

Resume Builder & Analyzer

Prof. Dr. RajneeshKaur
Sachdeo

Dept. of Computer Science and
Engineering
MIT Art Design and Technology
University
Pune, India

Rajneeshkaur.sachdeo@mit
university.edu.in

Shashwat Sharma

Dept. of Computer Science and
Engineering
MIT Art Design and Technology
University
Pune,India

shashwat2910@gmail.com

Aditi Shirbhate

Dept. of Computer Science and
Engineering
MIT Art Design and Technology
University
Pune,India

aditishirbhate711@gmail.c
om

Aditya Sethi

Dept. of Computer Science and
Engineering
MIT Art Design and Technology
University
Pune,India

adisethi2001@gmail.com

ABSTRACT

Resume analysis is a multi-step process that includes both automated and manual resume reviews. The goal is to extract key information about job applicants from resumes in the most efficient way possible. Every month, some big firms may receive thousands of resumes. Automated resume screening can mitigate the impact of any potential bias and ensure that only the most qualified candidates advance. How? Because, when using a chatbot to screen resumes, you can program it to only evaluate job-specific criteria, such as the specific skills required for the role. AI technology ensures that all submissions are automatically and promptly screened, giving all applicants a fair chance, even if they are not qualified for the job role. The sooner they are informed, the sooner they can proceed.

Keywords

Resume analysis, automated, AI technology, Resume Builder, Models

1. INTRODUCTION

A resume is a formal document that demonstrates a person's professional background and skills. In most cases, it is created to assist a candidate in finding a new job. A traditional resume includes sections for a professional summary, work history, and education. It functions similarly to your job search marketing document. A resume's purpose is to introduce yourself to potential employers, present your qualifications, and secure an interview. The goal of writing a resume is to present your experience, education, and skills in a standardized format that recruiters can easily read. Resume analysis is a multi-step process that includes both automated and manual resume reviews. The goal is to extract key information about job applicants from resumes in the most efficient way possible. Every month, some larger organizations may receive thousands of resumes. They scan incoming resumes into resume management software to reduce the cost of resume reviews. The scanned text is then searched for keywords, work history, education, and years of experience by the software. The system then ranks resumes based on the job criteria and displays a list of the most likely candidates to users. Resumes that are obviously unsuitable may never be seen by a human resources

representative. After the automated resume analysis (if any) is completed, the remaining resumes are examined by hand. Because some resumes are professionally prepared, it can be difficult to determine whether someone is the genuine article or has stitched together a sketchy background into a polished presentation based on the information presented. Some techniques for selecting the best resumes from the pile are listed below.

2. MOTIVATION

Historically, Human Resources departments have been responsible for conducting internal searches for potential candidates using methods such as internal promotion and traditional advertising. While these approaches have proven effective over the years, the rise of technology has introduced a new era of recruitment. Traditional methods are becoming obsolete due to their high costs and time-consuming nature, as well as their inefficiency in identifying highly specialized candidates. In modern recruitment, technology is utilized to locate potential candidates, typically through online job boards and social media sites such as LinkedIn. Other modern recruitment methods, such as online recruitment agencies and inbound marketing, may also be employed depending on the company's needs. Compared to traditional approaches, modern recruitment techniques save the Human Resources department a significant amount of time.

3. OBJECTIVES

The objective of this project is to make the recruitment process hassle free for Job Recruiters.

1. Builder & Analyzer
2. User can upload video resume
3. Get detail analysis of your resume
4. Score base Resumes.
5. Resume can be matched with a job description.
6. Resume generated report will be sent to HR

4. PROJECT

We approached the system development using the waterfall model depicted in the Based on this model, the required estimates have been stated in Annexure. In order to map our estimates with the steps in a waterfall model, we considered each phase separately and then stated the required estimates.

5. SECTIONS

The methodology should include the formulation of the problem as well as the processes used to solve the problem and prove or disprove the hypothesis. Illustrations can help to clarify ideas and support conclusions. How you resolved the issue. In order to solve the problem, methodology is used. This should include both computational and experimental details. If the work is computational, this section discusses the methodology used to solve the problem.

STEP 1: Creating a website for candidates and HR for Job Recruitments. 2)Image to text: OCR – Optical Character Recognition. 3)Analysis report: Panda’s profiling Report

STEP 2: Create a bucket to store the Resume, Video Resume and Certifications.

STEP 3: Apply Model to analysis the Candidates Profile for HR. 1) Keywords extractor: NLP- RAKE that is supported by NLTK 5) Correlation: NumPy, SciPy, Pandas singular value decomposition (SVD) 6) Job recommendation.

5.1 Block diagram of proposed seminar work

5.1.1 Candidate's View:

1. Candidates will be able to login or sign up.
2. The User will have an option to either upload his own resume or make one from our website.
3. Candidate will be able get the analysis with the help of our website
4. Users will also have a choice to upload the video resume.
5. If a user wants to apply to a company, he has to undergo a psychometric tests and then apply.

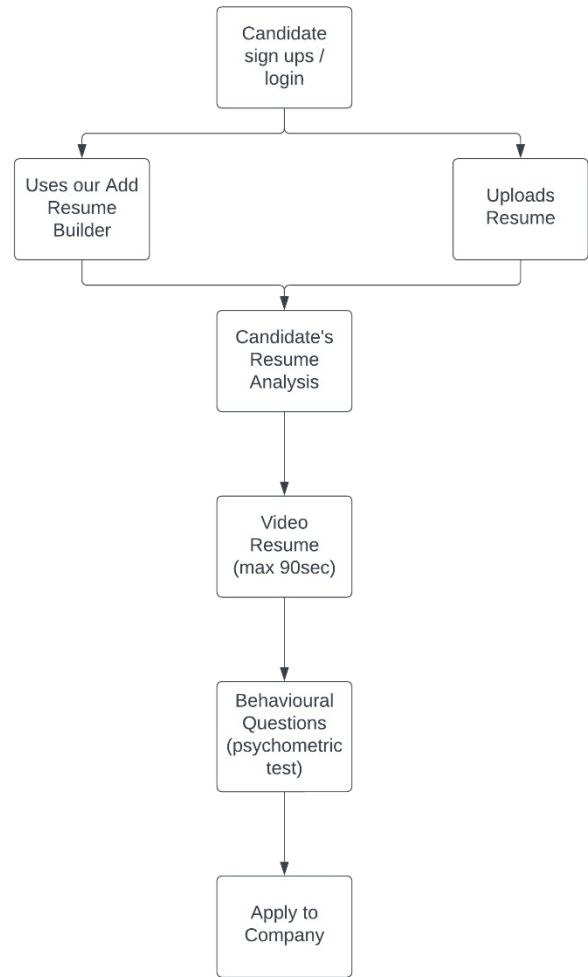


Figure 1: Candidates View

5.1.2 HR's View:

1. HR will be able to sign up/login
2. The list of candidates will be popped up in front
3. Best candidates resume will be at the top
4. The detailed report of video analysis, behavioral questions will be given
5. Then the interview process will be started

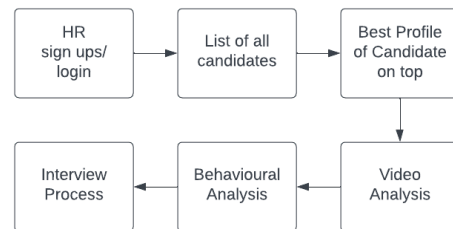


Figure 2: HR's View

5.1.3 Backend Process:

1. When the candidate logs in a unique Id will be generated
2. This UUID will be saved in the database
3. All the information which the candidate uploads will be saved.
4. With the help of the backend ML models the analysis will be done

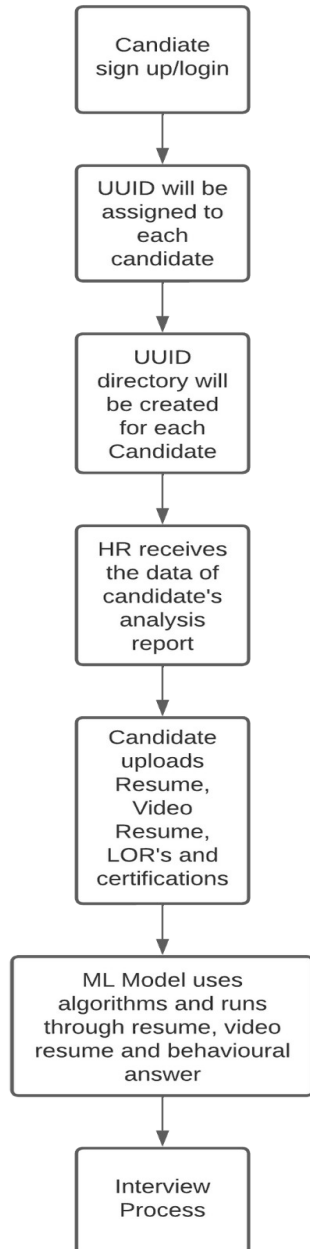


Figure 3: Projects's View

5.2 Models Used:

5.2.1 Docx2txt

Docx2txt is a software package that allows users to extract plain text from Microsoft Word documents (files with the .docx extension). It works by removing any formatting, styles, and images from the document and providing a clean, unformatted version of the text. This plain text version of the document can then be used for further analysis, such as keyword extraction or natural language processing (NLP). Docx2txt is a useful tool for researchers, data analysts, and anyone who needs to extract text from large numbers of Word documents.

5.2.2 NLP

Natural Language Processing (NLP) is a field of computer science and artificial intelligence (AI) that focuses on enabling computers to understand, interpret, and generate human language. It involves developing algorithms and models that can process and analyze natural language data, such as text and speech, in order to perform various tasks, including sentiment analysis, machine translation, text summarization, and chatbot development, among others. NLP combines techniques from computer science, linguistics, and mathematics to enable computers to "understand" human language and communicate with humans in a more natural way.

5.2.3 Sklearn

Scikit-learn (sklearn) is a popular open-source Python library that provides a range of tools for machine learning and data analysis tasks. It includes a variety of machine learning algorithms, such as classification, regression, clustering, and dimensionality reduction, as well as tools for preprocessing, model selection, and evaluation. Scikit-learn is designed to be user-friendly and efficient, making it a valuable tool for both beginners and experienced practitioners. It also offers integration with other Python libraries, such as NumPy, Pandas, and Matplotlib, for data manipulation and visualization. Scikit-learn is widely used in industry and academia for various applications, such as predictive modeling, image recognition, and natural language processing.

5.2.3 NumPy

NumPy is a Python library for scientific computing and numerical analysis that provides a powerful array data structure, which enables efficient manipulation of large multi-dimensional arrays and matrices. NumPy also provides a variety of mathematical functions and tools for working with these arrays, such as linear algebra operations, Fourier transforms, and random number generation.

NumPy arrays are homogeneous, meaning all elements of the array must be of the same data type. This allows for fast, memory-efficient operations and makes NumPy a popular choice for data analysis, machine learning, and scientific computing tasks. In addition, NumPy is also used as a foundation for many other scientific Python libraries, such as Pandas and Matplotlib. Overall, NumPy is a fundamental library in the Python data science ecosystem and is widely used in academia and industry.

5.3 Technologies Used:

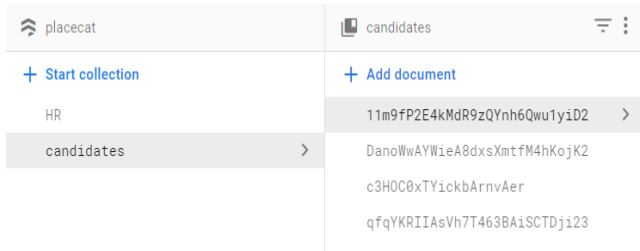
5.3.1 Firebase Authentication

Identifier	Providers	Created ↓	Signed In
saivedant169@gmail.com	📧	Apr 29, 2023	Apr 29, 2023
shashwatdev2910@gmail...	📧	Apr 29, 2023	Apr 29, 2023
shashwat2910@gmail.com	📧	Feb 28, 2023	May 1, 2023
adisethi2001@gmail.com	📧	Jan 23, 2023	Jan 24, 2023
madhuri.jadhav@mituniver...	📧	Dec 2, 2022	Dec 2, 2022
pranay@dosii.com	📧	Oct 15, 2022	Oct 15, 2022
rksharma9472@gmail.com	📧	Oct 15, 2022	Oct 15, 2022
monil.pvt04@gmail.com	📧	Oct 14, 2022	Oct 15, 2022

Figure 4: Firebase Authentication

Firebase Authentication is a service provided by Google that allows developers to easily authenticate users to their application. Each user registered in the application is assigned a unique identifier or UUID, which is used to identify them throughout the system. Additionally, Firebase Authentication requires that each user verify their email address before they can log in to the application. This verification process helps ensure that users are authentic and have access to a valid email address. Once a user has verified their email address, they can securely log in to the application using Firebase Authentication, and the system can grant or restrict access to specific features or data based on the user's permissions.

5.3.2 Firestore Database



placecat	candidates
+ Start collection	+ Add document
HR	11m9fP2E4kMdR9zQYnh6Qwu1yiD2 >
candidates >	DanoWwAYW1eA8dxsXmtfM4hKojK2
	c3H0C8xTYickbArnvAer
	qfqYKRiIAsVh7T463BAiSCTDji23

Figure 5: Firestore

Firestore is a NoSQL database provided by Google that allows developers to store and manage data for their applications. In this context, each user registered in the application is assigned a unique identifier or UUID, which is used to identify them throughout the system. The Firestore database likely contains a collection of user documents, with each document containing the user's basic details such as their name, email address, phone number, and any other relevant information. With Firestore, developers can securely store and manage this data, and use it to personalize the user's experience in the application. Additionally, Firestore allows for easy querying and filtering of data, making it simple to find specific user information when needed.

5.3.3 Firebase Storage

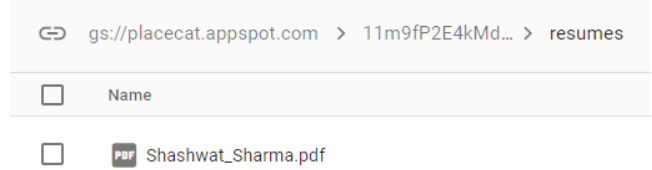


Figure 6: Firebase Storage

Firestore Storage is a cloud-based object storage service provided by Google that allows developers to store and serve user-generated content such as images, videos, and documents. In this context, the storage system likely has a directory for each user, with the directory name being the unique identifier or UUID of the user from Firebase Authentication. This approach helps ensure that each user's data is kept separate and secure. Within each user directory, the user can upload their resume, video resume, or any other relevant documents. Additionally, there may be a separate directory for company data, where job descriptions can be stored for easy access by recruiters and hiring managers. Firestore Storage allows for easy and secure storage of user-generated content, making it simple for developers to build powerful applications that rely on user-generated data.

5.3.4 ResumeScore Model

```
[8] content = (job_description , resume)

[9] from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer()
matrix = cv.fit_transform(content)

[10] from sklearn.metrics.pairwise import cosine_similarity
similarity_matrix = cosine_similarity(matrix)

print(similarity_matrix)

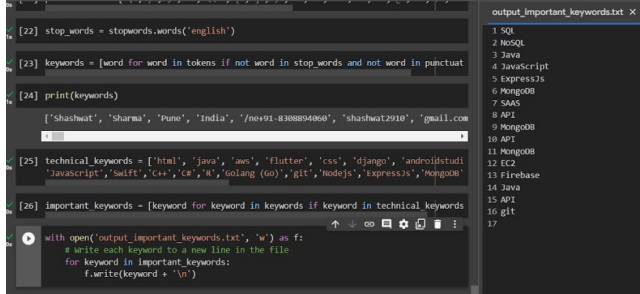
[[1. 0.48365354]
 [0.48365354 1.  ] ]

print('Resume matches by: '+str(similarity_matrix[1][0]*100)+'%')
Resume matches by: 48.36535391380342%
```

Figure 7: ResumeScore

The job of matching job descriptions with resumes can be a daunting task for recruiters and hiring managers. However, with the help of machine learning models, this process can be streamlined and automated. The Cosine similarity library measures the similarity of two texts by calculating the cosine of the angle between the vectors representing the texts. The result is a score in percentage that represents how closely the text of the resume matches the job description. This score can be used by recruiters and hiring managers to quickly assess whether a candidate is a good fit for a particular job, and to prioritize which resumes to review first. This ML model can help save time and resources while improving the overall quality of the hiring process.

5.3.5 SkillExtractor Model



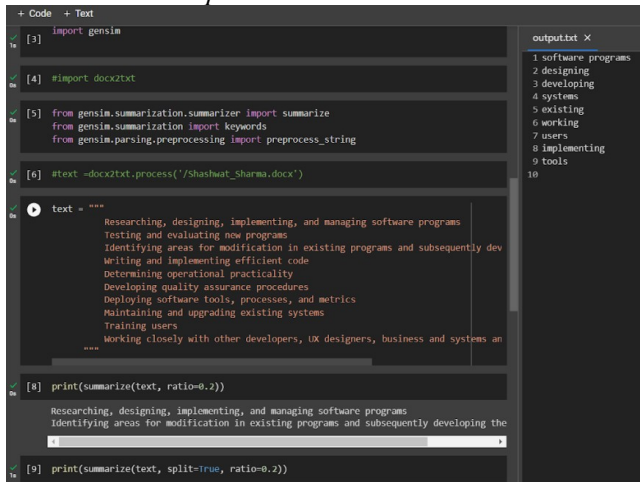
```
(22) stop_words = stopwords.words('english')
(23) keywords = [word for word in tokens if not word in stop_words and not word in punctuat
(24) print(keywords)
['shashwat', 'Sharma', 'Pune', 'India', '/ne491-83080406d', 'shashwat2910', 'gmail.com
(25) technical_keywords = ['html', 'java', 'aws', 'flutter', 'css', 'django', 'androidstudio
'javascript', 'swift', 'c++', 'c', 'k', 'colang (oo)', 'git', 'nodejs', 'expressjs', 'mongodb
(26) important_keywords = [keyword for keyword in keywords if keyword in technical_keywords

with open('output_important_keywords.txt', 'w') as f:
    # write each keyword to a new line in the file
    for keyword in important_keywords:
        f.write(keyword + '\n')
```

Figure 8: SkillExtractor

The process of extracting technical skills from a resume can be a time-consuming and challenging task for recruiters and hiring managers. However, machine learning models can be trained to automate this process and extract relevant skills from a resume quickly and accurately. One such model uses `textract`, `nlTK`, and `word_tokenize` libraries in Python to extract technical skills from a resume. The model begins by using `textract` to extract the text from the resume, which is then preprocessed using `nlTK`'s `word_tokenize` library to break the text into individual words. The model then applies various natural language processing techniques to identify and extract technical skills from the text, such as named entity recognition, part-of-speech tagging, and keyword matching. The resulting output is a list of relevant technical skills extracted from the resume, which can be used by recruiters and hiring managers to assess a candidate's qualifications and match them to specific job requirements. This ML model can help streamline the hiring process and improve the overall quality of candidate selection.

5.3.6 JobDescriptionSummarizer Model



```
(3) import gensim
(4) #import docx2txt
(5) from gensim.summarization.summarizer import summarize
from gensim.summarization import keywords
from gensim.parsing.preprocessing import preprocess_string
(6) #text =docx2txt.process('/Shashwat_Sharma.docx')

text = """
Researching, designing, implementing, and managing software programs
Testing and evaluating new programs
Identifying areas for modification in existing programs and subsequently dev
Writing and implementing efficient code
Determining operational practicality
Developing quality assurance procedures
Deploying software tools, processes, and metrics
Maintaining and upgrading existing systems
Training users
Working closely with other developers, UX designers, business and systems an

(8) print(summarize(text, ratio=0.2))

Researching, designing, implementing, and managing software programs
Identifying areas for modification in existing programs and subsequently developing the

(9) print(summarize(text, split=True, ratio=0.2))
```

Figure 9: JobDescriptionSummarizer

Summarizing a job description in terms of the required technical skills can be a helpful tool for recruiters and hiring managers to quickly assess whether a candidate is a good fit for a particular job. Machine learning models can be trained to analyze the text of a job description and extract the required technical skills, which can then be summarized in a concise and easy-to-understand format. The model might use natural language processing techniques such as named entity recognition and keyword matching to identify and extract technical skills from the text of

the job description. The resulting output is a summary of the required technical skills, which can be used to match candidates with the job requirements and streamline the hiring process. This ML model can help improve the efficiency of the hiring process while ensuring that candidates are being matched to the most appropriate jobs based on their technical skills.

6. PROJECT SCOPE

The proposed work aims to answer these questions

- Detailed library of taxonomies to identify candidate skills
- Machine learning-based parsing.
- Email integration.
- Process common file formats, i.e., PDF, Doc, Docx, HTML, RTF
- Support for collaborative hiring processes.
- Integration with ATS and other applications.
- Add on for the company's and user to easily access the website

7. FUTURE WORK

In the future, we can enhance our model by providing a more comprehensive analysis that includes an evaluation of a candidate's facial expressions and tonal quality of voice to assess their level of confidence. Video content analysis (VCA) or video analytics (VA) is a technique that automatically analyzes video to detect temporal and spatial events. For degree verification, we cannot rely solely on ground-level checks. Therefore, we plan to use third-party verification services like DigiLocker, which is an initiative under the Digital India program aimed at providing citizens with authentic digital documents. We aim to develop a module within our system that allows users to assess their resume skills and receive an evaluation after submitting their CV. Our future plans also include offering better pay scales and commission opportunities for our users, as well as implementing a detailed employee evaluation system to improve the efficiency of the hiring process for employers and clients seeking to hire freelance employees.

8. CONCLUSION

Finally, in today's job market, the use of resume builder and analyzer tools is becoming increasingly popular. These tools provide job seekers with a quick and easy way to create and analyze their resumes, allowing them to tailor their applications to specific job postings and stand out from the crowd.

We discovered that resume builder and analyzer tools can help job seekers save time, improve the quality of their resumes, and increase their chances of being noticed by potential employers through our research. It is important to note, however, that these tools should not be relied on entirely, as human input and judgment are still required in the job application process.

Furthermore, we have identified some limitations and potential biases that may exist in these tools and should be considered when using them. Job seekers must understand the algorithms used by these tools and manually review their resumes to ensure that they accurately reflect their qualifications and experiences.

Overall, we believe that resume builder and analyzer tools can be useful resources for job seekers; however, they should be used in conjunction with other job search strategies and with a critical

eye on their potential limitations.

9. REFERENCES

- [1] Rinki Tyagi, "Resume Builder Application" May 2020, DOI:[10.22214/ijraset.2020.5003](https://doi.org/10.22214/ijraset.2020.5003)
- [2] "RESUME BUILDER APPLICATION", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.8, Issue 3, page no.18-20,March-2021,Available:<http://www.jetir.org/papers/JETIREN06004.pdf>
- [3] Prasuna Pokharel,"RESUME PARSER USING NLP",July 2022, DOI:[10.13140/RG.2.2.10323.25127/1](https://doi.org/10.13140/RG.2.2.10323.25127/1), Affiliation: University of Wolverhampton
- [4] "RESUME BUILDER WITH CLASSIFIER" Rishil Shah*1, Vishal Patel*2, Hriday Lal*3 *1,2,3Student, Department Of Information Technology, Thakur Shyamnarayan Degree College,Mumbai,Maharashtra,India. Volume:04/Issue:05/May-2022
- [5] Ankita Satish Vaidya, Pooja Vasant Sawant, " Resume Analyzer An Automated solution to recruitment process",ISSN: 2321-0869 (O)2454-4698 (P), Volume- 3, Issue-8, August 2015.
- [6] Suhjit Amin, Fr.Conceicao Rodrigues Institute Of Technology, 2019. Web Application forScreening resume, IEEE DOI:10.1109/ICNTE44896.2019.8945869
- [7] Tejaswini K, Umadevi V, Shashank M Kadiwal,Sanjay Revanna, Design and Development ofMachine Learning based Resume RankingSystem (2021), DOI:<https://doi.org/10.1016/j.gltp.2021.10.002>.
- [8] Resume Screening Using Machine Learning and NLP : A Proposed System Bhushan Kinge 1, Shrinivas Mandhare2, Pranali Chavan3, S. M. Chaware4, doi :<https://doi.org/10.32628/CSEIT228240>
- [9] Pradeep Kumar Roy, Vellore Institute ofTechnology, 2019. A Machine learning approach for automation of resume recommendation system, ICCIDS 2019.10.1016/j.procs.2020.03.284