Enigma in Education
https://enigma.or.id/index.php/edu

e-ISSN: 3026-5827

ORIGINAL ARTICLE
Enigma in Education | Vol. 4, No. 2 (2026) | pp. 32–49

Aligning Undergraduate Data Science Program Learning Outcomes with Indonesian Industry Competency Demands: An Exploratory Bilingual Content Analysis Based on ESCO and SKKNI Frameworks

Iis Sugandhi[1*], Amir Serikova[2]

[1]*Department of Art Sciences, Enigma Institute, Palembang, Indonesia*
[2]*Department of Cultural and Heritage Preservation, Emeritus Research Center, Skopje, North Macedonia*

ARTICLE INFO

Article history
Received: August 15, 2026
Revised: September 18, 2026
Accepted: October 14, 2026
Available online: October 28, 2026
Published: November 1, 2026

Keywords
content analysis; curriculum alignment; data science education; ESCO; SKKNI

Corresponding author
Iis Sugandhi
iis.sugandhi@enigma.or.id

ORCID
0009-0009-6338-3532

DOI
https://doi.org/10.61996/edu.v4i2.134

ABSTRACT

Background. Program Learning Outcomes (PLOs) in undergraduate degrees describe broad, macro-level competencies. Online job advertisements do something else. They list narrow tasks and named software tools in fine detail. Reading literal keyword overlap as graduate competence can misdiagnose curriculum gaps.

Objective. To quantify how explicitly industry competency concepts appear in the PLO documents of Indonesian undergraduate Data Science programs, using a bilingual mapping to the European Skills, Competences, Qualifications and Occupations (ESCO v1.2.0) taxonomy and the Indonesian National Work Competency Standards (SKKNI Data Science).

Methods. An exploratory computational content analysis covered 485 deduplicated Indonesian job advertisements collected in March–April 2025: JobStreet (271 descriptions across 210 employer clusters) and Glints (214 skill-tag postings). These were set against 64 PLO statements from six universities (46 in four dated primary curricula). Twenty-five operational domains were formed from 21 SKKNI units (consolidated into 17 workflow domains) and eight ESCO skill groups. Uncertainty was estimated with 4,000 employer-cluster bootstrap replications. Role divergence was tested with Fisher's exact tests and Benjamini–Hochberg false discovery rate (FDR) correction.

Results. Employer demand led with database and digital storage (68.6%), computer programming (62.4%), teamwork (60.9%), and professional communication (60.5%). Demand-weighted PLO coverage ranged from 52.5% (95% CI: 50.9–54.1%) to 73.8% (95% CI: 72.2–75.6%). Core analytical domains appeared in 4/4 primary curricula. Model evaluation (18.1%) and deployment (13.3%) did not (0/4 explicit detection).

Conclusion. Observed gaps are consistent with textual granularity asymmetry rather than graduate incompetence, an explanation not tested here, and offer curriculum committees an empirical agenda for hierarchical alignment.

Copyright: © 2026 The Author(s). Published by Enigma Institute on behalf of Enigma in Education. The authors retain copyright of their work.

Access license: This is an open access article distributed under the terms of the Creative Commons Attribution–NonCommercial–ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, sharing, adaptation and reproduction in any medium, provided the original author and source are credited and any derivative work is distributed under the same license.

1. Introduction

1.1 Constructive Alignment and Granularity Asymmetry in Data Science Higher Education

Undergraduate Bachelor of Data Science (S1 Sains Data) programs have multiplied quickly across Indonesian higher education. Behind this growth lies a national priority: supplying analytical talent for digital economy transformation, industrial automation, and evidence-based public governance.[1,2] Do the new curricula match changing labor market expectations? Answering that needs careful pedagogical and methodological framing. Outcome-based education (OBE) and Biggs' theory of constructive alignment offer the frame.[3] A curriculum is relevant when its levels fit together vertically. At the top sit macro-level Program Learning Outcomes (PLOs, known nationally as Capaian Pembelajaran Lulusan or CPL). Beneath them come meso-level Course Learning Outcomes (CLOs or Capaian Pembelajaran Mata Kuliah, CPMK), micro-level weekly instructional plans (Rencana Pembelajaran Semester, RPS), learning experiences, and authentic performance assessments.[4–9] Computational curriculum studies often share a limitation. Investigators take raw keyword overlap between university curriculum documents and commercial job advertisements as a direct measure of educational quality or graduate readiness.[10–13]

Job postings are transactional documents. Human-resource specialists and hiring managers write them to screen for immediate operational readiness, so they lean toward specific commercial software packages, programming languages, and granular workflow tasks.[10,12,13] PLO statements work differently. They are normative, long-horizon commitments meant to stay stable across five-year curriculum cycles. Their purpose is to build mathematical reasoning, statistical inference, algorithmic problem-solving, professional ethics, and the capacity for lifelong learning.[14–17] Some text-mining studies set macro-level PLOs directly against recruitment descriptions and ignore this difference in granularity. In those studies, high-level curriculum summaries that omit literal product names such as Python, SQL,

Tableau, or Docker are often misread as an instructional deficit.[18–21] A sound evaluation should first separate foundational conceptual domains, operational workflow stages, and transient technological tools. Only then can it support recommendations for curriculum reform.[22–26]

Global data science education faces a structural tension. Enduring academic foundations must be balanced against industrial toolsets that change quickly.[14–19,27,28] Curricular guidelines agree on the answer. The Association for Computing Machinery (ACM) Data Science Task Force competencies, the Park City Math Institute (PCMI) undergraduate framework, and ten-year retrospectives on computing in statistics curricula all stress the same point: an undergraduate degree should build deep mathematical, inferential, computational, and ethical fluency, not short-lived vocational training in proprietary software syntax.[14–17] Studies of online job advertisements in North America, Europe, China, and the Commonwealth of Independent States show a consistent pattern. Employers list dozens of software libraries, cloud vendors, and business intelligence platforms as screening filters. Yet these surface-level keywords cluster into a concise set of underlying cognitive and operational competency domains.[10–13] Over-reacting to raw keyword counts has a cost. Institutions that crowd macro-level PLO statements with tool names or narrow procedural tasks risk fragmenting curricular coherence and speeding up obsolescence. Biggs' framework suggests a different division of labor. Macro-level PLOs state integrative graduate capabilities. Stage-specific operational procedures and technological ecosystems belong to meso-level course outcomes (CPMK) and micro-level laboratory syllabi (RPS).[20–26]

1.2 Bridging National Competency Standards (SKKNI) and International Skill Ontologies (ESCO)

Researchers need an objective layer between academic learning outcomes and the varied prose of the labor market. Standardized competency taxonomies and occupational ontologies are increasingly used for that purpose.[29–32] In Indonesia, the Ministry of Manpower issued Decree No. 299 of 2020 (Kepmenaker No. 299 Tahun 2020). It established the Indonesian National Work Competency Standards (Standar Kompetensi Kerja Nasional Indonesia or SKKNI) for the Information and Communication Category, Artificial Intelligence Sub-field of Data Science.[29] The SKKNI Data Science framework codifies 21 official competency units, indexed from J.62DMI00.001.1 through J.62DMI00.021.1. Together they trace the whole industrial data science lifecycle. The sequence runs from business objective formulation and project planning to data collection, exploratory analysis, data validation, object selection, cleaning, feature construction, labeling, integration, modeling scenario design, statistical and artificial intelligence modeling, model evaluation, deployment, maintenance, project review, and final technical reporting.[29,33–35] SKKNI gives an authoritative national workflow sequence. Job advertisements in mixed Indonesian and English behave differently. They often bundle overlapping functional tasks, and they ask for broader foundational and transversal skills that cross unit boundaries: general computer programming, database administration, statistical mathematics, visual communication, teamwork, critical problem-solving, and data privacy stewardship.[36,37]

The European Skills, Competences, Qualifications and Occupations (ESCO v1.2.0) classification, developed by the European Commission, adds a second layer. It supports structured cross-lingual grouping of concepts at the level-three skill hierarchy.[30] Joining the 21 SKKNI Data Science lifecycle units to eight complementary ESCO v1.2.0 skill groups produces a 25-domain bilingual crosswalk. With it, one can examine how Indonesian job market demands map onto undergraduate PLO statements.[29,30,38] SKKNI and ESCO supply the standardized conceptual vocabulary. The demand weights come from observed Indonesian job postings.[10–12] The labor market itself is not uniform. Data science work divides into distinct occupational tracks: Data Analytics and Business Intelligence (BI), Data Science and Artificial Intelligence (AI), Data Engineering, and Business Analysis. A fair assessment of curriculum alignment therefore calls for role-stratified disaggregation, together with cross-platform sensitivity analysis.[39–43]

This multi-tiered curricular architecture is written into Indonesian regulation. The Indonesian National Qualifications Framework (Kerangka Kualifikasi Nasional Indonesia or KKNI Level 6 for undergraduate Bachelor programs) and the National Higher Education Standards (Standar Nasional Pendidikan Tinggi or SN-Dikti) require every study program to formulate an explicit graduate profile (Profil Lulusan). Each program must also map its PLOs across foundational attitudes, general skills, specialized knowledge, and specific occupational skills.[33–37,44] The SKKNI Data Science standard links academic qualifications to workforce readiness through its 21 competency units.[29] A risk exists, however. Indonesian vocational and higher education alignment studies report that copying SKKNI unit titles straight into university PLO tables, without pedagogical synthesis, can yield rigid, checklist-driven curricula. Such curricula hide cross-cutting analytical and communicative competencies.[33–36] Pairing SKKNI with the ESCO v1.2.0 classification of the European Commission addresses the problem. The national workflow sequence is matched with internationally standardized level-three skill groups. Operational lifecycle stages and foundational academic pillars can then be evaluated together, in two languages.[29,30]

1.3 Study Rationale, Research Questions, and Analytical Objectives

Undergraduate Data Science programs in Indonesia have grown quickly since 2020. Systematic empirical studies are still scarce, however. Few studies map the explicit textual representation of SKKNI and ESCO competency domains in institutional PLO documents against job advertisements from several Indonesian platforms.[4–9,29,30,38] Under the Indonesian National Qualifications Framework (KKNI Level 6) and the National Higher Education Standards (SN-Dikti), PLOs are the foundational contract between an academic program and society. They guide the later design of course syllabi, laboratory practicums, and capstone assessments.[20–23] Competency representation can be evaluated strictly at the macro-level PLO tier. Doing so while accounting for employer-level clustering in job advertisements, and while keeping conceptual domains apart from commercial software brand names, gives curriculum committees an objective diagnostic baseline for constructive alignment.[45–48]

This exploratory cross-institutional computational content analysis addresses four research questions. (1) To what extent are the 25 operational competency domains demanded in the Indonesian data science job market explicitly represented in the official PLO statements of undergraduate Bachelor of Data Science programs? (2) How strongly do foundational analytical competencies and operational machine learning lifecycle stages (such as model evaluation, deployment, feature engineering, data cleaning, and monitoring) diverge in their explicit PLO representation? (3) How much do competency demands vary across occupational role categories (specifically Data Science/AI versus Data Analytics/BI)? How do literal software tool mentions differ between full-text job descriptions (JobStreet) and tag-based recruitment formats (Glints)? (4) How stable are demand-weighted curriculum alignment indices when senior job titles are excluded, role definitions are restricted, or cross-platform skill-tag weights are substituted? The answers provide an empirical baseline for multi-tier curriculum alignment audits in Indonesian higher education.

2. Methods

2.1 Study Design and Analytical Scope

The study used an exploratory quantitative computational content analysis based on bilingual concept vector

alignment.[10,29,30,38] The observational units were unique, deduplicated online job advertisements and individual institutional PLO statements. The primary analytical unit for curriculum comparison was the undergraduate degree program (study program level). Universities offering an undergraduate Bachelor of Data Science (S1 Sains Data) degree were chosen purposively. They represented public technical institutes, state universities, Islamic state universities, defense-affiliated state universities, and private outcome-based universities across Java and Sumatra.[4–9]

Both the labor market corpus and the curriculum documents were observed cross-sectionally. There was no longitudinal tracking and no direct testing of student competency. The analysis therefore measures explicit curricular concept representation, not instructional effectiveness or graduate employability. Text normalization, deduplication, bilingual dictionary matching, non-parametric cluster bootstrap resampling, and exact contingency table analyses were all run in Python 3.12 with pandas, NumPy, SciPy, and matplotlib.[38]

Treating curriculum alignment as a bilingual concept-vector comparison requires clear construct boundaries across the vertical hierarchy of Outcome-Based Education (OBE).[20–23] Where national qualification frameworks apply, curricular documents sit at three nested levels of abstraction. Macro-level Program Learning Outcomes (CPL) define program-wide graduate capabilities that are verified at degree completion. Meso-level Course Learning Outcomes (CPMK) specify the subject-specific competencies reached within semester-long courses. Micro-level weekly instructional plans (RPS) detail instructional topics, laboratory exercises, and software environments.[21,24–26] Evaluating alignment at the macro PLO tier asks one question: are high-demand occupational domains explicitly anchored in the top-level educational charter of a study program? Both the labor market demand distribution and each institution's PLO profile were represented as structured 25-dimensional vectors in a shared SKKNI–ESCO competency space. This separates conceptual domain coverage from transient instructional tooling, and it keeps the mathematics transparent across all institutional and occupational strata.[22,38]

2.2 Labor Market Corpus Selection, Eligibility Screening, and Role Stratification

Labor market data came from an open-access Indonesian data science recruitment corpus (Zenodo record 18162371; CC BY 4.0). It holds job advertisements collected between 17 March and 23 April 2025 from two major Indonesian online recruitment platforms, JobStreet Indonesia and Glints.[38] The corpus is a cross-sectional window on early 2025 recruitment demand in Indonesia's digital and analytical sectors. Primary outcome weighting used only verified job title and description text, not auxiliary salary or experience fields, to protect construct validity.[38]

The initial dataset held 852 job advertisement records. Of these, 433 came from JobStreet with full narrative job descriptions, and 419 came from Glints with structured skill-tag lists. A four-stage deterministic screening and deduplication protocol preceded the lexical analysis. First, platform identifiers were normalized using canonical job IDs on JobStreet and universally unique identifiers (UUIDs) on Glints. When the same identifier appeared in several collection batches, the most complete text record was kept. This step removed 173 duplicate records and retained 679 unique records.

Second, role eligibility was screened with standardized title rules. They limited the scope to four core data-related tracks: Data Analytics and Business Intelligence (BI), Data Science and Artificial Intelligence (AI), Data Engineering, and Business Analysis. Advertisements for clerical data entry, manual data annotation, general software engineering, network administration, or non-technical management outside these tracks were excluded, which removed 184 records. Third, a completeness check removed two records with descriptions shorter than 100 characters or unreadable skill-tag strings. Fourth, exact content deduplication within each platform removed eight cross-listed postings. They shared identical employer names, job titles, and body text under different identifiers. The protocol left 485 eligible job advertisements (56.9% of the initial records). The primary analytical corpus held 271 JobStreet full-text descriptions from 210 unique employer clusters, including 39 titles with explicit seniority markers such as Senior, Lead, Principal, Manager, or Head. The comparator corpus held 214 Glints skill-tag postings. The two platforms were analyzed in separate strata, because full prose descriptions and compact skill tags differ in baseline lexical density.

2.3 Undergraduate Data Science Curriculum Sample and Program Learning Outcome (PLO) Scope

The curriculum sample held 64 official Program Learning Outcome (PLO/CPL) statements from six Indonesian undergraduate Bachelor of Data Science (S1 Sains Data) programs.[4–9] Comparability at the macro-curricular tier under KKNI Level 6 and SN-Dikti standards was a priority. The analysis therefore used only official program-level outcome statements. Downstream course catalogs, semester learning plans, and external appendix tables were excluded.[4–9]

The six programs were sorted into two analytical cohorts. The criterion was whether the curriculum documents named a multi-year implementation window that overlapped the 2025 recruitment period. The primary dated cohort had four programs and 46 PLO statements: Institut Teknologi Sepuluh Nopember (ITS; 2023–2028 curriculum; 9 PLOs), Universitas Negeri Surabaya (UNESA; 2024–2028 Transformative Curriculum; 14 PLOs), Universitas Insan Cita Indonesia (UICI; 2023 Outcome-Based Education Curriculum; 14 PLOs), and Universitas Pembangunan Nasional Veteran Jakarta (UPNVJ; 2024 Undergraduate Curriculum; 9 PLOs).[4–7] Two further institutions published official PLO statements without effective year ranges. They were Universitas Islam Negeri Sumatera Utara (UINSU; 7 PLOs) and Universitas Pembangunan Nasional Veteran Jawa Timur (UPNVJT; 11 PLOs). These two formed a supplementary undated cohort (18 PLOs), used to check sensitivity to different institutional drafting styles.[8,9] University names are reported only to identify the analyzed public curricula and are not used for commercial ranking or accreditation adjudication.

2.4 Bilingual SKKNI–ESCO Taxonomic Crosswalk and Rule-Based NLP Dictionary Construction

To evaluate concept representation systematically, a two-layered taxonomic crosswalk was constructed by integrating the Indonesian SKKNI Data Science standard (Kepmenaker No. 299/2020, unit codes J.62DMI00.001.1 to J.62DMI00.021.1) with the European Commission's ESCO v1.2.0 Skills-Occupations Matrix Tables (level-three skill groups).[29,30] Because several adjacent SKKNI competency units describe functionally continuous operational tasks that employers routinely combine in recruitment prose, four pairs of overlapping units were consolidated into unified operational domains: defining business objectives (001) and determining technical goals (002) were combined into Business Objectives and User Needs; evaluating model results (014) and reviewing modeling processes (015) were combined into Model Evaluation and Review; planning deployment (016) and executing model deployment (017) were combined into Model Deployment; and maintaining deployed models (018) and monitoring model performance (019) were combined into Model Maintenance and Monitoring. Combining these four consolidated pairs with the remaining 13 individual SKKNI units yielded 17 workflow domains.

SKKNI centers on specific lifecycle tasks and assumes foundational academic prerequisites. Eight complementary domains based on ESCO v1.2.0 level-three skill groups were therefore added. They cover core mathematical, computational, transversal, and governance competencies: Computer

Programming, Database and Digital Storage, Statistics and Mathematics, Data Visualization and Presentation, Professional Communication, Teamwork and Collaboration, Problem Solving and Critical Thinking, and Data Security and Privacy. The result was 25 operational research domains that bridge national regulatory standards and international skill ontologies.[29,30]

Each of the 25 domains was operationalized with a bilingual (Indonesian and English) regular-expression dictionary of phrases and synonyms. Strict lexical disambiguation rules guarded against false positives. A generic reference to artificial intelligence or machine learning did not earn credit for Model Evaluation and Review, Model Deployment, or Model Maintenance and Monitoring. Credit required explicit evaluation metrics, validation protocols, production deployment, serving pipelines, or post-deployment monitoring phrases. Coordinated compound phrases in Indonesian curriculum prose were parsed so that every constituent concept was credited correctly. One example is the structure matching "pengumpulan, penyimpanan, dan pengelolaan data" (collection, storage, and management of data). Eleven specific software and technology tools were tracked on a separate analytical panel: SQL, Python, Tableau, R, Power BI, Microsoft Excel, Cloud AWS/Azure/GCP, Apache Spark and Hadoop, TensorFlow and PyTorch, Large Language Models [LLM] and NLP, and Git. This kept the absence of commercial product names in macro-level PLO statements from distorting the conceptual alignment scores.

2.5 Statistical Analysis, Cluster Bootstrap Resampling, and Sensitivity Protocols

A binary mention indicator $M_{i,d} \in \{0, 1\}$ was recorded for every job advertisement $i \in \{1, ..., 271\}$ in the primary JobStreet corpus and every competency domain $d \in \{1, ..., 25\}$. The empirical demand weight w_d for domain d was defined as the total number of JobStreet postings mentioning that domain, $w_d = \sum_{i=1}^{271} M_{i,d}$, with the cumulative demand weight across all 25 domains summing to $W = \sum_{d=1}^{25} w_d = 2{,}019$ domain mentions. Each prevalence proportion $p_d = w_d / 271$ carries a 95% Wilson score binomial confidence interval (CI). For each undergraduate program k, the binary curriculum representation indicator $C_{k,d} \in \{0, 1\}$ took the value 1 when at least one PLO statement in program k matched the bilingual dictionary rules for domain d, and 0 otherwise.

The demand-weighted concept alignment index A_k (expressed as a percentage) for program k was calculated as $A_k = 100 \times (\sum_{d=1}^{25} w_d C_{k,d}) / (\sum_{d=1}^{25} w_d)$, and the corresponding textual representation gap was defined as $G_k = 100 - A_k$. A complementary geometric metric was the cosine similarity between the 25-element empirical demand weight vector w and the binary program representation vector c_k, computed as $\cos(\theta_k) = (w \cdot c_k) / (||w||_2 \, ||c_k||_2)$. Some organizations posted several advertisements, which creates within-employer correlation. To account for it, 95% confidence intervals for A_k came from 4,000 non-parametric cluster bootstrap replications. Each replication resampled the 210 unique JobStreet employer clusters with replacement and kept every posting of a sampled employer. Institutional PLO vectors were held fixed. The intervals reflect employer-sampling variability only.

Occupational role divergence within the primary JobStreet corpus was evaluated by comparing domain mention frequencies between Data Science and AI postings (n = 75) and Data Analytics and BI postings (n = 130). Twenty-five two-by-two contingency tables were analyzed with two-sided Fisher's exact tests. Sample odds ratios (OR), exact 95% confidence intervals, exact p-values, and Benjamini–Hochberg false discovery rate (BH-FDR) adjusted q-values were reported, with statistical significance defined at $q < 0.05$. Four sensitivity analyses then recalculated A_k: (i) after excluding the 39 JobStreet postings with senior title markers (n = 232 non-senior postings); (ii) restricting demand weights exclusively to Data Science and AI postings (n = 75); (iii) replacing JobStreet prose weights with Glints skill-tag weights (n = 214) without cross-platform pooling; and (iv) computing alignment estimates for the two supplementary undated PLO documents (UINSU and UPNVJT).

The estimators for binomial prevalence, demand-weighted alignment, vector cosine similarity, employer-cluster bootstrap variance, and multiple-testing correction are specified below. For each competency domain $d \in \{1, ..., 25\}$ with observed mention count w_d across N = 271 primary JobStreet postings ($\hat{p}_d = w_d/N$), the 95% Wilson score confidence interval ($z_{\alpha/2} = 1.96$) was computed as $[\hat{p}_d + z^2/(2N) \pm z\sqrt{(\hat{p}_d(1 - \hat{p}_d)/N + z^2/(4N^2))}] / [1 + z^2/N]$. Standard Wald normal approximations ($\hat{p}_d \pm z\sqrt{(\hat{p}_d(1 - \hat{p}_d)/N)}$), collapse to zero width at boundary counts ($w_d = 0$) and distort coverage at low prevalence ($w_d \leq 5$). The Wilson score interval stays well-behaved across the entire empirical domain spectrum, from Data Labeling (0/271 = 0.0%, 95% Wilson CI: 0.0%–1.4%) and Modeling Scenario and Design (1/271 = 0.4%, 95% CI: 0.1%–2.1%) up to Database and Digital Storage (186/271 = 68.6%, 95% CI: 62.9%–73.9%). For institutional curriculum alignment, let $W = \sum_{d=1}^{25} w_d = 2{,}019$ denote the total domain mentions across the primary JobStreet corpus, and let $S_k = \sum_{d=1}^{25} w_d C_{k,d}$ denote the numerator sum of empirical demand weights for domains explicitly detected ($C_{k,d} = 1$) in program k, yielding $A_k = 100 \times S_k/W$. In the non-parametric cluster bootstrap (B = 4,000 replications), each replication $b \in \{1, ..., 4{,}000\}$ drew a simple random sample with replacement of G = 210 employer clusters $\{g_1^*, ..., g_{210}^*\}$, reconstructed the resampled demand weight vector $w_d^*(b) = \sum_{m=1}^{210} \sum_{i \in g(m)^*} M_{i,d}$, and recomputed $A_k^*(b) = 100 \times (\sum_{d=1}^{25} w_d^*(b) C_{k,d}) / (\sum_{d=1}^{25} w_d^*(b))$, from which the 2.5th and 97.5th empirical percentiles defined the 95% confidence interval. For the m = 25 role-stratified two-by-two contingency tables comparing Data Science/AI ($n_1 = 75$, with a_d mentions and $b_d = 75 - a_d$ non-mentions) against Data Analytics/BI ($n_2 = 130$, with c_d mentions and $d_d = 130 - c_d$ non-mentions), the sample odds ratio was calculated as $OR_d = (a_d d_d) / (b_d c_d)$ alongside exact hypergeometric two-sided p-values, and ordered p-values $p_{(1)} \leq p_{(2)} \leq ... \leq p_{(25)}$ were transformed into Benjamini–Hochberg step-up false discovery rate q-values via $q_{(r)} = \min_{s \geq r} \{\min(1, (25/s)p_{(s)})\}$.

3. Results

3.1 Sample Selection Attrition, Occupational Role Composition, and PLO Distribution

Across the initial recruitment corpus of 852 job advertisements (433 JobStreet and 419 Glints), identifier normalization removed 173 duplicate entries (−20.3%), leaving 679 unique platform records. Subsequent title screening excluded 184 out-of-scope occupational titles (−21.6%), text completeness checks removed 2 incomplete records (−0.2%), and within-platform exact text deduplication eliminated 8 cross-listed duplicates sharing identical company, title, and description text (−0.9%), yielding 485 eligible job advertisements (56.9% of initial records; Table 1 and Figure 1A). Within the primary JobStreet full-description corpus (n = 271), postings originated from 210 distinct employer clusters (mean 1.29 ± 0.84 [standard deviation, SD] postings per employer; median 1.0 [interquartile range, IQR: 1.0–1.0]), and 39 advertisements (14.4%) carried senior-level title designations, leaving 232 non-senior advertisements (85.6%).

Occupational role stratification revealed marked structural differences between the two recruitment platforms (Table 1 and Figure 1B). In the primary JobStreet corpus (n = 271), Data Analytics and BI constituted nearly half of all eligible postings (130/271; 48.0%), followed by Data Science and AI (75/271; 27.7%), Data Engineering (65/271; 24.0%), and Business Analysis (1/271; 0.4%). In contrast, the comparator Glints skill-tag corpus (n = 214) contained a higher proportion of Data Analytics and BI (120/214; 56.1%) and Business Analysis (36/214; 16.8%), alongside Data Engineering (38/214; 17.8%), but a substantially smaller share of Data Science and AI postings (20/214; 9.3%). Across the combined 485 eligible postings,

Data Analytics and BI accounted for 250/485 (51.5%), Data Engineering for 103/485 (21.2%), Data Science and AI for 95/485 (19.6%), and Business Analysis for 37/485 (7.6%), demonstrating that aggregate Indonesian data job postings are heavily weighted toward analytical and business intelligence roles rather than specialized machine learning research.

The institutional curriculum sample comprised 64 official PLO statements across six undergraduate Bachelor of Data Science programs (mean 10.67 ± 2.88 PLOs per program; median 10.0 [IQR: 9.0–13.25]; Table 2 and Figure 1C). The primary dated comparison cohort comprised 46 PLO statements (71.9% of the curriculum sample) across four institutions: ITS (9 PLOs; 14.1%; 2023–2028 curriculum), UNESA (14 PLOs; 21.9%; 2024–2028 curriculum), UICI (14 PLOs; 21.9%; 2023 OBE curriculum), and UPN Veteran Jakarta (9 PLOs; 14.1%; 2024 curriculum). The supplementary undated cohort accounted for the remaining 18 PLO statements (28.1%) across two institutions: UIN Sumatera Utara (7 PLOs; 10.9%) and UPN Veteran Jawa Timur (11 PLOs; 17.2%).

Table 1. Multi-stage sample selection criteria and occupational role stratification across the primary JobStreet and comparator Glints recruitment corpora (collected 17 March–23 April 2025).

Analytical Panel	Sample Screening Stage / Occupational Role Track	JobStreet Corpus (n, %)	Glints Corpus (n, %)	Combined Sample (N, %)
Panel A: Sample Selection	Initial job advertisements retrieved prior to screening	433 (100.0%)	419 (100.0%)	852 (100.0%)
Panel A: Sample Selection	Duplicate platform records removed via identifier normalization	—	—	−173 (−20.3%)
Panel A: Sample Selection	Unique job records retained for occupational screening	—	—	679 (79.7%)
Panel A: Sample Selection	Out-of-scope occupational titles excluded	—	—	−184 (−21.6%)
Panel A: Sample Selection	Incomplete text records (<100 characters or unreadable tags) excluded	—	—	−2 (−0.2%)
Panel A: Sample Selection	Cross-listed identical postings removed via exact text deduplication	—	—	−8 (−0.9%)
Panel A: Sample Selection	Final eligible analytical sample (210 JobStreet employer clusters; 39 senior titles)	271 (100.0%)	214 (100.0%)	485 (56.9% of initial)
Panel B: Occupational Roles	Data Analytics and Business Intelligence (BI)	130 (48.0%)	120 (56.1%)	250 (51.5%)
Panel B: Occupational Roles	Data Science and Artificial Intelligence (AI)	75 (27.7%)	20 (9.3%)	95 (19.6%)
Panel B: Occupational Roles	Data Engineering	65 (24.0%)	38 (17.8%)	103 (21.2%)
Panel B: Occupational Roles	Business Analysis	1 (0.4%)	36 (16.8%)	37 (7.6%)
Panel B: Occupational Roles	Total eligible job postings by platform (100.0% column sum)	271 (100.0%)	214 (100.0%)	485 (100.0%)

Note: JobStreet full-text job descriptions (n = 271 across 210 hiring organizations) and Glints discrete skill-tag postings (n = 214) were analyzed in separate strata without cross-platform pooling due to structural differences in lexical density between narrative prose and discrete skill tags. PLO = program learning outcome; BI = business intelligence; AI = artificial intelligence.

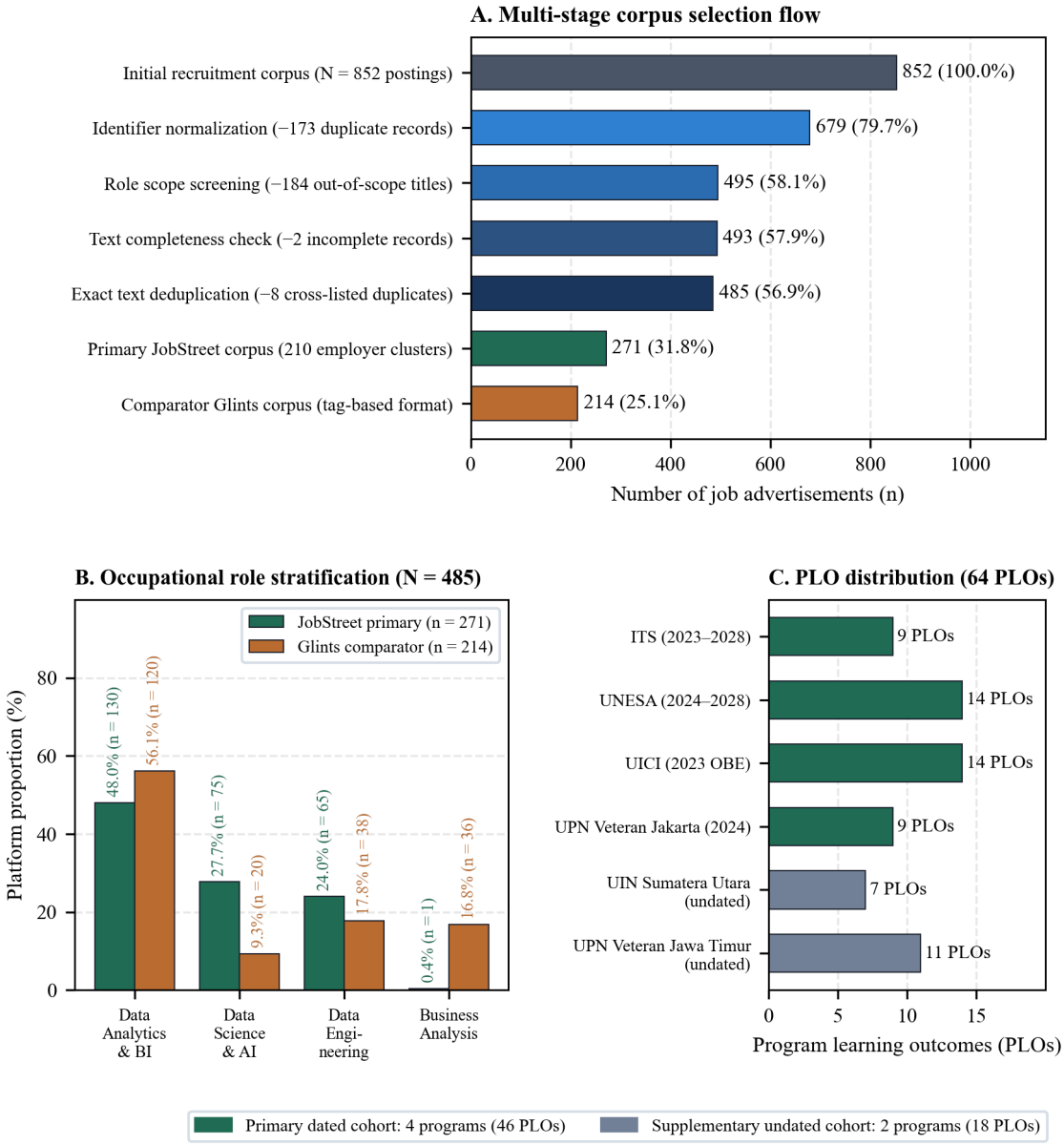


Figure 1. Multi-stage sample selection, occupational role stratification, and institutional Program Learning Outcome (PLO) distribution: (A) deterministic screening flow from 852 initial job advertisements (collected 17 March–23 April 2025) through identifier normalization (−173 duplicate records; 679 unique records), occupational title screening (−184 out-of-scope titles), text completeness verification (−2 records), and exact text deduplication (−8 cross-listed records) to 485 eligible postings (271 JobStreet primary full descriptions across 210 employer clusters; 214 Glints skill-tag postings), (B) occupational role stratification comparing JobStreet (48.0% Data Analytics & BI [n = 130], 27.7% Data Science & AI [n = 75], 24.0% Data Engineering [n = 65], 0.4% Business Analysis [n = 1]) against Glints (56.1% [n = 120], 9.3% [n = 20], 17.8% [n = 38], 16.8% [n = 36]), and (C) PLO statement counts across six Indonesian undergraduate Data Science programs (64 PLOs total: 46 PLOs in four dated primary curricula [ITS 9, UNESA 14, UICI 14, UPN Veteran Jakarta 9] and 18 PLOs in two undated supplementary curricula [UIN Sumatera Utara 7, UPN Veteran Jawa Timur 11]).

Table 2. Institutional profile, curriculum cohort classification, Program Learning Outcome (PLO) distribution, and core curricular focus of the six analyzed Indonesian undergraduate Bachelor of Data Science programs (64 PLOs total).

Institution (Undergraduate Study Program)	PLO State-ments (n, % of 64)	Curriculum Period	Analytical Cohort	Core Graduate Profile and Curricular Competency Focus
Institut Teknologi Sepuluh Nopember (ITS)	9 (14.1%)	2023–2028 Curriculum	Primary Dated Cohort	Statistical and mathematical modeling, algorithmic computer programming, big data integration, visual analytics, and scientific communication.
Universitas Negeri Surabaya (UNESA)	14 (21.9%)	2024–2028 Transforma-tive Curriculum	Primary Dated Cohort	Database management, exploratory data analysis, statistical and AI modeling, business problem alignment, teamwork, and data privacy.
Universitas Insan Cita Indonesia (UICI)	14 (21.9%)	2023 OBE Curriculum	Primary Dated Cohort	Outcome-based data science, business objective formulation, database storage, data quality validation, technical reporting, and ethics.
UPN Veteran Jakarta (UPNVJ)	9 (14.1%)	2024 Undergraduate Curriculum	Primary Dated Cohort	Mathematical statistics, database systems, AI modeling, data validation, information security, project management, and collaboration.

Institution (Undergraduate Study Program)	PLO Statements (n, % of 64)	Curriculum Period	Analytical Cohort	Core Graduate Profile and Curricular Competency Focus
Subtotal — Primary Dated Cohort (4 Programs)	46 (71.9%)	2023–2028 Active Windows	Primary Comparison Group	Evaluated for primary demand-weighted alignment, 4,000 employer-cluster bootstrap confidence intervals, and 25-domain coverage.
UIN Sumatera Utara (UINSU)	7 (10.9%)	Undated Institutional Curriculum	Supplementary Undated Cohort	Foundational mathematical statistics, exploratory data analysis, scientific problem solving, and professional communication.
UPN Veteran Jawa Timur (UPNVJT)	11 (17.2%)	Undated Institutional Curriculum	Supplementary Undated Cohort	Applied data science workflow, computer programming, database systems, statistical modeling, visualization, and cross-functional teamwork.
Total Analyzed Curriculum Sample (6 Programs)	64 (100.0%)	4 Dated + 2 Undated Programs	Complete Curriculum Sample	Represents macro-level graduate learning outcomes (KKNI Level 6 / SN-Dikti CPL) across six Indonesian public and private universities.

Note: To maintain analytical comparability at the macro-curricular tier (KKNI Level 6), evaluation was restricted strictly to official program-level learning outcome statements (CPL), excluding meso-level course catalogs and micro-level weekly syllabi. PLO = program learning outcome; ITS = Institut Teknologi Sepuluh Nopember; UNESA = Universitas Negeri Surabaya; UICI = Universitas Insan Cita Indonesia; UPNVJ = UPN Veteran Jakarta; UINSU = UIN Sumatera Utara; UPNVJT = UPN Veteran Jawa Timur; KKNI = Indonesian National Qualifications Framework.

3.2 Employer Demand Distribution Across 25 Competency Domains and Explicit PLO Coverage

When unweighted binary domain counts were examined across the 25 operational domains, the four primary dated programs detected a narrow range of 9 to 11 domains (mean 10.25 ± 0.96 domains; median 10.5 [IQR: 9.75–11.0]; Table 3): ITS and UICI each detected 11/25 domains (44.0%), UNESA detected 10/25 domains (40.0%), and UPN Veteran Jakarta detected 9/25 domains (36.0%). Across the 271 primary JobStreet descriptions, the bilingual NLP pipeline detected a cumulative total of 2,019 domain mentions (mean 7.45 ± 3.12 domains per advertisement; across the 25 domains, mean mention frequency was 80.76 ± 64.61 postings [29.80% ± 23.84%] and median was 78.0 [IQR: 20.0–141.0] postings [28.8% (IQR: 7.4%–52.0%)]). As detailed in Table 4 and Figure 2A, eight foundational and analytical domains were mentioned in more than half of all primary JobStreet postings: Database and Digital Storage ranked first at 186/271 (68.6%; 95% Wilson CI: 62.9%–73.9%), followed by Computer Programming at 169/271 (62.4%; 95% CI: 56.5%–67.9%), Teamwork and Collaboration at 165/271 (60.9%; 95% CI: 55.0%–66.5%), Professional Communication at 164/271 (60.5%; 95% CI: 54.6%–66.2%), Statistics and Mathematics at 157/271 (57.9%; 95% CI: 52.0%–63.7%), Data Visualization and Presentation at 148/271 (54.6%; 95% CI: 48.7%–60.4%), Statistical Modeling and AI at 141/271 (52.0%; 95% CI: 46.1%–57.9%), and Data Exploration and Analysis at 140/271 (51.7%; 95% CI: 45.7%–57.5%).

Intermediate demand prevalence was observed for Technical Reporting (132/271; 48.7%; 95% CI: 42.8%–54.6%), Problem Solving and Critical Thinking (104/271; 38.4%; 95% CI: 32.8%–44.3%), Data Integration (93/271; 34.3%; 95% CI: 28.9%–40.2%), Business Objectives and User Needs (78/271; 28.8%; 95% CI: 23.7%–34.4%), Data Collection (78/271; 28.8%; 95% CI: 23.7%–34.4%), and Data Quality and Validation (66/271; 24.4%; 95% CI: 19.6%–29.8%). Conversely, specialized SKKNI operational lifecycle stages exhibited lower explicit mention frequencies in job descriptions, including Model Evaluation and Review (49/271; 18.1%; 95% CI: 14.0%–23.1%), Model Deployment (36/271; 13.3%; 95% CI: 9.8%–17.8%), Feature Construction and Transformation (35/271; 12.9%; 95% CI: 9.4%–17.4%), Data Security and Privacy (30/271; 11.1%; 95% CI: 7.9%–15.4%), Data Cleaning (20/271; 7.4%; 95% CI: 4.8%–11.1%), Model Maintenance and Monitoring (13/271; 4.8%; 95% CI: 2.8%–8.0%), Project Planning and Management (10/271; 3.7%; 95% CI: 2.0%–6.7%), Data Object and Variable Selection (3/271; 1.1%; 95% CI: 0.4%–3.2%), Data Science Project Review (1/271; 0.4%; 95% CI: 0.1%–2.1%), Modeling Scenario and Design (1/271; 0.4%; 95% CI: 0.1%–2.1%), and Data Labeling (0/271; 0.0%; 95% CI: 0.0%–1.4%; Table 4).

Cross-tabulating employer demand against the four primary dated undergraduate curricula showed a clear contrast between core academic pillars and operational workflow stages (Table 4, Figure 2B, and Figure 3A). Four high-demand domains—Professional Communication (60.5% demand), Statistics and Mathematics (57.9%), Statistical Modeling and AI (52.0%), and Data Exploration and Analysis (51.7%)—achieved universal explicit representation across all four primary PLO documents (4/4; 100.0%). Five additional domains achieved explicit representation in three of the four curricula (3/4; 75.0%): Database and Digital Storage (68.6% demand; detected in UICI, UNESA, and UPNVJ), Teamwork and Collaboration (60.9%; ITS, UNESA, and UPNVJ), Problem Solving and Critical Thinking (38.4%; ITS, UICI, and UNESA), Data Quality and Validation (24.4%; UICI, UNESA, and UPNVJ), and Data Security and Privacy (11.1%; UICI, UNESA, and UPNVJ). Three domains were detected in two of the four curricula (2/4; 50.0%): Technical Reporting (48.7%; ITS and UICI), Business Objectives and User Needs (28.8%; UICI and UNESA), and Data Collection (28.8%; ITS and UICI). Four domains were explicitly detected in only one primary PLO document (1/4; 25.0%): Computer Programming (62.4%; ITS only), Data Visualization and Presentation (54.6%; ITS only), Data Integration (34.3%; ITS only), and Project Planning and Management (3.7%; UPNVJ only).

Table 3. Demand-weighted competency concept alignment, 4,000 employer-cluster bootstrap 95% confidence intervals, textual representation gaps, cosine similarity, and multi-scenario sensitivity across undergraduate Data Science programs.

Program Document (Cohort)	Domains Detected (k/25, %)	Primary Weighted Alignment (% [95% Bootstrap CI])	Textual Representation Gap (%)	Vector Cosine Similarity	Non-Senior Titles Only (n = 232)	DS & AI Role Only (n = 75)	Glints Skill-Tag Weights (n = 214)
ITS (Primary Dated; 9 PLOs)	11/25 (44.0%)	73.8% (72.2–75.6%)	26.2%	0.8762	74.3%	72.4%	73.8%
UICI (Primary Dated; 14 PLOs)	11/25 (44.0%)	63.2% (61.4–65.0%)	36.8%	0.7499	63.6%	60.4%	70.9%
UNESA (Primary Dated; 14 PLOs)	10/25 (40.0%)	61.0% (59.4–62.5%)	39.0%	0.7587	60.7%	63.0%	75.0%

Program Document (Cohort)	Domains Detected (k/25, %)	Primary Weighted Alignment (% [95% Bootstrap CI])	Textual Representation Gap (%)	Vector Cosine Similarity	Non-Senior Titles Only (n = 232)	DS & AI Role Only (n = 75)	Glints Skill-Tag Weights (n = 214)
UPN Veteran Jakarta (Primary Dated; 9 PLOs)	9/25 (36.0%)	52.5% (50.9–54.1%)	47.5%	0.6880	52.5%	52.9%	70.3%
Primary Cohort Summary (Mean ± SD; Median [IQR])	10.25 ± 0.96 (10.5 [9.75–11.0])	62.6% ± 8.76% (62.1% [58.9–65.9%])	37.4% ± 8.76%	0.7682 ± 0.0785	62.8% ± 9.01%	62.2% ± 8.05%	72.5% ± 2.26%
UIN Sumatera Utara (Supplementary Undated; 7 PLOs)	7/25 (28.0%)	41.8% (Point Estimate)	58.2%	0.6167	—	—	—
UPN Veteran Jawa Timur (Supplementary Undated; 11 PLOs)	14/25 (56.0%)	76.4% (Point Estimate)	23.6%	0.8891	—	—	—

Note: Primary weighted alignment equals 100 × (sum of JobStreet mention counts for domains detected in the program's PLOs) / 2,019 total domain mentions. 95% confidence intervals were computed via 4,000 non-parametric bootstrap replications resampling the 210 employer clusters on JobStreet. CI = confidence interval; SD = standard deviation; IQR = interquartile range; program abbreviations as in Table 2.

Table 4. Complete distribution of employer demand across the 25 operational competency domains in the primary JobStreet corpus (n = 271 postings; 2,019 total domain mentions), 95% Wilson confidence intervals, and institution-by-institution PLO detection.

Rank & Operational Competency Domain	JobStreet Mentions (n/271)	Prevalence (%)	95% Wilson Score CI (%)	Primary PLOs Detecting (k/4)	Institutional Detection Profile (ITS / UICI / UNESA / UPNVJ)
1. Database and digital storage	186/271	68.6%	62.9–73.9%	3/4 (75.0%)	0 / 1 / 1 / 1
2. Computer programming	169/271	62.4%	56.5–67.9%	1/4 (25.0%)	1 / 0 / 0 / 0
3. Teamwork and collaboration	165/271	60.9%	55.0–66.5%	3/4 (75.0%)	1 / 0 / 1 / 1
4. Professional communication	164/271	60.5%	54.6–66.2%	4/4 (100.0%)	1 / 1 / 1 / 1
5. Statistics and mathematics	157/271	57.9%	52.0–63.7%	4/4 (100.0%)	1 / 1 / 1 / 1
6. Data visualization and presentation	148/271	54.6%	48.7–60.4%	1/4 (25.0%)	1 / 0 / 0 / 0
7. Statistical modeling and AI	141/271	52.0%	46.1–57.9%	4/4 (100.0%)	1 / 1 / 1 / 1
8. Data exploration and analysis	140/271	51.7%	45.7–57.5%	4/4 (100.0%)	1 / 1 / 1 / 1
9. Technical reporting	132/271	48.7%	42.8–54.6%	2/4 (50.0%)	1 / 1 / 0 / 0
10. Problem solving and critical thinking	104/271	38.4%	32.8–44.3%	3/4 (75.0%)	1 / 1 / 1 / 0
11. Data integration	93/271	34.3%	28.9–40.2%	1/4 (25.0%)	1 / 0 / 0 / 0
12. Business objectives and user needs	78/271	28.8%	23.7–34.4%	2/4 (50.0%)	0 / 1 / 1 / 0
13. Data collection	78/271	28.8%	23.7–34.4%	2/4 (50.0%)	1 / 1 / 0 / 0
14. Data quality and validation	66/271	24.4%	19.6–29.8%	3/4 (75.0%)	0 / 1 / 1 / 1
15. Model evaluation and review	49/271	18.1%	14.0–23.1%	0/4 (0.0%)	0 / 0 / 0 / 0
16. Model deployment	36/271	13.3%	9.8–17.8%	0/4 (0.0%)	0 / 0 / 0 / 0
17. Feature construction and transformation	35/271	12.9%	9.4–17.4%	0/4 (0.0%)	0 / 0 / 0 / 0
18. Data security and privacy	30/271	11.1%	7.9–15.4%	3/4 (75.0%)	0 / 1 / 1 / 1
19. Data cleaning	20/271	7.4%	4.8–11.1%	0/4 (0.0%)	0 / 0 / 0 / 0
20. Model maintenance and monitoring	13/271	4.8%	2.8–8.0%	0/4 (0.0%)	0 / 0 / 0 / 0
21. Project planning and management	10/271	3.7%	2.0–6.7%	1/4 (25.0%)	0 / 0 / 0 / 1
22. Data object and variable selection	3/271	1.1%	0.4–3.2%	0/4 (0.0%)	0 / 0 / 0 / 0
23. Data science project review	1/271	0.4%	0.1–2.1%	0/4 (0.0%)	0 / 0 / 0 / 0
24. Modeling scenario and design	1/271	0.4%	0.1–2.1%	0/4 (0.0%)	0 / 0 / 0 / 0
25. Data labeling	0/271	0.0%	0.0–1.4%	0/4 (0.0%)	0 / 0 / 0 / 0

Note: Because a single job advertisement may mention multiple domains simultaneously, domain percentages do not sum to 100%. Across all 25 domains, mean ± SD mention frequency is 80.76 ± 64.61 postings (29.80% ± 23.84%), and median [IQR] is 78.0 [20.0–141.0] postings (28.8% [7.4–52.0%]). Binary profile indicates explicit phrase dictionary detection (1 = detected; 0 = not detected). CI = confidence interval; k/4 = number of the four primary dated PLO documents in which the domain was detected; ITS, UICI, UNESA, and UPNVJ as in Table 2.

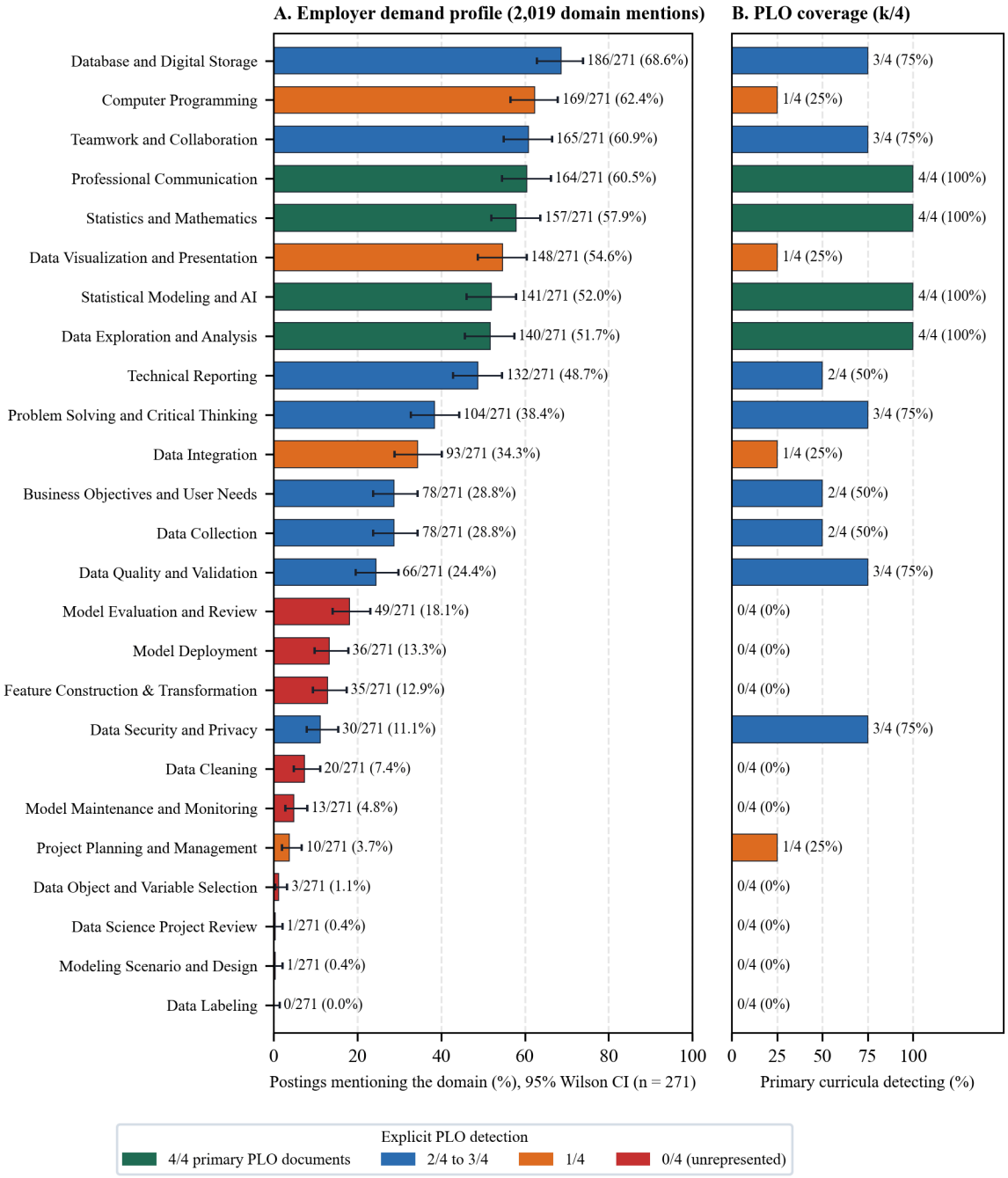


Figure 2. Empirical employer demand prevalence across the 25 SKKNI–ESCO operational competency domains in the primary JobStreet corpus (n = 271 postings; 2,019 total domain mentions) and explicit detection across the four primary dated PLO documents: (A) percentage of JobStreet advertisements mentioning each domain with horizontal error bars indicating 95% Wilson score binomial confidence intervals, color-coded by explicit institutional PLO detection tier, and (B) proportion of the four primary dated undergraduate curricula (k/4) explicitly representing each domain.

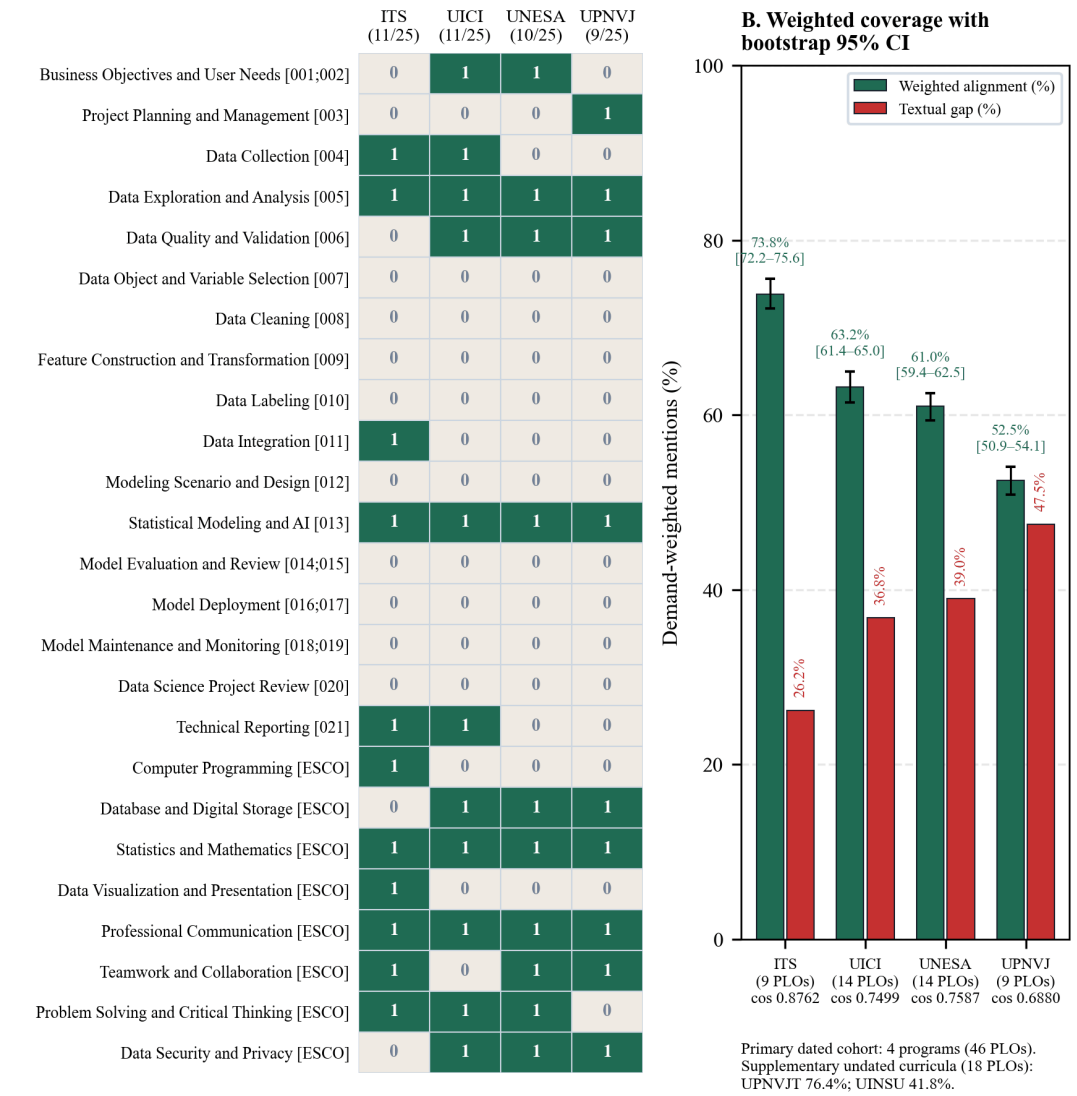


Figure 3. Institutional competency representation matrix and demand-weighted curriculum alignment across the four primary dated undergraduate Data Science programs (ITS, UICI, UNESA, and UPN Veteran Jakarta): (A) binary detection heatmap across the 25 SKKNI–ESCO domains ordered by the SKKNI workflow sequence (units 001–021) and ESCO v1.2.0 complementary groups (1 = explicitly detected via bilingual phrase dictionary; 0 = not detected), and (B) demand-weighted concept alignment percentages, 95% employer-cluster bootstrap confidence intervals (4,000 replications across 210 hiring organizations), textual representation gaps (100 − Alignment), and vector cosine similarities.

3.3 Demand-Weighted Alignment Indices, Bootstrap Uncertainty, and Unrepresented Lifecycle Stages

Weighting each competency domain by its empirical JobStreet mention frequency (W = 2,019) differentiated the undergraduate programs substantially based on whether high-prevalence computational and visualization domains were stated explicitly at the PLO level (Table 3 and Figure 3B). Demand-weighted concept alignment was highest for ITS at 73.8% (95% employer-cluster bootstrap CI: 72.2%–75.6%; textual representation gap: 26.2%; vector cosine similarity: 0.8762), followed by UICI at 63.2% (95% CI: 61.4%–65.0%; gap: 36.8%; cosine: 0.7499), UNESA at 61.0% (95% CI: 59.4%–62.5%; gap: 39.0%; cosine: 0.7587), and UPN Veteran Jakarta at 52.5% (95% CI: 50.9%–54.1%; gap: 47.5%; cosine: 0.6880). Across the four primary dated curricula, mean demand-weighted alignment was 62.6% ± 8.76% (median 62.1% [IQR: 58.9%–65.9%]), mean textual gap was 37.4% ± 8.76%, and mean cosine similarity was 0.7682 ± 0.0785.

Nine of the 25 competency domains (36.0%) exhibited zero explicit detection (0/4; 0.0%) across all four primary PLO documents (Table 4 and Figure 3A). As shown in Table 5 (Panel A) and Figure 4B, the five highest-demand unrepresented domains corresponded directly to operational machine learning engineering and data preprocessing stages codified in the SKKNI Data Science workflow: Model Evaluation and Review (SKKNI units 014 and 015; demanded in 49/271 postings or 18.1%, 95% CI: 14.0%–23.1%), Model Deployment (SKKNI units 016 and 017; 36/271 or 13.3%, 95% CI: 9.8%–17.8%), Feature Construction and Data Transformation (SKKNI unit 009; 35/271 or 12.9%, 95% CI: 9.4%–17.4%), Data Cleaning (SKKNI unit 008; 20/271 or 7.4%, 95% CI: 4.8%–11.1%), and Model Maintenance and Monitoring (SKKNI units 018 and 019; 13/271 or 4.8%, 95% CI: 2.8%–8.0%). The remaining four unrepresented domains—Data Object and Variable Selection (1.1%), Data Science Project Review (0.4%), Modeling Scenario and Design (0.4%), and Data Labeling (0.0%)—experienced minimal explicit mention rates in job advertisements as well.

Decomposing the demand-weighted alignment numerator $S_k = \sum_{d=1}^{25} w_d C_{k,d}$ relative to the total JobStreet demand denominator (W = 2,019 mentions) illustrates how specific institutional domain profiles shape overall alignment percentages in Table 3 and Figure 3B. For ITS (11/25 domains detected; $C_{ITS,d} = 1$ for Computer Programming [169], Teamwork [165], Professional Communication [164], Statistics and Mathematics [157], Data Visualization [148], Statistical Modeling and AI [141], Data Exploration and Analysis [140], Technical Reporting [132], Problem Solving [104], Data Integration [93], and Data Collection

[78]), the covered weight sum equals S_{ITS} = 169 + 165 + 164 + 157 + 148 + 141 + 140 + 132 + 104 + 93 + 78 = 1,491 (yielding 73.8% weighted alignment and 26.2% textual gap, with $\|c_{ITS}\|_2 = \sqrt{11} = 3.3166$ and cosine similarity 0.8762). For UICI (11/25 domains detected: Database and Digital Storage [186], Professional Communication [164], Statistics and Mathematics [157], Statistical Modeling and AI [141], Data Exploration and Analysis [140], Technical Reporting [132], Problem Solving [104], Business Objectives [78], Data Collection [78], Data Quality [66], and Data Security and Privacy [30]), the covered weight sum equals S_{UICI} = 1,276 (63.2% weighted alignment, 36.8% gap, cosine 0.7499). Although ITS and UICI both detect 11/25 domains (44.0% unweighted), ITS achieves a 10.6-percentage-point higher demand-weighted alignment (73.8% vs. 63.2%) because its PLOs explicitly state high-frequency computational and visual domains (Computer Programming [169], Teamwork [165], Data Visualization [148], and Data Integration [93]), whereas UICI explicitly states lower-frequency governance and business domains (Business Objectives [78], Data Quality [66], and Data Security and Privacy [30]). Similarly, UNESA (10/25 domains detected: Database [186], Teamwork [165], Communication [164], Statistics [157], Modeling [141], Exploration [140], Problem Solving [104], Business Objectives [78], Data Quality [66], and Security [30]) accumulates S_{UNESA} = 1,231 (61.0% alignment, 39.0% gap, cosine 0.7587), and UPN Veteran Jakarta (9/25 domains detected: Database [186], Teamwork [165], Communication [164], Statistics [157], Modeling [141], Exploration [140], Data Quality [66], Security [30], and Project Planning [10]) accumulates S_{UPNVJ} = 1,059 (52.5% alignment, 47.5% gap, cosine 0.6880).

Table 5. Operational lifecycle domains with highest employer demand that showed zero explicit detection (0/4) in the four primary undergraduate Data Science PLO documents, and role-stratified divergence between Data Science/AI (n = 75) and Data Analytics/BI (n = 130).

Panel A: Top Unrepresented Domains (0/4 PLOs)	JobStreet Mentions (n/271)	Demand Prevalence (%)	95% Wilson CI (%)	Explicit PLO Detection (k/4)
Model evaluation and review (SKKNI 014;015)	49/271	18.1%	14.0–23.1%	0/4 (0.0%)
Model deployment (SKKNI 016;017)	36/271	13.3%	9.8–17.8%	0/4 (0.0%)
Feature construction and transformation (SKKNI 009)	35/271	12.9%	9.4–17.4%	0/4 (0.0%)
Data cleaning (SKKNI 008)	20/271	7.4%	4.8–11.1%	0/4 (0.0%)
Model maintenance and monitoring (SKKNI 018;019)	13/271	4.8%	2.8–8.0%	0/4 (0.0%)
Panel B: Role Stratification (DS/AI vs. Analytics/BI)	**Data Science & AI (n = 75)**	**Data Analytics & BI (n = 130)**	**Odds Ratio (OR) [95% Exact CI]**	**Two-Sided Fisher p / BH-FDR q**
Statistical modeling and AI	65/75 (86.7%)	36/130 (27.7%)	OR = 16.97 (7.61–40.76)	p = 9.73 × 10-17; q < 0.001
Model deployment	28/75 (37.3%)	7/130 (5.4%)	OR = 10.47 (4.10–29.98)	p = 1.01 × 10-8; q < 0.001
Data visualization and presentation	27/75 (36.0%)	95/130 (73.1%)	OR = 0.21 (0.11–0.40)	p = 3.40 × 10-7; q < 0.001
Computer programming	60/75 (80.0%)	61/130 (46.9%)	OR = 4.52 (2.26–9.46)	p = 3.74 × 10-6; q < 0.001
Technical reporting	25/75 (33.3%)	87/130 (66.9%)	OR = 0.25 (0.13–0.47)	p = 5.04 × 10-6; q < 0.001

Note: In Panel B, two-sided Fisher's exact tests were conducted across all 25 domains with Benjamini–Hochberg false discovery rate (BH-FDR) adjustment; 9 of 25 comparisons yielded q < 0.05. An odds ratio (OR) > 1 indicates higher mention prevalence in Data Science/AI postings relative to Data Analytics/BI postings. OR = odds ratio; CI = confidence interval; q = Benjamini–Hochberg adjusted p value; DS = Data Science; AI = artificial intelligence; BI = business intelligence; k/4 as in Table 4.

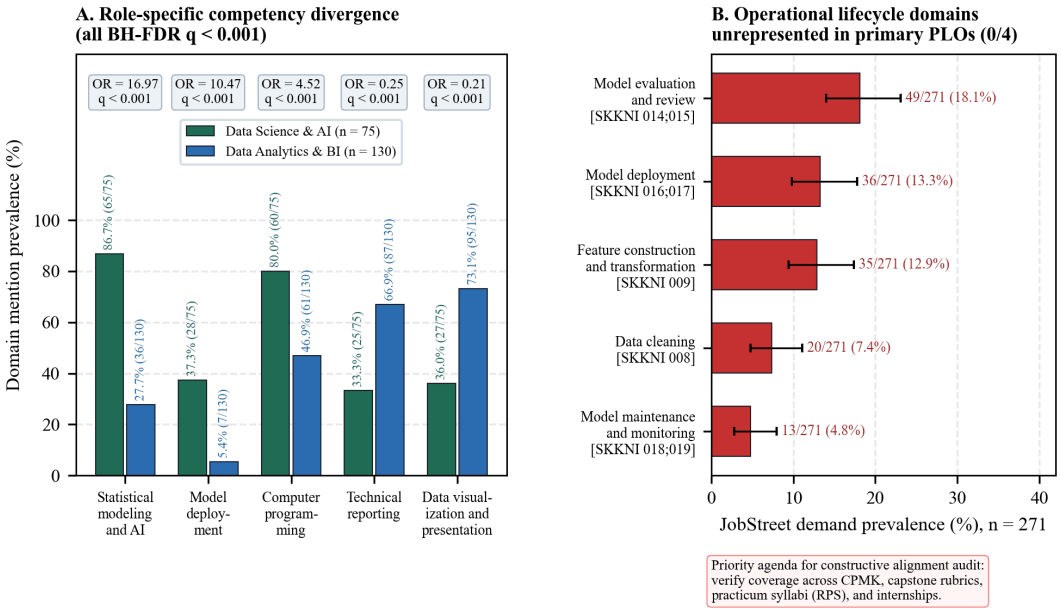


Figure 4. Occupational role divergence and unrepresented operational lifecycle stages in the primary JobStreet corpus: (A) comparison of domain mention prevalence between Data Science & AI postings (n = 75) and Data Analytics & BI postings (n = 130) across the five strongest discriminating domains (Statistical Modeling & AI [OR = 16.97], Model Deployment [OR = 10.47], Computer Programming [OR = 4.52], Technical Reporting [OR = 0.25], and Data Visualization & Presentation [OR = 0.21]; all two-sided Fisher's exact Benjamini–Hochberg FDR q < 0.001), and (B) the five highest-demand SKKNI operational workflow domains showing zero explicit detection (0/4) across the primary PLO documents (Model Evaluation & Review [18.1%], Model Deployment [13.3%], Feature Construction & Transformation [12.9%], Data Cleaning [7.4%], and Model Maintenance & Monitoring [4.8%]).

3.4 Occupational Role Divergence, Software Tool Granularity, and Multi-Scenario Sensitivity

Role-stratified contingency analyses comparing Data Science and AI postings (n = 75) against Data Analytics and BI postings (n = 130) within the primary JobStreet corpus demonstrated that 9 of the 25 domains differed significantly at a Benjamini–Hochberg FDR threshold of $q < 0.05$. As shown in Table 5 (Panel B) and Figure 4A, postings in the Data Science and AI track were substantially more likely to demand Statistical Modeling and AI (65/75 [86.7%] vs. 36/130 [27.7%]; OR = 16.97, 95% exact CI: 7.61–40.76, two-sided Fisher's exact $p = 9.73 \times 10^{-17}$, BH-FDR $q < 0.001$), Model Deployment (28/75 [37.3%] vs. 7/130 [5.4%]; OR = 10.47, 95% CI: 4.10–29.98, $p = 1.01 \times 10^{-8}$, $q < 0.001$), and Computer Programming (60/75 [80.0%] vs. 61/130 [46.9%]; OR = 4.52, 95% CI: 2.26–9.46, $p = 3.74 \times 10^{-6}$, $q < 0.001$). Conversely, Data Analytics and BI postings were significantly more likely to demand Data Visualization and Presentation (95/130 [73.1%] vs. 27/75 [36.0%]; OR = 0.21, 95% CI: 0.11–0.40, $p = 3.40 \times 10^{-7}$, $q < 0.001$) and Technical Reporting (87/130 [66.9%] vs. 25/75 [33.3%]; OR = 0.25, 95% CI: 0.13–0.47, $p = 5.04 \times 10^{-6}$, $q < 0.001$).

Evaluation of the 11-item software and technology tool panel confirmed a pronounced granularity contrast between recruitment advertisements and macro-level PLO statements (Table 6 and Figure 5A). In the primary JobStreet prose corpus (n = 271), SQL was explicitly requested in 161/271 postings (59.4%) and Python in 155/271 postings (57.2%), followed by Tableau (78/271; 28.8%), R (74/271; 27.3%), Power BI (73/271; 26.9%), Microsoft Excel (68/271; 25.1%), Cloud platforms AWS/Azure/GCP (66/271; 24.4%), Apache Spark and Hadoop (46/271; 17.0%), TensorFlow and PyTorch (27/271; 10.0%), Large Language Models and NLP (22/271; 8.1%), and Git (21/271; 7.7%). In the Glints skill-tag corpus (n = 214), SQL (107/214; 50.0%), Microsoft Excel (83/214; 38.8%), Python (53/214; 24.8%), and Power BI (38/214; 17.8%) dominated tag assignments, whereas specialized modeling tools appeared less frequently: Tableau (19/214; 8.9%), Spark and Hadoop (15/214; 7.0%), Cloud platforms (8/214; 3.7%), TensorFlow and PyTorch (4/214; 1.9%), R (3/214; 1.4%), Git (2/214; 0.9%), and LLM/NLP (1/214; 0.5%). Across all four primary dated undergraduate PLO documents, literal detection of these 11 commercial and open-source tool names was uniformly zero (0/4; 0.0%; Table 6).

Multi-scenario sensitivity testing demonstrated that curriculum alignment indices were highly robust to the removal of senior job titles but sensitive to role restriction and platform format (Table 3 and Figure 5B). Excluding the 39 senior-titled advertisements from JobStreet (n = 232 non-senior postings remaining) altered weighted alignment scores by at most 0.5 percentage points: ITS shifted from 73.8% to 74.3%, UICI from 63.2% to 63.6%, UNESA from 61.0% to 60.7%, and UPN Veteran Jakarta remained unchanged at 52.5% (cohort mean 62.8% ± 9.01%). Restricting demand weights exclusively to Data Science and AI postings (n = 75) yielded 72.4% for ITS, 60.4% for UICI, 63.0% for UNESA, and 52.9% for UPN Veteran Jakarta (cohort mean 62.2% ± 8.05%), reversing the relative rank order between UICI and UNESA. Substituting Glints skill-tag weights (n = 214), which emphasize general analytical and communication tags while under-representing specialized modeling stages, raised three of the four alignment scores (ITS stayed at 73.8%) and compressed all four into a narrow band between 70.3% and 75.0% (UNESA 75.0%, ITS 73.8%, UICI 70.9%, and UPNVJ 70.3%; cohort mean 72.5% ± 2.26%). Finally, in the supplementary undated curriculum cohort, UIN Sumatera Utara (7 PLOs; 7/25 domains; cosine 0.6167) achieved a primary weighted alignment of 41.8% (gap 58.2%), whereas UPN Veteran Jawa Timur (11 PLOs; 14/25 domains; cosine 0.8891) achieved 76.4% (gap 23.6%), spanning a 34.6-percentage-point range that underscores how strongly drafting specificity can shape text-based alignment metrics across the 25-domain SKKNI–ESCO crosswalk (Table 3 and Figure 5B).

Table 6. Cross-platform comparison of specific software and technology tool mentions in JobStreet full descriptions (n = 271), Glints skill tags (n = 214), and literal detection in the four primary undergraduate Data Science PLO documents.

Software / Technology Tool	Primary Functional Category	JobStreet Full Descriptions (n/271, %)	Glints Skill-Tag Postings (n/214, %)	Literal Detection in Primary PLOs (k/4)
SQL	Relational database querying & data warehousing	161/271 (59.4%)	107/214 (50.0%)	0/4 (0.0%)
Python	General-purpose data science & ML programming	155/271 (57.2%)	53/214 (24.8%)	0/4 (0.0%)
Tableau	Interactive business intelligence & visualization	78/271 (28.8%)	19/214 (8.9%)	0/4 (0.0%)
R	Statistical computing & econometric modeling	74/271 (27.3%)	3/214 (1.4%)	0/4 (0.0%)
Power BI	Enterprise dashboarding & business analytics	73/271 (26.9%)	38/214 (17.8%)	0/4 (0.0%)
Microsoft Excel	Spreadsheet analysis & business reporting	68/271 (25.1%)	83/214 (38.8%)	0/4 (0.0%)
Cloud Platforms (AWS / Azure / GCP)	Cloud data infrastructure & model hosting	66/271 (24.4%)	8/214 (3.7%)	0/4 (0.0%)
Apache Spark and Hadoop	Distributed big data processing	46/271 (17.0%)	15/214 (7.0%)	0/4 (0.0%)
TensorFlow and PyTorch	Deep learning & neural network frameworks	27/271 (10.0%)	4/214 (1.9%)	0/4 (0.0%)
Large Language Models (LLM) and NLP	Generative AI & natural language processing	22/271 (8.1%)	1/214 (0.5%)	0/4 (0.0%)
Git	Version control & collaborative software engineering	21/271 (7.7%)	2/214 (0.9%)	0/4 (0.0%)

Note: Specific software tools are analyzed on a separate panel from the 25 conceptual competency domains so that the absence of commercial or open-source product names in macro-level PLO statements is not misclassified as a curriculum gap. PLO = program learning outcome; k/4 as in Table 4; LLM = large language model; NLP = natural language processing.

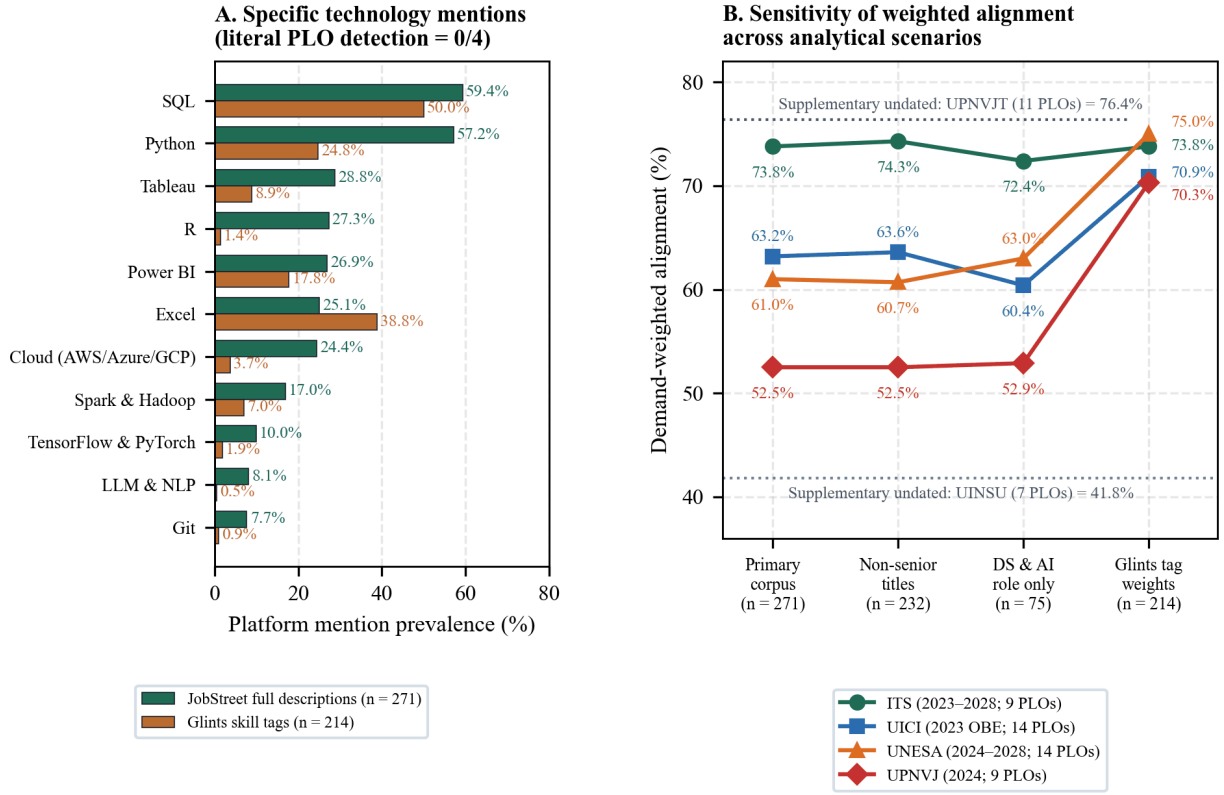


Figure 5. Cross-platform software tool granularity and multi-scenario alignment sensitivity analysis: (A) prevalence of 11 specific commercial and open-source technology tools in JobStreet full descriptions (n = 271) versus Glints skill tags (n = 214), all showing 0/4 literal detection in primary PLO statements (led by SQL [59.4% vs. 50.0%] and Python [57.2% vs. 24.8%]), and (B) sensitivity of demand-weighted alignment across four analytical scenarios (Primary JobStreet corpus [n = 271], Non-Senior JobStreet titles [n = 232], Data Science & AI role only [n = 75], and Glints skill-tag weights [n = 214]) alongside horizontal reference lines for the two supplementary undated curricula (UPN Veteran Jawa Timur [76.4%] and UIN Sumatera Utara [41.8%]).

Table 7. Thematic crosswalk linking the 25 operational research domains to Indonesian National Work Competency Standards (SKKNI Data Science, Kepmenaker No. 299/2020; prefix J.62DMI00) and European ESCO v1.2.0 level-three skill groups.

Operational Research Domain (English / Indonesian)	Related SKKNI Unit Code(s) (Prefix J.62DMI00.)	ESCO v1.2.0 Level-Three Skill Group Label	Primary Workflow Stage / Taxonomic Role
1. Business objectives and user needs (Tujuan bisnis dan kebutuhan pengguna)	001.1; 002.1	analysing business operations	SKKNI Stage 1: Business Understanding
2. Project planning and management (Perencanaan dan manajemen proyek)	003.1	directing, supervising and coordinating projects	SKKNI Stage 1: Project Scoping
3. Data collection (Pengumpulan data)	004.1	managing, gathering and storing digital data	SKKNI Stage 2: Data Acquisition
4. Data exploration and analysis (Eksplorasi dan analisis data)	005.1	analysing and evaluating information and data	SKKNI Stage 2: Exploratory Data Analysis
5. Data quality and validation (Kualitas dan validasi data)	006.1	managing, gathering and storing digital data	SKKNI Stage 2: Data Auditing
6. Data object and variable selection (Pemilihan objek dan variabel data)	007.1	managing, gathering and storing digital data	SKKNI Stage 3: Data Preparation
7. Data cleaning (Pembersihan data)	008.1	managing, gathering and storing digital data	SKKNI Stage 3: Data Preprocessing
8. Feature construction and transformation (Konstruksi fitur dan transformasi data)	009.1	managing, gathering and storing digital data	SKKNI Stage 3: Feature Engineering
9. Data labeling (Pelabelan data)	010.1	managing, gathering and storing digital data	SKKNI Stage 3: Ground-Truth Annotation
10. Data integration (Integrasi data)	011.1	managing, gathering and storing digital data	SKKNI Stage 3: Multi-Source Merging
11. Modeling scenario and design (Skenario dan desain pemodelan)	012.1	analysing and evaluating information and data	SKKNI Stage 4: Experimental Architecture
12. Statistical modeling and AI (Pemodelan statistik dan AI)	013.1	analysing and evaluating information and data	SKKNI Stage 4: Model Training & Estimation
13. Model evaluation and review (Evaluasi dan review model)	014.1; 015.1	evaluating systems, programmes, equipment and products	SKKNI Stage 5: Validation & Benchmarking
14. Model deployment (Deployment model)	016.1; 017.1	programming computer systems	SKKNI Stage 6: Production Deployment

Operational Research Domain (English / Indonesian)	Related SKKNI Unit Code(s) (Prefix J.62DMI00.)	ESCO v1.2.0 Level-Three Skill Group Label	Primary Workflow Stage / Taxonomic Role
15. Model maintenance and monitoring (Pemeliharaan dan monitoring model)	018.1; 019.1	evaluating systems, programmes, equipment and products	SKKNI Stage 6: Post-Deployment MLOps
16. Data science project review (Review proyek sains data)	020.1	directing, supervising and coordinating projects	SKKNI Stage 7: Project Retrospective
17. Technical reporting (Pelaporan hasil teknis)	021.1	presenting research or technical information	SKKNI Stage 7: Final Documentation
18. Computer programming (Pemrograman komputer)	ESCO Complementary	programming computer systems	Foundational Computational Pillar
19. Database and digital storage (Basis data dan penyimpanan digital)	ESCO Complementary	managing, gathering and storing digital data	Foundational Data Infrastructure Pillar
20. Statistics and mathematics (Statistika dan matematika)	ESCO Complementary	analysing and evaluating information and data	Foundational Mathematical Pillar
21. Data visualization and presentation (Visualisasi dan presentasi data)	ESCO Complementary	presenting research or technical information	Foundational Visual Analytics Pillar
22. Professional communication (Komunikasi profesional)	ESCO Complementary	communicating with colleagues and clients	Transversal Professional Competency
23. Teamwork and collaboration (Kerja sama dan kolaborasi)	ESCO Complementary	working in teams	Transversal Collaborative Competency
24. Problem solving and critical thinking (Pemecahan masalah dan berpikir kritis)	ESCO Complementary	solving problems	Transversal Cognitive Competency
25. Data security and privacy (Keamanan dan privasi data)	ESCO Complementary	protecting privacy and personal data	Governance & Ethical Stewardship Pillar

Note: The 21 official SKKNI Data Science competency units (Kepmenaker No. 299/2020, prefix J.62DMI00) were consolidated into 17 operational workflow domains where functional pairs overlapped in job advertisement prose (001+002, 014+015, 016+017, 018+019), and supplemented with 8 ESCO v1.2.0 foundational and transversal competency domains. SKKNI = Indonesian National Work Competency Standards; ESCO = European Skills, Competences, Qualifications and Occupations; ESCO complementary = domains not derived from SKKNI units.

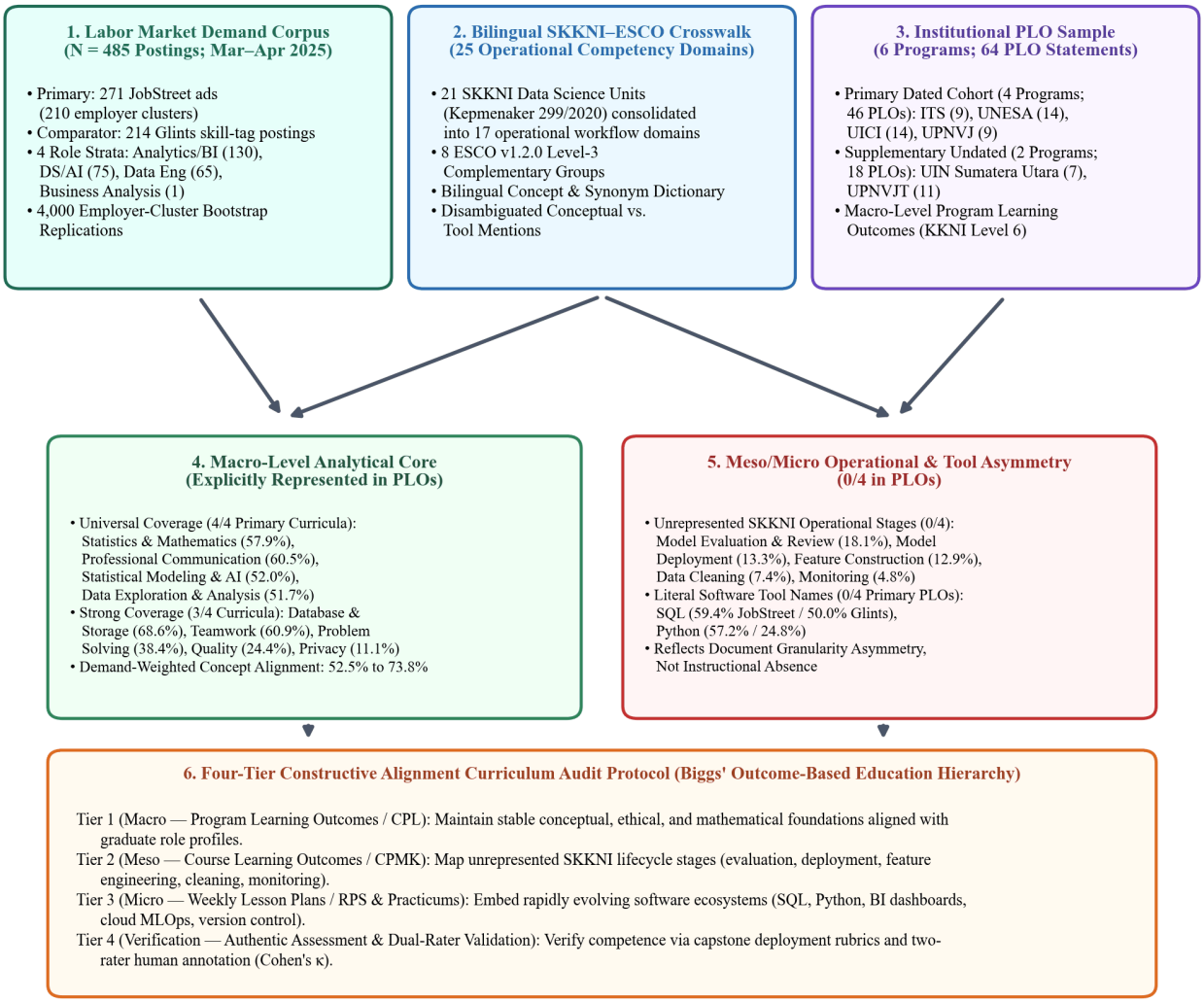


Figure 6. Bilingual SKKNI–ESCO Constructive Alignment Framework for Multi-Tier Data Science Curriculum Auditing: schematic architecture linking the deduplicated Indonesian job posting corpus (N = 485), the 25-domain SKKNI (Kepmenaker No. 299/2020) and ESCO v1.2.0 crosswalk, and the 64 PLO statements into a four-tier outcome-based education (OBE) audit protocol spanning Macro PLOs (CPL), Meso Course Learning Outcomes (CPMK), Micro Weekly Lesson Plans (RPS), and Authentic Capstone/Multi-Rater Verification.

4. Discussion

4.1 Principal Findings: Strong Analytical Foundations Versus Operational Lifecycle Asymmetry

This study examined how the PLO documents of Indonesian undergraduate Bachelor of Data Science programs represent industry competency demands. It used two frameworks together: the national SKKNI Data Science standards and the European ESCO v1.2.0 skill groups.[4–9,29,30,38] The main finding is clear. All four dated programs explicitly state the core mathematical, statistical, analytical, and communicative pillars that Indonesian employers demand. Four domains reached 100.0% (4/4) institutional representation: Statistics and Mathematics (57.9% job demand), Professional Communication (60.5%), Statistical Modeling and AI (52.0%), and Data Exploration and Analysis (51.7%). Five more reached 75.0% (3/4): Database and Digital Storage (68.6%), Teamwork (60.9%), Problem Solving (38.4%), Data Quality (24.4%), and Data Security and Privacy (11.1%). Demand-weighted concept representation in the primary dated curricula therefore ranged from 52.5% to 73.8%, and it reached 76.4% in the supplementary undated UPN Veteran Jawa Timur curriculum. Judged at the level of conceptual domains, not literal software keywords, Indonesian undergraduate Data Science programs explicitly represent 52.5% to 73.8% of weighted employer demand.[10–16]

The domain-by-domain breakdown also shows a structural asymmetry in explicit representation. Theoretical and analytical foundations are explicit in every PLO document. Operational machine learning lifecycle stages are not. Model Evaluation and Review (18.1% job demand), Model Deployment (13.3%), Feature Construction and Transformation (12.9%), Data Cleaning (7.4%), and Model Maintenance and Monitoring (4.8%) showed zero explicit detection (0/4) across the primary PLO documents (Table 5 and Figure 4B). International computing education literature calls this the operationalization or MLOps gap. Academic programs frame data science mainly around mathematical model estimation ($\hat{y} = f(X)$). Data engineering preprocessing, production deployment, containerization, and post-deployment drift monitoring stay implicit.[18,19,27,28,39–41] Within an outcome-based architecture, missing operational phrasing at the macro PLO tier does not mean students are never taught data cleaning, model validation, or deployment. It marks a hierarchical documentation boundary. The trail should be followed downward into meso-level Course Learning Outcomes (CPMK), laboratory syllabi (RPS), and capstone assessment rubrics (Figure 6).[20–26]

Why are the operational SKKNI workflow stages missing from macro-level PLO statements? Model Evaluation and Review (18.1%), Model Deployment (13.3%), Feature Construction (12.9%), Data Cleaning (7.4%), and Model Maintenance and Monitoring (4.8%) all fall into this group. An answer has to consider both how academic instruction is structured and how industrial machine learning operations (MLOps) work.[39–41] In traditional coursework, students may often use pre-cleaned, static benchmark datasets. Variable selection, schema harmonization, missing-value imputation, and feature scaling become short preliminary exercises, not core learning outcomes. Models may be graded on in-memory test-set accuracy, not on containerized production deployment, latency benchmarking, or post-deployment data-drift monitoring.[16–18,27,28,39–41] Drafting practice matters too. Faculty committees writing program-level CPL statements under national OBE guidelines may tend to fold all phases of the modeling lifecycle into umbrella phrases, such as "designing and applying statistical and artificial intelligence models to solve data problems." Under strict lexical matching, that phrasing credits Statistical Modeling and AI (SKKNI unit 013) and Data Exploration and Analysis (unit 005). It does not automatically credit evaluation (units 014–015), deployment (units 016–017), or maintenance (units 018–019). The 0/4 finding at the PLO tier thus marks a boundary of explicit representation. Curriculum teams can address it by checking whether these operational stages are taught and assessed in meso-level CPMK and capstone engineering modules.[20–26]

4.2 Occupational Role Divergence and the Risk of Single Composite Alignment Scores

The role-stratified contingency analysis (Table 5 Panel B and Figure 4A) gives a second insight. The Indonesian data science labor market splits into distinct occupational specializations. Data Analytics and Business Intelligence positions make up nearly half of all JobStreet openings (48.0%) and over half of Glints postings (56.1%). Their competency profile differs sharply from that of Data Science and AI roles (27.7% of JobStreet). Data Science and AI advertisements show nearly 17-fold higher odds of demanding Statistical Modeling and AI (OR = 16.97, $q < 0.001$). They show over 10-fold higher odds for Model Deployment (OR = 10.47, $q < 0.001$) and 4.5-fold higher odds for Computer Programming (OR = 4.52, $q < 0.001$). Data Analytics and BI postings show the reverse. Their odds are nearly five-fold higher for Data Visualization and Presentation (OR = 0.21, $q < 0.001$) and four-fold higher for Technical Reporting (OR = 0.25, $q < 0.001$).[11–13,42,43,45]

This specialization has direct implications for curriculum governance and accreditation policy in Indonesia. Suppose a program builds its graduate profile (Profil Lulusan) around business intelligence, visual analytics, and organizational decision support. Judging that program mainly against production MLOps deployment criteria in its top-level PLOs would not fit its institutional mission. A program that trains machine learning engineers or AI specialists faces the opposite problem, because generic data exploration statements are not enough for it. Sensitivity testing confirms the point (Table 3 and Figure 5B). Moving the demand weighting vector from aggregate JobStreet descriptions to Data Science/AI roles or to Glints skill tags changes the absolute alignment percentages and the rank order of institutions. UNESA, for example, rises from 61.0% to 75.0% under Glints tag weights and passes ITS at 73.8%. Single composite alignment scores should therefore not be read as institutional rankings. Alignment has to be judged against the graduate profile a program declares.[20,21,33–36,44]

The sensitivity gap between JobStreet full-text descriptions (n = 271) and Glints skill-tag postings (n = 214) shows how platform design shapes labor market analytics (Table 3, Table 6, and Figure 5).[10–13] JobStreet employers write multi-paragraph prose about day-to-day responsibilities. These advertisements averaged 7.45 detected domains and often named SQL (59.4%), Python (57.2%), Tableau (28.8%), R (27.3%), and cloud platforms (24.4%). On Glints, recruiters pick compact skill tags from standardized menus. The menus favor broad category labels such as general data analysis, business intelligence, communication, teamwork, SQL (50.0%), and Microsoft Excel (38.8%). Granular SKKNI workflow stages and specialized deep learning frameworks are rarely tagged (TensorFlow/PyTorch at 1.9% and LLM/NLP at 0.5%). When Glints tag frequencies replace the demand weights, universally represented analytical and communicative domains gain relative weight. All four primary curricula are compressed into a narrow 70.3%–75.0% band, and UNESA (75.0%) moves above ITS (73.8%). The contrast between platforms shows that alignment percentages depend jointly on the granularity of curriculum phrasing, on occupational role composition, and on the lexical density of the recruitment platform.[10,20,42,43,45–48]

4.3 Methodological Significance of the Bilingual SKKNI–ESCO Crosswalk and Constructive Alignment Protocol

Valid curriculum analytics depends on separating conceptual domains (Table 4 and Table 7) from specific software tool names (Table 6).[22,23,29,30,38] Scoring curricula on literal mentions of SQL (59.4% JobStreet demand) or Python (57.2%) would have given all four primary programs 0.0% alignment on those items (Figure 5A). Programs are likely to place fast-changing

software ecosystems in weekly practicum modules (RPS) and course-level outcomes (CPMK); this study did not examine those documents. Python, R, SQL, Tableau, Power BI, Apache Spark, TensorFlow, and Git are typical examples. Macro-level PLO statements then stay stable across multi-year accreditation cycles.[14–17,27,28] The 25-domain crosswalk pairs 21 SKKNI Data Science units (Kepmenaker No. 299/2020) with eight ESCO v1.2.0 level-three skill groups. It offers a standardized vocabulary that respects the abstraction hierarchy of higher education.[29,30]

The diagnostic findings can be turned into practice through a four-tier Constructive Alignment Curriculum Audit Protocol, which we propose here (Figure 6).[21,24–26] At Tier 1 (Macro/PLO), curriculum committees keep their conceptual, mathematical, statistical, and ethical commitments stable. They also clarify whether program-level computational and visualization competencies are stated explicitly or hidden under broader phrasing. At Tier 2 (Meso/CPMK), committees treat the five unrepresented high-demand SKKNI operational domains as a structured checklist: Model Evaluation and Review (18.1%), Model Deployment (13.3%), Feature Construction (12.9%), Data Cleaning (7.4%), and Model Maintenance and Monitoring (4.8%). The checklist tests whether explicit course-level outcomes exist for end-to-end data preprocessing and model validation. At Tier 3 (Micro/RPS), instructors map high-demand software tools (SQL, Python, BI dashboarding, and cloud version control) to concrete laboratory exercises. At Tier 4 (Verification), programs audit capstone project rubrics and internship portfolios to confirm that students show authentic operational competence. When curriculum mapping rules are validated, independent dual-rater human annotation adds to this audit, reporting Cohen's κ and precision/recall against a gold-standard corpus.[46–48]

Computational curriculum audits should support improvement, not reductive ranking. Two methodological safeguards help when text-mining indices are interpreted. Indonesian accreditation agencies can apply them, including the National Accreditation Board for Higher Education (BAN-PT) and the Independent Accreditation Agency for Informatics and Computer Programs (LAM INFOKOM). So can internal university quality assurance offices (Lembaga Penjaminan Mutu).[33–37,44] The first safeguard concerns document tracing. Automated dictionary alignment at the macro PLO level should be paired with a multi-tier trace across Course Learning Outcomes (CPMK), semester learning plans (RPS), practicum manuals, and capstone grading rubrics (Figure 6). Suppose an operational SKKNI competency such as Model Deployment or Data Cleaning is already taught and assessed in a mandatory course. The right response is then to strengthen vertical traceability links in the curriculum matrix, not to create redundant courses. The second safeguard concerns validation. Before any high-stakes policy use, bilingual curriculum dictionaries should undergo formal multi-rater expert validation. Independent domain specialists would annotate a stratified random sample of PLO–domain and posting–domain pairs on a four-level epistemic rubric: (1) Explicitly Represented, (2) Implicitly Entailed by Broader Construct, (3) Not Represented, and (4) Indeterminate/Ambiguous. Inter-rater reliability (Cohen's kappa [κ] and Krippendorff's alpha [α]) and precision, recall, and macro-F1 on a held-out evaluation set would then establish the psychometric construct validity that national curriculum benchmarking needs.[22,23,46–48]

4.4 Study Strengths and Methodological Limitations

The study has several strengths. Multi-stage deduplication reduced 852 recruitment records to 485 eligible postings. Macro-level PLO statements were kept clearly apart from meso-level course lists. The 210 employer clusters were handled through 4,000 bootstrap replications. Role comparisons used exact non-parametric tests with BH-FDR control, and sensitivity was tested across several scenarios. Several limitations also need attention when the findings are interpreted. First, the institutional sample is purposive: four primary dated curricula and two supplementary undated ones. The work is an exploratory multi-institutional analysis, not a probabilistic national census of Indonesian Data Science programs. Second, the recruitment corpus covers advertisements from two online portals (JobStreet and Glints) in one cross-sectional window (17 March–23 April 2025). It misses unadvertised hiring, civil service recruitment, and multi-year longitudinal shifts.

Third, rule-based bilingual phrase matching captures explicit lexical representation, not implicit semantic entailment. Highly abstract PLO phrasing, such as "developing intelligent systems," may produce false negatives for specific workflow stages. Generic terms may produce false positives when they are not disambiguated. Future studies should add independent multi-rater expert annotation to establish inter-rater reliability (κ) and out-of-sample classification performance. Fourth, the primary index scored only macro-level PLOs. Competencies taught and assessed in course syllabi (CPMK), weekly lesson plans (RPS), and capstone projects are absent from the macro-level numerator. Fifth, the differences between JobStreet and Glints reflect platform format (narrative prose versus discrete skill tags) and occupational role composition together. Sixth, employer-cluster bootstrap confidence intervals quantify resampling variability only for the recruitment corpus actually observed. Contingency tests across occupational tracks should be read as descriptive associations.

5. Conclusion

5.1 Summary of Findings and Future Directions

This exploratory bilingual content analysis shows a split. Foundational analytical, statistical, database, and communicative competencies, which the Indonesian job market demands, are strongly represented in undergraduate Data Science PLO documents. Demand-weighted concept coverage in the dated curricula ranged from 52.5% to 73.8%. Operational SKKNI workflow stages are another matter, because model evaluation (18.1%), model deployment (13.3%), feature construction (12.9%), data cleaning (7.4%), and model monitoring (4.8%) are not stated explicitly at the macro PLO level. Alignment estimates vary substantially across occupational role tracks and recruitment platform formats. Aggregate percentages should not be used as institutional rankings. Nor are they direct measures of graduate skill deficits. The findings instead give an empirically prioritized agenda for hierarchical CPL–CPMK–RPS constructive alignment audits, supported by multi-rater expert validation and authentic assessment verification in undergraduate Data Science education.

Declaration

Artificial Intelligence Use Statement

The authors declare no generative artificial intelligence were used in preparation of this manuscript.

Author Contribution Statement

Conceptualization, I.S.; methodology, I.S.; software, I.S.; validation, I.S. and A.S.; formal analysis, I.S.; investigation, I.S.; data curation, I.S.; writing – original draft, I.S.; writing – review and editing, I.S. and A.S.; visualization, I.S.; supervision, A.S.; project administration, I.S. All authors read and approved the final manuscript.

Funding Statement

None.

Conflict of Interest

None declared.

Ethics Approval and Consent to Participate

Not applicable. This computational content analysis used only publicly accessible institutional curriculum documents published on official university web portals and an open-access,

anonymized job advertisement dataset licensed under Creative Commons Attribution 4.0 International (CC BY 4.0) on Zenodo (record 18162371). No human participants, student academic records, patient data, or personally identifiable private records were collected or analyzed.

Consent for Publication

Not applicable. No identifiable individual participant images, clinical records, or private personal data are included in this manuscript.

Data Availability Statement

The primary labor market dataset analyzed in this study is publicly available on Zenodo at doi:10.5281/zenodo.18162371 under a CC BY 4.0 license. The official SKKNI Data Science standard (Kepmenaker No. 299 Tahun 2020) and the European Commission ESCO v1.2.0 Skills-Occupations Matrix Tables are publicly available from their respective portals. The derived 25-domain binary indicator matrices, the bilingual regular-expression dictionaries, and the Python 3.12 analysis scripts can be obtained from the corresponding author on reasonable request.

Acknowledgments

The author acknowledges the open-science data curation of Muhammad Ikhsan Zainuddin for depositing the Indonesian Data Science job posting corpus on Zenodo, and the curriculum development teams of the six Indonesian universities whose publicly accessible undergraduate Data Science curricula enabled this study.

References

1. Cunningham W, Moroz H, Muller N, et al. The demand for digital and complementary skills in Southeast Asia. Policy Research Working Paper No. 10070. Washington, DC: World Bank; 2022. doi:10.1596/1813-9450-10070
2. Saltz J, Armour F, Sharda R. Data science roles and the types of data science programs. Commun Assoc Inf Syst. 2018;43:615-624. doi:10.17705/1CAIS.04333
3. Biggs J. Enhancing teaching through constructive alignment. High Educ. 1996;32(3):347-364. doi:10.1007/BF00138871
4. Institut Teknologi Sepuluh Nopember. Undergraduate curriculum 2023–2028 of the Bachelor of Data Science Study Program [Internet]. Surabaya: Department of Statistics, ITS; 2023 [cited 2026 Oct 26]. Available from: https://www.its.ac.id/statistika/akademik/program-studi/s1-sains-data/kurikulum-2023-2028/
5. Universitas Negeri Surabaya. Transformative curriculum 2024–2028 of the Bachelor of Data Science Study Program [Internet]. Surabaya: UNESA; 2024 [cited 2026 Oct 26]. Available from: https://sindig.unesa.ac.id/kurikulum/s1-sains-data?id-kurikulum=11d182c4-9108-09b5-feea-914133687dbc
6. Universitas Insan Cita Indonesia. Outcome-based education (OBE) curriculum guideline for the Data Science Study Program [Internet]. Jakarta: UICI; 2023 [cited 2026 Oct 26]. Available from: https://uici.ac.id/wp-content/uploads/2023/11/PEDOMAN-KURIKULUM-BERBASIS-OBE-SAINS-DATA-UICI.pdf
7. Universitas Pembangunan Nasional Veteran Jakarta. Curriculum book of the Undergraduate Data Science Study Program 2024 [Internet]. Jakarta: Faculty of Computer Science, UPNVJ; 2024 [cited 2026 Oct 26]. Available from: https://new-fik.upnvj.ac.id/wp-content/uploads/2025/11/Buku-Kurikulum-Program-Studi-Sains-Data-Program-Sarjana-2024.pdf
8. Universitas Islam Negeri Sumatera Utara. Curriculum of the Undergraduate Data Science Study Program [Internet]. Medan: UINSU; 2025 [cited 2026 Oct 26]. Available from: https://sainsdata.uinsu.ac.id/kurikulum/
9. Universitas Pembangunan Nasional Veteran Jawa Timur. Learning outcomes of the Undergraduate Data Science Study Program [Internet]. Surabaya: UPNVJT; 2025 [cited 2026 Oct 26]. Available from: https://sada.upnjatim.ac.id/learning-outcome/
10. Tzimas G, Zotos N, Mourelatos E, et al. From data to insight: transforming online job postings into labor-market intelligence. Information. 2024;15(8):496. doi:10.3390/info15080496
11. Yang S, Abas A. Data science talent demand in China: a large-scale job posting analysis and implications for curriculum alignment. Int J Latest Technol Eng Manag Appl Sci. 2026;15(3):630-659. doi:10.51583/ijltemas.2026.150300052
12. Gurcan F, Soylu A, Khan AQ. Towards a sustainable workforce in big data analytics: skill requirements analysis from online job postings using neural topic modeling. Sustainability. 2025;17(20):9293. doi:10.3390/su17209293
13. Işığıçok E, Çelik S, Özdemir Yılmaz D. Analysis of skills and qualifications required in data scientist job postings based on the Pareto analysis perspective using text mining. Ekoist J Econom Stat. 2023;(39):10-25. doi:10.26650/ekoist.2023.39.1256697
14. Schwab-McCoy A, Baker CM, Gasper RE. Data science in 2020: computing, curricula, and challenges for the next 10 years. J Stat Data Sci Educ. 2021;29(sup1):S40-S50. doi:10.1080/10691898.2020.1851159
15. Hardin J, Horton NJ, Nolan D, et al. Computing in the statistics curricula: a 10-year retrospective. J Stat Data Sci Educ. 2021;29(sup1):S4-S6. doi:10.1080/10691898.2020.1862609
16. Adhikari A, Jordan MI. Interleaving computational and inferential thinking in an undergraduate data science curriculum. Harv Data Sci Rev. 2021;3(2):1-18. doi:10.1162/99608f92.cb0fa8d2
17. De Veaux RD, Agarwal M, Averett M, et al. Curriculum guidelines for undergraduate programs in data science. Annu Rev Stat Appl. 2017;4(1):15-30. doi:10.1146/annurev-statistics-060116-053930
18. Deb D, Fuad M. Integrating big data and cloud computing topics into the computing curricula: a modular approach. J Parallel Distrib Comput. 2021;157:303-315. doi:10.1016/j.jpdc.2021.07.012
19. Jaiswal K, Kuzminykh I, Modgil S. Understanding the skills gap between higher education and industry in the UK in artificial intelligence sector. Ind High Educ. 2024;39(2):234-246. doi:10.1177/09504222241280441
20. Aglusulos LJ. Beyond competency mapping: reframing curriculum–industry alignment through an outcome-based systems perspective. Ind High Educ. 2026;40(1):09504222261498315. doi:10.1177/09504222261498315
21. Derouich M. Ensuring outcome-based curriculum coherence through systematic CLO–PLO alignment and feedback loops. Discov Educ. 2025;4(1):486. doi:10.1007/s44217-025-00915-7
22. Snae Namahoot C, Snae Namahoot K, Nuntawong C, et al. ONTO-ALIGN: an ontology-based framework for evaluating curriculum alignment with educational standards in computer science. Interdiscip J Inf Knowl Manag. 2025;20:022. doi:10.28945/5600
23. Zhuang G, Lu X. A KeyBERT-enhanced pipeline for electronic information curriculum knowledge graphs: design, evaluation, and ontology alignment. Information. 2025;16(7):580. doi:10.3390/info16070580
24. Poolkrajang A, Papanai R. Reflective thinking on enhancing student competencies in learning management through outcome-based learning according to constructive alignment. Int J Eng Pedagog. 2024;14(6):4-22. doi:10.3991/ijep.v14i6.48409
25. Squalli Houssaini M, Aboutajeddine A, Toughrai I, et al. Development of a design course for medical curriculum: using design thinking as an instructional design method empowered by constructive alignment and generative AI. Think Skills Creat. 2024;52:101491. doi:10.1016/j.tsc.2024.101491
26. Lee N, Chen C. Constructive alignment in studio-based learning: enhancing design education through the zone of proximal development. Int J Technol Des Educ. 2026;36(2):753-775. doi:10.1007/s10798-025-10021-2
27. Mizgajski J, Szymczak A, Morzy M, et al. Return on investment in machine learning: crossing the chasm between academia and business. Found Comput Decis Sci. 2020;45(4):281-304. doi:10.2478/fcds-2020-0015
28. Lanubile F, Martínez-Fernández S, Quaranta L. Training future machine learning engineers: a project-based course on MLOps. IEEE Softw. 2024;41(2):60-67. doi:10.1109/MS.2023.3310768
29. Ministry of Manpower of the Republic of Indonesia. Decree of the Minister of Manpower of the Republic of Indonesia Number 299 of 2020 concerning the establishment of Indonesian National Work Competency Standards (SKKNI) for the information and communication category, artificial intelligence sub-field of data science [Internet]. Jakarta: Kemnaker RI; 2020 [cited 2026 Oct 26]. Available from: https://jdih.kemnaker.go.id/peraturan/detail/1979/keputusan-menaker-nomor-299-tahun-2020
30. European Commission. Skills-occupations matrix tables ESCO v1.2.0: level-three skill groups and occupation relations [Internet].

Brussels: European Commission; 2024 [cited 2026 Oct 26]. Available from: https://esco.ec.europa.eu/en/about-esco/publications/publication/skills-occupations-matrix-tables

31. Ojan MP, Lara-Navarra P, Sanz-Martos S. Bridging communication studies and employability: ESCO-based curriculum mapping and job-vacancy skill signals. Educ Sci. 2026;16(4):606. doi:10.3390/educsci16040606
32. Dakduk S, García-Chitiva MP, Correa JC. A case study on professional skills and accredited graduate programs of marketing. Appl Netw Sci. 2025;10(1):17. doi:10.1007/s41109-025-00696-w
33. Zuhdi M. Social construction of the outcome-based education curriculum in Indonesia's higher education: a public value perspective. J Public Value. 2024;8(1):37-46.
doi:10.53581/jopv.2024.8.1.37
34. Tjalla M, Sunubi AH. Challenges in the implementation of curriculum based on the Indonesian National Qualification Framework. JELITA. 2024;5(2):297-313. doi:10.56185/jelita.v5i2.698
35. Subiyantari AR, Gazali A, Handoyo SS, et al. Relevance of building engineering education curriculum towards SKKNI building and construction management competencies. J PenSil. 2024;13(3):299-313. doi:10.21009/jpensil.v13i3.48949
36. Setiyawan A, Ahyar H, Zakiyah F. Outcome-based education curriculum implementation in Arabic language education: a comparative study from a constructive alignment perspective. Edulab. 2026;10(2):70-85. doi:10.14421/edulab.2025.102.06
37. Chen K. Discussion on the curriculum system of data science and big data technology specialty based on OBE concept. Educ Rev USA. 2023;7(4):505-510. doi:10.26855/er.2023.04.021
38. Zainuddin MI. Dataset job in the field of data science in Indonesia from Glints and JobStreet [dataset]. Geneva: Zenodo; 2026 [collected 2025 Mar 17–Apr 23; cited 2026 Oct 26]. doi:10.5281/zenodo.18162371
39. Kramer J, Lu T. A reproducible framework for benchmarking machine learning operations (MLOps) infrastructures: comparing bare-metal and orchestrated machine learning workflows. Cureus J Comput Sci. 2025;2:e08693. doi:10.7759/s44389-025-08693-x
40. Manchana R. MLOps without borders: fostering synergy across data science, engineering, and operations. J Artif Intell Mach Learn Data Sci. 2024;2(2):1109-1118. doi:10.51219/jaimld/ramakrishna-manchana/261
41. Bade S. Demystifying MLOps: bridging the gap between machine learning and operations. Int J Inf Technol Manag Inf Syst. 2025;16(2):1376-1385. doi:10.34218/ijitmis_16_02_086
42. Alshameri F, Adhikari A. A content analysis of business job advertisements using data mining techniques. J Inf Syst Educ. 2025;36(3):316-330. doi:10.62273/vvie6840
43. Peslak A, Ceccucci W, Jones K, et al. Examining artificial intelligence job descriptions: an analysis of current job postings. J Inf Syst Appl Res Anal. 2026;19(2):25-35. doi:10.62273/ccjx2208
44. Samsuri T, Widiana IW, Lasmawan IW, et al. Transforming higher education curriculum based on the Indonesian National Qualifications Framework (KKNI) to address human resource challenges in the Industrial Revolution 4.0 era. Path Sci. 2025;11(6):2009. doi:10.22178/pos.119-11
45. Nastasa A, Maer Matei MM, Rughiniș C, et al. Skills and epistemic cultures in artificial intelligence research: evidence from job postings. Front Sociol. 2025;10:1655903.
doi:10.3389/fsoc.2025.1655903
46. Sheriff N, Sevukan R. Discovering research data management trends from job advertisements using a text-mining approach. J Inf Sci. 2026;52(2):484-500. doi:10.1177/01655515231193845
47. Ternikov A. Soft and hard skills identification: insights from IT job advertisements in the CIS region. PeerJ Comput Sci. 2022;8:e946. doi:10.7717/peerj-cs.946
48. Walek B, Pektor O. Data mining of job requirements in online job advertisements using machine learning and SDCA logistic regression. Mathematics. 2021;9(19):2475. doi:10.3390/math9192475