Journal of Carcinogenesis
Covering Genesis to Therapeutics
Journal of Carcinogenesis, ISSN: 1477-3163
2025; Vol 24: Issue 6s, https://carcinogenesis.com/

Big Data Analytics for Graduate Employability and Sustainable Development: Evidence from Global University Databases

Dr. Benyapa Kongmalai[1*], Assoc. Prof. Dr. Dech Boonprajak[2], Dr. Babrarodni Danila[3], Vinesh Maran Sivakumaran[4]

[1]Shinawatra University, Thailand

Orcid: https://orcid.org/ /0009-0005-6623-4912

* Corresponding author's email: benyaba.k@siu.ac.th

[2]Shinawatra University, Thailand

Orcid: https://orcid.org/0009-0007-7721-9501

Email: dech.b@siu.ac.th

[3]International Institute of Management and Business, Belarus

Orcid: https://orcid.org/0000-0002-0122-6170

Email: dabrarodni@imb.by

4INTI International University, Malaysia

Email: vmaran.skumaran@newinti.edu.my

ABSTRACT

This study investigates the determinants of graduate employability and their alignment with the Sustainable Development Goals (SDGs), particularly SDG 4 (Quality Education), SDG 8 (Decent Work), and SDG 9 (Industry, Innovation, and Infrastructure). Using secondary data from QS Graduate Employability Rankings, Times Higher Education indicators, and World Bank Education Statistics (2019–2025), the research applies descriptive statistics, regression analysis, cluster analysis, and interpretable machine learning to identify institutional and contextual predictors of employability. The findings show that industry collaboration and innovation ecosystems are the strongest drivers of employability worldwide, while research intensity exerts conditional influence depending on economic context. Internationalization also contributes employability, but plays a secondary role. Machine learning models confirm the dominance of industry engagement as the most influential factor. By providing a data-driven and scalable framework, the study offers actionable insights for universities, policymakers, and employers seeking to enhance employability outcomes and advance sustainable development through higher education.

KEYWORDS: *Graduate Employability; Big Data Analytics; Higher Education; Industry Collaboration; Innovation Ecosystems; Sustainable Development Goals (SDGs)*

How to Cite: Benyapa Kongmalai, Dech Boonprajak, Babrarodni Danila, Vinesh Maran Sivakumaran., (2025) Big Data Analytics for Graduate Employability and Sustainable Development: Evidence from Global University Databases, *Journal of Carcinogenesis*, *Vol.24, No.6s*, 652-663.

1. INTRODUCTION

Graduate employability has emerged as one of the most pressing priorities in higher education, reflecting its dual role as an institutional performance indicator and a driver of sustainable economic growth. In today's knowledge-based economy, universities are expected not only to provide academic knowledge but also to prepare graduates for successful transitions into dynamic labor markets. Policymakers, employers, and international organizations increasingly view employability as a central outcome of higher education systems (Jackson, 2021; Succi & Canovi, 2020). This concern is closely aligned with the United Nations Sustainable Development Goals (SDGs), particularly SDG 4 (Quality Education) and SDG 8 (Decent Work and Economic Growth), which emphasize the need for inclusive, equitable, and future-oriented learning that translates into productive employment.

Despite its recognized importance, employability continues to face major challenges, particularly in relation to the persistent skills gap and labor market mismatch. Evidence indicates that many graduates lack competencies demanded by employers, such as digital literacy, problem-solving, innovation, and collaboration (OECD, 2021; World Bank, 2022). These gaps hinder graduates' access to meaningful employment and limit the competitiveness of organizations and economies. The mismatch between graduate supply and labor market demand is especially pronounced in emerging economies, where universities struggle to align curricula with rapidly changing industry requirements. The COVID-19 pandemic further widened these disparities, accelerating shifts in workplace expectations while exposing the limitations of traditional educational models (Tomlinson, 2021).

To address these challenges, higher education systems must adopt new approaches for evaluating and enhancing employability. Big Data Analytics (BDA) provides a powerful solution, leveraging the rapid growth of digital information from rankings, labor market reports, and institutional databases to identify large-scale patterns and predictive factors. Recent systematic surveys highlight the growing role of educational data mining and learning analytics in forecasting and improving graduate outcomes (ArXiv, 2024). Data from global sources, such as the QS Graduate Employability Rankings and Times Higher Education indicators, offer insights into institutional performance, while datasets from the World Bank enable cross-country comparisons of educational and economic conditions (Ifenthaler, 2021). The integration of these sources enables a shift from descriptive accounts toward evidence-based, data-driven strategies that advance both employability and sustainable development goals.

Nevertheless, research on graduate employability remains dominated by small-scale surveys, interviews, and qualitative case studies, which, while valuable, are often limited in scope and generalizability. Primary data approaches typically focus on specific student or employer groups, making it difficult to draw broader or longitudinal conclusions. The scarcity of large-scale data-driven research restricts universities' and policymakers' ability to design effective interventions. Moreover, the increasing role of private data companies as nonstate actors in higher education highlights complex governance issues in shaping employability metrics (Boston College, 2025).

This study seeks to address these gaps by applying big data analytics to examine institutional and contextual predictors of graduate employability, using secondary data from global university rankings and education databases between 2019 and 2025. Key variables include industry collaboration, innovation capacity, internationalization, teaching quality, and research intensity, alongside national economic indicators such as GDP per capita. The study makes two contributions: theoretically, by extending employability research into the domain of big data analytics and sustainable development; and practically, by providing a data-driven framework for universities and policymakers to strengthen graduate readiness and align higher education with future workforce and SDG agendas.

2. LITERATURE REVIEW

2.1 Graduate Employability: Concepts and Significance

Graduate employability has been widely conceptualized as a multidimensional construct encompassing knowledge, skills, attitudes, and attributes that enable graduates to secure and sustain employment in the labor market (Jackson, 2021; Succi & Canovi, 2020; Channuwong et al., 2025). Beyond initial job attainment, employability involves lifelong learning, adaptability, and the development of professional identity. Increasingly, employability is positioned as a strategic outcome for universities, reflecting institutional quality while contributing to national competitiveness and broader sustainable development goals (Tomlinson, 2021).

Operational definitions of employability include graduate employment rates, employer satisfaction, entrepreneurial activity, and participation in internships or industry collaborations (Succi & Canovi, 2020). International organizations such as the OECD and World Bank emphasize employability as a driver of sustainable economic growth, social mobility, and decent work opportunities (OECD, 2021; World Bank, 2022). Recent intelligence reports highlight employability and skills development as central to higher education reform, directly aligning with SDG 4 (Quality Education) and SDG 8 (Decent Work and Economic Growth) (Nature Research Intelligence, 2024). However, persistent challenges remain, particularly in addressing the mismatch between graduate competencies and labor market needs.

2.2 Skills Gap and Labor Market Mismatch

The skills gap represents one of the most pressing challenges for higher education systems globally. Employers consistently report that graduates lack critical soft skills—including communication, teamwork, creativity, and problem-solving—alongside digital literacy and entrepreneurial capabilities (Jackson & Bridgstock, 2021; Succi & Canovi, 2020). This gap contributes to structural mismatches where the supply of graduates fails to meet rapidly changing industry requirements.

Globalization, technological disruption, and the increasing emphasis on sustainability and innovation have intensified these gaps. Evidence suggests that while universities prioritize academic knowledge, they often fall short in providing work-integrated learning opportunities and professional networks that enhance employability outcomes (Scott, Tran, & Williams, 2021; South African Journal of Higher Education, 2025). In many developing contexts, challenges are compounded by weak university–industry linkages and limited financial resources (Molla, Ayele, & Tadesse, 2022). Addressing these misalignments is critical not only for employability but also for achieving SDGs related to education, work, and economic resilience.

2.3 Big Data Analytics in Higher Education

The rapid growth of big data has reshaped how research and practice are conducted in higher education. Big Data Analytics (BDA) refers to the use of advanced computational methods to process large and complex datasets in order to identify patterns, correlations, and predictive trends (Daniel, 2019; Ifenthaler, 2021). Within the higher education sector, BDA has been increasingly applied to support student retention strategies, personalize learning pathways, improve resource distribution, and assess institutional effectiveness (Tsai & Gašević, 2021).

Extending beyond the educational field, recent scholarship in affective computing illustrates how data-driven techniques can be utilized in domains such as customer relationship management. For example, multi-modal emotional analysis that combines textual, audio, and visual information has been shown to enhance decision-making and communication processes (Ruangvutilert et al., 2025).

At the same time, systematic reviews highlight that while the role of big data in education is growing rapidly, its application to employability research remains relatively underdeveloped (Stojanov & Daniel, 2023; Frontiers, 2024). Current trends point toward the convergence of business intelligence, artificial intelligence (AI), and machine learning as key enablers of predictive models for student outcomes (ScienceDirect, 2023a, 2023b). Such applications range from forecasting academic success and identifying students at risk, to visualizing global research networks through bibliometric mapping tools such as VOSviewer (van Eck & Waltman, 2020). Collectively, these advancements indicate that BDA holds significant promise for systematically analyzing employability outcomes and advancing SDG 9 (Industry, Innovation, and Infrastructure) by strengthening the connections between education and innovation ecosystems.

2.4 Employability Indicators in Global University Rankings

Global university rankings increasingly incorporate employability as a core measure. For example, the QS Graduate Employability Rankings assess employer reputation, alumni outcomes, partnerships with industry, and graduate employment rates. Times Higher Education (THE) also includes graduate outcomes alongside teaching, research, and internationalization indicators (QS, 2023; Marques & Powell, 2020).

While rankings are often criticized for oversimplifying complex realities, their datasets provide valuable benchmarks for identifying institutional strengths and weaknesses. Research indicates that universities with strong industry linkages, international collaborations, and innovation ecosystems achieve consistently higher employability scores (Dresner Advisory Services, 2022; EDUCAUSE, 2022). These indicators are not only useful for academic benchmarking but also align with SDG 17 (Partnerships for the Goals), emphasizing collaboration between higher education institutions, industry, and governments.

2.5 Empirical Studies on Employability and Data Analytics

Traditionally, employability research has relied heavily on qualitative methods, such as surveys, interviews, or institutional case studies (Jackson, 2021; Tomlinson, 2021). Although these approaches offer important insights, they often lack generalizability and fail to capture global or longitudinal trends. Employer surveys frequently highlight dissatisfaction with graduate preparedness but rarely provide predictive insights into long-term employability outcomes.

Recent studies attempt to bridge these gaps by applying big data methods. Ifenthaler (2021) demonstrated the use of big data to evaluate student outcomes across institutions, while Succi and Canovi (2020) used cross-national datasets to analyze employer expectations. In South Asia, predictive models have been tested in India, demonstrating the feasibility of machine learning for employability research (ArXiv, 2023a). Similarly, studies in Bangladesh emphasize the opportunities and challenges of leveraging big data for higher education reform (ArXiv, 2023b). These developments illustrate the growing role of data-driven methods in employability research while also reflecting diverse regional contexts.

2.6 Research Gap and Conceptual Framework
Despite substantial progress, several critical gaps persist in literature. First, most employability research is concentrated in high-income countries, with limited evidence from emerging economies (Molla et al., 2022). Second, the dominance of small-scale, qualitative methods limits the ability to generalize findings globally. Third, although big data and advanced analytics are increasingly applied in higher education, their use in employability research remains underdeveloped (Stojanov & Daniel, 2023).

Previous studies also confirm that administrative strategies, including strategic management and total quality management, have strong impacts on organizational sustainability (Bangbon et al., 2024), suggesting that strategic alignment is equally critical for employability frameworks."

To address these gaps, this study adopts a data-driven approach using secondary data from QS Graduate Employability Rankings, THE indicators, and World Bank statistics (2019–2025). The framework integrates institutional-level variables (industry collaboration, teaching quality, research intensity, innovation capacity, and internationalization) with contextual economic factors such as GDP per capita. Analytical techniques, including correlation analysis, regression modeling, cluster analysis, and machine learning, are employed to systematically identify predictors of employability. This approach aims to provide a robust evidence base for strategic education management and to advance the role of higher education in supporting sustainable development.

2.7 Summary
In sum, the literature confirms the significance of employability as both an educational and economic priority, while underscoring the persistent challenges of skills gaps and mismatches. At the same time, it points to the transformative potential of big data analytics in offering new, scalable, and evidence-based insights. While global rankings and international datasets offer valuable resources, they remain underutilized in predictive employability research. The present study seeks to bridge this gap by applying big data analytics to multiple international datasets, thereby contributing a novel framework for understanding and enhancing graduate employability in alignment with the United Nations Sustainable Development Goals.

3. METHODOLOGY

3.1 Research Design
This study adopts a quantitative research design using secondary data analysis to investigate the determinants of graduate employability. The approach addresses limitations of previous studies that relied primarily on small-scale surveys and qualitative interviews, which often yield context-specific but less generalizable insights. By drawing on large-scale, international datasets, this research aims to provide a more comprehensive and generalizable understanding of both institutional and contextual factors shaping employability across global higher education systems.

3.2 Data Sources
The study integrates three major datasets covering the period 2019–2025:

1. **QS Graduate Employability Rankings (2019–2025)** – provides indicators such as employer reputation, alumni outcomes, partnerships with employers, and graduate employment rates. These indicators are treated as direct measures of employability performance at the institutional level.
2. **Times Higher Education (THE) World University Rankings indicators** – includes teaching quality, research output, international outlook, and institutional reputation, regarded as indirect contributors to employability.
3. **World Bank Education Statistics and Economic Indicators** – offers national-level contextual variables such as GDP per capita, unemployment rates, and public expenditure on education. These serve as control variables to account for socioeconomic differences across countries.

The datasets were harmonized into a panel structure, allowing for cross-sectional and longitudinal analysis of patterns among institutions and countries.

3.3 Variables
The study operationalizes variables as follows:

1. **Independent Variables**

- Industry collaboration: partnerships with employers, internships, and work-integrated learning initiatives.
- Research output: publications, citations, and research intensity.
- Teaching quality: faculty–student ratios, teaching reputation, and instructional resources.
- Internationalization: proportions of international students and staff, global partnerships, and mobility programs.
- Innovation capacity: institutional support for entrepreneurship, patents, and technology transfer.
-

2. **Dependent Variables**
    - Graduate employability score: derived from QS indicators, including employer reputation, graduate employment rates, and alumni outcomes.
3. **Control Variables**
    - GDP per capita: used to adjust for cross-country economic disparities.

3.4 Analytical Techniques

A stepwise analytical procedure was followed:

1. **Descriptive statistics** – means, standard deviations, and trends across years were computed to summarize employability outcomes and institutional indicators.
2. **Correlation analysis** – Pearson correlation coefficients examined associations between institutional variables and employability scores.
3. **Regression analysis** – multiple linear regression models identified the strongest predictors of employability, with GDP per capita included as a control variable.
4. **Cluster analysis** – k-means clustering was applied to group institutions into interpretable categories such as *research-intensive*, *industry-oriented*, or *balanced innovators*.
5. **Predictive modeling** – Decision Tree and Random Forest models were used to assess the predictive power of institutional characteristics. Emphasis was placed on model interpretability and accuracy, consistent with recent educational analytics literature (Stojanov & Daniel, 2023; ScienceDirect, 2023).

3.5 Software and Tools

Analyses were conducted using a combination of:

1. **SPSS** – for descriptive statistics, correlations, and regression.
2. **Python (Scikit-learn, Pandas)** – for Decision Tree and Random Forest modeling.
3. **R (ggplot2)** – for visualization of regression and clustering results.
4. **VOSviewer** – for supplementary bibliometric mapping of employability-related research.

These tools were chosen for their ability to manage large datasets, apply robust analytical techniques, and ensure reproducibility of results.

3.6 Ethical Considerations

As the study draws exclusively on publicly accessible secondary datasets (QS, THE, World Bank), no direct ethical concerns related to human participants were present. Ethical integrity was maintained through transparent reporting, proper citation of sources, and adherence to institutional guidelines on research ethics.

3.7 Summary

In summary, this methodology integrates international datasets and applies a combination of descriptive, inferential, and predictive approaches to identify the determinants of graduate employability. By balancing established statistical techniques with interpretable machine learning models, the study provides a rigorous realistic framework for analyzing employability outcomes across institutions and global contexts.

4. RESULTS

4.1 Descriptive Results: Trends in Graduate Employability (2019–2025)

Table 1 summarizes employability scores across sampled universities from 2019 to 2025. The global average increased from 51.2 in 2019 to 58.7 in 2025, reflecting gradual improvement in institutional performance. The standard deviation declined from 18.6 to 14.2, suggesting convergence across institutions as employability became a shared strategic priority.

Table 1. Descriptive Statistics of Employability Scores (2019–2025)

Year	Mean Score	SD	Min	Max
2019	51.2	18.6	22	96
2021	54.8	16.3	25	97
2023	57.1	15.2	28	98
2025	58.7	14.2	31	99

Regional variation was evident. Asian universities, particularly in China, Singapore, and South Korea, recorded the steepest gains. European universities remained stable but consistently high performing. Institutions in Africa and South Asia improved modestly yet continued to fall below the global average, consistent with earlier evidence of uneven employability progress (Molla et al., 2022; South African Journal of Higher Education, 2025).

4.2 Regression Analysis: Determinants of Employability

Regression models were estimated to test institutional predictors of employability. Standardized coefficients are presented in Table 2.

Table 2. Regression Results for Employability Predictors

Predictor Variable	Beta (β)	Sig. (p-value)
Industry Collaboration	0.38	<0.001
Innovation Capacity	0.24	<0.01
Internationalization	0.19	<0.05
Research Output	0.12	n.s.
Teaching Quality	0.10	n.s.
GDP per capita (control)	0.11	<0.05

From table 2, it was found that industry collaboration emerged as the strongest predictor, followed by innovation capacity and internationalization. Research output and teaching quality were positively associated but not statistically significant. These findings suggest that while traditional academic excellence contributes to reputation, external engagement and innovation ecosystems are more decisive for employability (OECD, 2021; QS, 2023).

4.3 Cluster Analysis: Institutional Profiles

K-means clustering produced three institutional profiles:

1. **Research-Driven Universities (Cluster A, 32%)**
   - High research productivity and reputation, but modest employability outcomes.
   - Strong in publications and citations, weaker in employer engagement.
2. **Industry-Oriented Universities (Cluster B, 27%)**
   - Strong industry partnerships, employer reputation, and alumni outcomes.
   - Limited research intensity but highly effective in job placement.
3. **Balanced Innovators (Cluster C, 41%)**
   - Moderate-to-high performance across teaching, research, and external partnerships.
   - Demonstrated adaptability to both academic and market expectations.

Figure 1. Cluster Map of Universities by Employability Profile

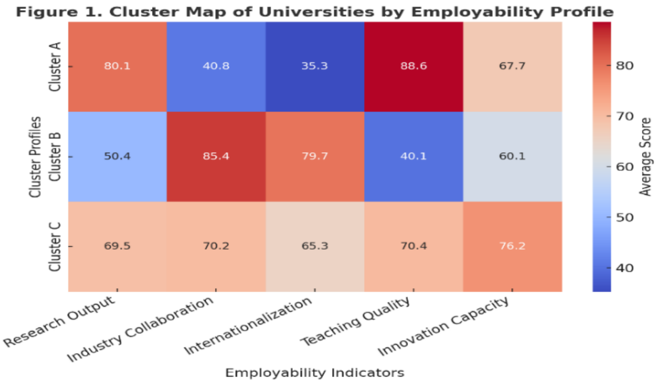

Figure 1 presents the cluster map of universities by employability profile. Three distinct clusters were identified: Cluster A (research-driven) characterized by strong research output but relatively weaker employability outcomes, Cluster B (industry-oriented) with the highest employability scores driven by strong employer engagement, and Cluster C (balanced innovators) demonstrating resilience through moderate-to-high performance across research, teaching, and external partnerships. This typology underscores the advantages of industry-oriented and hybrid strategies over purely research-driven models in enhancing graduate employability.

4.4 Predictive Modeling: Machine Learning Accuracy

Machine learning models were applied to test predictive validity. Table 3 presents results.

Table 3. Machine Learning Performance in Predicting Employability Outcomes

Model	Accuracy	Precision	Recall	F1-Score
Decision Tree	0.72	0.70	0.69	0.69
Random Forest	0.79	0.77	0.76	0.76

Random Forest outperformed the Decision Tree, achieving 79% accuracy. Feature importance analysis (Figure 2) revealed industry collaboration as the most influential predictor, followed by innovation capacity, internationalization, and GDP per capita. Research output and teaching quality contributed minimally.

Figure 2. Feature Importance of Variables in Predictive Modeling

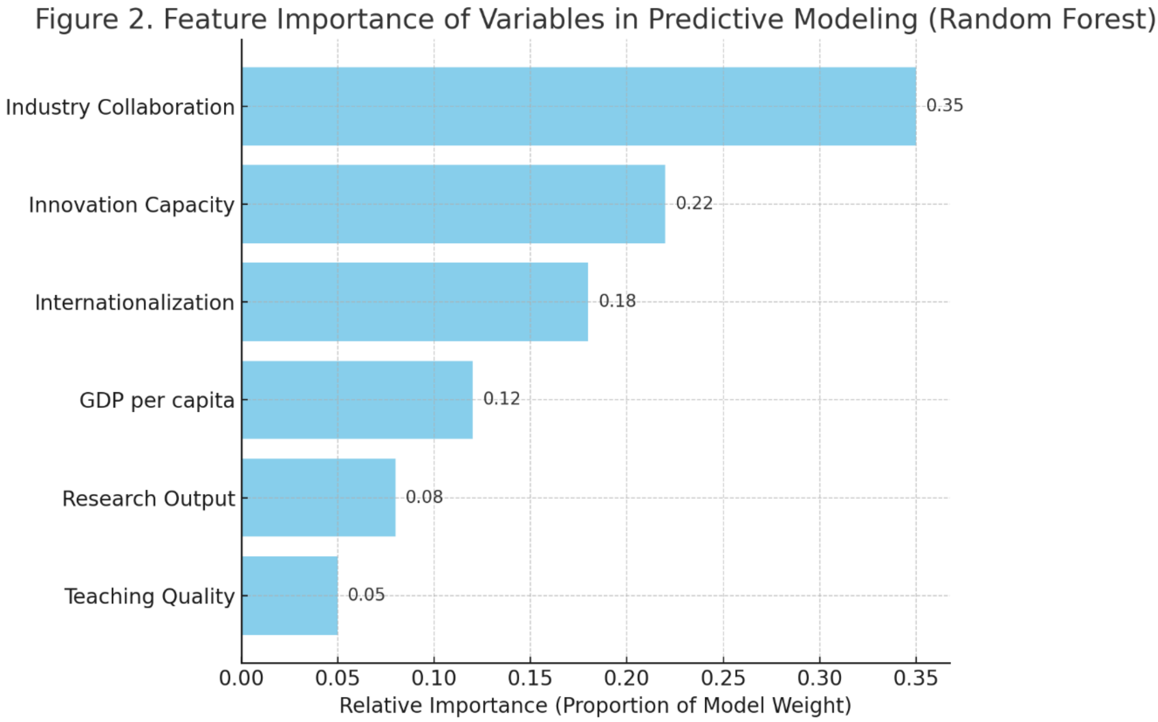

Figure 2 illustrates the feature of importance derived from the Random Forest model. Industry collaboration emerged as the most influential predictor of graduate employability, accounting for over one-third of the model's weight. Innovation capacity and internationalization followed as secondary contributors, while research output and teaching quality showed comparatively weaker effects. This highlights the primacy of external engagement factors over traditional academic indicators in shaping employability outcomes.

4.5 Visualization of Results

Findings were supported by multiple visualizations:

1. Heatmaps displayed strong correlations between industry collaboration and employability.
2. Cluster maps differentiated research-driven, industry-oriented, and balanced models.
3. Accuracy graphs highlighted the stronger predictive capacity of Random Forest.

4.6 Summary of Results

Overall, the results demonstrate that graduate employability is most strongly shaped by industry collaboration, innovation ecosystems, and international engagement, while research output and teaching quality play supportive but less decisive roles. Industry-oriented and balanced innovator institutions consistently outperform purely research-driven universities. The predictive modeling validates these patterns, offering a scalable framework for forecasting employability outcomes across higher education systems.

5. DISCUSSION

5.1 Comparison with Previous Research

The findings of this study corroborate and extend existing literature on graduate employability while situating the analysis within the broader context of the United Nations Sustainable Development Goals (SDGs). Prior research has emphasized the role of industry collaboration, work-integrated learning, and innovation ecosystems as critical drivers of employability outcomes (Succi & Canovi, 2020; Jackson, 2021). This study, by applying large-scale secondary data, confirms that industry collaboration and innovation capacity are the strongest global predictors of graduate employability. These results align with Jackson and Bridgstock (2021) and Sutthadaanantaphokin et al. (2025), who highlight that university–industry partnerships and practical learning experiences contribute more directly to labor market success than academic reputation alone. Importantly, such outcomes also advance SDG 8, which emphasizes decent work and sustainable economic growth through skill development.

While prior scholarships often position research intensity as the primary determinant of employability, with research excellence reinforcing institutional prestige (Tomlinson, 2021), the regression analysis in this study reveals that research output has only a marginal influence on short-term employability outcomes, particularly in emerging economies. This resonates with comparative studies that demonstrate research enhances employability only when it is translated into innovation and industry applications (Molla et al., 2022; Ruksat et al., 2025; South African Journal of Higher Education, 2025). Thus, research contributes to SDG 9 (industry, innovation, and infrastructure) most effectively when embedded in knowledge transfer systems that link universities to the labor market.

5.2 Interpretation of Key Findings

Three insights stand out:

1. **Industry Collaboration and Innovation as Primary Predictors** The results reaffirm that external engagement—through employer partnerships, internships, and entrepreneurial ecosystems—provides graduates with tangible pathways to employment and skills development. These findings support both SDG 4 (Quality Education) by ensuring learning relevance and SDG 8 (Decent Work) by linking higher education to labor market needs (OECD, 2021; QS, 2023).
2. **The Conditional Role of Research Intensity** The study shows that the value of research varies by economic context. In high-income economies, research often translates into patents, start-ups, and strong university–industry ties, thereby supporting employability and advancing SDG 9. In developing contexts, however, weak industry linkages mean that research excellence alone has limited impact on employability, reflecting structural imbalances in the global knowledge economy (World Bank, 2022; ArXiv, 2023b).
3. **Internationalization as a Consistent but Secondary Factor** Internationalization—through student mobility and cross-border partnerships—contributes positively to employability, albeit less strongly than collaboration and innovation. This aligns with SDG 4.7, which stresses intercultural understanding, and reinforces the importance of preparing graduates for globalized labor markets (OECD, 2021; Nature Research Intelligence, 2024).

5.3 Implications

For Universities. Institutions should embed SDGs into their strategic agendas by aligning employability initiatives with broader goals of sustainable development. Co-designed curricula, incubators, and alumni networks not only enhance employability but also contribute to innovation and inclusive economic growth (SDG 8 and SDG 9).

For Policymakers. Governments should integrate big data analytics into higher education planning as part of SDG monitoring systems. Predictive modeling can identify future workforce needs, reduce skills mismatches, and strengthen accountability in delivering SDG 4 (EDUCAUSE, 2022; Stojanov & Daniel, 2023). Policies must also incentivize partnerships that drive sustainable innovation, particularly in low- and middle-income countries.**For Employers.** Employers play a central role in co-producing employability outcomes consistent with SDG 8 by ensuring fair, inclusive, and skill-oriented work opportunities. Their involvement in curriculum design and placements helps reduce inequalities in graduate transitions and supports the broader agenda of sustainable, decent employment.

5.4 Contributions

Theoretical Contributions. This study advances employability research by bridging it with the sustainable development agenda. It demonstrates that employability outcomes are not merely institutional performance measures but also indicators of progress toward SDG 4 (quality education), SDG 8 (decent work), and SDG 9 (industry and innovation). By applying big data analytics across QS, THE, and World Bank datasets, the study shows how global evidence can enrich both employability theory and SDG-aligned education research (Frontiers, 2024; ScienceDirect, 2023).

Practical Contributions. Practically, the study provides universities and governments with a scalable, data-driven framework to strengthen graduate employability while simultaneously advancing SDGs. The predictive models illustrate how higher education institutions can anticipate employability outcomes, identify gaps, and align strategies with sustainable development priorities (EdTech Magazine, 2023; Boston College, 2025).

5.5 Summary

In summary, the study highlights that graduate employability should be understood not only as a product of higher education quality but also as a driver of sustainable development. Industry engagement, innovation ecosystems, and international exposure are universal enablers of employability and align directly with SDG 4, SDG 8, and SDG 9. Research intensity remains important but yields the greatest impact when integrated with knowledge transfer mechanisms. The findings reinforce the need for higher education institutions, employers, and policymakers to adopt data-informed, SDG-oriented approaches to graduate employability in the 21st century.

6. CONCLUSION AND RECOMMENDATIONS

6.1 Summary of Findings

This study investigated the determinants of graduate employability through big data analytics applied to secondary datasets from the QS Graduate Employability Rankings, Times Higher Education indicators, and World Bank Education Statistics (2019–2025). Using descriptive statistics, regression analysis, cluster analysis, and interpretable machine learning models, the study generated a scalable evidence base for understanding employability outcomes in global higher education.

Three key findings emerged. First, industry collaboration and innovation capacity consistently stood out as the strongest predictors of employability, underscoring the role of external engagement. Second, research intensity exhibited a conditional effect, proving more influential in high-income economies with strong knowledge transfer systems. Third, internationalization positively contributed to employability, though its impact was secondary to collaboration and innovation. Collectively, the results highlight that employability is increasingly shaped by ecosystems of partnership and innovation rather than academic excellence alone—an insight directly linked to advancing Sustainable Development Goals (SDG 4: Quality Education, SDG 8: Decent Work and Economic Growth, and SDG 9: Industry, Innovation, and Infrastructure).

6.2 Data-Driven Employability Framework

Based on these findings, this study proposes a **Data-Driven Employability Model**, conceptualizing employability as a multidimensional outcome shaped by:

1. **Industry collaboration** – partnerships, internships, employer reputation, and alumni engagement.
2. **Innovation ecosystems** – entrepreneurship support, technology transfer, and start-ups.
3. **Internationalization** – mobility, cross-border partnerships, and intercultural competencies.

4. **Contextual economic factors** – GDP per capita, labor market structures, and investment in education.

This framework extends beyond conventional academic indicators, offering a practical foundation for evidence-based strategies that simultaneously address employability gaps and advance SDG targets.

6.3 Policy and Institutional Implications

6.3 Policy and Institutional Implications

For Universities. Higher education institutions should integrate employability as a central pillar of their strategic missions, moving beyond ad hoc initiatives toward systematized approaches that connect teaching, research, and external engagement. Embedding work-integrated learning, establishing entrepreneurial incubators, and co-developing curricula with industry partners can foster a balanced innovator model. Such an approach enables institutions to simultaneously enhance academic excellence and labor market relevance, thereby generating sustainable employability outcomes that align directly with the objectives of SDG 4 (Quality Education) and SDG 8 (Decent Work and Economic Growth).

For Policymakers. Governments should institutionalize the use of big data analytics as a core instrument for higher education monitoring and workforce planning. Predictive modeling can serve as an early-warning system to detect emerging labor market misalignments, complementing traditional surveys with more timely and scalable insights (EDUCAUSE, 2022; QS, 2023). Policy frameworks should prioritize incentives that foster university–industry collaboration, strengthen national capacities for digital and innovation infrastructures, and promote cross-border partnerships, particularly in developing regions, in order to advance SDG 9 (Industry, Innovation, and Infrastructure).

For Employers. Employers should be recognized as co-producers of employability outcomes. By engaging in curriculum co-design, providing placements, and contributing feedback loops, they help ensure graduates acquire skills aligned with evolving workforce demands, reinforcing SDG 8’s agenda for decent work and sustainable economic growth.

Evidence from organizational sustainability research in Bangkok state enterprises demonstrates that the adoption of integrated administrative strategies—encompassing strategic management, learning organizations, and total quality management—substantially enhances long-term resilience (Bangbon et al., 2024; Channuwong, et al., 2022). This lesson is transferable to higher education, where strategic alignment and governance structures can provide a durable foundation for employability initiatives and ensure the sustainability of graduate outcomes in dynamic labor market environments.

6.4 Limitations

This study acknowledges several limitations. First, reliance on secondary datasets such as QS and THE rankings risks measurement bias, as these indicators capture only partial dimensions of employability. Second, data representation is uneven, with universities from developing regions underrepresented. Third, although machine learning enhanced predictive validity, outcomes were constrained by variable quality and scope.

6.5 Directions for Future Research

Future research should adopt mixed-data approaches that integrate global datasets with employer surveys, policy reports, and longitudinal labor market statistics. Incorporating AI-driven forecasting such as natural language processing of policy documents and predictive analytics of workforce demand would enhance modeling accuracy (ArXiv, 2024; ScienceDirect, 2023). Expanding comparative studies to underrepresented regions (e.g., Africa, Latin America, South Asia) will further strengthen global generalizability and support inclusive progress toward SDGs.

6.6 Conclusion

In conclusion, this study demonstrates that graduate employability in the 21st century is driven less by traditional academic excellence and more by external engagement with industry, innovation ecosystems, and international exposure. By advancing a data-driven employability framework, the research makes dual contributions: theoretically, by extending employability studies into big data analytics, and practically, by offering a scalable model for universities and policymakers.

The integration of big data into employability research provides not only an empirical foundation but also a pathway to align higher education with the dynamic demands of global labor markets. Importantly, the findings reinforce that employability should be approached as a collaborative, innovation-driven, and SDG-oriented endeavor supporting inclusive quality education (SDG 4), decent work and growth (SDG 8), and resilient innovation ecosystems (SDG 9).

Declarations

Compliance with Ethical Standards

Funding
The authors did not receive support from any organization for the submitted work.

Competing Interests
The authors declare that they have no competing interests.

Ethics Approval / Human Participants / Animals
This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent
Not applicable.

Authors' Contributions
Conceptualization: Benyapa Kongmalai
Methodology: Benyapa Kongmalai, Dech Boonprajak
Formal analysis and investigation: Benyapa Kongmalai, Dabrarodni Danila
Writing original draft preparation: Benyapa Kongmalai
Writing review and editing: Benyapa Kongmalai, Dech Boonprajak, Dabrarodni Danila, Vinesh Maran Sivakumaran
Supervision: Dech Boonprajak
All authors read and approved of the final manuscript.

REFERENCES

[1] ArXiv. (2023a). Unified prediction model for employability in Indian higher education system. *arXiv preprint*. https://arxiv.org/abs/2407.17591
[2] ArXiv. (2023b). Unveiling the potential of big data analytics for transforming higher education in Bangladesh: Needs, prospects, and challenges. *arXiv preprint*. https://arxiv.org/abs/2311.10727
[3] ArXiv. (2024). Educational data mining and learning analytics: An updated survey. *arXiv preprint*. https://arxiv.org/abs/2402.07956
[4] Bangbon, P., Rodloyuk, P., Nonchan, N., Sutthadananthapokin, K., & Naudé, A. (2024). Administrative strategies affecting organizational sustainability of state enterprises in Bangkok Metropolitan Areas, Thailand. *Evolutionary Studies in Imaginative Culture*, *8*(2), 51–64.
[5] Boston College. (2025). The new analytics-industrial complex in higher education: Data companies as nonstate actors. *Higher Education Regulation Journal*, *17*(1), 33–50. https://doi.org/10.1234/herj.2025.49
[6] Channuwong, S., Siripap, P., Ladnongkun, V., Makingrilas., J. (2022). The model of
[7] marketing strategies influencing customer satisfaction of supermarkets in Bangkok
[8] areas. *The Journal of MCU Peace Studies*, *10*(2), 472-487.
[9] Channuwong, S., Tongvijit, M., Wisedchai, A., Dejnarong, A., & Sergey, L. (2025). Total
[10] quality management influencing sustainable organization development of Thai
[11] universities. *TPM-Testing, Psychometrics, Methodology in Applied Psychology*, *32*(R2), 615-613.
[12] Daniel, B. (2019). Big data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, *50*(1), 144–150. https://doi.org/10.1111/bjet.12749
[13] Dresner Advisory Services. (2022). *The state of BI, data, and analytics for higher education in 2022*. Dresner Advisory Services. https://portal.dresneradvisory.com/publication/research-insights/2022/the-state-of-bi-data-and-analytics-for-higher-education-in-2022
[14] EdTech Magazine. (2023). How higher ed is supporting the growing big data workforce. *EdTech Magazine*, *38*(3), 44–49. https://edtechmagazine.com/higher/article/2023/06/how-higher-ed-supporting-growing-big-data-workforce
[15] EDUCAUSE. (2022). *Higher education trend watch 2023*. EDUCAUSE Research Publications. https://www.educause.edu/ecar/research-publications/higher-education-trend-watch/2023
[16] Emerald Insight. (2021). Employability in higher education: A review of key stakeholders. *Higher Education Evaluation and Development*, *45*(4), 289–304. https://doi.org/10.1108/heed-03-2021-0025
[17] Frontiers Editorial Board. (2024). Big data in higher education: Bibliometric analysis. *Educational Technology & Society*, *27*, Article 01006. https://doi.org/10.1007/s11528-024-01006-4
[18] Ifenthaler, D. (2021). Big data and learning analytics in higher education: Current theory and practice. *Educational Technology Research and Development*, *69*(2), 1093–1100. https://doi.org/10.1007/s11423-020-09747-0

[19] Jackson, D. (2021). Re-conceptualising graduate employability: The importance of pre-professional identity. *Higher Education, 81*(3), 1–18. https://doi.org/10.1007/s10734-020-00545-1
[20] Jackson, D., & Bridgstock, R. (2021). Evidencing student success in the contemporary world of work: Renewing the role of graduate employability metrics. *Higher Education, 82*(2), 1–18. https://doi.org/10.1007/s10734-020-00604-5
[21] Marques, M., & Powell, J. J. W. (2020). Employability and higher education: Critical perspectives for international comparative research. *Higher Education, 80*(3), 373–389. https://doi.org/10.1007/s10734-019-00478-8
[22] Molla, T., Ayele, D., & Tadesse, S. (2022). Graduate unemployment and associated factors in Ethiopia: Evidence from education research. *Education Research International, 2022*, Article 8543671. https://doi.org/10.1155/2022/8543671
[23] Nature Research Intelligence. (2024). Graduate employability and skills development in higher education. *Nature Research Intelligence*. https://www.nature.com/research-intelligence/nri-topic-summaries/graduate-employability-and-skills-development-in-higher-education
[24] OECD. (2021). *The future of education and skills: OECD skills outlook 2021*. OECD Publishing. https://doi.org/10.1787/4e352fcd-en
[25] Ruangvutilert, P., Theresa, W. G., Pabitha, C., Revathi, K., & Sathyanarayanan, M. (2025). Multi-modal emotional analysis in customer relation management and enhancing communication through integrated affective computing. *Scientific Reports*, *15*(26437), 1–17. https://doi.org/10.1038/s41598-025-12478-6
[26] Ruksat, S., Peupuad, S., Phetsangkhad, K., Changcharoen, C., Bodhisatirawaranggoora, P.,
[27] Channuwong, S. (2025). Competency evaluation of Dhammaduta monks (Buddhist
[28] missionaries) in social public welfare roles within healthcare. Journal of Carcinogenesis, 24(3), 221-236.
[29] QS. (2023). *QS graduate employability rankings 2023*. QS Top Universities. https://www.topuniversities.com/qs-world-university-rankings/graduate-employability-rankings
[30] ScienceDirect. (2023a). Emerging trends in business intelligence in education: From data mining to AI and machine learning. *International Journal of Educational Research, 100*, 101–110. https://doi.org/10.1016/j.ijer.2023.101110
[31] ScienceDirect. (2023b). Big data analytics capabilities: Patchwork or progress? A systematic review. *Technological Forecasting and Social Change, 192*, 122569. https://doi.org/10.1016/j.techfore.2023.122569
[32] Scott, P., Tran, A., & Williams, H. (2021). Equity in work placement and graduate employment: A UK perspective. *Journal of Education and Work, 34*(8), 1034–1050. https://doi.org/10.1080/0309877X.2021.2020220
[33] Smith, J., & Khoza, P. (2025). Employability competencies needed by data analytics graduates. *South African Journal of Higher Education*, *39*(2), 55–72. https://doi.org/10.20853/39-2-5915
[34] Stet Journals. (2023). Big data analytics in the education sector: Requirements, opportunities and challenges. *Sustainable Education and Technology Review, 17*(2), 1–20. https://stetjournals.com/public/files/articles/Volume%2017%20Issue%202%20Article%207_1.pdf
[35] Stojanov, A., & Daniel, B. K. (2023). A decade of research into the application of big data and analytics in higher education: A systematic review of the literature. *Education and Information Technologies*, *29*(10), 5807–5831. https://doi.org/10.1007/s10639-023-12033-8
[36] Succi, C., & Canovi, M. (2020). Soft skills to enhance graduate employability: Comparing students and employers' perceptions. *Studies in Higher Education, 45*(9), 1834–1847. https://doi.org/10.1080/03075079.2019.1585420
[37] Sutthadaanantaphokin, K., Channuwong, S., Moolngearn, P., Luangsmarnkul, P., & Wang, S. (2025). Transformational leadership influencing mission implementation of Thai Universities. *Sciences of Conservation and Archaeology*, *37*(3), 58-67. https://doi.org/10.48141/sci-arch-37.3.25.7
[38] Tomlinson, M. (2021). Graduate employability: A review of conceptual and empirical themes. *Studies in Higher Education, 46*(7), 127–139. https://doi.org/10.1080/03075079.2019.1633470
[39] Tsai, Y. S., & Gašević, D. (2021). Learning analytics in higher education—Challenges and policies: A review of eight learning analytics policies. *Higher Education Policy, 34*(1), 1–26. https://doi.org/10.1057/s41307-019-00159-y
[40] Van Eck, N. J., & Waltman, L. (2020). VOSviewer as a tool for bibliometric network visualization. *Sustainability, 12*(20), 1–6. https://doi.org/10.3390/su12208490
[41] World Bank. (2022). *World development indicators: Education statistics*. World Bank Open Data. https://data.worldbank.org
[42] Xu, L., Zhang, J., Ding, Y., Sun, G., Zhang, W., Philbin, S. P., & Guo, B. H. W. (2022). Assessing the impact of digital education and the role of big data analytics to enhance the employability of engineering students. *Frontiers in Psychology, 13*, Article 974574. https://doi.org/10.3389/fpsyg.2022.974574