Journal of Applied Science, Engineering, Technology, and Education *ISSN: 2685-0591 (online)*
Vol. 8 No. 2 (2026)
https://doi.org/10.35877/454RI.asci4868

An NLP-Based Intelligent Recruitment System Using Semantic Similarity for Ranking Interview Candidates

Silvester Dian Handy Permana*, Anies Lastiati, Umar Al Faruq

Department of Informatics, Faculty of Sains, Engineering and Design, Universitas Trilogi, Jakarta, Indonesia

Abstract

The increasing number of applicants competing for a single job vacancy presents significant challenges for Human Resources (HR) professionals in identifying the most qualified candidates for further evaluation and interview stages. Traditional manual screening processes are often time-consuming, subjective, and difficult to scale when handling large volumes of applications, particularly when applicants provide open-ended responses that require semantic interpretation. This study proposes and implements a Natural Language Processing (NLP)-based artificial intelligence (AI) model designed to automate and improve the candidate screening process by analyzing and evaluating applicants' textual responses. The research utilized data collected from 100 job applicants, which served as the primary dataset for model development and testing. The NLP framework employed semantic analysis techniques to assess the relevance, quality, and contextual meaning of applicant responses, enabling the system to generate recommendations regarding candidate suitability. Empirical evaluation was conducted by comparing the model's recommendations with manual assessments of applicant responses. The results demonstrate that the proposed model effectively identifies promising candidates and provides consistent recommendations that support HR decision-making. By reducing the workload associated with initial screening and improving the objectivity of candidate evaluation, the proposed AI-driven approach offers a practical solution for modern recruitment processes. The findings suggest that NLP-based recruitment systems can enhance talent acquisition efficiency and contribute to more accurate and data-driven hiring decisions in organizations.

Keywords: Natural Language Processing, Intelligent Recruitment System, Semantic Similarity, Interview Answer Analysis, Candidate Ranking

Received: 6 April 2026 Revised: 9 June 2026 Published: 31 August 2026

1. Introduction

The rise of digital recruiting platforms has resulted in organizations receiving more and more job applications (Gilch & Sieweke, 2020; Nabil Rakha Dwitya & Istiono, 2023). While digital recruiting platforms make online job applications accessible, they also make it difficult for recruiters to screen candidates, especially in the initial stages of the recruitment process (Hangartner et al., 2021). When it comes to initial recruitment processes and screening, employers have historically relied on more human, bias-prone processes like screening resumes and filtering out candidates. These processes are time consuming and difficult to implement at scale. Therefore, there is a strong need for recruitment systems that can evaluate candidates in a more automated way, while also ensuring accuracy and fairness (Agbasiere & Nze-Igwe, 2025; Mujtaba & Mahapatra, 2024).

In many recruitment systems, candidates are posed with a question that lacks a definitive answer to gain an understanding of the candidates' qualifications and to evaluate the scope and content of the candidate's response (Fisher et al., 2021). Responses to such questions can be particularly difficult to gauge, since most automated recruiting systems typically assess candidate responses based strictly on keywords rather than the meanings behind the words (Mashayekhi et al., 2024). These types of questions are often as follows: Describe a time when you held a relevant position, What position do you want to live out in the future, What motivates you professionally, What do you want to accomplish, in terms of development, at the organization.

* Corresponding author.
E-mail address: handy@trilogi.ac.id

The challenge of screening candidates and matching them with relevant jobs can now be addressed with the most recent advancements in Natural Language Processing (NLP) (García-Díaz et al., 2018; Huaze Xie; Mohd Anuaruddin Bin Ahmadon; Shingo Yamaguchi; Ichiro Toyoshima, 2019; Saatci et al., 2024). Models of semantic representation empower computers to comprehend contextual meaning and relationships, evaluate similarity among texts, and derive hidden meaning from unstructured information. While the application of Natural Language Processing (NLP) has gained traction in the recruitment sector—particularly for screening resumes and aligning candidates with job roles (Alsaif et al., 2022) current research remains largely confined to analyzing structured documents and extracting predefined skill sets (Devi & Anand, 2025; Konstantinidis et al., 2022). Consequently, there is a notable lack of focus on the systematic analysis of responses from open-ended interviews.

This is a significant oversight, as such qualitative data often yields deeper insights into a candidate's practical experience, underlying motivations, and long-term professional aspirations. This study introduces an intelligent, NLP-based recruitment framework designed to bridge this gap by evaluating and ranking candidates through a semantic similarity analysis of their interview responses against specific job descriptions. Moving beyond the limitations of rudimentary keyword matching, the proposed model emphasizes a semantic assessment that captures the sophistication and nuance of a candidate's discourse. By juxtaposing personal narratives regarding work history, role comprehension, and growth trajectories with institutional requirements, the system aims to provide a more objective and meaningful evaluation.

The necessity for such research is underscored by the rising demand for equitable and scalable hiring solutions. As automation becomes a corporate standard, developing digital tools capable of authentic natural language understanding is essential to mitigate manual burdens, reduce unconscious bias, and enhance the caliber of strategic hiring decisions (Hryn, 2025). The system developed in the study advances the understanding of the application of NLP techniques grounded in semantic similarity, particularly in the context of ranking candidates during the interview process, and provides human resource management and the recruitment sector, particularly automated recruitment, with an increase in the AI-centered tools available.

2. Methods

The development and assessment of an NLP-based intelligent recruitment system for ranking interview candidates have been carried out with an experimental and quantitative approach in this case. Current methodologies make it possible to analyze the open-ended written answers of candidates and quantify their semantic relevance to the respective job descriptions in an attempt to rank candidates objectively. The proposed system workflow includes data collection, pre-processing of text, semantic vector generation, calculation of semantic similarity to metrics, and assessment, ranking, and evaluation of candidates.

2.1. Data Collection and Dataset Construction

Data for this study was collected through an online recruitment procedure. Every candidate's response to a recruitment-related question is recorded. Each candidate has a unique candidate ID and is allocated responses to a series of open-ended interview questions. The question consists of key elements of recruitment that include, but are not limited to the following:

a. The candidate's work experience and background relevant to the position being applied for.
b. The candidate's understanding of job roles and responsibilities.
c. Motivation of the candidate to apply for a job.
d. Candidate's career ambitions and self-development.

Along with the responses from the candidates, the job description suitable for the position applied for is also collected and considered as a reference text (Fisher et al., 2021). The resulting dataset consists of paired text documents, candidate responses and job descriptions that are appropriate for semantic comparison purposes (Ajjam & Al-Raweshidy, 2025).

Table 1. Dataset collection by form

candidate_id	Q1	Q2	Q3	Q4	Q5
C001	I work in administration!	I want to try a new experience	This position is related to computers	I want to learn	I need basic training
C002	I have non-technical	I am interested in	I think this position	I want to	I need guidance

candidate_id	Q1	Q2	Q3	Q4	Q5
	experience!	working at an institution	is operational	improve soft skills.	
C003	I once worked in marketing.	I want a new challenge	This position is related to technology	I want to learn new things	I need training
C004	I have no IT experience	I want to work	This position is related to computers	I want to learn	I need training
C005	I work in the service sector	I want to change careers	This position is related to administration	I want to gain experience	I need mentoring.
C006	I have experience in a warehouse	I want a more stable job	This position is about data entry	I want to be more productive	I need training
C007	I worked in a shop	I want to work in an office	This position is related to computers.	I want to learn IT	I need a guide
C008	I have non-IT experience.	I want to develop myself.	This position is helpful for the team.	I want to achieve targets.	I need support.
C009	I am a fresh graduate (non-IT).	I want to start a career.	This position is an initial step.	I want to learn.	I need direction.
C010	I have experience in sales.	I want a new atmosphere.	This position is support-oriented.	I want to contribute.	I need training.
C011	I have basic computer knowledge.	I want to explore the IT field.	This position is about computer maintenance.	I want to be an expert.	I need access to materials.
….	….	….	….	…..	….
C100	I once helped with school administration.	I want to work in education.	This position helps students.	I want to be more organized.	I need a laptop.

2.2. Text Pre-processing

Candidate responses and related job descriptions undergo similar processing steps to achieve text uniformity, as well as reduce textual data noise (Hidayat et al., 2015; Patel & Thakkar, 2014; Rajesh et al., 2025). These steps include normalization of text (converting to lowercase), removal of punctuation and special characters, and tokenization (Nazir et al., 2024). Words are also lemmatized to enhance the text semantically and computationally to build better text embedding (Szymański et al., 2024). Building text embedding involves reducing large syntactic variations, and retaining most of the important semantic variations. The steps shown on Figure 1.

2.3. Semantic Representation Using NLP Models

Contextual semantic models aim to understand the meaning behind processed words. Using NLP semantic model embedding, text from a given candidate's responses and job descriptions is converted into a solid vector representation (Giachos et al., 2017; Naresh Kumar & Uma, 2021; Patil et al., 2023; Vanetik & Kogan, 2023). These vectors, generated through sentence-level encoding, are semantically embedded to describe meaning in the text.

The various text responses of the candidate are encoded separately, meaning that the responses are processed individually, and then, through the incorporation of the response text, are summed up to form a holistic representation of the candidate. Through this, the system can capture and reflect various evaluative dimensions – experience, motivation, and role understanding – into the evaluative system. The average confidence of keywords shown on Table 2.

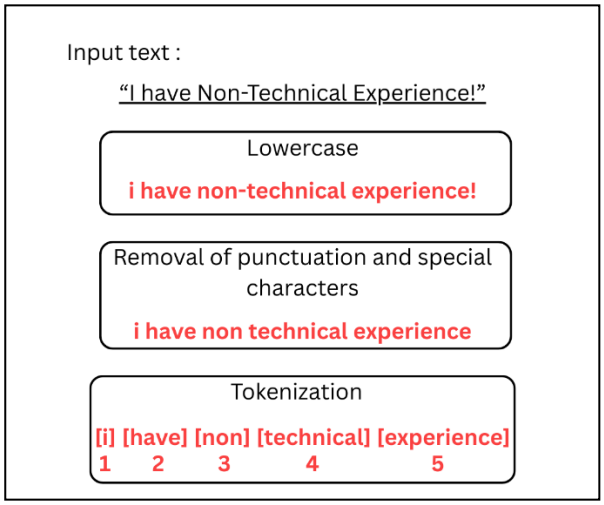

Figure 1. Steps in Preprocessing

Table 2. Average confidence of keywords

Keywords	Confidence
work	0.99
technology	0.99
support	0.98
maintain	0.95
reporting	0.95
digital	0.95
improvement	0.95
report	0.95
contribute	0.95
minor	0.95
damage	0.95
troubleshooting	0.95
facilities	0.95
handle	0.95
appropriate	0.95
network	0.95
documentation	0.95
expertise	0.95
guidance	0.94
capability	0.94

2.4. Similarity Calculation

The candidates' responses were compared with the job description to establish semantic similarities using cosine similarities (Ajjam & Al-Raweshidy, 2025). This is done by measuring the angular distance between the vectors. Given its efficiency in embedding calculations in large dimensions, as well as its ability to withstand variations in text length, cosine similarity is preferred (Azhari & Hernandez, 2016; Zhou et al., 2022). For each candidate, the similarity score for each response embedding to the job description embedding is calculated. The higher the similarity score, the better the candidate's response is in line with the job description (Figure 2).

Job Description : "Handle server infrastructure and network security"	
Text	Similarity
Candidate A: "Handle network firewall configuration "	**0.91** ✅
Candidate B: "Handle administrative documentation and office equipment maintenance"	**0.38** ❌

Figure 2. Similarity on Job Description

2.5. Candidate Assessment and Ranking Mechanism

To reach the final ranking, the similarity scores for a number of interview questions must be combined to form a candidate score (Tao, 2024). This is done through a weighted or unweighted incorporation scheme, depending on how significant each evaluative criterion is. The final score indicates the semantic proximity of the candidate's profile to the job description. The candidates are ranked in descending order based on scores, with candidates with higher scores more likely to get the job.

2.6. System Evaluation

The evaluation of the proposed recruitment system depends on the rating that the system generates against the assessment given by the human recruiter (Pathak & Pandey, 2025). The rating system and the rating of experts are correlated using evaluation metrics such as the Spearman rating correlation coefficient. In addition, in positive analysis, we explore situations where the semantic similarities of candidates, as captured by the system, overlap with relevant information that the system did not identify using a keyword filter. The proposed framework underscores the system's ability to foster objectivity and scalability, thereby providing robust support for strategic hiring selections.

3. Results and Discussion

The study introduces a specialized pipeline designed for the automated evaluation and ranking of job applicants. At the core of this system, IndoBERT (indobenchmark/indobert-base-p1) serves as the primary mechanism for semantic feature extraction, while Logistic Regression is employed as the classification head to categorize candidates (Tandi et al., 2025). The experimental phase utilized a curated dataset of candidate responses, consisting of 20 labeled samples distributed across three distinct suitability tiers: High, Medium, and Low. To ensure the integrity of the results and maintain a balanced representation of these classes, the dataset was partitioned using a stratified 80/20 train-test split.

The technical architecture specifically leverages the [CLS] token embedding derived from IndoBERT's final hidden state. This process generates a 768-dimensional vector representation that encapsulates rich, contextual, and semantic nuances from the concatenated text of the candidates' responses. By shifting away from traditional, rigid keyword-matching techniques, this method focuses on encoding the latent semantic correlations between a candidate's professional qualifications and the specific demands of a job role. Consequently, the system facilitates a more sophisticated and multidimensional profiling of potential hires.

3.1. Classification Performance

The classification performance of the trained model (Table 3), evaluated on a held-out test set (n = 20), is detailed in the following report. The model demonstrated robust predictive power, achieving an overall accuracy of 85%, complemented by a macro-averaged F1-score of 0.8 and a weighted-average F1-score of 0.84. These metrics suggest a significant capacity for generalization, even when considering the constraints of the dataset size. Nevertheless, while these preliminary results are promising, they should be interpreted with academic caution due to the relatively small sample population utilized in this study. The figure 3 showed the confusion matrix on data test.

Table 3. Classification performance

Class	Precision	Recall	F1-Score	Support
High	0.8	1	0.88	4
Low	0.87	0.87	0.87	8
Medium	0.85	0.75	0.80	8
Accuracy			0.85	20
Macro Avg	0.84	0.87	0.85	20
Weight Avg	0.85	0.85	0.84	20

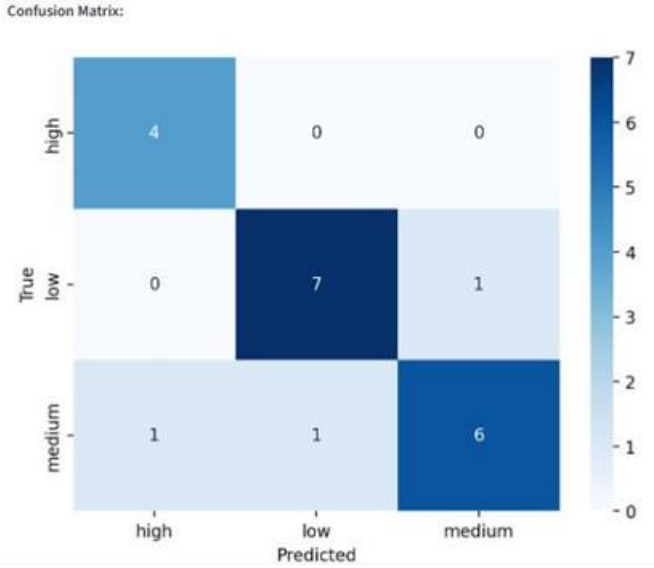

Figure 3. Confusion Matrix on Data Test

3.2. High-Suitability Candidates

In terms of specific class performance, the High category exhibited the most optimal recall, attaining a perfect score of 1.0. This result confirms the model's efficacy in successfully identifying every ground-truth “High-suitability” candidate within the test set, with zero omissions. Conversely, the recorded precision of 0.8 implies that a small fraction of candidates classified as High actually belonged to adjacent categories, resulting in a 20% false positive rate. This leads to an F1-score of 0.8, illustrating a robust yet slightly asymmetric balance between precision and recall.

The achievement of perfect recall in this specific tier holds substantial weight within a recruitment framework. From an organizational standpoint, the cost associated with overlooking a high-potential talent (false negative) typically outweighs the inconvenience of advancing a borderline candidate to the next stage of evaluation (false positive) (Frazzetto et al., 2025). Consequently, the model’s inclination toward inclusivity for this category is viewed as a beneficial characteristic, aligning with the strategic priorities of talent acquisition to ensure no top-tier prospects are prematurely excluded.

3.3. Low-Suitability Candidates

The Low suitability category demonstrated the most cohesive performance, where precision and recall converged at 0.8750, resulting in a balanced F1-score of 0.87. This statistical symmetry suggests that the model maintains a remarkably consistent and unbiased decision boundary when identifying candidates with lower job alignment. It appears that the IndoBERT embeddings effectively capture discriminative features that isolate these profiles from other tiers. This is particularly evident when candidate narratives exhibit a noticeable absence of domain-specific technical terminology or convey a restricted motivation toward professional growth, allowing the model to distinguish these responses with high reliability.

3.4. Medium-Suitability Candidates

Analysis of the Medium category revealed the least robust performance among the three classifications, yielding a precision of 0.8571 and a recall of 0.7500, with a corresponding F1-score of 0.80. The diminished recall indicates that approximately one-quarter of the true "Medium-suitability" candidates were misattributed, likely being absorbed into the adjacent High or Low categories. This observation is consistent with the inherent semantic ambiguity characteristic of the "Medium" tier, which, by its very nature, exists within an intermediate decision space—trapped between the more definitive markers of highly qualified and clearly unqualified candidates.

The difficulty in accurately classifying these candidates is further exacerbated by the overlapping semantic signatures found in their responses. For instance, a candidate possessing fragmented technical expertise paired with a strong motivation to learn generates embeddings that may gravitate toward either the High or Low boundaries within the latent representation space. This phenomenon aligns with existing literature on intermediate-class classification in NLP-driven HR systems, which identifies boundary ambiguity as a persistent challenge for embedding-based methodologies.

3.5. Comparative Analysis with Baseline Models

To provide a clearer context for the proposed model's efficacy, Table 3 offers a comparative evaluation against two established baselines. These include: (1) a lexical baseline utilizing TF-IDF paired with Logistic Regression, and (2) a semantic baseline employing Word2Vec in conjunction with a Support Vector Machine (SVM). To ensure a rigorous and unbiased comparison, all models were benchmarked using identical data partitions and subjected to strictly equivalent experimental parameters.

Table 4. Classification performance

Model	Accuracy	F1-Score (Macro)	Notes
TF-IDF + Logistic Regression	0.75	0.73	Baseline model
Word2Vec + SVM	0.78	0.76	Semantic baseline
IndoBERT + Logistic Regression (Proposed)	0.85	0.85	Best performance

The proposed IndoBERT-based model consistently exceeds the performance of both established baselines across every reported metric. Specifically, it demonstrates a substantial 10-percentage-point increase in accuracy over the TF-IDF baseline (0.75 → 0.85), while the macro F1-score reflects an even more pronounced improvement of 12.26 points (0.73 → 0.85). Although the margin of superiority over the Word2Vec + SVM baseline is more incremental, it remains significant, with consistent gains of 7 percentage points in accuracy and 9.36 points in the macro F1-score.

These performance gains can be primarily attributed to IndoBERT's extensive pre-training on large-scale Indonesian corpora, which equips the model with the ability to discern the intricate morphological, syntactic, and semantic nuances inherent to the Indonesian language (Table 4). In contrast to static word embeddings like Word2Vec which assign a fixed vector to each term IndoBERT generates contextualized representations that dynamically adjust based on the surrounding linguistic environment of each token. This adaptability results in the extraction of more discriminative, sentence-level features, ultimately enhancing the precision of the downstream classification process.

3.6. Implications for Recruitment Practice

The experimental findings validate the efficacy of IndoBERT-based semantic classification as a robust initial screening mechanism, particularly for specialized technical roles such as Laboratory Head positions. The system exhibits a sophisticated discriminative capacity, enabling the automated filtering of extensive applicant pools. This capability significantly alleviates the administrative burden on human resource departments while ensuring high recall for "High-suitability" candidates a critical factor in identifying top-tier talent early in the recruitment cycle.

Nevertheless, these results also emphasize the indispensable role of human oversight during the final stages of selection, especially for applicants categorized within the "Medium" suitability bracket. Given the inherent semantic ambiguity of this tier and the model's sensitivity to motivational discourse, automated scores should be utilized as a prioritized shortlist rather than a definitive selection tool. Consequently, a hybrid workflow where NLP-generated insights inform and guide, but do not supersede, professional recruiter judgment is recommended for practical implementation.

Furthermore, the model's current performance regarding candidates from non-IT backgrounds who demonstrate strong learning potential (recall: 0.61) indicates a potential bias against growth-oriented individuals. This suggests that the system may inadvertently undervalue applicants whose strengths lie in their future trajectory rather than current technical alignment. Future iterations of this framework should, therefore, integrate a dedicated potential assessment module. Such a component would explicitly weight transferable skills and learning curves alongside established technical competencies to ensure a more holistic and equitable evaluation.

4. Conclusion

Several constraints must be considered when interpreting the results of this study. Primarily, the reliance on a dataset of only 20 labeled samples significantly limits the statistical power of the evaluation metrics, making performance estimates highly susceptible to individual sample variance. Consequently, the generalizability of these findings to broader, real-world recruitment environments requires further empirical validation.

Furthermore, the inherent class imbalance comprising High (n=4), Medium (n=8), and Low (n=8) categories introduces an asymmetric evaluation landscape. This is particularly evident in the High category, where the limited support size means that precision and recall are disproportionately affected by a single misclassification. While a balanced class-weighting strategy was implemented during the Logistic Regression phase to mitigate this, future research should explore more robust data augmentation techniques, such as the Synthetic Minority Over-sampling Technique (SMOTE). Additionally, as the evaluation focused exclusively on Indonesian-language responses using IndoBERT, the model's efficacy in multilingual contexts specifically involving code-switching (Indonesian-English) or highly formal academic prose remains to be systematically determined.

Building upon these findings, several strategic pathways for future inquiry are proposed. Expanding the corpus to a minimum of 100 labeled samples per category would substantially enhance model reliability and facilitate the transition to more sophisticated architectures, such as end-to-end IndoBERT fine-tuning.

The integration of Named Entity Recognition (NER) also presents a significant opportunity to automate the extraction of key candidate attributes, including institutional background, job titles, and specific technology stacks. Such a multi-modal extension—combining structured resume data with unstructured textual responses would provide a more granular and holistic feature representation.

Finally, deploying the system within an active recruitment pipeline would allow for continuous feedback loops from HR practitioners. This real-world data could be used to iteratively refine the model's decision boundaries, ensuring closer alignment with organizational standards and a progressive reduction in systematic bias.

In summary, the proposed IndoBERT + Logistic Regression framework achieved an accuracy of 85% and a macro F1-score of 0.85, effectively outperforming both TF-IDF and Word2Vec baselines. The attainment of perfect recall for high-suitability candidates underscores the system's practical utility as an automated first-stage screening tool. While challenges such as small sample size and "Medium-class" ambiguity persist, this research establishes a foundational step toward more objective, NLP-driven talent acquisition.

Acknowledgements: This article used DeepL AI Translator for translation tool from Indonesian language to English, also used ChatGPT 5.2 for improved readiness content to make it language more easily to understand.

References

Agbasiere, C. L., & Nze-Igwe, G. R. (2025). Algorithmic Fairness in Recruitment: Designing AI-Powered Hiring Tools to Identify and Reduce Biases in Candidate Selection. *Path of Science*, *11*(4), 5001. https://doi.org/10.22178/pos.116-10

Ajjam, M.-H., & Al-Raweshidy, H. S. (2025). AI-driven semantic similarity-based job matching framework for recruitment systems. *Information Sciences*, *724*, 122728. https://doi.org/10.1016/j.ins.2025.122728

Alsaif, S. A., Hidri, M. S., Ferjani, I., Eleraky, H. A., & Hidri, A. (2022). NLP-Based Bi-Directional Recommendation System: Towards Recommending Jobs to Job Seekers and Resumes to Recruiters. *Big Data and Cognitive Computing*, *6*(4), 147. https://doi.org/10.3390/bdcc6040147

Azhari, A., & Hernandez, L. (2016). Brainwaves feature classification by applying K-Means clustering using single-sensor EEG. *International Journal of Advances in Intelligent Informatics*, *2*(3), 167–173. https://doi.org/10.26555/ijain.v2i3.86

Devi, S., & Anand, G. P. (2025). *Technical Skills Information Extraction From Resumes Using Advanced Natural Language Processing Model With Transfer Learning*. https://doi.org/10.21203/rs.3.rs-7744440/v1

Fisher, E., Thomas, R. S., Higgins, M. K., Williams, C. J., Choi, I., & McCauley, L. A. (2021). Finding the right candidate: Developing hiring guidelines for screening applicants for clinical research coordinator positions. *Journal of Clinical and Translational Science*, *6*(1).

Frazzetto, P., Muhammad, Fabris, F., & Sperduti, A. (2025). Graph Neural Networks for Candidate-Job Matching: An Inductive Learning Approach. *Data Science and Engineering*. https://doi.org/10.1007/s41019-025-00293-y

García-Díaz, V., Espada, J. P., Crespo, R. G., Pelayo G-Bustelo, B. C., & Cueva Lovelle, J. M. (2018). An approach to improve the accuracy of probabilistic classifiers for decision support systems in sentiment analysis. *Applied Soft Computing Journal*, *67*, 822–833. https://doi.org/10.1016/j.asoc.2017.05.038

Giachos, I., Papakitsos, E. C., & Chorozoglou, G. (2017). Exploring natural language understanding in robotic interfaces. *International Journal of Advances in Intelligent Informatics*, *3*(1), 10–19. https://doi.org/10.26555/ijain.v3i1.81

Gilch, P. M., & Sieweke, J. (2020). Recruiting Digital talent: the Strategic Role of Recruitment in Organisations' Digital Transformation. *German Journal of Human Resource Management: Zeitschrift Für Personalforschung*, *35*(1), 53–82. https://doi.org/10.1177/2397002220952734

Hangartner, D., Kopp, D., & Siegenthaler, M. (2021). Monitoring hiring discrimination through online recruitment platforms. *Nature*, *589*(7843), 572–576. https://doi.org/10.1038/s41586-020-03136-0

Hidayat, E. Y., Firdausillah, F., Hastuti, K., Dewi, I. N., & Azhari, A. (2015). Automatic Text Summarization Using Latent Drichlet Allocation (LDA) for Document Clustering. *International Journal of Advances in Intelligent Informatics*, *1*(3), 132–139. https://doi.org/10.26555/ijain.v1i3.43

Hryn, V. (2025). Implementation of CRM in Mass Recruitment: Efficiency and Impact on Employee Turnover. *Social Development: Economic and Legal Issues*, (6). https://doi.org/10.70651/3083-6018/2025.6.13

Huaze Xie; Mohd Anuaruddin Bin Ahmadon; Shingo Yamaguchi; Ichiro Toyoshima. (2019). Random Sampling and Inductive Ability Evaluation of Word Embedding in Medical Literature. *2019 IEEE International Conference on Consumer Electronics (ICCE)*. https://doi.org/10.1109/ICCE.2019.8662022

Konstantinidis, Maragoudakis, M., Magnisalis, I., Berberidis, C., & Peristeras, V. (2022). Knowledge-driven Unsupervised Skills Extraction for Graph-based Talent Matching. *Zenodo (CERN European Organization for Nuclear Research)*, 1–7. https://doi.org/10.1145/3549737.3549769

Mashayekhi, Y., Li, N., Kang, B., Lijffijt, J., & De Bie, T. (2024). A challenge-based survey of e-recruitment recommendation systems. *ACM Computing Surveys*, *56*(10). https://doi.org/10.1145/3659942

Mujtaba, D. F., & Mahapatra, N. R. (2024). Fairness in AI-Driven Recruitment: Challenges, Metrics, Methods, and Future Directions. In *arXiv*. https://doi.org/10.48550/arXiv.2405.19699

Nabil Rakha Dwitya, & Istiono, W. (2023). Consumer Segmentation of Emina Cosmetics Optimal and Relevant Approach of RFM+Lifetime Analysis. *SAGA: Journal of Technology and Information System*, *1*(3), 79–89. https://doi.org/10.58905/saga.v1i3.171

Naresh Kumar, K. E., & Uma, V. (2021). Intelligent sentinet-based lexicon for context-aware sentiment analysis: optimized neural network for sentiment classification on social media. *Journal of Supercomputing*, *77*(11), 12801–12825. https://doi.org/10.1007/s11227-021-03709-4

Nazir, S., Asif, M., Rehman, M., & Ahmad, S. (2024). Machine learning based framework for fine-grained word segmentation and enhanced text normalization for low resourced language. *PeerJ Computer Science*, *10*, e1704–e1704. https://doi.org/10.7717/peerj-cs.1704

Patel, U., & Thakkar, M. (2014). *An Efficient Exact Single Pattern Matching Algorithm*. *3*(5), 1955–1958.

Pathak, G., & Pandey, D. (2025). *AI Agents in Recruitment: A Multi-Agent System for Interview, Evaluation, and*

Candidate Scoring. https://doi.org/10.2139/ssrn.5242372

Patil, R., Boit, S., Gudivada, V., & Nandigam, J. (2023). A Survey of Text Representation and Embedding Techniques in NLP. *IEEE Access*, *11*, 36120–36146. https://doi.org/10.1109/access.2023.3266377

Rajesh, S. G., Madangarli, S. V., Pisharady, G. S., & Subrahmanyam, R. (2025). Enhancement of Virtual Assistants through MultiModal AI for Emotion Recognition. *IEEE Access*, 1. https://doi.org/10.1109/access.2025.3577664

Saatci, M., Kaya, R., & Ünlü, R. (2024). Resume Screening With Natural Language Processing (NLP). *Alphanumeric Journal*. https://doi.org/10.17093/alphanumeric.1536577

Szymański, J., Operlejn, M., & Weichbroth, P. (2024). Enhancing Word Embeddings for Improved Semantic Alignment. *Applied Sciences*, *14*(24), 11519. https://doi.org/10.3390/app142411519

Tandi, T. Y., Abidin, T. F., & Riza, H. (2025). Incorporation of IndoBERT and Machine Learning Features to Improve the Performance of Indonesian Textual Entailment Recognition. *Journal of Information Systems Engineering and Business Intelligence*, *11*(2), 173–186. https://doi.org/10.20473/jisebi.11.2.173-186

Tao, C. (2024). LLMs are Also Effective Embedding Models: An In-depth Overview. In *arXiv*. https://doi.org/10.48550/arXiv.2412.12591

Vanetik, N., & Kogan, G. (2023). Job Vacancy Ranking with Sentence Embeddings, Keywords, and Named Entities. *Information*, *14*(8), 468. https://doi.org/10.3390/info14080468

Zhou, K., Ethayarajh, K., Card, D., & Jurafsky, D. (2022, May). Problems with Cosine as a Measure of Embedding Similarity for High Frequency Words. *ACLWeb*.