A Data-Driven Exploration of Urban Systems: Unveiling the Mysteries of City Sustainability

Xiaolei Tian[a] , Jianwei Ren[b*]

Transportation Institute, Inner Mongolia University, Hohhot 010070, China

[a]1750143492@qq.com; [b*]rjw@imu.edu.cn

Abstract. In the face of the dual challenges of an aging population, negative growth, and the increasing frequency of extreme climate events, urban sustainability has become a key concern in contemporary development strategies. This study proposes a comprehensive assessment framework to tackle these challenges, utilizing the TOPSIS-Entropy weighting method to evaluate urban systems across three critical dimensions: housing, service provision, and resilience. Through careful data preprocessing, precise model construction, and multi-dimensional analysis, the framework identifies vulnerabilities within urban systems and provides actionable insights into enhancing sustainability. By comparing two representative cities, the study reveals key disparities in infrastructure, service distribution, and resilience capacities, offering a deeper understanding of the factors that influence sustainable urban development. The findings provide valuable guidance for policymakers, urban planners, and stakeholders, offering actionable recommendations to optimize resource allocation, improve service accessibility, and enhance cities' resilience to external shocks. This research underscores the crucial role of data-driven approaches in shaping long-term urban planning and promoting sustainable development.

Keywords: Urban sustainability, TOPSIS-Entropy method, Decision tree regression, resilience assessment, entropy weight method, urban development.

1 Introduction

China faces demographic challenges such as an aging population and negative growth, compounded by frequent extreme climate events and a global economic downturn. These factors challenge urban resilience and highlight the need for advanced data-based techniques to assess urban sustainability[1].

While previous studies have made valuable contributions, they have limitations. Bartke focused on housing prices but overlooked the complex interactions between housing and other urban factors, leading to an incomplete understanding of urban sustainability[2]. Quintal et al. examined urban service levels but failed to account for variations in service requirements across regions and demographic groups, limiting the generalizability of their findings[3]. Samarakkody et al. focused on urban resilience but neglected the dynamic nature of urban systems and the influence of external factors like

G. Zhao et al. (eds.), *Proceedings of the 2025 8th International Conference on Traffic Transportation and Civil Architecture (ICTTCA 2025)*, Atlantis Highlights in Engineering 39,
https://doi.org/10.2991/978-94-6463-793-9_46

technological advancements and policy changes[4]. Tang's research explored housing and services independently, disregarding their interdependencies[5].

This study addresses these gaps by employing a comprehensive framework that integrates the TOPSIS-Entropy weighting method. This framework evaluates multiple factors while considering the unique characteristics of different cities. By incorporating advanced data processing and modeling techniques, it enhances the accuracy and reliability of the evaluation, providing more targeted and actionable recommendations for urban development.

2 Methodology

2.1 Data Preparation, Feature Engineering and Model Selection

Data was sourced from platforms such as Aliyun Tianchi and Google Dataset. Missing values were handled through deletion or imputation, and outliers were treated using the IQR method[6]. Key features impacting housing prices were classified as primary (e.g., total households, greening rate) or secondary (e.g., property management fee). Features were encoded using One-Hot Encoding[7]. The study focused on two cities: City 1 (Changchun) and City 2 (Hohhot).

2.2 Housing Stock Estimation Model

The total housing stock for each neighborhood was estimated using the formula:

$$H = T \times \frac{1}{F} \times P \tag{1}$$

where T is the total number of households, F is the Floor Area Ratio and P is the Parking Ratio. Summing up neighborhood values gave the total city housing stock.

2.3 Urban Service Level Quantitative Model

Service-related data was categorized into major groups based on functional attributes, including basic living services, business and financial services, government and public services, educational and cultural services, medical and health services, transport and travel services, and other services. Facility density was calculated to quantify the spatial distribution of facilities in each area:

$$D = \frac{N}{A} \tag{2}$$

where N represents the number of facilities in a specific category, and A is the area of the region:$S_A = \pi \cdot r^2$. The service coverage area was calculated based on the service radius of each facility type to evaluate the coverage capacity. The Gini coefficient was employed to measure the equilibrium of facility distribution[8]:

$$G = 1 - \sum(f_i \cdot L_i) \tag{3}$$

where f_i is the cumulative facility share, and L_i is the cumulative regional area share.

2.4 Resilience and Sustainability Evaluation Model

Data related to infrastructure and services was collected, and key indicators, including infrastructure coverage, facility density, service diversity, and distribution balance, were identified. Infrastructure coverage was determined as:

$$CoverageRate = \frac{A_{covered}}{A_{city}} \times 100\% \tag{4}$$

where $A_{covered}$ represents the area covered by facilities, and A_{city} represents the total city area. Facility density was calculated as:

$$FacilityDensity = \frac{N_{facilities}}{A_{city}} \tag{5}$$

where $N_{facilities}$ represents the number of facility points.

Data was standardized to normalize indicators. The entropy weight method was used to calculate the weights of each indicator. First, the weight of the i-th sample under the j-th indicator was calculated as:

$$p_{ij} = \frac{x'_{ij}}{\Sigma\left(x'_{ij}\right)} \tag{6}$$

The information entropy e_j and the weight of the j-th indicator w_j were computed as:

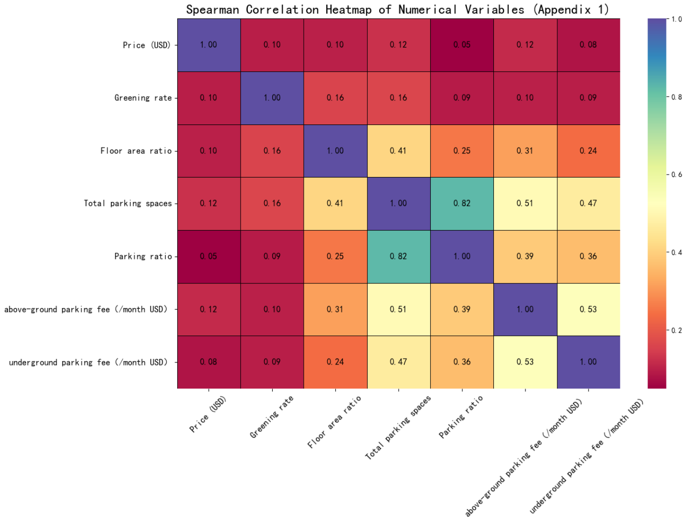
$$e_j = -k \sum \left(p_{ij} \cdot ln\left(p_{ij}\right)\right), w_j = 1 - e_j \tag{7}$$

Finally, the TOPSIS model ranked cities based on their sustainable development performance[9].

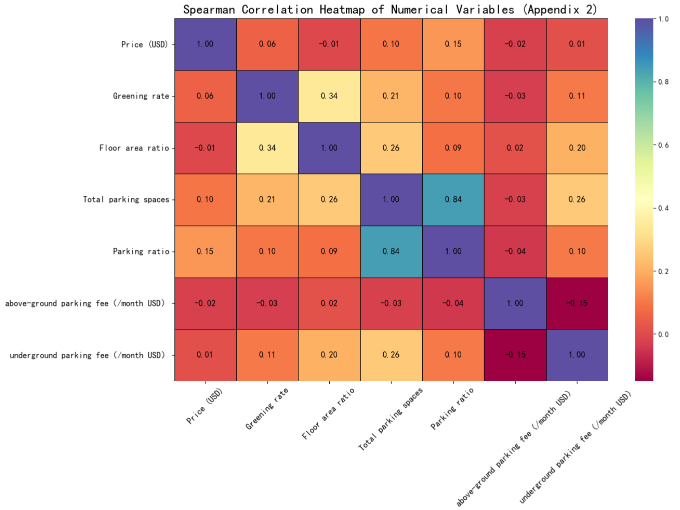
3 Results and Discussing

3.1 House Price and Stock Forecast Model

The Spearman correlation coefficient analysis showed that “Price (USD)” had a weak positive correlation with “Total parking spaces” and “Aboveground parking fee”, while “Parking ratio” and “Total parking spaces” had a strong correlation, as show in Fig 1.



(a) Housing Prices and Key Variables (Appendix 1)



(b) Housing Prices and Key Variables (Appendix 2)

Fig. 1. Housing Prices and Key Variables

Multiple regression models were applied and compared using the R^2 metric on training and test datasets. In both cities, the Decision Tree model outperformed others. In City 1 and City 2, the Decision Tree model achieved training set scores of 0.902349 and 0.863127 respectively. Before parameterization, the test set scores were 0.705278 for City 1 and 0.692811 for City 2, and after parameterization, the test set scores were 0.872 for City 1 and 0.879878 for City.

3.2 City Service Benchmarks: A Dual-City Study

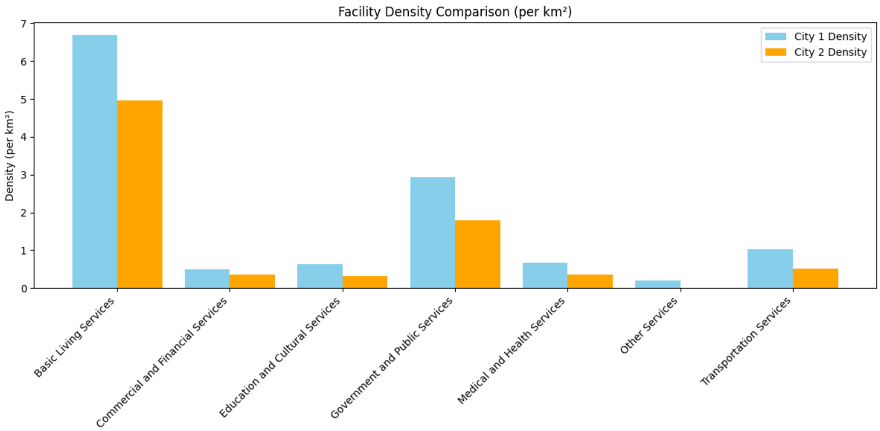


Fig. 2. Facility Density Comparison

City 1 had a higher service density than City 2 in most categories, especially in basic living and government public services, as show in Fig 2. However, Figure 3 reveals that in some service categories, like business and financial, education and cultural, and transport services in City 2, the distribution of service facilities was uneven, with high Gini coefficients, suggesting a need for better distribution balance.

City 1 had more resources, especially in basic living, medical, and transport services, with higher facility density. Its central layout efficiently served the population, but peripheral areas faced shortages, as indicated by low density and high Gini coefficients. City 2 had broader coverage but lower resource density, particularly in basic and medical services. This dispersed distribution struggled to meet high-density population needs, resulting in less efficient service, as show in Fig 4.

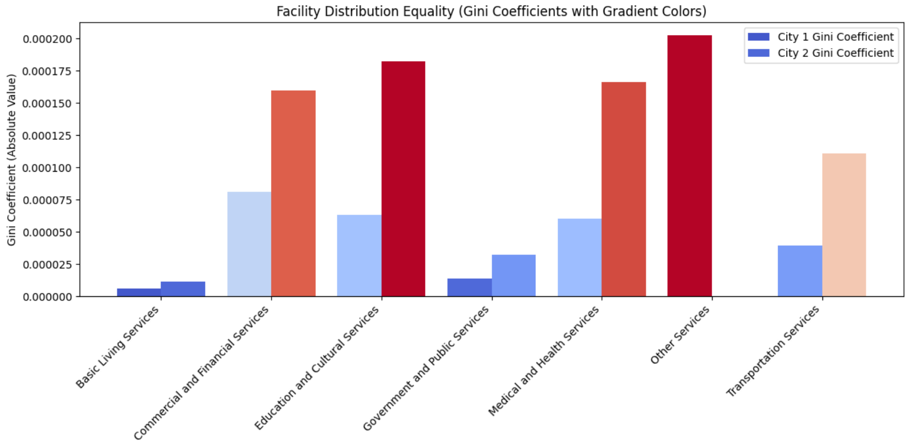
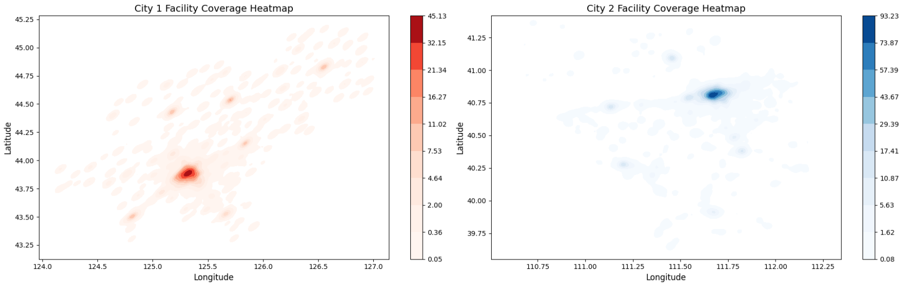


Fig. 3. Facility Distribution Equality



(a) City 1 Facility Coverage Heatmap (b) City 2 Facility Coverage Heatmap

Fig. 4. Facility Coverage Heatmap

3.3 Resilience and Sustainable Development Capacity Evaluation

In the resilience assessment model, infrastructure coverage, with a weight of 0.7464, is the dominant factor, while indicators such as service diversity and distribution balance have relatively lower weights. The TOPSIS evaluation shows that City 1 performs better in both extreme weather response and overall sustainable development, scoring 0.8923, compared to City 2's 0.8356. This indicates that City 1 has more optimized resource allocation and better facility development.

Regarding sustainable development capacity, City 1 is rich in resources for basic living, medical, and transportation services, with a high facility density. The layout in the central area can efficiently serve the concentrated population. However, peripheral areas suffer from resource shortages, as evidenced by low facility density and high Gini coefficients. City 2 has a wide geographical coverage but a low resource density, especially in basic and medical services. This leads to low service efficiency when meeting the needs of a high - density population.

Based on the results, City 1 should optimize resource allocation in its peripheral areas by adding public facilities. City 2 needs to enhance infrastructure and service facility density, especially in basic living and medical fields, and optimize layouts.

4 Conclusion

This study developed a framework to assess urban sustainability and resilience, analyzing two cities' housing, service levels, and resilience. Results showed City 1's overall sustainability advantages, while City 2 required improved infrastructure and resource allocation. Limitations include the small sample size of two cities, which may not represent all urban types, and the model's lack of dynamic data integration, affecting its accuracy. Future research should expand the sample to include diverse cities and integrate dynamic data like population flow and real-time policies to create a more accurate, time-sensitive evaluation model for urban development.

Acknowledgements

This work was supported by Innovation and Entrepreneurship Projects for College Students at Inner Mongolia University (202420462).

References

1. M. Zhang, Y. Yang, H. Li, and M. P. van Dijk, “Measuring urban resilience to climate change in three Chinese cities,” Sustainability, vol. 12, no. 22, p. 9735, 2020.
2. S. Bartke and N. Schwarz, “Urban green spaces and housing prices: An alternative perspective,” Sustainability, vol. 11, no. 13, p. 3707, 2019.
3. C. Quintal, M. Fuentes, R. Boudreau, et al., “Rural-urban and regional variations in aspects of caregiving, support services, and caregiver health in the USA: Evidence from a national survey,” BMJ Open, vol. 10, no. 9, p. e037383, 2020.
4. A. Samarakkody, D. Amaratunga, and R. Haigh, “Technological innovations for enhancing disaster resilience in smart cities: A comprehensive urban scholar’s analysis,” Sustainability, vol. 15, no. 15, p. 12036, 2023.
5. D. Tang, “Urban sustainability: Sustainable housing and communities,” Buildings, vol. 13, no. 7, p. 1732, 2023.
6. Z. Yuan and R. Hu, “Understanding the effects of influential factors on housing prices by combining extreme gradient boosting and a hedonic price model,” Land, vol. 10, no. 5, p. 533, 2021.
7. Train in Data, “One-hot encoding categorical variables,” Train in Data Blog, Published November 2022.
8. D. Zhang, G. Zhang, and C. Zhou, “Differences in accessibility of public health facilities in hierarchical municipalities and the spatial pattern characteristics of their services in Doumen District, China,” Land, vol. 10, no. 11, p. 1249, 2021.
9. H. He, Z. He, F. Gao, et al., “A comprehensive entropy-based evaluation of industrial sustainable development: An application to China,” PLoS ONE, vol. 15, no. 12, p. e0243586, 2020.