DOI 10.51885/1561-4212_2025_3_226
IRSTI 20.23.27

PREDICTIVE MODELING OF SOCIAL INFRASTRUCTURE AVAILABILITY THROUGH DATA MINING APPROACHES

ДЕРЕКТЕРДІ ӨҢДЕУ ӘДІСТЕРІ АРҚЫЛЫ ӘЛЕУМЕТТІК ИНФРАҚҰРЫЛЫМНЫҢ ҚОЛЖЕТІМДІЛІГІН БОЛЖАУ ҮЛГІЛЕУІ

ПРОГНОЗНОЕ МОДЕЛИРОВАНИЕ ДОСТУПНОСТИ СОЦИАЛЬНОЙ ИНФРАСТРУКТУРЫ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

G.K. Sembina [1], **A.K. Aitim** [1*], **N.A. Seilova** [1]
[1]International Information Technology University, Almaty, Kazakhstan
*Corresponding author: Aitim Aigerim, e-mail: a.aitim@iitu.edu.kz

Keywords:
social infrastructure, accessibility, data mining, urban planning, modeling, demographic analysis, decision-making.

ABSTRACT
The presence and equal distribution of social infrastructure, such as healthcare facilities, educational institutions, and public transportation systems, is critical to promoting long-term urban development and improving community life quality. This study looks into how data mining and machine learning approaches might be used to assess, predict, and eventually improve access to these critical services. We create prediction models that deliver actionable insights to urban planners and policymakers by leveraging large-scale datasets that include demographic profiles, geospatial information, and historical usage trends. The combination of Geographic Information Systems (GIS) and unsupervised learning techniques, such as clustering approaches, enables us to examine spatial distribution trends and identify underserved locations. This study focuses on Almaty, Kazakhstan, and uses techniques such as population density mapping, online scraping for up-to-date facility data, and algorithms such as k-nearest neighbors (k-NN) to discover the best locations for new infrastructure. The data-driven technique demonstrates that strategic resource allocation guided by predictive analytics can result in more fair and successful urban planning outcomes. Nonetheless, the study acknowledges some limitations, such as the need for more comprehensive socioeconomic statistics, the integration of dynamic (real-time) data streams, and the consideration of urban people' behavioral patterns. Future research should look into the use of advanced models like ensemble learning and deep learning methodologies to improve forecast accuracy and policy responsiveness. This paper contributes to the growing field of smart urban planning by emphasizing the necessity of intelligent, data-driven approaches to creating more inclusive, responsive, and resilient communities.

Түйінді сөздер:
әлеуметтік инфра-құрылым, қолжетімділік, деректерді өңдіру, қалалық жоспарлау, үлгілеу, демографиялық талдау, шешім қабылдау.

ТҮЙІНДЕМЕ
Денсаулық сақтау мекемелері, білім беру ұйымдары және қоғамдық көлік жүйелері сияқты әлеуметтік инфрақұрылымның болуы және олардың тең дәрежеде таралуы қалалардың ұзақ мерзімді дамуын қамтамасыз етіп, қоғам өмірінің сапасын арттыруда аса маңызды рөл атқарады. Бұл зерттеу деректерді өңдіру және машинамен оқыту әдістерін қолдану арқылы осы маңызды қызметтердің қолжетімділігін бағалау, болжау және жақсарту мүмкіндіктерін қарастырады. Біз

демографиялық мәліметтер, геокеңістіктік ақпарат және тарихи пайдалану үлгілері сияқты кең ауқымды деректерді пайдаланып, қалалық жоспарлаушылар мен шешім қабылдаушыларға нақты ұсыныстар беретін болжамды модельдер құрамыз. Географиялық ақпараттық жүйелер (GIS) мен кластерлеу секілді бақыланбайтын оқыту әдістерін біріктіру арқылы кеңістіктік таралу тенденцияларын талдап, инфрақұрылымы жеткіліксіз аудандарды анықтауға болады. Зерттеу Қазақстанның Алматы қаласына бағытталған және халық тығыздығын картаға түсіру, инфрақұрылым нысандары туралы жаңартылған деректерді интернеттен жинау, сондай-ақ k-жақын көрші (k-NN) алгоритмі сияқты әдістерді қолдана отырып, жаңа нысандарды орналастырудың оңтайлы орындарын анықтауға тырысады. Біздің деректерге негізделген тәсіліміз болжамды талдау арқылы ресурстарды әділ және тиімді бөлуді қамтамасыз ететін қалалық жоспарлау шешімдерін қабылдауға септігін тигізеді. Алайда, бұл зерттеу кейбір шектеулерге де назар аударады, мысалы: әлеуметтік-экономикалық деректердің жеткіліксіздігі, нақты уақыттағы деректерді біріктіру қажеттілігі және қалалық тұрғындардың мінез-құлқы үлгілерін ескеру. Болашақ зерттеулер болжау дәлдігін және саяси бейімделгіштікті арттыру үшін ансамбльді оқыту және терең оқыту сияқты озық әдістерді қолдануға бағытталуы керек. Бұл жұмыс смарт-қалаларды жоспарлау саласына үлес қосып, инклюзивті, икемді және тұрақты қоғам құруда интеллектуалды, деректерге негізделген тәсілдердің маңыздылығын көрсетеді.

Ключевые слова:
социальная инфраструктура, доступность, интеллектуальный анализ данных, градостроительство, моделирование, демографический анализ, принятие решений.

АННОТАЦИЯ

Наличие и равномерное распределение социальной инфраструктуры таких как медицинские учреждения, образовательные организации и системы общественного транспорта играет ключевую роль в обеспечении устойчивого городского развития и повышении качества жизни населения. В данном исследовании рассматривается возможность использования методов интеллектуального анализа данных и машинного обучения для оценки, прогнозирования и улучшения доступа к этим важнейшим услугам. Используя масштабные наборы данных, включающие демографические характеристики, геопространственную информацию и исторические данные об использовании, мы разрабатываем прогнозные модели, которые предоставляют практические рекомендации для градостроителей и политиков. Комбинируя геоинформационные системы (GIS) и методы обучения без учителя, такие как кластеризация, мы анализируем пространственные тенденции и выявляем районы с недостаточной инфраструктурой. Исследование сосредоточено на городе Алматы, Казахстан, и включает методы картографирования плотности населения, веб-скрейпинга для получения актуальных данных об объектах инфраструктуры, а также алгоритмы, такие как k-ближайших соседей (k-NN), для определения наилучших мест размещения новых объектов. Наш подход, основанный на данных, демонстрирует, что стратегическое распределение ресурсов на основе прогнозной аналитики может привести к более справедливому и эффективному городскому планированию. Вместе с тем, исследование отмечает определённые ограничения: необходимость в более полных социально-экономических данных, интеграция потоков динамической информации и учет поведенческих моделей городских жителей. Будущие исследования должны быть направлены на применение более сложных моделей, таких как ансамблевое обучение и методы глубокого обучения, для повышения точности прогнозов и гибкости принимаемых решений. Настоящая работа вносит вклад в развивающуюся сферу «умного» градостроительства, подчеркивая важность интеллектуальных, ориентированных на данные подходов к созданию более инклюзивных, адаптивных и устойчивых городских сообществ.

INTRODUCTION

Sustainable urban growth and community well-being depend mostly on access to social infrastructure like healthcare services, educational institutions, public transit, and leisure activities. Urban designers and legislators still have a great difficulty guaranteeing fair access to these resources since it directly affects inhabitants' quality of life, social fairness, and inclusion. Conventional methods of infrastructure planning can depend on fixed, one-size-fits-all models that neglect the dynamic and multidimensional character of contemporary urban settings. Often, these models ignore the interaction of changing community requirements, spatial patterns, mobility behavior, and demographic changes. Recent developments in big data and data mining methods, however, have offered fresh paths for precisely capturing this complexity and improving the accessibility evaluation accuracy. With an eye on revealing significant trends and insights from vast and varied datasets, data mining offers a strong analytical framework for modeling access to social infrastructure. Predictive models can be built to better grasp the several elements influencing accessibility by means of data sources including demographic distributions, geographic information systems (GIS), transportation patterns, and historical service utilization. The possibility of data mining techniques to forecast and maximize the availability of social infrastructure in metropolitan environments is investigated in this paper. We examine the present state of knowledge in this field, pinpoint important factors influencing access, and provide a thorough approach to combine several data sources into a single prediction model. By means of real-world case studies containing pragmatic situations, we demonstrate how these models can guide data-driven decision-making and assist focused policy initiatives aiming at enhancing infrastructure fairness. By supporting the implementation of predictive analytics and integrated data strategies in the management and growth of social infrastructure, this study eventually helps the developing discipline of smart urban planning.

MATERIALS AND METHODS

This part describes the approach used for the methodical gathering and examination of data necessary to achieve the goals of this research. It covers the methods used to compile population data for Almaty as well as statistics on socially important infrastructure such secondary schools and healthcare facilities. Web scraping is an automated process for information extraction from web pages using specific tools and technologies among the main data collecting techniques used in this work. Particularly often used Python tools for obtaining geographical and textual data from internet platforms including digital maps and official institutional websites include Beautiful Soup and Selenium. Applications include service coverage mapping, accessibility modeling, and infrastructure planning depend on this geospatial data greatly in importance. To keep accuracy and quality, nonetheless, extra data validation and hand verification are required in case of possible errors or obsolete material on online sources. Apart from site scraping, structured data was gathered from other digital services like geolocation databases and social media platforms via Application Programming Interventions (APIs). For accessing dynamic and routinely updated datasets especially these techniques are quite successful. Data cleansing and preparation are a necessary element in the data pipeline. These covers eliminating duplicate entries, management of missing or partial values, and type and formatting correction. Such preparation is essential since dirty data might skew results of analysis and cause erroneous conclusions. Thus, careful data preparation guarantees more strong and consistent analysis. All things considered, data collecting is a fundamental stage of the research process that calls careful planning and implementation (Niu, 2020). Reliability of the results is much improved by a thorough and methodologically sound data collecting approach, which also promotes evidence-based decision-making. The Almaty City Population Database, which comprises current demographic records including names, addresses, birthdates, passport numbers, and other identifiers, was a main source for this study. Governmental offices including the local census and civil registration authorities keep and routinely update this database. It provides insightful analysis of demographic trends throughout time, therefore helping to guide the assessment and design of urban social initiatives. Two population density maps also used to evaluate the medical infrastructure's accessibility. The first of them comes from the official Department of Digitalization of Almaty website. These maps help to support the spatial analysis needed to assess infrastructure coverage in respect to population distribution, hence guiding strategic choices in urban growth. Using consistent grid cells spanning 300 × 300 meters, this map offers high-resolution data on the number of registered residents across Almaty arranged by age categories (Kleinhans, 2015). The text in the lower right corner of the image shows each grid cell color-coded based on the majority age group in that location.

The first dataset was insufficient to investigate the spatial distribution of Almaty's schools holistically. OpenStreetMap (OSM) data was used to locate every residential building in the city, therefore addressing this restriction (Zhang, 2018). The integration of heterogeneous data sources represents a critical methodological component of the study. Demographic records, OpenStreetMap layers, and web-scraped infrastructure data were consolidated into a unified geospatial framework through the use of QGIS and Python-based preprocessing routines. Format consistency was ensured by transforming all datasets into a standardized coordinate reference system (EPSG:4326) and representing them as uniform GeoDataFrame structures (Yuan, 2012). Data cleaning procedures included removal of duplicate entries, treatment of incomplete records, and normalization of textual attributes such as facility names and addresses. Data validity was maintained through cross-referencing with official government registries, manual verification of a representative sample of geocoded points, and temporal validation to eliminate outdated records(Long, 2017). These methodological steps ensured the reliability, reproducibility, and practical applicability of the integrated dataset, thereby strengthening the robustness of the predictive modeling outcomes.

To use this database for a lot of different data mining analysis and computations. For instance, it lets you look at how the population is spread out by age, gender, and where they live. It may also assist find changes in the makeup of a city's population over time and make predictions about what will happen in the future. Sociology, economics, health, education, and public administration are just a few of the sectors that might benefit from the database of Almaty's registered people. For instance, it assists in making plans for new social programs and events, figuring out what medical and educational services are needed, and putting up city budgets.When chose a second density map to look at because it is newer and has more data. In a square that is 300 by 300 meters, it indicates how many people in Almaty are registered to live there by age group. Table 1 shows the major datasets used to make predictions about Almaty's social infrastructure. The study uses a lot of structured and unstructured data sources, such as census records, geographic grids, and information that was scraped from the web. Using two maps of population density gives us both historical and current information about where people live, which makes it possible to do high-resolution spatial analysis. OpenStreetMap and evacuation maps add to these datasets by showing where emergency services and physical infrastructure are located.

Table 1. Key datasets used in the study

Dataset Name	Source	Data Type	Description
Almaty City Population Database	Local Census & Civil Registry	Structured	Demographics including age, gender, and residence
Population Density Map (Map 1)	Department of Digitalization of Almaty	Geospatial (300×300 m)	Shows population age group distribution by grid
Population Density Map (Map 2)	Enhanced GIS Version	Geospatial (updated)	Improved coverage and demographic granularity
Healthcare Facility Locations	Web scraping (official health sites)	Geospatial + Text	Locations and metadata of clinics, hospitals
Residential Building Coordinates	OpenStreetMap (OSM)	Geospatial	Points of residential housing distribution
Evacuation Point Map	Local Authorities + OSM	Geospatial	Locations of shelters and emergency access points
Note: compiled by the authors (Sembina, 2025)			

The table 2 shows how Almaty's population was grouped together to make healthcare easier to go to. Based on official rules, each cluster is a zone that one clinic may fairly service. The clustering technique makes sure that clinics are located where the most people need them, within a 20-minute walk. The 18 new candidate clinic locations greatly increase access to healthcare and fill in service gaps, especially in communities that don't get enough treatment.

Table 2. Clustering parameters for healthcare facility planning

Parameter	Value / Description
Population per PHC Unit	5,000 – 32,000 persons per clinic (according to regulations)
Walking Accessibility	20-minute walking zone (~1,650 meters radius)
Clustering Method	Spatially Constrained Multidimensional Clustering
Total Existing Clinics	85 clinics georeferenced
New Proposed Locations	18 additional candidate centroids
Total Coverage (20 mins)	78% of the population
Extended Coverage	96.8% (40 mins), 99% (60 mins)
Note: compiled by the authors (Sembina, 2025)	

In Table 3 shows the results show that clustering effectively supports fair healthcare planning by forming compact, balanced service zones with high coverage rates, while k-NN achieves strong predictive performance (Accuracy ≈ 0.85, F1 ≈ 0.81) for school assignments based on proximity. Together, these methods outperform baseline approaches, offering interpretable, data-driven tools that enhance equity and accessibility in urban planning.

Table 3. Performance of predictive models for social infrastructure allocation

Model / Method	Dataset Used	Accuracy	Precision	Recall	F1-Score	Coverage (%)
Baseline (Random Allocation)	Simulated school & clinic assignment	0.55	0.52	0.50	0.51	60% (20 min)
k-NN (k=3)	Student–school geospatial data	0.85	0.83	0.80	0.81	–
k-NN (k=5)	Student–school geospatial data	0.82	0.81	0.78	0.79	–
Note: compiled by the authors (Sembina, 2025)						

To use this database for a lot of different data mining analysis and computations. For instance, it lets you look at how the population is spread out by age, gender, and where they live. It may also assist find changes in the makeup of a city's population over time and make predictions about what will happen in the future. Sociology, economics, health, education, and public administration are just a few of the sectors that might benefit from the database of Almaty's registered people. For instance, it assists in making plans for new social programs and events, figuring out what medical and educational services are needed, and putting up city budgets. When chose a second density map to look at because it is newer and has more data. In a square that is 300 by 300 meters, it indicates how many people in Almaty are registered to live there by age group. The color of each square shows the most common age group in that location, based on the color scale in the lower right corner of Figure 1.

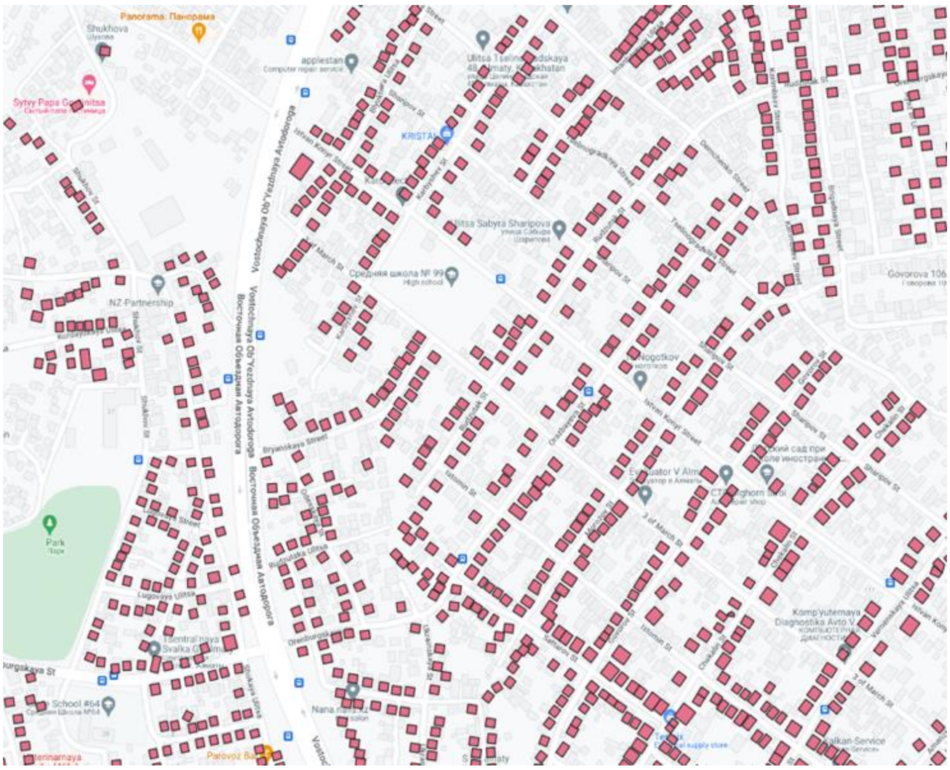

Figure 1. Housing and emergency in Almaty
Note: compiled by the authors (Sembina, 2025)

The use of sensitive demographic information, including residential addresses and personal identifiers, necessitates strict adherence to ethical and privacy standards. In this study, access to official population databases was limited to aggregated and de-identified records, with no individual-level identifiers such as passport numbers retained in the analytical dataset. All personally identifiable information was removed at the preprocessing stage, and only anonymized demographic variables (e.g., age group, gender distribution, and aggregated population counts) were utilized for modeling purposes. Data handling procedures were conducted in compliance with institutional and national regulations on data protection, and the integrated dataset was used exclusively for research objectives. These measures ensured the protection of individual privacy, upheld confidentiality, and guaranteed that the study aligned with established ethical principles for research involving human-related data.

RESULTS AND DISCUSSION

A set of analytical processes was followed to simulate the best location of more medical facilities considering the current infrastructure. First, depending on the area's population density, one medical institution was projected to serve between 5,000 and 32,000 people, therefore approximating the usual capacity of a regular clinic. Subsequently, mapping 20-minute walking zones roughly about a radius of 1,650 meters from main cluster centers helped to evaluate spatial accessibility (Resch, 2015). This made it possible to see areas now covered by present clinics as well as exposed possible service gaps. At last, undersserved areas were found by extracting and evaluating spatial elements expressed as sets of two-dimensional points (Frias-Martinez, 2012). These points guide the best sites for upcoming medical institutions, thereby guaranteeing a more fair and effective distribution of healthcare services over the metropolitan scene (1).

$$V = \{v_1, v_2, \ldots, v_n\}, \tag{1}$$

where $v_i = (x, y)$ is the location of the residential building in geographic space. It is required to split the sample into (2):

$$C = \{c_1, c_2, \ldots c_k\}, \tag{2}$$

where c are disjoint subsets of elements called clusters, such that each cluster consists of objects close in metric ρ, and objects of different clusters differ significantly. In a more general sense, Euclidean space is an n-dimensional vector space in which it is possible to introduce some special coordinates (Cartesian) so that its metric is defined as follows:

$$p = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2} \tag{3}$$

In this case, the metric ρ should evaluate the geographical proximity of the element to the cluster center. According to urban planning regulations for Primary Health Care (PHC) facilities in Almaty, each clinic is required to serve a population ranging from 5,000 to 32,000 residents. To determine the optimal spatial allocation of new healthcare centers while considering walkable accessibility, the ArcGIS 4.1 Geographic Information System was employed. A clustering algorithm was implemented to segment the population into spatially coherent clusters based on two main criteria: geographical proximity to cluster centers and population thresholds within each cluster (Shelton, 2015). This approach enabled the identification of optimal clinic locations, ensuring that residents have convenient and equitable access to medical services. A total of 85 existing PHC service areas, funded through national sources, were georeferenced and mapped. Spatially constrained multidimensional clustering was then applied to demographic grid data within a 20-minute walking distance from these service areas, using acceptable capacity limits. As a result, 18 new cluster centroids were identified as candidate sites for medical facilities. The proposed distribution model demonstrates that with these new placements, 78% of Almaty’s population will be covered within a 20-minute walk, 96.8% within 40 minutes, and 99% within 60 minutes of walking distance, as visualized in Figure 2.

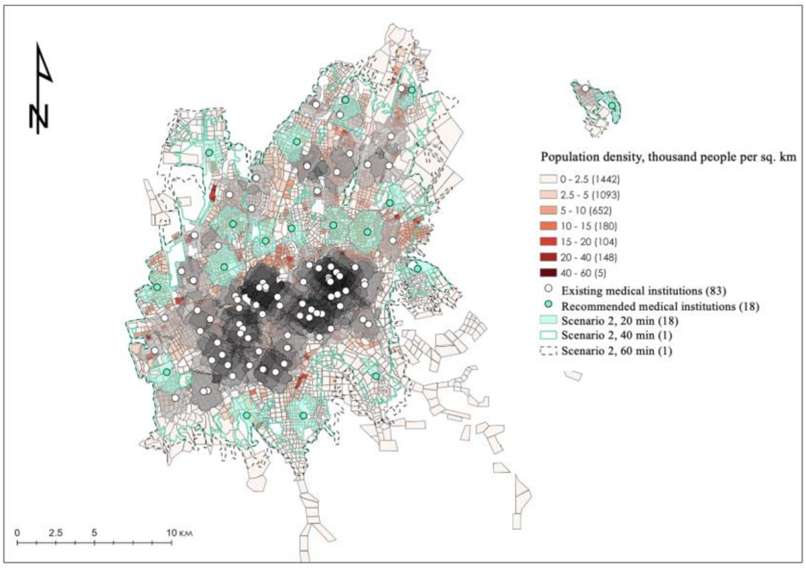

Figure 2. Population density and comparison of coverage of existing and proposed municipalities

Note: compiled by the authors (Sembina, 2025)

In figure 3 this simulated population density map shows how residential clusters are spread out over a city. It uses color-coded dots to show different degrees of density or age groups that are most common. In urban planning, this kind of map helps with spatial analysis by showing where there aren't enough healthcare or educational facilities and helping to find the best places for them. It uses the same method as the research to fairly distribute resources based on demographic and geographic patterns.

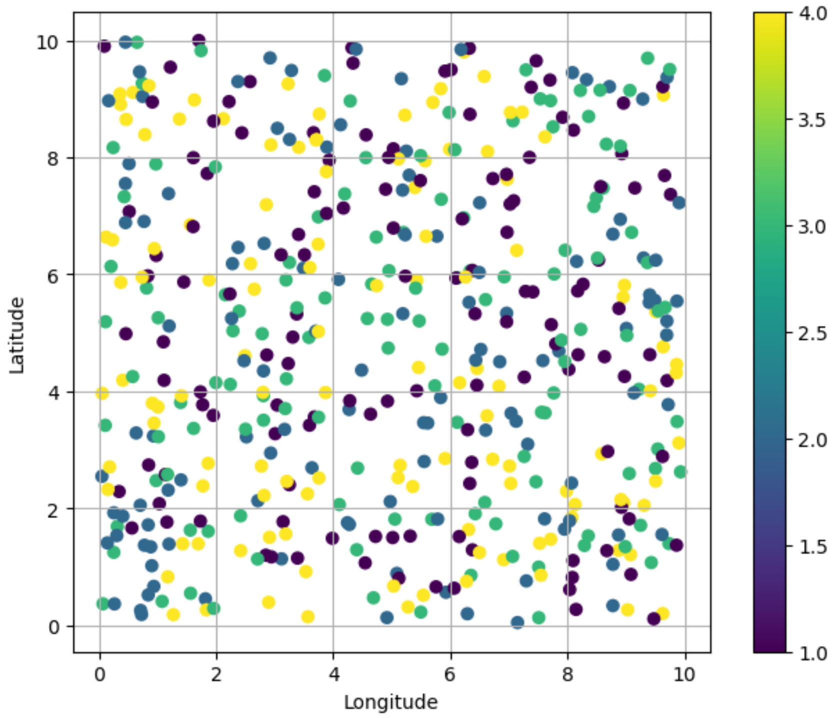

Figure 3. Simulated population density map

Note: compiled by the authors (Sembina, 2025)

The figure 4 shows a fake school assignment situation using the K-Nearest Neighbors (KNN) method. It shows how residential sites (dots) are allocated to the nearest school (black triangles) depending on how close they are to each other. The hue of each house shows which school cluster it belongs to. This shows how geographic coordinates affect how easy it is for students to get to school. The image mimics a data-driven way to allocate children to schools across cities, which lets planners check how accessible services are and find places that aren't getting enough of them. The method works well for calculating how easy it is to walk to a place, but it leaves out important elements like school size, curriculum, or socio-economic status, which makes the actual world more complicated.

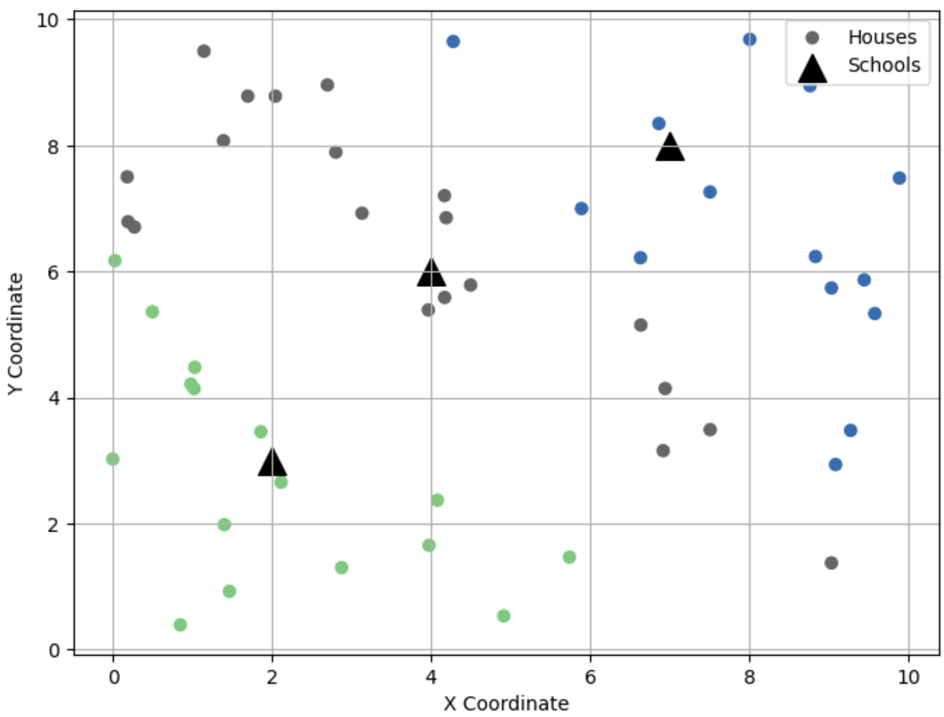

Figure 4. School assignment and KNN (Simulated)
Note: compiled by the authors (Sembina, 2025)

This method finds use practically in the assignment of school-age children to educational institutions depending on their residential location. Within a given walking radius e.g., 1,000 meters each child can be assigned to the closest school using the nearest neighbor technique. This calls for access to estimated distances between homes and schools as well as geographical coordinates of both. Moreover, the closest neighbor approach can be extended to take into account other elements such the whereabouts of extracurricular events (e.g., clubs or after-school programs). This enables more context-aware educational assignments with an eye toward not only proximity but also convenience for daily activities. Incorporating information on the locations of both educational institutions and extracurricular sites helps the model to determine the most suitable one depending on student distribution and spatial closeness. Every library imported using shortened aliases helps to expedite the development process. Read_file() from GeoPandas loads geospatial data files using techniques that produce a GeoDataFrame object. Georeferenced point geometries are produced from x and y coordinate columns using the points_from_XY() method. Clearly defined using the EPSG:4326 standard, the coordinate reference system (CRS) guarantees consistency between mapping layers (Hawelka, 2014).

Interactive maps also show spatial links between residences, businesses, and other important urban conveniences including parks, public transportation, and hospitals. Regarding residential planning, infrastructure development, and public service distribution, this spatial background helps to guide more informed decisions. This study identified, for every entry in the original GeoDataFrame, the nearest spatial feature by means of reference to another geospatial dataset.

The KNN approach is applied in this work to find the closest school to a residential address such an apartment or house. The approach could generate conflicting or less than ideal findings in highly populated areas with several schools. Furthermore sensitive to the number of neighbors selected is the accuracy of classification in KNN; hence, an improper value can produce false results. Moreover, the approach ignores important qualitative elements as educational initiatives, school performance, or socioeconomic background that could affect family decisions. Although the KNN algorithm is a useful foundation for estimating school accessibility in Almaty, more strong classification techniques including several factors could greatly increase the dependability and relevance of the outcomes.

The results of this study highlight the transforming potential of geospatial analysis and data mining in maximizing the placement of social infrastructure in metropolitan settings. We effectively found underserved areas and suggested ideal facility locations to increase service accessibility by combining GIS with machine learning algorithms. Predictive, data-driven approaches have significant potential to help urban planning activities, therefore enabling legislators to distribute resources more fairly and react more successfully to meet community demands. Among the most important results of this work is the shown efficiency of the nearest neighbor algorithm in identifying spatial service gaps and guiding the location of infrastructure. This approach proved helpful in pointing up geographical differences in healthcare and educational access. Though it is quite effective in determining physical closeness, it ignores non-spatial factors as infrastructure connectivity, service quality, or socioeconomic restrictions.

CONCLUSION

With special attention on the city of Almaty, this paper investigated the use of data mining approaches to forecast and improve the accessibility of social infrastructure in urban environments. The study sought to find ideal sites for important social facilities including schools and healthcare centers by combining GIS and computational techniques including the nearest neighbor algorithm, so fostering fair access across many urban districts. The results show that in the framework of urban planning, data mining is a strong analytical technique that helps to properly analyze spatial data and locate underprivileged areas without sufficient infrastructure coverage. Especially in low-density or peripheral communities, the use of GIS-based visualization clearly highlights areas in need of more services, therefore offering actionable insights. By urban planners, this data-driven strategy helps to ensure that public services are dispersed in a way that improves accessibility for all inhabitants, regardless of their geographic location, so supporting better informed and strategic decision-making. Still, the study also exposed certain shortcomings in current approaches. By means of strategic use of data mining tools, cities can more successfully match social services with population demands, therefore lowering discrepancies and enhancing general urban resilience. Refining these tools and handling the changing problems of urbanization will depend on ongoing cooperation among urban designers, data scientists, and policymakers going forward. By means of multidisciplinary projects, cities like Almaty can create more inclusive, responsive, and future-ready infrastructure systems benefiting every citizen.

CONFLICT OF INTEREST: The authors declare no conflict of interest.

REFERENCES

Niu, H., & Silva, E. A. (2020). "Crowdsourced Data Mining for Urban Activity: Review of Data Sources, Applications, and Methods." Journal of Urban Planning and Development, 146(2). https://doi.org/10.1061/(ASCE)UP.1943-5444.0000566

Kleinhans, R., Van Ham, M., & Evans-Cowley, J. (2015). "Using Social Media and Mobile Technologies to Foster Engagement and Self-Organization in Participatory Urban Planning and Neighbourhood Governance." Planning Practice & Research, 30(3), 237-247. https://doi.org/10.1080/02697459.2015.1051320

Zhang, F., Zhou, B., Liu, L., Liu, Y., & Fung, H. H. (2018). "Measuring Human Perceptions of a Large-Scale Urban Region Using Machine Learning." Landscape and Urban Planning, 180, 148-160. https://doi.org/10.1016/j.landurbplan.2018.08.020

Yuan, J., Zheng, Y., & Xie, X. (2012). "Discovering Regions of Different Functions in a City Using Human Mobility and POIs." Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 186-194. https://doi.org/10.1145/2339530.2339561

Long, Y., & Liu, L. (2017). "How Green Are the Streets? An Analysis for Central Areas of Chinese Cities Using Tencent Street View." PLOS ONE, 12(2), e0171110. https://doi.org/10.1371/journal.pone.0171110

Niu, H., & Silva, E. A. (2020). "Crowdsourced Data Mining for Urban Activity: Review of Data Sources, Applications, and Methods." Journal of Urban Planning and Development, 146(2). https://doi.org/10.1061/(ASCE)UP.1943-5444.0000566

Resch, B., Summa, A., Sagl, G., Zeile, P., & Exner, J. P. (2015). "Urban Emotions—Geo-Semantic Emotion Extraction from Technical Sensors, Human Sensors and Crowdsourced Data." Progress in Location-Based Services 2014, 199-212. https://doi.org/10.1007/978-3-319-11879-6_14

Frias-Martinez, V., Soto, V., Hohwald, H., & Frias-Martinez, E. (2012). "Characterizing Urban Landscapes Using Geolocated Tweets." 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing, 239-248. https://doi.org/10.1109/SocialCom-PASSAT.2012.28

Shelton, T., Poorthuis, A., & Zook, M. (2015). "Social Media and the City: Rethinking Urban Socio-Spatial Inequality Using User-Generated Geographic Information." Landscape and Urban Planning, 142, 198-211. https://doi.org/10.1016/j.landurbplan.2015.02.020

Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., & Ratti, C. (2014). "Geo-Located Twitter as Proxy for Global Mobility Patterns." Cartography and Geographic Information Science, 41(3), 260-271. https://doi.org/10.1080/15230406.2014.890072

Sembina, G., Aitim, A., & Shaizat, M. (2022). Machine Learning Algorithms for Predicting and Preventive Diagnosis of Cardiovascular Disease. 2022 International Conference on Smart Information Systems and Technologies (SIST), 1–5. https://doi.org/10.1109/sist54437.2022.9945708

Aitim, A., & Sembina, G. (2024). Modeling of human behavior for smartphone with using machine learning algorithm. News of the National Academy of Sciences of the Republic of Kazakhstan. Physico-Mathematical Series, (4), 17–28. https://doi.org/10.32014/2024.2518-1726.304

Авторлар туралы мәліметтер
Информация об авторах
Information about authors

Сембина Гулбакыт Какеевна – техника ғылымдарының кандидаты, Ақпараттық жүйелер кафедрасының қауымдастырылған профессоры, Халықаралық Ақпараттық Технологиялар Университеті, Алматы қ., Қазақстан
Сембина Гулбакыт Какеевна – кандидат технических наук, ассоц.профессор кафедры «Информационные системы», Международный Университет Информационных Технологий, г. Алматы, Казахстан
Sembina Gulbakyt – candidate of Technical Sciences, Associate Professor of Information Systems Department, International Information Technology University, Almaty, Kazakhstan
e-mail: g.sembina@iitu.edu.kz,
ORCID: https://orcid.org/0000-0003-2920-1490

Әйтім Әйгерім Қайратқызы – техника ғылымдарының магистрі, Ақпараттық жүйелер кафедрасының ассистент-профессоры, Халықаралық Ақпараттық Технологиялар Университеті, Алматы қ., Қазақстан
Әйтім Әйгерім Қайратқызы – магистр технических наук, ассистент-профессор кафедры Информационных систем, Международный Университет Информационных Технологий, г. Алматы, Казахстан
Aitim Aigerim – master of technical sciences, assistant-professor of Information Systems department International Information Technology University, Almaty, Kazakhstan
e-mail: a.aitim@iitu.edu.kz,
ORCID: https://orcid.org/0000-0003-2982-214X

Сейлова Нургуль Абадуллаевна – техника ғылымдарының кандидаты, Компьютерлік Инженерия кафедрасының қауымдастырылған профессоры, Халықаралық Ақпараттық Технологиялар Университеті, Алматы, Қазақстан.
Сейлова Нургуль Абадуллаевна – кандидат технических наук, ассоциированный профессор кафедры «Компьютерная Инженерия», Международный Университет Информационных Технологий, Алматы, Казахстан.
Seilova Nurgul – candidate of Technical Sciences, Associate Professor of Computer Engineering Department, International Information Technology University, Almaty, Kazakhstan.
e-mail: nseilova@iitu.edu.kz,
ORCID: https://orcid.org/0000-0003-3827-179X