BANCA D'ITALIA
EUROSISTEMA



Questioni di Economia e Finanza

(Occasional Papers)

Machine learning methods for the prediction of municipal financial distress: evidence from Italy

by Matteo Bergamaschi, Michele Cascarano, Anna Laura Mancini, Francesco Rinaldi and Giacomo Ziglio

November 2026

Number 1078

BANCA D'ITALIA
EUROSISTEMA

Questioni di Economia e Finanza

(Occasional Papers)

Machine learning methods for the prediction of municipal financial distress: evidence from Italy

by Matteo Bergamaschi, Michele Cascarano, Anna Laura Mancini, Francesco Rinaldi and Giacomo Ziglio

Number 1078 – November 2026

The series Occasional Papers *presents studies and documents on issues pertaining to the institutional tasks of Banca d'Italia and the Eurosystem. The* Occasional Papers *appear alongside the* Working Papers *series which are specifically aimed at providing original contributions to economic research.*

The Occasional Papers *include studies conducted within Banca d'Italia, sometimes in cooperation with the Eurosystem or other institutions. The views expressed in the studies are those of the authors and do not involve the responsibility of the institutions to which they belong.*

The series is available online at www.bancaditalia.it .

ISSN 1972-6643 (online)

Designed by the Printing and Publishing Division of Banca d'Italia

MACHINE LEARNING METHODS FOR THE PREDICTION OF MUNICIPAL FINANCIAL DISTRESS: EVIDENCE FROM ITALY

by Matteo Bergamaschi*, Michele Cascarano**, Anna Laura Mancini**, Francesco Rinaldi* and Giacomo Ziglio**

Abstract

This paper develops a data-driven framework for predicting severe financial distress in Italian municipalities. Using a balanced panel of 7,778 municipalities over 2009-23, it makes four contributions. First, it introduces daily treasury-advance data from SIOPE as a high-frequency indicator of liquidity stress. Second, it compares predictive exercises based on comprehensive financial statements, provisional budget forecasts and indicators developed by the Italian Court of Auditors. Third, it evaluates several machine learning (ML) methods against traditional econometric benchmarks. Fourth, it applies interpretability techniques to identify the factors underlying model predictions and support supervisory assessment. The results indicate that advanced ML models provide more accurate predictions than the current regulatory system for the early identification of financial distress. Predictive performance is generally stronger over two- and three-year horizons than for one-year-ahead forecasts. While daily treasury-advance indicators have limited standalone predictive power, they provide timely information on liquidity conditions and may strengthen early-warning systems when combined with budget forecasts or accounting data.

JEL Classification: C45, C53, H72, H74, G33.

Keywords: financial distress, Italian local government, data-driven approach, daily treasury advances, interpretability.

DOI: 10.32057/0.QEF.2026.1078

* Università di Padova.

** Banca d'Italia.

1 Introduction[1]

Municipal financial distress in Italy represents a significant and enduring challenge, disproportionately concentrated in the southern regions of the country. Since the legal framework governing the financial distress of local authorities was introduced, in 1989 for financial default (*dissesto*) and in 2012 for intermediate pre-default (*predissesto* or *riequilibrio finanziario pluriennale*), more than one thousand Italian municipalities have been subject to these procedures. As of 31 December 2024, 1,383 cases had been formally reported. Approximately 79% of these were concentrated in the southern regions and the islands (see Figure 1). This pronounced geographical clustering, coupled with high recurrence rates, indicates that the phenomenon is not merely episodic but symptomatic of deeper structural and systemic deficiencies.

Beginning in the early 2000s, the incidence of municipal financial distress appeared to decline. This was largely due to regulatory changes that relaxed some accounting constraints and allowed local governments to present formally balanced budgets. These improvements were largely formal: underlying structural weaknesses persisted, but accounting practices made them less visible in official budget balances. The reintensification of municipal financial difficulties from 2012 onward revealed these concealed vulnerabilities, a process likely accelerated by the introduction of the guided default (*dissesto guidato*) procedure (Legislative Decree 149/2011) and the intermediate pre-default mechanism (Decree-Law 174/2012, converted into Law 213/2012). These instruments reduced the discretion available to local administrations in delaying formal proceedings. By doing so, they created an institutional pathway through which fiscal imbalances that had previously remained hidden could surface. As a consequence, a share of the distress episodes observed from 2012 onwards reflects the revelation of fiscal tensions that had accumulated for several years, rather than a genuine deterioration in municipal finances.

The economic, administrative, and social costs associated with a declaration of financial distress are considerable. They often result in mandated tax increases, service reductions, and diminished public investment. Given the heavy concentration of the phenomenon in socioeconomically fragile municipalities, the development of robust and timely early detection and prevention mechanisms is essential to protect the continuity of public services and strengthen institutional resilience.

This paper contributes to the literature on fiscal surveillance in several ways. First, to the best of our knowledge, this is the first study to leverage high-frequency administrative data from the SIOPE system (*Sistema Informativo delle Operazioni degli Enti Pubblici*) on daily treasury advances (*anticipazioni di tesoreria*) as a dynamic indicator of prospective imbalances in municipal cash flows. Second, we conduct two complementary predictive exercises to assess whether liquidity-based indicators remain informative under different data-availability and comparability constraints. We develop and compare models using (i) full balance-sheet

[1]We thank G. Ivaldi and M. Francese for their useful comments, and S. Ranieri and A. Sparacino for their fruitful suggestions, patience, and for providing data on daily treasury advances from SIOPE. We also thank two anonymous referees for their careful reading and constructive suggestions.
The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of Banca d'Italia or the Eurosystem. All remaining errors are our own.

features from the OpenBDAP public finance data portal (*Banca Dati delle Amministrazioni Pubbliche*) and (ii) summary indicators developed by the Italian Court of Auditors (*Corte dei conti*). Where directly comparable configurations are available, we evaluate performance with and without the high-frequency SIOPE series; we also examine earlier-available combinations based on ReadyBDAP and SIOPE. This dual design allows a direct assessment of the consistency and incremental value of the liquidity measure. Third, the paper performs a systematic comparison of advanced predictive architectures, including Random Forest (RF), Extreme Gradient Boosting (XGBoost), Convolutional Neural Networks (CNNs), and Graph Neural Networks (GNNs). To our knowledge, this is the first application of GNNs to this specific municipal-distress setting. The model is used to explore whether geographically proximate municipalities display interdependent fiscal vulnerabilities. Finally, the study addresses model opacity, often raised in connection with machine learning applications, by integrating interpretability techniques that make model decisions transparent and actionable. These ensure that forecasts can effectively inform regulatory oversight and targeted intervention strategies, bridging methodological innovation with the practical needs of fiscal governance.

Our results indicate that several machine learning models substantially outperform traditional statistical approaches and the current rule-based regulatory benchmark in the experimental settings considered. The best-performing specifications, in particular Random Forest and selected neural-network configurations, reach median ROC AUC values around or above 0.87. When evaluated at threshold-optimised operating points, the leading models deliver F1-scores of 0.27-0.30, compared with 0.13 for the regulatory benchmark. Performance varies significantly across design configurations: predictions targeting two- and three-year horizons outperform one-year-ahead forecasts, while historical data beyond five years yield diminishing returns. A meaningful trade-off emerges between data richness and timeliness. Comprehensive annual balance sheets provide the strongest predictive accuracy. Budget forecasts and daily treasury data are less informative on their own, but they are available earlier and can support timelier screening, especially when used together.

To address concerns about the "black-box" nature of machine learning models, we complement our performance analysis with interpretability assessments using both model-native (Mean Decrease Impurity – MDI) and model-agnostic (Local Interpretable Model-Agnostic Explanations – LIME) techniques. These methods reveal that model predictions are grounded in economically meaningful fiscal indicators: in the pre-harmonisation period (2009–2015), models prioritize structural indebtedness and debt-service capacity as primary distress signals, while in the post-reform period they shift focus toward short-term liquidity indicators, particularly reliance on treasury advances and cash-flow management. The convergence of different interpretability methods on similar financial variables, together with their alignment with the Italian municipal accounting framework, suggests that advanced models can combine predictive accuracy with economically meaningful explanations. This would be an important precondition for their adoption in policy and supervisory contexts.

The remainder of the paper is structured as follows. Section 2 outlines the institutional framework governing financial distress in Italian municipalities and motivates the need for improved predictive approaches. Section 3 reviews the relevant literature. Section 4 describes

the data sources and sample construction as well as some descriptive statistics. Section 5 details the modeling strategy. Section 6 discusses the results, robustness checks, and the role of false positives. Section 7 discusses the interpretation of the results obtained from deep learning methods. Section 8 concludes by discussing policy implications and directions for future research.

Italian institutional terms and variable labels used throughout the paper are defined in Appendix D.

2 Institutional framework and motivation

In the Italian institutional context, municipal financial distress refers to the inability of a local government to meet its financial obligations and ensure uninterrupted provision of essential public services. The legal framework governing such conditions is articulated in the *Testo Unico delle leggi sull'ordinamento degli Enti Locali* (Legislative Decree No. 267/2000 – TUEL). The TUEL establishes a tiered system of intervention based on the severity of fiscal imbalance, comprising: structural deficit monitoring (*deficit strutturale*), intermediate pre-default procedures, and financial default, each regulated by specific provisions.

The system of fiscal stress indicators (*parametri obiettivo*) designed to detect structural vulnerabilities was overhauled by the Ministerial Decree of 28 December 2018 (subsequently updated by the Ministerial Decree of 4 August 2023) that replaced the model in place since 2009[2]. According to Ministero dell'Interno and Ministero dell'Economia e delle Finanze (2018), the reform sought to enhance diagnostic capabilities, align the system with harmonised public accounting principles (Legislative Decree No. 118/2011), and streamline monitoring obligations. The updated framework introduced eight indicators (seven synthetic and one analytical) evaluating budget rigidity, revenue performance, debt sustainability, and the presence of off-balance-sheet liabilities. Failure to comply with the identified standards leads to the municipality being classified as structurally deficient (*ente strutturalmente deficitario*) under the applicable public finance regulations and may entail financial sanctions.

The intermediate pre-default procedure, introduced in 2012, allows municipalities facing structural imbalances to avoid default by adopting a multi-year financial rebalancing plan lasting between four and twenty years, approved by the regional section of the Court of Auditors and the Ministry of the Interior. In cases of full financial default, an Extraordinary Liquidation Body (*Organo Straordinario di Liquidazione* – OSL) inherits and manages the liquidation of outstanding debts, while the local government retains ordinary administration functions.

[2]Prior to the 2018 reform, the structural deficit of Italian municipalities was assessed using a framework established by the Ministerial Decree of February 18, 2013. This system relied on ten objective parameters primarily focused on budgetary rigidity, cash flow strains, and revenue collection capacity (e.g., personnel cost thresholds, unreturned cash advances, and high levels of residual liabilities). A municipality was formally classified as structurally deficient if it failed at least five of these ten indicators. Importantly, although the 2018 reform falls within the 2016–2023 sub-period used in the empirical analysis, the resulting shift in the classification methodology does not affect the definition of our dependent variable, as the objective parameters are not used to identify the severe financial distress events considered in our analysis. Consequently, the reform does not affect the main empirical findings or the robustness of the results presented in this paper.

Over the past two decades, the support system for financially distressed municipalities has transitioned from unconditional bailouts to conditional performance-based assistance. Instruments such as liquidity advances, extraordinary contributions and simplified debt restructuring have been supplemented by regional support schemes and bilateral agreements with metropolitan cities. These arrangements combine financial transfers with binding commitments to fiscal and administrative reform.

2.1 Limitations of the current framework

Despite these advances, the current system continues to exhibit significant limitations, as reported by Corte dei conti (2025b). Although it includes a mechanism for identifying structural deficits based on predefined financial and administrative parameters, empirical evidence suggests that this monitoring approach has been largely ineffective in anticipating fiscal crises. Specifically, the indicators are applied ex post and rely on financial-statement information available with a substantial delay, referring to year $t-2$ relative to the reference period t. This delays the activation of corrective measures, thus reducing their effectiveness. The framework also lacks formal procedures to assess the financial impact on the municipality accounts of affiliated entities and municipal companies, which frequently contribute substantially to fiscal distress.

Additionally, the framework applies uniform thresholds across all municipalities, without accounting for differences in size, geographic location, or socio-economic conditions. All indicators are equally weighted, and the 50% activation threshold, which flags a municipality whenever at least half of the indicators exceed their respective thresholds, lacks empirical or theoretical justification. The corrective actions triggered are standardised and do not consider the diverse origins of fiscal imbalances, whether structural, liquidity-related, or governance-driven.

A confusion matrix for the period 2010–2023 (Table 1) reveals the extent of these weaknesses. Although overall accuracy exceeds 98%, this figure is misleading due to the pronounced class imbalance. The current regulatory framework exhibits low recall (10.7%) and precision (17.5%), indicating that it fails to identify most entities that eventually enter the intermediate pre-default procedure or financial default, while also generating a non-negligible number of false positives. The prevalence of false negatives is particularly concerning from a policy standpoint, as it implies that many at-risk municipalities are not flagged in time to implement preventive measures.

2.2 Recent statistics (2023–2024)

During the 2023–2024 period, the Italian local government system experienced a decrease in the overall number of financial distress procedures, which fell from 75 to 64 (see Corte dei conti (2025a) for further details). However, there was a notable increase in municipal insolvency cases, which reached 34 in a single year, the highest number recorded since 2020. In the same timeframe, 30 multi-year financial recovery (i.e., pre-default) procedures were initiated. The territorial distribution of these crises is highly concentrated in Southern Italy,

with Calabria, Campania, and Sicily accounting for over 70% of all cases. This highlights persistent structural fragility in these regions. As of December 31, 2024, there were 487 active procedures (227 insolvencies and 260 financial recoveries), involving 485 municipalities, with a clear predominance in the South. The average duration of insolvency procedures ranged from six to eight years, with only 18% resolved within the five-year period stipulated by law[3]. Financial recovery procedures are also characterised by lengthy administrative processes and a low success rate: only 12.7% are concluded positively, while more than one-third ultimately result in insolvency. Municipalities subject to formal financial-crisis procedures represent about 6% of all Italian municipalities. Nevertheless, the territorial concentration and persistence of these crises, especially in large urban centres and small municipalities in the South, make this phenomenon particularly significant for the sustainability of Italy's multi-level governance system.

3 Related literature

The prediction of financial distress in municipalities has long relied on econometric models and early-warning systems grounded in annual financial ratios, demographic indicators, and institutional variables.

Gregori and Marattin (2019) investigate the determinants of fiscal distress in Italian municipalities using a panel dataset covering 2000–2012. They estimate logit models with fixed effects, time dummies, macroeconomic controls, and lagged variables. Budget indicators, including loan repayments, current revenues, current spending, residuals, and personnel costs, are found to be relevant predictors, with debt repayment obligations playing a central role.

Similarly, Gorina (2018) develops a behaviour-based measure of local fiscal distress using information from annual financial reports to identify governments undertaking observable signals of distress such as layoffs, deferral of payments, or emergency budget cuts. Using a panel of nearly 300 U.S. cities and counties (2007–2012), the author finds that lower general fund balances, higher debt burdens, and greater reliance on volatile non-property-tax revenue significantly increase the likelihood of fiscal distress.

Padovani et al. (2024) examine the drivers of financial distress among Italian municipalities. The study concludes that a substantial share of the financial distress observed across numerous Italian municipalities can be attributed to the mismatch between allocated financial resources and the actual needs of the local population and territory. Both inadequate resources and past transfer cuts are found to significantly worsen municipal financial conditions, while municipal efficiency meaningfully reduces fiscal vulnerability, with this protective effect being consistent across different geographical areas and municipal sizes.

Recent research has increasingly turned to machine learning techniques to improve predictive accuracy.

Alaminos et al. (2018) develop a data mining-based model to predict municipal financial distress using a sample of 128 Spanish municipalities. The authors test Decision Trees,

[3]When extending the observation window to six years, thereby allowing for potential bureaucratic or administrative delays, the proportion increases to 20.3%.

Naive Bayes, Multi-Layer Perceptron (MLP), and Deep Belief Networks. The MLP classifier achieves the highest predictive accuracy, with the most influential predictors including short-term debt to own revenue, operating expenses to revenue, and political fragmentation.

A key contribution in this area is the work of Piermarini et al. (2023), which integrates domain expertise with machine learning models (SVM, Random Forest, and XGBoost) alongside a Logistic Regression baseline to improve predictive performance. Logistic Regression provides the best trade-off between interpretability and predictive performance, achieving perfect sensitivity on the test set and thus avoiding missed cases of financial distress. Although it generates several false positives, 61.5% of these subsequently experience distress within four years, indicating strong early-warning capability. Building on this foundation, the Italian Court of Auditors has recently developed MoDì (*Monitoraggio selettivo del Distress finanziario dei Comuni italiani*). This data-driven analytical framework is designed for the early detection of financial vulnerabilities in local governments, leveraging a supervised machine learning approach (for further details see Corte dei conti (2025b) and Corte dei conti (2026)).

Caravaggio et al. (2024) examine the limitations of the current Italian system and, using a rich dataset of financial, socioeconomic, demographic, and institutional variables for over 7,800 municipalities, show that Random Forest models significantly outperform traditional approaches in predicting crises between 2019 and 2024. The most relevant predictors concern the management of commercial debt, revenue collection, and structural socioeconomic factors.

Bianchi et al. (2024) apply binary classification models to financial indicators from OpenBDAP and socio-demographic variables from ISTAT (2016–2023). They find that parsimonious specifications, particularly Logistic Regression with LASSO variable selection, deliver the strongest predictive performance.

Deep learning techniques have also gained attention for their ability to process large datasets and uncover hidden patterns. Barboza et al. (2017) show that tree-based algorithms and neural networks can capture complex, non-linear relationships in corporate bankruptcy prediction. Tölö (2020) applies Recurrent Neural Networks (RNNs) to predict systemic financial crises, demonstrating the effectiveness of deep learning in capturing dynamic dependencies over time. Similarly, Holopainen and Sarlin (2017) compare various early-warning models and explore ensemble methods.

While most existing studies treat municipalities as independent entities, local governments are highly interconnected through economic, fiscal, and geographical linkages. Traditional models fail to capture these dependencies, potentially limiting their predictive power. Recent advances in graph-based machine learning offer a promising way to model these dependencies. To our knowledge, this paper provides the first application of Graph Neural Networks to municipal financial-distress prediction.

A critical challenge in the adoption of machine learning tools for financial distress prediction is interpretability. While advanced models often achieve high predictive accuracy, their complexity can make it difficult for policymakers to extract actionable insights. This study addresses that challenge by combining model-native and model-agnostic interpretability techniques, which can enhance the practical utility of machine learning for supervisory

purposes.

4 Data sources and sample construction

The unit of analysis is the municipality-year, observed over a balanced panel spanning 2009–2023.

Municipalities involved in mergers during the observation period are excluded to maintain panel balance and avoid discontinuities in financial reporting. Among the 441 excluded municipalities, only 3 (0.7%) were formally registered as being in intermediate pre-default or financial default in at least one year of the observation window, against 8.3% (643 out of 7,778) of municipalities retained in the panel. The substantially lower distress incidence among excluded municipalities confirms that the exclusion is driven primarily by administrative factors unrelated to fiscal stress and that the resulting sample does not systematically underrepresent municipalities at risk of distress. If anything, removing these observations slightly reduces class imbalance, because the excluded municipalities display a lower incidence of distress than the municipalities retained in the panel. The effect is nevertheless quantitatively negligible.

A small number of municipalities experienced more than one distress procedure over the observation period, either by re-entering pre-default after a previous episode or by transitioning from pre-default to formal default. We identify 175 such cases. In the supervised classification framework, both pre-default and default are coded as positive outcomes ($y = 1$) at the municipality-year level; repeated episodes therefore contribute multiple positive observations to the training set, consistent with the goal of capturing any manifestation of severe fiscal stress regardless of the specific legal classification involved.[4]

A further sample construction choice concerns the treatment of Italy's Special Statute Regions (SSRs). Municipalities in SSRs operate under distinctive institutional, fiscal, and regulatory frameworks that diverge markedly from those governing Ordinary Statute Regions (OSRs), potentially compromising the comparability of observations and the internal validity of distress predictions based on national accounting standards. At the same time, the autonomous Sicilian Region applies national legislation for local public finance. To assess the sensitivity of results to this boundary, we evaluated three alternative sample definitions: (i) all regions, including both SSRs and OSRs; (ii) OSRs only; and (iii) OSRs plus Sicily. Given the robustness of predictive performance across all three definitions, and to maximise the informational content of the training data – including the substantial volume of distress episodes concentrated in Sicily – the analysis proceeds on the full national dataset.

The balanced panel structure described above does not require that all municipality-years contain complete balance-sheet information: a municipality is retained in the panel as long as it is observed in every year of the window, regardless of whether its financial statements were submitted to BDAP in each of those years. Municipalities in formal distress frequently fail to submit financial statements for one or more consecutive years pending approval of

[4] We verify robustness by re-estimating the main models on a subsample that excludes municipality-years following a first distress declaration; results are qualitatively unchanged.

the *bilancio stabilmente riequilibrato*, a well-documented characteristic of BDAP coverage. Among the 643 municipalities formally registered as being in financial distress at any point over the observation period, 157 (24.4%) failed to submit a balance sheet in at least one year during their distress spell, accounting for 308 municipality-years with missing financial data out of 9,645 municipality-years belonging to municipalities that experienced distress at some point during the sample (3.2%). The corresponding figures for non-distressed municipalities are markedly lower: 732 entities and 930 municipality-years out of 107,025 (0.9%).

A closer inspection reveals that this differential is driven by the distress episode itself rather than by structural reporting weaknesses of the municipalities involved. The share of municipality-years with complete balance-sheet data is 99.1% for municipalities that never experienced a distress event, and 97.7% for municipalities that experienced at least one distress event when observed in years with no active distress procedure. These figures are nearly identical. The share falls to 92.7% only for municipality-years falling within an active distress procedure. Three implications follow. First, non-reporting is a consequence of distress status rather than a pre-existing characteristic, and is itself an informative signal that a well-designed classifier should retain rather than discard. Second, since missing balance sheets are concentrated in years following the activation of a formal distress procedure, after the outcome variable has already taken the value of one, non-reporting does not constitute a relevant omitted predictor: the information loss occurs predominantly where the distress event has already materialised, not in the pre-distress window where early-warning signals are most needed. Third, excluding these observations would introduce survivorship bias by conditioning the sample on the very outcome we aim to predict.

We accordingly retain all municipality-years and set missing balance-sheet values to zero in the baseline specification. We acknowledge that this imputation is not innocuous: replacing informative non-reporting with zeros may alter the information conveyed by missingness by making fiscally stressed municipalities superficially resemble financially sound ones. The limited prevalence of missingness documented above, however, bounds the severity of this attenuation: with over 92% of distress-episode observations remaining complete, the zeroed values affect a small share of the training signal rather than its core. We assess the robustness of our results by re-estimating the baseline models after replacing missing values with the last available balance sheet, thereby preserving the most recent observable fiscal profile of each municipality[5]. Results are qualitatively unchanged across both specifications, confirming that zero-imputation does not materially distort the overall predictive exercise. Performance metrics reported in Section 6 should therefore be interpreted in light of the data available to the monitoring system at each point in time. This mirrors operational surveillance conditions, where risk flags can only be generated from observable and timely submitted information.

The final dataset consists of a balanced panel of 7,778 municipalities observed over 2009–2023.

The temporal structure of the analysis is shaped by the harmonisation of public accounting standards introduced by Legislative Decree 118/2011, which became mandatory

[5]This last-observation approach is conceptually related to the data-construction strategy adopted by Bianchi et al. (2024).

for all local authorities from 2016 onwards. Because pre- and post-harmonisation financial documents are only partially comparable, the analysis is organised around two distinct sub-periods: 2009–2015 (pre-harmonisation) and 2016–2023 (post-harmonisation).

A further dimension of incomparability between the two sub-periods concerns the treatment of treasury advances within budget titles. In the pre-harmonisation framework, advances from the municipal treasury account were classified under the same budget titles as medium- and long-term loans (revenue title $E2_{P1}$ and expenditure title $S3_{P1}$), making it impossible to isolate them without manual intervention in the raw data. The harmonised framework introduces a dedicated classification for treasury advances, enabling their separate observation. This structural difference has two implications for the analysis. First, it reinforces the methodological rationale for treating the two periods separately rather than pooling them. Second, and more fundamentally, it highlights a key advantage of the high-frequency SIOPE data: the daily treasury advance records allow the model to identify liquidity stress independently of the budget classification in use, which is precisely the kind of information that can be recovered from transaction-level data but not from annual BDAP aggregates alone. The shift in importance from debt-title indicators (Period 1) to explicit advance indicators (Period 2) thus reflects not only a change in fiscal behaviour but also a gain in data resolution, and should be interpreted accordingly (see also section 7).

4.1 Outcome variable

The target variable draws on legally validated information from the **Ca' Foscari** database, which tracks all episodes of municipal financial default and intermediate pre-default since 1989. The legal codification of these events ensures cross-sectional consistency and temporal comparability, providing a credible and reproducible ground truth for supervised classification. Their principal limitations – rarity, delayed manifestation relative to underlying fiscal deterioration, and susceptibility to political or legal triggers – are well documented in the literature. We account for these limitations in the modeling strategy discussed in Section 5.

4.2 Predictive features

The feature space is constructed from three complementary layers of information, each capturing a distinct dimension of municipal fiscal behaviour.

OpenBDAP. The first layer draws on the digital platform maintained by the *Ragioneria Generale dello Stato* (RGS), which centralises the accounting records and financial statements submitted by sub-national public entities. Two datasets are constructed from this source:

- *FullBDAP* contains the complete set of first-level budgetary titles derived from management financial statements (*rendiconti della gestione*), fully disaggregated by stage of the financial management cycle. On the revenue side, it includes *accertamenti* (the formal, legally grounded recognition of a collectable claim) and *riscossioni* (representing actual cash collections); on the expenditure side, it comprises *impegni* (legally binding commitments that generate public liabilities) and *pagamenti* (the corresponding cash

disbursements). This dataset captures the entire progression from the legal recognition of receivables and obligations through to their ultimate cash settlement, enabling the construction of both accrual- and cash-based fiscal indicators;

- *ReadyBDAP* focuses exclusively on first-level revenue recognitions and expenditure commitments as reported in budget forecasts (*bilanci di previsione*), whose pre-outturn availability makes them particularly valuable for constructing predictors that can be operationalised prior to fiscal year closure[6].

From these sources, we derive annual aggregates and ratios describing recognition-to-collection performance, commitment-to-payment execution rates, the dynamics of budgetary residuals, and measures of administrative capacity. The harmonised post-2016 framework additionally enables the computation of the **ten financial risk indicators**[7] (hereafter named *Indicators*) proposed by the Corte dei conti (2021) and Padovani et al. (2024).

SIOPE. The second layer draws on the real-time cash-monitoring system for Italian local public administrations, operated within the institutional framework involving RGS and Banca d'Italia; it relies on technical standards defined with the involvement of the *Agenzia per l'Italia Digitale* (AgID). SIOPE records cash receipts and payments at the individual entity level. We use the transaction records that identify treasury advances, i.e. temporary credit lines provided by the treasurer to address cash shortfalls. Although permitted under current legislation, repeated or persistent recourse to such instruments is widely interpreted by Corte dei conti (2024) and Corte dei conti (2025a) as symptomatic of deeper fiscal difficulties. The daily granularity of SIOPE allows us to construct high-resolution indicators of liquidity stress. These include the average duration of advance spells, the average and total amounts of advances per capita, the maximum outstanding advance exposure per capita, and the utilisation rate relative to the statutory ceiling. As a result, SIOPE provides a window into municipal cash dynamics that cannot be obtained from annual accounting statements alone. Given the novelty of the dataset, Appendix A provides descriptive statistics on the dynamics of per capita treasury advances, disaggregated by period, geographical area, region, and municipal size class.

The informational advantage of daily SIOPE data over annual budget documents operates along two dimensions. First, the budget appropriation (*bilancio di previsione*) is prepared months before the reference fiscal year and reflects planning assumptions rather than realised cash dynamics; it cannot, by construction, capture within-year deterioration in liquidity conditions. SIOPE records, by contrast, are updated in real time and can reveal the trajectory of municipal cash balances throughout the year. Second, the granularity of daily data allows the construction of indicators that capture not only the total annual amount of advances but also their duration, cyclicality, and frequency of activation – dimensions that are invisible in annual aggregates but that supervisory bodies (notably the Italian Court of Auditors)

[6]For the ReadyBDAP dataset, the terms *impegni* and *accertamenti* refer to forecasted expenditure commitments and expected revenue recognitions as reported in the budget forecast, which is prepared prior to the reference fiscal year. These are provisional budget estimates and should not be confused with the legally binding *impegni* and formally recognised *accertamenti* reported in the management financial statement.

[7]The ten indicators constitute the financial distress scoring system for individual institutional bodies, named *Rating Finanziario* developed on the basis of the AIDA dataset.

have identified as the key diagnostic criteria distinguishing structural reliance on advances from normal seasonal treasury management. Annual indicators derived from management accounts, being available only after fiscal year closure (and subject to further administrative delays before appearing in BDAP), are too late for genuinely preventive action; the daily SIOPE flow provides a real-time window into the same underlying phenomenon.

ISTAT. The third layer incorporates demographic and geographic variables from ISTAT, including population size, and macro-area location, which enter the model both as controls and as potential structural determinants of cross-municipal variation in fiscal behaviour.

4.3 Dataset structure

Following the harmonisation of entity identifiers across sources and the resolution of municipal mergers and boundary changes, the final dataset comprises 7,778 municipalities and 116,670 municipality-year observations, of which 897 correspond to financial default or intermediate pre-default events. The resulting class imbalance, a defining feature of the prediction problem, is addressed through the sampling and weighting strategies described in Section 5.

5 Modeling

5.1 Problem formulation

The primary objective is to leverage the temporal structure of financial data to forecast distress events. We frame this as a supervised sequence modeling task: given a multivariate time series of features X for a given municipality and a binary distress indicator y (pre-default[8] or default), a single prediction instance is defined as:

$$X_{t_0:t_n} \rightarrow y_{t_{n+1}:t_{n+k}}.$$

where t_0 denotes the first year of the input window and t_n its final year. The input-window length, given by the number of years between t_0 and t_n, varies across experimental configurations as described in Section 5.2. This formulation emphasizes the temporal dimension of the problem. We use consecutive years of input data (X from t_0 to t_n) to predict financial distress occurrences over a subsequent consecutive window of k years (y from t_{n+1} to t_{n+k}), reflecting the gradual nature of municipal fiscal deterioration, thus allowing the model to capture the accumulation of distress signals over time rather than treating each year as an independent observation.

[8]It is worth noting that the legislative reconfigurations enacted since the 2018 Budget Law (Law no. 205/2017) targeting the intermediate pre-default procedure have gradually made this path more structurally attractive for municipalities (notably by extending the recovery periods up to 20 years) compared to an immediate declaration of default. However, since our target variable aggregates both pre-default and default under a single category representing severe financial distress, any strategic or procedural shift of a municipality from one framework to the other over time does not alter its underlying crisis status. This measurement strategy ensures that the classification of the dependent variable remains stable and consistent throughout the entire observation period, confirming that the empirical findings are unaffected by these shifting institutional incentives.

5.2 Data Construction and Inputs

Experiments are conducted separately for each of the two accounting periods, across different combinations of feature sets, data depths, and prediction horizons, which constitute the three key dimensions of the experimental design.

For **Period 1** (2009-2015), the four feature configurations evaluated are:

- *FullBDAP*
- *FullBDAP + Treasury advances*
- *Treasury advances*
- *ReadyBDAP + Treasury advances*

For **Period 2** (2016-2023), seven configurations were considered: the first four replicate those of Period 1, while the remaining three incorporate the Indicators dataset, available exclusively for this period.

- *FullBDAP*
- *FullBDAP + Treasury advances*
- *Treasury advances*
- *ReadyBDAP + Treasury advances*
- *Indicators*
- *Indicators + Treasury advances*
- *FullBDAP + Indicators + Treasury advances*

In each case, *data depth* is varied from 4 to 6 years and *target depth* from 1 to 3 years, yielding a total of 99 experimental settings[9], whose systematic analysis is reported in Section 6.

Structural heterogeneity across municipalities is addressed through multiple channels. Demographic and geographic controls from ISTAT (population, density, macro-area dummies) enter the feature set directly as predictors, allowing non-linear models to weight them differently depending on the fiscal context. The municipality-level panel structure further enables the models to learn differential patterns associated with size class and regional location without requiring explicit interaction terms. The potential influence of aggregate macro-shocks on Treasury advance dynamics, including fiscal consolidation (2011-2015), COVID-19 emergency transfers (2020-2021), and PNRR-related liquidity (2022-2024), is documented in Appendix A. To the extent that these shocks affect all municipalities simultaneously, their effect is reflected in the period-specific structure of the training data; municipality-specific responses are captured by cross-sectional variation in advance patterns documented in Tables 20 and 21.

[9] 4 datasets × 3 data depths × 3 target depths in Period 1 and 7 datasets × 3 data depths × 3 target depths in Period 2

5.3 Predictive Models

This section describes the six predictive architectures compared in the analysis and the main design choices adopted for each. Hyperparameters were selected using the training/validation portion of the sample, while the held-out test period was reserved for final evaluation. This separation limits the risk that reported test performance reflects tuning on the evaluation set. For a general explanation of how each model operates, we refer the reader to Appendix B. An exception is made for Graph Neural Networks, for which the Appendix provides a more detailed discussion of the methodology, given the novelty of their application in this context.

- **Logistic Regression** (see, e.g., Hosmer Jr et al. (2013); Bishop (2006)): We employ Logistic Regression with ℓ_2 regularization and the `sag` solver, chosen for its efficiency on large-scale datasets.
- **Classification Trees** (see, e.g., Breiman et al. (2017); Hastie et al. (2009)): Decision trees are trained without an explicit maximum depth constraint, using the Gini impurity as the splitting criterion and selecting the best split at each node.
- **Random Forest** (see, e.g., Breiman (2001)): The model consists of an ensemble of 500 decision trees, each trained using the Gini impurity and the best split strategy.
- **XGBoost** (see, e.g., Chen and Guestrin (2016)): We adopt a gradient boosting model with 100 estimators, a learning rate of 0.1, and a maximum tree depth of 5.
- **Convolutional Neural Networks (CNNs)** (see, e.g., Goodfellow et al. (2016)): The architecture comprises two stacked `Conv1D` layers followed by three fully connected layers. The model is trained using the Adam optimizer and binary cross-entropy loss, with dropout regularization to mitigate overfitting. This architecture directly exploits the three-dimensional structure of the data without requiring flattening.
- **Graph Neural Networks (GNNs)** (see, e.g., Gori et al. (2005); Bruna et al. (2014)): The model adopts a structure analogous to the CNN architecture, adding `Graph Convolutional` (GCN) layers. Due to the graph-based data representation, both the architecture and training methodology differ substantially from standard neural networks. These methodological aspects are discussed in greater detail in Appendix C.3.

5.4 Setting the prediction tasks

The classification workflow proceeds through four steps:

1. **Input Creation:** We construct the dataset using the feature sets described previously. Both the temporal depth of the input data and the length of the target window can be varied. Whenever the available data range exceeds the desired input depth, we augment the training set by including temporally shifted data from the same municipalities, each treated as distinct samples. This strategy increases the effective size of the training set and improves the robustness of the training process.

2. **Data Splitting:** The dataset is divided into training and test sets using an 80/20 ratio, while preserving the temporal ordering of the observations. Specifically, models are trained on earlier years and evaluated on later held-out years. Stratification is employed, where feasible, to preserve the proportion of distress events across the two partitions without violating the temporal split.

3. **Class Balancing:** Given the extreme rarity of financial distress, we compute and apply class weights during training. This increases the weight assigned to the minority class during model estimation and reduces the incentive to converge to a trivial "always-safe" classifier.

4. **Training and Evaluation:** Models are trained on the training set using the parameters described in the previous section and are then evaluated on a held-out test set. We consider robust threshold-independent metrics, such as ROC AUC and PR AUC, which are discussed in more detail in Section 6. In addition, we report standard metrics for imbalanced classification, including precision, recall, and the F1-score, enabling comparison with the Regulatory Indicators methodology currently in use.

Several of the fiscal series employed as predictors may exhibit persistence, low-frequency trends, or cyclical components tied to electoral or budget cycles. We do not conduct a formal battery of unit-root or autocorrelation tests; however, the modelling framework incorporates features that partially address these concerns. First, the input window design, feeding the models a multi-year sequence of levels and ratios, allows tree-based and neural network architectures to implicitly condition on recent trends, without requiring explicit pre-differencing. Second, the temporal train/test split (rather than a random split) ensures that the models are evaluated on genuinely out-of-sample future periods, which disciplines against overfitting to idiosyncratic cyclical patterns in the training window. Third, fiscal ratios normalised by current revenues or population are less susceptible to common trending than raw budget totals. A formal detrending exercise and stationarity analysis remain useful extensions of the present framework, particularly with a view to real-time implementation.

6 Results and their interpretation

In this section, we present the results of our extensive testing across various settings.

6.1 Detailed comparison of predictive models

To evaluate model performance in predicting financial distress, we use two complementary discrimination metrics: ROC AUC and PR AUC. ROC AUC provides a conventional benchmark for class separation, while PR AUC is especially informative under severe class imbalance, focusing specifically on the accuracy of identifying the minority (distressed) class; reporting both metrics thus offers a balanced assessment of predictive performance.

Median values (p50) are reported across all choices of data indicators, target depth, and data depth to mitigate the influence of outliers; clear differences emerge across models and

periods, as shown in Tables 4 and 5. In Period 1 (Table 4, Panel A), Random Forest achieves the highest median ROC AUC (0.873), followed by XGBoost (0.849) and Logistic Regression (0.826), while CNN and GNN perform moderately and Decision Tree ranks lowest. However, this ranking masks a critical trade-off between peak performance and reliability, revealed by the distributional spread: Random Forest exhibits a standard deviation of 0.255, nearly five times that of Logistic Regression (0.055). Its minimum value drops to 0.217, implying that while Random Forest has the highest potential accuracy, it is prone to severe performance deterioration under suboptimal configurations. Logistic Regression, with a tight distribution (min 0.693, max 0.889), serves as a stable and reliable baseline.

Period 2 (Panel B) confirms this pattern: Random Forest again leads (ROC AUC 0.895), followed by Logistic Regression (0.866), and the stability hierarchy is preserved with Logistic Regression displaying substantially lower variability than Random Forest (SD 0.046 vs. 0.230). The consistency of these rankings across both accounting regimes strengthens the conclusion that the choice between models involves a fundamental trade-off between peak discriminative power and reliability across specifications.

Median PR AUC values provide complementary insight (Table 5). In Period 1, XGBoost performs best (0.129), followed by Random Forest (0.108). CNN and Decision Tree display higher variability, resulting in lower medians despite occasionally high upper-tail values. Logistic Regression records one of the lowest PR AUC medians (0.079), reflecting limited sensitivity to rare positive cases. This highlights a limitation of standard linear models in this context: they provide a stable baseline but are less effective than non-linear alternatives in isolating the minority-class signal. In Period 2, Random Forest achieves the strongest PR AUC median (0.099), while XGBoost (0.097) and CNN (0.096) follow closely, suggesting that ensemble and non-linear methods retain their ability to capture the minority class signal even in a different temporal context. Logistic Regression and Decision Tree display moderate central performance (0.091), with the former once again confirming its limited capacity to discriminate rare positive cases under a linear boundary. GNN, despite showing a wide interquartile range and notable upper-tail values, records the lowest median (0.051), indicating high variability and inconsistent precision-recall performance across windows.

Overall, the median results suggest that Random Forest achieves the strongest median performance in several rich-data configurations, although other models, such as GNN, outperform it in specific PR AUC evaluations. Nevertheless, the broad dispersion of results for Random Forest highlights that its "superiority" is highly conditional on careful tuning.

6.2 Stability

The scatter plots in Figure 2 provide an additional validation of the preceding results, while also revealing differences in robustness across models. The joint distribution of ROC AUC and PR AUC scores confirms that Random Forest generally achieves the highest overall performance, clustering in the upper right region across both periods. However, the relatively broad ellipses surrounding its point clouds indicate that its performance is more sensitive to the choice of input indicators, target definitions, and data depth. This corroborates the wide interquartile ranges and high standard deviations observed in Table 4, suggesting that

Random Forest benefits substantially from careful tuning and, in practice, appears to require it.

By contrast, Logistic Regression exhibits the tightest ellipses, reflecting remarkable stability across specifications despite lower central performance; its results vary little, but consistently cluster in a lower-performing region. CNN, GNN, and, especially in Period 2, XGBoost strike a more balanced profile: they occupy intermediate-to-high performance regions while maintaining moderately compact ellipses, indicating a favourable compromise between accuracy and robustness.

Taken together, these plots show that while Random Forest remains the strongest performer, models such as GNN, CNN, and XGBoost display greater stability across specifications. This provides additional evidence that performance and robustness do not necessarily align and should be jointly considered.

6.3 Comparative analysis across configurations

We now examine how predictive accuracy varies along three key design dimensions: the prediction horizon (target depth), the amount of historical information provided to the models (data depth), and the dataset configuration. Tables 6–11 summarize these results for both periods. This analysis provides insights not only into which combinations yield the best discrimination and precision-recall performance, but also into the extent to which these patterns persist across fiscal cycles.

Target depth. Across both periods, the prediction horizon exerts a clear influence on model performance. ROC AUC distributions generally improve at two- and three-year horizons relative to one-year-ahead predictions (Table 6). In Period 1, medians increase from 0.752 (one year) to 0.826 (two years), while Period 2 shows an analogous progression from 0.829 to 0.840. This pattern is reflected in the means and percentiles, indicating a systematic increase in discriminative power as the temporal window widens. This pattern may reflect both the gradual accumulation of fiscal-distress signals and the broader definition of the positive class over longer target windows. We therefore interpret it as evidence of medium-term early-warning potential, rather than as proof that more distant events are intrinsically easier to predict.

PR AUC results are consistent with this pattern, although they should be interpreted together with changes in the prevalence of the positive class across target windows (Table 7). The mean PR AUC more than doubles from one-year to three-year horizons in Period 1 (0.052 to 0.122), and increases comparably from 0.059 to 0.120 in Period 2. Since PR AUC captures performance on the minority class, these findings suggest that distress signals accumulate and become more discernible well before the onset of fiscal distress. Thus, while short-term predictions remain challenging, two- and three-year horizons provide substantially more stable and accurate early-warning assessments.

Data depth. Differences across data depths are notably smaller. ROC AUC values remain tightly clustered for four-, five-, and six-year histories in both periods (Table 8). Median

values vary little within each period, and means lie within narrow intervals (around 0.72–0.73 in Period 1 and 0.75–0.76 in Period 2). Increasing the historical window from four to six years therefore produces, at most, marginal performance gains.

Similarly, PR AUC distributions show no systematic monotonic trend (Table 9). A five-year history often yields slightly higher mean values, particularly in Period 2, where it reaches 0.101. The differences are modest relative to those observed for target depth or dataset choice. Taken together, these results indicate that while models benefit from being trained on sufficiently long historical windows, the addition of extra years rapidly exhibits diminishing returns, with a five-year depth representing a reasonable balance between information richness and computational cost. This lack of significant improvement points to a form of information decay: fiscal data from beyond five years prior to the prediction window likely carries limited predictive content, as the institutional, political, and economic conditions prevailing in earlier periods may bear little resemblance to those driving current distress dynamics.

Dataset configuration. The comparison across dataset configurations reveals substantial differences in predictive performance, but also important nuances relevant for practical and policy-oriented applications. In both periods, models trained on the full BDAP dataset achieve the highest values of ROC AUC and PR AUC (Tables 10 and 11). In Period 1, BDAP attains a mean ROC AUC of 0.793 and a mean PR AUC of 0.119, clearly outperforming all other configurations. Period 2 confirms this dominance, with BDAP again providing the strongest results (mean ROC AUC of 0.800 and mean PR AUC of 0.116). These findings indicate that comprehensive annual financial statements remain the most informative source for predicting distress.

Nevertheless, two dataset families with earlier availability (Treasury Advances and ReadyBDAP) perform notably better than might be expected given their reduced informational content. The Treasury Advances dataset, derived from the real-time SIOPE system, captures high-frequency movements in liquidity needs and often anticipates short-term financial tension. The standalone Treasury Advances dataset has limited discriminatory power, with ROC AUC values close to 0.54 and PR AUC values around 0.05. Adding advances to FullBDAP does not improve average predictive performance in the present specifications, consistent with substantial overlap in the underlying liquidity signal. The main comparative advantage of SIOPE therefore lies in its timeliness and intra-year granularity rather than in additional end-of-year predictive content. Similarly, the ReadyBDAP datasets, based on forecasted expenditure commitments and expected revenue recognitions and available within the reference year, achieve intermediate performance levels. In Period 1, ReadyBDAP + Advances reaches a mean ROC AUC of 0.765 and a PR AUC of 0.089; in Period 2, the corresponding means are 0.772 and 0.096, respectively. Although these values remain below FullBDAP performance, they indicate that earlier-available datasets retain some useful predictive signal, especially in the upper tail of the performance distribution.

From a policy perspective, these results highlight a meaningful trade-off between informational richness and timeliness. FullBDAP datasets offer the highest predictive accuracy but become available only with significant delay, reducing their usefulness for real-time mon-

itoring and early intervention. By contrast, SIOPE-based advances enable high-frequency liquidity monitoring, while ReadyBDAP data provide earlier budget-forecast information than finalised accounts. Although their predictive performance is lower, the reduction is not large enough to negate their practical value in early-warning contexts. In settings where timely detection is crucial, such as continuous monitoring by oversight authorities, high-frequency risk flagging, or crisis-prevention dashboards, using Advances, ReadyBDAP, or their combination may be justified as a preliminary screening layer, to be followed by assessment using richer accounting information when available. The evidence suggests that these faster datasets can support preliminary risk flagging before final balance-sheet data become available, subject to calibration, validation and follow-up review by supervisory authorities.

The preceding analysis established that modern machine-learning (ML) methods outperform traditional approaches across a wide range of threshold-independent metrics, such as ROC AUC and PR AUC. However, these measures do not allow for a direct comparison with the existing system of Regulatory Indicators, which provides only binary classifications rather than predicted probabilities. To ground our results in current institutional practice, we complement the previous analysis with a threshold-based comparison using precision, recall, and the F1-score.

This comparison highlights a critical practical dimension: the adaptability of the decision threshold. Unlike the Regulatory Indicators, which rely on a fixed legislative rule, ML models allow policymakers to calibrate thresholds based on institutional needs, such as prioritizing the minimization of false negatives (missing a crisis) or false positives (unnecessary audits). Evaluating all methods at their optimal F1-score provides a conservative yet clear assessment of the potential gains offered by algorithmic oversight.

6.4 Comparison with the Regulatory Indicators

Table 14 reports the confusion-matrix metrics for the Regulatory Indicators alongside the best-performing ML models for both periods. The current regulatory framework achieves an F1-score of 0.133, characterised by a notably low recall of 0.107 and a moderate precision of 0.175. In contrast, the top ML architectures reach an F1-score of 0.272 in Period 1 and 0.297 in Period 2, representing improvements of approximately 105% and 123%, respectively. These gains suggest that ML methods can extract more predictive information from the available fiscal data than the existing rule-based system, at least under the evaluation design used here.

A structural shift is evident between the two periods in the way these scores are achieved. In Period 1, the optimal ML configuration (Logistic Regression) favors Recall (0.469) over Precision (0.192), identifying nearly half of the distress cases at the cost of a higher false-alarm rate. In Period 2, this relationship flips: the best model (CNN) achieves a higher Precision (0.423) but a lower Recall (0.229). This shift is consistent with the possibility that the post-2016 data structure supports more selective predictions (fewer false positives) but makes the identification of the full spectrum of at-risk municipalities more challenging. While the data do not explicitly reveal the cause of this flip, the finding underscores that the utility of ML depends on policy priorities – whether the objective is to achieve high recall

and flag a broad set of at-risk municipalities or to maximise the precision of distress signals generated by the system.

Finally, improvements generally persist even when models are not selected primarily for their F1-score. The last four rows of Table 14 report the performance of Random Forest and XGBoost models chosen for their superior PR AUC. Even under this alternative selection criterion, the models consistently outperform the regulatory baseline, suggesting that the relative advantage of the data-driven approach is not confined to a single model architecture or selection criterion.

7 Interpretability

A machine learning system that generates risk scores without explanatory information offers limited value to supervisory bodies. The Court of Auditors, the Ministry of the Interior, and the regional sections involved in reviewing and approving rebalancing plans need to understand the fiscal mechanisms driving a distress signal if they are to act on it. These actions might include deciding whether to convene an early audit, prioritising municipalities for technical assistance, or determining whether an observed deterioration reflects a transient shock or a structural trajectory toward default. Interpretability is therefore not an optional refinement but a precondition for institutional adoption.

We address this requirement using two complementary methods. MDI provides a global, model-internal ranking of variable importance derived from the Random Forest's training process, identifying which features consistently drove the classification across all municipalities and years. LIME complements this by generating, for each individual test observation, a local linear approximation of the CNN's decision boundary. As a result, it identifies which specific features pushed a particular municipality toward or away from a distress prediction. Together, these methods operate at different scales: MDI reveals the features that consistently drive model predictions that the model has generalised across the full dataset, while LIME pinpoints the case-specific signals that would inform an analyst reviewing an individual risk alert.

We restrict the interpretability analysis to the FullBDAP configuration for two reasons. First, it is the best-performing feature set in both periods (see Tables 10-11), so its internal logic is the most informative to unpack; explaining a weakly predictive configuration would yield unreliable attributions. Second, it provides the richest feature space within each accounting period and allows comparisons across model classes. Cross-period comparisons should still account for the change in budget classifications before and after harmonisation. The advance-related features discussed below are therefore those recorded in the financial statement; SIOPE captures the same phenomenon at daily frequency. The interpretability results should accordingly be read as identifying which signal matters, not which source delivers it first.

7.1 Interpretability of Random Forest results using native MDI

To interpret the internal logic of the Random Forest in predicting municipal financial distress, we utilize Mean Decrease Impurity (MDI). This model-native metric provides a computationally efficient global ranking of feature importance by measuring the contribution of each variable to the reduction of impurity across all decision trees (see Appendix C.1 for details on the methodology). Given the longitudinal nature of our dataset, we report the average MDI across all time steps to identify which fiscal dimensions most consistently contribute to the model's predictive accuracy.

The MDI analysis suggests that the model's predictive logic is closely related to the accounting frameworks of the two periods.

Period 1 (2009–2015): debt structure as the primary distress signal

In Period 1 (see Panels (a) of Figures 3 and 4), the most influential features are *Entrate derivanti da accensione di prestiti* observed in both the recognition and collection phases ($E2_{P1}\,REC$ and $E2_{P1}\,COLCY$ respectively), alongside *Totale spese per rimborso di prestiti* ($S3_{P1}COM$). Under the pre-harmonisation framework (DPR 194/1996), these titles also absorbed treasury advances, which had no separate classification until 2016. Their prominence is therefore consistent with two non-exclusive readings: structural dependence on external financing under weak borrowing constraints (Articles 203–204 TUEL), and a liquidity signal that the model detects but that the data cannot, at this resolution, separate from long-term debt. The recognition/collection gap for $E2_{P1}$ should be read with the same caveat: a lag between recognised loans and cash receipt can reflect genuine financing failure, but is also a physiological feature of CDP loans disbursed in construction-linked tranches. The MDI signal in Period 1 is best treated as a flag for further scrutiny rather than conclusive evidence, precisely because short- and long-term borrowing are not always separable in this period.

Period 2 (2016–2023): liquidity management under harmonised accounting

The top-ranked features shift to the treasury-advance cycle - *Anticipazioni da istituto tesoriere/cassiere* in both its recognition and collection phases ($E2_{P2}COLCY$ and $E2_{P2}REC$ respectively) and *Chiusura anticipazioni ricevute* ($S1_{P2}\,COM$ and $S1_{P2}\,PAY\,CY$) - reflecting the dedicated classification introduced by Legislative Decree 118/2011. A municipality that draws advances early in the year, extends them across quarters, and struggles to close them before 31 December exhibits the pattern the Court of Auditors has long flagged as chronic cash insufficiency rather than seasonal treasury management. Capital expenditure payments ($S4_{P2}\,PAY$) also rank prominently: sustained capital payment flows may indicate access to project-tied financing and functioning creditor relationships, while a sharp decline may signal liquidity constraints or the diversion of financing resources toward current imbalances, a pattern observed two to three years before formal default in several documented cases. One caveat: recognitions and collections for the same category can both rank highly simply because a single transaction generates multiple entries in the Italian recording system; analysts should avoid double-counting these as independent signals in summary risk scores. That this signal is the dominant one under harmonised accounting is precisely what motivates monitoring it in real time, before the financial statement records it (see Section 8).

Population: a structural rather than linear signal. Population (P) ranks in the Period 1

top-10 but disappears from Period 2 (see Panels (a) and (b) of Figure 3). Because MDI is non-directional, it cannot on its own establish the sign of this association, and Population appears in neither the pro-distress nor pro-stability LIME rankings in either period (see Figures 5 and 6). Its role in the Random Forest is therefore likely structural rather than linear: useful for partitioning municipalities into risk groups, but without a sign or magnitude stable enough to register as a first-order linear effect. This pattern is consistent with a threshold-like relationship between size and risk documented descriptively by Corte dei conti (2026): financial default incidence rises from 7.8% below 2,000 residents to 17.5% in the 20,000–60,000 bracket before levelling off. A relationship that increases with size up to a point and then plateaus is exactly what a tree-based model can exploit through repeated splits without it reducing to a single linear effect, though confirming this formally would require an explicit partial-dependence analysis, which we leave for future work. The disappearance of Population in Period 2 also fits the broader CY/PY pattern documented in Tables 15–16, whereby current-year (CY) indicators systematically outrank previous-year-based (PY) ones: CY variables measure the in-year conversion of recognitions into cash, a direct signal of administrative efficiency, while PY variables reflect a stock of accruals accumulated for heterogeneous reasons that dilutes their signal. Before 2016, the accounting framework offered no direct behavioural counterpart to size-driven managerial complexity, so Population may have served as a coarse proxy for this dimension of risk; after 2016, this role appears partly displaced by precise CY liquidity indicators ($E2_{P2}COL$ and $S1_{P2}COM$) and the capital payment rate ($S4_{P2}\,PAY$).

A further feature that ranks prominently in both periods is the Z4 geographic indicator, which identifies municipalities located in Southern Italy (see Panels (a) and (b) of Figures 3 and 4). Given the concentration of distress in the South documented in Section 2, its high MDI importance reflects the deep structural disadvantage of Southern municipalities (lower own-revenue capacity, higher dependence on central transfers, weaker administrative capacity). However, the fact that financial behavioural markers such as treasury advances rank alongside or above Z4 in Period 2 is an important qualification: the model is not simply predicting that Southern municipalities will default because they are Southern. Rather, it identifies the specific fiscal behaviours (such as chronic reliance on advances, compressed capital spending, deteriorating collection rates) that translate structural disadvantage into operational insolvency. This distinction matters for policy: it suggests that targeted interventions addressing cash flow management and revenue collection capacity could reduce distress risk even in structurally fragile contexts, without requiring the wholesale fiscal transfers that a purely geographic diagnosis would imply.

7.2 Interpretability of CNN Results Using LIME

To interpret the results of the Convolutional Neural Network (CNN) model, we employ Local Interpretable Model-Agnostic Explanations (LIME), as described by Ribeiro et al. (2016). While models such as Logistic Regression or Decision Trees offer more direct interpretability, they often require temporal features to be disaggregated across discrete time steps. This disperses information over multiple data points. In contrast, applying a version of the LIME

framework tailored for temporal data allows us to obtain a single interpretability score for each feature, aggregated across all time steps. This approach allows us to use one of the strongest predictive models while preserving transparency in the decision-making process. For a detailed overview of the LIME framework, we refer the reader to Appendix C.2.

For the purposes of this paper, given a specific classification task, defined by the combination of data depth, target depth, and dataset, we applied the following procedure to generate the interpretations:

1. We randomly sampled up to 1000 points from the test set; preliminary tests indicated no substantial difference in LIME behavior between positive and negative labels.

2. For each sampled point, we generated a local linear model.

3. From these local models, we extracted the coefficients for every feature. These coefficients indicate both the magnitude and the sign (positive for distress, negative for stability) of the feature's contribution to the prediction.

4. We averaged these coefficients across all sampled points to obtain an aggregated interpretation of feature importance for each dataset.

The resulting LIME score distributions (Figures 5, 6 and 7) demonstrate a strong alignment with the MDI analysis of the Random Forest[10].

Period 1: convergence with MDI on debt-driven fragility

In the pre-2016 period, the LIME profiles closely corroborate the MDI findings. The two features generating the largest average positive coefficients, those most consistently pushing the CNN toward a distress classification, are *Totale spese per rimborso di prestiti* ($S3_{P1}COM$) and *Entrate derivanti da accensione di prestiti* ($E2_{P1}REC$). Their joint prominence reflects the debt-driven fragility that characterised Italian municipal finance before harmonisation: municipalities that had accumulated large loan portfolios in the 1990s and early 2000s, often to fund inefficient public enterprises or to cover current deficits through accounting practices that deferred recognition of fiscal imbalances, entered the 2009–2015 period with debt service burdens that absorbed a disproportionate share of current revenues. The model captures this structural constraint irrespective of the specific year or the political cycle. A concrete illustration of this mechanism is provided by the wave of default declarations that followed the austerity measures of 2011–2013 (Legislative Decree 201/2011 and subsequent Spending Review provisions), which sharply reduced central government transfers to municipalities. Those with the highest debt service burdens, and correspondingly limited fiscal space, were the first to exhaust corrective options and file for default. The CNN's LIME scores, concentrated on debt repayment and new borrowing entries, are consistent with the diagnostic logic that fiscal analysts at the Italian Court of Auditors had identified retrospectively in those cases.

[10]Z4 does not appear in the top- or bottom-10 LIME rankings in either period, despite its high MDI importance. The two metrics are not comparable in this respect: LIME coefficients are averaged across all sampled test observations, including Northern municipalities for which Z4 is zero by construction, so its aggregate local coefficient is diluted, whereas MDI is computed over the full training set and rewards any variable that partitions the sample effectively.

Period 2: treasury advances as operational risk indicators

In the post-2016 period, *Anticipazioni da istituto tesoriere/cassiere* ($E2_{P2}COLCY$) and *Chiusura anticipazioni ricevute* ($S1_{P2}COM$) appear to be among the 10 largest positive contributions to distress predictions, confirming MDI findings. The negative LIME scores consistently associated with *Spese in conto capitale* ($S4_{P2}PAYCY$) reinforce its interpretation as a protective factor: municipalities that sustain capital investment flows have access to project-tied financing sources (EU structural funds, national cohesion programmes, extraordinary public works transfers) that reduce their dependence on short-term treasury credit. The gradual erosion of capital spending capacity, as captured by declining payments in the $S4_{P2}$ title, may therefore function as an early indicator of mounting fiscal adjustment pressures.

The cross-method convergence of MDI and LIME on the same set of financial indicators, across two different model architectures, two different accounting regimes, and two different methodological approaches, suggests that the models are capturing economically meaningful regularities rather than statistical noise. Equally important, the indicators they identify (debt service burdens, treasury advance reliance, capital payment capacity) are precisely those that fiscal analysts and judicial auditors already use as diagnostic criteria in manual case reviews. The models do not appear to rely on an unrelated or opaque logic. Their most important predictors are consistent with indicators commonly used in audit practice, such as debt-service pressure, treasury-advance reliance and capital-payment capacity.

8 Conclusions

In this paper we have developed a data-driven framework for predicting financial distress in Italian municipalities over the period 2009–2023, combining finalised balance sheet data from OpenBDAP, high-frequency daily liquidity data from SIOPE, demographic indicators from ISTAT, and legally validated crisis records from the Ca' Foscari database. When evaluated against both traditional econometric benchmarks and the current system of fiscal stress indicators, several machine learning models, especially Random Forest and selected neural-network specifications, outperform the existing institutional benchmark. The best-performing configurations reach median ROC AUC values around or above 0.87 and F1-scores of 0.27-0.30, compared with 0.13 for the benchmark. These gains are not merely statistical: they translate directly into fewer missed crises, reducing the number of false negatives that, under the current system, leave fiscally fragile municipalities without supervisory attention until their situation has deteriorated to the point of formal default or pre-default filing.

8.1 Key findings and their institutional significance

Three findings carry particular significance for the governance of Italian local public finance.

First, the prediction horizon matters more than the depth of historical data. Two- and three-year forecasts consistently outperform one-year-ahead predictions, reflecting the gradual, structural nature of municipal financial deterioration. This is not a purely technical finding: it implies that an early-warning system calibrated to a two-year horizon could,

in principle, alert supervisory bodies in time to activate the pre-default procedure before a municipality's fiscal situation reaches the point at which corrective action is no longer viable. The current system of fiscal stress indicators, which relies on lagged accounting information and fixed thresholds, is less well suited to this medium-term early-warning objective: it identifies distress after the fact rather than before it.

Second, a meaningful trade-off exists between informational richness and timeliness. Finalised OpenBDAP balance sheets yield the highest predictive accuracy, but they become available too late for genuinely preventive action. High-frequency SIOPE data and provisional ReadyBDAP accounts are individually less precise than finalised accounts, but they become available earlier and can support preliminary monitoring. Specifically, treasury advance data from SIOPE are continuously updated at the daily level, meaning that a monitoring system based on advance frequency, duration and per-capita amounts could be adapted to generate rolling risk scores for municipalities before fiscal year closure, subject to calibration, validation and data-governance arrangements. Such an architecture could help the Ministry of the Interior and the Italian Court of Auditors complement annual reporting cycles with more continuous liquidity monitoring, intervening at the first signs of sustained liquidity stress rather than after such stress has crystallised into accounting irregularities.

Third, the interpretability analysis confirms that the models' predictive logic aligns with the diagnostic frameworks already employed by fiscal oversight bodies. In the pre-2016 period, the dominant distress signals are concentrated in the debt and borrowing titles of the municipal budget classification — reflecting both the structural indebtedness of Italian municipalities in that period and the pre-harmonisation aggregation of treasury advances within the same budget category, which prevented their separate identification. In the post-2016 period, the harmonised accounting framework enables the explicit identification of treasury advance management and capital expenditure capacity as the primary distress signals — a distinction that was precluded by the aggregated classification of the pre-harmonisation period and that mirrors the emphasis placed by the regional sections of Italian Court of Auditors on these indicators in their annual municipal finance reports. This alignment is important not only as a validation of model quality but as evidence that the machine learning system can be explained to, and trusted by, the human analysts and magistrates whose institutional support is necessary for any reform of the surveillance framework.

8.2 Implications for the reform of fiscal surveillance

The current system of fiscal stress indicators was designed as a rule-based administrative filter, not a predictive instrument. Its limitations are structural, not merely technical. Incremental refinements within the existing architecture are unlikely to produce the performance gains documented in this paper. The evidence presented in this paper points to a more fundamental redesign, built around three principles.

The first is the importance of integrating real-time data flows into municipal financial oversight. The systematic incorporation of timely and granular transaction data, collected on a daily or near-daily basis, into a monitoring dashboard, alongside budget forecast data from OpenBDAP, would allow oversight bodies to track the evolution of municipal liquidity

stress on a rolling basis. This would flag anomalies as they emerge rather than document them after year-end closure. The value of this approach lies precisely in its timeliness: early-warning signals can only be acted upon if they are visible in real time, not retrospectively reconstructed from annual accounts. This may not require a change to the core TUEL framework, but it would require clear inter-institutional governance, data-sharing arrangements, audit trails, and operational rules for translating raw data flows into supervisory alerts.

The second principle is the differentiation of thresholds and weights by municipal characteristics. The current system applies identical parameters to a 500-resident Alpine municipality and a 200,000-resident Southern city. The modelling evidence presented here, particularly the evolution of feature importance across accounting regimes and the interaction between geographic location and fiscal behaviours, suggests that a well-calibrated risk scoring system should weight indicators differently depending on municipal size, geographic region, fiscal capacity, and accounting period. This is technically feasible: the Random Forest and XGBoost models already produce municipality-specific risk scores that implicitly encode these heterogeneities. Making these scores available to the regional sections of the Italian Court of Auditors – as an input to rather than a substitute for judicial assessment – could improve the targeting of pre-default interventions.

The third principle is the shift from binary classification to probabilistic risk stratification. The current system yields a binary output: a municipality either exceeds the threshold on four or more indicators, or it does not. This design entails a substantial loss of information. A municipality that exceeds four indicators by a small margin is treated identically to one that exceeds all eight by large margins. Similarly, a municipality that improves from six to four triggered indicators receives no recognition of its progress. Probabilistic risk scores, updated continuously as new data become available, would allow oversight bodies to rank municipalities by urgency, allocate technical assistance and monitoring resources proportionally to risk, and track trajectories over time, rewarding fiscal improvement and escalating attention for deteriorating cases.

8.3 Limitations and Future Research

Several limitations of the present analysis point toward productive extensions. The binary classification framework, predicting entry into formal distress procedures, captures only the most severe end of the municipal fiscal vulnerability spectrum. A large number of municipalities experience significant fiscal stress without ever filing for pre-default, either because they manage to stabilise through local adjustment, because they receive extraordinary central transfers, or because political and administrative inertia delays formal proceedings. Survival analysis or ordinal models calibrated on intermediate indicators of fiscal strain - such as persistent revenue shortfalls, growing expenditure arrears and declining collection rates - could extend the diagnostic reach of the framework to this broader population of at-risk entities.

On the network side, the GNN implementation currently uses geographic contiguity as the sole basis for municipal linkages. This is a useful first approximation, but Italian municipal finances are also interconnected through a richer set of structural ties. Inter-municipal consortia (*Comunità montane*, *Unioni di Comuni*) and shared service providers create ad-

ministrative and financial dependencies that do not necessarily follow territorial boundaries. Additionally, municipally owned enterprises (*società partecipate comunali*) introduce a distinct layer of financial interconnection: municipalities that jointly hold equity stakes in the same utility or service company share exposure to that entity's financial performance. As a result, stress or liquidity demands originating within a local public enterprise can simultaneously affect multiple municipal balance sheets regardless of geographic distance. Incorporating these network dimensions into the graph structure would allow the GNN to capture a more realistic topology of fiscal contagion, improving both the accuracy of stress propagation modelling and the interpretability of the resulting risk scores. Constructing multi-layered networks that encode these institutional and economic interdependencies, and learning their relative weights through attention-based GNN architectures, could reveal contagion dynamics that the current specification cannot capture.

A natural extension of the present framework would be to model the determinants of treasury advance reliance itself, rather than treating advances as a reduced-form predictor of distress. Advances at time t may depend systematically on fiscal variables observed at $t-k$: for instance, a sustained deceleration in capital expenditure payments ($S4_{P2}PAY$), by signalling the erosion of project-financing pipelines, may predict growing advance utilisation several quarters ahead. A two-stage framework, a first model predicting advance dynamics, a second translating advance profiles into distress risk, could deepen the early-warning capability of the system. We leave the development of such a framework to future research.

Feature engineering also offers substantial scope for future development. Applying natural language processing to the narrative sections of the budget appropriation and the outturn accounts (*rendiconto di gestione*), could extract qualitative signals about municipal fiscal strategy and risk awareness that are absent from numerical tables. Similarly, the integration of local labour market indicators, property market data, and political variables (electoral cycles, coalition fragility, mayoral tenure) could improve the model's capacity to distinguish between municipalities whose fiscal stress reflects transient shocks and those on a structural trajectory toward insolvency.

A practical challenge that this paper does not fully address is the governance of false positives. The models presented here significantly reduce false negatives relative to the current system of fiscal stress indicators, but partly at the cost of a higher false alarm rate. The two error types, however, are not symmetric in their consequences. A missed crisis (false negative) implies that a fiscally fragile municipality receives no supervisory attention until its situation has deteriorated to the point of formal filing, with attendant service disruptions, mandatory tax increases, and reputational damage to local governance. A false alarm (false positive), by contrast, triggers additional scrutiny of a municipality that ultimately proves financially sound: an administrative burden, but not an irreversible harm. Within reasonable margins, moderate over-signalling may be institutionally acceptable if the alerts trigger proportionate review rather than automatic sanctions, consistent with the practice of the Italian Court of Auditors's regional sections. That said, the optimal threshold is not invariant to institutional context: a full cost-benefit calibration of the administrative costs of false positives against the social costs of missed crises remains a prerequisite for operational deployment. The models presented here should accordingly be understood as tools for

differential risk stratification, directing supervisory attention toward the entities that most warrant deeper review, rather than as autonomous decision-making systems. Heterogeneity in territorial monitoring capacity across regions further strengthens the case for such targeted, data-driven prioritisation.

References

Alaminos, D., García, M., and Rodríguez, J. L. (2018). Data mining for municipal financial distress prediction. In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 155–162.

Barboza, F., Kimura, H., and Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83:405–417.

Bianchi, F., Crosato, L., and Liberati, C. (2024). La previsione della sofferenza finanziaria dei comuni italiani: un approccio data-driven. In *VII Rapporto Ca' Foscari sui comuni. 2024 – I comuni nella nuova governance europea della finanza pubblica*, pages 303–321. Castelvecchi.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, New York.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Breiman, L., Friedman, J., Olshen, R. A., and Stone, C. J. (2017). *Classification and regression trees*. Chapman and Hall/CRC.

Bruna, J., Zaremba, W., Szlam, A., and LeCun, Y. (2014). Spectral networks and locally connected networks on graphs. In *International Conference on Learning Representations (ICLR)*.

Caravaggio, N., Degni, M., Lagravinese, R., and Resce, G. (2024). L'utilizzo degli indicatori di bilancio per prevenire la criticità finanziaria dei comuni. In *VII Rapporto Ca' Foscari sui comuni. 2024 – I comuni nella nuova governance europea della finanza pubblica*, pages 1085–1107. Castelvecchi.

Chen, T. and Guestrin, C. (2016). 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.

Corte dei conti (2021). Prime analisi sulla qualità della spesa dei comuni. Technical Report Deliberazione n. 14/SEZAUT/2021/FRG, Sezione delle Autonomie, Corte dei conti.

Corte dei conti (2024). Esiti parziali dell'indagine sull'applicazione dei principi contabili in ordine alla verifica della consistenza della cassa e del fondo rischi da contenzioso (anno 2021). Technical Report Deliberazione n. 5/2024/VSG, Sezione Regionale di Controllo per l'Emilia Romagna, Corte dei conti.

Corte dei conti (2025a). Esiti finali dell'indagine sull'applicazione dei principi contabili in ordine alla verifica della consistenza della cassa e del fondo rischi da contenzioso - anno 2021. Technical Report Deliberazione n. 66/2025/VSG, Corte dei conti.

Corte dei conti (2025b). Relazione sulla gestione finanziaria degli enti locali: Comuni, province e città metropolitane – esercizi 2022–2024. Technical Report Deliberazione n. 14/SEZAUT/2025/FRG, Corte dei conti.

Corte dei conti (2026). Relazione al parlamento sui comuni in situazione di crisi finanziaria. Technical Report Deliberazione n. 3/SEZAUT/2026/FRG, Corte dei conti.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*, volume 1. MIT press Cambridge.

Gori, M., Monfardini, G., and Scarselli, F. (2005). A new model for learning in graph domains. In *Proceedings of the International Joint Conference on Neural Networks*, volume 2, pages 729 – 734 vol. 2.

Gorina, E. M. (2018). Local fiscal distress: measurement and prediction. *Public Budgeting & Finance*, 38(4):1–21.

Gregori, W. D. and Marattin, L. (2019). Determinants of fiscal distress in italian municipalities. *Empirical Economics*, 56(4):1386–1404.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer, New York, 2 edition.

Holopainen, J. and Sarlin, P. J. (2017). Toward robust early-warning models: a horse race, ensembles and model uncertainty. *Quantitative Finance*, 17(6):859–874.

Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied logistic regression*. John Wiley & Sons.

Ministero dell'Interno and Ministero dell'Economia e delle Finanze (2018). Nota metodologica: Definizione dei parametri obiettivi ai fini dell'individuazione degli enti locali in condizioni di deficitarietà strutturale nel triennio 2019–2021. Documento tecnico, Osservatorio sulla Finanza e la Contabilità degli Enti Locali. Allegato al Decreto del Ministro dell'Interno di concerto con il Ministro dell'Economia e delle Finanze.

Padovani, E., Porcelli, F., and Zanardi, A. (2024). The determinants of the financial distress of italian municipalities: How much is it due to inadequate resources? *International Tax and Public Finance*, 31:1494–1533.

Piermarini, D., Sudoso, A. M., and Piccialli, V. (2023). Predicting municipalities in financial distress: a machine learning approach enhanced by domain expertise. *arXiv preprint arXiv:2302.05780*.

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1135–1144. ACM.

Tölö, E. (2020). Predicting systemic financial crises with recurrent neural networks. *Journal of Financial Stability*, 49:100746.

Tables

Table 1: Benchmark performance of current regulatory indicators

	Predicted negative	**Predicted positive**
Actual negative	106,684	1,311
Actual positive	740	157

Notes: This table reports the confusion matrix for the regulatory indicators over the period 2010-2023. It shows the number of municipalities classified as distressed or non-distressed by the regulatory system compared to their actual financial condition, highlighting both false positives and false negatives.

Table 2: Data availability by type and period

Feature set	**2009–2015**	**2016–2023**
FullBDAP	7,706 (99.1%)	7,683 (98.8%)
ReadyBDAP	7,706 (99.1%)	7,683 (98.8%)
Indicators	–	7,683 (98.8%)
Treasury advances	7,778 (100.0%)	7,778 (100.0%)
ISTAT	7,778 (100.0%)	7,778 (100.0%)
Distress outcomes	40 (0.5%)	71 (0.9%)

Notes: Average values over the reference period.

Table 3: Summary of feature sets

Feature Set	**Features**	**Data Range**	**Source**
FullBDAP (Period 1)	30	2009–2015	OpenBDAP
ReadyBDAP (Period 1)	10	2009–2015	OpenBDAP
FullBDAP (Period 2)	42	2016–2023	OpenBDAP
ReadyBDAP (Period 2)	14	2016–2023	OpenBDAP
Indicators	10	2016–2023	OpenBDAP
Geographical and demographic data	6	2009–2023	ISTAT
Treasury advances	5	2009–2023	SIOPE

Notes: This table summarizes the feature sets used for model training across the two periods. It specifies the number of variables, their temporal coverage, and the data sources (BDAP, ISTAT, SIOPE). These feature sets are combined in different ways to assess model performance under various informational scenarios.

Table 4: Detailed ROC AUC Summary by Model (Period 1 and Period 2)

Model	min	p5	p25	p50	p75	p95	max	mean	SD
Panel A: Period 1									
CNN	0.372	0.437	0.742	0.815	0.856	0.871	0.881	0.752	0.146
Decision Tree	0.251	0.268	0.460	0.533	0.543	0.572	0.580	0.480	0.109
GNN	0.492	0.620	0.700	0.750	0.814	0.852	0.861	0.751	0.081
Logistic Regression	0.693	0.709	0.765	0.826	0.849	0.877	0.889	0.808	0.055
Random Forest	0.217	0.269	0.672	**0.873**	0.898	0.914	0.918	0.732	0.255
XGBoost	0.396	0.548	0.761	0.849	0.882	0.895	0.896	0.794	0.127
Panel B: Period 2									
CNN	0.392	0.537	0.777	0.841	0.868	0.906	0.920	0.799	0.121
Decision Tree	0.212	0.268	0.518	0.534	0.557	0.574	0.579	0.501	0.100
GNN	0.386	0.504	0.635	0.731	0.805	0.869	0.872	0.712	0.119
Logistic Regression	0.744	0.767	0.837	0.866	0.898	0.917	0.934	0.862	0.046
Random Forest	0.226	0.271	0.868	**0.895**	0.909	0.920	0.940	0.798	0.230
XGBoost	0.524	0.588	0.831	0.863	0.883	0.904	0.928	0.828	0.096

Notes: The table reports summary statistics across the relevant experimental configurations: 36 configurations in Period 1 and 63 in Period 2, for a total of 99 across both periods.

Table 5: Detailed PR AUC Summary by Model (Period 1 and Period 2)

Model	min	p5	p25	p50	p75	p95	max	mean	SD
Panel A: Period 1									
CNN	0.014	0.021	0.042	0.071	0.124	0.196	0.275	0.092	0.064
Decision Tree	0.004	0.021	0.055	0.091	0.115	0.155	0.178	0.089	0.044
GNN	0.008	0.011	0.042	0.057	0.074	0.125	0.133	0.059	0.030
Logistic Regression	0.028	0.030	0.058	0.079	0.092	0.110	0.147	0.074	0.028
Random Forest	0.005	0.005	0.048	0.108	0.164	0.208	0.228	0.112	0.064
XGBoost	0.018	0.032	0.073	**0.129**	0.176	0.202	0.207	0.125	0.059
Panel B: Period 2									
CNN	0.024	0.032	0.051	0.096	0.126	0.181	0.294	0.095	0.053
Decision Tree	0.004	0.019	0.062	0.091	0.130	0.164	0.179	0.092	0.045
GNN	0.006	0.009	0.024	0.051	0.092	0.165	0.238	0.066	0.051
Logistic Regression	0.026	0.034	0.063	0.091	0.114	0.191	0.224	0.095	0.046
Random Forest	0.005	0.009	0.058	**0.099**	0.129	0.181	0.201	0.097	0.052
XGBoost	0.024	0.034	0.058	0.097	0.124	0.172	0.207	0.098	0.043

Notes: The table reports summary statistics of the empirical distribution of PR AUC values across all experiments, grouped by model and period.

Table 6: Detailed ROC AUC Summary by Target Depth (Period 1 and Period 2)

Target Depth	min	p5	p25	p50	p75	p95	max	mean	SD
Panel A: Period 1									
1 Year	0.217	0.265	0.571	0.752	0.840	0.878	0.904	0.688	0.184
2 Years	0.291	0.304	0.649	**0.826**	0.861	0.896	0.900	0.741	0.177
3 Years	0.324	0.357	0.610	0.796	0.869	0.905	0.918	0.730	0.177
Panel B: Period 2									
1 Year	0.212	0.342	0.564	0.829	0.879	0.905	0.928	0.727	0.190
2 Years	0.272	0.392	0.668	0.832	0.878	0.916	0.928	0.755	0.169
3 Years	0.274	0.363	0.721	**0.840**	0.888	0.915	0.940	0.768	0.171

Notes: The table reports summary statistics of the empirical distribution of ROC AUC values across all experiments, grouped by target depth and period.

Table 7: Detailed PR AUC Summary by Target Depth (Period 1 and Period 2)

Target Depth	min	p5	p25	p50	p75	p95	max	mean	SD
Panel A: Period 1									
1 Year	0.004	0.007	0.028	0.044	0.066	0.133	0.200	0.052	0.038
2 Years	0.032	0.043	0.065	0.095	0.141	0.175	0.187	0.101	0.043
3 Years	0.028	0.045	0.080	**0.112**	0.164	0.208	0.275	0.122	0.056
Panel B: Period 2									
1 Year	0.004	0.006	0.032	0.051	0.086	0.131	0.162	0.059	0.039
2 Years	0.010	0.029	0.064	0.095	0.115	0.159	0.294	0.093	0.042
3 Years	0.014	0.044	0.085	**0.116**	0.149	0.199	0.238	0.120	0.047

Notes: The table reports summary statistics of the empirical distribution of PR AUC values across all experiments, grouped by target depth and period.

Table 8: Detailed ROC AUC Summary by Data Depth (Period 1 and Period 2)

Data Depth	min	p5	p25	p50	p75	p95	max	mean	SD
Panel A: Period 1									
4 Years	0.272	0.329	0.680	0.763	0.851	0.894	0.900	0.720	0.168
5 Years	0.217	0.318	0.603	**0.821**	0.856	0.896	0.917	0.730	0.186
6 Years	0.252	0.327	0.570	0.795	0.858	0.899	0.918	0.709	0.187
Panel B: Period 2									
4 Years	0.212	0.360	0.641	0.839	0.889	0.910	0.928	0.753	0.180
5 Years	0.250	0.335	0.578	**0.850**	0.883	0.917	0.940	0.751	0.183
6 Years	0.237	0.398	0.604	0.825	0.869	0.907	0.934	0.746	0.171

Notes: The table reports summary statistics of the empirical distribution of ROC AUC values across all experiments, grouped by input depth and period.

Table 9: Detailed PR AUC Summary by Data Depth (Period 1 and Period 2)

Data Depth	min	p5	p25	p50	p75	p95	max	mean	SD
Panel A: Period 1									
4 Years	0.005	0.024	0.044	0.079	0.128	0.178	0.187	0.089	0.052
5 Years	0.004	0.024	0.052	0.079	0.115	0.206	0.275	0.091	0.056
6 Years	0.005	0.013	0.057	**0.088**	0.143	0.198	0.228	0.095	0.056
Panel B: Period 2									
4 Years	0.004	0.023	0.054	0.087	0.116	0.178	0.207	0.088	0.046
5 Years	0.005	0.012	0.060	**0.097**	0.136	0.188	0.294	0.101	0.056
6 Years	0.005	0.019	0.044	0.088	0.113	0.159	0.224	0.084	0.045

Notes: The table reports summary statistics of the empirical distribution of PR AUC values across all experiments, grouped by data depth and period.

Table 10: Detailed ROC AUC Summary by Dataset (Period 1 and Period 2)

Dataset	min	p5	p25	p50	p75	p95	max	mean	SD
Panel A: Period 1									
Treasury advances	0.217	0.255	0.344	0.574	0.726	0.779	0.861	0.538	0.202
FullBDAP	0.531	0.535	0.758	**0.854**	0.876	0.898	0.918	0.793	0.125
FullBDAP + Treasury advances	0.504	0.535	0.738	0.828	0.862	0.902	0.917	0.782	0.121
ReadyBDAP + Treasury advances	0.496	0.526	0.702	0.808	0.849	0.893	0.902	0.765	0.124
Panel B: Period 2									
Treasury advances	0.212	0.238	0.300	0.588	0.732	0.799	0.860	0.541	0.212
FullBDAP	0.496	0.523	0.752	**0.872**	0.897	0.910	0.916	0.800	0.141
FullBDAP + Treasury advances	0.505	0.528	0.752	0.856	0.882	0.907	0.908	0.788	0.137
FullBDAP + Indicators + Treasury advances	0.517	0.534	0.731	0.860	0.888	0.921	0.940	0.793	0.137
Indicators	0.453	0.534	0.712	0.863	0.899	0.918	0.928	0.795	0.139
Indicators + Treasury advances	0.386	0.521	0.726	0.857	0.885	0.917	0.928	0.787	0.147
ReadyBDAP + Treasury advances	0.477	0.524	0.758	0.833	0.866	0.897	0.907	0.772	0.136

Notes: The table reports summary statistics of the empirical distribution of ROC AUC values across all experiments, grouped by dataset and period.

Table 11: Detailed PR AUC Summary by Dataset (Period 1 and Period 2)

Dataset	min	p5	p25	p50	p75	p95	max	mean	SD
Panel A: Period 1									
Treasury advances	0.005	0.007	0.030	0.055	0.074	0.095	0.116	0.053	0.028
FullBDAP	0.012	0.049	0.067	**0.108**	0.163	0.206	0.275	0.119	0.057
FullBDAP + Treasury advances	0.013	0.031	0.054	0.100	0.160	0.202	0.220	0.107	0.058
ReadyBDAP + Treasury advances	0.004	0.027	0.051	0.083	0.117	0.164	0.207	0.089	0.046
Panel B: Period 2									
Treasury advances	0.004	0.005	0.026	0.042	0.065	0.102	0.177	0.049	0.034
FullBDAP	0.004	0.040	0.075	**0.116**	0.157	0.211	0.238	0.116	0.055
FullBDAP + Treasury advances	0.019	0.041	0.091	0.100	0.126	0.173	0.209	0.105	0.044
FullBDAP + Indicators + Treasury advances	0.008	0.025	0.050	0.095	0.124	0.179	0.294	0.095	0.054
Indicators	0.007	0.031	0.059	0.099	0.130	0.167	0.192	0.095	0.045
Indicators + Treasury advances	0.006	0.025	0.055	0.090	0.114	0.151	0.175	0.086	0.040
ReadyBDAP + Treasury advances	0.009	0.033	0.067	0.093	0.124	0.169	0.192	0.096	0.041

Notes: The table reports summary statistics of the empirical distribution of PR AUC values across all experiments, grouped by dataset and period.

Table 12: Median ROC AUC scores by Dataset and Model (Period 1 and Period 2)

Dataset \ Model	CNN	DT	GNN	LR	RF	XGB
Panel A: Period 1						
Treasury advances	0.523	0.299	0.755	0.740	0.299	0.645
FullBDAP	0.858	0.538	0.796	0.859	0.879	0.874
FullBDAP + Treasury advances	0.843	0.539	0.745	0.828	0.899	0.872
ReadyBDAP + Treasury advances	0.797	0.532	0.742	0.822	0.886	0.864
Panel B: Period 2						
Treasury advances	0.572	0.286	0.694	0.781	0.282	0.636
FullBDAP	0.858	0.541	0.744	0.897	0.908	0.877
FullBDAP + Treasury advances	0.856	0.533	0.785	0.895	0.895	0.856
FullBDAP + Indicators + Treasury advances	0.826	0.552	0.726	0.879	0.908	0.885
Indicators	0.864	0.559	0.709	0.900	0.906	0.873
Indicators + Treasury advances	0.867	0.537	0.720	0.889	0.896	0.860
ReadyBDAP + Treasury advances	0.810	0.532	0.760	0.839	0.890	0.851

Notes: The table reports median ROC AUC scores across all experiments, grouped by dataset, model, and evaluation period.

Table 13: Median PR AUC scores by Dataset and Model (Period 1 and Period 2)

Dataset \ Model	CNN	DT	GNN	LR	RF	XGB
Panel A: Period 1						
Treasury advances	0.041	0.052	0.065	0.064	0.042	0.082
FullBDAP	0.163	0.101	0.065	0.101	0.163	0.162
FullBDAP + Treasury advances	0.112	0.104	0.053	0.084	0.163	0.166
ReadyBDAP + Treasury advances	0.079	0.097	0.055	0.079	0.126	0.125
Panel B: Period 2						
Treasury advances	0.043	0.036	0.062	0.063	0.014	0.052
FullBDAP	0.129	0.080	0.054	0.117	0.116	0.121
FullBDAP + Treasury advances	0.103	0.072	0.092	0.098	0.096	0.126
FullBDAP + Indicators + Treasury advances	0.096	0.117	0.034	0.087	0.099	0.124
Indicators	0.092	0.133	0.073	0.100	0.115	0.097
Indicators + Treasury advances	0.110	0.092	0.033	0.082	0.113	0.090
ReadyBDAP + Treasury advances	0.110	0.091	0.051	0.089	0.105	0.110

Notes: The table reports median PR AUC scores across all experiments, grouped by dataset, model, and evaluation period.

Table 14: Comparison of Regulatory Indicators and ML Models Using Precision, Recall, and F1 Score

Method	**Precision**	**Recall**	**F1 score**
Regulatory Indicators (2010–2023)	0.175	0.107	0.133
ML Best Model Period 1	0.192	0.469	0.272
ML Best Model Period 2	0.423	0.229	0.297
Random Forest Period 1	0.250	0.259	0.255
XGBoost Period 1	0.364	0.148	0.211
Random Forest Period 2	0.214	0.125	0.158
XGBoost Period 2	0.308	0.143	0.195

Notes: Precision, recall, and F1 score are computed from the confusion matrix of predicted vs. actual distress outcomes. Let TP, FP, and FN denote true positives, false positives, and false negatives, respectively. Then Precision $= \frac{\text{TP}}{\text{TP+FP}}$, Recall $= \frac{\text{TP}}{\text{TP+FN}}$, and F1 $= 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision+Recall}}$. Precision measures the accuracy of predicted distress cases, recall measures the percentage of actual distress cases correctly identified, and the F1 score provides a balanced summary combining both dimensions. "ML Best Model" identifies the configuration with the highest F1-score per period. Random Forest and XGBoost entries represent the best models of those specific classes based on their PR AUC scores.

Table 15: BDAP Dataset Features Summary Period 1 (2009–2015)

Feature Name	**Group**	**Label**
Total current expenditure	Commitments	$S1_{P1}$ COM
Total capital expenditure	Commitments	$S2_{P1}$ COM
Total debt repayment expenditure	Commitments	$S3_{P1}$ COM
Total expenditure for third-party services	Commitments	$S4_{P1}$ COM
Current expenditure	Cash payments (CY)	$S1_{P1}$ PAY CY
Capital expenditure	Cash payments (CY)	$S2_{P1}$ PAY CY
Debt repayment expenditure	Cash payments (CY)	$S3_{P1}$ PAY CY
Expenditure for third-party services	Cash payments (CY)	$S4_{P1}$ PAY CY
Current expenditure	Cash payments (PY)	$S1_{P1}$ PAY PY
Capital expenditure	Cash payments (PY)	$S2_{P1}$ PAY PY
Debt repayment expenditure	Cash payments (PY)	$S3_{P1}$ PAY PY
Expenditure for third-party services	Cash payments (PY)	$S4_{P1}$ PAY PY
Revenue from third-party services	Revenue recognitions	$E1_{P1}$ REC
Revenue from new borrowing	Revenue recognitions	$E2_{P1}$ REC
Revenue from asset disposals, capital transfers and credit repayments	Revenue recognitions	$E3_{P1}$ REC
Revenue from current grants and transfers	Revenue recognitions	$E4_{P1}$ REC
Non-tax revenues	Revenue recognitions	$E5_{P1}$ REC
Tax revenues	Revenue recognitions	$E6_{P1}$ REC
Revenue from third-party services	Cash collections (CY)	$E1_{P1}$ COL CY
Revenue from new borrowing	Cash collections (CY)	$E2_{P1}$ COL CY
Revenue from asset disposals, capital transfers and credit repayments	Cash collections (CY)	$E3_{P1}$ COL CY
Revenue from current grants and transfers	Cash collections (CY)	$E4_{P1}$ COL CY
Non-tax revenues	Cash collections (CY)	$E5_{P1}$ COL CY
Tax revenues	Cash collections (CY)	$E6_{P1}$ COL CY
Revenue from third-party services	Cash collections (PY)	$E1_{P1}$ COL PY
Revenue from new borrowing	Cash collections (PY)	$E2_{P1}$ COL PY
Revenue from asset disposals, capital transfers and credit repayments	Cash collections (PY)	$E3_{P1}$ COL PY
Revenue from current grants and transfers	Cash collections (PY)	$E4_{P1}$ COL PY
Non-tax revenues	Cash collections (PY)	$E5_{P1}$ COL PY
Tax revenues	Cash collections (PY)	$E6_{P1}$ COL PY

Notes: COM = Commitments; PAY = Cash Payments; REC = Revenue Recognitions; COL = Cash Collections; CY = Current Year (Accrual basis accounting); PY = Previous Years (Outstanding balances management).

Table 16: BDAP Dataset Features Summary Period 2 (2016–2023)

Feature Name	**Group**	**Label**
Closure of treasury advances	Commitments	$S1_{P2}$ COM
Debt repayment	Commitments	$S2_{P2}$ COM
Current expenditure	Commitments	$S3_{P2}$ COM
Capital expenditure	Commitments	$S4_{P2}$ COM
Expenditure for financial asset increases	Commitments	$S5_{P2}$ COM
Expenditure for third-party accounts and clearing entries	Commitments	$S6_{P2}$ COM
Closure of treasury advances	Cash payments (CY)	$S1_{P2}$ PAY CY
Debt repayment	Cash payments (CY)	$S2_{P2}$ PAY CY
Current expenditure	Cash payments (CY)	$S3_{P2}$ PAY CY
Capital expenditure	Cash payments (CY)	$S4_{P2}$ PAY CY
Expenditure for financial asset increases	Cash payments (CY)	$S5_{P2}$ PAY CY
Expenditure for third-party accounts and clearing entries	Cash payments (CY)	$S6_{P2}$ PAY CY
Closure of treasury advances	Cash payments (PY)	$S1_{P2}$ PAY PY
Debt repayment	Cash payments (PY)	$S2_{P2}$ PAY PY
Current expenditure	Cash payments (PY)	$S3_{P2}$ PAY PY
Capital expenditure	Cash payments (PY)	$S4_{P2}$ PAY PY
Expenditure for financial asset increases	Cash payments (PY)	$S5_{P2}$ PAY PY
Expenditure for third-party accounts and clearing entries	Cash payments (PY)	$S6_{P2}$ PAY PY
New borrowing	Revenue recognitions	$E1_{P2}$ REC
Treasury advances	Revenue recognitions	$E2_{P2}$ REC
Tax, contributory and equalisation current revenues	Revenue recognitions	$E3_{P2}$ REC
Revenue from reduction of financial assets	Revenue recognitions	$E4_{P2}$ REC
Non-tax revenues	Revenue recognitions	$E5_{P2}$ REC
Capital revenues	Revenue recognitions	$E6_{P2}$ REC
Revenue for third-party accounts and clearing entries	Revenue recognitions	$E7_{P2}$ REC
Current transfers	Revenue recognitions	$E8_{P2}$ REC
New borrowing	Cash collections (CY)	$E1_{P2}$ COL CY
Treasury advances	Cash collections (CY)	$E2_{P2}$ COL CY
Tax, contributory and equalisation current revenues	Cash collections (CY)	$E3_{P2}$ COL CY
Revenue from reduction of financial assets	Cash collections (CY)	$E4_{P2}$ COL CY
Non-tax revenues	Cash collections (CY)	$E5_{P2}$ COL CY
Capital revenues	Cash collections (CY)	$E6_{P2}$ COL CY
Revenue for third-party accounts and clearing entries	Cash collections (CY)	$E7_{P2}$ COL CY
Current transfers	Cash collections (CY)	$E8_{P2}$ COL CY
New borrowing	Cash collections (PY)	$E1_{P2}$ COL PY
Treasury advances	Cash collections (PY)	$E2_{P2}$ COL PY
Tax, contributory and equalisation current revenues	Cash collections (PY)	$E3_{P2}$ COL PY
Revenue from reduction of financial assets	Cash collections (PY)	$E4_{P2}$ COL PY
Non-tax revenues	Cash collections (PY)	$E5_{P2}$ COL PY
Capital revenues	Cash collections (PY)	$E6_{P2}$ COL PY

Continued on next page...

Table 16 – Continued from previous page

Feature Name	Group	Label
Revenue for third-party accounts and clearing entries	Cash collections (PY)	$E7_{P2}$ COL PY
Current transfers	Cash collections (PY)	$E8_{P2}$ COL PY

Notes: COM = Commitments; PAY = Cash Payments; REC = Revenue Recognitions; COL = Cash Collections; CY = Current Year (Accrual basis accounting); PY = Previous Years (Outstanding balances management).

Table 17: ISTAT Dataset Features Summary

Feature Name	Label	Description
Population	P	Total resident population of the municipality at the reference year
North-West	$Z1$	Binary indicator equal to 1 if the municipality is located in the North-West macro-area (Liguria, Lombardy, Piedmont, Aosta Valley)
North-East	$Z2$	Binary indicator equal to 1 if the municipality is located in the North-East macro-area (Emilia-Romagna, Friuli-Venezia Giulia, Trentino-Alto Adige, Veneto)
Centre	$Z3$	Binary indicator equal to 1 if the municipality is located in the Centre macro-area (Lazio, Marche, Tuscany, Umbria)
South	$Z4$	Binary indicator equal to 1 if the municipality is located in the South macro-area (Abruzzo, Basilicata, Calabria, Campania, Molise, Apulia)
Islands	$Z5$	Binary indicator equal to 1 if the municipality is located in the Islands macro-area (Sardinia, Sicily)

Notes: Geographic macro-areas follow the ISTAT classification of Italian regions.

The North-West ($Z1$) macro-area serves as the baseline category in models including geographic dummies.

Table 18: Treasury advances Dataset Features Summary

Feature Name	Label	Description
Average duration of treasury advance	$A1$	Average number of days between the opening and closure of a treasury advance spell within the fiscal year
Average treasury advance per capita	$A2$	Average amount of treasury advances requested by the municipality during the year, divided by resident population
Maximum cumulative advance per capita	$A3$	Maximum outstanding advance exposure during the year, divided by resident population
Utilisation rate	$A4$	Treasury advances drawn as a percentage of the maximum statutory advance amount the municipality is entitled to request
Total treasury advance per capita	$A5$	Total amount of treasury advances requested by the municipality during the year, divided by resident population

Table 19: Indicators Dataset Features Summary

Feature Name	Label	Description
Quality of budgetary surplus/deficit	$R1$	Available budgetary surplus or deficit (net of the doubtful debt fund – FCDE – and other mandatory surplus provisioning) divided by current expenditure commitments and medium-long-term debt service commitments.
Reliability of budgetary surplus/deficit	$R2$	Outstanding own-revenue arrears (residui) divided by the absolute value of the available budgetary surplus or deficit; measures the extent to which the surplus or deficit figure is backed by uncollected assessments.
Aged own-revenue arrears ratio	$R3$	Own-revenue arrears outstanding for more than 12 months (residui) divided by current-year own-revenue assessments; measures chronic weakness in collection capacity on tax revenues and tariffs and fees.
Structural current account balance	$R4$	Current revenue assessments (net of the FCDE accrual) divided by current expenditure and medium-long-term debt service commitments; the primary indicator of recurrent fiscal balance.

Continued on next page

Table 19 continued from previous page

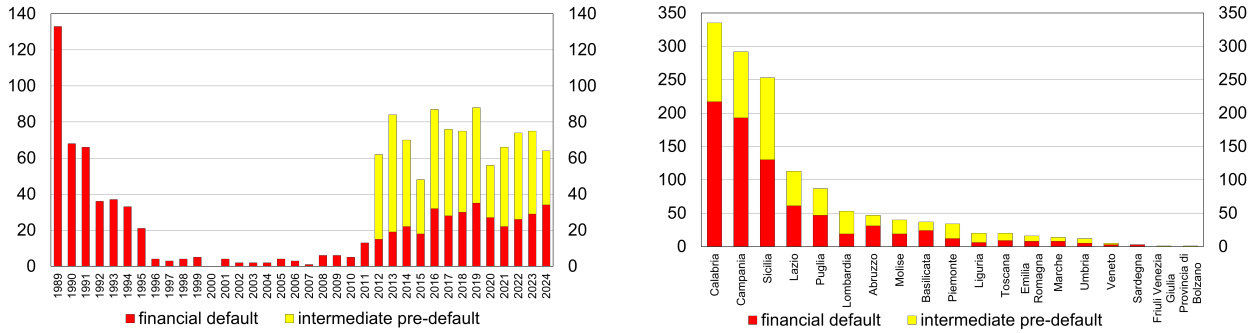
Feature Name	Label	Description
Personnel expenditure rigidity	$R5$	Personnel expenditure commitments divided by current revenue assessments; measures the share of current revenues absorbed by this rigid expenditure component.
Debt sustainability ratio	$R6$	Interest expenditure divided by current revenue assessments; monitors compliance with the statutory borrowing ceiling (debt burden limit) under Arts. 203–204 TUEL.
Theoretical debt repayment period	$R7$	Outstanding consolidated debt at year-end divided by annual principal repayment commitments on medium-long-term loans; measures the theoretical number of years to full debt extinction.
Liquidity advance utilisation ratio	$R8$	Fund for liquidity advances (FAL) revenue recognised (revenue title 7 under the harmonised accounting system) divided by current revenue assessments; measures recourse to State liquidity advances relative to own revenues.
Unreimbursed liquidity advances ratio	$R9$	Residual FAL liabilities outstanding at year-end (expenditure title 5 under the harmonised accounting system) divided by current revenue assessments; flags failure to reimburse liquidity advances within the statutory deadline.
Current expenditure arrears ratio	$R10$	Total outstanding payables on current expenditure (both CY and RY, including trade/commercial debts) divided by total current expenditure commitments; measures the backlog of unpaid current obligations.

Notes: CY = Current Year (Accrual basis accounting); PY = Previous Years (Outstanding balances management).

Indicators R1–R10 constitute the financial distress scoring system developed by the Italian Court of Auditors.

Figures

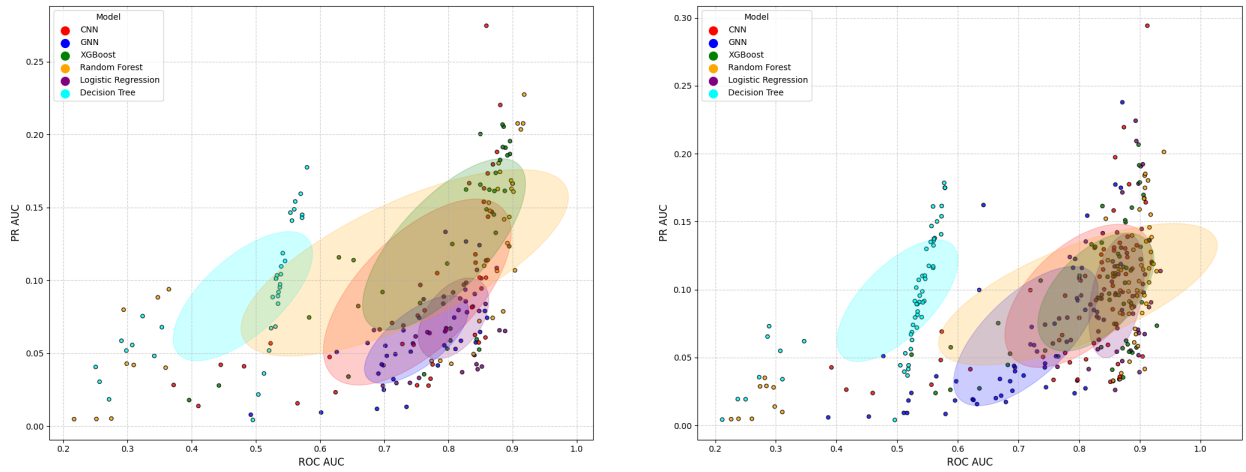
Figure 1: Number of financial distress procedures in Italian municipalities (1989–2024)



(a) By year (b) By region

Notes: Panel (a) shows the annual number of new *dissesto* (financial default) and *pre-dissesto* (intermediate pre-default) filings. Panel (b) reports the total cumulative count by region over the full period. Source: Ca' Foscari municipal distress database.

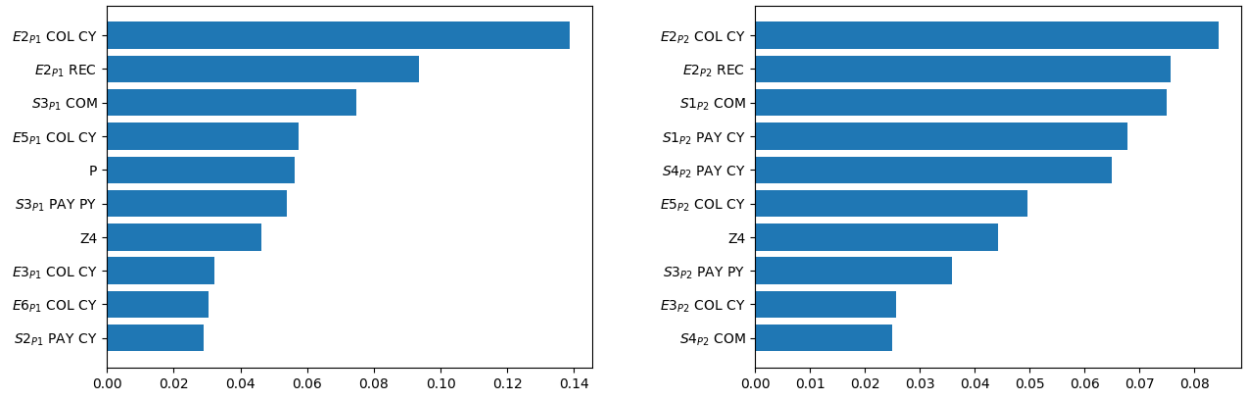
Figure 2: Joint distribution of ROC AUC and PR AUC by model and period



(a) Period 1 (2009–2015) (b) Period 2 (2016–2023)

Notes: Each point represents a single experimental configuration (dataset, data depth, target depth). Ellipses summarise the one-standard-deviation dispersion of the joint (ROC AUC, PR AUC) distribution for each model. They should be interpreted as descriptive dispersion regions rather than formal confidence regions. Models located in the upper-right region exhibit simultaneously high discrimination and high precision-recall performance.

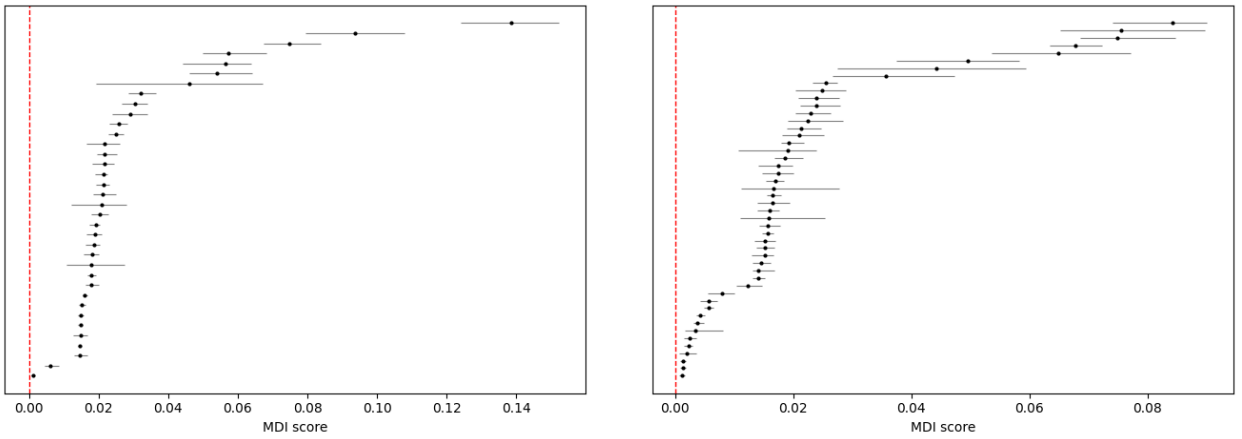
Figure 3: Mean Decrease Impurity (MDI) feature importance - Full BDAP dataset, top-10 features



(a) Period 1 (2009–2015) (b) Period 2 (2016–2023)

Notes: The figure reports the ten features with the highest average MDI score for the Random Forest trained on the Full BDAP dataset, averaged across all time steps. Feature codes follow the Italian municipal budget classification: See Tables 15-19 for variable notation. Higher MDI values indicate a greater contribution to reducing node impurity across the ensemble.

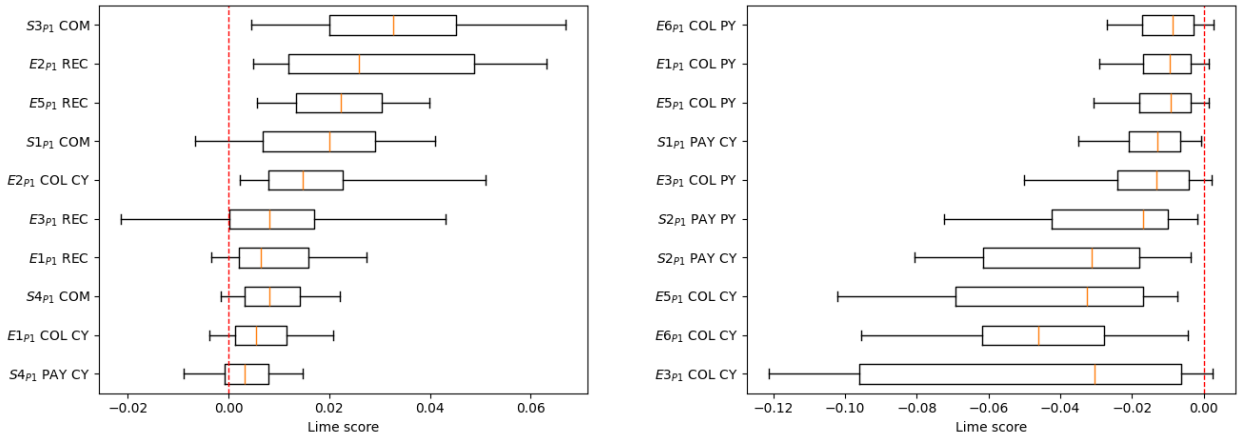
Figure 4: Mean Decrease Impurity (MDI) feature importance - Full BDAP dataset, all features



(a) Period 1 (2009–2015) (b) Period 2 (2016–2023)

Notes: The figure reports MDI scores for the full set of features available in the Full BDAP dataset, averaged across all time steps. Each bar corresponds to a distinct budgetary category. The distribution illustrates the relative concentration of predictive importance across revenue and expenditure titles within each accounting period. See Tables 15-19 for variable notation.

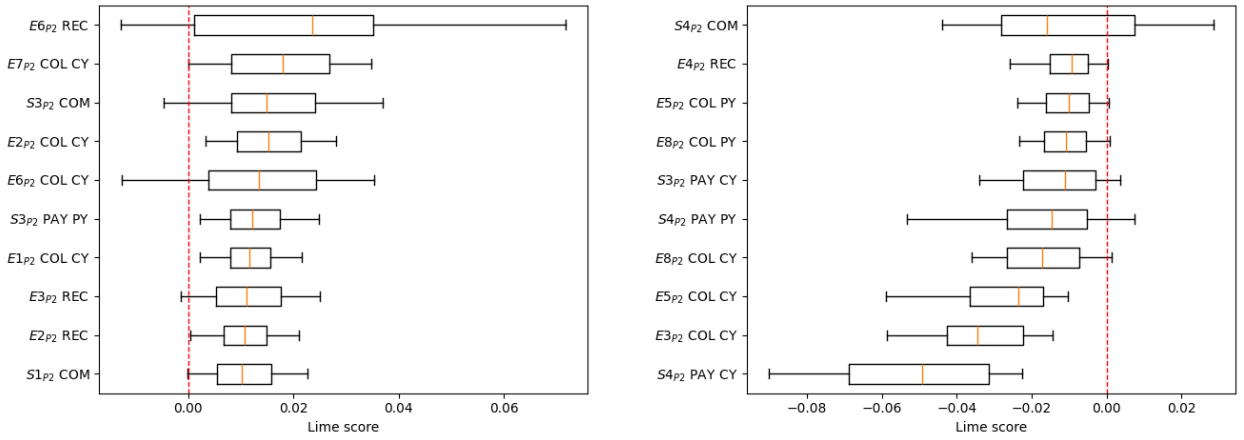
Figure 5: LIME feature importance scores - Full BDAP dataset, Period 1 (2009–2015)



(a) Top-10 features (pro-distress) (b) Bottom-10 features (pro-stability)

Notes: Average LIME coefficients from local linear approximations of the CNN's decision boundary, computed over up to 1,000 randomly sampled test observations. Positive values (Panel a) indicate features that push the model toward a distress prediction; by contrast, negative values (Panel b) indicate features associated with fiscal stability. Coefficients are averaged across all sampled points to obtain a representative global-local summary for the Full BDAP dataset in Period 1. See Tables 15-19 for variable notation.

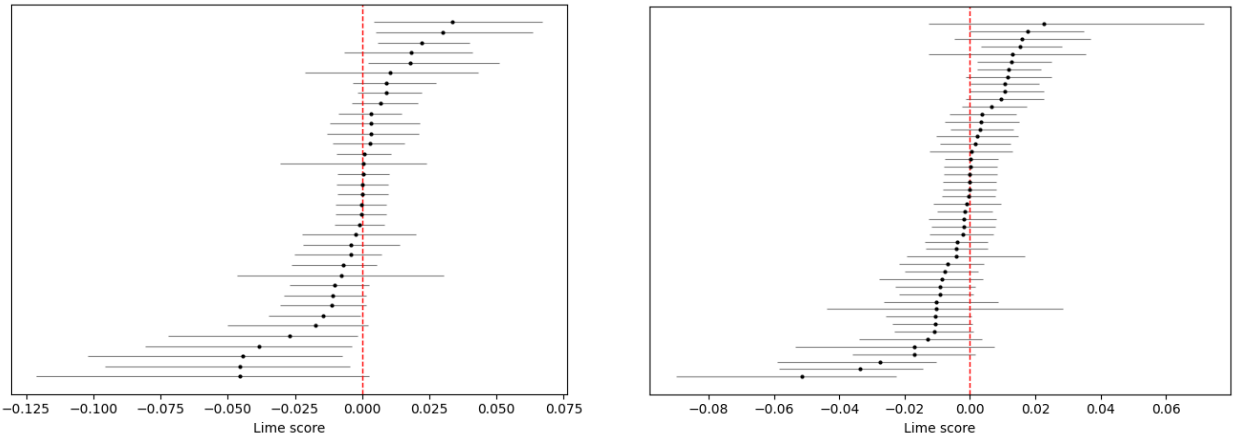
Figure 6: LIME feature importance scores - Full BDAP dataset, Period 2 (2016–2023)



(a) Top-10 features (pro-distress) (b) Bottom-10 features (pro-stability)

Notes: Average LIME coefficients from local linear approximations of the CNN's decision boundary, computed over up to 1,000 randomly sampled test observations. Positive values (Panel a) indicate features that push the model toward a distress prediction; by contrast, negative values (Panel b) indicate features associated with fiscal stability. Coefficients are averaged across all sampled points to obtain a representative global-local summary for the Full BDAP dataset in Period 2. See Tables 15-19 for variable notation.

Figure 7: Average LIME feature importance scores - Full BDAP dataset, all features



(a) Period 1 (2009–2015) (b) Period 2 (2016–2023)

Notes: Average LIME coefficients from local linear approximations of the CNN's decision boundary, computed over up to 1,000 randomly sampled test observations. Positive values indicate features that push the model toward a distress prediction; by contrast, negative values indicate features associated with fiscal stability. Coefficients are averaged across all sampled points to obtain a representative global-local summary for the FullBDAP dataset in each period. See Tables 15-19 for variable notation.

A Descriptive evidence on Treasury advances

This appendix provides descriptive evidence on the evolution of per capita treasury advances across Italian municipalities over the period 2008–2024. Treasury advances are short-term liquidity instruments granted by appointed treasury banks to municipalities when cash inflows are temporarily insufficient to meet payment obligations. Their systematic use serves as an indicator of fiscal strain and structural cash-flow imbalances in local public finance.

A.1 Temporal dynamics and macroeconomic context

The trajectory of treasury advances over the past two decades closely mirrors the succession of macroeconomic shocks and institutional reforms that have shaped Italian local public finance. Four distinct phases can be identified.

During 2008–2011, the initial impact of the global financial crisis and the progressive tightening of the Internal Stability Pact constrained local spending and investment. Despite the severity of the downturn, reliance on treasury advances remained relatively moderate at the national level (average of 518 euro per capita; see Table 20).

The period 2012–2015, marked by the intensification of fiscal austerity and the introduction of tighter balanced-budget rules following the Fiscal Compact, saw a generalised increase in treasury advances across most regions. As state transfers were sharply curtailed and municipalities faced rigid expenditure ceilings, short-term borrowing became a widespread coping mechanism. The national per capita average rose to 525 euro, with particularly pronounced growth in large urban centres (over 250,000 inhabitants), where the average jumped from 423 to 654 euro (see Table 21).

Between 2016 and 2019, despite the gradual relaxation of fiscal rules and the formal abolition of the Domestic Stability Pact (replaced by the more flexible fiscal balance rule), the use of treasury advances remained elevated and, in several regions, continued to grow. This persistence suggests that the dependence on short-term liquidity had acquired a structural character, rooted in chronic cash-flow mismatches, delayed revenue collection, particularly from property taxes and local service charges, and the backlog of payables accumulated during the austerity years. At the national level, the average stood at 505 euro per capita.

The most recent period, 2020–2024, introduced a new layer of heterogeneity. The COVID-19 pandemic and the subsequent implementation of the National Recovery and Resilience Plan (PNRR) produced divergent effects across territories. Extraordinary state transfers and liquidity support measures temporarily alleviated the financial pressure on some municipalities, contributing to a modest national decline in per capita advances (480 euro). However, other administrations, particularly those managing large and complex PNRR-funded projects, experienced renewed peaks in advances, driven by timing mismatches between expenditure commitments and reimbursement flows, as well as by cost overruns linked to inflationary pressures in the construction sector.

A.2 Geographical Heterogeneity

Table 20 shows significant heterogeneity across regions and macroareas. Over the entire 2008–2024 horizon, per capita treasury advances are highest in the South and Islands (555 euro), reflecting the structural fragility of local public finances in these areas, characterised by weaker fiscal capacity, higher dependence on state transfers, and lower administrative efficiency. North-Eastern municipalities display the lowest reliance on advances (300 euro per capita), consistent with their historically stronger fiscal management and more diversified revenue bases. Notably, the South and Islands exhibit a persistent declining trend, from 609 euro in 2008–2011 to 517 euro in 2016–2019, though the pace of decline slows markedly in the most recent period (504 euro in 2020–2024). This trajectory may partially reflect the gradual dissipation of the most acute phase of fiscal stress in Southern municipalities, though it could also be associated with changes in reporting behaviour and institutional oversight. By contrast, the North-West shows a broadly increasing profile through 2016–2019, possibly reflecting the growing fiscal pressure on medium-to-large municipalities in industrial regions as public investment in infrastructure and services expanded relative to available liquidity.

A.3 Heterogeneity by municipality size

Table 21 reveals a non-monotonic relationship between municipality size and per capita treasury advances. Small municipalities (below 5,000 inhabitants) record the highest per capita advances in the first three sub-periods, reflecting their structurally limited revenue-raising capacity, higher dependence on state transfers, and reduced administrative capacity to manage cash flows effectively. However, their reliance has declined over time, falling from 570 euro in 2008–2011 to 509 euro in 2020–2024. A markedly different dynamic characterises the largest municipalities (over 250,000 inhabitants), which exhibit a steep increase in per capita advances from 423 euro in 2008–2011 to a peak of 871 euro in 2016–2019, before partially declining to 703 euro in 2020–2024. This pattern likely reflects the growing complexity of large urban financial management, the scale of investment commitments undertaken, and the difficulty of aligning expenditure and revenue flows in environments characterised by large and diversified public service delivery. The partial reduction in the final period may be attributable to the extraordinary liquidity injections associated with COVID-19 emergency measures and PNRR advances, which disproportionately benefited large municipalities managing major investment programmes. Medium-sized municipalities (20,001–60,000 inhabitants) display the most favourable trend, with per capita advances declining steadily from 490 to 357 euro over the full horizon, suggesting improved financial management and a growing capacity to align revenue collection with expenditure commitments in this size bracket.

Table 20: Per capita total Treasury advances by region (2008–2024)

Region/Area	2008–2011	2012–2015	2016–2019	2020–2024	2008–2024
Italy	**518**	**525**	**505**	**480**	**511**
North-West	435	519	583	554	524
North-East	268	330	305	237	300
Centre	429	485	468	398	454
South and Islands	609	578	517	504	555
Piedmont	552	770	846	797	745
Aosta Valley	507	222	419	28	318
Lombardy	315	278	282	246	283
Province of Trento	68	295	224	160	233
Province of Bolzano	131	131	103	81	116
Veneto	243	396	398	294	345
Friuli-Venezia Giulia	272	124	181	96	182
Liguria	291	491	597	589	519
Emilia-Romagna	322	300	263	243	291
Tuscany	346	456	433	340	410
Umbria	461	583	537	297	495
Marche	335	402	393	200	366
Lazio	492	503	486	469	489
Abruzzo	433	552	530	521	516
Molise	470	503	432	262	451
Campania	527	515	384	415	463
Apulia	652	636	405	429	541
Basilicata	758	705	736	398	681
Calabria	784	628	588	556	647
Sicily	643	595	645	555	610
Sardinia	400	520	431	494	466

Notes: This table reports the average values of per capita total treasury advances requested by municipalities in each Italian region across four distinct periods: 2008–2011 (Global crisis and domestic fiscal tightening), 2012–2015 (Austerity intensification), 2016–2019 (Fiscal rules relaxation and accounting harmonization), and 2020–2024 (Pandemic and PNRR liquidity transfers). The Italy row provides the national average.

Table 21: Average per capita amount of total Treasury advances by population (2008–2024)

Population	2008–2011	2012–2015	2016–2019	2020–2024	2008–2024
0–5,000	570	564	569	509	557
5,001–20,000	457	431	399	427	427
20,001–60,000	490	474	391	357	436
60,001–250,000	660	636	530	533	599
Over 250,000	423	654	871	703	688

B A Brief Survey of the Prediction Models

B.1 Logistic Regression (Logit)

Logistic Regression serves as the fundamental linear baseline for binary classification. It models the conditional probability that an observation i belongs to the target class ($y_i = 1$) as a function of the feature vector X_i through the logistic (sigmoid) function:

$$P(y_i = 1|X_i) = \sigma(\beta^T X_i) = \frac{1}{1 + \exp(-(\beta_0 + \sum_{j=1}^{n} \beta_j X_{ij}))} \tag{1}$$

The parameter vector β is estimated by maximizing the log-likelihood function or, equivalently, minimizing the binary cross-entropy loss:

$$\mathcal{L}(\beta) = -\sum_{i=1}^{N} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \tag{2}$$

with $\hat{y}_i := P(y_i = 1 \mid x_i)$. From an econometric perspective, the primary utility of the Logit model lies in the direct interpretability of the coefficients β_j. These estimates represent the marginal change in the log-odds of the outcome variable associated with a unit change in the corresponding regressor, assuming a linear relationship in the log-odds space.

B.2 Decision Trees and Random Forests

Decision trees perform recursive binary partitioning of the feature space to minimize heterogeneity within terminal nodes. At each node n, the algorithm selects the specific feature j and splitting threshold t_n that maximize the reduction in Gini impurity, ΔI_n:

$$\Delta I_n = I_n - \left(\frac{N_{n,L}}{N_n} I_{n,L} + \frac{N_{n,R}}{N_n} I_{n,R} \right) \tag{3}$$

where $I_n = 1 - \sum p_k^2$ is the impurity at node n, and L, R subscripts denote the resulting left and right child nodes. While intuitive, single trees suffer from high variance.

Random Forests mitigate this by utilizing an ensemble of B independent trees. Each tree $\hat{f}_b$ is grown on a bootstrap sample of the training data (Bagging), and at each split, only a random subset of features is considered. For classification, the final ensemble prediction is obtained by averaging probabilities across trees:

$$\hat{f}_{rf}(x) = \frac{1}{B} \sum_{b=1}^{B} P_b(y = 1 \mid x) \tag{4}$$

This aggregation strategy reduces the variance of the estimator and allows the model to capture complex non-linear interactions and discontinuities in the feature space that linear models typically fail to identify.

B.3 XGBoost (Extreme Gradient Boosting)

XGBoost is a gradient boosting framework that constructs an ensemble of trees sequentially rather than independently. Each new tree f_t is trained to predict the residual errors of the existing ensemble, thereby iteratively reducing bias. The objective function at iteration t combines a differentiable convex loss function l and a regularization term Ω to penalize model complexity:

$$\text{Obj}^{(t)} = \sum_{i=1}^{N} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \tag{5}$$

where the regularization term is defined as:

$$\Omega(f_t) = \gamma L + \frac{1}{2}\lambda |w|^2, \tag{6}$$

with L denoting the number of leaves and w the vector of leaf weights. By utilizing a second-order Taylor expansion of the loss function, XGBoost optimizes split points efficiently. Furthermore, its sparsity-aware split finding algorithm makes it particularly robust when handling high-dimensional data with significant missing values or zero-inflated features.

B.4 Convolutional Neural Networks (CNN)

Convolutional Neural Networks are designed to process data with a grid-like topology. In the context of panel data or time-series analysis, 1D convolutions can be employed to extract temporal patterns from sequential inputs. An observation is represented as a matrix of shape $T \times F$, where T denotes the time steps and F the number of features per step.

A convolutional layer applies a set of learnable filters (kernels) w to the input sequence to produce feature maps:

$$h_j^l = \sigma\left(\sum_{c=1}^{F_{l-1}} (x_c^{l-1} * w_{c,j}^l) + b_j^l\right) \tag{7}$$

where $*$ denotes the convolution operation. By stacking multiple layers, the network learns a hierarchy of temporal features. Unlike standard autoregressive models which rely on predefined lags, CNNs automatically learn to detect relevant sequential dependencies—such as acceleration or volatility shifts—directly from the raw multivariate time series.

B.5 Graph Neural Networks (GNN)

Graph Neural Networks generalize deep learning operations to non-Euclidean data structures, making them suitable for modeling systems characterized by relational dependencies or network effects. The data is represented as a graph $\mathcal{G} = (V, E)$, where nodes V represent entities and edges E represent their relationships (e.g., spatial proximity, trade links).

The GNN updates the hidden state h_v of each node by aggregating information from its

local neighborhood $\mathcal{N}(v)$:

$$h_v^{(l+1)} = \text{UPDATE}^{(l)}\left(h_v^{(l)}, \text{AGGREGATE}^{(l)}\left(\{h_u^{(l)}, \forall u \in \mathcal{N}(v)\}\right)\right) \tag{8}$$

This formulation follows the message-passing neural network (MPNN) framework. The aggregation step allows the model to capture spatial spillovers and peer effects endogenously. In economic terms, this provides a non-linear, high-dimensional alternative to spatial autoregressive (SAR) or spatial error (SEM) models, capable of distinguishing between idiosyncratic shocks and systemic contagion.

C Interpretability methods

C.1 MDI

Mean Decrease Impurity (MDI), also referred to as *Gini importance*, is a model-specific interpretability technique for tree-based ensembles. MDI quantifies the importance of a feature j by calculating the sum of all impurity decreases in the forest due to that feature, weighted by the proportion of samples reaching the nodes where the splits occur. Formally, for an ensemble consisting of a collection of trees $T = \{t_1, \ldots, t_K\}$, the importance of feature j is:

$$\mathrm{MDI}_j = \frac{1}{K} \sum_{k=1}^{K} \sum_{n \in t_k : v(n) = j} p(n) \Delta I(n, t_n) \tag{9}$$

where $v(n)$ is the feature used to split node n, $p(n)$ is the fraction of samples reaching node n, and $\Delta I(n, t_n)$ is the decrease in impurity (typically Gini impurity for classification) resulting from the split at node n with threshold t_n. As a global metric, MDI captures the aggregate contribution of each covariate across the entire training set.

C.2 LIME

Local Interpretable Model-agnostic Explanations (LIME) is a technique designed to explain individual predictions of a complex model $f(\cdot)$ by approximating it locally with an interpretable surrogate model $g(\cdot)$. The explanation for a specific observation x is obtained by solving:

$$\xi(x) = \arg\min_{g \in G} \ \mathcal{L}(f, g, \pi_x) + \Omega(g) \tag{10}$$

where G is the class of interpretable models (such as linear regressions), $\mathcal{L}$ is a loss function (e.g., squared loss) that measures how closely g approximates f in the locality of x, and $\Omega(g)$ is a measure of the complexity of the surrogate model. The proximity measure π_x defines the local neighborhood around x, typically implemented via a kernel function on the distance between x and perturbed samples z.

Unlike MDI, LIME is model-agnostic and provides local coefficients that indicate the direction and magnitude of a feature's influence on a specific prediction. This allows for a granular validation of model behavior without assuming that the identified relationships are constant across the entire feature space.

C.3 GNN Methodology

The structural foundation of the GNN is the adjacency matrix A, which defines the topology of information flow between municipalities (see Appendix B.5).

In this paper we construct a geometric graph where nodes V represent municipalities and edges E represent geographic proximity. The pairwise distances are computed using centroid coordinates provided by the Italian National Institute of Statistics (ISTAT).

To define the connection weights, we employ a distance-based thresholding approach. We conjecture that the transmission of financial distress is governed by local economic spillovers rather than long-range interactions. Consequently, a fully dense graph would introduce noise and prohibitive computational costs. We define the sparse weighted adjacency matrix A_{uv} as:

$$A_{uv} = \begin{cases} \text{dist}(u,v)^{-1} & \text{if dist}(u,v) \leq d_{max} \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

We set the threshold $d_{max} = 25$ km. As illustrated in Figure 8, this choice provides a pragmatic balance between graph complexity and information retention. It results in a graph with 587,068 edges across 7,778 municipalities, corresponding to a sparsity ratio of approximately 0.0097 (roughly 1%). The use of inverse distances as weights ($w_{uv} = \text{dist}(u,v)^{-1}$) ensures that the message-passing mechanism prioritizes signals from the immediate geographic vicinity.

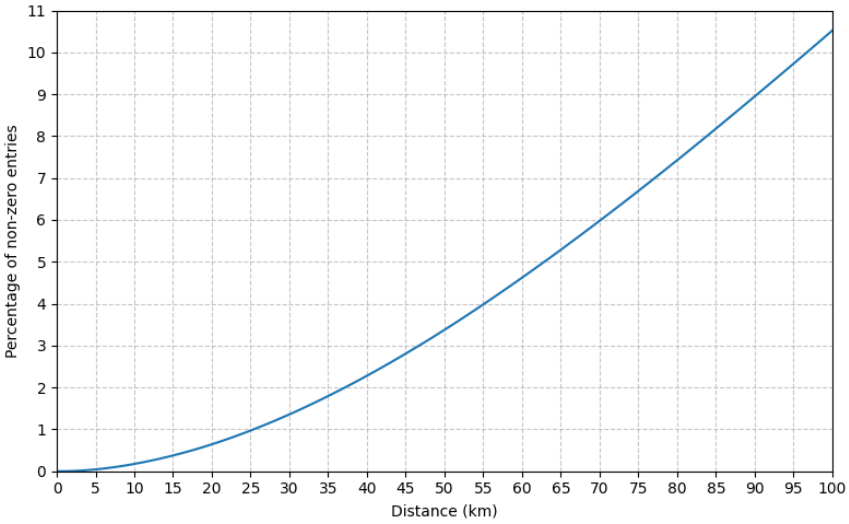


Figure 8: Relationship between graph complexity and the selected distance threshold.

The training optimization strategy for GNNs diverges fundamentally from standard architectures like CNNs due to these relational dependencies. In a CNN, samples are independent, allowing for efficient mini-batch Stochastic Gradient Descent (SGD). In our GNN framework, the hidden state of a municipality h_v depends on the states of its neighbors; essentially, the "receptive field" of a node expands with each layer of message passing.

While graph sampling techniques (e.g., neighbor sampling) exist to enable mini-batching, we observed empirically that they yield inferior classification performance, likely due to the loss of relevant peripheral connections during sampling. Therefore, we utilize Full-Batch

Gradient Descent, processing the entire graph structure ($N = 7,778$) simultaneously. This constraint necessitates a temporal validation strategy—splitting training and test sets by year—rather than a random node-level split, to preserve the integrity of the spatial network during training.

D Glossary of Italian institutional terms and variable labels

Italian term	English equivalent	Definition
Distress Procedures and Legal Framework		
Dissesto finanziario	Financial default	Formal declaration of municipal insolvency under Art. 244–269 TUEL (Legislative Decree 267/2000); triggers the appointment of an Extraordinary Liquidation Body (OSL) to manage pre-existing liabilities.
Predissesto/Riequilibrio finanziario pluriennale	Intermediate pre-default/Multi-year financial rebalancing	Procedure introduced by Decree-Law 174/2012 converted into Law 213/2012, and codified in Arts. 243-bis et seq. TUEL allowing municipalities facing structural imbalances to avoid formal default by adopting a multi-year rebalancing plan (4–20 years), subject to approval by the regional section of the Court of Auditors.
Dissesto guidato	Guided default	A court-mandated insolvency procedure for Italian local authorities (municipalities and provinces) experiencing severe, unaddressed financial instability. Introduced by Legislative Decree No. 149/2011, it functions as a coercive mechanism where central state authorities step in to force a local government into a formal declaration of financial distress when it fails to address its structural budget imbalances.
Deficit strutturale	Structural deficit	A condition of persistent structural imbalance in municipal finances, monitored through the system of fiscal stress indicators.

continued on next page

continued from previous page

Italian Term	English Equivalent	Definition
Ente strutturalmente deficitario	Structurally deficient entity	A municipality identified, under Arts. 242–243 TUEL, as having at least four of the fiscal stress indicators defined by the Ministry of the Interior above the relevant threshold values; representing a less severe level of financial difficulty than predissesto or dissesto, it entails central control on staffing and recruitment and a requirement to finance individual-demand, water, and waste-collection services at least partly through tariffs.
Bilancio stabilmente riequilibrato	Structurally balanced budget	The multi-year recovery plan that a municipality in dissesto must obtain approval for before resuming normal financial reporting.
Organo Straordinario di Liquidazione (OSL)	Extraordinary Liquidation Body	Body appointed upon a dissesto declaration to manage and liquidate the stock of pre-existing liabilities, while the municipal administration retains ordinary governance functions.
Parametri obiettivo	Fiscal stress indicators / Early-warning thresholds	System of eight financial indicators (introduced by the Ministerial Decree of 28 December 2018) used to identify municipalities in structural deficit; a municipality is flagged when at least four indicators exceed their respective thresholds.

continued on next page

continued from previous page

Italian Term	English Equivalent	Definition
Testo Unico degli Enti Locali (TUEL)	Consolidated Law on Local Authorities	Legislative Decree 267/2000; the primary legislative framework governing the organisation, functioning, and financial management of Italian local authorities, including the rules on financial distress.
Budget Documents and Accounting Cycle		
Rendiconto della gestione	Management accounts / outturn accounts	The final annual financial statement submitted by municipalities after fiscal year closure, reporting actual revenues and expenditures at each stage of the financial cycle (accertamento, riscossione, impegno, pagamento).
Bilancio di previsione	Budget appropriation / Expenditures and revenues forecast	The annual budget document prepared before the reference fiscal year, reporting planned revenues and expenditures; the basis for ReadyBDAP indicators.
Principio contabile armonizzato	Harmonised accounting standard	The unified public accounting framework introduced by Legislative Decree 118/2011, mandatory for all local authorities from 2016, replacing the previous DPR 194/1996 framework.
Revenue Concepts		
Accertamento	Revenue recognition	The formal, legally grounded recording of a collectable revenue claim in the management accounts; the first stage of the municipal revenue cycle under both pre- and post-harmonisation frameworks. In ReadyBDAP, refers to planned revenue recognitions as reported in the bilancio di previsione.

continued on next page

continued from previous page

Italian Term	English Equivalent	Definition
Riscossione	Cash collection	The actual receipt of funds corresponding to previously recognised revenue claims; the second and final stage of the revenue cycle. The gap between recognition and collection, the so-called revenue arrears, is a key indicator of revenue collection efficiency.
Entrate tributarie	Tax revenues	Revenues from municipal taxes, including the property tax (IMU) and the municipal surcharge on personal income tax (addizionale comunale IRPEF); classified under revenue title E6 in the pre-harmonisation framework.
Entrate extratributarie	Non-tax revenues	Revenues from fees, charges, and other non-tax sources; classified under revenue title E5 in the pre-harmonisation framework.
Entrate da contributi e trasferimenti correnti	Current transfer revenues	Revenues from current grants and transfers from other public administrations, including central government transfers (trasferimenti erariali) and the Municipal Solidarity Fund (FSC).
Entrate derivanti da accensione di prestiti	Loan revenues	Revenues from new borrowing (bond issuance, loans from CDP and other lenders); classified under revenue title E2 in the pre-harmonisation framework. Under the pre-harmonisation framework, this title also included treasury advances, making it impossible to isolate short-term from long-term borrowing without manual data intervention.

continued on next page

continued from previous page

Italian Term	English Equivalent	Definition
Entrate derivanti da alienazione, trasferimenti di capitale e riscossioni di crediti	Capital transfer revenues	Revenues from asset disposals, capital grants, and credit repayments; classified under revenue title E3 in the pre-harmonisation framework.
Anticipazioni da istituto tesoriere/cassiere	Treasury advances	Short-term credit lines provided by the appointed treasury bank to cover temporary cash shortfalls; classified under revenue title E2 in the post-harmonisation framework. Persistent or repeated recourse is interpreted by the Italian Court of Auditors as symptomatic of structural fiscal stress.
Expenditure Concepts		
Impegno	Expenditure commitment	The formal, legally binding recording of a public liability in the management accounts, corresponding to the first stage of the expenditure cycle. In ReadyBDAP, refers to planned expenditure commitments as reported in the *bilancio di previsione*.
Pagamento	Cash payment	The actual disbursement of funds corresponding to a previously registered expenditure commitment; the final stage of the expenditure cycle.
Spese correnti	Current expenditure	Expenditure for the ordinary functioning of municipal services, including personnel costs, procurement of goods and services, and interest payments; classified under expenditure title S1 in the pre-harmonisation framework and title S3 in the post-harmonisation framework.

continued on next page

continued from previous page

Italian Term	English Equivalent	Definition
Spese in conto capitale	Capital expenditure	Expenditure for investment purposes, including public works, asset acquisitions, and capital transfers; classified under expenditure title S2 in the pre-harmonisation and S4 in the post-harmonisation framework. A sustained reduction in capital payments is identified by the models as an early indicator of fiscal deterioration.
Spese per rimborso di prestiti	Debt repayment	Expenditure for the repayment of principal on outstanding loans and bonds; classified under expenditure title S3 in the pre-harmonisation framework. Under the pre-harmonisation framework, this title also included the repayment of treasury advances.
Chiusura anticipazioni ricevute	Closure of treasury advances	The repayment of treasury advances drawn during the fiscal year; under the harmonised framework (Legislative Decree 118/2011), advances must be fully extinguished before year-end and are classified under expenditure title S1.
Spese per il personale	Personnel expenditure	Expenditure for municipal employees' salaries, social contributions, and related costs; one of the main components of current expenditure and a key indicator of budget rigidity (rigidità della spesa).

continued on next page

continued from previous page

Italian Term	English Equivalent	Definition
Interessi passivi	Interest payments	Expenditure for interest due on outstanding debt obligations; used in the computation of the debt sustainability indicator (saturazione dei limiti di indebitamento, R6).
Rigidità della spesa	Expenditure rigidity	The share of current expenditure absorbed by difficult-to-cut items – personnel costs, debt service, and deficit repayment - relative to current revenue; structurally higher in southern municipalities and a key driver of their vulnerability to fiscal shocks.
Balance Sheet and Residual Concepts		
Residui attivi	revenue arrears / uncollected revenues	Revenues that have been formally recognised in previous fiscal years but not yet collected in cash.
Residui passivi	expenditure arrears / outstanding expenses / Unpaid commitments	Expenditure commitments formally registered in previous fiscal years but not yet paid in cash.
Fondo crediti di dubbia esigibilità (FCDE)	Doubtful Debt Fund	A mandatory provision, introduced under the harmonised accounting framework, set aside to cover the risk that recognised revenues will not be collected; its size is a direct indicator of revenue quality.

continued on next page

continued from previous page

Italian Term	English Equivalent	Definition
Fondo anticipazioni di liquidità (FAL)	Fund for liquidity advances	A mandatory provision, introduced alongside the FCDE under the harmonised accounting framework, set aside for the reimbursement of state-granted cash advances used by municipalities to settle overdue commercial debts; its growth contributed significantly to the widening of budget deficits in southern municipalities.
Avanzo di amministrazione	Available budgetary surplus	The cumulative balance of the municipality's financial accounts, net of provisions and committed funds; a positive balance indicates accumulated fiscal space.
Disavanzo di amministrazione	Budgetary deficit	A negative cumulative balance of the municipality's financial accounts; requires mandatory recovery through multi-year rebalancing plans.
Risultato di amministrazione	Budgetary surplus	The overall balance of the municipality's financial accounts at year-end, net of FCDE provisions and other mandatory accruals; the primary summary indicator of fiscal health under the harmonised framework.
Avanzo disponibile/Disavanzo	Available budgetary surplus / budgetary deficit	The fourth and residual component of the budgetary surplus, calculated as the difference between the total surplus and the sum of surplus provisioning, earmarked surplus, and surplus earmarked for investments; negative values indicate a budgetary deficit requiring recovery.
Treasury and Liquidity		

continued on next page

continued from previous page

Italian Term	English Equivalent	Definition
Anticipazione di tesoreria	Treasury advance	A short-term credit line granted by the appointed treasury bank (istituto tesoriere) to a municipality to address temporary cash shortfalls; must be repaid within the same fiscal year. Its duration, frequency, and amount are key inputs in the SIOPE-based feature set.
Anticipazione di liquidità	Liquidity advance	A longer-term liquidity support instrument provided by central government or CDP to municipalities with accumulated payment arrears, distinct from ordinary treasury advances.
Istituto tesoriere / cassiere	Appointed treasury bank	The financial institution contracted by a municipality to manage its cash flows, disburse payments, and provide treasury advances; its role is regulated under Art. 208–215 TUEL.
Cassa Depositi e Prestiti (CDP)	State savings and loans institution	The state-controlled financial institution that provides long-term loans to local authorities for investment purposes; CDP loans are typically disbursed in tranches tied to work-progress stages, generating physiological lags between accertamento and riscossione.
Institutional Bodies and Data Systems		
Corte dei conti	The Italian Court of Auditors	The supreme audit institution responsible for ex ante and ex post control of public accounts; its regional sections review and approve predissesto rebalancing plans and publish annual reports on the financial conditions of local authorities.

continued on next page

continued from previous page

Italian Term	English Equivalent	Definition
Ragioneria Generale dello Stato (RGS)	State Accounting Office	The central government body responsible for monitoring public finance, managing BDAP, and overseeing the implementation of harmonised accounting standards.
Banca Dati delle Amministrazioni Pubbliche (OpenBDAP)	Public Administrations Accounting Database	The digital platform managed by RGS that centralises budget and management account data for all Italian public administrations; the primary source of balance-sheet features in this paper.
Sistema Informativo delle Operazioni degli Enti Pubblici (SIOPE)	Treasury Operations Information System	A real-time cash monitoring system jointly managed by RGS, AgID, and Banca d'Italia, recording all treasury transactions of Italian public administrations at the individual entity-day level; the source of high-frequency treasury advance data in this paper.