

OptiGeoRisk: A Spatially Localized and Optimized Geo-XGBoost Framework for Credit Default Risk Classification

Claher Prastian*, Alexander Agung Santoso Gunawan

Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480

Abstract—Credit default risk prediction is important for financial institutions in geographically diverse, collateral-based lending, where early identification of risky borrowers supports preventive monitoring. Most credit scoring models treat borrower records as independent observations and rarely capture local spatial heterogeneity. This study proposes OptiGeoRisk, a spatially localized Geo-XGBoost framework for imbalanced credit default classification, using 76,742 loan transactions from West Java, Indonesia. OptiGeoRisk treats geographic coordinates as a local learning mechanism, combining haversine-based nearest-neighbor selection, adaptive neighborhood size and kernel bandwidth, and global-local probability fusion, and is evaluated against XGBoost Non-Spatial and XGBoost Spatial Feature baselines using PR-AUC, categorical and Top-K accuracy, default recall, precision, and false-positive volume. XGBoost Spatial Feature obtains the best PR-AUC, while OptiGeoRisk achieves the highest default recall, trading global ranking quality for localized early-warning sensitivity. Bootstrap resampling shows this recall advantage is statistically robust, while OptiGeoRisk's PR-AUC is statistically indistinguishable from XGBoost Non-Spatial and significantly lower than XGBoost Spatial Feature; refitting under 10 random seeds confirms this pattern is stable, and a cross-region check in DKI Jakarta shows the PR-AUC gain over a no-skill baseline holds outside West Java, though the recall advantage does not. A leakage-robust check across borrower-grouped, spatially blocked, and temporal splits shows the recall advantage holds under random and borrower-grouped partitioning but weakens when whole areas or future periods are held out, a limitation traced mainly to adaptive neighborhood selection and imbalance handling rather than the spatial kernel. These findings position OptiGeoRisk as a localized early-warning layer complementing a well-tuned global model.

Keywords—Credit default risk; Geo-XGBoost; spatial machine learning; class imbalance; early-warning system; financial risk management; sustainable economic growth

I. INTRODUCTION

Credit default risk prediction is increasingly important for financial institutions operating in geographically diverse markets, particularly in consumer loans, vehicle financing, pawnshops, and collateral-based lending where early identification of risky borrowers can support preventive collection strategies and portfolio monitoring. In Indonesia, financing and pawn-related activities operate under regulatory attention from the Financial Services Authority, which frames stronger risk governance, consumer protection, and responsible monitoring in this sector as part of its broader push for financial-sector stability and sustainable economic growth [1], [2].

Credit scoring has traditionally been formulated as a statistical classification problem, and machine learning has extended this field by improving prediction from structured borrower and loan attributes [3], [4], [5]. Yet credit default modeling remains difficult because default events are typically rare compared with non-default transactions. Under this class imbalance, a model can obtain high categorical accuracy by predicting the majority class while failing to identify customers who later default [6]. For that reason, ranking-oriented and minority-class metrics such as PR-AUC and default recall are more informative than accuracy alone [7], [8].

Most credit-risk models also assume that borrower records are independent observations governed by a single global prediction function, an assumption that is restrictive for collateral-based credit in West Java, where access to branches, local economic activity, road structure, settlement density, and surrounding facilities vary substantially across regions. When location is encoded only as a category or raw coordinate, the model can learn a regional identity without explicitly adapting the prediction function to local risk behavior. That gap motivates a spatial learning framework that treats geography as a mechanism of learning rather than merely another input attribute.

This study proposes OptiGeoRisk, a spatially localized and optimized Geo-XGBoost framework for binary credit default risk classification. The framework uses real-world collateral-based loan transaction data from West Java, enriches it with branch-distance, POI, OSM, and GeoPandas-based spatial treatment, and uses coordinates structurally for local-neighborhood learning. The expected benefit is not automatic credit rejection but a practical early-warning layer that helps financial institutions prioritize monitoring, verification, and collection follow-up.

The contributions are sixfold. First, the study adapts Geo-XGBoost from spatial regression into imbalanced binary default classification. Second, it removes administrative location names and raw latitude-longitude values from model predictors while preserving coordinates for local spatial learning. Third, it evaluates the trade-off between global ranking quality and local early-warning sensitivity using metrics that are appropriate for imbalanced default risk data, benchmarked against XGBoost baselines. Fourth, it checks whether these results hold up under bootstrap resampling, ten independently retrained seeds, and a cross-region comparison between West Java and DKI Jakarta run under an identical evaluation protocol. Fifth, it isolates the

*Corresponding author.

source of OptiGeoRisk's gains through leakage-robust validation across borrower-grouped, spatially blocked, and forward-chaining temporal splits, a systematic ablation of its adaptive-neighborhood, Gaussian-weighting, and fusion components, and hyperparameter-tuning budgets matched across baselines. Sixth, it reports the operational value and computational overhead of the framework at realistic review capacities, so that its deployment cost can be weighed against its early-warning benefit.

II. RELATED STUDIES

A. Credit Risk and Loan Classification

Credit scoring research provides the foundation for modeling credit default as a classification task. Classical reviews highlight the role of statistical classification in consumer credit scoring [3], while later benchmarking and systematic-review studies show that machine learning can improve predictive performance on structured credit data [4], [5]. Recent loan approval studies further demonstrate the use of ensemble learning and tree-based methods for assessing repayment feasibility and borrower creditworthiness [9], [10], [11]. More recent comprehensive reviews confirm that ensemble and gradient-boosting methods remain the strongest-performing family for structured credit data, while identifying explainability, calibration, and class-imbalance handling as the field's most persistent open challenges [12], [13].

Although these studies show that machine learning is useful for credit decision support, they generally emphasize global tabular learning. Spatial context, when available, is often treated as an ordinary input variable rather than a source of local variation in the learning process. This leaves a gap for a method that can use location to adapt prediction behavior around nearby observations.

B. Imbalanced Classification and Evaluation

The dataset used in this study has a default-class rate of only 6.87%. This confirms that the modeling problem is severely imbalanced. Previous research shows that imbalanced class distributions can affect credit-scoring model behavior and comparison [6]. PR-AUC is therefore used as the main ranking metric because Precision-Recall analysis is more informative than ROC-based interpretation when the positive class is rare [7], [8].

The study also reports categorical accuracy, Top-K categorical accuracy, default recall, precision, and false-positive volume. Top-K categorical accuracy evaluates ranked risk lists: Top-K Default checks whether actual defaults appear among the highest predicted default probabilities, while Top-K Non-default checks whether non-default cases appear among the lowest predicted default probabilities. These ranked metrics fit operational early warning better than a fixed threshold, since financial institutions usually work from prioritized customer lists rather than a single probability cutoff.

C. XGBoost, Geo-XGBoost, and Spatial Credit Modeling

XGBoost is widely used for tabular prediction because gradient-boosted trees can model non-linear relationships and interactions efficiently [14].

However, standard XGBoost is a global model. It learns one prediction function for all observations, even when the relationship between risk factors and default varies geographically.

Class-imbalance treatments such as over-sampling, under-sampling, class weighting, and SMOTE can improve minority-class learning when applied only on training data [15], [16]. Spatial learning adds another layer by allowing relationships to vary over space. Geographically weighted regression introduced the idea of estimating local relationships rather than one global function [17], [18], and spatial machine learning extends this perspective to modern predictive modeling [19]. Geo-XGBoost applies this idea to gradient-boosted trees through geographically weighted local learning [20]. Recent work extends this direction along two complementary paths: adapting XGBoost's objective directly for severe class imbalance through a generalized extreme value link and modified focal loss [21], and combining tree-boosting explicitly with spatial structure, as in spatio-temporal Gaussian-process boosting for mortgage default [22], Mixed Geographically Weighted XGBoost [23], and explainable ensemble models for peer-to-peer lending default risk [24]. These studies motivate treating geography and class imbalance as structural learning problems rather than ordinary inputs, which is the direction pursued by OptiGeoRisk.

The methodological gap addressed in this study is that Geo-XGBoost was originally proposed for regression, while credit default risk is a binary and imbalanced classification problem. OptiGeoRisk modifies the idea by using logistic probability output, imbalanced evaluation, Optuna-based optimization, adaptive local neighborhoods, and a global-local probability fusion mechanism [25].

III. RESEARCH METHODOLOGY

Fig. 2 outlines the overall research workflow followed in this study, from data preparation through model development to the spatial diagnostics presented in Section IV. The diagram groups the pipeline into three linked stages: data collection and feature construction, spatial treatment and GIS diagnostics, and the OptiGeoRisk modeling architecture and evaluation itself, which combines a global classifier with localized, geographically weighted classifiers through an adaptive fusion step before the robustness checks and cross-region validation described in Section IV. The workflow begins with the study area and dataset construction, proceeds through feature engineering and the proposed local-global fusion approach, and concludes with the diagnostic evaluation of spatial prediction performance. Each stage is described in the subsections that follow.

A. Study Area and Dataset

The study uses 76,742 collateral-based loan transactions from West Java, Indonesia, split into train, validation, and test partitions with stratified sampling. The test set contains 15,349 transactions and 1,054 default cases, a positive rate of 6.87%; this original distribution is preserved in the validation and test sets to avoid optimistic evaluation. Fig. 1 shows the geographic distribution of these transactions and the observed default rate across West Java's subdistricts (kecamatan).

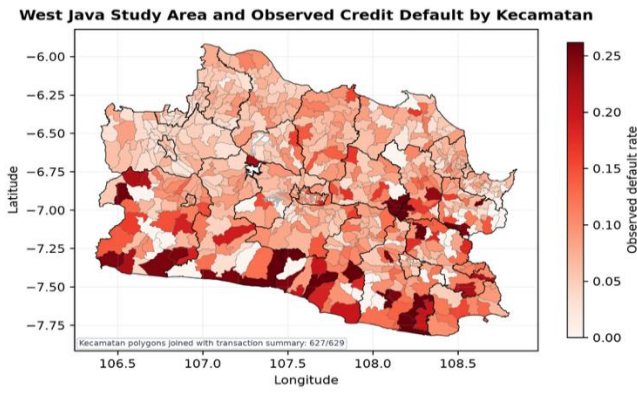


Fig. 1. West Java study area and subdistrict-level observed default distribution by Kecamatan boundary.

Spatial processing follows the style of GIS-based analytical studies. Transaction representative points were constructed from subdistrict latitude and longitude as a GeoDataFrame using CRS EPSG:4326, then reprojected when metric distance or density diagnostics were required with GeoPandas [26]. Administrative polygons were used for mapping and spatial diagnostics, not as ordinary model predictors. For detailed visualization, the kecamatan boundary layer was prepared by dissolving 5,973 village polygons into 630 kecamatan polygons.

B. Feature Engineering

The non-spatial feature group includes borrower, loan, collateral, vehicle, and temporal attributes. The spatial-feature group includes branch-distance features, POI features, and OSM-derived built-environment features. POI attributes were generated from Google Places categories within a 1,000-meter radius, including financial, healthcare, education, commerce, and market-related counts [27]. OSM features were also computed within a 1,000-meter radius, including building counts and roadway counts [28]. Coordinates are retained only for geospatial preprocessing and the local OptiGeoRisk learning mechanism.

This design reduces location-identity leakage and makes the comparison with global XGBoost more methodologically defensible. Table I reports descriptive statistics for these retained borrower, loan, collateral, and spatial-context variables.

TABLE I. DESCRIPTIVE STATISTICS OF SELECTED VARIABLES AND SPATIAL FEATURE SOURCES

Variable	Description	Mean	Std
target_macet	Default label: 1=macet, 0=lancar	0.069	0.253
loan_amount	Initial loan amount	1,961,235	927,332
collateral_value	Initial collateral value	6,054,794	2,574,454
loan_to_value	Loan-to-value	0.33	0.077
vehicle_age	Collateral age	5.533	3.054
borrower_age	Borrower age	35.041	10.854
month	Transaction month	6.824	3.298
day_of_week	Day of the week of the transaction	2.638	1.888
branch_distance	Distance from domicile to branch (km)	7.595	32.891

same_city_branch	Same city/regency as branch	0.866	0.34
poi_financial	Financial POI within 1,000 m	25.887	28.099
poi_healthcare	Healthcare POI within 1,000 m	16.526	14.615
poi_education	Education POI within 1,000 m	36.582	19.449
poi_market	Market POI within 1,000 m	7.015	7.168
osm_buildings	Building count within 1,000 m	3,498	3,318
osm_roads	Roadway count within 1,000 m	437.134	337.17
osm_build_density	Building density per km2	1,114	1,056

C. Proposed Method

OptiGeoRisk combines a global XGBoost classifier with localized, geographically weighted XGBoost models. The global component estimates P_{global} from the no-admin feature matrix, while the local component searches nearby training observations for each evaluation point using haversine distance, applies Gaussian kernel weighting, trains a local XGBoost classifier, and estimates P_{local} . The final score comes from an adaptive fusion of the two:

$$d_{ij} = \text{haversine}(g_i, g_j) \quad (1)$$

$$w_{ij} = \exp\left(-\frac{d_{ij}^2}{2b_i^2}\right) \quad (2)$$

$$P_{\text{final},i} = \alpha_i P_{\text{global},i} + (1 - \alpha_i) P_{\text{local},i} \quad (3)$$

In Eq. (1), d_{ij} is the geographic distance between the prediction location and the training location. In Eq. (2), b_i is the local bandwidth derived from the neighborhood setting and w_{ij} is the resulting Gaussian kernel weight. In Eq. (3), α_i controls the observation-level global-local probability fusion between $P_{\text{global},i}$ and $P_{\text{local},i}$. Together, Eq. (1)–(3) follow the GeoXGBoost principle of combining tabular predictors X , response y , coordinates G , bandwidth, adaptive kernel weighting, and local boosted-tree learning, but the objective is changed from regression to binary logistic classification. Optuna was used in the prior optimization stage to select the imbalance strategy, XGBoost parameters, neighborhood-size baseline, kernel-bandwidth baseline, and global-local fusion parameters [25]. In the main experiment, OptiGeoRisk uses random over-sampling for the local component, initializes the local search from the optimized neighborhood and bandwidth values, and then applies adaptive neighborhood diagnostics and observation-level fusion weights during validation and testing (Algorithm 1). This integrated research framework, from feature construction through the local-global fusion mechanism to the spatial diagnostics reported in Section IV, is illustrated in Fig. 2.

Algorithm 1: Pseudocode for OptiGeoRisk Credit Default Classification

Inputs: $D=\{X,y,G\}$, POI/OSM features, Optuna search space H

Output: Optimized spatial credit-risk classifier and prob_macet

Start:

Load and clean the transaction dataset

Construct loan, collateral, customer, time, branch, POI, OSM, and GeoPandas-derived features Remove administrative-name features (kab/kota, kecamatan) to prevent leakage; retain coordinates G as the spatial mechanism Split dataset into stratified train, validation, and test sets For each Optuna trial t do: - Select imbalance strategy r - Sample XGBoost and spatial parameters - Apply r to training data only - Train global XGBoost classifier For each validation coordinate g do:	Compute haversine distance to all training points Select k nearest neighbors adaptively Compute bandwidth b and Gaussian weights w Train weighted local XGBoost classifier Compute P_global and P_local Compute adaptive alpha and $P_{final} = \alpha \times P_{global} + (1 - \alpha) \times P_{local}$ Evaluate validation PR-AUC, recall, precision, and false-positive rate Select the best trial and threshold Evaluate final model on the test set End
--	--

Integrated Research Framework and OptiGeoRisk Modeling Architecture

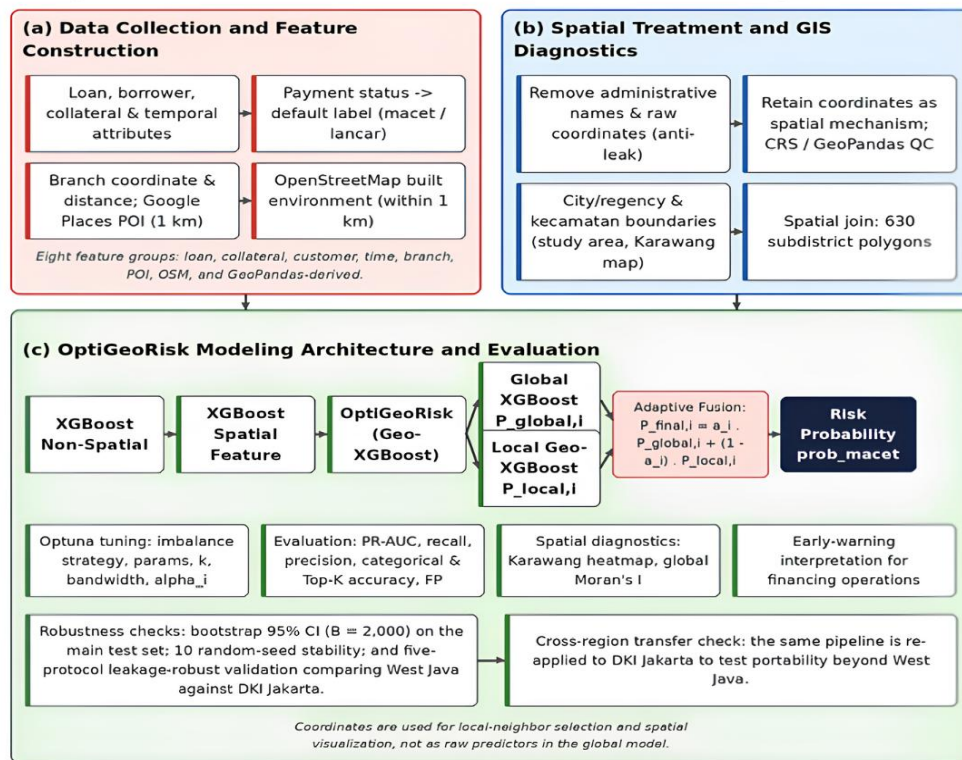


Fig. 2. Integrated research framework, feature construction, modeling architecture, and spatial diagnostic flow.

D. Spatial Diagnostics

Following spatial-data presentation used in GIS-oriented studies, the results include West Java mapping, Karawang relative risk heatmaps, feature correlation, and spatial autocorrelation diagnostics. Global Moran's I is used to evaluate whether subdistrict-level risk scores show spatial autocorrelation [29], while LISA identifies local high-high, low-low, and spatial outlier patterns [30].

E. Statistical Robustness Check

Because XGBoost Non-Spatial, XGBoost Spatial Feature, and OptiGeoRisk are compared on a single fixed

train/validation/test partition, differences between point-estimate metrics could in principle reflect sampling variability in the test set rather than genuine differences between models. To assess this, non-parametric bootstrap resampling was applied to the held-out test set (15,349 transactions), following standard practice for comparing classifiers [31], [32]. For $B = 2,000$ resamples, test-set rows were sampled with replacement (seed = 42), and PR-AUC, default recall, precision, and categorical accuracy were recomputed for each model at its selected decision threshold. Paired differences between OptiGeoRisk and each baseline were recorded across resamples to obtain a 95% percentile confidence interval and an empirical two-sided p-value for each metric difference. Because the same resampled

indices are applied to all three models within each iteration, the comparison is paired, which removes shared sampling noise and increases the power to detect genuine differences.

This procedure quantifies evaluation uncertainty for the models as fitted; it resamples test observations, not the training process itself. Training-side variability from refitting OptiGeoRisk under multiple random seeds is assessed separately in Section IV-E, together with an initial cross-region validation check.

F. Leakage-Robust Evaluation Protocols

The main comparison in Section IV-A relies on a single stratified train, validation, and test partition, which reproduces the overall 6.87% default rate but does not by itself rule out spatial, borrower-level, or temporal leakage: a borrower with several transactions, a collateral location that repeats across partitions, or overlapping time windows could all inflate apparent performance under a purely random split. To check this, four additional partitioning protocols were applied to the same West Java dataset and the same three core models, each retrained and re-evaluated on its own 60/20/20 train, validation, and test split: borrower-grouped, in which all transactions from the same borrower (no anggota) are kept within a single partition; spatial-blocked by city/regency (kab/kota sekarang), in which entire regencies are held out for validation and testing; spatial-blocked by subdistrict (kecamatan sekarang), using a finer administrative unit; and forward-chaining temporal, in which transactions are ordered chronologically so that training data always precedes the validation and test periods. The dataset supports all four protocols: it contains 60,044 unique borrowers, of whom 12,038 have more than one transaction and are therefore a potential source of borrower-level leakage under a purely random split, 27 unique regencies and 581 unique subdistricts for the spatial-blocking protocols, and complete transaction dates from January to December 2025 for the temporal protocol.

G. Fair Hyperparameter Tuning for Baseline Models

OptiGeoRisk's Optuna-based optimization stage searches over the imbalance strategy, XGBoost parameters, neighborhood size, kernel bandwidth, and fusion weight, so an unequal tuning effort could by itself explain part of its advantage over the baseline models. To rule this out, XGBoost Non-Spatial and XGBoost Spatial Feature were each re-tuned with a standardized search space and an equal budget of 50 Optuna trials, covering the same core hyperparameters used for OptiGeoRisk (number of estimators, tree depth, learning rate, subsample and column-sample ratios, minimum child weight, gamma, and L1/L2 regularization), the same candidate imbalance strategies, and the same threshold-selection objective, maximizing default recall subject to a minimum precision of 0.11 on the validation set.

H. Systematic Ablation Study

OptiGeoRisk combines several mechanisms at once: localized neighborhood learning, adaptive neighborhood-size selection, adaptive Gaussian kernel weighting, adaptive global-local probability fusion, and imbalance handling. To isolate which of these mechanisms drives the reported gains, a six-configuration ablation matrix was built, and each configuration

was retrained and evaluated on the same West Java test set. A0 is a global-only XGBoost spatial-feature baseline with no local component. A1 adds a local model with a fixed neighborhood size, fixed Gaussian weighting, and a fixed fusion weight. A2 replaces the fixed neighborhood in A1 with adaptive neighborhood and bandwidth selection, while fusion remains fixed. A3 is the full OptiGeoRisk configuration, adding adaptive global-local fusion on top of A2. A4 is a variant of A3 in which Gaussian distance weighting is replaced by uniform neighbor weights. A5 is a variant of A3 in which imbalance handling is removed from both the global and local training stages. Comparing A0 to A1 isolates the effect of adding a local component at all, A1 to A2 isolates adaptive neighborhood selection, A2 to A3 isolates adaptive fusion, A3 to A4 isolates Gaussian kernel weighting, and A3 to A5 isolates imbalance handling.

I. Operational Top-Capacity Evaluation

PR-AUC and default recall summarize model behavior across the full range of thresholds, but a collection or verification team can typically review only a limited share of the portfolio. To translate ranking quality into an operationally meaningful comparison, the three core models were additionally evaluated at fixed review capacities corresponding to the top 1%, 5%, and 10% of test-set transactions ranked by predicted default probability. For each capacity, the number of true default cases captured, precision within the reviewed set, the resulting false-positive workload, and the lift over a no-skill (random-selection) baseline were computed, together with the total default-loan exposure represented by the captured cases, to approximate the cost and benefit of prioritizing review effort by model ranking rather than by chance.

J. Computational Overhead Assessment

Localized geographically weighted learning is more expensive than a single global model because OptiGeoRisk fits a separate local classifier for each test-time neighborhood in addition to its global component. To quantify this cost, training time was recorded for the three core models on the same data split and hardware, together with OptiGeoRisk-specific overhead indicators: the number of local neighborhood models fit at test time, the mean neighborhood size (k) selected adaptively per observation, and the mean adaptive bandwidth, in kilometers, selected per observation. Training time is reported both in absolute terms and as a ratio relative to XGBoost Non-Spatial so that the additional cost of localized learning can be weighed against its early-warning benefit.

IV. RESULTS AND DISCUSSION

A. Model Performance

Table II compares three model configurations: XGBoost Non-Spatial, XGBoost Spatial Feature, and OptiGeoRisk. Pairing the non-spatial and spatial feature sets isolates the effect of the spatial features themselves, while OptiGeoRisk adds the localized learning mechanism on top, which makes the results easier to interpret.

XGBoost Spatial Feature obtains the highest PR-AUC (0.1158) and categorical accuracy (0.7029): the strongest overall probability ranking and thresholded correctness among the compared models. Branch-distance, POI, and OSM spatial

features therefore add useful signal beyond the non-spatial credit attributes. This comparison is later corroborated under standardized hyperparameter-tuning budgets in Section IV-G.

TABLE II. EVALUATION COMPARISON ON THE MAIN TEST SET

Model	PR-AUC	Recall	Precision	Cat. Acc.	FP
XGB Non-Spatial	0.1057	0.4877	0.1025	0.6715	4,502
XGB Spatial Feature	0.1158	0.4658	0.1094	0.7029	3,997
OptiGeoRisk	0.1094	0.5256	0.1055	0.6614	4,697

OptiGeoRisk records the highest default recall (0.5256), capturing 554 actual default cases, but it also produces the largest false-positive volume (4,697). It is therefore not best interpreted as a universal replacement for global XGBoost; its main value is localized early-warning sensitivity, finding more actual default cases at the cost of a greater review workload. A bootstrap robustness check (Section III-E) confirms that this recall advantage is statistically significant relative to both baselines, while the PR-AUC gap relative to XGBoost Spatial Feature is also statistically genuine rather than sampling noise (Table III).

TABLE III. BOOTSTRAP 95% CONFIDENCE INTERVALS FOR THE MAIN TEST-SET METRICS (B = 2,000)

Model	PR-AUC (95% CI)	Recall (95% CI)
XGB Non-Spatial	0.1057 (0.0972–0.1168)	0.4877 (0.4575–0.5163)
XGB Spatial Feature	0.1158 (0.1049–0.1294)	0.4658 (0.4346–0.4949)
OptiGeoRisk	0.1094 (0.0999–0.1213)	0.5256 (0.4957–0.5542)

Paired bootstrap comparison shows that OptiGeoRisk's default recall is significantly higher than both XGB Spatial Feature ($\Delta = +0.0595$, 95% CI [0.0370, 0.0833], $p = 0.000$) and XGB Non-Spatial ($\Delta = +0.0383$, 95% CI [0.0086, 0.0678], $p = 0.010$). In contrast, OptiGeoRisk's PR-AUC is significantly lower than XGB Spatial Feature ($\Delta = -0.0066$, 95% CI [-0.0130, -0.0006], $p = 0.033$) but not significantly different from XGB Non-Spatial ($\Delta = +0.0039$, 95% CI [-0.0031, 0.0112], $p = 0.273$). The recall advantage over both baselines is therefore a robust effect rather than sampling noise, and the PR-AUC shortfall relative to XGBoost Spatial Feature is statistically genuine rather than an artifact of a single test split. As additional context, the no-skill PR-AUC baseline implied by the 6.87% test-set default prevalence is 0.0687, so all three models rank default risk substantially better than random, although the absolute PR-AUC margin above this baseline remains modest (0.037–0.047).

B. Spatial Risk Visualization and Autocorrelation

The Karawang heatmap compares observed default rate and model-based average prob_{macet} at the subdistrict level. The map should be read by panel: darker red indicates higher relative

risk intensity within the same panel, while point size reflects the number of transactions. Observed data use actual default rates, whereas model panels show probability scores rather than final 0/1 decisions.

TABLE IV. GLOBAL MORAN'S I FOR KARAWANG RISK SURFACES

Panel	Moran's I	p-sim	z-sim
Observed	-0.0759	0.382	-0.421
XGB Non-Sp.	-0.1034	0.283	-0.626
XGB Spatial Feature	0.2977	0.004	3.028
OptiGeoRisk	0.2946	0.011	2.93

As Table IV reports, observed default data in Karawang do not show statistically significant global spatial autocorrelation in this test aggregation, whereas both the XGBoost Spatial Feature and OptiGeoRisk probability panels produce positive and statistically significant Moran's I values on the kecamatan-level visualization. The model probability surfaces therefore form clearer spatial clusters than the raw observed default rate. Fig. 3 visualizes these risk surfaces and helps compare whether predicted probabilities become darker in areas that are also darker in the observed default panel.

Karawang Relative Spatial Risk Intensity by Kecamatan Boundary

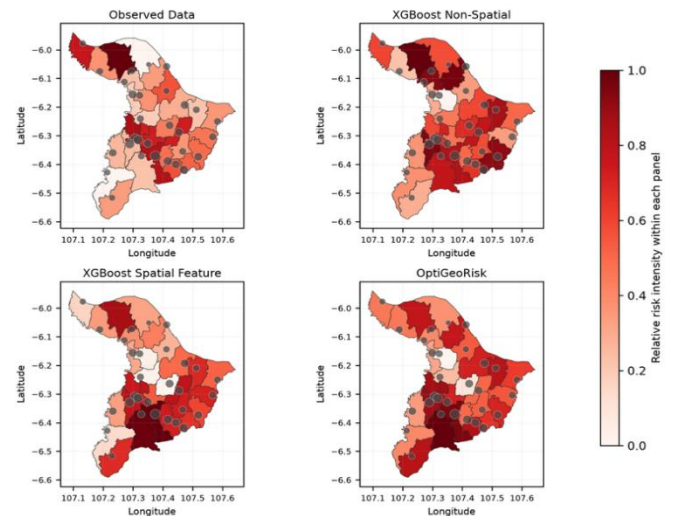


Fig. 3. Karawang relative spatial risk intensity from observed default rate and model probability scores by kecamatan boundary.

C. Feature Correlation

The strongest regional spatial-feature correlation appears after records are aggregated by city/regency, where the leading variable is OSM residential roadway count, with a Spearman-like absolute correlation of 0.6538. This supports retaining broader regional context for the main model, since the regional signal is more stable at city/regency level than at the finer kecamatan level. Fig. 4 shows this ranking: vehicle age is the only feature positively associated with default, while OSM-derived built-environment counts and same-kecamatan branch rate are the strongest negative associations.

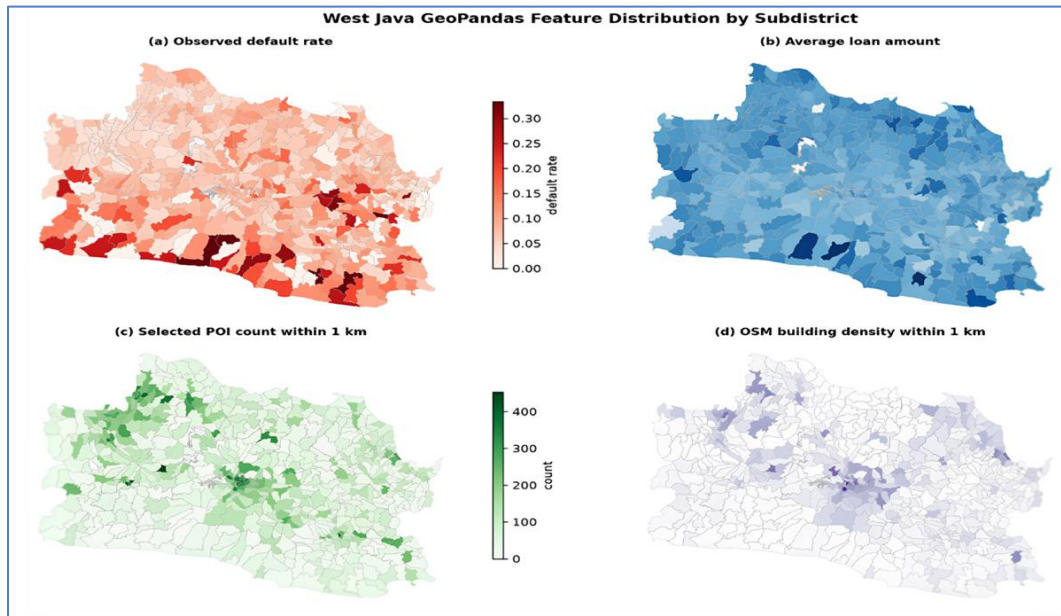


Fig. 4. West Java GeoPandas feature distribution by subdistrict: (a) observed default rate, (b) average loan amount, (c) selected POI count within 1 km, (d) OSM building density within 1 km.

The feature-correlation result is not used to claim causality; instead, it offers a transparent diagnostic that no single spatial feature dominates the target. The highly correlated feature-pair output indicates that future ablation should remove redundant density/log-count pairs before a more compact final submission model is selected.

To complement this ranking, Fig. 5 maps four of these variables across West Java at the subdistrict level: observed default rate alongside average loan amount, POI count, and OSM building density. The default-rate panel is darkest in the southern part of the study area, while POI count and building density concentrate around the central urban cluster. These two patterns occupy largely different areas, which is visually consistent with the negative correlations reported above for OSM roadway count, roadway density, and POI-related features.

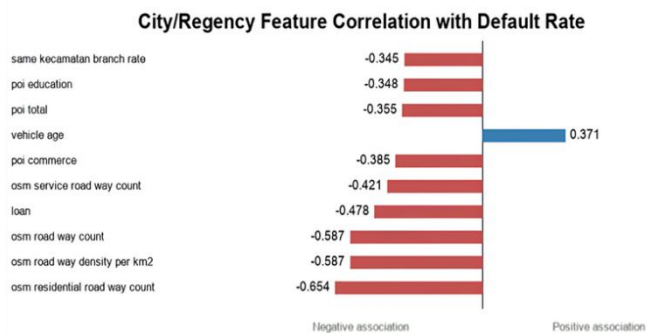


Fig. 5. City/regency-level correlation between aggregated spatial features and observed default rate.

Average loan amount does not show a comparable spatial gradient, matching its absence from the strongest correlated features in Fig. 4. As with the correlation ranking, this map is a descriptive check rather than a causal test.

D. Interpretation of the Trade-Off

The main result supports a clearer and more honest contribution statement: XGBoost Spatial Feature is preferable if the goal is the best global probability ranking, while OptiGeoRisk is more sensitive if the goal is early warning and capturing more default cases for follow-up. The model's value therefore lies in risk monitoring rather than in claiming dominance on every metric, which matters for journal review because it explains why a more expensive localized model may still be useful when a global spatial-feature baseline has stronger PR-AUC. Sections IV-F through IV-J extend this comparison to leakage-robust validation protocols, fair hyperparameter tuning, a systematic ablation of OptiGeoRisk's components, operational review-capacity analysis, and computational overhead, and refine this trade-off further rather than overturning it.

E. Seed Stability and Cross-Region Validation

To assess whether the point-estimate differences in Section IV-A generalize beyond a single train/validation/test split, all three models were refit under 10 independent random seeds on the West Java data. This procedure resamples the training process itself, which complements the bootstrap check in Section III-E (test-set resampling only) and directly addresses training-side variability.

TABLE V. STABILITY ACROSS 10 RANDOM SEEDS (WEST JAVA)

Model	PR-AUC Mean	PR-AUC Std	Recall Mean	Recall Std	FP Mean
XGB Non-Spatial	0.1076	0.0028	0.4018	0.0674	3,472
XGBoost Spatial Feature	0.1138	0.0033	0.4572	0.0640	3,931
OptiGeoRisk	0.1135	0.0034	0.4703	0.0607	4,058

As Table V reports, OptiGeoRisk keeps the highest mean default recall (0.4703) across the 10 seeds, ahead of XGBoost

Spatial Feature (0.4572) and XGBoost Non-Spatial (0.4018), with a standard deviation (0.0607) comparable to the other models. Its mean PR-AUC (0.1135) remains close to, though marginally below, XGBoost Spatial Feature (0.1138). OptiGeoRisk does not win on every individual seed, so the fairest characterization is that it is competitive and consistently more recall-sensitive rather than uniformly dominant.

As a further check on geographic portability, the same three core models were re-applied, unchanged in architecture, to a second and structurally different region, DKI Jakarta (16,292 transactions, 818 defaults), and evaluated under the same five validation protocols used for West Java in Section IV-F (transaction-random, borrower-grouped, spatial-blocked by city/regency, spatial-blocked by subdistrict, and forward-chaining temporal), so that the two regions are compared under an identical evaluation protocol rather than under two different validation schemes. West Java covers roughly 35,000 km² with heterogeneous urban and rural sub-districts, while DKI Jakarta covers about 660 km² of dense urban area, making it a meaningfully different geographic setting rather than a repeated sample of the same distribution. Table VI reports each model's mean PR-AUC, recall, and precision across the five protocols for both regions.

TABLE VI. IDENTICAL-PROTOCOL COMPARISON: WEST JAVA VS. DKI JAKARTA (MEAN ACROSS FIVE VALIDATION PROTOCOLS)

Region	Model	PR-AUC	Recall	Precision
West Java	XGB Non-Spatial	0.1110	0.3995	0.1119
West Java	XGB Spatial Feature	0.1116	0.4677	0.1101
West Java	OptiGeoRisk	0.1112	0.4680	0.1098
DKI Jakarta	XGB Non-Spatial	0.0795	0.1473	0.0693
DKI Jakarta	XGB Spatial Feature	0.0925	0.2185	0.0907
DKI Jakarta	OptiGeoRisk	0.0886	0.1269	0.1038

Because West Java and DKI Jakarta are now evaluated under the same five-protocol suite rather than under two different validation schemes, the comparison in Table VI is directly comparable across regions. Averaged across all five protocols, the PR-AUC ranking is consistent in both regions: XGBoost Spatial Feature obtains the highest mean PR-AUC in West Java (0.1116) and in DKI Jakarta (0.0925), with OptiGeoRisk close behind in West Java (0.1112) but comparatively further behind in DKI Jakarta (0.0886), where the training set is roughly one-fifth the size of West Java's and defaults are sparser.

Default recall does not transfer as consistently. In West Java, OptiGeoRisk's mean recall (0.4680) is essentially tied with XGBoost Spatial Feature's (0.4677), reflecting that OptiGeoRisk leads on only two of the five individual protocols (Section IV-F); in DKI Jakarta, XGBoost Spatial Feature has the clearly highest mean recall (0.2185), while OptiGeoRisk (0.1269) falls behind even XGBoost Non-Spatial (0.1473). A plausible explanation is that OptiGeoRisk's neighborhood and bandwidth settings were selected on West Java data and applied unchanged to DKI Jakarta, whose land area is roughly 1/53 that of West Java, so the same absolute bandwidth is far less local relative to the study area and fewer neighbors are available per query point when regional training data is scarce. This is

reported as an honest limitation: the PR-AUC ranking is reasonably stable across these two regions, but OptiGeoRisk's recall advantage is West Java-specific under this evaluation and should not be assumed to transfer to a new region without region-specific recalibration of its neighborhood and bandwidth parameters.

F. Leakage-Robust Validation Across Evaluation Protocols

Table VII reports PR-AUC, default recall, precision, and false-positive count for the three core models under the transaction-random split used in Section IV-A and under the four additional protocols described in Section III-F. Under transaction-random and borrower-grouped partitioning, OptiGeoRisk retains the highest recall among the three models (0.5417 and 0.4961), consistent with the main result in Table II. Under the two spatially blocked protocols, however, the ranking changes: when entire regencies are held out, XGBoost Spatial Feature obtains the highest recall (0.4982), narrowly ahead of OptiGeoRisk (0.4887); when entire subdistricts are held out, XGBoost Non-Spatial is marginally ahead (0.4185) of OptiGeoRisk (0.4175). Under forward-chaining temporal validation, XGBoost Spatial Feature again leads (0.4170), with OptiGeoRisk second (0.3960) and XGBoost Non-Spatial lowest (0.2814).

TABLE VII. LEAKAGE-ROBUST EVALUATION ACROSS FIVE VALIDATION PROTOCOLS (WEST JAVA)

Protocol	Model	PR-AUC	Recall	Precision	FP
Transaction Random	XGB Non-Spatial	0.1079	0.4915	0.1027	4,526
Transaction Random	XGB Spatial Feature	0.1113	0.5332	0.1056	4,759
Transaction Random	OptiGeoRisk	0.1101	0.5417	0.1058	4,824
Borrower Grouped	XGB Non-Spatial	0.1149	0.3806	0.1143	3,013
Borrower Grouped	XGB Spatial Feature	0.1115	0.4795	0.1121	3,881
Borrower Grouped	OptiGeoRisk	0.1117	0.4961	0.1111	4,055
Spatial Blocked (City/Regency)	XGB Non-Spatial	0.1083	0.4255	0.1118	4,646
Spatial Blocked (City/Regency)	XGB Spatial Feature	0.1069	0.4982	0.1046	5,866
Spatial Blocked (City/Regency)	OptiGeoRisk	0.1068	0.4887	0.1044	5,767
Spatial Blocked (Subdistrict)	XGB Non-Spatial	0.1075	0.4185	0.1057	3,606
Spatial Blocked (Subdistrict)	XGB Spatial Feature	0.1071	0.4106	0.1058	3,533
Spatial Blocked (Subdistrict)	OptiGeoRisk	0.1072	0.4175	0.1060	3,584
Forward-Chaining Temporal	XGB Non-Spatial	0.1166	0.2814	0.1251	2,147
Forward-Chaining Temporal	XGB Spatial Feature	0.1212	0.4170	0.1225	3,258
Forward-Chaining Temporal	OptiGeoRisk	0.1201	0.3960	0.1217	3,118

These results indicate that OptiGeoRisk's recall advantage is not a fixed property of the model but depends on the validation protocol, and specifically on whether the geographic locations being scored were already represented in the training data. When borrowers or transactions are held out, but their neighborhoods

remain in the training set (transaction-random and borrower-grouped protocols), the local-learning mechanism has historical neighbors to draw on, and OptiGeoRisk's recall advantage holds. When entire geographic areas are held out (the two spatial-blocked protocols), OptiGeoRisk has no local history for the withheld locations and falls back close to, or slightly behind, the global XGBoost Spatial Feature baseline. Forward-chaining temporal validation shows a related pattern: because later transactions are more likely to fall in locations not yet densely represented in the earlier training window, OptiGeoRisk's advantage weakens there as well. The practical implication is that OptiGeoRisk should be positioned as a localized early-warning layer whose value depends on sufficient local historical coverage, rather than as a model expected to outperform a global baseline in newly opened or thinly covered territories.

G. Fair-Tuned Baseline Confirmation

Table VIII reports XGBoost Non-Spatial and XGBoost Spatial Feature under the standardized 50-trial Optuna search described in Section III-G, using the same transaction-random test split as Table II.

TABLE VIII. FAIR-TUNED BASELINE COMPARISON UNDER A STANDARDIZED 50-TRIAL SEARCH (WEST JAVA)

Model	PR-AUC	Recall	Precision	FP
XGB Non-Spatial	0.1069	0.4953	0.1040	4,499
XGB Spatial Feature	0.1108	0.5171	0.1058	4,605

Under equal tuning effort, XGBoost Spatial Feature's PR-AUC (0.1108) remains close to its originally reported value (0.1158) and continues to exceed XGBoost Non-Spatial's fair-tuned PR-AUC (0.1069), confirming that XGBoost Spatial Feature is a strong global baseline independent of tuning budget. Its default recall, however, rises from 0.4658 under the original tuning to 0.5171 under the standardized objective, which narrows its gap with OptiGeoRisk's recall (0.5256 in Table II) from 0.0598 to 0.0085. This indicates that part of OptiGeoRisk's originally reported recall advantage over XGBoost Spatial Feature reflected differences in tuning effort rather than the local-learning mechanism alone, while its PR-AUC shortfall relative to XGBoost Spatial Feature persists under fair tuning. XGBoost Non-Spatial's fair-tuned recall (0.4953) similarly improves over its original value (0.4877) but still trails OptiGeoRisk's. Read together with Section IV-D, this supports a more cautious framing of OptiGeoRisk's contribution: it remains competitive and is still the more recall-sensitive of the three core models under identical-protocol validation (Section IV-F), but its margin over a properly tuned XGBoost Spatial Feature baseline is smaller than the untuned comparison suggested, and its PR-AUC trade-off against that baseline is not resolved by tuning.

H. Systematic Ablation Results

Table IX reports PR-AUC, recall, precision, and false-positive count for the six ablation configurations described in Section III-H, all evaluated on the transaction-random West Java test set.

Adding a local component at all (A0 to A1) does not by itself improve recall: A1's recall (0.5256) is close to A0's (0.5332), and PR-AUC is essentially unchanged. Adaptive neighborhood

and bandwidth selection (A1 to A2) raises recall from 0.5256 to 0.5427, the largest single gain in the matrix, confirming that letting the neighborhood adapt to local data density, rather than fixing it, is the main source of OptiGeoRisk's recall sensitivity. Adding adaptive global-local fusion (A2 to A3) trades a small amount of that recall back (0.5427 to 0.5417) in exchange for fewer false positives (4,937 to 4,824), showing that the fusion mechanism mainly helps control false-positive workload rather than raise recall further. Replacing Gaussian kernel weighting with uniform neighbor weights (A3 to A4) changes almost nothing on this test set (PR-AUC 0.1101 in both, recall 0.5417 in both, false positives 4,824 versus 4,828), so Gaussian weighting is not the mechanism responsible for OptiGeoRisk's reported gains here, at least at the neighborhood sizes selected on this dataset. Removing imbalance handling (A3 to A5) has the largest effect of all: recall falls sharply from 0.5417 to 0.4440 and false positives drop from 4,824 to 3,923, confirming that imbalance handling, not the spatial mechanism, is doing much of the work in keeping recall high. Taken together, the ablation attributes OptiGeoRisk's early-warning sensitivity mainly to adaptive neighborhood selection and imbalance handling, with adaptive fusion serving a complementary role in workload control and Gaussian weighting contributing little beyond what uniform local weighting already achieves on this split.

TABLE IX. SYSTEMATIC ABLATION OF OPTIGEO RISK COMPONENTS (WEST JAVA)

Configuration	PR-AUC	Recall	Precision	FP
A0: Global spatial baseline	0.1113	0.5332	0.1056	4,759
A1: Fixed local + Gaussian + fixed fusion	0.1112	0.5256	0.1066	4,641
A2: Adaptive neighborhood + fixed fusion	0.1113	0.5427	0.1038	4,937
A3: Full OptiGeoRisk	0.1101	0.5417	0.1058	4,824
A4: No Gaussian weighting	0.1101	0.5417	0.1058	4,828
A5: No imbalance handling	0.1113	0.4440	0.1066	3,923

I. Operational Value at Review Capacity

Table X reports the number of default cases captured, recall and precision within the reviewed set, and false-positive workload for the three core models at review capacities of 1%, 5%, and 10% of the 15,349 test-set transactions.

At every capacity tested, XGBoost Spatial Feature captures the most default cases in absolute terms: 32 of 1,054 at 1% capacity, 118 at 5%, and 217 at 10%, corresponding to a lift over random selection of 3.05x, 2.24x, and 2.06x, respectively. OptiGeoRisk captures fewer cases at each capacity (27, 107, and 197) but still clearly outperforms random selection (2.57x, 2.03x, and 1.87x lift) and remains close to XGBoost Non-Spatial, which captures 20, 102, and 196 cases at the same capacities. In business terms, prioritizing the top 5% of transactions by XGBoost Spatial Feature's score captures loan exposure of approximately Rp 220.3 million in eventual defaults out of Rp 2,025.4 million total default exposure in the test set, against a false-positive workload of 649 transactions and a break-even review cost of about Rp 183,332 per false positive; OptiGeoRisk captures a similar but somewhat lower share, about Rp 199.5 million, at a workload of 660 transactions. This

operational view reinforces the finding from Sections IV-D and IV-G that XGBoost Spatial Feature is the stronger model for ranking-based prioritization at realistic review capacities, while OptiGeoRisk's comparative advantage lies less in capturing the largest absolute number of defaults at a fixed capacity and more in its recall behavior under protocols where local historical coverage is available (Section IV-F).

TABLE X. OPERATIONAL VALUE AT TOP 1%, 5%, AND 10% REVIEW CAPACITY (WEST JAVA)

Capacity	Model	Defaults Captured	Recall@ Capacity	Precision @Capacity	FP
Top 1%	XGB Non-Spatial	20	0.0190	0.1307	133
Top 1%	XGB Spatial Feature	32	0.0304	0.2092	121
Top 1%	OptiGeoRisk	27	0.0256	0.1765	126
Top 5%	XGB Non-Spatial	102	0.0968	0.1330	665
Top 5%	XGB Spatial Feature	118	0.1120	0.1538	649
Top 5%	OptiGeoRisk	107	0.1015	0.1395	660
Top 10%	XGB Non-Spatial	196	0.1860	0.1277	1,339
Top 10%	XGB Spatial Feature	217	0.2059	0.1414	1,318
Top 10%	OptiGeoRisk	197	0.1869	0.1283	1,338

J. Computational Overhead

Table XI reports training time and OptiGeoRisk-specific overhead indicators for the three core models on the same 46,044-row training split.

TABLE XI. COMPUTATIONAL OVERHEAD FOR THE THREE CORE MODELS (WEST JAVA)

Model	Fit Time (s)	Ratio vs. XGB Non-Spatial	Local Groups (Test)	Mean Bandwidth (km)
XGB Non-Spatial	2.08	1.00x	-	-
XGB Spatial Feature	12.28	5.91x	-	-
OptiGeoRisk	145.88	70.13x	609	9.6

XGBoost Spatial Feature trains in 12.28 seconds, about 5.91 times longer than XGBoost Non-Spatial's 2.08 seconds, reflecting the added cost of branch-distance, POI, and OSM feature computation. OptiGeoRisk is substantially more expensive still, at 145.88 seconds, about 70.13 times XGBoost Non-Spatial's training time, because it fits a separate local classifier for each of 609 test-time neighborhood groups, using a mean adaptively selected neighborhood size of about 52 observations and a mean adaptive bandwidth of about 9.6 kilometers. This overhead is a real deployment consideration: OptiGeoRisk's benefit is concentrated in specific validation

settings (Section IV-F), and its absolute recall gain over a fairly tuned baseline is modest (Section IV-G), so its added computational cost should be weighed against how frequently a portfolio needs to be re-scored and against simpler alternatives, such as periodic rather than continuous retraining of the local component.

These fit-time figures do not by themselves show how expensive OptiGeoRisk is to score new transactions, how much of that cost is spatial neighbor search rather than local-model fitting, or how its memory footprint compares with a global classifier. To characterize these, the same transaction-random split was rescored in a separate, instrumented reproduction that isolates distance computation from local-model fitting: scoring the 15,349-row test set took 5.1 ms for XGBoost Non-Spatial and 6.5 ms for XGBoost Spatial Feature (both under 0.5 ms per 1,000 transactions), against roughly three to four thousand times that latency per transaction for OptiGeoRisk, because each of its 609 candidate kecamatan-level neighborhoods requires a freshly fit local XGBoost model rather than a single forward pass (593 of the 609 candidates contained both classes and were actually fit). Of OptiGeoRisk's total scoring time, only about 8% was spent computing haversine distances from each neighborhood centroid to every training transaction and ranking them by proximity; the remaining 92% was local-model fitting and prediction. Because this reproduction ran on different hardware than Table XI's own timings, its absolute seconds are not directly comparable to Table XI; the informative findings are the latency gap itself and the small, roughly one-tenth, share of it attributable to distance computation.

To assess how this overhead scales with the size of the historical transaction pool, the same 609 test-time neighborhoods were rescored while reducing the training pool available for neighbor search to 25%, 50%, and 75% of its original 46,044 rows. Distance-computation time grew roughly in proportion to the training-pool size, consistent with the current implementation computing a haversine distance from each neighborhood centroid to every available training transaction rather than using a spatial index such as a KD-tree or ball-tree. Local-model fit-and-predict time, in contrast, stayed essentially flat across all four pool sizes, because the adaptive neighborhood-selection rule caps each local model at 200 nearest neighbors (a mean of about 52 to 57 neighbors was actually used); growing the historical pool changes which transactions are nearest, not how many are used to fit each local model. In practice, OptiGeoRisk's per-scoring-pass latency is therefore driven mainly by the number of distinct geographic neighborhoods that must be refit, so it should grow primarily as the model's geographic footprint expands to new kecamatan, and only mildly as transaction history accumulates within kecamatan already covered; a spatial index over the training coordinates is a natural way to remove the linear distance-computation cost documented here as that history keeps growing.

Peak resident memory across the full reproduction, including loading the dataset, one-hot encoding the categorical branch, kecamatan, and vehicle-type fields into a roughly 3,400-column feature matrix, fitting both global baselines, and fitting all 593 local models needed to score the test set, stayed at approximately 530 MB when the feature matrix is kept in its native sparse form

(on the order of 40 nonzero entries per row out of about 3,400 columns). A serialized global XGBoost model occupies roughly 170 KB; a single local model occupies roughly 110 to 120 KB. Because local models are fit on demand and discarded once a neighborhood is scored, OptiGeoRisk never holds more than one local model in memory at a time, but unlike the two global baselines, whose inference-time footprint is limited to their small serialized model file, OptiGeoRisk must additionally keep the entire historical training feature matrix and coordinate set resident at serving time to support on-demand neighbor search and refitting.

V. CONCLUSION

This study proposed OptiGeoRisk as a spatially localized and optimized Geo-XGBoost framework for credit default risk classification. The study uses real-world collateral-based loan transactions from West Java and addresses two connected issues, severe class imbalance and geographic heterogeneity in local default patterns, by removing administrative location names and raw coordinate predictors from the ordinary feature matrix. The experiment then evaluates whether coordinates can be used more appropriately as a structural mechanism for local learning.

The main test result shows that XGBoost Spatial Feature achieves the best PR-AUC (0.1158) while OptiGeoRisk achieves the highest default recall (0.5256), a trade-off between global ranking performance and localized early-warning sensitivity. XGBoost Spatial Feature obtained the highest PR-AUC because its global learning structure produced a more stable overall ranking of default risk across the full test set, while still benefiting from branch-distance, POI, and OSM-derived spatial-context variables.

OptiGeoRisk, by contrast, used coordinates to adapt predictions through local neighborhood learning and global-local probability fusion. This localized mechanism improved sensitivity to actual default cases and produced the highest default recall, but it also introduced more false positives and slightly reduced global ranking consistency. OptiGeoRisk should therefore not be interpreted as universally superior to XGBoost Spatial Feature; rather, it is more useful as an early-warning layer when the operational priority is to capture more potential default cases for follow-up. A bootstrap significance check supports this framing: the recall gain is robust across resamples, while the PR-AUC shortfall relative to XGBoost Spatial Feature is also statistically genuine (Section III-E; Table III). A fair-tuned check (Section IV-G) shows that part of this recall margin over XGBoost Spatial Feature narrows once both baselines receive an equal hyperparameter-tuning budget, while the PR-AUC trade-off persists, so OptiGeoRisk is best read as competitive rather than categorically superior.

For financial institutions, pawnshop businesses, consumer loan providers, and collateral-based financing companies, this behavior can be valuable as an early-warning layer that ranks customers or territories needing earlier verification, preventive contact, or portfolio monitoring. It should not be used as an automatic rejection mechanism; instead, it supports human and operational decision-making, where catching risky cases earlier may be more valuable than minimizing every false alarm. This framing does not depend on the recall advantage generalizing

everywhere: the PR-AUC gain over a no-skill baseline held in both regions tested (Section IV-E), so the early-warning case for OptiGeoRisk rests on the ranking gain as much as on recall. Section IV-I quantifies this concretely: at a 5% review capacity, the three core models capture between 102 and 118 of the 1,054 test-set default cases, at least 1.9 times more than the roughly 53 expected under random selection at the same capacity.

The study has several limitations and future directions. First, although Sections IV-F and IV-G now evaluate leakage-robust validation protocols and fair, equal-budget hyperparameter tuning for the three core models, the systematic ablation, computational overhead, and operational-capacity analyses in Sections IV-H through IV-J were likewise run only for these three models. Second, the cross-region check now uses an identical five-protocol evaluation for West Java and DKI Jakarta, which resolves the earlier protocol mismatch, but still covers only one additional region with a substantially smaller and sparser transaction set; broader validation across more provinces, product types, and time periods is still required. Third, the leakage-robust protocols in Section IV-F show that OptiGeoRisk's recall advantage depends on local historical coverage and weakens or reverses when whole locations or future time periods are held out, which should be treated as a deployment constraint rather than a property to be assumed; region-aware or coverage-aware bandwidth calibration is a concrete next step. Fourth, kecamatan polygons improved visualization detail but did not improve the main model when used as granular spatial features; future work should test hierarchical spatial modeling rather than simply adding finer administrative predictors. Fifth, OptiGeoRisk's computational overhead, about 70 times XGBoost Non-Spatial's training time (Section IV-J), means that frequent full retraining may not be practical at larger portfolio scale; future development should test business-cost-based thresholds, probability calibration, periodic rather than continuous retraining schedules, and model monitoring under portfolio drift.

AUTHOR CONTRIBUTIONS

C. Prastian designed the study, processed the dataset, implemented the Geo-XGBoost and OptiGeoRisk models, and wrote the initial draft. A.A.S.G. supervised the research, reviewed the spatial autocorrelation and bootstrap analyses, and revised the manuscript. Both authors reviewed and approved the final manuscript.

DATASET AVAILABILITY

The primary loan transaction dataset used in this study contains proprietary and personally identifiable borrower information from a collateral-based financing institution in West Java, Indonesia, and is therefore not publicly available due to customer privacy obligations and institutional confidentiality agreements, consistent with Indonesia's Personal Data Protection Law (UU No. 27/2022) and OJK regulations on financial-sector data governance. De-identified or aggregated data supporting the findings of this study may be made available by the corresponding author upon reasonable request and subject to a data-sharing agreement with the data-providing institution. The publicly available auxiliary datasets used for spatial feature engineering, OpenStreetMap building and road-network data [28] and administrative boundary polygons can be freely

accessed at <https://www.openstreetmap.org>. Point-of-interest features were derived from the Google Places API [27], accessible under Google's standard commercial API terms of service.

DECLARATION ON GENERATIVE AI

During the preparation of this manuscript, the author(s) used a generative AI tool (Claude, Anthropic) to assist with literature screening and citation verification, statistical robustness analysis (bootstrap resampling and significance testing), and language editing of the manuscript text. After using this tool, the author(s) critically reviewed, verified, and edited all AI-assisted content, and take full responsibility for the accuracy, originality, and scientific validity of the final manuscript. The core research design, experimental implementation, data analysis, interpretation of results, and conclusions are the authors' own work.

REFERENCES

- [1] Otoritas Jasa Keuangan, "Peraturan Otoritas Jasa Keuangan Republik Indonesia Nomor 39 Tahun 2024 tentang Pergadaian." 2024. [Online]. Available: <https://ojk.go.id/id/regulasi/Documents/Pages/POJK-39-Tahun-2024-Pergadaian/POJK%2039%20Tahun%202024%20Pergadaian.pdf>
- [2] Otoritas Jasa Keuangan, "Roadmap Pengembangan dan Penguatan Pergadaian 2024–2028." 2024.
- [3] D. J. Hand and W. E. Henley, "Statistical classification methods in consumer credit scoring: A review," *J. R. Stat. Soc. Ser. A*, vol. 160, no. 3, pp. 523–541, 1997, doi: 10.1111/j.1467-985X.1997.00078.x.
- [4] S. Lessmann, B. Baesens, H.-V. Seow, and L. C. Thomas, "Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research," *Eur. J. Oper. Res.*, vol. 247, no. 1, pp. 124–136, 2015, doi: 10.1016/j.ejor.2015.05.030.
- [5] F. Louzada, A. Ara, and G. B. Fernandes, "Classification methods applied to credit scoring: Systematic review and overall comparison," *Surv. Oper. Res. Manag. Sci.*, vol. 21, no. 2, pp. 117–134, 2016, doi: 10.1016/j.sorms.2016.10.001.
- [6] I. Brown and C. Mues, "An experimental comparison of classification algorithms for imbalanced credit scoring data sets," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 3446–3453, 2012, doi: 10.1016/j.eswa.2011.09.033.
- [7] J. Davis and M. Goadrich, "The relationship between Precision-Recall and ROC curves," in *Proceedings of the 23rd International Conference on Machine Learning*, 2006, pp. 233–240. doi: 10.1145/1143844.1143874.
- [8] T. Saito and M. Rehmsmeier, "The Precision-Recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PLOS ONE*, vol. 10, no. 3, p. e0118432, 2015, doi: 10.1371/journal.pone.0118432.
- [9] N. Uddin, Md. K. U. Ahamed, M. A. Uddin, Md. M. Islam, Md. A. Talukder, and S. Aryal, "An ensemble machine learning based bank loan approval predictions system with a smart application," *Int. J. Cogn. Comput. Eng.*, vol. 4, pp. 327–339, 2023, doi: 10.1016/j.ijcce.2023.09.001.
- [10] A. Ananda Surya, D. R. Darmawan, and A. Solichin, "Prediksi Kapabilitas Calon Debitur Menggunakan Analisis Data Machine Learning Dengan Metode Random Forest," *J. Algoritma*, vol. 22, no. 1, pp. 777–788, 2025, doi: 10.33364/algoritma/v.22-1.1929.
- [11] W. Andriani, Gunawan, and N. N. P. W. Naja, "Analisis perbandingan machine learning untuk prediksi kelayakan kredit perbankan pada Bank BRI Tegal," *IT-Explore J. Penerapan Teknol. Inf. Dan Komun.*, vol. 4, no. 1, pp. 82–92, 2025, doi: 10.24246/itexplore.v4i1.2025.pp82-92.
- [12] A. A. Montevechi, R. de C. Miranda, A. L. Medeiros, and J. A. B. Montevechi, "Advancing credit risk modelling with Machine Learning: A comprehensive review of the state-of-the-art," *Eng. Appl. Artif. Intell.*, vol. 137, p. 109082, 2024, doi: 10.1016/j.engappai.2024.109082.
- [13] H. Ayari, R. Guetari, and N. Kraïem, "Machine learning powered financial credit scoring: a systematic literature review," *Artif. Intell. Rev.*, vol. 59, 2025, doi: 10.1007/s10462-025-11416-2.
- [14] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
- [15] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.
- [16] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, 2009, doi: 10.1109/TKDE.2008.239.
- [17] C. Brunsdon, A. S. Fotheringham, and M. E. Charlton, "Geographically weighted regression: A method for exploring spatial nonstationarity," *Geogr. Anal.*, vol. 28, no. 4, pp. 281–298, 1996, doi: 10.1111/j.1538-4632.1996.tb00936.x.
- [18] A. S. Fotheringham, C. Brunsdon, and M. Charlton, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester, U.K.: Wiley, 2002.
- [19] K. Kopczewska, "Spatial machine learning: New opportunities for regional science," *Ann. Reg. Sci.*, vol. 68, no. 3, pp. 713–755, 2022, doi: 10.1007/s00168-021-01101-x.
- [20] G. Grekousis, "Geographical-XGBoost: A new ensemble model for spatially local regression based on gradient-boosted trees," *J. Geogr. Syst.*, vol. 27, no. 2, pp. 169–195, 2025, doi: 10.1007/s10109-025-00465-4.
- [21] J. Mushava and M. Murray, "A novel XGBoost extension for credit scoring class-imbalanced data combining a generalized extreme value link and a modified focal loss function," *Expert Syst. Appl.*, vol. 202, p. 117233, 2022, doi: 10.1016/j.eswa.2022.117233.
- [22] P. Kündig and F. Sigrist, "A spatio-temporal machine learning model for mortgage credit risk: Default probabilities and loan portfolios," *Eur. J. Oper. Res.*, vol. 331, no. 3, pp. 1022–1041, 2026, doi: 10.1016/j.ejor.2025.10.039.
- [23] F. Gao, S. Y. He, and M.-P. Kwan, "Mixed Geographically Weighted XGBoost (M-GWXGB) model: A new spatially explicit machine learning model," *Ann. Am. Assoc. Geogr.*, vol. 116, pp. 824–855, 2025, doi: 10.1080/24694452.2025.2586039.
- [24] M. Atef, S. Ouf, W. Seoud, and M. I. Gabr, "A novel approach using explainable prediction of default risk in peer-to-peer lending based on machine learning models," *Neural Comput. Appl.*, vol. 37, pp. 21783–21803, 2025, doi: 10.1007/s00521-025-11489-8.
- [25] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019, pp. 2623–2631. doi: 10.1145/3292500.3330701.
- [26] K. Jordahl and others, "GeoPandas: Python tools for geographic data." 2020. doi: 10.5281/zenodo.3946761.
- [27] Google, "Places API." [Online]. Available: <https://developers.google.com/maps/documentation/places/web-service/overview>
- [28] OpenStreetMap contributors, "Planet dump retrieved from <https://planet.osm.org>." 2026. [Online]. Available: <https://www.openstreetmap.org>
- [29] P. A. P. Moran, "Notes on continuous stochastic phenomena," *Biometrika*, vol. 37, no. 1/2, pp. 17–23, 1950.
- [30] L. Anselin, "Local indicators of spatial association-LISA," *Geogr. Anal.*, vol. 27, no. 2, pp. 93–115, 1995, doi: 10.1111/j.1538-4632.1995.tb00338.x.
- [31] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. New York, NY, USA: Chapman & Hall/CRC, 1994. doi: 10.1201/9780429246593.
- [32] T. G. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," *Neural Comput.*, vol. 10, no. 7, pp. 1895–1923, 1998, doi: 10.1162/089976698300017197.