sustainability

MDPI

Article

Making Smart Cities Human-Centric: A Framework for Dynamic Resident Demand Identification and Forecasting

Wen Zhang [1,*], Bin Guo [2], Wei Zhao [1], Yutong He [1] and Xinyu Wang [2]

[1] School of Management, Xi'an University of Architecture and Technology, Xi'an 710055, China; zhaowei1806@xauat.edu.cn (W.Z.); heyutong@xauat.edu.cn (Y.H.)
[2] School of Public Administration, Xi'an University of Architecture and Technology, Xi'an 710055, China; gbin249640039@gmail.com (B.G.); wxy6@xauat.edu.cn (X.W.)
* Correspondence: wenzhang@xauat.edu.cn

Abstract

Smart cities offer new opportunities for urban governance and sustainable development. However, at the current stage, the construction and development of smart cities generally exhibit a technology-driven tendency, neglecting real resident demand, which contradicts the "human-centric" principle. Traditional top-down methods of demand collection struggle to capture the dynamics and heterogeneity of public demand. At the same time, government service platforms, as one dimension of smart city construction, have accumulated massive amounts of user-generated data, providing new solutions for this challenge. This paper aims to construct a big data-driven analytical framework for dynamically identifying and accurately forecasting core resident demand. The study uses Xi'an City, Shaanxi Province, China, as a case study, utilising user messages from People.cn spanning 2011 to 2023. These messages cover various domains, including urban construction, healthcare, education, and transportation, as the data source. The People.cn message board is China's most significant nationwide online political platform. Its institutionalised feedback mechanism ensures data content focuses on highly representative specific grievances, rather than the broad emotional expressions on social media. The study employs user messages from People.cn from 2011 to 2023 as its data source, encompassing urban construction, healthcare, education, and transportation. First, a large language model (LLM) was used to preprocess and clean the raw data. Subsequently, the BERTopic model was applied to identify ten core demand themes and construct their monthly time series, thereby overcoming the limitations of traditional methods in short-text semantic recognition. Finally, by integrating variational mode decomposition (VMD) with support vector machines (SVMs), a hybrid demand forecasting model was established to mitigate the risk of overfitting in deep learning when forecasting small-sample time series. The empirical results show that the proposed LLM-BERTopic-VMD-SVM framework exhibits excellent performance, with the goodness-of-fit (R2) on various demand themes ranging from 0.93 to 0.96. This study proposes an effective analytical framework for identifying and forecasting resident demand. It provides a decision-support tool for city managers to achieve proactive and fine-grained governance, thereby offering a viable empirical pathway to promote the transformation of smart cities from technology-centric to human-centric.

Keywords: smart cities; resident demand identification; resident demand forecasting; BERTopic model; VMD-SVM model

check for updates

Academic Editor: Enrico Ivaldi

Received: 16 September 2025
Revised: 13 October 2025
Accepted: 17 October 2025
Published: 23 October 2025

Citation: Zhang, W.; Guo, B.; Zhao, W.; He, Y.; Wang, X. Making Smart Cities Human-Centric: A Framework for Dynamic Resident Demand Identification and Forecasting. *Sustainability* **2025**, *17*, 9423. https://doi.org/10.3390/su17219423

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The development of smart cities has garnered significant attention in recent years as an effective approach to addressing multiple challenges arising from rapid urbanisation, including resource constraints, environmental pressures, and governance complexities, while also fulfilling the requirements of sustainable urban development [1]. It has emerged as a theoretical and technological paradigm for urban management, centred on establishing a data-driven intelligent ecosystem with residents at the core of its service provision [2]. By deeply integrating emerging technologies such as the Internet of Things (IoT), artificial intelligence (AI), and big data analytics [3], smart cities enhance operational efficiency, reduce operational costs, improve public service quality, and elevate residents' quality of life [4–6]. By applying information and communications technology (ICT) across urban systems, they deliver efficient services—including smart healthcare, smart education, smart transport, smart governance, public safety, and water management—to all urban stakeholders [7]. Globally, the AI and big data-driven "smart city" paradigm presents fresh opportunities for urban governance and sustainable development, demonstrating considerable potential in key areas such as optimising spatial allocation, intensifying operational efficiency, and refining management practices [8].

To systematically comprehend and evaluate the multifaceted nature of smart cities, academia has proposed various theoretical frameworks. Among the most influential is the six-dimensional model introduced by Giffinger et al. According to their research, a smart city should encompass six dimensions: Smart Economy, Smart Environment, Smart Governance, Smart Living, Smart Mobility, and Smart People [9]. Smart Governance is pivotal to steering high-quality development in smart cities [10]. Its purpose is to leverage technologies such as ICT to foster collaboration among diverse stakeholders, including government bodies and citizens [11]. This emphasises that smart city development must be underpinned by a people-centred philosophy, grounded in resident demand. Such an approach guides technology implementation within smart city projects, mitigating potential negative externalities [12]. Therefore, resident demand forms the foundational basis for developing a smart city, while intelligent technologies serve as the means of implementation [13]. However, since the early 21st century, smart city development has exhibited a pronounced technology-driven orientation, characterised by a widespread tendency to prioritise technology over actual demands [14]. Looking ahead, the construction and advancement of smart cities must strictly adhere to the fundamental principle of "people-centered development", focusing on the real challenges facing urban growth and genuinely addressing and fulfilling the core resident demand [15].

The evolution of smart cities has progressed through phases 1.0 to 3.0: Phase 1.0 primarily leverages technologies such as the Internet of Things and big data to enhance urban development and operational models; Phase 2.0 gradually emphasises concepts and vision over mere technological application; and Phase 3.0 stresses the integration of top-down and bottom-up approaches [16], highlighting public intelligence and participation to achieve a people-centred development pathway [17]. However, urban resident demand is diverse and dynamic. Traditional government-led, top-down demand collection methods often rely on predetermined research frameworks, making it difficult to comprehensively assess residents' genuine preferences and potentially leading to disconnects from actual demand [18,19]. Consequently, scientifically and accurately uncovering and effectively reflecting real resident demand has become a pressing issue in the current people-centred development of smart cities.

As one of the early dimensions of smart city development, the municipal administration platform has accumulated substantial operational data across numerous sectors—

including urban construction, elderly care services, and educational services—during its operation [20]. This data is both voluminous and multidimensional, particularly the proactive feedback submitted by urban residents. Such data reveals residents' preferences for public services. Compared to passively collected questionnaire data, it not only incurs lower collection costs but also more authentically reflects users' actual demands [21]. Reasonably analysing and mining this data can provide low-cost, high-quality guidance for decision-making, improve resource allocation within smart cities, clarify technical application scenarios, and promote proactive planning for smart city development [22,23]. This facilitates a shift in smart city development from a "technology-centric" to a "demand-centric" approach. However, such data lacks fixed format standards and exhibits pronounced unstructured and highly dynamic characteristics. This renders conventional data analysis methods ineffective for empowering resident demand identification based on this data. Consequently, extracting key, actionable insights from noisy and irregular data to study user demands is paramount. Even if resident demand can be identified through user-generated comments, the data is historical, and the identified demands are static. Guiding the people-centred transformation of smart cities based on static demands inevitably introduces lag. Therefore, establishing a framework for the dynamic identification and forecasting of resident demand is crucial for the people-centred transformation and development of smart cities. Integrating identified resident demand with the people-centred transformation objectives of smart cities constitutes another key focus of this paper.

Based on this, the research objective of this paper lies not only in accurately identifying existing resident demand, but also in achieving demand forecasting for the forward-looking, people-centred transformation of smart cities. This study will utilise user-generated comment data from urban government platforms as its research subject. It will employ large language models for pre-processing and cleaning the raw data, adopt the BERTopic thematic modelling method to identify user demands and their evolutionary trends, and construct demand trend time series with a monthly resolution. Given the diversity of demands, this study integrates variational mode decomposition (VMD) with support vector machines (SVMs) to establish a hybrid forecasting model, creating a forecasting framework adaptable to multiple resident demands. Based on this research, not only can current and historical resident demand be accurately identified, but future trends in diverse resident demand can also be forecasted. This organically integrates demand with the people-centred transformation of smart cities, providing forward-looking references for smart city planning and resource allocation. It advances the realisation of the Smart City 3.0 model, combining top-down and bottom-up approaches, driving the transition of urban governance from supply-driven to demand-oriented.

2. Literature Review

2.1. Demand Identification

In developing people-centred smart cities, accurately identifying and deeply understanding the real demands of urban residents constitutes the primary prerequisite and core task. Only by comprehensively grasping resident demand can smart cities' technological deployment and service provision truly align with their practical necessities, thereby avoiding misguided technological investments and resource misallocation. Consequently, identifying resident demand serves not only as the starting point for smart city construction, but also as the bridge connecting the dual principles of 'technology-enabled' and 'people-centred' approaches. Current research on smart city demand identification centres on four approaches: first, questionnaire-based surveys; second, the utilisation of technologies such

as the Internet of Things; third, methods grounded in social media mining; and fourth, approaches utilising government platforms.

As a traditional method for identifying requirements, questionnaire surveys can directly reflect residents' explicit preferences and expectations regarding specific services, facilities, or policies, providing government departments and public service providers with a relatively reliable reference basis. Some scholars have employed questionnaire surveys to conduct quantitative analyses of rural resident information demand, providing empirical references for libraries to deliver targeted literature and information resources [24]. Others have utilised structured questionnaires combined with structural equation modelling to compare satisfaction levels and demand priorities between urban and rural residents. This approach revealed that infrastructure and public services are key determinants of satisfaction and highlighted significant disparities in demand between urban and rural areas, indicating that demand identification must account for regional specificity and equity considerations [25].

Internet of Things (IoT) technology deploys extensive sensor networks at critical urban locations to monitor city operations in real time, generating vast volumes of structured and semi-structured data [26]. This provides an objective, continuous information foundation for identifying resident demand through data-driven approaches. By integrating information and communications technology (ICT) with big data analytics, latent patterns of resident demand can be extracted from this data deluge [27]. For instance, government departments utilise energy consumption data collected via IoT devices and big data analytics to identify residents' energy usage patterns and demand characteristics across different time periods and geographical areas, thereby formulating scientifically sound energy supply plans [28]. Retail enterprises utilise IoT devices in the commercial sphere to gather customer purchase order data, in-store movement trajectories, and product browsing paths. By analysing these to identify residents' consumption preferences and shopping demand patterns, they provide reliable decision-making support for pricing strategies, shelf display optimisation, and inventory management [29].

Social media data represent objective behavioural data and digital footprints that residents generate daily. By analysing this data, researchers can uncover latent demands that residents have not explicitly articulated but that manifest authentically through their actual actions and usage patterns. For instance, researchers utilised over 600,000 Twitter social media data points to achieve accurate short- and long-term traffic forecasting [30]. Another study examined the relationship between social media engagement and theatre demand using official theatre Facebook data to understand public theatre preferences [31]. Furthermore, a novel approach was developed to identify public travel demand by analysing Twitter social media data [32].

By contrast, the public–government interaction functions of administrative platforms can better serve smart city development by identifying resident demand. Citizens proactively consult government departments regarding practical concerns and difficulties in daily life. Based on crowdsourcing principles, this approach can more authentically and comprehensively reflect resident demand, offering novel insights for precise decision-making and continuous optimisation in the people-centred transformation of smart cities. For instance, one study analysed data from the People's Daily Online message board, employing inductive methods to categorise resident demand domains [33]. Another research project utilised US 311 hotline data, integrating geographical information to map and identify service request types across neighbourhoods, thereby providing theoretical foundations for governmental policy planning [34]. Singapore's OneService platform accepts daily feedback on residents' living concerns, employing chatbots to identify and categorise these feedback items [35].

Questionnaires and interviews play a crucial role in identifying resident demand. Their advantages lie in quantifiable outcomes and strong policy transferability, though they face challenges such as high collection costs and insufficient coverage in demand assessment. When employing IoT technology for demand identification, sensor deployment incurs high costs and limited coverage, predominantly applied in structured or semi-structured domains (energy, transport, etc.), making it difficult to address diverse public service demands comprehensively. This approach may overlook residents' implicit expectations and unmet demand. Using keyword identification methods, social media mining can recognise more authentic and granular resident demand. Based on such demands, it offers insights into the people-centred transformation of smart cities. However, issues such as significant data noise and divergent content direction constrain the analysis of resident demand. Because of their functional and institutionalised nature, government service platforms are valuable resources for identifying resident demand. They centralise feedback on public demand. However, existing research has yet to fully leverage the depth of data available from these platforms.

2.2. Forecasting Methods

Research has made considerable progress in demand forecasting for smart cities, primarily centred on utilising structured data generated by domain-specific Internet of Things (IoT) devices and sensors. For instance, researchers have employed machine learning or deep learning models to forecast urban water consumption and energy demand based on historical data from smart water meters and power grids [36–39]. Alternatively, advanced deep learning models such as graph neural networks have been applied to forecast parking demand using data from Internet of Vehicles (IoV) and IoT technologies [40]. Despite the limitations of these studies stemming from their reliance on structured IoT data, the patterns observed in time series forecasting still offer valuable insights for forward-looking demand projections. Traditional time series forecasting methods, exemplified by the Autoregressive Integrated Moving Average (ARIMA) model, are suitable for time series data exhibiting clear linear relationships. However, their forecasting capabilities become increasingly inadequate as data complexity increases. In contrast, machine learning and deep learning approaches generally outperform statistical models [41].

With advances in computational power and the development of big data technologies, computational intelligence methods and deep learning models have found widespread application in time series forecasting. Models such as support vector machines (SVMs) and long short-term memory networks (LSTM) have been extensively studied. Research on residential energy consumption forecasting based on RNNs and LSTMs has demonstrated their significant superiority over traditional methods in capturing temporal dependencies [42]. Concurrently, deep learning has experienced rapid advancements in time series forecasting. Review studies indicate that recurrent neural network models like LSTM demonstrate remarkable efficacy in handling complex patterns and long-term dependencies [43]. Transformer architectures and multimodal data fusion strategies have emerged as cutting-edge approaches [44]. These methods demand substantial data samples; otherwise, models risk overfitting, compromising forecasting performance. Support vector machines, owing to their modest data requirements and robust capability in capturing non-linear time series features, have also demonstrated commendable forecasting capabilities in time series forecasting [45].

However, single models exhibit forecasting limitations, leading to constrained accuracy improvements as time series complexity increases. Consequently, multi-method fusion strategies have emerged in time series forecasting. Research has proposed hybrid

deep learning models integrating CNN, LSTM, and Transformer for solar power generation forecasting, achieving outstanding performance [46]. Scholars have also constructed hybrid forecasting frameworks by combining traditional statistical models with LSTM or integrating LSTM with Transformer, enabling precise forecasting of complex demand patterns [47,48]. Research employing data preprocessing has similarly yielded notable improvements in forecasting performance. Modal decomposition methods, exemplified by VMD and EMD, can effectively enhance the forecasting capability of models [49]. Such fusion strategies for time series forecasting methods allow complementary strengths between approaches, offering novel technical insights for demand forecasting in people-centred transformation within smart cities.

Existing research indicates that time series forecasting offers a viable technical pathway for forecasting resident demand in smart cities. However, three limitations exist. Firstly, in terms of data paradigms, current studies exhibit a path dependency towards structured IoT data. This confines the analytical perspective to the singular domain of physical sensing and perpetuates a top-down approach in smart city demand forecasting, overlooking genuine, more socially driven demands originating from residents' proactive feedback. Secondly, existing studies often base demand forecasting on structured IoT data from a single domain, failing to consider multiple heterogeneous demands and lacking generalisation capabilities across demand categories. Resident demand should originate from residents themselves. Finally, at the methodological level, an applicability gap exists. Existing demand forecasting models based on structured data require a redesigned forecasting method construction when confronted with demand sequences extracted from text, which exhibit high noise and strong non-linear characteristics. Consequently, this study aims to analyse text data from residents' proactive feedback on government platforms to identify resident demand. It proposes a hybrid forecasting framework tailored to these data characteristics, addressing existing research limitations and advancing people-centred smart city development.

3. Research Methodology

3.1. Research Framework

This paper constructs a multi-model hybrid analysis framework comprising four principal stages to systematically identify and forecast resident demand, as illustrated in Figure 1. Firstly, large language models (LLMs) are employed to cleanse and analyse raw data, extracting stop words and eliminating high-frequency vocabulary unrelated to demand to ensure analytical accuracy. Subsequently, the BERTopic (version 0.16.0, available at: https://pypi.org/project/bertopic/, accessed on 10 September 2025) dynamic topic model identifies demand themes, modelling each demand entity as a demand time series at monthly resolution. Variational mode decomposition (VMD) then decomposes the cleaned complex time series into intrinsic mode functions (IMFs), eliminating high-frequency noise modes to yield purer demand time series. Finally, a hybrid sequential forecasting method is constructed, integrating VMD and support vector machines (SVMs). This approach takes the denoised demand time series as input and undergoes training, validation, and forecasting processes to achieve accurate forecasting of resident demand trends ultimately. This framework synthesises LLMs, signal processing, and machine learning methodologies to deliver precise forecasts adaptable to diverse resident demand patterns.

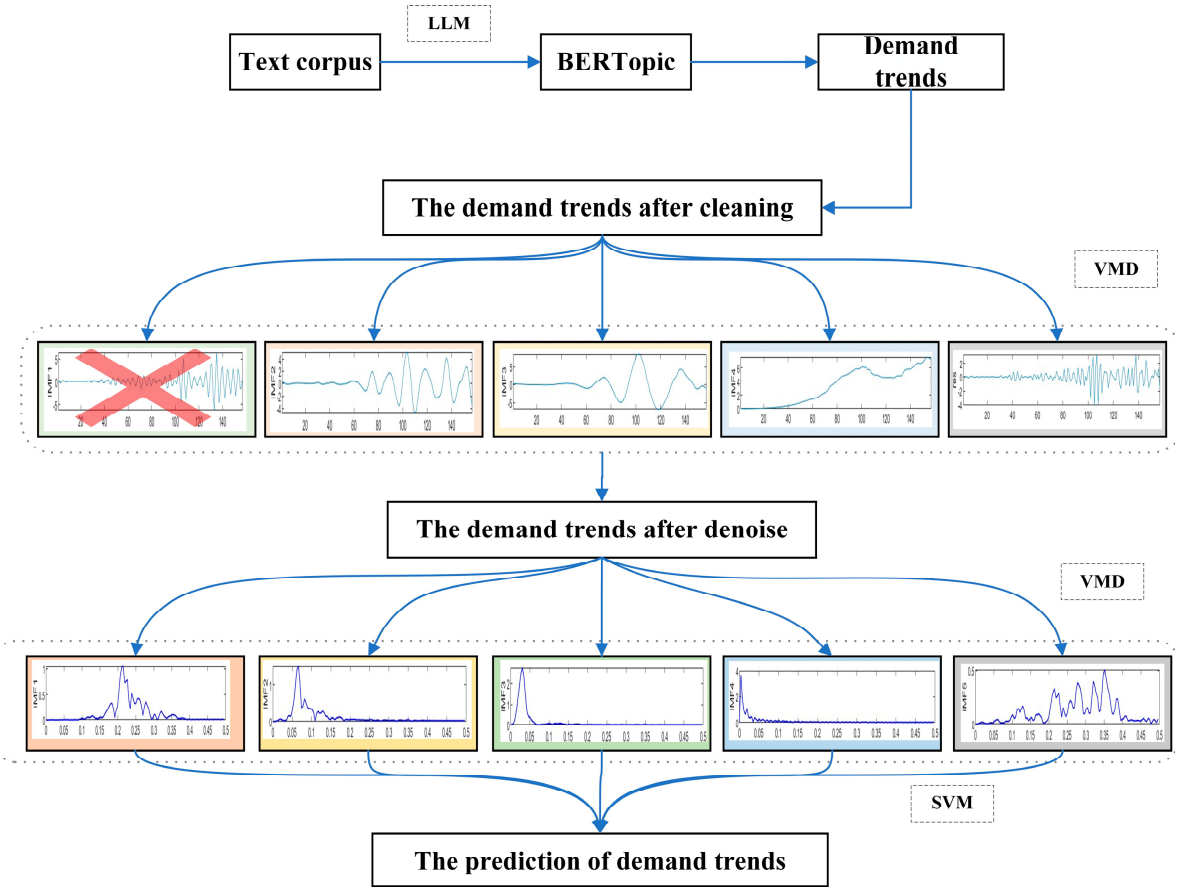


Figure 1. Research framework.

The selection of key models in this framework was based on the specific characteristics of the data and the research objectives. Firstly, in terms of topic identification, although traditional methods such as latent dirichlet allocation (LDA) are widely applied, their reliance on the bag-of-words assumption overlooks semantic relationships between words. Furthermore, they require a predefined number of topics, which limits their recognition capabilities when processing short texts concerning resident demand in this study [50]. In contrast, BERTopic proved to be more suited to this study's requirements due to its robust contextual semantic awareness and mechanism for dynamically determining topic counts. Regarding demand forecasting, compared to data-driven deep learning models like Transformers and LSTMs, this study deals with a limited sample size of monthly resident demand time series. Furthermore, complex models are prone to overfitting on small datasets, compromising forecasting accuracy. In contrast, support vector machines (SVMs) demonstrate robust stability and generalisation capabilities when handling small, non-linear datasets, rendering them a more suitable forecasting model for this research context. Based on these considerations, this paper employs BERTopic for resident demand topic identification and adopts a VMD-SVM stacking approach for demand forecasting.

Correspondingly, the data processing workflow for this study is illustrated in Figure 2. This diagram outlines the various stages from raw data collection to the final demand forecasting and evaluation, including detailed data information such as the volume of data and sample size for each phase.

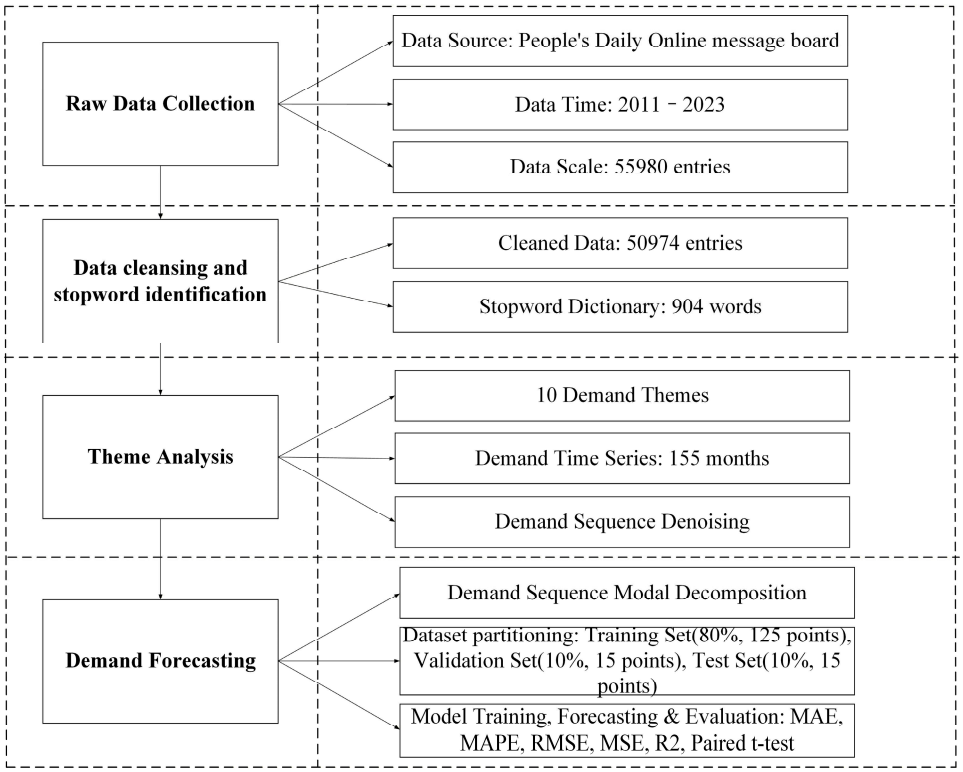


Figure 2. Data flow diagram.

3.2. Data Cleansing and Stopword Identification Based on Large Language Models

User-generated comments lack fixed formatting requirements or standards, resulting in a rather disparate and disorganised format that hinders accurate requirement identification. Leveraging its robust text comprehension and formatting capabilities, an LLM can standardise and cleanse user-generated comments through contextual corpus analysis. Specifically, the LLM can identify and retain core semantic content while eliminating redundant information, correcting formatting errors, and standardising expression styles. This process transforms unstructured user comments into structured, standardised text. This paper employs prompt engineering and utilises the following prompt words to cleanse the data samples: "Hello, I require data cleansing of the corpus to facilitate resident demand identification. Please remove symbols and meaningless characters, format the text, and ensure no semantic alterations occur." This guides the LLM in completing the text standardisation task.

Traditional stopword identification primarily relies on predefined stopword lists, making it difficult to adapt to the specific requirements of different domains and contexts. LLM-based stopword identification methods, through deep semantic understanding, can dynamically recognise words lacking substantive meaning within particular contexts. LLMs can intelligently determine whether a word contributes substantive semantic value in the current context based on the text's subject domain, linguistic style, and expressive conventions, as dictated by prompt requirements. For instance, in resident feedback texts, expressions such as "thank you," "leader," and "good day" bear no direct relevance to the actual request. However, as users frequently employ such polite phrases when submitting feedback, they should be treated as stopwords within the request recognition context. Given the substantial training costs associated with LLM models, this paper employs the pre-trained DeepSeek model based on prompt engineering, stopword identification, and summarisation of data samples. The groundwork is laid for accurately identifying the themes of resident demand. An interface written in Python 3.9 is used to call the LLM

service, utilising the prompt: "Hello, I require the identification of expressions unrelated to resident demand within the corpus, and the construction of a stopword list". Stopword recognition based on LLM provides a higher-quality data foundation for subsequent tasks such as text mining, topic modelling, and sentiment analysis, effectively enhancing the overall performance of natural language processing tasks.

3.3. BERTopic

BERTopic is a topic modelling algorithm based on pre-trained language models combining dimension reduction and clustering techniques. This advanced natural language processing tool automatically extracts meaningful topics from textual data [51]. Its core features include the following: Context-aware processing. Utilises text embeddings generated by pre-trained models such as BERT, preserving semantic and contextual information, making it suitable for handling short texts. Dynamic topic count. Automatically determines the optimal number of topics through clustering, eliminating the need for pre-specified values. Interpretability. Generates highly discriminative topic keywords using the c-TF-IDF algorithm, providing an intuitive representation of topic content. Flexible extensibility. Supports customisation of embedding models, dimensionality reduction methods, and clustering algorithms to accommodate diverse requirements. Visualisation and dynamic analysis. Provides topic distribution maps and hierarchical structure diagrams and supports topic evolution analysis over time series. Its core methodology can be divided into three key stages: text embedding, dimensionality reduction and clustering, and topic representation. This approach combines the strengths of deep learning with traditional statistical methods. The following section provides a detailed explanation of its principles.

3.3.1. Embedding

Objective: To convert text into high-dimensional semantic vectors while preserving contextual information.

Method: Employ pre-trained language models (such as BERT, Sentence-BERT, or RoBERTa) to generate text embeddings.

Mathematical formulation: For each document D_i, generate an embedding vector $e_i \in R^d$, where d denotes the embedding dimension.

Key advantage: Models like BERT utilise self-attention mechanisms to capture inter-word dependencies, resolving semantic ambiguity issues that traditional TF-IDF or LDA approaches cannot address.

3.3.2. Dimensionality Reduction and Clustering

(1) Dimensionality Reduction

Objective: Compress high-dimensional embedding vectors into a low-dimensional space to facilitate clustering.

Method: Employ the UMAP (Uniform Manifold Approximation and Projection) algorithm to preserve both local and global structural information.

Mathematical Principle: Maintain neighbourhood relationships between high-dimensional and low-dimensional spaces by optimising the following loss function:

$$L_{UMAP} = \sum_{i,j} \omega_{ij} \cdot \log\left(\frac{\mathrm{d}_{\mathrm{high}}(i,j)}{\mathrm{d}_{\mathrm{low}}(i,j)}\right) + (1-\omega_{ij}) \cdot \log\left(\frac{1-\mathrm{d}_{\mathrm{low}}(i,j)}{1-\mathrm{d}_{\mathrm{high}}(i,j)}\right) \quad (1)$$

where ω_{ij} denotes the similarity weight between samples i and j in the high-dimensional space.

(2) Clustering

Objective: To partition the reduced-dimensional vectors into clusters.

Method: Employing the HDBSCAN (Hierarchical Density-Based Spatial Clustering) algorithm.

Hierarchical clustering: Constructing a minimum spanning tree to merge clusters of proximate distance.

Density pruning: Automatically selecting clusters based on density stability while discarding low-density regions.

Key advantages: Requires no predefined cluster count; automatically handles outlier points.

3.3.3. Topic Representation

Objective: To generate interpretable keywords for each cluster, representing the thematic content.

Method: Employ the c-TF-IDF (Class-based TF-IDF) algorithm, whose formula is

$$\text{c-TF-IDF} = TF(t,k) \times \log\left(\frac{N}{DF(t)}\right) \tag{2}$$

where $TF(t,k)$ denotes the frequency of term t across all documents within cluster k, and $DF(t)$ represents the number of clusters containing term t.

Keyword extraction: Sort the c-TF-IDF matrix for each cluster by term weight and then select the top N terms as topic labels [52].

3.4. Variational Mode Decomposition

Variational mode decomposition (VMD), proposed by Konstantin Dragomiretskiy [53], is a method for signal decomposition and modal analysis. It enables the adaptive decomposition of input signals, generating several physically meaningful modal components alongside a residual component. From a mathematical modelling perspective, VMD constructs a constrained variational problem to deconstruct the target signal into a finite set of quasi-orthogonal eigenmode functions. Each modal component exhibits compact support in the frequency domain, effectively characterising oscillatory patterns across different temporal scales within the signal. The optimisation objective seeks the optimal solution for each mode within the Hilbert spectral space, ensuring that all components satisfy bandwidth constraints and signal reconstruction fidelity requirements simultaneously. The specific steps of VMD are as follows:

Solving the constrained variational model completes the signal decomposition work by constructing the constrained variational model as follows:

$$\begin{cases} \min\limits_{\{u_k\},\{\omega_k\}} \sum\limits_{k=1}^{K} ||\partial_t[(\delta(t) + \frac{j}{\pi t}) \times u_k(t)]e^{-j\omega_k t}||_2^2 \\ s.t. \sum\limits_{k=1}^{K} u_k(t) = f(t) \end{cases} \tag{3}$$

where $f(t)$ is the original signal and $u_k(t)$ is the Hilbert transform of $(\delta(\mathrm{t}) + \frac{j}{\pi t})u_k(t)$.

The second penalty factor α and the Lagrange multiplication operator $\lambda(t)$ are introduced to transform the above constrained variational model into an unconstrained variational model:

$$L(\{u_k\},\{\omega_k\},\lambda) = \alpha\sum_{k=1}^{K} ||\partial_t[(\delta(t) + \frac{j}{\pi t}) \times u_k(t)]e^{-j\omega_k t}||_2^2 + ||f(t) - \sum_{k=1}^{K} u_k(t)\Big||_2^2 + \left\langle \lambda(t), f(t) - \sum_{k=1}^{K} u_k(t) \right\rangle \tag{4}$$

By alternately updating u_k^{n+1}, ω_k^{n+1}, and λ_k^{n+1} with the alternating direction method of multipliers, the "saddle-point" expression of the Lagrangian expression is obtained as follows.

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2} \tag{5}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \tag{6}$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau_v(\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega)) \tag{7}$$

When the decomposition mode satisfies Equation (8), the VMD iteration stops.

$$\sum_{k=1}^{K} \frac{\left\| \hat{u}_k^{n+1} - \hat{u}_k^n \right\|_2^2}{\left\| \hat{u}_k^n \right\|_2^2} < \varepsilon \tag{8}$$

The VMD method has many applications in signal, image, audio, and other fields. It can be used for tasks such as signal analysis, signal noise reduction, modal analysis, vibration identification, etc., with good adaptivity and local characteristic preservation.

3.5. Support Vector Machine

The support vector machine (SVM) were formally introduced by Vapnik and Cortes in 1995 [54]. This supervised learning algorithm, grounded in statistical learning theory, finds extensive application in classification, regression, and anomaly detection tasks. Its core principle involves constructing an optimal hyperplane to achieve effective partitioning of data within high-dimensional feature spaces while simultaneously maximising the inter-class margin along the classification boundary to attain optimal generalisation performance.

The objective of an SVM is to establish a decision boundary between two categories and ensure maximum separation between classifications [55], that is, to construct a hyperplane within the sample space to distinguish training samples, such that the distance between training samples and the hyperplane is minimised. The hyperplane formula is as follows:

$$\omega\phi(X) + b = 0 \tag{9}$$

In the equation, ω denotes the weight coefficient, b represents the deviation quantity, and φ represents the nonlinear mapping function.

By introducing the Lagrange multiplier, Equation (9) is transformed into a dual problem, with the desired hyperplane represented as follows:

$$\sum_{i=1}^{N} \alpha_i y_i K(x, x_i) + b = 0 \tag{10}$$

In the equation, α_i denotes the Lagrange multiplier, $K(x, x_i)$ represents the kernel function, x_i is the *i-th* training sample, y_i indicates the positive or negative category of the *i-th* sample, and N denotes the number of training samples.

The commonly used kernel functions primarily include polynomial kernel functions, Gaussian radial basis function kernels, and Fourier kernels. The Gaussian radial basis function kernel is characterised by its rapid convergence, strong learning capability, and

widespread application. Hence, this paper employs it to construct a support vector machine model, expressed as follows:

$$K(x, x_i) = \exp\left(-\|x - x_i\|^2 / 2G^2\right) \tag{11}$$

In the equation, G denotes the kernel function parameter. G defines the radial influence range of the kernel function, thereby preventing overfitting or loss of learning capability.

4. Empirical Analysis

4.1. Research Area and Data Sources

Xi'an City in Shaanxi Province, situated in northwestern China, places high importance on smart city development. It has explicitly set the goal of accelerating the creation of a livable, resilient, and intelligent city, achieving significant results in this endeavour. Key government service platforms, such as the data openness platform, have been established. According to Xi'an's 2024 Information Technology Project Scenario Opportunity List, smart city development encompasses digital government construction alongside 15 key sectors, including social welfare, public services, urban management, healthcare, and transport operations. This renders the scientific identification and accurate forecasting of Xi'an's smart city development requirements particularly urgent.

This research utilises data from the People's Daily Online message board; the rationale for employing this as a data source lies in the fact that People's Daily Online stands as China's most significant nationwide online political platform [56], with its message board module (http://liuyan.people.com.cn/index.html, accessed on 10 September 2025) ranking among the country's foremost platforms for government–citizen interaction [33]. Over 2500 municipal and county-level officials, alongside 61 provincial party secretaries and governors, have publicly engaged with citizens through this platform, lending it considerable authority. Here, ordinary Chinese citizens can directly communicate with local and even central leadership, proactively voicing demands and suggestions regarding public services. Furthermore, the People's Daily Online message board operates with transparency and openness, with data readily accessible and low barriers to acquisition. Its value as a resource for studying China's social governance and public demands has gained widespread recognition within academic circles, and numerous research findings have been published to date [33,56–58]. Concerning comments from Xi'an residents in Shaanxi Province from 2011 to 2023, 55,980 entries about urban development, transportation, healthcare, and related matters were selected. The selection of the 2011–2023 timeframe stems from Shaanxi Province's October 2010 promulgation of the Interim Measures for Handling Messages from People's Daily Online Netizens Addressed to Principal Provincial Government Leaders, which institutionalised the People's Daily Online message board system. Consequently, this study commenced in 2011. As this study commenced in mid-2024, 2023 represented the most recent complete annual dataset available at that time; hence, it was designated as the data endpoint. Compared to other platforms such as Weibo and Douyin, the People's Daily Online comment board features anonymity and data anonymisation. Its institutionalised feedback mechanism ensures that comments typically constitute specific, targeted appeals to resolve issues rather than general expressions of sentiment. Consequently, it offers greater representativeness for analysing public demands.

4.2. Data Cleansing

This paper employs DeepSeek's R1 version, utilising Python 3.9 to develop an interface for service invocation, calling pre-trained official models, and employing prompt engineering to achieve data cleansing and stopword identification. Python 3.9 invokes DeepSeek

R1's official API, implementing data sample cleansing and structured processing based on data cleansing prompt words. Through LLM cleansing, manual inspection removed 5006 entries exhibiting ambiguous meaning or irrelevance to resident demand, significantly enhancing data quality. Concurrently, 5000 randomly selected data samples underwent stopword identification via Python 3.9 invocation of the pre-trained DeepSeek R1 API and stopword recognition prompts. Following deduplication and manual screening of these stopwords, an additional list of 158 stopwords was identified. This expanded the existing HIT stopword corpus (https://github.com/goto456/stopwords, accessed on 10 September 2025) from 746 to 904 stopwords relevant to this study, including 'my home,' 'Zhonghai,' 'already,' 'leadership,' and 'thank you,' laying a robust data foundation for subsequent thematic identification of requirements.

Regarding the efficacy of DeepSeek in data cleansing, its capabilities in format and symbol error detection have been validated, with DeepSeek achieving 86.6% accuracy [59] empirically validated its effectiveness and robustness, demonstrating DeepSeek's superior performance in data purification. Consequently, this paper focuses on conducting robustness testing specifically for the "stopword identification" phase. For stopword identification, this study selected popular large-model services, randomly choosing 100 entries from the predefined list as experimental samples to verify the robustness of the proposed method. The Jaccard coefficient [60] was employed to assess the similarity of stopword recognition across models, with the results presented in Table 1. DeepSeek R1 identified 57 stopwords, ChatGPT 4o recognised 54 stopwords with a Jaccard coefficient of 0.8814, Doubao identified 49 stopwords with a Jaccard coefficient of 0.8596, while Gemini detected 51 stopwords, yielding a coefficient of 0.8621. All Jaccard coefficients exceeded 0.85, indicating that the LLM-based approach employed herein demonstrates favourable generalisability and robustness for stopword identification.

Table 1. LLM comparative experiment.

Model Name	API Call URL	Identify The Number of Stop Words	Jaccard Coefficient
DeepSeek-R1	https://platform.deepseek.com/ (Accessed on 10 September 2025)	57	\
ChatGPT-4o	https://platform.openai.com/ (Accessed on 10 September 2025)	54	0.8814
Doubao-seed-1.6	https://api.volcengine.com/ (Accessed on 10 September 2025)	49	0.8596
Gemini-2.5-flash	https://ai.google.dev/gemini-api/ (Accessed on 10 September 2025)	51	0.8621

4.3. Theme Analysis

Over 50,000 user-generated comments were input into the BERTopic model for training, identifying 10 distinct demand themes. The characteristic distribution of words for each theme is shown in Figure 3. The ten demand themes are water, go to school, road planning, demolition transition, loan policy, park green space, heating, fire hazards, medical care, and public transportation.

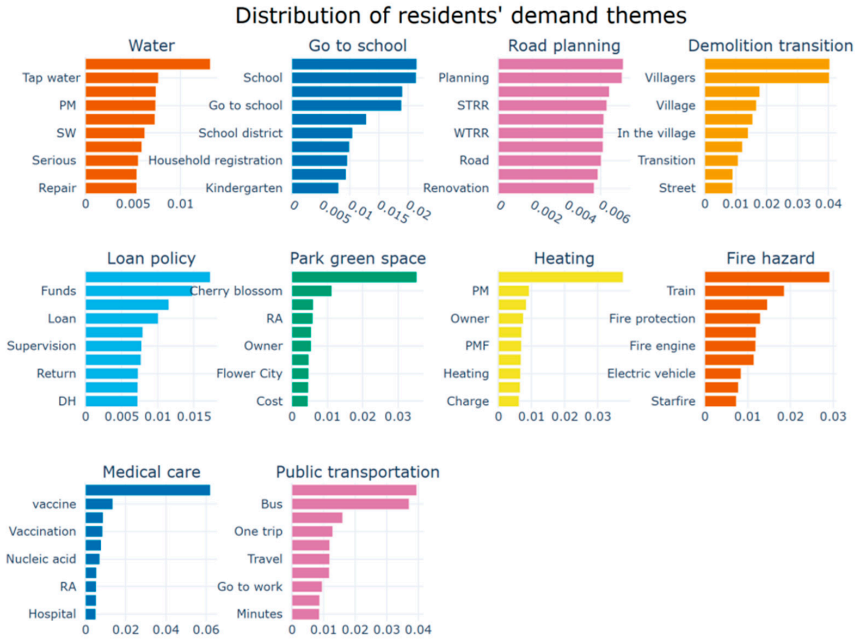


Figure 3. Resident demand theme distribution.

4.4. Trend Evolution of Resident Demand Sequences

This paper models the demand topics identified by BERTopic as a monthly time series, as illustrated in Figure 4. Analysis of the temporal evolution of resident demand reveals distinct phased characteristics and patterns of sudden fluctuations. From 2011 to 2018, resident demand remained at a relatively stable low level, with no category exceeding 50 units in intensity. From 2018 onwards, alongside the widespread adoption of information technologies, all categories of resident demand exhibited a marked upward trend. According to the 2018 Xi'an Municipal Government Work Report, that year saw accelerated development of Smart Xi'an, including establishing a city operations big data centre and a government service platform, alongside expedited advancement of public service digitalisation for citizens. Under government policy guidance and proactive promotion, resident demand feedback gradually shifted towards digital channels, culminating in two distinct peaks in 2020 and 2022. It is noteworthy that a low-level plateau persisted between 2020 and 2022, primarily due to pandemic control measures restricting residents' daily movements and consequently reducing related demand. The demand peaks in 2020 and 2022, both centred on schooling issues. Taking the surge in demand for school education around 2020 as an example, its primary driver stemmed from the uncertainty brought about by the 'reformative' nature of policy. As the enrolment demand of China's school-age population continued to rise, the 2019 'Opinions of the Central Committee of the Communist Party of China and the State Council on Deepening Educational Reform and Comprehensively Improving the Quality of Compulsory Education' and the 'Notice of the General Office of the Ministry of Education on Doing Well in the 2019 Enrolment Work of Ordinary Primary and Secondary Schools' mentioned initiatives such as 'simultaneous enrolment of citizens and migrant workers' children' and 'local enrolment of children accompanying migrant workers' to pursue educational equity. Subsequently, Shaanxi Province incorporated these elements into its key priorities, as these policy components were rarely discussed previously, such as local enrolment for migrant children," to promote educational equity. Subsequently, Shaanxi Province incorporated these priorities into its work agenda. As these policy elements were previously under-emphasised and reform-oriented, the 2020 Shaanxi Provincial Education Department Notice on Conducting 2020 General High School Admissions introduced concepts like simultaneous enrolment for citizens and local enrolment for the first time. This disrupted established pathways for advancement. The public lacked experience navigating decisions

under the new regulations, sparking intense debate and inquiries regarding foundational rules. Issues such as school district delineation, student household registration status, neighbourhood enrolment, and district affiliation became prominent topics, aligning with the thematic keywords under Theme 2: Enrolment Education in Figure 3.

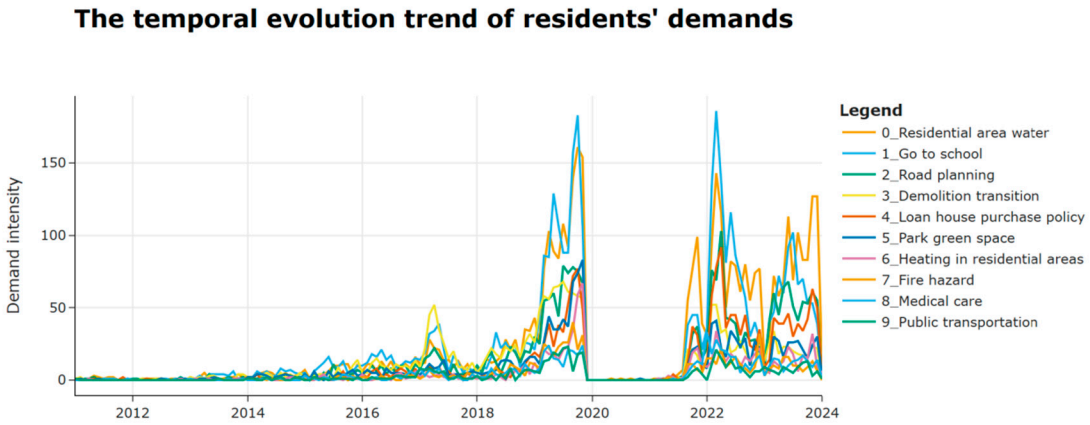


Figure 4. The temporal evolution trend of resident demand.

In contrast, the demand peak observed in 2022 reflected the public's refined demand for interpreting policy 'complexities' after initial adaptation to the reforms. The Chinese Ministry of Education issued the 'Notice of the General Office of the Ministry of Education on Further Improving the Enrolment Work of Ordinary Primary and Secondary Schools,' emphasising the need to enhance the scientific, institutionalised, and standardised nature of enrolment procedures, establish a long-term mechanism for equitable access, and further refine policies concerning 'simultaneous enrolment of citizens and non-citizens', 'neighbourhood enrolment', and other matters. The Shaanxi Provincial Education Department, adapting to local circumstances, subsequently issued the Notice of the Office of the Shaanxi Provincial Education Department on Doing Well in the 2022 Compulsory Education Enrolment Work, which deepened and refined issues such as 'citizen enrolment alongside students from other regions' and neighbourhood enrolment. 'Policy content became' multiple and complex, 'shifting public focus from' what it is 'to' how to implement it. The heated debate no longer centres solely on fundamental rules but delves into operational uncertainties, such as specific enrolment variations between different 'schools' and the transition between old and new 'hukou' policies, seeking optimal pathways within the complex policy framework. This research methodology for identifying resident demand enables more refined public sentiment monitoring, capturing citizens' demands and requirements. It establishes a robust theoretical foundation for analysing the driving mechanisms behind the formation of demand-related public sentiment. Concurrently, the associated conclusions can guide the application of technologies such as AI to empower smart city development, assign new responsibilities to smart city technologies, and promote the people-centred realisation of smart city construction.

4.5. Resident Demand Sequence Denoising

This study employed VMD to perform denoising on ten demand time series to enhance the quality of the demand time series. VMD decomposes the original time series into several intrinsic modal functions with distinct frequency characteristics through an adaptive decomposition process. As noise tends to concentrate in high-frequency modes, this paper adopted a strategy of retaining low-frequency modes while discarding the highest-frequency modes, thereby maximising the preservation of effective feature information in the demand time series. The denoising results are illustrated in Figure 5, where "origin" denotes the original sequence and "denoise" represents the denoised sequence. A comparative analysis indicates

that relative to the original demand time series, the denoised sequence more accurately represents the temporal trend of demand. VMD effectively filters out high-frequency noise and random fluctuations while preserving the data's primary trend characteristics and periodic components. This significantly enhances the sequence's smoothness, and the VMD-denoised demand time series better reflects the true data distribution patterns. Consequently, the forecasting model's generalisation capability and forecast stability are improved.

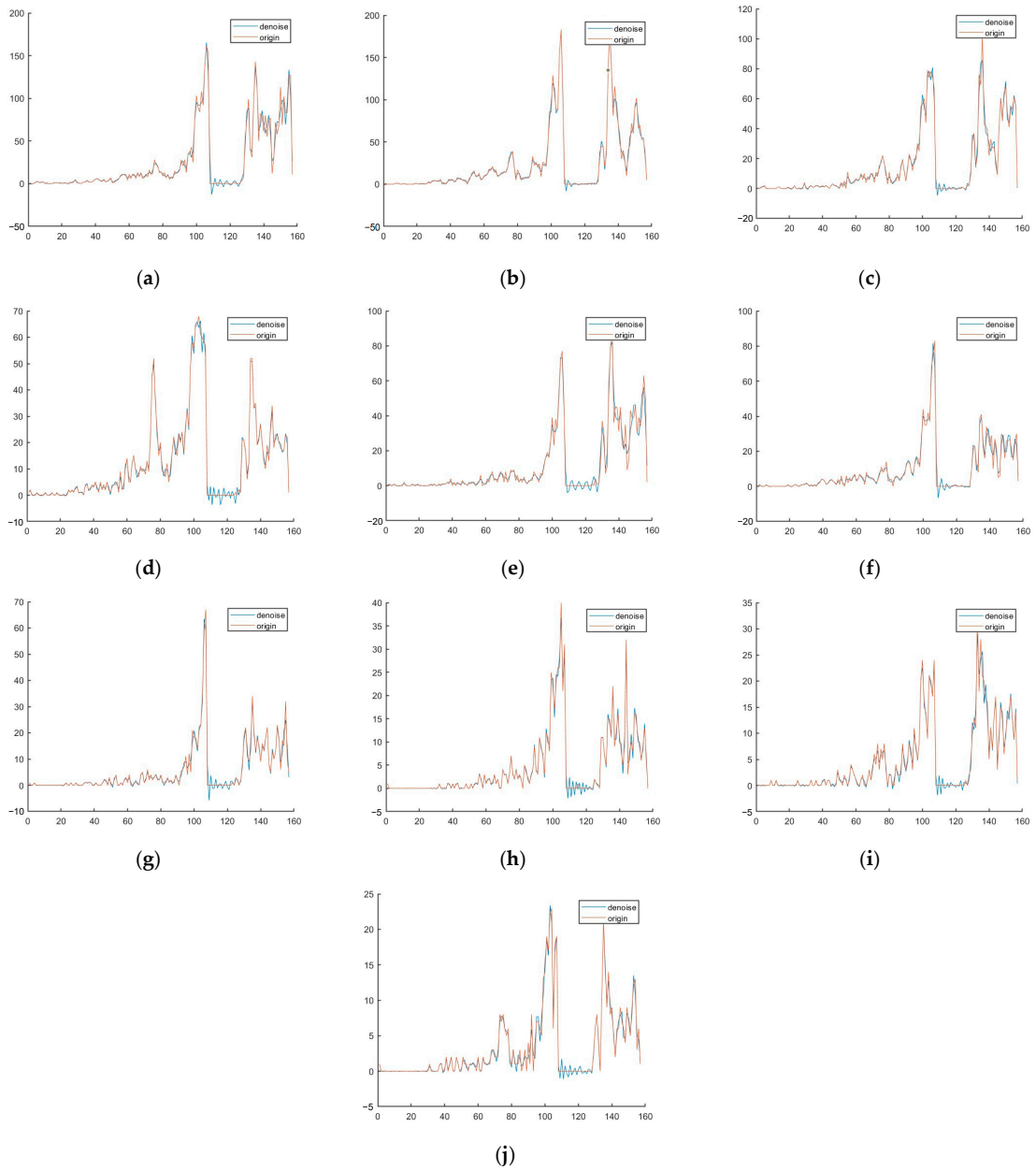


Figure 5. VMD denoising results. (**a**) Water; (**b**) Go to school; (**c**) Road planning; (**d**) Demolition transition; (**e**) Loan policy; (**f**) Park green space; (**g**) Heating; (**h**) Fire hazard; (**i**) Medical care; (**j**) Public transportation.

4.6. Resident Demand Sequence Modal Decomposition

Further decomposition of the ten denoised demand time series was performed using VMD, as illustrated in Figure 6. Each sub-figure illustrates the decomposition process for a single demand series. This paper decomposes the original demand time series into five eigenmode functions with distinct frequency characteristics. The decomposition results reveal markedly differentiated oscillatory patterns and frequency domain distributions

across the modal components: the high-frequency modes (positioned above) primarily capture rapid fluctuations within the demand series, encompassing residual noise and short-term random disturbances; mid-frequency modes effectively extract the periodic variation patterns of the demand time series, reflecting seasonal or policy-driven cyclical characteristics; and low-frequency modes (positioned at the bottom) accurately depict the long-term evolutionary trends and fundamental direction of the demand time series. Each modal component exhibits more regular and stable temporal characteristics following VMD decomposition. This enhanced regularity amplifies the periodic temporal features within the data, making them more readily captured and learned by the SVM. Consequently, the model's ability to recognise complex non-linear time-dependent relationships is improved, providing the SVM with richer, more discriminative temporal feature information. This establishes a robust data foundation for subsequent demand forecasting modelling.

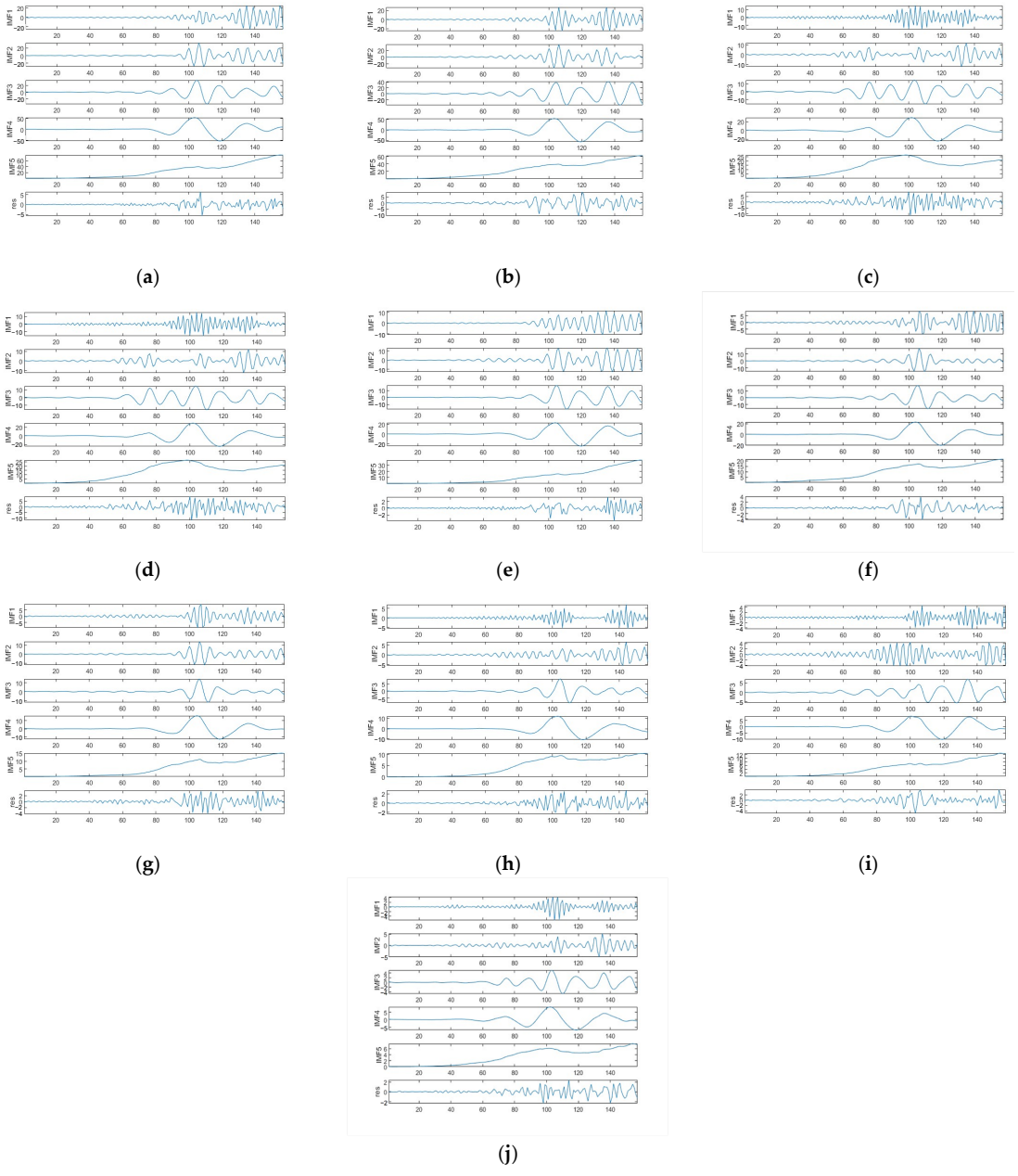
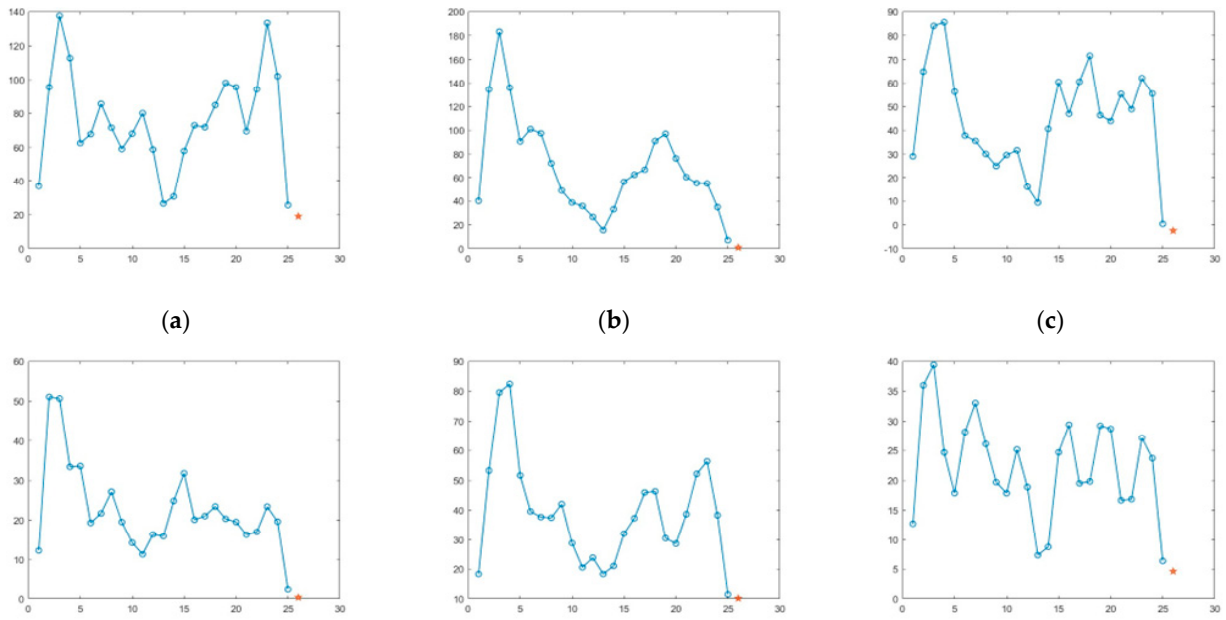


Figure 6. VMD decomposition results. (**a**) Water; (**b**) Go to school; (**c**) Road planning; (**d**) Demolition transition; (**e**) Loan policy; (**f**) Park green space; (**g**) Heating; (**h**) Fire hazard; (**i**) Medical care; (**j**) Public transportation.

4.7. Forecast Trends in Resident Demand

Based on the modal components derived from VMD decomposition, independent SVM forecast models were established for each component, fully utilising their specific time–frequency characteristics. Following forecast completion, the results from each component were integrated to yield the final composite forecast value. As illustrated in Figure 7, among the ten demand themes, all exhibited a declining trend except for Theme 8 (Fire Hazard) and Theme 9 (Medical Care), which demonstrated an upward trajectory. A deeper analysis of the causes reveals that the forecast time point was January 2024, coinciding with the period immediately preceding the Chinese New Year. During this festive season, traditional celebratory activities such as setting off firecrackers and releasing sky lanterns frequently increase fire hazards. However, the heightened festive atmosphere also leads to a significant rise in public awareness and the demand for fire safety safeguards. On 1 January 2024, Xi'an Municipality formally released the "Home-Based Medical Bed Service Guidelines", addressing healthcare demand for patients requiring prolonged bed rest due to illness, physical frailty, or inability to care for themselves. Following the policy's introduction, residents generated substantial inquiry demands regarding service eligibility criteria, reimbursement policy applicability, and application procedures. Using this rising demand trend as an example, we illustrate how demand identification and forecasting drive smart city transformation: traditional customs like fireworks pose safety management challenges in modern urban governance. Smart city initiatives deploy digital solutions—such as intelligent surveillance networks and emergency response systems—to safeguard festive security while preserving cultural traditions, enabling residents to celebrate the Spring Festival with peace of mind. For complex medical inquiries, smart city initiatives can introduce AI-powered intelligent Q&A platforms offering round-the-clock online services. These provide precise, one-to-one responses to residents' individual queries, streamlining administrative processes, enhancing the convenience and efficiency of public services, and advancing the development of people-centred smart cities.



(**a**) (**b**) (**c**)

(**d**) (**e**) (**f**)

Figure 7. *Cont.*

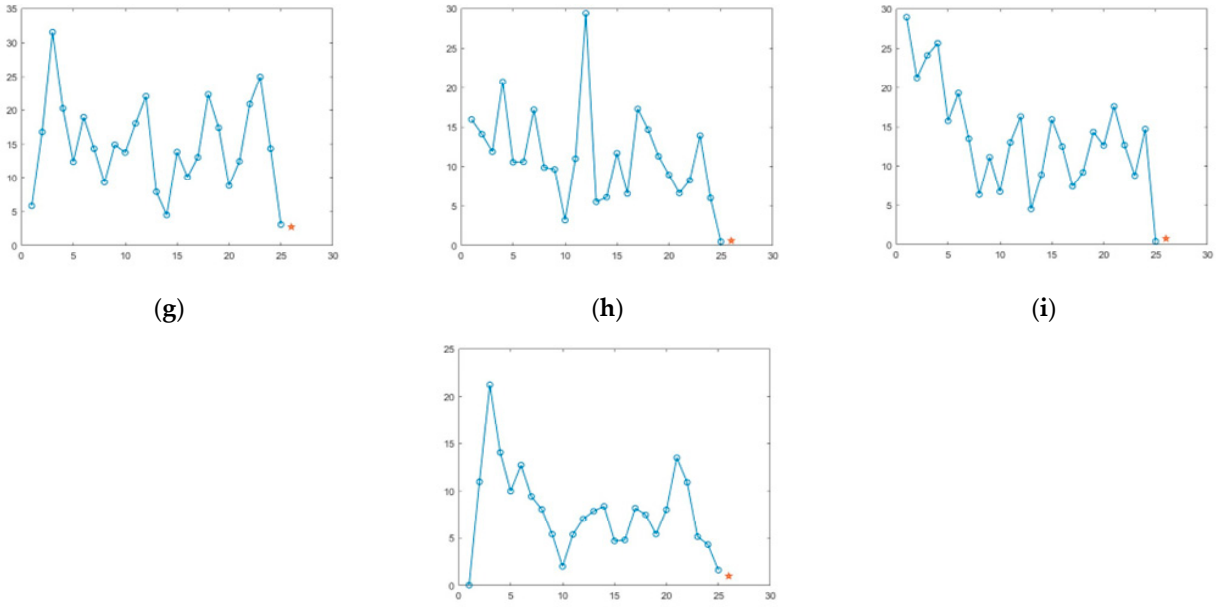


Figure 7. Forecast trends in resident demand. (**a**) Water; (**b**) Go to school; (**c**) Road planning; (**d**) Demolition transition; (**e**) Loan policy; (**f**) Park green space; (**g**) Heating; (**h**) Fire hazard; (**i**) Medical care; (**j**) Public transportation. The five-star symbols represent the forecast values.

4.8. Denoising Assessment of the Residential Demand Forecasting Model

To comprehensively evaluate model performance, this study employs a comparative analysis using standard metrics for machine learning forecast models, encompassing five core indicators: mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE), mean squared error (MSE), and coefficient of determination (R^2). Mainstream forecast algorithms from the current literature were selected as benchmark models, including SVM [61], VMD-SVM [62], BiLSTM [63], CNN-LSTM [64], and VMD-CNN-LSTM [65]. To ensure the fairness of the algorithms, this paper employs the Grey Wolf Optimisation (GWO) algorithm [66] to refine the hyperparameters of each algorithm. For the support vector machine (SVM) algorithm, optimisation targeted three hyperparameters: the regularisation parameter BoxConstraint, the kernel function coefficient KernelScale, and Epsilon [67]. For the BiLSTM, CNN-LSTM, and VMD-CNN-LSTM algorithms, optimisation focused on the learning rate, hidden layer nodes, dropout rate, and L2 regularisation [68]. A unified GWO strategy was applied to all models. The dataset was partitioned chronologically into training (80%), validation (10%), and test (10%) sets. Thirty iterations were conducted with minimising MSE as the objective function. The search space for hyperparameters and the optimal ranges for each model are detailed in Table 2.

Table 2. Model hyperparameter settings.

Hyperparameters	Search Space	Optimal Value Range					
		SVM	VMD-SVM	DN-VMD-SVM	BiLSTM	CNN-LSTM	VMD-CNN-LSTM
BoxConstraint	[0.01 5]	[0.8 4.3]	[1.2 3.2]	[0.8 3.7]	-	-	-
KernelScale	[0.01 30]	[4.7 23.4]	[3.3 19.4]	[6.0 21.9]	-	-	-
Epsilon	[0.001 1]	[0.2 0.9]	[0.1 0.7]	[0.2 0.8]	-	-	-
Learning rate ($\times 10^{-3}$)	[0.1 10]	-	-	-	[5 8]	[4 10]	[2 10]
Hidden layer nodes	[32 256]	-	-	-	[99 122]	[39 87]	[32 98]
Dropout ($\times 10^{-1}$)	[0 5]	-	-	-	[2 4]	[1 3]	[0.7 1]
L2 ($\times 10^{-3}$)	[0.1 10]	-	-	-	[3 4]	[2 4]	[2 4]

These were comprehensively compared with the proposed DN-VMD-SVM model across ten demand themes, with the results presented in Table 3. Analysis indicates that the proposed DN-VMD-SVM model achieved optimal performance across all ten demand theme forecast tasks, demonstrating outstanding forecasting stability. The R^2 values were stably distributed within the 0.93–0.96 range, indicating an exceptionally high degree of goodness-of-fit. Regarding error control, the model's MAE was maintained within 0.53–5.23, the MAPE remained at a low level of 0.07–0.80, the RMSE ranged between 0.61 and 5.96, and the MSE was distributed within 0.37–35.5. Compared to other models, DN-VMD-SVM demonstrated significant performance advantages and forecasting stability across all evaluation metrics. This indicates that the proposed model can be effectively applied to forecasting various resident demands.

Table 3. Resident demand forecasting results.

Demand Theme	Algorithm Name	MAE	MAPE	RMSE	MSE	R^2
Water	SVM	16.75	0.33	20.99	440.67	0.49
	VMD-SVM	8.21	0.16	10.22	104.40	0.91
	DN-VMD-SVM	5.23	0.80	5.96	35.50	0.96
	BiLSTM	39.49	0.63	44.57	1986.29	−1.30
	CNN-LSTM	38.65	0.50	44.99	2023.97	−1.34
	VMD-CNN-LSTM	41.61	0.64	43.61	1902.23	−1.20
Go to school	SVM	9.07	0.31	11.53	133.04	0.79
	VMD-SVM	5.83	0.19	6.57	43.20	0.94
	DN-VMD-SVM	4.16	0.13	4.71	22.14	0.96
	BiLSTM	28.49	0.68	32.45	1053.13	−0.67
	CNN-LSTM	37.52	1.02	42.19	1779.93	−1.82
	VMD-CNN-LSTM	29.41	0.62	32.53	1057.91	−0.68
Road planning	SVM	11.43	6.92	15.49	240.08	0.40
	VMD-SVM	4.35	0.19	5.49	30.17	0.92
	DN-VMD-SVM	3.30	0.20	3.95	15.63	0.96
	BiLSTM	21.07	1.89	27.13	736.29	−0.84
	CNN-LSTM	23.01	4.29	26.95	726.29	−0.82
	VMD-CNN-LSTM	24.41	0.71	26.45	699.74	−0.75
Demolition transition	SVM	5.07	0.67	7.37	54.31	−0.34
	VMD-SVM	1.52	0.10	1.91	3.66	0.93
	DN-VMD-SVM	1.22	0.07	1.55	2.39	0.94
	BiLSTM	6.90	0.83	8.91	79.35	−0.96
	CNN-LSTM	8.95	0.86	10.20	104.10	−1.57
	VMD-CNN-LSTM	4.29	0.33	5.07	25.68	0.37
Loan policy	SVM	7.97	0.30	8.80	77.42	0.52
	VMD-SVM	3.62	0.24	4.32	18.67	0.92
	DN-VMD-SVM	2.10	0.08	2.53	6.41	0.96
	BiLSTM	16.92	0.54	19.45	378.30	−1.35
	CNN-LSTM	26.16	0.83	28.44	808.86	−4.02
	VMD-CNN-LSTM	22.64	0.73	23.19	537.57	−2.33
Park green space	SVM	6.11	0.42	6.67	44.4695	0.22
	VMD-SVM	2.60	0.17	3.08	9.49	0.88
	DN-VMD-SVM	1.35	0.07	1.56	2.44	0.96
	BiLSTM	9.73	0.63	12.18	148.42	−1.61
	CNN-LSTM	7.53	0.57	9.50	90.17	−0.58
	VMD-CNN-LSTM	8.21	0.40	9.20	84.62	−0.49
Heating	SVM	4.72	0.42	5.44	29.61	0.29
	VMD-SVM	1.78	0.18	2.22	4.95	0.91
	DN-VMD-SVM	1.24	0.15	1.41	1.98	0.95
	BiLSTM	6.82	0.87	9.28	86.08	−1.07
	CNN-LSTM	6.83	0.78	8.88	78.93	−0.90
	VMD-CNN-LSTM	6.75	0.61	7.51	56.36	−0.36

Table 3. *Cont.*

Demand Theme	Algorithm Name	MAE	MAPE	RMSE	MSE	R^2
Fire hazard	SVM	6.02	Inf	8.37	70.08	−0.41
	VMD-SVM	1.91	Inf	2.31	5.36	0.89
	DN-VMD-SVM	1.07	0.27	1.34	1.81	0.96
	BiLSTM	6.38	1.63	8.65	74.78	−0.76
	CNN-LSTM	6.46	2.09	8.76	76.81	−0.81
	VMD-CNN-LSTM	3.80	0.45	5.31	28.20	0.34
Medical care	SVM	5.02	2.09	5.78	33.46	−0.61
	VMD-SVM	1.26	0.19	1.51	2.29	0.89
	DN-VMD-SVM	0.92	0.19	1.20	1.45	0.93
	BiLSTM	7.64	2.00	9.00	80.90	−2.89
	CNN-LSTM	7.27	1.37	8.62	74.25	−2.57
	VMD-CNN-LSTM	6.07	0.98	6.89	47.44	−1.28
Public transportation	SVM	2.34	0.44	2.75	7.56	0.04
	VMD-SVM	1.15	0.18	1.45	2.12	0.78
	DN-VMD-SVM	0.53	0.10	0.61	0.37	0.95
	BiLSTM	4.11	0.76	4.77	22.75	−1.90
	CNN-LSTM	3.97	0.85	5.01	25.11	−2.20
	VMD-CNN-LSTM	2.10	0.31	2.58	6.68	0.15

To further statistically validate the superiority of the DN-VMD-SVM model, this paper employs a paired *t*-test to assess the significance of differences in model forecasting performance [69]. The null hypothesis (H_0) posits no significant difference in forecast error between the DN-VMD-SVM model and the baseline model. The alternative hypothesis (H_1) asserts that the mean difference in forecast error between the DN-VMD-SVM model and the baseline model is significantly different [70]. The paired *t*-test requires the difference series between DN-VMD-SVM and the comparison model to follow a normal distribution [71]. Therefore, the Kolmogorov–Smirnov test was employed to assess the normality of model differences, with the results summarised in Table 4. The K-S-p value column indicates the *p*-value from the normality test. It was determined that $p > 0.05$ for all models across the ten demand themes, satisfying the normality requirement for paired *t*-tests. The paired *t*-test returns h-values and *p*-values: h = 0 and $p > 0.05$ indicate failure to reject H_0. In contrast, h = 1 and $p < 0.05$ rejects H_0, signifying a statistically significant difference between the models at the 95% confidence level. Table 4 presents the paired *t*-test results for DN-VMD-SVM versus other methods across the 10 demand themes. Here, the *t*-test (h) denotes the h value, and the *t*-test (*p*) denotes the *p* value. When h = 1 and $p < 0.05$, the null hypothesis is rejected. The results indicate that across 50 experimental groups, 47 yielded h = 1 and $p < 0.05$, representing a significant proportion of 94%. This confirms that the forecasting superiority of the proposed DN-VMD-SVM model in Table 3 is not coincidental, but demonstrates statistically significant robustness.

Table 4. Paired *t*-test comparison of the DN-VMD-SVM with other models.

Demand Theme	SVM			VMD-SVM			BiLSTM			CNN-LSTM			VMD-CNN-LSTM		
	K-S *p*-Value	*t*-Test (h)	*t*-Test (p)	K-S *p*-Value	*t*-Test (h)	*t*-Test (p)	K-S *p*-Value	*t*-Test (h)	*t*-Test (p)	K-S *p*-Value	*t*-Test (h)	*t*-Test (p)	K-S *p*-Value	*t*-Test (h)	*t*-Test (p)
Water	0.734	1	0.007	0.707	1	0.033	0.944	1	0.000	0.852	1	0.000	0.504	1	0.000
Go to school	0.853	1	0.027	0.658	1	0.016	0.516	1	0.000	0.952	1	0.000	0.998	1	0.000
Road planning	0.896	1	0.012	0.566	0	0.042	0.262	1	0.001	0.960	1	0.000	0.997	1	0.000
Demolition transition	0.907	1	0.018	0.659	0	0.543	0.383	1	0.002	0.814	1	0.000	0.999	1	0.001
Loan policy	0.651	1	0.000	0.614	1	0.008	0.453	1	0.000	0.986	1	0.000	0.932	1	0.000
Park green space	0.947	1	0.000	0.618	1	0.025	0.876	1	0.000	0.514	1	0.002	0.654	1	0.000
Heating	0.797	1	0.000	0.871	0	0.144	0.599	1	0.005	0.369	1	0.002	0.780	1	0.000
Fire hazard	0.758	1	0.005	0.323	1	0.032	0.250	1	0.003	0.810	1	0.004	0.541	1	0.007
Medical care	0.921	1	0.000	0.572	0	0.346	0.887	1	0.000	0.768	1	0.000	0.998	1	0.000
Public transportation	0.664	1	0.000	0.853	1	0.017	0.877	1	0.000	0.960	1	0.000	0.675	1	0.003

4.9. Discussion

Regarding demand identification, this paper analysed comments left by Xi'an residents on People's Daily Online. Findings indicate that people-centred smart city development should prioritise transport and mobility (road planning, public transport), infrastructure (heating, water, fire hazard), education (go to school), healthcare (medical care), policy and legal matters (loan policy, demolition transition), and environmental greening (park green space). This aligns with findings from Bernd's research in Germany [72] and Bosco et al.'s study in Italy [12], indicating that despite cultural and geographical differences, the fundamental demand of modern urban residents exhibits certain similarities and universality. However, unlike existing studies that often categorise demand into macro-level types, this paper employs the BERTopic thematic modelling method to identify resident demand themes from a finer-grained perspective. For instance, within infrastructure demands, distinct dimensions such as 'heating', 'water supply', and 'fire hazards' are distinguished, with each further subdivided into more granular thematic keywords.

The DN-VMD-SVM model proposed herein demonstrates commendable demand forecasting performance. As evidenced by Tables 3 and 4, the DN-VMD-SVM model significantly outperforms the VMD-SVM algorithm. As illustrated in Figure 4, the denoising algorithm moderates abrupt fluctuations in the raw demand time series data, thereby rendering the underlying trend more discernible. This stabilises the VMD-SVM training process, ultimately yielding more precise forecasts. This finding aligns with the conclusions drawn by Tang et al. [73]. The VMD algorithm accentuates features within demand time series data, facilitating learning by the SVM algorithm. Consequently, the overall forecasting performance of the VMD-SVM algorithm surpasses that of the SVM algorithm alone, which aligns with the findings of Meng et al. [74]. Although algorithms such as BiLSTM, CNN-LSTM, and VMD-CNN-LSTM demonstrate outstanding forecasting performance across various domains [75], they typically involve substantial parameter sets requiring extensive training samples to learn effectively and avoid overfitting. However, the demand time series lengths in this study are limited, falling far short of the ideal training scale for deep learning models. Under small-sample conditions, complex models tend to 'memorise' noise in training data rather than learning genuine patterns. The intricate structures of the limited-length demand time series in this study often prove difficult to capture fully, resulting in diminished generalisation capabilities and consequently poorer forecasting performance. This aligns with findings from Tuğrul [76] and Ahmadi [77]. This finding holds implications for forecasting modelling in smart city domains: Not all scenarios warrant the application of state-of-the-art deep learning techniques. Selecting appropriate models based on data characteristics often proves more effective than blindly pursuing complexity. This study advocates a 'data-feature-driven model selection' strategy. Simple yet robust hybrid models may represent a superior choice in urban governance scenarios with limited data. This conclusion is of particular importance for city administrators who are constrained by budgetary and data resource limitations. Rather than pursuing technological complexity indiscriminately, they can leverage lightweight, efficient forecasting frameworks—such as the one proposed herein—to proactively anticipate citizen demand. This enables a more precise and timely allocation of public service resources.

5. Conclusions and Outlook

5.1. Conclusions

This study aims to construct a data-driven analytical framework, termed LLM-BERTopic-VMD-SVM, to identify and forecast resident demand. This provides a viable pathway for transforming smart city development from technology-centric to people-centred. Empirical results demonstrate that the proposed LLM-BERTopic-VMD-SVM framework

exhibits outstanding performance in identifying and forecasting resident demand. Using data from Xi'an spanning 2011 to 2023 as an example, ten categories of resident demand were identified. A comparison with relevant international research reveals that the conclusions drawn from the demand analysis are largely consistent. This paper models each demand category as a time series. Considering the limited sample size of the sequences, DN-VMD-SVM is employed for demand forecasting. The forecast R^2 values for various demand themes remain between 0.93 and 0.96, indicating stable forecasting performance for heterogeneous demands. Comparisons with other mainstream forecasting methods using MAE, MSE, MAPE, RMSE, and paired *t*-tests against other mainstream forecasting methods. Compared to other mainstream forecasting methods, it demonstrates higher forecast accuracy and stability, showcasing good universality and generalisation capabilities in resident demand forecasting. This research not only establishes an effective framework for identifying and forecasting resident demand but also provides decision-support tools for urban administrators to implement proactive, refined governance. By achieving high-quality demand identification alongside forward-looking forecasting analysis, the framework offers a feasible, actionable empirical pathway to advance the transformation of smart cities from technology-centric to people-centric approaches.

5.2. Limitations

This study has the following limitations:

(1) The identification and forecasting of resident demand were conducted using data from a single source. Although satisfactory results were achieved for demand identification and forecasting accuracy, sampling bias may still exist.

(2) The study focuses on Xi'an City in Shaanxi Province, China. Consequently, its findings exhibit distinct characteristics influenced by Xi'an's governance structure, cultural context, resource endowments, and business landscape, thereby limiting the generalisability of the research conclusions.

(3) While establishing a comprehensive framework from demand identification to forecasting, the study's capacity to provide deeper causal explanations for underlying demand patterns remains to be enhanced.

5.3. Outlook

Given the aforementioned limitations, the research outlook for this paper is proposed as follows:

(1) Future research will incorporate multi-source heterogeneous data with enhanced temporal resolution, including real-time public sentiment data, social media data, and mobile communications data. Concurrently, support for multimodal data such as street view imagery, audio, and video will be expanded. This will enable the identification and forecasting of resident demand from a more comprehensive perspective, further advancing smart cities' people-centred development. Following the introduction of multi-source real-time and multimodal data—which originate from residents' daily activities—strict privacy protection mechanisms must be established. This requires residents' informed consent before data collection and employs techniques such as federated learning, differential privacy, and irreversible anonymisation during analysis. Such measures ensure accurate identification and forecasting of resident demand without compromising privacy. Given variations in data temporal resolution and modalities, selecting LLM models with multimodal processing capabilities and fine-tuning them enables effective handling of such data. Applying this framework achieves a more comprehensive perception of resident demand.

(2) Future research will extend the scope of this study to national and even global scales. Building upon the demand identification and forecasting framework presented herein, it will identify and forecast resident demand on larger scales, explore commonalities and differences in resident demand across varying contexts, conduct deeper analyses, reveal general patterns in demand formation, and further enhance the theoretical significance and practical value of this research. Regarding feedback channels for resident demand across global cities, adaptive fine-tuning of the LLM data preprocessing module within this framework must account for linguistic and cultural variations. Resident demand themes identified via BERTopic require validation and verification by local experts. Demand forecasting methods necessitate adaptive optimisation and adjustment based on demand characteristics to achieve more universally applicable forecasts.

(3) Future research may build upon this demand identification and forecasting framework by incorporating the double difference approach and causal analysis methods. Integrating policy and economic factors into the analysis could delve deeper into the correlations and causalities underlying demand formation across different geographical regions. This would aim to provide distinctive policy recommendations for the people-centred development of smart cities on a larger scale.

Author Contributions: Conceptualisation, W.Z. (Wen Zhang) and B.G.; data curation, W.Z. (Wei Zhao), Y.H. and X.W.; formal analysis, W.Z. (Wei Zhao); methodology, W.Z. (Wen Zhang); writing—original draft, W.Z. (Wen Zhang) and Y.H.; writing—review and editing, W.Z. (Wei Zhao) and B.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Hebei Province Major Science and Technology Support Plan Project (252D6101D).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author(s).

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

References

1. Blasi, S.; Ganzaroli, A.; De Noni, I. Smartening sustainable development in cities: Strengthening the theoretical linkage between smart cities and SDGs. *Sustain. Cities Soc.* **2022**, *80*, 103793. [CrossRef]
2. Kashef, M.; Visvizi, A.; Troisi, O. Smart city as a smart service system: Human-computer interaction and smart city surveillance systems. *Comput. Hum. Behav.* **2021**, *124*, 106923. [CrossRef]
3. Alahi, M.E.E.; Sukkuea, A.; Tina, F.W.; Nag, A.; Kurdthongmee, W.; Suwannarat, K.; Mukhopadhyay, S.C. Integration of IoT-Enabled Technologies and Artificial Intelligence (AI) for Smart City Scenario: Recent Advancements and Future Trends. *Sensors* **2023**, *23*, 5206. [CrossRef]
4. Mohammed, A.A.-J.; Burhanuddin, M.; Basiron, H.; Tunggal, D. Key enablers of IoT strategies in the context of smart city innovation. *J. Adv. Res. Dyn. Control Syst.* **2018**, *10*, 582–589.
5. Beneicke, J.; Juan, A.A.; Xhafa, F.; Lopez-Lopez, D.; Freixes, A.J.I.C.E.M. Empowering citizens' cognition and decision making in smart sustainable cities. *IEEE Consum. Electron. Mag.* **2019**, *9*, 102–108. [CrossRef]
6. Rejeb, A.; Rejeb, K.; Simske, S.; Treiblmaier, H.; Zailani, S. The big picture on the internet of things and the smart city: A review of what we know and what we need to know. *Internet Things* **2022**, *19*, 100565. [CrossRef]
7. Jindal, A.; Kumar, N.; Singh, M. A unified framework for big data acquisition, storage, and analytics for demand response management in smart cities. *Future Gener. Comput. Syst.* **2020**, *108*, 921–934. [CrossRef]
8. Shi, Z. Research on the Construction of Smart Cities from the People-Oriented Perspective. In Proceedings of the International Conference on Construction and Real Estate Management 2023, Xi'an, China, 23–24 September 2023; pp. 15–21.
9. Giffinger, R.; Fertner, C.; Kramar, H.; Meijers, E. City-ranking of European medium-sized cities. *Cent. Reg. Sci.* **2007**, *9*, 1–12.
10. Kaiser, Z.R.M.A. Smart governance for smart cities and nations. *J. Econ. Technol.* **2024**, *2*, 216–234. [CrossRef]

11. Pereira, G.V.; Parycek, P.; Falco, E.; Kleinhans, R. Smart governance in the context of smart cities: A literature review. *Inf. Polity* **2018**, *23*, 143–162. [CrossRef]
12. Bosco, G.; Riccardi, V.; Sciarrone, A.; D'Amore, R.; Visvizi, A. AI-driven innovation in smart city governance: Achieving human-centric and sustainable outcomes. *Transform. Gov. People Process Policy* **2024**, *ahead-of-print*. [CrossRef]
13. Mao, Y.-H.; Li, H.-Y.; Xu, Q.-R. The Mode of Urban Renewal Base on the Smart City Theory under the Background of New Urbanization. *Front. Eng.* **2015**, *2*, 261–265. [CrossRef]
14. Oh, J. Smart City as a Tool of Citizen-Oriented Urban Regeneration: Framework of Preliminary Evaluation and Its Application. *Sustainability* **2020**, *12*, 6874. [CrossRef]
15. Wang, S.; Chen, D.; Liu, L. The practice and prospect of smart cities in China's urbanization process. *Front. Urban Rural Plan.* **2023**, *1*, 7. [CrossRef] [PubMed]
16. Liu, Z.; Wu, J. A Review of the Theory and Practice of Smart City Construction in China. *Sustainability* **2023**, *15*, 7161. [CrossRef]
17. Gao, X. A Study on the Predicament and Breakthrough Path of China's Smart City Construction. *Urban Insight* **2018**, *04*, 123–129.
18. Huang, L.; Zheng, W.; Hong, J.; Liu, Y.; Liu, G. Paths and strategies for sustainable urban renewal at the neighbourhood level: A framework for decision-making. *Sustain. Cities Soc.* **2020**, *55*, 102074. [CrossRef]
19. Dogruyol, K.; Aziz, Z.; Arayici, Y. Eye of Sustainable Planning: A Conceptual Heritage-Led Urban Regeneration Planning Framework. *Sustainability* **2018**, *10*, 1343. [CrossRef]
20. Chen, Q.; Wang, W.; Wu, F.; De, S.; Wang, R.; Zhang, B.; Huang, X. A Survey on an Emerging Area: Deep Learning for Smart City Data. *IEEE Trans. Emerg. Top. Comput. Intell.* **2019**, *3*, 392–410. [CrossRef]
21. Zhao, J.; Xie, W.; Lu, X.; Zhao, J. User Demand Identification and Development Trend Prediction Based on RF-BERT and UGC. *J. Intell. Sci.* **2024**, *42*, 132–142.
22. Wolf, K.; Dawson, R.J.; Mills, J.P.; Blythe, P.; Morley, J. Towards a digital twin for supporting multi-agency incident management in a smart city. *Sci. Rep.* **2022**, *12*, 16221. [CrossRef]
23. Deng, T.; Zhang, K.; Shen, Z.-J. A systematic review of a digital twin city: A new pattern of urban governance toward smart cities. *J. Manag. Sci. Eng.* **2021**, *6*, 125–134. [CrossRef]
24. Yan, B.; Meng, W. An Empirical Study on Rural Residents' Document Information Needs from the Perspective of Accurate Identification of Cultural Needs—A Survey Based on Six Counties in Anhui Province. *Libr. Inf. Serv.* **2019**, *63*, 77–83. [CrossRef]
25. Ochoa-Rico, M.S.; Río, J.A.J.-d.; Romero-Subia, J.F.; Vergara-Romero, A. Study of citizen satisfaction in rural versus urban areas in public services: Perspective of a multi-group analysis. *Soc. Indic. Res.* **2024**, *171*, 87–110. [CrossRef]
26. Zaman, M.; Puryear, N.; Abdelwahed, S.; Zohrabi, N. A Review of IoT-Based Smart City Development and Management. *Smart Cities* **2024**, *7*, 1462–1501. [CrossRef]
27. Hashem, I.A.T.; Chang, V.; Anuar, N.B.; Adewole, K.; Yaqoob, I.; Gani, A.; Ahmed, E.; Chiroma, H. The role of big data in smart city. *Int. J. Inf. Manag.* **2016**, *36*, 748–758. [CrossRef]
28. Rathore, M.M.; Ahmad, A.; Paul, A.; Rho, S. Urban planning and building smart cities based on the Internet of Things using Big Data analytics. *Comput. Netw.* **2016**, *101*, 63–80. [CrossRef]
29. Caro, F.; Sadr, R. The Internet of Things (IoT) in retail: Bridging supply and demand. *Bus. Horiz.* **2019**, *62*, 47–54. [CrossRef]
30. Yang, B.; Guo, W.; Chen, B.; Yang, G.; Zhang, J. Estimating Mobile Traffic Demand Using Twitter. *IEEE Wirel. Commun. Lett.* **2016**, *5*, 380–383. [CrossRef]
31. Baldin, A.; Bille, T.; Mukkamala, R.R.; Vatrapu, R. The impact of social media activities on theater demand. *J. Cult. Econ.* **2024**, *48*, 199–220. [CrossRef]
32. Liao, Y.; Yeh, S.; Gil, J. Feasibility of estimating travel demand using geolocations of social media data. *Transportation* **2022**, *49*, 137–161. [CrossRef]
33. Zhang, X.; Liu, J.; Wang, X.; Ma, W.; Yan, P. Examining information needs of public data service users: A study based on the 'Message Board for Leaders of People's Daily in China'. *Inf. Res. Int. Electron. J.* **2024**, *29*, 701–723. [CrossRef]
34. Wang, L.; Qian, C.; Kats, P.; Kontokosta, C.; Sobolevsky, S. Structure of 311 service requests as a signature of urban location. *PLoS ONE* **2017**, *12*, e0186314. [CrossRef]
35. Miller, S.M.J.M.; Review, B. Singapore's AI applications in the public sector: Six examples. *Manag. Bus. Rev.* **2023**, *3*, 144–155. [CrossRef]
36. Narayanan, L.K.; Sankaranarayanan, S. IoT-based water demand forecasting and distribution design for smart city. *J. Water Clim. Change* **2020**, *11*, 1411–1428. [CrossRef]
37. Preciado, J.C.; Prieto, A.E.; Benitez, R.; Rodríguez-Echeverría, R.; Conejero, J.M. A High-Frequency Data-Driven Machine Learning Approach for Demand Forecasting in Smart Cities. *Sci. Program.* **2019**, *2019*, 8319549. [CrossRef]
38. Jindal, A.; Aujla, G.S.; Kumar, N.; Prodan, R.; Obaidat, M.S. DRUMS: Demand response management in a smart city using deep learning and SVR. In Proceedings of the 2018 IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, United Arab Emirates, 9–13 December 2018; pp. 1–6.

39. VE, S.; Shin, C.; Cho, Y. Efficient energy consumption prediction model for a data analytic-enabled industry building in a smart city. *Build. Res. Inf.* **2021**, *49*, 127–143. [CrossRef]
40. Xiao, X.; Peng, Z.; Lin, Y.; Jin, Z.; Shao, W.; Chen, R.; Cheng, N.; Mao, G. Parking prediction in smart cities: A survey. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 10302–10326. [CrossRef]
41. Kontopoulou, V.I.; Panagopoulos, A.D.; Kakkos, I.; Matsopoulos, G.K. A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks. *Future Internet* **2023**, *15*, 255. [CrossRef]
42. Ramos, P.V.B.; Villela, S.M.; Silva, W.N.; Dias, B.H. Residential energy consumption forecasting using deep learning models. *Appl. Energy* **2023**, *350*, 121705. [CrossRef]
43. Agarwal, K.; Dheekollu, L.; Dhama, G.; Arora, A.; Asthana, S.; Bhowmik, T. Deep learning-based time series forecasting. *Artif. Intell. Rev.* **2024**, *58*, 23. [CrossRef]
44. Manzoor, M.A.; Albarri, S.; Xian, Z.; Meng, Z.; Nakov, P.; Liang, S. Multimodality representation learning: A survey on evolution, pretraining and its applications. *ACM Trans. Multimed. Comput. Commun. Appl.* **2023**, *20*, 1–34. [CrossRef]
45. Sapankevych, N.I.; Sankar, R. Time series prediction using support vector machines: A survey. *IEEE Comput. Intell. Mag.* **2009**, *4*, 24–38. [CrossRef]
46. Salman, D.; Direkoglu, C.; Kusaf, M.; Fahrioglu, M. Hybrid deep learning models for time series forecasting of solar power. *Neural Comput. Appl.* **2024**, *36*, 9095–9112. [CrossRef]
47. Liu, Z.; Zhang, Z.; Zhang, W. A Hybrid Framework Integrating Traditional Models and Deep Learning for Multi-Scale Time Series Forecasting. *Entropy* **2025**, *27*, 695. [CrossRef] [PubMed]
48. Kabir, M.R.; Bhadra, D.; Ridoy, M.; Milanova, M. LSTM–transformer-based robust hybrid deep learning model for financial time series forecasting. *Sci* **2025**, *7*, 7. [CrossRef]
49. Zhou, J.; Xiao, M.; Niu, Y.; Ji, G. Rolling bearing fault diagnosis based on WGWOA-VMD-SVM. *Sensors* **2022**, *22*, 6281. [CrossRef]
50. Wang, Z.; Chen, J.; Chen, J.; Chen, H. Identifying interdisciplinary topics and their evolution based on BERTopic. *Scientometrics* **2024**, *129*, 7359–7384. [CrossRef]
51. Li, T.; Cui, L.; Wu, Y.; Pandey, R.; Liu, H.; Dong, J.; Wang, W.; Xu, Z.; Song, X.; Hao, Y. Unveiling and advancing grassland degradation research using a BERTopic modelling approach. *J. Integr. Agric.* **2025**, *24*, 949–965. [CrossRef]
52. Grootendorst, M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv* **2022**, arXiv:2203.05794.
53. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [CrossRef]
54. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
55. Huang, S.; Cai, N.; Pacheco, P.P.; Narrandes, S.; Wang, Y.; Xu, W. Applications of support vector machine(SVM)learning in cancer genomics. *Cancer Genom. Proteom.* **2018**, *15*, 41–51. [CrossRef]
56. Hu, D.; Wang, E.; Ye, Q.; Chen, S.; Gu, X. How do online public messages affect local government responsiveness in China? A multilevel analysis based on the message board for leaders. *J. Chin. Political Sci.* **2023**, 1–32. [CrossRef]
57. Chang, D.; Meng, T. Magnetic effect of government responsiveness on public participation: An empirical analysis based on the provincial Leadership Message Board in China. *Chin. Public Adm. Rev.* **2023**, *14*, 94–106. [CrossRef]
58. Zhang, C.; Ma, B.; Gan, Y.; Xu, H. The bottom of the heart of the property builder: Evidence from online messages of Chinese rural migrant workers. *Chin. Political Sci. Rev.* **2024**, *9*, 222–244. [CrossRef]
59. Aloufi, N.; Aljuhani, A. Empirical Evaluation of Prompting Strategies for Python Syntax Error Detection with LLMs. *Applied Sciences* **2025**, *15*, 9223. [CrossRef]
60. Travieso, G.; Benatti, A.; Costa, L. An Analytical Approach to the Jaccard Similarity Index. *arXiv* **2024**, arXiv:2410.16436. [CrossRef]
61. Otchere, D.A.; Ganat, T.O.A.; Gholami, R.; Ridha, S. Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models. *J. Pet. Sci. Eng.* **2021**, *200*, 108182. [CrossRef]
62. Wang, X.; Wang, S.; Guo, Y.; Hu, K.; Wang, W. Dielectric and geometric feature extraction and recognition method of coal and gangue based on VMD-SVM. *Powder Technol.* **2021**, *392*, 241–250. [CrossRef]
63. Zrira, N.; Kamal-Idrissi, A.; Farssi, R.; Khan, H.A. Time series prediction of sea surface temperature based on BiLSTM model with attention mechanism. *J. Sea Res.* **2024**, *198*, 102472. [CrossRef]
64. Kim, T.-Y.; Cho, S.-B. Predicting residential energy consumption using CNN-LSTM neural networks. *Energy* **2019**, *182*, 72–81. [CrossRef]
65. Boucetta, L.N.; Amrane, Y.; Chouder, A.; Arezki, S.; Kichou, S. Enhanced Forecasting Accuracy of a Grid-Connected Photovoltaic Power Plant: A Novel Approach Using Hybrid Variational Mode Decomposition and a CNN-LSTM Model. *Energies* **2024**, *17*, 1781. [CrossRef]
66. Hatta, N.M.; Zain, A.M.; Sallehuddin, R.; Shayfull, Z.; Yusoff, Y. Recent studies on optimisation method of Grey Wolf Optimiser (GWO): A review (2014–2017). *Artif. Intell. Rev.* **2019**, *52*, 2651–2683. [CrossRef]
67. Wainer, J.; Fonseca, P. How to tune the RBF SVM hyperparameters? An empirical evaluation of 18 search algorithms. *Artif. Intell. Rev.* **2021**, *54*, 4771–4797. [CrossRef]

68. Wei, Y.; Chen, Z.; Zhao, C.; Tu, Y.; Chen, X.; Yang, R. A BiLSTM hybrid model for ship roll multi-step forecasting based on decomposition and hyperparameter optimization. *Ocean Eng.* **2021**, *242*, 110138. [CrossRef]
69. Majhi, B.; Naidu, D.; Mishra, A.P.; Satapathy, S.C. Improved prediction of daily pan evaporation using Deep-LSTM model. *Neural Comput. Appl.* **2020**, *32*, 7823–7838. [CrossRef]
70. Kumar, K.; Haider, M.T.U. Enhanced Prediction of Intra-day Stock Market Using Metaheuristic Optimization on RNN–LSTM Network. *New Gener. Comput.* **2021**, *39*, 231–272. [CrossRef]
71. Zitti, M. Forecasting salmon market volatility using long short-term memory (LSTM). *Aquac. Econ. Manag.* **2023**, *28*, 143–175. [CrossRef]
72. Wirtz, B.W.; Müller, W.M.; Schmidt, F.W. Digital public services in smart cities—An empirical analysis of lead user preferences. *Public Organ. Rev.* **2021**, *21*, 299–315. [CrossRef]
73. Tang, J.; Chen, X.; Hu, Z.; Zong, F.; Han, C.; Li, L. Traffic flow prediction based on combination of support vector machine and data denoising schemes. *Phys. A Stat. Mech. Its Appl.* **2019**, *534*, 120642. [CrossRef]
74. Meng, E.; Huang, S.; Huang, Q.; Fang, W.; Wang, H.; Leng, G.; Wang, L.; Liang, H. A Hybrid VMD-SVM Model for Practical Streamflow Prediction Using an Innovative Input Selection Framework. *Water Resour. Manag.* **2021**, *35*, 1321–1337. [CrossRef]
75. Zhao, J.; Huang, Y.; Feng, J.; Xie, W.; Jain, K. Fusion of KANO theory and Attention-BiLSTM models for user demand analysis and trend prediction. *Inf. Fusion* **2025**, *122*, 103210. [CrossRef]
76. Tuğrul, T.; Hınıs, M.A.; Oruç, S. Comparison of LSTM and SVM methods through wavelet decomposition in drought forecasting. *Earth Sci. Inform.* **2025**, *18*, 139. [CrossRef]
77. Ahmadi, F.; Tohidi, M.; Sadrianzade, M. Streamflow prediction using a hybrid methodology based on variational mode decomposition (VMD) and machine learning approaches. *Appl. Water Sci.* **2023**, *13*, 135. [CrossRef]

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