SEMARAK ILMU PUBLISHING

Semarak International Journal of Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.my/index.php/sijaset/index
ISSN: 3030-5314

Leveraging Data Mining Techniques for Green Skill Development in TVET Curricula

Suhana Baharuddin[1], Zulhasni Abdul Rahim[2,*], Muhammad Saqib Iqbal[3]

[1] Malaysia-Japan International Institute of Technology (MJIIT), University Teknologi Malaysia (UTM), Kuala Lumpur, Malaysia
[2] NUST Business School, National University of Science & Technology (NUST), Islamabad, 44000, Pakistan

ARTICLE INFO

Article history:
Received 28 July 2025
Received in revised form 28 August 2025
Accepted 20 September 2025
Available online 3 October 2025

Keywords:
Data mining; green skills; curriculum development; sustainability policies; TVET; Natural Language Processing

ABSTRACT

The global transition toward sustainability requires a workforce equipped with specialised green skills, particularly within Malaysia's Technical and Vocational Education and Training (TVET) ecosystem, which includes Community Colleges, Polytechnics, Kolej Vokasional (KV), and institutions under the Malaysian Technical University Network (MTUN). However, existing curricula across these institutions often lag behind the evolving demands of green industries aligned with the Sustainable Development Goals (SDGs). This study employs a qualitative research approach, utilising Educational Data Mining (EDM) techniques, specifically Natural Language Processing (NLP), to systematically analyse a comprehensive body of sustainability-related textual data from policy documents, TVET frameworks, and industry reports. Using sentiment analysis, topic modelling, and keyword extraction, the study identifies and classifies essential green competencies demanded by industry. These include renewable energy systems, energy efficiency management, circular economy practices, sustainable resource management, and environmental compliance. The findings strongly align with SDG 4 (Quality Education), SDG 7 (Affordable and Clean Energy), SDG 9 (Industry, Innovation, and Infrastructure), and SDG 13 (Climate Action). The research underscores the importance of embedding data-driven curriculum design approaches across all tiers of TVET and recommends institutional collaboration, continuous capacity building, and proactive curriculum updates to ensure workforce readiness for sustainable economic transformation.

1. Introduction

Particularly in terms of training a workforce with competencies matched to environmental and economic transformation, the global move towards sustainable development has changed the expectations of businesses and the priorities of educational institutions. This change is significant in Malaysia within the Technical and Vocational Education and Training (TVET) sector, which strategically helps to upskill human capital for green development [11,18]. Among the several ministries in charge of TVET in Malaysia, this paper mainly addresses institutions under the Ministry

* *Corresponding author.*
E-mail address: zulhasni@utm.my

https://doi.org/10.37934/sijaset.6.1.3039

of Higher Education (MOHE)—that is, Polytechnics, Community Colleges, and the Malaysian Technical University Network (MTUN), comprising UTHM, UTeM, UMP, and UniMAP. These universities are fundamental for spreading sustainability competencies over national education systems and reflect the core of higher-level technical training [4,12].

Policy-level support for green economy transformation notwithstanding, there is still a significant curricular gap between industry needs and current course materials, especially about integrating focused green skills. Many institutional programs address sustainability, but they usually lack a methodical approach to embed cognitive (knowledge-based), practical (hands-on), and soft skill (attitudinal and behavioural) elements that are vital for generating graduates ready for green jobs. Current courses thus fail to equip students for dynamic roles in renewable energy, circular economy practices, sustainable resource management, and environmental compliance [13,18].

This paper systematically analyses educational frameworks, industry reports, and policy documents using EDM and NLP, bridging this gap. The study uses sentiment analysis, topic modelling, and keyword extraction to find and map green competencies pertinent to TVET courses under MOHE-governed direction. To guarantee fit with world priorities, the results are benchmarked against a set Sustainable Development Goals (SDGs)—especially SDG 4 (Quality Education), SDG 7 (Affordable and Clean Energy), SDG 9 (Industry, Innovation, and Infrastructure), and SDG 13 (Climate Action) [2,17, 24]

By providing strategic analysis for curriculum reform, stakeholder cooperation, and sustainable workforce development in Malaysia's higher TVET sector, this study aims to help design responsive and future-ready green curricula.

2. Review of the literature

2.1 Sustainability and Green Skills in TVET

Global development policies now revolve primarily around sustainability and environmental management, forcing businesses and academic institutions to change [15]. Given its direct influence on workforce readiness, TVET is critical in shaping the green economy. TVET systems have increasingly recognised the need to embed green skills—competencies that support sustainable environmental practices—within their curricula [3,10,15].

The ability of TVET to provide students with the practical skills necessary to address sustainability concerns is emphasised in the UNESCO roadmap for sustainable education [10]. Key competencies outlined include managing renewable energy systems, energy efficiency, sustainable resource management, and compliance with environmental standards [21]. For example, Polytechnic Primier Ungku Omar has initiated programs in renewable energy technology to support Malaysia’s green transition. At the same time, UTHM (Universiti Tun Hussein Onn Malaysia) offers modules on sustainable engineering practices under its MTUN mandate. Likewise, Community College Selandar integrates environmental conservation elements into its agriculture and landscape maintenance programs to enhance local sustainability relevance.

However, despite these initiatives, more focused and flexible solutions are still needed, as current industry expectations often exceed the capabilities developed in many existing programs [1]. This signals a clear need for continuous curriculum reform and more substantial alignment with real-time green skill demands from the industry.

2.2 Educational Data Mining and NLP in Curriculum Development

Educational Data Mining (EDM) refers to analytical approaches for processing educational data to inform curriculum enhancement, instructional strategies, and academic decision-making. Recent studies highlight how EDM tools such as predictive modelling, sentiment analysis, topic modelling, and text mining have supported data-driven curriculum design [13]. These methods help identify emerging competencies, gaps in current offerings, and evolving pedagogical needs [7].

This paper uses EDM with an emphasis on NLP techniques in the context of Malaysia's MOHE-governed TVET institutions, such as Polytechnics, Community Colleges, and MTUN universities. Analysing unstructured text data from policy documents and industry reports to derive competency-based insights calls especially for NLP. Given its character, NLP is especially useful in guiding the curriculum's Practical (skills-based) and Affective (soft skills and behavioural) domains.

For example, topic modelling may reveal the increasing demand for competencies like sustainability auditing or green entrepreneurship, which are not purely cognitive but involve applied practices and ethical awareness. Sentiment analysis further allows institutions to gauge how positively or negatively these competencies are framed across industry and policy texts, thus prioritising practical relevance and societal values.

By narrowing EDM and NLP applications toward practical competencies and affective soft skills, rather than solely theoretical content, TVET institutions can ensure that green curricula align more closely with workplace expectations and sustainability standards. A particular sub-domain of EDM, NLP, has shown especially success in textual data analysis from industrial reports, policy documents, and curricula. NLP techniques offer practical insights by methodically exposing theme patterns and critical competencies buried behind vast textual datasets [18].

NLP and text mining methods have been used in policy papers and sustainability documents in higher education, effectively identifying essential skills and topics for inclusion in courses [19]. For example, [18] used NLP to examine EU sustainability frameworks, exposing critical enabling competencies needed by the sector and guiding higher education curricula development.

2.3 Linking TVET with Sustainable Development Goals (SDGs)

Matching educational systems with the Sustainable Development Goals (SDGs) provides a methodical road towards solving modern-world problems. In particular, TVET courses can significantly affect various SDGs, namely SDG 4 (Quality Education), SDG 7 (Affordable and Clean Energy), SDG 9 (Industry, Innovation and Infrastructure), and SDG 13 (Climate Action) [8,15].

Good alignment requires identifying and including particular green skills in the curriculum. Studies highlight how TVET's industry trend responsiveness, especially regarding sustainability, directly improves graduates' employability and efficacy in green job sectors. Comprehensive, data-driven curriculum creation methods that constantly reflect changing industry standards and sustainability requirements (OECD, 2020) are needed in this alignment.

2.4 Existing Gaps in TVET Curricula

Although the need to include sustainability in TVET is becoming more acknowledged, current courses may remain fixed or poorly match the constantly changing needs of green sectors [20]. Ahmad [2] observed many recurring deficiencies, including poor use of data-driven approaches in curriculum updates, limited adaptation to technology improvements, and inadequate integration of

cross-disciplinary abilities. This curricular inertia risks generating a workforce unprepared for rapidly changing environmental problems.

Educational policy makers want to close these discrepancies by using cutting-edge analytical techniques, such as EDM and NLP, as strategic tools for spotting developing trends, gaps, and competencies in sustainability education. In addition, it is underlined that bridging curricular gaps requires cooperation between industry stakeholders, legislators, and educational institutions [18].

2.5 Summary and Research Justification

The examined literature emphasises the urgent need to use cutting-edge analytical techniques to ensure TVET courses remain within industry standards and meet environmental needs. Although EDM and NLP are valuable tools for curriculum research and improvement, their methodical implementation within the TVET framework and the integration of green skills remain very limited [10,16,17].

This study thus adds to the body of knowledge by using NLP-based data mining techniques to define the essential green capabilities modern companies need. It seeks to guide strategic curricular changes, thereby improving TVET's potential to produce a workforce ready for environmentally friendly economic transformations.

3. Methodology

3.1 Research Design

More significantly, this study uses advanced EDM techniques, NLP, and a qualitative, data-driven approach to identify and classify vital green skills needed for inclusion in TVET curricula (see Figure 1). This methodological approach is chosen because it can methodically examine vast amounts of textual material from reliable sources, including industry reports, educational frameworks, and sustainability policy papers.

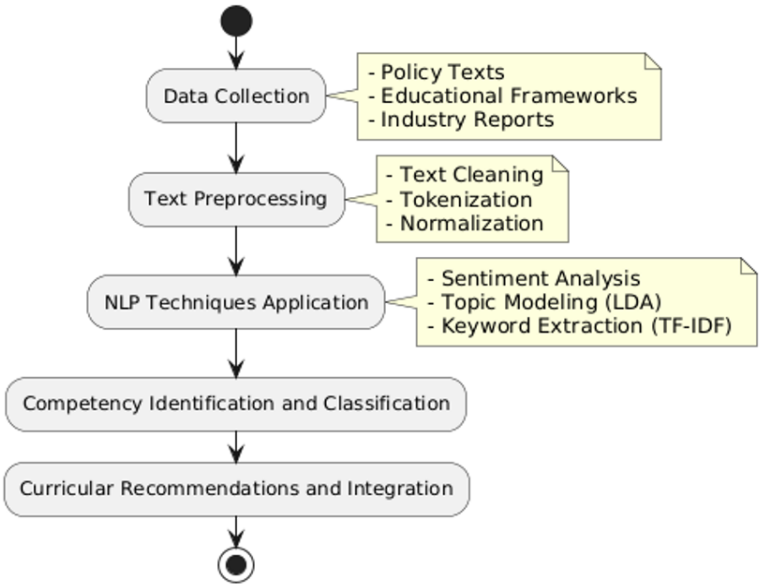

Fig. 1. Methodology flow chart

3.2 Data Collection

Three main types of sources provided the data for this project:

- Texts comprising national and global policy recommendations, including the UNESCO roadmap [14], and industry-specific sustainability laws.
- Existing TVET courses, competency requirements, and qualification systems are relevant for sustainability.
- Industry reports and whitepapers from important industrial players delineate the present and expected green technologies and sustainability competencies.

Published within the last five years, the documents chosen for the study guarantee relevance to the present sustainability criteria and competencies.

3.3 Data Analysis Techniques

Three main NLP-based techniques were used to study the textual data gathered thoroughly.

3.3.1 Examination of the sensation

Through sentiment analysis, the orientation and relevance of green competencies found in texts were determined, allowing the prioritisation of skills depending on their frequency of mention, contextual relevance, and sentiment polarity (positive, negative, neutral).

3.3.2 Subjects modelling

The literary corpus was searched for theme patterns using the latent Dirichlet allocation (LDA) topic modelling method, allowing the identification of prominent subjects related to green skill development. LDA enables topic groupings of text parts, precisely defining the necessary competencies.

3.3.3 Extracting keywords

Using NLP techniques, mainly the term frequency-inverse document frequency (TF-IDF) approach, keyword extraction sought important green skill terms and phrases stressed throughout the examined texts. This stage guaranteed a realistic representation of the capabilities sought by the industry.

3.4 Validation and Reliability

Triangulation and cross-referencing data from many categories of documents, policy texts, educational frameworks and industry reports were used to confirm the reliability and consistency of the findings. This procedure thoroughly covered important competencies and reduced possible biases or anomalies from single-source research.

3.5 Ethical Considerations

The ethical guidelines were strictly followed throughout the study. Since all the materials examined were freely available, questions about intellectual property rights or confidentiality were

allayed. Additionally, the correct reference techniques were rigorously followed to credit the sources of data and literature.

3.3 Data Analysis Techniques

Three main NLP-based techniques were used to study the textual data gathered thoroughly.

3.3.1 Examination of the sensation

Through sentiment analysis, the orientation and relevance of green competencies found in texts were determined, allowing the prioritisation of skills depending on their frequency of mention, contextual relevance, and sentiment polarity (positive, negative, neutral).

3.3.2 Subjects modelling

The literary corpus was searched for theme patterns using the latent Dirichlet allocation (LDA) topic modelling method, allowing the identification of prominent subjects related to green skill development. LDA enables topic groupings of text parts, precisely defining the necessary competencies.

3.3.3 Extracting keywords

Using NLP techniques, mainly the term frequency-inverse document frequency (TF-IDF) approach, keyword extraction sought important green skill terms and phrases stressed throughout the examined texts. This stage guaranteed a realistic representation of the capabilities sought by the industry.

3.4 Validation and Reliability

Triangulation and cross-referencing data from many categories of documents, policy texts, educational frameworks and industry reports were used to confirm the reliability and consistency of the findings. This procedure thoroughly covered important competencies and reduced possible biases or anomalies from single-source research.

3.5 Ethical Considerations

The ethical guidelines were strictly followed throughout the study. Since all the materials examined were freely available, questions about intellectual property rights or confidentiality were allayed. Additionally, the correct reference techniques were rigorously followed to credit the sources of data and literature.

4. Results

4.1 Identified Green Competencies in TVET Curricula

Using NLP tools, the analysis of industry reports, educational frameworks, and sustainability policy documents identified several key green capabilities essential for integration into TVET curricula. Policy and industry publications frequently highlighted renewable energy systems as crucial for sustainable infrastructure development aligned with SDG 7 (Affordable and Clean Energy). Similarly, energy efficiency management emerged prominently within sustainability policy documents, underscoring the industry's growing commitment to resource conservation and operational efficiency. Circular economy practices were widely acknowledged in educational frameworks and industry reports, indicating a clear shift toward sustainable consumerism, waste minimisation, and resource reuse. Furthermore, sustainable resource management was significantly featured in policy studies and educational materials, highlighting the importance of efficient resource allocation and utilisation. Lastly, environmental compliance and auditing appeared consistently in industry and policy reports, reflecting stringent regulatory standards and industry demands aligned closely with SDG 13 (Climate Action). Table 1 illustrates the identified green competencies and the corresponding SDGs.

Table 1
Identified green competencies and the corresponding SDGs

Green Competency	Relevant SDGs
Renewable energy systems	SDG 7, SDG 9, SDG 13
Energy Efficiency Management	SDG 7, SDG 13
Circular Economy Practices	SDG 9, SDG 13
Sustainable Resource Management	SDG 9, SDG 13
Environmental Compliance and Auditing	SDG 13

4.2 Results of the Sentiment Analysis

Sentiment analysis indicated a generally favourable attitude connected with including green capabilities in TVET courses throughout the examined texts, suggesting general acceptability and urgency in the educational, industrial, and policy-making domains. Most importantly, very positive attitudes were generated by renewable energy systems and circular economy ideas, suggesting their apparent significant influence and need.

4.3 Topic Modelling Insights

Different theme clusters produced by the LDA-based topic modelling amply show the regions of the focus of current and future industrial demands:

- Energy and Resource Management: Ideally, these subjects should immediately be prioritised for curricular development.
- Sustainability Auditing and Regulatory Compliance: Equipping students with the knowledge necessary to negotiate challenging environmental rules properly is essential.
- The industry emphasises the growing reliance on technical advances to solve environmental concerns, which are closely aligned with SDG 9 (Industry, Innovation and Infrastructure), innovation, and green technologies.

These thematic observations show that to remain relevant, current TVET programmes should prioritise curriculum modules covering these competencies.

4.4 Keyword Extraction Outcomes

Keyword extraction using TF-IDF underlined terms such as " renewable energy", "energy management", "circular economy", "resource efficiency", "environmental audit" and sustainability compliance" as most often occurring. The frequency of these keywords in many articles unequivocally shows the industry demand for these particular skills, confirming the need for their strong integration into TVET education (see Table 2 for a summary of data mining techniques).

Table 2
Summary of data mining techniques used

Technique	Purpose	Example application
Sentiment Analysis	Assess sentiment towards green skills integration.	Identification of positively-viewed skills
Topic Modelling (LDA)	Identify the main themes and competencies	Dominant sustainability themes
Keyword Extraction	Identify recurring key competencies	Key industry-demanded terms

5. Discussions

The study emphasises how urgently a proactive and responsive curriculum development plan based on data-driven methods is needed. Recognised green competencies in TVET courses promise significant benefits, such as improved employability, industry relevance, and meeting national and global sustainability targets.

Furthermore, the study supports constant and dynamic communication among educators, legislators, and business players to update and change courses to meet changing sustainability requirements. Using such cooperative initiatives, educational opportunities can remain flexible, relevant, and effective in equipping a qualified workforce for sustainable economic transitions.

Despite evident advantages, several challenges remain, including institutional inertia, resistance to regular curricular updates, and methodological changes. Technological limitations persist, with unequal access to advanced technologies essential for practical training in recognised green competencies. In addition, capacity development remains an issue, highlighting the need for institutional readiness and professional development for educators to deliver new curriculum modules effectively. Addressing these challenges requires strategic actions such as encouraging institutional support and ongoing professional training for teachers, establishing more substantial and more structured sector-academic partnerships to facilitate practical training and resource sharing, and developing supportive policy frameworks and incentives to accelerate the integration of identified green skills into TVET institutions.

6. Conclusions

Integrating sustainability-focused competencies, identified through advanced data mining methods such as NLP, into technical and vocational education and training (TVET) curricula is crucial to equipping a workforce aligned with SDGs 4,7,9, and 13. Key green competencies, including renewable energy systems, energy efficiency management, circular economy practices, sustainable resource management, and environmental compliance, were systematically classified, underscoring the strategic advantage of data-driven curriculum planning. To sustain this momentum, TVET

institutions should foster collaboration among educators, policymakers, and industry stakeholders, strengthen institutional capacity, use continuous data-driven updates, ensure equitable access to technology, and advocate for supportive policy frameworks, thus significantly enhancing the workforce's readiness for sustainable economic transitions.

References

[1] Acedo, Clementina. "Skills for inclusive and sustainable development: Perspectives from the Asia Pacific region and beyond." *Prospects* 44, no. 2 (2014): 137-140. https://doi.org/10.1007/s11125-014-9314-1

[2] Ahmad, Shabir, Sabina Umirzakova, Ghulam Mujtaba, Muhammad Sadiq Amin, and Taegkeun Whangbo. "Education 5.0: requirements, enabling technologies, and future directions." *arXiv preprint arXiv:2307.15846* (2023). https://doi.org/10.1007/s10639-023-11785-y

[3] Aryawati, N. P. A., Harahap, T. K., Yanti, N. N. S. A., Mahardika, I. M. N. O., Widiniarsih, D. M., Ahmad, M. I. S., Mattunruang, A. A. & Amali, L. M. (2023). Manajemen keuangan. Penerbit Tahta Media.

[4] Baharuddin, Suhana, Muhammad Saqib Iqbal, and Nooriza Ibrahim. "TVET Education for Community Education 5.0 in Malaysia." *Semarak International Journal of Innovation in Learning and Education* 4, no. 1 (2024): 27-41.

[5] Baharuddin, Suhana Binti, Zulhasni Abdul Rahim, and Muhammad Saqib Iqbal. "Impact of Data Mining Techniques and Self-Regulated Learning (SRL) in Predicting TVET Student Performance: A Review."

[6] Biggs, J. & Tang, C. (2021). Teaching for quality learning at university: What the student does (5th edn.). McGraw-Hill Education.

[7] Carlisle, Sheena, Karam Zaki, Mohamed Ahmed, Louise Dixey, and Emmet McLoughlin. "The imperative to address sustainability skills gaps in tourism in Wales." *Sustainability* 13, no. 3 (2021): 1161. https://doi.org/10.3390/su13031161

[8] Cebrián, Gisela, and Mercè Junyent. "Competencies in education for sustainable development: Exploring the student teachers' views." *Sustainability* 7, no. 3 (2015): 2768-2786. https://doi.org/10.3390/su7032768

[9] Kolb, D.A. (2022). Experiential Learning: Experience as the source of learning and development (2nd ed.). Pearson Education.

[10] LAMBINA, REGINA, RAMLEE MUSTAPHAB, and MP ROSLIN MD SHARIFC. "Identification of vocational well-being taxonomy for TVET sustainable growth: an initial review." *Journal of Contemporary Issues in Business and Government| Vol* 27, no. 2 (2021): 1434. https://doi.org/10.47750/cibg.2021.27.02.155

[11] A Lei, M., & Abidin, N. Z. (2024). Exploring the relationship between the Sustainable Development Goal (SDG) of the United Nations and Technical and Vocational Education and Training (TVET). International Journal of Advanced Research in Education and Society, 6(2), 591–601.

[12] Moore, Jonathan H., Luke Gibson, Zachary Amir, Wirong Chanthorn, Abdul Hamid Ahmad, Patrick A. Jansen, Calebe P. Mendes, Manabu Onuma, Carlos A. Peres, and Matthew Scott Luskin. "The rise of hyperabundant native generalists threatens both humans and nature." *Biological Reviews* 98, no. 5 (2023): 1829-1844.

[13] Murnawan, Murnawan, Sri Lestari, Rosalim Samihardjo, and Deshinta Arrova Dewi. "Sustainable educational data mining studies: Identifying key factors and techniques for predicting student academic performance." *Journal of Applied Data Sciences* 5, no. 3 (2024): 1325-1342. https://doi.org/10.47738/jads.v5i3.347

[14] Organisation for Economic Cooperation and Development (OECD). (2023). The role of education in climate resilience and the development of green skills. OECD Publishing. https://doi.org/10.1787/1c12a07b-en

[15] Paryono. "The importance of TVET and its contribution to sustainable development." In *AIP Conference Proceedings*, vol. 1887, no. 1, p. 020076. AIP Publishing LLC, 2017. https://doi.org/10.1063/1.5003559

[16] Peng, C., Liu, Z., Yang, W., & Lv, C. (2023). Green Vocational Education Implementation Path in Guangxi Under the Background of Rural Revitalisation: Taking Civil Engineering Major as an Example. Journal of Contemporary Educational Research, 7(12), 344-350. https://doi.org/10.26689/jcer.v7i12.5846

[17] Ramli, Syarina, Mohamad Sattar Rasul, and Haryanti Mohd Affandi. "Fourth Industrial Revolution (4IR)." *International Journal of Academic Research in Business and Social Sciences* 8, no. 9 (2018).https://doi.org/10.6007/ijarbss/v8-i9/4682

[18] Spada, Irene, Vito Giordano, Filippo Chiarello, Antonella Martini, and Gualtiero Fantoni. "Text mining on green policies for integrating sustainability in higher education." *The International Journal of Management Education* 23, no. 2 (2025): 101126.https://doi.org/10.1016/j.ijme.2024.101126

[19] Srivastava, Sidharth, Sandeep Kumar Dey, Savita Sharma, and Mark Ratilla. "Adoption of sustainable practices by Asian hotels: gaps in academia." *Worldwide Hospitality and Tourism Themes* 15, no. 3 (2023): 220-230. https://doi.org/10.1108/whatt-01-2023-0009

[20] Thevathas, V. A. P., & Zakaria, A. (2024). Relationship Between the Importance of Technical and Vocational Education and Training (TVET) Programmes and Their Contributions Towards Sustainability of Society's Development Among Kuala Lumpur Technical Secondary Students. International Journal of Academic Research in Business and Social Sciences, 14(7). https://doi.org/10.6007/ijarbss/v14-i7/21057

[21] United Nations Educational, Scientific and Cultural Organisation (UNESCO). (2023). Education for sustainable development: A roadmap. UNESCO Publishing. Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000374802

[22] World Bank. (2022). Climate education and workforce transformation: Policy strategies. World Bank Group. Retrieved from https://openknowledge.worldbank.org/handle/10986/37268

[23] World Economic Forum (WEF). (2023). Transitions to a green workforce and circular economy models. WEF Publishing. Retrieved from https://www.weforum.org/reports/green-workforce-transitions-and-circular-economy-models

[24] Safin, Rais S., and Evgeny A. Korchagin. "Vocational training of Russian students within educational cluster." *Rev. Eur. Stud.* 7 (2015): 168. https://doi.org/10.5539/res.v7n5p168