





Analysis of Business Intelligence Contributions in Educational Institutions: a Literature Review using Topic Modeling

Eduardo Luiz Alba ^{1*}, Gilson Adamczuk Oliveira ¹,
Matheus Henrique Dal Molin Ribeiro ¹,
Erick Oliveira Rodrigues ¹

¹Graduate Program of Production and Systems Engineering (PPGEPS)
Universidade Tecnológica Federal do Paraná, Pato Branco, Brazil.

*eduardoalba0@hotmail.com;

Abstract

Business Intelligence (BI) is a comprehensive concept encompassing tools, technologies, and methodologies designed for the discovery, analysis, and presentation of business information, aimed at facilitating decision-making processes. This concept has gained widespread popularity, particularly in light of the substantial surge in data generated by individuals in recent times. The advent of Big Data poses significant challenges for organizations, compelling them to adapt and enhance their decision-making processes to gain a competitive edge. Within an educational context, the scenario remains analogous, as decisions play a pivotal role in the development of institutions and student learning outcomes. In response to this, numerous authors have delved into the application of Business Intelligence in educational institutions, necessitating an exploratory literature review to identify key issues addressed in these works. This study conducts this exploration using Topic Modeling with Latent Dirichlet Allocation (LDA) to delineate the principal themes covered in scientific articles on this subject. The review was carried out using the Scopus database, encompassing a preliminary portfolio of 331 works, revealing four topics: I-Utilizing Business Intelligence to enhance the learning process and elevate students' performance in Higher Education, II-Employing Business Intelligence for the management and analysis of educational data, III-Integrating Business Intelligence and information technologies to foster innovation and collaboration in education and research, and IV-Exploring the intersection of business intelligence, data warehouse, and Big Data for the management of financial, academic, and sustainability aspects.

Keywords: Business Intelligence, Educational Institutions, Higher Education, Topic Modeling, Latent Dirichlet Allocation

1 Introduction

Educational institutions play a vital role in economic and social development, enhancing the quality of life for citizens and undertaking projects of considerable social significance [1]. Nevertheless, the current landscape is marked by intense competition for both business organizations and educational institutions, necessitating concerted efforts to maintain competitiveness and ensure their continued existence [2].

One approach to attain this objective is through effective information management, facilitating the visualization of the institution from diverse perspectives and enabling the comprehensive, rapid, and accurate management of organizational data [3].

Nevertheless, the substantial growth of data, commonly linked to as Big Data, has presented new challenges for organizations in extracting information and value from the data stored in their information systems. This is attributed to the difficulties in processing such extensive volumes, particularly considering that a significant portion of this data is unstructured [4], making it challenging to analyze using traditional methods, because their refers to information that lacks a predefined format or organization.

A search of the Scopus database yielded a comprehensive overview of the application of Business Intelligence (BI) in the educational field. This search identified 331 English-language works published between 2018 and 2023, highlighting the growing interest in this area and the increasing relevance of BI for the educational sector.

Due to the abundance of available work, extracting and comprehending this data can be a complex and time-consuming process, a initial effective approach should be to identify the topics of interest and subsequently examine the documents related to those topics, as recommended in the literature [5].

In this context, the present study puts forth an exploratory literature review, employing topic modeling with Latent Dirichlet Allocation (LDA) to delineate the primary topics identified in scientific papers concerning the contributions of Business Intelligence in educational institutions. The study addresses the research question: In what ways can Business Intelligence be incorporated into the routines of educational institutions?

This study contributes to the literature in two main aspects. First, by creating a comprehensive review of the literature on the contributions of Business Intelligence (BI) in educational institutions, providing an insightful and current overview of the primary trends and recent developments in the field. Second, by employing the topic modeling technique with Latent Dirichlet Allocation (LDA), the study addresses a gap in the existing literature, which lacked investigations utilizing this technique for reviewing the applications of BI in educational institutions.

2 Methodology

In conducting the literature review, the study adopted the framework proposed by [6]. This framework encompasses pre-processing, topic modeling, and post-processing activities, which will be detailed throughout this section. The proposed methodological framework for this study is illustrated in Figure 1.

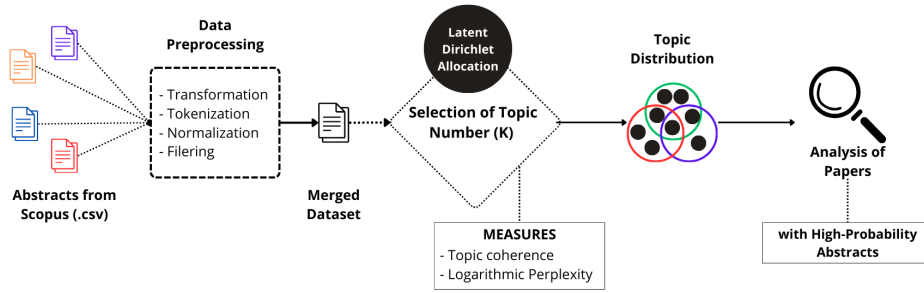


Fig. 1 Methodology framework proposed for the Literature Review with Topic Modeling.

2.1 Data Collection

The study started with the collection of papers through the Scopus database. Four keywords were integrated with the term ‘Business Intelligence’ in the search string to explore titles, abstracts, and keywords of the papers. The employed keywords were: ‘educat*’, ‘universit*’, ‘college’, and ‘high* education*’. The asterisk (*) at the end of keywords was used to expand the search, encompassing terminological variations, and enabling the capture of a broader spectrum of articles pertinent to the study.

Furthermore, the study applied the following selection criteria: a) papers written in English, and b) papers published between 2018 and 2023. A total of 331 articles, book chapters, and conference papers were retrieved.

In light of the significance attributed to linguistic coherence within the domain of topic modeling, as elucidated by [7], the selection process for inclusion within the literature review was informed by a preference for works composed in English. This choice was predicated upon their prevalence among the identified corpus, thereby obviating the need for inclusion of works in other languages to circumvent redundancy.

All retrieved articles from Scopus were exported in a comma-separated values (.csv) format to facilitate data manipulation and processing. This format allowed for seamless integration with the Orange Data Mining software, a user-friendly platform widely used for data analysis. In Orange, subsequent phases encompassing pre-processing and, ultimately, topic modeling, were undertaken.

2.2 Data Preprocessing

Data pre-processing is the initial step involved in Topic Modeling. In this study, this was carried out using the Orange Data Mining software, which utilizes its text mining tools, according to the steps below:

1. Conversion to lowercase and removal of accents (Transformation);
2. Division of data into smaller components (Tokenization);
3. Lemmatization of words (Normalization);
4. Removal of irrelevant words, numbers, and punctuation marks (Filtering);

This preprocessing stage is crucial to ensure the quality and accuracy of the data before proceeding to topic modeling. Since textual data lacks a well-defined structure, various inconsistencies may be encountered.

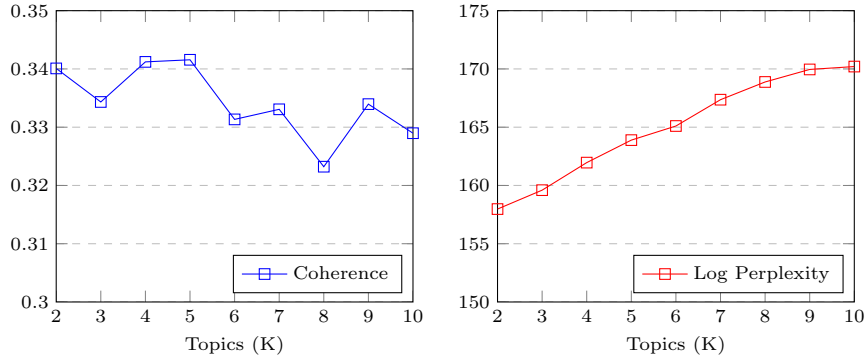


Fig. 3 Comparison of LDA algorithm measures.

As shown in the figure, the values of K that exhibited the highest coherence measure values were, in decreasing order, K=5, K=4, and K=2. Upon analyzing the content and the most frequent words of the topics generated by the K=5 experiment revealed redundancy across topics, hindering clear differentiation of themes addressed in the review. Furthermore, some topics exhibited limited relevance to the research question.

Consequently, a new analysis was conducted for the topics generated by the experiment with K=4, leading to the decision to adopt four topics for the literature review. This choice allowed for the exploration of diverse perspectives on the subject, facilitating the comparison and contrast of different ideas within each topic.

3.1 Business Intelligence to improve learning process and student performance in Higher Education

Table 1 presents the five articles with the highest probability in topic ‘a’. In this topic, were identified papers related to projects on technologies for activity management and learning and projects centered on the visualization of financial and academic data.

Table 1 Top five articles with the highest probability in the topic “a”.

Ref.	Title	Probability
[9]	Development of Students’ Digital Competence When Using the “Oracle” Electronic Portal	99,49%
[10]	Comprehensive learning system based on the analysis of data and the recommendation of activities in a distance education environment	99,44%
[11]	A business intelligence based solution to support academic affairs: case of Taibah University	99,38%
[12]	Service-Oriented Business Intelligence (SoBI) for Academic and Financial Data Integration in University	99,38%
[13]	PABED - A tool for big education data analysis	99,27%

In [9], a pedagogical technology based on learning management and creative activities of economics students was designed. This study revealed that the learning effectiveness of future economists was higher when utilizing the educational management model based on Business Intelligence.

In [10], a comprehensive learning system was proposed, converging Business Intelligence, expert systems, and learning management systems with other learning techniques to recommend various activities tailored to the needs of students.

In [11], a Business Intelligence solution to support academic affairs at Taibah University was described. Their approach involved data collection activities, multi-dimensional description, and visualization of dashboards and reports. Experiments conducted in SQL Server Data Tools indicated that the solution provides many necessary indicators for academic tasks, through data acquisition from diverse sources, development of multidimensional academic workflows, and the visualization of results through interactive dashboards and reports.

In [12], an implementation of Service-oriented Business Intelligence (SoBI) was carried out to integrate academic and financial data from Satya Wacana Christian University into a Data Warehouse. The results showed that, in this type of implementation, the use of dashboards to manage the data integration process proved to be of utmost importance.

In [13], a presentation of a Big Data analysis tool for the education sector was conducted, utilizing cloud technologies to compare undergraduate enrollment data for different academic years. The authors emphasized that the tool validates the use of cloud computing and Big Data technologies in education.

All five analyzed articles exhibited high adherence to the topic keywords, with marginal probabilities exceeding 99%. A common convergence point among these studies is their aim to demonstrate the transformative potential of Business Intelligence (BI) in Higher Education. Despite employing diverse approaches and methodologies, all studies converge on the objective of enhancing the learning process and student performance.

The articles explore various methods for collecting, integrating, and analyzing relevant data within the educational context. And recognizing the importance of clear and accessible communication, the authors emphasize the need to present data in a concise and visually appealing manner, utilizing charts, dashboards, and other tools.

3.2 Business Intelligence for the management and analysis of educational data

Table 2 lists the five articles with the highest probability in topic ‘b’. Within this topic, the papers made significant contributions to information management in university hospitals and solutions for the analysis of educational data across diverse sectors.

Table 2 Top five articles with the highest probability in the topic “b”.

Ref.	Title	Probability
[14]	An intelligent system for predicting a user access to a web-based e-learning system using web mining	99,27%
[15]	Improving Dental School Clinic Operations Using Business Intelligence Data	99,57%
[16]	CogStack - Experiences of deploying integrated information retrieval and extraction services in a large National Health Service Foundation Trust hospital	99,56%
[17]	Personality learning analytics system in intelligent virtual learning environment	99,49%
[18]	Village Business Intelligence (BI) Design to Support Social Welfare Intervention Programs by Using GIS Approach	99,41%

In [14], a system was proposed to predict user access to a web-based e-learning system using Web Mining techniques. In this work, the challenges of data analysis and its application in social networks were highlighted, with e-learning and e-commerce defined as important domains for these techniques.

In [15], a Business Intelligence-based information system was employed to assess the effects of computerization in a dental school clinic. Following the implementation of this system, it was found that appointment delays were reduced. An increase in the punctual start of student commitments was also observed. Thus, based on this analysis, this process can enhance the patient, student, and faculty experience.

In [16], a low-cost architecture for retrieval and extraction of structured and unstructured information was implemented at King's College Hospital, a university hospital. This architecture has already processed 300 million lines of clinical data to enhance services at the university hospital.

In [17], a concept for the architecture of a learning analysis system in an intelligent learning environment was proposed. According to the authors, the process should be employed as a means to collect data on students and learning, to enhance teaching and learning practices through data analysis.

In [18], a database and a Business Intelligence application were designed for a village in Indonesia, aiming to assist its employees and the government in decision-making processes. The government concerns over education, health, and nutrition were addressed in the study. Among its outcomes is the design of a dashboard produced with village data, including the database and applications.

Paralleling Topic a, the articles exhibited marginal probabilities exceeding 99%, indicating a strong adherence to the topic keywords. Despite focusing on diverse educational settings – ranging from e-learning systems and dental schools to intelligent learning environments – all converge on a central theme: educational data management through Business Intelligence.

This common objective manifests in the application of diverse methodologies, such as predicting user access in e-learning systems, enhancing a dental school clinic, architectures for retrieving structured and unstructured information from a university hospital, and collecting and analyzing data to optimize teaching practices. The adoption of BI solutions by higher education institutions and university hospitals can facilitate significant improvements in data management, student teaching and learning processes, and ultimately, patient quality of life.

3.3 Business Intelligence and information technologies for innovation and collaboration in education and research

Table 3 enumerates the five articles with the highest probability in topic 'c'. This topic encapsulates numerous interdisciplinary contributions of Business Intelligence within the realms of research and educational collaboration.

In [19], a system was proposed to address the lack of collaboration between universities and administrative issues. This system performs data collection, processing, and comparison, providing business analysis reports and dashboards. The results of the study indicate that the system can enable efficient collaboration in universities

Table 3 Top five articles with the highest probability in the topic “c”.

Ref.	Title	Probability
[19]	A way of developing collaboration between universities and businesses in a time of COVID-19	99,60%
[20]	On the relevance of self-service business intelligence to university management	99,46%
[21]	Critical examination using business intelligence on the gender gap in information technology in Brazil	99,22%
[22]	Predicting the post graduate admissions using classification techniques	92,04%
[23]	A Systematic Literature Review of Business Intelligence Technology, Contribution and Application for Higher Education	82,79%

and companies, facilitating analyses and the introduction of innovative trends through mutual cooperation.

In [20], the implementation of self-service Business Intelligence dashboards in a university was analyzed utilizing tools such as Power BI, QlikView, and Tableau. The aim was to evaluate the role of dashboards in performance management processes and their impact on the organization. The results indicate that dashboards brought about a shift in the organization procedures and outcomes, uncovering technical and organizational issues that were previously hidden.

In [21], Business Intelligence techniques were employed to analyze data from the computer and information technology market to characterize the role of women in these fields. The aim was to highlight that their involvement remains a challenge in Brazil. Based on this work, the results indicated that public policies should encourage young women to work in STEM fields.

The authors in [22] use data mining methods to support decision-making in educational organizations. The objective was to identify factors influencing student academic performance and the chances of admission to prestigious universities for postgraduate.

In [23], the authors use the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) systematically literature review method, to identify how Business Intelligence can improve the quality of education and research. The objective was to analyze the technologies, contributions, and applications of Business Intelligence in the educational sector. Applications in research, curriculum, assessment, behavior analysis, student enrollment, and resource management were addressed.

Although the constituent articles of this topic exhibited marginal probabilities slightly lower than others, surpassing 82%, in contrast to the 99% observed in the case of other topics, a robust thematic coherence centered on Business Intelligence (BI) in education and research persists. These articles highlight various applications of BI tools and techniques, promoting collaboration and innovation in these domains.

A common thread among these studies is the focus of BI on applications in academic research, promoting innovation and collaboration in education. Solutions are presented for collaboration and performance management between universities, analyses of gender representation in computer science fields, identification of factors influencing student academic success, and diverse applications in higher education, encompassing research, curriculum development, and resource management. These studies collectively showcase the transformative potential of BI in fostering innovation and collaboration in educational and research landscapes.

3.4 Business intelligence, data warehouse and big data for financial, academic and sustainability management

Table 4 outlines the five articles with the highest probability in topic ‘d’. This topic encompasses various contributions related to sustainability information management.

Table 4 Top five articles with the highest probability in the topic “d”.

Ref.	Title	Probability
[24]	A business intelligence framework for sustainability information management in higher education	99,63%
[25]	Building an Information Analysis System within a Corporate Information System for Combining and Structuring Organization Data (on the example of a University)	99,55%
[26]	The role of business intelligence in sustainability reporting for South African higher education institutions	99,45%
[27]	Building a Data Warehouse to Support Active Student Management: Analysis and Design	99,44%
[28]	Data warehouse system for multidimensional analysis of tuition fee level in higher education institutions in Indonesia	99,44%

In [24], a Business Intelligence (BI) framework was developed for the strategic management of sustainability information in higher education institutions (HEIs). The results indicated that the usability of the tool was positive, emphasizing the framework ability to overcome the constraints faced by HEIs in managing sustainability data.

In [25], a model was proposed to automate work with data within the corporate analytical system, using PowerBI technologies to create a metadatabase and solve data consolidation problems. The results present an efficient solution for organizational performance analysis, modernizing the generally accepted design scheme of the university’s corporate information system.

In [26], a study was conducted on how BI reports can support stakeholders in the disclosure of sustainability information. The study highlighted that the use of BI in South African HEIs is still at a low maturity level and that there is a need for investments in these technologies to assist in sustainability reporting processes.

In [27], an analysis and design of a data warehouse was conducted to integrate diverse operational databases, with the objective of fulfilling the reporting prerequisites for active students at XYZ University, as per the regulations stipulated by the Ministry of Research, Technology, and Higher Education of the Republic of Indonesia. The outcomes encompassed the formulation of the data warehouse design alongside the development of a dashboard facilitating the presentation of integrated information concerning active students in a visually intuitive and comprehensive manner.

In [28], a data warehouse (DW) system was developed for fee management in higher education institutions (HEIs) in Indonesia. The system integrates data from various sources and employs a Business Intelligence (BI) approach with four procedures: preparation, integration, analysis, and visualization. The results are presented in charts on a fee dashboard, providing insights into business performance.

Encouragingly, all articles analyzed in this section showed high probability scores with the target topic keywords, indicating a strong thematic connection within the chosen research area. Taking a closer look at the specific contributions, a common

thread emerges: all articles highlight the positive impact of integrating BI with data warehousing and big data.

Those works cover everything from structures for strategic management of sustainability information to data warehouse designs for active student management and applications in sustainability reports, corroborating with the transformative potential of BI in improving data-driven decision-making across multiple domains.

4 Conclusions

This study conducted an exploratory literature review using topic modeling with Latent Dirichlet Allocation. The primary objective of this review was to identify the main themes present in recent literature that discuss the contributions of Business Intelligence to the routines of educational institutions.

Over the employed methodology, four main topics were identified, addressing various contributions of Business Intelligence in educational institutions. These topics encompass applications aimed at enhancing student learning processes, managing and analyzing educational data, promoting innovation and collaboration in education and research, and supporting financial, academic, and sustainable management.

A total of 331 papers were collected from the Scopus database for topic modeling. Following the application of inclusion criteria, a content analysis was conducted on 20 of them, revealing that Business Intelligence is an emerging and significant research area in the educational context. It has demonstrated the potential to provide efficient solutions to the challenges faced by higher education institutions.

It is crucial to acknowledge that even when employing specific methods to define the topics, selecting only 20 articles does not guarantee that all the themes present in the entire corpus of 331 documents are represented. It is possible that less frequent or less prominent topics in the selected articles may be relevant to understanding the subject matter. Despite this limitation, as this is an initial study, the LDA provided us with a general overview of the research area and helped us to understand patterns and relationships between the topics.

For future work, the collected data will be utilized as a foundation for planning and constructing a Business Intelligence-based software solution to address the challenges faced by educational institutions.

References

- [1] Starostina, S.E., Kazachek, N.A., Tokareva, J.S.: The development of the education quality assurance system in the context of socio-economic growth of the cross-border region. *Mathematics Education* **11**(9), 3289–3300 (2016)
- [2] Santos, T.R., Ribeiro, M.W.S., Santos, W.B., Costa, L.R., Stédile, C.G.S.: A Systematic Review on the Implementation of Business Intelligence at Federal Universities, pp. 85–92 (2021)
- [3] Retnowardhani, A., Sardjono, W., Triana, Y.S.: Review Study of Business Intelligence to Support Strategic Decision Making, vol. 2019-July (2019)

- [4] Van Der Aalst, W.: Process Mining: Data Science in Action, 2nd edn. Springer, Heidelberg (2016)
- [5] Blei, D.M.: Probabilistic topic models. *Communications of the ACM* **55** (2012)
- [6] Asmussen, C., Møller, C.: Smart literature review: a practical topic modelling approach to exploratory literature review. *Journal of Big Data* **6**(1), 1–18 (2019) <https://doi.org/10.1186/s40537-019-0255-7>
- [7] Souza, M.d., Souza, R.R.: Modelagem de tópicos: Resumir e organizar corpus de dados por meio de algoritmos de aprendizagem de máquina. *Múltiplos olhares em ciência da informação* **9** (2019)
- [8] Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. *The Journal of Machine Learning Research* **3**, 993–1022 (2003)
- [9] Tolmachev, M., Korotaeva, I., Zharov, A., Beloglazova, L.: Development of students’ digital competence when using the “oracle” electronic portal. *European Journal of Contemporary Education* **11**(4), 1261–1270 (2022)
- [10] Villegas-Ch, W., Palacios-Pacheco, X., Buenaño-Fernández, D., Luján-Mora, S.: Comprehensive learning system based on the analysis of data and the recommendation of activities in a distance education environment. *International Journal of Engineering Education* **35**(5), 1316–1325 (2019)
- [11] Boulila, W., Al-kmal, M., Farid, M., Mugahed, H.: A business intelligence based solution to support academic affairs: case of taibah university. *Wireless Networks* **29**(3), 1051–1058 (2023)
- [12] Somya, R., Manongga, D., Ineke Pakereng, M.A.: Service-Oriented Business Intelligence (SoBI) for Academic and Financial Data Integration in University, pp. 1–5 (2018)
- [13] Khan, S., Shakil, K.A., Alam, M.: PABED - A Tool for Big Education Data Analysis, vol. 2019-February, pp. 794–799 (2019)
- [14] Sathiyamoorthi, V.: An intelligent system for predicting a user access to a web based e-learning system using web mining. *International Journal of Information Technology and Web Engineering* **15**(1), 75–94 (2020)
- [15] Nalliah, R.P., Miller, K., Stefanac, S.: Improving dental school clinic operations using business intelligence data. *Dental Education* **83**(6), 614–623 (2019)
- [16] Jackson, R., Kartoglu, I., Stringer, C., Gorrell, G., Roberts, A., Song, X., Wu, H., Agrawal, A., Lui, K., Groza, T., Lewsley, D., Northwood, D., Folarin, A., Stewart, R., Dobson, R.: Cogstack - experiences of deploying integrated information retrieval and extraction services in a large national health service foundation

- trust hospital. *BMC Medical Informatics and Decision Making* **18**(1) (2018)
- [17] Ruangvanich, S., Nilsook, P.: Personality Learning Analytics System in Intelligent Virtual Learning Environment, pp. 245–250 (2018)
- [18] Adi, S., Heripracoyo, S.: Village Business Intelligence (BI) Design to Support Social Welfare Intervention Programs by Using GIS Approach. (2018)
- [19] Seturidze, R., Topuria, N.: A way of developing collaboration between universities and businesses in a time of covid-19. *Kybernetes* (2021)
- [20] Arnaboldi, M., Robbiani, A., Carlucci, P.: On the relevance of self-service business intelligence to university management. *Journal of Accounting and Organizational Change* **17**(1), 5–22 (2021)
- [21] Gallindo, E.L., Cruz, H.A., Moreira, M.W.L.: Critical examination using business intelligence on the gender gap in information technology in brazil. *Mathematics* **9**(15) (2021)
- [22] Jeganathan, S., Parthasarathy, S., Lakshminarayanan, A.R., Ashok Kumar, P.M., Khan, M.K.A.: Predicting the Post Graduate Admissions Using Classification Techniques, pp. 346–350 (2021)
- [23] Santi, R.P., Putra, H.: A Systematic Literature Review of Business Intelligence Technology, Contribution and Application for Higher Education. (2018)
- [24] Scholtz, B., Calitz, A., Haupt, R.: A business intelligence framework for sustainability information management in higher education. *International Journal of Sustainability in Higher Education* **19**(2), 266–290 (2018)
- [25] Kopnova, O., Shaporeva, A., Iklassova, K., Kushumbayev, A., Tadzhitov, A., Aitymova, A.: Building an information analysis system within a corporate information system for combining and structuring organization data (on the example of a university). *Eastern-European Journal of Enterprise Technologies* **6**(2-120), 20–29 (2022)
- [26] Calitz, A., Bosire, S., Cullen, M.: The role of business intelligence in sustainability reporting for south african higher education institutions. *International Journal of Sustainability in Higher Education* **19**(7), 1185–1203 (2018)
- [27] Sutedja, I., Yudha, P., Khotimah, N., Vasthi, C.: Building a Data Warehouse to Support Active Student Management: Analysis and Design, pp. 460–465 (2018)
- [28] Yulianto, A.A., Kasahara, Y.: Data warehouse system for multidimensional analysis of tuition fee level in higher education institutions in indonesia. *International Journal of Advanced Computer Science and Applications* **11**(6), 541–550 (2020)