DOI: 10.15849/IJASCA.251130.16

Int. J. Advance Soft Compu. Appl, Vol. 17, No. 3, November 2025
Print ISSN: 2710-1274, Online ISSN: 2074-8523

Illuminating Summarization Efficacy: A Qualitative Analysis of Resume Condensation for Job Category Classification

Mohammed Maree

Department of Information Technology, Arab American University, Jenin, Palestine
e-mail: mohammed.maree@aaup.edu
ORCID: https://orcid.org/0000-0002-6114-4687

Abstract

Automatic text summarization is increasingly explored to manage the large influx of resumes in recruitment, potentially aiding downstream tasks like job category classification. While quantitative evaluations offer insights into summarizer performance, they often do not fully explain why certain methods excel or falter in preserving category-distinguishing information. This paper presents an in-depth qualitative analysis of summaries generated by representative extractive (Reduction, Random) and abstractive (BART, PEGASUS) techniques for resume analysis. We meticulously examine how well crucial information such as specific skills, job titles, and industry-related terms is retained or lost in the generated summaries, and how these characteristics correlate with the performance of a downstream job category classifier. Through illustrative examples and analysis of misclassification patterns, our findings reveal distinct strategies of information preservation and common pitfalls, offering a nuanced understanding beyond aggregate metrics. This qualitative perspective underscores the importance of information fidelity for specific downstream applications and provides actionable insights for selecting or developing summarization tools for effective resume processing in e-recruitment systems.

Keywords: *E-recruitment, extrinsic evaluation, information retrieval, natural language processing, qualitative analysis, resume analysis, text summarization.*

1 Introduction

The proliferation of online job portals has led to an overwhelming volume of resume submissions, posing a significant challenge for human resource departments [1], [11]. Manual screening is time-consuming, costly, and prone to human bias [2], [12]. Automated e-recruitment systems [13], often employing text summarization, aim to streamline this process by condensing resume content [3], [14]. These summaries can then serve as input for downstream tasks such as candidate screening or job category classification [4], [15].

Text summarization techniques broadly fall into two categories: extractive methods, which select salient sentences directly from the source text, and abstractive methods, which generate novel sentences to convey the core meaning [5], [16]. Modern abstractive models are often based on the powerful transformer architecture [17]. Previous work has quantitatively evaluated various summarization techniques for resumes using downstream

Received 1 September 2025; Accepted 1 November 2025

classification tasks [6, 30], highlighting performance differences in terms of accuracy and efficiency. For instance, certain extractive methods and some transformer-based abstractive models like BART showed promise, while others, notably PEGASUS pre-trained on general news, struggled to preserve critical information, a finding consistent with observations in similar studies [6, 30].

However, aggregate metrics such as F1-score or accuracy, while useful, do not fully elucidate the underlying reasons for these performance disparities. A crucial question remains: how do different summarization approaches affect the retention of category-distinguishing information within resumes? While quantitative studies tell us which models perform well, they do not explain why. This explanatory gap is the primary motivation for our work. This paper, therefore, seeks to answer: How do different summarization paradigms qualitatively alter the informational content of resumes, and how do these alterations directly impact the accuracy and error patterns of a downstream classification task?

To address this question, this study makes the following contributions:

- Demonstration of Information Fidelity as the Primary Determinant of Success: We establish a direct qualitative link between the preservation of specific entities (job titles, skills) and the performance of a downstream classifier, providing qualitative evidence consistent with plausible causal mechanisms.
- Identification of Failure Modes in Abstractive Summarization: We provide a clear diagnostic analysis of why a domain-mismatched abstractive model (PEGASUS) fails catastrophically, attributing it to over-abstraction that erodes category-distinguishing information—a critical insight for practitioners.
- Characterization of Summarizer-Specific Error Patterns: We illustrate how different summarization strategies (e.g., extractive vs. abstractive) lead to distinct and predictable error patterns in classification, providing a nuanced understanding beyond aggregate accuracy scores.
- A Framework for Qualitative, Extrinsic Evaluation: We present and apply a methodology for evaluating summarizers that prioritizes their utility in a specific application, offering actionable guidance for selecting models for real-world e-recruitment systems.

By moving beyond purely quantitative assessments, we aim to provide a richer understanding of the interplay between summarization strategies and their utility in real-world resume processing scenarios.

The rest of this article is organized as follows. Section 2 reviews the related literature on automated resume analysis, text summarization in recruitment, and evaluation methodologies. Section 3 details our methodology, including the dataset, the selection of summarization models for analysis, and the qualitative evaluation procedure. In Section 4, we present our qualitative findings, providing an in-depth analysis of the summaries generated by each model. Section 5 discusses the implications of these findings, highlighting key patterns and practical takeaways for e-recruitment systems. Finally, Section 6 concludes the paper with a summary of our contributions and suggestions for future research.

2 Related Work

Our research is situated at the confluence of three key domains: automated resume analysis, the application of text summarization in recruitment, and methodologies for NLP model evaluation.

2.1 Automated Resume Analysis and Screening

The automation of resume analysis has evolved from rudimentary keyword matching to NLP-driven pipelines that parse, represent, and classify full documents. Early work emphasized information extraction using rule-based or statistical techniques to pull structured fields (e.g., contact details, education, work history) from free text [26]. At the classification stage, representing the entire resume—rather than only extracted fields—has proven important for capturing role-specific semantics; for example, Zaroor et al. [1] highlight challenges and design considerations for job–resume classification at scale.

Recent work in representation learning and search has been complemented by feature selection (FS) and optimization methods that improve downstream model efficiency and generalization. In particular, Mustafa et al. introduce the Chernobyl Disaster Optimizer (CDO), a meta-heuristic designed for feature selection across high-dimensional problems, reporting competitive selection of informative subsets for classifiers [28]. While CDO is not resume-specific, such FS techniques are directly applicable to resume pipelines that rely on high-dimensional lexical or embedding features, where reducing redundancy can speed training/inference and mitigate overfitting. In our study, we similarly target efficient downstream processing, but we focus on summarization as a content-level reduction step: instead of selecting features after vectorization, we investigate whether condensing the text itself can preserve category-diagnostic content while reducing noise for the classifier.

Our contribution therefore complements prior resume analytics by examining summarization not as an end in itself, but as a strategic preprocessing stage that trades length for information fidelity—crucial for job-category prediction.

2.2 Text Summarization in Recruitment

Applying text summarization to resumes aims to reduce recruiter workload and accelerate automated screening without sacrificing essential signals. Both extractive and abstractive paradigms have been explored. Zaroor et al. [1] provide an early comparative perspective showing the viability of summarization in recruitment contexts. With the advent of Transformer models, more recent efforts have examined modern abstractive systems: for instance, Mercan et al. [2] evaluate BART/T5-style models for resume summarization, and general abstractive frameworks based on deep learning and semantic content control have been studied in broader settings [27]. Despite encouraging aggregate metrics, an enduring gap is qualitative understanding—which pieces of category-distinguishing information are preserved or lost, and why particular models fail in downstream tasks.

Complementary lines of research outside summarization can still inform recruitment workflows. Mustafa et al. propose a multi-objective memetic differential evolution method for text clustering, optimizing multiple criteria to improve cluster quality [29]. While clustering is unsupervised and not a summarization technique, its objectives—grouping documents by latent structure—are relevant for pre-clustering resumes, organizing skill taxonomies, or sampling representative documents for human review. Such clustering can be paired with summarization: clusters help discover structure and coverage, while summaries make cluster exemplars digestible for both humans and simple classifiers. Positioned against this landscape, our work contributes a qualitative, extrinsic evaluation of representative extractive and abstractive summarizers on resumes, explaining failure modes (e.g., over-abstraction) and linking information fidelity to job-category classification performance.

2.3 Evaluation of Text Summarization

The evaluation of summarization systems is traditionally bifurcated into intrinsic and extrinsic methods [16]. Intrinsic evaluation assesses the inherent quality of a summary (e.g., fluency, coherence) and its faithfulness to a human-written "gold standard" summary, often using n-gram overlap metrics like ROUGE [23]. While useful for general-purpose summarization research, this approach is often impractical in specialized domains like recruitment, where reference summaries are scarce or non-existent. More importantly, intrinsic scores may not correlate with a summary's usefulness for a specific goal [24]. Extrinsic evaluation resolves this by measuring a summary's utility on a downstream task, such as classification, question answering, or information retrieval [25]. This task-based paradigm has gained prominence as it provides a practical, goal-oriented measure of a model's real-world value. Our study wholly embraces an extrinsic and qualitative evaluation framework. By focusing exclusively on the impact of summarization on a downstream classification task, we assess the models based on their direct utility. Our qualitative deep dive complements this by dissecting the error patterns, providing a richer, more practical assessment of model suitability for deployment in high-stakes e-recruitment pipelines.

3 Problem Formulations and Methodology

3.1 Dataset and Preprocessing

The study utilizes a dataset comprising 2484 resumes, sourced from Kaggle [10]. While the original dataset provides resumes in raw PDF format, this research used the pre-extracted text available in the accompanying CSV file as the primary input for all summarization methods. This decision ensures consistency in the source text provided to each summarizer, removing variability that could arise from different PDF-to-text extraction processes. The CSV file contains columns for a unique ID, the resume text (Resume_str), and the ground truth category (Category), with 24 distinct job category labels (e.g., HR, Finance, IT, Sales). A sample structure is shown in Table 1. The absence of human-authored reference summaries for these resumes motivates the primary focus on extrinsic, task-based evaluation.

Table 1: Sample structure of the resume dataset CSV file.

ID	Resume_str (Excerpt)	Category
16852973	HR ASSOCIATE Summary HR Administrator...	HR
22323967	HR SPECIALIST Summary Dedicated...	HR
11847784	HR ASSISTANT Summary Resourceful...	HR
...	...	...

3.2 Underlying Quantitative Study Context

Our qualitative analysis builds upon experimental frameworks and quantitative findings from prior work on extrinsic summarization evaluation [6, 30]. For context, the underlying quantitative stage informing this work evaluated seven extractive methods (including classical approaches like Luhn [18] and graph-based methods like LexRank [19] and TextRank [20]) and five abstractive, transformer-based methods (such as BART [8], PEGASUS [9], and variants of T5 [21]). The primary extrinsic evaluation involved training a TF-IDF and Logistic Regression classifier to predict the job category based solely on the generated summary. To formalize this process, Algorithm 1 outlines the standardized extrinsic evaluation pipeline used to generate such quantitative results. This pipeline systematically assesses the utility of summaries produced by different models for the downstream classification task. For each summarization model, the algorithm first generates a complete set of summaries from the original resumes. These summaries are then converted into numerical feature vectors using the TF-IDF representation. Finally, a Logistic Regression classifier is trained and evaluated on these vectors, yielding a performance score (e.g., F1-score) that quantifies how well the information preserved in the summaries serves the classification goal. This systematic approach produces the quantitative scores (as seen in Table 2) that form the basis for our subsequent qualitative investigation into why certain models perform better than others.

Algorithm 1: An Exemplar Extrinsic Evaluation Pipeline

Input: Resumes $R = \{r_1, r_2, ..., r_n\}$

Input: Summarizer Models $M = \{m_1, ..., m_k\}$

Initialize: Results dictionary 'Perf'

for each model m_j in M do

  Initialize summary set $S_j = \emptyset$

  for each resume r_i in R do

    Generate summary $s_i = m_j(r_i)$

    Add s_i to S_j

  end for

  Vectorize all summaries in S_j using TF-IDF

  Split vectorized summaries into Train/Test sets

  Train Logistic Regression classifier on Train set

  Evaluate classifier on Test set to get performance p_j

  Store p_j in 'Perf' for model m_j

end for

Output: 'Perf' containing performance for each model

To provide a concrete basis for our qualitative analysis, Table 2 presents illustrative quantitative results, reflecting typical performance patterns observed in such studies [6, 30]. Key findings indicate that extractive methods like Reduction and the abstractive BART model often achieve high accuracy, while models like PEGASUS (XSum) perform poorly.

Table 2: Illustrative Quantitative Results for Downstream Classification.

Input for Classifier	F1-Score (Macro)	Accuracy
Full Resume (No Summary)	0.96	96.2%
Reduction Summary	0.94	94.5%
BART Summary	0.93	93.8%
Random Summary	0.45	46.1%
PEGASUS (XSum) Summary	0.12	15.3%

3.3 Selection of Summarization Models for Qualitative Analysis

For this qualitative analysis, we selected a representative subset of models that reflect the performance spectrum observed in quantitative studies [6, 30] to cover a range of performance and approaches:

- **Reduction:** A graph-based extractive method from the 'sumy' library [7], consistently among the top performers in classification accuracy.
- **BART (facebook/bart-large-cnn):** An abstractive transformer model [8], which also performed well in the classification task.
- **PEGASUS (google/pegasus-xsum):** An abstractive transformer model [9] specifically pre-trained for summarization, but which showed very low classification accuracy on resume summaries.
- **Random:** An extractive baseline that selects sentences randomly, serving as a lower-bound reference.

Summaries for extractive methods were typically set to 3 sentences, while abstractive summaries were generated with a maximum length of 150 tokens, consistent with common experimental setups. The selection is designed to cover the full spectrum of performance, from top-tier to baseline, as summarized in Table 3.

Table 3: Performance Tiers of Selected Models for Qualitative Analysis.

Model	Type	Tier	Rationale
Reduction	Extractive	High	Represents top-performing, computationally efficient extractive methods.
BART	Abstractive	High	Represents effective, modern abstractive approaches.
Random	Extractive	Low	Serves as a simple, non-semantic baseline for comparison.
PEGASUS	Abstractive	Very Low	Illustrates pitfalls of domain mismatch in abstractive models.

3.4 Qualitative Analysis Procedure

Our qualitative analysis involved the following steps:

1. **Sample Selection:** We randomly sampled a subset of 20 resumes from various job categories where classification outcomes diverged across models. For each sampled resume, we examined the summaries generated by the four selected models, focusing on instances where the classifier made correct predictions for some summaries and incorrect ones for others derived from the same source.
2. **Content Review:** Each summary was manually reviewed against its original resume text. We focused on:
    - **Keyword Preservation:** Presence or absence of critical job titles (e.g., "Software Engineer," "HR Manager"), technical skills (e.g., "Java," "SQL," "Agile"), soft skills, and industry-specific jargon.
    - **Information Distortion/Loss:** Identification of instances where key details were omitted, overly generalized, or (in the case of abstractive models) potentially hallucinated or misattributed.

- **Clarity and Coherence:** General readability, although this was secondary to information preservation for the classification task.

3. **Correlation with Classification Outcome:** We linked the observed summary characteristics to the success or failure of the downstream job category classifier. For misclassified summaries, we sought to understand how the summary's content (or lack thereof) might have led the classifier astray. For instance, if an IT resume summary omitted all technical keywords, it might be misclassified as a generic administrative role.
4. **Pattern Identification:** Across multiple samples, we looked for recurring patterns in how each summarization method handled resume content and how these patterns related to overall classification performance trends observed in similar quantitative evaluations.

3.5 Downstream Classification Task Formalization

The extrinsic evaluation pipeline can be formally described. For each summary, a feature vector was created using the Term Frequency-Inverse Document Frequency (TF-IDF) method. The TF-IDF score for a term *t* in a summary *s* from a collection of summaries *S* is given by:

$$\text{TF-IDF}(t, s, S) = \text{tf}(t, s) \times \text{idf}(t, S) \qquad (1)$$

where tf(*t, s*) is the frequency of term *t* in summary *s*, and the inverse document frequency is:

$$\text{idf}(t, S) = \log(|S| / (1 + |\{s' \in S : t \in s'\}|)) \qquad (2)$$

Each summary s_i is thus represented by a vector x_i of TF-IDF scores. These vectors are then used to train a Logistic Regression classifier. The classifier predicts the probability of a summary belonging to a specific category *c* using the sigmoid function:

$$P(y = c \mid x_i; W_c) = 1 / (1 + e^{\wedge}\{-(w_c^T x_i + b_c)\}) \qquad (3)$$

where w_c and b_c are the weight vector and bias for category *c*, learned during training. This formalization highlights the classifier's direct dependence on term frequencies (Eq. 1), underscoring why the preservation of discriminative keywords is paramount for its performance.

4 Qualitative Findings and Analysis

This section details our qualitative observations on how different summarization methods impact the preservation of category-distinguishing information in resumes. To illustrate these patterns while preserving anonymity, we present representative, synthesized examples that reflect the observed phenomena.

4.1 Reduction (High-Performing Extractive)

The Reduction method, which aims to select representative sentences while minimizing redundancy, generally produced summaries that retained crucial keywords and phrases.

- **Observation:** Summaries often included sentences directly stating job titles, key responsibilities, or core skills. This direct extraction proved effective for the TF-IDF based classifier. Example (Conceptual - Original: IT Resume):
- **Original Snippet:** "... extensive experience as a Senior Java Developer. Led a team in developing microservices using Spring Boot. Proficient in SQL, Docker, and AWS..."

- **Reduction Summary:** "Extensive experience as a Senior Java Developer. Proficient in SQL, Docker, and AWS."
- **Impact on Classification:** High accuracy. The presence of terms like "Java Developer," "Spring Boot," "SQL," "Docker," and "AWS" strongly signals an IT-related category.
- **Misclassification Insight:** When misclassifications occurred, it was often due to the resume itself being ambiguous, or the top 3 sentences selected missing a more dominant categorical signal present elsewhere in the resume. However, information loss due to the summarization process itself was less frequent compared to poorer performing models.

4.2 BART (High-Performing Abstractive)

BART, being an abstractive model, rephrased content rather than extracting it verbatim. It often managed to capture the essence of the resume effectively.

- **Observation:** BART summaries were generally fluent and could synthesize information from multiple parts of the resume. They often retained key professional roles and skill areas, albeit sometimes in a more generalized form. Example (Conceptual - Original: HR Resume):
- **Original Snippet:** "Managed full-cycle recruitment for technical and non-technical roles. Developed and implemented HR policies. Conducted employee on-boarding and performance reviews."
- **BART Summary:** "Human resources professional experienced in recruitment, policy development, and employee relations."
- **Impact on Classification:** Good accuracy. The summary, while abstractive, clearly points to "Human resources."
- **Misclassification Insight**: Occasionally, BART's abstractions could lead to a loss of specificity that might be crucial for distinguishing between closely related job categories (e.g., "Marketing Specialist" vs. "Digital Media Manager") if the generalization smoothed over unique identifiers. There's also a minor risk of slight factual drift if not well-calibrated, but this was not a dominant issue in the extrinsic task.

4.3 PEGASUS (Low-Performing Abstractive)

PEGASUS (XSum variant) consistently struggled to produce summaries useful for the classification task, a finding strongly supported by quantitative results from prior work [6, 30]. Our qualitative analysis reveals why.

- **Observation:** PEGASUS summaries were often extremely short and highly abstract, frequently omitting almost all specific keywords, job titles, skills, or quantifiable achievements. This is likely due to its pre-training on the XSum dataset, which favors very brief, single-sentence summaries of news articles. Example (Conceptual - Original: Finance Resume):
- **Original Snippet:** "Chartered Financial Analyst with 8+ years in investment banking. Specialized in M&A, financial modeling, and equity research. Proficient in Bloomberg Terminal."
- **PEGASUS Summary:** "This article is about a financial professional." or "An experienced professional shares their career journey."

- **Impact on Classification:** Very low accuracy. Such generic summaries provide virtually no discriminative features for the classifier. The term "financial professional" might offer a slight hint, but often summaries were even more generic.
- **Misclassification Insight:** The extreme level of abstraction and information loss is the primary reason for PEGASUS's poor performance. It effectively strips the resume of its unique identifiers, making it impossible for the classifier to assign it to a specific job category. The confusion matrices for PEGASUS in similar studies typically show widespread misclassifications across all categories.

4.4 Random (Baseline Extractive)

The Random baseline, as expected, showed poor performance, but its failure mode was different from PEGASUS.

- **Observation:** Randomly selected sentences sometimes, by chance, picked up relevant keywords. However, more often, they selected generic introductory or concluding sentences, or sentences detailing less critical aspects of the resume. Example (Conceptual - Original: Sales Resume):
- **Original Snippet:** "Dynamic Sales Executive with a proven track record of exceeding targets. Skilled in B2B sales, negotiation, and CRM software. Seeking a challenging role..."
- **Random Summary:** "Seeking a challenging role. My hobbies include hiking. References available upon request."
- **Impact on Classification:** Low accuracy, though typically better than PEGASUS because there's a chance of hitting useful sentences.
- **Misclassification Insight:** The inconsistency of information capture is the main issue. The summaries lack a coherent focus on the most relevant aspects of the resume.

5 Discussions

Our qualitative analysis complements the quantitative findings from prior extrinsic evaluations [6, 30] by providing a deeper understanding of how different summarization strategies impact the representation of resume content for a downstream classification task.

Information Fidelity is Key: The primary differentiator between high-performing and low-performing summarizers in this context was their ability to retain specific, category-distinguishing information. Extractive methods like Reduction, by their nature, preserve original phrasing, which is beneficial when keywords and exact job titles are critical. Abstractive models like BART can also perform well if their abstraction process correctly identifies and rephrases these core concepts without significant loss.

The Perils of Over-Abstraction: The case of PEGASUS (XSum) starkly illustrates the danger of applying a summarization model trained for one type of text and summary style (very short news headlines) to a vastly different domain (semi-structured resumes requiring detail preservation). The resulting over-abstraction rendered summaries almost useless for distinguishing job categories. This highlights the critical importance of model-task alignment or domain-specific fine-tuning for abstractive summarizers [22].

Nature of "Good" Summaries for Classification: For the TF-IDF and Logistic Regression classifier used, a "good" resume summary is one that is rich in specific nouns and verb phrases that act as strong categorical indicators (e.g., "Java," "financial modeling," "recruitment process"). While intrinsic metrics like ROUGE [23] measure fluency and overlap with reference summaries, they may not correlate well with performance on downstream tasks. This highlights the broader challenge in summarization evaluation, where intrinsic and extrinsic performance can diverge [24], [25].

Implications for E-Recruitment Systems:

- When selecting a summarizer for resume preprocessing, especially for tasks like initial categorization, methods that prioritize factual and keyword integrity (like top-tier extractive approaches or carefully chosen/tuned abstractive models) are preferable.
- Off-the-shelf abstractive models, particularly those pre-trained for high compression on dissimilar domains, should be used with caution and rigorously evaluated for information loss.
- Efficiency remains a crucial factor. The high information fidelity of methods like Reduction, coupled with their computational efficiency, makes them strong candidates for high-throughput systems. BART offers good fidelity but at a higher computational cost inherent to large transformer models [8].

This qualitative analysis, while insightful, is based on manual inspection of a subset of summaries. The interpretation can have a degree of subjectivity. A larger-scale human evaluation or more sophisticated automated methods for analyzing information loss could further substantiate these findings.

6 Conclusion

This paper provided a qualitative analysis into how different text summarization techniques affect the utility of resume summaries for job category classification. By moving beyond aggregate performance metrics, we have illuminated the specific ways in which high-performing methods like Reduction and BART preserve critical category-distinguishing information, and how low-performing methods like PEGASUS (XSum) fail due to excessive abstraction and information loss. Our findings underscore that for tasks like resume classification, the fidelity of the summary in terms of specific keywords, job titles, and skills is paramount. While abstractive summaries can offer fluency, their tendency to generalize or omit crucial details (if not appropriately tuned or selected) can be detrimental. Extractive methods, by their nature, often provide a more reliable representation of these key elements.

This qualitative understanding is crucial for practitioners in e-recruitment when choosing or developing summarization tools. It highlights the need for careful consideration of the trade-offs between summary style, information preservation, and the specific requirements of downstream NLP tasks. In the future work, we aim to explore hybrid summarization approaches or fine-tuning strategies for abstractive models specifically aimed at optimizing the retention of key entities within resumes.

Data availability

The resume dataset supporting the findings of this study is openly available on Kaggle. The dataset can be accessed via the link provided in reference [10].

References

[1] Zaroor, A., Maree, M., & Sabha, M. (2017). JRC: A Job Post and Resume Classification System for Online Recruitment. In *Proc. IEEE Int. Conf. Tools Artif. Intell. (ICTAI)*, (pp. 780-787). IEEE.

[2] Ö. B. Mercan, S. N. Cavsak, A. Deliahmetoglu and S. Tanberk, "Abstractive Text Summarization for Resumes With Cutting Edge NLP Transformers and LSTM," 2023 Innovations in Intelligent Systems and Applications Conference (ASYU), Sivas, Turkiye, 2023, pp. 1-6, doi: 10.1109/ASYU58738.2023.10296563.

[3] El-Kassas, W. S., Salama, C. R., Rafea, A. A., & Mohamed, H. K. (2021). Automatic Text Summarization: A Comprehensive Survey. *Expert Syst. Appl.*, 165, 113679.

[4] Chen, J., Niu, Z., & Fu, H. (2015). A Novel Knowledge Extraction Framework for Resumes Based on Text Classifier. In *Proc. Int. Conf. Web-Age Inf. Manag. (WAIM)*, (pp. 540-543). Springer.

[5] Giarelis, N., Mastrokostas, C., & Karacapilidis, N. (2023). Abstractive vs. Extractive Summarization: An Experimental Review. *Appl. Sci.*, 13(13), 7620.

[6] Supriyono, Aji Prasetya Wibawa, Suyono, Fachrul Kurniawan, A survey of text summarization: Techniques, evaluation and challenges, Natural Language Processing Journal, Volume 7, 2024, 100070, ISSN 2949-7191, https://doi.org/10.1016/j.nlp.2024.100070.

[7] Belica, M. et al. (2024). Sumy: Automatic text summarizer. *GitHub Repository*. Available: https://github.com/miso-belica/sumy

[8] Lewis, M., et al. (2020). BART: Denoising Sequence-to-Sequence Pre-Training for Natural Language Generation, Translation, and Comprehension. In *Proc. 58th Annu. Meet. Assoc. Comput. Linguist. (ACL)*, (pp. 7871-7880).

[9] Zhang, J., Zhao, Y., Saleh, M., & Liu, P. (2020). PEGASUS: Pre-Training with Extracted Gap-Sentences for Abstractive Summarization. In *Proc. 37th Int. Conf. Mach. Learn. (ICML)*, (pp. 11328-11339).

[10] Dutta, G. Resume Dataset. Kaggle. 2020. Available online: https://www.kaggle.com/datasets/gauravduttakiit/resume-dataset.

[11] Singh, Y., Goel, A., & Sharma, D. K. (2016). A survey on text mining techniques for resume analysis. *International Journal of Computer Applications*, 142(6), 1-5.

[12] Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *Proc. 2020 Conf. Fairness, Account., Transpar.*, (pp. 469-481).

[13] Faliagka, E., Tsakalidis, A., & Tzimas, G. (2012). An integrated e-recruitment system for automated personality mining and applicant ranking. *Internet Research*, 22(5), 551-568.

[14] Kute, V. B., Bhole, G. P., & Mahajan, S. B. (2022). Resume classification using abstractive text summarization and deep learning model. *Journal of King Saud University-Computer and Information Sciences*, 34(10), 9184-9193.

[15] Chakraborty, S., Roy, S., & Dutta, P. (2020). Resume classification using machine learning. In *Proc. 4th Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA)*, (pp. 1167-1172). IEEE.

[16] Nenkova, A., & McKeown, K. (2011). Automatic summarization. *Foundations and Trends in Information Retrieval*, 5(2-3), 103-233.

[17] Vaswani, A., et al. (2017). Attention is all you need. In *Adv. Neural Inf. Process. Syst.*, 30.

[18] Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of Research and Development*, 2(2), 159-165.

[19] Radev, D. R., Jing, H., Styś, M., & Tam, D. (2004). LexRank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research*, 22, 457-479.

[20] Mihalcea, R., & Tarau, P. (2004). Textrank: Bringing order into texts. In *Proc. 2004 Conf. Empir. Methods Nat. Lang. Process.*, (pp. 404-411).

[21] Raffel, C., et al. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*, 21(1), 5485-5551.

[22] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proc. 2019 Conf. North Am. Chapter Assoc. Comput. Linguist.*, (pp. 4171-4186).

[23] Lin, C. Y. (2004). ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out* (pp. 74-81).

[24] Fabbri, A. R., Kryściński, W., McCann, B., Xiong, C., Socher, R., & Radev, D. (2021). On the extrinsic and intrinsic evaluation of summarization. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, (pp. 4446-4456).

[25] Gholampour, H., & Zou, Y. (2021). Extrinsic evaluation of text summarization: A case study of its impact on stock market and fake news prediction. *arXiv preprint arXiv:2104.04859*.

[26] Maree, Mohammed, and Wala'A. Shehada. "Optimizing curriculum vitae concordance: A comparative examination of classical machine learning algorithms and large language model architectures." *AI* 5.3 (2024): 1377-1390.

[27] Kouris, Panagiotis, Georgios Alexandridis, and Andreas Stafylopatis. "Abstractive text summarization based on deep learning and semantic content generalization." In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy (2019).

[28] H. M. J. Mustafa, H. A. Shehadeh and A. Khasawneh, "Chernobyl Disaster Optimizer Algorithm for Feature Selection Problems," *2025 1st International Conference on Computational Intelligence Approaches and Applications (ICCIAA)*, Amman, Jordan, 2025, pp. 1-8, doi: 10.1109/ICCIAA65327.2025.11013112.

[29] Mustafa, H.M.J., Ayob, M., Shehadeh, H.A. *et al.* Multi-objective memetic differential evolution optimization algorithm for text clustering problems. *Neural Comput & Applic* **35**, 1711–1731 (2023). https://doi.org/10.1007/s00521-022-07888-w

[30] Priyanka, J.H., Parveen, N. An efficient resume skill extraction using deep feature-based AGT optimized K means clustering. *Multimed Tools Appl* **83**, 65967–65986 (2024). https://doi.org/10.1007/s11042-024-18220-0

Notes on contributors

Mohammed Maree is an associate professor of Information Technology at the Faculty of Information Technology, Arab American University of Palestine. He received the Ph.D. degree in Information Technology from Monash University. He has published articles in various high-impact journals and conferences, such as ICTAI, Knowledge-Based Systems, IEEE Access, Behaviour & Information Technology, Journal on Computing and Cultural Heritage, Information Development and the Journal of Information Science. He is also a Committee Member/Reviewer of several conferences and journals. He has supervised a number of Master's and PhD students in the fields of knowledge engineering, data analysis, information retrieval, natural language processing, and hybrid intelligent systems. He began his career as the Manager of Research and Development at gSoft Technology Solution Inc. Then, he worked as the Director of Research and QA with Dimensions Consulting Company. Subsequently, he joined the Faculty of Engineering and Information Technology (EIT), Arab American University, Palestine (AAUP), as a full-time Lecturer. Dr. Mohammed is currently the Dean of Faculty of Information Technology at the Arab American University.