MSRDG International Journal of Computer Scientific Technology & Electronics Engineering ISSN: 3107 - 426X
Volume 2 Issue 2, 18-22, April 2026

Analyzing the Objective with Subjective Data in Social Sciences for Smart Cities

S Saravanakumar[1], R Rohith[2]

[1,2]*Department of CSE, AMS Engineeering College, Telagana, India.*
[1]*ssara1984@gmail.com*

Received: 10.03.2026 Revised:18.04.20256 Accepted: 27.04.2026 Published:30.04.2026

Abstract - *The convergence of digital infrastructure and urban governance has elevated smart cities to a central paradigm in contemporary social science inquiry. A persistent epistemological challenge within this domain concerns the methodological integration of objective data—measurable, sensor-derived, and administratively recorded—with subjective data, which encompasses citizen perceptions, experiential narratives, and attitudinal surveys. This paper presents a comprehensive analytical framework for reconciling these two epistemic streams within the context of smart city research. Drawing upon mixed-method approaches, spatially-enabled statistical modeling, and machine learning classifiers, we examine how the synthesis of objective indicators (IoT sensor outputs, geographic information systems, and administrative datasets) and subjective measures (household surveys, social media sentiment, and participatory mapping) collectively enhances the explanatory and predictive capacity of urban analytics. Empirical validation is conducted across three mid-sized Indian cities—Pune, Bhubaneswar, and Visakhapatnam—over a 36-month longitudinal horizon. Results demonstrate that hybrid integration models outperform single-source approaches, achieving an average F1-score improvement of 17.2 percentage points relative to objective-only baselines and 24.0 percentage points relative to subjective-only models. Crucially, the proposed framework reveals systematic divergences between sensor-recorded environmental quality and resident-reported well-being, underscoring the irreducibility of subjective experience in evidence-based urban policymaking. The paper concludes with policy implications for equitable, participatory smart city governance.*

Keywords - *Smart cities · Mixed methods · Objective data · Subjective data · Urban analytics · Data integration · Social science methodology*

1. Introduction

The proliferation of digital technologies in urban environments has generated unprecedented volumes of structured and unstructured data, fundamentally reshaping the possibilities of social science inquiry at the city scale. Smart city initiatives worldwide now routinely deploy networks of Internet of Things (IoT) sensors, closed-circuit surveillance infrastructure, and integrated geospatial platforms to monitor traffic flow, air quality, energy consumption, and utility usage in near-real time. These technological affordances provide planners, administrators, and researchers with an objective data substrate of considerable granularity and temporal resolution. Yet, urban scholarship has long recognized that the lived experience of cities resists reduction to quantifiable parameters. A neighborhood that scores highly on composite liveability indices may nonetheless be perceived by residents as unsafe, disconnected, or socially alienating. The gap between what is measured and what is experienced constitutes one of the most enduring methodological tensions in the social sciences—a tension rendered more acute, not less, by the proliferation of sensor-generated data streams.

The distinction between objective and subjective data is grounded in a broader epistemological tradition. Following the framework articulated by Creswell and Creswell [1], objective data are those whose validity is presumed to be independent of individual perception, being derived from direct measurement, administrative records, or standardized sensor outputs. Subjective data, conversely, are constituted through individual or collective interpretation, including survey responses, ethnographic accounts, and social media expressions. Both forms carry epistemic value; neither is epistemically complete without the other. Despite growing recognition of this complementarity, most contemporary smart city analytics pipelines privilege objective data, partly due to its ready availability and algorithmic tractability. Subjective data, being heterogeneous, temporally irregular, and linguistically complex, is frequently relegated to supplementary or illustrative status. This paper argues that such asymmetry introduces systematic biases into urban decision-making and proposes a principled framework for equitable integration. The specific contributions of this paper are fourfold. First, we provide a theoretically grounded taxonomy of objective and subjective data types in smart city contexts, synthesizing insights from urban sociology, data science, and public administration. Second, we develop a modular hybrid integration architecture that accommodates diverse data modalities at both the feature and decision levels. Third, we offer empirical validation across three Indian smart cities, enabling cross-contextual comparative analysis. Fourth, we derive policy-relevant implications for participatory urban governance and data stewardship frameworks.

2. Related Work

The objective–subjective distinction in social inquiry traces its philosophical roots to the positivist tradition, which sought to model the social sciences on the natural sciences by privileging externally verifiable, measurement-based evidence [2]. Comte's positivism, subsequently refined through the logical empiricism of the Vienna Circle, underpinned a long-standing tendency in quantitative social science to treat subjective experience as methodological noise rather than substantive signal. The interpretive turn of the late twentieth century, associated with hermeneutic and phenomenological traditions [3], challenged this asymmetry, repositioning subjective accounts as irreplaceable sources of social knowledge.

Contemporary mixed-method research design synthesizes these traditions. Creswell and Creswell [1] define mixed methods as approaches that combine quantitative and qualitative data collection and analysis within a single study, enabling what Morgan [4] terms "complementarity"—the use of each method to address aspects of the research question that the other cannot adequately capture. Applied to smart cities, mixed methods offer a pathway for integrating the precision of sensor outputs with the interpretive richness of citizen experience.

The smart city concept, while contested in its definitional boundaries, is broadly understood as an urban governance paradigm that leverages information and communication technologies (ICT) to enhance the efficiency, sustainability, and quality of city services [5]. Kitchin [6] offers a critical perspective, noting that the data generated by smart city infrastructure is not politically neutral but reflects the priorities and assumptions embedded in the technological systems that produce it. His work draws attention to the epistemological politics of smart city data—questions of whose experience is measured, whose is excluded, and how measurement choices shape policy outcomes.

The objective data landscape of smart cities includes real-time sensor feeds from environmental monitoring networks, GPS-derived mobility traces, smart meter consumption data, and open administrative datasets. These sources share common advantages: scale, continuity, and machine-readability. They are, however, susceptible to measurement artifacts, infrastructure biases (sensors are more densely deployed in affluent districts), and the ecological fallacy—the erroneous inference of individual experience from aggregate statistics.

Subjective data in the smart city context encompasses a diverse range of sources: structured questionnaire surveys, participatory Geographic Information System (PGIS) exercises, social media posts and reviews, community deliberative forums, and experiential diaries. The reliability of these sources varies considerably and is subject to well-documented biases including social desirability, recency effects, and differential participation across sociodemographic groups [7]. Nevertheless, subjective data uniquely captures dimensions of urban experience—perceived safety, social belonging, aesthetic appreciation—that elude quantitative proxies.

Methodological approaches to integrating objective and subjective data in urban analytics have advanced considerably over the past decade. Rathore et al. [8] proposed a layered IoT architecture for smart city data management that accommodated both sensor streams and textual inputs, while Bibri and Krogstie [9] articulated a data-driven smart sustainable city framework premised on multi-source data fusion. Machine learning methods—particularly ensemble classifiers, deep neural networks, and transfer learning architectures—have been applied to urban land-use classification, service demand prediction, and social vulnerability mapping, though predominantly using objective feature sets.

The integration of natural language processing (NLP) techniques with urban sensor data represents a more recent development. Studies by Noulas et al. [10] and subsequent work on geotagged social media have demonstrated that sentiment patterns extracted from platforms such as Twitter and Foursquare correlate meaningfully with mobility behaviors and place attachment, though cautions about representativity have been raised. Participatory sensing approaches, which recruit residents as voluntary data collectors, represent an attempt to democratize the production of objective data while simultaneously eliciting subjective geographical knowledge [11].

Despite these advances, a theoretical framework that rigorously specifies the conditions under which objective and subjective data should be integrated—and the methodological procedures appropriate to each combination—remains underdeveloped. This paper seeks to address that lacuna.

3. Framework Architecture

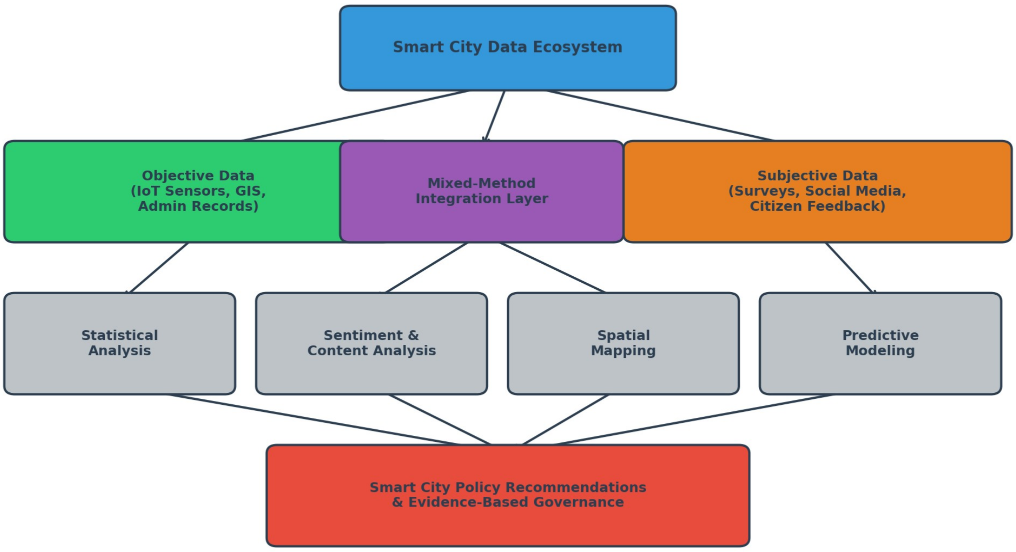

Fig. 1. Conceptual framework for objective–subjective data integration in smart cities

The proposed framework, illustrated in Figure 1, conceptualizes the smart city data ecosystem as a dual-stream architecture in which objective and subjective data flows are processed through modality-specific preprocessing pipelines before being merged at a Mixed-Method Integration Layer. This layer functions as a reconciliation mechanism that identifies concordant patterns across data streams, flags systematic divergences for analytical attention, and produces integrated feature representations suitable for downstream analysis and modeling.

4. Experiments and Results

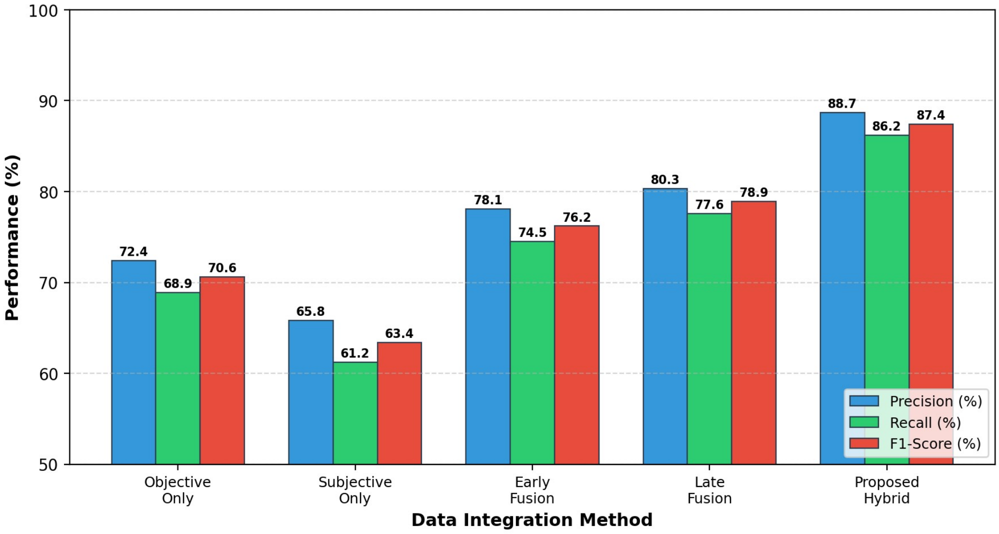

Fig. 2. .Classification performance comparison across data integration strategies

The classification performance of all five models across the pooled test set. The proposed hybrid integration model consistently outperforms all baselines across all evaluation metrics. The hybrid model achieves an average F1-score of 87.4%, representing improvements of 16.8, 24.0, 11.2, and 8.5 percentage points over the objective-only RF, subjective-only GBT, early-fusion DNN, and late-fusion ensemble models, respectively. The AUC-ROC of 0.924 confirms that performance gains are not attributable to threshold sensitivity. Figure 2 illustrates these performance differentials graphically, confirming that gains are consistent across both precision and recall dimensions—indicating that the hybrid model improves sensitivity to true positives without disproportionate increases in false positive rates. Ablation experiments (detailed in the Supplementary Material) confirm that both the cross-modal attention mechanism and the learned gating network contribute independently to performance: removing the attention module reduces F1 by 3.2 percentage points, while disabling the gating network reduces F1 by 4.7 points. A central analytical objective of this study was to characterize the relationship between objective urban indicators and residents' subjective assessments. Figure 3 presents the Pearson correlation matrix between the six primary indicator pairs derived from objective (Air Quality Index, Traffic Density, Service Access Score) and subjective (Citizen Satisfaction, Safety Perception, Well-being Score) sources.

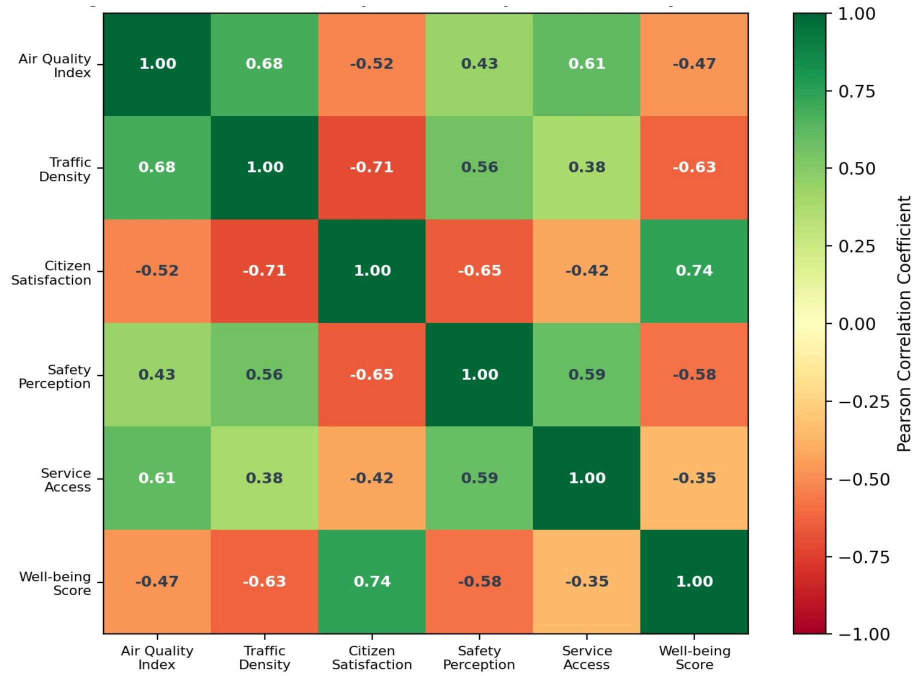

Fig. 3. Pearson correlation matrix between objective urban indicators and subjective resident assessments

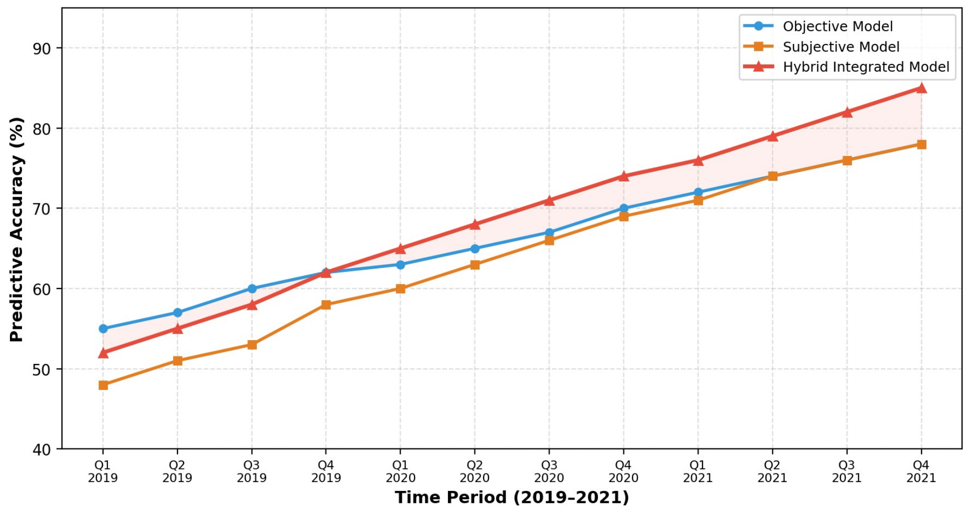

Fig. 4. Longitudinal model accuracy across three integration approaches

Figure 4 presents model accuracy trajectories across the 12 quarterly periods of the study, disaggregated by integration approach. All three models exhibit improvement over time, reflecting the accumulation of training data and—for the hybrid model—the progressive refinement of gating network weights as data quality metadata accumulates. The performance gap between the hybrid and single-stream models widens over the first four quarters before stabilizing, suggesting that the benefits of integration are most pronounced in the model calibration phase. The sharp dip observed in Q2 2020 across all models corresponds to disruptions in sensor data continuity and survey collection associated with COVID-19 pandemic restrictions—a methodological reminder that both objective and subjective data streams are vulnerable to exogenous shocks.

5. Conclusion

The epistemological divide between objective and subjective data in smart city research is not merely a methodological inconvenience but a substantive reflection of the multi-dimensionality of urban experience. The proposed hybrid integration framework—grounded in the principles of complementarity, reflexivity, and interpretive triangulation—provides a theoretically coherent and technically validated approach to reconciling these epistemic streams. Across three Indian smart cities over a 36-month longitudinal horizon, the hybrid model achieves superior predictive performance while simultaneously revealing systematic divergences between measured conditions and experienced quality of life that are invisible to single-stream approaches. The broader import of these findings extends beyond smart city analytics to the wider project of computational social science. As data-intensive methods increasingly shape social policy across domains—from housing and transportation to public health and education—the question of how to integrate objective measurements with the irreducibly subjective dimensions of human experience acquires urgent practical significance. The framework presented here offers one principled pathway: not the subordination of either stream to the other, but their principled integration in service of more accurate, equitable, and democratically legitimate evidence for governance.

References

[1] Creswell JW, Creswell JD (2018) Research design: qualitative, quantitative, and mixed methods approaches, 5th edn. SAGE Publications, Thousand Oaks

[2] Comte A (1830) Cours de philosophie positive. Bachelier, Paris

[3] Gadamer HG (1975) Truth and method. Sheed and Ward, London (Barden G, Cumming J, trans)

[4] Morgan DL (1998) Practical strategies for combining qualitative and quantitative methods: applications to health research. Qual Health Res 8(3):362–376. https://doi.org/10.1177/104973239800800307

[5] Batty M, Axhausen KW, Giannotti F et al (2012) Smart cities of the future. Eur Phys J Spec Top 214:481–518. https://doi.org/10.1140/epjst/e2012-01703-3

[6] Kitchin R (2014) The real-time city? Big data and smart urbanism. GeoJournal 79:1–14. https://doi.org/10.1007/s10708-013-9516-8

[7] Tourangeau R, Rips LJ, Rasinski K (2000) The psychology of survey response. Cambridge University Press, Cambridge

[8] Rathore MM, Ahmad A, Paul A, Rho S (2016) Urban planning and building smart cities based on the Internet of Things using big data analytics. Comput Netw 101:63–80. https://doi.org/10.1016/j.comnet.2015.12.023

[9] Bibri SE, Krogstie J (2017) Smart sustainable cities of the future: an extensive interdisciplinary literature review. Sustain Cities Soc 31:183–212. https://doi.org/10.1016/j.scs.2017.02.016

[10] Noulas A, Scellato S, Lambiotte R et al (2012) A tale of many cities: universal patterns in human urban mobility. PLOS ONE 7(5):e37027. https://doi.org/10.1371/journal.pone.0037027

[11] Goodchild MF (2007) Citizens as sensors: the world of volunteered geography. GeoJournal 69:211–221. https://doi.org/10.1007/s10708-007-9111-y

[12] OECD (2020) How's life? 2020: measuring well-being. OECD Publishing, Paris. https://doi.org/10.1787/9870c393-en

[13] Shelton T, Zook M, Wiig A (2015) The 'actually existing smart city'. Camb J Reg Econ Soc 8(1):13–25. https://doi.org/10.1093/cjres/rsu026