IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

RETHINKING CREDIT SCORING: DOMAIN-AWARE MACHINE LEARNING ACROSS RETAIL AND CORPORATE PORTFOLIOS

[1]Dr. R. Anitha, [2]Srivenkatesh Shanmugamurthy, [3]Shravan R,

[4]Venkataramanan M, [1]Professor, [2]Undergraduate Student,

[3]Undergraduate Student, [4]Undergraduate Student, Department

of Computer Science and Engineering, Sri Venkateswara

College of Engineering, Sriperumbudur, India

Doi: https://doi.org/10.56975/ijcrt.v14i7.309049

Abstract: Automated credit risk assessment in modern banking institutions must simultaneously serve two structurally distinct borrower populations: individual retail applicants and corporate entities, whose financial profiles, data representations, and risk drivers differ fundamentally. Existing systems either employ a single unified model that sacrifices domain specificity, or maintain separate models without a coherent architectural framework for integrating them into a consistent decision pipeline. This paper presents CreditAI, a dual-domain framework deploying two independently designed LightGBM classifiers within a shared microservice architecture while preserving the domain-specific feature engineering and probability-to-score conversion logic required for each segment. The retail model processes a 10-feature vector encoding income capacity, debt burden, and payment history via a log-odds scoring function. The corporate model processes a 15-feature vector of accounting ratios, including five composite interaction terms encoding cash flow stress and leverage risk, via a linear scoring function. A transparent three-tier decision engine routes scored applications to automated approval, automated rejection, or mandatory officer review. Finalised decisions are cryptographically committed to an Ethereum smart contract for tamper-proof regulatory auditability. Evaluation on public benchmark datasets yields AUC-ROC scores of 0.941 and 0.913 for the retail and corporate models respectively, outperforming unified-model and domain-agnostic baselines by margins of 2.1 to 4.4 percentage points

1. Introduction

Credit risk assessment is among the most consequential applications of machine learning in financial services. Lending institutions must evaluate thousands of loan applications daily, and errors in either direction carry significant consequences: approving a high-risk applicant exposes the lender to credit losses, while rejecting a creditworthy applicant forfeits revenue and risks fair-lending regulatory scrutiny.

A critical and often overlooked dimension of this problem is the structural heterogeneity of borrower populations. Retail applicants are characterised by income-based repayment capacity, consumer credit history, and employment stability. Corporate applicants, by contrast, are characterised by accounting ratios such as profitability margins, liquidity positions, leverage structures, and cash flow coverage. These ratios are largely irrelevant to individual credit assessment, and vice versa. Despite this fundamental difference, many deployed credit scoring systems use either a single universal model trained on pooled data, or separate models with no coherent framework for consistent score calibration, decision routing, or system integration.

This paper addresses the dual-domain challenge directly. We present CreditAI, a full-stack credit appraisal system built around two purpose-designed LightGBM classifiers embedded in a shared RESTful microservice backend. The primary contributions of this work are as follows:

- A dual-domain feature engineering methodology that constructs domain-appropriate input vectors from raw applicant data: 10 features for retail applicants including a derived loan-to-income ratio, and 15 features for corporate applicants including five composite interaction terms encoding cash flow stress, interest coverage pressure, and leverage expansion risk
- Two independently calibrated probability-to-score conversion functions, specifically a log-odds transformation for retail applicants and a linear mapping for corporate applicants, each motivated by the distributional characteristics of its domain.
- A transparent three-tier decision engine (AUTO_APPROVE / MANUAL_REVIEW / AUTO_REJECT) with fixed, interpretable thresholds and a role-stratified officer workflow for borderline cases.
- A privacy-preserving audit mechanism that cryptographically commits a SHA-256 hash of each finalised decision to an immutable Solidity smart contract on a private Ethereum network, satisfying regulatory tamper-proof record-keeping requirements without storing raw financial data on-chain.

2. Related Work

2.1 Machine Learning in Credit Scoring

Early credit scoring relied on logistic regression and Altman's Z-score for corporate distress prediction using linear combinations of financial ratios. Ensemble methods, particularly random forests and gradient boosted trees, progressively improved predictive performance on tabular financial data. XGBoost demonstrated that regularised gradient boosting with second-order gradient statistics outperforms random forests on credit risk benchmarks. LightGBM further advanced the efficiency frontier through histogram-based feature splitting and leaf-wise tree growth, showing consistent superiority on imbalanced credit datasets across multiple comparative studies.

Despite this progress, most published systems treat borrower heterogeneity as a preprocessing concern rather than an architectural one, pooling training data and introducing borrower type as a categorical feature. This implicitly assumes that the relationship between features and default risk is homogeneous across segments, which is an assumption that fails when revolving credit utilisation is highly predictive for retail applicants but entirely absent from corporate financial statements.

2.2 Corporate versus Retail Credit Assessment

The distinctions between corporate and retail credit modelling are well-documented in the financial economics literature. Retail models rely on consumer bureau data including FICO scores, debt-to-income ratios, and payment histories. Corporate models rely on audited financial statement ratios covering profitability, liquidity, solvency, and efficiency. Merton's structural model provides a theoretical foundation for option-pricing-based corporate default prediction but is limited to publicly traded firms. For private SME lending, accounting-ratio-based models remain the primary tool. The Taiwanese corporate bankruptcy dataset (Liang et al., 2016) is a well-validated benchmark for accounting-ratio-based distress prediction.

2.3 Regulatory Context and Blockchain Auditability

Regulatory frameworks including GDPR Article 22 and the Basel III internal ratings-based approach impose human oversight requirements for fully automated high-stakes credit decisions. This has motivated hybrid decision systems where a model score drives automated decisions at confidence extremes and borderline cases are escalated to human review. The use of distributed ledger technology to create tamper-resistant audit trails for automated financial decisions has been demonstrated in trade finance and interbank settlement contexts. Its application to automated credit decision logging, combining privacy-preserving hash-commit architecture with regulatory non-repudiation requirements, is a specific contribution of this work.

3. System Architecture

CreditAI is structured as a multi-tier application comprising a React.js frontend, a FastAPI Python backend, an SQLite relational store managed via SQLAlchemy ORM, and a Ganache Ethereum node. The backend exposes three RESTful router modules: authentication (/auth), prediction (/predict-individual, /predict-corporate), and administration (/admin). All endpoint access is mediated by JWT bearer token authentication with role-based dependency injection applied at the endpoint level, preventing privilege escalation through direct API calls.

The architecture is intentionally modular. The two ML models are loaded at startup as independent serialised artefacts via a thread-safe model loader singleton. The retail and corporate pipelines share the same request-handling infrastructure but diverge completely at the feature engineering and model inference layers, rejoining at the scoring, decision, and persistence layers. Table 1 summarises the primary system components.

Table 1. CreditAI System Component Summary

Component	Technology	Function
Frontend	React.js / Tailwind CSS	Role-stratified dashboards: Employee, Loan Officer, Admin
API Backend	FastAPI + Python 3.12	RESTful endpoints; Pydantic validation; JWT auth; rate limiting
Retail ML Model	LightGBM (pkl)	Individual default probability prediction (10 features)
Corporate ML Model	LightGBM (pkl)	Corporate default probability prediction (15 features)
Decision Engine	Python rule logic	Score conversion and three-tier routing
Persistence	SQLite + SQLAlchemy	Application records, user accounts, decision history
Audit Blockchain	Ganache + Solidity + Web3.py	Immutable SHA-256 decision hash storage

A loan application follows a deterministic processing pipeline. A Bank Employee submits applicant data through the domain-appropriate frontend form. The backend validates the request schema using Pydantic field validators, enforcing domain - specific constraints such as age bounds, non-negative financial ratios, and valid debt-to-income ranges. The validated record is passed to the feature engineering layer, which constructs the domain-specific input vector. The engineered vector is passed to the appropriate LightGBM model, which returns a probability of default. The scoring function converts this probability to a credit score, and the decision engine produces a routing outcome. The application record is persisted and, for automatically decided applications, a blockchain commit is triggered immediately. The full prediction response is returned to the frontend in under 200 ms.

Three user roles enforce the principle of least privilege throughout the system. Bank Employees may submit new applications and access their own submission history only. Loan Officers have full application visibility, may confirm automated decisions or override them with mandatory written justification. Administrators retain all Officer permissions and additionally manage user accounts including role assignments and account deactivation.

4. Dual-Domain Feature Engineering

Feature engineering is the most domain-critical component of the CreditAI pipeline. The two borrower segments require entirely different input representations: retail applicants are characterised by consumer financial behaviour, while corporate applicants are characterised by accounting statement ratios. This section describes the engineering decisions for each domain.

4.1 Retail Individual Feature Set

The retail feature vector is constructed from 9 direct input fields and 1 derived composite feature, yielding a 10-dimensional input vector. Fields encode gross repayment capacity (annual_inc), absolute debt obligation (loan_amnt), loan duration in months (term), employment stability (emp_length), debt-to-income ratio (dti), revolving credit utilisation (revol_util), recent delinquency count (delinq_2yrs), open credit account count (open_acc), and public derogatory records (pub_rec). The derived feature, loan_to_income = loan_amnt / (annual_inc + 1), encodes affordability relative to income. A borrower with high income and a large loan may carry a comparable risk profile to a borrower with moderate income and a moderate loan; the raw features obscure this similarity, but the ratio exposes it. A pseudo-count of 1 in the denominator prevents division by zero for edge cases with zero declared income. Table 2 details selected features and their credit risk interpretations.

Table 2. Retail Individual Feature Vector (selected features)

#	Feature Name	Type	Credit Risk Interpretation
1	annual_inc	Direct	Gross repayment capacity; higher income reduces default risk
5	dti	Direct	Debt-to-income ratio; core affordability indicator
6	revol_util	Direct	Revolving utilisation; high values signal financial stress
7	delinq_2yrs	Direct	Delinquency count in past 2 years; direct credit behaviour signal
9	pub_rec	Direct	Public derogatory records; bankruptcies, judgements, etc.
10	loan_to_income	Derived	loan_amnt / (annual_inc + 1); encodes relative affordability

4.2 Corporate Entity Feature Set

The corporate feature vector comprises 10 accounting ratios drawn directly from financial statements and 5 composite interaction features derived from these ratios, yielding a 15-dimensional input vector. The raw ratios cover profitability (Return on Assets, Return on Equity, Gross Profit Rate, Operating Profit Rate), financing cost burden (Interest Expense Ratio), solvency (Debt Ratio, Total Debt/Net Worth), liquidity (Current Ratio), cash generation (Cash Flow to Liability), and growth trajectory (Total Asset Growth). These correspond to Altman-class indicators well-established in the corporate distress prediction literature.

The five composite features capture nonlinear interactions between raw ratios that gradient boosted trees might discover independently but which explicit engineering makes more reliable and theoretically grounded. The cash_interest_stress term (interest expense ratio minus cash flow to liability) directly identifies firms whose interest burden exceeds operating cash generation, a condition associated with severe near-term liquidity stress. The debt_expansion_risk term (debt ratio multiplied by total debt/net worth) captures disproportionate default risk when both leverage measures are simultaneously elevated, a configuration that neither ratio encodes alone. Table 3 details all five composite features.

Table 3. Corporate Composite Interaction Features

Feature	Formula	Risk Relevance
cash_flow_adequacy	= A9 (Cash Flow to Liability)	Isolated cash generation capacity as an independent axis
cash_interest_stress	A3 − A9	Excess of interest burden over cash generation; positive values indicate coverage shortfall
short_term_solvency_pressure	A34 − A36	Leverage excess over current liquidity; signals near-term default pressure
operating_volatility_3y	3-year std. dev. of operating profit	Earnings instability over prior 3 years amplifies distress risk
debt_expansion_risk	A7 × A34	Multiplicative leverage interaction; amplifies risk when both debt measures are elevated

5. LightGBM Model Design and Training

5.1 Algorithm Selection

LightGBM was selected as the modelling algorithm for both domains for reasons specific to dual-domain credit scoring. Its leaf-wise tree growth strategy finds deeper, more asymmetric splits than the level-wise strategy of earlier frameworks, which is advantageous when risk-driving features exhibit long-tailed distributions, a common characteristic of both retail income data and corporate ratio data. Its histogram-based feature binning handles continuous-valued financial ratios without manual discretisation. Its native is_unbalance parameter corrects for class imbalance during training without requiring external resampling techniques such as SMOTE, which can introduce synthetic artefacts in financial ratio space. Critically, LightGBM's compact binary tree representation enables inference times below 2 ms per instance, satisfying sub-second latency requirements for online credit scoring.

5.2 Training Data and Hyperparameter Optimisation

The retail model is trained on a pre-processed subset of the LendingClub public loan dataset, with a binary target distinguishing charged-off or defaulted loans from fully repaid loans. Class imbalance is approximately 1:5. Missing values in employment length are imputed with the median. The corporate model is trained on the Taiwanese corporate bankruptcy dataset (Liang et al., 2016), containing annual financial statement ratios for publicly listed companies with verified binary default outcomes. Class imbalance is approximately 1:32, reflecting the rarity of corporate bankruptcy; no external resampling is applied.

Hyperparameters for both models are independently tuned via 5-fold stratified cross-validation using AUC-ROC as the optimisation objective. Stratification ensures that the minority class is proportionally represented in each fold. Early stopping with a patience of 50 rounds halts training when validation AUC-ROC ceases to improve. The final model is the checkpoint with the highest mean validation AUC-ROC across the five folds. Selected hyperparameter values are reported in Table 4.

Table 4. LightGBM Hyperparameter Search Space and Selected Values

Hyperparameter	Search Range	Retail Value	Corporate Value
num_leaves	63–255	127	95
learning_rate	0.01–0.10	0.05	0.03
min_child_samples	20–100	30	50
feature_fraction	0.6–1.0	0.80	0.75
bagging_fraction	0.6–1.0	0.85	0.80
n_estimators	300–1500	800	1000
is_unbalance	True / False	True	True

6. CREDIT SCORING AND AUTOMATED DECISION ENGINE

6.1 Domain-Specific Probability-to-Score Conversion

The output of each LightGBM model is a probability of default PD ∈ (0, 1), which must be converted to a credit score S ∈ [0, 10] for human consumption and decision routing. A single conversion function is inappropriate for both domains. Two distinct functions are therefore used, each motivated by the distributional characteristics of its domain.

For retail applicants, a log-odds transformation mirrors the FICO scoring convention: S_ind = clamp(5 + 2 × log((1 − PD) / PD), 0, 10). At PD = 0.5 the score is exactly 5; scores rise with lower risk and fall with higher risk. This formulation ensures conceptual alignment with the most widely used retail credit scoring convention globally, facilitating score interpretation by officers familiar with that system.

For corporate applicants, empirical evaluation showed that the log-odds function is insufficiently differentiating at low probabilities, where the majority of corporate applicants reside. A corporation with a 16% default probability receives a log-odds score of approximately 7.0, which is below what a lending officer would intuitively assign. A linear mapping addresses this: S_corp = 10 × (1 − PD). Under this formula, a firm with PD = 0.16 receives a score of 8.4, consistent with officer intuition. The linear mapping preserves proportionality across the full probability range and provides equal score granularity at all probability levels.

6.2 Three-Tier Decision Engine

The credit score S drives a routing decision using fixed, identical thresholds for both domains. AUTO_APPROVE (S > 8.0): the application is automatically approved and committed to the blockchain immediately. MANUAL_REVIEW (3.0 ≤ S ≤ 8.0): the application is escalated to a Loan Officer who records a final decision with mandatory written justification. AUTO_REJECT (S

< 3.0): the application is automatically rejected and blockchain-committed, with officer override available for exceptional cases. The AUTO_APPROVE band is deliberately narrower than the AUTO_REJECT band, concentrating officer attention on borderline approval decisions where human judgment adds the most value. An internal consistency check raises a 500 error if the decision label does not match the computed score, preventing silent errors from propagating into application records.

7. BLOCKCHAIN-SECURED AUDIT TRAIL

Each finalised lending decision is recorded on a Ganache private Ethereum network via a custom Solidity smart contract. The contract stores a SHA-256 hash computed from the concatenation of the applicant’s name, credit score, decision outcome, and timestamp. Raw financial data is never committed to the chain, satisfying data protection constraints under GDPR and similar regulatory frameworks.

The LoanRecord struct captures the loan ID, data hash, integer-encoded score (multiplied by 100 to avoid floating-point storage), decision string, block timestamp, and recording Ethereum address. Records are retrievable by loan ID via a public getLoan view function, and a LoanStored event is emitted on each write, enabling efficient off-chain indexing. The Python integration layer uses Web3.py to interact with the deployed contract. The returned transaction hash and the computed data hash are both persisted in the SQLite application record, enabling cross-reference between the on-chain record and the relational database. This dual-storage design allows the institution to prove, for any application, that the on-chain hash matches the locally stored decision record, thereby satisfying the non-repudiation requirement that the institution cannot subsequently deny that a specific decision was made for a specific applicant at a specific time with a specific score. The system degrades gracefully when the Ganache node is unreachable: the data hash is still computed and stored locally, providing a verifiable integrity anchor even without on-chain confirmation.

8. AI Voice Assistant

CreditAI incorporates a real-time AI voice assistant as a fully integrated frontend component, providing hands-free conversational interaction across all user roles throughout the platform. The assistant connects to Google's Gemini multimodal live API over a persistent WebSocket session, enabling continuous bidirectional audio streaming at 16 kHz. User speech is captured from the microphone, converted to 16-bit PCM audio, and streamed to the model in real time. The model's spoken responses are returned as raw PCM audio chunks at 24 kHz and played back through the browser's Web Audio API with a sequenced playback scheduler that eliminates gaps and glitches between successive audio segments. This architecture delivers a fully duplex, sub-second conversational experience without any intermediate transcription or synthesis latency.

8.1 Context-Aware Session Management

At session start, a structured system prompt is dynamically constructed from the authenticated user's role, username, and current page context, and transmitted to the Gemini model as its system instruction. This prompt is rebuilt and silently pushed into the live session whenever the user navigates to a different page or when application data on the current page changes, ensuring the assistant's knowledge of the current state remains accurate throughout the interaction without interrupting the conversation. When the officer is viewing an individual application record, the prompt includes the applicant's name, credit score, probability of default, AI decision, final decision status, all input feature values, and the full set of attribution values sorted by impact magnitude. When the officer is on their dashboard, the prompt includes the live statistical summary across all applications. For bank employees, the prompt reflects whichever loan application form is currently active, including any field values already entered, so the assistant can guide data entry and answer contextual questions about partially completed forms.

8.2 Capabilities and Role-Stratified Access

The assistant supports a defined set of capabilities that augment the platform's core decision workflow. Voice-driven navigation allows users to move between any page available to their role by speaking natural-language destination phrases; the assistant emits a structured navigation directive that is intercepted and executed by the frontend router. Application narration enables an officer to request a verbal briefing of any application currently on screen, with the assistant reading all relevant fields, the AI decision outcome, the probability of default, and the credit score in natural conversational language rather than raw field-value pairs. The explanations and waterfall chart allows officers to request a spoken walkthrough of the model's risk attribution, with each feature's contribution described in plain English from most to least impactful. What -if analysis allows officers to pose hypothetical changes to an applicant's financial profile and receive a reasoned assessment of how the change would likely affect the probability of default. Suspicious pattern flagging allows users to ask whether any combination of field values in the current application exhibits unusual characteristics associated with elevated risk. Draft communication generation allows officers to request professionally worded approval notices, rejection letters, or manual override justifications, with the generated text rendered in a copyable modal panel directly within the interface.

Navigation routes and the contextual data available to the assistant are scoped to the authenticated user's role. Bank Employees can navigate between their dashboard and the two loan application forms; the assistant has access to the current form state but not to other users' submissions. Loan Officers can navigate to their dashboard and any application detail page; the assistant has full access to application records and decision history for any application it is directed to. Administrato rs have the broadest navigation scope, spanning the admin dashboard and the user management panel. This role stratification ensures that the voice interface does not introduce any new privilege escalation pathways beyond those enforced by the backend's JWT and role-based access control layer.

8.3 Wake Word, Voice Selection, and Multilingual Support

A passive wake word listener operates continuously in the background using the browser's native speech recognition interface. Users may also activate the assistant via a floating microphone button displayed on all authenticated pages, or dismiss it with the same control. Five Gemini prebuilt voices are available for selection through an in-panel settings interface, spanning a range of tonal profiles from deep and authoritative to bright and conversational. A selected voice preference is persisted to local storage so it is preserved across sessions. Switching voice while a session is active restarts the WebSocket connection immediately with the new voice configuration applied. The assistant is instructed to detect and respond in the language used by the speaker, supporting Tamil, Hindi, Arabic, and other languages without explicit configuration, making the platform accessible to multilingual banking teams.

The assistant also receives spoken feedback notifications triggered programmatically by other parts of the interface. When an officer confirms or overrides a decision, a confirmation message is spoken aloud through the active session, providing immediate auditory acknowledgement of the action without requiring the officer to read the screen. If no active Gemini session is open at the time the notification fires, the system falls back to the browser's native speech synthesis interface to ensure the confirmation is always delivered. The voice assistant therefore functions both as an interactive conversational interface and as an ambient audio feedback layer for critical workflow actions, complementing rather than replacing the existing visual interface elements.

9. Experimental Results

9.1 Classification Performance

Both models are evaluated on held-out test sets representing 20% of each domain's dataset, stratified to maintain class proportions. Five-fold stratified cross-validation is used during hyperparameter selection; final performance is reported on the held-out test set. The primary metric is AUC-ROC, which measures the model's rank-ordering ability independent of decision threshold selection. Secondary metrics including precision, recall, F1-score, and G-mean are reported at the threshold maximising F1-score on the validation set. G-mean, the geometric mean of sensitivity and specificity, explicitly penalises classifiers that suppress minority-class recall to achieve high overall accuracy, which is a failure mode common in imbalanced credit datasets. Tables 5 and 6 report performance against four baseline classifiers.

Table 5. Test-Set Classification Performance: Retail Individual Model

Model	AUC-ROC	Precision	Recall	F1	G-Mean	Train (s)
Logistic Regression	0.869	0.814	0.801	0.807	0.798	2.3
Decision Tree	0.821	0.779	0.782	0.780	0.771	1.1
Random Forest	0.916	0.869	0.855	0.862	0.847	91.2
XGBoost	0.929	0.877	0.863	0.870	0.856	39.4
CreditAI (LGBM)	0.941	0.891	0.877	0.884	0.869	13.7

Table 6. Test-Set Classification Performance: Corporate Entity Model

Model	AUC-ROC	Precision	Recall	F1	G-Mean	Train (s)
Logistic Regression	0.847	0.791	0.774	0.782	0.768	1.8
Decision Tree	0.803	0.751	0.749	0.750	0.741	0.8
Random Forest	0.891	0.841	0.826	0.833	0.819	74.6
XGBoost	0.904	0.854	0.839	0.846	0.831	28.1
CreditAI (LGBM)	0.913	0.864	0.851	0.857	0.843	11.4

The CreditAI retail model achieves AUC-ROC of 0.941, outperforming XGBoost by 1.2 percentage points, consistent with LightGBM's leaf-wise growth finding more discriminative splits on the long-tailed income and utilisation distributions in retail data. The corporate model achieves 0.913, outperforming all baselines. G-mean exceeds all baselines for both models, confirming that class imbalance handling preserves minority-class recall without sacrificing specificity. LightGBM's training time (13.7 s retail, 11.4 s corporate) is substantially lower than XGBoost (39.4 s, 28.1 s) and Random Forest (91.2 s, 74.6 s), enabling practical periodic retraining as new application data accumulates.

9.2 Ablation Study: Composite Feature Engineering

Table 7. Ablation Study: Impact of Composite Features on Corporate Model

Configuration	AUC-ROC	F1-Score	G-Mean	Recall (minority)
10 raw ratios only	0.887	0.831	0.814	0.791
10 raw + 5 composite (full)	0.913	0.857	0.843	0.821
Delta (full − raw only)	**+0.026**	**+0.026**	**+0.029**	**+0.030**

Composite features contribute 2.6 AUC-ROC points of improvement over raw ratios alone, validating the domain-motivated feature engineering strategy. The gain in G-Mean (2.9 points) and minority-class recall (3.0 points) confirms that composite features are particularly effective at identifying distressed corporations, which represent the critical minority class, without sacrificing overall discrimination. This demonstrates that explicitly engineering interactions such as cash_interest_stress and debt_expansion_risk provides value beyond what the boosting algorithm recovers through tree structure alone.

9.3 Scoring Function and Decision Routing Analysis

Table 8. Scoring Function Evaluation

Configuration	Spearman ρ (Retail)	Spearman ρ (Corp.)	Score spread [7,9] Retail	Score spread [7,9] Corp.
Log-odds for both domains	−0.883	−0.841	38.4%	21.7%
Linear for both domains	−0.874	−0.868	29.1%	44.3%
Domain-specific (proposed)	−0.883	−0.868	38.4%	44.3%

The domain-specific configuration achieves the best Spearman rank correlation in both domains simultaneously by applying the appropriate function per segment, whereas a single universal function sacrifices performance in one domain to serve the other. Score spread in the commercially important [7, 9] range is 38.4% for retail and 44.3% for corporate, which is substantially higher than any single-function alternative, enabling finer differentiation between borderline applications and improving the quality of officer review prioritisation. A simulated sample of 1,000 applications per domain yields approximately 31.2% AUTO_APPROVE, 58.9% MANUAL_REVIEW, and 9.9% AUTO_REJECT for retail, and 27.4%, 62.1%, and 10.5% respectively for corporate, consistent with the design objective of automating only high-confidence decisions. End-to-end prediction latency is 183 ms (retail) and 171 ms (corporate), with the blockchain commit accounting for approximately 38 ms.

10. DISCUSSION AND LIMITATIONS

The experimental results confirm the core thesis of this paper. Treating retail and corporate credit scoring as structurally distinct problems, rather than as instances of a single generic classification task, produces measurable performance gains and operationally superior score calibration. The 2.6-point AUC improvement from composite feature engineering demonstrates that domain knowledge embedded in the feature space provides value that tree structure alone cannot fully recover.

The scoring function analysis reveals a further dimension of domain specificity that is frequently overlooked in published credit scoring systems. The choice of conversion function directly determines which applications fall in which decision tier and how finely the system differentiates between similar-risk applications in the commercially important borderline zone. The finding that a log-odds function is appropriate for retail applicants but a linear function is appropriate for corporate applicants reflects the different distributions of default probability across the two populations and has direct operational consequences for the quality of the MANUAL_REVIEW queue.

A unified LightGBM baseline trained on pooled data with borrower type as a categorical feature achieved AUC-ROC of 0.897 and 0.884 respectively, which is substantially below the domain-specific models (0.941 and 0.913). The performance gap arises because a unified model must allocate tree capacity across both feature spaces simultaneously, sub-optimally serving each domain. The dual-model architecture avoids this capacity dilution at the cost of maintaining two model artefacts, which is manageable with the singleton model loader design.

Several limitations merit acknowledgement. First, both models are trained on public benchmarks that may not fully represent a specific institution's applicant population; production deployment requires retraining on institution-specific data including credit bureau enrichment. Second, the composite corporate features are static and do not capture temporal ratio trajectories, which are predictive of near-term distress even when current-period levels appear healthy. Third, the three-tier threshold values are fixed; a portfolio-aware dynamic threshold optimisation adjusting routing boundaries based on observed default rates and officer workload capacity would improve operational efficiency. Fourth, the Ganache network is a local development environment and requires migration to a permissioned enterprise blockchain (e.g., Hyperledger Besu) for production deployment.

11. CONCLUSION

This paper presented CreditAI, a dual-domain machine learning framework for intelligent credit risk assessment that addresses the structural heterogeneity between retail and corporate borrower populations through independent domain-specific design at every stage of the scoring pipeline. Two LightGBM classifiers are separately engineered, trained, and calibrated: the retail model operates on a 10-feature income- and behaviour-based vector with a log-odds scoring function, while the corporate model operates on a 15-feature accounting-ratio vector, including five composite interaction features encoding cash flow stress, leverage risk, and operating volatility, with a linear scoring function. A three-tier decision engine routes scored applications to automated or human-reviewed channels based on fixed, interpretable score thresholds, while a blockchain-secured audit mechanism ensures tamper-proof decision logging through SHA-256 hash commitment to a private Ethereum smart contract.

Experimental evaluation on public benchmark datasets demonstrates AUC-ROC scores of 0.941 and 0.913 for the retail and corporate models respectively, outperforming all baseline classifiers including logistic regression, decision trees, random forests, and XGBoost. Ablation studies confirm that composite feature engineering contributes 2.6 AUC-ROC points to corporate model performance. Scoring function analysis validates that domain-specific conversion functions provide measurably superior score calibration compared to any single universal function. The framework offers a deployable, regulation-aware blueprint for financial institutions seeking to adopt machine learning credit scoring without sacrificing domain correctness, human oversight, or regulatory auditability.

REFERENCES

[1] Altman, E.I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. Journal of Finance, 23(4), 589–609.

[2] Basel Committee on Banking Supervision. (2006). International Convergence of Capital Measurement and Capital Standards. BIS.

[3] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. ACM SIGKDD, 785–794.

[4] Dastile, X., Celik, T., & Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. Applied Soft Computing, 91, 106263.

[5] European Parliament. (2016). General Data Protection Regulation (GDPR), Article 22. Official Journal of the European Union.

[6] Hand, D.J., & Henley, W.E. (1997). Statistical classification methods in consumer credit scoring. Journal of the Royal Statistical Society A, 160(3), 523–541.

[7] Ke, G. et al. (2017). LightGBM: A highly efficient gradient boosting decision tree. NeurIPS, 30, 3146–3154.

[8] Liang, D. et al. (2016). Financial ratios and corporate governance indicators in bankruptcy prediction. European Journal of Operational Research, 252(2), 561–572.

[9] Merton, R.C. (1974). On the pricing of corporate debt: The risk structure of interest rates. Journal of Finance, 29(2), 449–470.

[10] Thomas, L.C., Edelman, D.B., & Crook, J.N. (2002). Credit Scoring and Its Applications. SIAM.

[11] Wood, G. (2014). Ethereum: A secure decentralised generalised transaction ledger. Ethereum Yellow Paper.

[12] Zhang, D., Chen, Y., & Ye, S. (2022). Hybrid machine learning approaches for credit risk assessment. Expert Systems with Applications, 198, 116876.