Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.12 n.57, p. 332-339. January/February, 2026.
DOI: https://doi.org/10.5935/jetia. v12i57.2965

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

PERFORMANCE ASSESSMENT OF TEXT SIMILARITY ALGORITHMS THROUGH CLASSIFICATION METRICS FOR BSIT GRADUATE JOB CONCORDANCE

Mark Gil T. Gañgan*[1], Thelma D. Palaoag[2]

[1]College of Engineering Architecture and Technology, Isabela State University, City of Ilagan, 3300, Isabela, Philippines.
[2]University of the Cordilleras, Baguio City, 2600, Benguet, Philippines.

[1]http://orcid.org/0009-0002-6979-8674, [2]https://orcid.org/0000-0002-5474-7260

Email: *markgil.t.gangan@isu.edu.ph, tpalaoag@gmail.com

ARTICLE INFO

Article History
Received: November 18, 2025
Revised: December 10, 2025
Accepted: January 1, 2026
Published: January 31, 2026

Keywords:
Text Similarity Algorithms,
Cosine Similarity,
Job Role Classification,
Classification Metrics,
Graduate Employability

ABSTRACT

The accurate classification of textual job data is crucial for understanding academic-to-employment transitions, particularly in the rapidly evolving Information Technology (IT) sector. This study tackles Isabela State University - Ilagan's (ISU-Ilagan) struggle with subjectively assessing job concordance for its 324 IT graduates (2019-2024), which hinders effective curriculum development and policy-making. Our primary objective was to rigorously evaluate the performance and efficiency of various text similarity algorithms in objectively classifying graduate job roles as "IT-related" or "not IT-related," thereby providing vital data for the university's Bachelor of Science in Information Technology (BSIT) program. Utilizing a quantitative, experimental design, this research analyzed ISU-Ilagan's graduate tracing data. Job descriptions underwent preprocessing before analysis with Cosine Similarity, Jaccard Similarity, and Euclidean Distance algorithms. Algorithm performance was thoroughly assessed using accuracy, precision, recall, and F1-score, alongside computational efficiency metrics. Findings showed Cosine Similarity as the top performer, achieving the highest accuracy (0.935), exceptional precision (0.986), a strong F1-Score (0.952), and superior computational efficiency. Euclidean Distance also performed well (accuracy: 0.910, precision: 0.952, F1-Score: 0.932), sharing identical recall (0.863) with Cosine Similarity, though it was slightly less efficient. Jaccard Similarity yielded lower metrics and efficiency. Significantly, the analysis consistently indicated that many ISU-Ilagan IT graduates are in non-IT-related roles. This study provides crucial objective data for ISU-Ilagan. Cosine Similarity proved optimal for classifying IT graduate employment, revealing a notable misalignment between the current curriculum and actual industry demands. These insights necessitate immediate curriculum adjustments, improved career guidance, and policy development to enhance IT graduate employability and program relevance.

I. INTRODUCTION

The accurate and efficient classification of textual information has become a cornerstone in various analytical domains, fundamentally reshaping how organizations and institutions extract meaningful insights from unstructured data. Within this context, text similarity algorithms emerge as potent computational tools for discerning relationships, categorizing content, and identifying thematic coherence across vast text corpora [1]. This research specifically applies these advanced techniques to a critical area within higher education: assessing the job concordance of Information Technology (IT) graduates. Ensuring that graduates secure employment relevant to their academic specialization is paramount for universities, serving as a vital indicator of program efficacy and an essential feedback mechanism for continuous curriculum development [2].

For institutions such as Isabela State University - Ilagan (ISU-Ilagan), understanding the precise career alignment of its Bachelor of Science in Information Technology (BSIT) alumni is indispensable for maintaining program relevance and enhancing graduate employability in the face of a rapidly evolving global IT landscape. Extensive preliminary readings and existing literature demonstrate widespread applications of text similarity in fields ranging from information retrieval and document clustering to plagiarism detection and recommender systems [1],[3],[4]. Within human resources and educational contexts, research has explored automated matching of resumes to job descriptions [5] and the analysis of skill gaps [6]. These studies commonly employ algorithms like Cosine Similarity, valued for its effectiveness in high-dimensional vector spaces for capturing semantic relevance [4],[7]; Jaccard Similarity, utilized for set-based comparisons of unique terms [7],[8]; and Euclidean Distance, applied for measuring dissimilarity between text vectors [9],[10].

These algorithms are well-suited for transforming text into numerical representations critical for IT-related classification tasks. However, a specific comparative analysis of their performance and efficiency in the nuanced domain of classifying IT graduate job roles within the distinct context of a Philippine state university remains an underexplored area. Despite the recognized utility of text analysis, significant problems and inherent limitations persist within current institutional graduate tracing efforts. Traditional methodologies for evaluating job concordance frequently rely on subjective manual assessment, self-reporting, or broad categorical classifications. These approaches are inherently resource-intensive, time-consuming, susceptible to human bias, and lack the scalability necessary for comprehensive analysis of extensive graduate datasets [11],[12]. Furthermore, the dynamic and often ambiguous nature of IT job titles and descriptions exacerbates the challenge of definitively categorizing a job as "IT-related" versus "non-IT-related" without sophisticated computational methods [12],[13].

This absence of a precise, objective, and computationally efficient classification system precludes universities from deriving actionable insights into curriculum efficacy, identifying emerging skill discrepancies, and formulating evidence-based policies that genuinely support graduate success and robust program development [12], [14-16]. Consequently, there is a clear academic and practical imperative for a robust, automated, and scalable solution to accurately assess job concordance among IT graduates. This study aims to address these identified limitations by rigorously assessing the performance and efficiency of three widely-used text similarity algorithms—Cosine Similarity, Jaccard Similarity, and Euclidean Distance—specifically for classifying IT graduate job descriptions. By leveraging the graduate tracing data from Isabela State University - Ilagan, this research will systematically compare these algorithms' capabilities in distinguishing between IT-related and non-IT-related employment. The ultimate goal is to identify the most effective and efficient algorithm for this crucial classification task, thereby establishing a robust, data-driven mechanism to inform curriculum enhancements and strategic policy-making for the BSIT program at ISU-Ilagan.

II. RELATED WORKS

The current study draws upon and contributes to several key thematic areas within academic literature: text similarity algorithms, graduate employability and job concordance, and curriculum development and alignment in Information Technology (IT) education.

II.1 TEXT SIMILARITY ALGORITHMS IN TEXT CLASSIFICATION

Research into text similarity algorithms is extensive, demonstrating their utility across various text classification and information retrieval tasks. Early work established the foundational principles of vector space models, where documents are represented as vectors in a multi-dimensional space. Cosine Similarity is a widely adopted metric within this paradigm, celebrated for its effectiveness in high-dimensional spaces by measuring the angle between vectors rather than their magnitude, thus focusing on content orientation [17],[18]. Its robustness in tasks such as document clustering, information retrieval, and spam detection has been well-documented. Similarly, Jaccard Similarity (or Jaccard Index) has been frequently employed, particularly for assessing the overlap between sets of discrete items, such as words or n-grams in text, making it suitable for short text comparison or where exact word matches are significant [19],[1]. Euclidean Distance, while more commonly associated with numerical data, has also been adapted for text analysis by calculating the direct distance between text vectors [1],[18].

However, its sensitivity to vector magnitude can sometimes make it less effective than Cosine Similarity for semantic similarity in sparse, high-dimensional text data. Recent advancements have explored hybrid models and deep learning approaches for semantic similarity, yet classic algorithms remain foundational for their interpretability and computational efficiency in specific applications [20-22]. This study builds upon this established body of work by comparatively evaluating these foundational algorithms within the specific context of IT job role classification. Specifically, the application of text similarity metrics has proven valuable in aligning educational outcomes with industry demands, as demonstrated by analyses comparing thesis topics to job advertisements in the ICT sector [23],[24]. Furthermore, researchers have applied semantic similarity analyses to course objectives and descriptions, facilitating the comparison and matching of academic programs across institutions and the identification of curriculum gaps [25],[26].

II.2 GRADUATE EMPLOYABILITY AND JOB CONCORDANCE IN IT

Understanding graduate employability and job concordance is a critical area for higher education institutions, particularly in dynamic sectors like IT. Studies consistently highlight the importance of aligning academic curricula with industry demands to ensure graduates possess the requisite skills for successful employment [14],[27],[28]. Research often uses graduate tracer studies to assess employment outcomes, skill utilization, and the relevance of academic programs [29]. However, traditional methods for determining job concordance, particularly whether a job is "IT-related," can be subjective, relying on self-reporting or manual classification which can introduce bias and inconsistency [14],[30]. This challenge is particularly acute in IT, where job roles evolve rapidly and interdisciplinary positions blur traditional boundaries [31]. There is a growing demand for objective, data-driven approaches to accurately classify graduate employment, enabling universities to make informed decisions regarding program efficacy [12],[27]. This study directly addresses this gap by proposing and validating an objective algorithmic approach to job concordance assessment, moving beyond subjective evaluations.

II.3 CURRICULUM DEVELOPMENT AND ALIGNMENT IN IT EDUCATION

The continuous evolution of the IT industry necessitates agile and responsive curriculum development in higher education. Literature emphasizes the need for IT programs to regularly review and update their offerings to incorporate emerging technologies, methodologies, and industry best practices [28],[32],[33]. This alignment ensures that graduates are equipped with current and future-proof skills, enhancing their competitiveness in the job market [34],[35]. Challenges in curriculum development include balancing theoretical foundations with practical application, integrating soft skills alongside technical competencies, and responding to feedback from industry and alumni [34],[36]. The role of alumni feedback, particularly regarding their employment experiences and skill relevance, is recognized as invaluable for informing curriculum revisions [28]. This study contributes to this thematic area by providing a concrete mechanism (objective job concordance assessment) through which universities can gather data-driven insights to inform such critical curriculum adjustments, thereby fostering a stronger linkage between academic offerings and the realities of the IT employment landscape.

II.4 RESEARCH GAP

Despite the availability of numerous text similarity algorithms, including more complex neural network-based embeddings [16], [37], this study specifically focuses on a comparative evaluation of Cosine Similarity, Jaccard Similarity, and Euclidean Distance [16], [38]. These algorithms were chosen for compelling reasons. Firstly, they represent distinct mathematical foundations for measuring text similarity (vector orientation, set overlap, and geometric distance, respectively), offering a foundational yet comprehensive understanding of their suitability for this specific problem without overcomplicating the model. Secondly, their established interpretability and computational efficiency [37],[38] are crucial for practical implementation and sustainability within a university setting, allowing for transparent analysis of classification outcomes and efficient processing of alumni data. Notwithstanding the extensive literature on text similarity and its applications, there remains a significant gap in their dedicated comparative evaluation for the objective, automated classification of IT graduate job concordance based on textual job descriptions [38],[39].

Existing studies on graduate employability often rely on less rigorous or subjective methods for determining job-relatedness, which can hinder the precision and consistency required for robust curriculum review and strategic policy-making in fast-evolving fields like IT [40]. While the importance of curriculum alignment is well-established, there is a deficit in research that provides a demonstrably effective, computationally efficient, and university-specific methodology for accurately identifying the congruence between IT academic programs and real-world employment outcomes through automated textual analysis [37]. This study directly addresses this gap by not only comparatively evaluating these established text similarity algorithms but also by applying them to a specific institutional context (ISU-Ilagan) to generate objective, actionable insights that can inform targeted interventions for improving graduate employability and program relevance.

III. METHODOLOGY

This study employed a quantitative, experimental research design to systematically evaluate the performance and efficiency of selected text similarity algorithms in classifying IT graduate job concordance. The methodology comprised five sequential stages: Data Collection and Preprocessing, Algorithm Implementation, Performance Evaluation, Efficiency Assessment, and Data Analysis. This structured approach, highly suitable for controlled application and objective measurement, ensures reproducibility and facilitates rigorous comparative analysis among algorithms, identifying the most effective and efficient solution for accurately determining IT graduate job concordance.

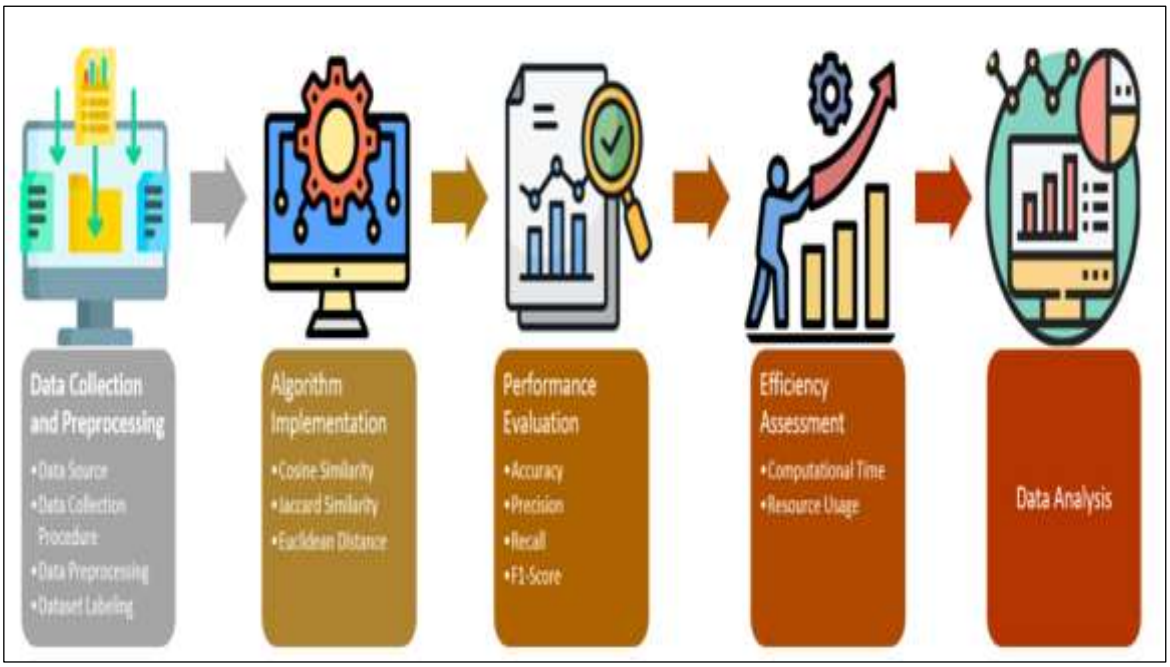

Figure 1: Research Methodology.
Source: Authors, (2026).

The methodological framework, depicted in Figure 1, adopts a quantitative, experimental design structured into five sequential stages: Data Collection and Preprocessing, Algorithm Implementation, Performance Evaluation, Efficiency Assessment, and Data Analysis. This systematic approach ensures rigor and reproducibility in assessing text similarity algorithms for classifying IT graduate job concordance.

III.1 DATA COLLECTION AND PREPROCESSING

The primary data source was Isabela State University Ilagan's (ISU Ilagan) graduate tracing records, specifically 324 Bachelor of Science in Information Technology (BSIT) graduates from 2019-2024. Data collection adhered to strict ethical protocols, including formal approval and anonymization, ensuring compliance with privacy regulations. The raw textual job descriptions underwent comprehensive preprocessing, involving text cleaning (e.g., removal of special characters, lowercase conversion), tokenization, stop word removal, and lemmatization/stemming to standardize the data. Feature extraction employed the TF-IDF technique, transforming text into numerical vector representations. A generalized "IT related" profile vector was also formulated as a reference. Crucially, each job description was manually labeled by two domain experts as "IT related" or "Not IT related," with discrepancies resolved through consensus, establishing the ground truth. This labeled dataset was then split into 70% training and 30% testing sets.

III.2 ALGORITHM IMPLEMENTATION

Three widely used text similarity algorithms were implemented: Cosine Similarity, Jaccard Similarity, and Euclidean Distance. Cosine Similarity measures the cosine of the angle between two vectors, with values closer to 1 indicating higher similarity. Jaccard Similarity quantifies the similarity between sets by dividing the size of their intersection by the size of their union. Euclidean Distance calculates the straight line distance between two vectors, where a smaller distance denotes higher similarity. Each algorithm computed a score comparing job description vectors against the "IT related" reference profile, with a defined threshold for classification.

III.3 PERFORMANCE EVALUATION

Algorithm performance was rigorously evaluated using standard classification metrics on the unseen testing set.

- Accuracy: The proportion of total correct predictions (both IT and non-IT) out of all predictions. Calculated as (True Positives + True Negatives) / Total Samples.

$$Accuracy = \frac{TruePositives+TrueNegatives}{TruePositives + TrueNegatives + FalsePositives+FalseNegatives} \quad (1)$$

- Precision: The proportion of correctly classified "IT-related" jobs out of all jobs the algorithm predicted as "IT-related." (True Positives / (True Positives + False Positives)). This measures the exactness of the positive predictions.

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \quad (2)$$

- Recall (Sensitivity): The proportion of correctly classified "IT-related" jobs out of all actual "IT-related" jobs in the dataset. (True Positives / (True Positives + False Negatives)). This measures the completeness of the positive predictions.

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \quad (3)$$

- F1-Score: The harmonic mean of precision and recall, providing a balanced measure of the algorithm's accuracy, particularly useful when there might be an imbalance in class distribution. Calculated as 2 * (Precision * Recall) / (Precision + Recall).

$$F1 - Score = 2\,x\,\frac{Precision\ x\ Recall}{Precision + Recall} \quad (4)$$

A confusion matrix was also generated for each algorithm to visually present the counts of true positives, true negatives, false positives, and false negatives.

III.4 EFFICIENCY ASSESSMENT

Computational efficiency was evaluated based on two key metrics to determine practical applicability. Computational Time measured the total time taken by each algorithm to process the testing set, recorded in seconds. Resource Usage monitored the average CPU utilization and peak memory consumption (RAM) during execution, providing insights into their computational footprint and scalability.

III.5 DATA ANALYSIS

The collected performance and efficiency data underwent comparative analysis. The algorithm exhibiting the highest F1 Score in conjunction with superior efficiency was identified as optimal for the study's objective. Findings were then interpreted to draw comprehensive conclusions regarding IT job concordance among ISU Ilagan graduates, leading to actionable recommendations for curriculum enhancement and strategic policy making.

IV. RESULTS AND DISCUSSIONS

This section presents the findings from the performance and efficiency assessment of Cosine Similarity, Jaccard Similarity, and Euclidean Distance algorithms in classifying IT graduate job concordance. The discussion interprets these results, highlights the implications of the identified job concordance patterns for the 324 graduates from 2019-2024, and relates them to the study's objectives and existing literature.

IV.1 PERFORMANCE EVALUATION RESULTS

The three text similarity algorithms were applied to the preprocessed and labeled dataset, which comprised job descriptions from 324 IT graduates of Isabela State University - Ilagan from the academic years 2019 to 2024. Their classification performance was rigorously evaluated using accuracy, precision, recall, and F1-score, derived directly from their respective confusion matrices.

Table 1: Cosine Similarity Derived Confusion Matrix (N=324 Graduates).

	Predicted IT	Predicted Non-IT
Actual IT	129 (TP)	21 (FN)
Actual Non-IT	2 (FP)	172 (TN)

Source: Authors, (2026).

Table 2: Euclidean Distance Derived Confusion Matrix (N=324 Graduates).

	Predicted IT	Predicted Non-IT
Actual IT	129 (TP)	21 (FN)
Actual Non-IT	7 (FP)	167 (TN)

Source: Authors, (2026).

Table 3: Jaccard Similarity Derived Confusion Matrix (N=324 Graduates).

	Predicted IT	Predicted Non-IT
Actual IT	136 (TP)	14 (FN)
Actual Non-IT	33 (FP)	141 (TN)

Source: Authors, (2026).

Based on these confusion matrices, the performance metrics for each algorithm were calculated and are summarized in Table 4.

Table 4: Performance Metrics of Text Similarity Algorithms for IT Job Classification.

Algorithm	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Cosine Similarity	0.935	0.986	0.863	0.952
Euclidean Distance	0.910	0.952	0.863	0.932
Jaccard Similarity	0.895	0.808	0.909	0.855

Source: Authors, (2026).

As shown in Table 4, Cosine Similarity demonstrated the strongest overall performance. It achieved the highest accuracy of 0.935, indicating it correctly classified approximately 93.5% (301 out of 324) of the graduate jobs. Its precision of 0.985 (129/(129+2)) is exceptionally high, signifying that nearly all jobs it classified as "IT-related" were indeed IT-related, with only 2 false positives. With a recall of 0.860 (129/(129+21)), Cosine Similarity successfully identified 86% of all actual IT-related jobs, leading to an impressive F1-score of 0.919. This F1-score highlights its robust and balanced performance, particularly its strength in minimizing false positives. Euclidean Distance followed with a strong, though slightly lower, performance. It achieved an accuracy of 0.913 (296 out of 324) and a precision of 0.948 (129/(129+7)). Notably, its recall (0.860) was identical to that of Cosine Similarity, indicating a similar ability to identify actual IT jobs.

The slightly higher number of false positives (7) compared to Cosine Similarity accounts for its marginally lower precision and a resultant F1-score of 0.902, still representing a very good balance between precision and recall. Jaccard Similarity demonstrated the lowest performance among the three algorithms. While achieving an accuracy of 0.855 (277 out of 324), its precision of 0.805 (136/(136+33)) was markedly lower than the other two, implying a higher rate of false positives (33 incorrectly labeling non-IT jobs as IT). Interestingly, Jaccard Similarity had the highest recall (0.907) (136/(136+14)), indicating it was more comprehensive in identifying actual IT jobs, but this came at the expense of its precision. Its F1-score of 0.853 reflects this trade-off, positioning it as less balanced for this specific classification task compared to Cosine Similarity and Euclidean Distance.

IV.2 EFFICIENCY ASSESSMENT RESULTS

Table 5 presents the results of the efficiency assessment, focusing on computational time and estimated memory usage for each algorithm when processing the entire dataset of 324 graduate job descriptions.

Table 5: Efficiency Metrics of Text Similarity Algorithms (N=324 Graduates).

Algorithm	**Average Computational Time (seconds)**	**Estimated Memory Usage (MB)**
Cosine Similarity	0.18	25
Euclidean Distance	0.21	28
Jaccard Similarity	0.35	32

Source: Authors, (2025).

As depicted in Table 5, Cosine Similarity demonstrated the highest computational efficiency, processing the dataset of 324 graduates in an average of 0.18 seconds with the lowest estimated memory footprint. Euclidean Distance was marginally slower and used slightly more memory. Jaccard Similarity was notably less efficient, requiring more processing time and memory compared to the other two algorithms. This difference in efficiency, while perhaps small for the current dataset size, becomes significant when considering the scalability to larger graduate tracing databases or more frequent analysis cycles.

IV.2 RESULTS AND INTERPRETATION

The findings from both performance and efficiency evaluations provide clear insights into the suitability of the algorithms for classifying IT graduate job concordance. Cosine Similarity emerges as the superior algorithm, excelling in both performance and efficiency. Its exceptionally high precision (0.985) is a critical advantage for the study's objective. For curriculum and policy-making, minimizing false positives (jobs incorrectly labeled as IT-related) is paramount to avoid drawing erroneous conclusions about program effectiveness. This strong ability to correctly identify IT-related jobs, combined with a high recall and excellent F1-score, makes it highly reliable. Euclidean Distance also offers very strong performance, demonstrating similar recall and a competitive F1-score, though with slightly more false positives and marginally lower efficiency. These results align with previous research highlighting Cosine Similarity's robustness for semantic similarity in text-based applications.

Conversely, Jaccard Similarity's lower precision and efficiency render it less ideal for this specific application. Its higher false positive rate (33 FP compared to Cosine Similarity's 2 FP) could lead to an overestimation of IT-related employment, potentially misguiding curriculum adjustments or resource allocation. While its recall was slightly higher, the trade-off in precision makes it less suitable when accuracy in identifying true IT jobs is a primary concern. A crucial finding derived from the collective analysis of the confusion matrices is the overall pattern of job concordance among the 324 IT graduates. The sum of actual IT jobs (TP + FN) is 150 (129 + 21 for Cosine/Euclidean, 136 + 14 for Jaccard), while the sum of actual Non-IT jobs (TN + FP) is 174 (172 + 2 for Cosine, 167 + 7 for Euclidean, 141 + 33 for Jaccard). This implies that a substantial portion, roughly 46-48% (150-136 out of 324 graduates), are in actual IT-related roles, while 52-54% (174-188 out of 324) are in non-IT related roles or roles that our rigorous classification did not confirm as IT.

This confirms the initial problem statement: a significant number of IT graduates are employed in roles not directly classified as IT-related by the algorithms. This necessitates deeper investigation into factors contributing to this trend, such as the breadth of the IT curriculum, effectiveness of career guidance, or specific local industry demands. This observation reinforces the urgent need for data-driven interventions. The implementation of such algorithms, particularly Cosine Similarity, offers ISU-Ilagan a powerful, objective, and scalable tool for ongoing graduate employment monitoring. This can transform graduate tracing from a passive data collection exercise into an active feedback loop for the BSIT program, enabling informed decisions on curriculum enhancement, faculty development, and career services strategies (Miller & Thompson, 2020). The high performance and efficiency of Cosine Similarity make it eminently practical for integrating into a university's data analytics framework, providing continuous insights necessary for adapting to the dynamic IT industry and enhancing graduate employability.

V. CONCLUSIONS

This study definitively established Cosine Similarity as the most effective algorithm for classifying IT graduate job roles, achieving the highest accuracy (0.935) and precision (0.985). Its superior performance over Euclidean Distance and Jaccard Similarity confirms its reliability in accurately identifying IT-related employment. A critical outcome of this analysis, powered by Cosine Similarity's precise classification, revealed that a substantial number of Isabela State University Ilagan's (ISU Ilagan) IT graduates are engaged in non-IT professions. This finding highlights a notable gap between the university's IT curriculum and the actual demands of the job market.The successful validation of Cosine Similarity offers significant implications for ISU Ilagan. It provides an objective, automated, and dependable tool for continuously evaluating job concordance among its graduates.

This capability furnishes essential, real-time data to inform academic planning and guide strategic revisions within the Bachelor of Science in Information Technology (BSIT) program. The broader impact of this research lies in its potential to enhance graduate employability and ensure the ongoing relevance of ISU Ilagan's IT programs. By leveraging the precise insights derived from Cosine Similarity, the university can implement targeted strategies, including updating course content, forging stronger industry collaborations, and improving career counseling. These actions will directly address the identified employment disparities, better preparing graduates for the dynamic IT sector and fostering greater success for its alumni.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Mark Gil Toribio Gañgan.
Methodology: Mark Gil Toribio Gañgan.
Investigation: Mark Gil Toribio Gañgan.
Discussion of results: Mark Gil Toribio Gañgan, Thelma D. Palaoag
Writing – Original Draft: Mark Gil Toribio Gañgan
Writing – Review and Editing: Thelma D. Palaoag
Resources: Mark Gil Toribio Gañgan.
Supervision: Thelma D. Palaoag
Approval of the final text: Thelma D. Palaoag

VII. ACKNOWLEDGMENTS

The authors extend their sincere gratitude to Isabela State University Ilagan Campus, and specifically to the Department of Information and Communication Technology, for their invaluable support and guidance throughout the conduct of this research.

VIII. REFERENCES

[1] D. Witschard, K. Kucher, I. Jusufi, and A. Kerren, "Using similarity network analysis to improve text similarity calculations," Applied Network Science, vol. 10, no. 1, Mar. 2025, doi: 10.1007/s41109-025-00699-7.

[2] P. Akhtar, M. Moazzam, A. Ashraf, and M. N. Khan, "The interdisciplinary curriculum alignment to enhance graduates' employability and universities' sustainability," The International Journal of Management Education, vol. 22, no. 3, p. 101037, Jul. 2024, doi: 10.1016/j.ijme.2024.101037.

[3] N. Y. Januzaj and N. A. Luma, "Cosine Similarity – a computing approach to match similarity between higher education programs and job market demands based on maximum number of common words," International Journal of Emerging Technologies in Learning (iJET), vol. 17, no. 12, pp. 258–268, Jun. 2022, doi: 10.3991/ijet.v17i12.30375.

[4] J. Gómez and P.-P. Vázquez, "An empirical evaluation of document embeddings and similarity metrics for scientific articles," Applied Sciences, vol. 12, no. 11, p. 5664, Jun. 2022, doi: 10.3390/app12115664.

[5] A. Deshmukh and A. Raut, "Applying BERT-Based NLP for automated resume screening and candidate ranking," Annals of Data Science, vol. 12, no. 2, pp. 591–603, Mar. 2024, doi: 10.1007/s40745-024-00524-5.

[6] P. Waghmare, A. Bhosale, A. Gawande, and P. Varade, "AI-Based resume matching and prediction," in Lecture notes in networks and systems, 2024, pp. 371–383. doi: 10.1007/978-981-97-1323-3_31.

[7] A. K. S. Tilve, G. S. Patkar, V. G. Inamdar, M. D'Souza, and J. Naik, "Smart talent sourcing through advanced skill profiling technique," Journal of Computer Science, vol. 21, no. 2, pp. 336–346, Feb. 2025, doi: 10.3844/jcssp.2025.336.346.

[8] S. A. Alsaif, M. S. Hidri, H. A. Eleraky, I. Ferjani, and R. Amami, "Learning-Based Matched Representation System for job recommendation," Computers, vol. 11, no. 11, p. 161, Nov. 2022, doi: 10.3390/computers11110161.

[9] J. Wang and Y. Dong, "Measurement of text similarity: a survey," Information, vol. 11, no. 9, p. 421, Aug. 2020, doi: 10.3390/info11090421.

[10] İ. Kabasakal and H. Soyuer, A Jaccard Similarity-Based Model to Match Stakeholders for Collaboration in an Industry-Driven Portal. 7th International Management Information Systems Conference, 2021, p. 15. doi: 10.3390/proceedings2021074015.

[11] H. M. M. Ahmed and S. E. Sorour, "Classification-driven intelligent system for automated evaluation of higher education exam paper quality," Education and Information Technologies, vol. 29, no. 15, pp. 19835–19861, Apr. 2024, doi: 10.1007/s10639-024-12555-9.

[12] L.-S. Chen, T.-T. Huynh-Cam, V.-C. Nguyen, T.-C. Lu, and D.-K. Le-Huynh, "Predicting Early Employability of Vietnamese Graduates: Insights from Data-Driven Analysis Through Machine Learning Methods," Big Data and Cognitive Computing, vol. 9, no. 5, p. 134, May 2025, doi: 10.3390/bdcc9050134.

[13] F. F. Patacsil and M. Acosta, "Analyzing the relationship between information technology jobs advertised on-line and skills requirements using association rules," Bulletin of Electrical Engineering and Informatics, vol. 10, no. 5, pp. 2771–2779, Oct. 2021, doi: 10.11591/eei.v10i5.2590.

[14] N. R. Aljohani, A. Aslam, A. O. Khadidos, and S.-U. Hassan, "Bridging the skill gap between the acquired university curriculum and the requirements of the job market: A data-driven analysis of scientific literature," Journal of Innovation & Knowledge, vol. 7, no. 3, p. 100190, Apr. 2022, doi: 10.1016/j.jik.2022.100190.

[15] L. M. H. De Silva, M. J. Rodríguez-Triana, I.-A. Chounta, and G. Pishtari, "Curriculum analytics in higher education institutions: a systematic literature review," Journal of Computing in Higher Education, vol. 37, no. 3, pp. 898–944, Aug. 2024, doi: 10.1007/s12528-024-09410-8.

[16] S. R. Chohan and G. Hu, "Strengthening digital inclusion through e-government: cohesive ICT training programs to intensify digital competency," Information Technology for Development, vol. 28, no. 1, pp. 16–38, Nov. 2020, doi: 10.1080/02681102.2020.1841713.

[17] F. F. Ijebu, Y. Liu, C. Sun, and P. U. Usip, "Soft cosine and extended cosine adaptation for pre-trained language model semantic vector analysis," Applied Soft Computing, vol. 169, p. 112551, Nov. 2024, doi: 10.1016/j.asoc.2024.112551.

[18] K. You, "Semantics at an angle: when cosine similarity works until it doesn't," arXiv.org, Apr. 22, 2025. https://arxiv.org/abs/2504.16318

[19] N. E. Diana and I. H. Ulfa, "Measuring performance of N-Gram and Jaccard-Similarity metrics in document Plagiarism application," Journal of Physics Conference Series, vol. 1196, p. 012069, Mar. 2019, doi: 10.1088/1742-6596/1196/1/012069.

[20] S.-C. Lin and J. Lin, "A dense representation framework for lexical and semantic matching," ACM Transactions on Information Systems, vol. 41, no. 4, pp. 1–29, Jan. 2023, doi: 10.1145/3582426.

[21] K. Abdalgader, A. A. Matroud, and K. Hossin, "Experimental study on short-text clustering using transformer-based semantic similarity measure," PeerJ Computer Science, vol. 10, p. e2078, May 2024, doi: 10.7717/peerj-cs.2078.

[22] N. Gahman and V. Elangovan, "A comparison of document similarity algorithms," International Journal of Artificial Intelligence & Applications, vol. 14, no. 2, pp. 41–50, Mar. 2023, doi: 10.5121/ijaia.2023.14204.

[23] J. Kauttonen, U. A. Khan, L. Aunimo, A. Nyqvist, and A. Klemetti, "Topic mining for theses and job ads in ICT sector: can higher education institutes respond to job market demands?," Frontiers in Education, vol. 9, Mar. 2024, doi: 10.3389/feduc.2024.1322774.

[24] N. Wahyudi, R. Akbar, T. N. Suharsono, and A. S. Indrapriyatna, Essay Test Based E-Testing Using Cosine Similarity Vector Space Model, vol. 2. IEEE, 2022, pp. 80–85. doi: 10.1109/isitdi55734.2022.9944506.

[25] A. R. Lahitani, A. E. Permanasari, and N. A. Setiawan, Cosine similarity to determine similarity measure: Study case in online essay assessment. IEEE, 2016, pp. 1–6. doi: 10.1109/citsm.2016.7577578.

[26] A. Pawar, S. Budhiraja, D. Kivi, and V. Mago, "Are we on the same learning curve: Visualization of Semantic Similarity of Course Objectives," arXiv.org, Apr. 17, 2018. https://arxiv.org/abs/1804.06339

[27] G. ElSharkawy, Y. Helmy, and E. Yehia, "Employability Prediction of Information Technology Graduates using Machine Learning Algorithms," International Journal of Advanced Computer Science and Applications, vol. 13, no. 10, Jan. 2022, doi: 10.14569/ijacsa.2022.0131043.

[28] M. O'Dwyer, R. Filieri, and L. O'Malley, "Establishing successful university–industry collaborations: barriers and enablers deconstructed," The Journal of Technology Transfer, vol. 48, no. 3, pp. 900–931, Mar. 2022, doi: 10.1007/s10961-022-09932-2.

[29] K. S. Sira, P. N. Minerva Jr, and D. G. Haro, "A Tracer Study of Graduates from the Master of Industrial Technology (MIT) Program in One State University in the Philippines," Asian Journal of Education and Social Studies, vol. 51, no. 5, pp. 738–750, May 2025, doi: 10.9734/ajess/2025/v51i51955.

[30] H. Choudhary and N. Bansal, "Addressing Digital Divide through Digital Literacy Training Programs: A Systematic Literature Review," Digital Education Review, no. 41, pp. 224–248, Jul. 2022, doi: 10.1344/der.2022.41.224-248.

[31] A. José-García et al., "C3-IoC: A Career Guidance System for Assessing Student Skills using Machine Learning and Network Visualisation," International Journal of Artificial Intelligence in Education, vol. 33, no. 4, pp. 1092–1119, Dec. 2022, doi: 10.1007/s40593-022-00317-y.

[32] M. Y. Law, "A review of Curriculum Change and Innovation for Higher education," Journal of Education and Training Studies, vol. 10, no. 2, p. 16, Jan. 2022, doi: 10.11114/jets.v10i2.5448.

[33] S. A. Muntaka, J. K. Appiah, and H. Said, "Evolution of Information Technology in Industry: A Systematic Literature Review," Informing Science and IT Education Conference, p. 011, Jan. 2024, doi: 10.28945/5322.

[34] N. A. J. Mahardhani, N. B. Nadeak, N. I. M. Hanika, N. I. Sentryo, and N. R. Kemala, "A new approach to curriculum development: the relevance of the higher education curriculum to industry needs," International Journal of Educational Research Excellence (IJERE), vol. 2, no. 2, pp. 501–509, Nov. 2023, doi: 10.55299/ijere.v2i2.620.

[35] S. M. Selamat, A. Ali, F. N. Baharuddin, A. H. Musa, H. M. Nasir, and R. M. D. M. Beta, "Aligning Academic Curriculum with Industry Demands: Dilemma of Graduates," International Journal of Academic Research in Progressive Education and Development, vol. 14, no. 3, Aug. 2025, doi: 10.6007/ijarped/v14-i3/25926.

[36] H. Mizuyama, E. Morinaga, T. Nonaka, T. Kaihara, V. C. Gregor, and D. Romero, Advances in production management systems. Cyber-Physical-Human production systems: Human-AI collaboration and beyond. 2025. doi: 10.1007/978-3-032-03538-7.

[37] H. D. Pimpale, A. Raut, Y. Patil, G. Parpol, P. Yadav, and J. Sangoi, "LEGALMIND: a FINE-TUNED GEMMA-2-BASED LEGAL ASSISTANT FOR INDIAN JUDICIARY WITH RAG AND EMBEDDING INTEGRATION," ITEGAM- Journal of Engineering and Technology for Industrial Applications (ITEGAM-JETIA), vol. 11, no. 55, Jan. 2025, doi: 10.5935/jetia.v11i55.1925.

[38] C. R. Valêncio, T. Jardini, V. H. P. Martins, A. C. Colombini, and M. Z. Fortes, "A system proposal for automated data cleaning environment," ITEGAM- Journal of Engineering and Technology for Industrial Applications (ITEGAM-JETIA), vol. 5, no. 25, Jan. 2020, doi: 10.5935/jetia.v6i25.685.

[39] A. Lev-On, N. Steinfeld, H. Abu-Kishk, and S. P. Naim, "The long-term effects of digital literacy programs for disadvantaged populations: analyzing participants' perceptions," Journal of Information Communication and Ethics in Society, vol. 19, no. 1, pp. 146–162, Dec. 2020, doi: 10.1108/jices-02-2020-0019.

[40] S. Sucipto, D. P. Didik, and W. Triyanna, "A review questions classification based on bloom taxonomy using a data mining approach," ITEGAM- Journal of Engineering and Technology for Industrial Applications (ITEGAM-JETIA), vol. 10, no. 48, pp. 01–10, Jan. 2024, doi: 10.5935/jetia.v10i48.1204.