Journal of Intelligent Financial Systems and Autonomous Management
(JIFSAM)

Volume 1 Issue 1
Article Number: JIFSAM20260005

Intelligent Cash-Flow Stress Detection for Small and Medium-Sized Enterprises Using Financial Ratios and Payment-Delay Signals

Sonal Devesh[1] and Priyameet Kaur Keer* [2]

[1]Department of Management Studies, CHRIST (Deemed to be University), Bengaluru, Karnataka 560029, India
[2]Department of Management Studies, New Horizon College of Engineering, Bengaluru, Karnataka 560103, India

Abstract

Small and medium-sized enterprises (SMEs) can remain profitable while becoming unable to meet near-term obligations because receivables arrive late, short-term liabilities accumulate, or operating cash flow weakens. This study develops a lightweight early-warning approach that combines conventional financial ratios with operational payment-delay signals. A reproducible Python process generated 3,000 clearly synthetic SME-period observations and probabilistically sampled a cash-flow-stress label from an explicit logistic data-generating mechanism. Logistic regression and a constrained decision tree were evaluated under stratified 70/15/15 training, validation, and untouched-test splits. Four required configurations compared ratio-only features with combined financial, payment, and control variables. On the test set, combined logistic regression achieved 0.780 accuracy, 0.790 precision, 0.663 recall, 0.868 specificity, 0.721 F1, 0.856 ROC-AUC, and 0.547 Matthews correlation coefficient. Its F1 and ROC-AUC exceeded the ratio-only logistic model, although bootstrap intervals overlapped. The combined decision tree did not improve over its ratio-only counterpart, demonstrating that additional variables do not automatically generalize. Operating cash-flow ratio, leverage, overdue-invoice percentage, customer and supplier delays, and interest coverage were prominent interpretable signals. The findings support human-reviewed screening rather than automated lending or rejection. Because the data and target are synthetic, real SME validation, temporal testing, calibration, fairness analysis, and cost-sensitive threshold governance remain necessary.

Keywords: Cash-Flow Stress; Small and Medium-Sized Enterprises; Financial Ratios; Payment Delays; Logistic Regression; Decision Tree; Interpretable Machine Learning

1. Introduction

Cash-flow stress is an operational liquidity condition in which an SME faces elevated risk of missing short-term obligations because liquid resources and incoming cash are insufficient relative to near-term commitments. It is narrower than formal bankruptcy, insolvency, credit default, long-run business failure, or a general decline in profitability. Prior SME research has similarly treated financial distress as difficulty covering liabilities rather than solely as bankruptcy [1], while empirical evidence shows that failure and distress can have different predictors and classification behavior [2].

*Corresponding Author: Priyameet Kaur Keer (priyameetkeer@gmail.com)
Received: 19 June 2026; Revised: 2 August 2026; Accepted: 29 August 2026; Published: 15 September 2026
© 2026 The Author(s).
This is an open access article licensed under a Creative Commons Attribution-NonCommercial 4.0 International Licence (CC BY-NC 4.0).

A profitable firm may therefore experience stress when receivables are delayed, short-term debt rises, or cash conversion deteriorates even though its income statement remains positive.

The distinction matters operationally. Bankruptcy and default are observable terminal or contractual events, whereas cash-flow stress is intended here as an earlier, reversible warning state. A screening model can therefore be useful before a missed payment occurs, provided that it does not represent its output as proof of insolvency. The prediction target is a binary analytical label for one SME-period, not a lifetime statement about the firm. This framing supports intervention such as receivables follow-up, cash-flow planning, revised payment schedules, or qualified review rather than automatic adverse action.

Financial statements provide leverage, liquidity, coverage, profitability, and turnover indicators, but annual or periodic ratios may lag day-to-day deterioration. In SME failure models, days in accounts receivable and trade-payable duration have shown discriminatory value alongside solvency and profitability variables [3]. Payment duration is nevertheless context-dependent rather than universally harmful [4]. Recent cross-country evidence links late payments to credit rationing and worse loan terms [5], while trade credit also conveys private and relational information between firms and financiers [6]. These findings motivate combining ratios with operational payment behavior while avoiding deterministic assumptions.

Lightweight models are appropriate where data, computing, and specialist capacity are limited. Logistic regression has produced competitive SME distress models [7], and decision trees have been used as interpretable classifiers in SME settings [8]. Both can be trained on ordinary hardware, audited without specialized accelerators, and summarized through coefficients or explicit splits. The aim is not to establish a universal best classifier, but to test whether a compact, transparent baseline can identify a probabilistically generated early-warning condition.

The study addresses four research questions: RQ1 asks whether lightweight models can identify SME cash-flow stress from financial and payment features; RQ2 asks whether payment-delay indicators add predictive value beyond ratios; RQ3 compares logistic regression and a constrained decision tree in performance, interpretability, and computational simplicity; and RQ4 identifies the strongest model contributions. The paper contributes a compact combined feature set, an explicit reproducible synthetic generator, a four-configuration ablation, interpretable outputs, and a governance-oriented workflow for human-reviewed early warning.

2. Related Work

Financial-distress and bankruptcy studies commonly transform accounting statements into liquidity, solvency, profitability, cash-flow, and activity ratios. These variables are attractive because their definitions are familiar to analysts and can be computed consistently when statements are available. Their limitations are equally important: ratios can be stale, correlated, influenced by accounting policy, and insensitive to the timing of individual receipts and payments. The present study therefore treats ratios as a necessary financial-health layer rather than a complete representation of short-term liquidity pressure.

Prediction methods range from classical statistical models to machine-learning classifiers. Decision-tree and survival models can perform differently across prediction horizons and error-cost assumptions [9]. Recent company-level evidence also shows that trees can serve as practical early-warning benchmarks [10], while recursive tree rules remain attractive when decision paths must be communicated to non-specialists [11]. Logistic regression offers a complementary linear probability model whose standardized coefficients can be inspected for direction and relative magnitude. Neither approach is assumed to dominate in advance.

Evidence from bankruptcy research cannot be transferred mechanically to cash-flow stress. Reviews caution that reported performance varies with country, target definition, sampling design, class balance, prediction horizon, and validation protocol [12]. A bankruptcy label may be legally defined, while a stress label may be operational and earlier in the deterioration process. Consequently, the paper does not compare its synthetic metrics with published accuracy values as if the tasks were equivalent. Prior work is used to motivate variables and model choices, not to claim external validity for the executed experiment.

Payment behavior adds a process-level view of the cash-conversion cycle. Average customer delay, overdue-invoice share, extreme delay, delay growth, supplier delay, receivables older than 60 days, and volatility can change before annual ratios fully reflect the deterioration. Yet these variables must not directly encode the outcome. A field such as 'account in default' or 'collection action taken' would create target leakage if used to predict stress. The selected signals describe timing and dispersion only, and the label is sampled probabilistically from several standardized inputs plus noise.

Interpretability is treated as a governance requirement rather than an afterthought. Surveys emphasize that consequential machine-learning use requires transparent, auditable reasoning [13], and responsible-AI taxonomies connect explainability with accountability and stakeholder needs [14]. Explanations can help users assess whether a particular prediction should

be trusted for action [15]. Accordingly, the study reports standardized logistic coefficients, odds ratios, a deliberately shallow tree, node-complexity controls, and impurity-based importance. These summaries remain descriptive rather than causal and must be interpreted alongside data quality and domain knowledge.

Table 1: Compact Feature Definitions

Feature	**Sym.**	**Definition**	**Expected link**
Current ratio	CR	Current assets/current liabilities	Lower → higher stress
Quick ratio	QR	Liquid current assets/current liabilities	Lower → higher stress
Cash ratio	CAR	Cash/current liabilities	Lower → higher stress
Operating cash-flow ratio	OCFR	Operating cash flow/current liabilities	Lower → higher stress
Working capital/assets	WCA	Working capital/total assets	Lower → higher stress
Debt/assets	DTA	Total debt/total assets	Higher → higher stress
Short-term debt ratio	STD	Short-term debt/total debt	Higher → higher stress
Interest coverage	ICR	Operating earnings/interest expense	Lower → higher stress
Net-profit margin	NPM	Net profit/revenue	Lower → higher stress
Revenue growth	RG	Period revenue growth rate	Lower → higher stress
Receivables turnover	RTR	Revenue/average receivables	Lower → higher stress
Payables turnover	PTR	Purchases/average payables	Context dependent
Customer delay	CPD	Mean days beyond invoice due date	Higher → higher stress
Invoices overdue	OIP	Fraction of invoices overdue	Higher → higher stress
Maximum delay	MAXD	Maximum invoice delay in days	Higher → higher stress
Delay increase	DINC	Change in average delay	Higher → higher stress
Supplier delay	SPD	Mean days supplier payments are late	Higher → higher stress
Overdue count	OIN	Number of overdue invoices	Higher → higher stress
Receivables $>$ 60 days	R60	Fraction older than 60 days	Higher → higher stress
Delay volatility	PDV	SD of invoice-delay days	Higher → higher stress

3. Methodology

The experiment uses a synthetic cross-sectional dataset of 3,000 SME-period observations generated with NumPy and pandas. Synthetic data were selected because no openly reusable dataset containing the complete requested combination of SME ratios, invoice delays, supplier delays, and a defensible cash-flow-stress target was adopted for this study. Every record is artificial and does not describe an actual company. The defined accounting period can be interpreted as one reporting interval for which statement ratios and aggregated payment behavior are simultaneously available.

The generator covers 12 financial variables: current, quick, cash, operating-cash-flow, working-capital-to-assets, debt-to-assets, short-term-debt, interest-coverage, net-profit-margin, revenue-growth, receivables-turnover, and payables-turnover ratios. Eight operational variables describe average customer delay, overdue-invoice percentage, maximum delay, increase in average delay, supplier delay, overdue-invoice count, receivables more than 60 days overdue, and delay volatility. Firm age, log-normal size and revenue, five sector categories, and seasonal demand are controls. Table 1 defines the full predictor set and expected directional relationships.

$$P(y_i = 1|x_i) = 1/[1 + \exp(-z_i)],\ \log[p_i/(1 - p_i)] = \beta_0 + \Sigma_j \beta_j x_{ij}. \tag{1}$$

Distribution families reflect each variable's support rather than attempting to reproduce a particular country. Revenue and firm size use log-normal draws; bounded proportions use beta draws; nonnegative payment delays use gamma-based draws; and ratios use normal, truncated-normal, beta, or log-normal draws as appropriate. Explicit clipping prevents negative liquidity ratios, payment delays below zero, proportions above one, and extreme coverage values outside documented bounds. Programmatic assertions stop execution if these range checks fail.

The binary target is $y_i = 1$ for cash-flow stress and $y_i = 0$ otherwise. A probabilistic label is sampled from $p_i = (1 + \exp(-z_i))^{-1}$. The latent score combines standardized debt-to-assets (+0.55), current ratio (-0.42), operating-cash-flow ratio (-0.62), average customer delay (+0.52), overdue-invoice percentage (+0.56), supplier delay (+0.39), delay volatility (+0.31), log interest coverage (-0.34), short-term debt (+0.24), working capital to assets (-0.22), an intercept of -0.52, and Gaussian noise with standard deviation 0.75. These coefficients encode plausible directions but are not fitted from real firms. Bernoulli sampling and latent noise ensure that the outcome is not a deterministic threshold and cannot be recovered perfectly from one predictor.

The data are split into 2,099 training, 451 validation, and 450 test observations with stratification and random seed 42. The test partition remains isolated until model choice is complete. Numeric median imputation and standardization, together with categorical most-frequent imputation and one-hot encoding, are fitted only inside scikit-learn pipelines [16]. This ordering prevents test-set statistics from influencing preprocessing. Full-row hashes are compared across partitions; the executed checks found no duplicate overlap between training, validation, and test sets.

Logistic regression uses L2 regularization and sufficient iterations for convergence. Candidate C values are 0.05, 0.1, 0.5, 1, 2, and 10. The decision-tree search is limited to Gini or entropy criteria, shallow maximum depths, and minimum split and leaf sizes that discourage highly specific branches. Validation F1 selects the configuration, with validation ROC-AUC as the tie-breaker. No class weight is applied because the generated 43/57 class distribution is not severely imbalanced. The untouched test set is evaluated once after selection.

Four experiments implement the ablation: logistic regression with ratios only; logistic regression with ratios, payment variables, and controls; a decision tree with ratios only; and a decision tree with the combined set. A majority-class dummy model is retained as a sanity check. The combined-versus-ratio comparison is interpreted separately for each algorithm because a feature set can interact differently with a linear model and a recursive partitioning model.

Evaluation uses accuracy, precision, recall, specificity, F1, ROC-AUC, MCC, and confusion matrices. The classification metrics are computed as

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP+FP}, \quad \text{Recall} = \frac{TP}{TP+FN}, \\ \text{Specificity} &= \frac{TN}{TN+FP}, \qquad F_1 = \frac{2PR}{P+R}. \end{aligned} \tag{1}$$

MCC is included because it summarizes all four confusion-matrix cells [17], while precision and recall complement ROC-AUC when class prevalence is unequal [18]. Zero denominators are handled safely. One thousand stratified bootstrap resamples of each fixed test prediction vector are used to obtain percentile intervals for F1 and ROC-AUC. These intervals express sampling uncertainty for the synthetic test set and should not be interpreted as population-level confidence guarantees.

Reproducibility controls include a fixed seed, recorded package versions, exported data-generating coefficients, saved selected hyperparameters, generated CSV outputs, manuscript-ready PNG/PDF figures, and an execution log. Leakage checks confirm that preprocessing is fitted after splitting, no target field enters the predictors, tuning is restricted to training and validation data, and payment variables represent behavior rather than a terminal label. The complete script regenerates every numerical table and figure.

Figure. 1 presents the end-to-end early-warning workflow before model results are interpreted.

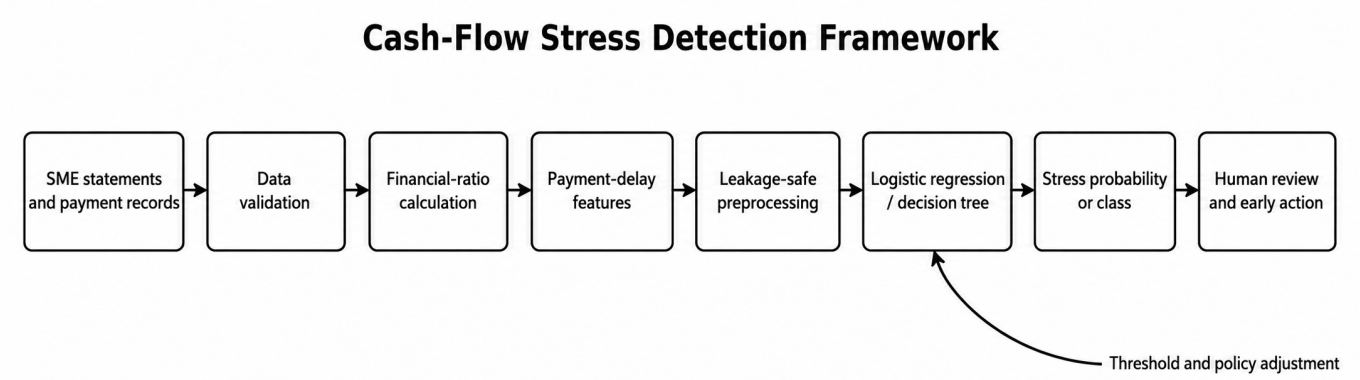


Figure 1: Governed cash-flow stress detection workflow.

Table 2: Synthetic Dataset Summary ($N = 3{,}000$)

Variable	Mean	SD	Min	Max
Current ratio	1.730	0.790	0.427	5.948
Operating CF ratio	0.161	0.217	−0.550	0.850
Debt/assets	0.486	0.208	0.050	0.950
Interest coverage	4.023	3.133	0.298	25.000
Customer delay (days)	23.100	14.443	0.000	116.426
Invoices overdue	0.280	0.174	0.000	0.898
Supplier delay (days)	15.498	10.658	0.000	78.671
Delay volatility (days)	9.460	6.933	0.000	54.298

4. Results

The generator produced 1,290 stressed observations (43.0%) and 1,710 non-stressed observations (57.0%). The stratified partitions preserved similar prevalence while retaining 2,099 training, 451 validation, and 450 test records. Table. 2 reports the executed summary for representative variables. For example, the displayed minima and maxima satisfy the range assertions, and no split-pair duplicate hashes were detected.

Validation selected C=0.1 for both logistic models. The ratio-only tree used Gini criterion, maximum depth 5, minimum split 70, and minimum leaf 20. The combined tree used entropy, maximum depth 6, minimum split 20, and minimum leaf 10. The latter tree achieved a higher validation F1 than the ratio-only tree, but its less restrictive structure also created more opportunities for unstable payment-feature splits.

Table 3: Untouched Test-Set Performance

Model	Acc.	Prec.	Rec.	Spec.	F1	AUC	MCC
LR-R	0.762	0.747	0.674	0.829	0.708	0.843	0.511
LR-C	0.780	0.790	0.663	0.868	0.721	0.856	0.547
DT-R	0.756	0.715	0.715	0.786	0.715	0.803	0.501
DT-C	0.720	0.674	0.674	0.755	0.674	0.787	0.428

Table 3 reports final untouched-test performance. Combined logistic regression produced the highest accuracy (0.780), precision (0.790), specificity (0.868), F1 (0.721), ROC-AUC (0.856), and MCC (0.547), while recall was 0.663. The ratio-only logistic model achieved 0.762 accuracy, 0.747 precision, 0.674 recall, 0.829 specificity, 0.708 F1, 0.843 ROC-AUC, and 0.511 MCC. Thus, adding payment and control variables increased most summary measures but reduced recall by 0.010.

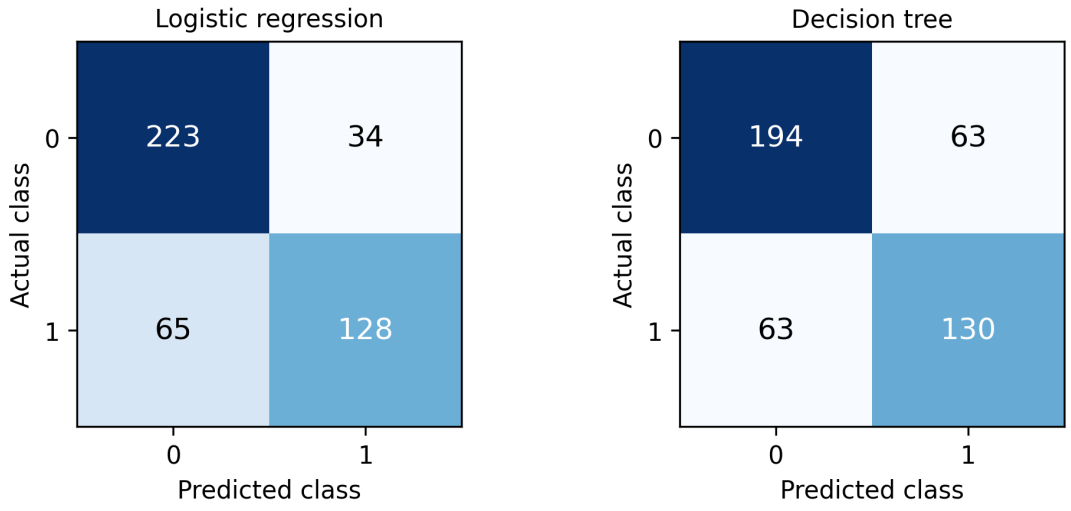


Figure 2: Test confusion matrices for combined-feature models.

The combined logistic F1 bootstrap interval was 0.669-0.771 and its ROC-AUC interval was 0.819-0.890. The corresponding ratio-only intervals were 0.651-0.760 and 0.804-0.878. Because these intervals overlap substantially, the observed improvement should be described as modest and uncertain. The experiment establishes that the additional signals can contribute within this generated sample, not that they would necessarily improve a model trained on real SMEs.

Table 4: Leading Interpretability Results

Variable	**Logit beta**	**OR**	**Tree imp.**
Operating cash flow ratio	−0.669	0.512	0.279
Debt to assets	+0.463	1.589	0.132
Pct invoices overdue	+0.430	1.538	0.114
Supplier payment delay days	+0.386	1.470	0.074
Interest coverage ratio	−0.268	0.765	0.080
Avg customer payment delay days	+0.379	1.461	0.067
Current ratio	−0.428	0.652	0.023
Short term debt ratio	+0.262	1.299	0.069

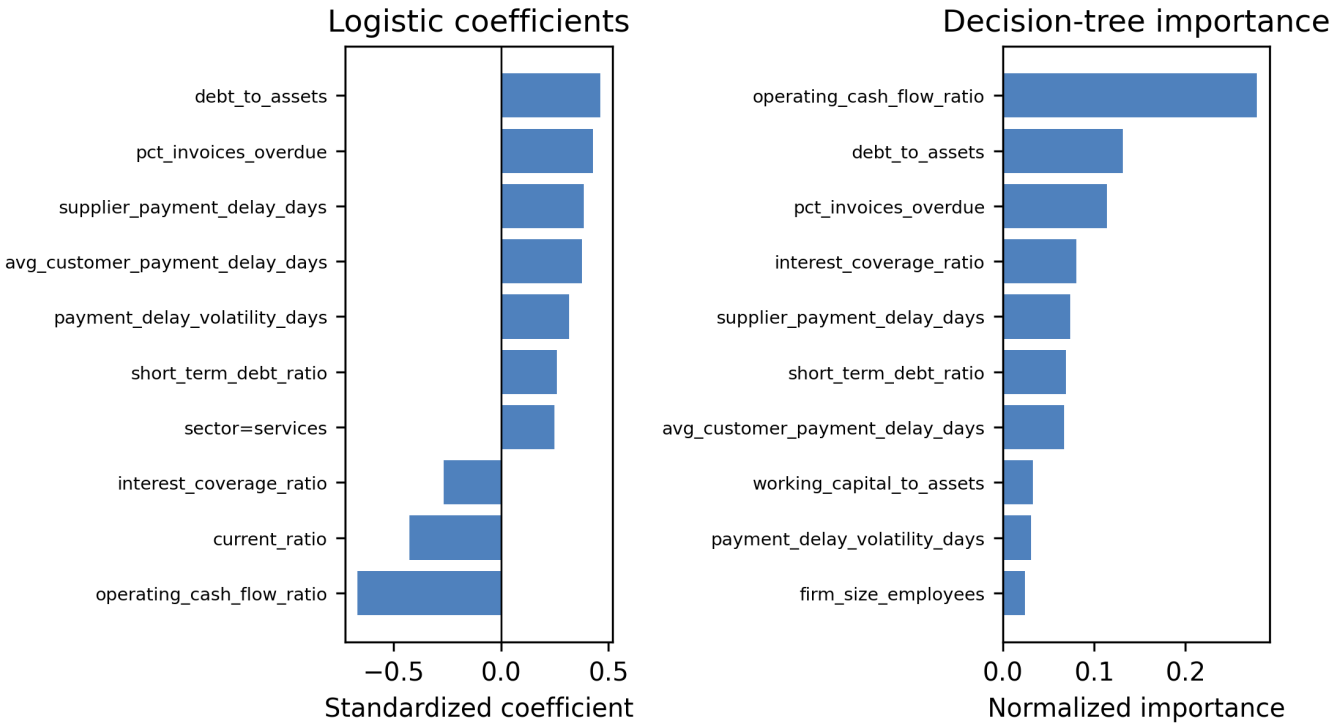


Figure 3: Combined-model coefficients and feature importance.

The ratio-only decision tree achieved 0.756 accuracy, 0.715 precision, 0.715 recall, 0.786 specificity, 0.715 F1, 0.803 ROC-AUC, and 0.501 MCC. The combined tree fell to 0.720 accuracy, 0.674 precision, 0.674 recall, 0.755 specificity, 0.674 F1, 0.787 ROC-AUC, and 0.428 MCC. This is an instructive generalization failure: extra payment variables improved validation F1 but did not yield stable test improvement. It also prevents the ablation conclusion from being overstated across algorithms.

Figure. 2 shows the combined-model confusion matrices. Logistic regression produced TN=223, FP=34, FN=65, and TP=128. The combined tree produced TN=194, FP=63, FN=63, and TP=130. The tree recovered two additional stressed cases but generated 29 additional false alarms. Which trade-off is acceptable cannot be decided from accuracy alone; it depends on the operational cost of a missed stress case versus unnecessary review.

Table. 4 and Figure. 3 summarize model interpretation. The largest absolute standardized logistic coefficient was operating-cash-flow ratio (-0.669), followed by debt-to-assets (+0.463), overdue-invoice percentage (+0.430), current ratio (-0.428), supplier delay (+0.386), and customer delay (+0.379). Their exponentiated coefficients provide odds-ratio interpretations per one standardized-unit change while holding other encoded features fixed. The tree likewise

prioritized operating cash flow (importance 0.279), leverage (0.132), overdue percentage (0.114), interest coverage (0.080), supplier delay (0.074), short-term debt (0.069), and customer delay (0.067). Directional consistency with the generator is reassuring, but correlated ratios can redistribute coefficient magnitude and impurity importance is not causal attribution.

5. Discussion

RQ1 is supported within the bounds of the synthetic experiment: both model classes exceeded the majority baseline, and all four configurations produced useful discrimination above chance. This result demonstrates technical feasibility for a compact feature set, not empirical effectiveness in a live SME portfolio. The observed metrics depend on the selected distributions, latent coefficients, noise level, prevalence, split, and hyperparameter grid.

For RQ2, payment-delay signals added modest test value to logistic regression but not to the decision tree. The linear model can aggregate several individually noisy operational signals into a smoother score. A tree, by contrast, may create local thresholds that fit validation-specific combinations. The combined tree's weaker test result shows why feature enrichment should be validated rather than assumed beneficial. It also illustrates the value of retaining a ratio-only baseline instead of reporting only the most feature-rich configuration.

RQ3 favors combined logistic regression as the stronger overall baseline in this sample. It provided the best AUC, F1, specificity, and MCC, used a single regularization parameter, and exposed signed coefficients. The ratio-only tree, however, achieved slightly higher recall and offers direct conditional rules. Interpretability therefore has multiple dimensions: logistic regression gives global additive direction and odds ratios, while a shallow tree gives threshold-based paths. Neither representation explains causality or guarantees that a local prediction is correct.

RQ4 identifies weak operating cash flow, high leverage, a large overdue share, reduced liquidity and coverage, and longer or more volatile payment delays as the principal synthetic warning signals. These findings mirror the encoded data-generating logic and confirm that the code, fitted model, and interpretation outputs are internally coherent. They should not be presented as newly discovered real-world effect sizes. Validation with observed SME histories is required to determine whether the same rankings persist across industries, countries, and economic cycles.

False negatives are especially important because they leave stressed SMEs unflagged; combined logistic regression missed 65 of 193 stressed test cases. False positives also matter because unnecessary escalation can consume analyst time, stigmatize firms, or indirectly affect credit terms. Threshold choice should therefore be governed through explicit cost scenarios rather than fixed automatically at 0.5. A lender, SME adviser, or internal finance team may select different operating points, but all should document the reason, approval authority, monitoring interval, and process for reverting a harmful threshold.

A practical workflow would ingest validated statements and payment records, compute versioned features, apply the locked preprocessing pipeline, return a probability and concise reasons, and route the result to qualified human review. The model must not automatically deny loans, terminate supplier relationships, restrict services, or trigger punitive collection. Reviewers need access to source-data corrections, confidence and limitation notices, and an appeal mechanism. Data confidentiality, encryption, role-based access, audit logs, cybersecurity testing, retention controls, periodic recalibration, drift monitoring, and sector-level fairness checks are deployment prerequisites. Figure 1 illustrates this feedback-controlled governance loop.

6. Limitations and Future Work

The dataset is synthetic, cross-sectional, and generated from simplified distributions and a known latent mechanism. It cannot reproduce country-specific accounting practices, sector cycles, macroeconomic shocks, strategic payment behavior, missingness, fraud, bookkeeping error, invoice disputes, or evolving reporting quality. The label is an analytical proxy rather than an observed inability to pay, and the selected coefficients determine part of the recoverable signal. Correlation among constructed variables can diffuse coefficient effects, while categorical controls and a single random split may reflect chance variation. No temporal, external, subgroup-fairness, probability-calibration, or cost-sensitive validation was performed. The bootstrap intervals condition on the generated test sample and do not address uncertainty in the synthetic mechanism. Impurity-based tree importance can favor variables with more candidate splits, and standardized logistic coefficients remain sensitive to collinearity and coding choices. The experiment also omits macroeconomic variables, financing access, payment terms, and sector-specific seasonality beyond one simplified indicator. Future work should use legally accessible longitudinal SME records, preserve invoice-level timing, define operational stress labels with accountants and domain experts, and separate prediction horizons. Studies should test across sectors, firm ages, sizes, countries, and economic regimes; quantify calibration, drift, and subgroup error rates; assess stability under missing and delayed data; and choose thresholds from explicit false-negative and false-positive

costs. Prospective shadow-mode evaluation should precede any decision support, followed by controlled human oversight and documented model-change governance.

7. Conclusion

This study executed a reproducible synthetic experiment for detecting SME cash-flow stress with financial ratios and payment-delay signals. Combined logistic regression provided the strongest overall untouched-test performance, but its improvement over ratios alone was modest and its bootstrap intervals overlapped. The combined decision tree did not generalize better than the ratio-only tree, showing that additional operational variables are not automatically beneficial. The coefficient and importance outputs provided transparent, directionally coherent explanations of the fitted models. These results support the feasibility of a low-cost, human-reviewed early-warning prototype rather than production readiness, universal generalizability, bankruptcy prediction, or automated credit approval. Real longitudinal validation, calibration, fairness and drift analysis, cost-sensitive thresholds, secure data governance, and qualified human judgment are essential before practical use.

Author Contributions

Sonal Devesh (S.D): Conceptualization, Investigation, Writing – original draft; Priyameet Kaur Keer (P.K.K): Conceptualization, Methodology, Software, Formal analysis, Validation, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The data that support the findings of this study are available from the corresponding author, P.K.K, upon reasonable request.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this work the authors did not use any generative AI or AI-assisted technologies.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] A. Malakauskas and A. Lakštutienė, "Financial Distress Prediction for Small and Medium Enterprises Using Machine Learning Techniques," *Engineering Economics*, vol. 32, no. 1, pp. 4–14, 2021, doi: 10.5755/j01.ee.32.1.27382.

[2] Y. Tong and Z. Serrasqueiro, "Predictions of Failure and Financial Distress: A Study on Portuguese High and Medium-High Technology Small and Mid-Sized Enterprises," *Journal of International Studies*, vol. 14, no. 2, pp. 9–25, 2021, doi: 10.14254/2071-8330.2021/14-2/1.

[3] Y. Zizi, M. Oudgou, and A. El Moudden, "Determinants and Predictors of SMEs' Financial Failure: A Logistic Regression Approach," *Risks*, vol. 8, no. 4, art. 107, 2020, doi: 10.3390/risks8040107.

[4] V. Lefebvre, "Trade credit, payment duration, and SMEs' growth in the European Union," *International Entrepreneurship and Management Journal*, vol. 19, no. 3, pp. 1313–1340, 2023, doi: 10.1007/s11365-023-00871-4.

[5] O. Kaya, "The impact of late payments on SMEs' access to finance: Evidence from credit rationing and loan terms," *Economic Modelling*, vol. 141, art. 106896, 2024, doi: 10.1016/j.econmod.2024.106896.

[6] P. Murro and V. Peruzzi, "Relationship lending and the use of trade credit: the role of relational capital and private information," *Small Business Economics*, vol. 59, no. 1, pp. 327–360, 2022, doi: 10.1007/s11187-021-00537-x.

[7] Y. Zizi, A. Jamali-Alaoui, B. El Goumi, M. Oudgou, and A. El Moudden, "An Optimal Model of Financial Distress Prediction: A Comparative Study between Neural Networks and Logistic Regression," *Risks*, vol. 9, no. 11, art. 200, 2021, doi: 10.3390/risks9110200.

[8] Y. Aker and A. Karavardar, "Using Machine Learning Methods in Financial Distress Prediction: Sample of Small and Medium Sized Enterprises Operating in Turkey," *Ege Academic Review*, vol. 23, no. 2, pp. 145–162, 2023, doi: 10.21121/eab.1027084.

[9] A. Gepp and K. Kumar, "Predicting Financial Distress: A Comparison of Survival Analysis and Decision Tree Techniques," *Procedia Computer Science*, vol. 54, pp. 396–404, 2015, doi: 10.1016/j.procs.2015.06.046.

[10] M. Ďurica, J. Frnda, and L. Svabova, "Artificial neural network and decision tree-based modelling of non-prosperity of companies," *Equilibrium. Quarterly Journal of Economics and Economic Policy*, vol. 18, no. 4, pp. 1105–1131, 2023, doi: 10.24136/eq.2023.035.

[11] I. Podhorská, J. Vrbka, G. Lăzăroiu, and M. Kováčová, "Innovations in Financial Management: Recursive Prediction Model Based on Decision Trees," *Marketing and Management of Innovations*, no. 3, pp. 276–292, 2020, doi: 10.21272/mmi.2020.3-20.

[12] B. Prusak, "Review of Research into Enterprise Bankruptcy Prediction in Selected Central and Eastern European Countries," *International Journal of Financial Studies*, vol. 6, no. 3, art. 60, 2018, doi: 10.3390/ijfs6030060.

[13] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine Learning Interpretability: A Survey on Methods and Metrics," *Electronics*, vol. 8, no. 8, art. 832, 2019, doi: 10.3390/electronics8080832.

[14] A. Barredo Arrieta et al., "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion*, vol. 58, pp. 82–115, 2020, doi: 10.1016/j.inffus.2019.12.012.

[15] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You?: Explaining the Predictions of Any Classifier," *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, pp. 1135–1144, 2016, doi: 10.1145/2939672.2939778.

[16] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[17] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, art. 6, 2020, doi: 10.1186/s12864-019-6413-7.

[18] T. Saito and M. Rehmsmeier, "The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets," *PLOS ONE*, vol. 10, no. 3, art. e0118432, 2015, doi: 10.1371/journal.pone.0118432.