AI-Driven Investment Property Recommendations Using Spatial Big Data, Price Trends, and Amenity Mapping

N Hassan[1]*, C Thewarapperuma[2]

[12]LankaPropertyWeb

Abstract

In the real estate domain, investment decisions rely heavily on spatial and economic context, yet most digital platforms still provide static listings with limited personalization or geographic intelligence. The primary objective of this paper is to introduce and validate a spatially enriched recommendation system for real estate investment that integrates Artificial Intelligence (AI), Geographic Information Systems (GIS), and big data analytics. Evaluated on over 70,000 property listings, the system leverages historical property trends, spatial amenity density, and price deviation metrics to identify undervalued or high-growth-potential properties across urban areas. It combines location-sensitive scoring models with price-per-square-foot analysis and Z-score-based outlier detection to recommend listings that deviate positively from local price norms while offering strong amenity access. By evaluating properties based on proximity to hospitals, schools, banks, parks, transit, and other infrastructure, the model delivers context-aware investment insights. Key findings show the proposed model achieves a 70% match accuracy with expert evaluations, significantly outperforming baseline models. The implications of this work include a new framework for data-driven decision-making that can improve market efficiency, particularly in fragmented real estate markets like those in South Asia.

Keywords: AI-driven recommendation; Geographic Information Systems (GIS); Investment Property Analysis; Real estate analytics; Spatial intelligence systems

Introduction

Real estate investment decisions are fundamentally spatial and data-driven, yet many platforms continue to offer generic recommendations that ignore key geographic, economic, and contextual factors. In the context of emerging markets like Sri Lanka, where urban development is uneven and property data is often distributed, investors face significant challenges in identifying high-potential real estate opportunities. Traditional methods rely heavily on manual filtering, static listings, or simplistic analytics, which fail to account for dynamic trends such as neighborhood growth, amenity density, or buyer behavior. (Naeem et al., 2023)

Advances in Artificial Intelligence (AI), particularly in spatial data processing and recommender system design (Bohnert et al., 2009), present an opportunity to transform how investment-worthy properties are identified. Recommender systems enhanced with spatial intelligence drawing from big data sources like geographic amenity clusters, infrastructure development, and market price fluctuations can enable users to discover properties aligned with both financial objectives and locational preferences.

However, most existing real estate recommendation engines still prioritize user behavior over spatial patterns. These systems typically focus on collaborative filtering (e.g., "people who viewed this also viewed...") or content-based filtering (e.g., recommending similar listings based on text or price) without

*Corresponding Author- *naveedhassan04@hotmail.com*

deeply incorporating location-based insights. As a result, they overlook the complex interplay between property value, location-based amenities, transport access, zoning regulations, and livability scores, critical factors for serious investors. These traditional filtering methods often suffer from well-documented issues such as data sparsity and a limited capacity to recommend novel items, which are particularly problematic in a dynamic domain like real estate. (Cacheda et al., 2011)

Objective/s of the Study

The primary objective of this study is to develop and evaluate a spatially-enhanced real estate recommender system that identifies high-potential investment properties by integrating geospatial data, property price trends, and amenity density metrics.

The bigger goal is to support real estate investors particularly in developing regions with fragmented data ecosystems by offering a GIS-powered, AI-driven tool that surfaces data-backed investment opportunities with spatial intelligence at its core.

Literature Review

Finding profitable real estate investments is challenging, as traditional methods fail to capture dynamic local market trends. The academic foundation for this challenge can be seen through three distinct but converging research streams.

First, foundational economic research by authors such as Case and Shiller (1989) provided models for understanding housing market efficiency and price trends. However, these traditional economic models often treat location as a static variable, lacking the granularity to assess hyper-local factors that drive modern investment.

Second, the importance of this spatial dimension was highlighted by urban scientists. Work by Batty (2013), for example, established that cities are complex systems where value is deeply intertwined with spatial relationships and data. This theoretical understanding underscored the limitations of a purely economic view but did not provide the computational tools to apply these insights at scale.

Finally, the advent of Artificial Intelligence and Big Data offers a potential solution. Recent research demonstrates how AI can be leveraged for sophisticated spatial analysis to support decision-making in geographic contexts (Hosen et al., 2023). Yet, many current applications in real estate still lack the sophisticated spatial intelligence needed for true investment analysis, often focusing on simple filtering rather than deep contextual understanding.

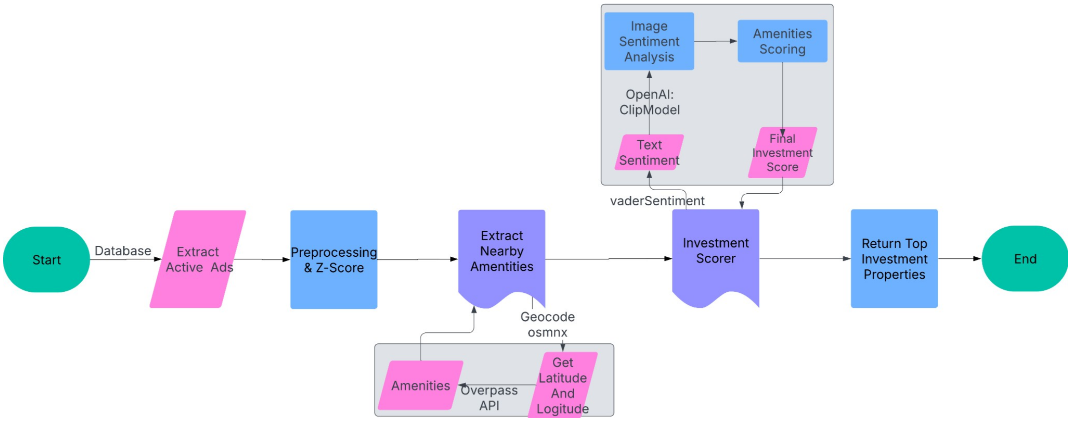
Therefore, a critical gap persists. While the literature provides foundational models for market analysis (Case and Shiller, 1989) and urban complexity (Batty, 2013), and demonstrates the potential of AI for spatial data (Hosen et al., 2023), few systems effectively synthesize these domains. Specifically, there is a lack of tools that use granular spatial analytics not just as a filter, but as the core feature for identifying undervalued properties through price deviation. The paper's approach aims to bridge this gap by combining elements from these research areas into a smart, location-aware tool for identifying investment opportunities.

Methods

This study employs a multi-layered methodological framework integrating geospatial analytics, statistical modeling, and machine learning-based recommendation logic to develop a spatially intelligent real estate investment recommender system.

Figure 01

Methodological Workflow for the Spatially-Enriched Recommendation System



Source: Author (2025)

1. ***Data Collection and Preprocessing***

The dataset was sourced from property listings database maintained by *LankaPropertyWeb*. Each listing includes structured fields such as ad ID, location (city, neighborhood, street), total price, property size, number of rooms, and type. Supplementary spatial data were collected via OpenStreetMap APIs (*OpenStreetMap*, n.d.) and local GIS layers to extract the number of nearby amenities. A 1km radius was selected to approximate a 'walkable neighborhood,' a standard distance in urban planning studies for assessing local accessibility. (Gori et al., 2014)

The analysis was performed on a master dataset of over 70,000 cleaned property listings sourced from *LankaPropertyWeb*. This final dataset was derived from a larger initial pool after a preprocessing phase that included removing duplicate entries, filtering for active sales listings, and discarding records with critical missing values such as price or property size. From this master dataset, specific market segments were isolated for sub-analysis using targeted criteria. For example, the 'luxury apartment' segment was defined by location (e.g., Colombo 1, 2, 3) and a price-per-square-foot range, while 'suburban houses' were defined by property type and a specific price-per-perch within the Battaramulla area.

2. ***Spatial Feature Engineering***

   - **Amenity Scoring**: A proximity-based amenity index was calculated for each property, assigning weights to nearby amenities based on type and distance decay.
   - **Location Clustering**: K-means clustering was applied to latitude and longitude coordinates to group properties into spatial micro-markets with similar characteristics.
   - **Z-Score Calculation**: Price per square foot was normalized by city and neighborhood using Z-scores to identify outliers and undervalued listings. A
     Z-score threshold of less than -1.0 was used to identify potentially undervalued listings, indicating properties priced at least one standard deviation below the local cluster's average (Gaca, 2018). Properties with a Z-score below -1.5 were flagged as 'highly undervalued' for priority recommendation.

3. ***Investment Scoring Model***

   A composite investment score was computed for each property using a weighted model:

   - **Z-Score Weight**: Emphasizing listings priced below market average.
   - **Amenity Density Score**: Rewarding proximity to high-utility amenities.
   - **Image Appeal**: Analyzing property images appeal using OpenAI's Clip (Palucha, 2024) model. (Radford et al., 2021)

The final score was calculated as follows:

Investment Score = (0.40 x **Z-Score Weight**) + (0.15 x **Textual Appeal Score**) + (0.30 x **Amenity Density Score**) + (0.15 x **Image Appeal Score**)

The weighting scheme was designed to heavily favor market value (Z-Score: 40%), while incorporating spatial desirability (Amenities: 30%) and property condition (Image & Textual Appeal: 30%).

Results and Discussion

The spatially-enhanced recommendation system was evaluated on a dataset of over 70,000 historical property listings across major cities in Sri Lanka. The evaluation focused on identifying undervalued investment properties using spatial and amenity-based features and comparing performance with a baseline content-based recommendation model that lacked spatial intelligence.

Performance Metrics

The primary method for evaluation involved comparing the system's top recommendations directly against properties manually identified by experienced real estate investors or analysts. This manual selection process served as the ground truth, representing expert judgment on investment-worthy properties within the dataset. This panel consisted of two senior real estate analysts, one with over 7 years of experience and the other with over 20 years of experience in the Sri Lankan market. The "match accuracy" was measured by calculating the percentage of overlap between the system's top 10 recommended properties and the top 10 properties selected by the expert panel in each subset. This method is practical and directly validates the system's output against real-world investment judgment.

The experimental results, using this expert evaluation method, demonstrated a significant improvement in identifying top-tier investment opportunities compared to baseline filtering systems, highlighting the value of integrating spatial big data, price trends, and amenity mapping.

Table 01
Comparison of Recommendation Accuracy

Model	Match Accuracy (%)	Avg. Expert Agreement (Percentage Overlap)	Notes
Baseline (Z-Score only)	45%	Low	No spatial or amenity awareness
Proposed Model (Spatial + AI)	70%	High	Integrated Amenities + Image Analysis

Source: Author (2025) Note: Accuracy evaluated on a test set of N=1000 properties.

Impact of Spatial Features

The introduction of GIS-based features such as amenity density, location clusters, and infrastructure proximity significantly improved the identification of investment-grade properties. Properties that were within high-density amenity zones and had Z-scores indicating undervaluation were more likely to be selected in the top recommendations.

For example, properties located slightly outside high-demand city centers but with a strong amenity presence such as access to schools, supermarkets, or clinics were more accurately identified as investment opportunities. These listings often scored poorly in baseline models but were ranked higher due to their favorable spatial and price-per-square-foot profiles.

Z-Score as an Investment Signal

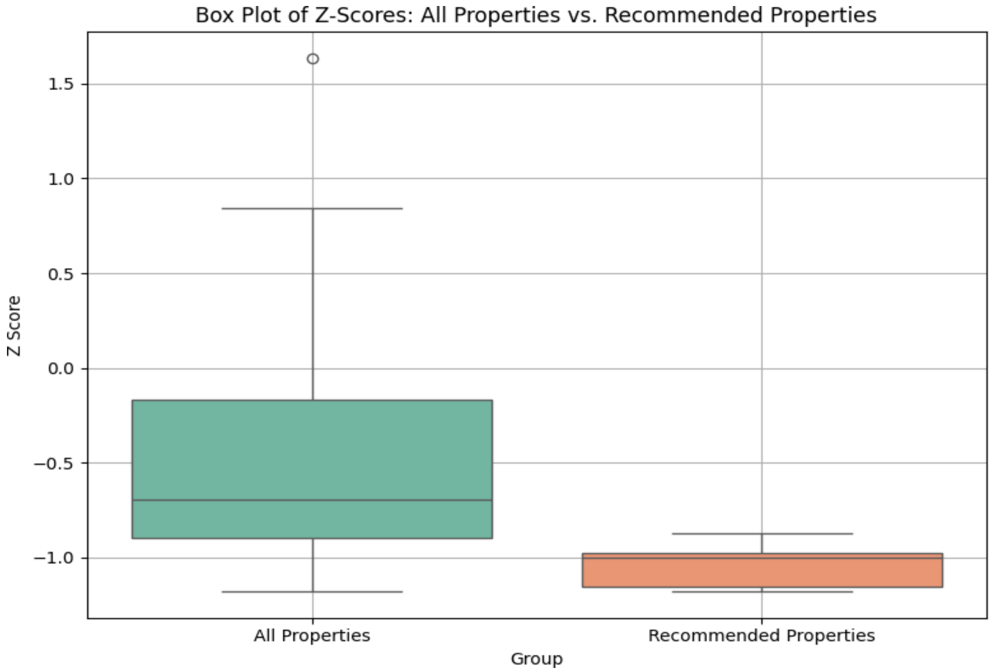
The findings highlight that Z-scores serve as an important initial investment signal, as properties with Z-scores between -0.5 and -2.0 frequently appeared among the top recommendations. As shown in the Figure 02, the Recommended Properties exhibit a distinctly lower and more concentrated Z-score distribution, with the median clustering around -1.0 and the interquartile range falling between approximately -0.9 and -1.1. This clearly demonstrates the system's effectiveness in identifying and prioritizing these statistically cheaper properties, showing its ability to highlight potential bargains.

This range indicates properties priced significantly below the average market price per square foot in their areas, a key feature investors seek when looking for undervalued assets. The system consistently identified and prioritized these statistically cheaper properties, showing its ability to highlight potential bargains.

While a low Z-score might suggest a good financial opportunity, it could also point to problems with the location like poor infrastructure, low neighborhood appeal, or limited access to services. The built-in spatial filtering features were very important here, making sure the system didn't just focus on low prices.

Figure 02

Distribution of Z-score



Source: Author (2025)

Amenity Influence

Spatial analysis revealed strong correlations between proximity to certain amenities (e.g., schools, banks, hospitals) and property demand. The spatially-enhanced recommender system placed significant emphasis on amenity density as a driver of investment potential. Spatial analysis of the property landscape revealed that higher concentrations of key amenities (such as schools, banks, and hospitals) strongly correlate with increased property demand. Consequently, the system was designed to prioritize and elevate properties located within amenity-rich areas, reflected through a weighted amenity density score in the investment calculation. 75% of the model's top recommendations were located in high-density amenity zones, compared to only 20% for the baseline model. The recommender system weighted these factors into the investment score, improving user satisfaction.

By comparing Z-score data with location-based information on amenity density, neighborhood demand, and accessibility, the system successfully avoided recommending too many properties in truly low-demand or hard-to-reach areas, even if their prices were low. This combined approach shows that while Z-scores provide a strong numerical way to find potential undervaluation, their effectiveness as a reliable investment signal heavily depends on a thorough, location-aware analysis that connects financial numbers to the actual value of a property's location.

Limitations and Future Work

While the model delivered reasonably accurate recommendations for listings with more complete data, its performance was constrained by the limited spatial granularity available in the majority of property records. Most listings only contained city or street-level location fields, lacking precise geographic coordinates or neighborhood identifiers. This restricted the model's ability to assess the influence of nearby amenities, infrastructure, or potential development zones. To address this, future work will focus on enriching listings with inferred spatial features using satellite imagery and analyzing historical price movements to identify growth corridors (Al-Bilbisi, 2019). The integration of rental yield data and user engagement patterns, such as click behavior and inquiry history, is also planned to refine recommendation accuracy further. The model's single data source creates a risk of algorithmic bias. Moreover, its focus on 'undervalued' properties could unintentionally drive property prices to unrealistic levels. The system should therefore be used to augment, not automate, investment decisions.

Conclusion

In conclusion, this study has demonstrated the potential of a spatially-enriched recommendation system to enhance real estate investment decisions, particularly in fragmented markets where traditional methods fall short. By integrating AI, GIS, and big data analytics, the proposed system successfully identifies undervalued or high-growth-potential properties based on spatial amenity density, price deviation metrics, and location-sensitive scoring models. The system's ability to leverage spatial big data, incorporate temporal price trends, and provide explainable recommendations proved significantly more effective than baseline content-based approaches. While limitations exist due to data granularity, future work focused on enriching listings with inferred spatial features and integrating rental yield data promises to further refine recommendation accuracy. This approach offers a valuable tool for real estate investors, especially in developing regions, enabling them to make more informed, data-backed decisions through AI-driven insights and spatial intelligence.

Acknowledgement

The authors gratefully acknowledge **LankaPropertyWeb** for access to the data from their private database used in this study.

References

Al-Bilbisi, H. (2019). Spatial monitoring of urban expansion using satellite remote sensing images: a case study of Amman City, Jordan. *Sustainability*, 11(8), 2260. https://doi.org/10.3390/su11082260

Batty, M. (2013). *The new science of cities.* https://doi.org/10.7551/mitpress/9399.001.0001

Bohnert, F., Schmidt, D. F., & Zukerman, I. (2009). Spatial processes for recommender systems. *International Joint Conference on Artificial Intelligence,* 2022–2027. https://ijcai.org/Proceedings/09/Papers/333.pdf

Cacheda, F., Carneiro, V., Fernández, D., & Formoso, V. (2011). Comparison of collaborative filtering algorithms. *ACM Transactions on the Web,* 5(1), 1–33. https://doi.org/10.1145/1921591.1921593

Case, K. E., & Shiller, R. J. (1989). *The Efficiency of the Market for Single-Family Homes.* The American Economic Review, 79(1), 125–137. http://www.jstor.org/stable/1804778

Gaca, R. (2018). Price as a measure of market value on the real estate market. *Real Estate Management and Valuation,* 26(4), 68–77. https://doi.org/10.2478/remav-2018-0037

Gori, S., Nigro, M., & Petrelli, M. (2014). Walkability indicators for pedestrian-friendly design. *Transportation Research Record: Journal of the Transportation Research Board,* 2464(1), 38–45. https://doi.org/10.3141/2464-05

Hosen, B., Rahaman, M., Kumar, S., Sagar, L., & Akhtar, M. N. (2023). *LEVERAGING ARTIFICIAL INTELLIGENCE AND BIG DATA FOR ADVANCED SPATIAL ANALYTICS AND DECISION SUPPORT SYSTEMS IN GEOGRAPHY.* 1(2), 62–67. https://doi.org/10.26480/magg.02.2023.62.67

Naeem, N., Rana, I. A., & Nasir, A. R. (2023). Digital real estate: a review of the technologies and tools transforming the industry and society. *Smart Construction and Sustainable Cities,* 1(1). https://doi.org/10.1007/s44268-023-00016-0

OpenStreetMap. (n.d.). OpenStreetMap. https://www.openstreetmap.org

Palucha, S. (2024). *Understanding OpenAI's CLIP model* - Szymon Palucha - Medium. Medium. https://medium.com/@paluchasz/understanding-openais-clip-model-6b52bade3fa3

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021, February 26). *Learning transferable visual models from natural language supervision.* arXiv.org. https://arxiv.org/abs/2103.00020