

CreditRisk.AI: An Explainable, Fair, and Production-Ready AI Platform for Consumer Credit Risk Assessment

Keshav Kumar
keshavkumarhf@gmail.com | +91 92668 26263

Submitted: February 16, 2026 | Version 1.0

Abstract

Credit risk assessment is a critical function in the financial industry, yet most deployed systems operate as black boxes, providing decisions without transparency into the reasoning process. This paper presents CreditRisk.AI, a complete, production-ready platform that combines six machine learning models (CatBoost, XGBoost, LightGBM, Random Forest, Gradient Boosting, Logistic Regression) with triple explainability through SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations), and counterfactual analysis. Unlike existing research that focuses on individual components in isolation, our system integrates the entire credit decisioning pipeline: comprehensive 50+ field application processing aligned with the 5 Cs of Credit framework, automated feature engineering, model inference, SHAP-based factor attribution, LIME cross-validation, actionable recommendations, FCRA-compliant adverse action notice generation, demographic fairness auditing via Fairlearn, and counterfactual 'what-if' analysis — all served through a production REST API with authentication, rate limiting, and a modern SaaS web interface. Experimental results on the German Credit dataset demonstrate that the CatBoost model achieves 79% AUC-ROC with 76% accuracy, while the SHAP and LIME explanations show 87% agreement on the top-5 influential features, validating the reliability of the dual-method approach. The platform addresses regulatory requirements across FCRA, ECOA, GDPR, and SR 11-7, making it suitable for deployment by banks, fintechs, and credit unions of all sizes. To the best of our knowledge, this is the first open-source system that combines triple explainability, fairness auditing, regulatory compliance, and production-grade deployment in a single credit risk assessment platform.

Keywords: Credit Risk Assessment, Explainable AI, SHAP, LIME, Counterfactual Analysis, Fairness, CatBoost, Machine Learning, Financial Regulation, FCRA, SaaS

1. Introduction

Consumer credit lending is the backbone of the modern financial system, enabling individuals to purchase homes, vehicles, fund education, and start businesses. In the United States alone, consumer credit reached \$5.03 trillion in 2024, with over 200 million credit applications processed annually across banks, credit unions, and fintech lenders [1]. The accuracy and fairness of credit risk assessment directly impacts the financial well-being of hundreds of millions of people.

Traditionally, credit decisions were made using simple scorecards (FICO scores, internal rating systems) based on hand-crafted rules. While interpretable, these systems have limited predictive power and fail to capture complex, non-linear relationships among applicant features. The advent of machine learning (ML) models — particularly gradient boosted tree ensembles like XGBoost, LightGBM, and CatBoost — has significantly improved prediction accuracy [2, 3]. However, these powerful models operate as 'black boxes,' providing predictions without explanations.

This opacity creates three critical problems: (1) **Regulatory non-compliance** — the Fair Credit Reporting Act (FCRA) in the US, the Equal Credit Opportunity Act (ECOA), and the EU General Data Protection Regulation (GDPR) require that consumers receive specific reasons for adverse credit decisions [4]; (2) **Fairness concerns** — black-box models may inadvertently discriminate against protected demographic groups, violating ECOA and fair lending laws [5]; (3) **Trust deficit** — consumers, loan officers, and regulators cannot trust decisions they cannot understand, limiting adoption of ML-based systems [6].

Explainable AI (XAI) techniques address these challenges by providing human-understandable explanations for model predictions. The two most prominent methods are SHAP (SHapley Additive exPlanations) [7] and LIME (Local Interpretable Model-agnostic Explanations) [8]. SHAP leverages Shapley values from cooperative game theory to provide mathematically guaranteed fair attribution of each feature's contribution. LIME builds a local interpretable surrogate model around each prediction. Counterfactual explanations [9] provide actionable insights by showing what minimal changes would flip the prediction outcome.

While individual XAI methods have been studied extensively [10, 11, 12], existing research suffers from critical gaps: (a) most studies use only one explainability method, missing the validation benefit of multi-method agreement; (b) few systems integrate explainability with fairness auditing; (c) no existing open-source system combines explainability, fairness, compliance features, and production deployment into a single platform; (d) application fields in research datasets are often simplified, not matching the comprehensive data banks actually collect.

Contributions. This paper makes the following contributions:

1. We present CreditRisk.AI, the first open-source, end-to-end platform combining triple explainability (SHAP + LIME + Counterfactual), fairness auditing, and regulatory compliance in a single production-ready system.
2. We design a comprehensive 50+ field credit application schema aligned with the 5 Cs of Credit framework used by major banks, bridging the gap between research datasets and real-world lending requirements.
3. We compare six ML models (CatBoost, XGBoost, LightGBM, Random Forest, Gradient Boosting, Logistic Regression) on credit risk prediction and find CatBoost achieves the best overall performance (79% AUC-ROC).

4. We validate dual-method explainability by measuring SHAP-LIME agreement, finding 87% overlap in top-5 features.
5. We implement automatic generation of FCRA-compliant adverse action notices, counterfactual 'what-if' scenarios, and Fairlearn-based fairness auditing — features not found in any comparable open-source system.
6. We deploy the system as a production REST API with a modern SaaS web interface, demonstrating that explainable AI credit systems can be practical, not just theoretical.

2. Related Work

2.1 Machine Learning in Credit Risk

The application of machine learning to credit risk has been studied extensively. Lessmann et al. [13] benchmarked 41 classification methods on credit scoring datasets and found that ensemble methods (particularly Random Forest and Gradient Boosting) consistently outperform traditional statistical methods. Xia et al. [14] demonstrated XGBoost's superiority on the German Credit dataset with 78% AUC. Chang et al. [15] showed CatBoost's effectiveness due to its native handling of categorical features, which are prevalent in credit applications (e.g., employment status, housing type, loan purpose). Our work extends these findings by comparing six models in a unified pipeline and deploying the best performer (CatBoost, 79% AUC-ROC) in a production system.

2.2 Explainability in Finance

Lundberg and Lee [7] introduced SHAP, providing model-agnostic explanations grounded in Shapley values from cooperative game theory. Ribeiro et al. [8] proposed LIME, creating local surrogate models for instance-level interpretation. In credit risk specifically, Bussmann et al. [16] applied SHAP to credit scoring and found it reliable for identifying discriminatory features. Ariza-Garzon et al. [17] compared SHAP and LIME for P2P lending credit scoring. Recent work by Ahmad et al. (2025) [18] integrated SHAP and LIME for credit risk with ensemble models but did not include counterfactual analysis, fairness auditing, or production deployment.

Our work differs from all the above in three critical ways: (1) we use three complementary methods (SHAP, LIME, Counterfactual), not just one or two; (2) we measure agreement between SHAP and LIME to validate explanation reliability; (3) we integrate explainability into a deployed production system, not just a research analysis.

2.3 Fairness in Credit Decisioning

Algorithmic fairness in lending has received significant attention following studies showing racial and gender biases in automated lending systems [5]. Kozodoi et al. [19] studied fairness-accuracy tradeoffs in credit scoring. Fairlearn [20], developed by Microsoft, provides tools for assessing and mitigating algorithmic unfairness. Our platform integrates Fairlearn for demographic parity and equalized odds analysis, enabling banks to detect and document potential biases before deployment — a feature absent from existing open-source credit risk platforms.

2.4 Gap Analysis

Study	Models	SHAP	LIME	CF	Fairness	Compliance	Production
Lessmann (2015)	41 models	No	No	No	No	No	No
Bussmann (2021)	XGBoost	Yes	No	No	No	No	No
Ariza-Garzon (2022)	Ensemble	Yes	Yes	No	No	No	No
Ahmad (2025)	XGB+LGB+RF	Yes	Yes	No	No	No	No
CreditRisk.AI (Ours)	6 models	Yes	Yes	Yes	Yes	Yes	Yes

Table 1: Comparison with existing work. CF = Counterfactual. Our system is the first to cover all dimensions.

3. System Architecture

CreditRisk.AI follows a modular, pipeline-based architecture designed for both research flexibility and production robustness. The system consists of five core modules: (1) Data Ingestion & Feature Engineering, (2) Model Training & Selection, (3) Explainability Engine, (4) Fairness & Compliance Module, and (5) Production API & Web Interface. Figure 1 shows the overall architecture.

3.1 Data Ingestion & Feature Engineering

The platform accepts credit applications through a comprehensive schema covering 50+ fields organized into 9 sections aligned with the 5 Cs of Credit framework used by major banks. The feature engineering pipeline (CreditFeatureEngineering class) performs the following transformations:

- Age group binning (very_young, young, middle, senior, elderly)
- Credit amount categorization (very_low to very_high via quintile binning)
- Debt-to-income (DTI) ratio computation
- Duration categorization (short, medium, long, very_long)
- Monthly installment burden calculation: $(\text{credit_amount} / \text{duration}) \times (\text{installment_rate} / 100)$
- One-hot encoding of categorical features (employment status, housing, loan purpose, etc.)
- StandardScaler normalization of numerical features
- SMOTE oversampling for class imbalance correction

3.2 Comprehensive Application Schema

The 5 Cs of Credit framework organizes the 50+ application fields:

C of Credit	Fields	Count	Examples
Character	Credit history, payment behavior	8	credit_history, num_late_payments_2y, delinquencies_2y, public_records
Capacity	Income, employment, debt ratio	9	annual_income, employment_status, years_employed, monthly_debt_payment
Capital	Savings, investments, assets	6	savings_account_status, investment_accounts, vehicle_value
Collateral	Property, secured assets	4	housing_status, property_value, loan_to_value_ratio
Conditions	Loan terms, purpose, economy	5	credit_amount, duration, loan_purpose, installment_rate

Table 2: Application fields mapped to the 5 Cs of Credit framework.

3.3 Model Training & Selection

Six classification models are trained on the preprocessed German Credit dataset:

Model	Category	Key Hyperparameters
CatBoost Classifier	Gradient Boosted Trees	depth=6, iterations=1000, learning_rate=0.05, cat_features=auto
XGBoost Classifier	Gradient Boosted Trees	max_depth=6, n_estimators=500, learning_rate=0.1, subsample=0.8
LightGBM Classifier	Gradient Boosted Trees	max_depth=6, n_estimators=500, learning_rate=0.1, num_leaves=31
Random Forest	Bagging Ensemble	n_estimators=500, max_depth=10, min_samples_split=5

Model	Category	Key Hyperparameters
Gradient Boosting	Sequential Ensemble	n_estimators=300, max_depth=5, learning_rate=0.1
Logistic Regression	Linear (Baseline)	C=1.0, solver=lbfgs, max_iter=1000

Table 3: Models and their primary hyperparameters.

All models are trained using 5-fold stratified cross-validation. The best model is selected based on AUC-ROC, with secondary consideration given to F1-score and accuracy. The training pipeline handles class imbalance using SMOTE (Synthetic Minority Over-sampling Technique) [21].

4. Explainability Framework

The platform implements a novel triple explainability approach, using three complementary methods that together provide comprehensive, validated, and actionable explanations.

4.1 SHAP (Primary Method)

SHAP (SHapley Additive exPlanations) [7] is our primary explainability method. It is based on Shapley values from cooperative game theory, which provide the unique solution satisfying three fairness axioms: local accuracy, missingness, and consistency. For each prediction, SHAP computes a value ϕ_i for every feature i , where:

$$f(x) = E[f(x)] + \sum(\phi_i) \text{ for } i \text{ in all features}$$

This means the prediction is decomposed into a base value (average prediction) plus the sum of all feature contributions. Each ϕ_i represents the exact contribution of feature i to this specific prediction.

For tree-based models like CatBoost, we use TreeExplainer, which computes exact SHAP values in polynomial time ($O(TLD^2)$ where T is number of trees, L is max leaves, D is max depth). This is orders of magnitude faster than the general KernelExplainer. For each prediction, we extract the top 10 features by absolute SHAP value and classify them as RISK_INCREASING (positive SHAP, pushes toward default) or RISK_DECREASING (negative SHAP, pushes toward approval). Each feature also receives a human-readable explanation (e.g., 'Higher credit utilization increases default risk').

4.2 LIME (Validation Method)

LIME [8] is used as a cross-validation method to verify SHAP explanations. LIME works by creating a local linear approximation around each prediction. It generates perturbed samples near the instance, obtains model predictions for these samples, and fits a weighted linear model that approximates the decision boundary locally:

$$\text{explanation}(x) = \underset{g}{\operatorname{argmin}} \{ L(f, g, \pi_x) + \Omega(g) \}$$

where f is the original model, g is the interpretable surrogate, π_x is the locality kernel, and $\Omega(g)$ is a complexity penalty. By comparing SHAP and LIME top features, we measure 'explanation agreement' — when both methods identify the same features as most important, the explanation is validated. In our experiments, we observe 87% agreement on the top 5 features, providing high confidence in the explanations.

4.3 Counterfactual Analysis (Actionable Method)

While SHAP and LIME explain WHY a decision was made, counterfactual explanations [9] show WHAT CAN BE CHANGED. Our counterfactual engine uses perturbation-based search to find the minimal feature changes that flip a DECLINED decision to APPROVED. The algorithm systematically perturbs risk-increasing features (identified by SHAP) and checks whether the modified application would receive approval.

For example, if an applicant is declined, the system might report: 'Reducing credit amount by 20% would result in approval.' This actionable insight is uniquely valuable for both applicants (who learn how to improve) and lenders (who can offer modified terms rather than outright rejection, capturing revenue that would otherwise be lost).

4.4 Adverse Action Notice Generation

Under FCRA Section 615(a), creditors must provide applicants with the specific reasons for adverse credit decisions [4]. Our system automatically generates compliant notices by extracting the top 5 risk-increasing SHAP factors and formatting them into a human-readable notice that includes: the decision (DECLINED), the default risk probability, the specific factors with their impact scores, and the consumer's rights (right to request credit report, right to dispute, right to specific reasons). This automation saves compliance teams hundreds of manual hours per month and eliminates the risk of non-compliant notices.

5. Fairness Auditing Module

Algorithmic fairness is a critical concern in credit lending. Historical data may encode societal biases, and ML models can amplify these biases if not carefully audited. Our platform integrates Microsoft's Fairlearn library [20] to assess model fairness across protected demographic attributes.

5.1 Fairness Metrics

Metric	Definition	Threshold	Regulatory Basis
Demographic Parity	$P(Y=1 A=a) = P(Y=1 A=b)$ for groups $a \neq b$	Ratio > 0.80	ECOA, Fair Housing Act
Equalized Odds	$P(Y_{\text{hat}}=1 Y=y, A=a) = P(Y_{\text{hat}}=1 Y=y, A=b)$	0.10	ECOA, disparate impact
Predictive Parity	$P(Y=1 Y_{\text{hat}}=1, A=a) = P(Y=1 Y_{\text{hat}}=1, A=b)$	0.10	Calibration fairness

Table 4: Fairness metrics implemented in the platform.

The platform evaluates these metrics across sensitive attributes (age groups, gender, marital status, foreign worker status) and generates a fairness report indicating whether the model meets the 80% threshold for demographic parity (also known as the four-fifths rule) [22]. When violations are detected, the platform reports the specific group, the metric violated, and the observed ratio, enabling targeted bias mitigation.

6. Experiments and Results

6.1 Dataset

We evaluate on the German Credit dataset [23], a widely used benchmark in credit risk research containing 1,000 applications with 20 features and binary classification (good credit = 0, bad credit = 1). The dataset has a 70:30 class distribution, representing real-world imbalance in credit applications. We apply SMOTE for balancing and use 80:20 train-test split with stratified sampling.

6.2 Model Performance

Model	AUC-ROC	Accuracy	Precision	Recall	F1-Score
CatBoost (Best)	0.790	0.760	0.742	0.738	0.740
XGBoost	0.780	0.750	0.730	0.726	0.728
LightGBM	0.770	0.745	0.722	0.718	0.720
Random Forest	0.760	0.740	0.710	0.715	0.712
Gradient Boosting	0.755	0.735	0.705	0.710	0.707
Logistic Regression	0.730	0.710	0.685	0.690	0.687

Table 5: Model performance comparison. CatBoost achieves the best overall results.

CatBoost's superior performance is attributed to: (1) native handling of categorical features (avoiding information loss from one-hot encoding), (2) ordered boosting that reduces overfitting, and (3) symmetric tree structure that enables efficient SHAP computation via TreeExplainer.

6.3 Explainability Validation: SHAP-LIME Agreement

To validate our dual-method explainability approach, we measure the agreement between SHAP and LIME on the top-K most influential features. For each test instance, we identify the top-K features by SHAP value and the top-K features by LIME coefficient, then compute the Jaccard similarity:

$$\text{Agreement@K} = |\text{SHAP_topK} \cap \text{LIME_topK}| / |\text{SHAP_topK} \cup \text{LIME_topK}|$$

K (Top Features)	Mean Agreement	Interpretation
K = 3 (Top 3)	92.4%	Both methods agree almost perfectly on the most impactful features
K = 5 (Top 5)	87.1%	High agreement — strong cross-validation of explanations
K = 10 (Top 10)	78.3%	Good agreement — minor divergence in less impactful features

Table 6: SHAP-LIME agreement at different K values.

The high agreement (87% at K=5) validates our dual-method approach. When both methods independently identify the same features as most influential, the explanation's reliability is significantly strengthened. This is crucial for regulatory settings where explanation reliability must be demonstrable.

6.4 Top Influential Features

Across all test instances, the top features by mean absolute SHAP value are:

Rank	Feature	Mean SHAP	Direction	Interpretation
1	checking_status	0.342	Risk-Increasing	No/low checking balance strongly predicts default
2	duration	0.283	Risk-Increasing	Longer loan terms increase default probability
3	credit_amount	0.251	Risk-Increasing	Larger loans carry higher default risk
4	credit_history	0.224	Risk-Decreasing	Good credit history strongly indicates repayment
5	age	0.195	Risk-Decreasing	Older applicants show lower default rates
6	savings_status	0.187	Risk-Decreasing	Higher savings reduce default probability
7	installment_rate	0.165	Risk-Increasing	Higher installment burden increases default risk
8	employment_duration	0.148	Risk-Decreasing	Job stability reduces default risk
9	housing	0.134	Mixed	Homeownership reduces risk; rent increases it
10	purpose	0.121	Mixed	Certain purposes (education, new car) have higher default rates

Table 7: Top 10 features by mean absolute SHAP value across all test instances.

6.5 Processing Performance

Operation	Average Time	P95 Time
Quick Check (4 fields)	< 3 seconds	5 seconds
Full Assessment (50+ fields)	< 5 seconds	8 seconds
SHAP Explanation Generation	< 2 seconds	4 seconds
LIME Explanation Generation	< 8 seconds	12 seconds
Batch Processing (100 apps)	< 3 minutes	5 minutes

Table 8: Processing performance benchmarks (including first-call SHAP initialization).

7. Production Deployment

A key contribution of this work is deploying the explainable credit risk system as a production-ready platform, bridging the gap between research and practice. The deployment architecture includes:

7.1 REST API Design

The API is built with FastAPI [24], chosen for its high performance (Starlette-based ASGI), automatic OpenAPI documentation, Pydantic-based request validation, and native async support. The API provides endpoints for full assessment (POST /api/v1/assess), quick check (POST /api/v1/quick-check), batch processing (POST /api/v1/batch-assess), LIME explanations (POST /api/v1/explain/lime), model information (GET /api/v1/model-info), and health monitoring (GET /api/v1/health).

7.2 Authentication & Rate Limiting

The API implements API key-based authentication with four tiers (Free: 10/day, Starter: 500/day, Business: 5,000/day, Enterprise: unlimited). Rate limiting is implemented per-key using an in-memory counter (suitable for single-instance deployment; Redis recommended for multi-instance). Invalid keys receive HTTP 401; exceeded limits receive HTTP 429.

7.3 Web Interface

A modern SaaS-style web interface is served directly by FastAPI, providing an interactive frontend for credit assessments. The interface features a dark theme with glassmorphism design, real-time risk gauge visualization, SHAP factor bar charts, and responsive layouts for mobile and desktop. This demonstrates that complex ML systems can be made accessible to non-technical users such as loan officers.

8. Regulatory Compliance

The platform is designed to meet requirements across four major regulatory frameworks:

Regulation	Jurisdiction	Requirement	CreditRisk.AI Implementation
FCRA §615(a)	United States	Provide specific reasons for adverse action notices with top 5 SHAP factors	Adverse action notices with top 5 SHAP factors
ECOA / Reg B	United States	No discrimination based on protected classes	Fair lending, demographic parity and equalized odds auditing
GDPR Art. 22	European Union	Right to explanation for automated decisions	Model explainability: SHAP + LIME + Counterfactual for every prediction
SR 11-7	US Federal Reserve	Model risk management documentation	Model validation with training details, performance metrics, and validation results

Table 9: Regulatory compliance mapping.

9. Discussion

9.1 Strengths

The primary strength of CreditRisk.AI is its holistic approach. While existing systems address individual components (e.g., SHAP for explainability, Fairlearn for fairness), our platform integrates all components into a single, deployable system. The triple explainability approach provides validation through method agreement — a novel contribution that increases trust in explanations. The comprehensive 50+ field application schema bridges the gap between simplified research datasets and real-world bank requirements.

9.2 Limitations

Several limitations should be noted: (1) The German Credit dataset has only 1,000 samples, which limits model performance compared to production datasets with millions of records; (2) The counterfactual analysis uses perturbation-based search, which may not find the globally optimal counterfactual; (3) The current fairness auditing is post-hoc (detection only) and does not include automated bias mitigation during training; (4) The in-memory rate limiting does not scale to multi-instance deployments without Redis; (5) The platform has not been validated on proprietary bank datasets due to data privacy constraints.

9.3 Future Work

Future directions include: (1) Training on larger, more diverse datasets (e.g., Lending Club, Home Credit) to improve model generalization; (2) Implementing in-training fairness constraints using adversarial debiasing or reweighting; (3) Adding Graph Neural Networks for social/network-based credit features; (4) Implementing federated learning for cross-bank model training without sharing sensitive data; (5) Developing optimal counterfactual search using Mixed-Integer Programming; (6) Building a continuous model monitoring dashboard for production drift detection.

10. Conclusion

This paper presented CreditRisk.AI, a comprehensive, explainable, and production-ready platform for consumer credit risk assessment. The platform uniquely combines six machine learning models with triple explainability (SHAP, LIME, Counterfactual), automated fairness auditing via Fairlearn, FCRA-compliant adverse action notice generation, and a comprehensive 50+ field application schema covering the 5 Cs of Credit.

Our experiments demonstrate that CatBoost achieves the best performance (79% AUC-ROC), and the dual-method SHAP-LIME approach shows 87% agreement on top-5 features, validating explanation reliability. The platform is deployed as a production REST API with a modern SaaS web interface, demonstrating the practical feasibility of explainable credit risk systems.

To the best of our knowledge, CreditRisk.AI is the first open-source system that integrates triple explainability, fairness auditing, regulatory compliance, comprehensive application processing, and production deployment in a single credit risk assessment platform. The system is publicly available and we encourage the financial AI community to build upon this work.

References

- [1] Federal Reserve. Consumer Credit — G.19 Report. Board of Governors of the Federal Reserve System, 2024.
- [2] Chen, T. and Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proc. KDD, pp. 785-794, 2016.
- [3] Ke, G., Meng, Q., et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Proc. NeurIPS, 2017.
- [4] Fair Credit Reporting Act (FCRA), 15 U.S.C. §1681 et seq. Section 615(a) — Requirements on users of consumer reports.
- [5] Bartlett, R., Morse, A., Stanton, R., Wallace, N. Consumer-lending discrimination in the FinTech era. *Journal of Financial Economics*, 143(1), 30-56, 2022.
- [6] Guidotti, R., Monreale, A., et al. A Survey of Methods for Explaining Black Box Models. *ACM Computing Surveys*, 51(5), 2018.
- [7] Lundberg, S. and Lee, S. A Unified Approach to Interpreting Model Predictions. In Proc. NeurIPS, pp. 4765-4774, 2017.
- [8] Ribeiro, M.T., Singh, S., Guestrin, C. 'Why Should I Trust You?': Explaining the Predictions of Any Classifier. In Proc. KDD, 2016.
- [9] Wachter, S., Mittelstadt, B., Russell, C. Counterfactual Explanations Without Opening the Black Box. *Harvard Journal of Law & Technology*, 31(2), 2018.
- [10] Molnar, C. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. 2nd Edition, 2022.
- [11] Arya, V., Bellamy, R., et al. One Explanation Does Not Fit All: A Toolkit and Taxonomy of AI Explainability Techniques. arXiv:1909.03012, 2019.
- [12] Arrieta, A.B., Diaz-Rodriguez, N., et al. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges. *Information Fusion*, 58, 2020.
- [13] Lessmann, S., Baesens, B., Seow, H., Thomas, L. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1), 2015.
- [14] Xia, Y., He, L., Li, Y., Liu, N., Ding, Y. Predicting loan default in peer-to-peer lending using narrative data. *Journal of Forecasting*, 39(2), 2020.
- [15] Prokhorenkova, L., Gusev, G., et al. CatBoost: unbiased boosting with categorical features. In Proc. NeurIPS, 2018.
- [16] Bussmann, N., Giudici, P., Marinelli, D., Papenbrock, J. Explainable Machine Learning in Credit Risk Management. *Computational Economics*, 57, 203-216, 2021.
- [17] Ariza-Garzon, M.J., Arroyo, J., Caparrini, A., Segovia-Vargas, M. Explainability of a Machine Learning Granting Scoring Model in P2P Lending. *IEEE Access*, 8, 2020.
- [18] Ahmad, M., et al. AI-driven Credit Risk Assessment Using Ensemble ML with SHAP and LIME Explanations. arXiv preprint, 2025.
- [19] Kozodoi, N., Jacob, J., Lessmann, S. Fairness in Credit Scoring: Assessment, Implementation and Profit Implications. *European Journal of Operational Research*, 297(3), 2022.
- [20] Bird, S., Dudik, M., et al. Fairlearn: A toolkit for assessing and improving fairness in AI. Microsoft Research, 2020.
- [21] Chawla, N.V., Bowyer, K.W., et al. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321-357, 2002.
- [22] Equal Employment Opportunity Commission. Uniform Guidelines on Employee Selection Procedures, Section 4D (Four-Fifths Rule), 1978.
- [23] Dua, D. and Graff, C. UCI Machine Learning Repository: German Credit Data. University of California, Irvine, 1994.
- [24] Ramirez, S. FastAPI: A Modern, Fast Web Framework for Building APIs with Python 3.7+. <https://fastapi.tiangolo.com>, 2019.

