

Article

Identifying Smart City Leaders and Followers with Machine Learning

Fangyao Liu ¹, Nicole Damen ², Zhengxin Chen ³, Yong Shi ³, Sihai Guan ^{1,*} and Daji Ergu ^{1,*}

¹ College of Electronic and Information, Southwest Minzu University, Chengdu 610093, China

² School of Interdisciplinary Informatics, University of Nebraska at Omaha, Omaha, NE 68182, USA; kafuinternational@gmail.com

³ College of Information Science and Technology, University of Nebraska at Omaha, Omaha, NE 68182, USA; zhengxinchen@unomaha.edu (Z.C.)

* Correspondence: gcihey@swun.edu.cn (S.G.); ergudaji@163.com (D.E.)

Abstract: Smart cities have been a popular topic for the city stakeholders. A smart city is the next urban lifestyle that citizens expect. Due to the hypercompetitive and globalized economy, many cities have already started or are about to start their smart city projects. There is no uniform benchmark to evaluate the smart cities' performance. Several organizations use their own indicators to evaluate smart cities worldwide or nationwide. This research paper leverages fuzzy logic to label smart city leaders and followers based on various organization's evaluation meta results and then uses machine learning techniques to identify the key characteristics of leaders and followers. Based on the training data performance, the Support Vector Machine (SVM) is used to predict who will be the next smart city leader or follower. According to the proposed prediction framework, we have successfully predicted 30 smart city leaders and 20 followers.

Keywords: smart city; fuzzy logic; machine learning; prediction



Citation: Liu, F.; Damen, N.; Chen, Z.; Shi, Y.; Guan, S.; Ergu, D. Identifying Smart City Leaders and Followers with Machine Learning. *Sustainability* **2023**, *15*, 9671. <https://doi.org/10.3390/su15129671>

Academic Editors: T. M. Indra Mahlia and Pg Emeroylariffion Abas

Received: 9 May 2023

Revised: 2 June 2023

Accepted: 6 June 2023

Published: 16 June 2023



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

1. Introduction

In 2009, the number of people living in urban areas (3.42 billion) surpassed the number living in rural areas (3.41 billion), and since then the world has become more urban than rural. In 2014, there were 7.2 billion people living on the planet (United Nations, 2014) [1]. It is estimated that by 2017, a majority of people were living in urban areas. The global urban population was expected to grow approximately 1.84% per year between 2015 and 2020, 1.63% per year between 2020 and 2025, and 1.44% per year between 2025 and 2030 (World Health Organization, 2014) [2].

The increasing population trend shows us the importance of arranging city resources. Smart city projects are one of the efficient solutions. The use of smart computing technologies makes the critical infrastructure components and services of a city—which include city administration, education, healthcare, public safety, real estate, transportation, and utilities—more intelligent, interconnected, and efficient (Washburn et al., 2010) [3]. There is a range of conceptual variation generated by replacing smart with other alternative adjectives such as digital city or sustainable city. Mills et al. (2022) [4] also give a definition of smart city from the perspective of big data, artificial intelligence, and other characteristics. Oke et al. (2022) [5,6] found that all the smart city leaders and followers themselves can help each other to overcome some challenges.

Smart city ranking is a useful performance evaluation method. There exist many smart city ranking results. The ranking results give all the city stakeholders an idea about how each smart city is making progress. The results also help stakeholders make decisions; for example, investors may decide which smart city project to invest in based on a reliable ranking result. Many companies, research institutes, and Non-Government Organizations

(NGOs) are working on smart city ranking or evaluation (Albino, Berardi, and Dangelico, 2015) [7]. They are typically displayed as a score or ranking index.

This research will use fuzzy logic and machine learning techniques to predict whether a smart city will be classified as a leader or a follower. This research starts with the current smart cities' leader or a follower classification summary and analysis through fuzzy logic and machine learning techniques. Based on the current smart cities classification result, some insightful rules and information will be extracted for future smart cities prediction.

Only limited smart cities were in the prediction list, due to the limitation of sampling framework, survey budgets, data accessibility, and others. More cities should be included in the prediction list in the future. Furthermore, different ranking results use different methodologies. For example, an organization may use survey methodology; another may use secondary data. These differences lead to "heteroscedastic" results.

Based on the accessible smart city ranking results, a smart city can be either classified as a leader or a follower. A fuzzy logic will be used to summarize the current smart city leaders and followers on the list. This research paper applies several machine learning algorithms to identify smart city leaders and followers by using some existing city indicators. The highest test accuracy algorithm will be used for additional smart city leader and follower predictions. Smart city progress issues will also be investigated based on the prediction.

In their assessment of smartest cities in the Gulf States, Woods et al. (2016) [8] define a smart city leader and follower as follows:

Smart City Leader: These cities have differentiated themselves through the clarity, breadth, and inclusiveness of their smart city vision and planning. They are also leading the way in implementing significant projects at both the pilot and increasingly full-scale levels.

Smart City Follower: These are cities that are beginning their smart city journeys. They may have made initial statements of intent and begun limited pilot projects and soloed operations, but they need to develop a more integrated view for city development and/or stronger leadership for their programs.

Thus, the research question is "What machine learning algorithm can accurately identify smart city leaders and followers based on existing city indicators, and how can this knowledge be used to analyze smart city progress issues?"

The research paper employs a combination of fuzzy logic and machine learning techniques to identify and predict smart city leaders and followers. The authors first use fuzzy logic to label cities as either leaders or followers based on evaluation meta results from various organizations. They then apply machine learning techniques to uncover the key characteristics of each group. Using the Support Vector Machine (SVM) algorithm, the authors use the training data's performance to predict which cities are likely to become smart city leaders or followers. The proposed prediction framework successfully predicted 30 smart city leaders and 20 followers.

2. Related Work

2.1. Call for Clarity

Amidst the multitude of efforts surrounding the notion of the smart cities, Hollands (2008) [9] formulates a critique on the usage of smart cities as a label. The call for clarification finds fertile soil in the research community, which assesses smart city research to be fragmented, divergent, and lacking unifying cohesion and intellectual exchange (Mora, Bolici, and Deakin, 2017) [10].

Hollands' (2008) [9] main critique is that the smart city label incorporates a wide range of fields (from IT to business to communities). However, it remains ambiguous in the ways in which these fields are connected to the smart city notion and to each other. This is exemplified by the way that "smart" can be replaced by a multitude of other adjectives, such as "creative" or "wise" cities, without increasing descriptive clarity. Although Hollands' considers this overlap in meaning to be problematic, Moir, Moonen, and Clark (2014) [11] point out that these slight differences may indicate a desire to highlight one of the specific

aspects of the smart city concept. They observe that smart cities are but one formulation of the more generic ‘future city’ term, which is used to “convey either environmental, social, economic or governance aims, or a hybrid of some or all of these elements” (p. 4). Additionally, the lack of cohesive understanding may also be due to the various different motivations that determine the choice of smart city label. Cities gravitate towards concepts that are most appealing to them in that moment, which may be influenced by factors such as geography and zeitgeist (Eremia, Toma, and Sanduleac, 2017) [12]. For example, after the 1950s, the most popular term in urban development was “sustainable city”, while “digital city” came up in the late 90s (Eremia et al., 2017) [12]. In 2009–2010, “smart city” became the dominant term with previously 132 documents published between 2002 and 2009 to more than 900 in 2010–2012 (Mora et al., 2017) [10].

The current discourse on future cities is distinctive for its global, positive, strategic, integrated, and evidence-led character (Moir et al., 2014) [11]. This is also noted by Hollands, who claims that the way that these labels “link together technological informational, transformations with economic, political and social-cultural change” (Hollands, 2008, p. 305) [10], which is generally positive in nature. With this positive connotation, cities are generally eager to use these labels in an effort to appear more positive as well. Thus, a rhetorical inflation occurs in which the label loses its actual meaning and reference to technological and infrastructural change in favor of marketing-fueled hype. This conflation of labels also occurs with words that might initially appear more neutral, such as “intelligent” or “digital”. These words similarly carry an optimistic assumption regarding urban development (i.e., a harmonious high-tech future) and can have multiple possible meanings (see (Komninos, 2013) [13] for four possible meanings of intelligent cities). The purpose of Hollands’ paper was to break down the usage of the label and its assumptions, thus creating an opportunity for other researchers reflect on and seek clarification of the notion of a smart city. For example, Allwinkle and Cruickshank (2011) [14] critically reflect on the concept of “smartness” and other arguments set forth by Hollands. More recently, Kitchin (2015) [15,16] contrasts Hollands’ arguments by arguing that the majority of the smart city literature actually appears to be non-ideological, commonsensical, and pragmatic. Still, he identifies several shortcomings that inhibit the growth of the smart city agenda. The first of which is in line with Hollands’ argument that there is a lack of shared understanding about the concept and initiatives. Kitchin (2015) [15,16] then extends it by claiming an overreliance on canonical and simplified examples and an absence of in-depth empirical case studies and comparative research in the literature.

In 2014, the European Parliament commissioned a report that maps the state of European smart cities. To do this, they first outlined what a smart city seeks to achieve (Manville, Europe, Millard, Institute, and Liebe, 2014, p. 17) [17]:

“A Smart City is quintessentially enabled by the use of technologies (especially ICT) to improve competitiveness and ensure a more sustainable future by symbiotic linkage of networks of people, businesses, technologies, infrastructures, consumption, energy and spaces”.

As such, their working definition is (Manville et al., 2014, p. 17) [17]:

“A Smart City is a city seeking to address public issues via ICT-based solutions on the basis of a multi-stakeholder, municipally based partnership. These solutions are developed and refined through Smart City initiatives, either as discrete projects or (more usually) as a network of overlapping activities”.

2.2. Smart City Characteristics

Since there is no commonly agreed-upon definition, substantial research effort is conducted on describing the characteristics of smart cities.

The most prominent scheme distinguishes six conceptually distinct characteristics related to a smart city: (1) smart governance, (2) smart people, (3) smart living, (4) smart mobility, (5) smart economy, and (6) smart environment (Giffinger, Fertner, Kramar, and Meijers, 2007) [18]. The European Parliament follows this scheme in the sense that in

order to qualify as a smart city strategy or initiative, it must exhibit at least one of these six characteristics. Other schemas approach the matter from different perspectives. For example, Chourabi et al. explored the literature from multiple fields to propose a framework containing eight core components of smart city initiatives: “(1) management and organization, (2) technology, (3) governance, (4) policy, (5) people and communities, (6) the economy, (7) built infrastructure, and (8) the natural environment” (2012, p. 2291) [19]. Interestingly, the authors caution against using these components to rank smart cities. Instead, they highlight these components as a supportive tool to understand and advance smart city strategies and initiatives. A similar approach was undertaken by Joshi, Saxena, Godbole and Shreya (2016) [20], who propose a six-pillar framework “SMELTS”: (1) social, (2) management, (3) economy, (4) legal, (5) technology, and (6) sustainability. In this framework, technology, economy, and legal are said to have a greater impact on and by the smart city initiatives, which then affect the social, management, and sustainability factors in the outer level [21].

2.3. Fuzzy Logic

The core idea behind fuzzy logic is that it aims to model the more imprecise reasonings used by humans when they make rational decisions, especially in an uncertain and imprecise environment. This is possible due to the human ability to use imprecise, inexact, incomplete, or unreliable knowledge to infer an approximate answer. Thus, fuzzy logic seeks to extend logical reasoning in the sense that if logic is the application of formal principles of reasoning, then fuzzy logic is the application of formal principles of *approximate* reasoning (Zadeh, 1998) [22]. Fuzzy logic is better equipped to handle the concept of a partial truth, because fuzzy logic views everything, including truth itself, as a matter of degree rather than a binary true or false. This does not mean that “fuzzy logic is fuzzy”; rather, it is a “precise logic of imprecision and approximate reasoning”. (Zadeh, 2008) [23]. Its principal facts are that it is logical, fuzzy-set-theoretic, epistemic, and relational (Dzitac, Filip, and Manolescu, 2017) [24]. By providing a mathematical means of representing vagueness, fuzzy logic models or sets are able to recognize, represent, manipulate, interpret, and utilize approximate information. This contrasts with more traditional Western Aristotelian logic systems, which tend to be more binary in approach. It initially drew mixed reactions as science, and engineering at the time did not consider the dullness of class boundaries [25]. Yet, the way that fuzzy logic seeks to formalize the human ability to reason and decide in situations of imperfect importation is one of the factors that has enabled fuzzy logic to be applied to many fields, from artificial intelligence and quantum particle physics to control engineering, robotics, and even natural languages.

3. Importance of Smart City Evaluation

As cities vary widely in their economic, geographical, socio-cultural, and historical make-up, smart city efforts require tailored approaches in order to satisfy the requirements of that particular city. Taking this into consideration, Pellicer et al. (2013) [26] take an innovation and development-based standpoint in which they divide current initiatives into those that feature newly formed cities versus efforts that seek to transform existing traditional cities into smart cities.

Within the smart transportation area, Refs. [27–33] proposed a comprehensive and practical framework for benchmarking cities with specific indicators according to the smartness of their transportation systems. This framework was developed through the (1) formulation of a proper concept of smartness in the context of urban transport system, which the authors view as one that utilizes self-operative and corrective technologies and systems in its operation and management, (2) the generation of a generic matrix of 66 indicators of smartness based on a systematic literature survey, and (3) calculating a composite smartness index (SI) of a city’s transportation system using the smartness indices. They then applied their framework to 26 major cities in the world to provide an illustrative example on how it might be applied by benchmarking smart transport cities

across the world. This study is illustrative in multiple ways. The first is with regard to the selection of the criteria or indicators used for analysis. The criteria for selection of these cities were to rank within the top 50 of a global infrastructure benchmarking study and have at least two million inhabitants. Of the 66 indicators identified, only 21 indicators were ultimately included due to a lack of available information on the other indicators. This reveals a concern with benchmarking studies because due to their reliability on secondary information sources (for reasons that are in many cases perfectly practical and sensible), they may be limited in the quality or generalizability of their results. The quality and availability of information are related to the second concern of the indicators. The authors ran their analysis with both equal and unequal weights assigned to the sub-systems and concluded that this had a strong influence on the resulting city rankings. A third difficulty concerns the relevancy of the results. The authors note that due to the speed at which technology and information changes, the accuracy of the benchmarking study may only be applicable for a short time period only. This is a valid concern and one that applies especially to the smart city field as smart city initiatives are constantly initiated and terminated [34].

In their study, Giffinger et al. (2007) [18] specifically focused on medium-sized cities in Europe. The discourse regarding city development is often discussed in a similar way that management literature discusses organizations: in broad sweeping terms that pertain more to the larger metropolises and multinationals than to the smaller medium-sized organizations and cities. While size may be an important differentiator, it is not the only or most important characteristic by which these entities differ. Giffinger et al. (2007) [18] observe that medium-sized cities often have less resources, organizing capacity, and critical mass than their larger counterparts, forcing them to have to be more selective competitive. Yet, comparisons between cities rely on similar metrics, no matter the size or circumstance. This is not to say that city rankings are identical. On the contrary, rankings are known to produce different results depending on their aims and resources as well as their data collection, processing, and analysis methods. Additionally, not all cities are included in the ranking, often due to issues with data access or quality. Therefore, although city rankings can be a useful tool to assess the attractiveness of urban regions and to identify city strengths and assets, cities are not always able to benefit from them. In an effort to alleviate some of these concerns, Giffinger et al. (2007) [18], based their ranking on a rather comprehensive selection method, sought to apply a more solid methodology that would better reflect the characteristics of medium-sized cities. In addition to their robust methodology, Giffinger et al. [18,35] also contributed to the smart city literature by identifying six characteristics by which smart cities can be understood: smart economy, smart people, smart governance, smart mobility, smart environment, and smart living. These six characteristics can be further described by 33 factors, each of which is further associated with 1–4 indicators for a total of 74 indicators.

4. Identifying Current Leaders and Followers with Machine Learning Algorithms

Below is the process of smart city leaders or followers' identification. At the beginning, we use fuzzy logic to summarize the smart city ranking results and then categorized them into two groups: smart city leaders or smart city followers. Next, use all the smart cities and their corresponding indicators as the training data set. Then, we apply several classic machine learning algorithms to this data set. Based on machine learning algorithm's accuracy performance, the highest accuracy rate algorithm will be used for future smart cities prediction. The below Figure 1 shows the detailed process.

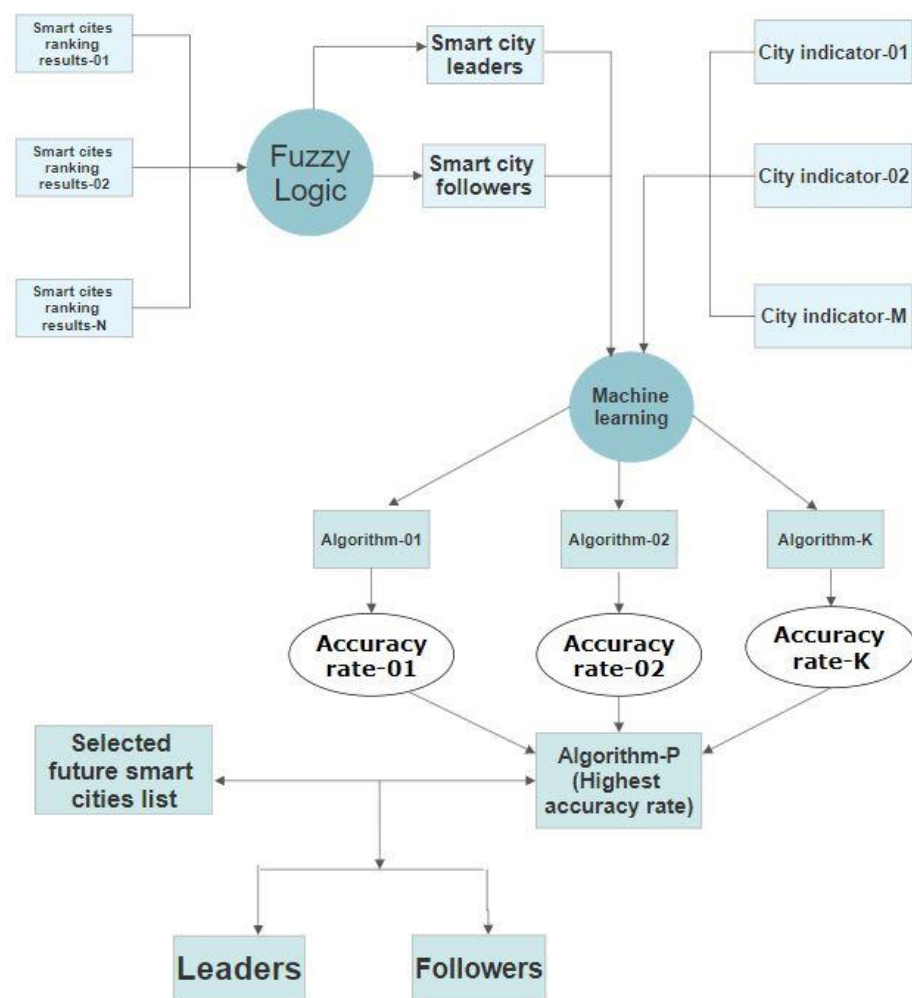


Figure 1. Schema overview of the leader/follower identification process.

5. Data Preparation

5.1. Fuzzy Leaders and Followers Classification

This research uses five organization ranking results for classification. These data sources are selected based on data availability, reputation, data quality, newspaper citation, and other factors. Some of them are institutes; some are companies or NGOs. A city may not be listed on all the ranking results. Table 1 displays the smart city ranking resource details.

Table 1. Smart city ranking resources.

Source Name	Year	Website	Category
Easypark (Stockholm, Sweden)	2018	www.easyparkgroup.com (accessed on 1 January 2020)	Company
Eden Strategy Institute (Singapore)	2018	www.edenstrategyinstitute.com (accessed on 1 January 2020)	Institute
Juniper Research (Chineham Park, UK)	2018	www.juniperresearch.com (accessed on 1 January 2020)	NGO
OTB Research Institute (Delft, The Netherlands)	2017	www.otb.tudelft.nl (accessed on 1 January 2020)	Institute
Navigant Consulting (Chicago, IL, USA)	2017	www.navigant.com (accessed on 1 January 2020)	Company

Different organizations rank smart cities differently. This data preparation applies fuzzy logic to make the leader or follower identifiable. Every ranking list will be divided into three levels. Table 2 shows the three levels and their relative locations.

Table 2. Three ranking levels.

Level Label	Relative Location
RANKING-HIGH	Top 30%
RANKING-MEDIUM	Middle 40%
RANKING-LOW	Low 30%

All the selected cities are assigned with corresponding levels. For example, Tokyo is assigned with “RANKING-HIGH”, “RANKING-MEDIUM”, and “RANKING- MEDIUM”.

Essentially, this will be the fuzzy set problem. A membership function will be used to quantify the grade of membership of the element in X to the fuzzy set.

$$\mu_A : X \rightarrow [0, 1]$$

where μ_A is the membership function, and X represents the universe of discourse while the fuzzy set is A . A Triangular function will be used here. There is a lower limit a , an upper limit b , and a value m , where $a < m < b$ also shown in Figure 2.

$$\mu_A = \begin{cases} 0, & x \leq a \\ \frac{x-a}{m-a}, & a < x \leq m \\ \frac{b-x}{b-m}, & m < x \leq b \\ 0, & x > b \end{cases}$$

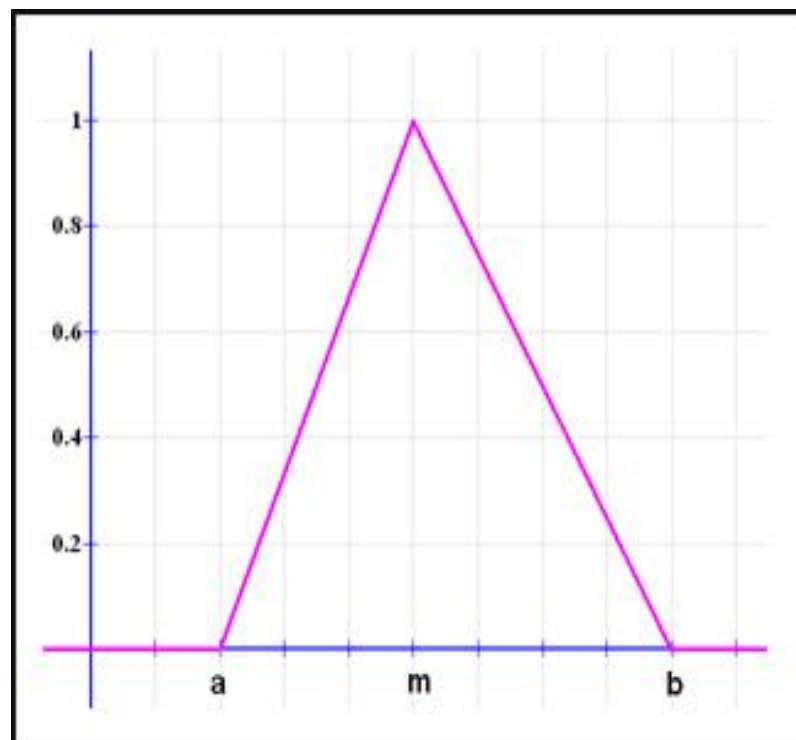


Figure 2. Membership function graph (Flir and Yuan, 1995) [36].

5.2. Defuzzification Process

Defuzzification is the process of converting a fuzzified output into a single crisp value with respect to a fuzzy set. There are many defuzzification methods, such as the Center of Sums Method (COS), Center of Gravity (COG)/Centroid of Area (COA) Method, Center of Area/Bisector of Area Method (BOA), and Weighted Average Method [Flir and Yuan, 1995] [36]. This research takes advantage of the Center of Sums (COS) method, which is one of the most commonly used methods for the defuzzification process. This method is defined as follows:

$$x^* = \frac{\sum_{i=1}^N x_i * \sum_{k=1}^n \mu_{A_k}(x_i)}{\sum_{i=1}^N \sum_{k=1}^n \mu_{A_k}(x_i)}$$

where n is the number of fuzzy sets, N is the number of fuzzy variables, and $\mu_{A_k}(x_i)$ is the membership function for the k -th fuzzy set.

As mentioned before, Tokyo is associated with “RANKING-HIGH”, “RANKING-MEDIUM”, and “RANKING- MEDIUM”.

$$A1 = [(1 - 0) + (0.75 - 0) + (1 - 0)] \times 1/3 = 0.917$$

$$A2 = [(0.75 - 0) + (0.25 - 0) + (-0.5 - 0)] \times 1/3 = 0.5$$

$$A3 = [(0.75 - 0) + (0.25 - 0) + (-0.5 - 0)] \times 1/3 = 0.5$$

The center of the area of the fuzzy set is let to say $\bar{x}_1 = (0.75 + 1)/2 = 0.875$, similarly, $\bar{x}_2 = 0.5$, $\bar{x}_3 = 0.5$

Now, the calculated defuzzification value $x^* = \frac{(A1\bar{x}_1 + A2\bar{x}_2 + A3\bar{x}_3)}{A1 + A2 + A3} = 0.68$

The next step is to give fuzzy classification results to all the smart cities on the list. The list is the top-ranking results, which means it is a leader’s smart city ranking results. Additionally, there are many other follower smart cities not on the list. So, based on a Delphi method adjustment, the fuzzy classification criteria are displayed in Table 3:

Table 3. Fuzzy classification criteria.

Defuzzification Value x^*	Fuzzy Classification	Classification Value
$x^* > 0.3$	Leader	0
$x^* \leq 0.3$	Follower	1

Based on the fuzzy classification criteria, Tokyo should be classified into the leader group. All the other cities can be found in the data set upon request.

5.3. Attribute Selection

After defining fuzzy smart city leader and follower, smart city attributes are also needed. This research selects smart city meta-data (Index) for modeling. All the meta-data relate to smart city dimensions, such as living quality, sustainability, and others. Different smart city concepts have different smart city dimensions. This research selects the most used dimensions. All the metadata are the latest. Most of them are year 2018 data; only a small portion are year 2017 data or earlier. Table 4 summarizes all the smart city meta-data (Index).

Table 4. Smart city meta-data.

Attribute Name	Description	Year	Source
GFCI	Global Financial Centers Index	2018	Long Finance
GLOBAL	Global Cities	2018	A.T. Kearney
ICIM	Cities in Motion Index	2018	IESE Business School Center for Globalization and Strategy & IESE Department of Strategy
LIVEABILITY	Global Livability Index	2018	The Economist Intelligence Unit
LIVING_QUALITY	Quality of Living City Ranking	2018	Mercer
SAFETY	Safe Cities Index	2017	The Economist Intelligence Unit
SUSTAINABLE	Sustainable Cities Index	2018	Arcadis
INNOVATION	Innovation Cities Index	2017	2thinknow
NETWORKED	Networked society city index	2016	Ericsson
GREEN	Green City Index	2012	Siemens AG

6. Models Building and Evaluation

Prescreening is used to ensure the modeling quality. During this process, NETWORKED is removed due to low correlation. Additionally, 30% of the current leader/follower smart cities were removed due to high missing data; only 93 smart cities stay on the data set. There are still some missing values. Moving average smoothing is used here as an efficient imputation method. This research also uses a Python package (sklearn fit_transform) to scale all the attribute values into a range between −1 and 1.

Four types of supervised learning algorithms were implemented in this research.

- o Logistic Regression;
- o KNN;
- o SVM;
- o Neural Network.

To perform machine learning on the smart city data set, we utilized the Scikit-learn and Pandas packages for Python (all the source code will be available on request).

6.1. Logistic Regression

Logistic regression is a classical machine learning classification algorithm that is used to predict the probability of a categorical dependent variable. This research is a two-class value problem, so logistic regression will be used for binary classification.

Below is an example of logistic regression equation:

$$Y = \frac{e^{b_0 + b_1 \times X}}{1 + e^{b_0 + b_1 \times X}}$$

where Y is the predicted output, b_0 is the bias or intercept term, and b_1 is the coefficient for the single input value (X).

Table 5 shows the logistic regression results (test size = 0.20):

Table 5. Logistic regression results.

	Precision	Recall	f1-Score	Support
0	1.00	0.56	0.71	9
1	0.71	1	0.83	10
accuracy			0.79	19
Macro avg	0.86	0.78	0.77	19
Weighted avg	0.85	0.79	0.78	19

6.2. KNN

The k-Nearest Neighbors algorithm (KNN) is a non-parametric method used for classification and regression (Altman, 1992). The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other. For example, for this research classification problem, leader is 0, and follower is 1.

$$Probability(Leader) = Count(Leader) / (Count(Leader) + Count(Follower))$$

$$Probability(Follower) = Count(Follower) / (Count(Leader) + Count(Follower))$$

Table 6 shows the logistic regression results (test size = 0.20):

Table 6. KNN results.

Number of Neighbors (K)	Accuracy
K = 4	0.68
K = 5	0.58
K = 6	0.79
K = 7	0.58
K = 8	0.63
K = 9	0.73

6.3. SVM

Support vector machine (SVM) is a supervised machine learning algorithm. It is powerful for classification problems. For two-dimensional data, there is more than one possible dividing line that can perfectly discriminate between the two classes. The best is accepted to be the hyperplane that creates the largest separation between the two classes, or the maximum margin.

The SVM can be described as the following equation:

$$\hat{Y} = \sum_{i=1}^n \lambda_i K(X, X_i)$$

The SVM is more powerful when it is associated with kernels, especially for the non-linear relationship classification more fit. The kernel projects data into higher-dimensional space defined by polynomials, Gaussian basis functions, or other functions.

This research uses four different kernel functions: sigmoid functions, radial-basis functions (RBF), polynomial, and linear. Table 7 illustrates all the four function names and their kernels.

Table 7. SVM functions.

Function	Kernel Function
Sigmoid	$K(X, Y) = \text{TANH}(\alpha X^T Y + C)$
RBF	$K(X, Y) = \text{EXP}\left(-\frac{\ X - Y\ ^2}{2\sigma^2}\right)$
Polynomial	$K(X, Y) = (\alpha X^T Y + C)^D$
Linear	$K(X, Y) = X^T Y + C$

Table 8 shows the SVM results (test size = 0.20):

Table 8. SVM results.

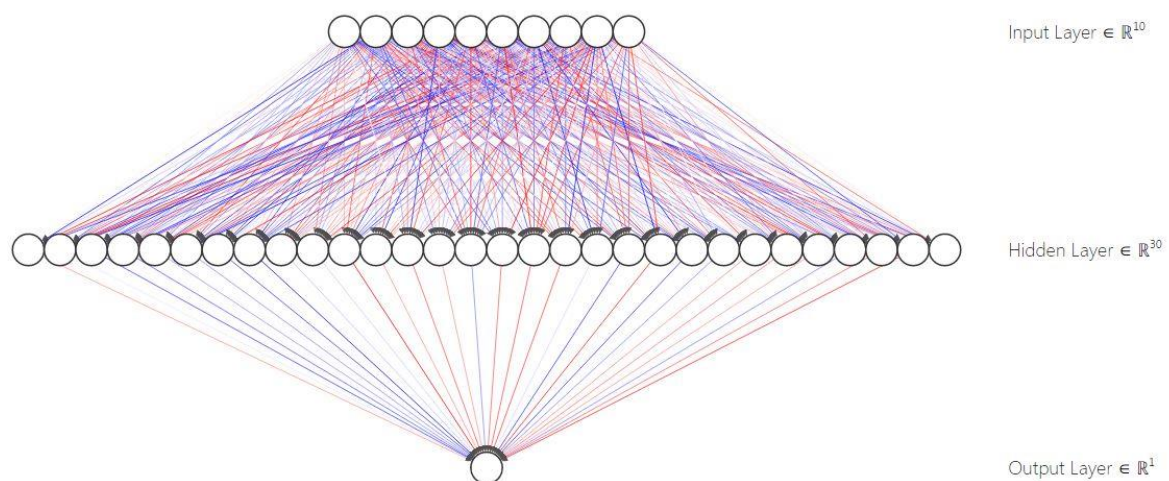
Kernel	C = 1
Sigmoid	0.88
RBF	0.79
Polynomial	0.83
Linear	0.73

6.4. Artificial Neural Network

An artificial neural network is a collection of connected units or nodes, which are inspired by the biological neural networks that constitute animal brains. In an artificial neural network, the signal at a connection between artificial neurons is a real number, and the output of each artificial neuron is computed by some non-linear function of the sum of its inputs.

This research uses a multilayer perceptron (MLP), which is one of the feedforward neural networks. Only three layers are created in this research (input layer, hidden layer, and output layer). For the hidden layer, each node uses the “RELU” activation function. For the output, “SIGMOID” is used as the activation function. The optimizer selects “ADAM”. For the input layer, there are 10 nodes. For the hidden layer, there are 30 nodes, and there is one node for output layer, because of a binary classification problem. The epoch size is 50.

The results of the accuracy of this artificial neural network are little lower than SVM. The highest rate is only 80% (test size = 0.2). Figure 3 is a graph showing this artificial neural network architecture design.

**Figure 3.** Artificial neural network design (Basheer and Hajmeer, 2000) [37].

7. Summary of Models Results

This research evaluates all the four models based on different parameters, such as testing size, Value of N, kernel function, and others. Based on the current smart city data set, the SVM with sigmoid kernel holds the highest accuracy for both 10% and 20% test size. In a word, the prediction will apply this model. Table 9 shows all the machine learning algorithms training and testing results.

Table 9. Machine learning training and testing results.

Algorithms	Condition-1	Condition-2	Test Size = 0.1	Test Size = 0.2
Logistic			0.79	0.85
KNN	K = 4		0.39	0.68
	K = 5		0.68	0.58
	K = 6		0.69	0.79
	K = 7		0.71	0.58
	K = 8		0.66	0.63
	K = 9		0.55	0.73
SVM	Sigmoid	C = 1	0.86	0.88
	RBF	C = 1	0.73	0.79
	Polynomial	C = 1	0.80	0.83
	Linear	C = 1	0.66	0.73
	Sigmoid	C = 2	0.80	0.82
	RBF	C = 2	0.63	0.88
	Polynomial	C = 2	0.75	0.80
	Linear	C = 2	0.69	0.66
Artificial Neural Network	1 Hidden layer	30 nodes	0.75	0.80

According to the above model results, the algorithm SVM (Kernel = Sigmoid, C = 1) has the highest performance score. For all the four algorithms, the results can vary if conditions change. Additionally, if the sample size was increased, the results could also be different. So, based on the above results, the SVM (Kernel = Sigmoid, C = 1) algorithm will be used for the future prediction task.

8. Smart Cities Leader and Follower Prediction

Prediction Results

Using the selected machine learning model (SVM), the 50 cities were predicted as being either a potential smart city leader or follower. Not all the cities were used for prediction. A lot of cities were abandoned due to the problem of missing data. All the cities are listed based on alphabetical order. Table 10 provides an overview of the cities predicted as leaders, while Table 11 lists the cities predicted as followers. Figures 4 and 5 depict the geographical location of the smart city leaders and followers in the world.

Table 10. Overview of cities predicted as leaders.

City	Country	City	Country	City	Country
Almaty	Kazakhstan	Dallas	United States	Minsk	Belarus
Bangkok	Thailand	Damascus	Syria	Munich	Germany
Bengaluru	India	Denver	United States	Muscat	Oman
Bern	Switzerland	Detroit	United States	Phnom Penh	Cambodia
Bogotá	Colombia	Frankfurt	Germany	Phoenix	United States
Brisbane	Australia	Hanoi	Vietnam	Pittsburgh	United States
Calgary	Canada	Kiev	Ukraine	Pretoria	South Africa
Casablanca	Morocco	Luanda	Angola	San José	United States
Colombo	Sri Lanka	Manchester	United Kingdom	Suzhou	China
Curitiba	Brazil	Minneapolis	United States	Tbilisi	Georgia

Table 11. Overview of cities predicted as followers.

City	Country	City	Country	City	Country
Accra	Ghana	Durban	South Africa	Miami	United States
Antwerp	Belgium	Edinburgh	United Kingdom	Medellin	Colombia
Baku	Azerbaijan	Glasgow	United Kingdom	Montevideo	Uruguay
Baltimore	United States	Guangzhou	China	Nagoya	Japan
Basel	Switzerland	Honolulu	United States	Nanjing	China
Belgrade	Serbia	Houston	United States	Rotterdam	The Netherlands
Cleveland	United States	Hyderabad	India		

**Figure 4.** Geographical illustration of where the smart city leaders and followers are located in Europe, Africa, Asia, and the Pacific.

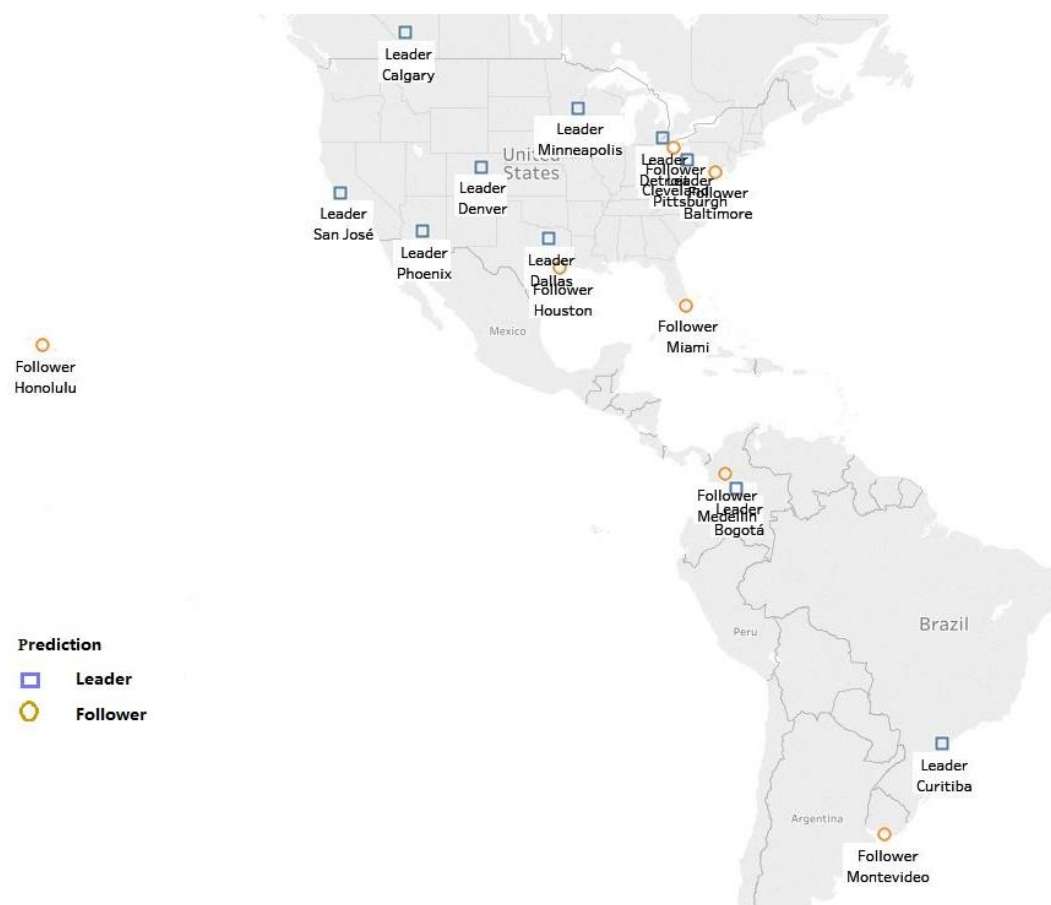


Figure 5. Geographical illustration of where the smart city leaders and followers are located in the Americas.

9. Results Validation

To evaluate the prediction results validation. We use the F-test to compare their internet infrastructure performance improvement. Internet infrastructure is a key factor for smart city projects. Better internet service could lift up the smart city project standards. The internet service plays a significant role in transforming financial, environment, and other aspects of urban life digitally. The International Data Corporation (IDC) states that smart city development uses smart initiatives combined with leverage technology investments across an entire city, with common platforms increasing efficiency, data being shared across systems, and IT investments tied to smart missions. All the tasks rely heavily on the internet service.

To evaluate the internet service improvement, we use the data from Existent Ltd. This company along with New America's Open Technology Institute, Google, Princeton University's PlanetLab, and other supporting partners released an annual worldwide broadband report for 2019, <https://www.cable.co.uk/broadband/speed/worldwide-speed-league/> (accessed on 1 January 2020). This report includes 207 countries' internet service data, such as ranking, mean download speed, distinct IPs tested, and others. This report also includes the data of the year 2018 for comparison purposes. We assume all the cities in the same country have the same internet service performance. To evaluate the performance improvement, we use the internet average download speed change rate from year 2018 to 2019. The formula is below:

$$\text{Performance improvement} = \frac{\text{mean download speed}_{2019} - \text{mean download speed}_{2018}}{\text{mean download speed}_{2018}} \%$$

The F-test is a classic method to evaluate two population data variations. The formula is below:

$$F - \text{test statistic} = \frac{S^2_{Leaders}}{S^2_{Followers}} = \frac{\text{performance improvement variation}_{Leaders}}{\text{performance improvement variation}_{Followers}}$$

After removing the 10% of the extreme data, we conducted an F-test. The F statistic is 2.2605. The p -value is 0.04225, which is less than $\alpha = 0.05$. This result means that under a 95% confidence level, we have sufficient evidence to say that the smart city leader group has a better internet service improvement than the follower group.

10. Discussion

This research used fuzzy logic and machine learning to predict whether smart cities can be categorized as either leaders or followers. This result contributes to a lot of practitioners and theory researchers. All the public and private stakeholders (urban planning department, citizen, and others) could take advantage of it according to their own goals. For example, investors could take advantage of this result of further technology investment decisions; policymakers could use the result and insights for urban planning; employees could take this result into consideration when they decide which city they should move to if they like smart city lifestyle.

This result also contributes to theory development. This smart city classification algorithm has proven a high accuracy based on testing data, which means that smart cities have a significant relationship with its basic elements, such as innovation, living quality, globalization, and others. For example, innovation has a significant positive relationship with smart city evaluation results.

The prediction results indicate that more than one leader or follower comes from the same country. For example, Guangzhou and Nanjing are followers that both come from China. Phoenix and Pittsburgh are both leaders from the United States. This points to the potential effect of peer effects on smart cities, similar to peer effects on classmates. This could be further investigated because if peer effects exist, then it could lead to both theoretical and applied urban planning contributions.

Another finding is that most follower locations are close to the coast, while the leader locations have no such relationship with coastal proximity. For example, in the United States, all the four followers (Cleveland, Baltimore, Miami, and Houston) are close to coast, while the leaders in the United States are nationwide, with some located on the west coast, some located on the east coast, and some located in the middle.

11. Conclusions and Future Work

The smart city prediction results provide a helpful framework of categorizing smart cities. This study has the following limitations. The highest accuracy is less than 90%, according to the experiments, which means there is room for improvement regarding Type I or Type II errors. It is conceivable that some smart city leaders have been mis-categorized as followers and vice versa. There are many reasons that lead to this bias in these errors. Firstly, the original feature data could be biased because most of their city sampling is not transparent. The data collected are not reliable. Secondly, there may be an issue with multi-collinearity. The predicted feature variable could be linearly predicted from others. For example, the livability could be related to safety features, which may adversely affect investment opportunities. Lastly, future experimental design should look into extracting more features, such as technology investments and economic features.

The prediction results are in binary categories, which means that the result is either leader or follower. A possible solution is presenting the results as quantitative values. If so, a scoring system should be developed. Current smart city evaluation methods are only ranking, expert scoring, focus group analysis, or any other qualitative methods. All the methods are biased due to the sample cities' selection transparency. Different evaluation methods have different city samples. Some samples only contain big cities while others just

include developed cities. These evaluation results are not reliable. It is necessary for an evaluation paradigm-shifting from smart city ranking to testing. For future work, we plan to propose a smart city testing framework. The testing framework should be similar to a quiz. Every city could do the quiz and then receive a score. The testing framework would ignore the city selection because all the cities can be tested. Additionally, by doing testing, the testing scores become comparable, either comparable to other cities or itself.

One of the future studies is about investigating the factors that impact smart city leaders or followers. This research shows that the highest accuracy is less than 90%, which means there is a large room for improvement. Those factors of smart city can make a difference. Currently, there is not enough deep investigation of those factors, either factors themselves, or factor interactions.

The difference between the smart city leader and follower prediction results should be further analyzed. For example, the current results assume that there is no significant relationship between smart city identification and Gross Domestic Product (GDP). The assumption can be either rejected or not rejected if further hypothesis testing has been conducted.

Another insight is about shifting the smart city follower to a leader. Being a smart city leader means that citizens have higher satisfaction about their urban lives. All the stakeholders have an agreed goal of shifting to a smart city leader. The actionable and meaningful plan should be further developed and reviewed.

Author Contributions: Conceptualization, F.L. and N.D.; methodology, Y.S.; software, S.G.; validation, Z.C., Y.S. and D.E.; formal analysis, F.L.; investigation, F.L.; resources, N.D.; data curation, N.D.; writing—original draft preparation, Y.S.; writing—review and editing, Z.C.; visualization, S.G.; supervision, D.E.; project administration, D.E.; funding acquisition, D.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been partially supported by grants from the National Natural Science Foundation of China #72174172, and #71774134; supported by “the Fundamental Research Funds for the Central Universities”, Southwest Minzu University (ZYN2022013).

Data Availability Statement: All the data are available upon request.

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

References

1. United Nations. Department of Economic and Social Affairs, Population Division. World Urbanization Prospects: The 2014 Revision, Highlights (ST/ESA/SER.A/352). 2014. Available online: https://www.google.com.hk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjxcPioML_AhWNU94KHbezCywQFnoECBYQAQ&url=https%3A%2F%2Fesa.un.org%2Fpubdpd%2Fwup%2Fpublications%2Ffiles%2Fwup2014-report.pdf&usq=AOvVaw1K8kVV-23-am3YUcFlpk4W (accessed on 1 January 2020).
2. WHO Expert Committee on the Selection, Use of Essential Medicines, & World Health Organization. *The Selection and Use of Essential Medicines: Report of the WHO Expert Committee, 2013 (Including the 18th WHO Model List of Essential Medicines and the 4th WHO Model List of Essential Medicines for Children)*; World Health Organization: Geneva, Switzerland, 2014; Volume 985.
3. Washburn, D.; Sindhu, U.; Balaouras, S.; Dines, R.A.; Hayes, N.; Nelson, L.E. Helping CIOs Understand “Smart City” Initiatives. *Growth* **2010**, *17*, 1–17.
4. Mills, D.; Pudney, S.; Pevcin, P.; Dvorak, J. Evidence-based public policy decision-making in smart cities: Does extant theory support achievement of city sustainability objectives? *Sustainability* **2022**, *14*, 3. [\[CrossRef\]](#)
5. AlAwadhi, S.; Scholl, H.J. Aspirations and realizations: The smart city of Seattle. In Proceedings of the 2013 46th Hawaii International Conference on System Sciences, Wailea, HI, USA, 7–10 January 2013; pp. 1695–1703. [\[CrossRef\]](#)
6. Oke, A.E.; Stephen, S.S.; Aigbavboa, C.O.; Ogunsemi, D.R.; Aje, I.O. Smart City Team Partnership. In *Smart Cities: A Panacea for Sustainable Development*; Emerald Publishing Limited: Bingley, UK, 2022.
7. Albino, V.; Berardi, U.; Dangelico, R.M. Smart cities: Definitions, dimensions, performance, and initiatives. *J. Urban Technol.* **2015**, *22*, 3–21. [\[CrossRef\]](#)
8. Woods, E.; Omara, H.; Ravens, S.; Citron, R. *Gulf States Smart Cities Index: Assessment of Strategy and Execution for 10 Cities (White Paper)*; Navigant Research: Boulder, CO, USA, 2016.
9. Hollands, R.G. Will the real smart city please stand up? *City* **2008**, *12*, 303–320. [\[CrossRef\]](#)
10. Mora, L.; Bolici, R.; Deakin, M. The first two decades of smart-city research: A bibliometric analysis. *J. Urban Technol.* **2017**, *24*, 3–27. [\[CrossRef\]](#)

11. Moir, E.; Moonen, T.; Clark, G. “The Future of Cities: What Is the Global Agenda?” *The Business of Cities*; UK Government: London, UK, 2014. Available online: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/429125/future-cities-global-agenda.pdf (accessed on 1 January 2020).
12. Eremia, M.; Toma, L.; Sanduleac, M. The smart city concept in the 21st century. *Procedia Eng.* **2017**, *181*, 12–19. [CrossRef]
13. Komninos, N. *Intelligent Cities: Innovation, Knowledge Systems and Digital Spaces*; Routledge: Oxford, UK, 2013. [CrossRef]
14. Allwinkle, S.; Cruickshank, P. Creating Smart-er Cities: An Overview. *J. Urban Technol.* **2011**, *18*, 1–16. [CrossRef]
15. Kitchin, R. Making sense of smart cities: Addressing present shortcomings. *Camb. J. Reg. Econ. Soc.* **2015**, *8*, 131–136. [CrossRef]
16. Kitchin, R.; Lauriault, T.P.; McArdle, G. Knowing and governing cities through urban indicators, city benchmarking and real-time dashboards. *Reg. Stud. Reg. Sci.* **2015**, *2*, 6–28. [CrossRef]
17. Manville, C.; Europe, R.; Millard, J.; Institute, D.T.; Liebe, A. Mapping Smart Cities in the EU. 2014. Available online: [https://www.europarl.europa.eu/RegData/etudes/etudes/join/2014/507480/IPOL-ITRE_ET\(2014\)507480_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/etudes/join/2014/507480/IPOL-ITRE_ET(2014)507480_EN.pdf) (accessed on 1 January 2020).
18. Giffinger, R.; Fertner, C.; Kramar, H.; Meijers, E. City-ranking of European Medium-Sized Cities. *Cent. Reg. Sci. Vienna UT* **2007**, *9*, 13.
19. Chourabi, H.; Nam, T.; Walker, S.; Gil-Garcia, J.R.; Mellouli, S.; Nahon, K.; Pardo, T.A.; Scholl, H.J. Understanding smart cities: An integrative framework. In Proceedings of the 2012 45th Hawaii International Conference on System Sciences, Maui, HI, USA, 4–7 January 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 2289–2297.
20. Joshi, S.; Saxena, S.; Godbole, T.; Shreya. Developing Smart Cities: An Integrated Framework. *Procedia Comput. Sci.* **2016**, *93*, 902–909. [CrossRef]
21. Anand, A.; Winfred Rufuss, D.D.; Rajkumar, V.; Suganthi, L. Evaluation of Sustainability Indicators in Smart Cities for India Using MCDM Approach. *Energy Procedia* **2017**, *141*, 211–215. [CrossRef]
22. Zadeh, L.A. Fuzzy logic. *Computer* **1998**, *21*, 83–93. [CrossRef]
23. Zadeh, L.A. Is there a need for fuzzy logic? *Inf. Sci.* **2008**, *178*, 2751–2779. [CrossRef]
24. Dzitac, I.; Filip, F.G.; Manolescu, M.-J. Fuzzy Logic Is Not Fuzzy: World-renowned Computer Scientist Lotfi A. Zadeh. *Int. J. Comput. Commun. Control.* **2017**, *12*, 748–789. [CrossRef]
25. Neirotti, P.; De Marco, A.; Cagliano, A.C.; Mangano, G.; Scorrano, F. Current trends in Smart City initiatives: Some stylised facts. *Cities* **2014**, *38*, 25–36. [CrossRef]
26. Pellicer, S.; Santa, G.; Bleda, A.L.; Maestre, R.; Jara, A.J.; Skarmeta, A.G. A Global Perspective of Smart Cities: A Survey. In Proceedings of the 2013 Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, Taichung, Taiwan, 3–5 July 2013; pp. 439–444. [CrossRef]
27. Deakin, M.; Al Waer, H. From intelligent to smart cities. *Intell. Build. Int.* **2011**, *3*, 140–152. [CrossRef]
28. Debnath, A.K.; Chin, H.C.; Haque, M.M.; Yuen, B. A methodological framework for benchmarking smart transport cities. *Cities* **2014**, *37*, 47–56. [CrossRef]
29. Iswan, M.; Khairul, K.; Siahaan, A.P.U. Fuzzy Logic Concept in Technology, Society, and Economy Areas in Predicting Smart City. *Int. J. Appl. Eng. Res.* **2016**, *2*, 176–181.
30. Lazaroiu, G.C.; Roscia, M. Definition methodology for the smart cities model. *Energy* **2012**, *47*, 326–332. [CrossRef]
31. Lombardi, P.; Giordano, S.; Farouh, H.; Yousef, W. Modelling the smart city performance. *Innov. Eur. J. Soc. Sci. Res.* **2012**, *25*, 137–149. [CrossRef]
32. Marsal-Llacuna, M.-L.; Colomer-Llinàs, J.; Meléndez-Frigola, J. Lessons in urban monitoring taken from sustainable and livable cities to better address the Smart Cities initiative. *Technol. Forecast. Soc. Chang.* **2015**, *90*, 611–622. [CrossRef]
33. Nam, T.; Pardo, T.A. Conceptualizing smart city with dimensions of technology, people, and institutions. In Proceedings of the 12th Annual International Digital Government Research Conference on Digital Government Innovation in Challenging Times, College Park, MD, USA, 12–15 June 2011; dg.o’11. ACM Press: New York, NY, USA, 2011; p. 282. [CrossRef]
34. Jucevičius, R.; Patašienė, I.; Patašius, M. Digital Dimension of Smart City: Critical Analysis. *Procedia-Soc. Behav. Sci.* **2014**, *156*, 146–150. [CrossRef]
35. Tian, Z.; Wang, J.; Wang, J.; Zhang, H. A multi-phase QFD-based hybrid fuzzy MCDM approach for performance evaluation: A case of smart bike-sharing programs in Changsha. *J. Clean. Prod.* **2017**, *171*, 1068–1083. [CrossRef]
36. Flir, G.; Yuan, B. *Fuzzy Sets and Fuzzy Logic*; Prentice Hall: Hoboken, NJ, USA, 1995; Volume 4.
37. Basheer, I.A.; Hajmeer, M. Artificial neural networks: Fundamentals, computing, design, and application. *J. Microbiol. Methods* **2000**, *43*, 3–31. [CrossRef] [PubMed]

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