International Journal of
Geo-Information

Article

Prediction of Commercial Street Location Based on Point of Interest (POI) Big Data and Machine Learning

Linghan Yao [1], Chao Gao [2,*], Yanqing Xu [3,4], Xinyue Zhang [5], Xiaoyi Wang [6] and Yequan Hu [5]

1 School of Architecture, South China University of Technology, Guangzhou 510641, China; 202221006096@mail.scut.edu.cn
2 School of Humanities, Chang'an University, Xi'an 710061, China
3 School of Architectural Science and Engineering, Yangzhou University, Yangzhou 225009, China; xuyq@yzu.edu.cn
4 School of Architecture, Southeast University, Nanjing 210096, China
5 Bartlett School of Architecture, University College London, London WC1E 6BT, UK; ucbvxz5@ucl.ac.uk (X.Z.); zczqhuj@ucl.ac.uk (Y.H.)
6 Aerospace Information Research Institute, Chinese Academy of Sciences, Suzhou 215000, China; wangxiaoyi@aircas.ac.cn
* Correspondence: 2018023004@chd.edu.cn; Tel.: +86-18202921671

Abstract: Identifying optimal locations for sustainable commercial street development is crucial for driving economic growth and enhancing social vitality in cities. This study proposes a data-driven approach to predict potential sites for commercial streets in Foshan City, China, utilizing Points of Interest (POI) big data and machine learning techniques. Decision tree algorithms are employed to quantitatively assess and predict optimal locations at a fine-grained spatial resolution, dividing the study area into 9808 grid cells. The analysis identifies 2157 grid cells as potential sites for commercial street development, highlighting the significant influence of Medical Care, Shopping, and Recreation and Entertainment POIs on site selection. The study underscores the importance of considering population base, human activity patterns, and cultural elements in sustainable urban development. The main contributions include providing a novel decision-support method for data-driven and sustainable commercial street site selection and offering insights into the complex interplay between urban land use, human activities, and commercial development. The findings have important implications for urban planning and policy-making, showcasing the potential of data-driven approaches in guiding sustainable urban development and fostering vibrant commercial areas.

Keywords: machine learning; decision tree; sustainable urban development; spatial analysis; urban planning; site selection; commercial street prediction

check for updates

Citation: Yao, L.; Gao, C.; Xu, Y.; Zhang, X.; Wang, X.; Hu, Y. Prediction of Commercial Street Location Based on Point of Interest (POI) Big Data and Machine Learning. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 371. https://doi.org/10.3390/ijgi13100371

Academic Editor: Wolfgang Kainz

Received: 30 August 2024
Revised: 14 October 2024
Accepted: 18 October 2024
Published: 21 October 2024

1. Introduction

"Commercial streets" refers to streets or areas in urban areas primarily designated for commercial activities. Commercial streets, as a vital form of urban commercial space, play an irreplaceable role in meeting residents' diverse needs for consumption, leisure, entertainment, and social interaction, while enhancing the city's commercial image, promoting urban economic development, and creating a vibrant urban environment [1,2]. In this study, "sustainable commercial street development" refers to an efficient strategy for commercial street growth, emphasizing scientific and rational site selection and layout to maximize market demand and consumer experience. It focuses on optimizing resource utilization and selecting the best locations through data analysis, ultimately achieving sustainable growth of commercial activities and promoting regional economic prosperity [3,4]. In the context of China's reform and opening-up policy and rapid urbanization, the construction of commercial streets has undergone a period of vigorous development [5,6]. However, in the development process, some commercial streets have encountered issues such as large-scale expansion, hasty construction, and homogenization, resulting in low foot traffic, high vacancy rates, and poor economic performance. These problems not only run counter to the original intentions of the government and developers but also result in a waste of valuable urban resources [7]. To avoid these issues from the outset, it is imperative to strengthen preliminary research on sustainable commercial street site selection, which plays a crucial role in fostering urban regeneration. Moreover, the spatial structure of commercial activities has been an essential research topic in urban geography and urban planning, and its rational layout is conducive to the efficient allocation of urban resources and the development of the urban economy. Thus, reducing the waste of urban resources and promoting sustainable urban development. With the acceleration of urbanization and the increasing diversification of consumer demands, the location decision of commercial streets, as an important carrier of urban economic activities, has become more and more complex and critical. Reasonable site selection can not only enhance the competitiveness of commercial facilities, but also optimize the allocation of urban resources and promote regional economic development. However, traditional site selection methods rely on empirical judgment and qualitative analysis, which are difficult to comprehensively and accurately reflect the complex relationship between commercial facilities and the surrounding environment.

Scholars around the world have long been studying commercial streets, constantly exploring their multidimensional characteristics, such as spatial layout, consumer behavior, urban vitality, and store-pedestrian interaction. Christaller proposed the central place theory, providing a theoretical foundation for the spatial layout of commercial streets [8]. In the initial stage of research, numerous scholars studied commercial spaces based on behavioral geography theory, focusing on consumer behavior [9,10]. Jacobs and Gehl explored the relationship between commercial streets and human behavioral activities from the perspective of urban vitality [11,12]. Takahashi utilized field survey data to reveal the correlation between shop density and pedestrian behavior [13].

Since the beginning of the 21st century, with the development of big data and machine learning technologies, commercial street research has encountered new opportunities, with an increase in empirical studies and continuous enrichment and innovation in research methods and perspectives. Sung employed multi-source data to quantitatively evaluate the spatial vitality of commercial streets [14]. Mehanna focused on historic shopping districts and explored strategies for the protection and renewal of traditional commercial spaces [13]. Ghazi characterized the key elements influencing the vitality of shopping districts from multiple dimensions [15]. Kickert using research methods such as independent regression models, pooled cross-section analysis, and spatial autoregressive probit models, found that the important factors affecting retail business closures in the United States, United Kingdom, and Netherlands were agglomeration, centrality, and connectivity [16]. Feilong Hao et al. analyze POIs in Changchun City to reveal the polycentric and stratified trends in retail shop location patterns and their interactions with factors, such as consumer behavior, retail formats, and governmental planning, which provide a basis for urban planning and the allocation of commercial facilities [17]. Colaco Rui et al. propose a new cluster analysis and a small number of variables based on a new commercial classification system, thus strengthening the link between commercial classification and location modeling [18]. Colaco Rui et al. also used cellular automata (CA), incorporating centrality indicators in spatial syntax as drivers, to simulate commercial land use changes in order to better understand future land use dynamics [19].

Among the studies targeting site selection, empirical location models, such as the MCI model [20], Geomarketing [21], Analytic Hierarchy Process model [22], etc., are effective for small-scale site selection but struggle with complex environments due to increased data and computational demands. To address these challenges, machine learning is being integrated to enhance the accuracy and efficiency of site-selection processes for larger scales. Yuxue Wang and colleagues develop a site-selection model for outdoor digital signage that combines spatial characteristics with machine learning techniques, enhancing placement accuracy and efficiency, as demonstrated in Beijing's Sixth Ring Road area [23].

Luyao Wang and colleagues propose a two-step hybrid model for site selection of small retail shops in China, combining spatial accessibility evaluation with market potential estimation to identify high-value locations with low competition, as demonstrated in their study conducted in Guiyang [24]. An increasing number of scholars are beginning to use machine learning methods, such as decision trees and neural networks, to address large-scale location issues [23–27].

Although existing studies have achieved rich results in the spatial layout, commercial structure, vitality shaping, and spatial distribution of commercial streets, there are still some limitations in site selection prediction. Traditional commercial street siting research mainly relies on qualitative analysis, lacking data support and quantitative methods, which makes it difficult to adapt to the increasingly complex urban planning and commercial development needs. Existing location studies mostly focus on analyzing the spatial distribution of commercial street locations, with little attention paid to predictive simulations. There is a lack of large-scale location prediction research based on machine learning.

With the rise of point-of-interest (POI) big data and the advancement of machine learning technology, the predictive simulation study of commercial street siting has gained new perspectives and methods [27,28]. On the one hand, massive POI data provides an important data source for characterizing the location characteristics of commercial streets; on the other hand, machine learning algorithms, especially decision tree models, with their superior classification and prediction performance, offer precise quantitative analysis tools for commercial street site selection [29–31]. This provides an opportunity to develop new methods for sustainable commercial street location prediction integrating big data and machine learning. In this study, hidden patterns in POI data are deeply mined by combining machine learning algorithms to improve the accuracy and science of commercial street location prediction.

However, existing research still needs to be deepened in some aspects: first, there is a lack of in-depth characterization of the internal structure and functional composition of commercial streets; second, the application of machine learning methods is still in the initial stage, and the interpretability and applicability of the models need to be improved; third, there is a lack of comparative analysis of commercial street site selection patterns in different cities and regions. In view of this, this study, based on a big data perspective, fully explores the multi-dimensional attributes of POI data to refine the characterization of the location characteristics of commercial streets. Furthermore, the study employs decision tree algorithms to improve prediction accuracy while generating simple and clear site selection rules, greatly enhancing the practicality of the model. The study focuses on a single city case, deeply analyzing the influencing mechanisms of commercial street site selection in a specific regional context, deepening the understanding of regional heterogeneity. The research results not only expand the theoretical horizon of sustainable commercial street site selection research but also provide new technical support and decision-making references for urban regeneration and commercial layout practices, having significant academic value and practical significance.

The remainder of the paper is organized as follows: Section 2 introduces the study area and methods. Section 3 discusses the analysis results. Finally, Sections 4 and 5 summarize the paper and provide suggestions for future research.

2. Materials and Methods

2.1. Research Area

This study focuses on Foshan City, located in the hinterland of the Pearl River Delta, Guangdong Province, China (110°00′–115°00′ E, 22°00′–25°00′ N), as shown in Figure 1 [32]. With a subtropical monsoon climate, Foshan City has a total area of 3797.99 square kilometers and a permanent resident population exceeding 9.6 million as of 2022. The city comprises five districts: Chancheng, Nanhai, Shunde, Gaoming, and Sanshui, with an altitude that decreases from the northwest to the southeast [33]. Foshan City was selected as the focus of this research due to its strong economic foundation, ranking among the top cities in China for GDP per capita, which creates a dynamic environment for commercial street development. Additionally, Foshan's innovative urban planning and commercial development strategies provide a robust framework for analyzing site selection. As a rapidly developing second-tier city, Foshan represents a common urban typology in China, making the insights from this study applicable to similar urban contexts both nationally and internationally. Furthermore, its advanced digital infrastructure ensures access to high-quality POI data, essential for our predictive model.

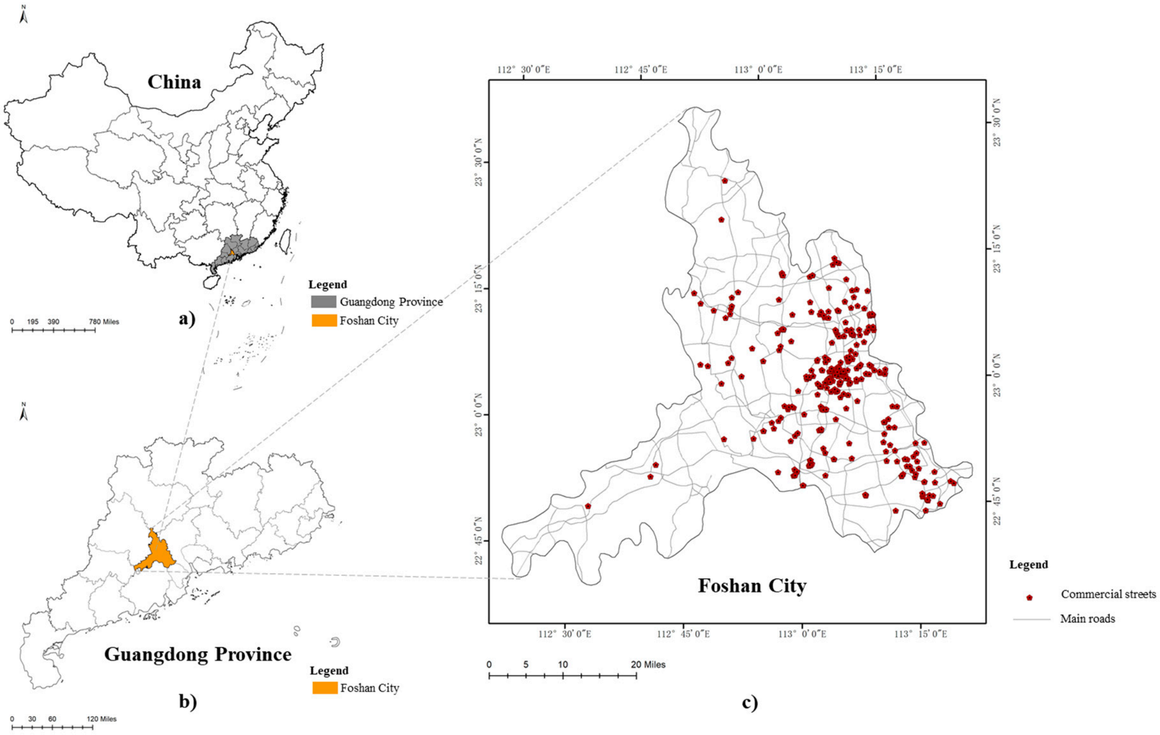

Figure 1. Study area. (**a**) Location of Guangdong Province in China. (**b**) Location of Foshan City in Guangdong Province. (**c**) Distribution of commercial streets in Foshan.

Our research focuses on addressing two key challenges: enhancing urban vitality and improving commercial infrastructure. As cities expand and evolve, there is a growing need to strategically plan commercial street development in ways that foster social and economic activity. In Foshan, like many rapidly urbanizing cities, identifying optimal locations for commercial streets is essential to boost foot traffic, stimulate local economies, and create vibrant, livable urban spaces. Additionally, improving the efficiency and effectiveness of commercial infrastructure planning can help balance development across different areas, preventing overconcentration and ensuring equitable access to commercial services.

According to data from Amap, a popular Chinese mapping and navigation application, Foshan City currently has 310 commercial streets. Amap's commercial street data are obtained through a combination of channels such as field collection, user feedback, as well as commercial cooperation and public data sources, and these data are widely recognized by the public as reliable and comprehensive. Whether a street is classified as a commercial street is determined by Amap based on a multi-dimensional analysis and algorithmic modeling of historical data, user behavior, official and corporate data, as well as user feedback and reviews. Since the actual boundaries of commercial streets are not clearly defined, we did not include street length information in the dataset. Instead, representative points are used to indicate the location of the commercial street. These representative points are either the center or key nodes of the street, determined through a comprehensive evaluation of government data, merchant registration information, street view data, and user-submitted data. Using Ripley's K function for analysis, we observed that the observed

K values consistently exceed the expected K values and fall above the confidence interval. This indicates a significant spatial clustering phenomenon in the distribution of commercial street locations in Foshan, particularly in the core commercial areas of the central and eastern parts of the city, as shown in Figure 2. These regions exhibit dense commercial activities, with commercial street locations highly concentrated, forming a clear commercial core. Conversely, the commercial streets in the city's edge areas show relative sparsity, where commercial development is limited and still in the developmental phase. The spatial distribution characteristics of commercial streets in Foshan provide important insights for future commercial development and urban planning, as identifying clustering phenomena can help optimize resource allocation and enhance the effectiveness of commercial street site selection. Table 1 presents an overview of five representative commercial streets located in different districts of Foshan City, showcasing the diverse characteristics and business types commonly found in urban commercial areas.

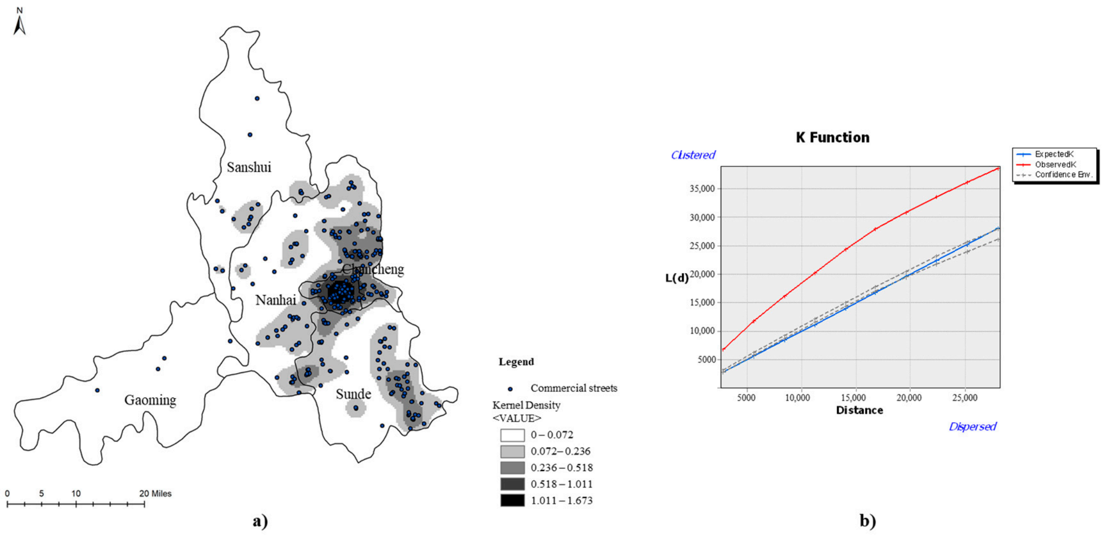

Figure 2. (**a**) Distribution of commercial streets in Foshan. (**b**) Analyze the K function of commercial streets.

Table 1. Overview of representative commercial streets in Foshan City.

Commercial Street Name	District	Length (m)	Main Business Types
Dali Commercial Pedestrian Street	Nanhai District	320	Retail, Catering
Qushuiwan Fengqing Commercial Pedestrian Street	Shunde District	230	Retail, Entertainment Leisure, Catering
Xijiang River Fresh Food Street	Sanshui District	640	Catering
Tea Culture Street	Chancheng District	250	Cultural and Artistic Industries, Catering
Qintian Hui ZhonghuanGold Street	Gaoming District	460	Catering, Retail, Entertainment and Leisure

2.2. Data Sources and Processing

This study generates feature variables based on POI big data, which is diverse and large in quantity, and can comprehensively and accurately characterize the environment. Moreover, POI big data can be updated quickly and is highly scalable, which is conducive to the expansion and comparative research in the spatial and temporal dimensions. The primary data source for this study was the Amap Open Platform. Utilizing Python programs and API interfaces, a total of 506,216 Points of Interest (POI) data records for Foshan City in 2022 were collected. The data fields included name, category, longitude, latitude, and district. After data deduplication and cleaning, 503,191 valid POI data records were obtained. Referring to the Amap POI classification system, the data were categorized into 14 main categories: Food, Company and Business, Shopping, Traffic Facility, Finance, Hotels, Education and Cultural Media, Scenic Spot, Automobile Service, Real Estate, Life and Service, Recreation and Entertainment, Medical Care, and Sports and Fitness. Table 2 provides a detailed breakdown of the POI types included in each category. The POI data were generated as point files with a unified spatial coordinate system.

Table 2. POI data classification.

Primary Category	Secondary Category	Number	Proportion
Food	Teahouse, Cake and Dessert Shop, Coffee, Foreign Food, Fast Food, Chinese Food, Others	76,098	15.12%
Company and Business	Factory, Company, Agriculture, Forestry, Fishery, Others	91,370	18.16%
Shopping	Department Stores, Convenience Stores, Supermarkets, Shopping Centers, Flowers, Birds, Fish and Insects, Home Appliances and Digital Appliances, Household Building Materials, Duty Free Shops, Markets, Sports and Cultural Goods, Others	153,773	30.56%
Traffic Facility	Subway, Airplane, Service Area, Port Terminal, Bus Station, Train, Port, Ferry, Toll Station, Parking Lot, Coach, Others	27,776	5.52%
Finance	ATM, Insurance, Investment Banking, Banking, Others	5083	1.01%
Hotels	Economy Hotel Chains, Hostels, Youth Hostels, Three-star Hotels, Four-star Hotels, Five-star hotels, Others	8251	1.64%
Education and Cultural Media	Museums, Adult Education, Archives, Higher Education, Radio and Television Broadcasting, Conventions and Exhibitions, Driving Schools, Science and Technology Museums, Scientific Research Institutes, Art Exhibitions, Training Units, Libraries, Palaces of Culture, Elementary School, Press and Publications, Art Organizations, Kindergartens, Vocational and Technical Education, Secondary Schools, Others	16,056	3.19%
Scenic Spot	Zoos, Parks, Squares, Red Tourism, Memorials, Attractions, Aquariums, Botanical Gardens, Religious, Others	4888	0.97%
Automobile Service	Charging Stations, Used Cars, Filling Stations, Gas Stations, Other Energy Stations, Auto Parts, Auto Repair, Car Sales, Car Care, Car Rental, Car Wash, Others	18,750	3.73%
Real Estate	Villa Areas, Industrial Parks, Industrial Buildings, Commercial and Residential Buildings, Community Centers, Dormitories, Office Buildings, Residential Areas, Others	16,619	3.30%
Life and Service	Lottery Ticket Sales, Telecommunication Business Halls, Public Toilets, Public Utilities, Beauty Salons, Photography and Printing, Logistics, Laundry, Bath and Massage, Information and Consultation Centers, Post Offices, Intermediaries, Others	61,675	12.26%
Recreation and Entertainment	KTV, Cinema, Vacation Retirement, Bar, Theater, Chess Room, Internet Cafe, Playground, Others	4869	0.97%
Medical Care	Animal Medical Care, Emergency Centers, Disease Prevention, Pharmaceutical Sales, Clinics, Specialty Hospitals, General Hospitals, Others	13,817	2.75%
Sports and Fitness	Bowling, Ice Sports, Golf, Outdoor Fitness, Fitness Centers, Basketball, Campgrounds, Equestrian and Horse Racing, Table Tennis, Water Sports, Billiards, Taekwondo, Tennis, Swimming, Badminton, Sports Complexes, Soccer, Others	4165	0.83%

China's mandatory national standard "Urban Residential Area Planning and Design Standard" (GB 50180-2018) stipulates that the walking distance of 15 min living circle should be 800–1100 m. Because the raster data of the spatial distribution of the population in the LandScan population dataset, an important database, is of 1 km accuracy, this study adopts a 1 km × 1 km scale grid to meet the Chinese national conditions and support the subsequent study. This makes the spatial representation consistent, helps the model better capture the differences between different regions, reduces the model complexity, and transforms the continuous geospatial features into discrete grid cells, which is more suitable for the characteristics of the decision tree algorithm. The study area was divided into 9808 grid cells of 1 km × 1 km using the fishnet creation tool in ArcGIS 10.7 software, as shown in Figure 3. Based on the spatial relationship between the grid and POI points, the POI category information and spatial composition within each grid were integrated into the grid attributes. The data were then exported and subjected to binary processing, assigning a value of 1 if a certain category of POI existed within a grid and 0 otherwise. This process resulted in a grid attribute matrix with the presence or absence of the 14 categories of POIs as feature variables and the presence or absence of commercial streets as the target variable. Grids containing commercial streets were defined as positive examples, while those without were defined as negative examples, yielding 255 positive examples and 9553 negative examples. In machine learning, especially when using decision tree algorithms for classification, an imbalance of positive and negative samples may lead to bias in the model's prediction by tending to predict a larger number of samples. Equal positive-to-negative ratio sampling enables the model to better learn the features of each class during training and improves the generalization ability of the model. Random sampling is achieved by generating random numbers using the RAND() function. Randomly selected 200 positive example grids and 200 negative example grids from the training dataset. The dataset was further divided into a training set and a test set in a 7:3 ratio.

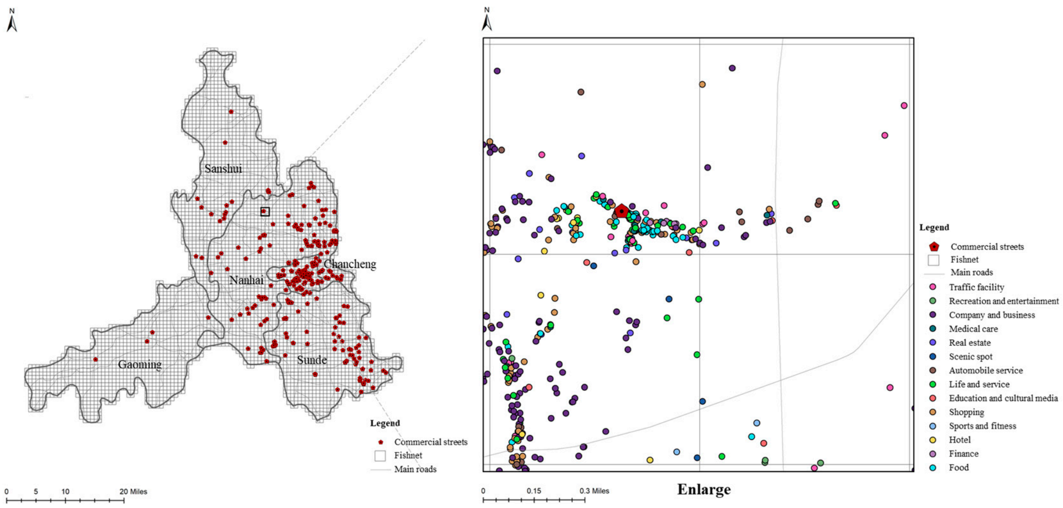

Figure 3. Foshan City grid division and local schematic.

2.3. Research Methods

This study utilizes POI big data to first categorize the data, extract features, and balance samples, constructing training and testing sets. Next, the ID3 algorithm is employed to generate a decision tree model, using information gain or entropy to split nodes, and the model is visualized. Based on this model, we predict 2157 potential commercial street grids in Foshan. Finally, by combining logical chains and factor analysis, we propose targeted policy recommendations to optimize the layout and site selection strategies for commercial streets.

Drawing upon existing research [34], this study constructs a commercial street site selection prediction model using POI big data and decision tree algorithms. The ID3

(Iterative Dichotomiser 3) algorithm, proposed by Quinlan in 1986 [35], is employed for model training. ID3 uses information entropy and information gain as splitting criteria and recursively selects the optimal attributes to construct a classification tree [13,30,36].

Information Entropy measures the uncertainty of a sample set, and its calculation formula is as follows [37,38]:

$$Ent(S) = -\sum_{i=1}^{m} P_i log_2 P_i \quad (1)$$

where S: the set of samples; P_i: the proportion of samples in category i; and m: the type of sample prediction results.

Information gain $Gain(S,a)$ is also an indicator that measures the purity of a sample set. The higher the information gain, the higher the purity of the child nodes after splitting. The calculation formula for information gain is as follows [39–41]:

$$Gain(S,a) = Ent(S) - \sum_{v=1}^{v} \frac{|S^v|}{|S|} Ent(S^v) \quad (2)$$

where S: the set of samples; a: a discrete attribute; and v: a branch node.

The ID3 algorithm compares the information gain of each attribute and selects the attribute with the highest gain as the splitting attribute of the current node. This process is repeated until the stopping condition is met, generating the final decision tree. By analyzing the relative importance of various influencing factors and their mechanisms, and comparing the model results with the current planning, the differences and reasons between the two are discussed. The research workflow is shown in Figure 4.

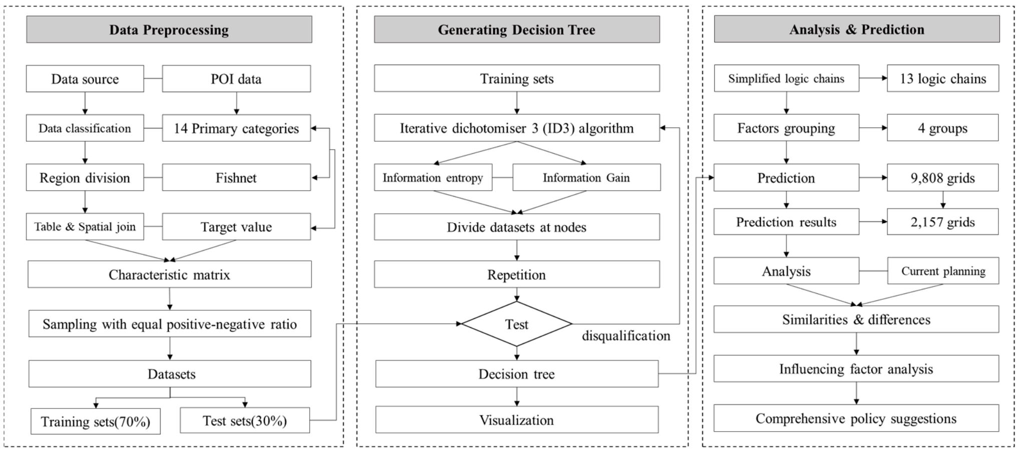

Figure 4. The research workflow.

3. Results

3.1. Model Design and Evaluation

In this study, the ID3 algorithm was employed to construct a decision tree model for identifying optimal sites for sustainable commercial street development in Foshan City, China [42]. The maximum depth of the tree was set to 5 to ensure a balance between model complexity and interpretability. The model's performance was evaluated by comparing the true values and predicted values on the test set, achieving an accuracy of 83%. This high accuracy indicates the credibility and reliability of the prediction results. Only 2022 data have been used to validate the model accuracy, and due to the long period of time required for the construction and development of the commercial street, we will validate the forecasting model again in another 5 or 10 years to optimize the forecasting model. Figure 5 presents the generated decision tree structure, which serves as the basis for the subsequent analysis of decision rules and influencing factors.

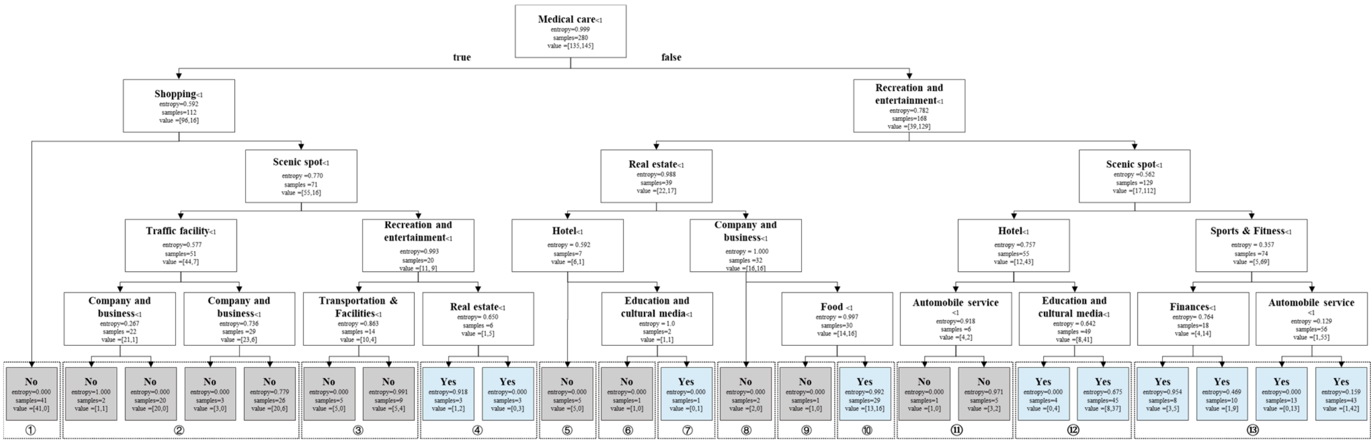

Figure 5. Decision tree structure based on POI data for predicting optimal commercial street locations.

3.2. Decision Rule Analysis

The hierarchical structure of the decision tree reflects the importance ranking of attributes, with attributes closer to the root node having a greater impact on classification decisions [42]. As shown in Figure 5, the root node is "Medical Care", indicating that Medical Care POIs are the primary factor influencing the location selection of commercial streets. This finding suggests that the presence of Medical Care POIs within a grid is a strong indicator of a considerable population base and pedestrian flow, which are key considerations in the location selection of commercial streets. Figure 6 illustrates the spatial distribution characteristics of Medical Care POIs in Foshan City, highlighting their convergence with the target audience of commercial streets. Medical Care is often an important part of a city's functionality and can reflect the level of infrastructure and services in the area. Areas with good healthcare services tend to attract more residents and tourists, thus promoting commercial activities and the formation of commercial streets. However, the detailed mechanism of its role in influencing the location of commercial streets needs to be further studied.

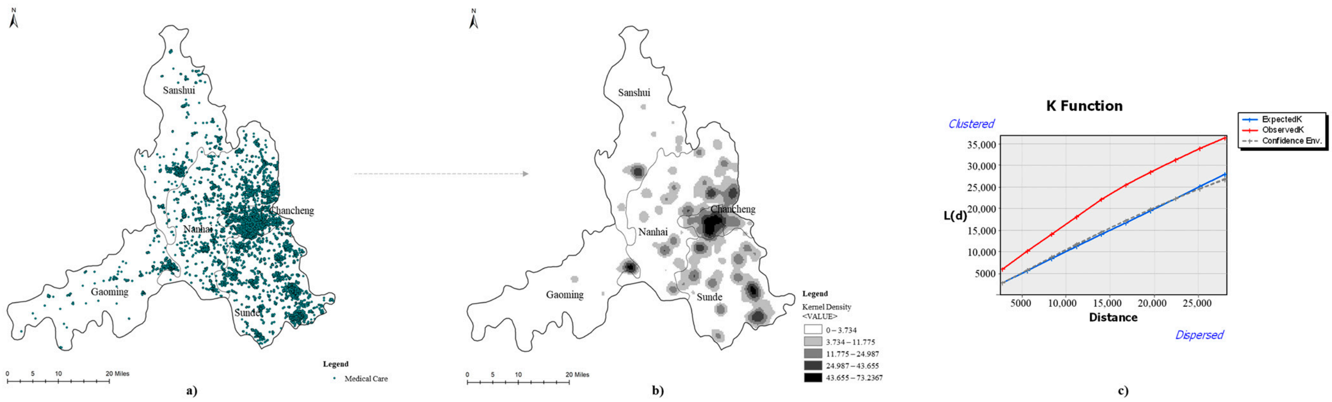

Figure 6. (a) Distribution of POI locations in the "Medical Care" category. (b) Kernel density analysis of POIs in the "Medical Care" category. (c) K function analysis of POIs in the "Medical Care" category.

On the second level of the decision tree, "Shopping" and "Recreation & entertainment" emerge as the next best bases for location selection when Medical Care POIs are absent or present, respectively. The presence of shopping venues and recreation and entertainment facilities not only corroborates the population size of the surrounding area but also provides retail business support and agglomeration effects for the construction of commercial streets.

The third and fourth levels of the decision tree reveal the positive influence of "Scenic spot", "Education & Cultural media", and "Food" on the location selection of commercial streets. These factors contribute to shaping the commercial atmosphere and are essential components of successful commercial streets. However, their importance is relatively lower compared to the hard indicators such as population base and consumption capacity.

3.3. Rule Chain Simplification

Based on the decision tree model, 13 logical rules for discriminating the location selection of commercial streets were derived, as shown in Figure 7. Among these rules, five support the construction of commercial streets, while 8 do not recommend location selection. The simplest rule, containing only two conditions, determines that a location is not suitable for developing a commercial street when there are neither Medical Care POIs nor Shopping POIs within a grid.

Serial number	Population base						Activity base			Cultural base		Facilities base			Commercial street
	Medical care	Company and business	Finance	Hotel	Real estate	Life and service	Food	Shopping	Recreation and entertainment	Education and cultural media	Scenic spot	Traffic facility	Automobile service	Sports and fitness	
④	No	/	/	/	/	/	/	Yes	Yes	/	Yes	/	/	/	Yes
⑦	Yes	/	/	Yes	No	/	/	/	No	Yes	/	/	/	/	Yes
⑩	Yes	Yes	/	/	Yes	/	Yes	/	No	/	/	/	/	/	Yes
⑫	Yes	/	/	Yes	/	/	/	/	Yes	/	No	/	/	/	Yes
⑬	Yes	/	/	/	/	/	/	/	Yes	/	Yes	/	/	/	Yes
①	No	/	/	/	/	/	/	No	/	/	/	/	/	/	No
②	No	/	/	/	/	/	/	Yes	/	/	No	/	/	/	No
③	No	/	/	/	/	/	/	Yes	No	/	Yes	/	/	/	No
⑤	Yes	/	/	No	No	/	/	/	No	/	/	/	/	/	No
⑥	Yes	/	/	Yes	No	/	/	/	No	No	/	/	/	/	No
⑧	Yes	No	/	/	Yes	/	/	/	No	/	/	/	/	/	No
⑨	Yes	Yes	/	/	Yes	/	No	/	No	/	/	/	/	/	No
⑪	Yes	/	/	No	/	/	/	/	Yes	/	No	/	/	/	No

Population base	Activity base	Facilities base	Cultural base	Yes(colorful)	No(gray)

Figure 7. Simplification of the decision tree logic chain for optimal commercial street site selection.

3.4. Classification of Influencing Factors

The analysis of the decision tree model reveals that the influence of various types of POIs on the location selection of commercial streets can be summarized into four dimensions: population base, consumption level, facility support, and cultural atmosphere. While 9 categories of POIs have a significant impact on the location selection of commercial streets, the effects of five categories, including Finance, Life and Service, Traffic Facility, Automobile Service, and Sports and Fitness, are not reflected in the model. This result differs from the views of some scholars who emphasize the importance of the transportation environment in determining the location of commercial areas [34]. The presumed reason for this discrepancy may be the relatively macro grid scale of 500 m $\times$ 500 m and the lack of detailed classification of Traffic Facility, which limits the explanatory power of transportation factors in the model. Figure 8 presents the spatial distribution characteristics of Traffic facility POIs in Foshan City.

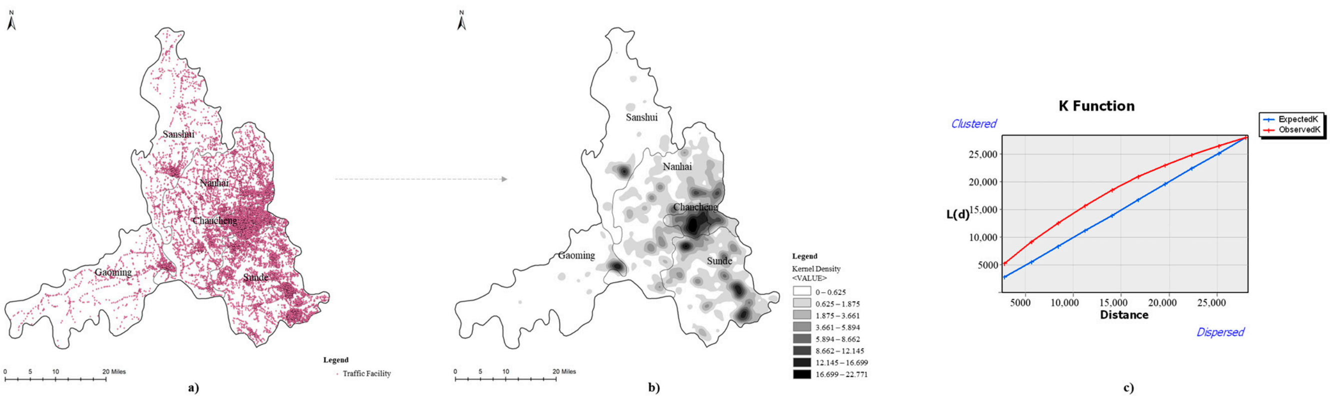

Figure 8. (**a**) Distribution of POI locations in the "Traffic Facility" category. (**b**) Kernel density analysis of POIs in the "Traffic Facility" category. (**c**) K function analysis of POIs in the "Traffic Facility" category.

3.5. Evaluation of Prediction Results

The trained decision tree model was used to predict the location selection of commercial streets for all 9808 grids in Foshan City. The results, depicted in Figure 9, show that 2157 grids are suitable for developing commercial streets. The result is only a preliminary prediction result, and in the actual application, it is necessary to combine the population resources, regional consumption structure, the surrounding business environment, environmental quality, cultural resources, and other factors to carry out secondary screening.

Figure 9. (**a**) Prediction results of location selection for commercial streets. (**b**) Comparison between the predicted site selection results of commercial streets and the current planning.

A comparison of the prediction results with the "Foshan City Territorial Spatial Master Plan (2020–2035)" [36] reveals that the mature conditions for constructing commercial streets are currently lacking in the middle section of the northern urban development axis, the western section of the southern urban development axis, and the northern section of the industrial development axis. Notably, there are significant deviations between the planned construction land for commercial streets in the northern Foshan area and the airport economic zone and the model prediction results.

These differences may be attributed to two factors: (1) the northern Foshan area and the airport economic zone are still in the initial stage of development, with supporting facilities yet to be improved; and (2) the scale of commercial land planned may not match the current population base and consumption demand. Considering the strategic positioning and future development plans of these areas, it can be inferred that the necessity and feasibility of large-scale development of commercial streets in the northern Foshan and airport areas at the current stage are debatable, and the commercial layout in the planning is more of a forward-looking arrangement for future needs. In contrast, the existing population size, consumption level, and public service level in the urban central area have a more significant supporting role for commercial streets, which highly aligns with the model's prediction results.

4. Discussion

The findings of this study provide valuable insights into the factors influencing the location selection of sustainable commercial streets in Foshan City, China. The results highlight the significant impact of Medical Care, Shopping, Recreation and Entertainment, and Scenic Spot POIs on the optimal siting of commercial streets. These findings align with the conclusions drawn by previous studies, which have emphasized the importance of factors such as human resources, consumer participation, regional consumption structure, surrounding commercial environment, environmental quality, and cultural resources in the development of street commerce [13,43–47]. However, the present study distinguishes itself from previous research by focusing on predicting the construction conditions of areas where sustainable commercial streets have not yet been built, rather than evaluating the development of existing commercial streets. By employing a data-driven approach using specific POI data, this study offers a novel method for identifying optimal sites for sustainable commercial street development and fostering urban regeneration. Our research emphasizes the broad impact of different categories of Points of Interest (POI) on the location of commercial streets. Integrating our new method with traditional approaches will provide a more comprehensive perspective for both conventional and innovative street development strategies.

The results of this study have guiding suggestions for the overall planning layout of commercial streets in urban planning practice, and have important use value for the overall spatial optimization layout. This study takes the administrative district as the research scope, and from the overall perspective, the machine learning algorithm based on spatial data mining can reduce the subjectivity of the planning technicians in the site selection, realize the data analysis to assist the planning decision, and make the site selection of the commercial street more scientific and global [48]. In the future, it can provide a research basis for the full realization of smart city intelligent planning.

Despite the significant contribution of this study, its limitations and the need for further research must be recognized. The distribution of commercial streets in Foshan City is mainly characterized by polycentricity, with traditional and modern commercial streets co-existing. The scope of the study is limited to Foshan City, which may affect the generalizability of the conclusions. Future research should explore the transferability of the proposed methodology to other urban contexts and conduct comparative analyses between different cities to identify the deeper reasons for the differences and similarities in urban findings.

Furthermore, the present study relies on POI data from a single year (2022), which may not capture the temporal variations in urban dynamics. Future research should investigate the potential impact of temporal changes in POI data on site selection decisions, as using the current status of POIs around existing commercial streets for prediction may produce large errors. Incorporating historical POI data and analyzing planned but not yet constructed commercial streets could lead to the development of more scientific and accurate decision tree prediction models.

To address these limitations, future research should optimize the proposed methodology by incorporating more data sources, such as socio-economic indicators, transportation networks, urban form, and other multi-source big data, in order to gain a more comprehensive understanding of the factors affecting the development of sustainable commercial streets. In addition, the robustness and applicability of the proposed methodology can be further enhanced by expanding the study area, comparing different cities, and analyzing the temporal changes in POI data for a more detailed, accurate, and generalizable study.

This study exclusively employs decision tree algorithms and does not incorporate other machine learning methods such as random forests, support vector machines, gradient boosting decision trees, neural networks, or spatial econometric models for further optimization of the site selection prediction model in Foshan. The primary reason is that the development cycle of commercial streets is relatively long, and changes in short-term predictions may not be significant. Additionally, since the main objective of this research is to explore the preliminary application of big data and machine learning techniques in commercial street site selection, using a simple and interpretable decision tree model facilitates a more intuitive understanding of the model's impact on site selection factors. In the future, we plan to incorporate various machine learning models after five or ten years, combining long-term observational data to improve prediction accuracy and fully consider dynamic factors in both time and space, aiming to provide more comprehensive decision support for the planning of commercial streets in Foshan.

5. Conclusions

This study contributes to the growing body of research on sustainable urban development by proposing a data-driven approach to identifying optimal sites for commercial street development. By leveraging the power of Points of Interest (POI) data and machine learning techniques, specifically the ID3 decision tree algorithm, we demonstrate the effectiveness of this approach in the context of Foshan City, China. The fine-grained spatial analysis, conducted at the grid cell level, allows for a comprehensive understanding of the factors influencing commercial street site selection and their complex interrelationships.

The research results indicate that commercial street locations in Foshan exhibit significant spatial clustering in the core commercial areas of the central and eastern parts of the city, while they are relatively sparse in the edge areas, where commercial development is limited. The results highlight the significant impact of Medical Care, Shopping, and Recreation and Entertainment POIs on the location selection of sustainable commercial streets. These findings underscore the importance of considering population base, human activity patterns, and cultural elements when planning for commercial development. The strong influence of Medical Care POIs on site selection decisions likely reflects the importance of the population base in supporting sustainable commercial streets. Furthermore, the presence of Shopping and Recreation and Entertainment POIs not only indicates a sufficient population base but also provides an essential foundation for the vibrancy and attractiveness of commercial streets. The influence of Scenic Spot POIs on site selection emphasizes the role of cultural resources in creating unique and appealing shopping destinations that contribute to urban regeneration.

The main contributions of this study are twofold. First, it presents a novel decision-support method for data-driven and sustainable commercial street site selection, showcasing the potential of machine learning techniques in urban planning. Compared with traditional methods, this study not only provides more refined and objective site selection recommendations, but also provides strong data support for urban planning and business decision-making. Second, it offers valuable insights into the complex interplay between urban land use, human activities, and commercial development, highlighting the need for a holistic approach to sustainable urban regeneration. Also, the findings have important implications for urban planners, policy-makers, and developers seeking to create vibrant and sustainable commercial streets that foster urban regeneration. By considering the influence of various POI categories and their spatial distribution, decision-makers can make informed choices regarding the location and design of commercial streets, ensuring their long-term success and positive impact on the urban fabric. Our study focuses on the broader influence of various Point of Interest (POI) categories on commercial street locations.

In conclusion, this study demonstrates the value of a data-driven approach to identifying optimal sites for sustainable commercial street development, contributing to the growing field of urban informatics and sustainable urban planning. By leveraging the power of POI data and machine learning techniques, we can make more informed decisions that foster urban regeneration and promote the creation of livable, vibrant, and sustainable cities. On the basis of many such studies, the theory and application of data-driven planning will be gradually improved, further approaching the beautiful vision of smart planning for smart cities.

Author Contributions: Data curation, Linghan Yao and Xiaoyi Wang; formal analysis, Chao Gao, Linghan Yao and Yanqing Xu; methodology, Linghan Yao, Yanqing Xu and Xinyue Zhang; validation, Chao Gao and Linghan Yao; visualization, Xiaoyi Wang, Yequan Hu and Xinyue Zhang; funding, Chao Gao and Yanqing Xu; writing—original draft, Linghan Yao and Yanqing Xu; writing—review and editing, Chao Gao, Xinyue Zhang and Yequan Hu. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 52408042; Jiangsu Natural Science Foundation, grant number BK20240931; China Postdoctoral Science Funding, grant number 2024M750430; and Jiangsu Provincial Department of Education Fund of Philosophy and Social Science, grant number 2023SJYB2054.

Data Availability Statement: Dataset available on request from the authors.

Acknowledgments: We thank Guangzhou Fangyu Technology Co., Ltd. and Guangzhou Sikan Surveying and Mapping Technology Co. We are grateful to several teachers of South China University of Technology for their guidance.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Hu, L.; Yan, J.; Zhu, Y.; Deng, J.; Chen, L.; Lu, S. Research on the Sustainable Design of Commercial Street Space Based on Importance Performance Analysis. *Buildings* **2022**, *12*, 2096. [CrossRef]
2. Zhang, L.; Xu, X.; Guo, Y. Comprehensive Evaluation of the Implementation Effect of Commercial Street Quality Improvement Based on AHP-Entropy Weight Method—Taking Hefei Shuanggang Old Street as an Example. *Land* **2022**, *11*, 2091. [CrossRef]
3. Longdill, P.C.; Healy, T.R.; Black, K.P. An Integrated GIS Approach for Sustainable Aquaculture Management Area Site Selection. *Ocean. Coast. Manag.* **2008**, *51*, 612–624. [CrossRef]
4. Aksoy, E.; San, B.T. Geographical Information Systems (GIS) and Multi-Criteria Decision Analysis (MCDA) Integration for Sustainable Landfill Site Selection Considering Dynamic Data Source. *Bull. Eng. Geol. Environ.* **2019**, *78*, 779–791. [CrossRef]
5. Zheng-Biao, H.E. Analysis on Development Strategy of China's Commercial Street: A Case Study of Hangzhou City. *CUBR* **2013**, *12*, 787–795. [CrossRef]
6. Lian, H.; Li, G. Correlation Analysis of Retail Space and Shopping Behavior in a Commercial Street Based on Space Syntax: A Case of Shijiazhuang, China. *Buildings* **2023**, *13*, 2674. [CrossRef]
7. Zhang, Y. How Does Cultural Planning Empower Historic District and Urban Regeneration—Case in the China Baroque Historical Street in Daowai District, Harbin. *Highlights Bus. Econ. Manag.* **2024**, *26*, 197–203. [CrossRef]
8. Christaller, W. *Die Zentralen Orte in Süddeutschland*, 3rd ed.; Wissenschaftliche Buchgesellschaft: Darmstadt, Germany, 1980; pp. 214–380.
9. Rushton, G. Analysis of Spatial Behavior by Revealed Space Preference. *Ann. Assoc. Am. Geogr.* **1969**, *59*, 391–400. [CrossRef]
10. Potter, R.B. Spatial Patterns of Consumer Behaviour and Perception in Relation to the Social Class Variable. *Area* **1977**, *9*, 153–156.
11. Gehl, J. *Livet Mellem Husene*, 1st ed.; Architect's Publisher: Copenhagen, Denmark, 1971; pp. 121–181.
12. Jacobs, J. *The Death and Life of Great American Cities*, 1st ed.; Random House: New York, NY, USA, 1961; pp. 214–380.
13. Mehanna, W.A.E.-H.; Mehanna, W.A.E.-H. Urban Renewal for Traditional Commercial Streets at the Historical Centers of Cities. *Alex. Eng. J.* **2019**, *58*, 1127–1143. [CrossRef]
14. Sung, H.; Lee, S. Residential Built Environment and Walking Activity: Empirical Evidence of Jane Jacobs' Urban Vitality. *Transp. Res. Part D Transp. Environ.* **2015**, *41*, 318–329. [CrossRef]
15. Ghazi, N.M.; Abaas, Z.R. Toward Liveable Commercial Streets: A Case Study of Al-Karada Inner Street in Baghdad. *Heliyon* **2019**, *5*, e01652. [CrossRef] [PubMed]
16. Kickert, C.; vom Hofe, R.; Haas, T.; Zhang, W.; Mahato, B. Spatial Dynamics of Long-Term Urban Retail Decline in Three Transatlantic Cities. *Cities* **2020**, *107*, 102918. [CrossRef]
17. Hao, F.; Yang, Y.; Wang, S. Patterns of Location and Other Determinants of Retail Stores in Urban Commercial Districts in Changchun, China. *Complexity* **2021**, *2021*, e8873374. [CrossRef]
18. Colaco, R.; de Abreu e Silva, J. Commercial Classification and Location Modelling: Integrating Different Perspectives on Commercial Location and Structure. *Land* **2021**, *10*, 567. [CrossRef]
19. Colaco, R.; e Silva, J.D.A. Commercial Land Use Change and Growth Processes—An Assessment of Retail Location in Lisbon, Portugal, 1995–2020. *J. Urban Manag.* **2024**, *13*, 157–170. [CrossRef]
20. Wieland, T. Market Area Analysis for Retail and Service Locations with MCI. *R J.* **2017**, *9*, 298–323. [CrossRef]
21. Baviera-Puig, A.; Buitrago-Vera, J.; Escriba-Perez, C. Geomarketing Models in Supermarket Location Strategies. *J. Bus. Econ. Manag.* **2016**, *17*, 1205–1221. [CrossRef]
22. Erbiyik, H.; Ozcan, S.; Karaboga, K. Retail Store Location Selection Problem with Multiple Analytical Hierarchy Process of Decision Making an Application in Turkey. In Proceedings of the 8th International Strategic Management Conference, Barcelona, Spain, 21–23 June 2012; Ozsahin, M., Zehir, C., Eds.; Elsevier Science Bv: Amsterdam, The Netherlands, 2012; Volume 58, pp. 1405–1414.
23. Wang, Y.; Li, S.; Zhang, X.; Jiang, D.; Hao, M.; Zhou, R. Site Selection of Digital Signage in Beijing: A Combination of Machine Learning and an Empirical Approach. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 217. [CrossRef]

24. Wang, L.; Fan, H.; Wang, Y. Site Selection of Retail Shops Based on Spatial Accessibility and Hybrid BP Neural Network. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 202. [CrossRef]
25. Lu, J.; Zheng, X.; Nervino, E.; Li, Y.; Xu, Z.; Xu, Y. Retail Store Location Screening: A Machine Learning-Based Approach. *J. Retail. Consum. Serv.* **2024**, *77*, 103620. [CrossRef]
26. Shan, Z.; Song, H. Research on Management Decision Based on Machine Learning: Taking the Decision of Location Selection of a Pharmaceutical Retail Enterprise as an Example. In Proceedings of the Fuzzy Systems and Data Mining V (FSDM 2019), Kitakyushu City, Japan, 18–21 October 2019; TallonBallesteros, A.J., Ed.; Ios Press: Amsterdam, The Netherlands, 2019; Volume 320, pp. 564–574.
27. Mazhi, K.Z.; Davi, A.; Suryana, L.E.; Dewi, W.R. Site Selection of Retail Shop Based on Spatial Analysis and Machine Learning. In Proceedings of the ICACSIS 2020: 2020 12th International Conference on Advanced Computer Science and Information Systems (ICACSIS), Depok, Indonesia, 17–18 October 2020; IEEE: New York, NY, USA, 2020; pp. 399–404.
28. Andrade, R.; Alves, A.; Bento, C. POI Mining for Land Use Classification: A Case Study. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 493. [CrossRef]
29. Liu, K.; Yin, L.; Lu, F.; Mou, N. Visualizing and Exploring POI Configurations of Urban Regions on POI-Type Semantic Space. *Cities* **2020**, *99*, 102610. [CrossRef]
30. Al-sharif, A.A.A.; Pradhan, B. A Novel Approach for Predicting the Spatial Patterns of Urban Expansion by Combining the Chi-Squared Automatic Integration Detection Decision Tree, Markov Chain and Cellular Automata Models in GIS. *Geocarto Int.* **2015**, *30*, 858–881. [CrossRef]
31. Delen, D.; Kuzey, C.; Uyar, A. Measuring Firm Performance Using Financial Ratios: A Decision Tree Approach. *Expert Syst. Appl.* **2013**, *40*, 3970–3983. [CrossRef]
32. Zhang, S.; Li, X. A Comparative Study of Machine Learning Regression Models for Predicting Construction Duration. *J. Asian Archit. Build. Eng.* **2023**, 1–17. [CrossRef]
33. Deng, Y.; Huang, Y.; Zhang, C.; Cheng, J.; Zhao, W. Guidelines for the Industrial Development of Historic Villages in China Based on Resource Potential Evaluation: 14 Cases in the Guangzhou and Foshan Area, Guangdong Province. *J. Asian Archit. Build. Eng.* **2023**, *22*, 932–944. [CrossRef]
34. Han, Y.; Jia, H. Simulating the Spatial Dynamics of Urban Growth with an Integrated Modeling Approach: A Case Study of Foshan, China. *Ecol. Model.* **2017**, *353*, 107–116. [CrossRef]
35. Wang, X.; Xiong, F.; Wang, Z.; Chen, S.; Zhang, Z.; Peng, L. Planning and Layout of Elderly Facilities Based on POI Big Data and Machine Learning—Wuhan City as an Example. *Econ. Geogr.* **2021**, *41*, 49–56.
36. Quinlan, J.R. Improved Use of Continuous Attributes in C4.5. *JAIR* **1996**, *4*, 77–90. [CrossRef]
37. Le, Q.V. Building High-Level Features Using Large Scale Unsupervised Learning. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 8595–8598.
38. Ben-Gal, I.; Trister, C. Parallel Construction of Decision Trees with Consistently Non-increasing Expected Number of Tests. *Appl. Stoch. Models Bus. Ind.* **2015**, *31*, 64–78. [CrossRef]
39. Bhatt, H.; Mehta, S.; D'mello, L.R. Use of ID3 Decision Tree Algorithm for Placement Prediction. *Int. J. Comput. Sci. Inform. Technol.* **2015**, *6*, 4785–4789.
40. An, Y.; Zhou, H. Short Term Effect Evaluation Model of Rural Energy Construction Revitalization Based on ID3 Decision Tree Algorithm. *Energy Rep.* **2022**, *8*, 1004–1012. [CrossRef]
41. Zhang, Y.-D.; Dong, Z.; Wang, S.-H.; Yu, X.; Yao, X.; Zhou, Q.; Hu, H.; Li, M.; Jimenez-Mesa, C.; Ramirez, J.; et al. Advances in Multimodal Data Fusion in Neuroimaging: Overview, Challenges, and Novel Orientation. *Inf. Fusion* **2020**, *64*, 149–187. [CrossRef]
42. Wang, S.; Celebi, M.E.; Zhang, Y.-D.; Yu, X.; Lu, S.; Yao, X.; Zhou, Q.; Miguel, M.-G.; Tian, Y.; Gorriz, J.M.; et al. Advances in Data Preprocessing for Biomedical Data Fusion: An Overview of the Methods, Challenges, and Prospects. *Inf. Fusion* **2021**, *76*, 376–421. [CrossRef]
43. Chanmee, S.; Kesorn, K. Semantic Decision Trees: A New Learning System for the ID3-Based Algorithm Using a Knowledge Base. *Adv. Eng. Inform.* **2023**, *58*, 102156. [CrossRef]
44. Xu, Y.; Rollo, J.; Jones, D.S.; Esteban, Y.; Tong, H.; Mu, Q. Towards Sustainable Heritage Tourism: A Space Syntax-Based Analysis Method to Improve Tourists' Spatial Cognition in Chinese Historic Districts. *Buildings* **2020**, *10*, 29. [CrossRef]
45. Howley, P.; Scott, M.; Redmond, D. Sustainability versus Liveability: An Investigation of Neighbourhood Satisfaction. *J. Environ. Plan. Manag.* **2009**, *52*, 847–864. [CrossRef]
46. Zhang, L.; Zhang, R.; Yin, B. The Impact of the Built-up Environment of Streets on Pedestrian Activities in the Historical Area. *Alex. Eng. J.* **2021**, *60*, 285–300. [CrossRef]
47. Lyons, M.; Brown, A. Erratum to: Pro-Poor Business Law? On MKURABITA and the Legal Empowerment of Tanzania's Street Vendors. *Hague J. Rule Law* **2016**, *8*, 429. [CrossRef]
48. Shang, H. The primary factor analysis for effects about the market selected location. *J. Beijing Inst. Financ. Commer. Manag.* **2006**, *2*, 23–25.