FINANCE AND RISK MANAGEMENT

Data-Driven Credit Risk and SME Portfolio Optimization: A Practical Framework for Regional and Community Financial Institutions—Evidence from U.S. SBA 7(a) Lending Data and Interpretable Portfolio Analytics

Tadiwanashe Lennon Kasuwa[1], Isheanesu Chiworeso[1], James Sydney[1], Mabasa Masunungure[2], Munashe Naphtali Mupa[3]

[1]George Washington University, Washington, DC, USA
[2]Suffolk University, Boston, MA, USA
[3] Hult International Business School

Published November 2026

Abstract

This study develops and tests a practical credit-risk and portfolio-optimization framework for regional and community financial institutions using official U.S. Small Business Administration (SBA) 7(a) loan data. The empirical sample contains 463,775 loans approved between fiscal years 2010 and 2021 and subsequently observed as either paid in full or charged off. Descriptive analytics, portfolio heat maps and an interpretable logistic-regression model are combined to identify risk patterns that are directly relevant for origination, pricing, monitoring and portfolio steering. The observed charge-off rate in the resolved-loan sample is 7.17%. New or startup businesses (8.20%) and unsecured exposures (9.62%) underperform established firms (6.18%) and secured loans (6.21%). Loans with maturities of 60 months or less exhibit the highest observed charge-off rate (17.59%), while 241-360 month credits perform materially better (1.26%). Sectoral differences are economically meaningful: transportation and warehousing, accommodation and food services, retail trade and construction show above-average risk, whereas health care and social assistance is materially lower. The out-of-time model achieves an area under the ROC curve of 0.846 on a 2019-2021 test set, and the top two predicted risk deciles capture 68.23% of observed defaults. The paper translates these findings into a governance-oriented operating model that combines interpretable scorecards, relationship-lending judgement, collateral discipline, sector and vintage monitoring, and early-warning dashboards. The contribution lies in linking SME credit analytics to portfolio governance in a form that is rigorous enough for risk management yet practical enough for regional and community institutions.

Keywords: SME finance; credit risk; community banks; portfolio optimisation; interpretable analytics; SBA 7(a); regional banking

1 Introduction

Small businesses remain central to the structure of the United States economy. According to the US Small Business Administration Office of Advocacy (2026), 99.9% of U.S. businesses are small, they employ 62.3 million people, and they account for 43.5% of gross domestic product. That macroeconomic footprint means that the quality of SME credit allocation is not merely a bank-level issue; it is an issue of local investment, employment resilience and economic continuity.

For regional and community financial institutions, SME credit decisions are especially consequential because they are frequently made close to the borrower, embedded in local market knowledge and tied to relationship-specific information. Berger and Udell (2006) argue that SME finance is best understood through lending technologies rather than a single universal model, while Holod et al. (2024) show that relationship-based small business lending continues to create measurable franchise value for community banks. The implication is not that analytics should replace relationships, but that institutions need disciplined ways to combine soft information with robust, auditable risk measurement.

This need has become more pressing as loan books digitise, regulatory scrutiny of model governance deepens, and portfolio managers are asked to do more than approve or decline credits. They must allocate risk capacity, identify concentration build-ups, calibrate collateral and tenor structures, and detect deterioration early enough to change outcomes. The literature on modern credit scoring demonstrates clear performance gains from more systematic data use (Altman and Sabato, 2007; Baesens et al., 2003; Lessmann et al., 2015), yet community institutions often still face a translation problem: how to convert predictive analytics into a portfolio operating model that remains explainable, proportionate and implementable.

This paper addresses that problem in two ways. First, it provides evidence from a large U.S. SBA 7(a) loan sample, using descriptive portfolio analytics, heat maps and an out-of-time predictive model to identify the most policy-relevant risk gradients. Second, it converts those empirical findings into a practical framework for regional and community institutions. The aim is not to offer a purely academic scoring exercise, but to build a decision architecture for SME lending that improves underwriting discipline without sacrificing relationship-based judgement.

2 Literature review

The classic SME finance problem is shaped by informational opacity. Smaller firms tend to have thinner audited histories, more concentrated cash flows and fewer alternative financing channels, which makes lender information processing central to credit availability (Berger and Udell, 2006). Relationship lending helps reduce this opacity by incorporating soft information gathered over time, while transaction-based lending uses more standardised hard information such as financial statements, collateral values or credit scores. The policy question for modern regional banks is therefore not whether to use data, but which lending technology is most appropriate for which borrower segment.

The empirical credit-scoring literature has steadily advanced from conventional scorecards toward richer machine-learning environments. Baesens et al. (2003) benchmarked classification methods for credit scoring and showed that predictive performance differences can be meaningful across algorithms. Altman and Sabato (2007) extended SME credit-risk modelling using U.S. data and demonstrated that SME-specific scoring frameworks outperform models imported mechanically from larger-firm contexts. Lessmann et al. (2015) later updated the benchmarking literature and confirmed that no single method dominates in all settings, a result that remains important for bank implementation because portfolio context and governance requirements matter as much as raw classification power.

More recent studies have pushed the frontier toward alternative data, non-linear methods and explainability. Bellotti and Crook (2009) used support vector machines to improve credit classification and feature discovery. Malekipirbazari and Aksakalli (2015) showed that random forests can materi-

ally improve risk assessment in digital lending contexts. Jagtiani and Lemieux (2019) demonstrated that alternative data and machine learning can improve lending decisions in fintech settings, although their evidence also implies that institutional context matters when these tools are transplanted into regulated banking environments. Li and Wu (2024) and Jacobs (2024) both highlight a central implementation challenge: the strongest predictive model is not automatically the most appropriate bank model when validation, accountability and supervisory scrutiny are taken seriously.

This tension between predictive accuracy and decision governance is especially important for regional and community institutions. Leo et al. (2019) observe that machine learning has clear potential across banking risk management but remains unevenly embedded in governance practice. Internal-control and decision-intelligence research points in the same direction. Mupa et al. (2024) emphasise the importance of risk management and internal controls in operational decision quality; Matsebula et al. (2025) argue for integrating predictive and managerial layers rather than treating analytics as a stand-alone technical artefact; and Taanisa et al. (2026) show how finance, audit and data-science integration can strengthen SME budget governance and risk-aware growth. These strands suggest a gap in the SME credit literature: the need for frameworks that explicitly connect lending analytics to portfolio governance, monitoring and managerial execution. This paper contributes in that space.

3 Data and methodology

3.1 Data source and sample construction

The empirical analysis uses the official SBA 7(a) FOIA loan files published by the US Small Business Administration (2026). The dataset is updated quarterly and, at the time of access, covered approvals since fiscal year 1991. For analytical relevance and seasoning consistency, the present study focuses on approvals from fiscal year 2010 through fiscal year 2021. Kaggle-hosted analytical mirrors of the same SBA extracts were reviewed during preliminary schema familiarisation, but final estimation used the official SBA files rather than secondary repostings.

To avoid mixing matured and unmatured outcomes, the study retains only loans whose reported status is either paid in full (PIF) or charged off (CHGOFF). The resulting resolved-loan sample contains 463,775 observations. This choice improves outcome clarity but does introduce a limitation: the sample describes observed outcomes among resolved loans rather than the ultimate lifetime default rate of all loans originated in a cohort. The analysis therefore interprets the default indicator as an observed charge-off outcome within the available resolution window.

3.2 Variables and feature engineering

The core variables are gross approval amount, SBA guaranteed amount, initial interest rate, term in months, NAICS sector, business age, processing method, collateral indicator, revolver status, jobs supported and state. The dependent variable equals one if the resolved loan was charged off and zero if it was paid in full. Several derived variables are created to make the portfolio analysis operationally meaningful: guarantee ratio, loan-size bands, term bands, business-age categories and sector groups based on NAICS. These transformations are designed to reflect how credit officers and portfolio committees actually segment books, rather than how a purely machine-centric workflow would encode the data.

3.3 Analytical design

The paper combines three layers of analysis. The first is descriptive portfolio analytics, used to establish the central risk gradients across borrower age, collateral, loan size, term, sector, state and processing method. The second is visual segmentation through heat maps, which helps reveal interaction effects that are less visible in marginal summaries. The third is an interpretable logistic-regression model used as an out-of-time benchmark for operational credit triage.

The model specification follows the standard logistic form:

$$P(\text{Default}_i = 1) = \frac{1}{1 + \exp\left(-\left(\beta_0 + \beta' X_i\right)\right)} \quad (1)$$

Numeric inputs include approval amount, guarantee ratio, interest rate, missing-interest flag, term, jobs supported and approval year. Categorical inputs include sector, business age, size band, term band, guarantee band, collateral status, revolver status and processing method. A 2010-2018 training window is used for estimation, with a tractable random training sample of 120,000 loans, while the full 2019-2021 sample is reserved for out-of-time testing. Logistic regression is deliberately preferred over black-box alternatives because regional and community institutions often need models that are easier to validate, explain and embed in committee governance (Jacobs, 2024; Li and Wu, 2024).

4 Empirical results

4.1 Portfolio profile

The resolved-loan sample has a mean gross approval of USD 362,445 and a median of USD 105,000. The mean guarantee ratio is 0.65, while the mean initial interest rate is 6.50%. The average term is 115.3 months, and the mean number of jobs supported is 10.6. The overall observed charge-off rate is 7.17%. Table 1 summarises these portfolio characteristics.

Table 1. Descriptive statistics for the resolved-loan sample.

Variable	Mean	Median	Std. dev.	25th pct.	75th pct.	N
Gross approval (USD)	362,445.36	105,000.00	672,693.72	35,000.00	350,000.00	463,775
SBA guarantee ratio	0.65	0.50	0.15	0.50	0.75	463,775
Initial interest rate (%)	6.50	6.00	1.58	5.50	7.25	463,775
Term (months)	115.26	84.00	75.90	76.00	120.00	463,775
Jobs supported	10.64	4.00	20.56	2.00	11.00	463,775
Observed charge-off indicator	0.07	0.00	0.26	0.00	0.00	463,775

Note: Gross approval and jobs supported are shown in natural units. The charge-off indicator equals 1 for charged-off loans and 0 for paid-in-full loans. Source: author's calculations using SBA 7(a) FOIA data.

Figure 1 shows the observed charge-off rate by approval year for the resolved-loan sample. The pattern suggests a rising stress profile for mid-to-late 2010s cohorts among loans that had already resolved by the extraction date, peaking around 2018 before falling for later vintages. Because later approval years have had less time to mature, the figure should be read as an observed resolved-loan pattern rather than as a final cohort-loss curve.

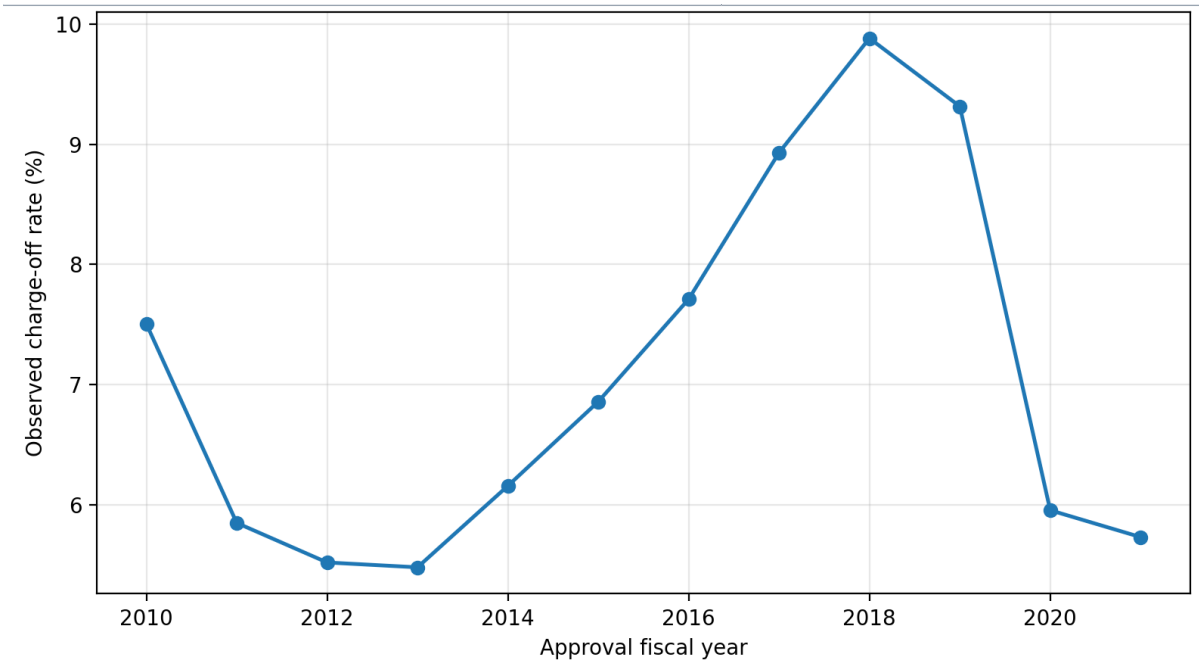


Figure 1. Observed charge-off rate by approval year (2010-2021 approvals). Source: author's calculations using SBA 7(a) FOIA data. The figure uses only loans that had resolved as paid in full or charged off by the extraction date.

4.2 Segment-level risk patterns

Table 2 demonstrates that observed risk is far from evenly distributed across the book. Established operating firms record a charge-off rate of 6.18%, compared with 8.20% for new or startup firms and 7.72% for firms aged two to four years. Collateral is also strongly associated with performance: unsecured loans show a 9.62% charge-off rate versus 6.21% for secured credits.

Tenor segmentation is particularly revealing. Loans with maturities of 60 months or less record a 17.59% observed charge-off rate, substantially above the 5.06% rate for 61-120 month loans and the 1.26% rate for 241-360 month facilities. Loan size also matters: exposures up to USD 150,000 show an 8.23% charge-off rate, while facilities between USD 1 million and USD 5 million show 3.87%. This does not imply that larger loans are safer in an absolute loss sense; indeed, mean charge-off severity rises from about USD 46,276 in the smallest band to more than USD 1,070,628 in the USD 1-5 million band. The implication is that default frequency and loss severity must be managed jointly.

Table 2. Selected segment-level risk patterns.

Dimension	Segment	Loans	Charge-off rate (%)	Average approval (USD)
Business age	Established/COO	227,416	6.18	395,689
Business age	New/Startup	126,769	8.20	412,576
Business age	Unknown	41,416	8.58	226,194
Business age	Young (2-4 years)	68,174	7.72	241,107
Collateral	Unsecured	131,240	9.62	129,959
Collateral	Secured	332,535	6.21	454,200
Term band	≤60	98,853	17.59	143,497
Term band	61-120	277,485	5.06	231,147

Dimension	Segment	Loans	Charge-off rate (%)	Average approval (USD)
Term band	121-180	21,930	2.67	579,498
Term band	181-240	11,936	4.79	744,561
Term band	241-360	53,568	1.26	1,272,582
Loan-size band	≤150k	275,172	8.23	58,457
Loan-size band	150k-350k	78,684	6.97	257,849
Loan-size band	350k-1M	65,663	5.21	609,958
Loan-size band	1M-5M	44,256	3.87	2,071,295
Loan-size band	>5M	0	–	–

Note: The table reports the number of resolved loans, observed charge-off rate and average approved amount for selected segment definitions. Source: author's calculations using SBA 7(a) FOIA data.

The interaction effects are visible in the heat maps. Figure 2 shows that the most vulnerable cells cluster among shorter-term credits, especially in the sub-USD 1 million range. Figure 3 shows that retail trade, accommodation and food services, construction, and transportation and warehousing remain relatively fragile across borrower-age categories, while health care and social assistance is materially more resilient. This kind of visual portfolio decomposition is useful because it highlights where a bank may need differentiated policies rather than blanket tightening across the entire SME book.

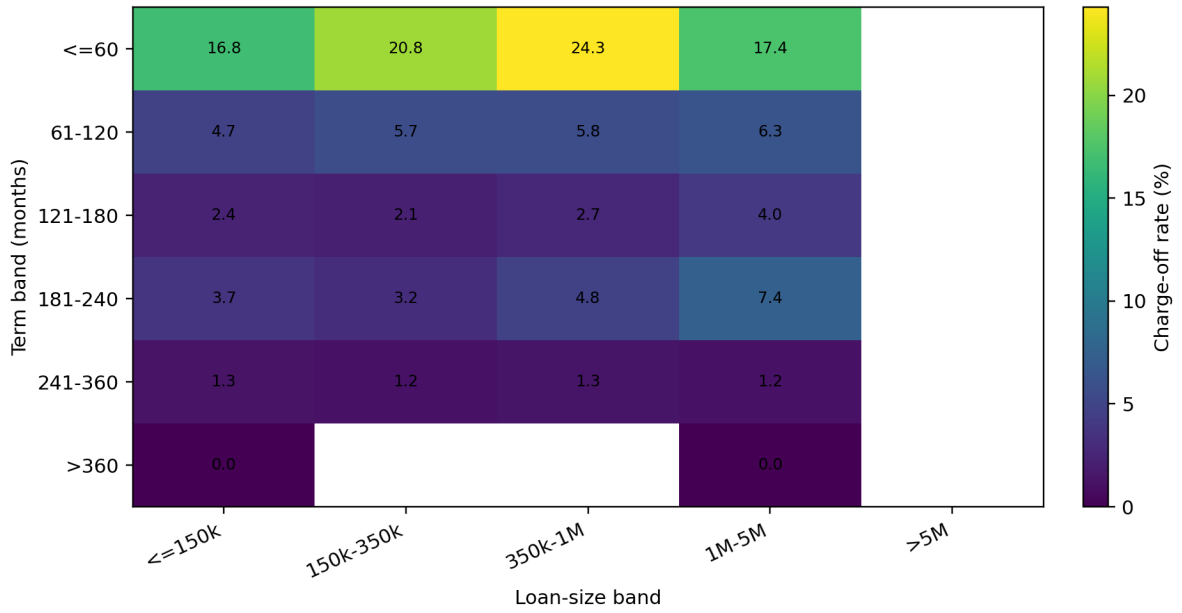


Figure 2. Charge-off heat map by term and loan size. Source: author's calculations using SBA 7(a) FOIA data. Blank cells denote sparse or unavailable combinations in the resolved-loan sample.

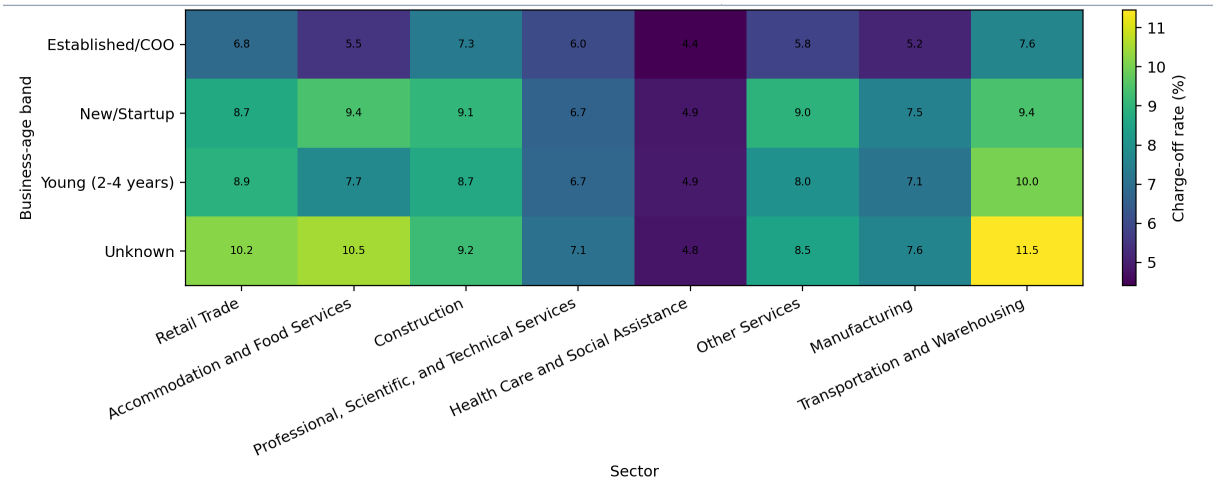


Figure 3. Charge-off heat map by business age and major sectors. Source: author's calculations using SBA 7(a) FOIA data.

4.3 Sectoral, geographic and program heterogeneity

Among the largest sectors by loan count, transportation and warehousing (8.74%), wholesale trade (8.20%), retail trade (7.98%) and construction (7.93%) sit above the portfolio average, whereas health care and social assistance records only 4.67%. Among high-volume states, Florida (11.21%), Texas (9.64%) and New York (8.97%) run above the sample average, while Minnesota (4.97%), Massachusetts (5.78%) and Ohio (6.08%) perform materially better within the high-volume group (Table 3).

Processing-method differences are also instructive. Community Express (17.55%), Community Advantage Initiative (16.56%) and Patriot Express Loans (11.01%) display substantially weaker performance than the core 7a General book (7.08%). This suggests that portfolio committees should not treat program labels as administrative metadata only; they can be material portfolio-risk signals.

Table 3. High-volume sector, state and processing-method heterogeneity.

Panel	Segment	Loans	Charge-off rate (%)	Average approval (USD)
Panel A: sectors	Retail Trade	62,841	7.98	368,968
Panel A: sectors	Accommodation and Food Services	58,724	7.86	514,938
Panel A: sectors	Construction	50,738	7.93	204,786
Panel A: sectors	Professional, Scientific, and Technical Services	46,088	6.39	268,939
Panel A: sectors	Health Care and Social Assistance	41,936	4.67	424,974
Panel A: sectors	Other Services	40,077	7.37	316,297
Panel B: states	CA	56,101	7.16	479,972
Panel B: states	TX	34,817	9.64	466,862
Panel B: states	NY	29,206	8.97	235,470
Panel B: states	OH	26,822	6.08	222,904
Panel B: states	FL	22,265	11.21	467,363

Panel	Segment	Loans	Charge-off rate (%)	Average approval (USD)
Panel B: states	MI	18,469	5.11	275,252
Panel C: processing methods	Community Express	4,940	17.55	23,626
Panel C: processing methods	Community Advantage Initiative	4,415	16.56	131,558
Panel C: processing methods	Patriot Express Loans	5,352	11.01	88,339
Panel C: processing methods	Small Loan Advantage Initiative	25,089	9.36	169,981
Panel C: processing methods	Export Express	1,197	9.36	160,568
Panel C: processing methods	Gulf Opportunity Pilot Loan Program	1,446	7.61	64,192

Note: Sectors and states are ranked by loan count within the resolved-loan sample. Processing methods are shown for approaches with at least 1,000 resolved loans. Source: author's calculations using SBA 7(a) FOIA data.

4.4 Predictive performance and portfolio lift

For operational use, a model need not be opaque to be useful. The logistic-regression benchmark produces an out-of-time AUC of 0.846 and an average-precision score of 0.330 on the 2019-2021 test sample. That level of discrimination is materially informative for triage, pricing and monitoring decisions, especially when combined with relationship-manager overlays. The decile analysis is even more persuasive operationally: the highest-risk decile alone captures 43.92% of defaults, while the top two deciles capture 68.23% and the top three capture 83.65% (Table 4 and Figure 4).

In practice, this means a regional institution does not need to push every borrower through an overly complex artificial-intelligence stack to gain substantial portfolio insight. An interpretable model can already concentrate a large share of future losses into a manageable monitoring population. The managerial question then becomes how to act on that lift through covenants, collateral reviews, sector flags, watchlist rules and relationship-based remediation.

Table 4. Model performance and risk-decile concentration.

Panel / decile	Description	Result
Panel A: performance metrics		
	AUC	0.846
	Average precision	0.330
	Brier score	0.175
	Accuracy @ 0.5	0.775
	Precision @ 0.5	0.220
	Recall @ 0.5	0.809
	F1 @ 0.5	0.345
Panel B: highest predicted risk deciles		

Panel / decile	Description	Result
Decile 1	7,188 loans | 32.18% charge-off	43.92% capture | 43.92% cumulative
Decile 2	7,187 loans | 17.81% charge-off	24.31% capture | 68.23% cumulative
Decile 3	7,187 loans | 11.30% charge-off	15.42% capture | 83.65% cumulative
Decile 4	7,187 loans | 3.65% charge-off	4.98% capture | 88.63% cumulative
Decile 5	7,187 loans | 2.27% charge-off	3.10% capture | 91.72% cumulative

Note: The model is estimated on a 2010-2018 training sample and evaluated on the full 2019-2021 out-of-time test set. Source: author's calculations using SBA 7(a) FOIA data.

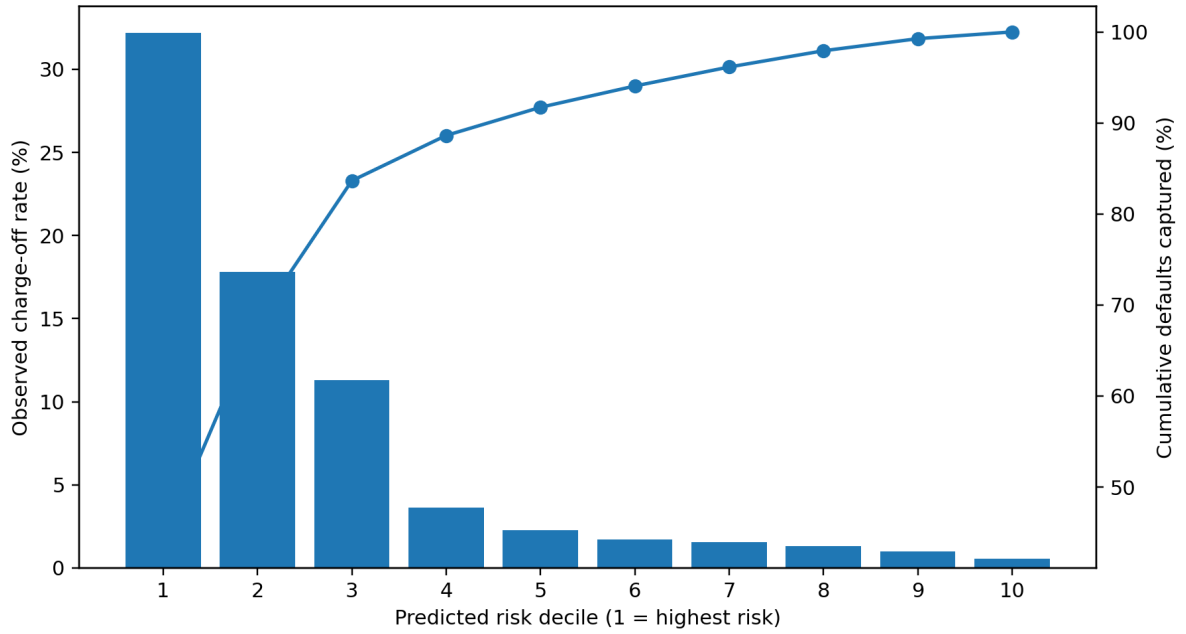


Figure 4. Portfolio lift by predicted risk decile (2019-2021 test set). Source: author's calculations using SBA 7(a) FOIA data. The bars show observed charge-off rates; the line shows cumulative defaults captured.

5 A practical framework for regional and community financial institutions

The evidence points to a simple but important principle: SME portfolio optimisation is not a one-off model-building exercise. It is a management system. Regional and community institutions need a workflow that integrates origination, relationship knowledge, collateral policy, concentration management and early-warning escalation. Table 5 sets out a practical operating model built from the empirical findings.

The first design choice is to preserve the complementarity between data and judgement. Holod et al. (2024) show that relationship-based small business lending continues to create value for community banks, especially in smaller credits where private information matters. The appropriate response is therefore a dual-track process: use interpretable analytics to standardise baseline risk assessment, then allow relationship officers to add or defend context-specific information through disciplined override protocols.

The second design choice is to separate frequency management from severity management. Small, unsecured and short-tenor loans in this sample default more often, whereas larger loans generate

more severe average losses when they fail. A mature portfolio framework therefore needs both a frequency lens and a severity lens, rather than one undifferentiated risk rank.

Table 5. Governance-oriented implementation framework for SME credit portfolios.

Framework layer	Managerial actions	Illustrative KPIs
1. Data foundation	Create a loan-level data mart combining origination, collateral, sector, borrower age, pricing, servicing and loss outcomes. Standardise definitions for defaults, restructures, modifications and recoveries.	Completeness rate; missing-value rate; data lag; reconciliation breaks
2. Dual-track origination	Use an interpretable score alongside relationship-manager judgement. Require explicit justification for overrides, especially for new firms, unsecured credits and high-risk sectors.	Override rate; override win/loss rate; approval-to-book ratio
3. Risk-based structure and pricing	Link tenor, collateral requirements, guarantee coverage, covenants and pricing to empirically observed risk bands rather than flat policy rules.	Risk-adjusted margin; secured mix; weighted average tenor
4. Early-warning monitoring	Deploy monthly dashboards covering payment stress, utilisation shifts, sector weakness, covenant breaches, collateral deterioration and relationship concentration.	Roll-rate; delinquency emergence; watchlist migration
5. Portfolio steering	Manage concentrations by sector, state, vintage, borrower age and program type. Set thresholds for fast-growing pockets before charge-offs rise materially.	Sector concentration; vintage loss rate; exposure-at-default mix
6. Governance and model risk	Validate scorecards regularly, test stability by vintage and geography, and present model performance to credit committees in plain language.	AUC drift; calibration drift; validation exceptions
7. Feedback and learning loop	Use post-mortems on defaults, restructures and recoveries to refresh policy, pricing and relationship-manager playbooks.	Policy update cycle time; lessons-closed rate; recovery improvement

Note: The framework is derived from the empirical results in this paper and from the wider literature on relationship lending, risk governance and interpretable analytics.

6 Discussion

Three conclusions emerge from the evidence. First, the riskiest parts of the book are not random. The concentration of higher charge-off rates among startups, unsecured credits, shorter maturities and certain service-oriented sectors suggests that policy settings can be improved without indiscriminate tightening. A bank that prices all SME credits in roughly the same way, or that monitors only at the aggregate portfolio level, is leaving useful information unexploited.

Second, explainability matters. The recent literature increasingly shows that highly complex models may offer incremental performance gains, but those gains are not always decisive once governance, validation and supervisory usability are considered (Li and Wu, 2024; Jacobs, 2024). For institutions that require committee buy-in, documentation discipline and consistent override governance, an interpretable model with strong decile lift can be more valuable than an opaque model with modestly better in-sample fit.

Third, implementation quality depends on internal controls. Mupa et al. (2024) argue that risk management and internal controls enhance the quality of organisational decisions, and Taanisa et al. (2026) show how decision intelligence can strengthen SME governance when finance, audit and data

science are integrated rather than siloed. Matsebula et al. (2025) similarly emphasise the value of combining predictive and managerial layers in risk platforms. In credit portfolios, that means model outputs should feed approval governance, covenant design, watchlist migration, remediation actions and post-mortem learning loops.

There is also a broader institutional point. Lending frameworks influence real economic resilience, not only bank loss statistics. In adjacent governance scholarship, Aror and Mupa (2025) show that institutional design choices can have significant downstream welfare consequences. The same logic applies to SME finance: poorly governed lending can destroy viable local businesses through weak screening or delayed intervention, while well-governed lending can support continuity, employment and regional economic stability.

The study nonetheless has limits. The SBA 7(a) program is a policy-mediated lending environment and cannot be treated as a complete mirror of all SME credit markets. The resolved-loan filter improves outcome clarity but introduces seasoning bias, especially for later vintages. Moreover, borrower financial-statement variables are not available in the public extract, which means the model cannot test the full set of accounting drivers emphasised in some corporate default studies. These limitations do not eliminate the value of the results, but they do caution against over-generalisation.

7 Conclusion

This paper set out to build a credit-risk and portfolio-optimisation framework for regional and community financial institutions that is both empirically grounded and operationally usable. Using 463,775 resolved SBA 7(a) loans approved between 2010 and 2021, the study shows that SME portfolio risk is systematically associated with borrower age, collateral structure, tenor, loan size, sector, geography and program channel. An interpretable logistic model delivers strong out-of-time discrimination, and the highest-risk 20% of the portfolio captures more than two-thirds of defaults. The practical implication is clear: community and regional institutions can materially improve underwriting discipline and portfolio monitoring without abandoning relationship lending or becoming dependent on black-box infrastructure. The most effective path is a governance-oriented architecture in which analytics, relationship knowledge, internal controls and portfolio steering work together.

Data availability

The empirical analysis uses the official SBA 7(a) FOIA files published by the U.S. Small Business Administration. Public Kaggle-hosted analytical mirrors of the same SBA extracts were consulted during preliminary data familiarisation, but final estimation used the official SBA source files. Derived figures and tables were produced from the cleaned research sample described in Section 3.

References

[1] Altman EI, Sabato G (2007). Modelling credit risk for SMEs: Evidence from the U.S. market. *Abacus*, 43(3), 332–357. https://doi.org/10.1111/j.1467-6281.2007.00234.x

[2] Aror E, Mupa MN (2025). Housing law and health outcomes. *World Journal of Advanced Research and Reviews*, 27(3), 1767–1774. https://doi.org/10.30574/wjarr.2025.27.3.3326

[3] Baesens B, Van Gestel T, Viaene S, Stepanova M, Suykens J, Vanthienen J (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54(6), 627–635.

[4] Bellotti T, Crook J (2009). Support vector machines for credit scoring and discovery of significant features. *Expert Systems with Applications*, 36(2), 3302–3308. https://doi.org/10.1016/j.eswa.2008.01.005

[5] Berger AN, Udell GF (2006). A more complete conceptual framework for SME finance. *Journal of Banking & Finance*, 30(11), 2945–2966.

[6] Holod D, Peek J, Torna G (2024). Relationship lending: That ship has not sailed for community banks. *Federal Reserve Bank of Boston Research Department Working Paper No*. 24-5. https://doi.org/10.29412/res.wp.2024.05

[7] Jacobs M Jr (2024). Benchmarking alternative interpretable machine learning models for corporate probability of default. *Data Science in Finance and Economics*, 4(1), 1–52. https://doi.org/10.3934/DSFE.2024001

[8] Jagtiani J, Lemieux C (2019). The roles of alternative data and machine learning in fintech lending: Evidence from the LendingClub consumer platform. *Financial Management*, 48(4), 1009–1029. https://doi.org/10.1111/fima.12295

[9] Leo M, Sharma S, Maddulety K (2019). Machine learning in banking risk management: A literature review. *Risks*, 7(1), 29. https://doi.org/10.3390/risks7010029

[10] Lessmann S, Baesens B, Seow H-V, Thomas LC (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1), 124–136. https://doi.org/10.1016/j.ejor.2015.05.030

[11] Li H, Wu W (2024). Loan default predictability with explainable machine learning. *Finance Research Letters*, 60, 104867. https://doi.org/10.1016/j.frl.2023.104867

[12] Malekipirbazari M, Aksakalli V (2015). Risk assessment in social lending via random forests. *Expert Systems with Applications*, 42(10), 4621–4631. https://doi.org/10.1016/j.eswa.2015.02.001

[13] Matsebula NC, Postigo Toledo C, Saungweme J, Masunungure M, Mupa MN (2025). Agile-predictive convergence: A new paradigm for smart investment and risk management platforms. *Iconic Research and Engineering Journals*, 9(1), 1247–1257.

[14] Mupa MN, Chiganze FR, Mpofu TI, Mangeya R, Mubvuta M (2024). The evolving role of management accountants in risk management and internal controls in the energy sector. *Iconic Research and Engineering Journals*, 8(2), 859–881.

[15] Nguyen H, Sagara N (2020). Credit risk database for SME financial inclusion. *ADBI Working Paper* 1111. Asian Development Bank Institute, Tokyo.

[16] Taanisa T, Mukwata NA, Sydney J, Ganyani L, Maturure RN, Chingezi E, Yelduora PG, Mupa MN (2026). From variance analysis to decision intelligence: Integrating finance audit and data science to improve SME budget governance and risk-aware growth. *World Journal of Advanced Research and Reviews*, 30(3), 666–675. https://doi.org/10.30574/wjarr.2026.30.3.1597

[17] US Small Business Administration (2026). 7(a) & 504 FOIA [dataset]. https://data.sba.gov/dataset/7-a-504-foia (accessed 30 June 2026).

[18] US Small Business Administration Office of Advocacy (2026). Frequently asked questions about small business 2026. https://advocacy.sba.gov/2026/02/03/frequently-asked-questions-about-small-business-2026/ (accessed 30 June 2026).