alphanumeric
https://alphanumericjournal.com/
Volume 12
Issue 02
Year 2024

Research Article

Resume Screening with Natural Language Processing (NLP)

Mehtap Saatçı [1], Rukiye Kaya [1] & Ramazan Ünlü [1]

[1] Faculty of Engineering, Abdullah Gül University, Kayseri, Türkiye

Abstract

This study addresses the difficulties employers face in screening the large number of resumes received for job positions. We aim to ensure fair evaluation of candidates, reduce bias, and increase the efficiency of the candidate evaluation process by automating the resume screening process. The proposed system uses NLP techniques to extract the relevant competencies from the resumes, focusing on the key skills required for specific positions. The competency sets taken for the positions were used. A case study was conducted for 123 job positions. Jaccard Similarity and Cosine Similarity measures were evaluated for the purposes of the study. Due to the fact that Cosine Similarity focuses on word frequency, Jaccard Similarity measure generates results more aligned with the purposes of the study. The extracted competencies are matched to predefined skill sets associated with various job positions using Jaccard Similarity. This approach assigns a similarity score to rank candidates by analyzing the presence or absence of specific words in their resumes in relation to the required competencies. This NLP-based system offers significant benefits such as saving time and other resources, increasing accuracy in candidate selection, and reducing bias by focusing only on competencies. The system's integration with LinkedIn enhances the effectiveness of the approach by facilitating seamless importation and analysis of resumes. Overall, this study demonstrates the potential of NLP in optimizing the resume screening process by providing a scalable, efficient, and unbiased solution for large organizations.

Keywords Natural Language Processing (NLP), Resume Screening, Jaccard Similarity, Cosine Similarity, Candidate Evaluation

Jel Codes C8, J21, M15

Contents

2024.12.02.MIS.04

Correspondence

R. Kaya rukiye.kaya@agu.edu.tr

Timeline

Submitted	Aug 21, 2024
Revision Requested	Oct 06, 2024
Last Revision Received	Oct 30, 2024
Accepted	Dec 23, 2024
Published	Dec 31, 2024

Citation

Saatçı, M., Kaya, R. & Ünlü, R. (2024). Resume Screening with Natural Language Processing (NLP), *alphanumeric*, 12 (2), 121-140. https://doi.org/10.17093/alphanumeric.1536577

1. Introduction

The recruiters post their requirements for the job positions they need to fill, and many applicants apply by uploading their resumes. Screening candidates' resumes manually consume a lot of time and energy hence causing huge hiring expenses. In addition to this, due to time limitations, human error, and bias, applicants cannot be evaluated accurately by a manual screening of resumes. These limitations cause missing out the right candidates or selecting the wrong applicants for the positions. Gan et al. (2024) emphasize the significance of automation in enhancing process efficiency and effectiveness of recruitment process. The authors accelerate the process by automation 11 times faster than the manual resume screening. On the other hand, FraiJ & László (2021) review the resume screening studies based on AI and present the advantageous outcomes of using AI in recruitment process. In this paper, we offer an effective Natural Language Processing (NLP)-based resume screening system to solve the issues recruiters face while screening the resumes. Pimpalkar et al. (2023) examine the implementation of NLP and ML approaches for automating the resume screening process and evaluate their effectiveness in comparison to existing methods. In the literature, NLP is widely adopted for resume screening process (Aydin et al., 2024; Bharadwaj et al., 2022; Suhas Chavare & Bhaskar Patil, 2023). Alamelu et al. (2021) use NLP to sort the best fit resumes and save time for the manual review of the resumes by 85%. In this study, we propose a system reviews resumes obtained via LinkedIn and other sources automatically. Bharadwaj et al. (2022) and Mehboob et al. (2022) also use NLP to review resumes from LinkedIn and GitHub profiles. We aim to ensure fair evaluation of candidates, reduce bias, and increase the efficiency of the candidate evaluation process by automating the resume screening process. The proposed system extracts the relevant competencies and qualifications from candidates' resumes, focusing on the key skills required for specific positions. Then it assesses the suitability of the candidate for the job position. It uses NLP techniques. Preprocessing steps include the removal of stopped words, lemmatization to reduce words to their basic forms, and tokenization to divide the text into manageable units. These steps ensure that the information extracted is accurate and standardized, facilitating effective comparison of candidate skills. In this study, the competency sets taken for the positions were used. A total of 123 positions were included. The extracted competencies are matched to predefined skill sets associated with various job positions using Jaccard Similarity.

In the literature, the studies employing NLP utilized various similarity measurements. Daryani et al. (2020) automatize the resume screening system by applying NLP with Vector Space Model and Cosine similarity measures. On the other hand, Aminu et al. (2023) utilize Cosine and Jaccard similarities with NLP. Cabrera-Diego et al. (2015) compare the performances of three similarity measures; Cosine Similarity, Dice Coefficient, and Jaccard Index, and two distance measures; Euclidean and Manhattan distances. They conclude that Jaccard Similarity and Dice Coefficient provide the best results among the other methods. Jaccard Similarity method provides a similarity score that helps rank candidates by comparing the presence or absence of words in the candidate's resume to the required competencies. We test both Jaccard Similarity and Cosine similarity in this study. However, considering the fact that Cosine similarity concentrates on word frequency and Jaccard Similarity better addresses the aims of the study in comparison to Cosine Similarity, we adopt Jaccard Similarity as similarity measure. This study involves the development of a user-friendly interface for HR staff. The system processes these resumes, extracts the relevant competencies, and ranks the candidates according to their suitability for the position. Results, including detailed similarity scores are displayed and can be saved for further analysis. The NLP-based automated resume screening system significantly saves the cost, time, and effort required for manual evaluations, provides more effective and unbiased candidate assessment process.

2. Literature Review

Resume screening is an essential part of the hiring process for every organization. It is time-consuming process and the effectiveness of the process is a big challenge, particularly for a large pool of applicants. The authors introduce an innovative framework to automate the resume screening, leveraging Large Language Models (LLMs) to enhance efficiency and effectiveness in recruitment processes. The framework employs LLM-based agents to summarize and grade resumes, optimizing decision-making such as job offers or interview selections. Evaluation using a constructed dataset shows the proposed framework is 11 times faster than manual methods. Fine-tuning of LLMs improves the F1 score to 87.73% for resume sentence classification and surpasses baseline performance in resume summarization and grading tasks (Gan et al., 2024). Aydin et al. (2024) examine the use of artificial intelligence (AI), virtual reality (VR), augmented reality (AR), and Metaverse technologies in human resources (HR) processes. A survey was used to collect the opinions of the participants in the study about the use of AI, VR, AR, and Metaverse technologies in HR processes. Participants give information about how these technologies can be used in HR management and their advantages and disadvantages. Chi-square analysis was used to analyze the relationship between participants' demographic characteristics (age, gender, professional experience, education level) and their opinions about these technologies. Findings show that 84.8% of the participants think that AI will be useful in HR processes, 65.8% think that VR will be advantageous in HR processes, 65.8% think that AR will be advantageous in HR processes, and 53.8% think that Metaverse will be useful in HR processes. In conclusion, this study revealed the potential of the researched technologies to increase efficiency in HR processes, improve recruitment processes, support training and development, and facilitate performance management. However, significant disadvantages such as ethical issues, data security risks, and acceptance difficulties have also been noted. It has been stated that future research should be carried out to eliminate these disadvantages.

FraiJ & László (2021) conducted a systematic literature review to present the impact of Artificial Intelligence (AI) on the recruitment process. In the study, related studies published from 2010 to 2020 were examined. It has been stated that studies in this field have increased in the last 5 years and focused on effective screening, human bias, and best candidate fit. It is stated that the use of artificial intelligence in the field of HRM increases the quality of recruitment processes saves time and effort for applicant screening and increases the quality of the output of hiring and provides unbiased selections.

Pal et al. (2022) focus on using Machine Learning algorithms like Naive Bayes, Random Forest, and SVM for classifying resumes, enabling the extraction of skills and presenting diverse capabilities aligned with specific job profiles. The study finds that Random Forest achieves the highest accuracy among the other algorithms. Furthermore, the research suggests extending these capabilities by integrating other features such as facial recognition, voice-to-text generation, and voice analysis into video interviewing processes. On the other hand, Lad et al. (2022) present a machine learning (ML) based CV recommendation system use methods include text feature extraction techniques such as TF-IDF and K-Nearest Neighbors (KNN) machine learning algorithms. In addition, the system is designed to be web-based and offers employers the opportunity to sort and review CVs according to job description through a user-friendly interface. It has been reported that future studies can be expanded by adding advanced features such as database integration and multiple file selection in the system. The utilization of the K-Nearest Neighbors (KNN) method for resume screening is widely recognized in the literature. Tejaswini et al. (2022) propose to match the CVs with job descriptions by using techniques like cosine similarity and the K-Nearest Neighbors (KNN) algorithm. Experimental results show promising performance with an average text parsing accuracy of 85% and a ranking accuracy of 92%. This automated approach not only saves time but also enhances the consistency and objectivity of candidate evaluation. Additionally, Anand & Dubey (2022) utilize semi-supervised learning, primarily K-nearest Neighbors, to achieve high accuracy. This approach automates the process of specifying requirements and ranking applicants, aiming to closely match the judgements of human experts. It promises to streamline the shortlisting of CVs from a large pool of applicants, ensuring consistency and effectiveness in candidate selection. Furthermore, the literature on resume screening also includes applications of deep learning. Li et al. (2023) focus on deep learning methods used in automated resume screening, specifically how word embeddings (numerical vectors representing words) can inadvertently reinforce biases from the training data. The study uses a publicly available dataset of more than 1,000 CVs collected from popular job posting websites in Singapore. 105 resumes from China, India, and Malaysia were selected for study with automatic and manual checks. The study reveals that word embedding-based job-resume matching algorithms may carry national origin bias, and this bias can cause major problems in the workforce. The paper presents several improved algorithms to solve this problem and conducts extensive experiments to evaluate the performance of the proposed algorithms in terms of fairness and accuracy. To reduce the effect of biased terms, p-ratio, and sigmoid-adjusted p-ratio methods were used. These methods aim to reduce bias by adjusting the weights of word embedding vectors. In conclusion, this article makes a significant contribution to reducing bias in AI-supported recruitment processes. The results show that the algorithms have great performance in terms of fairness measure and accuracy. It has been stated that future studies on this subject will increase, that it is necessary to eliminate natural bias instead of adjusting the weights of the terms, and that bias should also be investigated at other recruitment stages. In a related context, Spoorthi et al. (2023) suggest using an ensemble deep-learning model for the classification of resumes.

On the other hand, Chou & Yu (2020) present a machine-learning and text-mining-based technology to match the job vacancies and the job applicants. The proposed system recommends a talent list for the job positions. Another widely recognized approach in the literature is Named Entity Recognition (NER). Sajid et al. (2022) presents a resume parsing framework that aims to extract information from resumes with different text formats and layouts. The work includes three basic processing steps such as text extraction, text block classification and named entity recognition (NER). The extracted skills were then enriched using a specially developed ontology. As a result, the study focuses on more than one area to ensure diversity, and the most important contribution is the development of a special skill ontology for data enrichment.

Mohanty et al. (2023) also use The NER (Named Entity Recognition) method to extract information from CVs. Afterward, the CVs were trained with machine learning models (KNN, SVM, Decision Tree, Naive Bayes and XGBoost) and models were tested with cross-validation. As a result, it was stated that the XGBoost model was more successful than other models in the resume classification task with high accuracy and low standard deviation. As an alternative study, Li et al. (2020) present a novel resume classification approach for this purpose. They analyze 6,492 resumes from 24,933 job applications across 252 Clinical Research Coordinator (CRC) positions categorized into four experience levels. Each resume is meticulously annotated by experts to determine its most suitable CRC level, achieving a high inter-annotator agreement Kappa score of 61%. Using this annotated dataset, the study develops novel transformer-based classification models for two tasks: the first task classifies resumes into CRC levels (T1), and the second task assesses whether a resume matches a given job description (T2). The best-performing models, incorporating section encoding and multi-head attention decoding techniques, achieve accuracies of 73.3% for T1 and 79.2% for T2. Analysis indicates that prediction errors primarily occur between closely adjacent CRC levels, underscoring the practical utility of the models in real-world HR platforms. The literature presents numerous examples of the application of NLP and ML approaches for resume screening. Pimpalkar et al. (2023) discuss the use of Natural Language Processing (NLP) and Machine Learning (ML) technologies for automating resume analysis. The article also includes a comparison of existing analysis tools. As a result, it is emphasized that ML and NLP technologies can increase efficiency in resume analysis and improve the candidate selection processes. This study also addresses the research issues related to writing style, word choice, and grammar of unstructured written communication and the future potential of video resume analysis. Roy et al. (2020) utilize Content Based Recommendation using Cosine Similarity and k-Nearest Neighbours for resume recommendation system. However the proposed approaches yield poor accuracy. Kino et al. (2017) adopt Co-occurrence in text analysis for job matching but they provide results depending on limited key words. On the other hand, Ali et al. (2022) explore a range of ML algorithms and NLP methods to assess the effectiveness of the RCS, aiming to offer a solution that achieves superior accuracy and reliability across various scenarios. To highlight the effectiveness of NLP and ML in RCS, the study evaluated features using nine different ML classification models, including Support Vector Machine (SVM) variants (Linear, SGD, SVC, and NuSVC), Naïve Bayes models (Bernoulli, Multinomial, and Gaussian), K-Nearest Neighbor (KNN), and Logistic Regression (LR). The Term-Frequency-Inverse-Document-Frequency (TF-IDF) feature representation was found to be effective for the RCS. The models were assessed through metrics such as the Confusion Matrix, F-Score, Recall, Precision, and overall accuracy. The results showed that the SVM class of classifiers when using a One-Vs-Rest-Classification strategy for this multi-class resume classification task, achieved over 96% accuracy on a dataset of more than 960 resumes. These encouraging findings suggest that the NLP and ML methods applied in this study could be used to develop an effective RCS.

A comprehensive review of the literature demonstrates that Natural Language Processing (NLP) is the most widely adopted approach for resume screening, with the largest number of studies addressing this problem relative to other approaches. Pant et al. (2022) indicates that NLP presents significant benefits with text preprocessing, keyword extraction and ranking abilities among the other methods. The authors focus on using NLP techniques, specifically resume label character positioning, to extract technical skills. The approach involves utilizing a dataset of software engineering job requirements, employing regular expressions and word matching to retrieve candidate information. A novel technique called character positioning is introduced for extracting essential data from resumes. This methodology generates resume summaries based on the extracted information and calculates scores for recognized skills, education, and experience levels. In testing with five randomly selected software engineering resumes, an extraction accuracy rate of 33.59% was achieved. Daryani et al. (2020) developed an automatic resume screening system based on Natural

Language Processing (NLP). They extract relevant information from resumes and then match this information to job descriptions to recommend the most suitable candidates. Vector Space Model and Cosine similarity were used for the fitness rate in the study. In future studies, it is recommended to use data from social networks to increase the accuracy of the system and use collaborative filters for detailed evaluation. Similarly, Trinh & Dang (2021) present an automatic resume process tool based on NLP. The NER method was used in the information extraction phase and the fastText method was used in the embedding and classification phase. Finally, personal information extraction weights are added to calculate the final scores by measuring the similarity of the resumes with the job descriptions. Analysis of the data shows the compatibility of the resumes with the job descriptions, and visualizations and statistical analysis are presented to demonstrate the success of the study in practice. This practice is expected to enable recruiters to reduce their workload and achieve an efficient staffing process. The most important limitation of the study is that it remains at the experimental analysis level. It is recommended to improve data sets and reduce manual tasks during data preparation in future studies. On the other hand, Aminu et al. (2023) propose a two-staged automated recruitment system based on NLP. In the first stage, CVs are classified using the Support Vector Classifier (SVC) method. In the second stage, the person's resume is analyzed based on its similarity to the job description by Cosine and Jaccard Similarities. On the other hand, Lalitha et al. (2023) use NLP for the benefit of both applicants and employers. They simplified the job application process calculated the suitability of the applicant for the position by cosine similarity and presented the similarity percentage to the applicants. As a similar NLP approach for resume screening, Harsha et al. (2022) develop a resume filtering application using Python programming language and Natural Language Processing (NLP) techniques. The application has a modular structure, supports a wide variety of file formats, and was created in accordance with the Model View Controller (MVC) design pattern. Mehboob et al. (2022) develop a CV shortlisting tool. They conduct experiments on 222 CVs collected from LinkedIn profiles, an average similarity was obtained between 40.54% and 89.40% using the cosine similarity measurement. They propose to recommend the most suitable job titles according to the candidates' profiles in future research.

Alamelu et al. (2021) use NLP to filter the resumes that best fit to the job description. The proposed approach aims to reduce the time spent on reviewing resumes by 85% and can process more than 30 resumes per minute.

Satheesh et al. (2020) also propose to automate the resume screening process using advanced Natural Language Processing (NLP). By leveraging the Spacy Named Entity Recognition (NER) model, the system automatically extracts essential entities from resumes. It then generates a graphical representation displaying each resume's score. SpaCy, an open-source Python library for NLP, forms the foundation of the approach.

Bharadwaj et al. (2022) utilize a natural language processing (NLP) toolkit to achieve high accuracy in resume analysis. Future upgrades aim to automate 90% of the resume selection process by extracting candidate skills and work experience from LinkedIn and GitHub profiles. Recent literature includes numerous additional examples of the application of Natural Language Processing (NLP) approaches for resume screening (Harsha et al., 2022; Naveed et al., 2024; Suhas Chavare & Bhaskar Patil, 2023). As noted in the literature, numerous similarity measures are utilized within Natural Language Processing (NLP) approaches for resume screening. Cabrera-Diego et al. (2015) behave to determine the similarity of resumes in e-Recruitment systems. They aim to determine whether selected CVs are more similar to each other than the rejected CVs. A comparative analysis of three similarity measures; Cosine Similarity, Dice Coefficient, and Jaccard Index, and two distance measures; Euclidean and Manhattan distances are presented in this study. Jaccard Similarity and Dice Coefficient have emerged as the methods that give the best results and it is believed that there should be a specific vocabulary that will allow us to easily identify the CVs of candidates selected or not selected by the recruiter. They point out the use of different languages and methods in future research.

3. Resume Screening with Natural Language Processing (NLP)

Resume screening involves reviewing resumes to select the most suitable candidate for a position. Employers need to review thousands of resumes for a single position, and this is a process that consumes a considerable amount of time. Resume screening systems can be automated with Natural Language Processing (NLP) processes. Facilitated by NLP, the desired information can be extracted from resumes in a short time. Additionally, by implementing the NLP system, difficulties in the resume screening system can be avoided. Resumes prepared in various formats can be reviewed efficiently in a short time. Information such as the person's name, gender, and age can be examined in a way that is ignored by the system. In this way, a bias-free system is developed. Thanks to the NLP system, both time and resources can be saved in resume screening. The purpose of using NLP in this project is to automatically scan the resumes coming through the company's application channel for a specific position and rank the candidates starting from the most suitable one based on the position. By extracting relevant data from resumes through NLP, the candidate's suitability for the position is determined. In this project, the information extracted from resumes is the candidate's skills. Extracted skills relate to predefined skill sets associated with various job positions. In this way, candidate screening for positions can be accelerated and efficiency can be increased.

3.1. Dataset

The dataset used in the study consists of competency sets defined for positions. This dataset includes technical and competency-based skills required for each of 123 different positions in total. Each position has different competencies depending on specific job descriptions and job requirements. For example, the skills required for a "Systems Expert" position are generally technical skills such as "Field Engineering" and "Network Security".

The competencies for each position in the dataset consist of specific keywords. For example, while the "Systems Expert" position seeks skills such as "Network Security", "Microsoft Exchange Server" and "Server Hardware Architectures", the "Education Management Specialist" position considers skills such as "Machine Learning", "Python" and "Analysis" as critical. In this way, the skills sought for each position reflect both technical depth and application area.

The dataset also includes the frequency of competencies and skills that are repeated across multiple positions. For example, general usage skills such as "MS Office" are common competencies sought in more than one position. Thus, a wider competency pool was created by considering both position-specific and general skills.

As an example, the competencies required for three different positions are given in Table 1, Table 2 and Table 3.

Table 1. Required Competencies for Education Management Specialist

Education Management Specialist			
Active Directory and Core Services	Requirements Analysis	Modelling	SQL
Network Security	Visual Topics	Nagios	Swift
Analytical Thinking	Image Processing	Network Security	Teamwork
Analysis	IOS	Object-Oriented	Basic Software Engineering
Angular	Java	Objectives	Artificial Intelligence
C++	Kotlin	OS Security	Zabbix
Cluster Services	Linux	Game Engines Development	Simulation Software
Deep Learning Methods	Machine Learning	Platform	Mobile Device Applications
ELK	Microservice Architecture	Python	Integration

Table 2. Required Competencies for System Expert

System Expert	
Active Directory and Core Services	Server Hardware Architectures
Microsoft Exchange Server	Component Library
Network Security	Redux
Field Engineering	

Table 3. Required Competencies for Database Engineer

Database Engineer		
Ansible	MS SQL	RDBMS
Artificial Intelligence	MySQL	SQL
Big Data	NoSQL	SQL Server
Data Engineer	ORACLE	Data Modeling
Data Marts	POSTGRESQL	Database
Data Mining	PROGRAMMING	Database Infrastructure
Machine Learning	Progress	Database Architecture
Microsoft SQL Server	Python	Database Management

Another dataset used in the study consists of a total of 16 resumes (CVs). These resumes were collected from different industries and professional fields, and each CV contains the candidate's educational information, work experience, certifications, and other important information. The dataset was used to test the performance of the NLP-based automatic screening system, which specifically evaluates whether the candidates meet the qualifications and position requirements. The resumes in this dataset are collected from various sectors such as information technology (IT), marketing, finance, and engineering. This diversity provides an important test bed to see how the NLP-based system evaluates candidates from different sectors and to measure its performance. These resumes from different fields increase the general applicability of the system when evaluating candidates' sector-specific skills and competencies. The table below shows the position titles and sectors of the individuals included in the CV used in this study.

Table 4. Position Title and Sectors of Collected CVs

CV PDF Name	Position Title	Sector
A	Industrial Engineer	Manufacturing/Engineering
B	Industrial Engineer	Manufacturing/Engineering
C	Planning Specialist	Supply Chain/Logistics
D	Mechanical Engineer	Manufacturing/Engineering
E	Marketing Expert	Marketing/Advertising
F	Electrical Engineer	Electrical/Energy
G	Java Developer	Information Technology (IT)
H	Scrum Master	Information Technology (IT)
I	Data Analyst	Information Technology (IT)/Finance
J	Accounting Manager	Finance/Accounting
K	Finance Expert	Finance/Banking
L	Product Manager	Information Technology (IT)/Product Development
M	Database Engineer	Information Technology (IT)
N	Customer Service Specialist	Customer Service
O	Business Intelligence Analyst	Information Technology (IT)
P	Software Developer	Information Technology (IT)

Resumes are provided in PDF format. Some resumes are long and detailed, while others are more concise. This structural diversity allows the system to process candidate information correctly. Resumes also vary greatly in format, language use, and content, which is an important metric for understanding how well an NLP system performs even in complex situations. This dataset provides a test environment for evaluating and improving the performance of the system by reflecting the diversity of real-world resumes. It is especially useful for seeing how NLP-based screening and assessment systems process candidates from different industries and experience levels. Although the dataset includes a variety of fields, its small size (16 CVs) may limit the generalizability of findings. Future studies could expand the dataset to include more candidates from a broader range of industries to improve the robustness of the results.

3.2. Data Preprocessing

Received resumes should be suitable for automated resume screening process. To do this, we have implemented various preprocessing steps in order to extract competencies. In this section, we have briefly explained the details of those steps. Resumes received for resume screening must be made suitable for processing. Competencies are extracted from pre-processed resumes. The following steps were followed to pre-process the CVs.

3.2.1. Removing Stop Words

Stop words are common words that occur frequently in a language, but often carry little or no meaning. Examples of stop words in English include "the", "I", " in", etc. takes place. The words "and", "to" and "of" in the competencies were not removed in order not to miss the competency. Stop words appear frequently in texts, but they do not provide useful information. Stop words have been removed from resumes to reduce data size and focus on more meaningful words.

3.2.2. Lemmatization

Lemmatization aims to reduce words to their basic form. For example, it is the process of converting words to their original form or removing suffixes such as "-ing, -(e) s". In this project, the use of lemmatization, which takes into account the context and morphology of the word, was preferred instead of stemming. Thanks to this process, the standardization of words, and correct text analysis are ensured.

3.2.3. Tokenization

Tokenization is the process of breaking a text into smaller units. In this process, the text is divided into words or phrases. This process was done in order to manage the text more easily and to process and extract information. Since there are competencies that contain more than one word, a tokenization process that checks consecutive words has been carried out. In this way, competencies were effectively extracted.

3.2.4. Measurement of Similarity

At this stage, a similarity measurement selection was carried out to compare the similarity of the competencies extracted and determined for the position. Jaccard Similarity and Cosine Similarity measures, which are popularly used in text analysis, were tested and the results of both similarity measures were compared. It was observed that the similarities obtained as a result of Cosine Similarity, which focuses on word frequency, are not close to the correct result. Therefore, Jaccard Similarity was used for similarity comparison. It was observed that the similarity results obtained depending on the presence or absence of words within this measurement gave values close to reality.

3.2.4.1. Jaccard Similarity

Jaccard Similarity was used to compare the similarity of the competencies extracted and determined for the position. Zhang (2023) indicates that Jaccard similarity is a measure of similarity between two clusters, defined as the intersection size divided by the union size of the clusters. The formula used in this method is given in Eq. 1.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

Where $|A \cap B|$: the size of the intersection of sets A and B (the number of elements common to both sets). $|A \cup B|$: the size of the union of sets A and B (the total number of unique elements in both sets).

Jaccard similarity ranges from 0 to 1. 0 indicates no similarity (no elements in common) and 1 indicates complete similarity (all elements are common). Jaccard Similarity focuses on the presence or absence of elements that are the same in two sets. Their frequency is not taken into account for this similarity. For this reason, this similarity was preferred in the project. Higher Jaccard similarity scores indicate a greater degree of match between the candidate's skills and job requirements, potentially indicating a stronger fit for the position.

3.2.4.2. Cosine Similarity

Cosine Similarity was applied to compare the similarity of the competencies extracted and determined for the position, but this method was not used as it was observed that it was not compatible with the project. Cosine similarity is a metric used to measure the similarity between two non-zero vectors **[58]**. The formula used in this method is given in Eq. 2.

$$\cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} \tag{2}$$

Where A and B are the two vectors being compared. $A \cdot B$ represents the dot product of the vectors. $\|A\|$ and $\|B\|$ represent the magnitudes of the vectors.

The cosine similarity value varies between −1 and 1: A cosine similarity of 1 indicates complete similarity between the vectors. When evaluated in the context of the text, it shows that the two texts are quite similar. A cosine similarity of 0 indicates that the vectors have no similarity between them. In the context of the text this often indicates the difference. A cosine similarity of −1 indicates complete contrast. This means a complete difference between documents. Since cosine similarity focuses on word frequency in documents, this similarity was not preferred for the project.

3.3. Resume (CV) Screening System

An interface was created using Python code for the resume screening system with Natural Language Processing. The user is expected to select position information first. Then, the file containing the resumes should be selected. When the Select Folder button is pressed, a file is expected to be selected. At this stage, the file containing the CVs is selected. The file can contain any number of resumes. There is no limitation. CVs must be in `pdf` format. The system also expects CVs in English. After selecting the file containing the resumes and pressing the `Run Analysis` button, the results appear on the screen as shown in Figure 1.

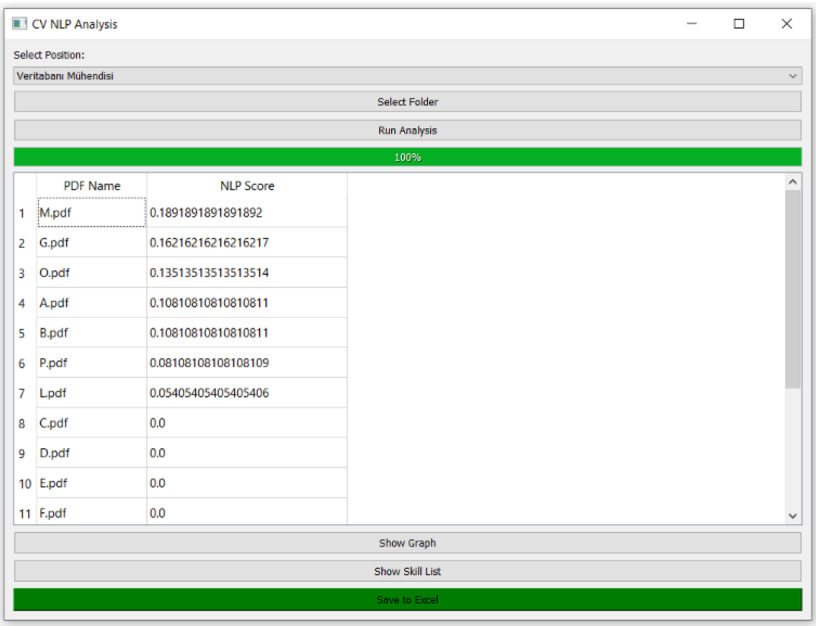

Figure 1. Run the Analysis Button and Similarity Score Results Screen in the Interface

When the `Show Graph` button is pressed, the bar graph opens and the ranking of the candidates is displayed on the graph as seen in Figure 2.

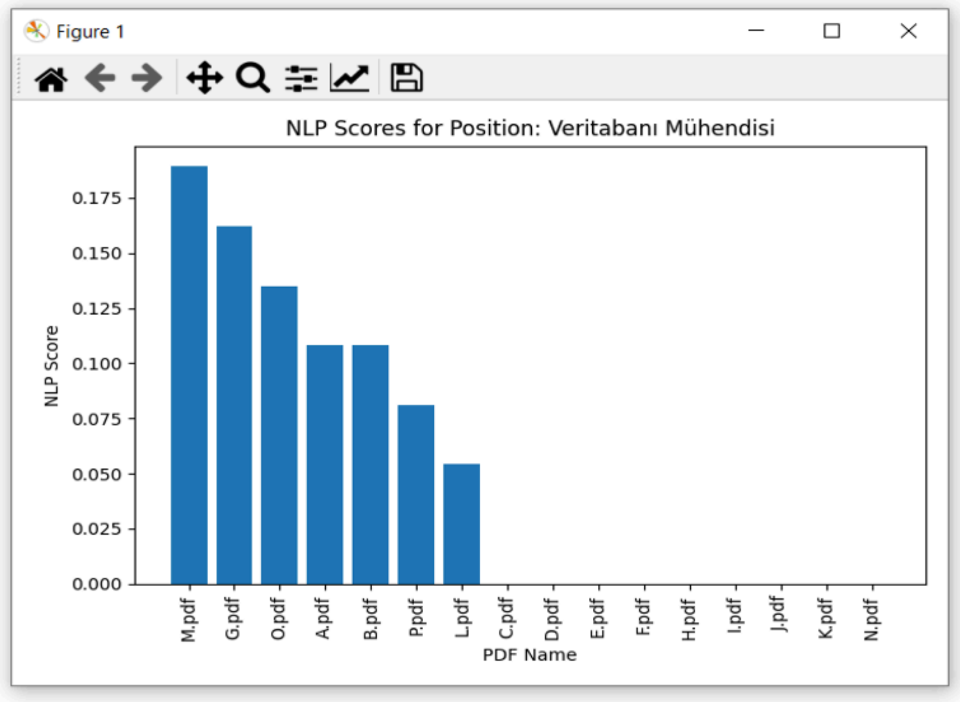

Figure 2. Similarity Scores of the CVs on the Graph

When the `Show Skill List` button is pressed, a screen containing the skills and similarity rates extracted from the CVs appears. As illustrated in Figure 3, this file can be saved as a `.txt` file from the `Save` section.

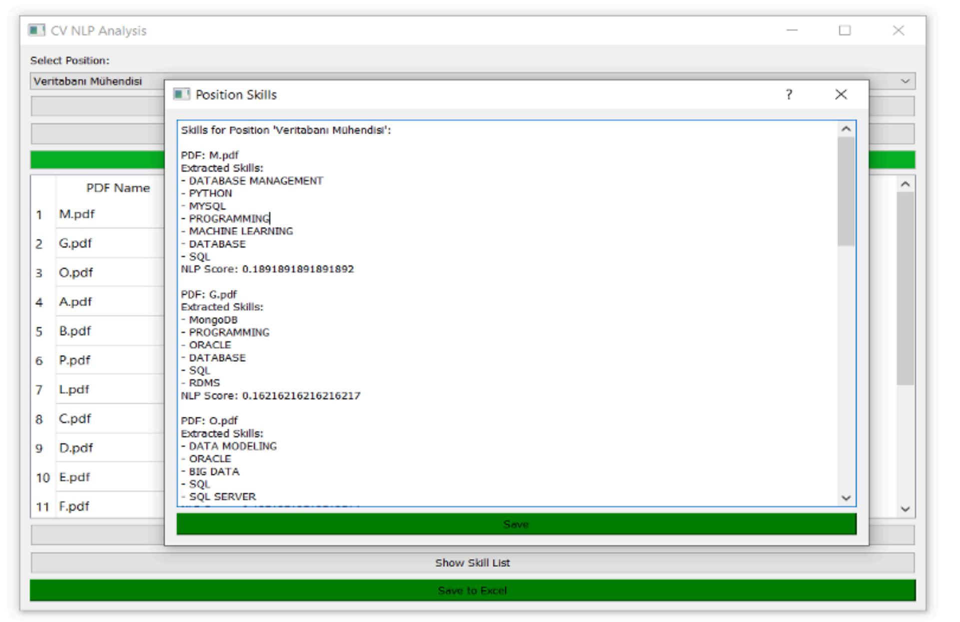

Figure 3. Extracted Skills from the CVs in the Interface

A total of 16 resumes were collected in the study, including fields such as information technology (IT), finance, marketing and engineering, and the results were recorded. This method allowed us to evaluate the performance of the system in a dataset containing resumes from other fields. As can be seen from the results, the results of those who did not have the relevant competence in the relevant position were recorded as 0. Although the dataset used in this project focused on 123 positions in technical and managerial fields, the methodology can be generalized to other sectors by simply adjusting the predefined competency sets for the relevant roles.

4. Results

4.1. Results of Resume Screening with Natural Language Processing (NLP)

Position-based results were recorded on a total of 16 resumes. In this section, the competencies extracted from the CVs, Cosine and Jaccard Similarity values were taken as results. The results are taken from the interface and presented in table form.

Table 5. Result of Extracted Competencies from CVs for Database Engineer Position

Extracted Competencies						
A	B	G	L	M	O	P
Machine Learning	MySQL	Oracle	Big Data	Machine Learning	ORACLE	SQL
MySQL	Database	Database	Python	MySQL	Data Modeling	Linux
Programming	Progress	RDMS		Database	SQL Server	Programming
Python	Python	SQL		SQL	SQL	
		Programming		Programming	Big Data	
		MongoDB		Database Management		
				Python		

In Table 5, it can be observed which competencies are extracted from the CVs named A, B, G, L, M, O, and P. Different numbers of competencies were extracted from each resume. For example, a total of 4 competencies were extracted from the resume named A, namely Machine Learning, MySQL, Programming, and Python. CVs that do not contain any of the information included in the qualification set are not included.

Table 6. Cosine Similarity Scores of the CVs for Database Engineer Position

PDF Name	Cosine Similarity Score	PDF Name	Cosine Similarity Score
O	0.481	D	0
M	0.461	E	0
G	0.382	F	0
L	0.278	H	0
B	0.276	I	0
P	0.234	J	0
A	0.147	K	0
C	0	N	0

In the Table 6, the Cosine Similarity results, indicate the suitability of CVs for the Database Engineer position. Candidates are ranked from the candidate with the highest score to the candidate with the lowest score. As can be seen from the table, there are a total of 7 people suitable for this position. The candidate with the highest score for the position is the one whose resume is named O. With a score of 0.481, the candidate has the highest similarity to the job description, making them the most suitable for the Database Engineer position among all candidates. Candidates M and G also show significant suitability with scores of 0.461 and 0.382, respectively. A cosine similarity score of 0 indicates that there is no similarity between the resumes of candidates and the job description.

Table 7. Jaccard Index Scores of the CVs for Database Engineer Position

PDF Name	Jaccard Index Score	PDF Name	Jaccard Index Score
M	0.189	D	0
G	0.162	E	0
O	0.135	F	0
A	0.108	H	0
B	0.108	I	0
P	0.081	J	0
L	0.054	K	0
C	0	N	0

Moreover, Table 7 shows the Jaccard Similarity results, indicate the suitability of CVs for the Database Engineer position. Candidates are ranked from the candidate with the highest score to the candidate with the lowest score. As can be seen from the table, there are a total of 7 people suitable for this position. Candidates with higher scores have CVs that share more common terms with the job description, indicating better suitability for the Database Engineer position. The candidate with the highest score for the position is the one whose resume is named M. With a score of 0.189, this candidate has the highest Jaccard similarity to the job description, making them the most suitable for the Database Engineer position among all candidates. Candidates G and O also show significant suitability with scores of 0.162 and 0.135, respectively.

For hiring decisions, focusing on candidates with higher scores will likely yield better results in terms of job fit and performance.

4.2. Comparison of the Similarity Score Results

When two similarity measurements are compared, it is observed that the rankings are different. As a result of Cosine Similarity, O is ranked first with a similarity rate of 0.481, while as a result of the Jaccard Index, M is ranked first with a similarity rate of 0.189. There are various methodological reasons why the ranking between the two similarities is different. Cosine similarity measures similarity by taking into account the frequency and rarity of words. It also considers the overall structure and context of the document content. Within this similarity, word frequencies in the document are important. Therefore, if there is a difference in length between documents or if the documents have different word counts, the results may be affected. Jaccard Index measures a similarity by only considering common elements (words) in the document. Within this similarity, the overall length or word frequency of the document's content is not important. On the other hand, since it does not take into account the frequency or rarity of words in the document, it cannot fully measure the semantic similarity between documents.

As explained, Cosine Similarity focuses on word frequencies in two texts, so it takes into account how often certain words (competencies) appear in resumes relative to the overall document length. By emphasizing word frequency, Cosine Similarity can assign higher similarity scores to resumes with irrelevant keyword occurrences. This can weaken matching accuracy and lead to inaccurate results, especially for technical roles where the presence of a skill rather than its frequency is important. A resume that frequently mentions a particular competency may rank higher even if the competency is not critical to the job. In contrast, Jaccard Similarity focuses on the presence or absence of key competencies that better align with the goal of the matching project, based on the candidates' required competency sets rather than their frequency. Since the job roles being assessed in this project require specific technical competencies, it is critical to match skills precisely. By taking into account whether a skill is present or not, Jaccard Similarity's binary approach provides more accurate and reliable results in such cases. At the same time, with larger candidate pools later in the project, Cosine Similarity may become less practical. Since resumes vary in length and complexity, longer resumes with more filler content (even if it is irrelevant to the job) can skew similarity scores. This is less of an issue with Jaccard Similarity, which treats all skills equally regardless of resume length and focuses solely on their presence or absence.

Considering the purpose of the project within the scope of this information, the similarity of both sets comes to the fore. It was concluded that the purpose of the project was the existence of words, not their frequency. For this reason, the Jaccard Index was preferred in the project to simply measure the word overlap (common competencies) between documents and to ensure that documents can be compared without being affected by factors such as length and word count. In addition, when the extracted competencies were analyzed, it was concluded that the number of competencies in both sets was important as a result of the Jaccard Index. The main reason why the Jaccard index is generally low is that this method only evaluates the existence of words and there are not enough common elements between the two compared sets. This occurs when the position set has a wide range of competencies. Having a narrow range of competencies in the position set will increase this value. As a result, the Jaccard Index may generally give a low value for these reasons, but the results should be evaluated according to this information.

To conclude, while Cosine Similarity is useful for analyzing general text similarity, it is less suitable for systems focused on exact qualification matching for job positions. The targeted approach of Jaccard Index provided higher accuracy in matching candidates' skills to predefined job requirements and was therefore chosen as the primary method in this project.

4.3. Performance and Scalability in Larger Candidate Pools

Although the resume screening system's results show how well Jaccard Index works to determine a candidate's appropriateness, more research is required to see how well the techniques will work with a wider pool of candidates and a range of employment opportunities. The computing time and complexity of similarity comparisons may grow in direct proportion to the number of candidates. But because Jaccard Index concentrates on the presence rather than the frequency of competences, it is still computationally efficient. Because it just needs to compare sets rather than compute complex vector representations, the method can be scaled to larger datasets.

Expanding the pool of potential candidates may result in more variation in the formats of resumes and the definition of competencies. While Jaccard Index performs well with exact keyword matching, it may struggle when candidates use synonyms or different expressions for the same competencies, reducing accuracy. Additionally, Jaccard Index may yield lower scores for positions with broader skill requirements. Therefore, while Jaccard Index may be effective for smaller datasets or positions with narrow skill requirements, its performance may need to be evaluated on larger and more diverse datasets. Future improvements could include incorporating NLP techniques such as synonym recognition or using more advanced models such as BERT to improve performance across large and diverse candidate pools.

The Jaccard Index performs particularly well when job descriptions and resumes include clear, defined technical competencies. For positions that require more abstract or soft skills (e.g., leadership, creativity), the Jaccard Index may not fully capture the nuances of candidates' qualifications. Semantic analysis is one of the strategies that can enhance performance for jobs requiring higher subjective abilities.

4.4. Applicability Across Different Sectors and Job Positions

The CV screening system developed in this study was first tested in technical and management roles where competencies are generally specific and well-defined. This method can be adapted to a wide range of sectors and positions, especially those involving technical and management roles. In order to adapt the system to different sectors and positions, the competencies expected within the position should be defined in advance and entered into the system. At the same time, additional techniques can be integrated to increase the generalizability of the system across different sectors. In particular, a thesaurus-based approach or word embeddings can be added to account for differences in how competencies are defined (e.g., "teamwork" vs. "collaboration"). This helps improve the similarity measurement in areas where competencies are more loosely defined. In some sectors and positions, it may be necessary to capture the relationship between keywords. In such cases, integrating the models with each other can increase the accuracy of the results. While the Jaccard Index was chosen for this project because of its focus on matching exact terms, combining it with Cosine Similarity or using context-aware models like BERT can allow the system to capture both the presence of keywords and the relationship between them. It is particularly useful for management or customer relations roles where these skills are often interconnected.

By measuring the overlap between two skill sets, Jaccard Index is particularly effective for positions where exact competency matches are critical, such as technical roles where specific competencies are clearly defined (e.g., software development, IT, engineering). However, its performance can vary depending on the nature of the job and how competencies are defined in different industries. In fields where specific tools, programming languages, or methodologies are clearly defined, such as engineering or IT, Jaccard Index can accurately assess candidate suitability. Since competencies in these industries are often explicitly listed on both resumes and job descriptions, Jaccard Index works well by focusing on the presence or absence of exact keywords. In creative industries like marketing and design or in positions that require more detailed competencies (e.g., leadership, communications), Jaccard Index may be less effective. These roles often involve soft skills or abstract competencies that can be expressed in a variety of ways. Because Jaccard Index does not take into account the context or frequency of terms, its performance may be limited when trying to match more subjective attributes. In conclusion, while Jaccard Index provides reliable results in technical and well-structured domains, it should be noted that it has limitations in capturing the semantic meaning or hierarchy of competencies. In more dynamic or creative domains, incorporating context-aware techniques together with Jaccard Index may yield better results. Future studies may investigate hybrid methods to increase applicability across sectors.

5. Discussion

The resume screening system with Natural Language Processing quickly and effectively handles large amounts of resumes submitted to human resources. Resumes are evaluated according to the competency words they contain, and their suitability for the position is obtained using the Jaccard similarity score. This system works for CVs in English. Turkish CV screening may also be added in future studies. Additionally, the use of different similarity scores can be evaluated, or alternative methods such as vectorization can be used to enhance the evaluation process. Considering the purpose of the project, which focuses on the existence of competency words rather than their frequency, the Jaccard Index was chosen for its simplicity in measuring word overlap and comparing documents without being affected by factors such as length and word count. Although the Jaccard Index generally yields low scores due to its focus on common elements, this method is suitable for identifying the presence of specific competencies. Future enhancements could include integrating multilingual capabilities, exploring additional similarity measures, and investigating advanced NLP techniques such as word embeddings and Named Entity Recognition (NER). Ethical considerations, including bias mitigation and transparency in evaluation, are also crucial for ensuring fair and accurate resume assessments. By addressing these aspects, the system can continuously evolve to provide more robust and comprehensive resume evaluations.

The increasing use of automated resume screening methods raises ethical questions that need to be addressed. In the system created in this project, information such as gender, age, race, or location is not extracted from the resume and therefore does not affect the ranking of candidates. The system focuses only on competency-based factors. Another important ethical issue is the lack of transparency in automated screening processes. Candidates and hiring managers may not fully understand how decisions are made, which can lead to concerns about fairness. Algorithms are often viewed as “black boxes” where internal decision-making is unclear. To address this, systems should be designed with explainability in mind. This means providing clear and understandable explanations for why certain candidates are ranked higher or lower. For example, if a candidate is ranked higher because of certain technical skills that match the job description, the system should be able to explain this reasoning. This transparency not only increases trust in the system, but also allows candidates and employers to understand and potentially appeal decisions if they feel they have been unfairly evaluated. Since the system created within this project focuses on whether the required competencies for the position are included in the CVs, the system can be explained. Another ethical concern in automatic screening is data privacy. Candidates should be informed about how their data is used in the screening process. In particular, it is important to ensure compliance with data protection regulations that require candidates’ personal information to be used transparently and securely.

Future work should include expanding the dataset to include resumes from a variety of industries to improve the generalizability of the screening system. Additionally, integrating multilingual support will allow the system to be globally applicable, enabling it to process resumes from a variety of cultural and linguistic backgrounds. While this project focused on technical skills, future developments should explore including soft skills, which are critical but harder to measure. Using NLP techniques such as sentiment analysis can help analyze qualitative characteristics, providing a more complete assessment of candidates, especially for senior or leadership roles. Ethical concerns such as algorithmic bias should be addressed in future work to ensure fairness in automated screening. Future improvements could allow recruiters to provide real-time feedback and adjust the system's skill weighting based on specific hiring needs. This would make the system more adaptable to different roles and industries, increasing its flexibility and usability.

6. Conclusion

To conclude, this study demonstrates the great potential of using Natural Language Processing (NLP) to create an effective resume screening process, especially for large organizations that can receive numerous applicants. The proposed system provides a scalable and efficient solution to the challenges faced by traditional manual screening methods by integrating Jaccard similarity to assess candidate suitability based on extracted competencies. By focusing on the presence of relevant competencies rather than their frequency, this approach ensures that candidates are evaluated more fairly and objectively, thereby reducing the biases that are common in manual processes.

The results of the study shows that the system can effectively identify and rank candidates based on their suitability for specific positions. This is demonstrated by the comparative analysis of the cosine and Jaccard similarity measures. Although the Jaccard index generally yields lower scores, it is more appropriate for the goals of the project because it provides a simple measure of the overlap between the qualifications of the candidates and the requirements of the job.

For future work, the proposed system could be enhanced by incorporating multilingual capabilities, exploring alternative similarity measures, and integrating more advanced NLP techniques, such as word embeddings, which offer advantages in high-dimensional spaces by representing each word as a continuous vector. This flexibility of the word embedding method allows the proposed system to be generalized across various sectors and positions that require a broader range of competencies. Another approach to enhance the system is through Named Entity Recognition (NER), a powerful method for identifying and categorizing entities. For example, in the sentence 'Ercan works as a Database Engineer,' NER can recognize 'Ercan' as a name and 'Database Engineer' as a position. These techniques not only leverage similarities but also employ additional mathematical approaches to create a more robust model. Overall, the proposed NLP-based approach represents a promising advancement in optimizing the recruitment process, offering a more effective, unbiased, and resource-efficient solution for modern human resource management. It is likely to attract further interest from researchers.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

Alamelu, M., Kumar, D., Sanjana, R., Sree, J., Devi, A., & Kavitha, D. (2021). Resume Validation and Filtration using Natural Language Processing. *2021 10th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON)*, 1–5. https://doi.org/10.1109/iemecon53809.2021.9689075

Ali, I., Mughal, N., Khan, Z. H., Ahmed, J., & Mujtaba, G. (2022). Resume Classification System using Natural Language Processing and Machine Learning Techniques. *Mehran University Research Journal of Engineering and Technology*, *41*(1), 65–79. https://doi.org/10.22581/muet1982.2201.07

Aminu, H., Yau, B. I., Zambuk, F. U., Nanin, E. R., Abdullahi, A., & Yakubu, I. Z. (2023). Salary Prediction Model using Principal Component Analysis and Deep Neural Network Algorithm. *International Journal of Innovative Science and Research Technology*, *8*(12), 1–11. https://doi.org/10.5281/ZENODO.10629245

Anand, A., & Dubey, M. S. (2022). CV Analysis Using Machine Learning. *International Journal for Research in Applied Science and Engineering Technology*, *10*(5), 1316–1322. https://doi.org/10.22214/ijraset.2022.42295

Aydin, O., Karaarslan, E., & Narin, N. G. (2024). *Artificial Intelligence, VR, AR and Metaverse Technologies for Human Resources Management*. https://doi.org/10.48550/ARXIV.2406.15383

Bharadwaj, S., Varun, R., Aditya, P. S., Nikhil, M., & Babu, G. C. (2022). Resume Screening using NLP and LSTM. *2022 International Conference on Inventive Computation Technologies (ICICT)*, 238–241. https://doi.org/10.1109/icict54344.2022.9850889

Cabrera-Diego, L. A., Durette, B., Lafon, M., Torres-Moreno, J.-M., & El-Bèze, M. (2015,). How Can We Measure the Similarity Between Résumés of SelectedCandidates for a Job?. *International Conference on Data Mining DMIN 2015*.

Chou, Y.-C., & Yu, H.-Y. (2020). Based on the application of AI technology in resume analysis and job recommendation. *2020 IEEE International Conference on Computational Electromagnetics (ICCEM)*, 291–296. https://doi.org/10.1109/ICCEM47450.2020.9219491

Daryani, C., Chhabra, G. S., Patel, H., Chhabra, I. K., & Patel, R. (2020, January). An Automated Resume Screening System Using Natural Language Processing and Similarity. *Ethics and Information Technology*. https://doi.org/10.26480/etit.02.2020.99.103

FraiJ, J., & László, V. (2021). Literature Review: Artificial Intelligence Impact on the Recruitment Process. *International Journal of Engineering and Management Sciences*, *6*(1), 108–119. https://doi.org/10.21791/ijems.2021.1.10.

Gan, C., Zhang, Q., & Mori, T. (2024). Application of LLM Agents in Recruitment: A Novel Framework for Automated Resume Screening. *Journal of Information Processing*, *32*(0), 881–893. https://doi.org/10.2197/ipsjjip.32.881

Harsha, T. M., Moukthika, G. S., Sai, D. S., Pravallika, M. N. R., Anamalamudi, S., & Enduri, M. (2022). Automated Resume Screener using Natural Language Processing(NLP). *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*, 1772–1777. https://doi.org/10.1109/icoei53556.2022.9777194

Kino, Y., Kuroki, H., Machida, T., Furuya, N., & Takano, K. (2017). Text Analysis for Job Matching Quality Improvement. *Procedia Computer Science*, *112*, 1523–1530. https://doi.org/10.1016/j.procs.2017.08.054

Lad, A., Ghosalkar, S., Bane, B., Pagade, K., & Chaurasia, A. (2022). Machine Learning Based Resume Recommendation System. *International Journal of Modern Developments in Engineering and Science*, *1*(3), 17–20. https://journal.ijmdes.com/ijmdes/article/view/17

Lalitha, B., Kadiyam, S., Kalidindi, R. V., Vemparala, S. M., Yarlagadda, K., & Chekuri, S. V. (2023). Applicant Screening System Using NLP. *2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA)*, 379–383. https://doi.org/10.1109/icidca56705.2023.10099953

Li, C., Fisher, E., Thomas, R., Pittard, S., Hertzberg, V., & Choi, J. D. (2020). Competence-Level Prediction and Resume & Job Description Matching Using Context-Aware Transformer Models. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 8456–8466. https://doi.org/10.18653/v1/2020.emnlp-main.679

Li, S., Li, K., & Lu, H. (2023). *National Origin Discrimination in Deep-learning-powered Automated Resume Screening*. 1–15. https://doi.org/10.48550/ARXIV.2307.08624

Mehboob, M., Ali, M. S., Ul Islam, S., & Sarmad Ali, S. (2022). Evaluating Automatic CV Shortlisting Tool For Job Recruitment Based On Machine Learning Techniques. *2022 Mohammad Ali Jinnah University International Conference on Computing (MAJICC)*, 1–4. https://doi.org/10.1109/majicc56935.2022.9994112

Mohanty, S., Behera, A., Mishra, S., Alkhayyat, A., Gupta, D., & Sharma, V. (2023). Resumate: A Prototype to Enhance Recruitment Process with NLP based Resume Parsing. *2023 4th International Conference on Intelligent Engi-*

neering and Management (ICIEM), 1–6. https://doi.org/10.1109/iciem59379.2023.10166169

Naveed, Z., Nisar, B., Saifullah, D. M., & Iqbal Baig, J. (2024). Resume Ranking Using Natural Language Processing. *Journal of Computers and Intelligent Systems*, *2*(1), 61–66. https://journals.iub.edu.pk/index.php/JCIS/article/view/2806

Pal, R., Shaikh, S., Satpute, S., & Bhagwat, S. (2022). Resume Classification using various Machine Learning Algorithms. *ITM Web of Conferences*, *44*, 3011. https://doi.org/10.1051/itmconf/20224403011

Pant, D., Pokhrel, D., & Poudyal, P. (2022). Automatic Software Engineering Position Resume Screening using Natural Language Processing, Word Matching, Character Positioning, and Regex. *2022 5th International Conference on Advanced Systems and Emergent Technologies (IC_ASET)*, 44–48. https://doi.org/10.1109/ic_aset53395.2022.9765916

Pimpalkar, A., Lalwani, A., Chaudhari, R., Inshall, M., Dalwani, M., & Saluja, T. (2023). Job Applications Selection and Identification: Study of Resumes with Natural Language Processing and Machine Learning. *2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, 1–5. https://doi.org/10.1109/sceecs57921.2023.10063010

Roy, P. K., Chowdhary, S. S., & Bhatia, R. (2020). A Machine Learning approach for automation of Resume Recommendation system. *Procedia Computer Science*, *167*, 2318–2327. https://doi.org/10.1016/j.procs.2020.03.284

Sajid, H., Kanwal, J., Bhatti, S. U. R., Qureshi, S. A., Basharat, A., Hussain, S., & Khan, K. U. (2022). Resume Parsing Framework for E-recruitment. *2022 16th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, 1–8. https://doi.org/10.1109/imcom53663.2022.9721762

Satheesh, K., Jahnavi, A., Iswarya, L., Ayesha, K., Bhanusekhar, G., & Hanisha, K. (2020). Resume Ranking based on Job Description using SpaCy NER model. *International Research Journal of Engineering and Technology*, *7*(5), 74–77.

Spoorthi, M., Indu Priya, B., Kuppala, M., Karpe, V. S., & Dharavath, D. (2023). Automated Resume Classification System Using Ensemble Learning. *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 1782–1785. https://doi.org/10.1109/icaccs57279.2023.10112917

Suhas Chavare, D., & Bhaskar Patil, A. (2023). Resume Parsing using Natural Language Processing. *GRENZE International Journal of Engineering and Technology*, *9*(1), 721–726.

Tejaswini, K., Umadevi, V., Kadiwal, S. M., & Revanna, S. (2022). Design and development of machine learning based resume ranking system. *Global Transitions Proceedings*, *3*(2), 371–375. https://doi.org/10.1016/j.gltp.2021.10.002

Trinh, Q., & Dang, T.-T. (2021). Automatic Process Resume in Talent Pool by Applying Natural Language Processing. *International Conference on Logistics and Industrial Engineering 2021*, 234–240.