ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct786

Data-Driven Credit Risk Assessment in U.S. Public Finance: A Review of Rating-Based, Market-Based, and Machine Learning Approaches Under Sparse Default and Delayed Disclosure

Godfred Yankson[1]; Mathew Oman-Amoako[2]; Sewando Erick Mkuchu[3]; Barnabas Anim[4]

[1]Auburn University, USA
[2]Accra Institute of Technology, Ghana
[3]Independent Researcher, Massachusetts, USA
[4]University of Bradford, UK

Publication Date: 2026/11/04

Abstract: Credit risk assessment in U.S. public finance faces two persistent constraints. Conventional predictive validation is limited by the rarity of municipal defaults and financial disclosure may not be timely with underlying fiscal conditions. This review compares three approaches to municipal credit risk assessment: rating-based, market-based, and machine-learning approaches. They are assessed using a common framework that includes information basis, reliability, timeliness, predictive usefulness, assessment purpose, constraint adaptation and institutional maturity. Rather than applying each paradigm's conventional evaluation standard, the review holds the two constraints constant and examines how each approach manages the resulting trade-offs. Ratings rely on institutional judgment to include information beyond structured data but have raised reliability issues with agency incentives. Market signals update continuously but also reflect liquidity, tax treatment, and dealer behavior. Machine learning models can use broader and earlier information sources but are severely constrained in their ability to be validated due to the sparsity of default labels. Across all three paradigms, a central distinction emerges between assessing ongoing creditworthiness and predicting the rare event of default. This distinction helps explain why apparent weaknesses may reflect a mismatch between method and assessment purpose rather than methodological failure. The review contributes a structured comparison of these paradigms around the joint constraints of sparse defaults and delayed disclosure, which existing research has largely examined in isolation.

Keywords: *Municipal Credit Risk, Credit Ratings, Market-Based Assessment, Machine Learning, Sparse Default Events, Delayed Disclosure, Public Finance.*

How to Cite: Godfred Yankson; Mathew Oman-Amoako; Sewando Erick Mkuchu; Barnabas Anim (2026) Data-Driven Credit Risk Assessment in U.S. Public Finance: A Review of Rating-Based, Market-Based, and Machine Learning Approaches Under Sparse Default and Delayed Disclosure. *International Journal of Innovative Science and Research Technology*, 11(10), 2848-2856. https://doi.org/10.38124/ijisrt/26oct786

I. INTRODUCTION

The municipal bond market functions as the primary capital market for state and local government issuers in the United States (Cestau et al., 2019), financing schools, water systems, transportation networks, and the broader infrastructure that public services depend on. Assessing credit risk in this market matters for the same reason it is unusually difficult to do well. Tens of thousands of distinct issuers participate in it, most of them small, most of them covered by no independent analyst research, and the market as a whole has only recently become tractable to systematic study following regulatory initiatives that expanded access to granular trade data.

This market poses two characteristics that make credit risk assessment more complicated than most other asset classes. The first is how rarely default actually occurs. Moody's Investors Service (2023) reports a five-year cumulative default rate for its rated municipal universe that has held near 0.08 percent since 2013, a rate low enough that the outcome any predictive method would need to learn from

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct786

is present in only a small fraction of the tens of thousands of issuers in the market. The second is the time lag in the availability of information on an issuer's financial condition. Timeliness of governmental financial reporting has drawn sustained attention as a problem distinct from private-sector disclosure delay (Edmonds et al., 2017), and current federal rulemaking under the Financial Data Transparency Act, with joint data standards that took effect in October 2026 and municipal-specific application still pending, is direct evidence that this constraint remains unresolved rather than historical (Financial Data Transparency Act Joint Data Standards, 2026; Securities and Exchange Commission, 2026).

These two constraints, sparse defaults and delayed disclosure, are the organizing problem that this review addresses, and are the reason to consider the three paradigms of assessing municipal credit risk, credit ratings, market-based signals, and machine learning models together, rather than in parallel. Existing scholarship tends to evaluate each paradigm on its own terms, against its own field's conventional success criteria: rating stability studies judged against rating transitions, market microstructure studies judged against pricing efficiency, and machine learning studies judged against classification accuracy on a constructed label. This separation makes it hard to know what a common issue in all three is: an outcome too rare to validate against directly, and information too delayed to assess current conditions with confidence. This review asks how rating-based, market-based, and machine-learning approaches trade off reliability, timeliness, and predictive usefulness once that joint constraint is held constant across all three, rather than asking which paradigm performs best by the standard its own literature applies.

A distinction carried through every part of this review, and one the existing literature frequently collapses, is that assessing an issuer's ongoing creditworthiness and predicting whether that issuer will default are not the same task. The first asks how a fiscal condition is trending given currently available information. The second asks whether a specific, rare, binary event will occur within a defined horizon. A method can be genuinely useful for the first purpose while contributing very little to the second, and judging all three paradigms by the rare-event standard alone risks understating what ratings and market signals actually do well.

This review proceeds by first establishing the public finance information environment and the two constraints in more detail, then examining each of the three paradigms in turn against a shared set of questions covering information basis, reliability, timeliness, predictive usefulness, constraint adaptation, and institutional maturity, before drawing the three together in a comparative synthesis that identifies where the paradigms genuinely differ and where the existing literature leaves real gaps still to be closed.

II. METHODOLOGY

- *Review Design and Comparative Framework*

This review contrasts three literatures, credit ratings, market-based signals, and machine learning approaches to municipal credit risk within a shared framework of seven dimensions: information basis, reliability, timeliness, predictive usefulness, assessment purpose, constraint adaptation, and applicability and maturity. Stability is considered part of reliability and not a separate dimension. These dimensions organize the argument within each section rather than dictating separately headed subsections throughout.

- *Search Strategy and Inclusion and Exclusion Criteria*

The three paradigms came from different research fields that often use different terminology, so the search was performed independently for each paradigm rather than using a single search string. Rating-based search terms included municipal credit rating, issuer-pay model, rating recalibration, and rating agency conflict of interest. Market-based terms included municipal bond spread, municipal bond liquidity, dealer markup, and municipal disclosure and trading. Machine-learning terms included machine learning credit risk, fiscal distress prediction, municipal bankruptcy prediction, and bond default prediction.

Search was conducted through general academic web search, drawing on peer-reviewed publishers including the Journal of Finance, Review of Financial Studies, Journal of Accounting Research, Journal of Accounting and Economics, and Management Science, working paper repositories including SSRN, Federal Reserve working paper series, and arXiv, and institutional sources including the SEC, GASB, MSRB, and GFOA. The primary review window is January 2015 to October 2026, selected to capture contemporary public-finance disclosure, credit-risk, and data-driven assessment research. GASB Statement 68 (Governmental Accounting Standards Board, 2012), effective for fiscal years beginning after June 15, 2014, provides context for the reporting environment during this period but does not dictate the window's boundaries. Two exceptions apply: a foundational exception for pre-2015 sources establishing a method or finding the argument depends on, applied to Ang, Bhansali, and Xing (2014) and Hájek (2011), and an institutional-source exception for current regulatory material regardless of date, applied to FDTA, SEC, GASB, and GFOA sources, since these establish the present state of a constraint this review treats as unresolved rather than historical.

Sources were included where they addressed U.S. municipal or public-finance credit risk directly, or, for machine learning, where a non-U.S. study was retained as an explicitly identified methodological analogue. Studies addressing consumer or corporate credit risk without transferable relevance, and papers addressing algorithmic performance without a credit-risk question, were excluded.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct786

- *Credit Risk Assessment in U.S. Public Finance: Context and Information Constraints*

The municipal bond market functions as the primary capital market for state and local government issuers in the United States (Cestau et al., 2019; Bergstresser, 2023). Recent regulatory initiatives requiring more granular trade data have made the market newly tractable to researchers, yet even with that improved data availability, trade-by-trade studies have documented large trading costs and persistent price dispersion, which are deviations from a single, consistent price for economically similar bonds across this market (Cestau et al., 2019). That persistence is itself informative: better data alone has not resolved the pricing frictions that characterize this market. Any credit risk assessment approach applied to this market inherits these structural conditions before it does anything else. The three paradigms this review examines, credit ratings, market-based signals, and machine learning models, are not simply competing techniques layered onto a neutral information environment. They are three different ways of coping with a market whose basic price and disclosure infrastructure has only recently been studied systematically.

It is important to note that the assessment of an issuer's creditworthiness and the prediction of whether an issuer will default are not the same, although the literature on all three paradigms often interchanges the two. Creditworthiness assessment asks how an issuer's fiscal situation is developing based on the information that is currently available. The question of default prediction is whether a specific, rare, binary event will happen within a certain time horizon. The distinction is not a claim from any one source but is an analytical framework of this review and is repeated in the three approach sections that follow.

That distinction matters because of how rarely default actually occurs in this market. Moody's Investors Service (2023) states that the average five-year cumulative default rate for its rated municipal universe, drawn from more than fifty thousand different state, local, and other municipal issuers, has remained around 0.08 percent since 2013, consistent with the trend over the entire period of the study, which started in 1970. Why defaults remain this rare is not fully settled by the data alone. One plausible mechanism, offered here as interpretation rather than as an established finding, is that general-purpose governments retain options including taxing authority, service reductions, and reserve draws that a corporation facing insolvency does not have in the same form, and that political incentives favor exhausting those options before a payment is missed. This review does not regard that mechanism as a fact, but rather as a hypothesis that needs to be tested in the context of the literature reviewed in subsequent sections.

Delayed disclosure compounds this problem rather than sitting alongside it as a separate concern. Timeliness of governmental financial reporting has drawn sustained academic and practitioner attention as a distinct problem from private-sector reporting delay (Edmonds et al., 2017). Using data drawn from nearly twelve thousand municipal audited financial reports, Carroll and Ciccarone (2024) find a direct correlation between audit timeliness and credit rating across governmental bond sectors and most revenue bond sectors, and report that median municipal audit completion times have slowed rather than improved over the past decade, reaching a twelve-year high. Four distinct moments are relevant for any assessment approach: when the underlying financial event happens; when the event is recorded; when the event becomes public; and when an assessment can actually include the event. In municipal finance, these moments can be separated by many months, a gap the corporate bond market, where audited financials are typically produced in sixty to ninety days against a municipal norm closer to one hundred forty to one hundred eighty, does not face at the same scale (Carroll & Ciccarone, 2024).

That constraint is not historical. The joint data standards required under the Financial Data Transparency Act became a final rule in June 2026, with an effective date of October 1, 2026 (Financial Data Transparency Act Joint Data Standards, 2026; Securities and Exchange Commission, 2026). That rule establishes machine-readable data standards but does not itself change what municipal issuers must disclose or when. That depends on a second phase of rulemaking applying these standards specifically to the municipal market through the Municipal Securities Rulemaking Board, a process the Government Finance Officers Association (n.d.) describes as still running on its own statutory timeline, with full issuer compliance not yet scheduled.

The two constraints do not operate independently, and their interaction is the organizing problem this review takes up. As this review argues, and given the default frequencies reported by Moody's Investors Service (2023), sparse defaults push toward approaches that require increased data and longer time series to compensate for the rarity of the outcome of interest. Delayed disclosure, whose extent is documented by Carroll and Ciccarone (2024), shortens the window of information that is both recent and reliably point-in-time valid. This tension between the two, rather than either constraint alone, is what the three approaches examined next are each built to manage differently.

- *Rating-Based Credit Risk Assessment*

Credit ratings remain a central institutional mechanism for assessing municipal creditworthiness, and their information basis differs fundamentally from the other two paradigms examined in this review. Ratings are produced by an analyst process that incorporates qualitative judgment, issuer communication, and proprietary criteria, not solely observable market data or structured financials. But that institutional role is accompanied by a structural characteristic that literature largely considers to be consequential and not just incidental to the institutional context. Rating agencies in the municipal market earn lower fees and face higher disclosure requirements than in corporate and structured finance markets, and Cornaggia et al. (2023) find that this fee structure is associated with distorted outcomes: using actual fees paid by Texas municipal issuers, they show that unexpectedly expensive ratings are more likely to be downgraded and unexpectedly inexpensive ratings more likely to be upgraded, an effect concentrated among issuers

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct786

who stand to lose access to AAA bond insurance. This is best understood as a reliability concern under the contemporaneous-information standard, that is, a rating can be defensible given the information available to the analyst yet still shaped by an incentive structure that has nothing to do with credit quality, rather than as evidence that ratings are unreliable in a general sense.

The clearest illustration of how much a rating level can move independent of underlying credit quality comes from the 2010 recalibration of Moody's and Fitch's municipal rating scales, a change that raised ratings without any corresponding change in issuer fundamentals. Beatty et al. (2019) report that this recalibration reduced yields for the impacted issuers, while also raising fees and market share for the agencies that adopted it, a pattern that suggests strategic incentives, but does not provide direct evidence that the model is generally distorted. Gillette et al. (2020) extend this finding in a way directly relevant to the disclosure constraint this review examines, which is that municipalities that received an exogenous rating upgrade from the recalibration subsequently reduced their voluntary disclosure of continuing financial information, an effect strongest among issuers monitored by investors who had relied most heavily on that disclosure, and absent where underwriters provided issuer-specific monitoring or where direct regulatory enforcement applied. Together, these two studies of the same event show that a rating change can alter the information environment it is meant to summarize.

Timeliness under delayed disclosure is where the institutional strength of ratings becomes a genuine limitation. Here it is important to separate two distinct problems the literature sometimes conflates, which include how quickly a rating reflects a changing credit condition, and how quickly the market that depends on a rating can actually access it. Zhang (2024) provides evidence for the second problem. A 2010 rule mandating that current ratings be made available on the Municipal Securities Rulemaking Board (MSRB)'s EMMA website was correlated with a 6 to 8 basis point decline in investor transaction costs, with the impact being strongest among retail investors on longer-maturity bonds where current ratings information previously was not displayed on EMMA. The rating itself was not changed. It was altered to show who views the rating and how fast. This is a visibility and access finding and not necessarily evidence that the underlying assessment of the analyst changes as quickly.

The distinction between assessing ongoing creditworthiness and predicting the rare event of default also affects how rating performance can actually be validated, given how infrequently default occurs. Instead of evaluating a single annual default observation, rating agencies and researchers look at rating transitions and default experience through broad levels of rating cohorts drawn from time. Moody's (2023) publishes one-year municipal rating-transition matrices tracking whether issuers remain in their initial rating category, migrate upward or downward, have their rating withdrawn, or default. The study also reports that Aaa-rated issuers remained Aaa one year later 94.9 percent of the time across the full 1970-2022 study period, a figure that fell to 92.8 percent over the more recent 2013-2022 window, evidence of somewhat greater rating volatility in the past decade even at the highest rating category. Over a five-year horizon, the average position of defaulted municipal credits relative to all rated credits was 97 percent. This means ratings correctly ranked all but a small fraction of eventual defaulters below higher-rated peers, compared to 84 percent for global corporate issuers on the same measure. This cohort-based approach does not solve the sparse-default problem, but it enables rating performance to be evaluated using stability, migration, and ordinal ranking of actual credit events, and not by individual annual default observations alone. Beatty et al. (2019) and Gillette et al. (2020) illustrate a complementary adaptation. Since default cannot be observed often enough to validate a rating change directly, both studies examine the institutional process and incentives surrounding the change, using the 2010 recalibration as an exogenous shock to separate the rating's movement from the underlying credit condition it was intended to track.

Taken together, the literature on rating-based assessment suggests an approach that is institutionally durable and capable of incorporating information no structured dataset would capture. It partly compensates for delayed disclosure through analyst judgment, issuer communication, and ongoing surveillance rather than waiting for a periodic filing cycle. But that compensation remains dependent on the availability and quality of issuer information reaching the analyst in the first place, and the paradigm's reliability is entangled with the economics of the agencies that produce it. Rating-based assessment should therefore not be judged solely by whether ratings predict defaults accurately. In municipal finance, ratings function as institutional summaries of ongoing creditworthiness, and their performance is more properly evaluated through rating transitions, market consequences, disclosure effects, and the quality of contemporaneously available information. This is the standard against which the next two sections evaluate market-based and machine-learning approaches in turn.

- *Market-Based Credit Risk Assessment*

Market-based assessment draws its information from a fundamentally different source than either institutional ratings or model-trained predictions, which includes observed prices and yields generated by actual trading activity. Schwert (2017) provides the clearest decomposition of what that price actually contains, using three complementary methods that consist of a transaction-based liquidity estimate, a credit default swap-based default estimate, and a quasi-natural experiment around pre-refunding events. This indicates that default risk accounts for 74 to 84 percent of the average municipal bond spread once tax-exempt status is accounted for. The price of default risk is high precisely because the risk is rare, implying a substantial risk premium is embedded even in ordinary trading. That is a striking finding given how rarely municipal default actually occurs. That finding on the level of the spread, however, is not the same question as what drives its movement, and this is where reliability becomes genuinely difficult to establish from market signals alone. Ang, Bhansali, and Xing (2014), studying the muni spread from

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct786

1995 to 2013, find that more than 90 percent of the variation in that spread over time is attributable to a time-varying liquidity component, not credit risk, which barely shifted across the 2008 financial crisis while liquidity effects moved sharply. Read alongside Schwert's finding, this means credit risk explains most of the spread's average level while liquidity explains most of its movement. This distinction matters directly for anyone trying to read a market signal for credit information: a widening spread on a given day is far more likely to reflect a liquidity shock than a genuine change in creditworthiness. Griffin, Hirschey, and Kruger (2023) add a further confound at the level of individual trades: municipal bonds show considerable pricing variation even for same-size trades of the same bond on the same day from the same dealer, with markups clustering around round price increments and dealers who price strategically showing systematically higher markups. And liquidity itself is not a stable background condition. Bagley, Gissler, Hiteshew, and Ivanov (2023) show that a discontinuous tax threshold in municipal bond holding periods causes institutional investors, mutual funds especially, to dispose of bonds as they approach that threshold in anticipation of future illiquidity, and that once a bond crosses it, trading activity falls and illiquidity-driven spreads rise substantially, with the effect amplified by monetary policy tightening. None of this is credit information. All of it moves the price a rating or a model might otherwise treat as a clean credit signal.

Timeliness is where market-based assessment's real advantage over an institutional process should show up. Prices can move continuously rather than waiting on a periodic review, but the evidence suggests this advantage depends heavily on what generates the price movement in the first place. Cuny, Li, Nakhmurina, and Watts (2025), examining the full universe of post-issuance municipal disclosures from 2009 to 2022, find that most municipal bonds simply do not trade in the weeks following a disclosure at all. Trading responds selectively in that investors trade more following credit-relevant disclosures, such as adverse credit events, and substantially less following routine annual financial statements, with the effect strongest for bonds that are large or already perceived as risky. The same study finds that credit rating agencies themselves respond to these disclosures, which suggests the market and the rating paradigm are not independent information channels but are pulling on some of the same underlying signals. Where trading does not follow a disclosure, Cuny et al. attribute the gap to a combination of prior informed trading, lack of timeliness, illiquidity, and investor information-processing constraints, meaning market timeliness is conditional on the type of information released and on structural features of the specific bond, not a general property of the market as a whole.

The creditworthiness-versus-default distinction applies here as it does to the other paradigms, but market-based signals face a validation problem specific to their reliance on trading itself. Ivanov, Zimmermann, and Heinrich (2025) show that governments fail to disclose material private debt placements 50 to 60 percent of the time, and that even where disclosure occurs, filings often omit contract terms necessary for accurate pricing, with noncompliant issuers significantly riskier than compliant ones. A market price can only reflect information that both exists and has actually traded, and validating a market signal against rare default outcomes runs into the same sparse-label problem the other two paradigms face, compounded by the fact that most individual CUSIPs trade only a handful of times a year or less, leaving long stretches where no price exists to validate against at all.

In the literature, adaptations to these two constraints are more about trying to work around the structural thinness of the market than about directly addressing it. Decomposition methods, separating credit from liquidity, tax, and frictional components rather than treating the raw spread as a single credit signal, are the primary response to the confounding problem, as in Schwert and in Ang, Bhansali, and Xing. A second, more limited adaptation follows from where the signal is strongest to begin with: Ivanov, Zimmermann, and Heinrich find that disclosure is especially informative for low-rated bonds and during market crises, which suggests market signals are most usefully relied upon exactly where and when other information sources are weakest, even though they remain an unreliable, always-on credit signal the rest of the time. Together, market-based assessment provides an advantage that ratings and machine-learning models don't through an evolving price that's built by real capital at risk. But that price is a joint product of credit risk, liquidity, tax treatment, and dealer behavior, and separating those components requires the kind of decomposition work this literature has only partially completed. Market-based assessment is consequently most useful as a timely but imperfect indicator of changing credit conditions, rather than as a complete or consistently reliable measure of individual default risk.

- *Machine Learning Credit Risk Assessment*

Machine learning applications to municipal credit risk sit in a different position from the other two paradigms examined in this review. The literature consists predominantly of academic studies, with comparatively limited evidence of operational deployment, and this asymmetry in maturity is itself one of this section's central findings, not a caveat to append at the end. Unlike the typical credit ratings and market pricing that are continuously generated by institutions whose daily activities are intertwined with the market, the ML in this context is still in the research phase, and this distinction influences how each of the other questions in this section should be interpreted.

The information basis of the ML approaches is also designed differently than what a rating analyst or a bond trader uses. Instead of using analysis or market prices, models are being developed to consume structured financial and statistical variables, and increasingly, unstructured texts. Hájek (2011) illustrates an early version of this: using financial statements and statistical reports from municipalities in Connecticut, with input variables selected through genetic algorithms, a neural network classified issuers into Moody's rating classes with a high accuracy rate using a limited subset of variables. This is a useful early result, but it is worth being precise about what it demonstrates. The model was trained to reproduce an existing

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct786

rating, not to make predictions about whether an issuer will default. That distinction, between modeling ongoing creditworthiness as ratings already summarize it and modeling the rare event of default itself, runs through the rest of this section and separates work that is comparatively easy to validate from work that is genuinely difficult to validate.

The problem of validation for reliability in ML applications differs from ratings and market signals in the same regard, because a model's output is only as reliable as the process of validation that is followed in order to ensure that it generalizes beyond the data it was trained on. Liu et al. (2025) illustrate the scale of this challenge directly. Applying six machine learning algorithms, with and without two undersampling techniques, to fiscal distress prediction across local governments in 49 U.S. states from 2015 to 2017, they find that an Exactly Balanced Bagging Classifier generally outperforms the other methods, including logistic regression. Its F1-scores range from about 53 to 55 percent, compared with a much wider and less stable range of 27 to 58 percent for logistic regression. Such variation indicates that model performance is sensitive to methodological choices, making reliability dependent not only on the data but also on the modeling and validation procedure used.

Timeliness under delayed disclosure is where ML's dependence on structured input data becomes most exposed. A model trained on financial statement variables inherits whatever lag exists between a fiscal event and its appearance in a filed report, and unlike an analyst, it generally cannot substitute judgment or issuer conversation for information that has not yet arrived. Some of the more recent work in this space is explicitly built around that limitation. Jordan (2021) shifts attention away from periodic financial statements by using text embeddings derived from bond issuance documents, combined with a neural network and synthetic oversampling, to predict default at the moment of issuance itself, without using bond ratings as an input, correctly flagging roughly nine of every ten eventual defaults at a false-positive cost below 0.1 percent among non-defaulting bonds.

The predictive value of ML methods is tightly linked to the severity of the underlying label density (sparseness), which is the main constraint the paradigm needs to address. Although conducted on Italian municipalities rather than U.S. local governments, Antulov-Fantulin et al. (2021) provide a useful methodological analogue for the sparse-label problem. Studying 7,795 municipalities between 2009 and 2016, they describe bankruptcy prediction as an extremely difficult task because only a small number of bankruptcy cases were available for model training. Their approach incorporates institutional and socio-demographic variables alongside financial data and evaluates performance using receiver operating characteristic and precision-recall curves rather than simple accuracy. The study finds that several non-financial features, including geographic region, carry substantial predictive weight, while municipal deficit and debt are among the most informative financial indicators, and the age and gender of council administrators also appear among the ten most important features overall. Their out-of-sample results have a high true-positive rate and low false-positive rate. Despite being based on a small number of positive cases to train the model, this demonstrates how much of the modeling work in this literature is organized around compensating for label scarcity instead of the prediction task itself.

Adapting to sparse defaults in this literature generally takes one of several forms. One approach uses undersampling or oversampling techniques to rebalance training datasets dominated by non-default observations, as demonstrated by Liu et al. (2025) and Jordan (2021). Another broadens the feature set beyond financial ratios to include institutional and demographic variables that may provide predictive information when financial data alone are insufficient, as in Antulov-Fantulin et al. (2021). A third approach avoids default prediction as the primary modeling objective and instead focuses on alternative forms of credit assessment. Saha et al. (2024), working in a large asset manager rather than an academic environment, rely on a similarity framework based on CatBoost instead of predicting default to determine bonds with similar risk profiles for relative pricing. This application is closer to the assessment side of the creditworthiness-versus-prediction distinction than to default prediction, and it is one of the few applications in this literature that has been applied to an application that involved back-testing against rule-based and heuristic benchmarks as a test of a practitioner setting.

Taken as a whole, ML approaches to municipal credit risk show real capacity to compensate for delayed disclosure, whether by using information available earlier in the disclosure timeline or by drawing on data sources that ratings and market pricing do not typically incorporate. They confront the sparse-label problem more directly than either of the other two paradigms, since a model's validity depends on learning a pattern from an outcome that occurs a fraction of a percent of the time.

III. COMPARATIVE SYNTHESIS

- *Comparative Findings Across the Seven Dimensions*

Information basis separates the three paradigms most cleanly. Ratings are based on an analyst's judgment, communication with the issuer, and proprietary criteria that may include information that is not captured in a structured dataset. Market signals are based on observed prices and yields, which are not a clean credit metric but are the result of a combination of credit risk, liquidity, tax treatment, and dealer behavior. Machine learning methods rely on structured financial and statistical data, and increasingly unstructured text, that are ingested without the intervention of a rating analyst's judgement or the price discovery a trading market carries out. When viewed closely, reliability takes a different form when it's not assumed. Rating reliability is also entangled with the economics of the agencies that produce the ratings, as seen in the distortions caused by the fees in Cornaggia et al. (2023) and the incentive effects around the 2010 recalibration. Market reliability is constrained by confounding of credit signal and liquidity, and dealer behavior as reported by Ang, Bhansali and Xing (2014) and Griffin, Hirschey, and Kruger (2023). ML reliability depends

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct786

heavily on modeling and validation choices, as shown directly by the wide performance range Liu et al. (2025) report across algorithms and resampling strategies applied to the same underlying data.

When it comes to timeliness, the same trend of conditional rather than absolute advantage is evident among the three. Ratings change based on continuous monitoring and are still dependent on information being provided to the analyst. Market prices can move continuously, yet Cuny et al. (2025) show that timeliness depends on the type of disclosure and the characteristics of the bond rather than holding as a general property of the market. ML timeliness depends on what the model consumes as input, illustrated by Jordan's (2021) use of issuance-time text rather than periodic financial statements to sidestep the usual reporting lag.

Predictive usefulness cannot be assessed without holding the creditworthiness-versus-default distinction fixed across all three paradigms. All three approaches, ratings, market prices, and ML models, have a greater capacity to track ongoing creditworthiness than to predict the rare event of default. Where ML studies attempt default prediction directly, as in Jordan (2021) and in Liu et al. (2025), they confront sparse labels as their central limitation. Ratings and market signals rarely attempt default prediction as a stated goal in the first place, which makes a like-for-like comparison across all three genuinely difficult.

Assessment purpose, closely related, shows that the three paradigms were not built to answer the same question. Ratings summarize institutional judgment about ongoing creditworthiness. Investor valuation is continually updated by trading and is reflected in market prices. ML models, where they target default directly, are built for prediction rather than summary assessment. To speak of all three as equivalent responses to the same question is to blur the distinction, not to illuminate it.

Constraint adaptation varies with the type of constraint each paradigm offers to adapt. Ratings adapt through analyst judgment and cohort-level validation using transition matrices rather than individual default observations. Market-based approaches evolve through decomposition, separating credit from liquidity and tax components rather than treating the raw spread as one signal. ML approaches adapt through resampling, broader feature sets, and validation metrics built for extreme imbalance. Applicability and maturity are the dimensions where the three paradigms differ most sharply, and that asymmetry is itself a finding rather than an incidental detail. Ratings and market pricing are established institutional practices that are continuously generated by institutions that are part of the daily functioning of the market. ML in this space remains predominantly academic, with Saha et al. (2024) being one of the few examples evaluated in a practitioner setting rather than solely through an academic exercise.

- *The Interaction Between Sparse Defaults and Delayed Disclosure*

The central argument of this review is that these two constraints interact rather than operate independently, and each paradigm manages that interaction differently. When defaults are sparse, the methods tend to require more data and longer histories to compensate for how rarely the outcome of interest occurs. Delayed disclosure reduces the time frame of the information that is both current and reliably valid at the time an assessment is made. Ratings absorb this tension through an institutional process that can wait on information without needing a large, labeled dataset, at the cost of reliability questions tied to agency incentives. Market signals absorb it through continuous price discovery that does not require waiting on periodic reporting, at the cost of confounding credit information with liquidity and frictions unrelated to credit. ML approaches confront the tension most directly, since a model's validity depends on learning a pattern from an outcome occurring a fraction of a percent of the time, using a training window shaped by whatever data happens to be available and current. There is no paradigm that resolves the joint constraint. Each trades one form of vulnerability for another, and the comparison across the seven dimensions above shows precisely where those trade-offs land.

- *Research Gaps*

These are candidate gaps, offered as directions the literature has not yet closed rather than as established findings.

There is an open gap in the synthesis between paradigms. No prior research emerged from the search for this review that synthesized ratings, market signals, and ML under one comparative framework, in the context of municipal finance or otherwise.

A joint-constraint gap follows from the argument that sparse defaults and delayed disclosure interact rather than operate separately, a treatment largely absent from work examining each constraint in isolation.

An assessment-purpose conflation gap persists across all three paradigms, where creditworthiness assessment and default prediction are frequently treated as the same task rather than distinguished explicitly.

A point-in-time validity gap follows from how difficult it remains to construct a dataset that is both current and free of look-ahead bias, given how unevenly municipal disclosure timing runs across issuers.

An applicability and maturity gap arises from the fact that very little of the ML literature has taken the next step beyond academic exercise to evaluate practitioner use, and therefore there is an open question of how these methods would work under the operational constraints of ratings and market pricing as they are used every day.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct786

IV. IMPLICATIONS AND FUTURE RESEARCH

➢ *Implications for Assessment Practice*

None of the paradigms studied in this review can stand alone. Ratings provide institutional durability and the ability to include information that is not available in any dataset but come with reliability concerns related to agency incentives. Market signals offer continuous updating, at the cost of confounding credit information with liquidity and frictions unrelated to credit. ML approaches offer the ability to use earlier or less conventional information than either of the other two, at the cost of validity resting on a label that fires too rarely to support straightforward testing. Those who use any one of these paradigms without the others are taking on the weaknesses of that paradigm, as well as its strengths.

Related work exploring AI-assisted tools for financial auditing, internal controls, and continuous auditing in other public-sector settings suggests this kind of data-driven oversight is an active area of development beyond credit risk assessment specifically, though these applications remain at an early stage and fall outside this review's direct scope (Amoakoh et al., 2026; Dankwah & Yeboah, 2026; Osei, 2026).

The practical implication follows directly from the distinction between assessing ongoing creditworthiness and predicting the rare event of default itself. The approach to assessment should be aligned with the specific question being asked. Ongoing creditworthiness assessment is better served by ratings and market signals, both of which are built for continuous or periodic reassessment rather than rare event prediction. If the question is specifically regarding default risk, the evidence presented here indicates that no matter which paradigm is applied, caution should be exercised because none of the three have been proven to be accurate when it comes to an outcome as sparse as municipal default. Treating a rating, a spread, or a model output as a default probability rather than as a creditworthiness signal risks a level of precision none of the underlying evidence supports.

However, the disclosure environment is not static either, and this has a direct impact on the weight any of these three paradigms can carry currently. With the Financial Data Transparency Act's data standards having taken effect in October 2026 and municipal-specific rulemaking still pending, the information environment underlying all three paradigms is likely to keep shifting before it stabilizes, which argues for treating current assessments of relative paradigm strength as conditional on a regulatory environment still in motion rather than as a fixed comparison.

➢ *Future Research Directions*

Four directions follow most directly from open questions in this literature. Joint-constraint studies would test the joint effect of sparse defaults and delayed disclosure, rather than each separately, and would directly assess whether the combination of the two has an empirical effect. Point-in-time validation would also require the creation of datasets that would reconstruct what was actually known at each point in time, rather than the look-ahead bias that as-filed data suggests. A cross-issuer-type comparison would determine whether the findings obtained from the general obligation and revenue bond sectors are consistent across the different types of issuers, as opposed to being limited to the sectors that are well represented in the existing literature. Practical ML deployment research would extend the small number of practitioner applications identified across this review into settings where model outputs are evaluated against the same operational standards, ratings, and market pricing already faced daily, rather than against academic benchmarks alone.

V. CONCLUSION

This review compared rating-based, market-based, and machine-learning approaches to credit risk assessment in U.S. public finance, where defaults are rare and disclosure can lag behind underlying fiscal conditions. Each paradigm responds differently. Ratings use institutional judgment to incorporate information beyond structured data but raise questions about incentives and reliability. Market prices update continuously but reflect liquidity and dealer behavior alongside credit risk. Machine-learning models can draw on broader information sources but face validation challenges because default events are sparse.

Against a common set of questions, the three paradigms do not answer the same question at all. Ratings summarize ongoing creditworthiness, market prices offer timely but noisy signals, and machine-learning models more often target patterns associated with distress or default directly. The paradigms also differ sharply in institutional maturity, with ratings and market pricing established practice and machine learning still largely confined to research settings.

Existing research has examined these approaches largely in isolation, without synthesizing them around the joint constraints of sparse defaults and delayed disclosure. This review addresses that gap. No single approach is universally superior. Each offers distinct value under different information and validation conditions.

REFERENCES

[1]. Amoakoh, C. K., Nwinyi, I. P., & Ampofo, I. (2026). AI-enabled assurance frameworks: A review of intelligent systems enhancing accuracy and independence in U.S. financial auditing. *EPRA International Journal of Economics, Business and Management Studies*, *13*(2), 154–164. https://doi.org/10.36713/epra26251

[2]. Ang, A., Bhansali, V., & Xing, Y. (2014). The muni bond spread: Credit, liquidity, and tax. *Columbia Business School Research Paper* (14-37).

[3]. Antulov-Fantulin, N., Lagravinese, R., & Resce, G. (2021). Predicting bankruptcy of local government: A machine learning approach. *Journal of Economic Behavior & Organization*, *183*, 681-699.

[4]. Bagley, J., Gissler, S., Hiteshew, K., & Ivanov, I. T. (2023). Pushing Bonds Over the Edge: Investor Demand and Municipal Bond Liquidity

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct786

[5]. Beatty, A., Gillette, J., Petacchi, R., & Weber, J. (2019). Do rating agencies benefit from providing higher ratings? Evidence from the consequences of municipal bond ratings recalibration. *Journal of Accounting Research*, *57*(2), 323-354.

[6]. Bergstresser, D. (2023). The municipal bond market. *Research handbook of financial markets* (pp. 301-330). Edward Elgar Publishing.

[7]. Carroll, D. A., & Ciccarone, R. (2024). Credit rating and geography: Examining the timeliness of municipal bond audits. University of Illinois Chicago Government Finance Research Center & Merritt Research Services.

[8]. Cestau, D., Hollifield, B., Li, D., & Schürhoff, N. (2019). Municipal bond markets. *Annual Review of Financial Economics*, *11*(1), 65-84.

[9]. Cornaggia, J., Cornaggia, K. J., & Israelsen, R. (2023). Rating agency fees: pay to play in public finance?. *The Review of Financial Studies*, *36*(5), 2004-2045.

[10]. Cuny, C., Li, K., Nakhmurina, A., & Watts, E. M. (2025). Muni Disclosure: All talk and no trade?. *Journal of Accounting and Economics*, 101797.

[11]. Dankwah, H. A. K., & Yeboah, M. M. L. (2026). Strengthening internal financial controls in U.S. government agencies through data-driven monitoring systems. *Journal of Economic, Finance Research and Review*, *2*(8), Article 2026-01.

[12]. Edmonds, C. T., Edmonds, J. E., Vermeer, B. Y., & Vermeer, T. E. (2017). Does timeliness of financial information matter in the governmental sector?. *Journal of Accounting and Public Policy*, *36*(2), 163-176.

[13]. Financial Data Transparency Act Joint Data Standards, 91 Fed. Reg. 12787 (June 25, 2026) (to be codified at 17 C.F.R. pt. 200). https://www.federalregister.gov/documents/2026/06/25/2026-12787/financial-data-transparency-act-joint-data-standards

[14]. Gillette, J. R., Samuels, D., & Zhou, F. S. (2020). The effect of credit ratings on disclosure: Evidence from the recalibration of Moody's municipal ratings. *Journal of Accounting Research*, *58*(3), 693-739.

[15]. Government Finance Officers Association. (n.d.). *Financial Data Transparency Act (FDTA)*.

[16]. Governmental Accounting Standards Board. (2012). *Statement No. 68 of the Governmental Accounting Standards Board: Accounting and financial reporting for pensions - an amendment of GASB Statement No. 27*.

[17]. Griffin, J. M., Hirschey, N., & Kruger, S. (2023). Do municipal bond dealers give their customers "fair and reasonable" pricing?. *The Journal of Finance*, *78*(2), 887-934.

[18]. Hájek, P. (2011). Municipal credit rating modelling by neural networks. *Decision support systems*, *51*(1), 108-118.

[19]. Ivanov, I. T., Zimmermann, T., & Heinrich, N. W. (2025). Limits of disclosure regulation in the municipal bond market. *Management Science*, *71*(10), 8452-8470.

[20]. Jordan, L. (2021). Bond default prediction with text embeddings, undersampling and deep learning. *arXiv preprint arXiv:2110.07035*.

[21]. Liu, R., Li, H., Yoon, K., & Vasarhelyi, M. A. (2025). Using machine learning algorithms to improve fiscal distress prediction models: The case of US local governments. *Journal of Information Systems*, *39*(3), 131-155.

[22]. Moody's Investors Service. (2023, July 19). *US municipal bond defaults and recoveries, 1970–2022* [Data report].

[23]. Osei, G. (2026). *Continuous auditing in decentralized public institutions: Analytical perspectives on big data, accountability, and governance*. Zenodo.

[24]. Saha, P., Lyu, J., Desai, D., Chauhan, R., Jeyapaulraj, J., Chu, P., ... & Mehta, D. (2024, November). Machine learning-based relative valuation of municipal bonds. In *Proceedings of the 5th ACM International Conference on AI in Finance* (pp. 634-642).

[25]. Schwert, M. (2017). Municipal bond liquidity and default risk. *The Journal of Finance*, *72*(4), 1683-1722.

[26]. Securities and Exchange Commission. (2026, June 8). *SEC establishes joint data standards as required under the Financial Data Transparency Act of 2022* [Press release].

[27]. Zhang, V. (2024). Municipal bond credit rating access and retail investors' transaction costs. *The Accounting Review*, *99*(1), 427-453.