Open Access Indonesia Journal of Social Sciences
Journal Homepage: https://journalsocialsciences.com/index.php/OAIJSS

The Velocity of Relevance: Mapping the Structural Divergence Between Labor Market Signals and University Curricula in Indonesia via Text Mining and Network Analysis

Bimala Putri[1*], Delia Tamim[2], Hesti Putri[3]

[1]Department of Islamic Education, Enigma Institute, Palembang, Indonesia

[2]Department of Administrative Law, Enigma Institute, Palembang, Indonesia

[3]Department of Informatics Science, CMHC Research Center, Palembang, Indonesia

ARTICLE INFO

Keywords:
Higher education policy
Labor market dynamics
Network analysis
Skills mismatch
Text mining

***Corresponding author:**
Bimala Putri

E-mail address:
bimala.putri@enigma.or.id

All authors have reviewed and approved the final version of the manuscript.

https://doi.org/10.37275/oaijss.v8i6.310

A B S T R A C T

The persistent disconnect between higher education outcomes and labor market demands, frequently termed the skills mismatch, remains a critical barrier to Indonesia's economic competitiveness in the Fourth Industrial Revolution. Traditional survey-based methodologies often lack the granularity to capture dynamic market shifts and technical nuances. This study employs a Big Data approach, utilizing automated web scraping to harvest N = 1,042,500 unique job advertisements from major Indonesian portals and N = 4,500 course syllabi from 50 top-tier Indonesian universities between 2023 and 2024. We applied Natural Language Processing, specifically Latent Dirichlet Allocation for topic modeling, and Social Network Analysis to calculate semantic overlap and centrality measures between industry demands and academic provision. We utilized the Overlap Coefficient to correct for corpus size imbalance. The analysis reveals a structural divergence: while 82% of job ads prioritize Digital Fluency and Agile Project Management, only 28% of curricula explicitly integrate these competencies. Network analysis identifies Data Analysis as a peripheral node in academic graphs but a central hub in industry networks with a Betweenness Centrality of 0.45. Conversely, theoretical constructs dominant in academia show weak linkage to employability clusters. In conclusion, the findings evidence a systemic velocity gap where industry requirements evolve three times faster than curriculum adaptation. We propose a dynamic, API-driven curriculum model to mitigate this asymmetry.

1. Introduction

The global economic architecture is currently undergoing a seismic transformation, driven by the confluence of cyber-physical systems, artificial intelligence, and ubiquitous connectivity—a paradigm often encapsulated as the Fourth Industrial Revolution (4IR).[1] This transition is not merely technological but fundamentally alters the ontology of labor, rendering traditional models of human capital accumulation increasingly precarious. For emerging economies, the exigencies of the 4IR present a paramount challenge: the rapid recalibration of workforce competencies to align with a production landscape characterized by volatility, uncertainty, complexity, and ambiguity (VUCA). In this context, the stability of the link between higher education and economic productivity has fractured, necessitating a rigorous re-examination of how human capital is developed, deployed, and signaled in the labor market.[2]

Indonesia serves as a critical case study within this global phenomenon. As the largest economy in Southeast Asia, Indonesia is currently navigating a pivotal demographic bonus, a period wherein the working-age population significantly outnumbers dependents.[3] Theoretically, this demographic structure offers a window of opportunity for accelerated economic growth and graduation from middle-income status. However, this bonus is a double-edged sword. It presents a stark binary outcome: if the surplus labor force is equipped with market-relevant competencies, it can catalyze a productivity miracle; conversely, if this population enters the workforce without the requisite skills to navigate the digital economy, the demographic bonus risks devolving into a demographic disaster, characterized by structural unemployment, underemployment, and rising inequality. The critical variable determining this trajectory is the alignment between the supply of skills generated by higher education institutions (HEIs) and the demand for skills emanating from the industrial sector.[4]

The phenomenon of skills mismatch—the discordance between the competencies possessed by graduates and those required by employers—has thus emerged as a central concern for policymakers and educators alike. While skills mismatch is a global issue, its manifestation in Indonesia is particularly acute due to the speed of the country's digital transformation relative to the institutional inertia of its education system.[5] Current academic literature differentiates this mismatch into three distinct dimensions. First is the vertical mismatch, often termed over-education, where graduates hold credentials that exceed the requirements of their roles, leading to wage penalties and job dissatisfaction. Second is the horizontal mismatch, where graduates are employed in fields unrelated to their study programs, suggesting a failure in career guidance or curriculum relevance. Third, and perhaps most insidious in the digital age, is the qualitative mismatchor skills gap. This refers to a situation where a graduate may hold the correct degree (Computer Science) and work in the correct field (Software Development) but lacks the specific, granular technical competencies (proficiency in cloud architecture or agile methodologies) required to be productive.

Traditionally, the diagnosis of these mismatches has relied on established social science methodologies, primarily employer satisfaction surveys, graduate tracer studies, and focus group discussions. While these methods provide valuable qualitative insights into employer sentiment and graduate trajectories, they suffer from significant epistemological and practical limitations in the context of the 4IR.[6] First, survey-based methods are inherently retrospective; they capture data on skills shortages often months or years after the shortage has impacted the market. In an era where software frameworks and digital tools evolve on six-month cycles, such temporal lag renders policy recommendations obsolete before they are implemented. Second, these methods are plagued by subjectivity and perception bias. An employer's complaint regarding a lack of communication skills is often semantically ambiguous—it could refer to English proficiency, presentation ability, or digital etiquette—making it difficult for universities to translate into curriculum reform. Finally, traditional methods are constrained by sample sizes and broad categorizations. Surveys often ask about generic skill categories (IT skills), failing to capture the granular, rapid-fire evolution of specific technical requirements (React.js vs. Angular) that determine actual employability.

Consequently, there is a pervasive informational asymmetry between the labor market and the higher education sector. Industry demands act as dynamic, high-frequency signals that evolve in real-time, while university curricula operate as static, low-frequency documents governed by multi-year accreditation cycles and bureaucratic rigidity. This disconnect creates a velocity gap, where the rate of change in skill demand significantly outpaces the rate of curriculum adaptation. Policymakers and curriculum developers are effectively navigating a dynamic terrain with static maps, leading to a persistent structural divergence between what is taught in the lecture hall and what is bought in the labor market.[7]

Despite the urgency of this issue, there is a paucity of empirical research that utilizes the tools of the digital era—specifically computational social science and Big Data analytics—to quantify this gap at a macroscopic scale.[8] The majority of existing studies on skills mismatch in Indonesia remain rooted in small-scale, survey-based approaches, limiting their generalizability and their ability to detect emerging trends. There is a critical need to move beyond perception-based metrics toward data-driven methodologies that can process the vast unstructured text data generated by the labor market. Job advertisements, when viewed through the lens of data science, are not merely recruitment tools but rich repositories of labor market signaling. They represent the explicit, real-time expression of industry demand. Similarly, university course syllabi and lesson plans (Rencana Pembelajaran Semester) represent the codified supply-side intent of the education sector.

This study addresses these methodological and theoretical limitations by leveraging Big Data analytics to perform a comparative analysis of these two distinct corpora. By treating job advertisements and university curricula as comparable text data, we can move beyond anecdotal evidence to mathematically map the topology of the skills ecosystem. This approach allows for a direct, granular comparison of the language used by industry versus the language used by academia. It enables the identification of specific lexical gaps—where the two sectors speak different languages—and structural gaps—where the connectivity and prioritization of skills differ.[9]

The novelty of this research lies in its rigorous methodological integration of unsupervised machine learning and network topology analysis to visualize the semantic distance between the labor market and higher education. Unlike previous studies that may have employed simple keyword counting, this study utilizes Latent Dirichlet Allocation (LDA) to uncover hidden thematic structures and Social Network Analysis (SNA) to model the relationships between skills. By constructing skill networks, we can analyze the centrality of specific competencies, identifying which skills act as bridge nodes that connect different domains of knowledge and which act as peripheral nodes that are isolated from market demand. Furthermore, this study innovates by employing the Overlap Coefficient rather than standard similarity metrics to statistically correct for the massive volume imbalance between the millions of data points generated by the labor market and the thousands of documents produced by universities, ensuring a robust and unbiased comparison.[10]

The aim of this study is to quantify the specific dimensions of the skills mismatch in the digital economy sector of Indonesia and identify high-centrality skills that act as bridge nodes between academic preparation and professional employability. By doing so, we seek to provide an empirical basis for API-based curriculum reform—a model where core academic foundations remain stable, but peripheral skill modules are updated dynamically in response to real-time market data. To guide this investigation, we posit two central research questions. First, is the skills mismatch primarily lexical, meaning the divergence stems from different terminology for the same concepts, or is it structural, indicating a fundamental difference in how skills are clustered and prioritized? Second, does the velocity of relevance—defined as the rate at which new skills permeate the network—differ significantly between the industrial and academic sectors? Through these inquiries, this study aspires to contribute a new methodological framework for educational policy in the age of Big Data.

2. Methods

To address the systemic disconnect between labor market signaling and educational supply, this study adopts a quantitative, exploratory research design situated within the emerging field of Computational Social Science. Unlike traditional educational research that relies on perception-based surveys or limited qualitative case studies, this research leverages the volume and velocity of Big Data to reconstruct the semantic topology of the Indonesian skills ecosystem. The methodological workflow is structured according to the Knowledge Discovery in Databases (KDD) process, a standard framework in data mining that ensures rigor and reproducibility. This iterative process encompasses five distinct phases: Selection, Preprocessing, Transformation, Data Mining, and Evaluation. By adhering to this framework, we move beyond anecdotal observation to provide a mathematically grounded mapping of the skills mismatch, treating text not merely as linguistic strings but as quantifiable data points that reveal latent structural relationships between industry demands and academic intent.

The data acquisition phase involved the construction of two distinct, high-dimensional corpora representing the Demand (Industry) and Supply (Academia) sides of the labor market equation. To capture the real-time, granular evolution of skill requirements, we developed a custom Python-based web scraping architecture utilizing Selenium for dynamic content rendering and BeautifulSoup for DOM parsing. This automated retrieval agent was deployed to harvest job advertisements from three primary digital employment platforms dominant in the Indonesian market: JobStreet, LinkedIn Indonesia, and Kalibrr. These platforms were selected for their high market penetration and their tendency to host formal sector jobs relevant to the digital economy.

The temporal window for data collection spanned a 24-month period from January 1, 2023, to December 31, 2024. This timeframe was chosen to capture the post-pandemic stabilization of labor trends and the rapid integration of generative AI technologies into job descriptions. The initial harvest underwent a rigorous de-duplication process using hashing algorithms to identify and remove identical postings across platforms (such as a job posted simultaneously on LinkedIn and JobStreet). Post-cleaning, the final industry corpus consisted of N = 1,042,500 unique job postings. To ensure sector-specific relevance, the scraper was configured to target five key sectors critical to Indonesia's Fourth Industrial Revolution roadmap: finance, technology, engineering, creative industries, and administration. This massive dataset serves as a proxy for real-time labor market signaling, capturing not just job titles but the specific technical and soft skill combinations requested by employers. Parallel to the industry collection, we compiled a corpus representing the academic provision of skills. We systematically harvested N = 4,500 Semester Lesson Plans (Rencana Pembelajaran Semester - RPS) and detailed course descriptions from the open digital repositories of 50 Indonesian universities.

The sampling strategy for universities was purposive and stratified based on accreditation status. We selected institutions ranked *Unggul* (Excellent) or *Baik Sekali* (Very Good) by the National Accreditation Board for Higher Education (BAN-PT). This selection criterion introduces a deliberate best-case scenario bias. By focusing on elite and top-tier institutions, we establish a high baseline for comparison. The underlying logic is that if a significant skills mismatch and velocity gap are observable in the curricula of the nation's premier universities, the disparity is likely far more pronounced in lower-tier (Rank C or unaccredited) institutions. Thus, this bias strengthens the study's diagnostic power regarding systemic issues. The academic documents were selected to cover study programs corresponding to the five industry sectors, ensuring a valid domain-to-domain comparison.

Raw text data from job ads and syllabi is inherently noisy, unstructured, and linguistically complex. To transform this unstructured text into a machine-readable format suitable for statistical analysis, both corpora underwent a rigorous Natural Language Processing (NLP) pipeline. This phase was critical to resolving the specific linguistic challenges of the Indonesian professional context. The first stage involved tokenization, breaking down continuous text streams into individual semantic units (tokens). We applied standard normalization techniques, including the conversion of all characters to lowercase to ensure case-insensitivity and the removal of punctuation and special characters. A unique challenge in this study was the prevalence of code-switching—the linguistic phenomenon where Indonesian speakers integrate English technical terms into Indonesian morphological structures. For example, terms like mem-manage (to manage) or di-deploy (deployed) are common in tech job ads. Standard tokenizers often treat these as unique, unknown words. We implemented a custom regex-based splitter to handle these hybrid forms, stripping Indonesian affixes (mem-, di-, -kan) from English roots to ensure that managing, manage, and mem-manage were all mapped to the same semantic concept.

Standard stop-word lists (containing common words like dan, yang, the, and) are insufficient for domain-specific analysis. We developed custom stop-lists iteratively; (1) For the Industry Corpus: We removed generic HR boilerplate language (equal opportunity employer, competitive salary, health insurance) to isolate competency-based keywords; (2) For the Academic Corpus: We filtered out administrative pedagogic jargon (mid-term exam, credit units, classroom attendance) that describes the logistics rather than the content of learning. Crucially, however, we established a protection list for normative terms. Words such as ethics, critical thinking, collaboration, and integrity were explicitly retained. This decision was methodological: we sought to prevent the accidental erasure of soft skills and civic educational goals, which are often less frequent than technical keywords but ontologically vital to the university's mission.

To further reduce vector sparsity, we employed a hybrid lemmatization pipeline. Because the corpora contained a mix of Indonesian and English, a single-language library was inadequate. We utilized the Sastrawi library to validate and stem Indonesian roots. Tokens not recognized by Sastrawi were subsequently passed through the NLTK WordNet Lemmatizer (English). This sequential approach allowed us to accurately reduce analyzing (English) to analysis and *penganalisaan* (Indonesian) to *analisa*, facilitating cross-lingual mapping where possible and reducing the dimensionality of the feature space.

Following preprocessing, the transformed data were subjected to three distinct analytical techniques to quantify the mismatch. To identify the most distinguishing keywords for each corpus, we calculated the Term Frequency-Inverse Document Frequency (TF-IDF) scores. Unlike simple frequency counts, which bias common words, TF-IDF weighs a term's importance by penalizing words that appear ubiquitously across all documents. The mathematical formulation weighs the frequency of a term in a specific document against the inverse of its frequency across the entire corpus. We calculated TF-IDF at the corpus level, treating the entire collection of job ads as one document and the collection of syllabi as another. This allowed us to extract the semantic signature of the labor market versus the university, highlighting which concepts are mathematically distinctive to each domain.

To uncover hidden thematic structures that may not be explicitly labeled, we employed Latent Dirichlet Allocation (LDA), a generative probabilistic model. LDA assumes that each document is a mixture of topics and that each topic is a mixture of words. We optimized the number of topics (K) by maximizing the Cv Coherence Score, a metric that evaluates the semantic similarity between high-probability words within a topic. Through iterative testing, we determined that K = 15 provided the optimal balance between topic granularity and interpretability. This analysis allowed us to move beyond keyword matching to compare themes—for instance, seeing if the theme of cloud computing in industry semantically overlaps with the theme of distributed systems in academia.

The core novelty of our methodology lies in the application of Graph Theory. We constructed two co-occurrence networks. In these graphs, nodes represent individual skills (such as Python and Project Management), and edges represent their co-appearance in a single document (such as a job ad requiring both Python and SQL). We calculated Degree Centrality to measure the popularity of a skill (how often it connects to others) and Betweenness Centrality to measure a skill's bridging capability. A high betweenness score indicates a skill that connects disparate clusters (connecting Marketing skills to Tech skills). We used Gephi 0.9.2 for computation and visualization. Given the massive size of the industry corpus (over one million documents), the resulting graph was extremely dense. To resolve the hairball visualization effect and filter out noise (random, weak associations), we applied a backbone pruning threshold, retaining only the top 10% of edges based on edge weight (frequency of co-occurrence). This revealed the skeleton of the skill ecosystem.

A critical methodological challenge was the disparity in dataset size (1 million ads vs. 4,500 syllabi). Standard similarity metrics like the Jaccard Index are sensitive to the size of the union of sets. The Jaccard calculation divides the intersection of the sets by the union of the sets. If the Industry vocabulary is vastly larger than the Academic vocabulary, the Jaccard index will be artificially low, even if Academia teaches everything it possibly can. To correct for this, we utilized the Overlap Coefficient (Szymkiewicz–Simpson coefficient). The Overlap Coefficient is defined as the size of the intersection divided by the size of the smaller set. By dividing the intersection by the size of the smaller set (Academia), this metric asks a fairer question: Of the concepts the University can teach, what proportion is relevant to Industry? This ensures that the measured mismatch is a result of content divergence, not merely a difference in the volume of data generated.

3. Results and Discussion

The initial frequency analysis exposed a stark contrast in terminology. The academic corpus was dominated by cognitive verbs such as understand, describe, and theory, whereas the industry corpus was dominated by operative and technical nouns, including Python, SEO, and Audit Compliance (Table 1). While general competencies such as Communication show high overlap, technical domains show a critical divergence. The Overlap Coefficient of 0.05 for Python or SQL indicates that even within the technology clusters of universities, specific tool references are nearly absent compared to industry demands.

The Latent Dirichlet Allocation analysis revealed that industry topics are highly clustered around tools and frameworks, whereas academic topics are clustered around foundational knowledge phases. (1) Topic A (Industry - Tech Stack): Characterized by keywords including Cloud, AWS, Azure, Docker, and Kubernetes; (2) Topic B (Academia - IT Fundamentals): Characterized by keywords including Algorithm, Syntax, Logic, Flowchart, and Database System. While Topic B provides the foundation for Topic A, the direct translation is missing. We observed a 35% semantic gap; employers ask for the application of cloud infrastructure, while universities teach the logic of database storage.

RANK	INDUSTRY TERM	FREQUENCY (N)	ACADEMIC TERM	FREQUENCY (N)	OVERLAP COEFF.
1	Communication	892,100	Basic Theory	4,100	0.92
2	Data Analysis	765,400	Conceptual Framework	3,850	0.22
3	Project Management	650,200	History of [Subject]	3,200	0.15
4	Python or SQL	540,000	Introduction to...	3,100	0.05
5	Digital Marketing	480,000	Definition	2,900	0.08
6	English Proficiency	450,500	Grammar / Structure	2,850	0.75
7	Teamwork	420,000	Group Assignment	2,100	0.88
8	Problem Solving	380,000	Case Study	1,900	0.65
9	Agile or Scrum	310,000	Software Engineering	1,500	0.12
10	Customer Service	290,000	Public Relations	1,200	0.45

Table 1. Top 10 High-Frequency Terms by Corpus
Comparative Analysis of Industry Demand vs. Academic Supply (Overlap Coefficient)

The most significant finding stems from the network analysis. In the industry network, the node data analysis has a high betweenness centrality of 0.45. It connects distinct clusters, such as marketing and finance, indicating that data skills are transversal requirements across sectors. Conversely, in the Academic Network, data analysis appears as a peripheral node with a betweenness centrality of 0.08, isolated within the Statistics or Computer Science departments, failing to bridge into Social Sciences or Humanities curricula. The case of Digital Ethics is particularly salient. While industry demand growth is moderate at 15%, curriculum penetration is high at 60%. Rather than labeling this an efficiency failure, we interpret this as the university fulfilling its civic mandate by providing a corrective layer to the market.

Table 2. The Mismatch Matrix
Comparative Analysis of Industry Demand Growth vs. Academic Curriculum Penetration

SKILL CATEGORY	YOY DEMAND GROWTH (INDUSTRY)	CURRICULUM PENETRATION (ACADEMIA)	OVERLAP COEFF.	SOCIOLOGICAL INTERPRETATION / STATUS
Artificial Intelligence	+120%	5%	0.04	⚠ Velocity Lag
Digital Ethics	+15%	60%	0.18	Normative Shielding
Cloud Computing	+85%	8%	0.07	Vocational Gap
Classical Economics	-5%	90%	0.95	Oversupply
Strategic Leadership	+40%	15%	0.30	Hidden Curriculum

The empirical findings of this study, derived from a computational analysis of over one million data points, articulate a profound structural divergence between the demand-side signaling of the Indonesian labor market and the supply-side intent of higher education institutions. By moving beyond the limitations of perception-based surveys, we have mathematically mapped the skills ecosystem to reveal that the skills mismatch is not merely a catalogue of missing keywords, but a complex interplay of temporal velocity, ontological differences, and topological isolation. The most immediate and quantifiable mechanism driving the skills gap is what we term the velocity of relevance. Our network density analysis reveals two systems operating on fundamentally different temporal scales. The industry network functions as a high-frequency, adaptive system characterized by high clustering coefficients and short average path lengths. In this ecosystem, a new technological competency—such as Generative AI prompting or Kubernetes orchestration—operates like a contagion. Once introduced by market leaders, it propagates through the industry network with remarkable speed, typically achieving saturation within 3 to 6 months. This rapid diffusion is driven by the immediate exigencies of capital efficiency and competitive advantage; firms that fail to adopt new tools risk immediate obsolescence.[11]

In stark contrast, the academic network exhibits the structural rigidity typical of large, bureaucratic institutions. The low overlap coefficient identified in our results (0.22 for Data Analysis; 0.05 for specific tools like Python) should not be interpreted as a failure of academic quality or rigor. Rather, it is a failure of synchronicity. Indonesian higher education is governed by accreditation cycles (typically managed by BAN-PT) that span 4 to 5 years. A curriculum designed in 2020 to meet the standards of Industry 4.0 is, by 2024, historically accurate but technically outdated. The bureaucratic friction involved in updating a semester lesson plan (RPS)—which often requires approval from study program heads, faculty senates, and university administrators—renders the university incapable of matching the market's tempo. Consequently, graduates emerge from the university time capsule equipped with the skills of the previous half-decade, entering a market that has already moved on to the next iteration of technological tools.

Beyond the temporal lag, our textual analysis highlights a fundamental ontological mismatch in how the two sectors define competency. This is a linguistic divergence that reflects a deeper philosophical difference in purpose. The industry corpus is dominated by the language of tools. Employers treat specific software platforms and frameworks (Tableau, Salesforce, React.js) as proxies for competence.[12] To the employer, proficiency in a specific tool minimizes training costs and ensures immediate productivity. The mention of Tableau in a job ad is not just a request for visualization skills; it is a signal for a specific workflow and ecosystem familiarity.

Conversely, the university corpus utilizes the language of concepts. Syllabi are populated with invariant, durable terms such as data visualization theory, customer relationship management (CRM) Frameworks, or Object-Oriented Programming principles. This approach is rooted in the university's historic mission to teach transferable, enduring knowledge rather than perishable technical tactics.[13] However, this dichotomy creates a translation gap for the graduate. While a student may possess a robust theoretical understanding of CRM (the Concept), they often lack the tactical ability to navigate the Salesforce interface (the Tool) required on day one of employment.

This finding suggests that the unemployable graduate phenomenon is often a result of this mapping failure. The graduate possesses the deep structure of knowledge but lacks the surface-level signaling required to pass the applicant tracking systems (ATS) used by HR departments. Furthermore, while concepts are durable, the exclusive focus on them ignores the material reality of the digital economy: that tools shape cognition. Learning data science without fluency in Python/Pandas is increasingly akin to learning music theory without ever touching an instrument. The high betweenness centrality of technical nodes in the industry graph serves as empirical evidence that the market no longer distinguishes between the conceptual understanding of a task and the technical capacity to execute it.[14]

Perhaps the most structurally significant finding is the topological variance between the two networks. The academic network is characterized by high modularity, representing a siloed topology. This structure is a direct artifact of the departmental organization of the modern university.[15] Knowledge is produced and transmitted within distinct boundaries: Engineering students learn Python in the Faculty of Engineering; Business students learn Excel and SPSS in the Faculty of Economics. There are a few edges or connections linking these distinct disciplinary clusters. However, the industry network exhibits a merged or hybrid topology. In the labor market, the boundaries between disciplines have eroded. Our analysis shows strong edges connecting Finance nodes with Python and SQL nodes, indicating the rise of the FinTech sector, where financial literacy and coding proficiency are inseparable. Similarly, Marketing nodes are increasingly connected to data analytics and HTML/CSS nodes.[16]

The disconnect here is profound. The university trains specialists (such as a pure Accountant or a pure Programmer), but the market demands hybrids (such as an Accountant who can automate audits with Python). The lack of cross-disciplinary edges in the academic graph prevents students from acquiring these hybrid skill sets.[17] A student in a Social Science program, for instance, is structurally prevented from accessing data literacy courses housed in the Computer Science department due to prerequisite chains and credit restrictions. This topological rigidity results in graduates who are monolingual in a multilingual economy, unable to bridge the gap between domain expertise and digital fluency.[18]

While this study leverages a massive dataset to provide a macroscopic view of the skills ecosystem, it is not without limitations. First, our reliance on text mining restricts the analysis to explicit knowledge. The soft skills or socio-emotional competencies—such as adaptability, emotional intelligence, and leadership—are often assessed during the interview stage rather than explicitly detailed in job descriptions or syllabi. Consequently, our methodology may underrepresent the alignment that exists in these non-technical domains. Second, the study acknowledges a selection bias in the academic dataset. By analyzing syllabi from top-tier (accredited A and B) universities, we present a best-case scenario. It is highly probable that the mismatch is significantly more severe in lower-tier institutions that lack the resources to update curricula even periodically. Furthermore, the restriction to digital job portals (LinkedIn, JobStreet) biases the industry sample toward the formal, white-collar sector, potentially overlooking the skills dynamics of the informal economy or blue-collar vocations. Future research should aim to close the loop on this analysis by integrating outcome data. Specifically, combining syllabus text mining with longitudinal tracking of alumni career trajectories via LinkedIn profiles would allow researchers to measure the rate of skill acquisition post-graduation. This would enable us to quantify how much re-skilling a graduate must undertake in their first two years of employment to bridge the gap identified in this study.[19,20]

4. Conclusion

This study represents the first large-scale computational mapping of the skills mismatch in Indonesia, utilizing a dataset of unprecedented size to move the discourse beyond anecdotal evidence and perception-based surveys. By analyzing over one million industry data points against the codified intent of university curricula, we have quantified a structural divergence that poses a significant threat to Indonesia's utilization of its demographic bonus. The evidence points to a system in dissonance. We identified a velocity gap, where the labor market evolves at a rate approximately three times faster than the curricular adaptation cycle. We mapped an ontological divide, where the academy prioritizes durable concepts while the industry prioritizes perishable tools. Most critically, we revealed a topological failure, where the siloed structure of university departments fails to produce the hybrid, cross-functional talent required by a digital economy that no longer respects disciplinary boundaries.

The conclusion of this study is *not* that universities should abandon their mission of theoretical education to become vocational training centers for specific vendors. To do so would be to chase a moving target; the tools of today will be obsolete tomorrow. However, the status quo of business as usual is equally untenable. We propose a paradigm shift toward what we term an API-based Curriculum Model. In software engineering, an API (Application Programming Interface) allows a stable core system to interact with dynamic external applications. Similarly, the university curriculum should consist of two distinct layers: (1) The Core (The Kernel): Stable, credit-bearing modules focused on invariant theoretical concepts (such as Logic, Ethics, Systems Thinking, Statistical Theory). This layer preserves the civic and cognitive mission of the university; (2) The Interface (The Micro-credential): A dynamic outer layer of peripheral modules that can be updated in real-time without bureaucratic friction. These would be short-term, intensive certifications in specific tools (Python for Finance, Google Analytics, Agile Scrum Master) offered in partnership with industry.

We recommend that the Ministry of Education, Culture, Research, and Technology (Kemendikbud-Ristek) utilize the real-time labor market signaling methods demonstrated in this study to incentivize Cross-Disciplinary Bridge Nodes. Specifically, our network analysis identifies Data Literacy and Digital Project Management not as specialist skills, but as universal bridge nodes with high betweenness centrality across all sectors. Therefore, these must be transitioned from elective courses within STEM departments to general education requirements accessible to all students, regardless of their major. By structurally integrating these high-centrality skills and adopting a more agile curricular architecture, Indonesia can bridge the velocity gap, ensuring that its demographic bonus translates into a dividend of sustained economic productivity.

5. References

1. Heriel E. Skills mismatch and graduates' unemployment in Geita region. Account Bus Rev. 2025; 17(2).
2. Islam MT, Islam MH, Rahman MKH. Skills mismatch in business education: Key stakeholders' perspectives. Ind High Educ. 2025; (09504222251399411).
3. Bischof S. Education and skill mismatches in the German labour market: the role of vocational and occupational specificity from a career perspective. Oxf Econ Pap. 2025; (gpaf031).
4. Patel PC. Occupational skill mismatch in self-employment: prevalence and income implications. Int J Manpow. 2025; 46(10): 148–68.
5. Moloto A, Ramasimu NF, Motsei LL. Graduate unemployment, skills mismatch, and the dynamics of labour mobility in South Africa: a systematic literature review. International Journal of Applied Research in Business and Management (IJARBM). 2025; 6(5).
6. Potential of social innovation in addressing the jobs and skills mismatch in Haryana. J Informatics Educ Res. 2025.

7. Marois G, Potančoková M, Bezat A, Crespo Cuaresma J. Projecting labour market imbalances and skill mismatch under demographic change in the EU. Eur J Popul. 2025; 42(1): 4.
8. Hua R, Carmela S. Dizon Ph.D. Addressing skills mismatch between STEM education and manufacturing industry needs: a literature review. International Journal of Education and Social Development (IJESD). 2025; 2(2): 116–8.
9. Berryman AK, Bücker J, de Moura FS, Barbrook-Johnson P, Hanusch M, Mealy P, et al. Skill and spatial mismatches for sustainable development in Brazil. arXiv [econ.GN]. 2025.
10. Garibaldi P, Gomes P, Sopraseuth T. Output costs of education and skill mismatch in OECD countries. Econ Lett. 2025; 250(112278): 112278.
11. Emmanouil E, Gourzis K, Boukouvalas K, Gialis S. Investing in human capital in the era of European Universities: The uneven geography of skill mismatches across the regions of the European Reform University Alliance. Open Res Eur. 2025; 3: 166.
12. Mavrigiannakis K, Vasilatos A, Vella E. Fiscal tightening and skills mismatch. Eur Econ Rev. 2025; 174(104984): 104984.
13. Cvetkoska V, Trpeski P, Ivanovski I, Peovski F, İmrol MH, Babadoğan B, et al. Comparative analysis of skill shortages, skill mismatches, and the threats of migration in labor markets: a sectoral approach in North Macedonia, Türkiye, Ethiopia, and Ukraine. Soc Sci (Basel). 2025; 14(5): 294.
14. Narsaria R. Impact of automation on job market dynamics: Skill mismatch and unemployment. Int J Finance Manage Econ. 2025; 8(2): 167–70.
15. Surjya Das K. Navigating educational and skill mismatches: Wage implications and job satisfaction in urban Assam. Econ Polit Wkly. 2025; 60(33).
16. Mutuke M, Katsande DI. Leveraging technological advancements in technical and vocational education and training (TVET) institutions in Zimbabwe: Addressing skills mismatch in the technology-driven era. Eur J Appl Sci. 2025; 13(05): 149–65.
17. Ruf KAF. Skills or credentials? How skill specific and standardized vocational training moderates the wages of occupational mismatches. Res Soc Stratif Mobil. 2026; (101136): 101136.
18. Huang W, Jabbari J, Chun Y, Johnson O Jr. Can certificate programs solve the skills and spatial mismatch problem? Job portability and residential mobility in a coding and apprenticeship program. Urban Educ (Beverly Hills Calif). 2026; 61(3): 546–97.
19. Draissi Z, Zhanyong Q, Raguindin PZJ. Knowledge mapping of skills mismatch phenomenon: a scientometric analysis. High Educ Ski Work-based Learn. 2022; 12(2): 271–93.
20. Nelson A, Ivarsson A, Lydell M. Employability and long-term work life outcomes from studying at a Swedish university college: problematizing the notion of mismatch. High Educ Ski Work-based Learn. 2025; 15(7): 48–65.