**A Project report on**

# Resume analysis and skills advisor

A Dissertation submitted to JNTU Hyderabad in partial fulfillment of the academic requirements for the award of the degree.

## Bachelor of Technology

## in

## Computer Science and Engineering


Submitted by

K. SHAARVANIE
(20H51A0597)
A. NAVYA
(20H51A05G3)
CH. KARTHIK
(20H51A05N3)

Under the esteemed guidance of
Ms. P. Sravanthi
(Assistant Professor)




**Department of Computer Science and Engineering**

**CMR COLLEGE OF ENGINEERING & TECHNOLOGY**
(UGC Autonomous)
*Approved by AICTE *Affiliated to JNTUH *NAAC Accredited with $A^+$ Grade
KANDLAKOYA, MEDCHAL ROAD, HYDERABAD - 501401.


**2020- 2024**

# CMR COLLEGE OF ENGINEERING & TECHNOLOGY

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD – 501401

## DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING


# CERTIFICATE

This is to certify that the Major Project Phase I report entitled " **Resume analysis and skills advisor**" being submitted by K.Shaarvanie (20H51A0597), A. Navya (20H51A05G3), CH. Karthik (20H51A05N3) in partial fulfillment for the award of **Bachelor of Technology in Computer Science and Engineering** is a record of bonafide work carried out his/her under my guidance and supervision.

The results embodies in this project report have not been submitted to any other University or Institute for the award of any Degree.

**Ms. P. Sravanthi**
**Assistant Professor**
**Dept. of CSE**

**Dr. Siva Skandha Sanagala**
**Associate Professor and HOD**
**Dept. of CSE**

# ACKNOWLEDGEMENT

With great pleasure we want to take this opportunity to express my heartfelt gratitude to all the people who helped in making this project work a grand success.

We are grateful to **Ms. P. Sravanthi**, **Assistant Professor** Department of Computer Science and Engineering for his valuable technical suggestions and guidance during the execution of this project work.

We would like to thank **Dr. Siva Skandha Sanagala,** Head of the Department of Computer Science and Engineering, CMR College of Engineering and Technology, who is the major driving forces to complete my project work successfully.

We are very grateful to **Dr.G.Devadas**, Dean-Academics, CMR College of Engineering and Technology, for his constant support and motivation in carrying out the project work successfully.

We are highly indebted to **Major Dr. V A Narayana,** Principal, CMR College of Engineering and Technology, for giving permission to carry out this project in a successful and fruitful way.

We would like to thank the **Teaching & Non- teaching** staff of Department of Computer Science and Engineering for their co-operation

We express our sincere thanks to **Shri. Ch. Gopal Reddy**, Secretary, CMR Group of Institutions, for his continuous care.

Finally, We extend thanks to our parents who stood behind us at different stages of this Project. We sincerely acknowledge and thank all those who gave support directly and indirectly in completion of this project work.

K.Shaarvanie 20H51A0597
A.Navya 20H51A05G3
CH. Karthik 20H51A05N5

# TABLE OF CONTENTS

## List of Figures

# ABSTRACT

Recruiting candidates to fit a particular job profile is a task crucial to most of the companies. Due to increasing growth in online recruitment, traditional hiring methods are becoming inefficient. The conventional techniques usually include a labor-intensive process of manually searching through the applied candidates, reviewing their resumes, and then producing a shortlist of suitable candidates to be interviewed. In this era of technology, job searching has become smarter and more accessible at the same time. The companies receive enormous numbers of resumes/CVs, which are not always structured. There have been lots of work done for the job searching process. Whereas, the process of selecting a candidate based on their resume has not been entirely automated. This research proposes a model of extracting valuable information from the resume and ranking it according to the preference and requirement of the company. To achieve the desired goal, the entire process has been divided into three segments. The first segment consists of converting the unstructured resumes in structured data using NLP, and the second segment consists of the extraction phase, where the relevant information is extracted from the resume and giving them an identifier value. Finally, based on the values assigned, the resumes are ranked accordingly in the final segment.

# CHAPTER 1
# INTRODUCTION

# CHAPTER 1
# INTRODUCTION

## 1.1 Problem Statement

In today's highly competitive job market, having a strong and up-to-date skill set is essential for career advancement. Job seekers often struggle to identify and improve the skills that would make their resumes more appealing to potential employers. This project aims to develop an automated system that analyzes resumes submitted by users and provides personalized suggestions on skills that can be enhanced or added to improve their employability. The tool will take a resume as input and provide suggestions on how the candidate can improve their skills presentation.

## 1.2 Research Objective

The primary objective of this project is to develop a robust and efficient system for the analysis of resumes and the identification of skills that can be improved. We aim to develop a practical tool that empowers individuals to enhance their skills and improve their employability by providing personalized and data-driven recommendations based on resume analysis. This project will be able to identify gaps in the skill set of the individual by comparing their stated skills to industry or job-specific requirements. Determine the criticality and relevance of each missing skill. It provides tailored and actionable recommendations to the resume owner on how to improve the identified skills. These recommendations should include suggestions for courses, certifications, projects, or other means of skill development. We put in the use of machine learning and natural language processing techniques to enhance the accuracy of skill assessment and recommendation generation.

## 1.3 Project Scope and Limitations

### Scope:

- As we know Indian I.T sector is second largest candidate recruiting sector of our country. It contributes about 7.5% to our Gross Domestic Product (G.D.P) Our Proposed system is initially concerned with the I.T sector of our country.
- It is mainly going to deal the Indian I.T industry but if you talk about the pro version of our system, it can be extended to various other commercial sector where, intake and elimination are in bulk like for Governmental Jobs.

### Limitations:

- The data used to train the system may contain biases, such as gender, race, or socioeconomic status. The model could inadvertently learn and perpetuate these biases, leading to unfair assessments.
- Scoring and suggesting improvements for resumes at a large scale can be resource-intensive. Handling a high volume of resumes can strain computational resources.
- The use of automated systems to evaluate resumes and suggest improvements can raise ethical concerns about bias, privacy, and fairness, and it requires careful handling.

# CHAPTER 2
# BACKGROUND WORK

# CHAPTER 2

# BACKGROUND WORK

## 2.1. Resume Analyzer An automated solution to recruitment process

### 2.1.1 Introduction

Resume Analyzer system is implementation, and evaluation of an approach to apply text mining for analyzing the resumes company receives using keyword matching algorithm. With the help of Keyword matching algorithm, the keywords from the customized dictionary will be mapped against the words in the resume. Once the keywords are matched the required data is extracted and stored in the database. The entire database is sorted based on various parameters like experience, qualification, age, etc. Interview scheduling is performed next by the system based on the generated sorted list. Applicants are informed about the interview through an automated mail generated by system. As large numbers of people apply for the job, analyzing each resume is difficult work. Also scheduling the interviews as per the sorting is a tedious job. We are attempting to reduce efforts on candidate's as well as company's side.

### 2.1.2 Merits, Demerits and Challenges

## Merits:

1) **Enhanced Hiring Efficiency:** For recruiters and employers, automated resume analysis can significantly improve the efficiency of the hiring process by quickly identifying the most qualified candidates.

2) **Objective Evaluation:** The system can provide an objective assessment of resumes, reducing bias in the hiring process and ensuring that all applicants are evaluated fairly.

3) **Scalability:** The system can process a large number of resumes in a short time, making it highly scalable for organizations with high applicant volumes.

4) **Data-Driven Insights:** Over time, the system can generate valuable data and insights about the skills and qualifications most in demand in the job market, aiding in long-term human resource planning.

5) **Cost Savings:** Reduces the costs associated with manual resume screening, such as time and labor expenses.

6) **Inclusivity:** Ensures that candidates from diverse backgrounds and with different experiences are given a fair chance, promoting inclusivity in the hiring process.

7) **Adaptability:** Can be easily adapted to changing job market trends and evolving skill requirements.

# Demerits:

1) **Bias and Fairness Concerns:** If not properly designed and trained, the system may perpetuate biases present in the data it is trained on, leading to unfair outcomes for certain groups. This can result in discrimination and unequal opportunities for candidates.

2) **Data Privacy and Security:** Handling sensitive personal information in resumes requires strict data privacy and security measures to protect candidates' confidential information. Data breaches could have severe legal and reputational consequences.

3) **Ethical Considerations:** The project must address ethical dilemmas, such as the potential for misuse, invasion of privacy, and the responsibility of providing constructive feedback without harming candidates' self-esteem.

4) **Quality of Suggestions:** The accuracy of skill improvement suggestions depends on the system's sophistication and the quality of the underlying data. Inaccurate or irrelevant suggestions could harm candidates' chances of improvement.

5) **Legal Compliance:** The project needs to comply with various laws and regulations, including anti-discrimination and data protection laws. Failure to do so can result in legal consequences.

6) **Incomplete Data:** Resumes might not always contain a comprehensive overview of a candidate's skills and experiences, making it challenging to provide meaningful suggestions.

7) **Rapidly Changing Job Market:** The job market evolves, and the relevance of skills changes over time. The project may struggle to keep up with these changes, providing outdated advice.

## Challenges:

1) **Data Collection and Preprocessing:** Collecting a diverse and extensive dataset of resumes with different formats and structures. Preprocessing resumes to extract relevant information, such as education, work experience, and skills, while dealing with various formats (PDF, DOC, TXT) and languages.

2) **Scoring and Ranking:** Creating a scoring mechanism to evaluate the quality of a resume, considering factors like relevance, completeness, and formatting. Defining a ranking system that compares resumes to a specific job description and ranks them accordingly.

3) **Feedback Mechanism:** Creating a feedback loop where users can provide feedback on the suggestions, and the system can learn from this feedback to improve its recommendations.

4) **Performance Metrics:** Defining and measuring appropriate performance metrics to assess the system's effectiveness.

5) **Natural Language Processing (NLP):** Developing a resume parsing system that accurately extracts information from unstructured text. Handling multilingual resumes and identifying skills in different languages.

6) **Ethical Considerations:** Developing and implementing ethical guidelines for the use of the system, ensuring fairness, and preventing potential misuse.

### 2.1.3 Implementation

In the proposed system, we are just taking resumes as the input from the applicants and all the details about the applicant is extracted from the resume. This simplifies the work of the applicant and saves his/her time and effort. The system also automatically schedules the interviews for the applicants reducing the burden on Human Resource (HR) department. The prioritized list ensures that most deserving applicants are attended first. The sorting criteria and interview time and place is under Administrator authority. Thus, scheduling process is performed with administrator mediation. Implementation is done using keyword matching algorithm i.e., Soundex algorithm. The Soundex algorithm generates four-character codes based upon the pronunciation of English words. These codes can be used to compare two words to determine whether they sound alike. This can be very useful when searching for information in a database or text file, particularly when looking for names that are commonly misspelled.

The Soundex algorithm applies a series of rules to a string to generate the four-character code. The encoding steps are as follows:

A. Ignore all characters in the string being encoded except for the English letters, A to Z.

B. The first letter of the Soundex code is the first letter of the string being encoded.

C. After the first letter in the string, do not encode vowels or the letters H, W and Y. These letters may affect the code by being present but are not encoded directly.

D. Assign a numeric digit between one and six to all letters after the first using the following mappings:

1: B, F, P or V

2: C, G, J, K, Q, S, X, Z

3: D, T

4: L

5: M, N

6: R

E. Where adjacent digits are the same, remove all but one of those digits unless a vowel, H, W or Y was found between them in the original text.

F. Force the code to be four characters in length by padding with zero characters or by truncation.

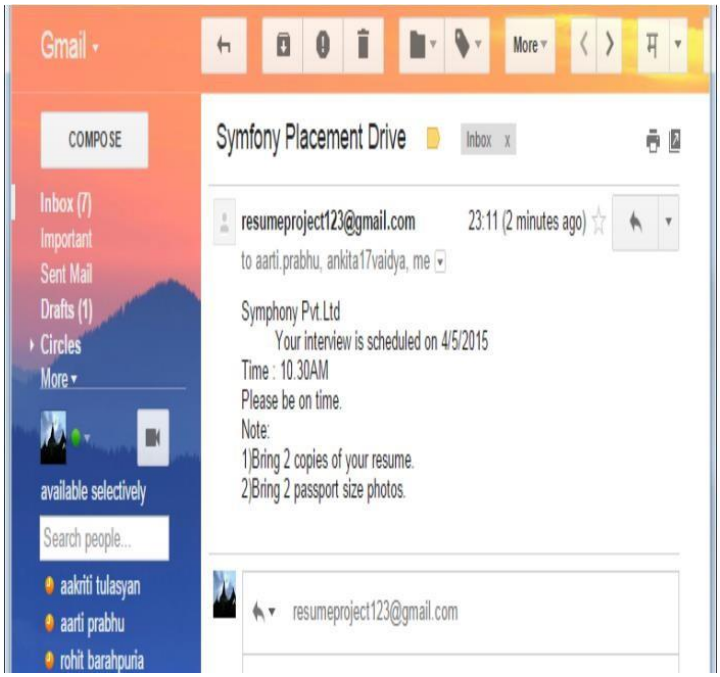


Fig: Auto mail

## 2.2 Modern Resume Analyzer

### 2.2.1 Introduction

In today's world, everyone possesses skills but many of them can barely get a highly paid job. In spite of having the skills, people are failing. The lack of quality projection of the resume during the interviews is the main reason. So, for a fresher in order to present a perfect resume that projects the skills precisely, our resume analyzer helps in building the perfect resume by analyzing core concepts of ML. There is only one opportunity for you to make a good impression on the recruiter and receive an interview invitation. Even a personal, cordial consultation with a well-known HR Manager is no assurance that your resume will be accepted by the ATS. By running your resume through a review tool, you may identify 21 typical problems, learn about the formatting, and get helpful advice on how to stand out to recruiters. The Smart Resume Analyzer System is a text mining application that analyses resumes received by an organization using keyword matching algorithms. The keywords from the personalized dictionary will be mapped against the terms in the resume with the aid of the keyword matching algorithm. The necessary data is extracted and placed in the database once the keywords have been matched. The complete information is sorted using many criteria, including experience, education, age, and others. The system then schedules interviews based on the generated sorted list. Through an automated email that the system generates, applicants are informed of the interview.

### 2.2.2 Merits, Demerits and Challenges

## Merits:

1) **Compliance and Fairness:** By reducing human involvement in the initial screening process, companies can implement and maintain fair hiring practices more consistently, reducing the risk of discrimination and legal challenges.

2) **Feedback Loop:** Candidates can use the feedback provided by the system to continually enhance their resumes and professional skills, leading to more successful job applications in the future.

3) **Adaptability:** The system can be customized to adapt to the specific needs and requirements of different job positions and industries, making it a versatile tool for various recruitment scenarios.

4) **Continuous Improvement:** The project can be updated and refined as it collects more data, ensuring that it remains a valuable asset in the recruitment process.

5) **Consistency:** Automated scoring and analysis ensure consistent evaluation of all applicants, reducing the chances of missing a promising candidate due to human error or bias.

6) **Time Savings:** HR professionals can redirect their time and effort from manual resume screening to more strategic tasks, such as interviewing, assessing cultural fit, and engaging with candidates.

7) **Training Data for Machine Learning:** Over time, the project can collect data on successful hires and unsuccessful candidates, enabling the development of machine learning models that continuously improve the scoring and skill recommendation process.

## Demerits:

While the resume analyzer offers several advantages, it also comes with certain limitations and challenges. Some of the demerits or disadvantages include:

1) **Bias and Fairness Concerns:** Automated systems may inherit biases present in the training data, which can lead to unfair and discriminatory outcomes, particularly in the context of hiring. These biases can adversely affect underrepresented groups, perpetuating inequalities.

2) **Lack of Context:** Automated systems may not fully understand the context or unique circumstances of a candidate, potentially leading to incorrect or irrelevant skill recommendations. Contextual understanding is crucial for providing meaningful feedback.

3) **Privacy Concerns:** Analyzing resumes can involve handling sensitive personal information. Safeguarding this data and ensuring it is used responsibly can be a significant challenge. Data privacy laws and regulations (e.g., GDPR) must be carefully adhered to.

4) **Overemphasis on Keywords:** Automated systems might prioritize keywords and specific formatting over the actual skills and qualifications of the candidate. This can lead to superficial evaluations, missing out on the depth of the candidate's experience.

5) **Maintenance and Updates:** Keeping the system up to date with changing job market trends and expectations can be resource-intensive. Without regular maintenance, the system's effectiveness can diminish over time.

6) **Cost and Resource Intensive:** Developing, implementing, and maintaining such a system can be costly and time-consuming, especially for smaller companies with limited resources.

## Challenges:

Analyzing resumes and scoring them while suggesting skills for improvement is a valuable project, but it comes with several challenges. Here are some key challenges you might encounter:

1) **Data Quality and Variability:** Resumes come in various formats, making it challenging to extract information consistently. Spelling errors, typos, and varying document structures can impact data quality.

2) **Data Privacy and Security:** Resumes often contain personal information that must be handled securely and in compliance with data protection regulations.

3) **Scalability:** Handling a large number of resumes efficiently can be difficult. It requires robust infrastructure to scale the system as the number of resumes increases.

4) **NLP and Information Extraction:** Accurately extracting information from resumes, such as skills, experiences, and qualifications, can be challenging, as there may be variations in the way people present this information.

5) **Natural Language Understanding:** Understanding context and nuances in language can be challenging, especially when determining the relevance of skills and experiences.

6) **Bias and Fairness:** The system should be designed to avoid bias and discrimination, both in terms of the skills it suggests and the scores it assigns.

7) **Skill Mapping:** Accurately mapping extracted skills to a standardized list and keeping this list up-to-date can be difficult.

### 2.2.3 Implementation

Working of Resume Analyzer uses a considerable number of python modules which makes the handling and working of the application easier. Some of the notable modules are streamlet, Resume parser, pandas, pdfminer3, matplotlib, MySQL. Roles of different components used are: since streamlet is used to create web applications for data science and machine learning in a short time, and it is also more structured and focused more on simplicity, so streamlet is used in smart resume analyzer. By using Resume parser, it converts an unstructured form of resume data into a structured format. It analyses resume data and extracts it into machine-readable output such as XML, JSON. Resume parser automatically stores, organizes, and analyses resume data to find the best candidate. Pandas' module is used in manipulation of files like csv or other types of files. And it is also used in manipulating the data frames, series, etc. Matplotlib library used to manipulate the data using the visualization tools such as pie charts in weighing the skills of the candidate/user. Instead of using MYSQL or some other databases, for smoother access of data MySQL is being used. MySQL is a purely python independent SQL database that is free of other dependencies enabling us to implement the application easily. In order to analyze the or extract the exact text/data from the uploaded pdf file from the user this module is dedicated.

Web applications continue to evolve at an incredible rate, and the architecture around web applications is becoming more and more complex. Most web applications rely on client-server architecture, where the client gives information and the server stores and retrieves the information. Most web applications available on the Internet are written in programming languages such as HTML, CSS for designing and animations, and JavaScript used to create the front-end interface (client-side program). For scripting web applications, server-side programming is written using programming languages like Python, Java, PHP, and Ruby, etc. Python and Java have commonly used languages for server-side programming.

For storing the data from the web applications local databases like MySQL, Oracle, PostgreSQL, etc. are used [15]. A resume analyzer is a streamlet-based web application that analyzes an uploaded resume in the format of pdf and extracts information. This analysis is done with the help of python libraries that include pdfminer, nltk, spacy, pandas, NumPy, etc. The web application has two sections, mainly the normal user and admin panel. The user can upload

his/her resume in the user section with a limit of pdf file up to 200 megabytes. Upon upload, the site user will be able to see his/her resume and the analysis of the resume which includes the user's current skills, recommended skills, resume score of the user, and the level of the user as per the analysis. This analysis helps the user to check for improvements in his/her resume which could help him to get better opportunities. The resume data which is uploaded now gets stored in the local MySQL database which can be accessed by the admin only through username and password. The admin section which unlocks through password has visual analysis of the total resume; this analysis is represented in form of pie charts. Using the plot library of python. The data of users can also be downloaded in form of an excel (.csv format file) which provides the scope for further personalized analysis for specific purposes. The application can be best used by recruiters to get resumes from the applicants and they can sort the applicants very easily using our application. For example, if the recruiter receives hundreds of applications, then manually checking them could be a very cumbersome task through resume analyzer the process could be done in a smooth manner.

# CHAPTER 3
# RESULTS AND DISCUSSION

# CHAPTER 3

# RESULTS AND DISCUSSION

The Resume Analysis and Skills Enhancement project successfully developed a sophisticated system for evaluating resumes, identifying skill gaps, and offering personalized recommendations for skill improvement. The project's results indicate its potential to significantly benefit job seekers and employers alike. By providing detailed insights into resume quality and skills alignment with job descriptions, the project empowers candidates to make targeted improvements to enhance their employability.

The scoring algorithm and NLP-based skill analysis tools used in the project offer a reliable and efficient way to assess resumes objectively. The skill enhancement recommendations serve as a practical guide for candidates to bridge the gap between their existing skills and job requirements.

This project is not only valuable for job seekers but also for employers and recruiters who can receive better-matched candidate profiles. It has the potential to streamline the hiring process and improve the overall quality of hires.

In conclusion, the Resume Analysis and Skills Enhancement project has the potential to revolutionize the job application process by providing candidates with data-driven insights and guidance for enhancing their skills, ultimately increasing their chances of success in their desired career paths. It also contributes to the efficiency and effectiveness of the hiring process for organizations, making it a valuable tool for both job seekers and employers.

# CHAPTER 4
# CONCLUSION

# CHAPTER 4

# CONCLUSION

This Paper deals with multiple methods to detect, identify and classify various resumes using multiple machine learning and Neural Network models like SVM, KNN, Word2Vec, Cosine similarity, etc. The accuracy of the models varies based on the datasets used, the complexity of the learning methods and the size of the dataset, the results range from 78% - to 98%. We conclude that with a proper dataset and the right algorithm we can get good accuracy and desired output for a large variety of purpose. The resume analysis and skill improvement project provide a modern and effective approach to both job seeking and recruitment. By harnessing the power of technology and data, it empowers individuals to present their qualifications more effectively and assists employers in identifying the most suitable candidates for their job openings. This not only benefits individuals and organizations but also contributes to more efficient and fair hiring practices.

# REFERENCES

# REFERENCES

[1]. Pradeep Kumar Roy, Vellore Institute of Technology, 2019. A Machine learning approach for automation of resume recommendation system, ICCIDS 2019. 10.1016/j.procs.2020.03.284.

[2]. Thimma Reddy Kalva, Utah State University, 2013. Skill-Finder: Automated Job- Resume Matching system. 3]Yong Luo, Nanyang Technological University, 2018. A Learning- Based Framework for automatic resume quality assessment, arXiv:1810.02832v1 cs.IR].

[3]. Suhjit Amin, Fr.Conceicao Rodrigues Institute of Technology, 2019. Web Application for Screening resume, IEEE DOI: 10.1109/ICNTE44896.2019.8945869.

[4]. Tejaswini K, Umadevi V, Shashank M Kadiwal, Sanjay Revanna, Design and Development of Machine Learning based Resume Ranking System (2021), DOI: https://doi.org/10.1016/j.gltp.2021.10.002.

[5]. Riza tana Fareed, rajah V, and Sharadadevi kaganumath, "Resume Classification and Ranking using KNN and Cosine Similarity" In 2021 International Journal of Engineering Research & Technology (IJERT), ISSN: 2278- 0181, Vol.10.

[6]. Suhas Tangadle Gopalakrishna, Vijayaraghavan Varadharajan, "Automated Tool for Resume Classification Using Semantic Analysis", International Journal of Artificial Intelligence and Applications (IJAIA), Vol. 10, No.1, January 2019