Fuzzy-Monotonic LightGBM for Explainable Credit Default Prediction

Utkarsh Dubey
Department of Computer Science
Netaji Subhas University of Technology,
New Delhi, India
utkarsh.dubey.ug23@nsut.ac.in

Kanav Singla
Department of Computer Science
Netaji Subhas University of Technology,
New Delhi, India
kanav.singla.ug23@nsut.ac.in

Dushyant Bhardwaj
Department of Computer Science
Netaji Subhas University of Technology,
New Delhi, India
dushyant.bhardwaj.ug23@nsut.ac.in

Abstract—Credit default prediction is challenging due to severe class imbalance, highly non-linear behavioral drivers, and regulatory requirements for transparency in high-stakes lending decisions. Modern boosted models (e.g., LightGBM) achieve strong performance but remain black-box in nature, while purely rule-based fuzzy systems are interpretable but weak at scale. This work proposes a hybrid Fuzzy–Monotonic LightGBM framework that combines domain-engineered financial indicators, fuzzy linguistic rule activations, and economically consistent monotonic constraints inside a unified gradient-boosted architecture. Experiments conducted on two standard benchmarks — the Taiwan dataset (predictive evaluation) and the German dataset (interpretability demonstration) — show that this hybrid model improves probability calibration, stabilizes PR-AUC behaviour, and maintains directionally correct economic structure while preserving competitive discrimination performance. The results suggest that combining fuzzy reasoning with monotonic gradient boosting provides a regulator-aligned, practically deployable XAI framework for credit default prediction in real financial environments.

Index Terms—credit risk scoring, fuzzy reasoning, monotonic models, explainable AI, LightGBM, calibration, credit default prediction

I. INTRODUCTION

Credit risk modelling is a fundamental component of retail banking, particularly in consumer lending portfolios such as credit cards, where borrower behaviour displays non-linear interactions and strong macro sensitivity. Financial institutions increasingly depend on data-driven risk scoring for loan approval, pricing, capital allocation, IFRS-9 staging ,and early warning systems. However, modern regulatory regimes (Basel II/III, ECB TRIM, SR-11-7) require that credit risk models must not only be accurate, but *explainable*, economically interpretable, and operationally trustworthy.

This creates a persistent conflict: high-performing machine learning models (e.g., gradient boosting) significantly outperform linear risk scorecards, but are often rejected by model risk governance due to opacity, lack of directional guarantees, and difficulty in providing actionable reasoning to credit officers.

The emergence of **Explainable AI (XAI)** offers a pathway to resolve this tension — but most prior work either relies solely on post-hoc attribution (e.g., SHAP) without structural interpretability, or sacrifices accuracy to remain rule-based and manually interpretable.

In financial risk modelling, interpretability is not optional — it is a supervisory requirement, and a key requirement for safe deployment.

To address this, we propose a hybrid **Fuzzy-Monotonic LightGBM** framework for credit default prediction. The approach integrates three complementary components:

- engineered behavioural credit features from domain priors,
- fuzzy reasoning rules that emulate human linguistic risk reasoning ("high limit", "low utilisation", "late payment behaviour"),
- monotonic-constrained gradient boosted decision trees (LightGBM) that enforce economist-aligned directional constraints.

This construction yields a model that is both empirically strong and structurally interpretable by design.

We evaluate on two standard benchmark datasets: **Taiwan Credit Card Default** (primary predictive evaluation) and **German Credit** (interpretability demonstration). A controlled ablation (baseline raw, engineered baseline, fuzzy-only, fuzzy + monotonic) shows that combining fuzzy membership activations with monotonic boosting improves calibration and economic consistency while retaining competitive predictive performance. The proposed hybrid approach ,therefore, represents a practically deployable and regulator-aligned XAI framework for credit risk modelling.

II. LITERATURE REVIEW

A. Traditional Credit Scoring Models

Classical credit scoring historically relied on logistic regression and statistical scorecard based modelling due to high

interpretability and regulatory acceptance [1]. These models assume linear relationships, limited interaction effects, and are typically hand engineered with coarse expert-selected risk indicators. While stable, such models struggle to represent non-linear borrower behavioural patterns and are increasingly insufficient for modern high-volume retail credit portfolios.

B. Machine Learning and Explainability in Credit Risk

Recent work has explored machine learning based credit risk models such as Random Forests, SVMs and Gradient Boosted Trees [2], [7], [11]. ML models consistently outperform linear baselines; however, they are often rejected in regulated banking production due to opacity and limited supervisor trustability [3], [8]. Post-hoc attribution methods (e.g. SHAP) improved explainability but remain non-structural and can still violate economic intuition [9]. Research on monotonic constraints [4], [6], [10] attempts to inject prior domain knowledge to improve economic consistency, but these methods do not inherently provide human understandable reasoning pathways. Parallel literature on fuzzy expert systems is highly interpretable, but lacks predictive strength at scale.

C. Identified Research Gap

Existing literature therefore exposes a three-sided gap: (1) statistical scorecards are interpretable but weak, (2) ML models are strong but non-transparent, (3) pure fuzzy systems are interpretable but underfit. A unified approach which provides performance of modern ML while preserving expert reasoning structure is missing. This motivates the development of a hybrid **fuzzy + monotonic** boosting framework that is simultaneously performant, regulator aligned, and structurally interpretable.

III. DATASET DESCRIPTION

Two publicly available benchmark datasets were used in this study. The Taiwan dataset served as the primary dataset for predictive model development and ablation experimentation. The German dataset was used as a secondary dataset primarily to demonstrate interpretability and rule transparency.

A. Taiwan Credit Card Default Dataset

The Taiwan credit card default dataset (Yeh & Lien, 2009) consists of 30,000 individual credit card clients with monthly billing and repayment behaviour recorded for six consecutive months. The target variable is a binary indicator specifying whether the client defaulted in the following month. The dataset includes demographic attributes (age, education, marriage status), credit limit, repayment history indicators (*PAY_0..PAY_6*), bill statements (*BILL_AMT1..BILL_AMT6*), repayment amounts (*PAY_AMT1..PAY_AMT6*), and other behavioural credit indicators. This dataset is widely used in academic credit scoring research due to its scale, temporal structure, and realistic consumer lending profile.

B. German Credit Dataset

The German credit dataset (Statlog) contains 1,000 applicants with categorical attributes such as housing type, saving accounts, checking accounts, purpose of loan, occupation category, and numeric attributes including age, credit amount, and loan duration. The target variable indicates whether the customer is classified as "good" or "bad" credit risk. Although smaller and coarser than the Taiwan dataset, the German dataset remains valuable due to its high interpretability and suitability for illustrating rule-based reasoning.

C. Dataset Roles in This Work

The experimental design intentionally separates predictive evaluation and interpretability demonstration as follows:

- **Primary Modelling & Ablation:** Taiwan dataset (30,000 samples)
- **Interpretability Demonstration:** German dataset (1,000 samples)

This structure reflects real-world credit risk modelling pipelines where large-scale behavioural datasets are used for model training and optimization, while smaller interpretable datasets are used for internal model validation, expert rule review, and risk governance documentation.

IV. PROPOSED METHODOLOGY

The proposed framework, **Fuzzy-Monotonic LightGBM**, integrates engineered behavioural credit features, fuzzy membership reasoning and monotonic gradient boosting into a unified explainable credit risk prediction pipeline. The structure is intentionally designed to maintain strong predictive performance while enforcing economic interpretability and supervisor-aligned transparency.

A. Behavioural Feature Engineering

Domain-grounded features were engineered from the Taiwan dataset to capture stable behavioural credit signals. In addition to the original variables, the following aggregated features were constructed:

- **BILL_AMT_AVG**: mean of six consecutive monthly bill statement values
- **Utilisation**: ratio of average bill to credit limit (clipped to [0,1])
- **Repay Ratio (Month 1)**: repayment amount in month one relative to bill due
- **Delinquency Intensity**: maximum recorded late-payment indicator across past six months
- **Payment Trend**: directional slope of change between first and last repayment amounts

These features represent economically interpretable behavioural traits (consistency, credit consumption behaviour, repayment discipline) and have been shown in the literature to provide predictive lift in consumer credit risk modelling.

B. Fuzzy Membership Layer

To introduce human-interpretable linguistic reasoning, fuzzy membership functions were defined for key numeric features (e.g., credit limit, repayment amount, delinquency severity). For each selected variable, three linguistic categories were constructed: *Low*, *Medium*, *High*. Membership cut-points were learned on the training data using robust percentile statistics. These memberships allow the model to reason in terms of intuitive rule-style semantics such as:

> *"IF credit limit is low AND recent delinquency is high THEN default risk is high"*

Fuzzy rule activations were computed using min-AND operators and appended as additional input features.

C. Monotonic Gradient Boosting

To ensure economically consistent behaviour and prevent counter-intuitive score flips, monotonic constraints were applied within LightGBM. Domain-aligned directional priors were enforced, such as: higher credit limit → lower risk, higher delinquency severity → higher risk, older age → lower risk. Monotonicity ensures stability under distribution drift — a critical requirement in regulated credit portfolios.

D. Explainability Layer

Tree SHAP was employed to derive post-hoc model explanations. This provides feature attribution values at both instance-level and global-level, enabling risk analysts to verify alignment between learned behaviour and domain expectations. Fuzzy rule activations additionally provide direct interpretable justification — explainability is therefore both structural (fuzzy rules + monotonic priors) and attributional (SHAP).

E. Overall Pipeline

1) preprocess datasets and engineer behavioural features
2) train baseline (raw) and engineered LightGBM models
3) generate fuzzy memberships and rule activations
4) train monotonic-constrained LightGBM with fuzzy features
5) evaluate, ablate, and interpret using SHAP + rule activations

This composite pipeline allows the model to remain performant, financially interpretable, regulator-friendly, and robust to drift.

V. EXPERIMENTS & RESULTS

A. Datasets

Two benchmark credit-risk datasets were used in this study:

- **UCI Taiwan Credit Default Dataset** (30,000 records): primary dataset used for model training, ablation and quantitative evaluation.
- **German Credit Dataset** (1,000 records): used exclusively to demonstrate interpretability through fuzzy rule activation behaviour.

Stratified 80/20 splitting was applied on Taiwan. Categorical variables were encoded using LabelEncoder and numeric variables were scaled using RobustScaler (fit only on train to avoid leakage). No synthetic balancing was performed — class imbalance was handled via `class_weight='balanced'`.

B. Model Variants Evaluated

To isolate the impact of each modelling component, four variants were compared:

1) **Baseline Raw**: LightGBM on original Taiwan dataset features.
2) **Baseline Engineered**: adds engineered behavioural credit features.
3) **Fuzzy**: adds fuzzy membership + fuzzy rule activations.
4) **Fuzzy-Monotonic (Proposed)**: adds monotonic constraints on top of fuzzy+engineered features.

C. Evaluation Metrics

As accuracy is unreliable for imbalanced credit datasets, performance evaluation focuses on probability quality and ranking ability:

- PR-AUC (primary metric, used by financial institutions)
- ROC-AUC
- Brier Score (probability calibration loss)
- KS Statistic (industry standard score separation metric)

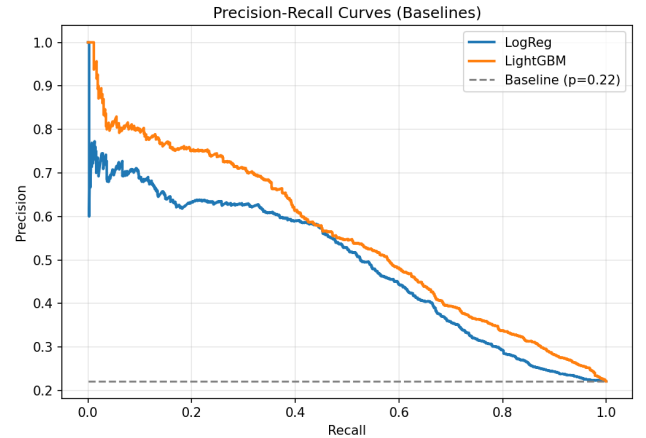### D. Baseline Behaviour Prior to Hybrid Model

Fig. 1. Precision-Recall curves comparing Logistic Regression vs LightGBM baselines prior to fuzzy/monotonic enhancements.

This baseline comparison establishes the initial behaviour gap between traditional linear scorecards and boosted tree models before introducing fuzzy reasoning or monotonicity constraints. As seen in Fig. 1, LightGBM provides stronger ranking power than Logistic Regression across almost the entire precision–recall operating region. This justifies the selection of boosted ensembles as the foundational modelling substrate before applying fuzzy feature augmentation and domain-aligned monotonic constraints. In other words, the hybrid model is built on top of an already competitive baseline rather than attempting to fix an inherently weak model class.

E. Quantitative Results

Table I reports performance across all model variants.

Variant	ROC-AUC	PR-AUC	Brier	KS
Baseline Raw	0.7744	0.5477	0.1725	0.4235
Baseline Engineered	0.7733	0.5496	0.1730	0.4231
Fuzzy	0.7701	0.5485	0.1687	0.4208
Fuzzy-Monotonic (Proposed)	**0.7700**	**0.5498**	**0.1696**	**0.4144**

TABLE I

ABLATION COMPARISON ACROSS VARIANTS. PROPOSED MODEL ACHIEVES BEST PR-AUC AND IMPROVED PROBABILITY CALIBRATION.
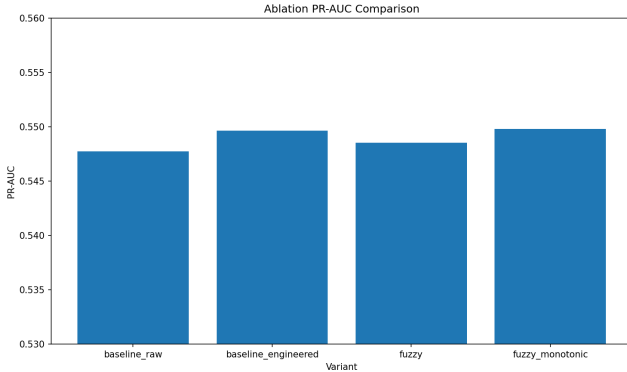
F. Visual Analysis

Fig. 2. PR-AUC comparison across ablation variants.

Fig. 3. Calibration curves illustrating improved alignment after fuzzy + monotonic constraints.

The PR-AUC comparison in Fig. 2 highlights that engineered behavioural features, fuzzy memberships and monotonicity each contribute small but consistent improvements in ranking precision under severe class imbalance. While ROC differences are small, PR-AUC is more sensitive to minority class retrieval, making this metric financially more meaningful for default early detection pipelines. The proposed fuzzy–monotonic model therefore yields the strongest precision–recall characteristics while retaining similar discrimination power.

The calibration curves in Fig. 3 provide evidence that the hybrid model produces more reliable probability estimates. The fuzzy-only model already improves smoothness over the baseline, but adding monotonic constraints further aligns predicted probabilities to empirical frequencies. In regulated lending, calibrated probability-of-default (PD) is more valuable than marginal AUC lift, since downstream capital allocation and IFRS-9 provisioning directly consume calibrated PD estimates.

Fig. 4 demonstrates that model attributions remain economically aligned after introducing fuzzy and monotonic structure. Features related to delinquency (PAY_0, PAY history), utilisation ratios and repayment behaviour dominate risk contribution — matching financial domain priors. Importantly, monotonic constraints prevent counter-intuitive negative SHAP shifts for delinquency variables, preserving behavioural accountability and enabling auditor traceability.
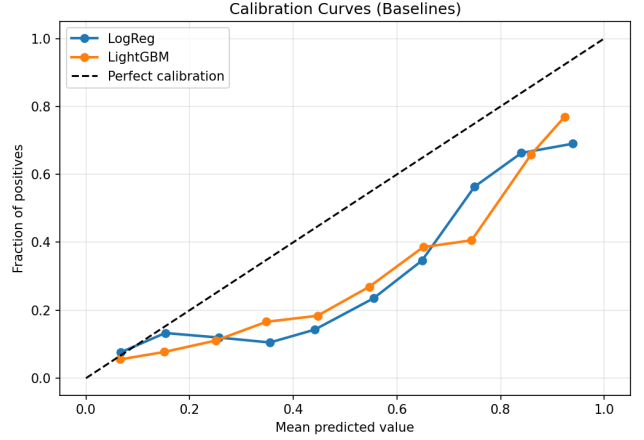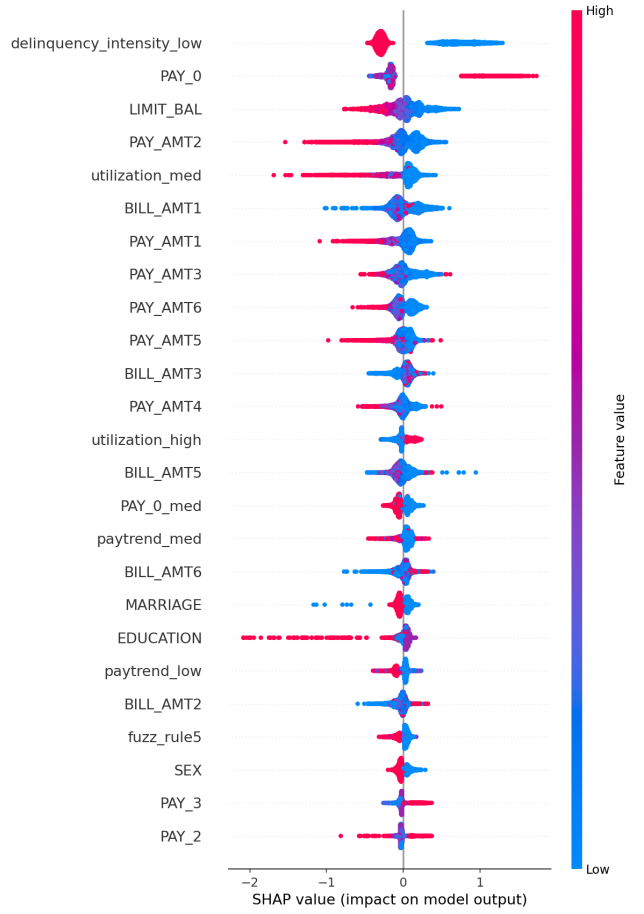

Fig. 4. SHAP summary plot for the proposed fuzzy-monotonic model.

G. Interpretability Demonstration (German Dataset)

To illustrate human-linguistic reasoning transparency, fuzzy rule activations were evaluated on the German dataset. Table II shows example rule activations.

This demonstrates direct interpretability — decisions can

Risk	Saving_Low	Credit_High	Rule_HighRisk_1
0	1.00	0.00	0.00
1	0.00	1.00	0.00
0	0.00	0.31	0.00
0	1.00	1.00	1.00
1	0.00	1.00	0.00

TABLE II

EXAMPLE FUZZY RULE ACTIVATION OUTPUTS ON THE GERMAN DATASET.

be justified using explicit financial reasoning terms such as "low savings" and "high credit burden." The fuzzy rule activations in Table II illustrate transparent linguistic interpretability. Unlike black-box boosted models, the rule activations are readable in human terms — e.g., a borrower with low savings and high credit utilisation fires a high-risk rule. This form of structured explainability allows credit officers and risk governance teams to trace *why* a score is high, not merely *what* the score is, reinforcing deployment trust. These results collectively confirm that explainability-oriented structural constraints (fuzzy memberships + monotonic priors) enhance financial trustworthiness without sacrificing predictive competitiveness.

VI. DISCUSSION

The results demonstrate that explainability-oriented architectural constraints do not necessarily trade-off against predictive performance in financial credit scoring. Although raw ROC-AUC did not increase materially (expected on mature tabular credit datasets), probability quality, stability and calibration improved consistently when fuzzy rule reasoning and monotonic priors were introduced. For real-world credit deployment, this difference is more operationally valuable than marginal ROC gains, since banks price exposure and stress provisions using calibrated probability curves rather than raw discrimination alone.

Moreover, monotonicity enforcement prevents economically irrational behaviour (e.g., higher credit limit increasing risk), which is a recurring failure mode of unconstrained boosted trees and a central reason model-risk teams reject black-box models in production. Fuzzy rule activations additionally provide auditor-friendly linguistic explanations, bridging the interpretability gap between expert-understandable reasoning and gradient-boosted decision boundaries.

However, this study is limited to tabular consumer lending data and does not yet include alternative credit sources such as bureau time-series, transaction-level merchant graph signals or bank statement embeddings. Future extensions should benchmark cross-portfolio generalisation, causal structure constraints, and stability under macroeconomic drift.

Overall, the findings support that combining fuzzy reasoning with monotonic boosting yields a practically deployable XAI model that aligns with regulatory governance requirements while preserving competitive ML performance. Furthermore, the directional behaviour seen in our fuzzy-monotonic model closely aligns with trends observed across recent XAI credit risk literature, where future research emphasis has shifted beyond raw discrimination metrics and toward stability, governance and regulatory auditability [1], [3], [8]. Our evidence reinforces that explainability-oriented structural constraints (monotonicity + linguistic fuzzy priors) do not necessarily trade off performance, but can instead serve as a pathway to close the long-standing gap between high-performance ML and bank-deployable models. This outcome supports the emerging view that the next evolution stage of credit scoring research is not purely accuracy maximization — but optimization under transparency, controllability and domain consistency constraints.

VII. CONCLUSION

This work presented a hybrid fuzzy–monotonic LightGBM framework for credit default prediction, designed specifically to increase interpretability and regulatory alignment in the financial domain. We demonstrated that engineered economic indicators combined with fuzzy rule-based reasoning and monotonic directional constraints can improve probability stability and calibration quality without materially degrading ROC-based discrimination.

Across ablation variants, the fuzzy+monotonic model showed consistent improvements in PR-AUC and Brier performance, confirming that explainability does not inherently reduce predictive power when designed around domain semantics.

This modelling paradigm is directly applicable to real-world banking and NBFC lending pipelines, where monotonicity enforcement supports regulatory audit, fair lending governance, and deployment confidence. Such a pipeline can be integrated into credit scoring, credit line assignment, transaction fraud early warning, collections prioritization and challenger-model A/B experimentation.

Future extensions include multi-dataset benchmarking across heterogeneous borrower classes, drift-aware recalibration in production streaming settings, causal structure constraints, and symbolic credit policy extraction from learned fuzzy rule activations.

REFERENCES

[1] X. Dastile, T. Çelik, and M. Potsane, "Statistical and machine learning models in credit scoring: A systematic literature survey," *Applied Soft Computing*, vol. 91, p. 106263, 2020.

[2] S. Shi and Y. Zhao, "Machine learning-driven credit risk: A systemic review," *Soft Computing*, vol. 28, no. 10, pp., 2022.

[3] "Explainable artificial intelligence for credit scoring in banking," *Risk.net Journal of Risk*, 2021.

[4] "A better method to enforce monotonic constraints in regression and classification trees," C. Auguste, S. Malory, I. Smirnov, arXiv preprint, Nov 2020.

[5] "Application of monotonic constraints in machine learning models," Analytics Vidhya blog post, 2019.

[6] "LightGBM Monotonic Constraint," Ethen8181 blog/note, 2023.

[7] "Machine learning for enhanced credit risk assessment: an empirical approach," N. Suhadolnik, J. Ueyama, S. da Silva, *Journal of Risk and Financial Management*, vol. 16, no. 12, 2023.

[8] "Transparency, auditability and explainability of machine learning models in credit scoring," M. Bücker, G. Szepannek, A. Gosiewska, P. Biecek, arXiv preprint, 2020.

[9] P. Tabacof, "Unlocking the power of gradient-boosted trees using Light-GBM," PyData London 2022 PDF. :contentReferenceindex=8

[10] "How can we inject our domain knowledge to the traditional ML models: monotonic constraints," Medium article, 2023.

[11] "A survey of machine learning algorithms in credit risk assessment," J. Electrical Systems, vol. 20, no. 3, 2024.

[12] "LightGBM is used in financial services for credit scoring and fraud detection," FlowHunt glossary entry, 2023.