

Article

Not peer-reviewed version

Resume2Vec: Transforming Applicant Tracking Systems with Intelligent Resume Embeddings for Precise Candidate Matching

[Ravi Varma Kumar Bevara](#) ^{*} , [Nishith Reddy Mannuru](#) , [Sai Pranathi Karedla](#) , [Brady D. Lund](#) , [Ting Xiao](#) ^{*} , [Harshitha Pasem](#) , [Sri Chandra Dronavalli](#) , [Siddhanth Rupeshkumar](#)

Posted Date: 23 January 2025

doi: [10.20944/preprints202501.1707.v1](https://doi.org/10.20944/preprints202501.1707.v1)

Keywords: resume screening; applicant tracking system (ATS); large language models(LLM); transformer models; embeddings; BERT; GPT, Llama, job description; recruitment



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Resume2Vec: Transforming Applicant Tracking Systems with Intelligent Resume Embeddings for Precise Candidate Matching

Ravi Varma Kumar Bevara ^{1,*}, Nishith Reddy Mannuru ¹, Sai Pranathi Karedla ²,
Brady Lund ¹, Ting Xiao ^{2,3,*}, Harshitha Pasem ¹, Sri Chandra Dronavalli ¹,
and Siddhanth Rupeshkumar ²

¹ Department of Information Science, University of North Texas, Denton, TX, USA

² Department of Computer Science and Engineering, University of North Texas, Denton, TX, USA

³ The Anuradha and Vikas Sinha Department of Data Science, University of North Texas, Denton, TX, USA

* Correspondence: ravivarmakumarbevara@my.unt.edu (R.V.K.B.); ting.xiao@unt.edu (T.X.)

Abstract: Conventional Applicant Tracking Systems (ATS) encounter considerable constraints in accurately aligning resumes with job descriptions (JD), especially in handling unstructured data and intricate qualifications. We provide Resume2Vec, an innovative method that utilizes transformer-based deep learning models, including encoders (BERT, RoBERTa, and DistilBERT) and decoders (GPT, Gemini, and Llama), to create embeddings for resumes and job descriptions, employing cosine similarity for evaluation. Our methodology integrates quantitative analysis via embedding-based evaluation with qualitative human assessment across several professional fields. Experimental findings indicate that Resume2Vec outperforms conventional ATS systems, achieving enhancements of up to 15.85% in Normalized Discounted Cumulative Gain (nDCG) and 15.94% in Ranked Biased Overlap (RBO) scores, especially within the mechanical engineering and health and fitness domains. Although conventional ATS exhibited slightly superior nDCG scores in operations management and software testing, Resume2Vec consistently displayed a more robust alignment with human preferences across the majority of domains, as indicated by RBO metrics. This research demonstrates that Resume2Vec is a powerful and scalable method for matching resumes to job descriptions, effectively overcoming the shortcomings of traditional systems while preserving a high alignment with human evaluation criteria. The results indicate considerable promise for transformer-based methodologies in enhancing recruiting technology, facilitating more efficient and precise candidate selection procedures.

Keywords: resume screening; applicant tracking system (ATS); large language models(LLM); transformer models; embeddings; BERT; GPT, Llama, job description; recruitment;

1. Introduction

In today's rapidly evolving job market, organizations face significant challenges in identifying and recruiting top talent within a complex digital landscape. The digital transformation of recruitment has led to an exponential increase in job applications, making traditional, manual resume screening both inefficient and unreliable. With talent pools growing exponentially, companies face the dual demands of maintaining competitive hiring practices and ensuring that the candidates selected possess the precise skills necessary for specialized roles. This challenge is amplified during economic downturns, where efficiency and strategic hiring are essential to ensuring candidates align closely with organizational needs. As organizations seek ways to streamline hiring processes without sacrificing candidate quality, there is an urgent need for intelligent systems capable of performing nuanced assessments in a time-efficient manner. Traditional recruitment methods, which often rely on manual processes, are struggling to keep pace, as they lack the capacity to evaluate the nuanced qualifications required for today's specialized roles while processing vast amounts of unstructured data.

To address these challenges, Applicant Tracking Systems (ATS) have been widely adopted to streamline recruitment processes by automating and organizing candidate selection. While ATS platforms provide a degree of efficiency, their traditional reliance on keyword matching to filter candidates limits their effectiveness, as this method fails to capture context and often overlooks qualified candidates who do not use exact keywords. The keyword-centric model further risks introducing systemic biases, as certain terms may favor specific groups or backgrounds, which can inadvertently undermine diversity and inclusion efforts within organizations. As a result, many ATS solutions miss out on understanding the broader context of candidates' experiences and skills, limiting their ability to accurately assess suitability for specialized roles. Format dependency and difficulty handling varying resume structures further exacerbate these issues, leading to potentially inaccurate assessments of candidate suitability and biases in the selection process [1].

The integration of advanced natural language processing (NLP) techniques, especially transformer-based models, presents a compelling solution to enhance ATS capabilities. These technologies, including encoders (BERT, RoBERTa, and DistilBERT) and decoders (GPT, Gemini, and Llama), enable ATS systems to move beyond keyword-based screening by generating neural embeddings that capture the semantic relationships and context within resumes and job descriptions. Embedding techniques allow systems to consider the full scope of a candidate's qualifications and experiences, recognizing similar competencies expressed in varied terms or across different contexts. This shift offers organizations a more holistic view of a candidate's qualifications, supporting recruitment processes that are more precise, inclusive, and scalable even with large volumes of applications [2]. Unlike traditional methods, embeddings can handle unstructured data and adapt to various resume formats, addressing critical limitations in conventional ATS approaches [3].

This paper introduces Resume2Vec, an innovative framework designed to transform applicant tracking systems using intelligent resume embeddings. Resume2Vec leverages transformer models to analyze and match resumes with job descriptions more accurately, allowing organizations to optimize talent acquisition processes by enhancing matching accuracy, reducing bias, and ensuring operational efficiency. Resume2Vec's embedding-based approach goes beyond the surface-level text by identifying the underlying intent and relevance of candidates' experiences to the requirements of specific roles. By evaluating the effectiveness of different transformer architectures, such as encoders (BERT, RoBERTa, and DistilBERT) and decoders (GPT, Gemini, and Llama), Resume2Vec demonstrates improvements over conventional ATS in terms of nuanced candidate evaluation and performance scalability, offering a robust and dynamic solution for modern recruitment challenges. Through comprehensive experimentation, our research highlights the advantages of embedding-based approaches, showcasing their consistency and adaptability across large-scale applications, making Resume2Vec a significant advancement in the field of recruitment technology.

2. Related Work

Applicant Tracking Systems (ATSs) are utilized by recruiters to oversee job applications. These systems range in their functionalities, supporting various elements of recruiting, including application tracking, candidate assessment, and engagement during the hiring process [4].

The automation of resume screening and candidate matching has gained critical importance as organizations face an overwhelming volume of job applications. The introduction of various advanced natural language processing (NLP) techniques has demonstrated potential in addressing these recruitment challenges. However, each approach brings distinct contributions and limitations, creating an opportunity for my proposed framework, Resume2Vec, to build on these advancements and address their shortcomings in scalability, contextual accuracy, and bias mitigation.

A competence-level classification model was developed using context-aware transformers to categorize resumes by experience levels for Clinical Research Coordinator roles [5]. This model effectively sorts candidates based on expertise but remains confined to a single, narrowly defined job role, limiting its application across diverse categories. Resume2Vec extends beyond this limitation by

leveraging transformers to capture the semantic context within resumes and job descriptions, making it adaptable to a broader range of domains and job roles.

To address recruitment biases, several researchers have explored techniques to “degender” resumes by removing gender indicators, thereby reducing bias in automated screening [6]. While this approach reduces explicit gender biases, it compromises candidate information by excessively obfuscating resumes. In contrast, Resume2Vec maintains the richness of candidate data by integrating embedding techniques that capture nuanced professional skills while incorporating bias mitigation strategies without sacrificing contextual depth.

An automated screening solution using Sentence-BERT (S-BERT) was introduced to rank resumes based on semantic similarity to job descriptions [7]. Although S-BERT reduces dependency on keywords, it struggles with large datasets due to computational overhead. Resume2Vec optimizes transformer architectures to handle extensive datasets efficiently, allowing for scalable deployment across larger organizations.

A resume shortlisting method using DistilBERT and YOLOv5, which focuses on efficiently parsing specific skills from resumes, was proposed in another study [8]. While their model improves resume parsing accuracy, it primarily extracts isolated skills, limiting its capability to assess holistic candidate qualifications. Resume2Vec addresses this by embedding the entire work history, qualifications, and experience of candidates, offering a more comprehensive assessment for accurate job matching.

The ResumeAtlas framework, developed using BERT and Gemini, demonstrates high accuracy in categorizing resumes across 43 job categories [9]. However, it lacks a direct matching mechanism with job descriptions, reducing its real-world application in recruitment. Resume2Vec bridges this gap by combining resume embeddings with job-specific requirements, ensuring a precise alignment of candidates’ competencies with role demands.

The evolution of ATS capabilities is further illustrated by studies employing SVM and XGBoost for resume screening and ranking based on company requirements [10]. While effective with structured data, their reliance on uniform resume formats limits adaptability with unstructured resumes. Resume2Vec surpasses this constraint by utilizing transformers capable of processing both structured and unstructured resumes, enhancing versatility across industries.

In examining AI’s impact on ATS, researchers have outlined advancements in parsing, ranking, and bias mitigation but highlighted the limitations of keyword-based approaches [11]. Resume2Vec advances beyond keyword dependency by leveraging semantic embeddings, providing a holistic assessment of candidates’ qualifications, and ensuring compatibility with varied resume formats.

A Siamese Sentence-BERT model, conSultantBERT, was introduced to emphasize cross-lingual matching for multilingual data compatibility [12]. Although it addresses language diversity, it is limited by supervised embeddings without industry-wide scalability. Resume2Vec’s embedding techniques enable consistent, large-scale performance across various industries and languages, facilitating broad applicability.

In tackling data sparsity, synthetic resumes were generated using a combination of real and synthetic data to improve BERT and feedforward neural network classification accuracy [13]. While this approach primarily benefits training, Resume2Vec’s optimized embeddings demonstrate robust performance in actual candidate-job matching, reducing dependency on synthetic data.

Keyword-based ranking using cosine similarity has been shown to be efficient, but it fails to capture the complete context within resumes [14]. Resume2Vec addresses this shortcoming by using embeddings that reflect a candidate’s full qualification spectrum, providing a more contextually aware selection process.

For models with sequential dependencies, an LSTM-based resume screening model was proposed, enhancing accuracy but struggling with large datasets due to high computational costs [15]. Resume2Vec uses transformer-based architectures, which handle long sequences more efficiently, making it suitable for processing high volumes of resumes.

A human-in-the-loop system leveraging ESCOXML-R+ for multilingual resume matching was introduced, combining manual corrections by recruiters with fine-tuned LLMs [16]. While innovative, the dependency on human input introduces inefficiencies and potential biases. Resume2Vec automates the process, leveraging intelligent embeddings that generalize across diverse datasets and requirements, eliminating the need for manual corrections.

A stacked model using KNN, SVC, and XGBoost demonstrated improved resume classification and ranking performance but relied heavily on feature engineering [17]. In contrast, Resume2Vec employs transformer-based contextualized embeddings that dynamically encode semantic relationships, outperforming models dependent on manual feature extraction.

A classification system based on knowledge-base-assisted matching and conceptual segmentation was also developed, effectively routing resumes to occupational categories [18]. However, its reliance on predefined knowledge bases like O*NET introduced high runtime complexity and error-prone segmentation. Resume2Vec eliminates this dependency by mapping resumes and job descriptions directly into a shared vector space, enabling faster and more precise classification.

CNNs with Glove embeddings have been used to classify resumes into hierarchical categories [19]. While CNNs offer hierarchical representations, they lack the contextual awareness provided by transformer models, leading to suboptimal performance when handling complex or ambiguous resume content. Resume2Vec's transformer-based embeddings encapsulate contextual and sequential nuances, delivering superior classification granularity.

Finally, studies have highlighted the ethical concerns and risks of bias perpetuation in LLMs used for candidate screening, underscoring the need for fairness in AI-driven recruitment [20]. Resume2Vec incorporates fairness-driven embedding techniques, leveraging balanced training datasets and fine-tuning to minimize bias.

Frameworks such as those emphasizing summarization and grading of resumes offer efficiency but lack the depth of semantic comparison between resumes and job descriptions [21]. Resume2Vec, however, uses tailored embeddings for alignment tasks, ensuring comprehensive matching and precise ranking.

These existing studies highlight advancements in automated resume screening and candidate matching but reveal significant limitations in handling unstructured data, scaling for large datasets, and providing unbiased candidate evaluations. Resume2Vec addresses these issues by integrating neural embeddings for nuanced semantic analysis, enabling efficient, large-scale deployment while supporting unbiased and precise candidate selection. This framework represents a meaningful advancement in recruitment technology, offering enhanced context, inclusivity, and scalability essential for meeting modern recruitment demands.

3. Methodology

This study employed a comprehensive mixed-methods approach to analyze the mapping of resumes to job descriptions (JDs). The methodology consisted of two phases: Phase 1 focused on qualitative human evaluations, while Phase 2 assessed the performance of traditional applicant tracking systems (ATS) and the proposed deep learning-based methods. Phase 1 established a foundational understanding of human decision-making in resume evaluations, which was later compared with quantifiable scoring mechanisms in Phase 2.

The overall architecture of the proposed system is depicted in Figure 1. The pipeline begins with the collection of resumes from secondary sources, including Kaggle and job descriptions through web scraping. Human evaluators performed manual assessments to establish a baseline ranking for resumes across various job descriptions. Simultaneously, traditional ATS methods processed the resumes to generate relevance scores.

The enhanced method, powered by transformer-based models such as BERT, RoBERTa, DistilBERT, GPT-4.0, Gemini, and Llama, converts both resumes and job descriptions into embeddings. These embeddings were computed to capture nuanced relationships between resumes and job descriptions.

Using cosine similarity, the enhanced method evaluated and ranked resumes based on their alignment with job descriptions. Results from the traditional ATS and the enhanced method were compared through metrics such as Normalized Discounted Cumulative Gain (nDCG) and Ranked Biased Overlap (RBO) to measure the quality of rankings. The architecture also facilitated a comparative analysis to identify the most optimal resumes for each job description, as shown in the results section.

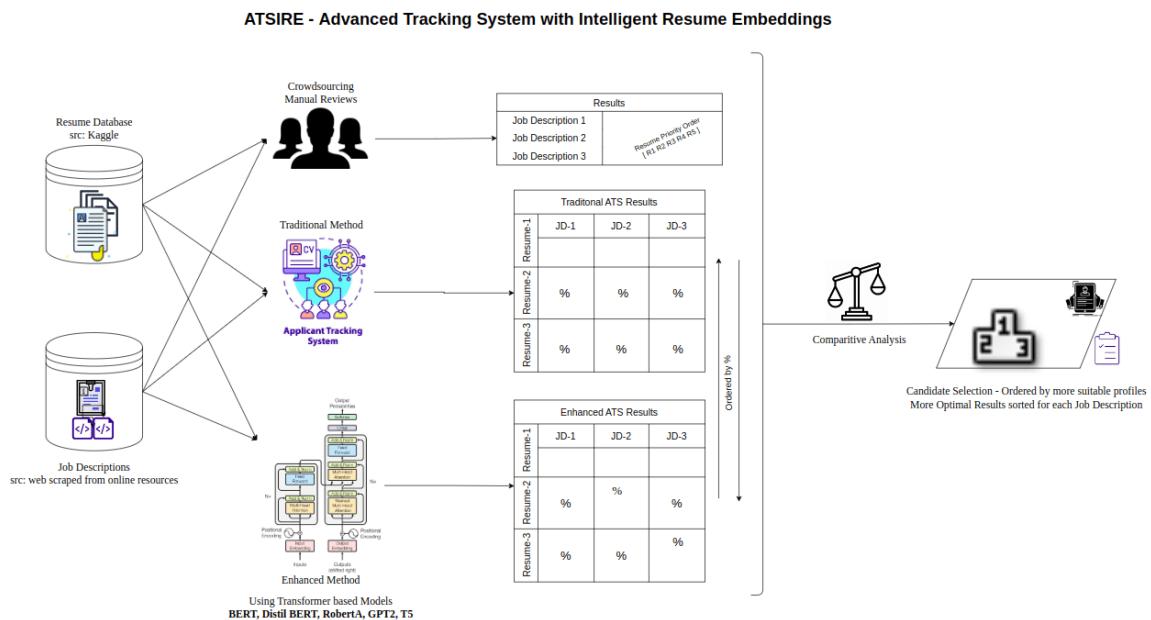


Figure 1. Architecture of the Proposed System for Resume-JD Mapping

Data Collection

The dataset for this study was compiled from two primary sources. Job descriptions (JDs) obtained via web scraping and resumes acquired from publicly accessible platforms and secondary data sources. The data gathering process was meticulously crafted to guarantee diversity across professional sectors, pertinence to actual recruitment scenarios, and coherence with the study's aims.

Job descriptions were acquired through automated web scraping methods. LinkedIn, a prominent professional networking platform, was chosen as the major source because of its vast array of job postings across many industries and locations. Data extraction was conducted utilizing Python packages, such as Selenium and BeautifulSoup, facilitating effective browsing and acquisition of pertinent postings. The scraping method focused on particular job categories, like Health and Fitness, Operations Management, and Data Science, while also covering listings for onsite, remote, and hybrid work environments. The essential information gathered for each job posting comprised job title, firm name, comprehensive job descriptions, employment type, location, posting date, and a link to the original listing. This systematic data collection guaranteed that the information encompassed a wide range of positions and industry-specific needs.

Resumes were obtained from secondary platforms, primarily from Kaggle. Kaggle was chosen for its varied and publicly accessible datasets, which comprise anonymized resumes featuring comprehensive qualifications, professional experiences, and skill sets. The incorporation of resumes from Kaggle yielded a representative dataset for evaluation, as it exhibited diverse levels of proficiency across several disciplines. To maintain the integrity and usability of the data, stringent cleaning and validation protocols were implemented. Duplicate posts and incomplete items were eliminated from job descriptions. Particular emphasis was placed on preserving formatting consistency and ensuring that the extracted information appropriately reflected the designated job categories. To ensure quality entries with missing or incomplete fields were omitted from the dataset for resumes.

This methodology for data collection, including scraping techniques and secondary data platforms, yielded a high-quality dataset that facilitates the efficient assessment of transformer-based approaches for matching resumes to job descriptions.

Data Preprocessing

To ensure the quality and consistency of the dataset, a multi-step text preprocessing pipeline was implemented. This process was designed to clean, standardize, and prepare the text data for downstream analysis. The key steps in the cleaning process are outlined below:

1. Removal of HTML Tags: HTML elements present in the scraped job descriptions were removed to retain only the core textual content.
2. Text Standardization: Text was converted to lowercase to ensure uniformity across the dataset.
3. Stopword Removal: Common stopwords were removed using the Natural Language Toolkit (NLTK) [23]. Additionally, domain-specific stopwords were included to improve relevance.
4. Special Characters and Non-ASCII Removal: Non-ASCII characters and special symbols were filtered out to maintain compatibility and readability.
5. Handling Numbers: While irrelevant digits (e.g., phone numbers) were removed, meaningful numeric data such as percentages, monetary values, and years were preserved.
6. Email Address Removal: To maintain privacy and reduce noise, email addresses were stripped from the data.
7. Punctuation Removal: The punctuation was removed using tokenization techniques to focus on alphanumeric content.

This pipeline ensured that both resumes and job descriptions were effectively cleaned and standardized, addressing common issues such as unstructured formatting, irrelevant content, and noise. The preprocessing steps were essential for preparing the data for subsequent qualitative and quantitative analyses, enabling accurate and reliable evaluation.

Phase 1: Qualitative Human Evaluation

In the first phase, a structured qualitative ranking methodology was employed to capture the nuanced decision-making processes inherent in resume evaluations across five distinct professional domains. Participants were recruited from a diverse pool of undergraduate, graduate, and professional students at a major, U.S.-based Research University. This selection was motivated by their unique perspectives as both job seekers and future hiring managers, as well as their diverse academic backgrounds. Students in professional programs brought valuable industry experience and current knowledge of emerging field requirements, contributing to a balanced and comprehensive evaluation framework [24]. Forty participants were recruited across all domains in the dataset, which originally consisted of ten categories. For the qualitative evaluation, five specific domains including Health and Fitness, Data Science, Software Testing, Mechanical Engineering, and Operations Management were selected. These domains were chosen to ensure the evaluation process reflected consistent expertise and technical specificity while maintaining a manageable scope. Each of the selected domains represented distinct skill requirements, enabling a focused and meaningful analysis of resume evaluation methodologies. Ten anonymized resumes per domain were assessed against two carefully crafted job descriptions, designed to reflect contemporary industry standards and requirements.

The evaluation process was conducted via Qualtrics and utilized a forced-choice ranking system, where participants ranked resumes from 1 to 10. This method emphasized primary criteria, including but not limited to relevance of professional experience, alignment of skills with job requirements, and overall candidate suitability. The forced-choice approach has been demonstrated to mitigate response biases and produce more discriminating results in recruitment contexts [25]. Individual rankings were aggregated using the Borda Count method, which assigns scores based on rank positions and produces a collective ordering. The Borda score $B(c)$ for a candidate c was calculated as:

$$B(c) = \sum_i (n - r_i(c)) \quad (1)$$

where n is the total number of candidates (10), and $r_i(c)$ represents the rank assigned by participant i . Aggregated rankings were validated through Kendall's coefficient of concordance (W), which ranged from 0.68 to 0.82 across domains, indicating substantial agreement among evaluators. To ensure thoughtful and deliberate evaluations, participants were required to justify their top five rankings. Optional comment fields captured additional insights, such as the interpretation of technical skills versus practical experience. These qualitative evaluations served as the human benchmark against which automated methods in Phase 2 were compared.

Phase 2: Quantitative Evaluation and Model Selection

For this phase, six transformer models were utilized to generate embeddings for resumes and job descriptions (JDs): BERT, RoBERTa, DistilBERT, GPT-4.0, Gemini, and Llama. These models were chosen for their proven capability in capturing nuanced relationships in textual data, with encoder models focusing on extracting contextual embeddings and decoder models enhancing generative capabilities. The inclusion of encoder-decoder models, such as GPT-4.0 and Gemini, ensured that contextual understanding and sequence-level nuances were preserved, as discussed in [26]. The choice of these models was further motivated by their ability to handle both structured and unstructured data, critical for robust resume-JD matching.

Cosine similarity was selected as the primary metric for comparing embeddings due to its effectiveness in measuring the angular distance between vectors. This metric is particularly suited for high-dimensional embeddings as it emphasizes the direction of vectors rather than their magnitude, making it robust to variations in embedding scales. Mathematically, cosine similarity between two vectors \mathbf{u} and \mathbf{v} is defined as:

$$\text{Cosine Similarity} = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|} \quad (2)$$

where $\mathbf{u} \cdot \mathbf{v}$ represents the dot product of the vectors, while $\|\mathbf{u}\|$ and $\|\mathbf{v}\|$ denote their magnitudes. This metric was chosen due to its wide applicability in text similarity tasks and its prior use in bias analysis studies [27].

The preprocessing steps involved cleaning the dataset and converting the textual data into embeddings using the six selected transformer models. Scatter plots were generated for both resumes and JDs to visualize the embeddings and assess their ability to capture contextual nuances and domain-specific information.

As shown in Figures 2 and 3, the embeddings exhibited clustering patterns based on domain similarities, indicating that the models effectively captured contextual relationships.

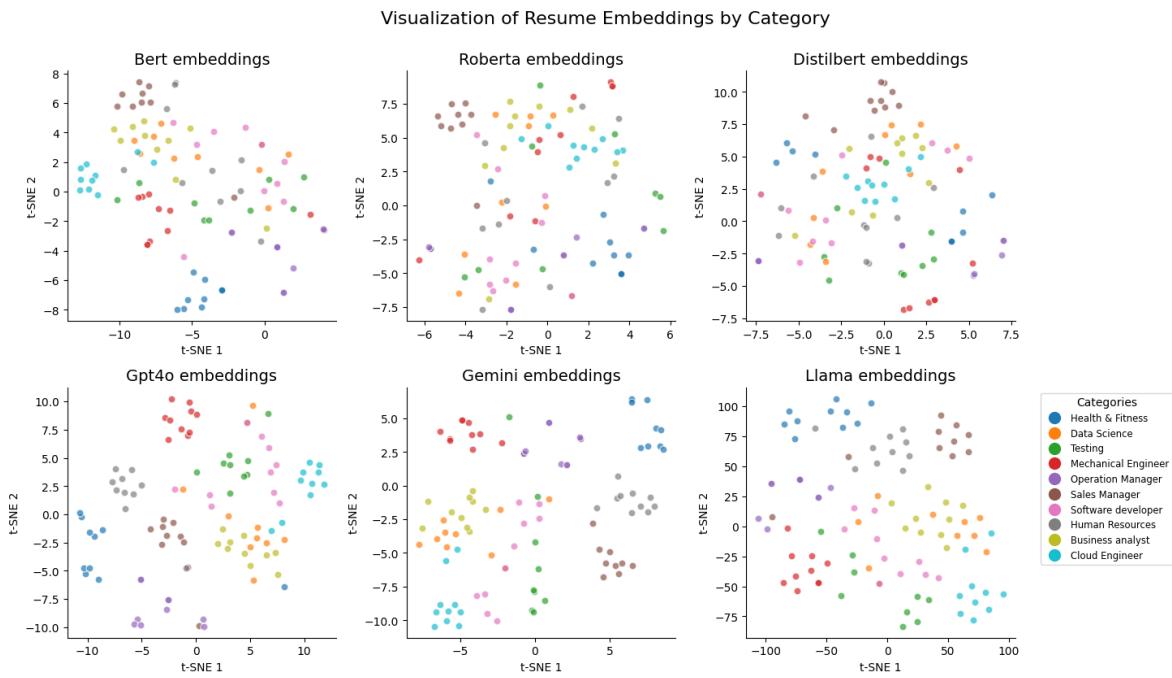


Figure 2. Scatter Plot of Resume Embeddings Across Domains

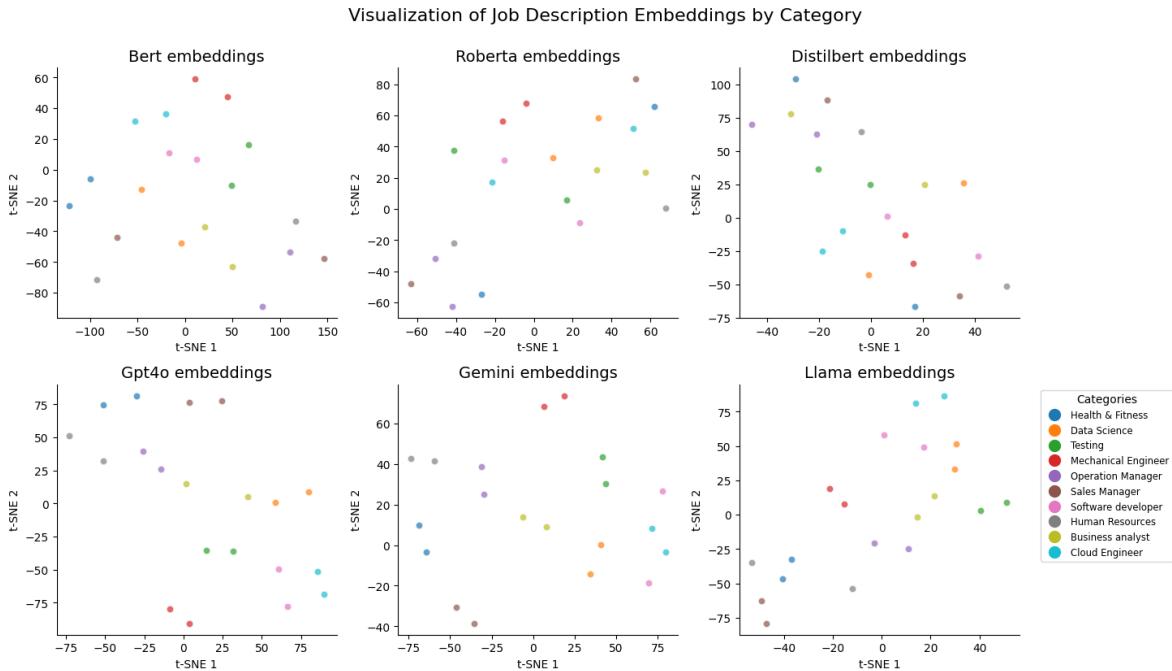


Figure 3. Scatter Plot of Job Description Embeddings Across Domains

To evaluate the performance of different transformer models, a downstream classification task was conducted, inspired by the Census2Vec framework [28]. This task involved predicting the category of resumes based on embeddings, with three embedding options considered: resume embeddings, JD embeddings, and stacked embeddings. Resume embeddings were chosen due to their higher data availability for testing. Accuracy was used as the evaluation metric, and the results indicate that while GPT-4.0 achieved the highest accuracy using SVM (90.5%), Llama demonstrated the best overall performance across other classification algorithms, particularly excelling in Random Forest, Gradient Boosting, and Logistic Regression, as shown in Table 1.

Table 1. Accuracy Comparison of Transformer Models Across Classification Algorithms. Bolded values indicate the highest accuracy achieved in each column for the corresponding classification algorithm.

Model	SVM	Random Forest	Gradient Boosting	Logistic Regression
BERT	78.5%	94.5%	91.3%	95.2%
RoBERTa	70.5%	93.5%	92.5%	94.5%
DistilBERT	83.0%	94.5%	89.5%	95.5%
GPT-4.0	90.5%	91.0%	55.0%	89.5%
Gemini	91.0%	94.0%	73.5%	90.5%
Llama	83.5%	95.5%	95.5%	95.5%

To validate the results, data augmentation was performed, and embeddings were analyzed with and without Principal Component Analysis (PCA). As shown in Figures 4 and 5, PCA improved the performance of some models, but Llama consistently delivered the best results across all scenarios.

Model Accuracy Comparison - Without PCA

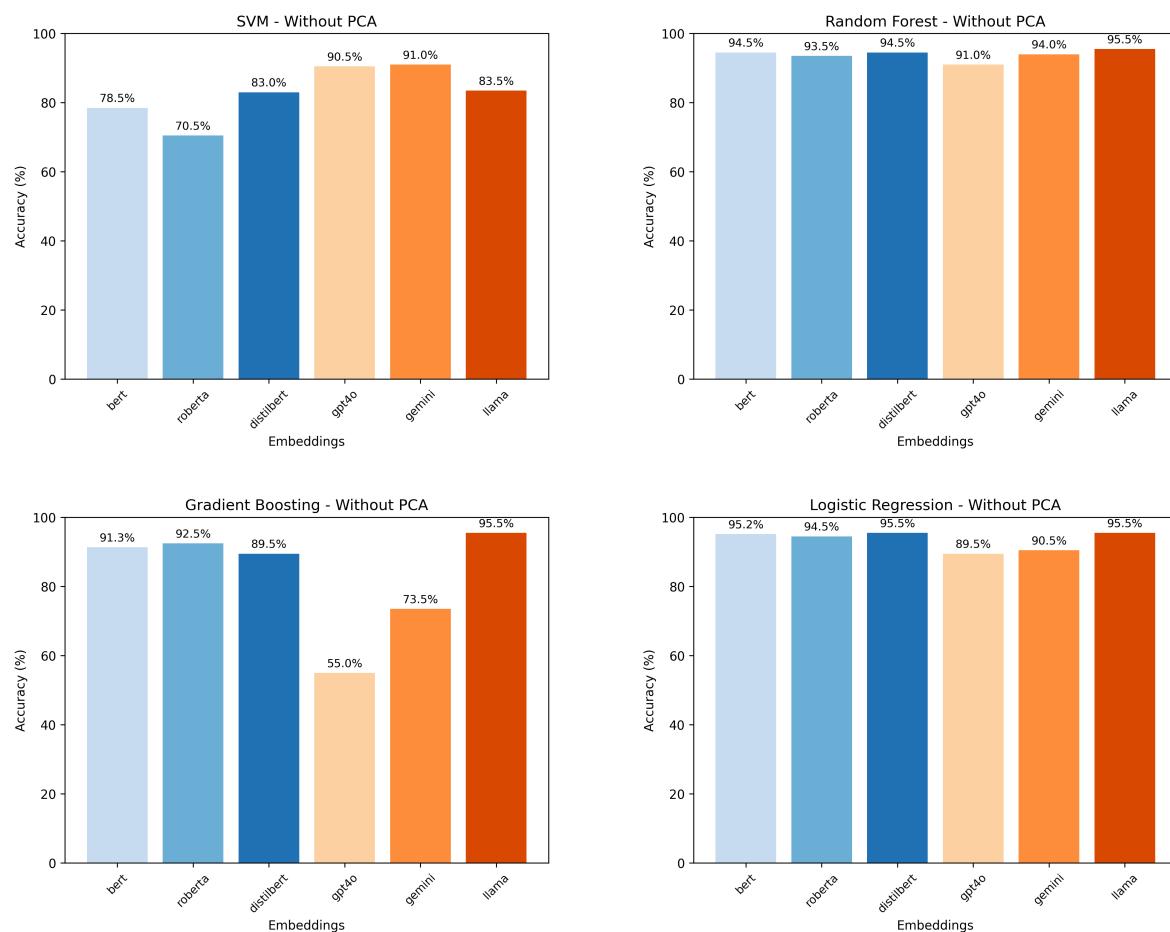


Figure 4. Model Accuracy Comparison Without PCA

Model Accuracy Comparison - With PCA

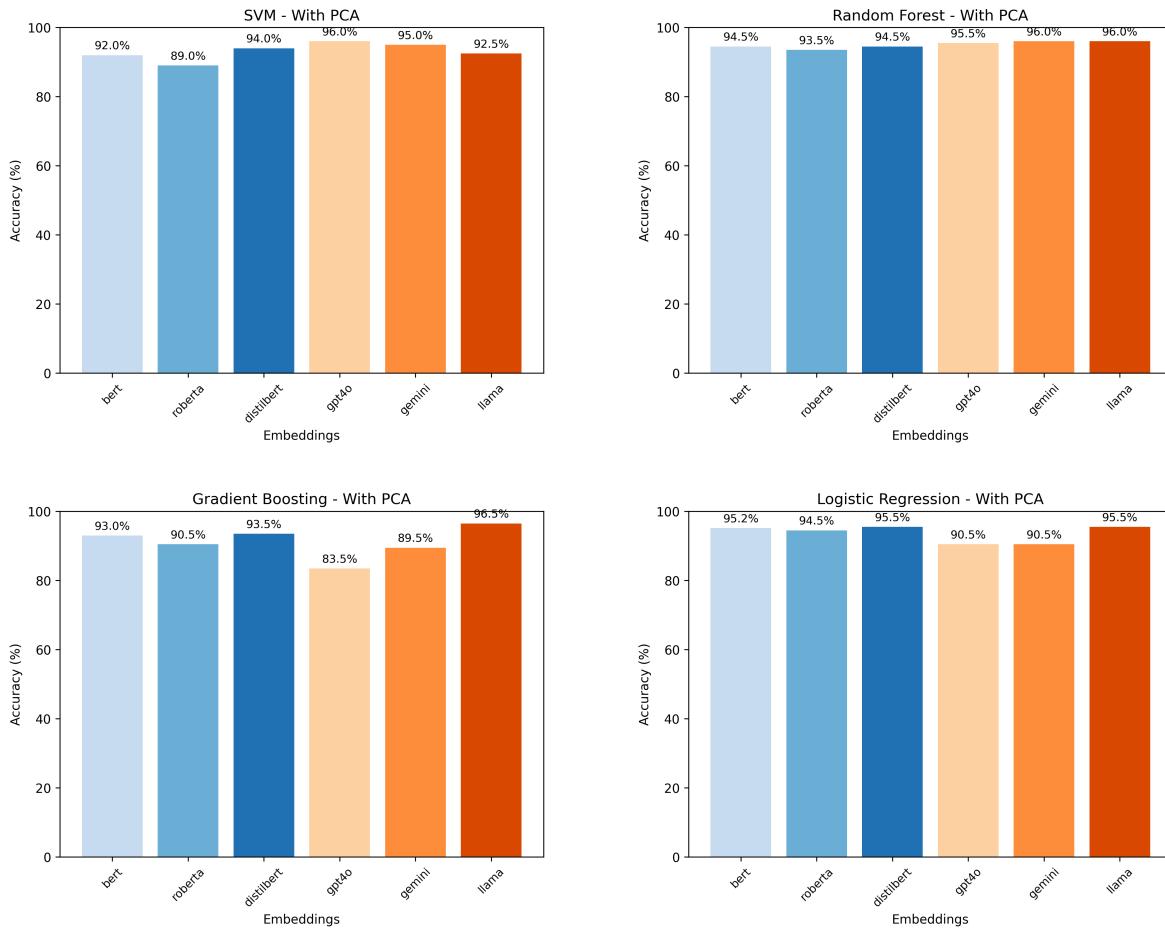


Figure 5. Model Accuracy Comparison With PCA

Using the Llama model, human-evaluated data was also converted to embeddings, and cosine similarity was computed to rank resumes. The final ranking order was compared across three scenarios: human evaluation, traditional ATS, and the proposed method. Metrics such as Normalized Discounted Cumulative Gain (nDCG) and Ranked Biased Overlap (RBO) were employed to evaluate the performance.

The nDCG metric, calculated as:

$$\text{nDCG} = \frac{\text{DCG}}{\text{IDCG}}, \quad \text{where } \text{DCG} = \sum_{i=1}^p \frac{2^{\text{rel}_i} - 1}{\log_2(i+1)} \quad (3)$$

measures the quality of ranked lists. Similarly, RBO, a ranking similarity measure, provided further insights into alignment with human rankings [29]. The results, as shown in table 2 & table 3, demonstrated that the proposed method outperformed traditional ATS in most domains.

Table 2. Comparison of nDCG Scores Across Domains

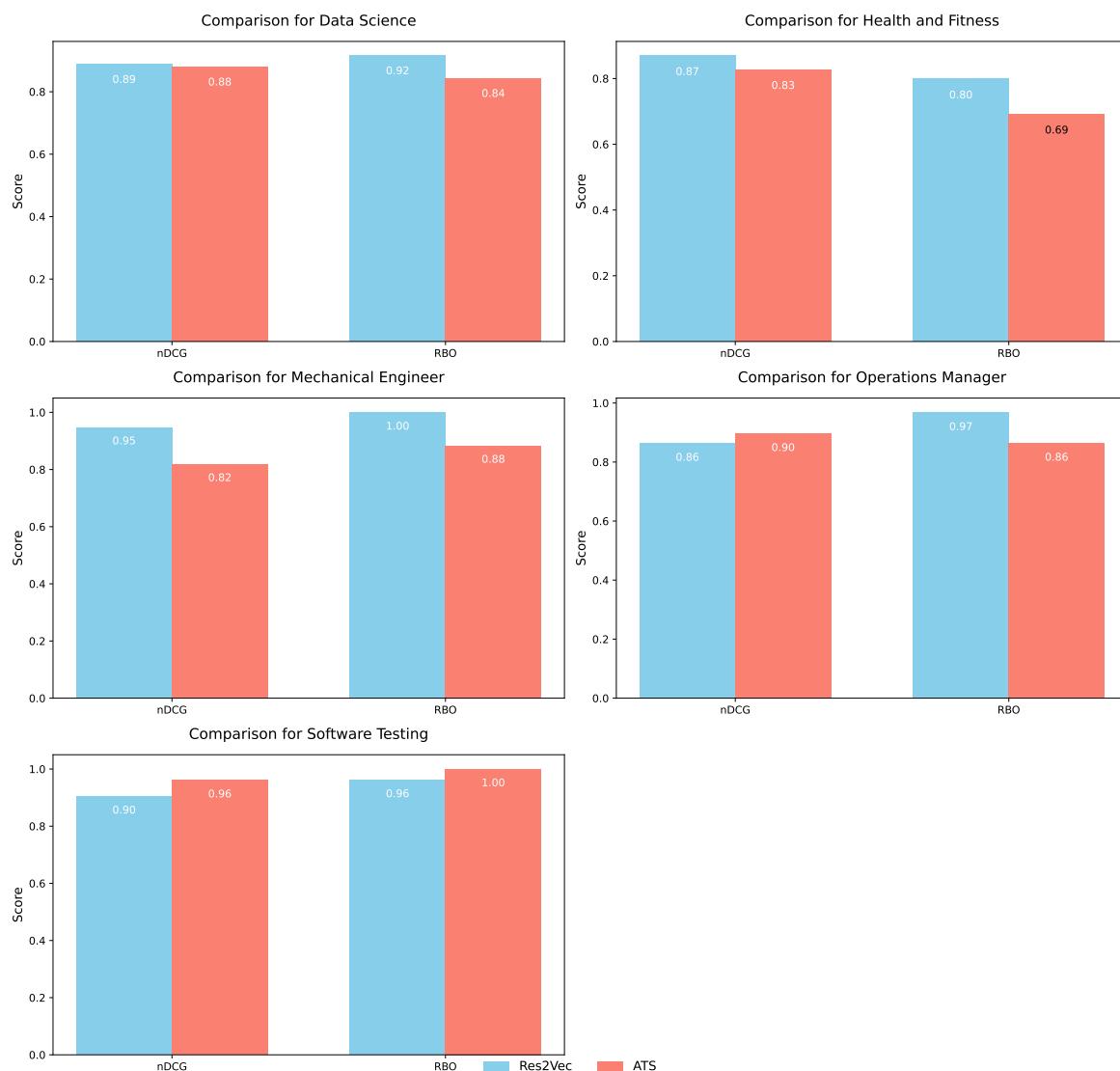
Domain	nDCG (ATS)	nDCG (Proposed)
Data Science	0.88	0.89
Health and Fitness	0.83	0.87
Mechanical Engineering	0.82	0.95
Operations Manager	0.90	0.86
Software Testing	0.96	0.90

Table 3. Comparison of RBO Scores Across Domains

Domain	RBO (ATS)	RBO (Proposed)
Data Science	0.84	0.92
Health and Fitness	0.69	0.80
Mechanical Engineering	0.88	1.00
Operations Manager	0.86	0.97
Software Testing	1.00	0.96

4. Results and Evaluation

The evaluation of Resume2Vec and ATS was conducted using two ranking metrics, namely Normalized Discounted Cumulative Gain (nDCG) and Rank-Biased Overlap (RBO), across five distinct job categories: Data Science, Health and Fitness, Mechanical Engineering, Operations Management, and Software Testing. The results provide insight into the comparative performance of the two approaches and highlight the strengths of Resume2Vec in most categories, as shown in Figure 6.

**Figure 6.** Comparison of Resume2Vec and ATS performance across various metrics (nDCG and RBO) for different job categories.

In the Data Science category, Resume2Vec demonstrated a slight advantage in both metrics. It achieved an nDCG score of 0.89 compared to ATS's score of 0.88, reflecting a marginal improvement

of 1.14%. For the RBO metric, Resume2Vec outperformed ATS with a score of 0.92, representing a significant 9.52% improvement over ATS's score of 0.84. These results indicate that Resume2Vec aligns better with human rankings in this domain.

For Health and Fitness, Resume2Vec exhibited a stronger advantage. Its nDCG score was 0.87, outperforming ATS's score of 0.83 by 4.82%. The improvement was even more pronounced in the RBO metric, where Resume2Vec achieved a score of 0.80, outperforming ATS's score of 0.69 by 15.94%. These findings highlight Resume2Vec's ability to consistently produce rankings that are more closely aligned with human judgments in this category.

The largest performance gains for Resume2Vec were observed in the Mechanical Engineering category. It achieved an nDCG score of 0.95, which was 15.85% higher than ATS's score of 0.82. In terms of RBO, Resume2Vec achieved a perfect score of 1.00, reflecting a 13.64% improvement over ATS's score of 0.88. These results underscore the effectiveness of Resume2Vec in capturing human preferences in highly technical fields like Mechanical Engineering.

In the Operations Management category, ATS showed a marginal advantage in the nDCG metric, achieving a score of 0.90 compared to Resume2Vec's score of 0.86. This difference represents a slight 4.44% improvement for ATS. However, Resume2Vec still outperformed ATS in the RBO metric, achieving a score of 0.97 compared to ATS's score of 0.86, marking a 12.79% improvement. This marginal edge for ATS in nDCG could be attributed to the domain's inherent nature, where operational workflows and performance evaluations may favor ATS's ability to prioritize direct relevance. In contrast, Resume2Vec excels in preserving the relative ranking order, as evidenced by its consistently higher RBO scores. This suggests that Resume2Vec is better suited for tasks requiring fine-grained rank alignment, while ATS may provide value in environments where specific job-requirement matching is paramount.

In the Software Testing category, ATS again demonstrated a slight advantage in nDCG, achieving a score of 0.96 compared to Resume2Vec's score of 0.90, resulting in a 6.25% improvement. Despite this, Resume2Vec achieved a perfect score of 1.00 in the RBO metric, reflecting a 4.17% improvement over ATS's score of 0.96. The stronger nDCG performance of ATS in this category might stem from how resumes and job descriptions in software testing often emphasize keyword matches or technical skills explicitly listed in job requirements. ATS's reliance on keyword relevance likely contributed to this outcome. However, Resume2Vec's superior RBO score indicates that it provides better alignment with overall human preferences, which are not solely dictated by keyword matches but rather by broader contextual factors.

Overall, the results show that Resume2Vec consistently outperformed ATS in the RBO metric across all categories. This suggests that Resume2Vec is better at maintaining agreement with human preferences in ranking tasks. While Resume2Vec also demonstrated superior nDCG scores in most categories, ATS showed slight improvements in nDCG in two domains: Operations Management and Software Testing. These instances of ATS's better performance are likely influenced by domain-specific characteristics, such as the explicitness of technical requirements or the prevalence of structured, keyword-driven evaluation criteria in resumes and job descriptions. However, these advantages are narrow and do not diminish Resume2Vec's broader strengths. Resume2Vec consistently demonstrated its ability to align with more nuanced human judgments, as reflected in its consistently higher RBO scores.

Furthermore, the findings suggest that Resume2Vec is a highly effective approach for ranking tasks, particularly in scenarios where maintaining the relative order of rankings is critical. Even in domains where ATS shows slight advantages, these are limited to specific metrics and contexts, whereas Resume2Vec offers more comprehensive and reliable performance across multiple job categories. Its performance across multiple domains underscores its potential for broader adoption in resume-ranking applications, while also pointing to opportunities for further optimization in specific contexts.

5. Discussion

The findings of this study underscore the transformative potential of Resume2Vec in redefining modern recruitment processes by addressing critical limitations of conventional Applicant Tracking Systems (ATS). By employing transformer-based architectures such as BERT, RoBERTa, DistilBERT, GPT-4, Gemini, and Llama, Resume2Vec demonstrates a clear advantage in aligning candidate profiles with job descriptions through semantic embedding techniques. Unlike traditional ATS platforms that rely heavily on keyword matching, Resume2Vec captures nuanced contextual relationships, enabling a more holistic assessment of candidates' qualifications.

Resume2Vec's performance gains were particularly pronounced in technical domains like Mechanical Engineering and Health and Fitness, where it outperformed traditional ATS by significant margins in metrics including Normalized Discounted Cumulative Gain (nDCG) and Rank-Biased Overlap (RBO). These findings highlight Resume2Vec's ability to process complex and unstructured resumes effectively. Additionally, Resume2Vec's embeddings consistently aligned with human preferences, as indicated by higher RBO scores across all evaluated domains. This alignment with human judgment underscores the framework's capability to mitigate the systemic biases often associated with keyword-based ATS systems. By focusing on neural embeddings, Resume2Vec promotes diversity and inclusion, ensuring that candidates are evaluated based on the full spectrum of their qualifications rather than narrowly defined terms.

However, the study also identified specific contexts where Resume2Vec demonstrated room for improvement. In domains like operations management and software testing, traditional ATS systems marginally outperformed Resume2Vec in nDCG scores. This outcome can be attributed to domain-specific characteristics, where job descriptions often rely on explicitly stated keywords or technical skills, making ATS's keyword matching slightly more effective. This highlights the need for further optimization in embedding techniques to enhance their adaptability in domains where structured, keyword-driven evaluations dominate.

The scalability of Resume2Vec is another critical advantage. By leveraging transformer-based architectures capable of processing large-scale datasets, Resume2Vec supports efficient recruitment workflows without sacrificing accuracy or alignment with human preferences. Moreover, its ability to handle unstructured data and adapt to diverse resume formats makes it highly versatile across industries. This versatility positions Resume2Vec as a robust tool for organizations navigating competitive talent acquisition landscapes.

Despite its successes, several areas warrant future exploration. Enhancing the interpretability of embeddings, particularly in sensitive domains, could improve recruiter trust and adoption. Additionally, integrating multimodal data—such as combining resumes with interview transcripts or performance reviews—could further refine the accuracy of candidate evaluations. Addressing computational efficiency in handling large datasets and expanding the framework's capabilities for multilingual and cross-cultural contexts also represent valuable avenues for improvement.

In conclusion, the discussion reaffirms Resume2Vec's potential to revolutionize recruitment by providing an inclusive, scalable, and context-aware solution for matching candidates to job descriptions. While challenges remain in specific contexts, the framework's adaptability and performance gains highlight its promise as a next-generation recruitment tool, paving the way for more equitable and efficient hiring practices.

6. Conclusions

Resume2Vec represents a significant advancement in recruitment technology, addressing the limitations of conventional ATS platforms through the integration of transformer-based models like BERT, RoBERTa, DistilBERT, GPT-4, Gemini, and Llama. By employing neural embeddings to capture semantic relationships between resumes and job descriptions, Resume2Vec demonstrates superior performance in aligning candidate profiles with organizational requirements. Its ability to process

unstructured data and adapt to varying resume formats underscores its versatility across diverse professional domains.

The framework's consistent outperformance of traditional ATS in metrics such as nDCG and RBO highlights its capacity to deliver nuanced, human-aligned evaluations. While domains like Mechanical Engineering and Health and Fitness showcased the strongest gains, areas such as operations management and software testing revealed opportunities for further optimization. These findings reinforce Resume2Vec's role in fostering diversity, reducing bias, and enhancing recruitment efficiency on a large scale.

Future research may explore deeper integrations of multimodal data, improved computational efficiency, and expanded applications to multilingual and cross-cultural contexts. By addressing these areas, Resume2Vec can continue to refine its capabilities, ensuring its relevance in an ever-evolving job market. As organizations prioritize inclusive and scalable hiring practices, Resume2Vec offers a transformative solution to meet modern recruitment challenges with precision and adaptability.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (IRB) of the University of North Texas (protocol code IRB-24-372, approved on August 5, 2024).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

References

1. Albaroudi, E., Mansouri, T., & Alameer, A. (2024). A Comprehensive Review of AI Techniques for Addressing Algorithmic Bias in Job Hiring. *AI*, 5(1), 383-404.
2. Strohmeier, S., & Piazza, F. (2015). Artificial intelligence techniques in human resource management—a conceptual exploration. *Intelligent Techniques in Engineering Management: Theory and Applications*, 149-172.
3. Le, Q., & Mikolov, T. (2014, June). Distributed representations of sentences and documents. In International conference on machine learning (pp. 1188-1196). PMLR.
4. AL-Qassem, A. H., Agha, K., Vij, M., Elrehail, H., & Agarwal, R. (2023, March). Leading Talent Management: Empirical investigation on Applicant Tracking System (ATS) on e-Recruitment Performance. In 2023 International Conference on Business Analytics for Technology and Security (ICBATS) (pp. 1-5). IEEE.
5. 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. *arXiv preprint arXiv:2011.02998*.
6. Parasurama, P., & Sedoc, J. (2021). Degendering Resumes for Fair Algorithmic Resume Screening. *arXiv preprint arXiv:2112.08910*.
7. Deshmukh, A., & Raut, A. (2024). Enhanced Resume Screening for Smart Hiring Using Sentence-Bidirectional Encoder Representations from Transformers (S-BERT). *International Journal of Advanced Computer Science & Applications*, 15(8).
8. James, V., Kulkarni, A., & Agarwal, R. (2022, December). Resume Shortlisting and Ranking with Transformers. In International Conference on Intelligent Systems and Machine Learning (pp. 99-108). Cham: Springer Nature Switzerland.
9. Heakl, A., Mohamed, Y., Mohamed, N., Elsharkawy, A., & Zaky, A. (2024). ResuméAtlas: Revisiting Resume Classification with Large-Scale Datasets and Large Language Models. *Procedia Computer Science*, 244, 158-165.
10. Vishaline, A. R., Kumar, R. K. P., VVNS, S. P., Vignesh, K. V. K., & Sudheesh, P. (2024, July). An ML-based Resume Screening and Ranking System. In 2024 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT) (pp. 1-6). IEEE.
11. Minvielle, L. (2025, January 8). How AI is shaping applicant tracking systems. Geekflare. <https://geekflare.com/guide/ats-and-ai/>
12. Lavi, D., Medentsiy, V., & Graus, D. (2021). consultantbert: Fine-tuned siamese sentence-bert for matching jobs and job seekers. *arXiv preprint arXiv:2109.06501*.

13. Skondras, P., Zervas, P., & Tzimas, G. (2023). Generating Synthetic Resume Data with Large Language Models for Enhanced Job Description Classification. *Future Internet*, 15(11), 363.
14. Pendhari, H., Rodricks, S., Patel, M., Emmatty, S., & Pereira, A. (2023, December). Resume Screening using Machine Learning. In 2023 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI) (pp. 1-5). IEEE.
15. Bharadwaj, S., Varun, R., Aditya, P. S., Nikhil, M., & Babu, G. C. (2022, July). Resume Screening using NLP and LSTM. In 2022 International Conference on Inventive Computation Technologies (ICICT) (pp. 238-241). IEEE.
16. Kavas, H., Serra-Vidal, M., & Wanner, L. (2024, August). Using Large Language Models and Recruiter Expertise for Optimized Multilingual Job Offer–Applicant CV Matching. In Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (pp. 8696-8699).
17. Ransing, R., Mohan, A., Emberi, N. B., & Mahavarkar, K. (2021, December). Screening and Ranking Resumes using Stacked Model. In 2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT) (pp. 643-648). IEEE.
18. Zaroor, A., Maree, M., & Sabha, M. (2017, November). JRC: a job post and resume classification system for online recruitment. In 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI) (pp. 780-787). IEEE.
19. Nasser, S., Sreejith, C., & Irshad, M. (2018, July). Convolutional Neural Network with Word Embedding Based Approach for Resume Classification. In 2018 International Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR) (pp. 1-6). IEEE.
20. Koh, N. H., Plata, J., & Chai, J. (2023). BAD: BiAs Detection for Large Language Models in the context of candidate screening. arXiv preprint arXiv:2305.10407.
21. Gan, C., Zhang, Q., & Mori, T. (2024). Application of LLM agents in recruitment: A novel framework for resume screening. arXiv preprint arXiv:2401.08315.
22. Bevara, R. V. K. (2024). Scraping LinkedIn for Job Descriptions: A Data Scientist's Guide. *Medium*. Retrieved from <https://medium.com/@ravivarmakumarbevara/scraping-linkedin-for-job-descriptions-a-data-scientists-guide-337dc91de618>.
23. Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. *O'Reilly Media, Inc.*
24. Willis, K. L., Holmes, B., & Burwell, N. (2022). Improving Graduate Student Recruitment, Retention, and Professional Development During COVID-19. *American Journal of Educational Research*, 10(2), 81–84. <https://doi.org/10.12691/education-10-2-2>.
25. Hung, S.-P., & Huang, H.-Y. (2022). Forced-Choice Ranking Models for Raters' Ranking Data. *Journal of Educational and Behavioral Statistics*, 47(5), 603–634. <https://doi.org/10.3102/10769986221104207>.
26. Shahriar, S., Lund, B. D., Mannuru, N. R., Arshad, M. A., Hayawi, K., Bevara, R. V. K., ... & Batool, L. (2024). Putting GPT-4.0 to the Sword: A Comprehensive Evaluation of Language, Vision, Speech, and Multimodal Proficiency. *Applied Sciences*, 14(17), 7782.
27. Bevara, R. V. K., Mannuru, N. R., Karedla, S. P., & Xiao, T. (2024). Scaling Implicit Bias Analysis Across Transformer-Based Language Models Through Embedding Association Test and Prompt Engineering. *Applied Sciences*, 14(8), 3483.
28. Bevara, R. V. K., Wagenvoord, I., Hosseini, F., Sharma, H., Nunna, V., & Xiao, T. (2024). Census2Vec: Enhancing Socioeconomic Predictive Models with Geo-Embedded Data. *Intelligent Systems Conference*, pp. 626-640. Springer Nature Switzerland.
29. Bevara, R. V. K., Xiao, T., Hosseini, F., & Ding, J. (2023, October). Bias Analysis in Language Models using An Association Test and Prompt Engineering. In 2023 IEEE 23rd International Conference on Software Quality, Reliability, and Security Companion (QRS-C) (pp. 356-363). IEEE.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.