*International Journal of Research in Engineering and Science (IJRES)*
*ISSN (Online): 2320-9364, ISSN (Print): 2320-9356*
*www.ijres.org Volume 11 Issue 3 ǁ March 2023 ǁ PP. 409-418*

# Smart Resume Analyser

Ms. Y. Sowjanya[1] | Mareddy Keerthana[2] | Pulluri Suneeksha [2] | Dorgipati Sai Sri Harsha[2]
*[1]Assistant Professor at Department of IT, Anurag Group of Institutions, Hyderabad*
*[2]Department of IT, Anurag Group of Institutions, Hyderabad*
*Corresponding Author Email Id: sowjanyait@cvsr.ac.in*

***Abstract-*** *The goal of resume screening is to identify the top applicants for a position and to inform users of their resume score and areas for improvement.The literature on existing approaches has been analyzed, and it has been discovered that the traditional systems like manual screening may result in false assumptions and the wasting of human potential, but they lack robustness in terms of processing, accuracyand efficiency. To acquire accurate results, software must use machine learning and natural language processing techniques to match and rate the candidates in real-time by ranking their resumes. The input would be the applicants resumes and output would be a ranked candidate's resumes list on the admin side and suggestions on the user side. Instantaneous real-time output results are acquired by employing natural language processing techniques. In the proposed system authors used Cosine Similarity, TF-IDF and Mong techniques of NLP for string matching. This system has the following benefits: security, interpretability, high accuracy, lightweight model, and quick processing. It could be utilized in Multi national companies, government organizations, and administrative agencies where numerous resumes must be reviewed daily for several openings. According to experimental findings, this system has a text parsing accuracy of 85% and a ranking accuracy of 92%.*

***Keywords—****Natural Language Processor, Term frequency inverse documentary frequency (TF-IDF), Cosine Similarity, Latent Dirichlet allocation (LDA), Exploratory data allocation.*

---

Date of Submission: 09-03-2023 Date of acceptance: 22-03-2023

---

## I. INTRODUCTION

Any recruiter will find it difficult to choose the best prospects from a vast pool of applicants for that employment vacancy. The chore of manually sorting through thousands of resumes to find the best candidates for the position is incredibly challenging for recruiters. Although the methods employed by job websites have produced some accuracy and precision, one of the main drawbacks is the intricacy of the time component. The time complexity for getting the results is very significant if every candidate resume is compared to every other job posting provided on the online recruitment site. In the last several years, more than 50,000 e-recruitment websites have been created. These online recruitment services creators have employed a variety of strategies to find potential applicants for a specific job profile. Some of them have been successful in using approaches for categorizing resumes of applicants into different groups for each job posting provided by each employer. These methods attempt to match each applicant's resume with the specific job posting. To determine which resumes are closest to the specified job description, top candidates could be sorted using Content-based Recommendation, cosine similarity, and KNN [1]. Although the methods employed by these websites have produced great levels of accuracy and precision, one drawback is the time required to find potential candidates for a given job description. Some of them have been successful in using approaches for categorizing resumes of applicants into different groups for each job posting provided by each employer. These strategies attempt to match each resume with the specific job posting. High accuracy and precision have been achieved by these sites' procedures, however one of the main drawbacks is the intricacy of the time component.

The strategy covered in this paper involves applying machine learning to prepare the dataset for a certain kind of work post. Additionally, the usage of section-based segmentation is suggested for data extraction utilising Natural Language Processing (NLP). This software uses Mong instead of generic processes for string matching, cosine similarity, overlap coefficient, and natural language [2]. The candidate's resume will only be matched to job openings that they are interested in and have applied to, which will decrease the time complexity and increase the web application's time efficiency.

Additionally, the recruiter for that specific organisation will be the only one to see the results of the resume matching of all applicants. They will have the choice to study the candidate's resume as well as receive results of the most qualified applicants who are already listed on the application website, according to this intelligent-based strategy.

## II. LITERATURE SURVEY

**Resume Classification and Ranking using KNN and Cosine Similarity:**The KNN Algorithm is used to categorise the resumes into the appropriate categories, and Cosine Similarity is used to determine how well the candidate's resume matches the job description. The resumes are then sorted in accordance with their classifications.

**Web Application for Screening Resume:** Semi-supervised learning is used by the suggested system, which is now being put into use, to attain high accuracy. The recruiter’s workload will be lighter because they won't have to manually go through the extensive pool of applicants' resumes.

**A Machine Learning approach for automation of Resume Recommendation system:**The screening and shortlisting processes may be greatly facilitated by an automated method of Resume Classification and Matching, and the candidate selection and decision-making processes would undoubtedly be sped up.

## III. RELATED WORK

**Existing system:** The existing system uses Naïve Bayes algorithm. This method uses traditional machine learning and has lower accuracy rates. It is less effective and inaccurate data. The potential of people may be lost due to this system. The recruiters cannot differentiate between the similarity of the resumes of the candidates as they might have taken a copy of someone else’s resume [4]. There are other approaches being tried, such KNN, however they are ineffective for huge datasets. Another popular approach is text parsing; however it can only be applied to resumes with structure. If the outcome is positive, the resume matched the job description, then the test of hypothesis gives a statistical representation by comparing the variable data to the specified data. Manually checking the resumes can be time-consuming for the recruiters [5]. Although semi-supervised learning based on machine learning can be utilised, it can only forecast the quality of the resume based on the job description.

**Proposed system:** Our approach uses, machine learning and natural language processing (NLP) techniques to evaluate resumes contextually. Artificial intelligence is employed along with these tools to go beyond keywords. Following resume screening, the software assesses candidates in real-time based on the employment requirements of the recruiter. This online application seeks to organize the resumes by comparing the resumes that best fits the specified Job Descriptions that is intelligently read as input. This software uses NLP for the instantaneous comparison and ranking of provided resumes [6]. In order to match candidates with job descriptions, we perform the extraction of abilities and relevant criteria from resume material, which is a different approach from other tasks. This application uses Mong instead of generic processes for string matching and cosine similarity for overlapping co-efficient. Job seekers will be able to submit their resumes and apply for any open positions via the interactive web application. The applicants can view the skills that they need to attain additionally in order to meet the recruiter's criteria. Moreover, the user is provided with sources that can acknowledge them regarding the required skills along with the resume building tips for increasing their resume score. The resumes can then be scored, and the scores can be sorted from best match to worst match. Only the firm recruiter who is interested in choosing the top prospects from a vast pool of applicants is given access to this ranking [7]. The recruiter can also view the insights like predicted field and the level of applicant from all resumes uploaded by using analytic statistical methods. The best candidates for that particular job vacancy can then be selected using the calculated ranks. By moving through all these mechanisms our method provides accurate results with greater efficiency, precision, and accuracy. This is done with the intention of saving recruiters at any organization time and effort from having to read through and evaluate hundreds of resumes.

**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 of considering 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. This is equivalent to the cosine, which has a value of unity when the segments subtend a zero angle and a value of zero when the segments are perpendicular. The attribute vectors A and B for text matching are typically the term frequency vectors of the documents. Equation 4 can be used to compare documents and use cosine similarity as a way to standardize document length.

$$\cos\theta = \frac{\overrightarrow{a.\vec{b}}}{|a||\vec{b}|} = \frac{\sum_{i-1}^{n} ai\ bi}{\sqrt{\sum_{i-1}^{n} ai^2 . \sum_{i-1}^{n} bi^2}} \quad (4)$$

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 Item2 represent the words that were translated into the vectors that were contained in the resume and the job description.

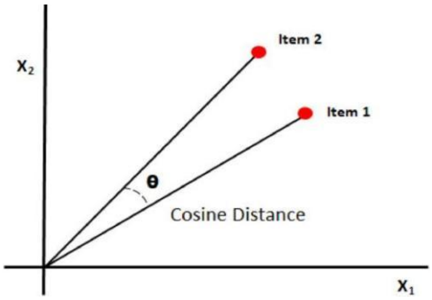


Fig 1: Cosine similarity

**NLP:**The JSON file is then sent through the NLP pipeline after the text annotations in the pdf file are converted to JSON format. This NLP pipeline is then used to train the model [13]. Using the NLP framework SpaCy, it can be trained. SpaCy is a framework that was developed for general data rather than for particular datasets, like a resume. In this method, rather than manually entering every word to create the dataset, semi-supervised learning is utilised to label the significant data in the ZIP file of PDF resumes.

After the extraction step, the resumes need to undergo pre-processing.
**Pre-processing:** In this process, the resumes sent as input would be narrowed down to remove any unusual or unnecessary characters. During cleaning, all distinct characters, numbers, and words with only one letter are removed. Following these procedures, our dataset was free of unique characters, numbers, and single-letter words [15].
The data is masked in the following ways:

1. Masking the strings like \w
2. Masking the escape letters such as \n
3. Masking each and every number
4. Substituting an empty string for all single-letter words

**Removing Stop Words:** Stop words like and, the, was, and others frequently exist in sentences and restrict the process of making predictions, thus they are eliminated. Each phrase corresponds to a stop word in NLTK after tokenizing the input words and filtering out the stop words. The processes are continued until the full resume has been parsed. When terms from the stop word list appear, they are cleared.

**Stemming:** Stemming, a technique for reducing word inflection to its basic forms, entails mapping a collection of words to a single stem, despite the stem itself not being a recognised word in the language. The stem (root) of the word is where inflectional (changing/deriving) affixes such (-ed, -ize, -s, -de, -ing, mis). For instance, words like “coding”, “codes” and “coded” will be mapped to “code”.

**Lemmatization:** Lemmatization ensures that the underlying word is correctly connected with the language by reducing derivative phrases [16]. The standard stages of lemmatization involve turning the corpus of text into a list of words and creating a corpus concordance that contains all of the word list entries in the corpus in the order in which they appear. Relate the word forms to their lemmas using the concordance.

Later, the essential libraries are being installed that help in the processing the upcoming tasks. After pre-processing the data chunking and named entity recognition are performed using the spacy module.

**Chunking:** By grouping brief phrases with parts of speech tags, the method of "chunking" seeks to give sentences more structure. Chunking combines parts of speech tags with regular expressions to provide a result as a set of chunk tags like Noun Phrase (NP), Verb Phrase (VP), etc. because parts of speech tags alone cannot provide information about the structure of the sentence or the real meaning of the text. It is also known as shallow parsing and entails building a parse tree with a maximum of one level of information from roots to leaves. This prevents the requirement for a full parse tree by ensuring that there is more information than simply the word's part of speech. This step gives a brief summarization of the entire resume into smaller parts and reduced time consumption.

**TF-IDF:** One of the factors used to calculate each word's final score is frequency. Without getting into the numbers, TF-IDF scores on word frequency try to highlight more intriguing phrases, like those that are prevalent 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. Without getting into the numbers, TF-IDF scores on word frequency try to highlight more intriguing phrases, like those that are prevalent inside a text but not across texts. This is used to find the predicted field, according to the frequency of skills. For instance, if the resume has many accomplishments in any field like iot then the user’s predicted field would be the same.

Later, we are using cosine similarity which gives information about how similar are two objects. In our case, it is used to compare the applicant resumes with the job descriptions. We have now retrieved the user's information. Now, we will recommend courses and certificates based on the user's current skills, just like if the user has machine learning skills, it will recommend free and paid courses and certificates.
A new module titled YouTube recommendations has been added. The two kinds of videos will now be recommended to the user by our system. One is for interview preparation, and the other is for resume preparation.

**Dataset:** Collecting the Resumes was the most important task in this process. Around fifty resumes were collected from Kaggle belonging to Java Developer and Project Manager Roles which were in doc and docx formats. The resumes collected were then converted into pdf format using bash script for easy handling of the data.

**Input:** Input on the user side is the job applicant resume which will be further pre-processed to remove any special characters or stop words. During cleaning, all distinct characters, numbers, and words with only one letter are removed. After all these steps we have a clean data collection. The login information must be entered on the admin page in order to access the admin page. The outcome of the uploaded resumes is available on the admin page.

**Output:** The user-side application shows the necessary skills that users' resumes lack. Even the links to the courses that would help applicants increase their knowledge are provided. Even on the user side, the resume is visible so that person can try to create a better resume. On the dashboard, there are also resume-building suggestions.
On the admin side, a highly ranked candidate's resume that most closely matches the job description is returned. Since they would be able to quickly review a huge number of resumes with the right fit, the technique would help the recruiter expedite profile shortlisting while also ensuring the validity of the shortlisting process, something a human would not be able to do in almost real-time [18]. The level of candidates for the requisite talent is acknowledged by a variety of statistical data that is given.

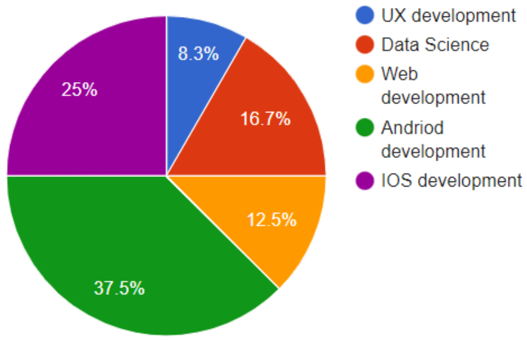


Fig 6: Pie chart for predicted field

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.

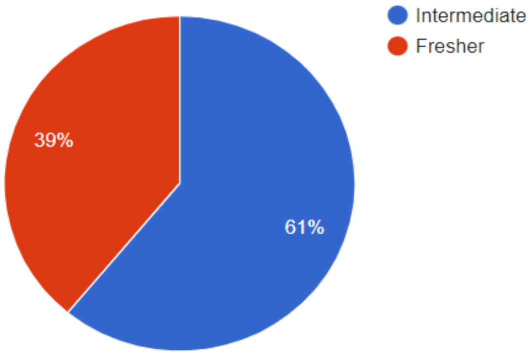


Fig 7: Pie chart for user experienced level

**Intelligent Recruitment System Using NLP:** The paper's main objective is to extract data from resumes and conduct the necessary analysis on the data to turn it into information that recruiters can use. This intelligent recruitment system uses Natural Language Processing. As a result, the Resume Parser would assist recruiters in quickly choosing the most qualified individuals, saving them both time and effort [3]. This can be efficient and useful for the company who are hiring the candidates.

**ResumeNet: A Learning-Based Framework for Automatic Resume Quality Assessment:** A neural-network model that incorporates text processing techniques is created in the proposed system to forecast each resume's quality. We suggest many modifications of the model to address the label deficiency problem in the dataset, either by using the pair/triplet-based loss or by introducing some semi-supervised learning strategy to take use of the large amount of unlabelled data. The proposed basic paradigm and its variations are both universal and simple to use.

**An Empirical Study of Artificial Intelligence and its Impact on Human Resource Functions:** In the Delhi/NCR region, 115 HR experts from various IT sectors participated in this study. The association between these two variables was found to be positive using the multiple regression method to test the hypothesis, proving that greater use of AI at work improves HR functional performance. However, there is a strong correlation between AI and both innovativeness and usability, which implies that AI has an impact on HR through both innovations and usability.

**A Hybrid Approach to Conceptual Classification and Ranking of Resumes and Their Corresponding Job Post:** What has been presented is a hybrid approach that uses conceptual-based classification of resumes and job postings and automatically ranks candidate resumes (that fall under each occupational category) to their corresponding job postings in order to address the issues with the previously highlighted techniques.

**Design and Development of Machine Learning based Resume Ranking System:** The model uses cosine similarity to find the resumes that are the most comparable to the job description supplied and the KNN algorithm is used to pick and rank the resumes based on job descriptions in huge quantities. KNN stores all the available data and classifies a new data point based on the similarity. The resumes are ranked according to the nearest similarity to the job description.

**Natural Language Processing based Jaro- The Interviewing Chatbot:** By presenting a chatbot that conducts interviews by evaluating the candidates resumes, the suggested system, JARO, expedites the interview process towards an objective decision-making process. The chatbot then develops a series of questions to be asked to the candidate. The technology will have functions like automated interviewing and resume analysis for analysing the resumes better.

## IV. VALGORITHMS

**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.

$$TF - IDF(t,d) = TF(t,d) * IDF(t,d) \qquad (1)$$

$$TF(t,d) = \frac{freq\ (t.d)}{\sum freq\ (ti,d)} \qquad (2)$$

$$IDF(t) = log(\frac{N}{count\ (t)}) \qquad (3)$$

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
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.

**KNN:**KNN is a non-parametric lazy learning method. A database of data points divided into various clusters is used to generate conclusions for new samples. KNN avoids making assumptions about the underlying data distribution in favor of focusing on item feature similarity. KNN determines the distance between the target resume and every other resume in its database while inferring information about a resume, ranks the distances, and presents the top K most comparable resume options. A database of data points divided into various clusters is used to generate conclusions for new samples. KNN avoids making assumptions about the underlying data distribution in favor of focusing on item feature similarity [11]. KNN can be used for different clusters which generate required results. Before computing the cosine similarity between the job description and resumes, the method unifies the cleaned resume data and job description into a single set of data.

## V. METHODOLOGY

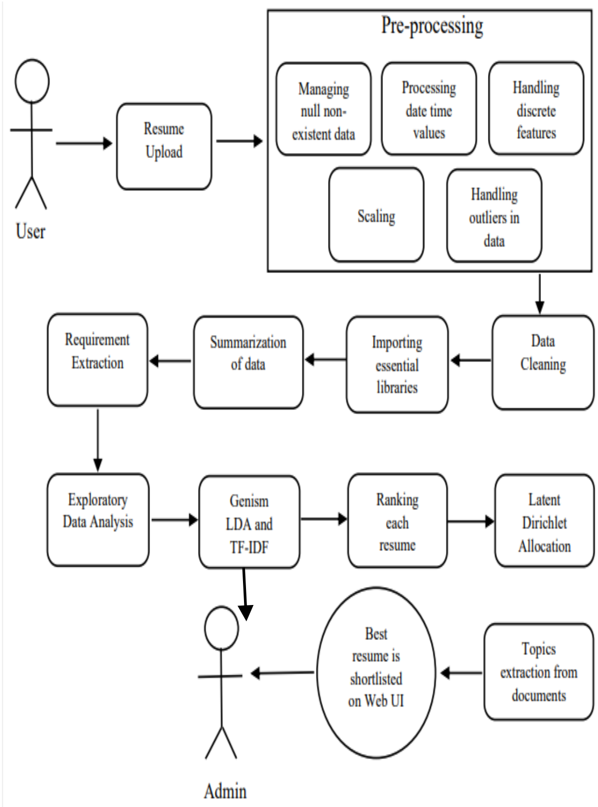


Fig 2: System Architecture

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.

**PDF Extraction**
In the first step of the process, the user uploads their resume in PDF format, further the PDF extraction is performed automatically. In this step the user’s data is abstracted from the resume by default. Text Extracting is the module in which it will fetch the text information from the resume. This text data will be used as language processing to further tasks like recommendations and fetching the user’s personal information.

**Named entity recognition:** Thistechnology that takes disorganised text chunks and groups them into specified categories, such as names of people, names of firms, contact information, educational background, and abilities, to extract the pertinent information. Our goal is to apply a similarity model to assess the similarity between the categorised resume data and the requirements provided by the recruiters after classifying the unstructured resume data into such diverse sets of categories. We c iv te our own algorithms that can extract the entities into required name. The details in the resume are classi ıd categorized under the name entities that will get saved into the database.

**Exploratory Data Analysis (EDA):** It is a method of data analysis that makes use of visual methods. With the use of statistical summaries and graphical representations, it is used to identify trends, patterns, or to verify assumptions. We are investigating a dataset to find patterns and anomalies (outliers) and developing hypotheses based on our knowledge of the dataset is the process of exploratory data analysis (EDA). In exploratory data analysis (EDA), data sets are analysed and trends are found by using statistics and visualisations. We are using this method to encourage data analysis before making any assumptions. It can assist in finding glaring errors, better understanding data patterns, spotting outliers or unusual occurrences, and discovering intriguing relationships between the variables.

**Gensim LDA:** Latent Dirichlet Allocation (LDA), a tool and method for topic modelling, classifies or categorises the text in a document and the words per subject using models based on Dirichlet distributions and processes. The application uses Latent Dirichlet Allocation to find hidden themes in the data, classify the data according to the themes found, and then use the classification to arrange/summarize/search the documents [9]. The application then deals with determining a candidate's resume score. A rank list will be created based on the scores each candidate's resume earns, with the candidate with the higher score being listed higher than the candidate with the lower score.By making it simple for recruiters to find resumes by organising the list in order of relevance to the job.

**KNN:** The CVs that are closest to the job description are found using KNN. To begin, we scaled the job descriptions and resumes with an open-source program called "gensim." Within the allotted number of words, this library produces a summary of the text provided. In this model, KNN is used to find resumes that are the most similar to the job description supplied or resumes that are a close match to the job description. This library was used to create a summary of the job descriptions and resumes to bring them to a similar word scale. KNN was then used to find CVs that closely matched the job descriptions.

The next phases are the recommendations, where the prediction of field, career recommendation, course recommendation, YouTube recommendations and data analytics are done.

## VI. EXPERIMENTAL RESULTS

There are two modules in smart resume analyser, user a v lmin while each having its own special functions.

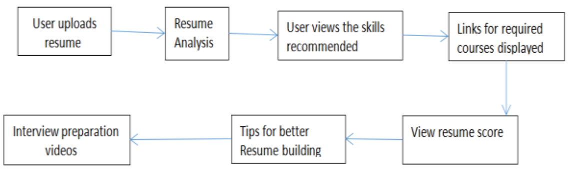


Fig 3: User work flow

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 are additionally given. These sources can be monitored by the admin which can be modified according to their requisite. The tips that can increment the resume rank are stipulated along with the score.

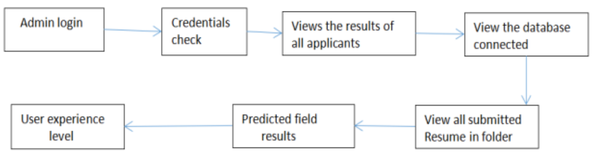


Fig 4: Admin work flow

The necessary credentials must be supplied in order to access admin module's features. Admin can view report of all the applicant resumes , which includes applicant name, ID, and email address. Additionally, the predicted field and experience level are listed, which are computed by the skills and score. Along with these recommended courses, recommended skills, and actual talents are all provided. Each resume has a time stamp, and details get automatically uploaded on to the connected database, while the resumes get saved to the destination folder.

| ID | Name | Email_ID | resume_score | Timestamp | Page |
|---|---|---|---|---|---|
| 1 | 0135New York | info@resumekraft.com+1-202-555-0135New | 60 | 2023-02-15_12:51:41 | 2 |
| 2 | art director | hello@allisonbeer.com | 20 | 2023-02-15_12:56:20 | 1 |
| 3 | 0135New York | info@resumekraft.com+1-202-555-0135New | 60 | 2023-02-15_13:02:10 | 2 |
| 4 | 0135New York | info@resumekraft.com+1-202-555-0135New | 60 | 2023-02-15_13:02:58 | 2 |
| 5 | 0135New York | info@resumekraft.com+1-202-555-0135New | 60 | 2023-02-15_16:04:27 | 2 |
| 6 | art director | hello@allisonbeer.com | 20 | 2023-02-15_16:32:26 | 1 |
| 7 | art director | hello@allisonbeer.com | 20 | 2023-02-15_16:32:39 | 1 |
| 8 | 0135New York | info@resumekraft.com+1-202-555-0135New | 60 | 2023-02-15_16:34:58 | 2 |
| 9 | Android Developer | info@qwikresume.com | 40 | 2023-02-15_16:39:34 | 2 |
| 10 | Android Developer | info@qwikresume.com | 40 | 2023-02-15_16:40:53 | 2 |
| 11 | Android Developer | info@qwikresume.com | 40 | 2023-02-15_16:41:26 | 2 |
| 12 | 0135New York | info@resumekraft.com+1-202-555-0135New | 60 | 2023-02-15_16:42:26 | 2 |
| 13 | art director | hello@allisonbeer.com | 20 | 2023-02-15_16:42:57 | 1 |
| 14 | 0135New York | info@resumekraft.com+1-202-555-0135New | 60 | 2023-02-15_16:48:00 | 2 |

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.

## VII. FUTURE ENHANCEMENTS

I. Future development for this approach entails mining applicant social networking data (such as their Facebook, LinkedIn, and GitHub pages) and using this social beh vi data in conjunction with the content of resumes to produce even better suggestions.

II. Using a collaborative filtering-based technique is another option, which can match the present applicant with a job based on how well other candidates who are similar to them (neighbours) are rated for it.

III. As a part of our ongoing research, we intend to leverage the information gleaned from resumes of applicants to dynamically create user profiles that will be used to suggest jobs to job seekers.

IV. Involving subject matter experts, such as HR professionals, will aid in the development of a more accurate model, and the HR professionals' feedback will aid in iteratively improving the model.

V. The use of latent semantic analysis (Berry, M., 2001) in calculating the semantic similarity of the documents and comparing it with the outcomes of the term frequency-based similarity technique is another area of future development.

VI. The shortlisted applicant’s personalities can be examined utilising the social media data included in their resumes. This evaluation will help determine whether the candidate's personality, as seen in his or her social life, aligns with the demands of the position.

## VIII. CONCLUSION

In this paper, we investigated the crucial yet understudied issue of automatic resume quality assessment (RQA). The classification of the applicant's resume is a laborious, time-consuming, and resource-wasting process. We have developed a machine learning-based automated algorithm to address this problem by recommending HR the resumes of qualified candidates based on the provided job description. With the use of Natural Language Processing techniques our algorithm was able to screen and shortlist the most qualified candidates [20].When displaying the top-selected resume on Web user interface, Latent Dirichlet Allocation produced incredibly precise results. By using job descriptions as input the best-fit resume can be selected by matching the skills in the resume of the applicant. Along with this, various visualization and data analytics are provided on the admin side. The linear classifiers are used to provide a statistical model of skill similarity. The user is provided with a proper guidance about the skills they need to upgrade and the sources for getting acknowledged. They are also equipped with essentials tips that can enhance their resume. This indeed, helps the user to know their potential and improve their proficiency in the fields that the company is looking for in them.

## REFERENCES

[1]. Rajath V; Riza Tanaz Fareed; Sharadadevi Kaganurmath, “Resume Classification and Ranking Using KNN And Cosine Similarity”, international JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY, Volume 10, Issue 08, AUGUST 2021

[2]. Akshay Kulkarni; Adarsha Shivananda; Anoosh Kulkarni, “Creating a resume Parsing, Screening and Shortlisting System”, Natural Language Processing Projects pp 125-155, December 2021

[3]. Sujit Amin; Nikita Jayakar; Sonia Sunny; Pheba Babu; M. Kiruthika; Ambarish Gurjar, “Web Application for Screening Resume”, 2019 International Conference on Nascent Technologies in Engineering (ICNTE), January 2020

[4]. Jitendra Purohit; Aditya Bagwe; Rishabh Mehta; Ojaswini Mangaonkar; Elizabeth George, “Natural Language Processing based Jaro-The Interviewing Chatbot”, 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), August 2019

[5]. Tejaswini K; Umadevi V; Shashank M Kadiwal; Sanjay Revanna,” Design and Development of Machine Learning based Resume Ranking System”, Global Transitions Proceedings, October 2021

[6]. 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.

[7]. Suhjit Amin, Fr. Conceicao Rodrigues Institute of Technology, 2019.Web Application for Screening resume, IEEE

[8]. Tejaswini K, Umadevi V, Shashank M Kadiwal, Sanjay Revanna, Design and Development of Machine Learning based Resume Ranking System (2021).

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

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

[11]. Centre for Monitoring Indian Economy Pvt Ltd. (CMIE),2022. The unemployment rate in India.

[12]. Jabri, Siham, Azzeddine Dahbi, Taoufiq Gadi, and Abdelhak Bassir. "Ranking of text documents using TF-IDF weighting and association rules mining." In 2018 4thinternational conference on optimization and applications (ICOA), pp. 1-6. IEEE, 2018.

[13]. Jagan Mohan Reddy D, Sirisha Regella., “Recruitment Prediction using Machine Learning”, IEEE Xplore, 2020.

[14]. 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].

[15]. Maheshwary, S., Misra, H., 2018. Matching resumes to jobs via deep Siamese network, in: Companion Proceedings of The Web Conference 2018, International World Wide Web Conferences Steering Committee. pp. 87–88.

[16]. Suhas H E, Manjunath AE, “Differential Hiring using a Combination of NER and Word Embedding”, In 2020 International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Vol.9.

[17]. Y. Luo, Y. Wen, T. Liu, and D. Tao, “Transferring knowledge fragments for Learning distance metric from a heterogeneous domain,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018.

[18]. Adopting a Strategic Approach to Matching People to Jobs Chalette Renee Griffin International Journal of Business and Management, 2018.

[19]. Dr. K. Satheesh, A. Jahnavi, L Aishwarya, K. Ayesha, G Bhanu Shekhar, K. Hanisha, “Resume Ranking based on Job Description using SpaCy NER model”. (2020). International Research Journal of Engineering and Technology.

[20]. Arvind Kumar Sinha, Ashwani Kumar, Md. Amir Khusru Akhtar., Resume Screening using Natural Language Processing and Machine Learning a Systematic Review”, (2019)., Machine Learning and Information Processing: Proceedings of ICMLIP.