| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

Intelligent Job Match

1. V Rama Krishna Assistant professor CSE department Anurag University,

2. Kodumuri Likhitha 20EG105225 AU, 3. S. Nikhila Reddy 20EG105245 AU,

4. Dutta Saurabh 20EG105212 AU

ABSTRACT: In today's digital era, the volume of resumes received by organizations has increased significantly, necessitating the need for automated resume analysis systems. This project presents a robust and efficient system for parsing and analyzing resumes using natural language processing (NLP) techniques and machine learning algorithms. The system leverages tools such as spaCy, NLTK, and scikit-learn to extract key information from resumes, including contact details, skills, work experience, education, and certifications. It incorporates advanced algorithms like TF-IDF (Term Frequency-Inverse Document Frequency) and cosine similarity to rank resumes based on job requirements, skill matches, and relevancy scores. The system provides an intuitive user interface for uploading resumes, entering job descriptions, and generating ranked lists of potential candidates. Evaluation metrics such as accuracy, precision, and recall are used to assess the system's performance, ensuring high-quality candidate shortlisting and efficient recruitment processes. Overall, this system streamlines the resume screening and analysis process, saving time and resources while improving the quality of candidate selection in recruitment workflows.

KEYWORDS: Resume Analysis, nltk, Job Description Parsing, Skill Extraction, Experience Level Determination, Candidate Ranking, Similarity Score Calculation, TF-IDF (Term Frequency-Inverse Document Frequency), Cosine Similarity

I. INTRODUCTION

Intelligent Job Match represents a significant leap in recruitment technology, leveraging state-of-the-art Natural Language Processing (NLP) techniques to redefine how organizations approach talent acquisition. By deeply analyzing resumes, this cutting-edge system assigns nuanced scores to job roles, evaluating candidates based on a thorough assessment of their skills and professional backgrounds. Through sophisticated NLP algorithms, the platform identifies and categorizes keywords within resumes, creating insightful clusters that reflect candidates' expertise across various sectors and domains. Unlike traditional resume parsers, Intelligent Job Match goes beyond surface-level assessments, providing not just an overall suitability score but also offering targeted insights, predictions, and analytics.

For job seekers, the platform acts as a valuable guide, highlighting areas for skill improvement and suggesting strategies to enhance their resumes. On the recruiter's side, it streamlines the hiring process by providing comprehensive candidate evaluations and facilitating data-driven decisions.

This technology empowers both parties with actionable intelligence, promoting better matches and elevating the overall quality of talent acquisition processes in today's competitive job market. Intelligent Job Match heralds a new era of recruitment efficiency, where data-driven insights and intelligent recommendations drive better hiring outcomes.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

II. RELATED WORK

The suggested method is a resume ranking software that makes use of Mong, Cosine Similarity, and Overlapping Coefficient Natural Language as well as Natural Language Processing and Machine Learning. After filtering, it offers a ranking and suggests the best resume for a specified text job description. Rather than processing the entire document, the supplied information is summarised after data cleansing. Exploratory data analysis, or EDA, is a method used to develop trends and visual representations for the supplied data collection [14]. To determine the frequency of terms that are used to meet the job description, TF-IDF is employed. The Web UI is used to extract the subjects from the document collection and display them. The user side of the dashboard allows users to upload their resumes in PDF format and access information on the abilities needed to land a job. Even the courses and advice for improved development are given to the users. The administrator side logs into the portal using the credentials and has access to data on the applicants' skill levels and resume scores. Both the user and the admin benefit from a hassle-free resume screening due to the complete process.

Keyword Extraction and Resume Parsing: In the first step of the process, user-uploaded PDF resumes are automatically processed to extract text content. This data extraction utilizes PDF parsing libraries and advanced Natural Language Processing (NLP) techniques such as Named Entity Recognition (NER) and syntactic parsing. These techniques help identify and categorize relevant details like skills, experiences, and education from the resumes. Following data extraction, a scoring mechanism is employed to assess the relevance of extracted skills to specific job roles, improving candidate screening efficiency. Integrated NLP models like spaCy and NLTK play a crucial role in accurately identifying key entities within the resume content. This approach streamlines resume analysis, facilitates informed decision-making in candidate selection, and optimizes recruitment workflows to handle large volumes of resumes effectively.

Matching Algorithm: The matching algorithm compares keywords extracted from the job description with content from each resume, evaluating their relevance. It assigns a score or weight to each matched keyword, reflecting its importance and alignment with job requirements. This process involves analyzing both the frequency and context of keywords within the resume content. The algorithm may utilize techniques like TF-IDF (Term Frequency-Inverse Document Frequency) to measure keyword importance. By considering the specificity and uniqueness of matched keywords, the algorithm enhances accuracy in identifying suitable candidates. This approach optimizes the screening process by quantifying the degree of match between resume content and job expectations. It aids recruiters in efficiently shortlisting candidates who closely match the desired skill set and experience criteria. Overall, the matching algorithm improves the effectiveness and precision of candidate evaluation in the recruitment workflow.

Threshold Setting: The matching algorithm compares keywords extracted from the job description with content from each resume, evaluating their relevance. It assigns a score or weight to each matched keyword, reflecting its importance and alignment with job requirements. This process involves analyzing both the frequency and context of keywords within the resume content. The algorithm may utilize techniques like TF-IDF (Term Frequency-Inverse Document Frequency) to measure keyword importance. By considering the specificity and uniqueness of matched keywords, the

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

algorithm enhances accuracy in identifying suitable candidates. This approach optimizes the screening process by quantifying the degree of match between resume content and job expectations. It aids recruiters in efficiently shortlisting candidates who closely match the desired skill set and experience criteria. Overall, the matching algorithm improves the effectiveness and precision of candidate evaluation in the recruitment workflow.

Real-time Advice Machine: The real-time recommendation system integrates seamlessly with job matching platforms, leveraging its trained model to evaluate incoming resumes swiftly. By applying advanced algorithms, it accurately matches candidate profiles with relevant job roles in real-time. This system provides intelligent recommendations based on skills, experience, and qualifications, enhancing the efficiency of the recruitment process. For applicants, it offers personalized job suggestions that align with their expertise and career aspirations. Simultaneously, recruiters benefit from predictive analytics, gaining insights into candidate-job fit and potential hiring trends. The system's real-time capabilities ensure timely and informed decision-making, improving overall recruitment outcomes and enhancing user experience on the platform.

III. PROPOSED ALGORITHM

TF-IDF: The most popular technique for calculating word frequencies is called TF-IDF. This stands for "Term Frequency - Inverse Document" Frequency, which is one of the factors used to calculate a word's ultimate score [10]. Without getting into the numbers, TF-IDF word frequency scores seek to highlight more intriguing phrases, such as those that are repeated inside a text but not across texts. The TF-IDF Vectorizer can encode new frequency weightings, learn new terminology, tokenize texts, and invert frequency weightings.

Term Frequency: How frequently a word appears in a document is referred to as its term frequency.

Inverse Document Frequency: Downscale phrases that regularly appear in documents are referred to as having an inverse document frequency.

$$F - DF\,(,) = F\,(,) * DF\,(,) \tag{1}$$

$$F(.) = \frac{rq(.)}{\sum rq(i,)} \tag{2}$$

$$DF() = o\left(\frac{}{on()}\right) \tag{3}$$

A. Where, freq (t, d) is the number of times the word t appears in document d N is the count of unique words in document d

B. TF (t, d) is the portion of the frequency of term t in document d

Using equation (1) (2) and (3) the TF-IDF is calculated; a term's higher TF-IDF score, which is determined using the formulas above, indicates that it is more relevant to the content. We modelled the resumes and JD into a vector space in our system. This is done by assembling and translating a dictionary of terminology from the papers. A dimension of the vector space is assigned to each sentence. We created the TF- IDF matrix for the resumes and the job query using the Count Vectorizer and the TF- IDF matrix.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

Cosine Similarity: Cosine similarity is a metric for comparing two non-zero vectors in an inner product space. Its value is the same as the inner product of the identical vectors normalized to have the same length or the cosine of the angle between them. The cosine of 0 degrees is 1, and any angle between (0,] radians) has a cosine that is less than 1. Thus, rather than considering the magnitude, the comparison is based on orientation: two vectors with the same orientation have a cosine similarity of 1, two vectors oriented at 90 degrees from one another have a similarity of 0, and two vectors opposite have a similarity of -1, regardless of magnitude. The cosine similarity is quite helpful in positive space since the result is cleanly constrained in presentation style [0,1] [0,1]. The name is derived from the phrase "direction cosine": in this example, unit vectors are maximally "dissimilar" if they are orthogonal and maximally "similar" if they are parallel. that is equivalent to the cosine, which has a price of cohesion whilst the segments subtend a 0 attitude and a cost of zero while the segments are perpendicular. the characteristic vectors a and b for text matching are normally the term frequency vectors of the files. Equation 4 can be used to compare documents and use cosine similarity to standardize document length.

$$\cos = \frac{\vec{\cdot}\vec{}}{\|\vec{}\|\vec{}\|} = \frac{\sum_{i-1}^{n} i\,i}{\sqrt{\sum_{i-1}^{n} i^2 . \sum_{i-1}^{n} i^2}} \qquad (4)$$

The below figure represents cosine similarity where, ai and bi, respectively, are the parts of the vectors A and B. Below is a representation of the general Cosine Distance/Similarity formula [12]. In this instance, X1 represents the resume, X2 represents the job description that was provided by the resume, and Item1 and Item 2 represent the words that were translated into the vectors that were contained in the resume uploaded by the applicant and the job description of the recruiter.

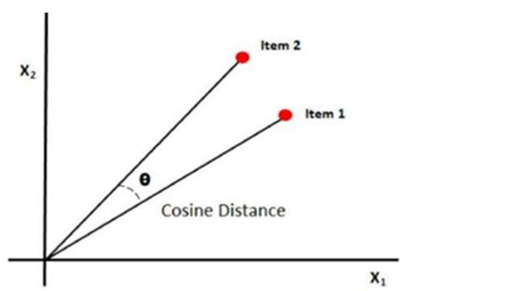


Fig. 2. Cosine similarity

NLP: The JSON train is also transferred through the NLP channel after the textbook reflections in the pdf train are converted to JSON format. This NLP channel is also used to train the model. Using the NLP frame SpaCy, it can be trained. SpaCy is a frame that was developed for general data rather than for datasets, like a capsule. In this system, rather than manually entering every word to produce the dataset, semi-supervised literacy is employed to marker the significant data in the ZIP train of PDF resumes. Homonyms, homophones, affront, expressions, conceits, exceptions to the rules of alphabet and operation, and changes in judgment structure are just many exemplifications of the irregularities in mortal language that take humans times to learn but that programmers must educate natural language-driven operations to fete and understand directly from the morning if those operations are to be useful.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

IV. SIMULATION RESULTS

There are two modules in Intelligent Job Match, user and admin while each having its own special functions.

Resumes are uploaded by the applicants, which later undergo few pre-processing procedures. The talents are extracted from the curriculum vitae and compared to the obligatory competencies. The missing capabilities are mentioned under the recommended skills where, the access to sources that avail in development.

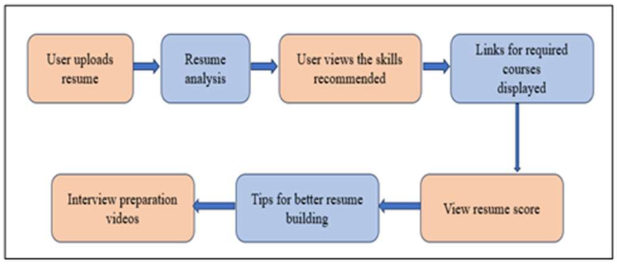


Fig. 3. User work flow

Access to the admin module's features requires providing the necessary credentials. Admins can access a comprehensive report of all applicant resumes, including applicant name, ID, and email address. The report also displays predicted fields and experience levels determined by skills and scores.

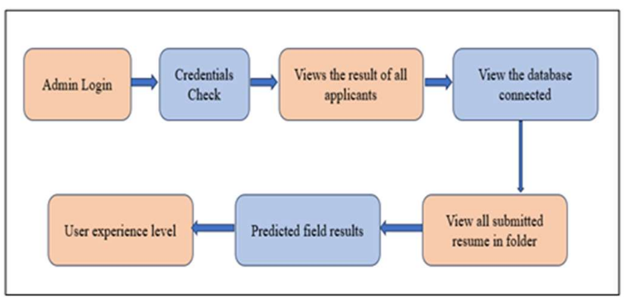
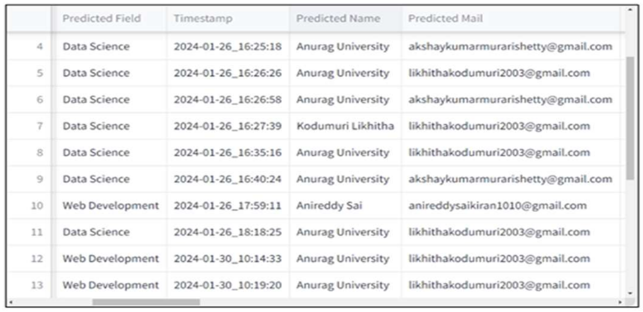


Fig. 4. Admin work flow

Furthermore, recommended courses, skills, and actual talents are listed. Each resume is timestamped, and details are automatically uploaded to the connected database, with the resumes saved in a designated destination folder. This streamlined process enhances efficiency and provides valuable insights for effective candidate management.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |



	Predicted Field	Timestamp	Predicted Name	Predicted Mail
4	Data Science	2024-01-26_16:25:18	Anurag University	akshaykumarmurarishetty@gmail.com
5	Data Science	2024-01-26_16:26:26	Anurag University	likhithakodumuri2003@gmail.com
6	Data Science	2024-01-26_16:26:58	Anurag University	akshaykumarmurarishetty@gmail.com
7	Data Science	2024-01-26_16:27:39	Kodumuri Likhitha	likhithakodumuri2003@gmail.com
8	Data Science	2024-01-26_16:35:16	Anurag University	likhithakodumuri2003@gmail.com
9	Data Science	2024-01-26_16:40:24	Anurag University	akshaykumarmurarishetty@gmail.com
10	Web Development	2024-01-26_17:59:11	Anireddy Sai	anireddysaikiran1010@gmail.com
11	Data Science	2024-01-26_18:18:25	Anurag University	likhithakodumuri2003@gmail.com
12	Web Development	2024-01-30_10:14:33	Anurag University	likhithakodumuri2003@gmail.com
13	Web Development	2024-01-30_10:19:20	Anurag University	likhithakodumuri2003@gmail.com

Fig. 5. Resume Report

The users field forecasts and level of experience are statistically represented as a pie chart. The below pie chart is plotted against the applicant's aptitudes to predict the field they might prosper, this avails recruiters in understanding which domain are most users familiar with.

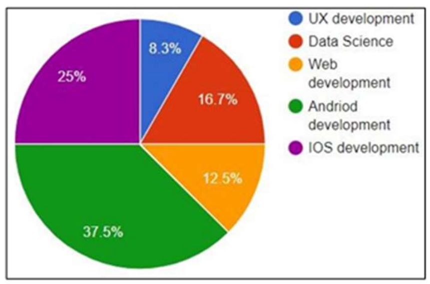


Fig. 6. Pie Chart for Predicted Field

The below cited pie chart is about the user’s experience level which is determined by the job description and qualifications mentioned in resume. The admin utilizes the experience level to narrow down the type of employee they are seeking for based on the job description.

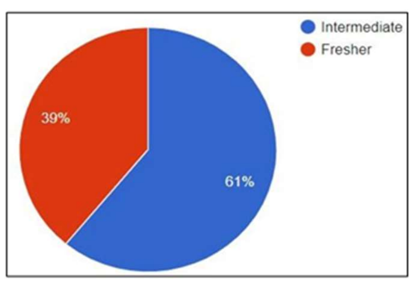


Fig. 7. Pie Chart for User Experience level

The process of bulk resume analysis and ranking involves several key steps that contribute to efficient candidate evaluation and ranking based on job relevance.

Initially, the system allows for the bulk upload of resumes, streamlining the input of candidate data. Upon upload, each resume undergoes data processing to extract vital information such as skills, experiences, and educational background.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

A ranking algorithm is then applied to assess the similarity between each resume and the provided job description. This algorithm calculates a similarity score, often leveraging techniques like TF-IDF (Term Frequency-Inverse Document Frequency) and Cosine Similarity. Based on these similarity scores, resumes are ranked in descending order, with the most relevant candidates receiving higher ranks.

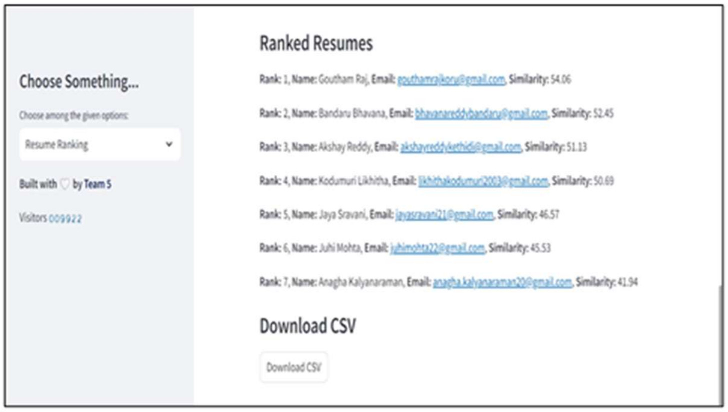
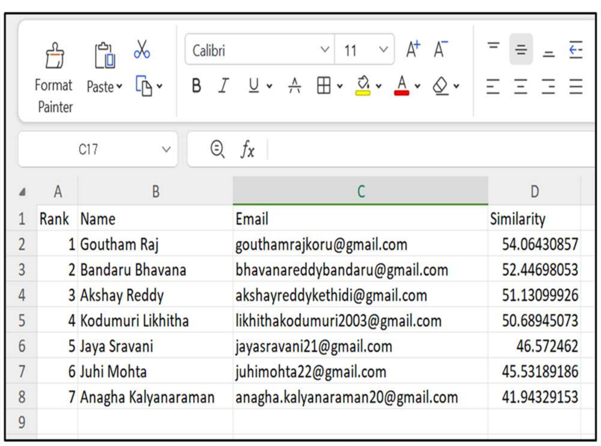


Fig. 8. Output of resumes are ranked in descending order

To ensure accuracy and avoid redundancy, a deduplication process may be employed, eliminating duplicate entries based on candidate names or email addresses. The ranked resumes are presented visually through tables or charts, offering recruiters a clear overview of candidate rankings and their corresponding similarities to the job description.

Moreover, the system generates downloadable reports, typically in CSV format, containing ranked lists of resumes along with pertinent details such as candidate names, email addresses, similarity scores, and ranks. These reports serve as valuable references for recruiters during the candidate selection process, facilitating informed decision-making and optimizing the recruitment workflow. Overall, this structured approach enhances recruitment efficiency and assists recruiters in identifying top candidates based on their qualifications and relevance to job requirements.



Rank	Name	Email	Similarity
1	Goutham Raj	gouthamrajkoru@gmail.com	54.06430857
2	Bandaru Bhavana	bhavanareddybandaru@gmail.com	52.44698053
3	Akshay Reddy	akshayreddykethidi@gmail.com	51.13099926
4	Kodumuri Likhitha	likhithakodumuri2003@gmail.com	50.68945073
5	Jaya Sravani	jayasravani21@gmail.com	46.572462
6	Juhi Mohta	juhimohta22@gmail.com	45.53189186
7	Anagha Kalyanaraman	anagha.kalyanaraman20@gmail.com	41.94329153

Fig. 9. Contents of csv file

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

V. Conclusion and Future Work

The challenges faced by both recruiters and job seekers in the current job market underscore the critical necessity for innovative solutions like automated resume screening and analysis systems. Recruiters grapple with the overwhelming volume of resumes, particularly for specialized roles like Data Analysts, which strains HR teams. Manual review processes are not only time-consuming but also prone to errors, risking the oversight of vital skills and qualifications.

On the flip side, job seekers struggle to pinpoint suitable roles that match their expertise and aspirations amidst a sea of job listings with varying requirements. This leads to confusion and inefficiency in the application process, prompting applicants to scatter their efforts across multiple positions, diminishing their chances of securing an ideal job match.
The proposed automated resume screening and analysis system tackles these challenges by harnessing advanced technologies such as PDF parsing, Natural Language Processing (NLP), and machine learning algorithms. By automating the extraction of crucial information from resumes and implementing a scoring mechanism based on skill relevance, the system streamlines the recruitment process for recruiters. It enables swift identification of qualified candidates, saving time and resources while ensuring a more precise screening process.

For job seekers, the system offers a targeted approach to job searching by recommending roles closely aligned with their skills and qualifications. This reduces the time and effort spent on applying to irrelevant positions, enhancing the likelihood of securing an ideal job fit. Ultimately, the proposed system aims to bridge the gap between recruiters and job seekers, fostering a more efficient and effective recruitment process that benefits both parties and contributes to a streamlined and productive hiring ecosystem.

References

[1] Web Application for Screening Resume- https://ieeexplore.ieee.org/document/8945869

[2] Recruitment Prediction using Machine Learning- https://ieeexplore.ieee.org/document/9276955

[3] Business Intelligence Assistant for Human Resource Management for IT Companies- https://ieeexplore.ieee.org/document/9325471

[4] Screening and Ranking Resumes using Stacked Model- https://ieeexplore.ieee.org/document/9707977

[5] Automated Resume Screener using Natural Language Processing(NLP)- https://ieeexplore.ieee.org/document/9777194