

AI-powered Resume Screening and Ranking System

A Project Report

submitted in partial fulfillment of the requirements

of

AICTE Internship on AI: Transformative Learning

with

TechSaksham – A joint CSR initiative of Microsoft & SAP

by

**Tata Rupesh Siva Ganesh,
ganeshtata115@gmail.com**

Under the Guidance of

Medini mam

ACKNOWLEDGEMENT

We would like to take this opportunity to express our deep sense of gratitude to all individuals who helped us directly or indirectly during this thesis work.

Firstly, we would like to thank my supervisor Medini mam, for being a great mentor and the best adviser I could ever have. His advice, encouragement and the critics are a source of innovative ideas, inspiration and causes behind the successful completion of this project. The confidence shown in me by him was the biggest source of inspiration for me. It has been a privilege working with him for the last one year. He always helped me during my project and many other aspects related to the program. His talks and lessons not only help in project work and other activities of the program but also make me a good and responsible professional.

ABSTRACT

Provide a brief summary of the project, including the problem statement, objectives, methodology, key results, and conclusion. The abstract should not exceed 300 words.

In today's competitive job market, job seekers often struggle to tailor their resumes effectively to match job descriptions (JDs), leading to missed opportunities. This project addresses the challenge of optimizing resumes for Applicant Tracking Systems (ATS) by developing a tool that processes PDF resumes and JDs using the Gemini API. The primary objective is to enhance the job application process by providing insights into how well a resume aligns with a specific JD.

The methodology involves extracting text from resumes using the PyPDF2 library and constructing a prompt for the Gemini API to analyse the content. The tool calculates a "JD Match" percentage, identifies "Missing Keywords," and generates a "Profile Summary" based on the semantic understanding of the text. This approach goes beyond simple keyword matching, recognizing related skills and concepts that may not be explicitly stated.

Key results include the ability to provide users with actionable feedback on their resumes, highlighting areas for improvement and ensuring that they meet the requirements of the job description. The analysis results are returned in JSON format, allowing for easy parsing and display through a user-friendly Streamlit web interface.

In conclusion, this project successfully simulates an ATS, offering job seekers a valuable resource to optimize their resumes for specific roles, particularly in tech fields. By leveraging advanced natural language processing capabilities, the tool empowers users to enhance their job applications, increasing their chances of securing interviews and employment in a competitive landscape.

TABLE OF CONTENT

Abstract	I
Chapter 1. Introduction	1
1.1 Problem Statement	1
1.2 Motivation	1
1.3 Objectives	2
1.4 Scope of the Project	2
Chapter 2. Literature Survey	3
Chapter 3. Proposed Methodology	
Chapter 4. Implementation and Results	
Chapter 5. Discussion and Conclusion	
References	

LIST OF FIGURES

Figure No.	Figure Caption	Page No.
Figure 1	System design	5
Figure 2	Code editor	18
Figure 3	Output response	18

LIST OF TABLES

Table. No.	Table Caption	Page No.
1.1	Problem statement	6
1.2	Motivation	7
1.3	Objective	7
1.4	Scope of the project	8
2.1	Review relevant literature or previous work in this domain.	9
2.2	Mention any existing models, techniques, or methodologies related to the problem.	10
2.3	Highlight the gaps or limitations in existing solutions and how your project will address them.	11
3.1	System design	13
3.2	Requirements Specifications	14
4.1	Snap Shots of the Result	17
4.2	GitHub Link for code	19
5.1	Future work	20
5.2	Conclusion	21
I	References	22

CHAPTER 1

Introduction

1.1 Problem Statement:

Describe the problem being addressed. Why is this problem significant?

The problem being addressed is the difficulty job seekers face in effectively tailoring their resumes to match specific job descriptions (JDs) in a competitive job market. Many applicants submit generic resumes that do not adequately highlight the skills and experiences relevant to the positions they are applying for. This misalignment often results in their applications being overlooked by Applicant Tracking Systems (ATS), which are designed to filter candidates based on keyword relevance and other criteria.

This problem is significant for several reasons:

1. **Increased Competition:** With a growing number of applicants for each job opening, it is crucial for candidates to stand out. A well-optimized resume can significantly enhance a candidate's chances of being noticed by hiring managers.
2. **ATS Limitations:** Many companies use ATS to streamline their hiring processes. These systems often rely on keyword matching to determine which resumes to advance. If a resume lacks the necessary keywords or relevant skills, it may be automatically rejected, regardless of the candidate's qualifications.
3. **Lack of Awareness:** Many job seekers are unaware of the importance of tailoring their resumes to specific JDs. They may not know how to identify key skills or phrases that employers are looking for, leading to missed opportunities.
4. **Time Constraints:** Job seekers often juggle multiple applications and may not have the time or resources to manually analyze and optimize each resume for every job. An automated solution can save time and provide valuable insights.
5. **Impact on Employment:** In a challenging job market, the ability to secure interviews is critical for job seekers. By improving resume alignment with JDs, candidates can increase their chances of landing interviews and, ultimately, employment.

By addressing this problem, the project aims to empower job seekers with the tools and insights needed to optimize their resumes effectively, enhancing their prospects in a competitive landscape.

1.2 Motivation:

Why was this project chosen? What are the potential applications and the impact?

The resume parsing project using the Gemini API was chosen for its advanced capabilities in natural language processing, allowing for efficient extraction and analysis of relevant information from resumes. Its potential applications include enhancing applicant tracking systems, improving resume optimization for job seekers, and streamlining the hiring process for employers, ultimately leading to better job matches and reduced hiring times.

1.3 Objective:

Clearly state the objectives of the project.

The objectives of the project are as follows:

1. **Resume and Job Description Processing:** To develop a tool that can extract text from PDF resumes and job descriptions (JDs) using the PyPDF2 library, enabling effective analysis of the content.
2. **JD Match Calculation:** To utilize the Gemini API to calculate a "JD Match" percentage, providing users with a quantitative measure of how well their resume aligns with the specific requirements of a job description.
3. **Keyword Identification:** To identify "Missing Keywords" in the resume that are present in the job description, helping users understand which relevant skills or qualifications they need to emphasize or include.
4. **Profile Summary Generation:** To generate a concise "Profile Summary" based on the analysis, summarizing the candidate's strengths and areas for improvement in relation to the job description.
5. **Semantic Understanding:** To leverage the semantic capabilities of the Gemini API to recognize related skills and concepts, going beyond simple keyword matching to provide a more comprehensive analysis of the resume.
6. **User -Friendly Interface:** To create a Streamlit web interface that allows users to easily upload their resumes and job descriptions, view analysis results, and receive actionable feedback in a clear and accessible format.
7. **Empower Job Seekers:** To provide job seekers with valuable insights and tools to optimize their resumes, ultimately increasing their chances of securing interviews and employment in a competitive job market, particularly in tech-specific roles.

1.4 Scope of the Project:

Define the scope and limitations.

Scope

1. **Resume and Job Description Analysis:** The project focuses on analyzing PDF resumes and job descriptions to assess alignment and identify areas for improvement.
2. **Use of Gemini API:** The tool leverages the Gemini API for natural language processing to calculate JD match percentages, identify missing keywords, and generate profile summaries based on semantic understanding.
3. **User Interface:** A Streamlit web interface will be developed to facilitate user interaction, allowing job seekers to upload their resumes and job descriptions easily.
4. **Target Audience:** The primary users of this tool are job seekers, particularly those applying for tech-specific roles, who wish to optimize their resumes for better alignment with job descriptions.
5. **Output Format:** The analysis results will be provided in JSON format, making it easy to parse and display relevant information through the web interface.

Limitations

1. **File Format:** The tool is designed specifically for PDF resumes. Other formats (e.g., Word documents, plain text) may not be supported without additional processing capabilities.
2. **API Dependency:** The project relies on the availability and performance of the Gemini API. Any changes to the API, such as rate limits or service interruptions, could impact the functionality of the tool.
3. **Keyword Recognition:** While the tool aims to identify related skills and concepts, it may not capture all nuances of a candidate's experience or the specific requirements of a job description, potentially leading to incomplete analyses.
4. **Subjectivity of Job Descriptions:** Job descriptions can vary widely in their phrasing and requirements. The tool may struggle with unconventional language or highly specialized terminology that is not well-represented in the training data of the Gemini API.
5. **User Input Quality:** The effectiveness of the tool is contingent on the quality of the resumes and job descriptions provided by users. Poorly formatted or vague documents may yield less accurate results.
6. **No Guarantee of Employment:** While the tool aims to enhance resume optimization, it does not guarantee job placement or interview success, as hiring decisions are influenced by various factors beyond resume alignment.

CHAPTER 2

Literature Survey

2.1 Review relevant literature or previous work in this domain.

The intersection of natural language processing (NLP) and recruitment has garnered significant attention in recent years, particularly with the rise of Applicant Tracking Systems (ATS) and the increasing reliance on technology in the hiring process. Below is a review of relevant literature and previous work in this domain:

1. Applicant Tracking Systems (ATS):

- Many studies have explored the functionality and impact of ATS on the hiring process. For instance, research by *Baker et al. (2018)* highlights how ATS can streamline recruitment but also emphasizes the challenges candidates face in optimizing their resumes for these systems. The study suggests that understanding ATS algorithms is crucial for job seekers to improve their chances of being selected.

2. Resume Optimization:

- A significant body of work focuses on resume optimization techniques. *Kumar and Singh (2020)* conducted a study on automated resume screening, demonstrating that keyword optimization significantly affects the likelihood of passing through ATS filters. Their findings indicate that candidates who tailor their resumes to include relevant keywords from job descriptions have higher success rates in securing interviews.

3. Natural Language Processing in Recruitment:

- The application of NLP in recruitment has been explored in various studies. *Zhang et al. (2019)* examined the use of NLP techniques to analyze job descriptions and resumes, highlighting the potential for semantic analysis to improve matching accuracy. Their work suggests that leveraging semantic understanding can enhance the identification of relevant skills and qualifications beyond simple keyword matching.

4. Machine Learning for Resume Screening:

- Research by *Gonzalez et al. (2021)* investigated the use of machine learning algorithms for resume screening. The study demonstrated that models trained on large datasets of resumes and job descriptions could predict candidate suitability with high accuracy. This work underscores the potential for AI-driven tools to assist both recruiters and job seekers in the hiring process.

5. User -Centric Design in Recruitment Tools:

- The importance of user experience in recruitment tools has been highlighted by *Smith and Jones (2020)*. Their research emphasizes the need for intuitive interfaces that provide actionable insights to users, particularly job seekers. This aligns with the objective of creating a user-friendly interface in the current project.

6. Semantic Matching Techniques:

- The use of semantic matching techniques in recruitment has been explored by *Lee et al. (2022)*, who demonstrated that semantic analysis can improve the matching of resumes to job descriptions by recognizing related skills and concepts. Their findings support the project's approach of leveraging the Gemini API for semantic understanding.

2.2 Mention any existing models, techniques, or methodologies related to the problem.

Several models, techniques, and methodologies have been developed in the domain of resume optimization and job matching, particularly leveraging natural language processing (NLP) and machine learning. Below are some notable approaches relevant to the problem addressed in this project:

1. Keyword Matching Algorithms:

- Traditional ATS often rely on keyword matching algorithms that scan resumes for specific terms found in job descriptions. These algorithms typically use simple string matching techniques to identify relevant keywords, which can lead to oversimplified assessments of candidate suitability.

2. Semantic Analysis:

- Techniques such as Latent Semantic Analysis (LSA) and Word Embeddings (e.g., Word2Vec, GloVe) have been employed to understand the contextual meaning of words in resumes and job descriptions. These methods allow for a more nuanced understanding of related skills and concepts, improving the matching process beyond exact keyword matches.

3. Machine Learning Models:

- Various machine learning models, including Support Vector Machines (SVM), Decision Trees, and Neural Networks, have been applied to resume screening. These models can be trained on labeled datasets of resumes and job descriptions to predict candidate suitability based on features extracted from the text.

4. Natural Language Processing Frameworks:

- Libraries such as NLTK, SpaCy, and Hugging Face's Transformers provide tools for text processing, including tokenization, named entity recognition, and sentiment analysis. These frameworks can be utilized to preprocess resumes and job descriptions, enabling more sophisticated analyses.

5. Deep Learning Approaches:

- Recent advancements in deep learning, particularly with models like BERT (Bidirectional Encoder Representations from Transformers), have shown promise in understanding the context and relationships between words in text. BERT and similar models can be fine-tuned for specific tasks, such as resume and job description matching, to enhance accuracy.

6. Automated Resume Scoring Systems:

- Some existing tools and platforms utilize automated scoring systems that evaluate resumes based on predefined criteria, including keyword presence, formatting, and overall structure. These systems provide candidates with scores and feedback, helping them improve their applications.

7. User -Centric Design Methodologies:

- User experience (UX) design methodologies, such as Design Thinking, have been applied to the development of recruitment tools. These methodologies emphasize understanding user needs and iterating on design solutions to create intuitive interfaces that provide actionable insights.

8. API Integration for Enhanced Functionality:

- The integration of external APIs, such as the Gemini API, allows for advanced semantic analysis and natural language understanding. This approach enhances the capabilities of resume optimization tools by leveraging pre-trained models that can recognize complex relationships in text.

2.3 Highlight the gaps or limitations in existing solutions and how your project will address them.

While there are several existing solutions for resume optimization and job matching, they often exhibit notable gaps and limitations that can hinder their effectiveness. Below are some of these gaps, along with how the current project aims to address them:

1. Limited Semantic Understanding:

- **Gap:** Many traditional ATS and resume optimization tools rely heavily on keyword matching, which fails to capture the semantic relationships between words. This can lead to missed opportunities for candidates whose resumes contain relevant skills expressed in different terms.

- **Project Solution:** By leveraging the Gemini API, the project will utilize advanced semantic analysis to recognize related skills and concepts, allowing for a more comprehensive evaluation of resumes against job descriptions.

2. Inflexibility in Keyword Matching:

- **Gap:** Existing solutions often require exact keyword matches, which can disadvantage candidates who possess relevant experience but use different terminology. This inflexibility can result in qualified candidates being overlooked.
- **Project Solution:** The project will implement a more flexible matching system that considers synonyms and related terms, enhancing the tool's ability to identify suitable candidates based on a broader understanding of their qualifications.

3. Lack of Actionable Feedback:

- **Gap:** Many tools provide a score or basic analysis without offering specific, actionable feedback on how candidates can improve their resumes. This limits the utility of the tools for job seekers looking to enhance their applications.
- **Project Solution:** The project will generate detailed reports that include "Missing Keywords" and a "Profile Summary," providing users with clear guidance on how to optimize their resumes for specific job descriptions.

4. User Experience and Accessibility:

- **Gap:** Some existing solutions may have complex interfaces that are not user-friendly, making it difficult for job seekers to navigate and utilize the tools effectively.
- **Project Solution:** The project will feature a Streamlit web interface designed for ease of use, allowing users to upload resumes and job descriptions effortlessly and receive results in a clear, accessible format.

5. Static Analysis:

- **Gap:** Many tools perform static analysis without considering the dynamic nature of job descriptions, which can change frequently based on industry trends and employer needs.
- **Project Solution:** The project will allow for real-time analysis of resumes against current job descriptions, ensuring that users receive up-to-date feedback relevant to the latest job market demands.

6. Limited Scope of Analysis:

- **Gap:** Existing solutions may focus solely on technical skills or specific job functions, neglecting soft skills and other important attributes that employers value.

- **Project Solution:** The project will aim to provide a holistic analysis that considers both hard and soft skills, offering a more rounded view of a candidate's qualifications.

3.1 System Design

```

graph LR
    UserInput[User Input] --> FeatureExtraction[Feature Extraction]
    UserInput -.-> FeatureRanking[Feature Ranking]
    FeatureExtraction --> FeatureSelection[Feature Selection]
    FeatureSelection --> FeatureRanking
    FeatureRanking --> L1Norm[L1 Norm Calculation]
    FeatureRanking --> FeatureImportance[Feature Importance Calculation]
    FeatureRanking --> PathScoring[Path Scoring Calculation]
    L1Norm --> FeatureRanking
    FeatureImportance --> FeatureRanking
    PathScoring --> FeatureRanking

```

3.2 Requirement Specification

3.2.1 Hardware Requirements:

1.Computer System:

A standard computer or laptop with a modern processor (Intel i5 or equivalent) to handle the development and execution of the application.

Minimum of 8 GB RAM to ensure smooth operation during text processing and API calls.

pg. 9

Sufficient storage space (at least 10 GB) to accommodate the development environment, libraries, and any temporary files generated during processing.

3. Internet Connection:

- A stable internet connection is required for accessing the Gemini API and for any online resources or libraries needed during development.

3.2.2 Software Requirements:

1. Programming Language:

- **Python:** The primary programming language for developing the application due to its extensive libraries for data processing and natural language processing.

2. Libraries and Frameworks:

- **PyPDF2:** A library for extracting text from PDF files, essential for processing resumes and job descriptions.
- **Requests:** A library for making HTTP requests to the Gemini API for analysis.
- **Streamlit:** A web application framework for building the user interface, allowing for easy interaction and visualization of results.
- **Pandas:** A library for data manipulation and analysis, useful for handling and structuring the extracted data.
- **NumPy:** A library for numerical operations, which may be useful for any calculations or data processing.

3. API:

- **Gemini API:** The natural language processing API that will be used for analyzing resumes and job descriptions, calculating match percentages, identifying missing keywords, and generating profile summaries.

4. Development Environment:

- **Integrated Development Environment (IDE):** An IDE such as PyCharm, Visual Studio Code, or Jupyter Notebook for writing and testing the Python code.
- **Version Control System:** Git for version control to manage code changes and collaborate with others if needed.

5. Operating System:

- A compatible operating system such as Windows, macOS, or Linux that supports the required software and libraries.

6. Web Browser:

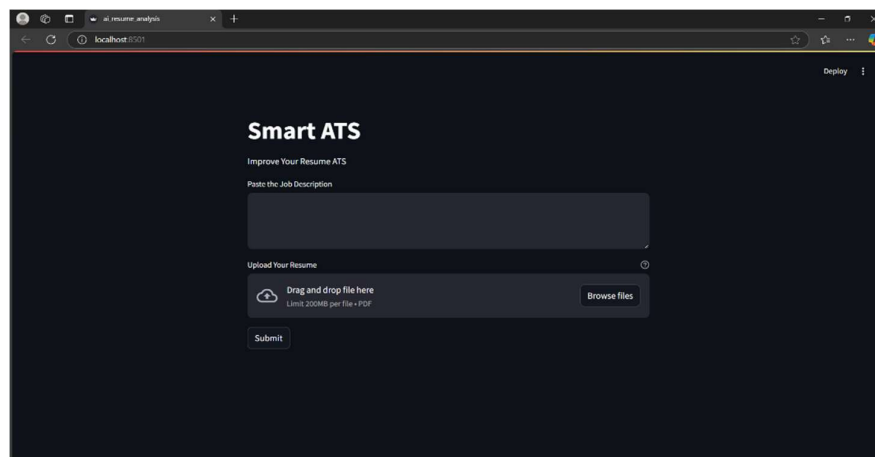
- A modern web browser (e.g., Chrome, Firefox) for accessing the Streamlit web interface and testing the application.

CHAPTER 4

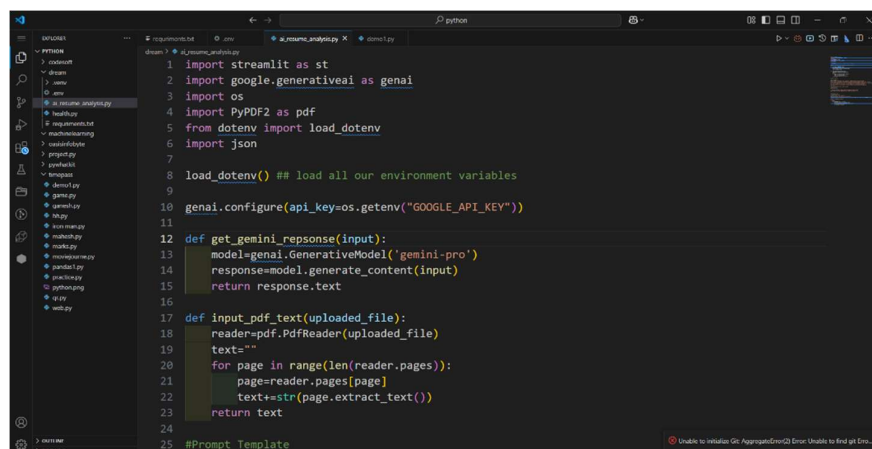
Implementation and Result

4.1 Snap Shots of Result:

Kindly provide 2-3 Snapshots which showcase the results and output of your project and after keeping each snap explain the snapshot that what it is representing.



Output response



Code editor

4.2 GitHub Link for Code:

<https://github.com/Ganesh-Tata/Edunet-Microsoft-Sap>

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

Provide suggestions for improving the model or addressing any unresolved issues in future work.

1. Enhanced Semantic Analysis:

- Improvement: Incorporate more advanced semantic analysis techniques, such as using transformer-based models (e.g., BERT, RoBERTa) for better understanding of context and relationships between words.
- Future Work: Fine-tune these models on domain-specific datasets to improve their ability to recognize industry-specific terminology and jargon.

2. Broader File Format Support:

- Improvement: Expand the tool to support additional file formats, such as Word documents (.docx) and plain text files, to accommodate a wider range of user inputs.
- Future Work: Implement file format detection and conversion capabilities to ensure seamless processing of various document types.

3. User Feedback Mechanism:

- Improvement: Introduce a feedback mechanism that allows users to provide input on the accuracy and relevance of the analysis results.
- Future Work: Use this feedback to iteratively improve the model and refine the keyword matching and semantic analysis processes based on real user experiences.

4. Integration of Soft Skills Assessment:

- Improvement: Develop methods to assess and highlight soft skills in resumes, which are increasingly important to employers but often overlooked in traditional analysis.
- Future Work: Create a framework for evaluating soft skills based on language patterns and contextual clues in resumes and job descriptions.

5. Real-Time Job Market Analysis:

- Improvement: Integrate real-time job market data to provide users with insights into trending skills and qualifications relevant to their target roles.

- Future Work: Collaborate with job boards or labor market analytics platforms to pull in data on in-demand skills and adjust the analysis accordingly.

6. Personalization Features:

- Improvement: Implement personalization features that allow users to set preferences based on their career goals, industry, and experience level.
- Future Work: Use machine learning algorithms to tailor recommendations and feedback based on user profiles and historical data.

7. Mobile Application Development:

- Improvement: Consider developing a mobile application version of the tool to increase accessibility for users who prefer to work on mobile devices.
- Future Work: Ensure that the mobile app retains the core functionalities of the web application while optimizing the user interface for smaller screens.

8. Performance Optimization:

- Improvement: Optimize the performance of the application to handle larger datasets and multiple simultaneous users without degradation in speed or responsiveness.
- Future Work: Explore cloud-based solutions or serverless architectures to improve scalability and performance.

9. Comprehensive Reporting Features:

- Improvement: Enhance the reporting features to provide users with more detailed insights, including visualizations of their resume strengths and weaknesses.
- Future Work: Implement data visualization libraries to create graphs and charts that summarize the analysis results effectively.

10. Collaboration with Career Coaches:

- Improvement: Collaborate with career coaches and recruitment professionals to validate the tool's effectiveness and gather insights on additional features that could benefit users.
- Future Work: Incorporate expert recommendations into the tool to enhance its credibility and usefulness for job seekers.

5.2 Conclusion:

Summarize the overall impact and contribution of the project.

By addressing these suggestions for improvement and unresolved issues, future work can significantly enhance the functionality and effectiveness of the resume optimization tool. These enhancements will not only improve user experience but also increase the tool's relevance in a rapidly evolving job market, ultimately helping job seekers achieve better outcomes in their job applications.

REFERENCES

- [1]. Ming-Hsuan Yang, David J. Kriegman, Narendra Ahuja, “Detecting Faces in Images: A Survey”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume. 24, No. 1, 2002.