Sensitivity Analysis: Improving Inclusive Credit Scoring Algorithm through Feature Weight and Penalty-based Approach

Reni Sulastri*, Aaron Yi Ding*, and Marijn Janssen*
*Delft University of Technology
Delft, The Netherlands
{r.sulastri, aaron.ding, M.F.W.H.A.Janssen}@tudelft.nl

2025 Eleventh International Conference on eDemocracy & eGovernment (ICEDEG) | 979-8-3315-9809-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICEDEG65568.2025.11081606

***Abstract*—Recent advancements in artificial intelligence (AI) and machine learning (ML) have enhanced credit scoring systems by improving prediction accuracy, fairness, and transparency. However, limited attention has been given to financial inclusion, particularly how algorithmic adjustments can improve access for underserved borrowers. This study addresses this gap by evaluating how parameter tuning in credit scoring models impacts borrower classifications and promotes inclusivity in online lending systems. Using a micro-enterprise loan dataset from Indonesia, we simulate three approaches: Feature Weight Adjustment, Penalty-Based Models, and a novel Hybrid Feature Penalty Tuning (HFPT) that combines both. Each method is evaluated using custom metrics that measure changes in borrower distribution, inclusion gains, and reclassification risks. Results show that HFPT consistently improves inclusive outcomes without increasing high-risk concentrations. These findings underscore the importance of regulatory oversight and transparent algorithm design, as small changes can easily influence the outcome of models' parameters. For governments, adopting clear guidelines for implementing and evaluating such models is crucial to ensure responsible lending practices. The flexibility of algorithm design, as manifested in this work, offers a pathway for policymakers to balance inclusivity with risk management in credit scoring systems.**

***Index Terms*—financial inclusion, credit scoring, machine learning, sensitivity analysis**

I. Introduction

Artificial intelligence (AI) and machine learning (ML) have reshaped credit scoring methodologies, especially in prediction accuracy [1]–[3]. Despite these advancements, a critical gap persists: ensuring that these systems facilitate financial inclusion, particularly for marginalized micro-enterprises that have important roles in economic growth in developing regions [4], [5]. This research addresses this issue by exploring data-driven methodologies to improve inclusion in lending systems using the micro-enterprises dataset in Indonesia. In Indonesia, informal sectors, including micro-enterprises, represent approximately 97% of the workforce and contribute 61% to the national GDP [6]. Despite the critical contribution, their access to credit remains low, with credit-to-GDP ratios of 7%, identified as one of the lowest in the world [7].

979-8-3315-9809-9/25/$31.00 © 2025 IEEE

Inclusion is not only about access, but also ensuring equal benefits for all segments of society [8]. Technological advancement improved financial inclusion by lowering the cost of reaching underserved populations [9]. For example, improvement in credit scoring algorithms was achieved by introducing profit scoring [10], [11] and poverty scoring [12]. Exclusion from formal systems limits access to resources for growth and sustainability [6].

While prior research has predominantly concentrated on *fairness* and *nondiscrimination* within machine learning frameworks, there remains a need for inclusion studies. This study fills this void by demonstrating how fine-tuning the model parameters provides more options to improve inclusions by improving credit opportunities. The primary research question is: "*How can adjustments of parameters in credit scoring models improve financial inclusion for micro-enterprises in online lending systems*?". We hypothesize that such adjustments, particularly in feature weights and penalty configurations, can increase inclusion by reclassifying borrowers into lower-risk categories without significantly increasing high-risk concentrations.

This study contributes to the broader conversation on how algorithmic systems can support inclusive digital transformation. In contexts where access to credit remains highly unequal, small adjustments in model parameters can have significant implications for who gets access and who remains excluded. By making these impacts visible and measurable, the study provides a foundation for developing socially responsive scoring models. These insights are relevant to the design of digital infrastructures that are fair, accessible, and accountable.

This paper is structured as follows: Section 2 reviews the literature on financial inclusion in credit scoring; Section 3 details the empirical methodology; Section 4 presents the simulation results; and Section 5 concludes with key insights and recommendations.

II. Literature Review

AI has increasingly been used to enhance credit scoring systems by addressing challenges in data preparation and

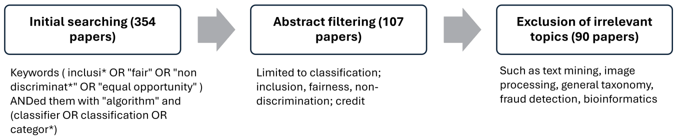

Fig. 1. Literature Review

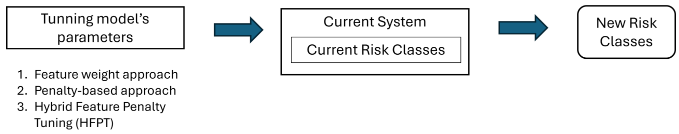

Fig. 2. Simulation Approach

model development. *In the modeling development stage*, AI enhances the accuracy of default prediction [3], [13], formulates individual credit risks [14], and determines interest rates [15], [16]. These improvements help refine risk assessments, allowing lenders to extend credit more confidently to a broader range of borrowers, including those without traditional credit histories. Additionally, AI-driven models can calculate profit [11], and assist non-expert lenders in making informed investment decisions [17]. *In the data preparation stage*, diverse datasets—such as social media activity, spending behavior, and even energy consumption—can help create more comprehensive profiles of individuals who might be excluded from traditional credit assessments. By integrating AI with big data analytics [18], lending institutions can collect and analyze nontraditional data to improve the inclusion of their credit scoring models. These models can better capture the financial behaviors of underserved populations. Moreover, *in the output analytics stage*, AI models assist in monitoring individual payment behavior [19]; monitoring credit performance [20]; detecting industrial risks based on individual behavior [21] and financial risk [22]; and detecting fraudulent activities [23].

However, despite rapid technological advancements in financial lending systems, the issue of inclusion—ensuring that individuals previously excluded from access to credit are brought into formal financial services—remains underexplored. Inclusion in financial lending is not simply about extending credit to more people; it involves identifying challenges preventing individuals from accessing financial services. These challenges include unconvincing profiles, geographical isolation, low financial literacy, and a lack of trust in formal systems [24]. While technological advancement holds promise, there is limited research on how such technologies can promote inclusion rather than merely replicate traditional exclusionary practices [25], [26].

The literature review in Fig. 1 examines inclusion, fairness, and non-discrimination as key values for equitable access in credit lending. A keyword search ((inclusi OR ”fair” OR ”non discriminat” OR ”equal opportunity”) AND (”algorithm” AND (classifier OR classification OR categor*))) was conducted on academic databases such as scopus, focusing on English-language publications from 2009 to 2023 in computer science and financial lending. Of the 354 papers identified, 107 were selected for abstract filtering, and 90 were included after excluding irrelevant topics like sentiment analysis and bioinformatics.

61% of the reviewed papers focus on fairness, while only 3% explicitly address inclusiveness, highlighting a significant gap in the literature. Fairness ensures equitable treatment, whereas inclusion aims to expand access for underrepresented groups. Despite the limited focus on inclusiveness, key studies provide valuable insights. [26] investigate the use of machine learning to predict bank account ownership in Africa, emphasizing education's role in enhancing financial inclusion. However, this research centers explicitly on general financial inclusion rather than credit inclusion. [25] review alternative credit scoring, advocating for context-specific data to include ‘credit invisible’ populations in developing regions such as, South America, Sub-Saharan Africa, and Southeast Asia. Lastly, [27] propose regulatory updates to address biases in U.S. credit lending, linking algorithmic fairness and discrimination laws.

Approximately 7% of papers address inclusiveness within fairness, such as integrating fairness constraints during model training [28] and addressing class imbalances to avoid disproportionate misclassification of underserved groups [29], [30]. 61% of the reviewed literature focuses on fairness, with an additional 11% addressing related values such as non-discrimination, explainability, transparency, accountability, robustness, and privacy. Privacy concerns also play a crucial role in classification models. For instance, [31] integrate differential privacy with fairness mechanisms to protect individual data while ensuring equitable outcomes, a key consideration in credit scoring. Similarly, [32] use data encryption in SVM training to address both privacy and fairness, ensuring the security of sensitive information.

The literature review highlights AI and machine learning in credit scoring and identifies a gap in addressing inclusiveness for underserved populations. Although fairness and alternative data have been explored, the role of algorithmic sensitivity in promoting inclusion remains under-examined. The next section presents the methodology used to address this gap.

III. Research Methodology

This study investigates how sensitivity in model parameter tuning impacts risk classification and inclusivity. Fig. 2 illustrates the approach, which applies three methods—*feature weight adjustment*, *penalty-based*, and *hybrid feature penalty tuning (HFPT)*—to adjust parameters, resulting in new borrower classifications that reflect potential loan approvals.

A. Data Collection

The dataset from Indonesian lending institutions contains anonymized borrower profiles and complete labeled classifications, requiring no imputation. Originally planned for 2021, the data was updated to 2022 to avoid COVID-19-related anomalies. It focuses on micro and ultra-micro enterprises, which dominate the workforce but remain underserved. Simulations were used to systematically vary model parameters, such as feature weights and penalties, since their individual effects cannot be isolated in real-world systems. This setup allows controlled testing of how tuning affects borrower classifications and inclusion outcomes

B. Approaches

This study investigates borrower classification methods by exploring extensive model variations, aiming to identify patterns that enhance financial inclusion. Three primary methods were examined: Feature Weight Adjustment, Penalty-Based Models, and HFPT. *Feature Weight Adjustment* modifies the importance of key features, particularly delayed debt payment (`OVER_TIME`) and interest payment (`OVER_INT`). *Penalty-based models* applies penalties based on predefined rules.

HFPT, the core method, integrates both approaches to enable broader exploration. A total of 629 model variations were tested, exploring various combinations of penalty types, penalty values, feature configurations, and reduction factors. HFPT leverages 17 *penalty types*, ranging from narrowly focused configurations to broader setups. *Penalty values* were tested at four levels (500-5000) to understand the impact of varying penalty intensities. *Feature reductions* were applied at factors of 0.1, 0.5, and 1, modifying the weights of `OVER_TIME` and `OVER_INT`. HFPT's novelty lies in its ability to uncover patterns unattainable by either method alone, offering a comprehensive and flexible approach to balancing inclusivity with risk management across diverse scenarios.

C. Measurement Metrics

The experiment assumes that a borrower's risk classification reflects future creditworthiness: lower-risk classification implies a higher likelihood of credit approval, while a higher-risk classification indicates greater barriers. The table below summarizes the metrics used, including descriptions and formulas.

Metric 1 compares borrower counts across each risk category in the baseline and adjusted models, highlighting any inclusivity-related shifts. *Metrics 2* and *3* measure reclassifications into lower and higher risk categories to capture improvements and unintended risk escalations. The Inclusivity Ratio *(Metric 4)* quantifies the balance of reclassifications toward lower versus higher risk categories.

D. Experiment setup and Baseline model

The lending system under investigation classifies borrowers into five risk categories based on financial performance. This study evaluates how model tuning can alter these classifications to enhance inclusivity.

In Fig. 3, the x-axis represents risk categories, and the y-axis depicts borrower scores. Categories 0 and 4 show broader distributions, reflecting higher variability in borrower characteristics, while Categories 2 and 3 have narrower ranges, indicating greater consistency. Category 0 represents the lowest risk, and Category 5 the highest. This study assumes that a lower-risk classification increases the likelihood of future credit approval. Therefore, inclusiveness, in this context, refers to the model's ability to move more borrowers into lower-risk categories, enhancing their chances for future credit access.

TABLE I
INCLUSIVITY METRICS FOR DATA-DRIVEN SIMULATION

Metric	Description and Formula
1. Distribution Comparison	Observes borrower distribution across risk classes in the baseline and adjusted models. This helps understand concentration shifts following model adjustments.
2. Shift to Lower Risk Classes	Measures the number of borrowers reclassified into lower-risk categories in the adjusted model, indicating increased inclusivity. Formula: $$\text{Lower Shift} = \sum_{\text{all classes}} \left(1 \text{ if } R_{\text{adj}} < R_{\text{base}}\right)$$
3. Shift to Higher Risk Classes	Measures the number of borrowers initially classified as low-risk in the baseline model who are reclassified to higher-risk categories in the adjusted model. Formula: $$\text{Higher Shift} = \sum_{\text{all classes}} \left(1 \text{ if } R_{\text{adj}} > R_{\text{base}}\right)$$
4. Inclusivity Ratio	Calculates the ratio of lower-risk shifts to higher-risk shifts. Formula: $$\text{Inclusivity Ratio} = \frac{\text{Lower Shift}}{\text{Higher Shift}}$$

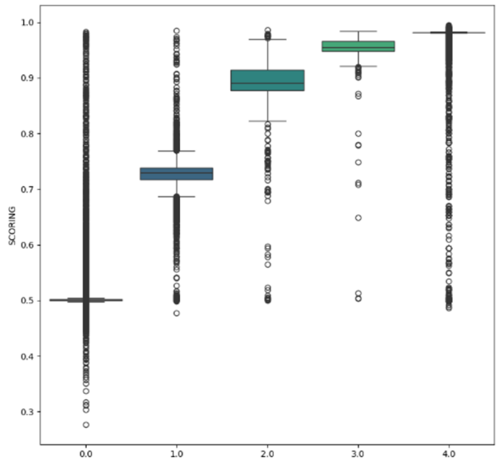

Fig. 3. Risk Classification of Borrowers from Selected Loan Samples

To test the hypothesis, a baseline model was constructed as a benchmark for evaluating the adjusted models. The model was validated to ensure reliability and provide a credible comparison. PyCaret, an open-source Python library, was used to streamline the machine-learning workflow, including model selection, training, and evaluation. Data preprocessing involved cleaning the dataset by addressing missing values with mean or mode imputation. A pipeline standardized numerical features and transformed categorical features using `one-hot encoding` or `LabelEncoder`.

Machine learning algorithms, including Gradient Boosting

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
gbc	Gradient Boosting Classifier	0.9974	0.9993	0.9974	0.9974	0.9974	0.9926	0.9926	222.7033
xgboost	Extreme Gradient Boosting	0.9973	0.9995	0.9973	0.9973	0.9973	0.9923	0.9923	179.1567
catboost	CatBoost Classifier	0.9970	0.9995	0.9970	0.9970	0.9970	0.9914	0.9914	142.6000
rf	Random Forest Classifier	0.9965	0.9980	0.9965	0.9965	0.9965	0.9900	0.9900	12.3733
dt	Decision Tree Classifier	0.9950	0.9952	0.9950	0.9950	0.9950	0.9858	0.9858	3.8067
lightgbm	Light Gradient Boosting Machine	0.9906	0.9920	0.9906	0.9922	0.9911	0.9735	0.9736	6.9400
et	Extra Trees Classifier	0.9897	0.9980	0.9897	0.9894	0.9895	0.9706	0.9706	13.9267
knn	K Neighbors Classifier	0.9653	0.9781	0.9653	0.9672	0.9661	0.9023	0.9025	61.8933
ada	Ada Boost Classifier	0.8858	0.9869	0.8858	0.9135	0.8887	0.6859	0.7037	10.2000
ridge	Ridge Classifier	0.8369	0.0000	0.8369	0.8852	0.8575	0.5903	0.5996	1.0133
lda	Linear Discriminant Analysis	0.8093	0.8959	0.8093	0.8913	0.8445	0.5363	0.5487	2.6433
dummy	Dummy Classifier	0.7914	0.5000	0.7914	0.6263	0.6992	0.0000	0.0000	1.0933
svm	SVM - Linear Kernel	0.6760	0.0000	0.6760	0.8610	0.7186	0.4298	0.4774	1.6700
nb	Naive Bayes	0.2549	0.9553	0.2549	0.9280	0.3126	0.1463	0.3065	2.5333
qda	Quadratic Discriminant Analysis	0.1144	0.5600	0.1144	0.9199	0.1310	0.0860	0.2125	1.1633

Fig. 4. Performance Comparison of Machine Learning Models for Baseline

Classifier (GBC), XGBoost, CatBoost, Random Forest, and Decision Tree, were tested using PyCaret's model comparison tools. Models were evaluated on metrics such as accuracy, AUC, recall, precision, F1 score, Kappa, and MCC. As shown in Fig. 4, GBC achieved the highest accuracy (0.9974) but required over two hours per iteration, making it impractical for large-scale analysis. XGBoost, with slightly lower accuracy (0.9973), required only 1–2 hours per iteration, making it more suitable for early experiments.

As experiments progressed, XGBoost's processing time became a limitation due to the complexity and number of scenarios. To address this, CatBoost was adopted as a more efficient alternative, delivering comparable performance (accuracy 0.9970, AUC 0.9995) with significantly faster training times (1–2 minutes per iteration vs. XGBoost's 1–2 hours). After switching to CatBoost, we re-ran all experiments conducted under XGBoost to ensure consistency and reproducibility of the results. This validation process showed that the transition to CatBoost did not introduce discrepancies, and the model continued to perform robustly.

IV. Empirical Simulations

A. Method 1: Feature Weight Adjustment

The effects of adjusting weights for `OVER_TIME` (delayed payments) and `OVER_INT` (remaining interest) were analyzed using Metrics 1–4. Each *reduction factor* reflects how much emphasis is reduced on these features, with lower factors increasing the likelihood of reclassification into lower-risk classes. The analysis of **Metric 1** shows that reducing `OVER_TIME` weights, particularly at factors of 0.1 and 0.25, significantly shifted borrowers from classes 3 and 4 to lower-risk categories. In contrast, adjustments to `OVER_INT` resulted in minimal reclassifications across all factors, highlighting `OVER_TIME` as a key variable for improving inclusivity.

Metrics 2 and 3 further highlight the sensitivity impact. For instance, at a reduction factor of 0.1, `OVER_TIME` adjustments moved over 48,127 borrowers (over 100,000 test data) to lower-risk categories, with minimal reclassification to higher-risk groups. By contrast, adjustments to `OVER_INT` alone

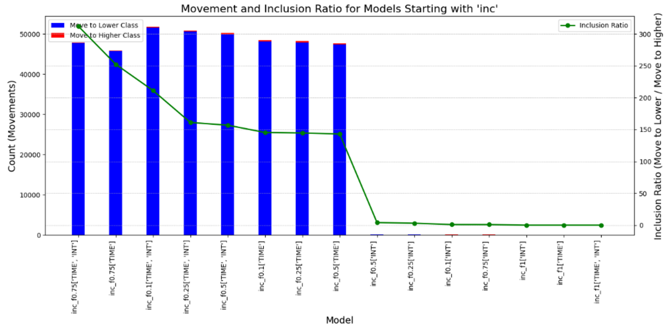

Fig. 5. Movement summary and inclusion ratio for feature weight approach

had limited impact across all reduction factors, underscoring the dominant influence of `OVER_TIME`. Models incorporating `OVER_TIME` adjustments achieved high inclusion ratios, exceeding 200 at reduction factors of 0.1 while maintaining minimal shifts to higher-risk classes.

This strong inclusivity performance is visually reinforced in Fig. 5 (**Metric 4**), where movements to lower-risk classes are significantly greater for scenarios involving `OVER_TIME`. Conversely, scenarios involving `OVER_INT` or reduction factors of 1 (no adjustment) resulted in negligible reclassifications, with inclusivity ratios close to 1.

The sensitivity analysis demonstrates that reducing the weight of `OVER_TIME` significantly shifts borrowers into lower-risk categories, achieving inclusion ratios exceeding 200 and improving access to credit for microenterprises. However, adjustments to `OVER_INT` show minimal impact, highlighting the dominant influence of `OVER_TIME`. Despite these improvements, the inclusion ratio alone does not fully capture the trade-offs in borrower distribution. For example, `inc_f0.75['TIME', 'INT']` yields the highest inclusion ratio (312.03) but results in higher borrower concentrations in risky classes compared to `inc_f0.1['TIME', 'INT']`, which has a lower inclusion ratio (211.22) but a more balanced risk distribution. This demonstrates the inherent challenge for policymakers to balance inclusivity with risk management based on their specific objectives.

The method's limitations become evident when addressing the complexity of choosing appropriate parameters, as small changes in feature weights can lead to drastically different outcomes. Additionally, relying solely on feature adjustments may not provide sufficient flexibility to accommodate diverse risk profiles across borrower segments. These gaps underline the need for Experiment 3 (HFPT), which integrates penalties and feature adjustments to offer a more comprehensive approach. HFPT is expected to provide policymakers with more apparent options by systematically exploring the interaction of penalties and feature.

B. Method 2: Penalty-based models

The penalty-based models were tested across various penalty types and penalty values. Experiments began by testing *penalty values* incrementally, starting from 1, 10, 20, and so on, to observe their effect on borrower classification. Initial trials revealed that penalty values below 100 had minimal impact on reclassification patterns. Significant changes were observed only when penalty values exceeded 100, prompting the selection of a focused range: 5000, 4000, 2000, 1500, 500, 100, and 1, to ensure the analysis captures both the upper and lower bounds of penalty intensity. Penalties are selectively applied based on *penalty types*, targeting only specific classes to preserve risk integrity.

Each configuration's penalty weight w_i for class C is determined by:

$$w_i = \begin{cases} \lambda, & \text{if } C \notin \text{excluded\_classes} \\ 1, & \text{if } C \in \text{excluded\_classes} \end{cases}$$

where λ represents the *penalty value*.

The experiments involve seventeen *penalty types* —such as `Pen0`, `Pen0-1`, `Pen0-2`, `Pen0-3`, `Pen4`, `Pen4-1`, `Pen4-2`, and `Pen4-3` —to target specific risk classes. Higher *penalty values* (e.g., 5000) apply a stronger push towards non-penalized classes, while lower values (e.g., 100) test the model's sensitivity to less aggressive interventions. This structured approach allows us to analyze how different penalty intensities and configurations impact borrower reclassification. We develop $17 \times 8 = 136$ model variations to evaluate the simulations.

The penalty-based approach reveals distinct patterns of borrower reclassification depending on the *penalty types* and *penalty values*. **Metric 1** , which examines the distribution of borrowers across risk classes, shows that high penalty values (e.g., 5000–1500) often result in concentrated borrower movements within penalized classes. For instance, `Pen4`-based configurations (e.g., `Pen4_VAL5000`) demonstrate marked retention of borrowers in Class 4. In contrast, `Pen0-2` and `Pen0-1` at high penalties show significant redistribution into lower-risk classes.

Metrics 2 and 3 further quantify movements to lower- and higher-risk classes, providing insights into reclassification dynamics. For example, `Pen0-2_VAL5000` reclassifies over 40,000 borrowers to lower-risk categories while limiting upward movements to only 337 borrowers. On the other hand, `Pen4_VAL5000` shifts over 80,000 borrowers to higher-risk categories, with minimal reclassification to lower-risk classes, leading to undesirable outcomes that contradict inclusivity goals. These trends confirm that penalties targeting lower-risk classes (e.g., `Pen0-2`, `Pen0-1`) are more effective in promoting financial inclusion than those focusing on higher-risk categories.

Metric 4, the inclusivity ratio, highlights the comparison between movements to lower- and higher-risk classes. A high inclusivity ratio, such as 120.00 in `Pen0-2_VAL5000`, indicates that penalties effectively prioritize downward reclassification while minimizing upward shifts. In contrast, `Pen4_VAL5000` achieves an inclusivity ratio close to 0, underscoring its failure to promote inclusion. Notably, while high inclusivity ratios suggest success, they must be evaluated alongside Metrics 1, 2, and 3 to ensure that the movements are meaningful and aligned with system stability.

In conclusion, the Penalty-Based approach investigates how applying penalties to specific risk classes influences borrower reclassification. While high-penalty configurations, such as `Pen0-1_VAL5000`, demonstrated strong redistributive effects by reclassifying over 59,295 borrowers into lower-risk categories with minimal upward movement, they often caused over-concentration in the penalized classes (e.g., Classes 0 and 1) while depleting other risk categories entirely. This lack of balance limits their applicability for broader inclusion goals. Moderate-penalty setups, like `Pen0-2_VAL500`, achieved a more balanced redistribution, reclassifying 40,348 borrowers while preserving diversity across risk classes. Narrow penalties targeting high-risk categories, such as `Pen4_VAL5000`, consistently failed to enhance inclusivity, trapping borrowers in high-risk classes and yielding inclusion ratios close to zero.

These results highlight two key challenges with penalty-based models: the difficulty in determining optimal penalty configurations and the tendency of high penalties to over-concentrate borrowers in specific classes. Furthermore, relying solely on penalties limits the model's ability to address nuanced borrower characteristics, often leaving inclusivity goals partially unmet. These limitations emphasize the need for a hybrid approach, like HFPT, which integrates penalty adjustments with feature modifications to address over-concentration issues and explore a broader range of solutions to improve inclusivity while maintaining system stability.

C. Method 3: Hybrid Feature Penalty Tuning (HFPT)

The HFPT method combines Feature Weight Adjustment and Penalty-Based Models to improve inclusion. A total of 629 models are evaluated, incorporating various combinations of *penalty types*, *penalty values*, *feature reduction factors*, and *feature configurations*. To simplify the analysis, we focus on a subset of representative models highlighting distinct patterns in borrower reclassification outcomes. These models were selected based on their performance across the four measurement metrics.

Table II presents the selected models, categorized into dominant inclusion, balanced redistribution, or extreme outcomes. **Dominant inclusion** models prioritize reclassifying borrowers into lower-risk categories, often achieving high inclusion ratios but sometimes at the expense of class diversity. **Balanced redistribution** models maintain proportional borrower distributions across risk classes while improving inclusivity. In contrast, **extreme outcomes** highlight scenarios where misaligned configurations lead to over-concentration in specific risk categories, undermining inclusivity goals.

Metric 1 evaluates how HFPT approach redistributes borrowers across risk classes. Dominant inclusion models, such as *Pen0-2_VAL500_f0.5_TIME_INT* and *Pen0_VAL5000_f0.1_TIME_INT*, emphasize shifting borrowers to the lowest-risk class, Class 0. For instance, *Pen0-2_VAL500_f0.5_TIME_INT* reallocates 78,511 borrowers to Class 0 while leaving Classes 3 and 4

TABLE II
METRICS SUMMARY FOR BASE MODEL AND ADJUSTED CONFIGURATIONS OF HFPT

Category	Model Name	0	1	2	3	4	Total Model	Lower (M2)	Higher (M3)	No Change	Incl. Ratio
Basemodel	-	20253	21200	18930	20071	20196	100650	0	0	100650	-
Dominant	Pen0-2_VAL500_f0.5_TIME_INT	78511	2664	19475	0	0	100650	80397	0	20253	80397/e
	Pen0_VAL5000_f0.1_TIME_INT	47363	35282	43	59	17903	100650	55400	1	45249	55400.00
	Pen0_VAL500_f0.5_TIME_INT	38268	42817	244	487	18834	100650	50676	17	49957	2980.94
	Pen0_VAL1000_f0.1_TIME_INT	34300	48262	108	71	17909	100650	49708	55	50887	903.78
	Pen0-1_VAL500_f0.1_TIME_INT	44389	38279	24	51	17907	100650	51086	57	49507	896.25
	Pen3-0_VAL1000_f0.1_TIME_INT	25008	57409	207	236	17790	100650	44941	55	55654	817.11
	Pen3_VAL500_f0.1_TIME_INT	74681	7322	321	373	17953	100650	60387	88	40175	686.22
	Pen1-0_VAL500_f0.5_TIME_INT	42145	39081	203	398	18823	100650	50076	57	50517	878.53
Moderate	Pen3_VAL1000_f0.5_TIME	53666	24254	2393	1399	18938	100650	59596	415	40639	143.60
	Pen3-0_VAL500_f1_INT	20221	21258	18977	20565	19629	100650	611	63	99976	9.70
	Pen0-3_VAL1000_f0.1_INT	20211	21345	18937	21207	18950	100650	1368	76	99206	18.00
	Pen4_VAL500_f0.1_TIME	23354	57664	461	177	18994	100650	42947	672	57031	63.91
	Pen4-3_VAL500_f0.5_TIME_INT	25934	54100	322	791	19503	100650	44535	391	55724	113.90
Extreme	Pen4_VAL5000	28	11	0	0	100611	100650	11	80427	20212	0.00014
	Pen4-2_VAL5000_f0.5_INT	0	0	60197	21021	19432	100650	1000	41901	57749	0.02
	Pen4-3_VAL5000_f0.1_TIME_INT	0	0	0	82762	17888	100650	2311	60383	37956	0.04
	Pen3_VAL5000	56	35	0	100551	8	100650	20212	60336	20102	0.33
	Pen2_VAL5000_f0.5_INT	28	35	100395	16	176	100650	40089	41525	19036	0.97
	Pen1_VAL5000_f0.1_INT	2	100545	0	0	103	100650	59094	20289	21267	2.91

empty. *Pen0_VAL5000_f0.1_TIME_INT* similarly moves 47,363 borrowers into Class 0, significantly reducing the population in higher-risk categories. While these models achieve substantial inclusion, the lack of representation in Classes 3 and 4 suggests potential over-concentration, which may oversimplify the risk structure.

Moderate redistribution models achieve a more balanced borrower distribution. *Pen3_VAL1000_f0.5_TIME* redistributes 53,666 borrowers into Class 0 but retains over 1,399 and 19,000 borrowers in Classes 3 and 4, ensuring the system remains proportional. This balance makes such models suitable for scenarios where inclusivity must align with systemic diversity across all classes. Extreme outcomes highlight the risks of over-penalization. *Pen4_VAL5000*, for example, leaves nearly all borrowers—100,611 individuals—classified in Class 4, effectively negating inclusivity goals. Similarly, *Pen4-2_VAL5000_f0.5_INT* concentrates borrowers in Classes 2, 3, and 4, leaving Classes 0 and 1 completely unpopulated. These cases underscore how poorly calibrated penalties can fail to achieve meaningful reclassification. The analysis demonstrates that dominant inclusion models effectively maximize reclassification into lower-risk categories, while moderate redistribution models balance inclusivity and diversity. Conversely, extreme outcomes reveal the pitfalls of misaligned penalty configurations, emphasizing the importance of tuning models to ensure proportional redistribution and inclusivity.

Metrics 2 and 3 assess the reclassification dynamics of borrowers under the HFPT approach, focusing on the number of individuals moving to lower-risk (Metric 2) and higher-risk (Metric 3) categories. **Dominant inclusion models** show exceptional performance in moving borrowers to lower-risk classes while minimizing upward reclassifications. *Pen0-2_VAL500_f0.5_TIME_INT* achieves the highest Metric 2 score, with 80,397 borrowers reclassified into lower-risk categories and zero upward movements. Similarly, *Pen0_VAL5000_f0.1_TIME_INT* records 55,400 downward movements, with only one borrower shifted upward, highlighting its inclusivity potential. *Pen1-0_VAL500_f0.5_TIME_INT* also excels with 50,076 downward reclassifications and only 57 upward movements.

Moderate redistribution models strike a balance between inclusivity and risk diversity. For instance, *Pen3_VAL1000_f0.5_TIME* achieves 59,596 downward movements while limiting upward shifts to 415. These configurations prevent over-concentration in lower-risk classes. Another example, *Pen4-3_VAL500_f0.5_TIME_INT*, facilitates 44,535 lower movement, with 391 higher movement. Such models are ideal for scenarios requiring inclusivity without sacrificing diversity. **Extreme outcomes** highlight the risks of poorly tuned models. *Pen4_VAL5000* and *Pen4-2_VAL5000_f0.5_INT* demonstrate high upward shifts, with 80,427 and 41,901 borrowers moving to higher-risk classes, respectively. Similarly, *Pen3_VAL5000* records 60,336 upward shifts, overshadowing its 20,212 downward movements. These configurations fail to promote inclusion and instead exacerbate the concentration of borrowers in higher-risk categories.

The combined analysis of Metrics 2 and 3 demonstrates the critical trade-offs in borrower movement. While dominant inclusion models maximize downward reclassification, moderate redistribution models provide a balanced approach that maintains diversity. Conversely, extreme configurations emphasize the importance of careful calibration to avoid counterproductive upward shifts and over-concentration in higher-risk classes.

The inclusion ratio (Metric 4) compares the number of borrowers moved to lower-risk classes (Metric 2) against those moved to higher-risk classes (Metric 3).

Dominant inclusion models demonstrate exceptional inclusion ratios, confirming their superior capacity for inclusive reclassification. *Pen0_VAL5000_f0.1_TIME_INT* achieves a remarkable inclusion ratio of 55,400, attributed to its substantial downward movements (55,400) and near-zero upward shifts (1). Similarly, *Pen0_VAL500_f0.5_TIME_INT* records an inclusion ratio of 2,980.94, reflecting its ability to reclassify 50,676 borrowers downward with only 17 upward shifts. **Moderate redistribution models** present balanced inclusion ratios, supporting proportional borrower distribution across risk classes. For example, *Pen3_VAL1000_f0.5_TIME* achieves an inclusion ratio of 143.60, driven by 59,596 downward movements against 415 upward reclassifications. *Pen4-3_VAL500_f0.5_TIME_INT* records a moderate ratio of 113.90, underscoring its effectiveness in achieving inclusivity while preserving class diversity.

Extreme outcomes exhibit low inclusion ratios, reinforcing their inefficacy in fostering inclusivity. *Pen4_VAL5000* and *Pen4-2_VAL5000_f0.5_INT* register near-zero ratios (0.00014 and 0.02, respectively), as their upward movements vastly outnumber downward reclassifications. Similarly, *Pen3_VAL5000* produces an inclusion ratio of 0.33, further highlighting the detrimental effects of high penalties in these configurations. These results emphasize the importance of calibrated penalties and feature adjustments to avoid counterproductive reclassification patterns. Metric 4 provides critical insights into the trade-offs between inclusivity and system stability. While dominant inclusion models achieve extraordinary ratios, moderate redistribution models balance inclusivity with risk class diversity. Conversely, extreme configurations underscore the risks of poorly designed models that fail to enhance credit access or achieve equitable borrower distribution.

Key Takeaway: Our experimental results demonstrate that our proposed HFPT with moderate penalty configurations combining with feature adjustment is the most effective for achieving inclusivity in credit scoring systems. These configurations of HFPT consistently overcome extreme penalties, which often result in concentrated borrower distributions in high-risk classes and minimal inclusivity. As revealed by models with overly narrow penalties (e.g., targeting Class 4) which consistently failed to support equitable reclassification, we shall be cautious of the critical role of penalty moderation in avoiding over-penalty effects since it may exacerbate risk class concentration and undermine inclusivity goals.

Importantly, broader and moderate penalties, particularly those targeting Classes 0 through 3, provided the most balanced outcome. This observation underscores the importance of aligning penalty and feature configurations with policy objectives, as illuminated by HFPT. Decision-makers should avoid extreme penalties, focus on models that prioritize reclassification volume and diversity, and consider the trade-offs between inclusivity and risk management. These insights provide a pathway for refining credit scoring policies to expand access while ensuring stability.

V. Conclusions and Recommendations

This study evaluated three approaches—Feature Weight Adjustment (FWA), Penalty-Based Models, and Hybrid Feature-Penalty Tuning (HFPT)—to improve credit scoring inclusion through data-driven sensitivity analysis. Across experiments, even minor parameter changes significantly impacted borrower classification and inclusion outcomes, underscoring the importance of precise tuning to balance inclusion and risk. HFPT emerged as the most effective approach, balancing inclusivity and risk management by integrating penalties and feature adjustments. Unlike FWA, which depends on highly sensitive features or penalty-based models that risk over-concentration, HFPT enables more flexible and proportional borrower reclassification across diverse scenarios. HFPT's structured outcomes categories (*dominant inclusion*, *balanced redistribution*, and *extreme scenarios*) provide policymakers with actionable insights to tailor configurations to specific goals, enhancing its adaptability.

FWA improves borrower reclassification by amplifying key features, such as delayed payments, but it struggles with limited adaptability, often overemphasizing high-sensitivity parameters. Penalty-Based Models redistribute borrowers through penalty configurations but frequently result in over-concentration in specific risk categories under high penalty values, limiting their applicability for broader inclusivity goals. HFPT combines the flexibility of FWA with the systemic reach of penalties, achieving a proportional and equitable redistribution of borrowers across risk categories.

HFPT offers policymakers actionable frameworks to align credit scoring systems with risk appetite. Moderate penalties combined with feature adjustments achieved balanced reclassifications, reducing upward shifts while promoting inclusivity. Configurations leading to concentrated outcomes provide insights into the trade-offs between inclusivity and systemic stability. HFPT empowers policymakers to refine credit inclusion strategies based on broader financial and risk management goals.

While this study focuses on Indonesia's lending system, which limits its generalizability, the HFPT framework remains adaptable to other credit ecosystems through flexible tuning of features and penalties. However, it is important to recognize that algorithmic adjustments, if not carefully managed, can result in unintended biases or discriminatory outcomes. To prevent unintended biases, we recommend clear guidelines for feature selection and careful calibration of penalties to avoid disproportionate impacts on vulnerable borrowers. Regulatory oversight should be incorporated to monitor the ethical application of these algorithms, with an emphasis on transparency in decision-making processes and regular audits.

Future research should address the reliance on Indonesian datasets and the computational scalability of HFPT. Expanding datasets, incorporating alternative socio-economic factors, and conducting longitudinal studies will help validate long-term impacts. Collaboration between policymakers and financial institutions is vital to implement robust regulatory frameworks, ensuring transparent and equitable lending practices.

This study underscores the transformative potential of parameter tuning in credit scoring models. By adopting HFPT and strategically adjusting parameters, financial institutions can expand credit access for underserved groups while maintaining stability, creating a more inclusive and resilient financial ecosystem.

Beyond its technical contributions, this study encourages reflection on how algorithmic decisions affect financial access in emerging markets. The approach here supports the design of inclusive digital systems that are responsive to local challenges, particularly in ensuring that underserved communities are not further marginalized by automated decision-making.

REFERENCES

[1] Ariza-Garzón, M. J.; Segovia-Vargas, M. J.; Arroyo, J, "Risk-return modeling in the P2P lending market: Trends, gaps, recommendations, and future directions," *Electronic Commerce Research and Applications*, 2021. https://doi.org/10.1016/j.elerap.2021.101079.

[2] Jagtiani, J, "Fintech Lending: Financial Inclusion, risk pricing, and alternative information," *Risk Pricing, and Alternative Information*, 2017. https://doi.org/10.21799/frbp.wp.2017.17.

[3] Lessmann, S; Baesens, B; Seow, H.V.; Thomas, L.C., "Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research," *European Journal of Operational Research*, vol. 247, pp. 124–136, 2015. https://doi.org/10.1016/j.ejor.2015.05.030.

[4] Beck, T, "Fintech and financial inclusion: Opportunities and pitfalls," *ADBI Working Paper Series*, No. 1165, 2020.

[5] Tambunan, T, "Financial Inclusion, P2P Lending, and MSMEs: Evidence From Indonesia," *Fintech Development for Financial Inclusiveness*, pp. 60–81, 2022. https://doi.org/10.4018/978-1-7998-8447-7.ch005.

[6] Situmorang, S.H., "The challenges of fintech inclusion and digitization of SMEs in Indonesia," *Fintech Development for Financial Inclusiveness*, pp. 118–134, 2022. https://doi.org/10.4018/978-1-7998-8447-7.ch008.

[7] Azis, I.J., "Listen and Design: On Micro, Small and Medium Enterprises," , 2024. https://doi.org/10.1007/978-981-97-3248-7.

[8] Women's World Banking, "From Exclusion to Empowerment: The Meaning of Financial Inclusion," June, 2023. https://www.womensworldbanking.org/insights/from-exclusion-to-empowerment-the-meaning-of-financial-inclusion/

[9] Hannig, A; Jansen, S, "Financial Inclusion and Financial Stability: Current Policy Issues," *Asian Development Bank Institute Working Paper Series*, No 259, December 2010. https://doi.org/10.2139/ssrn.1729122.

[10] Ye, X; Dong, L.A; Ma, D, "Loan evaluation in P2P lending based on Random Forest optimized by genetic algorithm with profit score," *Electronic Commerce Research and Applications*, pp. 23–36, 2018. https://doi.org/10.1016/j.elerap.2018.10.004.

[11] Xia, Y; Liu, C; Liu, N "Cost-sensitive boosted tree for loan evaluation in peer-to-peer lending," *Electronic Commerce Research and Applications*, Vol 24, pp. 30–49, 2017. https://doi.org/10.1016/j.elerap.2017.06.004

[12] Bumacov, V; Ashta, A; Singh, P, "Poverty scoring and financial inclusion of the poor," *Strategic Change, John Wiley & Sons, Ltd*, pp. 555–562, 2017. https://doi-org.tudelft.idm.oclc.org/10.1002/jsc.2166.

[13] Aleksandrova, Y, "Comparing Performance of Machine Learning Algorithms for Default Risk Prediction in Peer to Peer Lending," *TEM Journal 10(1)*, pp. 133–143, February 2021.

[14] Fahner, G, "Design of palatable credit scorecards as a highly automated analytic service: By combining machine learning with domain expertise," *IEEE International Conference on Services Computing*, pp. 850–851, 2014. doi: 10.1109/SCC.2014.120.

[15] Pişkin, M; Kuş, M.C, "Islamic Online P2P Lending Platform," *Procedia Computer Science*, Vol 158, pp. 415–419 2019. https://doi.org/10.1016/j.procs.2019.09.070.

[16] Bitvai, Z; Cohn, T, "Predicting peer-to-peer loan rates using Bayesian non-linear regression," *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015. https://doi.org/10.1609/aaai.v29i1.9515

[17] Guo, Y; Zhou, W; Luo, C; Liu, C; Xiong, H, "Instance-based credit risk assessment for investment decisions in P2P lending," *European Journal of Operational Research 249*, pp. 417—426, 2016. https://doi.org/10.1016/j.ejor.2015.05.050

[18] Hindistan, Y.S.; Kıyakoğlu, B.Y.; Rezaeinazhad, A.M.; Korkmaz, H.E.; Dağ, H, "Alternative Credit Scoring and Classification Employing Machine Learning Techniques on a Big Data Platform," *4th International Conference on Computer Science and Engineering* , pp. 1–4, 2019. DOI: 10.1109/UBMK.2019.8907113

[19] Kim, J.Y.; Cho, S.B., "Predicting repayment of borrows in peer-to-peer social lending with deep dense convolutional network, "*Expert Systems 36*, 2019. https://doi-org.tudelft.idm.oclc.org/10.1111/exsy.12403.

[20] Liu, S; Wu, S, "Multiangle P2P Borrower Characterization Analytics by Attributes Partition Considering Business Process, "*IEEE Intelligent Systems*, vol. 35, no. 3, pp. 96–105, 2020. DOI: 10.1109/MIS.2020.2986973

[21] He, F; Li, Y; Xu, T; Yin, L; Zhang, X, "A Data-Analytics Approach for Risk Evaluation in Peer-to-Peer Lending Platforms, "*IEEE Intelligent Systems*, vol. 35, no.3, pp. 85–95, 2020. DOI: 10.1109/MIS.2020.2971946.

[22] Yoon, Y; Li, Y; Feng, Y, "Factors affecting platform default risk in online P2P lending business: an empirical study using Chinese online P2P platform data, "*Electronic Commerce Research, 19* pp. 131–158, 2019. https://doi.org/10.1007/s10660-018-9291-1.

[23] Xiong, J; Tu, M; Zhou, Y, "Using weighted similarity to assess risk of illegal fund raising in online P2P lending, "*International Journal of Digital Crime and Forensics*, Vol 10 Issue 4, 2018. DOI: 10.4018/IJDCF.2018100105.

[24] Sulastri, R; Janssen, M, "Challenges in designing an inclusive Peer-to-peer (P2P) lending system, "*Proceedings of the 24th Annual International Conference on Digital Government Research*, pp. 55–65, 2023. https://doi.org/10.1145/3598469.3598475.

[25] Njuguna, R; Sowon, K, "Poster: A scoping review of alternative credit scoring literature, "*Proceedings of the 4th ACM SIGCAS Conference on Computing and Sustainable Societies*, pp. 437–444, 2021. https://dl.acm.org/doi/10.1145/3460112.3471972.

[26] Ismail, Q.F.; Al-Sobh, E.S.; Al-Omari, S.S.; Yaseen, T.M.B; Abdullah, M.A., "Using machine learning algorithms to predict the state of financial inclusion in Africa, "*12th International Conference on Information and Communication Systems (ICICS) IEEE*, pp. 317–323, 2021. DOI: 10.1109/ICICS52457.2021.9464590.

[27] Kumar, I.E.; Hines, K.E.; Dickerson, J.P., "Equalizing Credit Opportunity in Algorithms, "*Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 357–368, 2022. https://dl.acm.org/doi/10.1145/3514094.3534154.

[28] Verhaeghe, H; Nijssen, S; Pesan, G; Quimper, C.G.; Schaus, P, "Learning Optimal Decision Trees Using Constraint Programming, "*Constraints, 25*, pp. 226–250, 2020. https://doi.org/10.1007/s10601-020-09312-3.

[29] Tang, J; Li, J; Xu, W; Tian, Y; Ju, X; Zhang, J, "Robust Cost-Sensitive Kernel Method with Blinex Loss and Its Applications in Credit Risk Evaluation, "*Neural Networks, 143*, pp. 327–344, 2021.https://doi.org/10.1016/j.neunet.2021.06.016.

[30] Drichel, A; Meyer, U, "Making Use of NXt to Nothing: The Effect of Class Imbalances on DGA Detection Classifiers, "*Proceedings of the 15th International Conference on Availability, Reliability and Security*, pp. 1–9, 2020. https://doi.org/10.1145/3407023.3409190.

[31] Ding, J; Zhang, X; Li, X; Wang, J; Yu, R; Pan, M, "Differentially Private and Fair Classification via Calibrated Functional Mechanism, "*Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 622–629, 2020. https://doi.org/10.1609/aaai.v34i01.5402.

[32] Park, S; Byun, J; Lee, J, "Privacy-Preserving Fair Learning of Support Vector Machine with Homomorphic Encryption, "*Proceedings of the ACM Web Conference*, pp. 3572–3583, 2022. https://doi.org/10.1145/3485447.3512252.

[33] Salimi, B; Rodriguez, L; Howe, B; Suciu, D, "Interventional Fairness: Causal Database Repair for Algorithmic Fairness, "*Proceedings of the International Conference on Management of Data* , pp. 793–810, 2019. https://doi.org/10.1145/3299869.3319901.