International Journal of Soft Computing and Engineering (IJSCE)
ISSN: 2231-2307 (Online), Volume-14 Issue-3, July 2024

K-Means Clustering for Smart Cities: An Empirical Study in Indian Context

Biswajit Biswas, Sayantan Ghosh

Abstract: ***Cities are the concourse of people where they live, meet with new people in their daily life, exchange their thoughts and ideas with close one, and earn better livelihoods so that everyone can access proper education, better health facilities, and other essential services and live in a hazards free peace full environment. The Idea behind smart cities is to provide data-driven solutions to social problems, especially in a densely populated area of the country. The population continues to grow and urbanization is causing an explosion in urban areas. Advanced data-supported solutions with the help of technology can be the best solution for simple Governance. Smart City is a data-driven solution for managing people and city resources effectively that helps monitor and manage resources. It helps to make social and economic decisions intelligently. The authors studied and found out the present problem which is lacking between the basic infrastructure for community service particularly in India's small-to-medium cities and the existing smart city development plan***

Keywords: Smart City, Dendrogram, K-Means Clustering, Discriminant Analysis

I. INTRODUCTION

A Smart City uses different modern techniques and technology to provide better societal facilities. A city developed in healthcare, power plants, information systems, education, community service, transportation, crime and security detection, smart banking, and data supply networks is called a Smart city. Mainly this concept is based on IoT (Internet of Things), ICT (information communication technology), and other technological development system. A smart City is a city that is fully developed to manage the urban flow and real-time data collection. The objectives behind the smart city are to improve the city's life functions using smart technology and promote economic growth [7].

In this study, authors selected a few indicators that elaborate on the way of services distribution, and infrastructure and gradually govern the prospective areas for development under the smart city mission. Authors have taken indicators like population density, economy per capita income, healthcare, and education which classify the cities into different segments and compare all the cities using a dendrogram.

Manuscript received on 04 July 2024 | Revised Manuscript received on 09 July 2024 | Manuscript Accepted on 15 July 2024 | Manuscript published on 30 July 2024.

***Correspondence Author(s)
Biswajit Biswas***,** Department of Business Administration, University of Kalyani, Kolkata (West Bengal), India. E-mail: biswajit.biswas0012@gmail.com, ORCID ID: 0000-0002-5302-7675
Sayantan Ghosh, Performance-io, Kolkata (West Bengal), India. E-mail: sayantang28@gmail.com

It finds out the similarities between one or more variables and that categorization explains the structures of urban systems and supports those cities for development in the selected areas which they are lying behind.

II. OBJECTIVE OF THE WORK

A. To comprehensively recognize the multiplicity of difficulties those cities face and determine the standard technology-based approach that addresses the difficulties.

B. To do cluster analysis which is used to classify the cities and draw the dendrogram which is utilized to characterize the classified cities.

C. To do urban typology analysis indicates the huge discrepancies in urban centers located in various geographical regions throughout India.

III. LITERATURE REVIEW

Anticipating pedestrian activity is a compulsory task for maintaining a secure and energy-efficient urban atmosphere. It is possible to get useful information by strategically placing sensors throughout a city. We can identify the typology surrounding those sensors knowing their average activity on different days in a week. Clustering methods perform excellently for these purposes, forming groups of items with high intra-cluster similarity and low inter-cluster dissimilarity. The rapid growth of IoT changed the scenario of various public services positively, in this way the information is provided dynamically and real-time decisions are made. As a result, citizens' lives have become safer, and more convenient, and environmental issues can be addressed more effectively. Sensors in a smart city act as data collectors, collecting a massive amount of data by sensing various environmental parameters and events such as traffic incidents or pedestrian mobility. The increased activity of these devices necessitates proper maintenance and intelligent distribution in order to avoid various problems related to safety [9]. An approach has been attempted to classify the topology building depending on human interaction. In this study, authors have done the classification of topology buildings depending on human interaction and spatial-temporal population density [8].

To construct typology significant empirical work has been done on urban areas which aims to categorize urban environments based on their characteristics [2] City classification is a simplification process in which data collection is required to observe and analyze the patterns [1]. Typology is used to simplify complex information and provide a framework for identifying similarities and differences among cities.

It helps to know the factors that lead to differences in socioeconomic performance among groups or clusters of cities [3]. Typology-building exercises are useful because they allow an understanding of the complicated structure of cities and regions. Many researchers with a multiplicity of aims and outcomes used the urban typology method. In the past population and spatial expansion were used to define urban typology and economic function socioeconomic characteristics, and infrastructure distribution were performed in this process [5]. Giffinger et al applied in their work the urban typology concept in both socioeconomic and sustainable smart city metrics in an innovative way. It helps to build an evidence-based framework for smart city policymaking [6][15][16].

The goal of typology creation is to group indicators data into groups with similar features and also provides insight into how the groupings differed from each other [13][14]. Agglomerative hierarchical clustering and partition clustering are commonly utilized in the literature to accomplish this. To eliminate the bias authors used the agglomerative hierarchical clustering and drawn urban typology [4]. In smart cities moving toward cashless transactions, taken by GoI in the recent past, forced small business owner to adapt software systems for their business requirements providing extra benefit [10][11][12].

IV. RESEARCH GAP

Technology has been kept aside rather than studying how technology is improving the daily lives of people and tasks are getting more complex where previously humans had to work the whole day and all the days in a weak. How people can find out the best solution for their living. In this study, the related research gap is addressed

V. METHODOLOGY

India's cities need better urban planning and technology-aided management. Furthermore, many ideas are merely written down. As a result, the Smart Cities Mission can be viewed as a chance for the Governments to build and convey an intelligent idea while overcoming previous barriers. Table I. provides the constitution for non-governmental organizations that will carry out various activities including quarterly job reviews and the creation of a forum for the exchange of ideas.

Table I: Activities of Job Review and Exchange of Ideas

Entity	Composition	Role
National Level Apex Committee	Representatives from the Ministry of the City, as well as affiliated ministries and organizations.	Review and approve plans, approve funding releases, recommend mid-course corrections, and conduct quarterly activity reviews.
State Level High Powered Steering Committee	Representatives of State and Local government departments.	Provide a forum for the sharing of ideas and the evaluation of smart city proposals.
City Level Smart City Advisory Forum	Representatives of district to local government areas, technical experts, and local NGOs.	Provide guidance and facilitate collaboration among parties.

Source: MoUD, 2015

A. Cluster Analysis

To expose the natural group among the 49 selected cities authors have done an agglomerative hierarchical cluster analysis. The authors used SPSS tools for statistical analysis and drew a dendrogram to demonstrate the clustering. A dendrogram is an effective tree graph to identify the stop point in the clustering process. This hierarchical clustering process continued until all of the cities were united into a single cluster. It is very difficult to select specific parameters for the cluster analysis of distance and linkage measurements. Authors used the Ward method t implement the squared Euclidean distance among the observations.

VI. RESULT AND INTERPRETATION

A summary of case processing is shown in Table II. Which contained the number of valid cases, missing cases, and distance measures that took in squared Euclidean distance.

Table II. Summary of Case Processing

Cases					
Valid		**Missing**		**Total**	
Number	**Percentage**	**Number**	**Percentage**	**Number**	**Percentage**
49	100.00%	0	0.00%	49	100.00%

*a. Squared Euclidean Distance used

Table III(a). Agglomeration Schedule

Stages	Combined Cluster		Coefficients	First Appears Cluster Stage		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	11	21	.001	0	0	14
2	7	17	.002	0	0	27
3	40	45	.003	0	0	5
4	5	43	.004	0	0	16
5	40	41	.006	3	0	16
6	18	29	.010	0	0	25
7	38	44	.013	0	0	27
8	4	46	.017	0	0	26
9	9	42	.021	0	0	21
10	15	48	.026	0	0	13
11	1	26	.034	0	0	29
12	2	47	.042	0	0	20
13	3	15	.050	0	10	22
14	11	39	.059	1	0	19
15	6	36	.070	0	0	30
16	5	40	.081	4	5	20
17	8	30	.092	0	0	37
18	16	25	.103	0	0	22
19	11	27	.119	14	0	24
20	2	5	.138	12	16	37
21	9	32	.161	9	0	33
22	3	16	.185	13	18	36
23	10	35	.209	0	0	29
24	11	31	.240	19	0	31
25	18	49	.271	6	0	30

Table III (b). Agglomeration Schedule Continuation

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
26	4	24	.303	8	0	39
27	7	38	.338	2	7	31
28	22	33	.381	0	0	38
29	1	10	.430	11	23	33
30	6	18	.496	15	25	44
31	7	11	.564	27	24	36
32	14	37	.639	0	0	43
33	1	9	.736	29	21	42
34	12	19	.836	0	0	40
35	13	23	.936	0	0	41
36	3	7	1.042	22	31	44
37	2	8	1.150	20	17	39
38	20	22	1.266	0	28	42
39	2	4	1.427	37	26	48
40	12	28	1.594	34	0	47
41	13	34	1.840	35	0	46
42	1	20	2.109	33	38	43
43	1	14	2.393	42	32	45
44	3	6	2.679	36	30	45
45	1	3	3.257	43	44	46
46	1	13	4.343	45	41	47
47	1	12	5.732	46	40	48
48	1	2	7.715	47	39	0

The authors used a statistical method to expose the natural groups among the 49 cities and the agglomeration schedule is shown in Table 3(a) and III (b)**.** It is a numerical representation of cluster analysis. In the output section agglomeration schedule implies the proximity matrix. It shows how hierarchical cluster analysis groups the cases progressively. In this schedule, each row is interpreted as a stage where two cases are combined and form a cluster with respect to distance and linkage sections. In the schedule listing all stages are aggregated until only one cluster remains after the last stage.

The number of stages is less than one from the number of clustered cases. The authors give an example where 48 stages are presented as the sample consists of 49 belongings. In Table III**(a)**cases 11 and 21 are combined at the first stage as the squared Euclidean distance among these two stages is the smallest out of all pairs that means coefficients are small (approximately .001) in the first stage and it slowly goes up on the progression of the schedule. The purpose of this schedule is to guide the researcher to find out easily where two clusters are combined to create a group while increasing coefficient values.

Retrieval Number: 100.1/ijsce.B811413020724
DOI: 10.35940/ijsce.B8114.14030724
Journal Website: www.ijsce.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

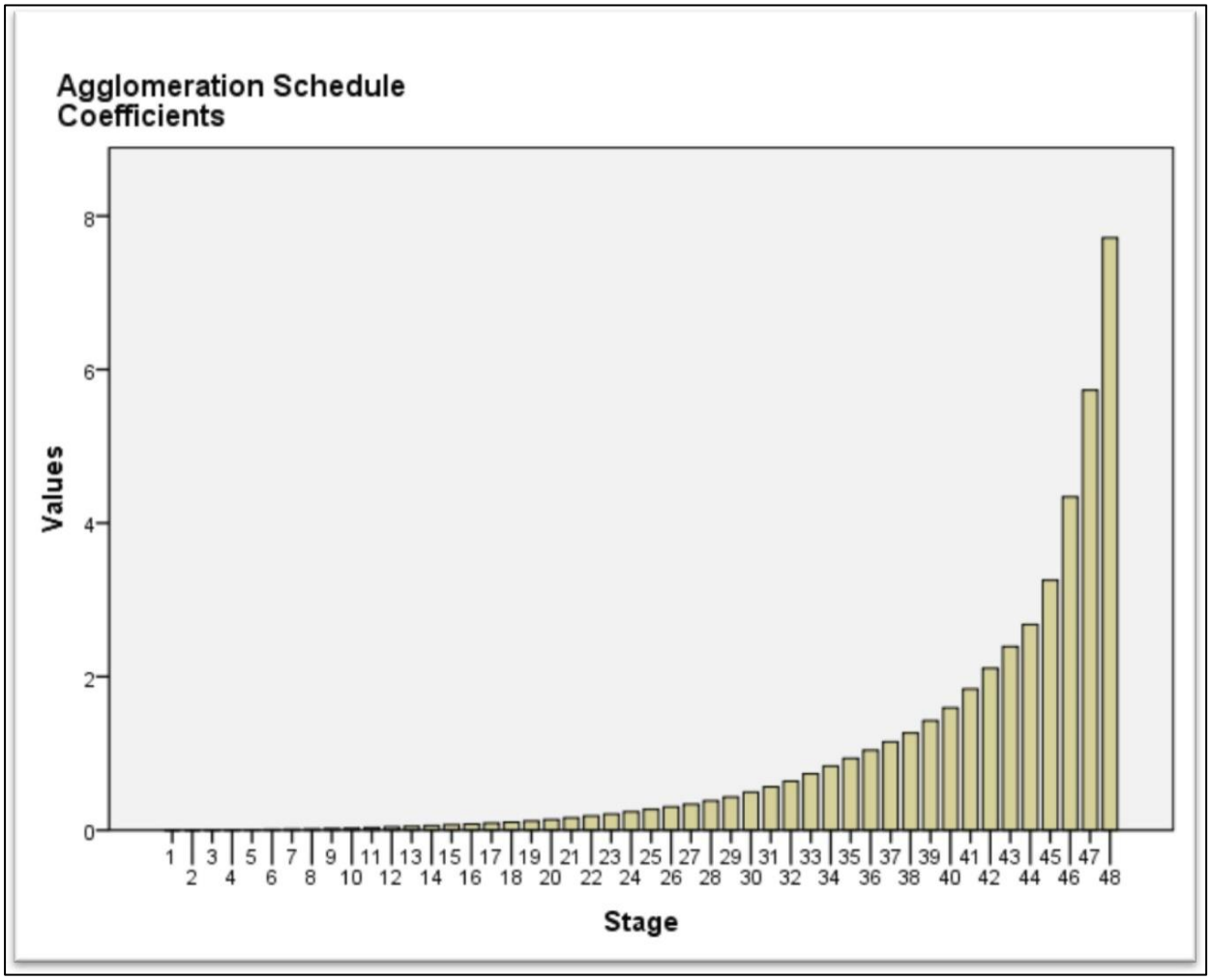

Figure 1. Agglomeration Schedule Coefficients

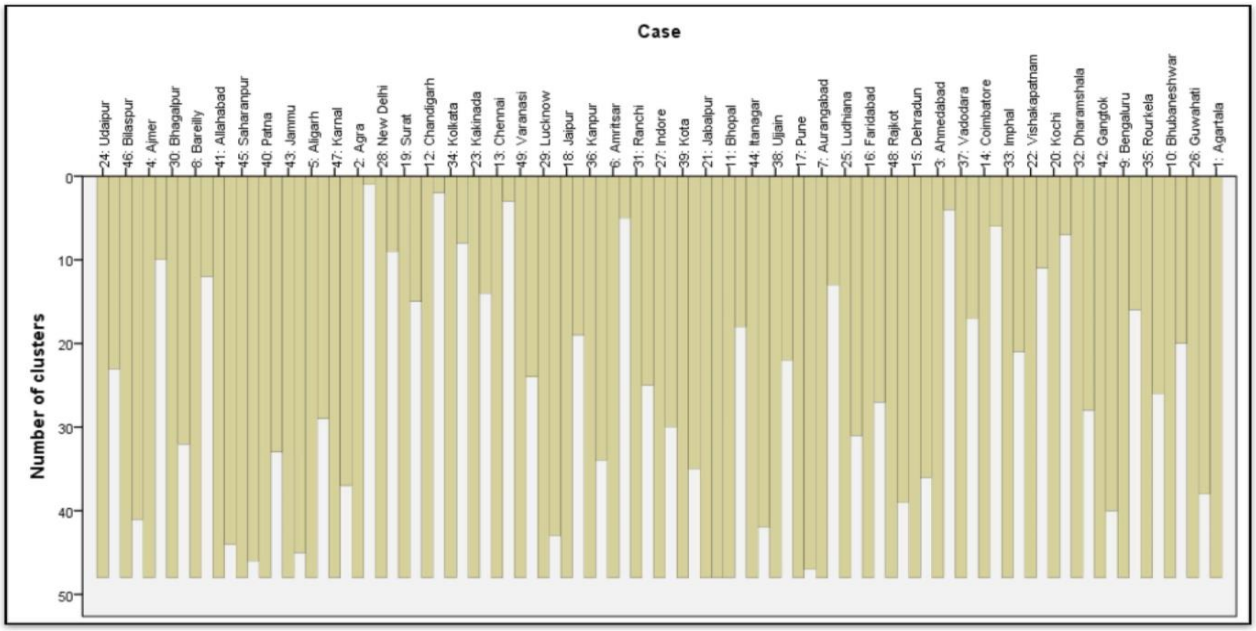

Figure 2: Number of Clusters

Always it is not an easy solution to calculate the difference of the coefficients from the agglomeration schedule. In Figure: 1 a visual representation is made which plots the coefficient values in the form of a bar. The agglomeration schedule is plotted in (Figure: 2) and displays the combination of two cases. The dark bars represent the cases and the point where the space between two cases becomes shaded represents that cases are joined together.

Retrieval Number: 100.1/ijsce.B811413020724
DOI: 10.35940/ijsce.B8114.14030724
Journal Website: www.ijsce.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Table IV. Membership of Clusters

Case	4 Clusters	3 Clusters	2 Clusters	Case	4 Clusters	3 Clusters	2 Clusters
1:Agartala	1	1	1	26:Guwahati	1	1	1
2:Agra	2	2	2	27:Indore	1	1	1
3:Ahmedabad	1	1	1	28:New Delhi	3	3	1
4:Ajmer	2	2	2	29:Lucknow	1	1	1
5:Aligarh	2	2	2	30:Bhagalpur	2	2	2
6:Amritsar	1	1	1	31:Ranchi	1	1	1
7:Aurangabad	1	1	1	32:Dharamshala	1	1	1
8:Bareilly	2	2	2	33:Imphal	1	1	1
9:Bengaluru	1	1	1	34:Kolkata	4	1	1
10:Bhubaneshwar	1	1	1	35:Rourkela	1	1	1
11:Bhopal	1	1	1	36:Kanpur	1	1	1
12:Chandigarh	3	3	1	37:Vadodara	1	1	1
13:Chennai	4	1	1	38:Ujjain	1	1	1
14:Coimbatore	1	1	1	39:Kota	1	1	1
15:Dehradun	1	1	1	40:Patna	2	2	2
16:Faridabad	1	1	1	41:Allahabad	2	2	2
17:Pune	1	1	1	42:Gangtok	1	1	1
18:Jaipur	1	1	1	43:Jammu	2	2	2
19:Surat	3	3	1	44:Itanagar	1	1	1
20:Kochi	1	1	1	45:Saharanpur	2	2	2
21:Jabalpur	1	1	1	46:Bilaspur	2	2	2
22:Vishakapatnam	1	1	1	47:Karnal	2	2	2
23:Kakinada	4	1	1	48:Rajkot	1	1	1
24:Udaipur	2	2	2	49:Varanasi	1	1	1
25:Ludhiana	1	1	1	_____	_____	_____	_____

In Table IV. it shows the cluster membership that is interconnected among each cluster. It keeps a record that is performing how many clusters depending on the coefficients described in agglomeration schedule coefficients.

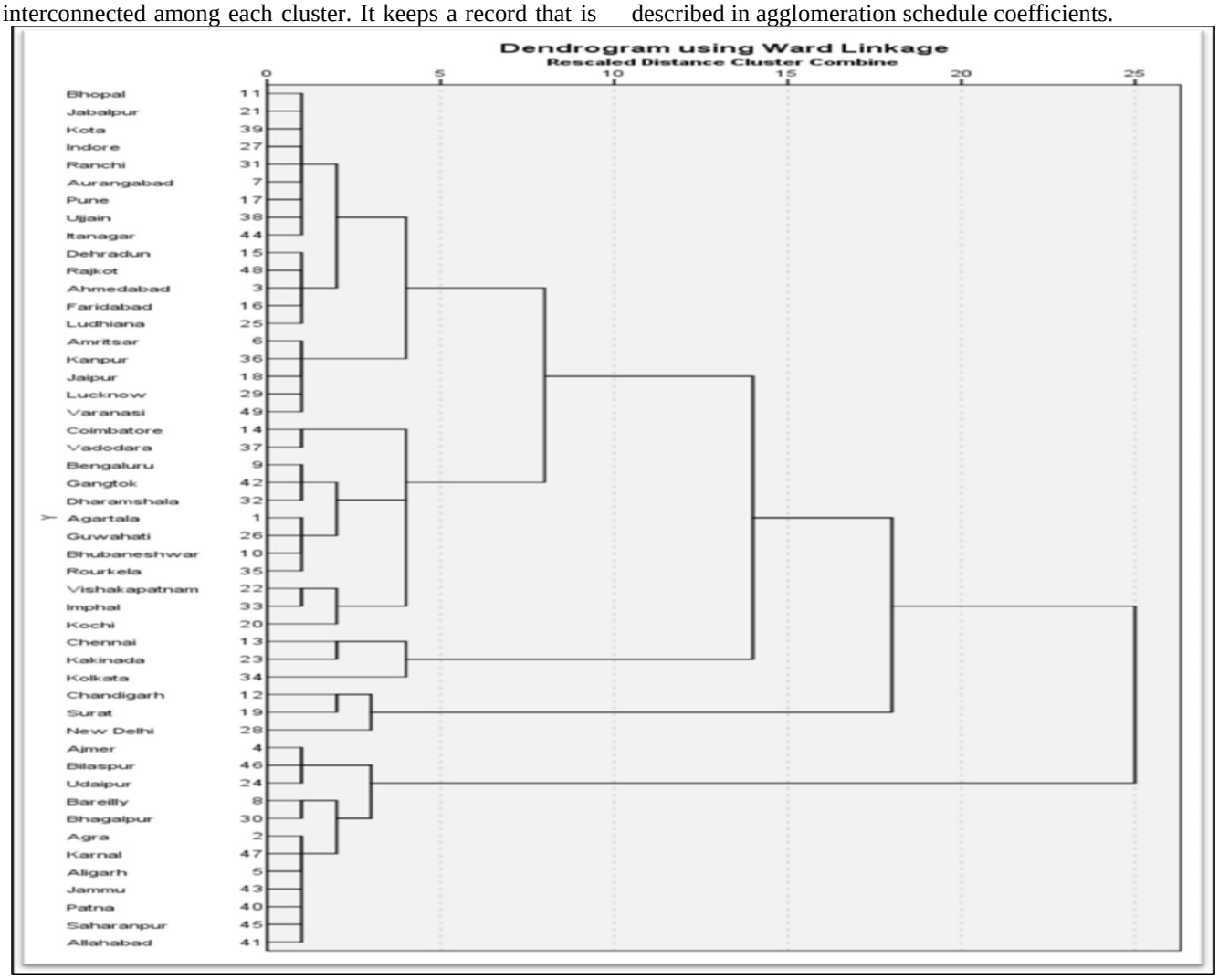

Figure: 3 Dendrogram of Selected Cities

In Figure 3 a Dendrogram of Selected Cities is created which illustrates the cluster analysis and gives a visual representation of the clustering process. It always appears at the end of the SPSS output. Understanding the dendrogram, from left to right the clusters which are more similar to each other are grouped at first. Vertical lines in the dendrogram indicate the stages of the agglomeration schedule and the distance between the two clusters.

VII. CONCLUSION

Over the last decade, the global debate on smart cities has been stimulating public interest. Researchers seek creative approaches to analyze cities and discover chances for developing smart solutions in the face of sweeping urbanization trends and growing complexities of urban concerns. This study used Key Performance Indicators (KPIs) to create urban typologies among 49 cities in India for demonstrating their readiness to become a smart city. To expose the status and changes in urban performance authors split the functions of cities into a few indicators and quantified those with the help of collected data. This is not determined by the present smartness of the city. Instead, authors focused on identifying critical characteristics and planned areas that required more attention on the variously capacitated cities in order to facilitate smart city transformation. The authors draw a typology of cities that depends on agglomerative hierarchical cluster analysis, which differs from the traditional ranking approach. In this study the essential determinants of urban prosperity and identified by discriminant analysis. The authors also used a dendrogram to make a typology of urban city classification.

VIII. RECOMMENDATIONS

A. This study will be helpful to recognize in an organized manner the diversity of challenges the cities are facing and evaluate whether the perspective view of adopting technology will be possible to justify whether those cities will be considered smart cities or not. The government can focus on optimizing the city functions and promote the growth in the economy by building the quality of life.

B. Smart cities must provide multi-specialty hospitality as well as smart healthcare using modern technology and focus on the medical ecosystem in an intelligent way.

C. A smart city enlarges the traditional industry and looks into creative people to create a path for promoting industrial development by creating a new paradigm for the upcoming generation.

IX. LIMITATION

A. Demand and Delivery: Only by attracting and retaining technical and artistic talent can a Smart City exist. Through the destruction of invention and innovation, these individuals are crucial in the continual rejuvenation of the economic infrastructure. They serve as the foundation for new campaigns, enterprises, and a climate conducive to technological advancement. As old activities fade away, expertise is required to assist in the development of new enterprises and occupations. As a result, major cities fight for this talent.

B. Social Unity, Inclusion, and Unity: While smart solutions have the potential to link individuals and enhance social cohesion, the concern is that the advantages of a smart city will not be shared equitably by all members of our society. There are three major causes behind this. • For starters, certain groups may not have access to current digital communications and equipment• Second, increased knowledge of risk in the emergence of big data sets might put unity under strain. • Finally, smart solutions can be utilized by groups to organize and build "digital gated communities," which can cause social cohesiveness and inclusion.

C. Resilience: Ability to anticipate and adapt to change conditions, as well as to endure and recover swiftly from disruptions caused by deliberate attacks, accidents, or natural dangers. Increased reliance on digital technology is one of the implications of digitalization. Serious societal and economic disruption might occur if important digital infrastructure fails or is harmed by a malicious attack. The challenge is to make smart solutions and digital infrastructure more resilient.

X. FUTURE SCOPE

A. Smart Travel: Smart solutions can be implemented to increase the usage of parking spaces, each parking place has a sensor that determines whether the parking place is free or not. It can offer the drivers real-time information on available free parking places near locations.

B. Smart Tourism & Leisure: Many of people can congregate in a small area during major city events. New technology can be utilized to obtain insight into crowd behavior and make recommendations for better crowd control strategies. The system may be arranged with the help of Wi-Fi, automatic census cameras, and GPS tracker sensors that can track the individual mobile of the driver. One hour ahead of time, advanced analytics is employed to anticipate the amount of local individuals in a certain area.

DECLARATION STATEMENT

In this research work authors used publicly available data across websites. This work does not contain any studies with human participants or animals performed by any of the authors. Further, the first author is attached to the Department of Business Administration, University of Kalyani, and 2nd Author doing a corporate job in performance io as an SEO executive consequently they used the infrastructure of their university and personal system to carry on this research activity. Consequently, there is no conflict of interest involved in this case. There is no internal or external funding agency to complete this research work.

Funding	No, I did not receive.
Conflicts of Interest	No conflicts of interest to the best of our knowledge.
Ethical Approval and Consent to Participate	No, the article does not require ethical approval and consent to participate with evidence.
Availability of Data and Material	Not relevant.
Authors Contributions	All authors have equal participation in this article.

REFERENCE

1. Mikelbank, B. A. (2004) 'A typology of U.S. suburban places', Housing Policy Debate, 15(4), doi: https://doi.org/10.1080/10511482.2004.9521527
2. Jones, K. J. and Jones, W. C. (1970) 'Toward A Typology of American Cities', Journal of Regional Science, 10(2), doi: https://doi.org/10.1111/j.1467-9787.1970.tb00047.x
3. Baum, S. et al. (2006) 'Advantage and disadvantage across Australia's extended metropolitan regions: A typology of socioeconomic outcomes', Urban Studies, 43(9), doi: https://doi.org/10.1080/00420980600831759
4. Hill, E. W., Brennan, J. F. and Wolman, H. L. (1998) 'What is a central city in the United States? Applying a statistical technique for developing taxonomies, Urban Studies, 35(11), doi: https://doi.org/10.1080/0042098983962
5. Clark, R. D. and Roche, J. P. (1984) 'Functional Typologies of Metropolitan Areas: An Examination of Their Usefulness', Urban Studies, 21(1), doi: https://doi.org/10.1080/00420988420080061
6. Giffinger, R. et al. (2007) Smart cities Ranking of European medium-sized cities, Vienna University of Technology. doi: https://doi.org/10.1016/S0264-2751(98)00050-X
7. IBM (2008) A Smarter Planet: The Next Leadership Agenda, Council on Foreign Relations. Council on Foreign Relations. Available at: https://www.cfr.org/event/smarter-planet-next-leadership-agenda
8. L. Zhuo, Q. Shi, C. Zhang, Q. Li, and H. Tao, ``Identifying building functions from the spatiotemporal population density and the interactions of people among buildings," *ISPRS Int. J. Geo-Inf.*, vol. 8, no. 6, p. 247, May 2019 https://doi.org/10.3390/ijgi8060247
9. Garmendia-Orbegozo. A, Noye.S, Anton. M.A and David Nuñez-Gonzalez. J (2022), Graph-Based Learning for Building Prediction in Smart Cities, IEEE Access, Digital Object Identifier 10.1109/ACCESS.2022.3169890 https://doi.org/10.1109/ACCESS.2022.3169890
10. Biswas, B., Bhadra, S., Sanyal, M.K., Das, S. (2018). Cloud Adoption: A Future Road Map for Indian SMEs. In: Bhateja, V., Coello Coello, C., Satapathy, S., Pattnaik, P. (eds) Intelligent Engineering Informatics. Advances in Intelligent Systems and Computing, vol 695. Springer, Singapore. https://doi.org/10.1007/978-981-10-7566-7_51
11. Biswas, B., Sanyal, M. K., & Mukherjee, T. (2021). Feedback Analysis for Digital Marketing in India: Empirical Study on Amazon.in, Flipkart, and Snapdeal. *International Journal of Online Marketing (IJOM), 11*(1), 78-88. http://doi.org/10.4018/IJOM.2021010105
12. Sanyal, M.K., Biswas, B., Roy, S., Bhadra, S. (2018). A Proposed Model to Integrate Business Intelligence System in Cloud Environment to Improve Business Function. In: Bhateja, V., Nguyen, B., Nguyen, N., Satapathy, S., Le, DN. (eds) Information Systems Design and Intelligent Applications. Advances in Intelligent Systems and Computing, vol 672. Springer, Singapore. https://doi.org/10.1007/978-981-10-7512-4_29
13. https://www.india.gov.in/spotlight/smart-cities-mission-step-towards-smart-india
14. https://smartcities.gov.in/
15. Sharma.S, Nanda.M, Goel.R, Jain.A, Bhushan.M, Kumar.A,(2019). Smart Cities using Internet of Things: Recent Trends and Techniques. In International Journal of Innovative Technology and Exploring Engineering (Vol. 8, Issue 9S, pp. 24–28). https://doi.org/10.35940/ijitee.i1004.0789s19
16. Sharma, K., & Tayal, S. (2019). Indian Smart City Ranking Model. In International Journal of Recent Technology and Engineering (IJRTE) (Vol. 8, Issue 2, pp. 4820–4832). https://doi.org/10.35940/ijrte.b2472.078219

AUTHORS PROFILE

Dr. Biswajit Biswas, *B. Tech, MBA, UGC-NET Ph.D* Assistant Professor in the Department of Business Administration, University of Kalyani. India. He has Seven years of research and teaching experience at postgraduate and Undergraduate levels in the domain of Business Administration. His expertise is in Business Intelligence, web data mining, Soft computing, Principle of Management, Strategic Management, and Digital Marketing. He has published many research articles in International journals and conferences. He also visited many prestigious academic Institutes in India and abroad. He has attached several teaching institutes including undergraduate & postgraduate levels.

Mr. Sayantan Ghosh is a Technical SEO Executive at Performance-io India, leveraging his expertise in the intersection of technology and marketing. With an MBA from the University of Kalyani specializing in Information Technology and a Bachelor's in Computer Science and Engineering from Chandigarh University, Mr. Ghosh has cultivated a deep understanding of both business administration and technical facets in the digital realm. Beyond his IT proficiency, Mr. Ghosh holds a profound fascination for the Stock Market, fostering an avid interest in analyzing market trends and predicting future trajectories.