K-Means Clustering algorithms in Urban studies: A Review of Unsupervised Machine Learning techniques

Bochra hadj kilani[1]

[1]University of Carthage ,Lab GADEV

January 16, 2024

Abstract

The use of unsupervised machine learning techniques, specifically K-means clustering algorithms, in urban studies has gained significant attention in recent years. These techniques have proven to be valuable in analyzing and understanding various aspects of urban design, such as land use patterns, transportation systems, and population distribution. This article aims to provide a comprehensive review of the application of K-means clustering algorithms in urban studies. The findings of this review demonstrate the wide range of applications of K-means clustering in urban studies, from identifying distinct land use categories to understanding the spatial distribution of social amenities. Furthermore, it is revealed that the use of K-means clustering in urban studies allows for the identification and characterization of hidden patterns and similarities among urban areas that might not be immediately apparent through traditional analysis methods. Overall, the use of K-means clustering algorithms provides a valuable tool for urban planners and researchers in gaining insights and making informed decisions in urban design.

keywords: k-means clustering,unsupervised Machine Learning ,urban studies

Introduction

In classification, unsupervised clustering algorithms are applied at early stages. These algorithms, such as k-means clustering, multiple kernel k-means clustering, and fuzzy c-means clustering, are used to classify pixels or produce pseudo-label training samples for subsequent classifiers. However, these existing clustering algorithms have significant disadvantages. Some limitations of the k-means clustering technique include difficulties in determining the value of k, sensitivity to the initial centroid value, and sensitivity to the size of the data (Na et al., 2010) . Another challenge with k-means clustering is its computational cost and scalability, particularly when dealing with large datasets. Additionally, existing clustering software for large datasets often relies heavily on methods designed for continuous data and specifically on k-means clustering (Vera and Macías, 2021) .However, these existing clustering algorithms have significant disadvantages. Some limitations of the k-means clustering technique include difficulties in determining the value of k, sensitivity to the initial centroid value, and sensitivity to the size of the data(SAPUTRA et al., 2020) . Another challenge with k-means clustering is its computational cost and scalability, particularly when dealing with large datasets.

Additionally, existing clustering software for large datasets often relies heavily on methods designed for continuous data and specifically on k-means clustering(Khandare and Alvi, 2016)

These limitations can pose challenges in the field of urban design, where accurate and efficient clustering techniques are crucial for analyzing and understanding complex spatial data. Unsupervised clustering algorithms, such as k-means clustering, can be valuable tools in urban design for analyzing and categorizing spatial data. However, it is important to consider the limitations of k-means clustering when applying it to urban design.(Zhang and Xia, 2009) .One such limitation is the difficulty in determining the optimal value of k, which represents the number of clusters to be formed. This determination is subjective and requires prior knowledge or expertise in the field of urban design. Another limitation is the impact of the initial centroid value on the final clustering result. The initial centroid value can influence the final clustering outcome, and choosing an inappropriate initial centroid value may result in suboptimal clustering results.

1-K-Means Clustering in Urban Design: An Overview

In the field of urban design, k-means clustering is commonly used as an unsupervised clustering algorithm to analyze and categorize spatial data. However, while k-means clustering can be a valuable tool, it is important to be aware of its limitations. These limitations include challenges in determining the optimal value of k, the sensitivity to the initial centroid value, and its computational cost and scalability. These limitations can affect the accuracy and effectiveness of k-means clustering in urban design applications.

Furthermore, existing clustering software for large datasets often relies heavily on methods designed for continuous data and specifically on k-means clustering, which may not be suitable. for mixed-type data commonly found in urban design. One solution to address these limitations is to explore alternative clustering algorithms that are better suited for mixed-type data and offer improved scalability. One such algorithm is the fuzzy k-means clustering technique, which allows data points to belong to multiple clusters with varying degrees of membership. This can provide more flexibility in capturing the complex relationships and patterns within urban design data. Furthermore, the Mahout project offers alternative clustering algorithms, such as the spectral clustering algorithm that involves running k-means on eigenvectors of the graph.(Dhillon et al., 2004)

Another approach is the spectral clustering algorithm, which involves running k-means on the eigenvectors of the graph Laplacian of the original data. This algorithm can handle both continuous and categorical data effectively and has the advantage of being able to handle mixed-type data effectively. In conclusion, k-means clustering has been widely used in urban design for spatial data analysis and categorization.(Lu et al., 2022) However, it is important to consider the limitations and challenges associated with k-means clustering in urban design applications. One of the main challenges is determining the optimal value of k, which represents the number of clusters. This decision can significantly influence the clustering outcome and may require trial and error or domain knowledge .the sensitivity of k-means clustering to the initial centroid value can lead to suboptimal clustering results. Choosing an inappropriate initial centroid value can cause the algorithm to converge to a local minimum instead of the global minimum, affecting the accuracy and effectiveness of the clustering. Additionally, the computational cost and scalability of k-means clustering can pose Unsupervised Machine Learning: Basic Concepts and Challenges limitations when dealing with large and complex urban design datasets.(Ran et al., 2021)

To overcome these limitations, alternative clustering techniques such as spectral clustering can be explored. Spectral clustering, unlike k-means clustering, is capable of handling mixed-type data effectively and offers improved scalability. Spectral clustering is useful when the clusters have a non-linear shape, and it can handle noisy data better than k-means. Another limitation of k-means clustering in urban design is its assumption that clusters are convex in shape. This assumption may not hold true for all urban design scenarios, as there may be non-convex clusters present. , while k-means clustering is a widely used method in urban design for its simplicity and interpretability, it is crucial to consider its limitations(Na et al., 2010)

2-The Role of K-Means in Unsupervised Machine Learning

K-means clustering plays a fundamental role in unsupervised machine learning, particularly in tasks such as data segmentation and pattern recognition. Its iterative distance-based approach allows for the automatic categorization of data into distinct groups based on their similarities. One of the key advantages of K-means clustering is its ability to handle large datasets efficiently. By partitioning the data into K clusters, the algorithm simplifies the analysis process and reduces the computational complexity. This scalability makes it a valuable tool in various real-world applications, including image segmentation and data mining. Moreover, K-means clustering is known for its simplicity and interpretability. The clearly defined clusters allow researchers and practitioners in urban design to easily identify and understand the distinct characteristics of different data points or observations. (Khandare and Alvi, 2016) This leads to valuable insights and informed decision-making in the field of urban design.

However, it is important to acknowledge the limitations of the K-means clustering algorithm in the context of urban design. One major drawback is its sensitivity to noise and discrete points. This means that even a small amount of outliers or irregular data points can significantly impact the clustering results, leading to inaccuracies in the analysis. Another limitation is the requirement to determine the hyperparameter K in advance. This can be challenging, especially when dealing with complex urban design datasets where the optimal number of clusters may not be obvious. Incorrectly specifying K can result in suboptimal clustering results and obscure the true patterns within the data. The initialization of the K-means algorithm is often random, which means that different initializations can lead to different cluster assignments. Therefore, researchers and practitioners need to be cautious in interpreting the results and should consider running the algorithm multiple times with different initializations to ensure the stability of the clustering solution.(Khandare and Alvi, 2016)

3-Applying K-Means Clustering in Urban Design

In the field of urban design, K-means clustering has proven to be a valuable methodology for analyzing and understanding various aspects of urban environments. For example, In the city of Barcelona, the city council extracts different urban typologies as shown on the map below, as basis for management decisions and planning for different sectors, in particular for waste management. They map different characteristics of the built environment such as the width of the street, the grid, and the number of households and portals along the roads, along with other parameters. With this method they have extracted 16 different urban typologies.(nei) The advantage of using K-means clustering in urban design is its ability to handle large datasets efficiently and provide concise results. K-means clustering allows researchers and practitioners in urban design to gain insights into the distinct characteristics of different elements or sectors within urban environments.

These insights can inform decision-making processes related to urban planning, transportation, and environmental sustainability. However, it is important to acknowledge the limitations of the K-means clustering algorithm in the context of urban design. These limitations include the challenge of determining the optimal number of clusters (K) and the sensitivity of the algorithm to initial centroid values(Zhang and Xia, 2009)

4-Case Studies of K-Means Clustering in Urban Planning: Bibiliometric analysis methodology

Bibliometric analysis is a rigorous and commonly used method for analyzing large sets of scientific data. It allows for a detailed examination of the evolutionary nuances within a specific field and highlights areas that are emerging. Several studies have applied K-means clustering in the field of urban planning to better understand different aspects of urban environments. Using the keywords "urban" and "Kmean, " we attempted to create a connection graph among authors who have worked on this topic. To achieve this, we utilized the "ResearchRabbit" app to extract a corpus of 50 articles. Our first step involved extracting the links and co-citations between authors. We then generated a timeline graph that highlights the first authors to have worked on this theme.

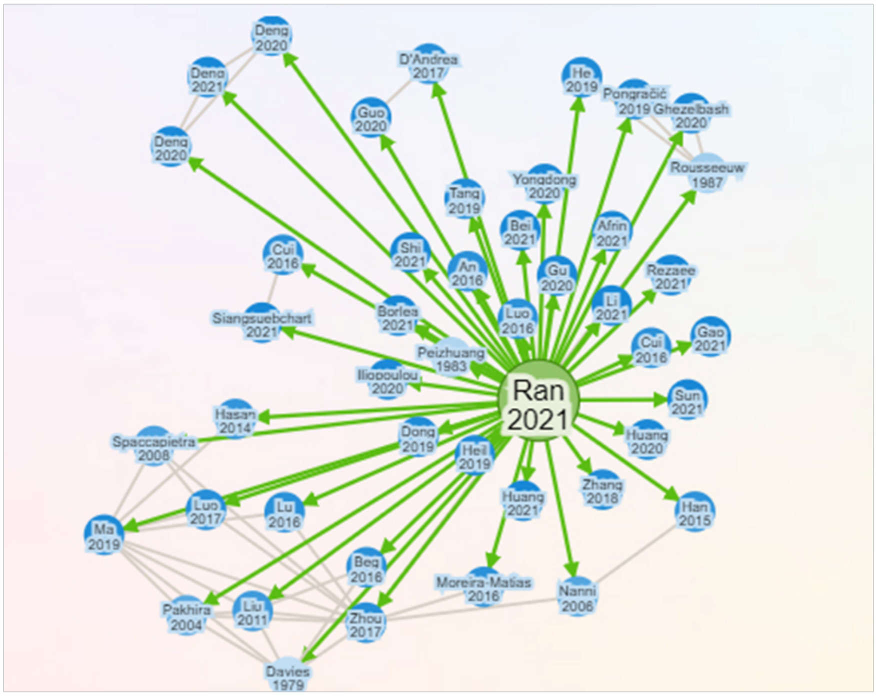

Figure 1: Network graph by first author (davies,1979)and (Ran,2021) © the author,2023

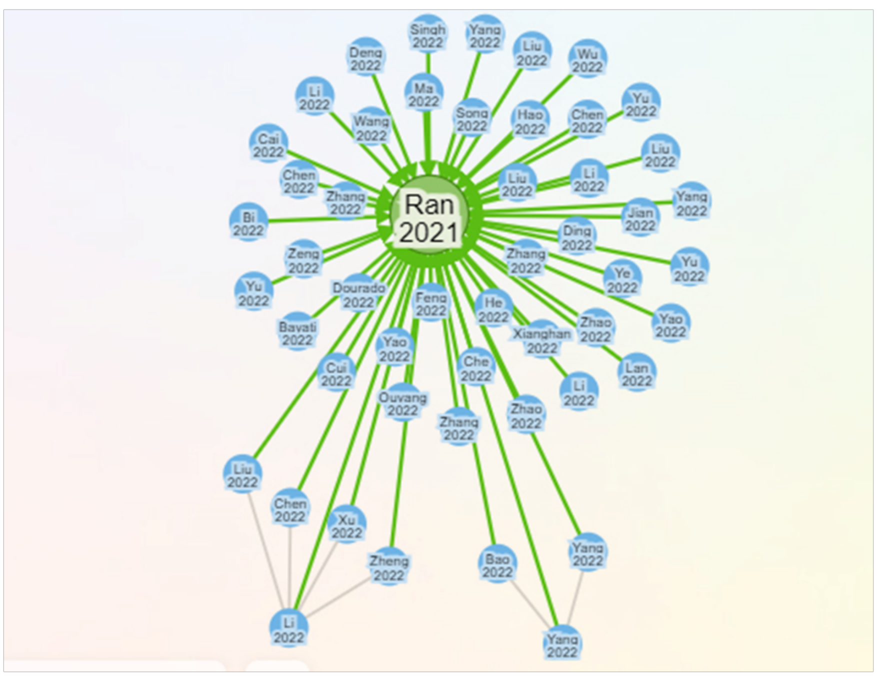

Figure 2: Network graph by last author (Ran,2021) citation © the author ,2023

We selected 9 articles for our study based on scientific discipline and number of citations.

Our focus was on the use of the K-means methodology in the urban environment, with a particular focus on urban planning, urban transport, and urban geography. These articles align well with our research theme.

These papers collectively discuss the application of K-means clustering in urban design. (Chang et al., 2017) The paper proposes a novel methodology by using Principle Component Analysis (PCA) and K-means clustering approach to find important features of the urban identity from public space. (Ran et al., 2021) introduces a novel K-means clustering algorithm with a noise algorithm to capture urban hotspots, addressing the challenges of determining the number of clusters and initializing the center cluster. (Duan et al., 2023) presents a hybrid heuristic initialization approach, combining a fuzzy system-particle swarm-genetic algorithm, to improve the initial clustering centers for K-means clustering in locating urban hotspots. Overall, these papers highlight the potential of K-means clustering as a valuable tool in urban design for identifying important features of urban identity and capturing urban hotspots.

Table 1: List of Articles

Authors	Article Title	Reference	Date
Ran, X., Zhou, X., Lei, M., Tepsan, W., & Deng, W.	A novel k-means clustering algorithm with a noise algorithm for capturing urban hotspots.	Applied Sciences, 11(23), 11202	2021
Chang, M. C., Bus, P., & Schmitt, G.	Feature extraction and k-means clustering approach to explore important features of urban identity	16th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 1139-1144)	2017
Jeong, J., So, M., & Hwang, H. Y	Selection of vertiports using k-means algorithm and noise analyses for urban air mobility (uam) in the Seoul metropolitan area	Applied Sciences, 11(12), 5729	2021
Zuo, C., Liang, C., Chen, J., Xi, R., & Zhang, J	Machine Learning-Based Urban Renovation Design for Improving Wind Environment: A Case Study in Xi'an, China.	China. Land, 12(4), 739	2023
İşeri, O. K., & Dino, İ. G	Building Archetype Characterization Using K-Means Clustering in Urban Building Energy Models	International Conference on Computer-Aided Architectural Design Futures (pp. 222-236). Singapore: Springer Singapore	2021
Liu, L., Peng, Z., Wu, H., Jiao, H., Yu, Y., & Zhao, J	Fast identification of urban sprawl based on K-means clustering with population density and local spatial entropy	Sustainability, 10(8), 2683	2018
Xu, H., Ma, C., Lian, J., Xu, K., & Chaima, E.	Urban flooding risk assessment based on an integrated k-means cluster algorithm and improved entropy weight method in the region of Haikou, China	Journal of hydrology, 563, 975-986	2018
Bhuyan, P. K., & Krishna Rao, K. V	Defining LOS criteria of urban streets using GPS data: k-means and k-medoid clustering in Indian context	Transport, 27(2), 149-157	2012
Yuhui, P., Yuan, Z., & Huibao, Y	Development of a representative driving cycle for urban buses based on the K-means cluster method	Cluster Computing, 22, 6871-6880	2019

(İşeri and Dino, 2022) describes implementing the k-means clustering algorithm in an Urban Building Energy Modeling (UBEM) framework to decrease the overall computational expenses of the simulation process. The work conducts two comparative analyses to assess the possibility of using the k-means clustering algorithm for UBEM. The k-means clustering algorithm's performance was evaluated by utilizing observations from the training data set with design parameters and performance goals. The second analysis examined prediction accuracy with various selection rates (5% and 10%) from clusters that the k-means clustering algorithm partitioned.

The study then comparatively analyzed the predicted and simulation-based calculated results of the selected observations. Analyses demonstrate that the k-means clustering algorithm can proficiently generate performance predictions with archetype characterization for Urban Building Energy Modeling (UBEM).

Keywords include archetype characterization, k-means clustering, and urban building energy modeling.(İşeri and Dino, 2022)

(Yuhui et al., 2018) developed a representative driving cycle for urban buses in Fuzhou city based on real-world driving data and the K-means clustering algorithm, (Xu et al., 2018)The study outcomes provide a novel approach for flood risk assessment and can provide valuable information for urban flood management.

(Liu et al., 2018)utilized K-means clustering with gridded population density and local spatial entropy. The results and comparison with open population data and mobile phone data confirm the assumption and indicate that the accuracy of source population data will constrain the precision of output identification. Urban sprawl is predominantly influenced by population and surrounding unevenness, as concluded in this article.

There are also some limitations in the results of some research , such as(Xu et al., 2018) the accuracy of classifying flood risk is limited by the availability of data and a more comprehensive index system should be developed in a future study. In the study, socio economic factors such as population density and gross domestic product density are not considered due to data constraints.

(Ran et al., 2021) amount of GPS data was too small to effectively reflect the distributions and the relationships of urban hotspots, and it was difficult to effectively avoid specific buildings in the city, even if the optimal results of urban hotspots are used in urban road planning. The text mentions several machine learning algorithms used to capture vehicle trajectory patterns, but it does not mention the number of algorithms.

5- Discussion:

5-1 The insights gained from the clustering analysis

The insights gained from the clustering analysis helped inform infrastructure improvements and traffic management strategies. Overall, K-means clustering is a powerful tool in urban design that allows for data-driven analysis and decision making.

In the field of urban design, one of the advantages of using k-means clustering is its ability to handle large datasets efficiently and provide concise results. By identifying distinct neighborhood types, analyzing land use patterns, and understanding vehicle movement patterns, urban planners can make informed decisions and interventions to improve the overall urban environment. The utilization of the k-means clustering algorithm in urban design allows for a comprehensive analysis of various factors such as population density, housing type, land use patterns, and CO2 emissions .

5-2 Challenges and Limitations of Using K-Means in Urban studies

While k-means clustering is a widely used and effective technique in urban design, it does have some challenges and limitations. One limitation of k-means clustering is that it assumes that the data points within a cluster are spherical and have the same variance. This assumption may not hold true in complex urban environments where there are diverse patterns and variations in the data. Another challenge is the determination of the optimal number of clusters. The k-means algorithm requires the user to specify the number of clusters beforehand, and selecting an inappropriate number of clusters can lead to biased and misleading results.

This decision is often subjective and may require trial and error or a priori knowledge of the data. Overall, while k-means clustering is a valuable tool in urban design, it is important for researchers and practitioners to be aware of its limitations and challenges and explore alternative clustering algorithms like "Fuzzy clustering techniques" or hybrid approaches that may better handle the complexities and variations in urban design data.(Sharma et al., 2023)

Another challenge lies in determining the optimal number of clusters. This decision is often subjective and requires careful consideration. Selecting an inappropriate number of clusters can result in misleading insights and hinder the understanding of urban design patterns. Additionally, the sensitivity of k-means clustering to the initial placement of cluster centroids can affect the clustering outcome and result in inconsistency or instability.

To address this sensitivity, researchers have proposed different techniques for initializing the centroids in k-means clustering. One popular method is the k-means++ initialization which aims to select the initial centroids in a way that promotes better convergence and reduces the likelihood of getting trapped in local optima.

By using k-means++ initialization, the algorithm iteratively selects centroids that are far apart from each other, resulting in a more representative initial placement. to overcome the limitations of k-means clustering when clusters are of different sizes, densities, or when outliers are present in the data, alternative clustering algorithms can be considered. One such algorithm is the DBSCAN.(Bao et al., 2023) DBSCAN is a density-based clustering algorithm that does not make assumptions about the shape and size of the clusters. This makes it more robust in handling complex urban design data with varying cluster characteristics. while k-means clustering is a widely used and effective tool in urban design analysis, it is crucial to acknowledge its limitations and challenges. By considering alternative clustering algorithms, such as DBSCAN, and utilizing techniques like k-means++ initialization, researchers can enhance the performance of k-means.

6-The Future of Unsupervised Learning in Urban Planning

The future of unsupervised learning in urban planning holds great potential for advancing our understanding of complex urban systems. By leveraging unsupervised learning algorithms, such as K-means clustering and DBSCAN, urban planners can extract valuable insights from large datasets and uncover hidden patterns and relationships within urban environments. With the growing availability of data from various sources, including social media, sensor networks, and public records, unsupervised learning techniques can help urban planners make evidence-based decisions and design more efficient and sustainable cities. These techniques can be applied to various urban planning tasks, such as identifying areas of high crime rates, analyzing transportation patterns, and understanding the spatial distribution of different land uses.

Moreover, advancements in machine learning and data science have led to the development of more sophisticated and specialized unsupervised learning algorithms tailored specifically for urban design and planning. These algorithms take into account the unique characteristics and challenges of urban data, such as heterogeneity, spatial dependencies, and temporal dynamics.

By incorporating these algorithms into urban planning workflows, planners can gain a deeper understanding of urban systems and make more informed decisions. The use of unsupervised learning algorithms in urban planning has the potential to revolutionize the field by providing valuable insights and aiding in evidence.

Topic	Keywords
Insights from Clustering Analysis	Infrastructure improvements, traffic management, data-driven analysis, concise results.
Challenges and Limitations	Spherical clusters assumption, optimal number determination, sensitivity to centroids.
Addressing K-Means Challenges	K-means++ initialization, alternative algorithms like DBSCAN.
Future of Unsupervised Learning	Advancing urban understanding, extracting insights from diverse datasets, specialized algorithms.

Table 2: Keywords for Various Topics

Conclusion:

Machine learning, natural language processing (NLP), and text mining play crucial roles in advancing urban research by enabling the extraction of valuable insights from vast amounts of textual data.(kilani, 2023)

While k-means clustering has been widely used in urban design analysis, it is crucial to acknowledge its limitations and challenges. To address these limitations, researchers are actively working on developing a more robust and efficient k-means algorithm that can handle complex urban data. One approach to enhancing k-means clustering is by integrating it with other clustering algorithms, such as DBSCAN or K-Distributions for Clustering Categorical Data.

By combining multiple clustering techniques, researchers can leverage the strengths of each algorithm and overcome the weaknesses of k-means clustering alone. Additionally, advancements in data mining and machine learning can also contribute to the development of newer types of clustering algorithms specifically designed for analyzing big data in urban planning. The fusion of unsupervised self-organizing neural networks with k-means clustering can also enhance the capabilities of urban design analysis.

This hybrid approach would leverage the unsupervised learning capabilities of self-organizing neural networks to automatically identify patterns and structures within urban data, while k-means clustering would provide a more interpretable and quantifiable representation of those patterns. In summary, the use of k-means clustering and other unsupervised learning algorithms in urban design offers a promising methodology for analyzing and understanding complex urban systems.

This review paper illustrates how the K-means clustering algorithm can assist urban designers in identifying urban identity characteristics of public spaces. The same clustering reveals public space features, signifying their significance. These features, related to spatial layout and place quality, inform urban planners to create more appropriate urban patterns.

References

Use of k-means clustering to extract urban typologies in neighborhoods. https://www.iaacblog.com/programs/k-means-clustering-for-urban-typologies/. URL https://www.iaacblog.com/programs/k-means-clustering-for-urban-typologies/. Accessed on Fri, December 01, 2023.

Lewen Bao, Zonglin Liu, Rui Miao, Zhihua Chen, Bo Zhang, Peng Guo, and Yuze Ma. Spatiotemporal clustering analysis of shared electric vehicles based on trajectory data for sustainable urban governance. *Journal of Cleaner Production*, 412, aug, pages=137373 2023. ISSN 0959-6526. doi: 10.1016/j.jclepro.2023.137373. URL http://dx.doi.org/10.1016/j.jclepro.2023.137373.

Mei-Chih Chang, Peter Bus, and Gerhard Schmitt. Feature Extraction and K-means Clustering Approach to Explore Important Features of Urban Identity. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, dec 2017. doi: 10.1109/icmla.2017.00015. URL http://dx.doi.org/10.1109/ICMLA.2017.00015.

Inderjit S. Dhillon, Yuqiang Guan, and Brian Kulis. Kernel k-means: spectral clustering and normalized cuts. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD04. ACM, aug, collection=KDD04 2004. doi: 10.1145/1014052.1014118. URL http://dx.doi.org/10.1145/1014052.1014118.

Suling Duan, Shanlin Jiang, Huan Dai, Luping Wang, and Zhenan He. The applications of hybrid approach combining exact method and evolutionary algorithm in combinatorial optimization. *Journal of Computational Design and Engineering*, 10(3), apr, pages=934–946 2023. ISSN 2288-5048. doi: 10.1093/jcde/qwad029. URL http://dx.doi.org/10.1093/jcde/qwad029.

Anand Khandare and A. S. Alvi. *Survey of Improved k-means Clustering Algorithms: Improvements, Shortcomings and Scope for Further Enhancement and Scalability*, page 495–503. Springer India, 2016. ISBN 9788132227526. doi: 10.1007/978-81-322-2752-6_48. URL http://dx.doi.org/10.1007/978-81-322-2752-6_48.

bochra hadj kilani. Méthodologie d'Approche du Text Mining et du NLP pour la Recherche Urbaine: revue de la littérature. *Center for Open Science*, nov 2023. doi: 10.31219/osf.io/bf8xs. URL http://dx.doi.org/10.31219/osf.io/bf8xs.

Lingbo Liu, Zhenghong Peng, Hao Wu, Hongzan Jiao, Yang Yu, and Jie Zhao. Fast Identification of Urban Sprawl Based on K-Means Clustering with Population Density and Local Spatial Entropy. *Sustainability*, 10(8), jul, pages=2683 2018. ISSN 2071-1050. doi: 10.3390/su10082683. URL http://dx.doi.org/10.3390/su10082683.

Xinhai Lu, Zhenxing Shi, Jia Li, Junhao Dong, Mingjie Song, and Jiao Hou. Research on the Impact of Factor Flow on Urban Land Use Efficiency from the Perspective of Urbanization. *Land*, 11(3), mar, pages=389 2022. ISSN 2073-445X. doi: 10.3390/land11030389. URL http://dx.doi.org/10.3390/land11030389.

Shi Na, Liu Xumin, and Guan Yong. Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm. In *2010 Third International Symposium on Intelligent Information Technology and Security Informatics*. IEEE, apr 2010. doi: 10.1109/iitsi.2010.74. URL http://dx.doi.org/10.1109/IITSI.2010.74.

Xiaojuan Ran, Xiangbing Zhou, Mu Lei, Worawit Tepsan, and Wu Deng. A Novel K-Means Clustering Algorithm with a Noise Algorithm for Capturing Urban Hotspots. *Applied Sciences*, 11(23), nov, pages=11202 2021. ISSN 2076-3417. doi: 10.3390/app112311202. URL http://dx.doi.org/10.3390/app112311202.

Danny Matthew SAPUTRA, Daniel SAPUTRA, and Liniyanti D. OSWARI. Effect of Distance Metrics in Determining K-Value in K-Means Clustering Using Elbow and Silhouette Method. In *Proceedings of the Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019)*, siconian-19. Atlantis Press, 2020. doi: 10.2991/aisr.k.200424.051. URL http://dx.doi.org/10.2991/aisr.k.200424.051.

Akash Sharma, Sunil K. Singh, Eshita Badwal, Sudhakar Kumar, Brij B. Gupta, Varsha Arya, Kwok Tai Chui, and Domenico Santaniello. Fuzzy Based Clustering of Consumers' Big Data in Industrial Applications. In *2023 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, jan 2023. doi: 10.1109/icce56470.2023.10043451. URL http://dx.doi.org/10.1109/ICCE56470.2023.10043451.

J. Fernando Vera and Rodrigo Macías. On the Behaviour of K-Means Clustering of a Dissimilarity Matrix by Means of Full Multidimensional Scaling. *Psychometrika*, 86(2), may, pages=489–513 2021. ISSN 1860-0980. doi: 10.1007/s11336-021-09757-2. URL http://dx.doi.org/10.1007/s11336-021-09757-2.

Hongshi Xu, Chao Ma, Jijian Lian, Kui Xu, and Evance Chaima. Urban flooding risk assessment based on an integrated k-means cluster algorithm and improved entropy weight method in the region of Haikou, China. *Journal of Hydrology*, 563, aug, pages=975–986 2018. ISSN 0022-1694. doi: 10.1016/j.jhydrol.2018.06.060. URL http://dx.doi.org/10.1016/j.jhydrol.2018.06.060.

Peng Yuhui, Zhuang Yuan, and Yang Huibao. Development of a representative driving cycle for urban buses based on the K-means cluster method. *Cluster Computing*, 22(S3), jan, pages=6871–6880 2018. ISSN 1573-7543. doi: 10.1007/s10586-017-1673-y. URL http://dx.doi.org/10.1007/s10586-017-1673-y.

Chen Zhang and Shixiong Xia. K-means Clustering Algorithm with Improved Initial Center. In *2009 Second International Workshop on Knowledge Discovery and Data Mining*. IEEE, jan 2009. doi: 10.1109/wkdd.2009.210. URL http://dx.doi.org/10.1109/WKDD.2009.210.

Orçun Koral İşeri and İpek Gürsel Dino. *Building Archetype Characterization Using K-Means Clustering in Urban Building Energy Models*, page 222–236. Springer Singapore, 2022. ISBN 9789811912801. doi: 10.1007/978-981-19-1280-1_14. URL http://dx.doi.org/10.1007/978-981-19-1280-1_14.