



Closing Time: Predicting Next-Quarter Establishment Closures for the Census Bureau

Project Lead: Gavin An

Analysts: Vedika Baradwaj, Ludovico Graber, Valerie Gutierrez, and Parth Halani*

Spring 2026

Abstract

The U.S. Census Bureau asked whether commercial and public data could identify small-business establishments at risk of closing in the next calendar quarter. We construct an establishment-quarter panel by linking SafeGraph point-of-interest closures to Rhetorik firmographics, county labor-market and housing series, Quarterly Workforce Indicators, and matched Small Business Administration loan records over 2019Q3–2025Q2. We estimate twenty-four gradient-boosted models formed from two outcomes, four feature specifications, and three algorithms (CatBoost, XGBoost, and NGBoost). On held-out test data, the highest F_1 scores are approximately 0.36 for closure during the panel window and 0.27–0.28 for closure in the following quarter. Static establishment attributes predict comparatively well on their own. Quarterly data alone perform worse in most specifications, and the combined static, quarterly, and lag specification performs best. Predictive performance is limited by commercial closure labels, the predominance of static firm-level fields, and reliance on county- or industry-level economic measures. Results should also be read in light of the COVID-19 period the panel spans. Replication files are available at <https://github.com/water6973/CREL-Census>.

*All of the listed authors study at the University of Chicago. Special thanks to Rebecca Hutchinson at the U.S. Census Bureau for the support and patience throughout this project, and to Darshan Kommanapalli for the opportunity with CREL.

Contents

1	CREL and Project Overview	3
1.1	Charles River Economics Labs	3
1.2	Partner Organization	3
1.3	Project Background	3
1.4	Motivation	4
2	Literature Review	4
3	Data and Methodology	5
3.1	Datasets	5
3.1.1	SafeGraph Global Places	5
3.1.2	Rhetorik Company Data (Dewey)	5
3.1.3	County macroeconomic and labor-market series	6
3.1.4	SBA 7(a) and 504 FOIA loans	6
3.2	Cleaning	6
3.3	Model	7
3.4	Limitations	8
3.5	Replication materials	9
4	Results	9
5	Conclusion	11

1 CREL and Project Overview

1.1 Charles River Economics Labs

Charles River Economics Labs is a group of undergraduate researchers at the University of Chicago, Harvard, and Stanford that completes pro bono economics research for governments and other social enterprises. The organization is advised by graduate researchers and faculty members at Harvard and the University of Chicago. Its past and current research partners include the United Nations Development Programme, the Department of Transportation, World Bank, the US Department of State, and the Census Bureau, among others.

1.2 Partner Organization

The U.S. Census Bureau is the federal government’s principal statistical agency for population and economic measurement. It produces official statistics on businesses, households, and the national economy that inform federal programs, state policy, and private-sector planning. Programs such as Business Dynamics Statistics summarize establishment openings and closings at the national and industry level, but timely, location-specific distress is often visible only after closures enter administrative records. For this project, Census researchers asked whether models could rank establishments by the chance of closing in the coming calendar quarter, using commercial geospatial data and public economic indicators together with the Bureau’s business-dynamics work.

1.3 Project Background

CREL partnered with the Census Bureau in Spring 2026 on a pilot closure-prediction study. The main task is binary classification at the establishment-quarter level: given a small business with certain firm- and location-specific features in quarter t , predict whether it closes in quarter $t + 1$. The label is `CLOSED_NEXT_QUARTER`, built from SafeGraph `CLOSED_ON` dates on each point of interest (POI).

The modeling pipeline links SafeGraph Global Places to firmographic records from Rhetorik (via Dewey Data), then enriches each location with county-level labor-market, housing, income, and GDP measures, Quarterly Workforce Indicators (QWI), and matched Small Business Administration (SBA) loan fields where a federal credit record can be tied to the establishment. We train gradient-boosted

tree models—CatBoost, XGBoost, and NGBoost—on a panel spanning 2019Q3 through 2025Q2.

1.4 Motivation

Establishment closures affect local employment, tax bases, and retail corridors, yet they are often detected only after the fact in traditional business registers. Earlier warnings could help the Census Bureau prioritize follow-up, validate alternative data sources, and study how macroeconomic shocks propagate to the storefront level. Prior work on short-horizon commercial-location closure has reported only modest predictive performance, and our benchmarks are consistent with that pattern. Gradient-boosted trees accommodate mixed numerical and categorical predictors, tolerate missing values, and model nonlinear interactions between local economic conditions and firm-level characteristics.

2 Literature Review

Gradient boosting fits a sequence of small trees to the loss gradient (Friedman, 2001). We use three implementations—XGBoost (Chen and Guestrin, 2016), CatBoost (Prokhorenkova et al., 2018), and NGBoost (Duan et al., 2020)—on the panel in Section 3.

A growing applied literature uses gradient-boosted models for business survival and related outcomes. Starakiewicz and Wojcik (2025) predict restaurant survival from nationwide Google Maps data in Poland with XGBoost. Snow (2019) train LightGBM on Yelp records to forecast restaurant exits one to two years ahead. Zhang and Luo (2023) use Yelp reviews and consumer-posted photos to predict U.S. restaurant survival up to three years forward. Vallapuram et al. (2021) study Foursquare location data and find that gradient-boosted decision trees outperform logistic regression and neural networks on several feature sets. Duprey et al. (2022) show that Google Places openings and closures can be tracked in near real time, which supports the use of commercial POI feeds as outcome proxies when administrative registers lag. Parallel work on financial distress reaches similar conclusions. Smith and Alvarez (2022) apply XGBoost to Spanish bankruptcy from accounting ratios. Schalck and Yankol-Schalck (2021) show that XGBoost outperforms dynamic probit and Lasso models for French SME failure. Overall, these prior studies show a wide spread in absolute performance, with AUCs anywhere between 0.6 and 0.95 depending on the quality of available data.

3 Data and Methodology

3.1 Datasets

Seven external sources are merged into the establishment-quarter panel: SafeGraph Global Places (outcomes and location), Rhetorik company data (firmographics), BLS county unemployment, FHFA house prices, BEA county income and GDP, Census Quarterly Workforce Indicators, and SBA 7(a)/504 FOIA loans. Join keys include `PLACEKEY`, county FIPS, two-digit NAICS, and calendar quarter. We include a source when it plausibly signals demand, cost pressure, credit access, or competitive conditions at the establishment, county, or industry level. Static POI and firmographic fields capture scale, sector, and location quality. County macro and QWI series capture local business cycle and labor-market stress. SBA fields proxy formal credit relationships. We exclude outcome fields (`CLOSED_ON`, `STATUS`, `CLOSED_NEXT_QUARTER`) from predictors to avoid leakage. Vendor fields such as operating hours, website URLs, and phone numbers were dropped after inspection (see Section 3.2) because they were sparse, inconsistently coded, or duplicated signals already in NAICS, revenue, and employee counts. Features we would have preferred but could not obtain at establishment-quarter frequency include quarterly revenue and payroll from tax records, lease terms, commercial rent indices at the storefront level, and administrative exit filings. Section 3.4 discusses these gaps further.

3.1.1 SafeGraph Global Places

SafeGraph Global Places provides point-of-interest records and building footprints for non-residential locations (SafeGraph, 2024). Each POI carries a unique `PLACEKEY`, name, address, coordinates, NAICS code, brand tags, and closure status (and if applicable, date). Building polygons add, among other things, approximate size in square meters and whether a parking lot is included on-site. Closure in the next calendar quarter is derived from `CLOSED_ON` relative to quarter end dates.

3.1.2 Rhetorik Company Data (Dewey)

The Rhetorik company registry, accessed through Dewey Data (Rhetorik, 2025), supplements SafeGraph with founding dates, revenue, employee counts, legal-entity type, and a small-business flag. We link records to POIs by address, name similarity, and distance (Section 3.2) and use Rhetorik

founding dates for panel age where opening dates are missing. Exit timing remains anchored to SafeGraph.

3.1.3 County macroeconomic and labor-market series

County-level unemployment (BLS LAUS) ([U.S. Bureau of Labor Statistics, 2024](#)), house prices (FHFA HPI) ([Federal Housing Finance Agency, 2024](#)), and income or GDP (BEA) ([U.S. Bureau of Economic Analysis, 2024](#)) are aggregated to calendar quarters and merged on COUNTY_FIPS. Quarterly Workforce Indicators ([U.S. Census Bureau, 2024](#)) supply county–industry–quarter employment and earnings, matched to each establishment’s NAICS prefix and size and age buckets. Eight-quarter rolling means of QWI and macro fields provide lagged predictors for next-quarter models and were selected because they yielded the highest SHAP values when compared to other possible lagged features.

3.1.4 SBA 7(a) and 504 FOIA loans

Public SBA loan records ([U.S. Small Business Administration, 2024](#)) cover the 7(a) guaranty and 504 certified development company programs. After linkage to POIs, the panel retains static fields such as loan status, approval fiscal year, and gross approval amount. The 7(a) program guarantees bank-originated small-business loans. The 504 program finances fixed assets through a bank–CDC–borrower capital stack.

3.2 Cleaning

The training dataset was built by merging SafeGraph Global Places to Rhetorik company records. A match required the same leading street number and five-digit ZIP code, standardized names scoring at least ninety on RapidFuzz’s weighted ratio, and haversine distance of no more than two hundred fifty meters between coordinates. When more than one company qualified for the same POI, the POI row was discarded. Synthetic POIs and non-small businesses were excluded. Spot checks supported the match quality, yielding roughly 1.2 million pairings from 24.3 million POI rows.

After linkage, we filtered to establishments with a usable founding date. Rhetorik founding dates are much more complete than SafeGraph opening dates for single-location firms (roughly 1.14 million out of 1.21 million compared to 2,000, with only 179 establishments having SafeGraph dates only),

so founding date anchors panel age, and roughly 70,000 rows without a start date were dropped. We derived county FIPS from coordinates and two-digit NAICS prefixes from full codes. SBA 7(a) and 504 FOIA loans were merged with the same address, name, and distance rules, attaching about 36,000 locations. When multiple loans matched one establishment, we kept the most recent. We judged repeated-loan history too sparse to model reliably without inflating dimensionality (only 8,063 establishments took out more than one loan), and `loanstatus` on the latest record should still capture the financial standing of the firm. A natural extension with more data would add loan count or total inflation-adjusted exposure.

Quarterly panels expand each establishment from 2019Q3 through 2025Q2 (twenty-four quarters). Membership begins in the first quarter whose end date is on or after the founding date and ends in the quarter of closure or at the window end if still open. Rows for establishments that closed before the panel start or fall in excluded industries are omitted. Each establishment-quarter carries `QUARTER`, age at quarter end, status, and `CLOSED_NEXT_QUARTER` as the supervised target. Company fields that remained low-signal based on SHAP value analysis despite feature engineering (e.g., hours, website, and phone number) were removed to save memory.

QWI statistics were joined on county, industry, quarter, and establishment size and age buckets. Roughly 92.9% of rows received a match. Rows with NAICS sector 92 was dropped due to sparse QWI coverage. Eight-quarter rolling means of selected QWI and county-macro variables provide lagged features for prediction at $t + 1$ without leakage from quarter $t + 1$ itself. The merged analysis file contains on the order of 27 million establishment-quarters and 63 modeling features. For the panel-period task, establishments are assigned to train (70%), validation (15%), or test (15%), with one row per location. For the next-quarter task, training quarters end in 2023Q4, validation runs through 2024Q2, and test covers 2024Q3–2025Q2.

3.3 Model

We estimate two related binary outcomes on the same establishment panel. The panel-period task asks whether an establishment has a recorded closure date during the study window (`CLOSED_ON` set), with an establishment-level split so that no location appears in both training and test sets. The next-quarter task asks whether the location closes in the following calendar quarter (`CLOSED_NEXT_QUARTER`), with a temporal split so that test quarters are strictly later than training

quarters.

For each outcome we estimate twelve models by crossing three gradient-boosted algorithms with four feature specifications, for twenty-four models in total. The static specification includes time-invariant establishment and location fields (POI geography, Rhetorik firmographics, SBA loan attributes, and related columns listed under **STATIC** in `feature_categories.json`). The quarterly specification includes only fields that vary by calendar quarter (listed under **QUARTERLY**), such as establishment age in the panel and county- and industry-level economic measures. The quarterly-plus-lags specification adds eight-quarter rolling means of selected quarterly series (**LAGGED**). The full specification uses static, quarterly, and lag columns together. The replication code labels these four specifications `static_only`, `quarterly_only`, `quarterly_lagged`, and `full`.

For the next-quarter task we omit the categorical quarter label from the quarterly blocks and instead include sine and cosine seasonality terms based on calendar quarter. This prevents the model from overfitting on historical shocks such as COVID-19 while still capturing the effects of seasonality (e.g., do businesses close more often in the winter?) Hyperparameters are chosen by random search on the validation set (eight trials per model). To prevent leakage, validation quarters are strictly later than training quarters and earlier than test quarters. Furthermore, within the training and validation sets, the model cannot see year-quarter: only calendar quarter. Test metrics include average precision, precision and recall at the validation-tuned F_1 threshold, and precision among the highest-scored 1%, 2%, and 5% of test rows.

3.4 Limitations

The main limitations are data-related. Closure outcomes and locations come from commercial point-of-interest records, which are less accurate than establishment-level data the Census Bureau could produce from its own registers. Rhetorik firmographics are static within the panel, so the models do not observe quarterly changes in firm-level employment or revenue from that source. The panel also covers the COVID-19 period, when closure rates and local economic conditions were unusual relative to a typical business cycle.

Beyond these specific sources, the project had limited quarterly depth and limited establishment-level coverage. Most predictors are county- or industry-aggregated, lagged, or measured only once per establishment, and the static fields that do exist are not uniformly high quality. Few datasets meet

the combination of requirements this application imposes, including establishment-level geography, frequent updates, linkage across vendors without heavy anonymization, accessibility to student researchers, and accuracy comparable to federal statistics. These results describe what the linked commercial and public files support. They should not be read as an upper bound on what the Census Bureau could obtain with administrative ground truth.

3.5 Replication materials

Replication files are posted at <https://github.com/water6973/CREL-Census>. The repository contains one merged panel (`establishment_quarter_panel.parquet`, Git LFS), the column manifest `feature_categories.json`, and a single training entry point `train_closure_models.py`. The command `python scripts/train_closure_models.py -v2` reads that parquet and retrains all twenty-four models in Section 4. The module `panel_sampling.py` constructs the train, validation, and test splits. Licensed vendor extracts are not redistributed.

4 Results

This section reports out-of-sample predictive performance for the two binary outcomes defined in Section 3.3. The panel-period test set comprises approximately 172,000 establishments (one row per location), with a closure prevalence near 13%. The next-quarter test set comprises approximately 4.3 million establishment-quarters in 2024Q3–2025Q2, with a closure prevalence near 0.5%. Each row in the tables below corresponds to one of twenty-four fitted models (two outcomes, four feature specifications, and three algorithms).

We rank establishments by predicted closure probability. On the validation set we select the threshold that maximizes the F_1 score and apply that threshold unchanged to test scores. Table 1 and Table 2 report test precision, recall, and F_1 at this operating point. Figure 1 reports precision–recall curves that do not depend on a single threshold.

As shown in Table 1, models trained on static firmographics and location attributes alone achieve test F_1 scores near 0.35 for the panel-period outcome. Models trained on quarterly data alone reach lower F_1 scores, near 0.31–0.32. Adding eight-quarter rolling means to the quarterly block does not close that gap. The full specification, which combines static fields, quarterly data,

Table 1: Panel-period closure: test precision and recall at the validation-tuned threshold that maximizes F_1 (raw scores)

Features	XGBoost			NGBoost			CatBoost		
	P	R	F_1	P	R	F_1	P	R	F_1
Static only	0.272	0.514	0.355	0.255	0.537	0.345	0.267	0.527	0.354
Quarterly only	0.246	0.468	0.322	0.232	0.472	0.311	0.241	0.478	0.321
Quarterly + lags	0.249	0.460	0.323	0.219	0.508	0.306	0.237	0.490	0.319
Full	0.269	0.542	0.359	0.268	0.516	0.353	0.282	0.506	0.362

Panel-period outcome: **CLOSED_ON** during 2019Q3–2025Q2. Validation F_1 threshold applied to test scores by algorithm.

Table 2: Next-quarter closure: test precision and recall at the validation-tuned threshold that maximizes F_1 (raw scores)

Features	XGBoost			NGBoost			CatBoost		
	P	R	F_1	P	R	F_1	P	R	F_1
Static only	0.174	0.493	0.258	0.160	0.542	0.248	0.171	0.521	0.258
Quarterly only	0.219	0.264	0.240	0.190	0.368	0.251	0.210	0.305	0.249
Quarterly + lags	0.212	0.297	0.247	0.196	0.392	0.262	0.206	0.328	0.253
Full	0.223	0.326	0.265	0.197	0.476	0.278	0.211	0.386	0.272

Outcome is **CLOSED_NEXT_QUARTER**. Temporal holdout on calendar quarters. Threshold tuned on validation F_1 .

and lags, attains the highest F_1 in each column. CatBoost reaches 0.362 and the other algorithms are close behind. Panel-period models recover more closure events at the validation-tuned threshold (recall near 0.50–0.54) than next-quarter models, with precision near 0.27–0.28.

Table 2 reports the next-quarter outcome, which is harder in this panel. Static-only models reach test F_1 near 0.25–0.26. Quarterly-only models show a different precision–recall mix. Precision rises relative to static-only runs for XGBoost and CatBoost, but recall falls, so F_1 remains near 0.24–0.25. This is most likely because our quarterly-data is not firm-specific and therefore only weakly separates individual storefronts. The quarterly-plus-lags specification improves recall for NGBoost in particular. The full specification again performs best. NGBoost attains the highest next-quarter F_1 (0.278) and CatBoost and XGBoost are near 0.27. These magnitudes are modest in absolute terms, which is consistent with predicting a rare event one quarter ahead using commercial closure labels.

Figure 1 displays precision–recall curves for each combination of outcome and algorithm. The four curves within a subplot correspond to the static-only, quarterly-only, quarterly-plus-lags, and full feature specifications. In many panel-period subplots, the quarterly-only curve lies below the static-

only curve. Time-varying county and industry series therefore do not replace establishment-level attributes when static fields are omitted. The full curve typically lies furthest from the prevalence baseline. Markers on each curve show the validation-tuned F_1 point on the test set.

Across algorithms, differences are small relative to differences across feature specifications. CatBoost, XGBoost, and NGBoost rank establishments similarly once the feature set is fixed. No estimator uniformly dominates the others on every metric in Table 1 and Table 2. The comparison is therefore best read as a measure of what each data block contributes when the learning algorithm is held constant.

5 Conclusion

This project set out to assess whether commercial geospatial and economic data can support short-horizon prediction of small-business establishment closures for the U.S. Census Bureau. We linked SafeGraph point-of-interest closures to Rhetorik firmographics, matched SBA loan records, and county-level labor-market and housing data, and expanded the linked file into an establishment-quarter panel from 2019Q3 through 2025Q2. We then estimated twenty-four boosted-tree classifiers defined by two outcomes, four feature specifications, and three algorithms.

The panel-period outcome asks whether an establishment records a closure date at any point during the study window. Under the establishment-level test split, the best models reach test F_1 scores near 0.36 when static, quarterly, and lagged predictors are used together. Models trained on static data alone perform nearly as well as this full specification in several columns of Table 1, which indicates that time-invariant location and firmographic attributes account for much of the measurable signal in the commercial files. Models trained on quarterly data alone, without static fields or lags, perform worse. Adding eight-quarter rolling means to the quarterly block improves some estimates but does not match the full specification.

The next-quarter outcome corresponds to the question posed by the Census Bureau: whether a location open in quarter t closes in quarter $t + 1$. This task is harder. Test F_1 scores peak near 0.27–0.28 in the full specification, and precision at the validation-tuned threshold remains limited. The ranking exercise is nevertheless informative. It shows that external data can be joined to form a large panel and scored one quarter ahead with standard software. It also shows where the data are

Test precision--recall curves by feature set (raw scores). Markers: validation F_1 -optimal threshold applied to test.

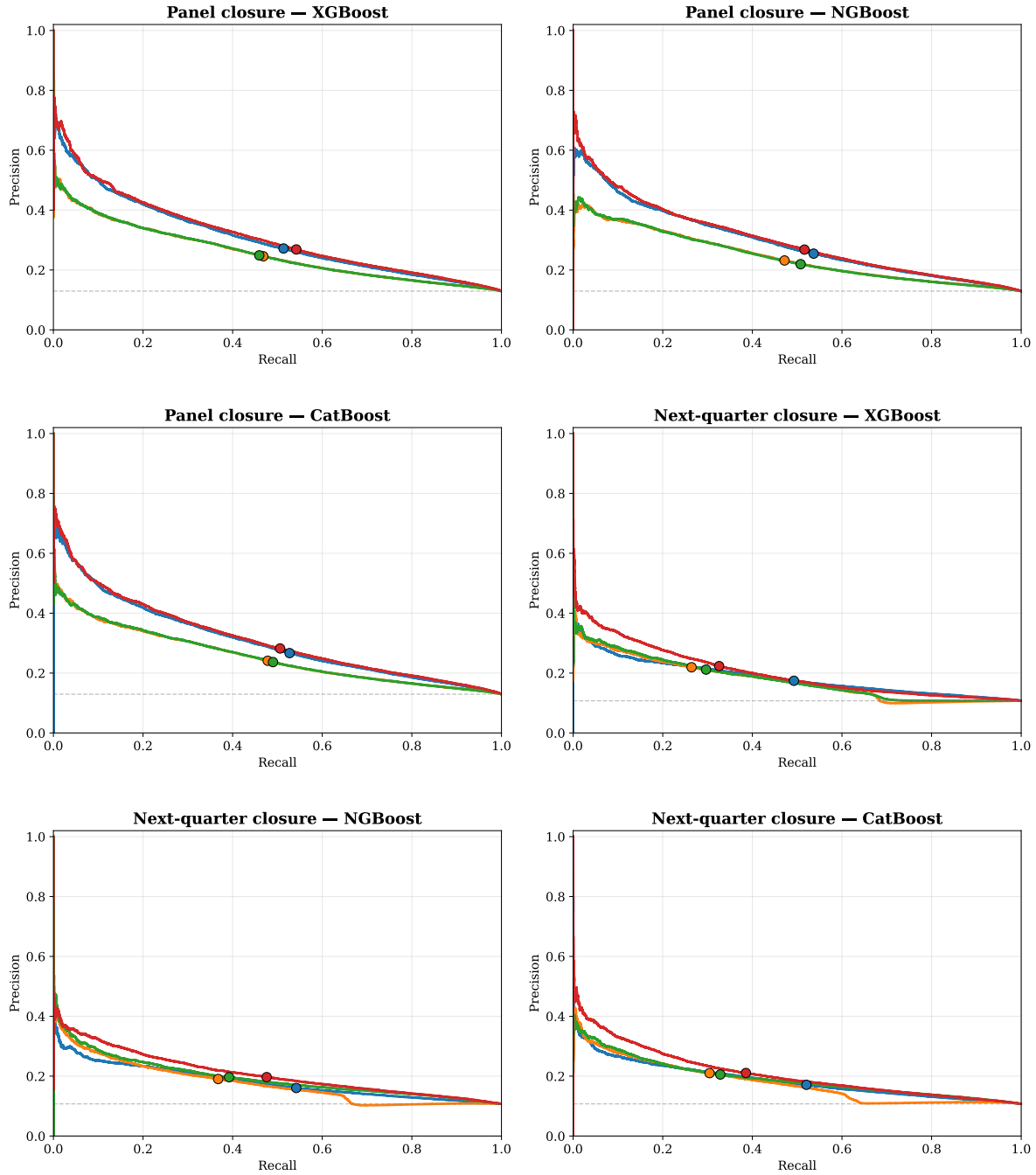


Figure 1: Test precision--recall curves (raw scores). Each subplot is one outcome and one algorithm. Blue: static data only. Orange: quarterly data only. Green: quarterly data plus eight-quarter lags. Red: static, quarterly, and lag data combined. Markers: validation-tuned F_1 on the test set. Dashed lines: test-set closure prevalence.

thin, particularly the reliance on POI closure dates and county-aggregated economic series rather than establishment-level administrative records.

Across CatBoost, XGBoost, and NGBoost, differences in test metrics are small relative to differences across feature specifications. Expanding the feature set from quarterly data alone to the combined static–quarterly–lag specification matters more than switching among algorithms. We therefore emphasize data linkage and feature construction over estimator choice when interpreting the benchmarks.

Replication materials at <https://github.com/water6973/CREL-Census> contain the merged training panel and the script needed to reproduce all twenty-four models. A natural extension for the Census Bureau would be to replace or calibrate commercial closure labels with administrative records, refresh firm-level covariates at quarterly frequency where possible, and evaluate models on internal ground truth rather than on held-out commercial data alone.

References

- Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. ACM, 2016.
- Tony Duan, Anand Avati, Daisy Yi Ding, Khanh K. Thai, Sanjay Basu, Andrew Ng, and Alejandro Schuler. NGBoost: Natural gradient boosting for probabilistic prediction. In *Proceedings of the 37th International Conference on Machine Learning*, pages 2690–2700. PMLR, 2020.
- Thibaut Duprey, Daniel E. Rigobon, Artur Kotlicki, Philip Schnattinger, Soheil Baharian, and Thomas R. Hurd. Business closures and (Re)Openings in real-time using Google Places: Proof of concept. *Journal of Risk and Financial Management*, 15(4):183, 2022. doi: 10.3390/jrfm15040183.
- Federal Housing Finance Agency. House price index datasets. <https://www.fhfa.gov/data/hpi>, 2024.
- Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5):1189–1232, 2001.
- Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey

- Gulin. CatBoost: Unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems 31*, pages 6638–6648, 2018.
- Rhetorik. Company data. <https://doi.org/10.82551/9HSG-6162>, 2025. Dataset distributed via Dewey Data.
- SafeGraph. Places documentation. <https://docs.safegraph.com/docs/places>, 2024.
- Christophe Schalck and Meryem Yankol-Schalck. Predicting French SME failures: New evidence from machine learning techniques. *Applied Economics*, 53(51):5948–5963, 2021. doi: 10.1080/00036846.2021.1934389.
- Matthew Smith and Francisco Alvarez. Predicting firm-level bankruptcy in the Spanish economy using extreme gradient boosting. *Computational Economics*, 59(1):263–295, 2022. doi: 10.1007/s10614-020-10078-2.
- Derek Snow. Predicting global restaurant facility closures, 2019. SSRN Working Paper 3420490.
- Tomasz Starakiewicz and Piotr Wojcik. Predicting restaurant survival using nationwide Google Maps data. *Knowledge-Based Systems*, 313:113198, 2025. doi: 10.1016/j.knosys.2025.113198.
- U.S. Bureau of Economic Analysis. Regional economic accounts: GDP and personal income by county. <https://www.bea.gov/data/gdp/gdp-county-metro-and-other-areas>, 2024.
- U.S. Bureau of Labor Statistics. Local area unemployment statistics. <https://www.bls.gov/lau/>, 2024.
- U.S. Census Bureau. Quarterly workforce indicators. <https://lehd.ces.census.gov/data/qwi/>, 2024.
- U.S. Small Business Administration. 7(a) and 504 FOIA loan data. <https://data.sba.gov/datasets/7-a-504-foia>, 2024.
- Anish K. Vallapuram, Nikhil Nanda, Young D. Kwon, and Pan Hui. Interpretable business survival prediction, 2021. arXiv:2109.12370.
- Mengxia Zhang and Lan Luo. Can consumer-posted photos serve as a leading indicator of restaurant survival? evidence from Yelp. *Management Science*, 69(1):25–50, 2023. doi: 10.1287/mnsc.2022.4359.