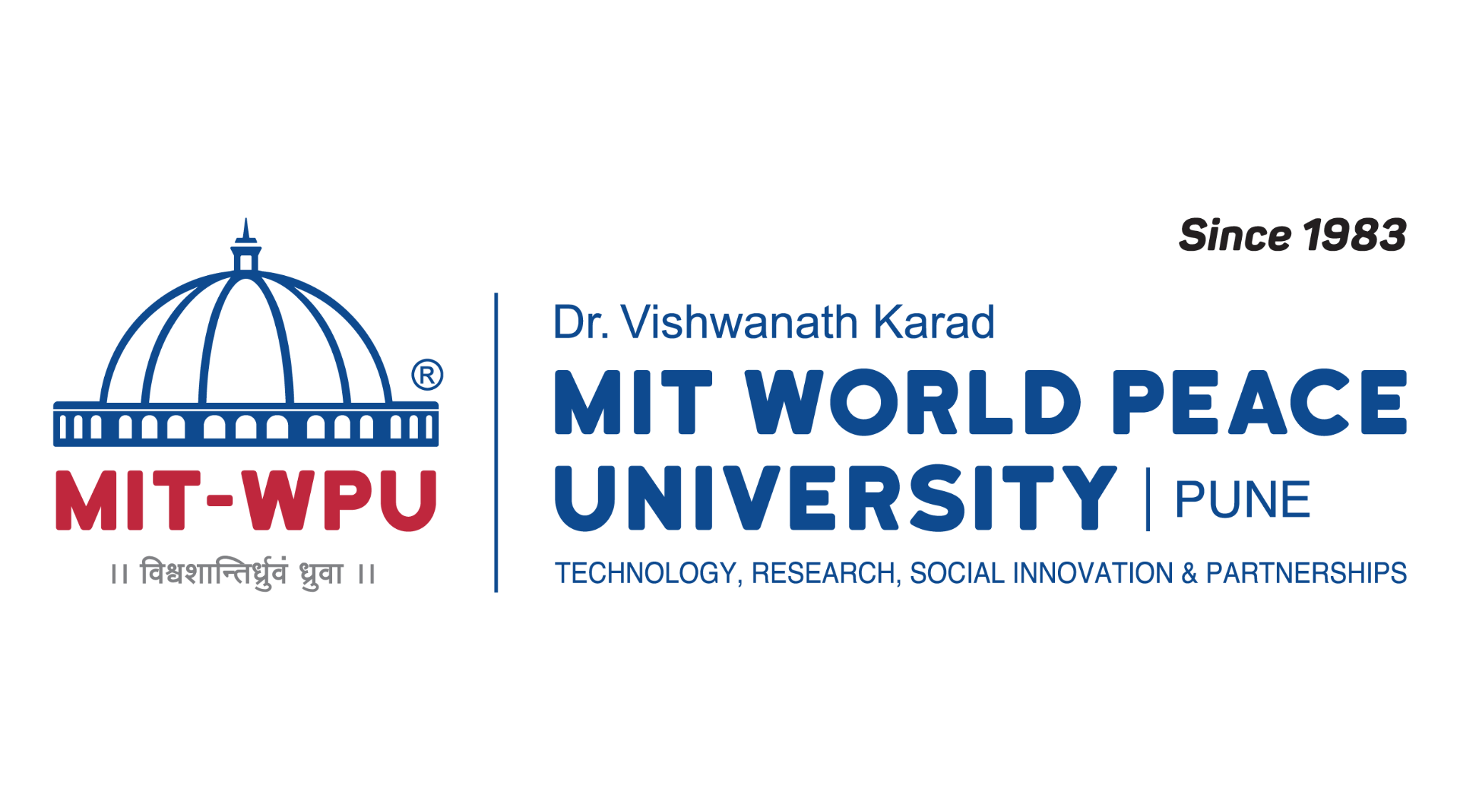
****

**MIT-World Peace University (MIT-WPU)**

**Faculty of Engineering**

**School of Computer Science & Engineering**

**Department of Computer Engineering & Technology**

**AI-Powered Resume Parsing and Recommendation System**

**Avani Halbe - 1032222238**

**Tiya Das - 1032222252**

**Riya Kondawar - 1032222226**

**Saee Patil - 1032222225**

**Saie Kale - 1032222245**

Guided by

**Prof. Chaitali Chandankhede**

**AI-Powered Resume Parsing and Recommendation System**

# **1. Abstract**

In today’s competitive job market, students and professionals must create resumes that effectively highlight their skills and qualifications. Traditional resume screening is time-consuming and often lacks objectivity. The increasing volume of job applications makes it challenging for recruiters to manually assess each resume with precision. To address this challenge, we have developed an AI-powered Resume Parsing and Recommendation System designed to enhance resume quality and streamline candidate evaluation. The system automates resume extraction, parsing, and enhancement using Natural Language Processing (NLP) and Machine Learning (ML) techniques, ensuring accuracy, efficiency, and fairness in the hiring process.

Our system focuses on three key aspects: (1) Efficiently extracting and structuring data from resumes across various formats (PDF, DOCX, TXT), (2) Utilizing NLP algorithms to identify and rank essential skills, qualifications, and experiences, and (3) Enhancing resumes through industry-specific keyword optimization, increasing alignment with employer expectations. Additionally, the system provides personalized recommendations for skills enhancement and job matching based on applicant profiles. By integrating AI-driven techniques such as Named Entity Recognition (NER), TF-IDF vectorization, and deep learning models like BERT, the system ensures a precise and unbiased evaluation of candidates. Furthermore, it offers actionable insights into the strengths and weaknesses of a resume, helping users refine their job application materials based on data-driven feedback.

By optimizing resumes for Applicant Tracking Systems (ATS) and improving candidate-job fit predictions, our solution significantly enhances hiring efficiency. The proposed system empowers students and job seekers to present themselves professionally, ultimately increasing their chances of securing interviews and job offers. Organizations and HR professionals benefit from the automation and intelligence embedded in the system, reducing hiring time and minimizing bias in candidate evaluation. With an ever-evolving job market, our AI-powered system continuously adapts to emerging industry trends, ensuring that resumes remain relevant and compelling.

**Keywords:** Natural Language Processing (NLP), Resume parsing, Resume analysis, TF-IDF, Named Entity Recognition (NER), Applicant Tracking system (ATS)

# 

# **2. Introduction**

In today’s competitive job market, an efficient and automated resume parsing and recommendation system is crucial for both job seekers and recruiters. Traditional hiring processes often involve manually screening resumes, which is time-consuming and prone to errors. The integration of Natural Language Processing (NLP) and Machine Learning (ML) techniques enables intelligent resume parsing, extracting key information such as skills, experience, and qualifications. This paper presents an **AI-powered Resume Parsing and Recommendation System** that leverages NLP for structured data extraction and ML algorithms for ranking and recommendation. The system processes resumes in multiple formats, enhances their readability, and provides tailored suggestions to improve a candidate’s job prospects. By incorporating Named Entity Recognition (NER), TF-IDF vectorization, and deep learning models like BERT, our approach ensures high accuracy in resume analysis while maintaining an interactive and user-friendly experience via a **Streamlit-based web interface**.

# 

# 

# **3. Objectives**

1. To automate resume parsing and extract structured information such as skills, experience, education, and certifications.
2. To develop an AI system capable of evaluating candidates by matching their qualifications against job descriptions.
3. To provide NLP-based suggestions for resume enhancement, ensuring alignment with industry trends and employer expectations.
4. To rank and score candidates based on predefined parameters such as skills, experience, and certifications.
5. To enable recruiters to filter candidates based on customizable criteria such as location, experience level, and technical expertise.
6. To enhance recruitment efficiency by reducing manual screening time while ensuring fair and unbiased candidate evaluation.

# 

# 

# **4. Literature Review**

Several research studies have explored AI-driven resume parsing and job recommendation systems, demonstrating the significance of Natural Language Processing (NLP) and Machine Learning (ML) in modern recruitment processes. The automation of resume screening and candidate evaluation has gained increasing attention due to its potential to enhance efficiency and reduce hiring biases. Below are some key studies that form the foundation of our project development.

Jaiswal et al. (2024) propose an NLP-based job recommendation system that leverages Cosine Similarity and Multinomial Naïve Bayes for profession prediction. The system web scrapes job postings, extracts skill requirements, and ranks jobs based on their match with a candidate’s resume. By utilizing text-based similarity techniques, the study demonstrates how job seekers can be matched to the most suitable roles based on their resume content. Additionally, the researchers highlight the challenges of industry-specific jargon and varying resume formats, emphasizing the need for adaptive NLP models that can standardize and interpret diverse resume structures. Their research underscores the importance of feature engineering in improving resume-job matching accuracy.

Artajaya et al. (2024) introduce a machine learning approach that utilizes NLP to analyze resumes, rank them based on recruiter-defined constraints, and store structured information in a knowledge base. The system retrieves additional details from candidates' LinkedIn and GitHub profiles to enhance evaluation accuracy. This study highlights the significance of integrating external sources to validate and enrich candidate data, making the resume analysis more comprehensive. The researchers explore techniques such as Named Entity Recognition (NER) to extract meaningful insights from resumes and propose a hybrid scoring mechanism combining rule-based filtering with machine learning models. The system’s ability to dynamically adapt scoring criteria based on recruiter preferences makes it an advanced tool for modern hiring.

Mohanty et al. (2023) develop an AI-based resume parser using deep learning techniques like Doc2Vec and recurrent neural networks (RNN). Their system emphasizes candidate scoring based on technical proficiency and project experience. By leveraging word embeddings and sequential learning models, the study showcases a powerful method for extracting contextual meanings from resumes. The research further delves into classification models for predicting a candidate’s suitability for different job roles. Their approach provides an in-depth comparison between traditional keyword-based extraction techniques and modern deep learning methodologies, demonstrating the superiority of context-aware systems in resume parsing.

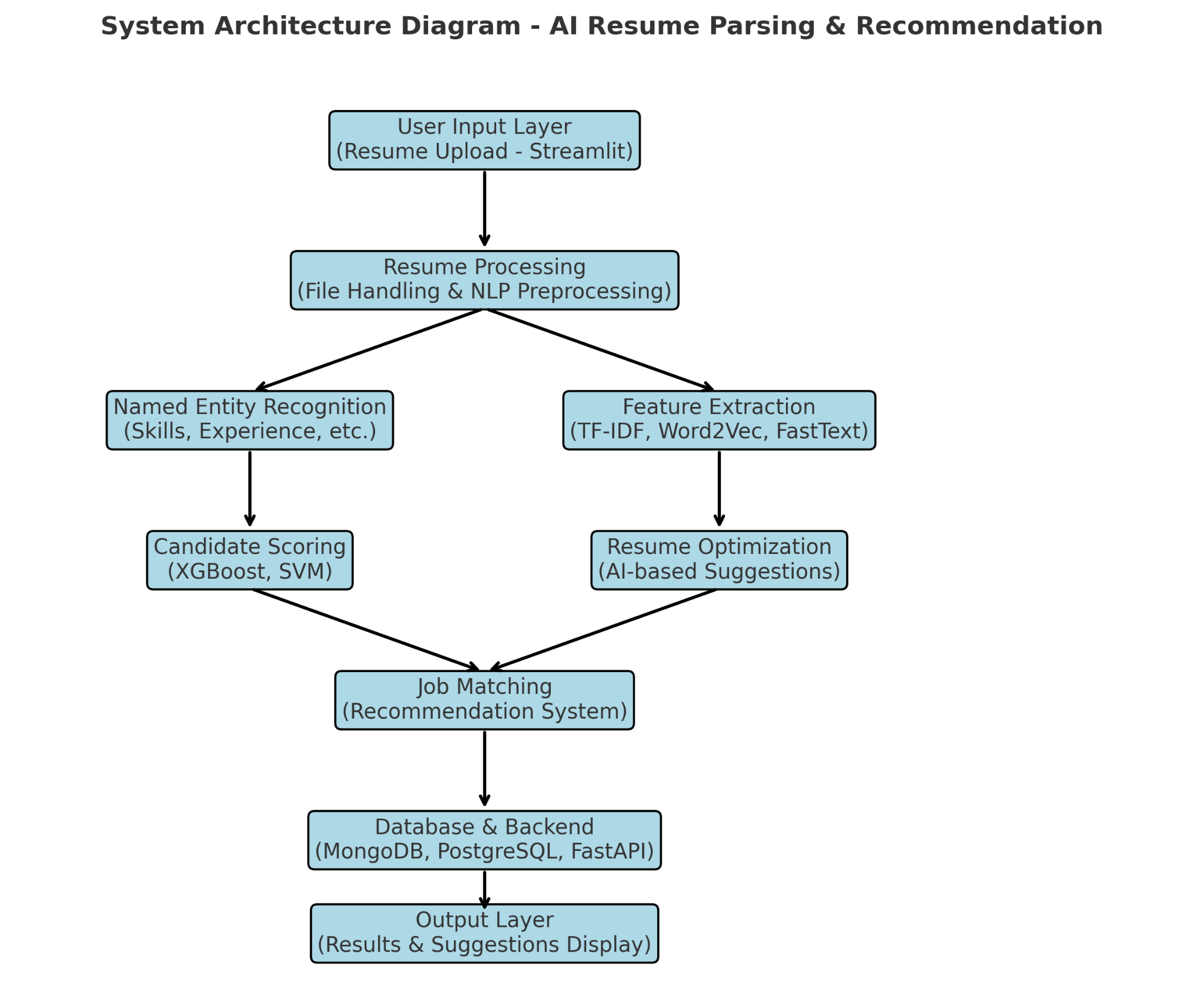
Additional research by Pundir et al. (2024) presents an advanced resume recommendation system using deep learning models for skill-based similarity analysis. The study employs transformer-based architectures such as BERT to enhance the contextual understanding of resume content. The researchers propose an adaptive learning mechanism where the system continuously improves by incorporating recruiter feedback, ensuring that resume evaluations become more precise over time. The paper also explores various feature selection techniques to optimize job matching, making it an essential reference for AI-driven recruitment systems.

Mishra et al. (2024) investigate the effectiveness of AI-powered models in providing intelligent resume recommendations and feedback. Their system incorporates Generative Pre-trained Transformers (GPT) to offer personalized suggestions for resume enhancement. The study highlights the benefits of using AI-generated feedback to align resumes with industry expectations, making candidates more competitive in the job market. Furthermore, they explore sentiment analysis in recommendation systems, identifying potential biases in resume evaluations and proposing strategies to mitigate them.

These studies highlight the growing importance of NLP in recruitment, providing foundational insights for our project’s development. By analyzing the strengths and limitations of existing systems, we aim to build an AI-powered Resume Parsing and Recommendation System that integrates advanced NLP and ML techniques, ensuring comprehensive, fair, and efficient candidate evaluations. Our approach will leverage the latest AI advancements to address challenges such as diverse resume formats, evolving job market trends, and recruiter-specific preferences, ultimately contributing to the modernization of hiring processes.

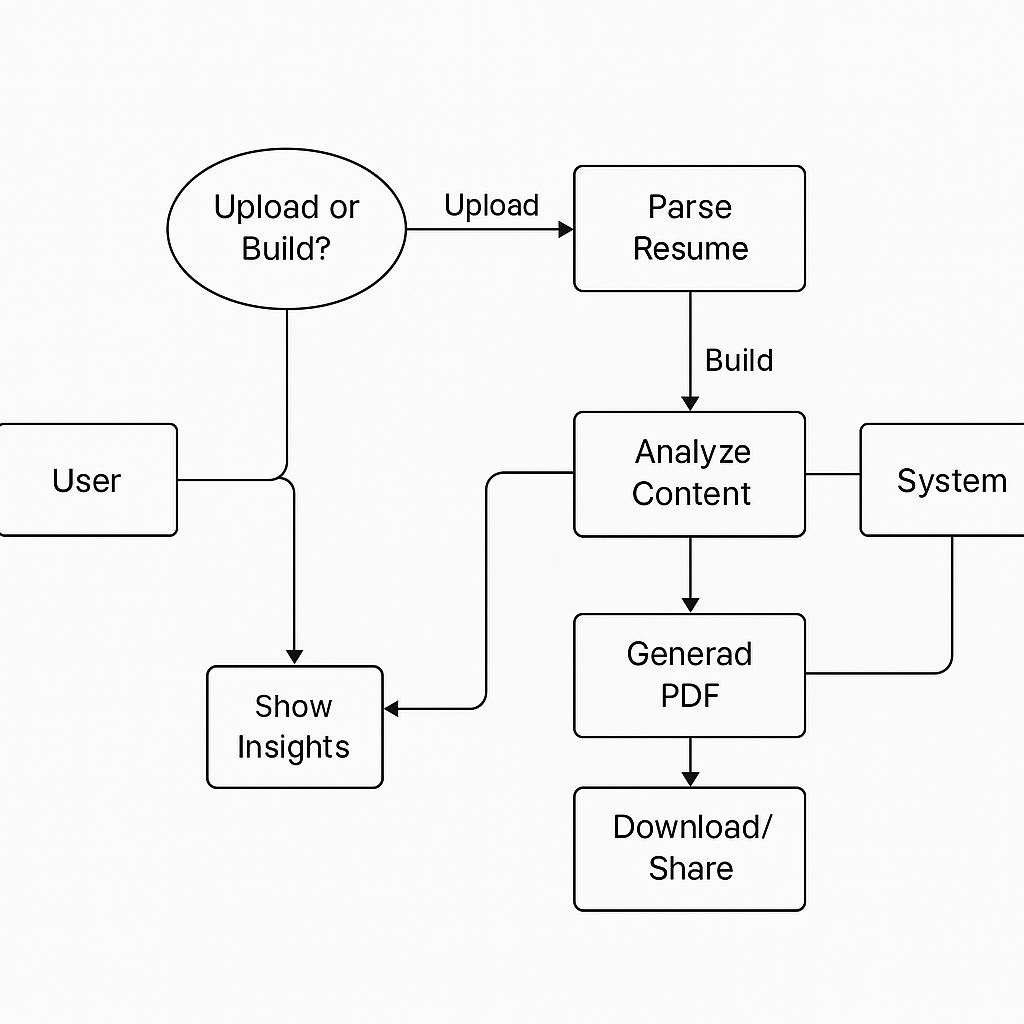
**5. Diagrams**

**5.1 System Diagram**

****

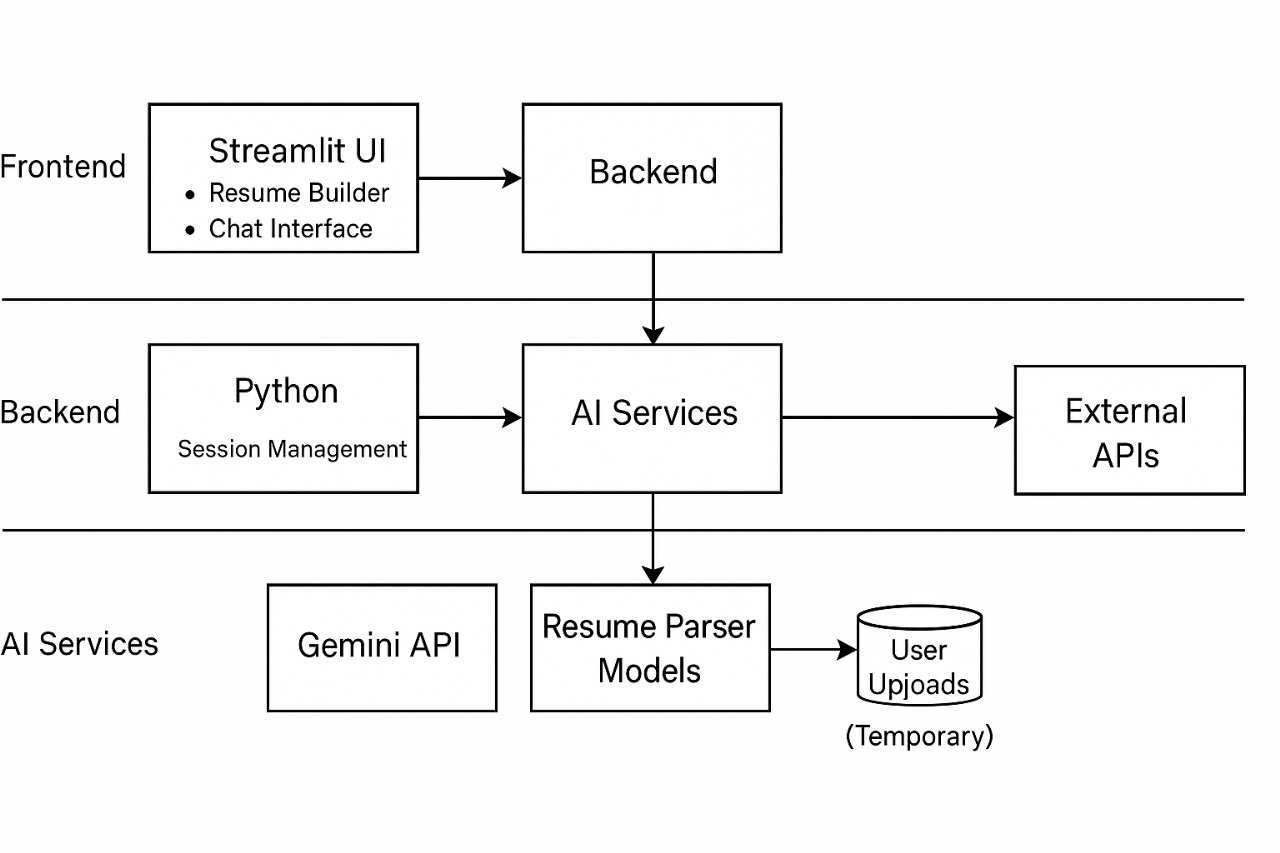
**System Architecture Diagram - AI Resume Parsing & Recommendation**

**5.2. Data flow diagram**

****

**Data flow diagram**

**5.3. Component diagram**

****

# **6. Methodology**

# **6.1 Resume Data Extraction**

Effective resume parsing begins with efficient data extraction. Our system utilizes various libraries and tools to extract textual content from different file formats:

1. PDF Resumes
2. DOCX files
3. TXT files

To ensure high-quality text extraction, the system incorporates **preprocessing techniques** such as:

1. Stop word removal
2. Lemmatization
3. Tokenization
4. Sentence segmentation
5. Whitespace normalization

### **6.2. Named Entity Recognition (NER) for Resume Parsing**

A key challenge in resume processing is extracting relevant information such as **candidate name, skills, experience, certifications, and education**. Our system employs **Named Entity Recognition (NER)** techniques:

1. spaCy NER
2. BERT-based NER
3. Regular Expressions (Regex)

This step ensures that resume content is structured and ready for further evaluation.

### **6.3. Feature Extraction and Candidate Scoring**

To evaluate candidates, our system extracts relevant features and assigns scores based on various criteria:

* **TF-IDF Vectorization**: Measures the importance of words in resumes and job descriptions, ensuring that relevant keywords are given higher importance.
* **Word Embeddings**: Utilizes models like Word2Vec, FastText, and BERT for better semantic understanding of resume content, improving keyword matching and candidate-job alignment.
* **Job Fit Prediction Models**: Implements machine learning models such as:
  + **Random Forest**: Ranks candidates based on weighted skill and experience parameters.
  + **XGBoost**: Handles complex resume-job matching scenarios by analyzing feature interactions.
  + **Support Vector Machines (SVM)**: Classifies resumes into categories based on their similarity to job descriptions.
* **Scoring Mechanism**: Assigns scores based on:
  + **Technical Skills & Certifications**
  + **Experience & Project Work**
  + **Soft Skills & Industry Keywords**
  + **Overall Resume Completeness**

### **6.4. Personalized Resume Recommendations**

A crucial feature of our system is **providing actionable recommendations** to improve resumes:

* **GPT-based NLP Model**: Analyzes missing keywords and suggests improvements based on job market trends.
* **Industry-Specific Enhancements**: Compares resumes to high-performing job applications in similar industries to provide optimization suggestions.
* **Rule-Based System**: Identifies gaps in certifications and domain expertise,recommending additional training or coursework.
* **Resume Formatting Suggestions**: Ensures ATS compliance by recommending improvements in layout, structure, and keyword density.

### **6.5. Deployment and User Interface**

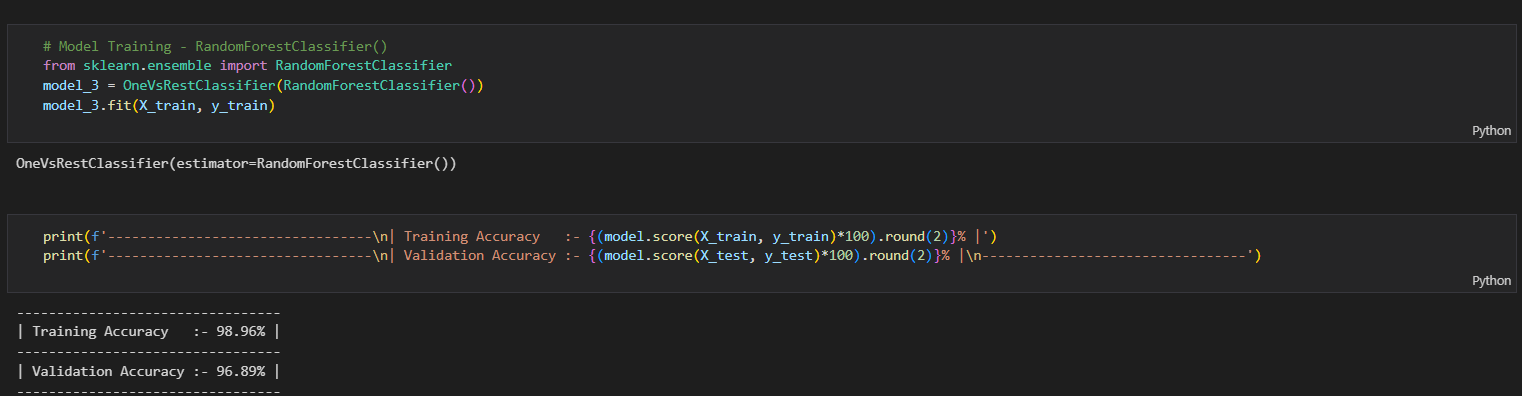
For seamless interaction, the system includes an intuitive **user interface (UI)** and backend infrastructure:

* **Streamlit-based UI**: Enables users to upload resumes, view parsed data, and receive recommendations in an interactive dashboard.
* **FastAPI Backend**: Manages API requests for resume analysis, NLP processing, and database storage.
* **Database Management**:
  + **MongoDB**: Stores unstructured resume data, parsed text, and metadata.
  + **PostgreSQL**: Manages structured job postings, parsed resume data, and scoring metrics.
* **Scalability Considerations**:
  + The system is designed to handle large-scale enterprise-level recruitment needs.
  + Optimized for cloud deployment using **Docker & Kubernetes** for containerized execution.

**6.6. Machine Learning algorithms**

* **Logistic Regression:** Used for classification tasks like skill matching, job fit prediction.
* **Random Forest Classifier:** Used for robust classification with decision trees.
* **K nearest neighbor:** For resume similarity matching.
* **TF-IDF (Term Frequency-Inverse Document Frequency)**: Used for text processing and feature extraction from resumes.

**6.7. Accuracy**

****

## **Unique Selling Points (USP)**

Our system stands out in the recruitment industry with the following unique advantages:

### **1. Automated Screening**

* Extracts and evaluates resumes with **minimal human intervention**, reducing the time spent on manual screening.
* Utilizes **AI-driven models** to ensure fair and unbiased candidate evaluations.

### **2. Personalized Resume Feedback**

* Provides **real-time improvement suggestions** based on employer expectations and job market trends.
* Ensures resumes align with **Applicant Tracking System (ATS) compliance**, improving job application success rates.

### **3. Advanced NLP & Machine Learning Integration**

* Utilizes state-of-the-art **AI models like GPT, BERT, and TF-IDF** for precise parsing and job matching.
* Enhances entity extraction using **deep learning-powered NER models**.

### **4. Scalability & Adaptability**

* **Designed for enterprise-level recruitment** processes, supporting bulk resume processing.
* Adapts to industry-specific requirements, allowing recruiters to customize ranking parameters.

### **5. Job Fit Prediction**

* **Ranks candidates objectively** based on their qualifications, skills, and experience compared to job descriptions.
* Employs **multi-layered AI scoring models** to enhance job-candidate matching accuracy.

### **6. Continuous Learning and Improvement**

* Uses **recruiter feedback** to refine AI models, ensuring better predictions over time.
* Adapts to evolving **job market trends**, dynamically updating skill relevance and recommendations.

By integrating advanced AI technologies, our system revolutionizes recruitment, making it **faster, smarter, and more efficient**.

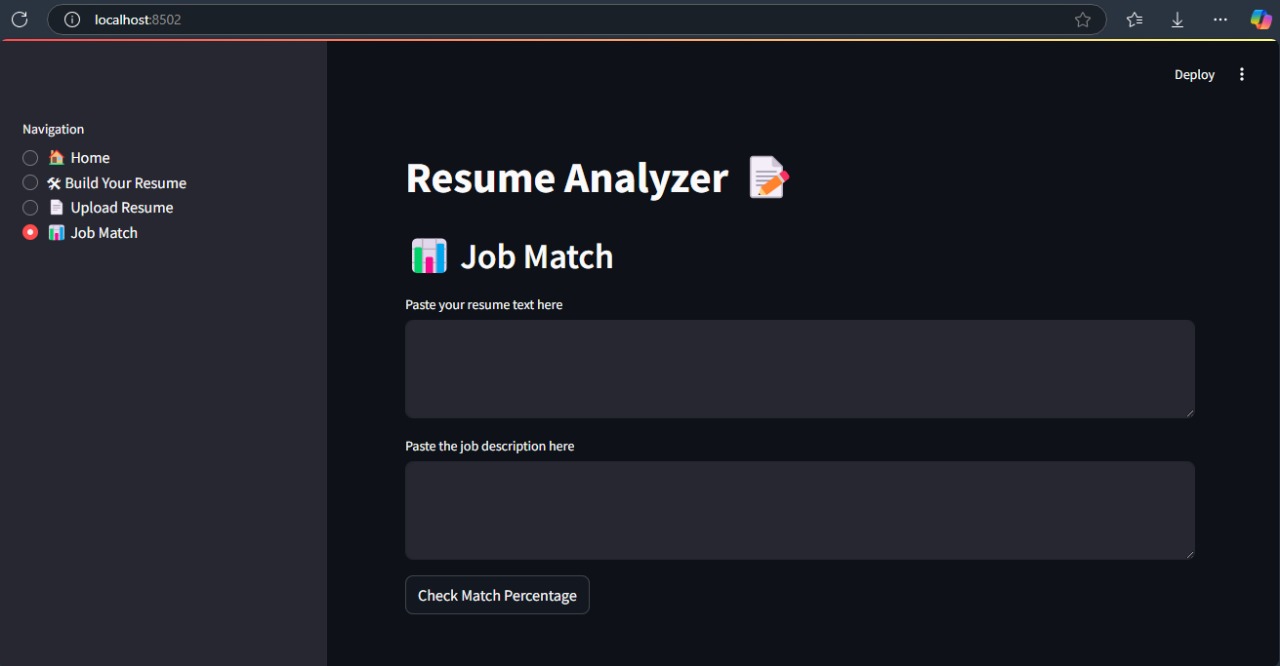
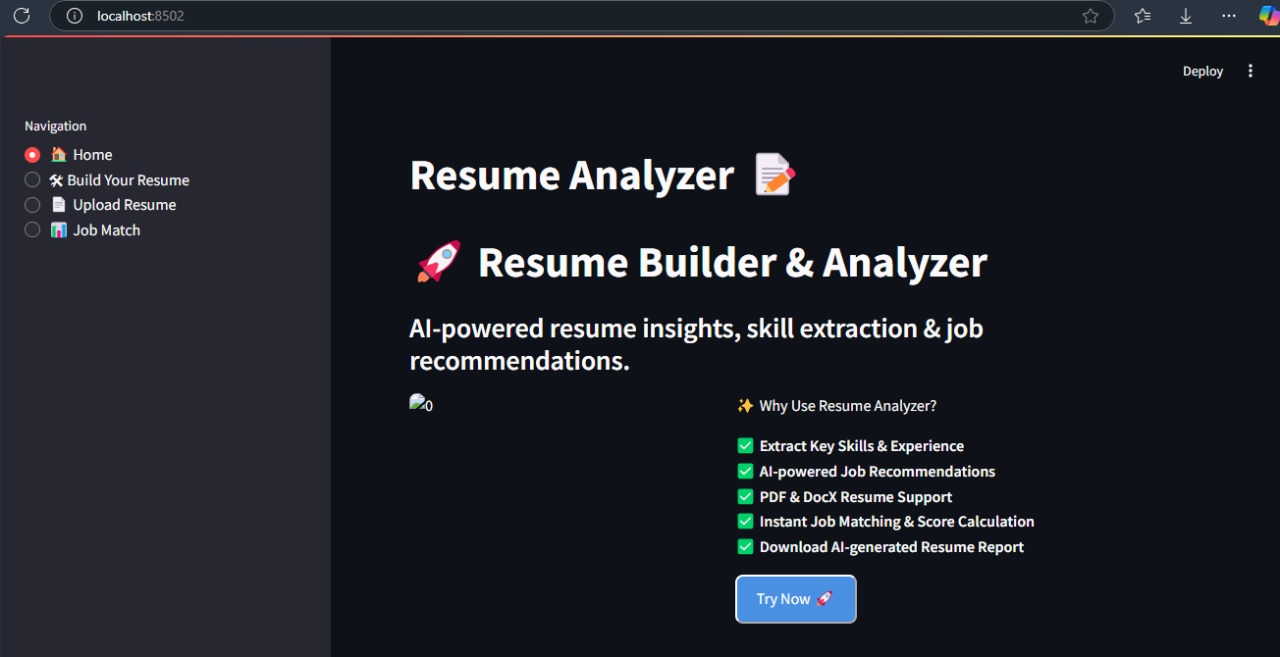
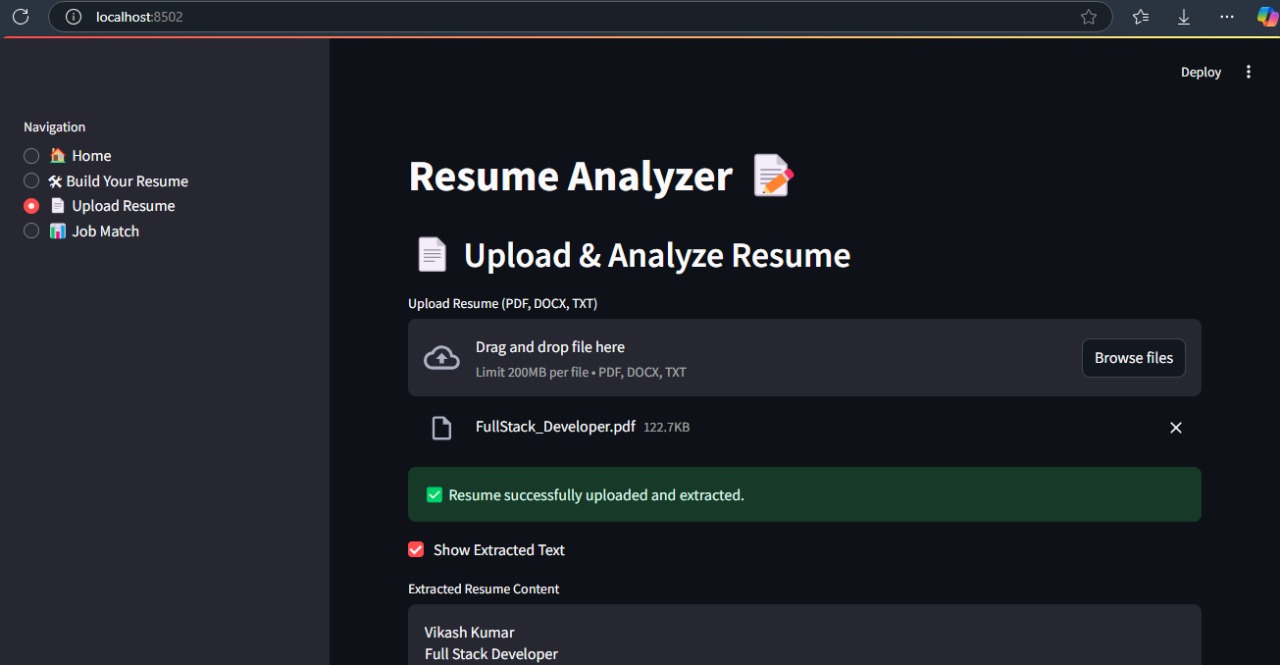
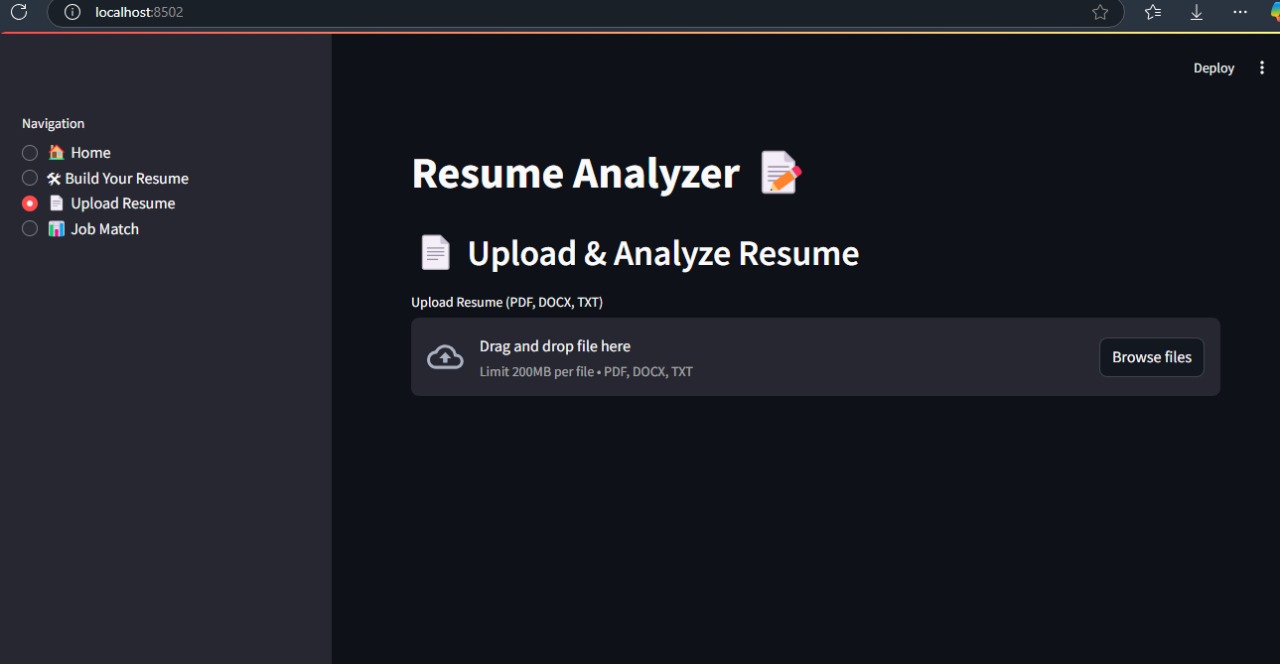
| **Process** | **Method Used** | **Algorithm/Model** |
| --- | --- | --- |
| **Resume Data Extraction** | File Parsing and Preprocessing | PDF, DOCX, TXT Parsers, Tokenization, Stop-word Removal, Lemmatization |
| **Named Entity Recognition (NER)** | NLP-based entity extraction | spaCy NER, BERT-based NER, Regular Expressions (Regex) |
| **Feature Extraction** | Text Vectorization for Resume & Job Matching | TF-IDF, Word Embeddings (Word2Vec, FastText, BERT) |
| **Candidate Scoring** | Machine Learning-Based Evaluation | Random Forest, XGBoost, Support Vector Machines (SVM) |
| **Resume Ranking & Job Fit Prediction** | AI-powered Resume-Job Matching | Logistic Regression, K-Nearest Neighbors (KNN) |
| **Personalized Resume Recommendations** | AI-driven Resume Enhancement Suggestions | GPT-based NLP Model, Rule-Based System |
| **Applicant Tracking System (ATS) Optimization** | Keyword Matching & ATS Compliance Checking | TF-IDF, Resume Formatting Rules |
| **Deployment & User Interface** | Web-based Resume Analysis and Recommendation | Streamlit UI, FastAPI Backend, MongoDB & PostgreSQL for DB |

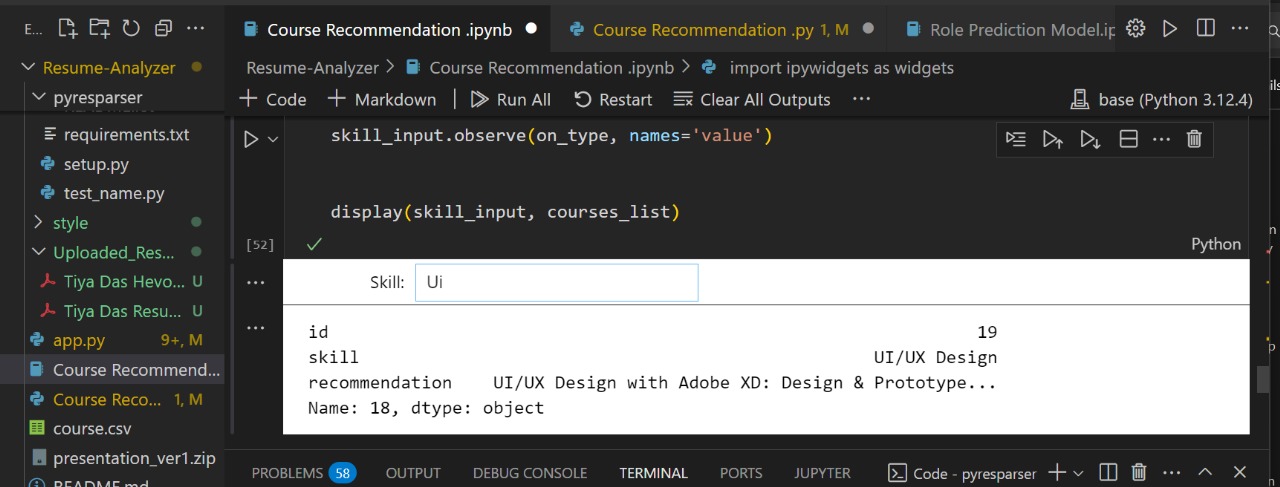
## 

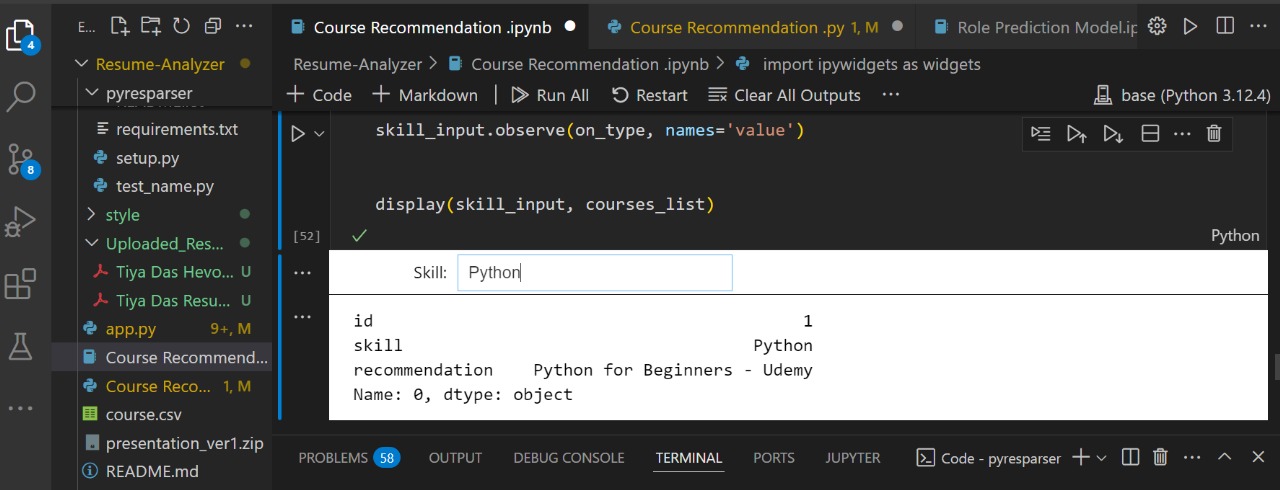
## 

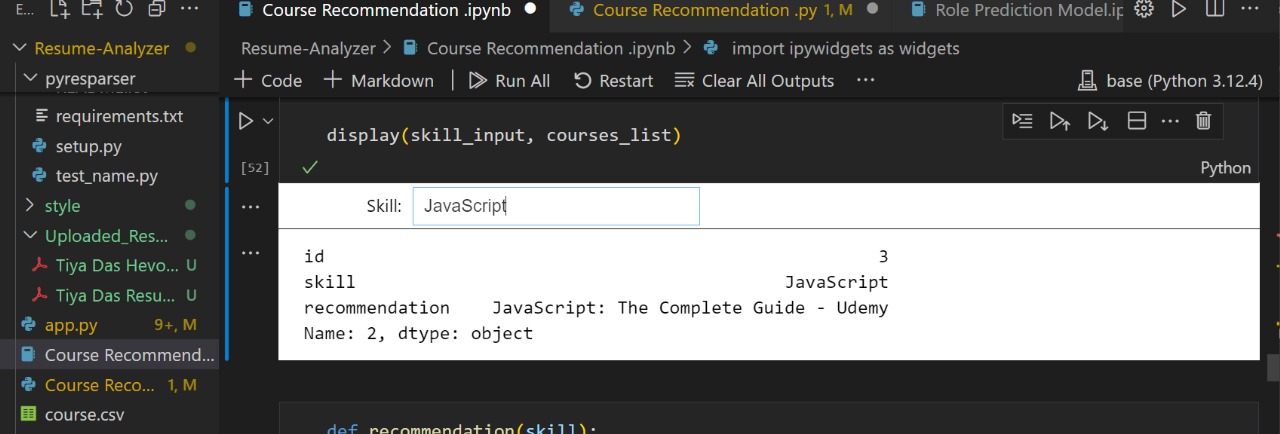
## 

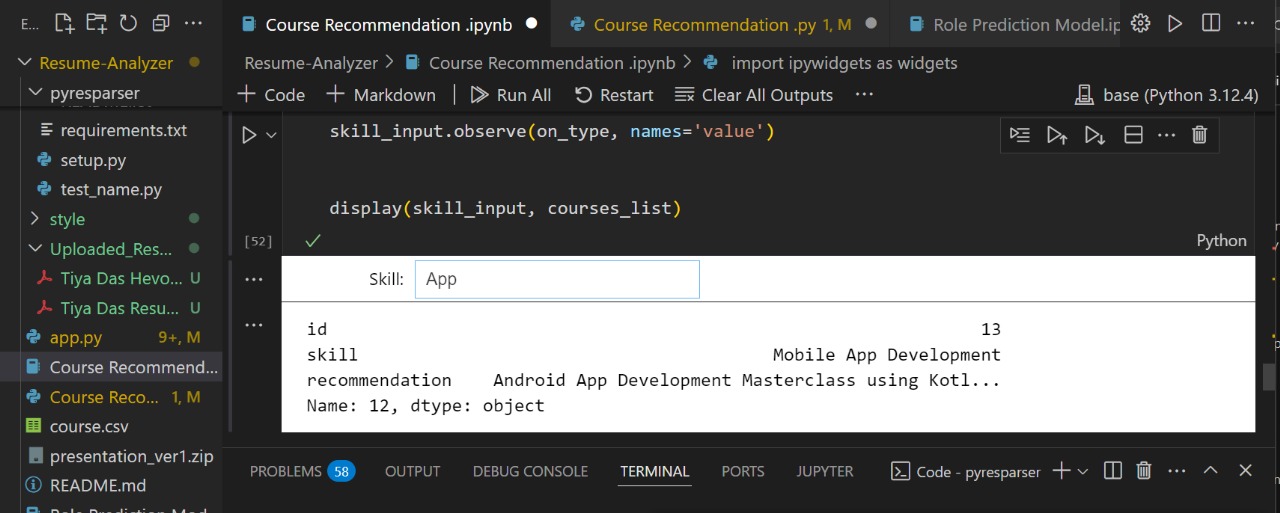
## **6. Results**











# 

# 

# **7. Future Scope**

# **Improved NLP Techniques** – Future advancements in NLP models, including transformer-based architectures (e.g., BERT, GPT), can enhance resume parsing and job recommendation accuracy.

# **Integration with AI-powered ATS** – The adoption of AI-powered Applicant Tracking Systems (ATS) can streamline automated resume screening and provide real-time feedback to candidates.

# **Bias-Free and Fair Screening** – Enhancing fairness and transparency in AI-driven resume screening by incorporating explainable AI (XAI) techniques and debiasing strategies.

# **Personalized Job Recommendations** – Future systems can leverage deep learning and reinforcement learning to refine job recommendations based on user feedback.

# **Multilingual Resume Processing** – Expanding NLP capabilities to support multiple languages for a more inclusive recruitment process.

# **Real-time Skill Gap Analysis** – AI can suggest upskilling programs to candidates based on their resume analysis and market job trends.

### 

### **8. Limitations**

# **Limited Training Data** – Most existing resume screening systems are trained on limited datasets, leading to biased or inaccurate results.

# **Contextual Understanding Issues** – NLP models may struggle with understanding domain-specific jargon, abbreviations, and nuanced language used in resumes.

# **Lack of Standardization** – Variability in resume formats (PDF, DOC, TXT) and inconsistent data structures can affect parsing efficiency.

# **Privacy and Security Concerns** – Handling sensitive personal data requires strict compliance with GDPR, CCPA, and other data protection regulations.

# **Dependence on Keyword Matching** – Many existing systems rely on keyword-based matching rather than semantic understanding, reducing the effectiveness of recommendations.

### 

### 

### **9. Challenges**

# **Handling Unstructured Data** – Extracting meaningful information from unstructured resumes remains a challenge due to diverse writing styles.

# **Bias and Ethical Concerns** – AI models can unintentionally inherit biases from training data, leading to discrimination in hiring decisions.

# **Scalability and Performance** – Ensuring real-time processing of large volumes of resumes without compromising accuracy remains a challenge.

# **Human-AI Collaboration** – Striking a balance between automation and human decision-making is necessary to ensure fair recruitment outcomes.

# **Domain-Specific Adaptation** – Resume recommendation systems need to adapt to different industries and job roles, requiring industry-specific training data.

# 

# 

# **10. Conclusion**

The AI-Powered Resume Parsing and Recommendation System is a transformative solution in modern recruitment, leveraging cutting-edge NLP and machine learning techniques to automate resume analysis. By efficiently extracting, structuring, and evaluating key resume data, the system significantly enhances hiring efficiency while mitigating human biases in candidate selection.

One of the key advantages of this system is its ability to ensure precise entity recognition, skill extraction, and job-role matching. By utilizing advanced AI models, including Named Entity Recognition (NER), BERT, and TF-IDF vectorization, the system accurately identifies and categorizes relevant details such as skills, experience, education, and certifications. This data is then structured into a standardized format, allowing for seamless integration with Applicant Tracking Systems (ATS) and HR management platforms.

Designed for scalability and adaptability, the AI-powered system can accommodate a wide range of recruitment needs, from small businesses to large enterprises. It streamlines candidate evaluation by ranking resumes based on a predefined set of criteria, including job-role fit, experience level, and skill set compatibility. Additionally, the system ensures that recruiters can filter and sort candidates efficiently, improving the overall hiring process.

Beyond parsing and ranking, the system offers personalized resume recommendations, providing job seekers with valuable feedback on how to optimize their resumes for better job market alignment. Using a combination of rule-based logic and AI-driven insights, it suggests improvements such as industry-specific keyword enhancements, missing certifications, and skill development areas.

Future improvements to the system will focus on refining deep learning models to further enhance candidate-job fit predictions. This includes incorporating more sophisticated semantic analysis techniques to better understand job descriptions and candidate profiles. Additionally, plans are in place to expand system functionalities by integrating interview scheduling, automated recruiter-candidate matching, and comprehensive applicant tracking features. These advancements will further streamline the hiring workflow, making the recruitment process faster, fairer, and more efficient.

# 

# 

# **11. References**

1. G. Jaiswal, A. Uttam, D. D. Dubey, and P. K. Mall, “Resume Analyser and Job Recommendation System Based on NLP,” in *Proc. IEEE ICDT*, 2024.
2. H. Artajaya, J. Giancarlos, J. V. Moniaga, and A. Chowanda, “Job Recommendation System based on Resume using NLP and Distance-based Algorithm,” in *Proc. IEEE AIMS*, 2024.
3. S. Mohanty, A. Behera, S. Mishra, A. Alkhayyat, D. Gupta, and V. Sharma, “Resumate: A Prototype to Enhance Recruitment Process with NLP-based Resume Parsing,” in *Proc. IEEE ICIEM*, 2023.
4. R. S. Pundir, A. Dhasmana, U. Karakoti, A. Sikder, S. Sharma, and M. Manchanda, “Enhancing Resume Recommendation System through Skill-based Similarity using Deep Learning Models,” in *Proc. IEEE ICICT*, 2024.
5. A. Mishra, S. Singh, and R. C. Jisha, “AI-Powered Model for Intelligent Resume Recommendation and Feedback,” in *Proc. IEEE ICCCNT*, 2024.
6. G. Navarro, “Fair and Ethical Resume Screening: Enhancing ATS with JustScreen the ResumeScreeningApp,” *J. Inf. Technol. Cybersecurity Artif. Intell.*, vol. 2, no. 1, pp. 1–7, 2025.
7. M. Saatci, R. Kaya, and R. Ünlü, “Resume Screening with Natural Language Processing (NLP),” *Alphanumeric J.*, vol. 12, no. 2, pp. 121–140, 2024.
8. Z. Naveed, B. Nisar, M. Saifullah, and J. I. Baig, “Resume Ranking Using Natural Language Processing,” *J. Comput. Intell. Syst.*, vol. 2, no. 1, 2024.
9. I. Ali, G. Mujtaba, Z. H. Khand, Z. Ali, and S. Khan, “Resume Classification System using Natural Language Processing & Machine Learning Techniques,” *Int. J. Comput. Sci. Netw. Secur.*, vol. 24, no. 7, pp. 108–117, 2024.
10. L. Zhang, X. Wu, and Y. Liu, “Automated Resume Screening Using Transformer-based NLP Models,” in *Proc. IEEE BigData*, 2024.
11. K. Patel, N. Shah, R. Mehta, and A. Desai, “AI-driven Resume Evaluation System Using BERT and Deep Learning,” in *Proc. IEEE ICAIIC*, 2024.
12. M. Ahmed, A. Hussain, and S. Rehman, “A Hybrid Machine Learning Approach for Resume Screening and Job Matching,” in *Proc. IEEE ICMLA*, 2023.
13. J. Brown, T. Lee, and P. Smith, “Ethical AI in Recruitment: Reducing Bias in Automated Resume Screening,” *AI Ethics J.*, vol. 5, no. 1, pp. 45–60, 2024.
14. Y. Chen, W. Sun, and H. Wang, “Resume Parsing and Ranking System Based on Named Entity Recognition and Deep Learning,” in *Proc. IEEE ICDM*, 2024.
15. R. Verma, A. Pandey, and S. Ghosh, “NLP-powered Resume Analysis for Effective Candidate Shortlisting,” in *Proc. IEEE ICACCI*, 2023.