EVALUATION AND RECOMMENDATION ALGORITHMS FOR URBAN AREAS BASED ON REVIEWS AND STATISTICS

Fabian-Florin NEAGU [1 *]

This paper presents a novel application designed to evaluate and recommend urban areas based on user reviews and statistical indicators. The originality of the work consists in the development of two main algorithms: one for categorizing and scoring reviews using both text and image analysis and another for personalizing urban area recommendations based on user profiles. These algorithms dynamically adapt to recent feedback, ensuring updated, relevant insights for users considering relocation. The proposed approach automates and refines the urban analysis process, offering a valuable tool for decision-making in urban mobility.

Keywords: urban area evaluation, recommendation algorithm, user reviews analysis, text and image processing, quality of urban life, relocation decision support, data-driven urban insights.

1. Introduction

In recent decades, the accelerated pace of urbanization has profoundly transformed the structure and functionality of modern cities, giving rise to increasingly complex urban ecosystems [1]. This rapid expansion has accentuated significant disparities between different urban zones, particularly in terms of infrastructure quality, access to public services, mobility and the availability of green spaces [2]. Consequently, a growing number of citizens are seeking digital tools capable of offering meaningful assistance in evaluating and comparing city areas, especially in the context of relocation or long-term settlement decisions [3]. However, conventional evaluation systems, which are often reliant on static indicators or general-purpose rankings, frequently fail in capturing the actual, real-time conditions of a neighborhood or in representing the lived experiences of its residents [4], [5].

Contemporary research in the field of smart city development increasingly emphasizes the value of combining objective data with subjective user perceptions in order to assess the true quality of urban life [6]. Integrating statistical metrics with user-generated content such as reviews, photos and other forms of feedback has been shown to provide a more personalized understanding of urban dynamics

[1]* Eng., M.Sc., Dept. of Automatic Control and Computers, NUST POLITEHNICA Bucharest, Romania, Corresponding author, e-mail: fabian_neagu@yahoo.com

[7]. Nevertheless, despite these advancements, most existing applications still rely on rigid evaluation filters, predefined scoring models or non-adaptive structures that fail to reflect the context-dependent nature of urban environments [8].

In response to these limitations, this paper introduces an evaluation framework composed of two interconnected algorithms that analyze user-generated content and generate personalized urban recommendations. The first algorithm combines text and image analysis to extract relevant features and compute weighted scores for various quality-of-life categories, which are continuously updated as new reviews are submitted to keep the evaluation current and community-driven. The second algorithm builds a user profile based on individual preferences and compares it with the latest category scores, producing a ranked list of urban areas that best match the user's lifestyle and overall living priorities.

As a result, by automating the evaluation of urban environments and providing customized, data-driven recommendations that incorporate both objective measures and authentic human feedback, this work contributes a scalable tool designed to support more informed and confident relocation decisions.

2. Motivation

In the context of rapidly evolving urban environments, making the decision to relocate to a specific area of a large city often involves uncertainty, due to the lack of personalized, reliable and up-to-date information about the quality of life in different neighborhoods [4]. In other words, citizens are frequently left to rely on subjective impressions, informal feedback or scattered reviews that do not necessarily reflect their individual priorities or lifestyle requirements [7].

The application proposed in this work addresses this issue by automating the process of analyzing user-generated reviews, both textual and visual, in order to extract concrete and relevant insights related to various quality-of-life indicators such as air quality, traffic congestion, green spaces, public transportation and access to education. These aspects, while essential in everyday life, are rarely evaluated in a structured and dynamic way that can assist people in making informed relocation choices [8].

The motivation behind this project lies in the need for a more intelligent, data-driven mechanism that can process large volumes of subjective feedback and translate it into meaningful, category-based evaluations of urban zones. Rather than providing general or pre-defined rankings, the system evaluates each zone dynamically, adjusting its scores based on the latest reviews submitted by users, ensuring the information remains current and reflective of real urban conditions.

Furthermore, the system extends beyond simple data aggregation by integrating a personalized recommendation module that interprets user preferences collected through a structured form and matches them with the characteristics of each city sector. This enables the application not only to inform but also to guide users toward areas that best align with their lifestyle, such as commuting needs or access to education. Ultimately, the solution bridges the gap between fragmented urban feedback and the real needs of residents, providing a structured tool that supports informed decision-making in complex urban environments.

On the other hand, from a developmental standpoint, the current implementation should be viewed as a proof of concept and a minimum viable product (MVP) designed to validate the proposed dual-algorithm architecture and assess the feasibility of multimodal analysis for urban evaluation. Rather than providing a final or large-scale deployment, this stage focuses on demonstrating the system's functional coherence and its potential to integrate textual and visual user feedback into a unified, adaptive decision-support framework.

3. Related work

In recent years, the evaluation of urban areas through data-driven and participatory approaches has gained significant attention in the context of smart city research [7]. In addition, various digital platforms have been developed to collect and analyze citizen-generated information for assessing quality-of-life indicators such as mobility, air quality and access to green spaces [8]. However, most existing systems focus either on aggregating data or on visualizing static indicators, without providing adaptive, personalized recommendations [9].

One relevant initiative is Data4City platform, a civic data hub that combines environmental sensor feeds with citizen-reported observations to estimate urban quality indicators such as pollution levels and public cleanliness [10]. Comparing this solution with the one proposed in this study, it can be highlighted that while Data4City platform effectively integrates objective and subjective data sources, it primarily serves as an aggregation and visualization tool, lacking a personalization component or real-time adaptive scoring.

Another example is MeinGrün, an application developed for the cities of Dresden and Heidelberg, which supports citizens in discovering urban green spaces through map-based exploration and filtering by individual preferences [11]. Although highly useful in promoting urban sustainability, the application focuses mainly on spatial accessibility to parks and green zones, without extending to broader aspects such as infrastructure, traffic or education.

Moreover, recent research trends also point toward the integration of generative artificial intelligence in urban analytics, aiming to improve citizen engagement and context-sensitive data interpretation [12], [13]. In this direction, the proposed study plans to gradually introduce responsible generative AI components focused on transparency, governance and controlled generation in order to assist with review summarization and anomaly detection, while keeping the focus on strict ethical and oversight principles.

Compared to these existing approaches, the solution presented in this paper introduces a more holistic and dynamic framework that combines textual and visual user feedback with statistical indicators. Furthermore, by integrating a personalized recommendation mechanism based on user profiles, the proposed system not only evaluates but also guides relocation decisions, adapting to recent community input and ensuring relevance for each user's lifestyle and priorities.

4. Algorithms

Before presenting the algorithmic extensions and ongoing development, it is important to mention that the current version of the platform should be regarded as a proof of concept, primarily tested on Bucharest's six administrative sectors. In other words, this initial implementation focuses on validating the system architecture and ensuring the functional coherence of the algorithms. Because of this limited geographical and linguistic scope, the keyword dictionary and object–category mapping tables currently cover the most relevant terms and visual elements for the local context and will be further expanded and adapted for deployment in other urban environments. In addition, future work will extend the dataset to multiple cities, refining the mappings to reflect regional and cultural differences and enabling broader scalability across diverse urban contexts.

Another point to consider is that at this stage, the system primarily relies on rule-based association mechanisms, a design choice that ensures transparency, interpretability and control during the early development phase. As part of its ongoing evolution, work has already begun on transitioning toward model-driven intelligence. An initial dataset has been collected to support the training of machine learning components, including NLP embedding models for semantic text interpretation and deep-learning architectures for visual context analysis. As a result, these upcoming enhancements aim to make the platform more adaptive and data-driven while preserving the responsible AI principles and governance standards that guide its development.

4.1 Review Association Algorithm Based on Image Analysis

The first core component of the proposed system is the automatic interpretation of user-submitted images associated with urban zone reviews, a process designed to extract meaningful visual indicators that reflect the conditions and characteristics of the reviewed areas. This component relies on object detection techniques based on pretrained deep learning models, specifically accessed through the Deep Java Library (DJL), which enables the identification of relevant elements within user-uploaded images.

Moreover, by using this automated recognition process, the system is able to associate visual content with specific urban evaluation categories. In other words, for each object detected, the algorithm references an internal mapping table that links object classes to quality-of-life dimensions, such as traffic, public transport or environmental factors. This association allows the platform to determine which aspects of the city are visually referenced in the review and to incrementally adjust the score of the relevant categories for the zone in question. In addition, this mechanism ensures that updates are based not only on textual reviews but also on the contextual information embedded in images. Likewise, in order to maintain fairness and consistency, each new review contributes equally to the category score calculation by applying a balanced average formula that considers the current score, the number of existing reviews and the score from the new input.

Table 1 below presents a representative subset of the associations established between the visual objects identified in user-submitted images and the predefined urban quality-of-life categories employed by the image analysis algorithm. These mappings serve as the foundation for assigning visual input to relevant evaluation criteria, enabling the system to interpret contextual clues within each image. In addition, it is important to note that this table does not represent an exhaustive list as the associations can be further expanded and refined to incorporate new object types and more nuanced classification rules as the system evolves or adapts to new urban contexts.

Table 1

Object-to-Category Mapping Table

Detected Object	**Mapped Urban Category**
bench, flower, plant, tree, grass, bush, shrub, leaf, park, garden, fountain, lawn, hedge, pond, riverbank, flower bed, trash can, recycling bin, gazebo, lamp post, walking trail, open field, wooden fence, picnic table	Air Quality / Green Spaces
car, truck, bus, van, motorcycle, bicycle, traffic light, road sign, crosswalk, street, bridge, tunnel, construction barrier, cone	Traffic / Roads & Infrastructure
backpack, desk, school, book, notebook, pencil, classroom, blackboard, student, library	Educational Facilities
tram, train, metro, subway, bus stop, station sign, railway, bike rack, ticket machine, timetable board, platform, turnstile	Public Transportation

Each detected object can be associated with one or more quality-of-life categories, allowing the system to capture overlapping aspects of urban living reflected in a single image. For instance, if the algorithm identifies both a bus and a traffic light within the same image, it will contribute simultaneously to the evaluation of both the „Public Transportation” and „Roads & Infrastructure” categories for the corresponding area. As a result, this flexible association model ensures a more comprehensive and realistic interpretation of the visual context.

Once the relevant categories have been identified, the algorithm advances to the score update phase. In this step, each category's score for the reviewed area is recalculated by integrating the new information into the existing dataset. To preserve fairness and consistency across all user contributions, the system applies a weighted averaging method that combines the new score with the cumulative one, ensuring that every review contributes proportionally to the overall evaluation.

The formula used for updating category scores is as follows:

$$newScore = \frac{oldScore\ x\ nrOfReviews + currentReviewScore}{nrOfReviews + 1} \quad (1)$$

- oldScore - previous score of the identified category in the zone,
- nrOfReviews - number of previous reviews that contributed to that score,
- currentReviewScore - user-assigned overall rating in the new review.

This method guarantees that each individual review contributes equally to the overall evolution of category scores, thereby preventing any single input from exerting a disproportionate influence on the final evaluation of an urban area. Moreover, since the algorithm is triggered automatically upon the submission of a review, the entire system remains dynamic and responsive, continuously updating the scores to reflect the most recent user experiences and real-time urban conditions.

In conclusion, this algorithm enhances the platform by transforming unstructured visual input submitted by users into structured evaluation data. Through this process, it reinforces the system's multi-modal assessment capabilities and plays a central role in ensuring that urban zone evaluations remain accurate, timely and grounded in the lived reality of residents.

4.2. Review Association Algorithm Based on Textual Analysis

The second algorithm integrated into the system complements the visual analysis component by processing the textual content of user-submitted reviews. While image recognition provides valuable environmental context, relying on it exclusively may lead to incomplete or ambiguous associations, so in order to increase the accuracy and coverage of urban zone evaluation, this algorithm interprets written feedback to detect references to specific urban conditions.

From the functional perspective, this mechanism operates by scanning both the title and the description of each review, identifying relevant keywords that are matched against an internal predefined dictionary. In other words, each keyword is linked to one or more quality-of-life categories and when a match occurs, the corresponding category is marked as relevant, triggering an update to its score for the reviewed zone using the same averaging logic as in the first algorithm. Moreover, in order to improve linguistic flexibility, the keyword detection system also supports variations such as plural forms, derivational suffixes or other natural language constructs.

Table 2 presents a representative sample of the keyword-to-category associations used by the system. This subset illustrates the underlying implementation logic, while the full dictionary encompasses a wider range of expressions to capture the diversity of user language.

Table 2

Sample Keyword-to-Category Mapping Table

Urban Category	Example Keywords
Air Quality	air, pollution, smog, emissions, clean air, fresh air, breathable, environment, plants, nature, toxins, dust, air quality, odor, smell
Green Spaces	park, tree, green area, garden, flower, lawn, playground, forest, plants, quiet place, open space, walking path, benches, relaxing area
Roads & Infrastructure	road, pavement, street, highway, intersection, bridge, tunnel, asphalt, damaged road, potholes, construction, maintenance, lighting, crosswalk, pedestrian zone
Public Transport	bus, tram, subway, metro, train, station, bus stop, ticket, timetable, delay, crowded, public transport, commute, route, platform
Traffic	car, congestion, traffic jam, rush hour, vehicle, gridlock, travel time, average speed, traffic light, signaling, noise, driving, heavy traffic
Educational Facilities	school, kindergarten, faculty, university, campus, classroom, amphitheater, library, teacher, student, exam, learning center, laboratory, education building

When relevant keywords are detected within a review, the system proceeds to update the scores of the associated quality-of-life categories using the same weighted averaging method described in the previous sections. This unified approach ensures consistency between the processing of visual and textual inputs, maintaining fairness across all forms of user contributions. As a result, each review, regardless of format, produces an equal and proportionate influence on the ongoing assessment of urban zones.

As a result, by integrating this layer of textual analysis, the system significantly enhances the reliability of its evaluations, particularly in cases where visual input alone may be insufficient or unavailable. In addition, it allows the platform to capture a wider range of urban experiences as expressed directly by users through natural language, reinforcing the overall responsiveness and inclusivity of the evaluation process. Furthermore, the development of AI-driven language models is already underway to improve this component, enabling more accurate automated review classification in future iterations of the system.

4.3 Personalized Urban Recommendation Algorithm

The recommendation system implemented in this application is based on a profile-matching algorithm that analyzes user preferences to identify urban zones that align with the individual's needs. In other words, the purpose of this module is to assist users in selecting areas that best match their lifestyle and expectations.

The algorithm begins by processing a form completed by the user, which captures their preferences and priorities such as transport access, green spaces or air quality. Based on these inputs, a compatibility score is calculated for each urban zone, indicating how closely its characteristics match the user's stated needs. The characteristics of each zone are derived from dynamically updated category scores on the platform's evaluation page, ensuring that recommendations are based on current and contextually relevant data. To further enhance accuracy, each question in the form is linked to specific urban quality categories, allowing the system to tailor recommendations more effectively. For example, a person working remotely may value air quality and noise control, while someone with a daily commute is likely to prioritize public transport and traffic conditions.

Finally, once the form is submitted, the algorithm evaluates each sector individually and for each question, the associated category is identified and the sector's score in that category is added to its total compatibility score. Likewise, this process is repeated across all questions and all sectors, resulting in a ranked list based on calculated compatibility.

In functional terms, the algorithm works within the following steps:

- For every sector, initialize a score to zero
- For each user answer, determine the relevant urban category
- Add the sector's rating in that category to its cumulative score
- After processing all answers, sort the sectors by score in descending order

To sum up, the personalized recommendation algorithm provides tailored suggestions for urban zones by combining user preferences with real-time evaluation data. By aligning individual priorities with dynamically updated category scores, the system delivers relevant, context-aware recommendations that support informed decisions and help users find areas suited to their lifestyle.

5. System integration

To ensure the coherence and effectiveness of the proposed architecture, all three algorithms such as visual analysis, textual analysis and personalized recommendation have been fully integrated into a unified system. The platform offers a dynamic evaluation interface where users can submit reviews, either in text or image form and receive personalized suggestions based on their lifestyle preferences. In addition, each algorithm contributes to a cumulative quality score database, which evolves continuously as new data is added by the community.

Moreover, the functionality of the system has been structured around the following interaction steps:

- The user selects a zone and submits a review containing text and an image
- The image is analyzed by the DJL-based visual detection module to identify urban objects and map them to predefined categories
- The text review is processed in parallel to extract keywords linked to the same quality-of-life categories
- The combined results update the zone's category scores through a weighted averaging mechanism
- If the user requests relocation suggestions, the recommendation engine compares their profile with the updated scores for each sector

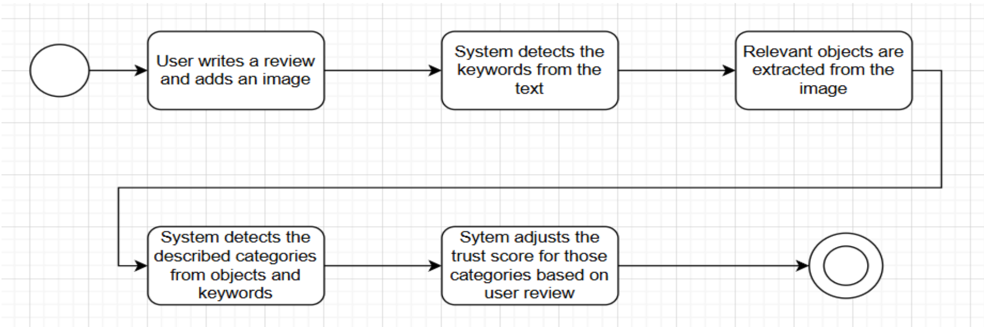

Fig. 1. Flow for analyzing a new review

As a result, from an application perspective, the proposed architecture ensures full automation of data processing and category evaluation, together with real-time responsiveness to new content. In addition, it integrates subjective user feedback with data-driven logic, creating a personalized experience that aligns urban assessments with individual lifestyles. Likewise, the modular structure also simplifies future expansion and supports integration with broader urban planning platforms and open data repositories.

6. Evaluation

The evaluation process aimed to assess how effectively the prototype analyzes multimodal user feedback and generates relevant urban recommendations. To achieve this, a dataset of approximately 120 user reviews, each containing both descriptive text and representative images, was collected from Bucharest's six administrative sectors. The entries were manually labeled into quality-of-life categories defined in the previous chapters providing a reliable reference for comparison and validation. As a result, this manual annotation allowed the evaluation to measure not only classification accuracy but also the system's ability to interpret and align subjective user input with objective urban characteristics.

Table 3

Performance indicators

Performance metric	**Value**
Accuracy - text module	0.82
Accuracy - image module	0.78
Accuracy - overall	0.8
Precision - recommendation module	0.82

The assessment procedure involved several iterations in which new reviews were progressively introduced to test the adaptability and consistency of the scoring mechanism. The text classification module achieved an accuracy of 82%, while the image-based analysis reached 78%, leading to an overall multimodal classification accuracy of around 80%. In addition, the scoring process remained stable as new data was incorporated, demonstrating that the weighted averaging method effectively balanced recent and historical inputs, preventing bias toward frequent contributors or newly added content.

On the other hand, the recommendation module achieved a precision of 0.82, indicating a strong correspondence between user preferences and the suggested urban sectors. This result suggests that the profile-matching logic performs reliably even in a small-scale deployment, confirming the potential for scaling the approach to larger datasets and more diverse linguistic contexts.

Overall, the findings demonstrate that the MVP operates consistently, validating both its multimodal analysis pipeline and its recommendation strategy and provide a solid analytical foundation for the next development phase involving large-scale data integration and generative AI components.

7. Ethical and Data-Quality Considerations

Given that the proposed system functions as a web-based platform where users can submit both textual and visual reviews, ensuring data integrity and privacy represents a fundamental requirement. To maintain the quality and reliability of the collected information, all submissions pass through a dedicated validation process before being included in the computation of area scores. This process is coordinated by an administrator who verifies the authenticity, relevance and tone of each entry to ensure that only credible and meaningful reviews contribute to the evaluation. Entries identified as duplicates, containing inappropriate language or deemed irrelevant are excluded from the dataset, preserving the overall consistency of the analysis.

This moderation workflow helps maintain a transparent and balanced scoring mechanism, resistant to manipulation or spam. At the same time, the platform incorporates strict access-control and anonymization measures that protect user data and uphold privacy throughout the aggregation and processing stages. By filtering low-quality or biased inputs, the system reinforces the credibility of its outputs and promotes fairness across all user contributions, ultimately enhancing the trustworthiness of the generated insights.

Looking toward future development, the integration of AI-assisted moderation tools is planned as part of the platform's next phase. These enhancements will be implemented under responsible AI principles, emphasizing transparency, explainability and continuous human oversight within a clear governance framework. In particular, tools such as Microsoft Presidio and similar privacy-oriented solutions are being explored to support automatic content screening and sensitive data protection. Nonetheless, the current version remains a proof of concept, primarily focused on validating the core architecture and operational workflow. Future iterations will build upon these foundations by embedding responsible AI mechanisms and expanding automated moderation capabilities in line with ethical and regulatory standards.

8. Conclusions

In conclusion, the personalized urban recommendation algorithm presented in this work demonstrates a structured and context-aware approach to aligning individual user needs with the real characteristics of city zones. By transforming subjective preferences expressed through a guided questionnaire into measurable compatibility scores, the system offers a practical and scalable method for generating urban relocation suggestions tailored to each user's lifestyle. Although the current implementation functions as a proof of concept and a minimum viable product (MVP) developed to validate the proposed architecture, the obtained results confirm the feasibility and consistency of the multimodal evaluation framework.

The core strength of the algorithm lies in its ability to dynamically integrate updated evaluation data, derived from crowdsourced reviews, into the recommendation process. This ensures that the recommendations reflect not only the user's expectations but also the current state of each urban sector as perceived and reported by other residents. Such integration of real-time community feedback with profile-based analysis provides a more accurate and responsive alternative to static recommendation models.

As part of the platform's future development, which has already begun, the second phase of the study focuses on integrating generative artificial intelligence within digital urban communities. This stage introduces natural language processing (NLP) models for advanced semantic analysis of user reviews and automated content summarization, supported by a curated dataset currently in progress. In line with responsible AI principles, the new components will include clear governance mechanisms, content control measures and human oversight to ensure transparency, accountability and ethical alignment throughout the system's evolution.

In summary, the proposed framework represents an early yet significant step toward data-driven urban decision support. By combining transparency, adaptability and responsible AI integration, the platform lays the groundwork for a new generation of intelligent, citizen-centered systems capable of enhancing the way people evaluate and interact with their urban environments

REFERENCES

[1] J. G. B. Wesz, L. I. G. Miron, I. Delsante, P. Tzortzopoulos, Urban quality of life: a systematic literature review, Urban Science, Vol. 7, Iss. 2, p. 56, 2022

[2] A. Dalia, J. Ruzevicius, Quality of life and its components' measurement, Engineering Economics, Vol. 52, Iss. 2, pp. 43–48, April 18 2007

[3] E. Cesario, Big data analytics and smart cities: applications, challenges and opportunities, Data Mining and Management, Vol. 6, 2023

[4] R. Volk, M. Rambhia, E. Naber, F. Schultmann, Urban resource assessment, management and planning tools for land, ecosystems, urban climate, water and materials—a review, Sustainability, Vol. 14, Iss. 12, Art. 7203, 2022.

[5] L. Zhang, M. Chen, "AI-driven decision support for smart city mobility and environmental monitoring," Sustainable Cities and Society, vol. 74, 2021.

[6] R. Volk, M. Rambhia, F. Schultmann, "Urban resource assessment and citizen participation in data-driven planning," Sustainability, vol. 14, 2022.

[7] S. Poria, A. Cambria, "Multimodal sentiment and urban experience mining: a survey," Information Fusion, vol. 91, 2023.

[8] F. Wang, Z. Li, "Smart city data fusion using multimodal learning techniques: a comprehensive review," IEEE Internet of Things Journal, vol. 8, no. 23, 2021.

[9] K. Yao, H. Sun, "Deep learning-based urban recommender systems: combining social data and spatial context," IEEE Access, vol. 9, 2021.

[10] A. Marques, L. Ferreira, "User-centric approaches in smart city recommender systems," Sensors, vol. 23, no. 8, 2023.

[11] MeinGrün Project, "Green Urban Mobility and Open Data for Cities," Fraunhofer Institute for Industrial Engineering, 2022.

[12] J. Kim, "Integrating NLP embeddings in urban analytics and citizen feedback systems," Computers, Environment and Urban Systems, vol. 110, 2024.

[13] M. Dignum, "Responsible Artificial Intelligence: designing for human values," AI & Society, vol. 38, 2023.