Application of GIS-Based Spatial Data Mining for Managing Outstanding Receivables in Water and Wastewater Companies: A Case Study of Khuzestan Province (Shush and Dasht-e Azadegan Zones)

Zeinab Obeidavi[1*], Adel Harbavi[1], Saber Alidadi[1], Hossein Obeidavi[1]

[1] Water and Wastewater Company of Khuzestan Province, Ahvaz, Iran - obeidavizeinab38@gmail.com, adelharbavi2026@gmaul.com, alidadis@yahoo.com, hosseinobeidavi@gmail.com

Keywords: Outstanding Receivables, Spatial data mining, GIS, Debt Management.

Abstract

Effective water pricing and timely bill collection are essential to prevent resource waste and ensure the financial stability of water and wastewater companies. This study analyses the spatial distribution and aging patterns of outstanding receivables in Shush and Dasht-e Azadegan, Khuzestan Province, using GIS-based spatial data mining techniques implemented in ArcGIS 10.8.1 in order to support more effective, data-driven debt management strategies. A dataset of over 90,800 water subscriptions, collected over a six-month period, was used to integrate spatial and financial data. Results indicate that 77.08% of outstanding receivables are of relatively low financial value (less than 10 million Iranian Rials), whereas 22.92% of subscribers carry heavy and high-risk debts. Optimized Hotspot Analysis identified statistically significant clusters ($p < 0.05$) of high-value and aged debts in Susangerd, Abu-Homeizeh, and Kut-Seyed Naeem, highlighting localized financial risks. In contrast, Shush and Bostan emerged as cold spots ($p < 0.05$), reflecting lower concentrations of overdue receivables. Overall, receivables aging analysis revealed that 54.19% of debts are recent (one year or less), with the highest short-term debt concentrations in Shush (67.25%) and Bostan (56.49%), both showing significant spatial clustering ($p < 0.05$). Although Alvan reported 64.72% of debts as short-term, the limited presence of cold spots suggests that favorable age profiles do not necessarily translate into spatially coherent financial stability. However, several cities exhibited disproportionately high levels of long-term debts, reflecting persistent inefficiencies in collection practices. These results underscore the importance of integrating spatial and temporal analyses to support targeted, data-driven receivables management strategies.

1. Introduction

Water, as one of the most valuable natural resources, is essential for sustaining life. However, its economic valuation remains challenging and is frequently underestimated in policy and management. In recent years, concerns about water scarcity and the misvaluation of water have intensified in many countries, particularly in Iran, due to factors such as population growth and climate change. Accurate pricing plays a fundamental role in preventing resource waste. Consequently, deviations from the true economic value of water are recognized as major factors contributing to inefficient usage and increased waste. Mispricing of water significantly hinders strategic planning and informed decision-making aimed at maximizing social welfare through its efficient and equitable allocation (Frone and Frone, 2012; European Environment Agency, 2013; Marston et al. 2021; Bastani et al. 2022). In the absence of a fully developed pricing mechanism for water, the failure to collect existing costs from subscribers is likely to result in excessive consumption and diminished public sensitivity to the ongoing water scarcity crisis. Accordingly, the enforcement of water billing serves as a fundamental instrument in demand management and the mitigation of water waste, and is one of the core responsibilities of water and wastewater utilities at the national level.

The collection of outstanding water bills from indebted subscribers, as a sub-process within the broader effort to reduce outstanding receivables, involves a variety of interactions. Managing these interactions often requires advanced expertise and technologies capable of rapidly mining large datasets to uncover hidden patterns. These insights, in turn, facilitate more effective and efficient management of the receivables reduction process.

With the increasing volume of spatial and temporal data stored in databases, extracting valuable insights poses a major challenge in the absence of advanced data analytics tools. As a result, decision-making processes often rely more on subjective interpretations by managers and users than on empirical evidence and data-driven insights. This issue is largely due to limited access to robust analytical systems. In response to this challenge, recent advances in data collection and storage have brought growing attention to data mining as a key technique in data analysis. Data mining is the process of extracting implicit, previously unknown, and potentially useful patterns and rules from large datasets (Ebrahimkhani et al., 2011; Han and Kamber, 2001; Piatetsky-Shapiro and Frawley, 1991; Shekhar et al. 2011). Spatial data mining, a specialized subfield, focuses on discovering valuable spatial patterns within geospatial databases.

* Corresponding author

In recent years, the integration of spatial data mining techniques with Geographic Information Systems (GIS) has attracted significant interest within the scientific community due to its potential to improve decision-making and resource management across various domains. For instance, Kuria et al. (2012) employed a GIS-based approach to improve water supply, distribution networks, and billing systems. Similarly, Igwe et al. (2023) examined the effectiveness of GIS technologies in electricity billing and revenue management.

In this study, given the abundance of spatial-descriptive data on outstanding receivables and the analytical capabilities of GIS-based spatial data mining, this approach has been employed to analyse spatial distribution patterns in the western regions of Khuzestan Province, specifically Shush and Dasht-e Azadegan. This study area represents a subset of the regions managed by the Khuzestan Province Water and Wastewater Company. Khuzestan, the fifth most populous province in Iran, with over one million water service subscriptions, faces substantial challenges related to unpaid receivables. These challenges primarily stem from the province's vast geographical area and dispersed population, especially in rural areas. Consequently, GIS-based spatial data mining has become an indispensable strategy for identifying, analysing, and ultimately reducing outstanding receivables. This study thus aims to evaluate and demonstrate the effectiveness of this approach.

2. Materials and Methods

2.1 Study Area

The study area encompasses the western regions of Khuzestan Province. According to the most recent administrative divisions established by the Khuzestan Province Water and Wastewater Company, the province is divided into 13 operational zones. Among these, Shush and Dasht-e Azadegan have been selected as the focal areas for this research (Figure 1). The Shush zone includes two counties. Shush County comprises the cities of Shush, Hor, and Fath ol-Mobin. Karkheh County consists of the cities of Alvan and Shavur. Similarly, the Dasht-e Azadegan zone encompasses two counties. Dasht-e Azadegan County comprises the cities of Susangerd, Bostan, Abu Homeizeh, and Kut-e Seyed Naeem. Hoveyzeh County consists of the cities of Hoveizeh and Rofayye. In total, these two zones contain more than 90,800 subscriptions, approximately 28,300 (31%) of which are rural.

2.2 Data preparation and analysis method

In this study, to assess the status of outstanding receivables in the study area, subscription payment data from the selected zones were extracted from the subscription database covering a six-month period. To prepare the spatial layer representing subscription locations, in addition to existing spatial data, location data collected by meter reading personnel during four reading cycles throughout the six-month study period were utilized. This approach aimed to maximize the inclusion of subscriptions in the analysis (Figure 1).

The overall workflow of the GIS-based spatial data mining process used in this study is illustrated in Figure 2, encompassing study area selection, data collection and preprocessing, data integration, statistical and spatial analysis, and interpretation of results (Figure 2).

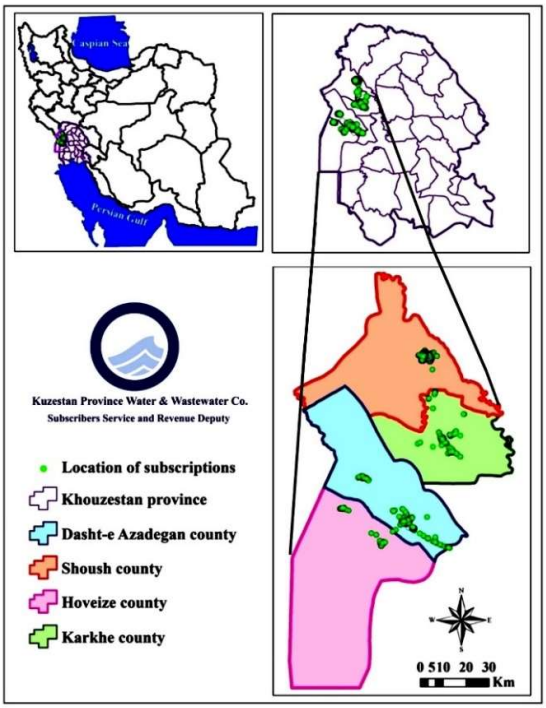

Figure 1. Study Area and Spatial Distribution of Water Subscriptions in Khuzestan Province, Iran.

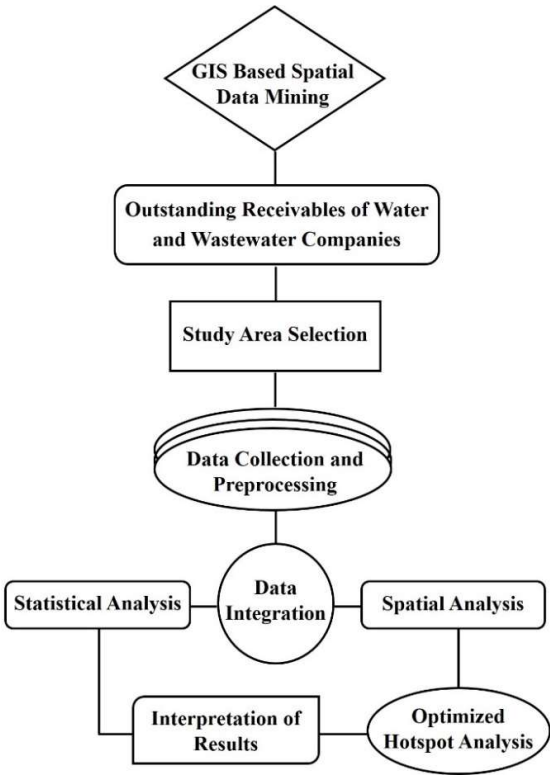

Figure 2. Workflow of the GIS-based spatial data mining process for analysing outstanding receivables in water and wastewater companies.

Subsequently, payment data were integrated with the spatial layer of subscriptions, resulting in a comprehensive dataset for analysing outstanding receivables in the study area. After integrating all available data, the final dataset was examined through both descriptive and spatial dimensions. Descriptive analysis was carried out using SPSS 27.0.1, while spatial analysis was performed in ArcGIS 10.8.1. To explore spatial patterns and relationships, spatial data mining techniques were applied using ArcGIS. Specifically, spatial classification methods were used to categorize spatial units based on both spatial and non-spatial attributes. As one of the core techniques in spatial data mining, spatial classification plays a pivotal role in uncovering hidden spatial patterns by leveraging spatial context and geographic dependencies. Compared to non-spatial classification methods, it offers enhanced accuracy and effectiveness in pattern recognition by incorporating spatial correlations. Moreover, it facilitates the integration of heterogeneous data types, allowing researchers to extract more precise and meaningful insights from complex spatial datasets (Malek et al., 2010; Estivill-Castro, 2002; Han et al., 2011).

In addition, Optimized Hotspot Analysis, a statistical spatial technique implemented within ArcGIS, was employed to detect statistically significant spatial clusters representing high (hotspots) and low (coldspots) values of outstanding receivables. This method improves upon traditional hotspot analysis by automatically determining the optimal scale of analysis and appropriate spatial parameters, such as distance thresholds, thereby enhancing the robustness and reliability of results. The analysis calculates the Getis-Ord Gi* statistic for each spatial feature. This statistic quantifies the degree of clustering of high or low values relative to the overall study area. By controlling for multiple testing and spatial dependence, Optimized Hotspot Analysis enables a statistically robust and spatially interpretable delineation of localized spatial clusters, supporting targeted decision-making and resource allocation in managing outstanding receivables (Esri, 2022; Getis and Ord, 1992; Wang and Luo, 2020).

3. Results

Based on the analysis, 77.08% of water utility subscribers in the study area have outstanding receivables less than 10 million Iranian rials, indicating that the majority of these receivables are relatively low in value and may be more manageable through routine collection efforts. In contrast, 22.92% of indebted subscribers owe amounts exceeding this threshold, representing a higher financial risk due to the magnitude of their individual receivables (Table 1).

The spatial distribution of outstanding receivables across the study area is depicted in Figure 3. The results reveal localized concentrations of high-value debts in cities such as Susangerd, Bostan, Abu Homeizeh, and Rofayye. This initial observation, based on frequency percentages and a visual inspection of the classified raw data, was corroborated by a formal hotspot analysis, which identified statistically significant clusters at both the 95% and 99% confidence levels in cities including Susangerd, Bostan, and Abu Homeizeh. These clusters thus represent critical financial risk zones, characterized by a high density and substantial value of outstanding receivables (Figure 4).

No.	Categories of receivables (million rial)	Abundance Percent
1	Less than 10	77.08
2	10–20	8.19
3	20–30	6.56
4	30–40	2.88
5	40–50	1.99
6	50 or more	3.30

Table 1. Classification of Outstanding receivables and Their Abundance Percent in the Study Area.

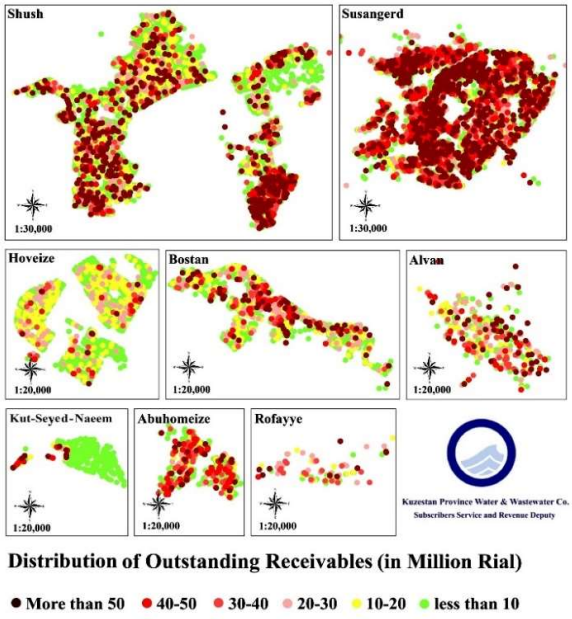

Figure 3. Spatial Distribution of Outstanding receivables (in Million Rial) in the Study Area.

Table 2 presents a detailed age distribution of outstanding receivables across the studied cities. Overall, 54.19% of outstanding receivables are aged one year or less, indicating a predominance of short-term outstanding receivables within the study area. The proportion of outstanding receivables decreases with age, as 19.5% are aged two years, 12.54% are aged three years, 5.25% are aged four years, and 8.52% are long-overdue (five years or older). Marked inter-city variations are observed. Shush (67.25%), Alvan (64.72%) and Bostan (56.49) exhibit the highest proportions of short-term outstanding receivables, reflecting more effective debt collection practices and superior receivables management. Conversely, Kut-Seyed Naeem (34.59%), Abu Homeizeh (23.48%), and Rofayye (20.61%) hold substantial proportions of long-overdue outstanding receivables (aged five years or more), highlighting potential challenges in recovering long-term debts within these cities.

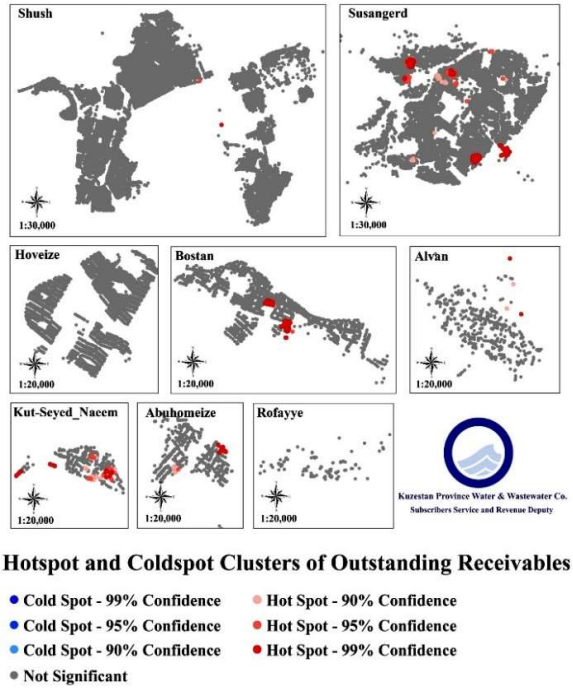

Figure 4. Spatial Distribution of Hotspots and Coldspots of Outstanding Receivables in the Study Area.

receivables Aging (yrs) / City names	1 yrs or less	2 yrs	3 yrs	4 yrs	5 yrs or more
Abu Homeizeh	38.65	17.81	12.53	7.52	23.48
Alvan	64.72	20.79	5.83	4.63	4.02
Bostan	56.49	18.12	15.05	6.73	3.64
Rofayye	32.51	15.53	17.27	14.08	20.61
Susangerd	44.76	21.72	18.57	5.75	9.21
Shush	67.25	17.84	6.64	2.43	5.84
Kut- Seyed Naeem	28.94	14.40	13.32	8.75	34.59
Hoveizeh	54.60	17.77	11.58	7.22	8.82
Overall	54.19	19.5	12.54	5.25	8.52

Table 2. City-wise and Overall Distribution of Receivables Aging (%) in the Study Area.

Figure 5 visually illustrates the spatial distribution of outstanding receivables by age group, supporting the numerical findings. Figure 6 identifies statistically significant hot and cold spot clusters across the studied cities.

The results revealed prominent hot spots at 95% and 99% confidence levels in cities such as Susangerd, Kut-Seyed Naeem, and Abu-Homeizeh, indicating concentrated zones of high-value, aged outstanding receivables. These patterns suggest notable financial risks and potential underlying issues such as

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-W8-2025-577-2026 | © Author(s) 2026. CC BY 4.0 License.

management inefficiencies or adverse socio-economic conditions.

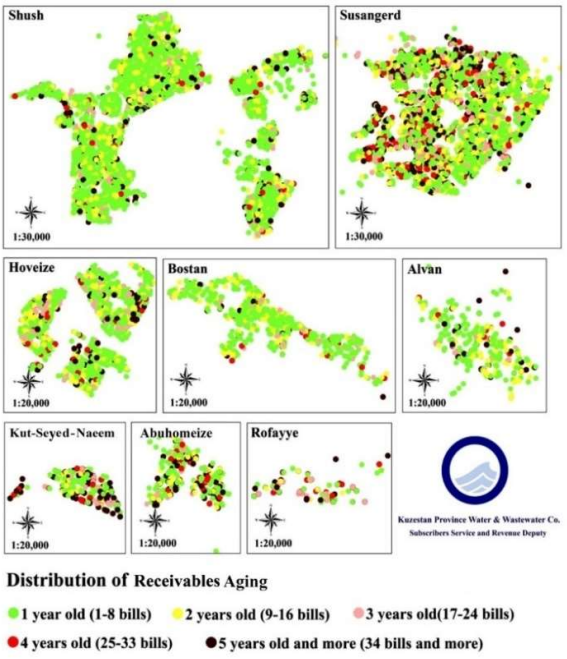

Figure 5. Spatial Distribution of Outstanding Receivables based on Aging Classification in the Study Area.

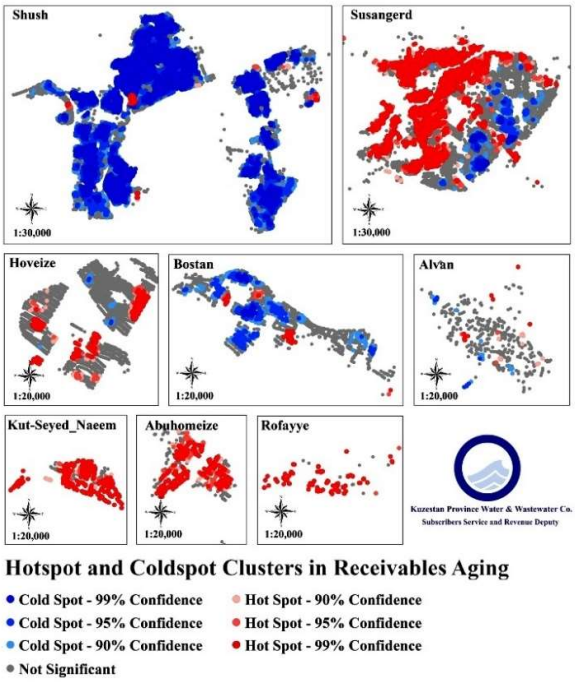

Figure 6. Hotspot and Coldspot Clusters of Receivables Aging in the Study Area.

Conversely, cities such as Shush and Bostan exhibited statistically significant cold spots at 95% and 99% confidence levels, indicating more effective management of outstanding receivables and lower financial risk in these areas. Although Alvan reported a high proportion of short-term outstanding receivables (64.72%) as shown in Table 2, the optimized hotspot analysis identified only limited and scattered cold spots, with most areas lacking statistical significance. This suggests that a favorable receivables age profile does not necessarily correspond to spatially coherent patterns of reduced financial risk. These findings highlight the importance of integrating both temporal and spatial analyses when evaluating receivables performance and underscore the need for tailored debt management strategies that reflect the unique context of each city.

4. Discussion and Conclusion

This study presents a successful application of GIS-based spatial data mining techniques to assess outstanding receivables within water and wastewater companies. This analytical approach reveals hidden spatial dependencies and offers a more precise understanding of the underlying patterns compared to purely descriptive methods, thereby enhancing the effectiveness of strategic and operational decision-making processes.

The findings indicate that the distribution of outstanding receivables across much of the study area does not exhibit statistically significant spatial clustering. This spatial uniformity may reflect consistent performance in receivables collection and administrative procedures. The observed spatial uniformity in outstanding receivables across much of the study area likely reflects the relatively homogeneous management structure within the regional offices of the Khuzestan Water and Wastewater Company. However, the absence of spatial clustering should not be interpreted as definitive evidence of managerial efficiency, as variations in consumption levels, subsidy policies, and socio-economic or demographic factors may also influence these patterns. Furthermore, the identified spatial clusters of high-value overdue receivables in specific localities highlight the need for tailored, region-specific management strategies.

Nonetheless, certain areas within the region do exhibit statistically significant clusters of outstanding receivables. Notably, hot spots were detected in cities such as Susangerd, Bostan, Abu Homeizeh, and Kut-Seyed Naeem, where high concentrations of large-value receivables are observed. These clusters may point to localized inefficiencies, socio-economic challenges, or weaknesses in debt collection processes. Such findings underscore the importance of spatial pattern analysis in evaluating the effectiveness of receivables management strategies.

The significance of spatial analysis in the public sector has been emphasized in various studies. For instance, Zhou et al. (2023) conducted spatial clustering analyses to assess the distribution of local government bonds in China. Similarly, Nakaya et al. (2005) highlighted the importance of spatial patterns as key indicators in

evaluating public health service effectiveness. Longley et al. (2015) also explored the role of spatial analytics in optimizing public service delivery. Collectively, these studies demonstrate that spatial pattern analysis serves as a powerful tool for operational control and resource management across diverse domains.

From a temporal perspective, the results reveal a predominance of short-term receivables in the study area, with the proportion of receivables decreasing as their age increases. This trend may reflect partially successful collection efforts during the first one or two years. However, the high share of long-overdue receivables (five years or more) in cities such as Kut-Seyed Naeem, Abu Homeizeh, and Rofayye points to serious challenges in recovering long-term debts in these areas. Hotspot and coldspot analyses based on receivables aging corroborate these observations. While cities such as Susangerd and Kut-Seyed Naeem are identified as hot spots of high-value, aged receivables, Shush and Bostan emerge as cold spots, reflecting lower concentrations of long-term debts.

A noteworthy finding is that, despite having a high proportion of short-term receivables (64.72%), the city of Alvan does not demonstrate statistically significant cold spot patterns. This suggests that a favorable receivables aging profile does not necessarily correspond to spatially coherent zones of reduced financial risk. This reinforces the importance of integrating both temporal and spatial analyses when assessing receivables performance. Indeed, hidden spatial dependencies within financial datasets play a critical role in identifying high-risk areas. Overall, these insights highlight that the combination of temporal (aging of receivables) and spatial (geographic distribution) analyses provides a powerful framework for developing more effective strategies in managing outstanding receivables.

This study is subject to several limitations. First, the analysis relies on administrative datasets, which may omit informal or unreported transactions, potentially leading to underestimation or bias in the results. Second, due to data constraints, important socio-economic variables, such as household income, employment status, and subscriber behavior, were not incorporated into the spatial modeling framework. This omission may limit the explanatory power of the models. Third, the cross-sectional nature of the analysis restricts the ability to capture dynamic changes and temporal trends in receivables patterns. To address these limitations, future research should utilize longitudinal datasets and integrate a broader range of contextual indicators to improve model robustness, explanatory capacity, and policy relevance.

Based on these findings, future research should employ dynamic spatio-temporal models to track the evolution of receivables over time. Moreover, integrating advanced spatial analysis techniques may provide deeper insights into the local socio-economic and administrative factors influencing the emergence of hot and cold spots. Furthermore, water and wastewater companies are encouraged to leverage spatial data mining and clustering methods to design tailored debt recovery policies that address the unique conditions of each city and region.

References

Bastani, M., Salehnia, N., Mirbagheri, S.H., 2022. Estimating the Economic Value of Water by Calculating Mashhad Citizens' Willingness to Pay for Excess Water in the Agricultural Sector for Drinking. *Iranian Journal of Irrigation and Drainage*, 15(6), 1459-1468.

David Ndegwa Kuria, Douglas Engoke Musiega, Moses Murimi Ngigi, Simon Kibue Ngugi, 2012. An open-source Geographic Information System (GIS) approach to water supply management, distribution, and billing. *International Journal of Computer Engineering Research*, 3(4), 55-62.

Ebrahimkhani, S., Afzali, M., Shokoohi, A., 2011. Prediction and analysis of road accident factors using data mining algorithms. *Danesh-e Entezami Zanjan Quarterly*, 1(1), 111-127. (In Persian)

Esri, 2022. Optimized Hot Spot Analysis (Spatial Statistics). ArcGIS Pro Documentation: Hot Spot Analysis (Getis-Ord Gi*).

Friedl, B., Getzner, M., Hawelka, B., 2018. Spatial analysis of utility payment behavior: Evidence from water services in Sub-Saharan Africa. *Utilities Policy*, 55, 87-96. doi.org/10.1016/j.jup.2018.09.004.

Getis, A., Ord, J.K., 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24(3), 189-206. doi.org/10.1111/j.1538-4632.1992.tb00261.x.

Han, J., Kamber, M., 2001. Data Mining: Concepts and Techniques. San Diego: Academic Press.

Han, J., Kamber, M., Pei, J., 2011. Data Mining: Concepts and Techniques (3rd ed.). Morgan Kaufmann.

Malek, J., Laurini, R., Wemmert, C., 2010. Spatial Classification: A Survey. *International Journal of Geographical Information Science*, 24(2), 193-212.

Nakaya, T., Fotheringham, A.S., Brunsdon, C., Charlton, M., 2005. Geographically Weighted Poisson Regression for Disease Mapping. *Statistics in Medicine*, 24(17), 2695-2717. doi.org/10.1002/sim.2116.

Nayak, R., Ghosh, S., Bhattacharyya, S., 2021. Detection of water utility fraud using data mining techniques. *Expert Systems with Applications*, 167, 114134. doi.org/10.1016/j.eswa.2020.114134.

Shekhar, S., Chawla, S., Lu, C., 2011. Spatial Data Mining: A Survey. *ACM Computing Surveys*, 43(1), 1-46.

Sengupta, M., Raj, T., 2017. Urban water billing and collection: Challenges and solutions. *Journal of Water Management*, 5(2), 101-112.

Stanley Igwe, Sunday Adetona, Obiageli Ngwu, 2023. The GIS Application for Electricity Distribution Billing and Revenue Management. *International Journal of Innovative Science and Research Technology*, 8(8).

Wang, F., Luo, W., 2020. Spatial Data Mining and Its Application in Urban Planning. *International Journal of Geographical Information Science*, 34(8), 1615-1632. doi.org/10.1080/13658816.2020.1720895.

Wang, F., Zhou, X., 2012. Data mining for utility billing analysis: A case study. *Expert Systems with Applications*, 39(8), 7029-7036.

Zhou, Q., Zhang, Y., Zhao, Z., 2020. Prediction of water consumers' payment default using machine learning: A case study in smart metering system. *Water*, 12(6), 1672. doi.org/10.3390/w12061672.

Zhou, W., Wong, D.W., 2020. Real-time spatial analytics for smart city water management. *Computers, Environment and Urban Systems*, 80, 101438.

Zhou, Y., Chen, Y., Liu, J., 2023. Spatial clustering and influencing factors of local government bond issuance in China: A spatial econometric approach. *T*

he Professional Geographer, 75(3), 399-416. doi.org/10.1080/00330124.2023.2300803