Explainable Machine Learning for Mapping Informal Urban Centers

J.T. M. S. da Silva

Relatório Técnico - IC-PFG-25-53
Projeto Final de Graduação
2025 - Dezembro

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE COMPUTAÇÃO

Explainable Machine Learning for Mapping Informal Urban Centers

João Theophilo Morais Santos da Silva *

12/2025

Abstract

Informal urban centers represent a persistent challenge for urban planning in Brazil due to their heterogeneity, structural precariousness, and the scarcity of standardized, high-resolution data. In this study, we propose an interpretable machine learning approach to classify informal urban centers across six Brazilian hubs using an Explainable Boosting Machine (EBM). A spatially structured dataset was built by intersecting census-based socioeconomic indicators with the official mapping of informal urban centers, producing a binary classification framework. To evaluate the robustness of the model under spatial heterogeneity, we implemented two training strategies: a weighted model incorporating a city-propensity adjustment and a standard unweighted model, each tested on a held-out hub. Results for the Porto Alegre test hub show that the weighted model achieves an AUC of 0.826, outperforming the standard model at 0.818. Feature importance rankings indicate consistent relevance of demographic density, income indicators, and local infrastructure. The findings highlight the potential of interpretable models to support policy-oriented spatial classification tasks, providing both predictive accuracy and transparent insights that can guide targeted urban interventions.

1 Introduction

Informal urban centers represent one of the most persistent challenges in contemporary urban planning, particularly in countries marked by unequal development and heterogeneous urban trajectories such as Brazil. These areas—commonly characterized by irregular land occupation, insufficient access to public services, and heightened social vulnerability—demand monitoring strategies capable of supporting targeted policy interventions. Traditional approaches for identifying informal centers often rely on labor-intensive surveys, incomplete administrative records, or remote sensing data that may fail to capture the socioeconomic nuances that define precarious urban conditions.

Recent advances in machine learning have opened new possibilities for automating the detection and characterization of informal urban centers using heterogeneous data sources. However, challenges remain regarding the generalization of predictive models across regions

*Institute of Computing, University of Campinas, 13081-970 Campinas, SP

with distinct historical, demographic, and infrastructural patterns. Distribution shifts between cities—or "hubs"—can lead to biased models whose performance degrades when evaluated on unseen urban contexts. This issue is particularly relevant in Brazil, where the socioeconomic and spatial configurations of informal centers vary widely from one region to another.

In response to this challenge, this study proposes a methodology for classifying informal urban centers using census-based socioeconomic indicators aggregated over uniform spatial grid cells. The approach combines interpretable machine learning—through the use of the Explainable Boosting Machine (EBM)—with a hub-level evaluation scheme designed to assess the robustness of predictions across heterogeneous cities. Two modeling strategies are compared: (i) a propensity-weighted model that adjusts the training process based on the similarity of samples to the target city, and (ii) a standard unweighted model trained without sample reweighting. The goal is to evaluate whether accounting for inter-city distribution differences can improve predictive performance and interpretability when identifying informal centers.

The results of this work contribute to ongoing efforts to integrate machine learning into urban policy analysis, offering insights into how demographic, infrastructural, and spatial indicators interact in the identification of informal urban centers, and how model generalization can be strengthened when transferring knowledge across distinct urban realities.

2 Problem Definition and Background

2.1 Definition of Informal Urban Center

Informal urban centers in Brazil are areas characterized by irregular land occupation, precarious housing conditions, and limited access to essential public services. A key challenge in defining these settlements is their high degree of heterogeneity, which stems from distinct regional contexts and historical patterns of urbanization. As a result, informal areas can vary widely even within the same city. These centers include favelas, informal settlements and recently formed precarious districts. Common issues include restricted access to infrastructure, limited mobility, and heightened exposure to social, health, and safety vulnerabilities. Accurate classification of informal urban centers is essential for identifying these areas across cities with diverse spatial structures. Such classification supports more effective urban planning and helps guide targeted interventions to improve living conditions in underserved communities.

2.2 Dataset

The dataset used in this study combines population census information aggregated into 10,000 m^2 geographical cells for six Brazilian urban hubs[1], together with an additional spatial dataset containing labels for informal urban centers derived from qualitative field research[2]. The census data, obtained from a private dataset due to data protection requirements, was aggregated to the same cell structure for each hub and includes economic, infrastructural, and geographical variables. Labels for informal urban centers were obtained from previous studies, which characterize different types of informal settlements across the hubs of Porto Alegre, Marabá, Brasília, Juazeiro do Norte, Recife, and Belo Horizonte. These hubs typically consist of a primary central city surrounded by satellite municipalities or districts. A key attribute in this dataset is column *V5a*, used as the target variable in this work. This column identifies the type of informal settlement associated with each geographic unit, with the following possible categories:

1. favela or spontaneous occupation;
2. irregular or clandestine subdivision;
3. housing Complex;
4. district or town;
5. municipal Headquarters;
6. occupation by traditional populations;
7. other (specify in the column alongside);
8. mixed (without predominance of any type)

These labels were assigned to the census cells through a geographic left join, generating the target variable used in this study. While the structure of the dataset allows both a binary formulation (informal vs. non-informal) and a multi-class formulation (type of informal settlement), this work adopts the binary approach, treating all categories of informal settlements as a single positive class and all other areas as negative.

2.3 Related Works

Recent advances in data availability and machine learning have strengthened the connection between data science applications and urban-planning needs. Most existing work on informal-settlement detection focuses on remote sensing imagery, often using image-based classification methods. Early approaches relied on object-oriented analysis to detect structural and textural patterns in urban landscapes [5]. With the rise of deep learning, Convolutional Neural Networks (CNNs) have been applied to high-resolution imagery, achieving strong performance when trained and tested within the same urban context [6]. More recent multimodal approaches integrate imagery with auxiliary datasets—such as population-density time series or mobility and street-network indicators—to improve detection and boundary prediction of informal areas [4, 7].

Beyond image-based approaches, other studies highlight the importance of socioeconomic and territorial characteristics in understanding informality. An extensive survey across 157 Brazilian municipalities used an ensemble k-modes clustering method to identify four distinct typologies of informal settlements, illustrating the heterogeneity of these territories and the need for context-aware models [3]. Furthermore, the increasing interest in interpretable machine learning has motivated the use of models that balance predictive accuracy with transparency. In particular, Generalized Additive Models with pairwise interactions (GA2M) [8] have demonstrated that intelligible models can approach the performance of high-complexity black-box models. Together, these works demonstrate the growing effort to combine multiple data sources, interpretable modeling, and computational methods to support the identification and characterization of informal urban centers.

3 Proposed Methodology

The evaluation is conducted by holding out an entire hub (Porto Alegre) as the test set to simulate real-world generalization to cities not present during model training.

3.1 Dataset Preparation

The initial stage consisted of preparing the spatial datasets used to generate the feature matrix for modeling. Two primary geospatial layers were utilized: (i) a uniform grid of approximately one-thousand–square-meter cells for six Brazilian urban hubs, and (ii) a polygonal dataset representing mapped Informal Urban Centers (IUCs), each containing a categorical variable ("V5a") describing the type of informal settlement. Both datasets were loaded using GeoPandas, with careful alignment of coordinate reference systems to ensure overlay compatibility.

To associate each grid cell with potential IUC information, a spatial overlay operation was performed. First, the intersection between the IUC polygons and the grid geometries was computed, retaining only the intersecting portions and extracting the IUC label associated with each intersected cell. Because some IUC polygons intersect multiple cells, which could lead to duplicated identifiers, only one unique IUC label was sampled per grid cell using a groupwise sampling strategy. After removing geometries and retaining only the semantic attributes, the full dataset was reconstructed by merging the intersecting labeled cells with the remaining non-intersecting cells, producing a unified table in which every grid cell is represented once.

The final product of the preparation stage was an aggregated dataset containing socioeconomic, infrastructural and demographic variables for each grid cell across the six hubs, along with a target variable indicating the presence or absence of informal settlements.

Although the underlying dataset contains multiple categories of informal settlements (e.g., favelas, housing complexes, irregular subdivisions), this study adopts a binary classification perspective, distinguishing only whether a grid cell corresponds to an Informal Urban Center (target = 1) or not (target = 0). All categories in the "V5a" attribute were consolidated into a single positive class, reflecting the practical goal of detecting informal settlements broadly rather than differentiating among specific types. Grid cells without intersection with any IUC polygon were labeled as negative cases.

This binary formulation simplifies the modeling objective and aligns with the research goal of evaluating the model's ability to generalize across heterogeneous urban environments. To assess how well the model generalizes to unseen urban environments, the training set was defined by excluding the target city entirely. All remaining hubs constituted the pool from which the model learned socioeconomic and infrastructural patterns associated with informal settlements. The target city—e.g., Porto Alegre—was held out as a pure test set.

3.2 Propensity-Based Sample Reweighting

A significant methodological challenge arises from the heterogeneity between hubs. Each city exhibits distinct socioeconomic distributions, spatial signatures, and sampling characteristics. Directly training a model by pooling all remaining hubs risks overweighting the most populous cities and underrepresenting those with fewer observations. To mitigate this imbalance and promote fairer cross-city generalization, a propensity-score weighting scheme was implemented.

For each city selected as the evaluation target, a binary indicator variable was created to specify whether each observation originated from that city. Using this indicator as the response variable, a Random Forest classifier was trained to estimate the probability that any given geographical cell belonged to the target city based solely on its feature profile. These estimated probabilities were then interpreted as propensity scores and assigned as sample weights during model training. The incorporation of these weights serves a dual purpose: it mitigates the risk of the classifier overfitting to cities with larger sample sizes and rebalances the influence of the training data so that cells whose characteristics more closely resemble those of the target city contribute more strongly to parameter estimation. In this way, the weighted training process ensures that the learned model more faithfully approximates the distributional structure of the target city, enabling a more realistic evaluation of cross-city generalization.

This reweighting method is particularly relevant when evaluating generalization performance in a leave-one-city-out setup, where the target city is entirely excluded from the training data.

3.3 Model Training Using Explainable Boosting Machines (EBM)

The predictive model adopted in this study is the Explainable Boosting Machine (EBM), a modern glass-box learning algorithm that extends the framework of Generalized Additive Models (GAMs) by incorporating a small number of pairwise interaction terms [8]. Unlike traditional black-box approaches such as deep neural networks or ensemble tree models, EBMs are explicitly designed to remain transparent while achieving competitive predictive performance. This makes them particularly suitable for studies involving urban inequality, territorial vulnerability, and public policy, where model interpretability is not only desirable but often required for accountability and deployability.

EBMs achieve interpretability through their structure: the predicted outcome is modeled as the sum of independently learned shape functions, each describing the marginal effect of a single feature, plus optional two-feature interaction terms chosen only when they significantly improve model quality. Because each component of an EBM is either one-dimensional (for single features) or two-dimensional (for interactions), the full model can be visualized and inspected by domain specialists. These visual explanations take the form of intuitive plots showing how changes in a given covariate affect the predicted probability of a cell being classified as an informal urban center. Such transparency is especially valuable in the Brazilian context, where informal-settlement classification involves complex socioeconomic, infrastructural, and spatial indicators that must be interpretable by planners, policymakers, and local researchers.

The appropriateness of EBMs for this dataset stems from both methodological and substantive considerations. First, the high-dimensional yet structured feature space—composed of demographic densities, economic variables, mobility indicators, and infrastructural measures—aligns well with EBM's ability to learn smooth, non-linear effects without overfitting. Second, the heterogeneity across urban hubs requires a model capable of capturing hub-specific relationships while remaining globally interpretable, allowing analysts to compare how features behave differently across regions. Third, because part of the classification task concerns understanding why certain areas are likely to be informal, not merely predicting whether they are, interpretability is a methodological requirement rather than a convenience.

Training the EBM involved using the feature matrix derived from all hubs except the one withheld for testing, along with the propensity-score sample weights described in Section 3.2. These weights guide the model to account for distributional differences between hubs, improving its ability to generalize to the held-out city.

3.4 Hyperparameter Optimization via Grid Search

To identify the optimal configuration of the EBM model, a multi-dimensional grid search was conducted using 3-fold cross-validation, weighted by the propensity scores. The search space explored parameters including learning rate, number of bins, number of leaves, maximum boosting rounds, and interaction depth. The objective function for selection was the Area Under the ROC Curve (AUC), which is appropriate for binary detection tasks with imbalanced classes.

The best-performing hyperparameters—defined by the highest cross-validated AUC—were then used to instantiate a final EBM classifier trained on the full training dataset (excluding the target city).

3.5 Model Evaluation on the Held-Out City

The final stage consisted of evaluating the tuned model on the target city. Predictions were computed using the `predict_proba` function to obtain continuous estimated probabilities of belonging to an informal settlement. The principal performance metric was the AUC, which summarizes the model's ability to rank positive cases above negative ones regardless of classification threshold.

By testing on a completely unseen city, this evaluation reflects the model's robustness to geographic distribution shift and its capacity to generalize across urban environments with distinct development patterns, socioeconomic compositions, and spatial structures.

3.6 Baseline Modeling Pipeline Without Sample Weights

In addition to the propensity-weighted approach, a second modeling pipeline was developed—an unweighted baseline algorithm—to isolate the impact of sample weighting on cross-hub generalization. This additional algorithm follows the same structure: splitting the dataset into training hubs and a held-out test hub, performing hyperparameter tuning via grid search, training an EBM model with the best parameters, and finally computing the ROC-AUC score on the test hub. By removing the influence of the weighting scheme, this secondary pipeline serves as a control model that illustrates how much predictive improvement can be attributed specifically to the weighting strategy.

This baseline algorithm was fully implemented and executed, producing a trained EBM model for the same test hub used in the weighted approach. Both models—weighted and unweighted—were saved, evaluated, and inspected using global explainability visualizations provided by the interpret package.

4 Results

This section presents the empirical findings obtained from the two modeling strategies developed in this study: (i) the weighted propensity-based model, in which samples were reweighted according to their estimated probability of belonging to the target hub, and (ii) the standard model, trained without the use of sample weights. In both cases, the Explainable Boosting Machine (EBM) served as the classifier, and the city of Porto Alegre was held out as the test hub to assess the generalization capability of the models under cross-hub transfer.

4.1 Model Performance

The weighted model achieved an Area Under the ROC Curve (AUC) of 0.826 when predicting informal urban centers in the test hub. This value indicates a strong discriminatory capacity, especially considering the cross-hub distribution shift present in the dataset. The inclusion of sample weights appears to have improved the model's robustness in handling inter-hub heterogeneity, yielding better separation between informal and non-informal urban cells during evaluation.

In contrast, the standard model, trained without sample-weighting and relying solely on the training hub's feature distributions, achieved a slightly lower AUC of 0.818 on the same test split. Although the performance remains high, the decrease relative to the weighted model suggests that the reweighting strategy provided a small but meaningful improvement when confronting distributional differences between hubs.

4.2 Feature Importance Analysis

Beyond overall predictive performance, the Explainable Boosting Machine provides interpretable importance rankings for each feature. These rankings offer insight into which socioeconomic and infrastructural indicators contribute most to identifying informal urban centers.

For the weighted model, the five most influential features were, in descending order of importance:

1. NDenPop – Population density per km^2
2. DomNBanHab – Average number of bathrooms per resident
3. RenRespMedia – Average income of household head
4. NDemDom – Household density per km^2
5. Vias50m – Percentage of the analysis unit's area located within 50 meters of the motor-vehicle road

For the standard model, the relative importance of features showed a slightly different ordering:

1. DomNBanHab
2. Vias50m
3. RenRespMedia
4. NDenPop
5. NResp30 – Percentage of household heads under 30 years old

This divergence between the models indicates that the weighting procedure not only improves performance but also affects how the model interprets the underlying structure of the data.

5 Discussion

The comparative evaluation of the weighted and standard modeling strategies provides important insights into the challenges associated with detecting informal urban centers across heterogeneous Brazilian urban hubs. The use of sample weights based on a propensity model—designed to mitigate the distribution shift between the target city (Porto Alegre) and the remaining hubs—yielded a marginal improvement in predictive performance, with an AUC of 0.826 compared to 0.818 for the standard EBM model. Although numerically modest, this increase suggests that weighting observations to account for hub-level disparities may contribute to a more robust generalization when applying the model to an unseen geographic context.

The analysis of feature importance further illustrates differences in how each model internally represents the structure of the classification task. In the weighted EBM, variables related to population density (NDenPop) and the number of bathrooms per resident (DomNBanHab) emerged as the most influential predictors, followed by income (RenRespMedia), household density (NDemDom), and proximity to road networks (Vias50m). This ordering is consistent with qualitative descriptions of informal settlements: densely populated areas with lower median income levels typically align with patterns of precarious urban occupation.

6 Conclusion

This work investigated the use of Explainable Boosting Machines for the binary classification of Informal Urban Centers across multiple Brazilian urban hubs. By comparing a standard training strategy with a weighted approach designed to mitigate inter-hub distribution differences, the results indicate that both models achieved strong predictive performance, with a slight advantage for the weighted version (AUC = 0.826 versus 0.818). Beyond performance, the analysis of feature importance revealed consistent patterns across models, highlighting the relevance of population density, household bathroom availability, income characteristics, and road-network proximity in the identification of informal areas. These findings demonstrate not only the feasibility of applying transparent machine-learning models to complex urban-socioeconomic data, but also the value of weighting strategies in improving generalization to previously unseen hubs.

Despite these positive results, an important limitation of the present study is that model evaluation was carried out using a single held-out hub—Porto Alegre. While this choice provides an initial perspective on cross-hub generalization, it does not capture the full variability of socioeconomic structures, urban morphologies, and patterns of informality present across Brazil's diverse urban regions. Evaluating the model sequentially on all hubs, or employing a leave-one-hub-out validation framework, would provide a more comprehensive assessment of robustness and distributional sensitivity. Furthermore, hub-specific idiosyncrasies may influence both the learned feature effects and the effectiveness of the weighting strategy, suggesting that broader evaluation is necessary to confirm the general applicability of the method.

Future work should therefore focus on expanding the evaluation to all available hubs and incorporating more systematic domain-adaptation techniques to enhance prediction performance. These steps would strengthen the reliability of EBM-based frameworks for supporting urban-planning interventions and deepen the understanding of informal-urban dynamics across heterogeneous territories.

References

[1] Data from multiple sources integrated into a 100×100 m grid, compatible with the ibge statistical grid (2016), referring to occupied areas (human settlements) in the research hubs (belo horizonte, brasília, juazeiro do norte, marabá, porto alegre, and recife). https://drive.google.com/file/d/1OvAEn1gLM-5Ni0dGj8yMZcHTHF6T4OF/view?usp=sharing. Acessado em: 7 dez. 2025.

[2] Data: mesh of the areas mapped as informal urban cores in the research hubs (belo horizonte, brasília, juazeiro do norte, marabá, porto alegre, and recife).

https://drive.google.com/file/d/1LvAJT43XMiwPCWT$_IQAkYatEGO8cP_w0/view?usp=sharing.Acessadoem:7dez$.2025.

[3] L. F. B. da Cunha, F. da F. Feitosa, A. Paez, and C. M. D. de Pinho. Exploring the diversity of informal settlements in brazil: a clustering approach to qualitative field survey data. In *XXIII GEOINFO*, São José dos Campos, SP, Brazil, 2022.

[4] Runyu Fan, Jun Li, Wei Song, Wei Han, Jun Yan, and Le Wang. Urban informal settlements classification via a transformer-based spatial-temporal fusion network using multimodal remote sensing and time-series human activity data. *International Journal of Applied Earth Observation and Geoinformation*, 111:102831, 2022.

[5] P. Hofmann, J. Strobl, T. Blaschke, and H. Kux. Detecting informal settlements from quickbird data in rio de janeiro using an object based approach. In *Lecture Notes in Geoinformation and Cartography*, pages 531–553. Springer Berlin Heidelberg, 2007.

[6] Nicholaus Mboga, Claudio Persello, J. R. Bergado, and A. Stein. Detection of informal settlements from vhr images using convolutional neural networks. *Remote Sensing*, 9(11):1106, 2017.

[7] Chuanji Shi, Yifan Zhang, Jiaming Wang, Ximin Guo, and Qunying Zhu. Multimodal urban areas of interest generation via remote sensing imagery and geographical prior. *arXiv preprint arXiv:2401.06550*, 2024.

[8] Johannes Gehrke Yin Lou, Rich Caruana and Giles Hooker. Accurate intelligible models with pairwise interactions.