Cross-Organizational Transferability of Early Default Prediction Models in Indonesia's Motorcycle Leasing Industry

Ivan Kristiawan[1], Twin Yoshua R. Destyanto[*2,3], Jeremy Laurence Bañez[4]

[1]CV Mitra Sistem Solusi, Yogyakarta, Indonesia, 55161

[2]Department of Industrial Engineering, Universitas Atma Jaya Yogyakarta, Sleman, Indonesia, 55281

[3]Department of Theological Studies, Faculty of Theology, Sekolah Tinggi Teologi Bethel Indonesia Jakarta, Jakarta, Indonesia, 10260

[4]Industrial Engineering Department, Bulacan State University, City of Malolos, Bulacan, Philippines, 3000

E-mail: mitrasistemsolusi@gmail.com[1], twin.destyanto@uajy.ac.id*[2], jeremylaurence.banez@bulsu.edu.ph[3]

Abstract. The motorcycle-leasing sector in Indonesia is critical for consumer financing, yet firms face persistent difficulty in predicting early installment defaults because most credit-risk models are built for single organizations, forcing companies to repeatedly rebuild models when policies or operations change, which increases costs and delays risk detection. This study examines whether early default prediction models developed in one motorcycle-leasing company can be transferred to others by applying a hierarchical framework that integrates feature engineering, behavioral clustering, and supervised classification. The model was trained on 113,222 fiduciary contracts (2011–2025) from a Batam-based firm and tested on two external firms in Batam and Jakarta using Logistic Regression, Random Forest, and LightGBM. Results show substantial performance decline for the second Batam firm but relatively stable performance in Jakarta, indicating that organizational policy differences matter more than regional factors. Fine-tuning with sufficient local data improves performance, while limited data creates instability. The study provides a practical foundation for scalable and transferable credit-risk modeling in emerging markets.

Keywords: credit risk prediction; transfer learning; domain adaptation; motorcycle leasing; clustering.

1. Introduction

The motorcycle-leasing sector in Indonesia plays a vital role in consumer financing, yet it faces persistent challenges in predicting early installment defaults that threaten operational stability. There are certain cases involving the poor performance of leasing customers in paying their motorcycle loan payments. These include late payments [1], [2], failed transfers [3], and even defaults [4], [5], [6]. Such issues should be prevented from the beginning of the leasing admission process.

Leasing companies have implemented certain preventive measures, including desk interviews, field surveys, and reviewing bank statements submitted during the admission process. However, such prior actions still have the potential for misjudgment in approval decisions. The submitted documents, survey results, and provided bank statements can be easily conditioned by the applicants. Therefore, a decision support system is needed to help leasing stakeholders make informed judgments about approval. Recent studies in credit-risk prediction have shown that feature-transformation ensembles, machine learning, and deep learning models can effectively capture non-linear default patterns in credit card and loan portfolios [10], [12], [13]. At the same time, explainable AI techniques such as SHAP have been proposed to interpret complex credit-scoring models and to highlight the contribution of behavioral and financial attributes to the final risk score [11].

Most existing credit-risk models in this sector are designed for single organizations [7], [8], [9], which limits their ability to generalize across firms that differ in management policies, penalty structures, and regional characteristics. Recent work on transfer learning and domain-adaptation frameworks for credit scoring indicates that such cross-domain shifts can substantially degrade predictive performance unless the model is recalibrated to the target environment [14], [15]. As a result, institutions often rebuild models from scratch, which increases costs and delays the detection of risks. This research addresses the gap by examining whether a predictive model trained on historical fiduciary contracts from one company can be applied effectively to other firms with different behavioral and operational settings. Hence, this study aimed to (1) assess the extent to which behavioral and structural features can be generalized across companies; (2) measure the impact of organizational-policy drift on predictive accuracy; and (3) propose a methodological guideline for adaptive credit-risk modeling in heterogeneous financial environments. Specifically, it assesses the cross-organizational transferability of early default prediction models within the motorcycle leasing industry in Batam and Jakarta.

2. Method

2.1. Dataset

Data Sources and Collection Period

This study used proprietary fiduciary motorcycle-leasing datasets from three Indonesian finance companies, accessed directly from their operational databases with formal academic permission. The data consist of historical transactional records generated through routine business operations, with no prospective or manual data collection.

Company A, based in Batam, operates on MongoDB and provides 113,222 contracts from 2011–2025, offering a comprehensive thirteen-year view of repayment behavior. Company B, also in Batam, uses MySQL, from which approximately 3,000 contracts from 2022–2025 were extracted, reflecting a shorter and evolving operational period. Company C, located in Jakarta, operates on MongoDB and contributes about 1,000 contracts from 2024–2025, representing an early-stage company with developing policies.

All datasets were standardized into a unified JSON schema to harmonize variables across companies and imported into a Google Colab environment for preprocessing, feature engineering, clustering, and classification.

Dataset Profile

The unified dataset combines contractual attributes and early repayment records obtained directly from the operational databases of the three companies (Table 1). The raw variables describe each financing contract, including pricing and installment details, tenor, capital and interest components, total receivables, key dates, and outcome indicators such as repossession status and remaining months at default.

Each record also captures detailed information from the first three installment cycles, including due and payment dates, early or late payment days, fines, discounts, outstanding penalties, and payment amounts. These variables allow reconstruction of early repayment patterns and assessment of timeliness, penalty exposure, and payment discipline. All data are used in their original stored form prior to transformation or feature engineering.

Table 1. Dataset features.

No.	Feature name	Description
1	number	Sequential row number of the contract
2	price	Cash price of the motorcycle
3	downPayment	Initial down payment paid by the customer
4	remainingReceivables	Remaining receivable balance after down payment
5	monthlyInstallments	Agreed monthly installment amount
6	tenor	Total number of installment months
7	amountOfReceivables	Total financed receivables over the tenor
8	capitalInstallments	Principal portion of each monthly installment
9	interestInstallments	Interest portion of each monthly installment
10	date	Contract start date
11	firstInstallmentDate	Scheduled date of the first installment
12	lastInstallmentDate	Scheduled date of the last installment
13	finalDueDate	Final overall due date of the contract
14	remainingInstallmentMonths	Number of installments remaining at extraction time
15	installmentDefault	Flag indicating repossession/default status
16	remainingMonthDefault	Months remaining at the time of default
17	firstDueDate	Due date of the first installment
18	firstPaymentDate	Actual payment date of the first installment
19	firstLateDayPaymentDifference	Number of days the first payment was late
20	firstAdvanceDayPaymentDifference	Number of days the first payment was early
21	firstFine	Fine charged on the first installment
22	firstInstallmentDiscount	Discount or rebate on the first installment
23	firstFineDebt	Unpaid fine from the first installment carried forward
24	firstPayment	Amount paid for the first installment
25	secondDueDate	Due date of the second installment
26	secondPaymentDate	Actual payment date of the second installment
27	secondLateDayPaymentDifference	Number of days the second payment was late
28	secondAdvanceDayPaymentDifference	Number of days the second payment was early
29	secondFine	Fine charged on the second installment
30	secondInstallmentDiscount	Discount or rebate on the second installment
31	secondFineDebt	Unpaid fine from the second installment is carried forward
32	secondPayment	Amount paid for the second installment
33	thirdDueDate	Due date of the third installment
34	thirdPaymentDate	Actual payment date of the third installment
35	thirdLateDayPaymentDifference	Number of days the third payment was late
36	thirdAdvanceDayPaymentDifference	Number of days the third payment was early
37	thirdFine	Fine charged on the third installment
38	thirdInstallmentDiscount	Discount or rebate on the third installment
39	thirdFineDebt	Unpaid fine from the third installment carried forward

No.	Feature name	Description
40	thirdPayment	Amount paid for the third installment

Data Accessibility

The datasets used in this study are proprietary and were sourced directly from the internal databases of three finance companies. Although the data are not publicly available, we obtained explicit permission to use the data for academic purposes through our roles as system developers and maintainers of their Enterprise Resource Planning (ERP) systems and database infrastructures. Companies A and C operate on MongoDB databases, while Company B uses a MySQL relational database. After extraction, all datasets were anonymized by removing personal identifiers and sensitive company labels, retaining only the contract-level and payment-level information necessary for modelling. The anonymized datasets were stored securely and not shared externally to maintain client confidentiality. Consequently, the data cannot be published or redistributed, and its accessibility remains limited solely to the authors within the scope of this research.

2.2. Model framework

The modeling framework adopted in this study is designed to convert raw contract and payment data into structured behavioral representations, extract latent repayment patterns, and evaluate predictive performance across different companies. The framework consists of interconnected components that operate sequentially: feature engineering, behavioral clustering, supervised classification, and cross-company transfer learning. Together, these components form the architectural backbone of the study's methodological approach.

Overall Architecture

The overall modeling architecture begins with the transformation of raw operational data into analytically meaningful variables . Once the datasets from all companies are standardized, we construct derived behavioral and structural features to capture repayment dynamics that are not directly visible in the raw fields. These enhanced representations serve as inputs to an unsupervised clustering model, which identifies latent behavioral groups. The cluster assignments are then incorporated into a series of supervised classification models to predict installment default. The final stage of the framework includes optional local fine-tuning to ensure the model aligns with company-specific policies, penalty structures, and data distributions prior to operational use.

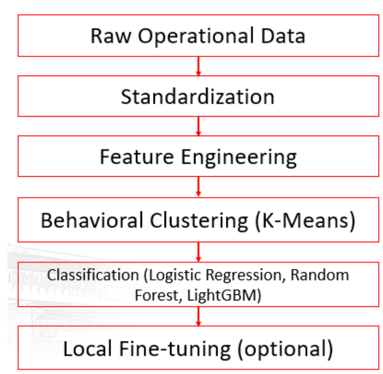

Figure 1. Design Framework

Feature Engineering Framework

Feature engineering transforms the raw contractual and payment records into behavioral indicators that more accurately represent early repayment patterns Table 2. Feature Engineering.. Using operational fields such as payment dates, late or early days, fines, and installment amounts, we programmatically construct structural and behavioral features in a Google Colab environment. These include measures such as the down-payment-to-price ratio, average delay, delay variability, early payment ratio, total payments, payment-to-due ratio, and trends in delay between installments. Penalty-related indicators, such as fine ratios and aggregated fine measures, capture the customer's exposure to financial penalties.

Together, these engineered features summarize key aspects of customer behavior, including financial commitment, payment discipline, consistency, penalty burden, and cash flow stability. This compact behavioral representation forms the basis for the clustering stage and enhances the feature space used in subsequent classification modeling.

Table 2. Feature Engineering.

No	Feature name	Formula	Meaning
1	dp_to_price_ratio	downPayment/price	Risk appetite/initial financial commitment
2	avg_delay_days	Mean (late_day_1, late_day_2, late_day_3)	Average lateness across the first 3 installments
3	std_delay_days	Std (late_day_1, late_day_2, late_day_3)	Payment consistency/variability of lateness
4	early_payment_ratio	count(advance_day > 0) / 3	Tendency to pay installments early
5	total_fines	fine_1 + fine_2 + fine_3	Total penalties charged in first 3 installments
6	total_payments	pay_1 + pay_2 + pay_3	Total cash paid across first 3 installments
7	payment_to_due_ratio	total_payments / (3 * monthlyInstallments)	Completeness of early repayment
8	first_delay_trend	late_day_3 - late_day_1	Behavior change (improving or worsening delays)
9	fine1_ratio	firstFine / monthlyInstallments	Fine percentage for installment 1
10	fine2_ratio	secondFine / monthlyInstallments	Fine percentage for installment 2
11	fine3_ratio	thirdFine / monthlyInstallments	Fine percentage for installment 3
12	avg_fine_ratio	Mean (fine1_ratio, fine2_ratio, fine3_ratio)	Average penalty burden relative to the installment
13	total_fine_ratio	Sum (fine1_ratio, fine2_ratio, fine3_ratio)	Total penalty burden relative to installment

Clustering Framework

Clustering is integrated into the modeling framework as a behavioral segmentation layer that groups customers based on early repayment patterns [1], [2]. The engineered features—such as delay statistics, acceptable ratios, and payment completeness—are used as inputs to unsupervised models to capture latent similarities in repayment behavior. Two approaches are employed: K-Means, a centroid-based method, and Gaussian Mixture Model (GMM), a probabilistic model. Prior studies highlight K-Means' effectiveness in segmenting heterogeneous borrower populations due to its interpretability, scalability, and stability on large transactional datasets [16], [17], [18], while GMM is well-suited for modeling overlapping and multi-modal borrower behaviors through soft cluster membership and uncertainty representation [19], [20]. Both methods are applied to the same feature space to generate alternative behavioral groupings.

The resulting cluster assignments are then converted into additional variables for the classification stage. For each contract, the assigned cluster (or most likely cluster in GMM) serves as a compact behavioral label summarizing early repayment behavior, which is combined with operational variables and engineered features as inputs for the default prediction models.

Classification Framework

The prediction of installment default is performed using three supervised learning models with complementary characteristics. Logistic Regression serves as the interpretable baseline model, allowing for the examination of linear relationships between features and default risk. Random Forest introduces non-linear decision boundaries and provides robustness through ensemble learning. LightGBM, a gradient-boosted decision tree model optimized for tabular data, offers high predictive capacity and efficient computation.

Each classifier is trained on the full set of operational variables, engineered behavioral features, and the K-Means model's cluster assignments. The models are evaluated using accuracy, precision, recall, F1 score, and AUC to provide a comprehensive assessment of performance. For Random Forest and LightGBM, feature importance scores are extracted to facilitate behavioral interpretation and cross-company comparison.

Model Transfer and Fine-Tuning Architecture

A central component of the framework is the evaluation of model transferability across companies. After training the full model on Company A's dataset, it is applied directly to Companies B and C without any retraining. This step measures the degree to which behavioral structures learned from one firm generalize to others that may differ in policy, customer demographics, or regional characteristics.

To further investigate adaptation, the models are then fine-tuned using the smaller datasets from Companies B and C. This enables us to investigate whether introducing a limited amount of local data can mitigate performance degradation. The fine-tuning procedure also provides insight into how feature importance shifts across companies and whether model behavior aligns or diverges when exposed to new environments.

Analysis Method

The analysis starts by training and validating classification models on Company A's dataset, chosen as the primary reference due to its long historical span and large sample size. Baseline performance is evaluated using accuracy, precision, recall, F1 score, and AUC on held-out test sets, with behavioral cluster labels incorporated alongside operational variables.

The trained models are then applied to Companies B and C to assess cross-company transferability under different operational conditions. Performance changes are measured using the same metrics, followed by fine-tuning with each company's local data to evaluate performance recovery. Feature importance from tree-based models is subsequently analyzed to identify shifts in behavioral emphasis, supporting the assessment of the framework's portability and limitations.

3. Results and Discussion

3.1. Clustering Results

Both K-Means and Gaussian Mixture Model (GMM) were evaluated for behavioral segmentation of early payment patterns

Table **3**. The K-Means algorithm achieved a Silhouette coefficient of 0.50, a Calinski–Harabasz score of 23,218.86, and a Davies–Bouldin index of 1.29, outperforming the GMM model across all criteria. These results indicate that K-Means formed more compact and well-separated behavioral clusters, while GMM produced more overlapping boundaries due to its probabilistic nature. Consequently, K-Means was selected as the preferred clustering algorithm for subsequent classification and interpretability analysis.

Table 3. Clustering results comparison

No	Model	Clusters	Silhouette	Calinski-Harabasz	Davies-Bouldin
1	K-Means	3	0.5004	23,218.8649	1.2913
2	GMM	3	0.1747	11,188.8864	2.3349

3.2. Classification Performance

Classification Results

The comparative results show how the three algorithms complement one another Table 4. Logistic Regression produced the highest precision (0.987) but a lower recall (0.853), indicating a conservative classifier that avoids false positives. Random Forest improved recall to 0.867 while maintaining a similar accuracy level of 0.977, and the K-Means cluster feature emerged as the second most influential predictor, reinforcing the value of the behavioral segmentation step. LightGBM achieved the strongest AUC of 0.985 and the highest recall of 0.912, capturing the largest share of true defaulters, although this came with a reduction in precision to 0.863. These results suggest that gradient-boosting models are better able to capture the complex repayment patterns that linear and bagging models only partially account for. Taken together, the results demonstrate that the hierarchical framework, which integrates feature engineering, behavioral clustering, and ensemble learning, is highly effective for early default prediction in fiduciary motorcycle-leasing data

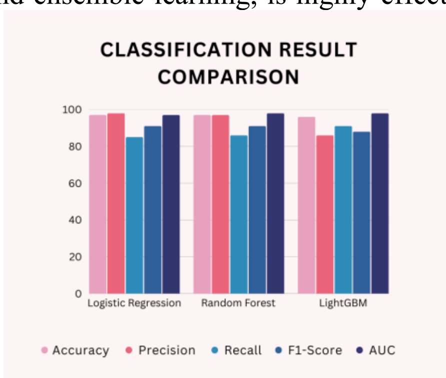

Figure **2**.

Table 4. Classification results comparison.

Model	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	0.9771	0.9865	0.8526	0.9147	0.9777
Random Forest	0.9772	0.9711	0.8674	0.9163	0.9819
LightGBM	0.9665	0.8633	0.9116	0.8868	0.9853

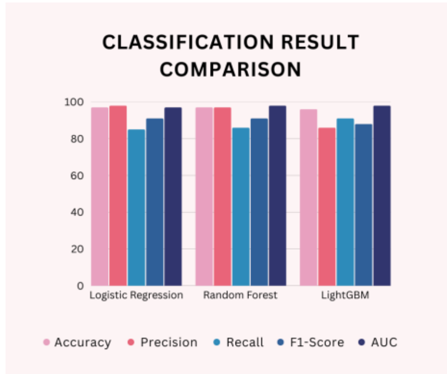

Figure 2. Classification results comparison.

Feature Importance Results

The feature-ranking results from Random Forest and LightGBM indicate that both models identify early-repayment behavior as the dominant driver of default prediction, but with differing priority structures Table **5**. Random Forest places the highest importance on std_delay_days, kmeans_cluster, and first_delay_trend, indicating that payment consistency, behavioral segmentation, and improvement or deterioration across installments are the strongest behavioral signals. LightGBM, in contrast, prioritizes payment_to_due_ratio, total_payments, and dp_to_price_ratio, emphasizing overall repayment completeness, early cashflow strength, and the customer's initial financial capacity. Secondary features, such as total fines, average delays, and fine ratios, contribute modestly across both models, while cluster assignment remains influential, even in gradient-boosted settings. Overall, the rankings suggest that repayment stability, behavioral grouping, and early cash flow performance collectively form the core determinants of default risk across both modeling approaches.

Table 5. Feature importance comparison.

Ranking	Random Forest		LightGBM	
	Feature	Importance	Feature	Importance
1	std_delay_days	0.260993	payment_to_due_ratio	1,703
2	kmeans_cluster	0.199051	total_payments	1,699
3	first_delay_trend	0.128745	dp_to_price_ratio	1,618
4	avg_delay_days	0.105767	total_fines	947
5	payment_to_due_ratio	0.103337	first_delay_trend	849
6	dp_to_price_ratio	0.085221	std_delay_days	812
7	total_payments	0.056627	avg_delay_days	630
8	total_fines	0.027407	avg_fine_ratio	532
9	total_fine_ratio	0.012981	early_payment_ratio	231
10	avg_fine_ratio	0.012194	total_fine_ratio	172
11			kmeans_cluster	137

3.3. Classification Performance

Model Performance without Fine-Tuning for Company B data

When the Company A models were directly applied to Company B data without retraining, all classifiers experienced AUC reductions greater than 19% and accuracy declines exceeding 15% Table 6. Such performance degradation surpasses the 10 percent non-generalizability threshold, indicating that the Phase 1 models had over-specialized to the behavioral patterns and contractual characteristics unique to Company A. This suggests that the model captured institution-specific repayment rules rather than universal behavioral signals. In other words, the Phase 1 pipeline exhibited overfitting at the organizational level.

Table 6. Comparison model performance of Company A and Company B without fine-tuning

Model	Company A AUC	Company B AUC	Δ Change	Company A Accuracy	Company B Accuracy	Δ Change
Logistic Regression	0.9777	0.7314	−25.2 %	0.9771	0.7209	−26.2 %
Random Forest	0.9819	0.7732	−21.3 %	0.9772	0.777	−20.5 %
LightGBM	0.9853	0.7929	−19.5 %	0.9665	0.8134	−15.8 %

Model Performance without Fine-Tuning for Company C data

The cross-regional evaluation on Company C yielded a notably stable performance (Accuracy ≈ 0.91, F1 ≈ 0.79) despite no retraining Table 7. This outcome demonstrates that the proposed clustering–classification pipeline captures a core set of behavioral and structural repayment dynamics that remain invariant across locations governed by similar fiduciary frameworks. The earlier generalization failure in Company B is thus attributed not only to model overfitting, but also to institutional policy drift. Conversely, the results from Company C indicate that when regulatory and operational conditions are reasonably consistent, the model can be successfully transferred across regions, which is a crucial step toward scalable, nationwide behavioral credit-risk modeling.

Table 7. Comparison Model Performance Company A and Company C without Fine-Tuning.

Model	Company A (Batam)	Company C (Jakarta, no fine-tuning)	Δ Accuracy	Δ F1 Score	Δ AUC
Logistic Regression	Acc 0.977 | F1 0.9147 | AUC 0.9777	Acc 0.9029 | F1 0.7699 | AUC 0.8934	−7.4 %	−15.9 %	−8.4 %
Random Forest	Acc 0.977 | F1 0.9163 | AUC 0.9819	Acc 0.9156 | F1 0.7929 | AUC 0.9018	−6.1 %	−12.3 %	−8.0 %
LightGBM	Acc 0.9665 | F1 0.8868 | AUC 0.9853	Acc 0.8158 | F1 0.6710 | AUC 0.8906	−15.1 %	−21.6 %	−9.5 %

3.4. Model Performance with Fine-Tuning

Model Performance with Fine-Tuning for Company B data

After fine-tuning using Company B data, all models improved, especially Random Forest (F1 +22%) and LightGBM (AUC +0.07), proving that domain adaptation restores model generalization Table 8.

Table 8. Comparison Model Performance Company A and Company B with Fine-Tuning.

Model	Company A (Original)	Company B (Before Fine-Tuning)	Company B (After Fine-Tuning)
Logistic Regression	Acc 0.9771 | F1 0.9147 | AUC 0.9777	Acc 0.7209 | F1 0.5132 | AUC 0.7314	Acc 0.7929 | F1 0.5980 | AUC 0.8136
Random Forest	Acc 0.9772 | F1 0.9163 | AUC 0.9819	Acc 0.7770 | F1 0.4425 | AUC 0.7732	Acc 0.8283 | F1 0.6634 | AUC 0.8353
LightGBM	Acc 0.9665 | F1 0.8868 | AUC 0.9853	Acc 0.8134 | F1 0.6306 | AUC 0.7929	Acc 0.8485 | F1 0.6809 | AUC 0.8651

Feature Importance Comparison for Company B Fine-Tune

The feature-importance results for Random Forest show a clear shift in predictive structure when the model trained on Company A is applied to Company B

Table **9**. Behavioral predictors such as std_delay_days, first_delay_trend, and kmeans_cluster lose much of their influence, while financial aggregation metrics, including total_fines, total_payments, and payment_to_due_ratio, emerge as the dominant drivers. This shift suggests fundamental differences in organizational policy between the two companies: default outcomes in Company A are strongly behavioral, driven by lateness patterns and repayment consistency, whereas in Company B, default appears to be shaped more by structural or policy-based mechanisms such as fine accumulation or internal rules governing payment completeness. As a result, a model trained on Company A implicitly assumes that behavioral delay patterns signal default risk, but this assumption does not transfer well to Company B, where accounting or policy thresholds exert a stronger influence on default classification.

Table 9. Feature Importance Random Forest Company B.

Rank	Company A Feature	Company B Feature (after fine-tuning)	Change
1	std_delay_days	payment_to_due_ratio	Behavioral to Structural shift
2	kmeans_cluster	total_payments	Behavioral segmentation lost impact
3	first_delay_trend	total_fines	Delay pattern replaced by penalty volume
4	avg_delay_days	total_fine_ratio	Consistency replaced by fine accumulation
5	payment_to_due_ratio	avg_fine_ratio	Same metric rises as dominant factor
10	avg_fine_ratio	dp_to_price_ratio (0%)	Risk appetite became negligible

The LightGBM results reinforce the observation that behavioral segmentation learned from Company A does not transfer effectively to Company B, as the importance of kmeans_cluster declines from a meaningful contributor to nearly zero Table 10. In its place, structural payment features, particularly payment_to_due_ratio, total_fines, and total_payments, dominate the predictive hierarchy. This pattern suggests that customers in Company B exhibit repayment behavior shaped by different underlying dynamics, likely influenced by stricter scheduling rules, more centralized penalty enforcement, or automated billing processes that reduce behavioral variability. As a result, the LightGBM findings are consistent with the Random Forest results in showing that default outcomes in Company B are driven more by structural characteristics than by individual behavioral traits. This further illustrates the limited portability of behavioral models across companies without some form of domain adaptation.

Table 10. Feature Importance LightGBM Company B.

Ran k	Company A Feature	Company B Feature (after fine-tuning)	Shift
1	payment_to_due_ratio	payment_to_due_ratio	Stable top feature
2	total_payments	total_payments	Stable structural predictor
3	dp_to_price_ratio	0.0 (irrelevant)	Lost significance
4	total_fines	3rd place (↑)	Increased influence
5	first_delay_trend	5th (≈ stable)	Still meaningful
6–7	std_delay_days, avg_delay_days	mid-table (~equal)	Consistent but reduced weight

Model Performance with Fine-Tuning for Company C data

Fine-tuning on Company C did not yield measurable performance gains; LightGBM even deteriorated Table 11. This outcome is consistent with small-sample fine-tuning dynamics: with only ≈ 1,100 records, model variance dominates added information. Random Forest maintained moderate stability (F1 ≈ 0.76), confirming that ensemble bagging remains more robust under data scarcity, whereas LightGBM's gradient boosting over-fitted to the majority class.

Table 11. Comparison Model Performance Company A and Company C with Fine-Tuning.

Model	Company A (Original)	Company C (Before Fine-Tuning)	Company C (After Fine-Tuning)
Logistic Regression	Acc 0.977 | F1 0.9147 | AUC 0.9777	Acc 0.9029 | F1 0.7699 | AUC 0.8934	Acc 0.8869 | F1 0.7059 | AUC 0.8618
Random Forest	Acc 0.9772 | F1 0.9163 | AUC 0.9819	Acc 0.9156 | F1 0.7929 | AUC 0.9018	Acc 0.9095 | F1 0.7561 | AUC 0.8967
LightGBM	Acc 0.9665 | F1 0.8868 | AUC 0.9853	Acc 0.8158 | F1 0.6710 | AUC 0.8906	Acc 0.7692 | F1 0.0000 | AUC 0.9035

Feature Importance Comparison for Company C Fine Tune

Jakarta data emphasize payment-schedule structure (payment_to_due_ratio, total_payments) more than behavioral noise Table 12. Customers are likely to follow stricter automated schedules, thereby reducing variability in delay-based signals.

Table 12. Comparison Model Performance Company A and Company C with Fine-Tuning.

Rank	Company A Model	Company C Fine-tuned	Comment
1	std_delay_days (0.26)	payment_to_due_ratio (0.28)	Shift from behavioral consistency to structural compliance.
2	kmeans_cluster (0.20)	total_payments (0.17)	Cluster lost dominance; repayment volume is more relevant.
3	first_delay_trend (0.13)	kmeans_cluster (0.14)	Cluster recovered some weight, showing partial behavioral transfer.
4–6	Delay & fine features	Fines & ratios ≈ 0.05 each	A greater balance among secondary features helps model spreads of attention.
10	avg_fine_ratio (0.01)	dp_to_price_ratio (0.04)	Down-payment risk relevance rises again (regional purchasing style).

LightGBM over-focuses on one dominant global feature (payment_to_due_ratio), ignoring subtle variance, a classic symptom of data scarcity + imbalance. With only ~200 defaults, the gradient boosting trees saturate early, hence precision = 0 (Table 13).

Table 13. Feature Importance LightGBM Company C.

Rank	Company A Model	Company C Fine-tuned	Change
1	payment_to_due_ratio (1,703)	payment_to_due_ratio (2,944)	Still the universal top predictor.
2–4	total_payments, dp_to_price_ratio, total_fines	avg_fine_ratio, total_payments, total_fines	Fine ratio rises; structural volume falls slightly.
6–8	Delay metrics	Delay metrics (mid-tier)	Temporal behavior still matters but weaker.
10–11	total_fine_ratio, kmeans_cluster (≈100–130)	0	Cluster behavior makes no contribution to regional homogeneity.

The performance comparison reveals clear differences in model transferability across the three companies (Table 14), with Figure 3 illustrating the before–and–after fine-tuning effects for Companies B and C. The baseline model trained on Company A's 13-year dataset showed very high internal validity (accuracy and AUC > 0.97), reflecting stable repayment behavior within a mature institutional setting. However, direct transfer to Company B—operating in the same region but under different managerial policies—resulted in a substantial performance drop, with F1 scores decreasing by 25–40%, driven by policy drift in penalty rules, tolerance windows, and enforcement practices. Fine-tuning using Company B's 3,000 local records recovered approximately 15–20% of the lost F1 score, indicating partial realignment to local policy semantics.

In contrast, transfer to Company C across regions caused only a modest performance decline (~10% F1 reduction), suggesting similar fiduciary behavior and contractual structures between Jakarta and Batam. Nonetheless, fine-tuning on Company C's limited dataset (~1,000 records) produced unstable or negative effects, particularly for LightGBM, due to increased variance and overfitting.

Overall, these results show that model transfer is viable only when operational definitions and penalty structures are aligned, fine-tuning is effective with sufficiently large target datasets (≥3,000 records), and models trained on source firms should be treated as behavioral priors rather than decision tools without proper local calibration, to avoid embedding policy bias.

Table 14. Comparison Model Performance Company A, B, C.

Model/Company	Accuracy	Precision	Recall	F1	AUC
Company A (Original)	0.97–0.98	0.97–0.99	0.85–0.87	0.91–0.92	0.97–0.98
Company B (No Fine-Tune)	0.72 – 0.81	0.50 – 0.70	0.32 – 0.57	0.44 – 0.63	0.73 – 0.79
Company B (After Fine-Tune)	0.79 – 0.85	0.65 – 0.82	0.55 – 0.61	0.60 – 0.68	0.81 – 0.86
Company C (No Fine-Tune)	0.82 – 0.92	0.57 – 0.91	0.70 – 0.82	0.67 – 0.79	0.89 – 0.90
Company C (After Fine-Tune)	0.77 – 0.91	0.00 – 1.00	0.00 – 0.61	0.00 – 0.76	0.86 – 0.90

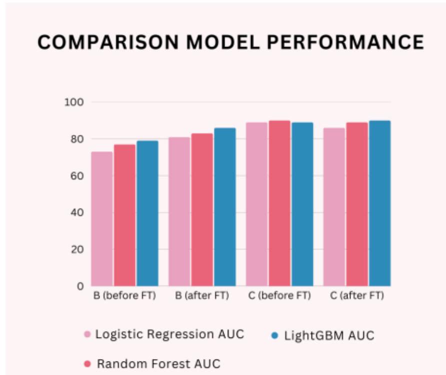

Figure 3. Comparison Model Performance.

4. Conclusion

This study demonstrates a clear hierarchy of model transferability in fiduciary credit prediction. The baseline model trained on Company A achieved very high internal performance, with accuracy and AUC above 0.97, indicating that the proposed hierarchical framework effectively captures stable repayment behavior in a mature institutional setting. However, when applied directly to Company B, performance declined sharply, with AUC reductions exceeding 19% and F1 scores dropping by approximately 25–40%, confirming substantial organizational policy drift. After fine-tuning with around 3,000 local records from Company B, model performance recovered by roughly 15–20% in F1, showing that moderate domain adaptation can meaningfully realign behavioral indicators with local operational rules.

In contrast, cross-regional transfer to Company C was considerably more stable: without fine-tuning, accuracy remained around 0.91 and AUC around 0.89–0.90, representing only an 8–10% decline, suggesting that similar fiduciary structures across regions support natural generalization. Nevertheless, fine-tuning on Company C's small dataset (≈1,000 records) led to unstable results, particularly for LightGBM, where F1 deteriorated to zero, indicating that sample scarcity increases model variance and overfitting risk.

Overall, pre-trained behavioral models can serve as useful priors for new leasing firms, but they must be recalibrated with sufficient local data before operational deployment; otherwise, they risk embedding source-company policy bias into credit decisions.

5. References

[1] P. Prameswari and T. F. Alfiany, "Juridical Analysis of Over Credit in Motor Vehicle Credit Agreement with Fiduciary Guarantee in View of Undang-Undang Nomor 42 of 1999 on Fiduciary Guarantee," Journal of Law, Politic and Humanities, vol. 4, no. 6, pp. 2390–2401, Oct. 2024, doi: 10.38035/JLPH.V4I6.700.

[2] S. J. Lesmana and R. Ristiyana, "Legal Counseling Regarding the Limits of Debt Collector and Leasing Authority in the Execution of Motor Vehicles," Help: Journal of Community Service, vol. 1, no. 4, pp. 246–255, Mar. 2025, doi: 10.62569/HJCS.V1I4.135.

[3] C. Tuccillo, "Termination for non-performance in transactional leasing: the latest interventions, fair compensation, blockchain and residential real estate leasing. Reflections from Cass., 28 January 2021, no. 2061 and Cass., 14 October 2021, no. 28023," European Journal of Privacy Law and Technologies, vol. 2023, no. 1, pp. 174–207, 2023, doi: 10.57230/EJPLT223CT.

[4] H. Adjie and E. Sinaga, "Implementation of Motorcycle Rental Agreement," Budapest International Research and Critics Institute-Journal (BIRCI-Journal), vol. 4, no. 4, pp. 9353–9362, Nov. 2021, doi: 10.33258/BIRCI.V4I4.2947.

[5] N. F. Hidayanti, A. Hulaimi, and Z. Ariani, “Vehicle Trade of Defaulted Finance Lease Companies in Mataram City Sharia Economic Perspective,” Indonesian Interdisciplinary Journal of Sharia Economics (IIJSE), vol. 7, no. 3, pp. 7882–7896, Dec. 2024, doi: 10.31538/IIJSE.V7I3.5693.

[6] R. Fujinuma and Y. Asahi, “Default Factors in Motorcycle Sales in Developing Countries,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 13305 LNCS, pp. 324–336, 2022, doi: 10.1007/978-3-031-06424-1_24.

[7] O. Triki and F. Abid, “Convertible lease risk spread modeling with correlation,” Decisions in Economics and Finance 2024 48:1, vol. 48, no. 1, pp. 747–773, Oct. 2024, doi: 10.1007/S10203-024-00490-W.

[8] S. J. Francis, H. G. C. Hansamali, M. Ganeshamoorthy, and R. Abarna, “Exploring Credit Risk Management and Financial Performance: Evidence from People’s Leasing and Finance PLC,” Wayamba Journal of Management, vol. 15, no. 2, pp. 86–100, May 2025, doi: 10.4038/WJM.V15I2.7630.

[9] Y. C. Chang, K. H. Chang, and W. T. Chen, “Vehicle leasing credit risk assessment modeling by applying extended logistic regression,” Journal of Intelligent and Fuzzy Systems, vol. 45, no. 4, pp. 5211–5222, Oct. 2023, doi: 10.3233/JIFS-231344.

[10] H. Li et al., “A novel method for credit scoring based on feature transformation and ensemble model,” PeerJ Comput Sci, vol. 7, p. e579, 2021, doi: 10.7717/PEERJ-CS.579.

[11] L. O. Hjelkrem and P. E. de Lange, “Explaining Deep Learning Models for Credit Scoring with SHAP: A Case Study Using Open Banking Data,” Journal of Risk and Financial Management 2023, Vol. 16, Page 221, vol. 16, no. 4, p. 221, Apr. 2023, doi: 10.3390/JRFM16040221.

[12] V. Chang, S. Sivakulasingam, H. Wang, S. T. Wong, M. A. Ganatra, and J. Luo, “Credit Risk Prediction Using Machine Learning and Deep Learning: A Study on Credit Card Customers,” Risks 2024, Vol. 12, Page 174, vol. 12, no. 11, p. 174, Nov. 2024, doi: 10.3390/RISKS12110174.

[13] X. Zhang et al., “Data-Driven Loan Default Prediction: A Machine Learning Approach for Enhancing Business Process Management,” Systems 2025, Vol. 13, Page 581, vol. 13, no. 7, p. 581, Jul. 2025, doi: 10.3390/SYSTEMS13070581.

[14] H. Suryanto, A. Mahidadia, M. Bain, C. Guan, and A. Guan, “Credit Risk Modeling Using Transfer Learning and Domain Adaptation,” Front Artif Intell, vol. 5, p. 868232, May 2022, doi: 10.3389/FRAI.2022.868232/BIBTEX.

[15] F. Xu and R. Zhang, “Explainable Domain Adaptation Learning Framework for Credit Scoring in Internet Finance Through Adversarial Transfer Learning and Ensemble Fusion Model,” Mathematics 2025, Vol. 13, Page 1045, vol. 13, no. 7, p. 1045, Mar. 2025, doi: 10.3390/MATH13071045.

[16] Y. Suh, “Machine learning based customer churn prediction in home appliance rental business,” Journal of Big Data 2023 10:1, vol. 10, no. 1, pp. 41-, Apr. 2023, doi: 10.1186/s40537-023-00721-8.

[17] M. Ay, L. Özbakır, S. Kulluk, B. Gülmez, G. Öztürk, and S. Özer, “FC-Kmeans: Fixed-centered K-means algorithm,” Expert Syst. Appl., vol. 211, p. 118656, Jan. 2023, doi: 10.1016/j.eswa.2022.118656.

[18] A. Kozina et al., “The default of leasing contracts prediction using machine learning,” Procedia Comput. Sci., vol. 225, no. 10, pp. 424–433, Jan. 2023, doi: 10.1016/j.procs.2023.10.027.

[19] D. Ghosh and S. Chatterjee, “Journal of Strategic Marketing Customer segmentation for stigmatized products using point-of-sale big-data and Gaussian-Mixture-Model: an application in female-hygiene market Customer segmentation for stigmatized products using point-of-sale big-data and…,” 2025, doi: 10.1080/0965254X.2025.2562444.

[20] E. Y. Jin, W. Jiang, and Z. Zheng, “A novel method for testing adverse selection with IoT data: Evidence from China’s auto insurance market,” Decis. Support Syst., vol. 200, p. 114576, Jan. 2026, doi: 10.1016/j.dss.2025.114576.