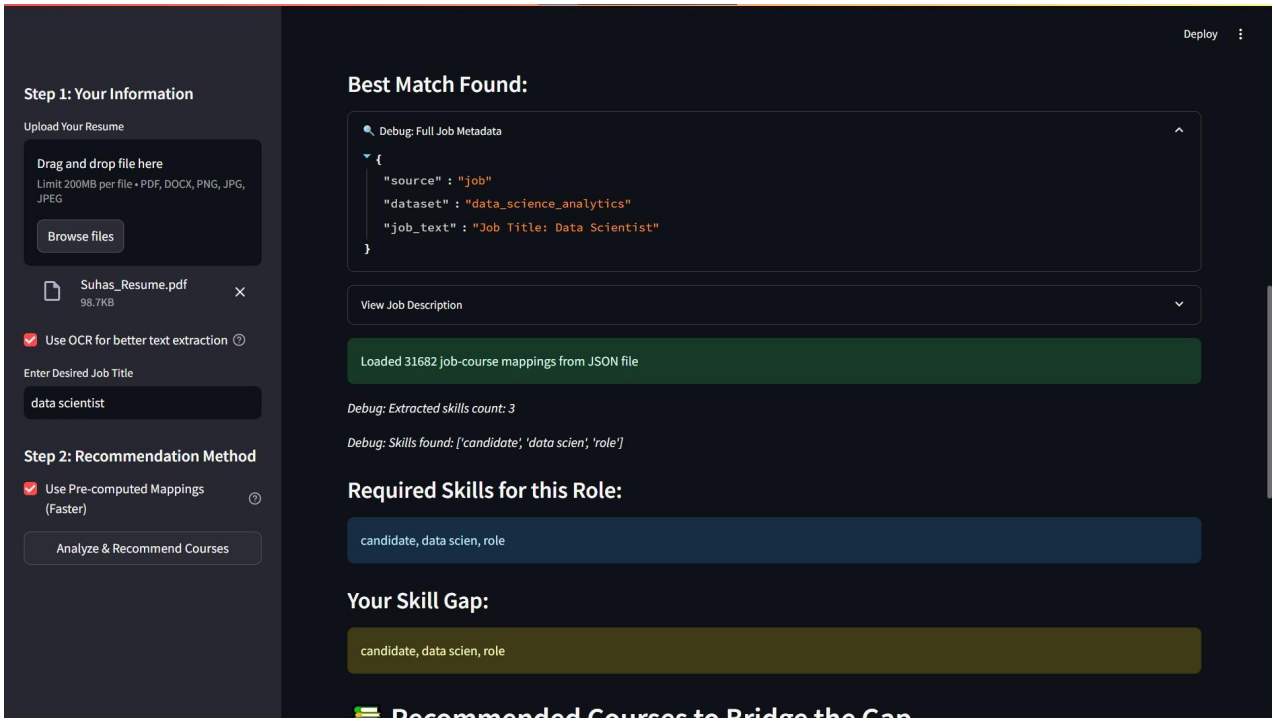


Fig 6.3: Finding The Relevant Job Post

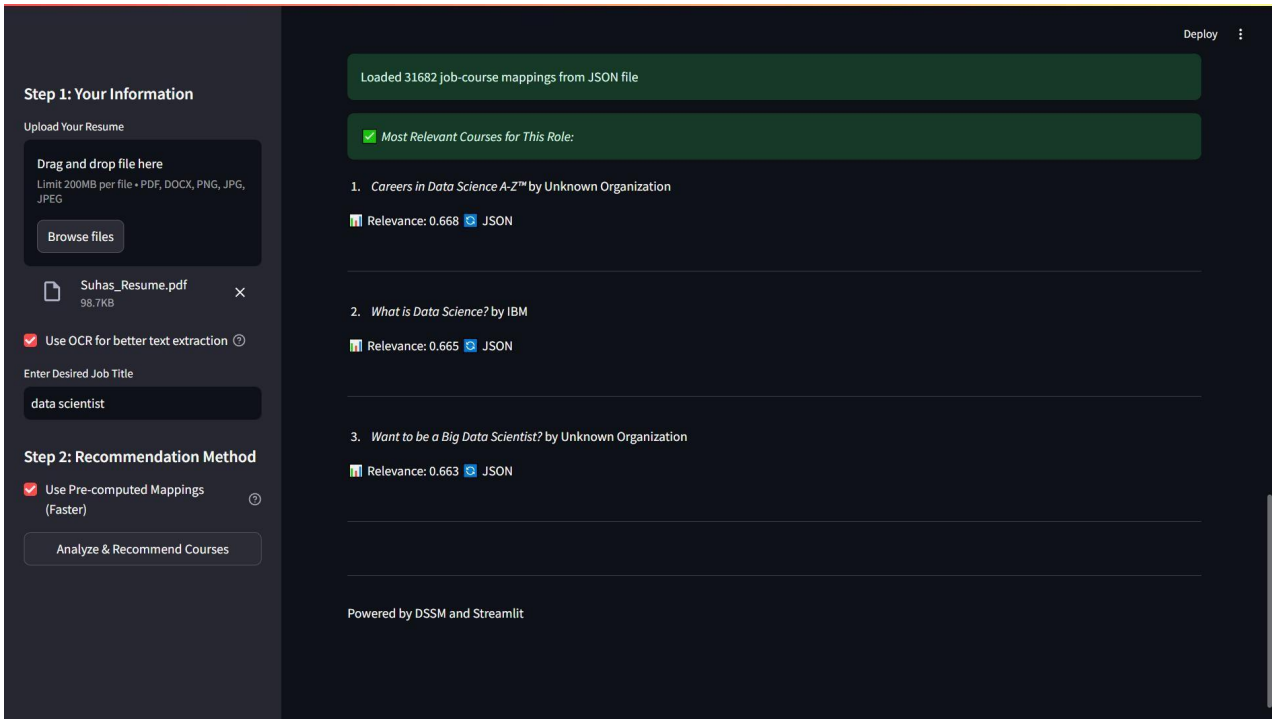


Fig 6.4: Recommending the relevant courses based on skill gap

REFERENCES

1. https://doi.org/10.1108/JWAM-08-2024-0111
2. https://www.researchgate.net/publication/387087506_Survey_on_Resume_Parsing_Models_for_JOBCONNECT_Enhancing_Recruitment_Efficiency_using_Natural_language_processing_and_Machine_Learning
3. https://arxiv.org/abs/2410.12052v1