

Article

Exposing Emerging Trends in Smart Sustainable City Research Using Deep Autoencoders-Based Fuzzy C-Means

Anne Parlina ¹, Kalamullah Ramli ^{1,*} and Hendri Murfi ²

¹ Department of Electrical Engineering, Universitas Indonesia, Depok, Jawa Barat 16424, Indonesia; anne.parlina@ui.ac.id

² Department of Mathematics, Universitas Indonesia, Depok, Jawa Barat 16424, Indonesia; hendri@ui.ac.id

* Correspondence: kalamullah.ramli@ui.ac.id; Tel.: +62-811-8606-457

Abstract: The literature discussing the concepts, technologies, and ICT-based urban innovation approaches of smart cities has been growing, along with initiatives from cities all over the world that are competing to improve their services and become smart and sustainable. However, current studies that provide a comprehensive understanding and reveal smart and sustainable city research trends and characteristics are still lacking. Meanwhile, policymakers and practitioners alike need to pursue progressive development. In response to this shortcoming, this research offers content analysis studies based on topic modeling approaches to capture the evolution and characteristics of topics in the scientific literature on smart and sustainable city research. More importantly, a novel topic-detecting algorithm based on the deep learning and clustering techniques, namely deep autoencoders-based fuzzy C-means (DFCM), is introduced for analyzing the research topic trend. The topics generated by this proposed algorithm have relatively higher coherence values than those generated by previously used topic detection methods, namely non-negative matrix factorization (NMF), latent Dirichlet allocation (LDA), and eigenspace-based fuzzy C-means (EFCM). The 30 main topics that appeared in topic modeling with the DFCM algorithm were classified into six groups (technology, energy, environment, transportation, e-governance, and human capital and welfare) that characterize the six dimensions of smart, sustainable city research.

Keywords: smart city; sustainable city; smart sustainable city; text mining; topic detection; topic modeling



Citation: Parlina, A.; Ramli, K.; Murfi, H. Exposing Emerging Trends in Smart Sustainable City Research Using Deep Autoencoders-Based Fuzzy C-Means. *Sustainability* **2021**, *13*, 2876. <https://doi.org/10.3390/su13052876>

Academic Editor: Gwanggil Jeon

Received: 12 January 2021

Accepted: 3 March 2021

Published: 7 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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

The increasing number of scholarly books, articles, and other publications on smart cities shows the topic's emergence as a research domain. A city is a center of population, commerce, and culture with its own government and administration. Cities must develop strategic plans to define the path of innovation and must prioritize the essential aspects of building smart, sustainable cities. The current coronavirus disease 2019 (COVID-19) pandemic has made cities practice smart governance to overcome a crisis that has shaken their stability in various areas. Big cities should broaden their views, avoid short-term planning, and innovate more frequently to improve their services' quality and efficiency. Moreover, they must consider all social factors and actors and ensure that the public and businesses are involved [1]. The concept of a sustainable smart city as a means to improve citizens' quality of life is becoming necessary and relevant for both decision-makers and academics. Evolving research interests, the quick transition of research topics, and a broad coverage make the field of smart, sustainable cities attractive in trend studies.

Analyzing technological trends has always been challenging. A trend is the general tendency of a set of data to change [2], and trend analysis is the process of capturing the history and current state of data to predict the future. Scientific and technological information is disseminated through literature such as scientific papers, technical reports, and patents. Therefore, the analysis of the scientific literature to identify trends and topics that emerge in the fields of science and technology has become more important than ever.

The methods for identifying research topic trends can be broadly divided into qualitative methods (such as evaluation by experts, the Delphi method, and literature reviews) and quantitative methods (such as bibliometrics and machine-learning techniques). Qualitative techniques to extract significant results from large amounts of data tend to be costly and time consuming. They also invite the possibility of bias as the researchers' subjective opinions or values may be reflected in the research. Moreover, the objective evaluation of a field covering a wide range of research topics for several decades can be a daunting task even for top experts in the field [3].

In recent years, meanwhile, data science and machine learning have developed rapidly and have been applied in various fields, such as health [4–6], education [7,8], transportation [9,10], tourism [11,12], energy [13,14], economics [15], and government [16]. Because of the tremendous growth of digital collections in recent times, machine learning has become a useful method for analyzing unstructured data. Due to rapid advances in computer science and machine learning, various algorithms have been proposed to help researchers confidently deduce information and hidden patterns from text, particularly in scientific publications. Topic modeling is one machine-learning technique for topic detection. A topic model provides a way to identify the hidden thematic structure and topics in extensive text collections based on the terms contained in each record. It can be used to automate the process of discovering topics in digital documents for document clustering and classification, and can be applied in lexicography-oriented tasks to discover new meanings of words [17]. Beyond these applications, previous works have used topic modeling for various other tasks, such as the detection of unique and trending topics in newspapers or social media to reveal common social trends [18–20] and the analysis of business reports and other documents to modify business objectives and improve services [21,22]. It has also been used for malware detection [23], recommender systems [24,25], and summarization [26,27]. In a similar vein, topic modeling has been widely used to uncover research trends and innovation in various fields by analyzing scientific publications and patents [28–35].

The most widely used text-mining algorithms for detecting research topic trends are latent Dirichlet allocation (LDA) [28,36–39], clustering based methods [40–43], and non-negative matrix factorization (NMF) [44–46]. Although machine-learning techniques have been widely used for topic detection in scientific articles, conventional machine-learning methods have a limited ability to process raw data [47]. Meanwhile, deep learning is known for its automatic learning capabilities, specifically learning high-level abstraction through a hierarchical architecture [48]. Deep learning is a representation-learning method that allows machines to fill in raw data and automatically find the representation needed for detection or classification. Deep learning consists of multiple representation levels or layers obtained from a simple but nonlinear module arrangement that, starting from raw data input, transforms data into a representation at a higher and slightly more abstract level [47].

Deep learning is a powerful, state-of-the-art machine-learning technique that learns from multiple layers, which are representations or features of data, to make predictions [49]. Along with its application in various domains, deep learning is becoming the primary machine-learning technique for processing unstructured data, such as images or text [50]. So far, deep learning has not been employed to identify trends in research topics. Therefore, this study introduces a new algorithm based on deep learning and fuzzy clustering, termed deep autoencoders-based fuzzy C-means (DFCM) [51], to analyze the intellectual structure of scientific publications. The study aims to:

- examine the performance of a novel method based on deep learning and fuzzy C-means clustering in analyzing the research topic trend;
- determine and analyze the current major research themes and gaps, focusing on smart, sustainable cities; and
- synthesize an understanding of the smart, sustainable city concept and characteristics.

This study shows that the proposed new method, based on deep learning and clustering to identify research topic trends, performs well and even outperforms other standard

topic-modeling algorithms. Topic modeling with this algorithm captured topics that have emerged in smart and sustainable city research during the period 1990–July 2020.

The remainder of this paper is structured as follows. Section 2 discusses previous work and how this paper differs from it. Section 3 describes the concept and model of smart, sustainable cities as well as labels related to the smart, sustainable city. Section 4 explains the DFCM algorithm, and the materials and methods are described in Section 5. Section 6 summarizes the analysis results, and Section 7 presents the discussion. Section 8 notes the limitations of this work and possible directions for future research. Finally, the conclusions are given in Section 9.

2. Literature Review

Previous studies related to topic analysis in the field of the smart and sustainable city are summarized in Table 1 below.

Table 1. Previous works on topic analysis in smart and sustainable city research.

Title and Reference	Objectives	Methods
Exploring the Nature of the Smart Cities Research Landscape [52]	To synthesize the smart city concept's emerging understanding and determine the themes, types, and significant research gaps.	Combination of bibliometrics research mapping and visualization techniques with content analysis
Combining co-citation clustering and text-based analysis to reveal the main development paths of smart cities [53]	Mapping and analyzing the development paths, making them visible and understandable within the smart city research landscape.	Combination of two-hybrid bibliometrics techniques, namely co-citation clustering and text-based analysis
Trajectory of urban sustainability concepts: A 35-year bibliometric analysis [54]	To summarize the evolution of smart city concepts and analyze the composition of each city concept and the core issues addressed by each city type.	Co-word analysis
Sustainable-smart-resilient-low carbon-eco-knowledge cities; making sense of a multitude of concepts promoting sustainable urbanization [55]	To investigate the conceptualization and relationship of each city category.	Keywords and city categories co-occurrence analysis
A Comparison of a Smart City's Trends in Urban Planning before and after 2016 through Keyword Network Analysis [56]	To investigate keywords regarding the smart city concept and capture the smart city's urban planning trends.	Keyword network analysis
Visualization and Analysis of Mapping Knowledge Domain of Urban Vitality Research [57]	To assess the trends and map the knowledge domain in urban vitality research.	Systematic bibliometric analysis and keywords co-occurrence analysis
Bibliometric Analysis on Smart Cities Research [58]	To obtain a comprehensive overview of the smart cities research's characteristics in terms of the number of publications, most influential authors, institutions, sources and countries, and future research directions.	Bibliometric analysis
A study on the research trends for smart city using topic modeling [59]	To analyze scientific publications and identify research trends regarding the smart city.	Topic modeling with latent Dirichlet allocation (LDA) algorithm
Sustainable development of smart cities: a systematic review of the literature [60]	To analyze scientific papers that focus on smart city concepts and environmental sustainability to understand these two relationships.	Systematic literature review

It is both attractive and essential to explore, map, and understand how topics in academic publications evolve. As seen in Table 1, most topic analysis studies in scientific publications related to smart and sustainable cities have used the bibliometrics

method [52–58]. Other methods used in previous studies include topic modeling [59] and systematic literature reviews [60].

This research was conducted to study the concepts and categories related to smart, sustainable cities and to find the dominant topics in smart, sustainable city research by analyzing scientific publications using a novel deep-learning-based topic-detection algorithm. The method used in this study is a novel approach that has never before been applied to similar studies.

Topic detection is the process of analyzing words in textual datasets to find hidden topics. The two most popular standard methods for topic detection are LDA, which uses a probabilistic approach, and NMF, which uses a matrix decomposition approach [61–64]. In addition to these two approaches, clustering-based topic detection methods are also widely used [43,65–67]. The clustering method has an advantage compared to the previous two techniques. It can process data with a negative representation, making it possible to combine it with dimensional reduction or representation learning [51]. The combination of dimensional method reduction with clustering can solve the sparsity problem [43] and significantly reduce computation time [68]. Topic detection for forecasting technology, among others, is done by a combination method of dimensional reduction approaches with clustering, namely singular value decomposition (SVD) and principal component analysis (PCA) with k-means [43]. Moreover, a combination of SVD and k-means was used for topic detection on Twitter [68].

The clustering method will classify documents or textual data based on similarity in topics so that the centroid or cluster center can be interpreted as a topic [41,69,70]. Clustering algorithms can be classified into two types: hard clustering and soft clustering [71]. Hard clustering maps each data point to exactly one cluster. On the other hand, in the soft clustering method, each data point can be a member of many clusters with a certain probability [72]. K-means is one of the most popular hard clustering algorithms because of its simplicity and efficiency [73]. The k-means algorithm groups textual data or documents into k clusters, where each cluster represents one topic. In hard clustering, it is assumed that one document has only one topic because one document can only be a member of exactly one cluster [69]. The assumption that a document has only one topic is weak, considering that generally, a document consists of a mixture of several topics. This assumption is also different from topic detection with LDA and NMF, which assume that each textual datum is a mixture of several topics [51,74]. A document can be grouped into several clusters with a certain degree of membership using the soft clustering method, which can be interpreted as having several topics. Therefore, one of the most popular soft clustering algorithms, fuzzy c-means (FCM) [75], is considered as the basis for the topic detection method [40,69,74,76].

However, the FCM algorithm has a weakness: it can only work well on low-dimensional data [77]. Meanwhile, textual data are data with high dimensions. For data with high dimensions, randomized initialization will cause FCM to tend to group data into one cluster or converge to a single-center known as the center of gravity of the entire data text [77]. Therefore, in high-dimensional textual data, FCM will only produce one centroid or one topic. To overcome this problem, FCM needs to be combined with dimensional reduction techniques such as SVD. The fuzzy clustering process is conducted after the data is transformed into lower eigenspace using the SVD technique. This method of combining SVD with FCM is known as eigenspace-based fuzzy c-means (EFCM) [40]. Previous studies have shown that the EFCM performance regarding the topic recall and interpretation in topic detection in several datasets is between LDA and NMF [74].

Currently, deep learning is known as a powerful technique in processing extensive data. The study of combining deep learning with clustering has also been an active research area [78–82]. However, most of the clustering techniques used in this combined approach of deep learning and clustering are hard clustering and research that combines deep learning with soft clustering is still rare. Therefore, a novel topic detection algorithm based on deep learning and fuzzy clustering is introduced in this study. Moreover, this algorithm's performance in topic detection from the scientific publication dataset is examined and

compared with the performance of LDA, NMF, and EFCM. The adoption of deep autoencoder (DAE) as a representation learning method and dimensional reduction combined with fuzzy c-means is expected to improve the EFCM algorithm's performance in terms of topic coherence.

3. Smart and Sustainable City Concept and Models

3.1. Smart Sustainable City Concepts

Initiatives to realize urban infrastructure and services to create better environmental, social, and economic conditions and increase cities' attractiveness and competitiveness have been pursued in recent decades. The emergence of various city concepts and categories such as green cities, digital cities, intelligent cities, resilient cities, and others in the policy discourse reflects this development [55]. Among those city types and concepts, the most persistent and eminent are smart cities and sustainable cities [54].

In general, these concepts indicate the desired aspects and characteristics of a city. Table 2 shows the concepts and types of the cities and their definitions. The city categories are not interchangeable as they have significant conceptual differences, even though the categories are related. The sustainable city is one of the most frequently emerging city concepts, which can be considered an umbrella term covering the ecological, economic, and social dimensions of the pillars of sustainable development. Meanwhile, the smart city concept that emphasizes information and communication technology to provide high-quality services and community empowerment has also received increasing attention recently [55].

Table 2. The city labels and their definitions.

Label	Definition
cyber city	"The cybercity is a virtual 3D city, which links the 3D Geographical Information System (GIS) with other city information and stores in computers. It is a virtual city consisting of various electronical data stored in the computer or in the internet. Key technologies: internet, remote sensing, GIS technology, virtual reality. One of its important application areas is in the urban designing and planning." [83]
digital city	"A digital city has at least two plausible meanings: (a) a city that is being transformed or re-oriented through digital technology and (b) a digital representation or reflection of some aspects of an actual or imagined city." [84]
eco city	"Eco-city is considered as a rural–urban transition process, to develop an integral system and concern about social, economic and environmental aspects. Rural issues should be also taken into account during this process, so as to improve the harmony and fairness among rural and urban residents." [85]
green city	"Green cities are defined as those that are environmentally friendly. The greening of cities requires some, or preferably all, of the following: (1) controlling diseases and their health burden; (2) reducing chemical and physical hazards; (3) developing high quality urban environments for all; (4) minimizing transfers of environmental costs to areas outside the city; and (5) ensuring progress towards sustainable consumption." [86]
information city	"Information city is defined as a large Internet-based site offering a range of online services, including access to social environments, community services, municipal information, and e-commerce to its info habitants. Its boundaries are potentially unlimited, scaling as far as the available computation and storage capacity allow in order to manage huge volumes of content and millions of users simultaneously." [87]
intelligent city	"Intelligent cities are territories with high capacity for learning and innovation, which is built-in the creativity of their population, their institutions of knowledge creation, and their digital infrastructure for communication and knowledge management." [88]
knowledge city	"Knowledge cities can be referred to as cities in which both the private and the public sectors value knowledge, nurture knowledge, spend money on supporting knowledge dissemination, and discovery and harness knowledge to create products and services that add value and create wealth." [89]

Table 2. Cont.

Label	Definition
learning city	“A learning city is a complex social construct. Its development means enabling learning at all city levels (inhabitants and their families, organisations and city administration through networks). The collective learning of individuals and their participation in partnership networks are especially important.” [90]
resilient city	“A resilient city is a sustainable network of physical systems and human communities.” [91]
smart city	“Smart cities are all urban settlements that make a conscious effort to capitalize on the new information and communications technology (ICT) landscape in a strategic way, aiming for (i) environmental sustainability, (ii) urban system functionality, (iii) quality of life for all, (iv) knowledge-based development and (v) community-driven development. In this sense, their basic components are the urban setting, ICTs, people and communities and a strategic approach towards one or more of the previous aims. Without one of these components, a city cannot be regarded as a fully-fledged ‘smart’ city.” [92]
sustainable city	“Sustainable cities work towards an environmentally, socially, and economically healthy and resilient habitat for existing populations, without compromising the ability of future generations to experience the same.” [93]
ubiquitous city	“Ubiquitous city defined as the advanced city that is able to play the role of providing increased convenience of life and greater quality of life, systematic city management, the innovation of various city functions through systematic management, and the generation of various markets, by offering a ubiquitous service to the city space using the advanced information and communication infrastructure.” [94]
virtual city	“Virtual city is an application of virtual reality in geo-science, and it should have three main characteristics: city simulation, interaction and network sharing of city information.” [95]
wired city	“Wired cities refer literally to the laying down of cable and connectivity (not in itself necessarily smart).” [96]

The combination of smart and sustainable concepts for a city has recently become a new concept and is increasingly used in the scientific literature [97–103]. A city can be sustainable without being smart. Likewise, smart technology can be used without paying attention to sustainable development [97]. However, a city cannot be truly smart without being sustainable [104]. The concept of a smart sustainable city emerged when information and communication technologies applied for urban sustainability [97]. Höjer and Wangel [105] define a smart sustainable city based on the Brundtland [106] concept, as follows:

“A Smart Sustainable City is a city that meets the needs of its present inhabitants; without compromising the ability for other people or future generations to meet their needs, and thus, does not exceed local or planetary environmental limitations; and where this is supported by ICT.”

3.2. The Dimensions of a Smart and Sustainable City

Smart sustainable cities have lately evolved to represent multidisciplinary interests [107]; the concepts of smart cities and sustainable cities are themselves the result of evolution from other types of cities such as knowledge-based cities, intelligent cities, digital cities, and other more complex models [108]. The coverage of smart and sustainable indicators in several city models is represented as the dimensions of smart and sustainable cities. Experts use various terms to describe a smart and sustainable city’s coverage and aspects, including dimensions, components, areas, elements, and pillars. Table 3 shows the smart city dimensions, while Table 4 presents the sustainable city’s dimensions from various sources.

Table 3. Smart city dimensions.

No.	Source	Dimensions/Elements/Pillar/Aspects	Number of Dimensions
1.	Citiasia Center for Smart Nation (CCSN) [109]	Smart governance; smart branding; smart economy; smart living; smart society; smart environment.	6
2.	Cities in Motion (2020) [1]	Economy; human capital; social cohesion; environment; governance; urban planning; international projection; technology; mobility; transportation.	9
3.	Angelidou 2017 [92]	Technology, ICTs, and the Internet; human and social capital development; entrepreneurship promotion; global collaboration and networking; privacy and security; locally adapted strategies; participatory approach; top-down coordination; explicit and workable strategic framework; interdisciplinary planning.	10
4.	European smart cities 4.0 (2015) [83]	Smart economy; smart mobility; smart environment; smart people; smart living; smart governance.	6
5.	Hassan and Aggarwal 2020 [110]	People and communities; leadership and organization; advanced machinery; policy context; control; economy; infrastructure; natural surroundings.	8
6.	Dameri 2017 [111]	Technologies; governance; urban area; environment; people.	5
7.	Adio-Moses and Oladiran 2016 [112]	Smart governance; smart energy; smart building; smart mobility; smart infrastructure; smart technology; smart healthcare; smart citizen.	8

Table 4. Sustainable city dimensions.

No.	Source	Dimensions/Elements/Pillar/Aspects	Number of Dimensions
1.	Shmelev and Shmeleva 2009 [113]	Sustainable energy; quality of life, health; sustainable transport; psychology of interaction with the environment; environmental conscience, behavior; democratic participation; sustainable transport; material flows, waste management; green space, biodiversity; landscape architecture, eco-design, modernization; preservation of the natural and cultural heritage.	10
2.	Jenks and Jones 2010 [114]	Land use and built form, environmental energy conservation, recycling and re-use, communication, and transport.	4
3.	Allen 2009 [115]	Economic sustainability, social sustainability, ecological sustainability, sustainability of the built environment, political sustainability.	5

4. Topic Model Algorithms

4.1. Clustering and Deep Learning for Topic Detect

A topic is the primary idea, subject, or theme of a discussion, conversation, or text data. It can also be defined as a set of words that tend to convey similar or related context. A topic may have different granularity, such as in a sentence, paragraph, article, or all digital library collections [116]. A topic model is a type of algorithm that aims to identify the hidden thematic structure in extensive text collections. Topic models have received much attention from researchers since they were first proposed. Despite this model initially focusing on text mining and information retrieval, it has been successfully applied for analyzing various type of data sources, including videos and images [117]. Many topic modeling algorithms have been developed and implemented to obtain topic models from document collections, including non-negative matrix factorization (NMF) [44–46], latent Dirichlet allocation (LDA) [28,36–39], and clustering-based methods [40–43].

This paper introduces a new method for topic detection, which is a combination of clustering and deep learning methods. The clustering method classifies text data that have similar and related topics into clusters. The center or centroid of each group represents the topic. The clustering method used in this study is fuzzy c-means (FCM) [75]. In general, FCM works well for low-dimensional data but fails to address high-dimensional textual data. FCM has poor performance in high-dimensional data because all the centroids resulting from the clustering process run to the dataset's center point. As a consequence, the clustering results converge to the same cluster (center of gravity) [77]. One of the ways to overcome this problem is to transform text data into a lower dimension. In previous studies, dimension reduction was carried out using the truncated singular value decomposition (SVD) method [40,69,74,76]. Fuzzy c-means processing on reduced data dimensions, i.e., eigenspace using the truncated SVD method, is known as eigenspace-based FCM (EFCM) [40].

Deep learning is a machine learning technique suitable for processing extensive data with high accuracy [118], which allows computational models consisting of several layers of processing to study data representations with various abstraction levels [47]. One of the deep learning architectures representing data in low dimensions is deep autoencoders (DAE) [80]. In this study, DAE is used as a data reduction method to overcome weaknesses and develop FCM performance in topic detection. This novel method is called deep autoencoders-based fuzzy c-means (DFCM).

4.1.1. Fuzzy C-Means

Fuzzy c-means is a clustering algorithm that aims to partition data points into clusters, where each member in the clusters has a minimal distance to the center of the cluster (centroids). Clustering with FCM allows each data point to have a certain degree of membership in multiple clusters so that FCM is classified in the soft clustering algorithm. In this algorithm, one data point updates many centroids based on their membership in the cluster [74].

Suppose $A = \{a_1, a_2, \dots, a_n\}$ is a set of data, where n is the number of data points, and $\mu = \{\mu_1, \mu_2, \dots, \mu_c\}$ is a set of centroids or centers of the clusters, where c is the number of clusters and $c \geq 2$. Each data point's cumulative weighted distance to the centroids can be expressed as an objective function (J):

$$J(U, Q, A, c, z) = \sum_i^c \sum_k^n (u_{ik})^w \|a_k - \mu_i\|^2, \quad (1)$$

where z is a fuzzification constant ($z > 1$), $U = [u_{ik}]$ is a membership matrix, and u_{ik} is the membership value of data point a_k in cluster i representing the inverse of the distance between the k th data point and the i th centroid. The membership values have some constraints:

$$0 \leq u_{ik} \leq 1 \quad (2)$$

$$\sum_{i=1}^c u_{ik} = 1, k = 1, 2, \dots, n \quad (3)$$

$$0 \leq \sum_{k=1}^n u_{ik} \leq n, i = 1, 2, \dots, c. \quad (4)$$

Based on the Lagrange multiplier theory, the Lagrange function of J in Equation (1) can be formed as follows:

$$L = \sum_i^c \sum_k^n (u_{ik})^z \|a_k - \mu_i\|^2 + \lambda \left(\sum_{i=1}^c (u_{ik}) - 1 \right), \quad (5)$$

where λ is a Lagrange multiplier. The optimum condition of the objective function can be achieved if u_{ik} and μ_i are optimal. The optimal u_{ik} and μ_i values are obtained by differentiating the Lagrange J function for each parameter iteratively, where the differentiation equation is set to 0. The value of u_{ik} in each iteration is:

$$u_{ik} = \left(\frac{\lambda}{z \|a_k - \mu_i\|^2} \right)^{\frac{1}{z-1}}, \quad (6)$$

while the value of μ_i is:

$$\mu_i = \frac{\sum_{k=1}^n ((u_{ik})^z a_k)}{\sum_{k=1}^n (u_{ik})^z}. \quad (7)$$

These two optimization steps are iterated until the termination criteria are reached—for example, the maximum number of iterations, and insignificant changes to the objective function J , membership u_{ik} , or centroid.

The fuzzy c-means algorithm is described in more detail in Algorithm 1.

Algorithm 1 Fuzzy c-means

Input: A, c, z , max iteration number (T), error threshold (ϵ)

Output: u_{ik}, μ_i

Set $t = 0$

1. Initialize μ_i
 2. Update $t = t + 1$
 3. Compute u_{ik}
 4. Compute μ_i
 5. If a stopping criterion, i.e., $t > T$ or $\|J_t - J_{t-1}\| < \epsilon$, is fulfilled then stop; otherwise, go
 6. back to step 3.
-

4.1.2. Deep Autoencoders (DAE)

A deep autoencoder is a form of deep neural network that functions to reduce the dimensions of the dataset [119]. DAE uses a deep neural network architecture (DNN) to build the model and has a processing layer like DNN, namely an input layer, hidden layer, and output layer. In DAE architecture, the number of neurons in the output layer is equal to the neurons' number in the input layer. This means that DAE uses the input data feature as the target. Moreover, DAE models with at least one hidden layer with fewer neurons than the input layer can be used to predict the input data itself and study the data more concisely by storing as much information as possible, even though the number of neurons in the hidden layer is reduced. Thus, data features can be represented on a lower dimension.

DAE architecture consists of three components, namely the encoder, code, and decoder. The encoder consists of layers that are used to transform the input data to lower dimensions. Furthermore, the code section is the part that has a hidden bottleneck layer—that is, a layer that has fewer neurons than the input layer and which represents the data in the

lowest dimension. The decoder section consists of layers whose structure is symmetrical to the encoder section and is used to transform data back from the lower dimensions to the original dimensions (the same dimensions as the input data). The decoder's architecture is symmetrical with the encoder's part because the complexity (the number of layers and neurons) required for the decoder to return the data to its original dimension is equal to the complexity required for the encoder to represent data at low dimensions.

One method of constructing DAE is to use greedy-layer-wise pretraining to build the layers, where each layer is constructed by denoising autoencoder [120]. A denoising autoencoder is a modified autoencoder to reconstruct deliberately tampered inputs to get a more robust and stable representation of data at low dimensions.

Suppose $A = \{a_1, a_2, \dots, a_N\}$ is a set of textual data, with $a_i \in \mathcal{R}^{D_x}, \forall i = 1, 2, \dots, N$. The denoising autoencoder, which consists of two layers, can be notated as follows:

$$\tilde{a}_i \sim \text{dropout}_1(a_i) \quad (8)$$

$$h_i = f_1(\tilde{a}_i, w_1) \quad (9)$$

$$\tilde{h}_i \sim \text{dropout}_2(h_i) \quad (10)$$

$$y_i = f_2(\tilde{h}_i, w_2), \quad (11)$$

where w_1 and w_2 are the weights, f_1 and f_2 are the activation functions, and $\text{dropout}_1()$ and $\text{dropout}_2()$ are methods for randomly ignoring multiple numbers of neuron output during training. A loss function is a quantity that training attempts to minimize [121]. Training or learning in deep autoencoders can be interpreted as an optimization process to find a combination of model parameters that minimize the loss function $\mathcal{L}(a, w)$ for a specific set of training data samples, a_i , and an appropriate target, y_i . There are many types of loss function, but the two most commonly used are mean squared error (MSE) and cross entropy [122]. In this study, MSE is used as a loss function to solve real value quantity prediction problems. MSE can be calculated as the average of the squared differences between the actual and expected output as follows [123].

$$\mathcal{L}(a, w) = \frac{1}{N} \sum_{n=1}^N (y_i - a_i)^2 \quad (12)$$

In the next process, h_i is used as input data for the next layer. The appropriate autoencoder weight is obtained by training each denoising encoder. The next stage is to retrain the autoencoder to minimize reconstruction loss for all layers.

The DAE learning process is further explained in Algorithm 2.

Algorithm 2 Deep Autoencoders Model Learning

Input: Data A, the number of neurons of code p, the number of the autoencoder's layers m

Output: $\text{encoder}(w), \text{decoder}(w)$

1. $\text{autoencoder}(p)$ initialization.
 2. $h_n^{(0)} = x_n$
 3. FOR $i = 1$ to m
 4. Fitting the denoising encoder's i -th layer:
 5. $\text{deAutoencoder}(w^{(i)}), w^{(i)} = \min_w \mathcal{L}(h_n^{(i-1)}, w), \forall n$
 6. $h_n^{(i)} = \text{deEncoder}(h_n^{(i-1)}), \forall n$
 7. The autoencoder weight corresponds to the weight obtained from denoising autoencoder initialized: $\text{autoencoder}((w^{(i)}), \forall i$
 8. Retraining the autoencoder $\text{autoencoder}(w), w = \min_w \mathcal{L}(x_n, w), \forall n$
-

4.1.3. Deep Autoencoders-Based Fuzzy C-Means

Deep autoencoders-based fuzzy c-means (DFCM) [51] is the application of FCM by first doing dimensional reduction using DAE. Suppose we have a dataset $A = \{a_1, a_2, \dots, a_N\}$ where $a_i \in R^{D_x}, \forall i = 1, 2, \dots, N$, and p is the dimensions of the code (new data representation). The process begins with DAE learning, starting from the encoder to the decoder. The transformation of the data into a new data representation with a lower dimension is performed using the encoder. The transformation process is notated as follows:

$$\tilde{A} = \text{encoder}(A, p). \quad (13)$$

Suppose the dataset that has been reduced in dimension is as follows: $\{a_1, a_2, \dots, a_N\}$ where $a_i \in R^{D_z}, \forall i = 1, 2, \dots, N$, and c is the number of clusters. Clustering with FCM can now be performed on reduced dimensional datasets. The clustering process with FCM produces centroids, which are the centers of the clusters. The process to get the centroid from each cluster is as follows:

$$\tilde{\mu}_i = \text{FCM}(\tilde{A}, c, z, T, \varepsilon). \quad (14)$$

Let $\{\tilde{\mu}_1, \tilde{\mu}_2, \dots, \tilde{\mu}_c\}$ be the centroids, where $\tilde{\mu}_i \in R^{D_z}, \forall i = 1, 2, \dots, c$, and c is the number of clusters. These centroids are the representation of the topics. However, the centroids that represent the topic in this low dimension have no meaning. These centroids will become meaningful if transformed back into their original dimensions. Therefore, it is necessary to transform centroids from the low dimension to the initial dimension by utilizing the layers of the decoder portion of the DAE, as follows:

$$\mu_i = \max(0, \text{decoder}(\tilde{\mu}_i)). \quad (15)$$

Centroids in the initial dimensions obtained from the transformation process are $\{\mu_1, \mu_2, \dots, \mu_c\}$, where $\mu_i \in R^{D_x}, \forall i = 1, 2, \dots, c$, and $\max()$ is a function that delivers a maximum value between 0 and each element of $\text{decoder}(\tilde{\mu}_i)$. Centroids in the original dimension can now be understood as topics from a dataset. The whole process is described in more detail in Figure 1 and Algorithm 3.

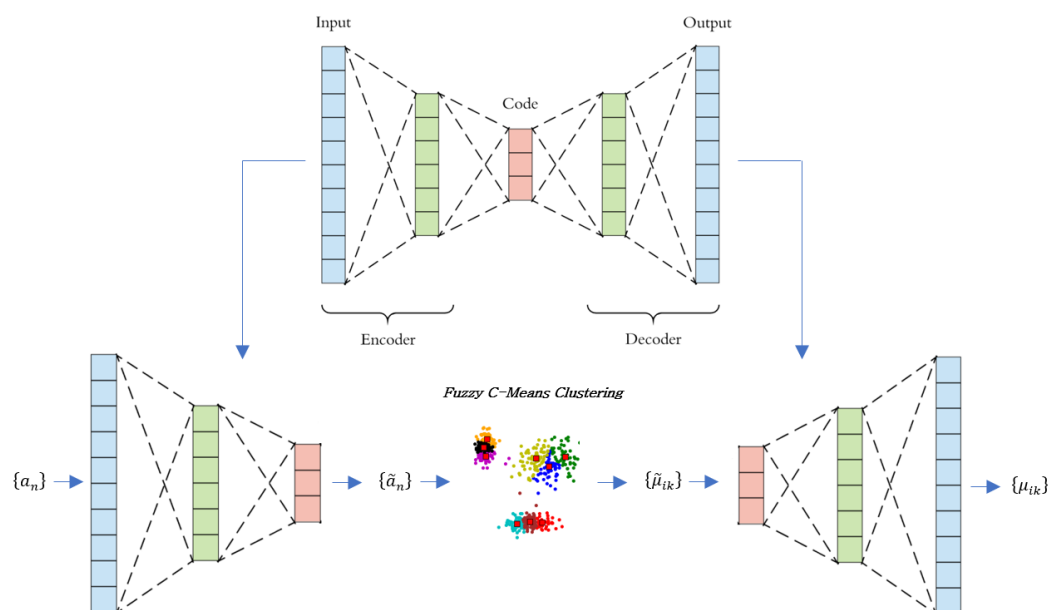


Figure 1. The illustration of deep autoencoders-based fuzzy c-means [51].

Algorithm 3 Deep Autoencoders-Based Fuzzy C-means [51]**Input:**Data $A \{a_1, a_2, \dots, a_N\}$,DAE parameters: the number of neurons in code layer p ,FCM parameters: number of cluster (c), degree of fuzzy (z), max iteration number (T), error threshold (ϵ), initial objective function value J_0 **Output:** μ_i

1. Autoencoder construction: $encoder, decoder = DAE(A, p)$
2. Data A transformation: $\tilde{A} = encoder(A)$
3. Clustering data \tilde{A} with FCM: $\tilde{\mu}_i = FCM(\tilde{A}, c, z, T, \epsilon)$, $i = 1, 2, \dots, c$
4. Calculate the centroids in original dimensions: $\mu_i = \max(0, decoder(\tilde{\mu}_i))$, $i = 1, 2, \dots, c$

5. Materials and Methods

This study was conducted in three stages, which are depicted in Figure 2 and described as follows.

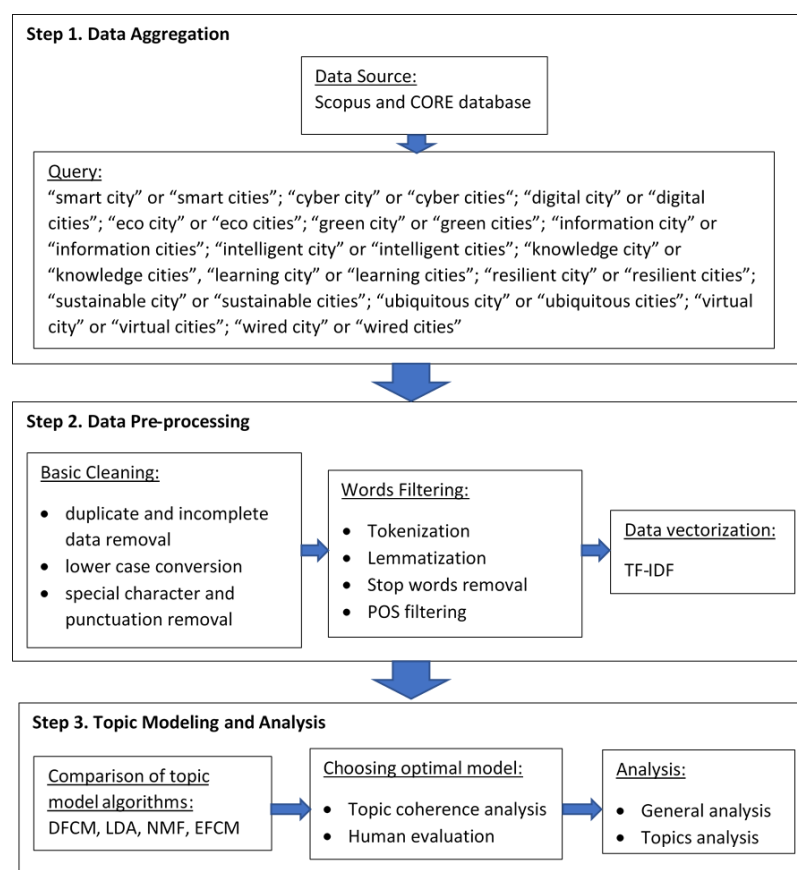


Figure 2. Research stages and framework.

5.1. Data Aggregation

City categories related to smart and sustainable cities were used as search queries (Figure 1) to ensure the comprehensiveness of the data aggregation. All the source documents were retrieved with the keyword search performed in SCOPUS and CORE databases in July 2020. Scopus claims to be the largest database of abstracts and citations from peer-reviewed literature, covering various types of scientific publications in the form of scientific journals, conference proceedings, and books. It provides a comprehensive overview of global research results in science, technology, the arts and humanities, social sciences, and medicine [124]. Meanwhile, CORE (Connecting REpositories) claims to be the world's most extensive full-text collection of scientific papers for machine processing [125]. It

provides a global aggregator service that collects the metadata and full text of open-access scientific papers from journals and repositories worldwide [126].

The data were collected using 14 queries consisting of various city labels, as shown in Figure 1. Aggregating the data from the Scopus database resulted in 28,824 raw data points, and 26,042 raw data points were retrieved from the CORE database. Only the title and abstract of each scientific publication were used in this study.

5.2. Data Preprocessing

Data preprocessing was carried out to prepare the raw text data for further processing [127]. It aimed to establish the corpus and lexicon of all the retrieved documents. A corpus is a body of text used for statistical analysis in natural language processing while a lexicon is a vocabulary list (words and/or phrases) [128]. The first step in this stage included the basic process of cleaning the raw data as well as standard preprocessing in natural processing languages. Incomplete and duplicate data were eliminated, and all the text data were converted to lowercase. Furthermore, special characters and punctuation marks were removed. The next step was to convert the cleaned text into a well-defined collection of linguistic terms. The entire text was split into single words through the tokenization process, and each word was reduced to the root word by considering the vocabulary with the lemmatization process. Lemmatization can improve the topic-modeling results but also slows the computation [129]. Stop words, that is, common words that have no meaning or are less meaningful than other keywords, were removed. Removing stop words can sharpen the focus on essential words [127], reduce feature size, and improve accuracy [130,131]. The types of words considered in this study were limited to nouns, verbs, adverbs, and adjectives through part of speech (POS) filtering. Filtering the text data based on the part of speech can enhance the quality of topic-modeling results [132].

After going through the above steps, the text data needed to be converted into a numeric vector to be used as input for machine-learning algorithms. This process is known as vectorization or feature extraction. Term frequency–inverse document frequency (TF-IDF) was applied as a feature extraction method. The importance of the words or features in a text document is weighted based on their frequency across multiple documents.

The preprocessing steps for the text were executed using Python programming language and Jupyter Notebook. After going through the preprocessing stage, the cleaned data comprised 25,451 documents.

5.3. Topic Modeling and Analysis

5.3.1. Comparison of Topic Detection Algorithms

The standard quantitative method to measure a topic's interpretability is to calculate the coherence scores of words constructing the topic. In this study, the coherence values of topics generated by the DFCM algorithm were compared with coherence values of topics generated by other topic modeling algorithms, namely NMF, LDA, and EFCM. NMF and LDA are standard topic detection algorithms that have been widely used to identify research topic trends, while EFCM is the predecessor algorithm to DFCM. The calculation of the coherence value in this study uses the Word2Vec Topic Coherence (TC-W2V) [63]. Given a topic t , which consists of N words $\{d_1, d_2, \dots, d_N\}$, the topic t 's TC-W2V value is formulated as follows:

$$TC - W2V = \frac{1}{\binom{N}{2}} \sum_{j=2}^N \sum_{i=1}^{j-1} \text{similarity}(wv_j, wv_i), \quad (16)$$

where wv_j and wv_i are the vectors of the words d_j and d_i , formed by the *word2vec* model. The closer the word wv_j to the word wv_i , the greater the similarity value. This means that the greater the TC-W2V value, the better the interpretation of the topic.

5.3.2. Results Analysis

In addition to comparing the coherence value of the topics generated from the algorithms mentioned above, various analyses were also carried out on data related to smart, sustainable, and other types of cities to determine research trends and the characteristics of smart and sustainable cities. Among these were the analysis of the total number of publications per year, the number of publications per city label per year, the percentage of publications in each city category, and prominent research topics.

6. Results Analysis

6.1. General Analysis

Figure 3 illustrates the growth in the number of publications related to smart and sustainable cities and all categories of other cities since 1990–2020. The data in 2020 (collected in July 2020) numbered less than half of the data in 2019. The number of publications shows a rapid increase since 2010 and continues to increase exponentially until 2019. The total number of publications is 53,998. The percentage of the number of publications for each category is shown in Figure 4. The smart city category took first place with a total number of publications of 27,934 (51.73%), followed by the sustainable city category, with a total number of publications of 10,464 (19.38%). The green city has a total number of publications of 3499 (6.48%). The fourth and fifth positions are occupied by resilient cities and digital cities, with total publications of 2581 (4.78%) and 2018 (3.74%), respectively.

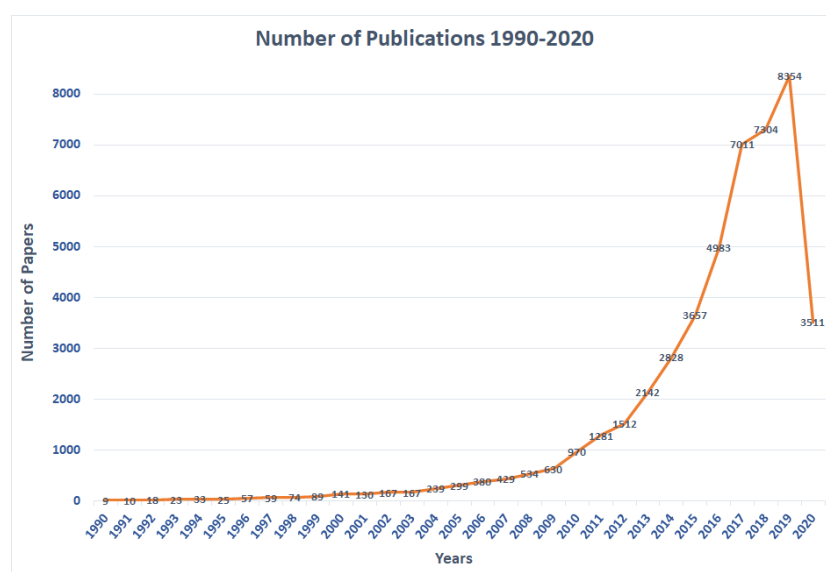


Figure 3. Growth in the number of publications on all city categories from 1990 to 2020.

Figure A1 presents the growth in the number of publications for each city category from 1990 to July 2020. As explained above, the smart city and sustainable city categories have total publications that far outnumber other city categories, so that the graph of the number of publications of the two categories is placed on a different *y*-axis. Figure 5 depicts explicitly the number of publications per year from the smart and sustainable city category. This graph shows that the concept and scientific publications related to sustainable cities emerged earlier than the idea of smart cities. Publications regarding smart cities began to appear in 1997. However, their numbers continued to increase, so that in 2013 scientific publications related to smart cities had the same number as sustainable cities and then shot up far beyond it.

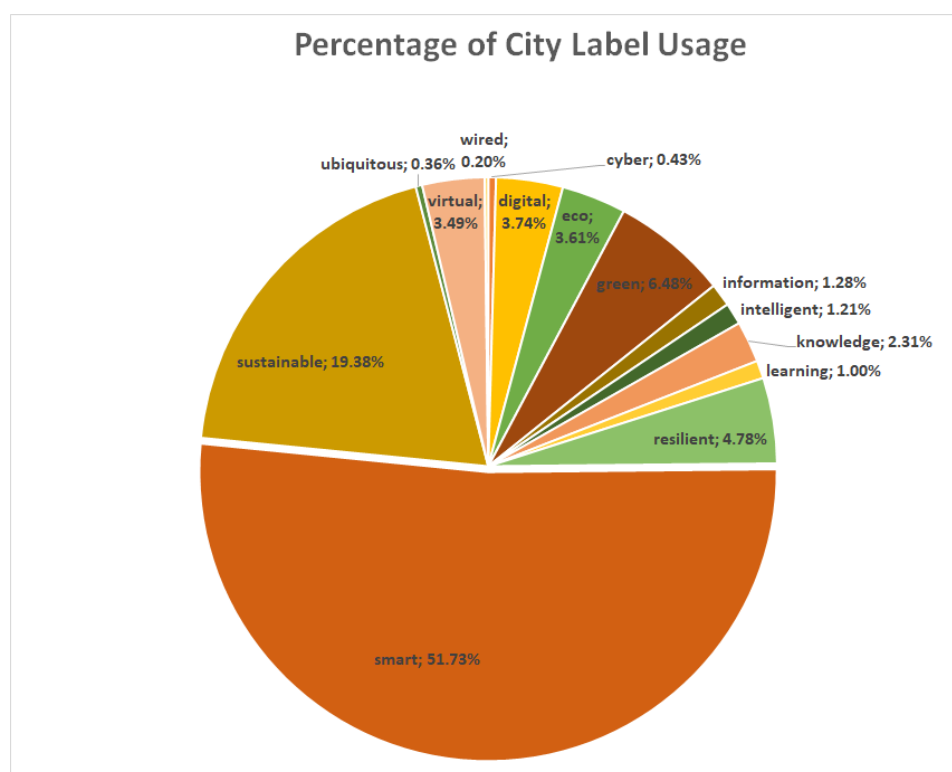
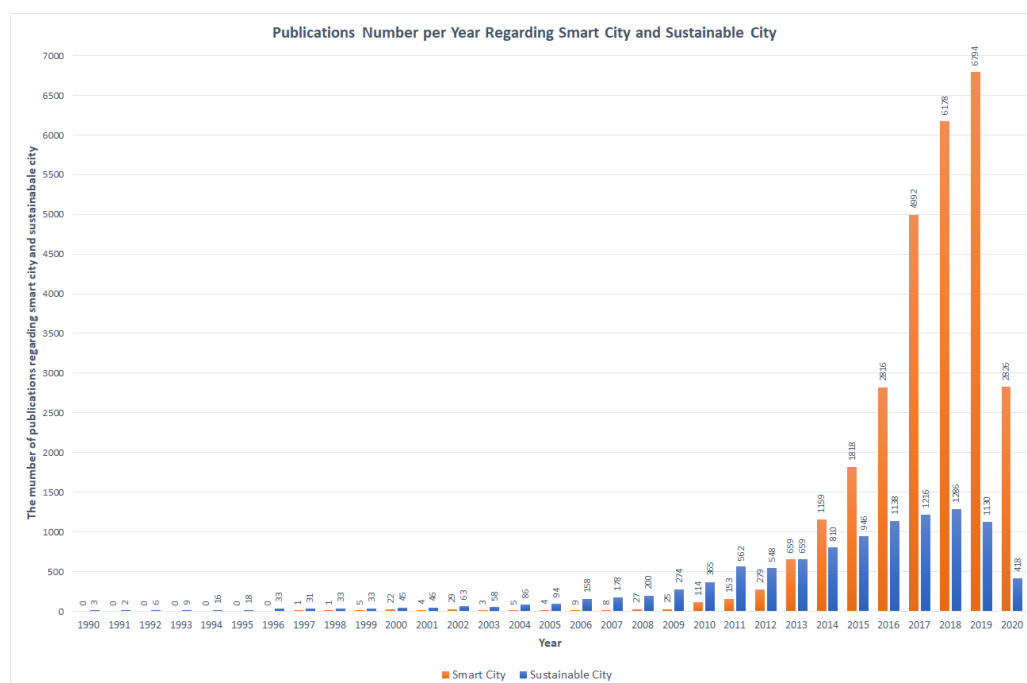


Figure 4. Percentage of city labels' (cyber, digital, eco, green, information, intelligent, knowledge, learning, resilient, smart, sustainable, ubiquitous, virtual, and wired) usage in publications.



$c \in \{10, 20, \dots, 90, 100\}$. The coherence value of topics generated by EFCM tends to be stable but occupies the lowest position compared to the coherence value of topics generated by the other three algorithms. The topics generated by LDA for the number of topics $c < 70$ had a higher coherence value than EFCM, while the number of topics $c > 70$ was almost the same as EFCM.

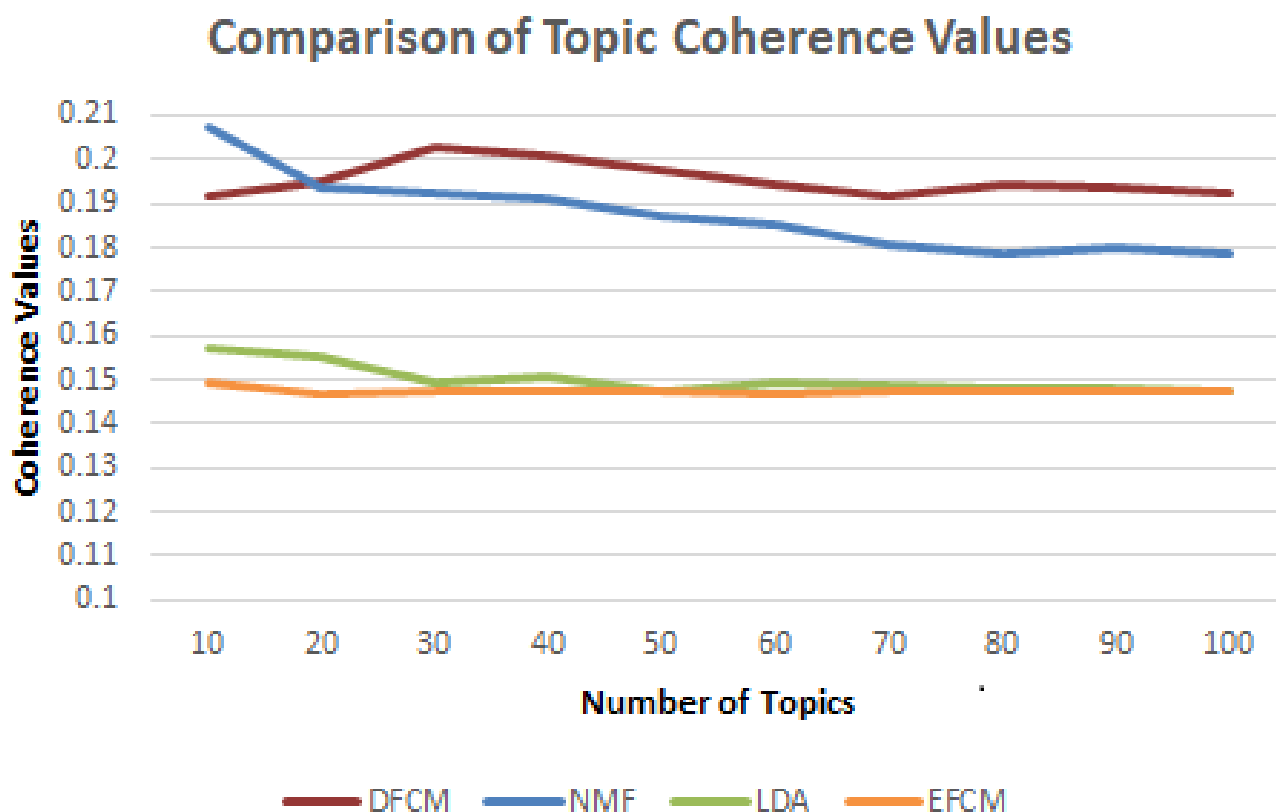


Figure 6. Coherence value of each algorithm comparison.

The coherence values of the DFCM and NMF algorithms' topics have higher coherence values compared to LDA and EFCM. The coherence value of topics generated by NMF shows the highest value (0.2072) when the number of topics is 10, and still has a higher coherence value than DFCM on a number of topics $c < 20$, but tends to continue to decline with a higher number of topics so that the coherence value is at or below the coherence score of the topics generated by DFCM. On the other hand, the coherence value of the topics generated by DFCM tends to be low when the number of topics is < 20 but higher than the coherence value of the topics generated by the NMF when the number of topics is > 20 .

6.3. Topic Analysis

This section examines smart and sustainable city research domain topics by analyzing scientific publications from Scopus and CORE databases. The selection of topics was based on the highest coherence value in the experiments. After running the DFCM model with 30 topics, the top 20 words in each cluster were taken for analysis (see Appendix B). The 30 topics were then further categorized into six dimensions: technology, energy, environment, transportation, e-governance, and human capital and welfare, as shown in Table 5.

Table 5. Categorization of research topics in scientific publications related to smart and sustainable cities.

No.	Dimension	Research Topics
1.	Technology	Wireless sensor network; machine learning; cloud, edge, and fog computing; semantic web; big data analytics; privacy and security; computer vision; Internet of things (IoT); virtual reality.
2.	Energy	Power plant; smart building; renewable energy.
3.	Environment	Disaster mitigation; greenhouse effect; climate change; pollution; waste management; green space; water resource management.
4.	Transportation	Traffic management; vehicle ad hoc network; smart parking; public transport.
5.	E-governance	E-government; politic and economy; citizen participation.
6.	Human capital and welfare	Education; housing; tourism, food, and agriculture.

7. Discussion

7.1. Smart and Sustainable City Milestones

The idea of sustainability appeared long ago in 1972 in “A Blueprint for Survival” [133]. Awareness of the need to conserve ecosystems and resources to create sustainable societies was the main driver of this idea [134]. Several subsequent events led to the emergence of the smart and sustainable city concept. First was the Kyoto Protocol [135] in 1997, which was signed by 192 parties. Its main goal was to limit CO₂ emissions and protect the environment around the world. The Kyoto Protocol was one of the main drivers of interest in smart cities and played a role in encouraging states and cities to design and implement environmental policies [136]. Meanwhile, the use of the word “smart” for city labels is thought to have originated from the use of the term “smart” for electronic devices that can combine mobile telecommunications, Internet connection, and data processing to provide real-time digital information services to users. This was marked by the launch of the iPhone in 2007, which was known as the first smartphone [137].

The smart city concept became popular after IBM introduced the smarter planet concept in 2008 [104]. The smarter planet vision promoted by IBM is driven by instrumentation, connectedness, and intelligence. These three concepts are expected to make industry, infrastructure, society, and the economy more productive, responsive, and efficient. The smarter planet system covers various domains such as a smarter power grid, smarter water treatment system, smarter food supply system, smarter health system, and smarter traffic system [138]. IBM was starting a new business in this sector and focused on providing solutions in communications, healthcare, energy, transportation, etcetera. Many ICT companies worldwide have followed IBM’s idea of supplying new smart projects for urban problems [136].

In 2010, the European Union formulated a strategy to transform the EU into a smart, sustainable, and inclusive economy that generates high productivity, employment, and social cohesion, which is contained in the Europe 2020 Strategy. This strategy has three priorities: smart growth in innovation, education, and digital society; sustainable growth in climate, energy, mobility, and competitiveness; and inclusive growth emphasizing employment, skills, and fighting poverty. Moreover, this strategy focuses on five targets in different areas to be achieved by the end of 2020 [139].

The signing of the 2016 Paris Agreement, in which 174 countries vowed to reduce climate change, was also a milestone in terms of developing a smart and sustainable city. This was the first international agreement with a legal effect that limited global warming

to below 2 °C compared to pre-industrial levels, and sought to pursue a limit of 1.5 °C. It demonstrated international cooperation and was a starting point for tackling climate change, reducing emissions, and implementing sustainable development [140]. The targets adopted by the Paris Agreement will guide future actions on mitigation and adaptation, emission reduction, and technological development, mainly the green information and communication technologies essential to tackling climate change and maintaining sustainable development [141].

However, it cannot be denied that, apart from the social and political factors mentioned above, technology plays a vital role in the development of a smart and sustainable city. The development of the internet and ICT infrastructure since 2000 and the increasingly widespread use of smartphones initiated electronic services in various fields such as health, education, transportation, energy, government, environmental management, etc. The growth of data and information is evident and continuous, so big data technology, machine learning, cloud computing, and the Internet of things (IoT) have emerged, which have increasingly driven the development of research in the field of smart cities. The development of these ICT technologies is the main factor causing the number of publications regarding smart cities to start increasing exponentially since 2010 (Figure 5).

7.2. Research Topics in Scientific Publications Related to Smart and Sustainable Cities

The topic analysis results from scientific publications related to smart and sustainable cities with deep learning-based modeling algorithms show the main research topics developing in this field. Technology is one of the main drivers of progress in the development of smart cities [53]. The smart cities concept tends to emphasize the transformative role of digital technology and ICT to increase economic growth and social benefits and support sustainable goals [112]. Figure 7 shows an illustration of the 17 sustainable goals formulated by the United Nations (UN). The United Nations Sustainable Development Goals (SDGs) consist of a series of targets and goals that make smart city development the main focus [112]. In the first group of topics, namely technology, nine subtopics often arise in smart and sustainable city research, namely IoT, wireless sensor networks, big data analytics, machine learning, cloud, edge and fog computing, privacy and security, computer vision, virtual reality, and the semantic web.



Figure 7. United Nations (UN) sustainable development goals [142].

The energy sector is one of the goals in SDGs (affordable and clean energy). This study's topic analysis results indicate three subtopics in the energy topic category, namely power plants, renewable energy, and smart buildings. While the environmental category is not explicitly stated as one of the SDGs, its subtopics are listed in several SDGs, namely water resource management (clean water and management, life underwater); climate change (climate action); and greenhouse effect, pollution, waste management, green space, and disaster mitigation (life on the land).

Transportation is one of the critical sectors of sustainable urban development. There is a close relationship between energy use in transportation and a city's physical characteristics, such as density, size, and the amount of open space [143]. Transportation includes a complex economic, social, and technical system, making it challenging to handle comprehensively. The transport sector consumes nonrenewable resources: energy, ecosystems, atmospheric carbon-loading capacity, and individual time. However, treatment to reduce depletion of one of these things can lead to worsening of the others [144]. The topic analysis in this study found four subtopics in the transportation topic category: traffic management, vehicle ad hoc network, smart parking, and public transport.

E-governance uses ICT tools at various levels of interaction in internal government operations and between government and public sector to simplify and enhance democratic governance [145]. Through e-governance, government services, information exchange, and business transactions between governments, businesses, and citizens will be made available in a more convenient, efficient, and transparent manner [146]. This study found three subtopics related to e-governance: e-government, politics and economy, and citizen participation.

The last topic category is human capital and welfare, including the subtopics of education, housing, tourism, food, and agriculture. One of the goals of sustainable development is health and well-being. However, the health sector has not yet emerged as a topic in this experiment. Based on this study's results, it can be concluded that research in the health sector related to smart and sustainable cities has not been widely conducted.

7.3. COVID-19 and Smart Sustainable City Research

The COVID-19 pandemic, a global health emergency, has caused suffering and economic chaos and changed people's lives worldwide. Thousands have died, and millions more have been infected and fallen ill during the outbreak. More work is needed to tackle various health crises by providing more efficient health system funding, increasing access to doctors and health facilities, and improving hygiene and sanitation, thereby saving more lives. The pandemic is a major milestone for health emergency preparedness and investment in public health services [147].

Cities are at the forefront of dealing with the pandemic and its lasting effects [148]. The financial crisis and the reduction in oil prices caused the world to adopt a new order in months [149], and the COVID-19 health crisis has become a crisis in urban access, transportation, infrastructure, and public services. COVID-19 endangers public health and affects many fields related to sustainability, including the economy, transportation, lifestyle, sociocultural matters, and community structure. If not handled properly, the crises that cities are likely to endure could jeopardize and undermine broader sustainable urban development efforts [148].

The current conditions caused by COVID-19 have changed the global lifestyle through lockdown policies, restrictions on industrial operations, and travel bans. Schools and universities are moving to an online mode, and most other sectors are adopting work-from-home policies so that people are increasingly forced to stay indoors. As a result, there has been a reduction in emissions and energy consumption [150]. This could be seen as a positive impact of COVID-19, but, of course, no one expects an environmental improvement in this way.

Most governments around the world are focused on fighting COVID-19 and dealing with the effects of the pandemic. Given that cities are centers of economic and population

activity, many researchers are struggling to understand, explain, and provide recommendations on various issues related to COVID-19, such as the mechanism of the virus's spread; the impact of the pandemic on the environment, society, and the economy; and necessary recovery and adaptation plans and policies. The impact of COVID-19 on cities is related to four major themes: environmental quality; transportation and urban design; city governance and management; and socioeconomic concerns [151]. Moreover, the pandemic also affects the design, improvement, and implementation of technology [152]. Researchers from various disciplines are increasingly working to cope with pandemics, contributing to several areas of multidisciplinary research, such as the information system community's role in realizing digital sustainability capabilities and achieving the SDGs [153].

8. Limitation and Further Study

This research has several limitations and challenges to its validity that can be addressed in future studies. First, the data aggregation for this research was conducted in August 2020, so no data were collected after July 2020. Meanwhile, the COVID-19 outbreak began in December 2019 after emerging in Wuhan, China. More data are needed to capture how the pandemic has affected academic research, particularly research on smart, sustainable cities. Second, the detection and identification of major research themes in a particular domain require the analysis and knowledge of a domain expert while, in this study, the determination of topic names for each word group was based solely on our interpretation. Future work needs to involve experts from the field of smart, sustainable cities to sharpen the analysis of the results. Third, although the deep-learning technique provides researchers an effective way to make inferences about a large volume of complex data, it is often criticized for being a black box model with unknown and untraceable predictions [154]. Fourth, the analysis of the topics in this study used only abstract data and titles from scientific publications, but a comparison of coherence values and human topic ranking from topic modeling using the LDA algorithm for abstract data and full-text data shows that topics generated from full-text data have better coherence and ranking values [155]. Therefore, the analysis of full-text data with the DFCM algorithm and the comparison of those results with abstract data will be of interest in future studies.

9. Conclusions

This study introduces a novel deep-learning-based modeling algorithm, namely deep auto-encoders-based fuzzy c-means (DFCM). Comparing the implementation of the DFCM algorithm with other standard topic-modeling methods, such as NMF, LDA, and EFCM, shows that the topics (words clusters) generated by the DFCM algorithm have a higher coherence value for a number of topics greater than 20. However, for a number of topics less than 20, the topic coherence value generated by the DFCM algorithm is lower than the topic coherence value generated by the NMF. The DFCM algorithm analyzes the main topics in scientific publication data on smart, sustainable cities and related city categories.

The concept of a sustainable smart city is receiving increasing attention worldwide in response to the challenges of sustainability and urbanization as well as the increasingly sophisticated use of information and communication technology. Topic modeling with this algorithm produced 30 topics, which were then further categorized into six topic groups: technology, energy, environment, transportation, e-governance, and human capital and welfare. The six categories of topics reveal the dimensions of a smart, sustainable city that often appear in research. Technology, particularly ICT, is the primary driver of a smart city and plays a role in achieving sustainability goals. The topic analysis results have covered the sustainable aspects listed in the SDGs except for the health sector. The conditions of the COVID-19 pandemic, however, suggest that research in this area will increase rapidly. The impact of the COVID-19 outbreak on research on smart, sustainable cities has also been discussed.

Author Contributions: A.P. conducted the experiments and prepared the original draft. H.M. provided the methods, while K.R. supervised the research and writing process. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partly funded by the Universitas Indonesia through the PUTI Q2 scheme under contract number NKB-4278/UN2.RST/HKP.05.00/2020.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

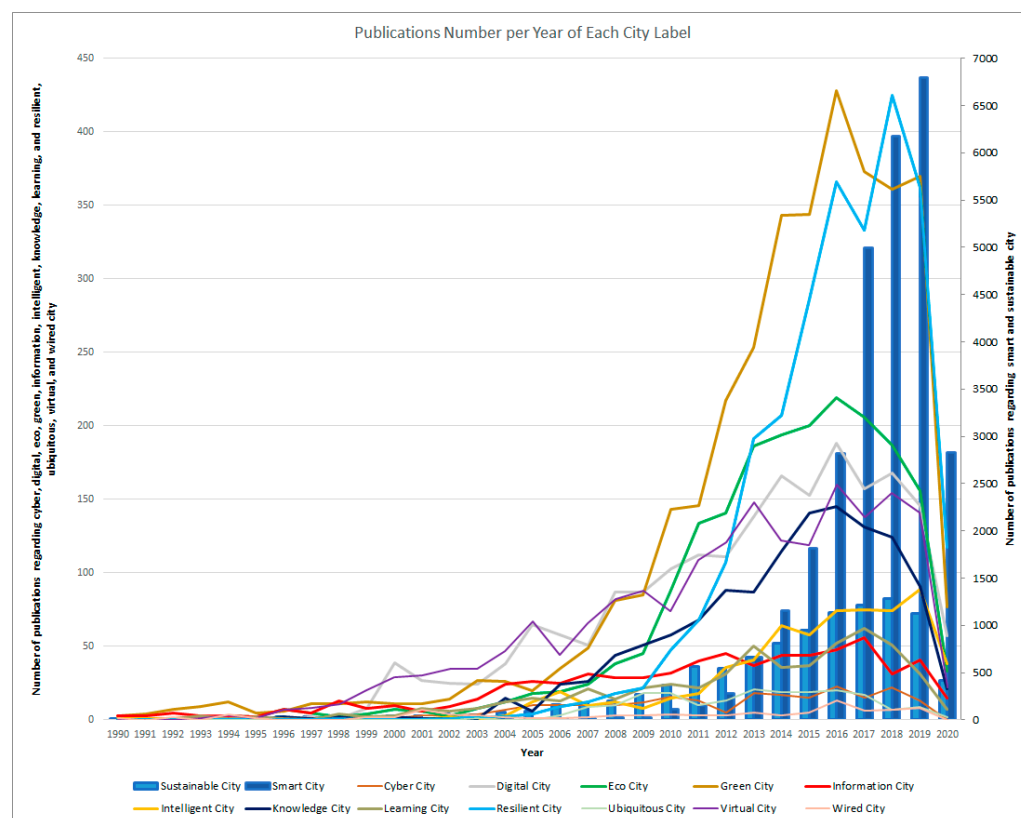


Figure A1. The growth of scientific publications related to each city category, 1990–2020.

Appendix B

Table A1. The topic labels and 20 top words.

No.	Topic Label	20 Top Words
1	Power plant	energy, power, renewable, grid, electricity, load, storage, demand, generation, electric, source, control, solar, battery, distribution, consumption, optimization, voltage, supply, plant
2	E-government	government, citizen, governance, innovation, technology, digital, model, initiative, public, policy, strategy, open, implementation, management, local, case, ICT, report, municipal, infrastructure
3	Disaster mitigation	disaster, risk, resilience, flood, resilient, natural, city, vulnerability, hazard, earthquake, urban, management, area, recovery, assessment, reduction, decision, spatial, emergency, mitigation
4	Traffic management	traffic, road, vehicle, time, real, intelligent, control, congestion, accident, smart, transportation, city, optimization, safety, route, speed, driver, street, noise, management

Table A1. Cont.

No.	Topic Label	20 Top Words
5	Wireless sensor network	node, wireless, network, sensor, energy, thing, internet, consumption, IoT, software, simulation, WSN, performance, device, application, lifetime, radio, battery, protocol, transmission
6	Machine learning	detection, recognition, deep, neural, image, machine, classification, feature, learning, accuracy, convolutional, method, object, time, processing, network, real, algorithm, segmentation, model
7	Smart building	thermal, building, temperature, heat, wind, material, energy, comfort, solar, heating, surface, air, indoor, effect, performance, roof, ventilation, condition, environmental, residential
8	Education	student, education, school, learning, university, teaching, teacher, project, knowledge, course, high, program, college, engineering, institution, campus, cultural, science, technology, creative
9	Cloud, edge, and fog computing	cloud, application, device, IoT, computing, architecture, edge, internet, resource, platform, fog, software, processing, infrastructure, real, time, environment, heterogeneous, communication, data
10	Greenhouse effect	climate, emission, energy, heat, building, cool, gas, change, retrofit, greenhouse, heating, mitigation, consumption, green, reduce, reduction, effect, build, city, impact
11	Semantic web	semantic, web, information, ontology, datum, application, user, platform, open, service, smart, technology, software, domain, management, integration, source, heterogeneous, context, knowledge
12	Tourism	tourism, business, tourist, destination, smart, industry, management, model, technology, information, innovation, sector, cultural, value, education, company, process, activity, economy, heritage
13	Big data analytics	datum, data, big, smart, application, management, analytic, city, processing, source, platform, analysis, technology, software, real, collection, time, large, mining, storage
14	Privacy and security	privacy, datum, security, application, device, blockchain, smart, authentication, user, IoT, access, data, secure, encryption, control, issue, personal, trust, protection, architecture
15	Pollution	air, pollution, monitoring, monitor, quality, noise, control, light, indoor, lighting, time, real, sensor, sense, cost, environmental, health, environment, pollutant, level
16	Waste management	waste, solid, management, collection, disposal, municipal, recycling, garbage, urban, material, treatment, landfill, sustainable, city, smart, transportation, area, bin, container, household
17	Smart parking	parking, smart, car, time, driver, lot, slot, application, transportation, sensor, intelligent, city, space, solution, spot, real, congestion, occupancy, user, management
18	Green space	green, urban, ecological, landscape, city, planning, environmental, area, plan, project, environment, sustainable, natural, ecosystem, nature, cultural, park, approach, construction, strategy
19	VANET (Vehicle ad hoc network)	traffic, transportation, intelligent, vehicular, vehicle, communication, congestion, road, control, information, management, time, technology, mobile, application, VANET, real, light, safety, network

Table A1. Cont.

No.	Topic Label	20 Top Words
20	Water resource management	water, urban, area, resource, management, environmental, ecological, population, natural, planning, wastewater, park, quality, supply, plant, environment, nature, soil, treatment, tree
21	Food and agriculture	food, housing, sustainable, environmental, urban, agriculture, social, economic, local, project, production, sustainability, income, area, settlement, country, market, land, affordable, population
22	Climate change	resilience, climate, city, change, resilient, urban, adaptation, plan, planning, strategy, project, approach, local, process, vulnerability, concept, build, challenge, action, risk
23	Politic and economy	policy, economic, economy, political, country, region, global, regional, government, growth, state, social, market, national, informal, settlement, report, rural, crisis, income
24	Computer vision	image, object, model, scene, real, computer, feature, texture, time, camera, visual, segmentation, dimensional, recognition, motion, reconstruction, render, map, simulation, vision
25	Renewable energy	energy, solar, thermal, heat, wind, heating, temperature, cool, model, electricity, consumption, comfort, power, demand, load, building, renewable, plant, optimization, storage
26	Internet of things	internet, thing, IoT, smart, monitoring, monitor, software, intelligent, datum, platform, model, real, architecture, management, time, open, application, control, environment, home
27	Citizen participation	citizen, city, participation, innovation, smart, public, open, technology, project, service, information, government, process, decision, business, initiative, policy, technological, lab, local
28	Virtual reality	virtual, user, reality, game, environment, interaction, design, experience, interface, augment, technology, world, computer, interactive, digital, human, application, real, agent, physical
29	Public transport	transport, mobility, transportation, public, city, bus, policy, car, freight, travel, passenger, mode, project, logistic, transit, infrastructure, bicycle, area, accessibility, sustainable
30	Housing	housing, sustainable, project, city, planning, plan, social, affordable, settlement, local, income, area, policy, residential, market, house, resident, economic, estate, neighborhood

References

- Berrone, P.; Ricart, J.E. *IESE Cities in Motion Index 2020*; IESE Business School University of Navarra: Barcelona, Spain, 2020; ISBN 2019104679.
- Robb, A. Identifying Trends, Patterns & Relationships in Scientific Data—Video & Lesson Transcript Study. Available online: <https://study.com/academy/lesson/identifying-trends-patterns-relationships-in-scientific-data.html> (accessed on 23 December 2019).
- Kang, H.J.; Kim, C.; Kang, K. Analysis of the Trends in Biochemical Research Using Latent Dirichlet Allocation (LDA). *Processing* **2019**, *7*, 379. [\[CrossRef\]](#)
- Ijaz, M.F.; Attique, M.; Son, Y. Data-Driven Cervical Cancer Prediction Model with Outlier Detection and Over-Sampling Methods. *Sensors* **2020**, *20*, 2809. [\[CrossRef\]](#) [\[PubMed\]](#)
- Ali, F.; El-Sappagh, S.; Islam, S.R.; Kwak, D.; Ali, A.; Imran, M.; Kwak, K.-S. A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. *Inf. Fusion* **2020**, *63*, 208–222. [\[CrossRef\]](#)
- Li, F.-Q.; Wang, S.-L.; Liu, G.-S. A Bayesian Possibilistic C-Means clustering approach for cervical cancer screening. *Inf. Sci.* **2019**, *501*, 495–510. [\[CrossRef\]](#)
- Villegas-Ch, W.; Román-Cañizares, M.; Palacios-Pacheco, X. Improvement of an Online Education Model with the Integration of Machine Learning and Data Analysis in an LMS. *Appl. Sci.* **2020**, *10*, 5371. [\[CrossRef\]](#)
- Chung, J.Y.; Lee, S. Dropout early warning systems for high school students using machine learning. *Child. Youth Serv. Rev.* **2019**, *96*, 346–353. [\[CrossRef\]](#)

9. Truong, D. Using causal machine learning for predicting the risk of flight delays in air transportation. *J. Air Transp. Manag.* **2021**, *91*, 101993. [\[CrossRef\]](#)
10. Boukerche, A.; Wang, J. Machine Learning-based traffic prediction models for Intelligent Transportation Systems. *Comput. Networks* **2020**, *181*, 107530. [\[CrossRef\]](#)
11. Sánchez-Medina, A.J.; Eleazar, C. Using machine learning and big data for efficient forecasting of hotel booking cancellations. *Int. J. Hosp. Manag.* **2020**, *89*, 102546. [\[CrossRef\]](#)
12. Zhang, K.; Chen, Y.; Li, C. Discovering the tourists' behaviors and perceptions in a tourism destination by analyzing photos' visual content with a computer deep learning model: The case of Beijing. *Tour. Manag.* **2019**, *75*, 595–608. [\[CrossRef\]](#)
13. Sharmila, P.; Baskaran, J.; Nayanatara, C.; Maheswari, R. A hybrid technique of machine learning and data analytics for optimized distribution of renewable energy resources targeting smart energy management. *Procedia Comput. Sci.* **2019**, *165*, 278–284. [\[CrossRef\]](#)
14. Shapi, M.K.M.; Ramli, N.A.; Awalin, L.J. Energy consumption prediction by using machine learning for smart building: Case study in Malaysia. *Dev. Built Environ.* **2021**, *5*, 100037. [\[CrossRef\]](#)
15. Tao, X.; Yang, H. Analysis of real-time changes in financial exchange rates based on machine learning and complex embedded systems. *Microprocess. Microsyst.* **2020**, *103493*, 103493. [\[CrossRef\]](#)
16. Lima, M.S.M.; Delen, D. Predicting and explaining corruption across countries: A machine learning approach. *Gov. Inf. Q.* **2020**, *37*, 101407. [\[CrossRef\]](#)
17. Lau, J.H. Improving the Utility of Topic MODELS: An Uncut Gem Does Not Sparkle. Ph.D. Thesis, The University of Melbourne, Melbourne, VIC, Australia, 2013.
18. Capela, F.D.O.; Ramirez-Marquez, J.E. Detecting urban identity perception via newspaper topic modeling. *Cities* **2019**, *93*, 72–83. [\[CrossRef\]](#)
19. Ali, F.; Kwak, D.; Khan, P.; El-Sappagh, S.; Ali, A.; Ullah, S.; Kim, K.H.; Kwak, K.-S. Transportation sentiment analysis using word embedding and ontology-based topic modeling. *Knowl.-Based Syst.* **2019**, *174*, 27–42. [\[CrossRef\]](#)
20. Pinto, S.; Albanese, F.; Dorso, C.O.; Balenzuela, P. Quantifying time-dependent Media Agenda and public opinion by topic modeling. *Phys. A Stat. Mech. Appl.* **2019**, *524*, 614–624. [\[CrossRef\]](#)
21. Robinson, S. Temporal topic modeling applied to aviation safety reports: A subject matter expert review. *Saf. Sci.* **2019**, *116*, 275–286. [\[CrossRef\]](#)
22. Bastani, K.; Namavari, H.; Shaffer, J. Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints. *Expert Syst. Appl.* **2019**, *127*, 256–271. [\[CrossRef\]](#)
23. Lou, S.; Cheng, S.; Huang, J.; Jiang, F. TFDroid: Android Malware Detection by Topics and Sensitive Data Flows Using Machine Learning Techniques. In Proceedings of the 2019 IEEE 2nd International Conference on Information and Computer Technologies (ICICT), Kahului, HI, USA, 14–17 March 2019; pp. 30–36.
24. Gao, Z.; Fan, Y.; Wu, C.; Tan, W.; Zhang, J.; Ni, Y.; Bai, B.; Chen, S. SeCo-LDA: Mining Service Co-Occurrence Topics for Composition Recommendation. *IEEE Trans. Serv. Comput.* **2019**, *12*, 446–459. [\[CrossRef\]](#)
25. Liu, D.-R.; Chou, Y.-C.; Jian, C.-T. Online Recommendation Based on Collaborative Topic Modeling and Item Diversity. In Proceedings of the 2018 7th International Congress on Advanced Applied Informatics (IIAI-AAI), Yonago, Japan, 8–13 July 2018; pp. 7–12.
26. Li, H.; Zhu, J.; Ma, C.; Zhang, J.; Zong, C. Read, Watch, Listen, and Summarize: Multi-Modal Summarization for Asynchronous Text, Image, Audio and Video. *IEEE Trans. Knowl. Data Eng.* **2018**, *31*, 996–1009. [\[CrossRef\]](#)
27. Nagwani, N.K. Summarizing large text collection using topic modeling and clustering based on MapReduce framework. *J. Big Data* **2015**, *2*, 6. [\[CrossRef\]](#)
28. Liu, G.; Nzige, J.H.; Li, K. Trending topics and themes in offshore construction(OSC) research: The application of topic modelling. *Constr. Innov.* **2019**, *19*, 343–366. [\[CrossRef\]](#)
29. Reisenbichler, M.; Reutterer, T. Topic modeling in marketing: Recent advances and research opportunities. *J. Bus. Econ.* **2019**, *89*, 327–356. [\[CrossRef\]](#)
30. Jiang, H.; Qiang, M.; Lin, P. Finding academic concerns of the Three Gorges Project based on a topic modeling approach. *Ecol. Indic.* **2016**, *60*, 693–701. [\[CrossRef\]](#)
31. Jiang, H.; Qiang, M.; Lin, P. A topic modeling based bibliometric exploration of hydropower research. *Renew. Sustain. Energy Rev.* **2016**, *57*, 226–237. [\[CrossRef\]](#)
32. Momeni, A.; Rost, K. Identification and monitoring of possible disruptive technologies by patent-development paths and topic modeling. *Technol. Forecast. Soc. Chang.* **2016**, *104*, 16–29. [\[CrossRef\]](#)
33. Figuerola, C.G.; Marco, F.J.G.; Pinto, M. Mapping the evolution of library and information science (1978–2014) using topic modeling on LISA. *Science* **2017**, *112*, 1507–1535. [\[CrossRef\]](#)
34. Moro, S.; Cortez, P.; Rita, P. Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. *Expert Syst. Appl.* **2015**, *42*, 1314–1324. [\[CrossRef\]](#)
35. Choi, H.; Oh, S.; Choi, S.; Yoon, J. Innovation Topic Analysis of Technology: The Case of Augmented Reality Patents. *IEEE Access* **2018**, *6*, 16119–16137. [\[CrossRef\]](#)
36. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
37. Amado, A.; Cortez, P.; Rita, P.; Moro, S. Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis. *Eur. Res. Manag. Bus. Econ.* **2018**, *24*, 1–7. [\[CrossRef\]](#)

38. Westgate, M.J.; Barton, P.S.; Pierson, J.C.; Lindenmayer, D.B. Text analysis tools for identification of emerging topics and research gaps in conservation science. *Conserv. Biol.* **2015**, *29*, 1606–1614. [\[CrossRef\]](#)
39. Sun, L.; Yin, Y. Discovering themes and trends in transportation research using topic modeling. *Transp. Res. Part C Emerg. Technol.* **2017**, *77*, 49–66. [\[CrossRef\]](#)
40. Muliawati, T.; Murfi, H. Eigenspace-based fuzzy c-means for sensing trending topics in Twitter. In Proceedings of the 2nd International Symposium on Current Progress in Mathematics and Sciences 2016, Jawa Barat, Indonesia, 1–2 November 2016.
41. Petkos, G.; Papadopoulos, S.; Kompatsiaris, Y. Two-level message clustering for topic detection in Twitter. *CEUR Workshop Proc.* **2014**, *1150*, 49–56.
42. Tu, H.; Ding, J. An Efficient Clustering Algorithm for Microblogging Hot Topic Detection. In Proceedings of the 2012 International Conference on Computer Science and Service System, Washington, DC, USA, 11–13 August 2012; pp. 738–741.
43. Jun, S.; Park, S.-S.; Jang, D.-S. Document clustering method using dimension reduction and support vector clustering to overcome sparseness. *Expert Syst. Appl.* **2014**, *41*, 3204–3212. [\[CrossRef\]](#)
44. Abuhay, T.M.; Nigatie, Y.G.; Kovalchuk, S.V. Towards Predicting Trend of Scientific Research Topics using Topic Modeling. *Procedia Comput. Sci.* **2018**, *136*, 304–310. [\[CrossRef\]](#)
45. Abuhay, T.M.; Kovalchuk, S.V.; Bochenina, K.; Mbogo, G.-K.; Visheratin, A.A.; Kamps, G.; Krzhizhanovskaya, V.V.; Lees, M.H. Analysis of publication activity of computational science society in 2001–2017 using topic modelling and graph theory. *J. Comput. Sci.* **2018**, *26*, 193–204. [\[CrossRef\]](#)
46. Abuhay, T.M.; Kovalchuk, S.V.; Bochenina, K.O.; Kamps, G.; Krzhizhanovskaya, V.V.; Lees, M.H. Analysis of Computational Science Papers from ICCS 2001–2016 using Topic Modeling and Graph Theory. *Procedia Comput. Sci.* **2017**, *108*, 7–17. [\[CrossRef\]](#)
47. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Tul, Q.; Ali, M.; Riaz, A.; Noureen, A.; Kamran, M.; Hayat, B.; Rehman, A.; Ain, Q.T. Sentiment Analysis Using Deep Learning Techniques: A Review. *Int. J. Adv. Comput. Sci. Appl.* **2017**, *8*, 426–433. [\[CrossRef\]](#)
49. Zhang, L.; Wang, S.; Liu, B. Deep learning for sentiment analysis: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, 1–25. [\[CrossRef\]](#)
50. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; The MIT Press: London, UK, 2016; ISBN 9780262035613.
51. Murfi, H.; Rosaline, N.; Hariadi, N. Deep autoencoder-based fuzzy C-means for topic detection. *arXiv* **2021**, arXiv:2102.02636.
52. Ojo, A.; Dzhusupova, Z.; Curry, E. Exploring the Nature of the Smart Cities Research Landscape. In *Public Administration and Information Technology*; Springer International Publishing: New York, NY, USA, 2016; pp. 23–47.
53. Mora, L.; Deakin, M.; Reid, A. Combining co-citation clustering and text-based analysis to reveal the main development paths of smart cities. *Technol. Forecast. Soc. Chang.* **2019**, *142*, 56–69. [\[CrossRef\]](#)
54. Fu, Y.; Zhang, X. Trajectory of urban sustainability concepts: A 35-year bibliometric analysis. *Cities* **2017**, *60*, 113–123. [\[CrossRef\]](#)
55. de Jong, M.; Joss, S.; Schraven, D.; Zhan, C.; Weijnen, M. Sustainable-smart-resilient-low carbon-eco-knowledge cities; making sense of a multitude of concepts promoting sustainable urbanization. *J. Clean. Prod.* **2015**, *109*, 25–38. [\[CrossRef\]](#)
56. Min, K.; Yoon, M.; Furuya, K. A Comparison of a Smart City's Trends in Urban Planning before and after 2016 through Keyword Network Analysis. *Sustainability* **2019**, *11*, 3155. [\[CrossRef\]](#)
57. Shi, J.-G.; Miao, W.; Si, H. Visualization and Analysis of Mapping Knowledge Domain of Urban Vitality Research. *Sustainability* **2019**, *11*, 988. [\[CrossRef\]](#)
58. Guo, Y.-M.; Huang, Z.-L.; Guo, J.; Li, H.; Guo, X.-R.; Nkeli, M.J. Bibliometric Analysis on Smart Cities Research. *Sustainability* **2019**, *11*, 3606. [\[CrossRef\]](#)
59. Park, K.C.; Lee, C.H. A study on the research trends for smart city using topic modeling. *J. Internet Comput. Serv.* **2019**, *20*, 119–128.
60. Trindade, E.P.; Hinnig, M.P.F.; Da Costa, E.M.; Marques, J.S.; Bastos, R.C.; Yigitcanlar, T. Sustainable development of smart cities: A systematic review of the literature. *J. Open Innov. Technol. Mark. Complex.* **2017**, *3*, 11. [\[CrossRef\]](#)
61. Pitichotchkphokhin, P.; Chuangkru, P.; Kalakan, K.; Suntisrivaraporn, B.; Leelanupab, T.; Kanungsukkasem, N. Discover Underlying Topics in Thai News Articles: A Comparative Study of Probabilistic and Matrix Factorization Approaches. In Proceedings of the 2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Phuket, Thailand, 24–27 June 2020; pp. 759–762.
62. Chen, Y.; Zhang, H.; Liu, R.; Ye, Z.; Lin, J. Experimental explorations on short text topic mining between LDA and NMF based Schemes. *Knowl.-Based Syst.* **2019**, *163*, 1–13. [\[CrossRef\]](#)
63. O'Callaghan, D.; Greene, D.; Carthy, J.; Cunningham, P. An analysis of the coherence of descriptors in topic modeling. *Expert Syst. Appl.* **2015**, *42*, 5645–5657. [\[CrossRef\]](#)
64. Mifrah, S.; Benlahmar, E.H. Topic modeling coherence : A comparative study between LDA and NMF models using COVID ' 19 corpus. *Int. J. Adv. Trends Comput. Sci. Eng.* **2020**, *9*, 5756–5761. [\[CrossRef\]](#)
65. Naud, A.; Usui, S. Exploration of a collection of documents in neuroscience and extraction of topics by clustering. *Neural Networks* **2008**, *21*, 1205–1211. [\[CrossRef\]](#) [\[PubMed\]](#)
66. Jayabharathy, J.; Kanmani, S.; Parveen, A.A. Document clustering and topic discovery based on semantic similarity in scientific literature. In Proceedings of the 2011 IEEE 3rd International Conference on Communication Software and Networks, Xi'an, China, 27–29 May 2011; pp. 425–429.
67. Abuaiadah, D. Using Bisect K-Means Clustering Technique in the Analysis of Arabic Documents. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* **2016**, *15*, 1–13. [\[CrossRef\]](#)

68. Nur'Aini, K.; Najahaty, I.; Hidayati, L.; Murfi, H.; Nurrohmah, S. Combination of singular value decomposition and K-means clustering methods for topic detection on Twitter. In Proceedings of the 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Depok, West Java, 10–11 October 2015; pp. 123–128.
69. Mursidah, I.; Murfi, H. Analysis of initialization method on fuzzy c-means algorithm based on singular value decomposition for topic detection. In Proceedings of the 2017 1st International Conference on Informatics and Computational Sciences (ICICoS), Semarang, Indonesia, 15–16 November 2017; pp. 213–218.
70. Madlock-Brown, R. A Framework for Emerging Topic Detection in Biomedicine. Ph.D. Thesis, University of Iowa, Iowa City, Iowa, 2014.
71. Bora, D.J.; Gupta, A.K. A Comparative study Between Fuzzy Clustering Algorithm and Hard Clustering Algorithm. *Int. J. Comput. Trends Technol.* **2014**, *10*, 108–113. [\[CrossRef\]](#)
72. Lucic, M.; Bachem, O.; Krause, A. Strong coresets for hard and soft bregman clustering with applications to exponential family mixtures. In Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, AISTATS 2016, Cadiz, Spain, 9–11 May 2016.
73. Parlina, A.; Ramli, K. Performance Comparison of Clustering Algorithms on Scientific Publications. *Adv. Sci. Lett.* **2017**, *23*, 3730–3732. [\[CrossRef\]](#)
74. Murfi, H. The Accuracy of Fuzzy c-Means in Lower-Dimensional Space for Topic Detection. In *Smart Computing and Communication. SmartCom 2018*; Qiu, M., Ed.; Lecture Notes in Computer Science; Springer Champ: Cham, Switzerland, 2018; ISBN 9783030057558.
75. Bezdek, J.C.; Ehrlich, R.; Full, W. FCM: The fuzzy c-means clustering algorithm. *Comput. Geosci.* **1984**, *10*, 191–203. [\[CrossRef\]](#)
76. Sutrisman, R.T.; Murfi, H. Analysis of Non-Negative Double Singular Value Decomposition Initialization Method on Eigenspace-based Fuzzy C-Means Algorithm for Indonesian Online News Topic Detection. In Proceedings of the 2018 6th International Conference on Information and Communication Technology (ICoICT), Bandung, Indonesia, 3–4 May 2018; pp. 55–60.
77. Winkler, R.; Klawonn, F.; Kruse, R. Fuzzy C-Means in High Dimensional Spaces. *Int. J. Fuzzy Syst. Appl.* **2011**, *1*, 1–16. [\[CrossRef\]](#)
78. Song, C.; Liu, F.; Huang, Y.; Wang, L.; Tan, T. Auto-encoder Based Data Clustering. In *Constructive Side-Channel Analysis and Secure Design*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 117–124.
79. Song, C.; Huang, Y.; Liu, F.; Wang, Z.; Wang, L. Deep auto-encoder based clustering. *Intell. Data Anal.* **2014**, *18*, S65–S76. [\[CrossRef\]](#)
80. Xie, J.; Girshick, R.; Farhadi, A. Unsupervised deep embedding for clustering analysis. In Proceedings of the ICML 2016: 33rd International Conference on Machine Learning (ICML-2016), New York, NY, USA, 19–24 June 2016; pp. 740–749.
81. Guo, X.; Gao, L.; Liu, X.; Yin, J. Improved Deep Embedded Clustering with Local Structure Preservation. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017; pp. 1753–1759.
82. Guan, R.; Zhang, H.; Liang, Y.; Giunchiglia, F.; Huang, L.; Feng, X. Deep Feature-Based Text Clustering and Its Explanation. *IEEE Trans. Knowl. Data Eng.* **2020**, *14*, 1. [\[CrossRef\]](#)
83. European Smart Cities 4.0. Available online: <http://www.smart-cities.eu/index.php?cid=2&ver=4> (accessed on 18 August 2020).
84. Schuler, D. Digital Cities and Digital Citizens. In *Constructive Side-Channel Analysis and Secure Design*; Springer International Publishing: New York, NY, USA, 2002; pp. 71–85.
85. He, B.-J.; Zhao, D.-X.; Gou, Z. Integration of Low-Carbon Eco-City, Green Campus and Green Building in China. In *Green Energy and Technology*; Springer International Publishing: New York, NY, USA, 2019; pp. 49–78.
86. UNEP towards a Green Economy: Pathways to Sustainable Development and Poverty Eradication; United Nations Environment: Nairobi, Kenya, 2011.
87. Ferguson, D.; Sairamesh, J.; Feldman, S. Open frameworks for information cities. *Commun. ACM* **2004**, *47*, 45–49. [\[CrossRef\]](#)
88. Komninos, N. The architecture of intelligent cities: Integrating human, collective and artificial intelligence to enhance knowledge and innovation. In Proceedings of the 2nd IET International Conference on Intelligent Environments (IE 06), Athens, Greece, 5–6 July 2006; Institution of Engineering and Technology (IET): London, UK; pp. 13–20.
89. Edvardsson, I.R.; Yigitcanlar, T.; Pancholi, S. Knowledge city research and practice under the microscope: A review of empirical findings. *Knowl. Manag. Res. Pr.* **2016**, *14*, 537–564. [\[CrossRef\]](#)
90. Jucevičienė, P. Sustainable Development of the Learning City. *Eur. J. Educ.* **2010**, *45*, 419–436. [\[CrossRef\]](#)
91. Godschalk, D.R. Urban Hazard Mitigation: Creating Resilient Cities. *Nat. Hazards Rev.* **2003**, *4*, 136–143. [\[CrossRef\]](#)
92. Angelidou, M. The Role of Smart City Characteristics in the Plans of Fifteen Cities. *J. Urban Technol.* **2017**, *24*, 3–28. [\[CrossRef\]](#)
93. ICLEI Local Governments for Sustainability Sustainable City. Available online: <http://old.iclei.org/index.php?id=35> (accessed on 26 October 2020).
94. Shin, D.; Nah, Y.; Lee, I.-S.; Yi, W.S.; Won, Y.-J. Security Protective Measures for the Ubiquitous City Integrated Operation Center. In Proceedings of the 2008 Third International Conference on Broadband Communications, Information Technology & Biomedical Applications, Gauteng, South Africa, 23–26 November 2008; pp. 239–244. [\[CrossRef\]](#)
95. Fan, W.; Shi, Y.; Peng, Z.; Liu, S. Research on Application of VRML in Virtual City Construction. In Proceedings of the 2009 International Joint Conference on Artificial Intelligence, Pasadena, CA, USA, 11–17 July 2009; pp. 598–601. [\[CrossRef\]](#)
96. Hollands, R.G. Will the real smart city please stand up? *City* **2008**, *12*, 303–320. [\[CrossRef\]](#)
97. Macke, J.; Sarate, J.A.R.; Moschen, S.D.A. Smart sustainable cities evaluation and sense of community. *J. Clean. Prod.* **2019**, *239*, 118103. [\[CrossRef\]](#)
98. Kumari, A.; Tanwar, S. Secure Data Analytics for Smart Grid Systems in a Sustainable Smart City: Challenges, Solutions, and Future Directions. *Sustain. Comput. Inform. Syst.* **2020**, *28*, 100427. [\[CrossRef\]](#)

99. Majumdar, S.; Subhani, M.M.; Roullier, B.; Anjum, A.; Zhu, R. Congestion prediction for smart sustainable cities using IoT and machine learning approaches. *Sustain. Cities Soc.* **2021**, *64*, 102500. [CrossRef]
100. Singh, S.; Sharma, P.K.; Yoon, B.; Shojafar, M.; Cho, G.H.; Ra, I.-H. Convergence of blockchain and artificial intelligence in IoT network for the sustainable smart city. *Sustain. Cities Soc.* **2020**, *63*, 102364. [CrossRef]
101. Ahad, M.A.; Paiva, S.; Tripathi, G.; Feroz, N. Enabling technologies and sustainable smart cities. *Sustain. Cities Soc.* **2020**, *61*, 102301. [CrossRef]
102. Shafiq, M.; Tian, Z.; Bashir, A.K.; Jolfaei, A.; Yu, X. Data mining and machine learning methods for sustainable smart cities traffic classification: A survey. *Sustain. Cities Soc.* **2020**, *60*, 102177. [CrossRef]
103. Zahmatkesh, H.; Al-Turjman, F. Fog computing for sustainable smart cities in the IoT era: Caching techniques and enabling technologies—An overview. *Sustain. Cities Soc.* **2020**, *59*, 102139. [CrossRef]
104. Yigitcanlar, T.; Kamruzzaman, M.; Foth, M.; Sabatini-Marques, J.; Da-Costa, E.; Ioppolo, G. Can cities become smart without being sustainable? A systematic review of the literature. *Sustain. Cities Soc.* **2019**, *45*, 348–365. [CrossRef]
105. Höjer, M.; Wangel, J. Smart Sustainable Cities: Definition and Challenges. In *New Advances in Information Systems and Technologies*; Springer Science and Business Media LLC: Secaucus, NJ, USA, 2015; pp. 333–349.
106. Brundtland, G.H. World Commission on environment and development. *Environ. Policy Law* **1985**, *14*, 26–30. [CrossRef]
107. Bibri, S.E.; Krogstie, J. Smart sustainable cities of the future: An extensive interdisciplinary literature review. *Sustain. Cities Soc.* **2017**, *31*, 183–212. [CrossRef]
108. Dewalska-Opitek, A. Smart city concept—the citizens’ perspective. In Proceedings of the International Conference on Transport Systems Telematics, Ustron, Poland, 22–25 October 2014; pp. 331–340.
109. Citiasia Center for Smart Nation (CCSN) Mastering Nation’s Advancement from Smart Readiness to Smart City. Available online: <https://docplayer.info/38729269-Citiasia-center-for-smartnation-smart-nation-mastering-nation-s-advancement-from-smart-readiness-to-smart-city-powered-by-smart-nation-i-1.html> (accessed on 10 July 2020).
110. Hassan, S.I.; Agarwal, P. Analytical approach to sustainable smart city using IoT and machine learning. In *Big Data, IoT, and Machine Learning: Tools and Applications*; Agrawal, R., Paprzycki, M., Gupta, N., Eds.; CRC Press: Boca Raton, FL, USA, 2020; pp. 277–294. ISBN 9780367512224.
111. Dameri, R.P. Using ICT in Smart City. In *Advances and New Trends in Environmental Informatics*; Springer International Publishing: New York, NY, USA, 2016; pp. 45–65.
112. Adio-Moses, D.; Oladiran, O. Smart city strategy and sustainable development goals for building construction framework in Lagos. In Proceedings of the 2016 International Conference on Sustainable Development (ICSD), New York, NY, USA, 21–22 September 2016.
113. Shmelev, S.E.; Shmeleva, I.A. Sustainable cities: Problems of integrated interdisciplinary research. *Int. J. Sustain. Dev.* **2009**, *12*, 4–23. [CrossRef]
114. Jenks, M.; Jones, C. (Eds.) *Dimensions of the Sustainable City*; Springer Science+Business Media, LLC: Secaucus, NJ, USA, 2010; ISBN 9781402086465.
115. Allen, A. Sustainable cities or sustainable urbanisation? *UCL’s J. Sustain. Cities* **2009**, *1*, 1–2.
116. Zhai, C.; Massung, S. *Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining*; ACM: New York, NY, USA, 2016.
117. Xiong, H.; Cheng, Y.; Zhao, W.; Liu, J. Analyzing scientific research topics in manufacturing field using a topic model. *Comput. Ind. Eng.* **2019**, *135*, 333–347. [CrossRef]
118. Jan, B.; Farman, H.; Khan, M.; Imran, M.; Islam, I.U.; Ahmad, A.; Ali, S.; Jeon, G. Deep learning in big data Analytics: A comparative study. *Comput. Electr. Eng.* **2019**, *75*, 275–287. [CrossRef]
119. Bodyn, L. Exploration of Deep Autoencoders on Cooking Recipes. Ph.D. Thesis, Universiteit Gent, Hyères, France, 2017.
120. Vincent, P.; Larochelle, H.; Jaoie, I.; Bengio, Y.; Manzagol, P.-A. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.* **2010**, *11*, 3371–3408.
121. Bhat, M.R.A.; Kundroo, M.A.; Tarray, T.; Agarwal, B. Deep LDA: A new way to topic model. *J. Inf. Optim. Sci.* **2019**, *41*, 823–834. [CrossRef]
122. Bishop, C.M. Pattern recognition and machine learning. In *Information Science and Statistics*; Springer: New York, NY, USA, 2006; pp. 21–24. ISBN 9780387310732.
123. Pedrycz, W.; Chen, S. (Eds.) *Deep Learning: Concepts and Architectures*; Springer Nature Switzerland AG: Cham, Switzerland, 2020; ISBN 9783030317553.
124. About Scopus. Available online: <https://www.elsevier.com/en-gb/solutions/scopus> (accessed on 6 November 2020).
125. CORE Dataset. Available online: <https://core.ac.uk/services/dataset/> (accessed on 6 November 2020).
126. Knoth, P.; Pontika, N. Aggregating research papers from publishers’ systems to support text and data mining: Deliberate lack of interoperability or not? In Proceedings of the Workshop on Cross-Platform Text Mining and Natural Language Processing Interoperability (INTEROP 2016), Portorož, Slovenia, 12 July 2016.
127. Kulkarni, A.; Shivananda, A. *Natural Language Processing Recipes: Unlocking Text Data with Machine Learning and Deep Learning Using Python*; Apress: New York, NY, USA, 2019; ISBN 2006062298.
128. Cakir, M.U.; Guldamlasioglu, S. Text Mining Analysis in Turkish Language Using Big Data Tools. In Proceedings of the 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC), Atlanta, GA, USA, 10–14 June 2016; pp. 614–618.

129. Lau, J.H.; Newman, D.; Baldwin, T. Machine Reading Tea Leaves: Automatically Evaluating Topic Coherence and Topic Model Quality. In Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, Gothenburg, Sweden, 26–30 April 2014; pp. 530–539.
130. Amarasinghe, K.; Manic, M.; Hruska, R. Optimal stop word selection for text mining in critical infrastructure domain. In Proceedings of the 2015 Resilience Week (RWS), Philadelphia, PA, USA, 18–20 August 2015.
131. Zaman, A.N.K.; Matsakis, P.; Brown, C. Evaluation of stop word lists in text retrieval using Latent Semantic Indexing. In Proceedings of the 2011 Sixth International Conference on Digital Information Management, Melbourne, Australia, 26–28 September 2011; pp. 133–136.
132. Lindahl, A. Linguistics and Theory of Science Topic Modeling for Analysis of Public Discourse—Enriching Topic Modeling with Linguistic Information to Analyze Swedish Housing Policies. Ph.D. Thesis, University of Gothenburg, Gothenburg, Sweden, 2017.
133. The Ecologist Vol 2 (1), January 1972 A blueprint for survival. Available online: <https://www.resurgence.org/magazine/ecologist/issues1970-1979.html> (accessed on 5 September 2020).
134. Basiago, A.D. The search for the sustainable city in 20th century urban planning. *Environment* **1996**, *16*, 135–155. [CrossRef]
135. Breidenich, C.; Magraw, D.; Rowley, A.; Rubin, J.W. The Kyoto Protocol to the United Nations Framework Convention on Climate Change. *Am. J. Int. Law* **1998**, *92*, 315–331. [CrossRef]
136. Cocchia, A. Smart and digital city: A systematic literature review. In *Smart City*; Dameri, R.P., Rosenthal-Sabroux, C., Eds.; Springer International Publishing Switzerland: Cham, Switzerland, 2014; pp. 13–43. ISBN 978-3-319-06159-7.
137. Dameri, R.P.; Cocchia, A. Smart city and digital city: Twenty years of terminology evolution. In Proceedings of the X Conference of the Italian Chapter of AIS, ITAIS, Milan, Italy, 14 December 2013; pp. 1–8.
138. IBM IBM100—Smarter Planet. Available online: <https://www.ibm.com/ibm/history/ibm100/us/en/icons/smarterplanet/> (accessed on 10 November 2020).
139. *A European Strategy for Smart, Sustainable and Inclusive Growth*; European Commission Europe: Brussels, Belgium, 2010.
140. Gao, Y.; Gao, X.; Zhang, X. The 2 °C Global Temperature Target and the Evolution of the Long-Term Goal of Addressing Climate Change—From the United Nations Framework Convention on Climate Change to the Paris Agreement. *Engineering* **2017**, *3*, 272–278. [CrossRef]
141. Wu, J.; Thompson, J.; Zhang, H.; Prasad, R.V.; Guo, S. Green communications and computing networks [Series Editorial]. *IEEE Commun. Mag.* **2016**, *54*, 106–107. [CrossRef]
142. United Nations United Nations Sustainable Development – 17 Goals to Transform Our World. Available online: <https://www.un.org/sustainabledevelopment/> (accessed on 20 November 2020).
143. Banister, D.; Watson, S.; Wood, C. Sustainable cities: Transport, energy, and urban form. *Environ. Plan. B Plan. Des.* **1997**, *24*, 125–143. [CrossRef]
144. Goldman, T.; Gorham, R. Sustainable urban transport: Four innovative directions. *Technol. Soc.* **2006**, *28*, 261–273. [CrossRef]
145. Marzooqi, S.A.; Nuaimi, E.A.; Qirim, N.A. E-Governance (G2C) in the public sector: Citizens acceptance to e-government systems—Dubai’s case. In Proceedings of the Second International Conference on Internet of things, Data and Cloud Computing, New York, NY, USA, 22–23 March 2017; pp. 1–11.
146. Maheshwari, A.K. Application of big data to smart cities for a sustainable future. In *Handbook of Engaged Sustainability*; Marques, J., Ed.; Springer International Publishing: Cham, Switzerland, 2018; pp. 945–968. ISBN 978-3-319-71312-0.
147. United Nations Sustainable Development Goals. Clean Water and Sanitation. Available online: <https://www.un.org/sustainabledevelopment/water-and-sanitation/> (accessed on 23 August 2020).
148. *UN Policy Brief: COVID-19 in an Urban World*; United Nations: New York, NY, USA, 2020; pp. 1–30.
149. Tahir, M.B.; Batool, A. COVID-19: Healthy environmental impact for public safety and menaces oil market. *Sci. Total. Environ.* **2020**, *740*, 140054. [CrossRef] [PubMed]
150. Elavarasan, R.M.; Shafiullah, G.; Raju, K.; Mudgal, V.; Arif, M.; Jamal, T.; Subramanian, S.; Balaguru, V.S.; Reddy, K.; Subramaniam, U. COVID-19: Impact analysis and recommendations for power sector operation. *Appl. Energy* **2020**, *279*, 115739. [CrossRef]
151. Sharifi, A.; Khavarian-Garmsir, A.R. The COVID-19 pandemic: Impacts on cities and major lessons for urban planning, design, and management. *Sci. Total. Environ.* **2020**, *749*, 142391. [CrossRef] [PubMed]
152. Sein, M.K. The serendipitous impact of COVID-19 pandemic: A rare opportunity for research and practice. *Int. J. Inf. Manag.* **2020**, *55*, 102164. [CrossRef] [PubMed]
153. Pan, S.L.; Zhang, S. From fighting COVID-19 pandemic to tackling sustainable development goals: An opportunity for responsible information systems research. *Int. J. Inf. Manag.* **2020**, *55*, 102196. [CrossRef] [PubMed]
154. Buhrmester, V.; Münch, D.; Arens, M. Analysis of explainers of black box deep neural networks for computer vision: A survey. *arXiv* **2019**, arXiv:1911.12116.
155. Syed, S.; Spruit, M. Full-Text or Abstract? Examining Topic Coherence Scores Using Latent Dirichlet Allocation. In Proceedings of the 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Tokyo, Japan, 19–21 October 2017; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA; pp. 165–174.