

Leakage-Aware Explainable AI for Credit Risk Assessment of Perishable Food Supply-Chain SMEs Using Food Shelf-Life, Postharvest Loss and Digital Banking Data in Nigeria

Kosisochukwu Henry Ukpabi¹, Abdullahi Mohammed Ibrahim¹, Zainab Abdulsalam Jamilu²

¹Department of Computer Science, Federal University Dutse

²Department of Computer Science, Jigawa State Polytechnic Dutse

³Department of Nutrition and Dietetics, Jigawa State Polytechnic Dutse

Abstract — Small and medium-sized enterprises (SMEs) operating in perishable food supply chains face intertwined operational and financial risks: inventory can deteriorate before sale, postharvest losses reduce cash available for repayment, and many firms have thin or irregular formal credit histories. This study proposes a leakage-aware, explainable artificial intelligence (AI) framework that integrates food shelf-life, postharvest-loss, commodity-price, and digital banking features to assess credit risk for Nigerian perishable-food SMEs. Because linked Nigerian lender-level repayment data and firm-level shelf-life measurements were not available for this research, the empirical demonstration uses a clearly labelled dataset rather than claiming results from real borrowers. A reproducible simulation of 5,000 SME-loan observations was specified, with 70%/15%/15% chronological train/validation/test partitions, a financial-only baseline, and an integrated model using LightGBM-style gradient boosting. Logistic regression and random forest were included as benchmarks; SHAP-style feature attribution was used for interpretability. The simulated results illustrate the intended evaluation, not observed Nigerian banking performance: the financial-only baseline achieved a test ROC-AUC of 0.742 and PR-AUC of 0.356, while the integrated model achieved ROC-AUC 0.801 and PR-AUC 0.421. At a threshold selected on validation data, recall increased from 0.61 to 0.70, while Brier score improved from 0.176 to 0.158. The largest simulated contributors were recent repayment behaviour, cash-flow volatility, estimated remaining shelf life, postharvest-loss rate, and debt-service coverage. A random split produced more optimistic estimates than chronological evaluation, demonstrating why time-aware validation is essential. The main contributions are (i) a domain-specific framework linking food deterioration to credit risk, (ii) a leakage-control protocol for temporal financial data, (iii) an explainability and calibration evaluation plan, and (iv) a transparent synthetic benchmark that can be replaced by real data. These results are illustrative only; empirical conclusions require validation using consented, anonymised lending outcomes and measured food-quality data.

Keywords— Credit risk; explainable AI; food technology; shelf-life prediction; postharvest loss; digital banking; SME finance; temporal data leakage; Nigeria

I. INTRODUCTION

Credit risk assessment is central to responsible lending, but conventional models can perform poorly when borrowers operate in volatile sectors and have limited formal financial histories. Perishable-food supply-chain SMEs including produce aggregators, wholesalers, cold-chain operators, processors, and market distributors face a distinctive combination of risks (Janeiro, 2025). Their repayment capacity depends not only on sales and cash flow but also on inventory quality, time-to-sale, storage conditions, transport delays, and commodity-price changes. Food that loses marketable quality can quickly become a financial loss, shrinking working capital and weakening debt-service capacity (Cortes et al., 2026). AI-

based credit scoring can learn nonlinear interactions among repayment behaviour, transaction patterns, and business indicators (Zhang, 2024). However, credit models must be assessed for calibration, stability, fairness, and operational validity, not only headline accuracy. A common threat is temporal data leakage: variables recorded after a lending decision or after the default outcome can inadvertently enter training features, producing performance that would not be available at the time of decision (Chen, 2026). Random train-test splits can also mix future and past observations and overstate generalisation under changing economic conditions. Recent work on temporal leakage and explainable credit modelling reinforces the need for chronological evaluation and stable explanations (Petrović et al., 2021).

A parallel stream of food-technology research uses computer vision, storage sensors, and machine learning to estimate food freshness and remaining shelf life (Cortes et al., 2026). For example, recent work on tomato shelf-life prediction has fused visual information with temperature data, while experimental studies have modelled tomato quality changes during storage (Senge et al., 2025). Nigeria's National Agricultural Sample Survey (NASS) provides a potential contextual source for agricultural production and reported harvest/postharvest losses; the FAO Food Loss and Waste Database offers a wider collection of reported food-loss estimates (Bureau, 2025). These resources do not, by themselves, provide linked SME loan outcomes, so they must not be misrepresented as a complete credit-scoring dataset.

This study addresses the intersection by proposing a framework that augments financial credit-risk indicators with food-quality and inventory-risk signals. This research has two purposes: define a testable research design and demonstrate its evaluation through a synthetic experiment. The simulated results are explicitly illustrative and must not be interpreted as findings from Nigerian banks or actual food businesses. The aim is to design and demonstrate a leakage-aware, explainable AI framework for credit-risk assessment of SMEs handling perishable food in Nigeria.

RQ1: Do food shelf-life and postharvest-loss features improve simulated credit-default discrimination beyond financial and digital-transaction features? RQ2: How does chronological validation change estimated model performance relative to random splitting? RQ3: Which features contribute most to predictions, and are those contributions operationally plausible and stable enough to support human review?

II. RELATED WORK AND RESEARCH GAP

Credit scoring has progressed from statistical models such as logistic regression to tree ensembles and other machine-learning methods capable of representing nonlinear relationships. Yet a stronger score is not sufficient for responsible deployment: the model must be calibrated, tested on later periods, and explainable to credit officers and affected businesses. Research on temporal leakage in credit risk warns that random splitting and post-outcome features can yield misleading performance (Passalacqua, 2021). Work on explainable dynamic credit assessment similarly identifies the difficulty of integrating external time-series data without leakage or opaque interactions (Ghadimpour & Ebrahimi, 2022).

Food shelf-life prediction is increasingly based on non-destructive sensing. A 2026 study described a smartphone-based tomato shelf-life model combining image-based freshness classification and temperature features, reporting promising results in its own experimental setting (Shehzad, 2025). A 2025 study also examined tomato quality during storage using laboratory measurements and machine learning [4]. These results support the feasibility of generating food-condition features, but performance from a particular crop, cultivar, or storage environment cannot be assumed to transfer directly to Nigerian markets (Christodoulou et al., 2025).

For Nigerian context, the NASS 2023 catalogue includes harvest quantities, reported losses, loss causes, and state/location descriptors in agricultural household data (Catalogue, 2025). The FAO Food Loss and Waste Database aggregates evidence across commodities, value-chain stages, and geographies (Congress, n.d.). These sources can inform feature definitions, sampling plans, and contextual analysis, but they do not substitute for loan-level outcomes. Digital lending research in Nigeria also highlights the importance of data protection and responsible use of alternative data (Messinis et al., 2024).

The research gap is therefore not the generic use of AI for lending or AI for food quality. It is the test of whether contemporaneously available food-condition indicators add incremental predictive value to credit-risk models for perishable-food SMEs, under leakage-resistant validation, with explanations and calibration assessed alongside discrimination. To the best of the present design, this specific linked-data question remains to be tested using an appropriately governed Nigerian dataset; no claim is made that a comprehensive systematic review has established absolute novelty.

III. METHODOLOGY

Study design. The study uses a quantitative predictive-modelling design with a synthetic proof-of-concept experiment. The target unit is a loan application or loan facility to an SME whose principal activity is the aggregation, storage, transport, wholesale, or processing of perishable food. The prediction point is the date of credit decision. The outcome is a binary default indicator defined prospectively for example, reaching 90 or more days past due within 12 months after origination. The definition must be aligned with the participating lender's policy and applied consistently.

Data sources and availability. A real implementation would use four linked sources: (a) consented and anonymised lender

data containing application-time financial features and subsequent repayment outcomes; (b) digital banking summaries such as transaction inflows/outflows, cash-flow volatility, balance trends, and returned-payment frequency; (c) food-quality observations from images and storage sensors, including temperature, relative humidity, visible defect class, and measured remaining shelf life; and (d) commodity prices and postharvest-loss context from field measurements and public sources such as NASS and FAO. Public agricultural datasets may support context and feature design, but must not be joined to individual borrowers without a defensible matching basis and lawful authority.

The numerical results below are based on a dataset of 5,000 records. Simulated variables included loan amount, business tenure, debt-service coverage ratio (DSCR), cash-flow volatility, recent repayment history, digital transaction regularity, commodity-price volatility, remaining shelf life (days), postharvest-loss rate (%), storage-temperature excursions, and default within 12 months. The data-generating process intentionally assigned higher default probability to weaker repayment history, lower DSCR, greater cash-flow volatility, shorter shelf life, and higher loss rates, while allowing correlated predictors. This design tests the pipeline; it cannot establish real-world effect sizes.

Feature construction. Financial-only features comprised loan amount, business age, DSCR, recent delinquency, inflow/outflow summaries, and transaction regularity. The integrated model added estimated remaining shelf life, postharvest-loss rate, temperature-excursion frequency and commodity-price volatility. Features were computed only from information available at the prediction date. Data captured after approval, after a missed payment, or after the defined default window were excluded. Shelf-life estimates used for scoring would need to be produced by a separately validated food-quality model and timestamped before the credit decision.

Models. Logistic regression served as an interpretable baseline. Random forest and gradient-boosted trees represented nonlinear alternatives. The primary integrated model was a LightGBM-style gradient-boosting classifier. Hyperparameters were selected on the validation period only. Class imbalance was handled through class weights or threshold selection, not by duplicating future test observations. Model comparisons used identical chronological splits and the same outcome definition.

Leakage-aware validation. Records were ordered by application date and partitioned into 70% training, 15%

validation, and 15% test observations by time. The training period was used for fitting; validation data supported hyperparameter and threshold selection; the final test period was used once for reporting. Preprocessing, missing-value imputation, scaling, encoding, and feature selection were fitted within the training data and then applied to later partitions. A random split was run only as a diagnostic comparison to demonstrate optimism that can occur when temporal structure is ignored. Where borrower IDs exist, a group-aware check should ensure that the same borrower does not inadvertently appear across partitions in a way that leaks repeated-facility information. Metrics and explainability. Discrimination was assessed using ROC-AUC and precision-recall AUC (PR-AUC), particularly relevant when defaults are relatively uncommon. Calibration was assessed with the Brier score; thresholded classification was summarised with recall and precision. A real study should additionally report confidence intervals using bootstrap resampling that respects time blocks, subgroup metrics, and decision-cost analyses. SHAP values were used in the synthetic demonstration to illustrate local/global attribution. Explanations describe model associations, not causal effects; correlated features can share or exchange attribution.

IV. PROPOSED FRAMEWORK

Figure 1 presents the proposed end-to-end pipeline. The key control is the decision-time feature boundary: only evidence available on or before the loan decision can enter the model.



Figure 1. Proposed Leakage-Aware Explainable AI Pipeline

V. RESULTS

Important reporting note: all numerical values in this section are synthetic illustrative outputs created to demonstrate how the analysis could be reported. They were not obtained from actual Nigerian bank records, an online borrower-level dataset, or a completed food-storage experiment. They must be replaced by results from a reproducible run on validated data before the manuscript is submitted as an empirical study.

Synthetic cohort

The simulated dataset contained 5,000 loan observations, of which 14.0% were assigned the default outcome. The temporal split allocated 3,500 observations to training, 750 to validation, and 750 to testing. Food-related features were generated with plausible dependencies on storage conditions and commodity handling, and default probabilities were generated from a logistic process involving both financial and food-condition signals.

Table 1. Synthetic dataset summary

Variable	Illustrative value
Total loan observations	5,000
Training observations (earliest 70%)	3,500

Validation observations (next 15%)	750
Test observations (latest 15%)	750
Overall simulated default prevalence	14.0%
Median remaining shelf life	4.5 days
Median postharvest-loss rate	8.0%
Primary commodity scenario	Perishable produce (tomato-focused design)

The values in Table 1 are simulation assumptions, not population estimates for Nigerian SMEs. A tomato-focused implementation is proposed because recent shelf-life modelling research provides a plausible starting point for the food-quality component

Predictive performance

Table 2 shows the intended comparison under chronological testing. The integrated model improves simulated ROC-AUC, PR-AUC, recall, and Brier score over the financial-only baseline. The magnitude of these differences is predetermined by the illustrative data-generating process and should not be interpreted as evidence that food features will improve a real lender's model by the same amount.

Table 2. Illustrative chronological test performance

Model	Feature set	ROC-AUC	PR-AUC	Recall*	Brier score
Logistic regression	Financial + digital banking	0.716	0.321	0.57	0.184
Random forest	Financial + digital banking	0.735	0.344	0.60	0.179
Gradient boosting	Financial + digital banking	0.742	0.356	0.61	0.176
Integrated gradient boosting	Financial + digital + food	0.801	0.421	0.70	0.158
Recall is shown at a decision threshold selected on validation data. Values are illustrative simulation outputs.					

The expected interpretation is that shelf-life and postharvest-loss indicators may carry incremental information about cash-generation risk in businesses where inventory deterioration directly affects revenue.

However, an apparent gain could also arise from confounding—for example, better-managed firms may both store food more effectively and manage cash flow better. Real validation must therefore assess incremental value beyond business size, seasonality, commodity mix, region, and other available confounders.

Leakage diagnostic

Table 3 contrasts a random split with chronological testing for the integrated model. The random split yields higher apparent performance in this illustrative scenario because neighbouring periods share patterns and the distribution is less challenging than a future-period test.

This does not imply that every random split always inflates results; it shows why the validation scheme must reflect the intended deployment setting.

Table 3. Illustrative validation-scheme comparison

Evaluation scheme	ROC-AUC	PR-AUC	Interpretation
Random 70/15/15 split	0.836	0.468	Optimistic diagnostic; temporal order ignored
Chronological 70/15/15 split	0.801	0.421	Primary estimate for future-period use
Chronological split with post-decision leakage field	0.914	0.612	Invalid demonstration; excluded from valid model comparison

The leakage-field row is a deliberately invalid diagnostic showing how a feature derived from post-decision repayment status could create implausibly high metrics. It is not a candidate model and must never be reported as valid predictive performance. In practice, leakage tests should include timestamp audits, target-correlated field reviews, and checks for outcome-derived variables.

Explainability

In the simulated integrated model, the illustrative global ranking placed recent repayment history and DSCR among the strongest financial signals, followed by cash-flow volatility, remaining shelf life, postharvest-loss rate, and temperature excursions. This ordering is plausible by design, not an empirical discovery. SHAP dependence plots should be inspected for non-monotonic effects and interactions, but no causal claim should be inferred. If short shelf life is strongly correlated with storage temperature or business type, attribution may be distributed among these features.

Table 4. Illustrative top feature groups

Rank	Feature group	Expected relationship in simulation	Interpretation for review
1	Recent repayment behaviour	More delinquencies → higher risk	Direct indicator of past repayment difficulty
2	DSCR / cash-flow coverage	Lower coverage → higher risk	Less capacity to service debt
3	Cash-flow volatility	Higher volatility → higher risk	Uncertainty in operating cash generation

4	Remaining shelf life	Fewer days → higher risk	Less time to sell inventory before deterioration
5	Postharvest-loss rate	Higher loss → higher risk	Lower saleable inventory and potential revenue
6	Temperature excursions	More excursions → higher risk	Possible deterioration risk; requires sensor validation
7	Commodity-price volatility	Higher volatility → higher risk	Uncertainty in inventory value and sales margin

Calibration, Robustness, and Fairness

The integrated model's illustrative Brier score of 0.158 compared with 0.176 for the financial-only gradient-boosting baseline suggests improved probability accuracy in the synthetic test set. A real deployment should also report calibration curves, expected calibration error, and reliability by time period. Robustness should be checked under high inflation, transport disruption, unusual weather, and sudden commodity-price changes. Fairness audits should test whether error rates differ by region, business size, gender of business owner where lawfully and ethically collected, and formal/informal business status; these variables should be used carefully for auditing, not automatically as predictive inputs.

VI. DISCUSSION

This demonstration suggests a testable hypothesis: physical food-condition information can complement financial records when inventory deterioration is a direct driver of SME cash-flow risk. For a lender financing tomato wholesalers or other perishable-food businesses, shelf-life estimates may help identify cases in which inventory is unlikely to be sold before quality declines. Postharvest-loss rates may similarly indicate lower recoverable inventory value. The value of these features is likely to vary by commodity, season, storage infrastructure, and the structure of the loan.

The proposed integration also creates practical challenges. First, food shelf-life estimates are uncertain and depend on cultivar, maturity, handling, temperature, humidity, packaging, and transport. A food model trained in a laboratory or another

country may not generalise to Nigerian market conditions. Second, a lender may not have direct access to reliable sensor data for each borrower. Third, loan default is influenced by many factors unrelated to food quality, including interest rates, household shocks, market access, fraud, and loan design. A responsible system should therefore treat food indicators as additional evidence, not as a substitute for standard underwriting.

The leakage-aware design is a central methodological contribution. Features must be frozen at the decision timestamp; imputation and tuning must not use future observations; and the final test set should represent a later period. Random splitting can still be useful for exploratory work but should not be the primary evaluation when the model is intended to score future loan applications. Model comparisons should include calibration and decision costs, since a modest AUC improvement may not translate into improved lending decisions if false positives are costly or probabilities are poorly calibrated.

The study has important limitations. The performance values are synthetic, not empirical; the assumed 14% default rate is not an estimate for Nigeria; and the assumed model gains cannot establish predictive utility. Public food-loss and agricultural survey datasets provide contextual information but are not linked to digital banking transactions or SME defaults. The proposed framework therefore remains a research design awaiting real-world validation. The synthetic data-generation process is intentionally simplified and cannot reproduce all dependencies, missingness, selection bias, and distribution shifts found in real lending portfolios.

VII. CONCLUSION

This research proposes a leakage-aware, explainable AI framework for assessing credit risk among Nigerian SMEs handling perishable food. The approach integrates conventional financial and digital-banking indicators with food shelf-life, postharvest-loss, temperature-excursion, and commodity-price features. It combines chronological validation, baseline comparisons, calibration, SHAP-based explanation, and governance checks to support a transparent evaluation. In the demonstration, the integrated model achieved illustrative ROC-AUC of 0.801 and PR-AUC of 0.421, compared with 0.742 and 0.356 for the financial-only gradient-boosting baseline. These figures are not findings from real borrowers and must not be used to claim demonstrated performance in Nigeria. The appropriate conclusion is narrower: the proposed design is

testable and specifies how incremental value could be assessed without temporal leakage.

The next research step is to build a governed, linked dataset combining application-time lender records with measured food-condition indicators for a defined commodity and region. The final study should preregister the outcome definition, report time-based test results with uncertainty intervals, compare against strong financial-only baselines, and assess calibration, fairness, and operational cost. Only then can the framework support evidence-based recommendations for Nigerian banks, microfinance institutions, and responsible digital lenders.

Data availability statement

No real borrower-level dataset was analysed for this manuscript. All numerical results are synthetic and illustrative. Potential contextual sources include the Nigeria National Agricultural Sample Survey 2023 catalogue and the FAO Food Loss and Waste Database. These sources do not contain the linked lender outcomes required to validate the proposed credit-risk model. A future study should publish code and a data dictionary where legally permissible, while keeping personal and confidential banking records protected.

REFERENCES

1. Bureau, N. (2025). *Nigeria - National Agricultural Sample Survey 2023 National Agricultural Sample Survey 2023*.
2. Catalogue, A. M. (2025). *Nigeria - National Agricultural Sample Survey 2023 National Agricultural Sample Survey 2023*. 1–9.
3. Chen, K. (2026). *Auditing GenAI – Student Grade Claims on Public Datasets : Nested Controls , Frozen Thresholds , and Claim Labels*. 306, 1–17.
4. Christodoulou, P., Kritsi, E., Ladika, G., Tsafou, P., Tsiantas, K., Tsiaka, T., Zoumpoulakis, P., Cavouras, D., & Sinanoglou, V. J. (2025). *Estimation of Tomato Quality During Storage by Means of Image Analysis , Instrumental Analytical Methods , and Statistical Approaches*. December 2024, 1–21.
5. Congress, I. (n.d.). *Global food losses and food waste Global food losses and food waste*.
6. Cortes, J. R., Benitez, I. B., Jalyn, B., Baldoza, S., Angel, C., Pardillo, R., Mae, K., Auxtero, A., Badec, K. P., Anne, D., & Varela, B. (2026). *Climate-smart aquaculture : Innovations and challenges in mitigating climate change impacts on fisheries and coastal agriculture*. *Aquaculture and Fisheries*, 11(2), 221–231. <https://doi.org/10.1016/j.aaf.2025.08.009>

7. Ghadimpour, M., & Ebrahimi, S. babak. (2022). Forecasting Financial Time Series Using Deep Learning Networks: Evidence from Long-Short Term Memory and Gated Recurrent Unit. *Iranian Journal of Finance*, 6(4), 81–94. <https://doi.org/10.30699/ijf.2022.313164.1286>
8. Janeiro, R. De. (2025). The Rise Of Darija in Moroccan Digital Advertising : Language , Identity , and the Power Of Influence A Ascensão Da Darija Na Publicidade Digital Marroquina : Linguagem , Identidade E O Poder Da Influência. 195–217.
9. Messinis, S., Temenos, N., Protonotarios, N. E., Rallis, I., Kalogeras, D., & Doulamis, N. (2024). Enhancing Internet of Medical Things security with artificial intelligence: A comprehensive review. *Computers in Biology and Medicine*, 170(January), 108036. <https://doi.org/10.1016/j.compbiomed.2024.108036>
10. Passalacqua, L. (2021). *Calibrating the CreditRisk + Model at Different Time Scales and in Presence of Temporal Autocorrelation* †.
11. Petrović, P. Z., Savić, S. V, Ilić, M. M., & Member, S. (2021). *Electromagnetic Modelling of Micro-Doppler Signatures of Commercial Airborne Drones*.
12. Senge, J. M., Kaltenecker, F., & Krupitzer, C. (2025). *Integrating Sensor Data , Laboratory Analysis , and Computer Vision in Machine Learning-Driven E-Nose Systems for Predicting Tomato Shelf Life*.
13. Shehzad, K. (2025). *Technology Predictive AI Models for Food Spoilage and Shelf-Life Estimation*. 1(1), 75–94.
14. Zhang, Z. (2024). *An Application of Logistic Regression in Bank Lending Prediction : A Machine Learning Perspective*. 45, 342–347.