

Article

Latent Dirichlet Allocation in Public Procurement Documents Analysis for Determining Energy Efficiency Issues in Construction Works at Polish Universities

Anna Pamula ^{1,*} , Zbigniew Gontar ² , Beata Gontar ¹  and Tetiana Fesenko ³ 

¹ Department of Computer Science, Faculty of Management, University of Lodz, 90-136 Łódź, Poland; beata.gontar@uni.lodz.pl

² Institute of Information Systems and Digital Economy, Collegium of Economic Analysis, SGH Warsaw School of Economics, 02-554 Warszawa, Poland; zgonta@sgh.waw.pl

³ Department of Electronic Computers, Kharkiv National University of Radio Electronics, 61166 Kharkiv, Ukraine

* Correspondence: anna.pamula@uni.lodz.pl

Abstract: This paper presents a comprehensive analysis of public procurement documents in the domain of university buildings taken from the e-procurement platform, particularly focusing on their transformation towards more efficient energy consumption. Using a corpus of the titles of the public procurement documents from 2020 to 2022, we used Latent Dirichlet Allocation (LDA) for topic modeling to understand the key thematic areas of focus. The methodology presented in this study incorporated a bifurcated approach. This two-stage procedure began with preprocessing and dictionary creation from the corpus of titles of procurement documents. Following this, the Latent Dirichlet Allocation (LDA) model was employed for topic extraction and trend analysis, thereby providing a comprehensive understanding of the thematic progression in procurement practices over time. Our analysis revealed a shift in emphasis from modernization towards the adoption of energy-saving technologies as well as a growing focus on broader sustainability initiatives. However, a less prevalent topic was adherence to cooling & heating systems, suggesting potential areas for improvement in procurement practices. These findings contribute to the growing body of knowledge on sustainable procurement in university buildings and offer valuable insight for universities to enhance their energy efficiency strategies.

Keywords: public procurement; energy efficiency; topic modeling; Latent Dirichlet Allocation



Citation: Pamula, A.; Gontar, Z.; Gontar, B.; Fesenko, T. Latent Dirichlet Allocation in Public Procurement Documents Analysis for Determining Energy Efficiency Issues in Construction Works at Polish Universities. *Energies* **2023**, *16*, 4596. <https://doi.org/10.3390/en16124596>

Academic Editor: Vincenzo Costanzo

Received: 1 March 2023

Revised: 18 May 2023

Accepted: 1 June 2023

Published: 8 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Background and Research Gap

The global energy landscape has undergone substantial changes in recent years, driven largely by increasing awareness of climate change and a desire for more sustainable practices. The building sector, including university buildings, has been identified as a major consumer of energy, making it a key area for improvements in energy efficiency. Higher education institutions, particularly large-scale universities, typically manage a diverse portfolio of buildings, each with its unique architectural structure, age, and condition. This portfolio often spans a broad historical spectrum; it encompasses venerable structures from centuries past, edifices erected in the post-war era of the previous century with expected lifespans of just a few decades, and modern constructions, which are increasingly designed with cutting-edge, energy-efficient technologies. This heterogeneity presents both challenges and opportunities when considering the transformation towards more efficient energy consumption.

Predominantly, universities embark on sustainable initiatives to bolster their reputations and satisfy the demands of diverse stakeholder groups. This includes local communities, financial backers, and environmentally conscious youth who are particularly

concerned about climate change issues. To address these expectations, universities have established dedicated units tasked with crafting, executing, and conveying Corporate Social Responsibility (CSR) strategies [1,2]. At the operational level, these strategies should prominently manifest in procurement processes, particularly in the acquisition of goods and services related to the maintenance and repair of the university's extensive infrastructure.

In the existing literature, a considerable amount of research has been devoted to energy efficiency in the context of residential and commercial buildings. However, university buildings present a unique case due to their diverse functions. They encompass aspects of both commercial and residential buildings and include specialized facilities such as laboratories, libraries, and sports centers, each with distinct energy needs and consumption patterns.

Despite the unique energy profiles of university buildings, there remains a significant gap in the literature regarding energy efficiency in this context. Even fewer studies have examined the role of procurement practices in achieving energy efficiency goals for these institutions. Procurement documents offer a wealth of data regarding the design, construction, and maintenance of university buildings, which has remained largely untapped in the existing literature.

Moreover, given the importance of universities in society, both as major energy consumers and as influential institutions shaping future generations, it is crucial that they lead by example in the transition towards more sustainable practices. Therefore, there is an urgent need for research in this area to guide the transformation of university buildings towards more efficient energy consumption.

This study aims to fill this research gap by conducting an analysis of public procurement documents that contractors are obliged to publish in the domain of Polish university buildings, with a view to identifying opportunities for improving energy efficiency.

1.2. Motivation

Sustainable public procurement policies and practices vary from country to country, as evidenced by Walker and Brammwer [3] or by Pacheco-Blanco and Bastante-Ceca [4] who analyzed the visibility of green public procurement initiatives in public universities from internal and external perspectives by analyzing selected case studies.

As public administrations, universities should purchase products, services, and buildings with high energy performance that are compatible with economic efficiency, economic feasibility, and sustainability. The research shows that universities publish Energy Efficiency (EE) strategies focusing more on cost savings and compliance, while renewable energy strategies are more linked to public support and environmental awareness [5]. There are a limited number of scientific studies considering green procurement practices in universities; some of which have already been partially explored and published by Pacheco-Blanco and Bastante-Ceca [4] and by Filho et al. [6]. The conducted analysis of the literature indicated a gap in the literature regarding the possibility of wider monitoring of real initiatives at the operational level in the field of green public procurement conducted by universities. This article involves text mining to automatically analyze energy efficiency issues. This method was effectively used to a similar extent by Haddadi et al. [7], who examined public procurement of the Moroccan ICT, and by Tavana et al. [8], who applied this technique to provide a systematic review of the literature prioritizing the fields of energy and sustainability, and by Abdelaziz et al., who examined energy consumption in buildings [9].

The domain of university buildings presents unique opportunities for research. As educational institutions, universities have a substantial role in leading societal change, including the transition to energy-efficient practices. These institutions are not only significant consumers of energy but also influential actors in shaping the behaviors and attitudes of future generations towards sustainability.

Research on text mining has been conducted for many years. In the literature there are still gaps that describe the use of these techniques in developing applications on real documents. The authors of this paper assume that there remains a contextual gap in the

case of geographic area [10], taking into account operations management research [11], which can be important to investigate. Each country may have different paths that reflect the country's reality and culture, which was presented by Young et al. [12]. Finding patterns in tenders allows for understanding what the current procurement market looks like [13].

Moreover, the analysis of public procurement documents provides a rich source of data to understand the current practices and challenges in this transformation process. These documents detail the procurement of materials, techniques, and technologies used in the construction and maintenance of university buildings, providing insight into the economic, regulatory, and technical aspects of energy efficiency implementations.

1.3. Research Questions

This paper seeks to answer the question: How can the analysis of public procurement documents inform and improve the transformation of university buildings towards more efficient energy consumption?

To reflect the energy efficiency activities in public procurement at the operational level among Polish higher education institutions, the following research questions were considered:

- Q1. What kind of EE topics can be recognized in public procurement documents?
- Q2. Is it possible to notice the differentiation of topics in different regions?
- Q3. Is it possible to determine the dynamics of the reported topics?

1.4. Research Objectives

The objectives of this research were:

- To analyze the current procurement practices in relation to the energy efficiency of university buildings;
- To identify gaps and potential areas for improvement in these practices;
- To propose recommendations for more sustainable and efficient procurement strategies.

1.5. Paper Structure

The next parts of the paper are organized as follows: Sections 2–4 review the relevant literature on energy efficiency in public buildings and public procurement. Section 5 details the methodology used for the analysis of procurement documents. Sections 6 and 7 present the findings of the analysis. Section 8 discusses these findings and their implications. Finally, Section 9 concludes the paper and offers suggestions for future research.

2. Energy Efficiency and Public Buildings

It was noticed by the European Commission (EC) more than 10 years ago that about 40% of final energy consumption (and 36% of greenhouse gas emissions) was recorded in buildings. Therefore, buildings have been identified as one of the key areas of action for EE [14]. Since the problem was identified by the EC, several initiatives have been introduced, such as regulations, guidelines, and provisions. It was also noted that it is crucial to put more emphasis on EE in the public sector, including public procurement, renovation of public buildings, and encouraging higher construction standards. The EU has taken action by showing that environmental and energy considerations are being considered in buildings occupied by public authorities and buildings frequented by the public. For that, the Energy Performance of Buildings Directive (2010/31/EU, as amended 2018/844/EU) was introduced and the Commission established a set of standards and accompanying technical reports called Energy Performance of Buildings Standards (EPB standards), which are managed by the European Standardization Committee (CEN/CENELEC). Another initiative—General Block Exemption Regulation (GBER)—identifies support for investments, including electric vehicle charging devices/points in buildings; an integrated on-site power plant producing electricity or heat from renewable sources, including related on-site storage facilities; and investments leading to the digital transformation of buildings to enhance their smart readiness [9]. The Commission indicated that EU member states carry out energy efficient renovations of at least 3% of the total floor area of buildings owned

and occupied by central governments [15]. To support the owners of buildings, several best practices have been published. Among them, the following examples of retrofitting techniques listed by the EU Green Best Practice Platform [16] are changing insulation materials, increasing insulation thickness and improving the environmental performance of roofs, using more efficient glazing sashes and frames, using internal and external shading devices, improving airtightness by improving doors, adding fast-acting doors, sealing air leakages and creating buffer sections, and maintenance. A valuable initiative in which many public building owners were involved was public–private partnerships programs that offered wide support for refurbishment and helped to optimize energy savings performance contracts [17]. The potential of this initiative seems to be untapped by most universities.

The improvement of energy efficiency is a part of the sustainable development policy. The EE area is intensively explored by scientists. Relationships between management, energy efficiency, and sustainable development have already been the subject of research by Polish and foreign researchers. The debate on energy efficiency is multi-threaded, touching on financial, economic, social, and technical issues, barriers, and effects [18].

Among the areas that improve the energy efficiency of building use, the literature lists processes and equipment related to heating, cooling, ventilation, and building lighting. The investments analyzed in this area include thermal modernization, insulating walls ceilings and floors, replacing windows, purchasing equipment that generates energy or heat from renewable sources and storage equipment, or installing smart devices and apps to improve the efficiency of energy consumption in the building [19]. The literature also indicates methods for identifying the impact of key building elements on the environment throughout the life cycle of the element [20]. Currently, universities in Poland are struggling with several issues related to planning their major renovations and implementing energy efficient modernizations. Only a few of the buildings have energy performance or audit documentation.

3. Public Procurement Regulations

Sustainable and green public procurement is the instrument aimed at activating the sustainable development of public institutions, including activities related to energy efficiency. As public and educational institutions, universities play an important role in promoting sustainable development and implementing its principles in conducting their activities [21]. In the case of the Polish economy, the share of public procurement in the GDP is about 10%, amounting to about 200 billion PLN [22].

The main Polish legal act on the basis of which public procurement of contractors for construction works is conducted is the Act of 11 September 2019, the Public Procurement Law (Journal of Laws of 2021, item 1129, 2269).

The technical parameters of products and technical conditions of buildings are regulated by the Act of 20 May 2016 on energy efficiency, which is compliant with Directive 2012/27/EU of the European Parliament and the Council on Energy Efficiency, amendments to Directives 2009/125/EC and 2010/30/EU, and repealing Directives 2004/8/EC and 2006/32/EC. Another regulation that affects GPP is the Act of 29 August 2014 on the energy performance of buildings, which is the implementation of Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010. To facilitate the understanding of the description of the subject of the contract, in addition to the requirement of using unambiguous, exhaustive, and understandable terms, the names and codes specified in the Common Procurement Vocabulary are used. Contracting entities may describe the parameters concerning the demand for primary energy and the heat transfer coefficient in relation to the current standard, pointing to more optimal parameters, and may (but do not have to) take them into account in the tender evaluation criteria [22].

The contracting authorities are obliged to publish contract notice in the Official Journal of the European Union for the contract value above EU thresholds and to place the notice in the Public Procurement Bulletin for contracts below this threshold. After publication, the contracting authority may additionally place the notice on a different platform.

4. Literature Review

The trends in sustainable public procurement practices of universities is a relatively new field, and it has not been much described in the scientific literature.

Researchers consider green public procurement as an important policy tool for the implementation of sustainable consumption in institutions of higher education. Universities play a responsible role in forming and promoting a “green” culture of behavior among students and other stakeholders. The authors of [4] found that 21.5% of universities managed to implement various initiatives in practice by implementing green procurement procedures. It was also found that the environmental criteria that universities included in the specifications of public procurement contracts did not have a clear correlation with the mission of the institution.

A previous study [23] considered the possibility of improving the implementation of environmental criteria in decision-making procedures during tenders and concluding contracts through the prism of assessing the degree of inclusion of environmental criteria. Five approaches were proposed for the employees of public procurement departments to make a decision on choosing a “green” supplier, in particular: 1—ignore environmental criteria; 2—include environmental criteria as a part of existing other criteria; 3—include environmental criteria in the tender documentation (before the beginning of the selection process); 4—integrate environmental content into traditional procurement criteria; 5—include separate environmental criteria as mandatory in the system of public procurement rules.

According to research conducted by the European University Association, 38% of 305 investigated institutions had a sustainable procurement process in place across the institution and 48% had one at least partially implemented [24]. This literature demonstrated that a lot of work had been done in the area of recognizing sustainable public procurement or green procurement at the universities [12,25–27].

Green procurement studies focus mainly on a specific group of products (e.g., food, energy, water, equipment, paper products, books) [28–31] or around evaluating current states and developing frameworks and best practices for sustainable supply chains [25,27,32,33]. Part of the research considers the behavior and perceptions of individual buyers (in the case of universities—students) towards buying green products [32,34].

These topics are studied using various methods and tools. Most of the applied research methods used to analyze sustainable public procurement (SPP) and procurement processes adopt questionnaires and manual or expert reviews. This type of research shows certain limitations owing to the shortcomings of using a survey method, as surveys present perceptions rather than performed actions. This approach was taken by Bala et al. [25], who identified barriers against widespread adoption of green purchasing practices and proposed actions to overcome them based on the case of Universitat Autònoma de Barcelona, and by Elser and Michael [28], who investigated the roles of administrative support employees and minor procurement decisions in the sustainable supply chain based on purchasing paper products. The same method was applied by Heravi et al. [27] for developing a green university framework for the University of Tehran. Brazilian researchers [33] investigated the relationship between environmental training and the adoption of sustainable procurement in three Brazilian public/state universities. The literature also shows that over the last years, some researchers have applied text mining to heterogeneous, multilingual procurement documents, e.g., Zhang et al. in the healthcare sector [35] and Modrušan, Rabuzin and Mrcsic for the analysis of the detection of suspicious or corrupt procedures in the public procurement process [36]. In China, wide research using 42,369 public procurement documents for 2015–2020 was performed to find implementation patterns of SPP [37]. Choi et al. proposed a tool for identifying and managing risk provisions for ITB and contract documents [38]. A wider overview of research articles on the application of text mining techniques in the field of supply chain risk management was reported by Shash et al. [39].

Numerous studies were conducted and published on the peculiarities of implementing green public procurement in institutions of higher education, which created the basis for developing effective methods of applying a system of environmental indicators for procurement decision-making in universities. A more thorough systematic review of the application of green technologies for university procurement can be obtained with the use of bibliographic databases and relevant programs, in particular, VOSviewer.

The Scopus scientific-metric database was selected to search for publications. For the search query, the terms “green public procurement”, “universities”, and “sustainability” yielded 12 documents. Processing of bibliographic data using the VOSviewer program made it possible to create a scientometric map (Figure 1). The VOSviewer program selected 117 terms and grouped them into ten clusters.

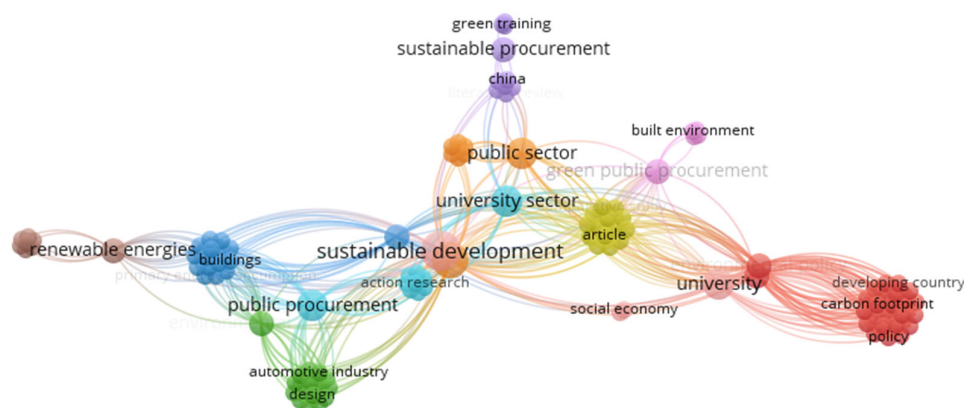


Figure 1. Cluster visualization map for co-occurring keywords of a preliminary literature review of green public procurement in universities.

Cluster #1 (marked in red) included 25 keywords, such as “carbon footprint”, “competition”, “energy consumption”, “environmental labeling”, “environmental management”, “environmental policy”, “government”, “life cycle”, “organization”, and “sustainability standards”. Further, this cluster was summarized as the “development of environmental management principles”.

Cluster #2 (marked in green) contained 16 keywords, such as “design”, “environmental impact”, “environmental impact reduction”, “environmental levels”, “extended producer responsibility”, “functional economy”, “policy instruments”, “policy measures”, and “product–service system”. The keywords in this cluster were summarized as “influencers of green procurement”.

Cluster #3 (marked in blue) had 16 keywords, such as “buildings”, “climate change”, “cogeneration plants”, “current distribution”, “ecological footprint”, “energy conservation”, “energy efficiency”, “energy policy”, “energy utilization”, and “environmental protection”. This cluster was summarized as “energy-saving infrastructure creation policy”.

Cluster #4 (marked in yellow) had 14 elements, such as “budget”, “comparative study”, “conservation of natural resources”, “environmental criteria”, “environmental economics”, “human experiment”, “organization and management”, and “organizational policy”. This cluster was briefly described as “environmental criteria system for green procurement”.

Cluster #5 (marked in purple) contained 10 keywords, such as “organizational policy”, “green training”, “higher education”, “public sector universities”, “structural equation modeling”, “sustainable human resource management”, “sustainable procurement”, and “sustainable supply chain management”. This cluster was summarized as “formation of green public procurement execution competence”.

Cluster #6 (marked in green) contained 10 elements, such as “action research”, “living lab”, “local”, “public procurement”, “public service”, and “university sector”. This cluster was grouped as “investigation of features of green procurement”.

Cluster #7 (marked in orange) had 9 keywords, such as “governance approach”, “innovation ecosystems”, “logit models”, “open innovation”, “policy approach”, “public sector”, “social policy”, and “sustainability”. This cluster was summarized as “company-oriented management”.

Cluster #8 (marked in brown) had 8 keywords, such as “acoustic performance”, “energy consumption”, “green buildings”, “lighting performance”, “overall building performance”, “renewable energies”, “thermal comfort”, and “thermal performance”. This cluster was summarized as “technical characteristics of a green university building”.

Cluster #9 (marked in pink) had 5 keywords, such as “built environment”, “green building certification systems”, “green public procurement”, “mandatory standards”, and “sustainable development goals”. This cluster was summarized as “established system of public green procurement in the university”.

Cluster #10 (marked in beige) had 4 keywords, such as “social economy”, “sustainable development”, “sustainable public procurement”, and “university”. This cluster was summarized as “socio-economic values from public green procurement at the university”.

The results of the keyword-sharing analysis indicated that research on green public procurement in universities has mainly focused on establishing the basic principles of creating a sustainable procurement system and analyzing the success of environmental initiatives in universities. In general, the sequence of clusters could be presented as a trajectory of the technological maturity of the university regarding the construction, introduction, and development of the green public procurement system.

Figure 2 presents an overlay of the visualization networks generated from VOSviewer, showing the trends in research topics in terms of co-occurring keywords over the period 2010–2023.

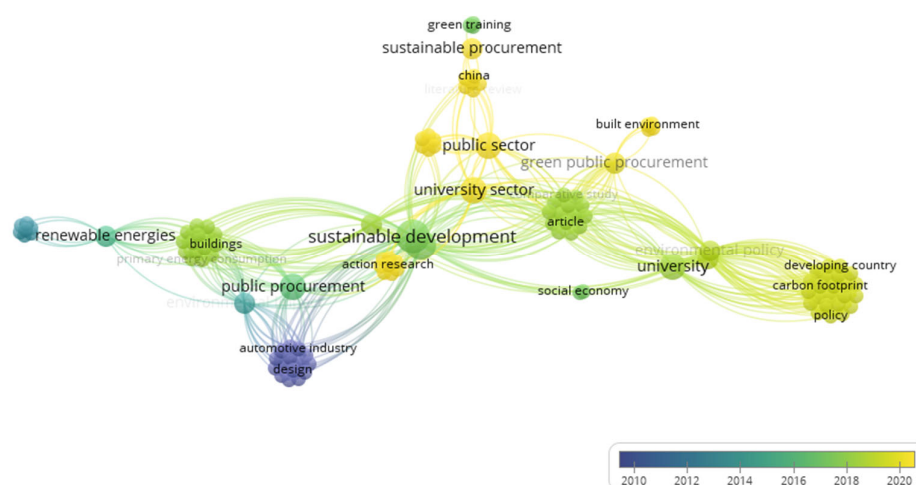


Figure 2. Overlaying visualization networks of research topics.

Sometimes the affiliations and countries/territories of the authors of the documents were important components of the analysis of bibliometric data. The leaders, in terms of the number of publications on the subject of “management of green public procurement in universities”, were Italy [40–42], China [43,44], and Spain [45,46]. The scientometric map containing 23 terms was constructed for these publications (Figure 3).

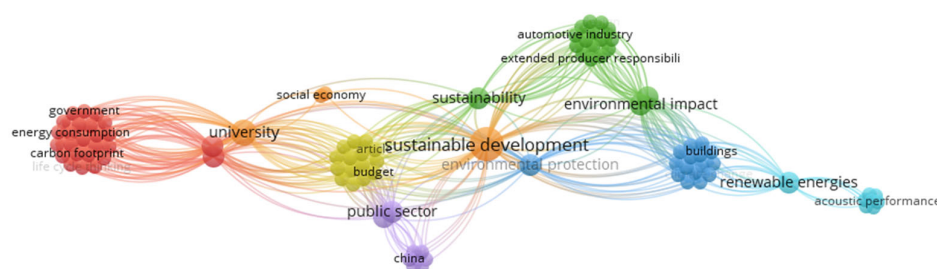


Figure 3. Scientometric map of terms in publications from Italy, China, and Spain.

5. Materials and Methods

This section outlines the methodology used in this study to analyze public procurement documents concerning the transformation of university buildings towards more efficient energy consumption.

Existing research has substantiated that the examination of openly accessible public procurement documents, obtained from e-procurement platforms, can facilitate the identification of themes that resonate with universities in the realm of sustainable construction. This study specifically focused on the domain corresponding to Cluster #8, as unveiled in the literature review section. This cluster was characterized by eight salient keywords, namely: acoustic performance, energy consumption, green buildings, lighting performance, overall building performance, renewable energies, thermal comfort, and thermal performance. These terms formed the point of our investigation into sustainable procurement practices in the university building sector.

The general scheme of analysis shown in Figure 4 illustrates the overall flow of the methodology employed in this study. It provides a visual representation of the key stages and processes involved in analyzing the public procurement documents in the domain of transforming university buildings towards more efficient energy consumption.



Figure 4. General scheme of analysis.

The methodology began with data collection from the e-procurement platform, followed by data preprocessing to clean and prepare the titles of documents for analysis. This included steps such as removing duplicates, filtering out irrelevant titles of documents, and text cleaning.

Once the data were preprocessed, the Bag of Words model was applied to represent the titles of the documents numerically, capturing the frequency of words within each title of the document. This representation served as the foundation for the subsequent analysis.

Next, the Latent Dirichlet Allocation (LDA) model was utilized for topic modeling, allowing the extraction of salient topics related to energy efficiency in university buildings. The identified topics provided insight into the key thematic areas presented in the titles of the procurement documents [47].

The trend analysis stage examined the changes and patterns in energy efficiency practices over time. This involved analyzing the identified topics related to energy efficiency in university buildings and visualizing the trends to understand the evolution of procurement practices and energy efficiency initiatives.

5.1. Data Collection

For this study, our primary data source was the public procurement documents related to university buildings. These documents were retrieved from the Public Procurement Bulletin (e-procurement platform), which serves as a comprehensive repository of procurement information for public entities. This platform was chosen due to its wide usage

and the richness of the data it contains, making it a valuable resource for understanding procurement practices.

Our data collection focused on procurement documents over a period of three years, spanning from 2020 to 2023. This timeframe was selected to provide a current and relevant view of procurement practices, while also allowing for the identification of trends over time.

In early 2021, the Public Procurement Office implemented new regulations and launched a revamped e-procurement platform, which also incorporated the Public Procurement Bulletin. For the purposes of our research, we sourced documents from the year 2020 from the platform's prior iteration. Documents published in the 2021–2022 period were gathered from the newly established platform.

The documents collected included requests for proposals (RFPs) of various types of procedures examined, including open procedure, restricted procedure, competitive negotiated procedure, competitive dialogue, and direct procurement. These documents provided detailed information on the goods and services being procured, the requirements and criteria for selection, and the terms of the contracts.

The Polish versions of documents were analyzed. EU threshold amounts in public procurement related to construction works for the time period 2020–2023 are presented in Table 1.

Table 1. Topic analysis—EU threshold amounts in public procurement for the time 2020–2023.

Domain of Public Procurement	EU Threshold Valid until 31 December 2021	EU Threshold Applicable in 2022 and 2023
Construction Works	EUR 5,350.00–PLN 22,840,755 The average euro exchange rate in public procurement valid until 31 December 2021. 4.2693	EUR 5,382,000–PLN 23,969,275 The average euro exchange rate in public procurement for 2022 and 2023 4.4536

There are around 360 universities in Poland (362 registered in the academic year 20/21), of which 137 are public universities. It must be noted that higher education institutions in Poland have different types and names (college, university, academia, etc.), so queries to obtain data in both cases applied the following criteria:

- The type of contract was equal to 'construction work';
- The name of the contractor included the text was 'university', 'polytechnic', 'academy' or 'school of higher education';
- The date of contract notice publication was between 1 January 2020 and 31 December 2022.

In total, we collected and analyzed 6573 documents. This dataset formed the basis of our text analysis and topic modeling in the subsequent stages of our methodology.

5.2. Data Preprocessing

The collected procurement documents underwent a series of preprocessing steps to prepare them for analysis. These steps included removing duplicate projects to eliminate redundancy, filtering out documents with missing or generic titles, and performing text cleaning operations. Text cleaning involved removing unnecessary characters, converting the text to a consistent case, and removing non-relevant sections of the document. Tokenization and stop word removal were applied to break the text into individual words and remove common stop words, while lemmatization was employed to reduce words to their base or root form.

The following steps were applied to the selected documents:

- Removal of Duplicate Projects

To eliminate redundancy, we identified and removed duplicate projects within the dataset. This was achieved by comparing key project attributes such as project titles, contract numbers, and procurement entities. Only one instance of each unique project was retained, ensuring that duplicate projects did not bias the subsequent analysis.

- **Filtering Irrelevant Documents**

To focus on relevant information, we filtered out documents with missing titles or generic titles such as ‘construction works’. These documents were deemed less informative for our study and could introduce noise into the analysis. By excluding them, we aimed to enhance the quality and accuracy of the data.

- **Text Cleaning**

To prepare the textual content of the documents, we performed text cleaning procedures. This involved removing unnecessary characters, such as special symbols and punctuation marks, and converting the text to a consistent case (e.g., lowercase). Additionally, we removed any non-relevant sections of the document that were not pertinent to the analysis, such as contact details or legal disclaimers.

- **Tokenization and Stop Word Removal**

The preprocessed text was then tokenized, which involved breaking the text into individual words or terms. We also removed common stop words, such as ‘the’, ‘and’, and ‘is’, as these words typically do not carry significant meaning for the analysis and can introduce noise.

- **Lemmatization**

To reduce words to their base or root form, we applied lemmatization. This process transformed words with different inflections or variations into a standardized representation. For example, ‘buildings’ and ‘building’ would both be lemmatized to ‘building’.

As a result of data preprocessing, a total of 3367 documents were identified and selected for inclusion in our study (Table 2). This represented a subset of the initial 6573 documents available on the platform.

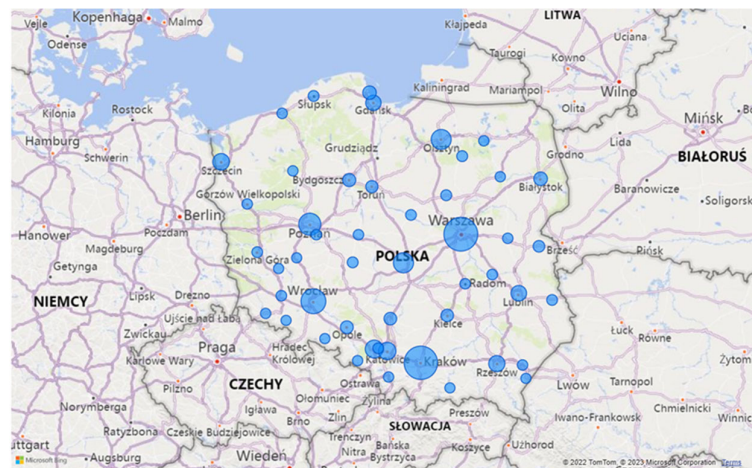
Table 2. Document output classified according to year of publication.

Year	2020	2021	2022	Total
Number of documents found	2949	1044	2580	6573
Number of selected documents	1148	1015	1204	3367

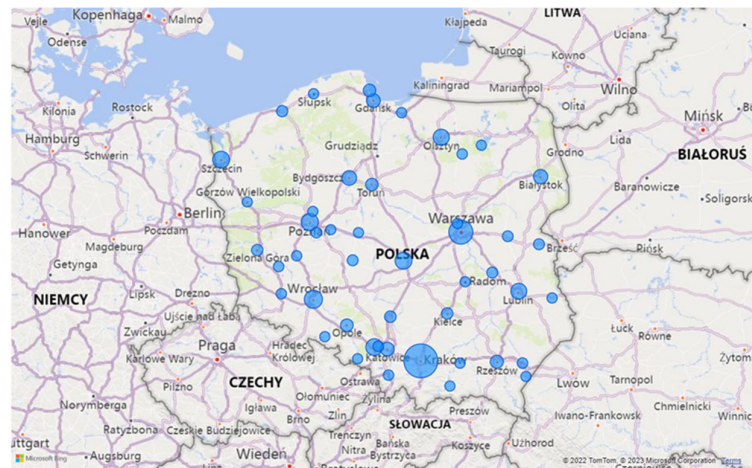
The highest numbers of offers were registered in Silesian, Mazovian, Lesser Poland, and Lower Silesian provinces. This distribution reflected the number of academic centers registered in Polish cities. The detailed data of selected objects can be seen in Table 3. The cartograms of the analyzed documents are shown in Figure 5 (a map plot with bubble sizes proportional to the number of documents).

Table 3. Number of documents by province.

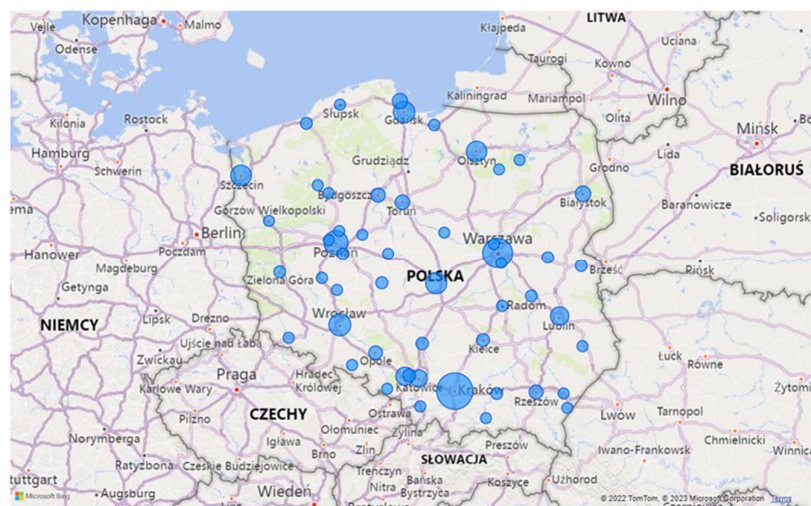
Region	2020	2021	2022
Lower Silesian	109	70	80
Kuyavian-Pomeranian	36	56	49
Lublin	58	63	68
Lubusz	8	12	8
Lodz	69	55	72
Lesser Poland	160	194	186
Masovian	179	120	147
Opole	20	22	22
Subcarpathian	46	29	24
Podlaskie	26	34	33
Pomeranian	67	52	107
Silesian	132	105	126
Świętokrzyskie	17	11	14
Warmian-Masurian	76	55	73
Greater Poland	95	73	117
West Pomeranian	50	64	78



(a) 2020



(b) 2021



(c) 2022

Figure 5. Cartograms of analyzed documents in 2020, 2021, and 2022.

5.3. Dictionary Creation

As a part of our data preprocessing, we created a dictionary of words from our corpus of procurement documents. After initial text cleaning steps, such as removing non-relevant characters, numbers, and punctuations, we tokenized the text in each document, splitting the text into individual words or terms.

We then removed common stop words from the data using a dictionary of common stop words developed specifically for this study. Stop words are words that occur very frequently but do not carry significant meaning, such as “the”, “and”, “is”, etc. Additionally, we applied lemmatization to reduce words to their base or root form (e.g., “buildings” became “building”).

Following these steps, we compiled a list of all unique words appearing in the corpus and applied text conversion using the n-gram model using 1-g, 2-g, and 3-g schema creating our dictionary (Table 4).

Table 4. Topic analysis—excerpt of list of n-grams and lemmatization handling.

Base Form of the Word	Possible Forms of the Word	Possible n-g
Photovoltaic	Photovoltaics pv Photovoltaically Photovoltaical	Photovoltaic effect Photovoltaic power system

The dictionary played a crucial role in our study, as it formed the basis for the Bag of Words model and the subsequent LDA topic modeling. It allowed us to quantify the procurement documents and analyze them from the perspective of topic prevalence.

5.4. Bag of Words Model

To analyze the textual content of the preprocessed procurement documents, we employed the Bag of Words model, a widely used technique in natural language processing.

The Bag of Words model treats each document as a “bag” or collection of words (n-g), disregarding the order and structure of the words. It represents each document as a numerical vector, where each dimension corresponds to a unique word or term (n-g) in the entire corpus of documents. The value in each dimension indicates the frequency of that word (n-g) in the document (Table 5).

Table 5. Topic analysis—excerpt from the dictionary related to energy efficiency constructions applied in the study and the number of times they occurred in the titles of the procurement documents.

Keyword	Number of Appearances
Transformer Station	26
Photovoltaic Installation	33
Air Conditioning	103
Elevator	61
Lighting System	63
Window Woodwork	76
Ventilation System	77
Heating And Cooling	24
Sanitary Facilities	31

In our study, we created a dictionary of unique words (n-g) from the preprocessed titles of the documents, assigning each word (n-g) a numerical identifier. The lengths of the titles of green procurement documents ranged between 2 and 108 words. This dictionary served as a reference for encoding the documents into the Bag of Words representation. For each document, we counted the occurrence of each word (n-g) from the dictionary and assigned the respective frequency as the value in the corresponding dimension of the document vector.

The resulting Bag of Words representation allowed us to quantitatively analyze the presence and frequency of specific words or terms in each document. This representation formed the basis for further topic modeling and trend analysis.

By employing the Bag of Words model, we were able to transform the textual procurement data into a numerical format that facilitated the exploration of energy efficiency

themes and the examination of procurement practices in the context of transforming university buildings towards more efficient energy consumption.

5.5. Latent Dirichlet Allocation (LDA) Model Setting

Topic modeling is an active domain of Natural Language Processing (NLP) research that can be described as an unsupervised machine learning technique used to find topics in a given text. Its task is to discover the meaning of a given text without having to study it as a whole, i.e., to find the semantic meaning in the content of the text. The most popular algorithms used in topic modeling are Latent Semantic Analysis, Non-Negative Matrix Factorization, Probabilistic Latent Semantic Analysis, and Latent Dirichlet Allocation.

Latent Dirichlet Allocation (LDA), applied in the study, is an algorithm that creates groups of words (topics) on the basis of their co-occurrence in a text and then assigns a given label to the created topic with a certain probability. The model is named after a German mathematician of French origin, Johann Dirichlet (1805–1859). Johann Dirichlet's scientific achievements were used in 2003 by David Blei, Andrew Ng, and Michael I. Jordan [47] to conduct topic analysis using Machine Learning techniques, which turned out to be the beginning of a trend that influenced the increased popularity of using the model in topic analysis.

LDA is a generative probabilistic model that assumes each document is a mixture of a certain number of topics, and each topic is characterized by a distribution of words. It allowed us to understand the hidden thematic structure in the document collection, thereby revealing the key themes in the procurement of university buildings related to energy efficiency.

In this study, we employed Latent Dirichlet Allocation (LDA) as a topic modeling technique to analyze the energy efficiency practices in university buildings. LDA is built upon several key assumptions that guide the topic modeling process and enable us to uncover the underlying thematic structure within the procurement documents. These assumptions played a crucial role in our analysis and interpretation of the topics related to energy efficiency.

Firstly, we leveraged the assumption of thematic similarity, which posits that documents discussing similar topics tend to use similar word groups. By applying LDA, we aimed to identify common word groups within the documents that reflected shared themes related to energy efficiency in university buildings. This allowed us to uncover the latent thematic structure and understand the key topics prevalent in the procurement documents.

Secondly, we relied on the assumption of word groupings, which suggests that words frequently appearing together within documents indicate their association with a specific topic. Through LDA, we identified these cohesive word groups, enabling us to delineate distinct topics within the corpus. This assumption helped us to uncover the semantic relationships and dependencies among words, facilitating a more comprehensive understanding of the energy efficiency discourse.

Furthermore, we incorporated the assumption of a latent document–topic distribution. This assumption suggests that each document exhibits a probability distribution of belonging to different topics. By inferring this distribution using LDA, we gained insight into the prevalence and importance of various energy efficiency topics within individual documents and across the entire corpus. This allowed us to discern the thematic composition of the procurement documents and assess the relative emphasis on different energy efficiency aspects.

Lastly, we considered the assumption of a latent word–topic distribution. This implies that each word has a probability of belonging to different topics. By estimating this distribution using LDA, we could determine the topic proportions for each word, facilitating the inference of topics within the documents. This assumption enabled us to capture the variability of word usage across different energy efficiency themes and enhanced our understanding of the nuanced topics discussed in the procurement documents.

By incorporating these LDA assumptions, we aimed to uncover the underlying themes and trends in energy efficiency practices in university buildings, contributing to a deeper understanding of sustainable transformations in this domain. The utilization of LDA as a topic modeling technique allowed us to extract meaningful insight and derive actionable recommendations based on the identified topics.

Our LDA model was implemented using the Gensim and pyLDAvis libraries in Python due to their efficiency and flexibility. The LDA model had several parameters that need to be selected. The most important were the number of topics and the Dirichlet hyperparameters, which controlled the distribution of topics within the documents and words within topics.

After several iterations and evaluations, we decided on a model with 15 topics for our analysis, where 15 seemed to provide a balance between granularity and coherence.

We employed the Gibbs sampling technique for the inference in the LDA model due to its effectiveness and relative simplicity. We set the hyperparameters, alpha and beta, to default symmetric values (alpha influences the document–topic density, and beta influences the topic–word density).

We evaluated the quality of our topics using coherence scores, which provided a quantitative measure of the semantic interpretability of the topics. The model with the highest coherence score was selected as the optimal model.

After training the model, we analyzed the resulting topics and interpreted them in the context of our research objectives.

In the Results section, we present our findings from the LDA topic modeling.

5.6. Trend Analysis

The trend analysis of topics evolution aimed to understand how the identified topics related to energy efficiency in university buildings changed over time. Within each specific time period (2020, 2021, and 2022), the distribution of topics was analyzed to understand their prevalence and importance. The frequency or proportion of documents assigned to each topic was calculated, and the significance of each topic within the time period was examined. This allowed for the identification of the most prominent and relevant topics during each specific time period.

To visualize the trends in topic evolution, line graphs were employed. These visual representations showcased the changes in the prevalence or prominence of topics across the different time periods, enabling a clear understanding of the evolving focus and relevance of topics related to energy efficiency in university buildings.

6. Results

Following the application of the LDA model to the procurement documents and subsequent analysis, several key topics emerged which pertained directly to energy efficiency in university buildings. In this section, the structure of the domain related to public procurement documents is presented. The conducted LDA analysis allowed us to identify 15 clusters. Three of them were rejected due to low interest. It was necessary to reject certain clusters based on two key criteria: redundant and sparse topics.

- **Redundant topics:** In our analysis, we noticed that two clusters contained overlapping themes. Clusters had similar keywords such as “installation”, “renovation”, and “adaptation”. Given the high similarity between these clusters, we decided to keep only one representative cluster to avoid redundancy and ensure a more concise and interpretable set of topics.
- **Sparse topics:** In our analysis, we discovered a cluster that had limited document coverage, meaning it was mentioned in only a small subset of the procurement documents and therefore lacked sufficient evidence and statistical support to make meaningful conclusions. Therefore, we chose to disregard this sparse topic in our analysis, as it may not have provided reliable insight into energy efficiency practices in university buildings.

By considering both the redundancy and sparsity of topics, we refined our topic analysis to focus on the most distinctive and representative clusters. This ensured that our findings were robust, relevant, and effectively captured the essential energy efficiency themes prevalent in the procurement documents.

The decision to reduce the number of clusters in the analysis had implications for the number of documents included in the final analysis. The branch of analyzed documents was also not assigned to any cluster. This branch of documents did not exhibit a clear thematic association with any specific topic or cluster identified through the analysis. The existence of this branch of unassigned documents highlighted the complexity and diversity of the dataset. It emphasized that not all documents could be easily classified into predefined topics or clusters and that there may be nuances, outliers, or emerging themes that require further exploration or refinement in the analysis. The final number of classified documents was equal to 2766. Table 6 presents the obtained clusters.

Table 6. Topic analysis for 2020–2022.

Cluster	Keywords	Label	2020	2021	2022
1	['reconstruction of the energy generation system', 'room adaptation', 'energy media distribution']	Energy Generation System	81	80	73
2	['elevator installation', 'replacement of window joinery', 'bathroom renovation', 'renovation of roofing', 'wall insulation', 'installation works', 'assembly of a new lift']	Reducing Energy Consumption	83	86	102
3	['application of sun protection film', 'replacement of window joinery', 'assembly', 'antenna', 'installation', 'technics', 'reconstruction', 'adaptation', 'computers']	Thermal Protection Of Windows	42	50	45
4	['façade renovation', 'input', 'exchange', 'customization', 'installation', 'upgrade', 'chimney', 'electrical', 'modernization', 'technology']	Complex Reconstruction	61	54	52
5	['installation', 'upgrade', 'extension', 'adjustment', 'rebuilding', 'power supply', 'electrics', 'photovoltaics', 'air conditioning', 'heating']	Photovoltaic Roof System	77	49	69
6	['rebuilding', 'installation', 'adaptation', 'sanitation', 'upgrade', 'replacement', 'center', 'network', 'technological']	Energy Savings	47	39	50
7	['installation', 'roofing replacement', 'upgrade', 'customization', 'rebuilding', 'works', 'ventilation', 'air conditioning']	Ventilation System	52	38	49
8	['façade', 'replacement', 'lighting installation', 'upgrade', 'works', 'reconstruction', 'coverage']	Lighting System	64	52	52
9	['upgrade', 'customization', 'IT', 'centre', 'elevation', 'reconstruction of installation']	Electrical Installation	67	53	68
10	['replacement of window joinery', 'roofing replacement', 'ceramic tiles', 'installation', 'replacement', 'thermal insulation']	Thermal Insulation Works	86	108	98
11	['modernization of lighting', 'centre', 'adjustment', 'modernization of sanitary facilities', 'installation', 'upgrade', 'lighting', 'sanitation', 'electricity installation', 'mounting']	Sanitary Facilities	96	67	121
12	['air conditioning', 'central heating installation', 'room modernization', 'roof renovation', 'modernization of installation']	Cooling & Heating System	93	71	96

Topic 1: Energy Generation Systems

The first topic encompassed n-grams such as “energy generation systems”, “room adaptation”, and “energy media distribution”. This topic suggested a strong emphasis on the procurement of energy-efficient technologies. The topics within this cluster revolved around various energy generation technologies and strategies employed in university buildings to enhance energy efficiency. The frequent appearance of this topic across multiple procurement documents indicated a growing trend towards the adoption of energy-saving technologies in university buildings.

Topic 2: Reducing Energy Consumption

The second topic included n-grams such as “elevator installation”, “window joinery replacement”, “bathroom renovation”, and “installation works”. This topic reflected considerations around the design and architectural elements of university buildings that could contribute to the reduction of energy consumption.

Topic 3: Thermal Protection of Windows

The third topic incorporated n-grams such as “sun protection film”, “window joinery replacement”, “installation”, and “assembly”. This topic implied a focus on adhering to enhancing the thermal efficiency and insulation properties of windows in university buildings. By analyzing the discussions within this cluster, researchers and practitioners can gain valuable knowledge to inform decision-making, design considerations, and energy efficiency initiatives related to windows in university buildings.

Topic 4: Complex Reconstruction

The fourth topic encompassed terms such as “façade renovation”, “modernization”, “upgrade”, and “customization”. This topic represented broader sustainability initiatives in universities, indicating that energy efficiency was being viewed within the larger context of reconstruction and renovation of complex university buildings.

Topic 5: Photovoltaic Roof System

The fifth topic encompassed terms such as “photovoltaic”, “air conditioning”, “heating system”, and “customization”. This cluster represented a specific aspect of energy efficiency in university buildings, focusing on the integration of photovoltaic roof systems and the optimization of air conditioning and heating systems. The topics within this cluster revolved around sustainable energy generation through photovoltaic technology and the efficient management of indoor climate control.

Topic 6: Energy Savings

The sixth topic encompassed terms such as “upgrade”, “extension”, “adjustment”, and “installation”. This cluster focused on energy-saving measures and sanitation practices in university buildings. It encompassed topics related to the efficient use of energy resources and the implementation of sustainable practices to promote cleanliness and hygiene within the building environment.

Topic 7: Ventilation Systems

The seventh topic encompassed terms such as “ventilation”, “air conditioning”, “roofing replacement”, and “installation”. This cluster represented a significant aspect of energy efficiency in university buildings, focusing on ventilation systems and their role in maintaining indoor air quality and optimizing energy consumption. The documents assigned to this cluster consistently discussed various aspects related to ventilation systems, including their design, installation, operation, and maintenance.

Topic 8: Lighting Systems

The eighth topic encompassed terms such as “lighting installation”, “replacement”, “upgrade”, and “works”. This cluster emerged as a prominent theme in the analysis of energy efficiency practices in university buildings. This cluster encompassed topics and discussions related to lighting technologies, strategies, and improvements aimed at enhancing energy efficiency and sustainability.

Topic 9: Electrical Installation

The ninth topic encompassed terms such as “IT”, “upgrade”, “center”, and “customization”. This cluster represented a distinct thematic group within the analysis of energy efficiency practices in university buildings. It encompassed topics related to electrical installations, technology infrastructure, and their impact on energy efficiency.

Topic 10: Thermal Insulation Works

The tenth topic encompassed terms such as “roofing replacement”, “windows replacement”, and “thermal installation”. This cluster encompassed topics related to the improvement of thermal insulation in university buildings and the reconstruction or renovation works. This cluster represented a key aspect of energy efficiency practices in university buildings and focused on enhancing the building envelope’s performance to reduce heat loss or gain and improve energy efficiency.

Topic 11: Sanitary Facilities

The eleventh topic encompassed terms such as “lighting modernization”, “sanitary facilities modernization”, and “lighting”. This cluster encompassed topics and keywords related to the design, functionality, and maintenance of sanitary facilities within university buildings. It focused on aspects such as plumbing systems, restroom amenities, hygiene standards, and accessibility. The topics within this cluster shed light on the considerations and practices associated with ensuring efficient and sustainable sanitary facilities in university buildings.

Topic 12: Cooling & Heating Systems

The twelfth topic encompassed terms such as “air conditioning”, “central heating”, and “room modernization”. This cluster represented a prominent topic within the analysis of energy efficiency practices in university buildings. This cluster encompassed discussions and procurement documents related to the design, installation, and optimization of cooling and heating systems in university facilities.

Through the distribution analysis of these topics, it was observed that Topics 10 and 11 were the most prevalent across the documents, suggesting a strong focus on integrating thermal insulation works and considering the modernization of sanitary facilities of university buildings.

7. Trend Analysis

To understand the evolution of topics over time, we conducted a trend analysis for the past three years. This was achieved by segmenting our dataset into three subsets corresponding to each year and applying the same LDA model to each subset. Figure 6 presents the dynamics of the classified documents according to the cluster in the considered period. The clusters with the highest number of documents are highlighted in green.

Cluster	2020	2021	2022	Trend	Total
1	81	80	73		234
2	83	86	102		271
3	42	50	45		137
4	61	54	52		167
5	77	49	69		195
6	47	39	50		136
7	52	38	49		139
8	64	52	52		168
9	67	53	68		188
10	86	108	98		292
11	96	67	121		284
12	93	71	96		260
other clusters	89	82	124		295
Total	938	829	999		2766

Green background - top 3 topics (highest number of documents)

Figure 6. Classified documents according to clusters.

2020: In the 2020 procurement documents, the dominant topics were “Sanitary Facilities” (Topic 11) and “Cooling & Heating Systems” (Topic 12). It seemed that during this year, the primary focus was on the modernization of sanitary facilities and transforming cooling and heating systems into more energy-efficient ones.

2021: The subsequent year, 2021, saw a significant increase in the prominence of “Reducing Energy Consumption” (Topic 2) and “Thermal Insulation Works”. This indicated a shift in focus towards procuring and implementing advanced energy-saving technologies.

2022: The most recent year, 2022, showed a balanced emphasis on both “Reducing Energy Consumption” (Topic 2) and “Sanitary Facilities” (Topic 11). This suggested a strong focus on integrating energy-efficient technologies and considering the modernization of sanitary facilities of university buildings.

Overall, the trend analysis revealed an evolving focus on energy efficiency in university building procurement practices. Starting from an emphasis on energy-efficient facilities installation, there was a noticeable shift towards adopting advanced energy-saving technologies and broader sustainability initiatives. This trend demonstrated the commitments of universities to continuous improvement in energy efficiency and sustainability.

Figure 7 presents the dynamics of the classified documents according to the region in the considered period. The regions with the highest number of documents are highlighted in blue. The dynamic nature of the graph suggested that the distribution of documents across regions changed over time, potentially indicating variations in energy efficiency practices, policy focus, or regional priorities. The scope of the study included data from a three-year period. Although this is a short period of time, it allowed us to draw a general trend, which for most regions in the last year of the study showed an increasing interest in energy efficiency.

















Region	2020	2021	2022	Trend	Total
Greater Poland	70	59	98		227
Kuyavian-Pomeranian	32	52	46		130
Lesser Poland	134	164	166		464
Lodz	58	45	60		163
Lower Silesian	94	61	77		232
Lubelskie	50	51	52		153
Lubusz	7	11	7		25
Masovian	151	97	124		372
Opolskie	19	18	17		54
Podlaskie	23	30	27		80
Pomeranian	51	37	82		170
Silesian	110	93	107		310
Subcarpathian	36	22	18		76
Świętokrzyskie	15	9	10		34
Warmian-Masurian	52	25	49		126
West Pomeranian	36	55	59		150
Total	938	829	999		2766

Figure 7. Classified documents according to region.

Figure 8 presents a detailed frequency analysis of topics (clusters) according to region. The detailed frequency analysis in Figure 8 shed light on the regional variations in the prominence of energy efficiency topics within university buildings. It allowed for a comparative assessment of the adoption, emphasis, or challenges associated with these topics across different geographical contexts. Differentiation between regions was identified in the sense of clusters. The majority of the universities in the regions made investments related to clusters 2, 10, 11, and 12. The smallest numbers of orders were observed for clusters 6, 7, and 8. The study did not include causal analysis, but it can be assumed that this differentiation may be due to infrastructure and the availability of funding sources.

Region ↓ Cluster # →	1	2	3	4	5	6	7	8	9	10	11	12
Greater Poland	28	45	11	9	16	10	11	13	10	21	19	18
Kuyavian-Pomeranian	10	19	5	12	16	6	10	6	7	10	10	8
Lesser Poland	37	40	25	32	38	13	34	21	36	39	51	43
Lodz	14	15	7	9	10	7	6	12	10	14	26	25
Lower Silesian	16	7	10	14	15	9	16	13	17	43	28	21
Lubelskie	12	17	10	6	21	8	8	6	15	19	12	11
Lubusz	6	5	1	2	3	1		2		1	2	1
Masovian	32	27	13	19	18	36	15	38	37	41	39	30
Opolskie	4	11	6	6	3	2	4		1	9	2	3
Podlaskie	5	9	9	3	5	1	5	5	5	8	9	5
Pomeranian	18	14	4	11	8	10	5	24	8	9	18	18
Silesian	27	30	8	25	19	15	6	14	18	39	27	34
Subcarpathian	4	2	3	6	3	10	3	3	4	13	7	9
Świętokrzyskie	2	3	3	2	4	1		4	1	2	5	1
Warmian-Masurian	6	11	7	3	4	4	10	3	9	10	16	26
West Pomeranian	13	16	15	8	12	3	6	4	10	14	13	7

Figure 8. Classified document according to clusters and regions.

8. Discussion

Our research aimed to analyze public procurement documents concerning the transformation of university buildings towards more efficient energy consumption. We set out to understand current procurement practices, identify gaps and potential areas for improvement, and propose recommendations for sustainable procurement strategies. Our findings have shed considerable light on these areas.

The prominence of “Thermal Insulation Works” (Topic 10) and “Sanitary Facilities” (Topic 11) across the procurement documents aligned with our research objectives and corroborates existing literature. Studies, such as those by Vallati, DiMatteo and Fiorini (2023) [48] and Piacentini (2017) [49] highlighted the importance of integrating thermal insulation works and sanitary facilities considerations into the design and architecture of buildings.

The increased emphasis on “Reducing Energy Consumption” (Topic 2) in 2022 was a promising sign of universities broadening their focus beyond modernization to wider sustainability goals. This finding was consistent with the broader societal trend towards sustainability, as noted by Filho et al. (2023) [50].

Our trend analysis revealed a shift in focus over the three years from modernization towards the adoption of energy-saving technologies and sustainability initiatives. This evolution suggested that universities were adapting their procurement strategies in response to advancements in technology and changing societal expectations.

However, our analysis also identified potential areas for improvement. Despite the importance of “Cooling & Heating Systems” (Topic 12), this topic was less dominant in the procurement documents. This could suggest a gap in current procurement practices, where cooling and heating strategies are not as prioritized as they could be. This aligned with the findings of Matana Junior et al. (2023) [51], who noted that air conditioning retrofit was among the most common practices related to energy efficiency to reduce consumption.

Based on these findings, we recommend that universities give greater consideration to cooling and heating systems in their procurement strategies. This could include providing additional data-driven energy savings and incorporating investment in digital electricity infrastructure and software into procurement processes—“Energy Savings” (Topic 6).

In conclusion, our analysis has provided valuable insight into the procurement practices of universities regarding energy efficiency. These findings contribute to the existing body of literature and provide a foundation for future research and improvements in this area.

The conducted topic analysis indicated the areas in which universities place orders for renovation work concerning EE. Differentiation of the topic according to regions was identified (Figure 6). The biggest reflections of EE practices were observed in Silesian, Masovian, and Lesser Poland. An increasing trend in most of the regions could be observed (9 out of 16). The largest numbers of initiatives were identified in cluster numbers 2,

10, and 11 (Figure 7). Differentiation between regions was identified in the sense of clusters (Figure 8).

A certain limitation of the study was the analysis of the document titles. Obtaining access to full documentation could affect the perspective of the discussed results. However, these documents, especially those published in earlier years, did not have the same structures, so their analysis would be difficult. Nevertheless, the authors assume that the areas would remain similar because, as indicated in the data description, many topics contained a detailed scope of work.

The discovered model of knowledge extraction from the public procurement documents is interesting from the point of view of universities' contribution to sustainable practices at the operational level of management. Universities are not focused on a limited number of investments but make many complex investments in different areas concerning energy efficiency. However, an exhaustive analysis of this domain was not possible and ones should be aware of the limitations in university practices. The conducted analysis allowed us to conclude that text mining, particularly LDA analysis, was an appropriate tool for verifying the operational activities of universities in the field of EE.

9. Conclusions

This study set out to analyze public procurement documents in the domain of university buildings, focusing specifically on their transformation towards more efficient energy consumption. Leveraging the Latent Dirichlet Allocation (LDA) model, we successfully extracted salient topics from these documents and examined the trends over a three-year period.

The results obtained and described above provided positive answers to the three research questions posed. The first question was investigated by the Latent Dirichlet Allocation method and showed current and potential topics of EE issues dealt with by universities. We also observed a diversity of topics between different regions of the country, which positively answered research question number two.

It must be added that this research was focused on the actual state of EE operational practices at the universities, the reasons for which remained beyond the scope of this study.

It can therefore be concluded that at the level of operational activities, the adopted strategic policies of universities in the field of sustainable development were reflected in the procurement documents, and thus, universities meet the expectations of stakeholders in this regard. The study period included three years of data. Despite this, in most cases, it was possible to observe an increasing trend in reported tenders in individual clusters, which positively answered the third research question.

In general, our findings revealed a clear shift in focus from modernization towards the adoption of energy-saving technologies and sustainability initiatives. This evolution is a promising indicator of universities' commitment to improving their energy efficiency and sustainability.

However, our analysis also exposed potential areas for improvement. Despite the importance of cooling and heating systems, enhanced with additional data-driven energy savings and incorporating digital electricity infrastructure and software, this topic was less prevalent in the procurement documents. EU member states are developing action plans for green procurement. In many cases, they include a general commitment to procure sustainable products and services. Associations are carrying out similar activities, and several recommendations and practices are also given in the academic literature. Following these guidelines will allow universities to achieve a higher degree of maturity in green procurement.

In this article, the problem of deconstructing the text of university public procurement documents concerning energy efficiency in construction projects was discussed. The proposed method was a two-step approach. In the first step, a literature review was performed to identify the structure of relevant research and to identify the domain for the

detailed analysis. In the second step, LDA analysis was performed to obtain the topics structure of the domain followed by trend analysis.

Despite these insights, our study is not without limitations. The procurement documents analyzed were represented only by titles, and our findings may not be generalizable to all document contents. Furthermore, while our LDA model was effective in extracting topics, it did not capture the relationships between different topics, which could be explored in future research.

In conclusion, our research contributes valuable insight into the procurement practices of universities regarding energy efficiency, adding to the existing body of knowledge in this field and providing a foundation for further research. As universities continue to strive for greater energy efficiency and sustainability, we hope that our findings will inform and enhance their procurement strategies.

Author Contributions: A.P.—performed conceptualization, methodology, supervision, writing, and project administration; T.F.—performed conceptualization, methodology, formal analysis, validation, writing, and visualization; B.G.—performed methodology, editing, data collection & cleaning; Z.G.—performed conceptualization, methodology, writing, formal analysis, visualization, and validation. All authors have read and agreed to the published version of the manuscript.

Funding: The publication was financed by resources from the Faculty of Management, University of Lodz, Poland.

Data Availability Statement: Data were obtained from the Public Procurement Bulletin, <https://bzp.uzp.gov.pl/> accessed on 1 February 2023.

Conflicts of Interest: The authors declare no conflict of interest. The authors declare that they have no known competing financial interest or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Sedlacek, S. The role of universities in fostering sustainable development at the regional level. *J. Clean. Prod.* **2013**, *48*, 74–84. [\[CrossRef\]](#)
2. Zilahy, G.; Huisingh, D.; Melanen, M.; Phillips, V.D.; Sheffy, J. Roles of academia in regional sustainability initiatives: Outreach for a more sustainable future. *J. Clean. Prod.* **2009**, *17*, 1053–1056. [\[CrossRef\]](#)
3. Walker, H.; Brammer, S. The relationship between sustainable procurement and e-procurement in the public sector. *Int. J. Prod. Econ.* **2012**, *140*, 256–268. [\[CrossRef\]](#)
4. Pacheco-Blanco, B.; Bastante-Ceca, M.J. Green public procurement as an initiative for sustainable consumption. An exploratory study of Spanish public universities. *J. Clean. Prod.* **2016**, *133*, 648–656. [\[CrossRef\]](#)
5. Giovanelli, L.; Rotondo, F.; Ezza, A. Business models for integration of sustainability in universities: An explorative analysis of Italian state universities. *J. Clean. Prod.* **2021**, *324*, 129227. [\[CrossRef\]](#)
6. Filho, W.L.; Skouloudis, A.; Brandli, L.L.; Salvia, A.L.; Avila, L.V.; Rayman-Bacchus, L. Sustainability and procurement practices in higher education institutions: Barriers and drivers. *J. Clean. Prod.* **2019**, *231*, 1267–1280. [\[CrossRef\]](#)
7. El Haddadi, T.; El Haddadi, O.; Mourabit, T.; El Allaoui, A.; Ben Ahmed, M. Automatic analysis of the sustainability of public procurement based on Text Mining: The case of the Moroccan ICT markets. *Clean. Responsible Consum.* **2021**, *3*, 100037. [\[CrossRef\]](#)
8. Tavana, M.; Shaabani, A.; Santos-Arteaga, F.J.; Vanani, I.R. A Review of Uncertain Decision-Making Methods in Energy Management Using Text Mining and Data Analytics. *Energies* **2020**, *13*, 3947. [\[CrossRef\]](#)
9. Abdelaziz, A.; Santos, V.; Dias, M.S. Machine Learning Techniques in the Energy Consumption of Buildings: A Systematic Literature Review Using Text Mining and Bibliometric Analysis. *Energies* **2021**, *14*, 7810. [\[CrossRef\]](#)
10. Ebneyamini, S.; Moghadam, M.R.S. Toward Developing a Framework for Conducting Case Study Research. *Int. J. Qual. Methods* **2018**, *17*, 1609406918817954. [\[CrossRef\]](#)
11. Voss, C.; Tsikriktsis, N.; Frohlich, M. Case research in operations management. *Int. J. Oper. Prod. Manag.* **2002**, *22*, 195–219. [\[CrossRef\]](#)
12. Young, S.; Nagpal, S.; Adams, C.A. Sustainable Procurement in Australian and UK Universities. *Public Manag. Rev.* **2016**, *18*, 993–1016. [\[CrossRef\]](#)
13. Swords, M. Finding Patterns in Procurements and Tenders Using a Graph Database. 2019. Available online: <http://kth.diva-portal.org/smash/get/diva2:1415996/FULLTEXT01.pdf> (accessed on 27 February 2023).
14. EC (European Commission). Recovery And Resilience Facility—State Aid. Guiding Template: Energy Efficiency in Buildings. Available online: https://ec.europa.eu/competition/state_aid/what_is_new/template_RFF_energy_efficiency_in_buildings.pdf (accessed on 27 February 2023).
15. ECEEE (European Council for an Energy Efficient Economy). Energy Performance of Buildings. Available online: <https://www.eceee.org/policy-areas/Buildings/> (accessed on 27 February 2023).

16. GBPC (Green Best Practice Community). Best Environmental Management Practice for the Public Administration Sector. Chapter 3.2.2. Improving the Energy Efficiency of Public Buildings.pdf. Available online: https://greenbestpractice.jrc.ec.europa.eu/sites/default/files/2019-12/Best_environmental_management_practice_report_public_administration_02.pdf (accessed on 27 February 2023).
17. EPEC (Expertise Centre). Guidance on Energy Efficiency in Public Buildings. 2012. Available online: <https://www.eib.org/en/publications/epec-guidance-on-energy-efficiency-in-public-buildings> (accessed on 27 February 2023).
18. Reddy, B.S. Barriers and drivers to energy efficiency—A new taxonomical approach. *Energy Convers. Manag.* **2013**, *74*, 403–416. [\[CrossRef\]](#)
19. Forsström, J.; Lahti, P.; Pursiheimo, E.; Rämä, M.; Shemeikka, J.; Sipilä, K.; Tuominen, P.; Wahlgren, I. Measuring Energy Efficiency Indicators and Potentials in Buildings, Communities and Energy Systems. 2012. Available online: <https://www.vttresearch.com/sites/default/files/pdf/tiedotteet/2011/T2581.pdf> (accessed on 27 February 2023).
20. Tarantini, M.; Loprieno, A.D.; Porta, P.L. A life cycle approach to Green Public Procurement of building materials and elements: A case study on windows. *Energy* **2011**, *36*, 2473–2482. [\[CrossRef\]](#)
21. Ralph, M.; Stubbs, W. Integrating environmental sustainability into universities. *High. Educ.* **2014**, *67*, 71–90. [\[CrossRef\]](#)
22. UZP Public Procurement Office (uzp.gov.pl)—Informacja o Polskim i Europejskim Rynku Zamówień Publicznych. Available online: <https://www.uzp.gov.pl/baza-wiedzy/analizy-systemowe/informacje-o-rynku-zamowien-publicznych-na-podstawie-ogloszen> (accessed on 15 January 2023).
23. Kozuch, A.C.; von Deimling, C.; Eßig, M. Implementing green public procurement: A replication study. *J. Clean. Prod.* **2022**, *377*, 134424. [\[CrossRef\]](#)
24. Bunesco, L.; Estermann, T. Greening in European Higher Education Institutions a Governance, Funding and Efficiency Perspective. European University Association December 2021. Available online: <https://eua.eu/downloads/publications/gaf%20greening-final.pdf> (accessed on 15 May 2023).
25. Bala, A.; Muñoz, P.; Rieradevall, J.; Ysern, P. Experiences with greening suppliers. The Universitat Autònoma de Barcelona. *J. Clean. Prod.* **2008**, *16*, 1610–1619. [\[CrossRef\]](#)
26. Thurston, M.; Eckelman, M.J. Assessing greenhouse gas emissions from university purchases. *Int. J. Sustain. High. Educ.* **2011**, *12*, 225–235. [\[CrossRef\]](#)
27. Heravi, G.; Aryanpour, D.; Rostami, M. Developing a green university framework using statistical techniques: Case study of the University of Tehran. *J. Build. Eng.* **2021**, *42*, 102798. [\[CrossRef\]](#)
28. Elser, N.; Michael, J. Sustainability in Higher Education Procurement: The Role of Employee Paper Purchasing Decisions. In *Educating the Sustainability Leaders of the Future. World Sustainability Series*; Springer: Cham, Switzerland, 2023; pp. 473–491. [\[CrossRef\]](#)
29. Kretschmer, S.; Dehm, S. Sustainability transitions in university food service—A living lab approach of locavore meal planning and procurement. *Sustainability* **2021**, *13*, 7305. [\[CrossRef\]](#)
30. Brindley, C.; Oxborrow, L. Aligning the sustainable supply chain to green marketing needs: A case study. *Ind. Mark. Manag.* **2014**, *43*, 45–55. [\[CrossRef\]](#)
31. Cong, D. Application of text mining in library book procurement. *MATEC Web Conf.* **2017**, *100*, 02044. [\[CrossRef\]](#)
32. Foong, L.M.; Ibrahim, B.; Mukhtar, M.I.; Sern, L.C. Go green practice: A study on TVET higher learning institution students. *J. Tech. Educ. Train.* **2021**, *13*, 97–101. Available online: <https://penerbit.uthm.edu.my/ojs/index.php/JTET/article/view/7882> (accessed on 23 February 2023).
33. Aragão, C.G.; Jabbour, C.J.C. Green training for sustainable procurement? Insights from the Brazilian public sector. *Ind. Commer. Train.* **2017**, *49*, 48–54. [\[CrossRef\]](#)
34. Naz, F.; Oláh, J.; Vasile, D.; Magda, R. Green Purchase Behavior of University Students in Hungary: An Empirical Study. *Sustainability* **2020**, *12*, 10077. [\[CrossRef\]](#)
35. Zhang, Z.; Jasaitis, T.; Freeman, R.; Alfrjani, R.; Funk, A. Mining Healthcare Procurement Data Using Text Mining and Natural Language Processing—Reflection from an Industrial Project. *arXiv* **2023**, arXiv:2301.03458. [\[CrossRef\]](#)
36. Modrušan, N.; Mršić, L.; Rabuzin, K. Intelligent Public Procurement Monitoring System Powered by Text Mining and Balanced Indicators. In *Data Management Technologies and Applications. DATA 2020. Communications in Computer and Information Science 2021*; Hammoudi, S., Quix, C., Bernardino, J., Eds.; Springer: Cham, Switzerland, 2021; Volume 1446. [\[CrossRef\]](#)
37. Cao, F.; Li, R.; Cao, X. Implementation of sustainable public procurement in China: An assessment using quantitative text analysis in large-scale tender documents. *Front. Environ. Sci.* **2022**, *10*, 947962. [\[CrossRef\]](#)
38. Choi, S.J.; Choi, S.W.; Kim, J.H.; Lee, E.-B. AI and Text-Mining Applications for Analyzing Contractor’s Risk in Invitation to Bid (ITB) and Contracts for Engineering Procurement and Construction (EPC) Projects. *Energies* **2021**, *14*, 4632. [\[CrossRef\]](#)
39. Shah, S.M.; Lütjen, M.; Freitag, M. Text Mining for Supply Chain Risk Management in the Apparel Industry. *Appl. Sci.* **2021**, *11*, 2323. [\[CrossRef\]](#)
40. Baricco, M.; Tartaglino, A.; Gambino, P.; Dansero, E.; Cottafava, D.; Cavaglià, G. University of Turin performance in UI GreenMetric Energy and Climate Change. *E3S Web Conf.* **2018**, *48*, 03003. [\[CrossRef\]](#)
41. Asdrubali, F.; Buratti, C.; Cotana, F.; Baldinelli, G.; Goretti, M.; Moretti, E.; Baldassarri, C.; Belloni, E.; Bianchi, F.; Rotili, A.; et al. Evaluation of Green Buildings’ Overall Performance through in Situ Monitoring and Simulations. *Energies* **2013**, *6*, 6525–6547. [\[CrossRef\]](#)

42. Ceschin, F.; Vezzoli, C. The role of public policy in stimulating radical environmental impact reduction in the automotive sector: The need to focus on product-service system innovation. *Int. J. Automot. Technol. Manag.* **2010**, *10*, 321–341. [[CrossRef](#)]
43. Ma, L.; Ashraf, R.U.; Haq, M.A.U.; Fan, X. Hurdles on the Way to Sustainable Development in the Education Sector of China. *Sustainability* **2023**, *15*, 217. [[CrossRef](#)]
44. Wiloso, E.I.; Nazir, N.; Hanafi, J.; Siregar, K.; Harsono, S.S.; Setiawan, A.A.R.; Muryanto; Romli, M.; Utama, N.A.; Shantiko, B.; et al. Life cycle assessment research and application in Indonesia. *Int. J. Life Cycle Assess.* **2019**, *24*, 386–396. [[CrossRef](#)]
45. Tolosa, A.E.; Irizar, M.Z.; Elena, M.T. An approach to consequences of market orientation in the Basque social Economy fostering public service [Aproximación a las consecuencias de la orientación al mercado en el servicio público de fomento de la Economía Social vasca]. In *CIRIEC-España Revista de Economía Pública, Social y Cooperativa*; CIRIEC: Valencia, Spain, 2017; Volume 89, pp. 55–79.
46. Fuentes-Bargues, J.L.; Ferrer-Gisbert, P.S.; González-Cruz, M.C. Analysis of Green public procurement of works by Spanish public universities. *Int. J. Environ. Res. Public Health* **2018**, *15*, 1888. [[CrossRef](#)]
47. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
48. Vallati, A.; Di Matteo, M.; Fiorini, C.V. Retrofit Proposals for Energy Efficiency and Thermal Comfort in Historic Public Buildings: The Case of the Engineering Faculty's Seat of Sapienza University. *Energies* **2023**, *16*, 151. [[CrossRef](#)]
49. Piacentini, R.V. A whole-building, integrated approach for designing a high-performance, net-zero-energy and net-zero-water building. In *Mediterranean Green Buildings and Renewable Energy: Selected Papers from the World Renewable Energy Network's Med Green Forum*; Springer: New York, NY, USA, 2017; pp. 931–940. [[CrossRef](#)]
50. Filho, W.L.; Vidal, D.G.; Dinis, M.A.P.; Lambrechts, W.; Vasconcelos, C.R.P.; Molthan-Hill, P.; Abubakar, I.R.; Dunk, R.M.; Salvia, A.L.; Sharifi, A. Low carbon futures: Assessing the status of decarbonisation efforts at universities within a 2050 perspective. *Energy Sustain. Soc.* **2023**, *13*, 9. [[CrossRef](#)]
51. Júnior, S.M.; Frandoloso, M.A.L.; Brião, V.B. The role of HEIs to achieve SDG7 goals from Netzero campuses: Case studies and possibilities in Brazil. *Int. J. Sustain. High. Educ.* **2023**, *24*, 462–480. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.