Dipartimento
di Economia
Working Papers

n. 289
2026

Mara Giua,
Francesca Micocci,
Giulia Valeria Sonzogno

ENHANCING IMPLEMENTATION SUCCESS IN COHESION POLICY. A MACHINE LEARNING APPROACH

ISSN 2279-6916 Working Papers

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Dipartimento
di Economia
Working Papers

WORKING PAPERS
Dipartimento di Economia
Università degli Studi Roma Tre
Via Silvio D'Amico, 77 - 00145 Roma
Tel. 0039-06-57335655 fax 0039-06-57335771
workpapers.economia@uniroma3.it
https://economia.uniroma3.it/

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Enhancing Implementation Success in Cohesion Policy. A Machine Learning Approach

Mara Giua[*], Francesca Micocci[†], Giulia Valeria Sonzogno[‡]

February 2026

Abstract

We test the hypothesis that Cohesion Policy (CP) underperformance, measured as projects' delays and (Output Indicators) target failures, is systematically driven by project-level features of the CP implementation architecture rather than by contextual conditions alone: using Italian project-level data (2014–2020) in a Machine Learning approach, we show how governance arrangements in terms of programme type, programmers, activation procedures and beneficiary combine with underlying contextual conditions in predicting projects' outcomes. Successful policy configurations avoiding underperformance can be adopted in an evidence-based perspective by combining some of the existing policy tools and accounting for the socio-economic context upstream.

JEL Classification: O18; R11; R58; C53; C55

Keywords: Cohesion Policy; European Union; Policy implementation; Machine Learning

Funding Acknowledgement: This research has been developed within the PRIN 2022 project *STARGATE –Strengthening TARgeting and Guidance with Actionable and Timely Evidence* (CUP F53D23003220006).

---

[*]Department of Economics and Rossi-Doria Centre, Roma Tre University, Via Silvio D'Amico 77 - 00145 Rome, Italy. mara.giua@uniroma3.it

[†]Department of Economics, Roma Tre University, Via Silvio D'Amico 77 - 00145 Rome, Italy. francesca.micocci@uniroma3.it

[‡]Department of Economics, Roma Tre University, Via Silvio D'Amico 77 - 00145 Rome, Italy. giuliavaleria.sonzogno@uniroma3.it

1 Introduction

The economic impact of the European Union's (EU) Cohesion Policy (CP) hinges on the existing socio-economic conditions of the territories in which it operates and the policy architecture through which it is implemented. While the role of the former has received extensive academic and policy attention, the latter remains comparatively underexplored (see Crescenzi and Giua (2024) for a review).

This paper addresses this gap by reconstructing the architecture of CP implementation at the project level and investigating how it shapes policy (under)performance.

Our analysis relies on project-level data for Italy's 2014–2020 programming period— Italy is the only Member State with open-access project-level data, which allows us to identify around 100 project-level features of policy implementation, alongside municipality-level context variables. We use Machine Learning (ML) algorithms (random forest) to predict projects' underperformance - measured by projects' delays (actual end date delayed with respect to the expected end date) and projects' unsuccess (unreached targets of the projects' output indicators) - and find out how it is associated with these implementation features.

The use of ML, still underutilized in the analysis of CP implementation, makes it possible to achieve a powerful predictive capability, surpassing traditional methods by capturing non-linear relationships in high-dimensional policy data (Chernozhukov et al., 2022). To assist us in the interpretation of the prediction results, once the top 15 predictors of delays and target failures are identified, SHapley Additive exPlanations (SHAP values) by Lundberg et al. (2020) will be computed to understand how each features of the implementation architecture influences underperformance, whether positively or negatively, and whether its effect is linear or non-linear. Then, dedicated prediction exercises will be performed, comparing Southern and Northern Italy[1] and contrasting interventions below €10,000 with those exceeding €100,000. Finally, a series of robustness tests will corroborate the main results. Overall, our analysis makes it possible to identify what project configuration should be avoided — and, conversely, what should be pursued — to enhance policy effectiveness in the implementation phase.

[1]Under the 2014–2020 EU programming period, Italian Southern and Islands regions were classified on the basis of GDP per capita relative to the EU average as either *Less Developed Regions* (LDRs) or *Transition Regions*. Specifically, *Less Developed Regions* (GDP per capita below 75% of the EU average) included Basilicata, Calabria, Campania, Puglia, and Sicilia; while *Transition Regions* (GDP per capita between 75% and 90%) comprised Abruzzo, Molise, and Sardegna. This classification determined both the allocation of CP funds and the intensity of support received.

With high predictive accuracy, our model confirms that underperformance in CP is not random but is a result of how the policy is put into practice: governance arrangements and choices in terms of Programme type (EU vs. Nationally-based), level (National vs. Regional), type of programmers and beneficiary, of activation procedures and the concentration/dispersions of roles and responsibilities, lead to a more or less successful implementation, which turns into higher or lower policy's impacts. Therefore, strengthening implementation mechanisms is no merely a matter of administrative efficiency, but a strategic priority for the EU.

Overall, this paper makes three main contributions. First, it offers the first systematic reconstruction of the architecture of CP implementation at the project level, moving beyond the traditional lens and enriching it, shifting the analytical focus to how projects are delivered. Second, it leverages a comprehensive ML approach that achieves high predictive accuracy to analyse the policy implementation mechanisms. Finally, it contributes to the debate around the shift of EU policies toward a performance-based approach. In particular, our contribution identifies actionable policy configurations at project level to respond to the urgency, complexity, and performance demands of the EU's evolving policy landscape. In doing so, it delivers actionable recommendations for an evidence-based reform of the post-2027 CP.

2 Implementation architecture and project performance: evidence and gap

According to the ongoing negotiations on the forthcoming 2028–2035 EU financial framework, CP will signal a structural shift towards more results-oriented and time-sensitive interventions,[2] following the model launched by the Next Generation EU (NGEU) programme.[3] In this evolving landscape, implementation has gained renewed prominence as a central determinant of overall policy performance.[4]

However, the literature is silent on how implementation choices can lead to successful or poor project outcomes. So far, it has privileged the analysis of the causal impacts achieved by the policy in terms of growth and employment, showing how they are conditioned by socio-economic underlying conditions. In this respect, prominent attention has been paid to the conditioning role played by local institutional quality

[2]European Commission (2025), COM(2025) 571.
[3]European Commission (2020), COM(2020) 456.
[4]European Commission (2022), COM(2022) 518.

and administrative capacity, with evidence confirming how strong administrations improve planning and delivery, while weak bureaucracies undermine outcomes, especially in lagging regions (see Polverari (2023) for a review).

More recently, a growing body of research on CP in Italy focuses on implementation outcomes at the project level, showing how they depend on differences in governance arrangements.[5] A causal analysis focusing on infrastructural projects shows that otherwise identical projects experience significant financial slowdowns when they are implemented in a Nationally-based rather than in an EU CP framework (Celli et al., 2025). Coherent findings are obtained in an ML approach when looking at the EU vs Nation-based CP projects in terms of implementation delay (Coco et al., 2025), and by analyzing the probability of delay for NGEU-like projects (Crescenzi et al., 2021). Focusing on the 2007-2013 policy cycle Del Monte et al. (2022) show that where the quality of institutions is low, "project duration was generally higher when projects were managed by municipalities, the lowest level of government. Projects managed by central government authorities were the shortest, and projects managed by regional and provincial governments were between the two" (Del Monte et al. (2022), p. 2134). In a more recent study, they conclude that on average, a 1 percent increase in administrative capacity shortens project duration by 3 percent (Del Monte et al., 2025).

Although limited and circumstantiated to specific policy subsets, this project-level evidence is particularly relevant to support the idea that CP outcomes depend not only on the underlying context but also on how policy is delivered in practice. These project-level studies integrate a large body of literature that already addressed governance issues of CP implementation from more aggregated perspectives. By comparing the GDP impact of EU vs Nationally-based CP funds Coppola et al. (2020) found that the former is greater than the latter. Crescenzi and Giua (2016) compared top-down vs. bottom-up approaches in EU policies, concluding that in more disadvantaged regions the former lead to better outcomes than the latter. The centralisation vs. decentralisation trade-off in CP has also been extensively addressed, with evidence showing how, when local administrative capacity is limited, decentralised policies are associated with low absorption rates, with effectiveness varying across contexts and policy domains (Santos et al., 2025). In particular, while Regional Operational Programmes (ROPs) managed by Regional Authorities allow for stronger territorial em-

[5]In this stream of literature, the case of Italy is largely represented thanks to the availability of open data on the universe of CP projects as provided by the *OpenCoesione* web portal.

beddedness and adaptation to local needs, National Operational Programmes (NOPs) steered by central administrations rely on more standardised procedures and tighter coordination, which can improve implementation performance, particularly in weaker institutional contexts (OECD, 2020). More recently, the OECD highlights how also who manages and who delivers interventions have a direct effect on implementation outcomes. Beneficiaries differ widely in administrative and compliance capacity, and many—especially at the local level—face persistent difficulties in planning, project design and regulatory compliance, with direct effects on execution (OECD, 2025). A relevant role of project manager characteristics is also highlighted by Bachtrögler-Unger (2024). When a project is managed by multiple beneficiaries, coordination costs can negatively impact on projects outcome and on the final achievement of the measure: by analyzing the impact of a measure funded by the 2007-203 European Regional Development Fund (ERDF) to foster collaborative industrial research in Italy, Crescenzi et al. (2020) found a negative conditioning impact of projects managed by large partnerships on firm's value added and employment. Multiple-beneficiary and multiple-location projects are also associated with a higher probability of delays in the descriptive analysis of Crescenzi et al. (2021). Among CP projects, the small ones benefit from simpler procedures and greater operational flexibility, while the larger ones face more complex procurement requirements, stricter compliance rules, and higher coordination costs.[6] As far as project activation mechanisms, negotiated procedures have been described as capable of emphasising early stakeholder involvement to improve alignment between policy supply and territorial demand, notably within Smart Specialisation Strategies (Foray et al., 2011). By contrast, it has been highlighted how competitive calls, while enhancing transparency and participation, may inadvertently exclude less-organised actors and exacerbate territorial disparities (Barca, 2009). Finally, the overall policy effectiveness turns out to be higher when infrastructure investments are combined with complementary "soft" measures, such as business support and technical support (Cristofoletti et al., 2023).

This paper will identify project-level features accounting for all the implementation dimensions that the above mentioned literature discussed as capable of influencing the policy outcome. It will therefore build the overall architecture of the CP implementation investigating how that drives CP underperformance in terms of delays and unsuccess.

[6] European Commission (2018), COM(2018) 375.

3 Data and Empirical Strategy

Our empirical analysis will identify underperforming projects and the extent to which the features of the CP implementation architecture contribute to the probability of the project delay (actual end date of the project > expected end date) or target failure (failure to meet the targets associated with the project's Output Indicators).

In particular, we adopt ML algorithms that allows flexible modeling of potentially non-linear interactions between a high-dimensional set of predictors and the outcomes of interest (Chernozhukov et al., 2022) to identify the associations between policy outcomes (delays and target failures) and a set of variables including around 100 project-level features capturing different dimensions of the policy implementation architecture, and around 30 municipality-level variables accounting for territorial features and administrative endowment of the municipality the project is located.[7]

We refer to the 2014-2020 CP projects located in Italy and we retain only completed projects, i.e., those with a reported end date, with a single beneficiary. This yields a final dataset of $635,662$ projects, corresponding to a total public investment of €38.6 billion. In the dataset, each observation corresponds to a project i implemented in municipality j.

To start, we identify the project-level variables that build the architecture of the CP implementation (Appendix A).

We select variables that distinguish projects belonging to the European CP from those belonging to the Nationally-based CP, as well as projects belonging to National or Regional Programs. We construct a variable accounting for the industry (e.g., Production, Education, Health) and the type of project programmer (e.g., Ministries, Regional Authorities, Municipalities, Metropolitan Cities, Private Entities). Moving along the chain of the project implementation, we account for the project beneficiary industry (e.g., firms' economic sector) and type (e.g., among public ones: Ministries, Regions, Municipalities; among private ones: Entrepreneurs, Cooperatives, Consortia, Limited Companies, Foundations, Individuals). We consider whether the programmer and the beneficiary coincide or not and the share of the eventual private co-financing. Specific variables account for the projects' activation procedures (e.g., circular, public notice, tenders, negotiated procedures, direct identification). A number of variables account for the project's nature (e.g., infrastructures, incentives, purchase of goods) and theme (e.g., firms' competitiveness, environment, transport,

[7] Project-level data are drawn from the *OpenCoesione* web portal.

labour, inclusion), and the main fund financing the project. When a project involves multiple financial sources/programs/locations is also accounted for. In addition, we include variables on the financial amount of the project, its expected duration, and the year of activation.

At the municipal level, we account for variables capturing local administrative features—such as absenteeism rates among municipal employees, staff turnover, the average educational attainment of municipal staff—as well, the mayor characteristics, as broader territorial attributes, including city type, population density, geographic remoteness, and regional fixed effects.[8]

Then, we frame our analysis as a predictive classification problem, where Y_{ij} denotes a binary outcome equal to 1 if project i in municipality j either experienced delays or failed to meet its targets, and 0 otherwise; and X_i is a vector of features of the project-level implementation architecture and W_j is a vector of municipality-level characteristics. We are interested in estimating the conditional probability of underperformance given these covariates.

$$f(x_i, w_j) = \Pr(Y_{ij} = 1 \mid X_i = x_i, W_j = w_j), \tag{1}$$

where the function $f(\cdot)$ is estimated using supervised ML. The model minimizes a loss function L over a class of admissible functions $\mathcal{F}$, subject to a regularization constraint that penalizes complexity to reduce overfitting:

$$\arg\min_{f \in \mathcal{F}} \sum_{i,j} L(f(x_i, w_j), y_{ij}) \quad \text{subject to} \quad R(f) \leq c. \tag{2}$$

The *project delay* outcome is defined through the binary indicator D_{ij}, as follows:

$$D_{ij} = \begin{cases} 1 & \text{if } t_i^{\mathcal{A}} > t_i^{\mathcal{E}} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where $t_i^{\mathcal{A}}$ denotes the actual (observed) end date of project i, and $t_i^{\mathcal{E}}$ the expected (planned) end date. The indicator D_{ij} thus takes the value 1 if the project experienced a delay, and 0 otherwise.

The second outcome variable captures whether the project failed to meet at least

[8] Population density and municipality classifications are drawn from ISTAT, administrative indicators from Cerqua et al. (2025), and information on mayors in office at the time of project activation from the Italian Ministry of the Interior.

one of its predefined output targets. We refer to this dimension as *target failure*, or *project unsuccess*, and define the corresponding binary variable U_{ij} as:

$$U_{ij} = \begin{cases} 1 & \text{if } \exists\, I_{ik} \in \{I_i\} \text{ such that } I_{ik}^{\mathcal{A}} < I_{ik}^{\mathcal{E}} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Here, $\{I_i\}$ denotes the set of Output Indicators associated with project i, and for each indicator k, $I_{ik}^{\mathcal{A}}$ denotes the actual reached target, and $I_{ik}^{\mathcal{E}}$ refer to expected to reach target. The variable U_{ij} is equal to 1 if at least one Indicator falls short of its planned target, and 0 otherwise. See Appendix B for a comprehensive discussion of the Output Indicators.

To ensure independence between training and testing samples, we allocate entire municipalities to either set. Specifically, we randomly assign 70% of municipalities to the training set and the remaining 30% to the testing set. Each project is then grouped according to the municipality in which it is implemented.[9] This procedure guarantees that no project appears in both sets, thereby preventing information leakage across samples.[10]

Having tested a series of linear and non-linear ML model we select Random Forest as our baseline algorithm (Breiman, 2001) (see **??** for details on comparative estimations and methodological details). As shown in Table 1, Random Forest combines strong predictive accuracy with high specificity and robust Precision–Recall performance, making it well suited to settings with pronounced class imbalance and complex non-linear relationships.

SHAP values will then decompose the predicted outcome into additive contributions from each feature, providing an interpretable and rigorous assessment of the contribution of each feature to the prediction (Lundberg et al., 2020). They will reveal not only which features are most predictive of underperformance, but also whether their influence is positive or negative and whether they follow a linear or non-linear pattern. In addition to the analysis conducted on the full sample, we investigate the associations between the projects' features and outcomes also by splitting the sample between small (below €10,000) and large projects (above €100,000) and by

[9]In the rare cases where a project spans multiple municipalities, we exclude it from the test set if any associated municipality is part of the training sample.

[10]While projects within a municipality may be correlated, we treat observations as conditionally independent within each subset. This simplifying assumption is supported by the large number of predictors and the richness of project-level data.

Table 1: Prediction accuracy of the baseline models - Random Forest

Model	**Accuracy**	**Specificity**	**Balanced Accuracy**	**ROC AUC**	**PR AUC**	**Share of positives**	**N. obs**
Delay	0.9167957	0.8919885	0.9043921	0.9629817	0.841726	0.1256109	138,125
Target Failure	0.8542808	0.9122167	0.8832487	0.9394975	0.76878	0.2082462	138,125

Note: We report standard measures of prediction accuracies (by column) for our baseline models predicting respectively the delay and the target failure of a project i implemented in municipality j (by row).

distinguishing projects located in Southern vs. Centre–Northern regions.

In terms of empirical applications of ML, our work aligns with a growing prediction-oriented literature that applies ML methods to EU policies, highlighting their suitability for analyzing heterogeneous outcomes across territories and steps of the policy cycle. For instance, Carrieri et al. (2026) employ ML to identify the territorial and institutional determinants of low absorptive capacity; Coco et al. (2025) focus on project implementation to predict delays in infrastructural projects; Caravaggio et al. (2025) apply ML to investigates the predictability of prospective beneficiaries of EU development funding; and Di Stefano and Resce (2025) use predictive modeling to assess which local governments may face challenges in utilizing available funding, focusing on open calls for childcare services within the Italian NGEU-funded Plan.

4 Results

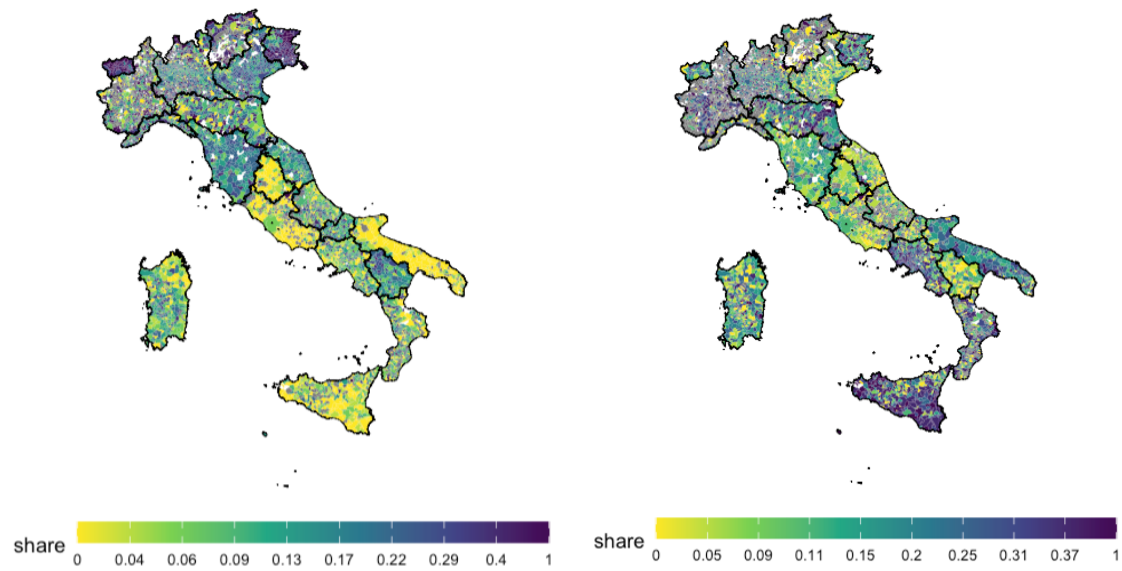
4.1 The Geography of Underperformance

Figure 1 shows the municipal-level variation in the share of delayed and unsuccessful projects. A highly heterogeneous territorial distribution of delays and target failures emerges at the local level, beyond the North-South divide and regional patterns. There is substantial heterogeneity among municipalities within the same region, confirming the significant role played by the socio-economic contexts in shaping project outcomes. However, many of the regional patterns persist at the municipal level. This might reflect the conditioning role of regional institutional quality and administrative capacity on the policy outcome, but also partly captures the role of the policy implementation architecture, given the significant regional specificity of project configuration choices.

Figure 1: Distribution of the share of underperforming projects at the municipal level

(a) Share of delayed projects – number
Municipality distribution (Deciles)

(b) Share of target failures – number
Municipality distribution (Deciles)



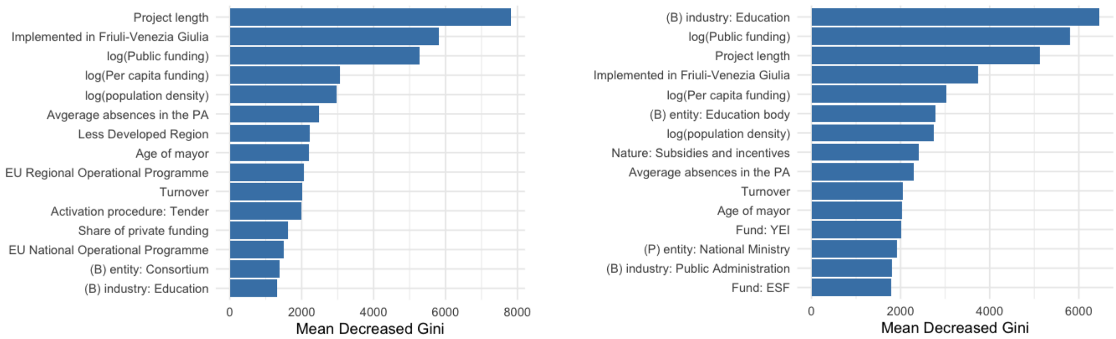
Note: The map on in panel (a) displays the share of delayed projects by municipality, while the map in panel (b) shows the corresponding share of unsuccessful projects. Due to the high degree of heterogeneity, values are grouped into deciles of the distribution. The legend bar indicates the minimum and maximum observed shares at the extremes, with intermediate values corresponding to the relevant decile cutoffs. Municipalities shown in white are those for which data are missing, largely due to changes in administrative boundaries that occurred during the 2014–2021 period.
Source: OpenCoesione (version 31/12/2024).

4.2 What Drives Underperformance

Here we show the most important features driving projects' underperformance; in the following paragraph, we will discuss the direction of the predicted associations as computed by SHAP values.

Figure 2: Top 15 predictors of project delay and target failure

(a) Delay (b) Target failure



Note: This figure shows the most important features identified by the Random Forest model in predicting project underperformance. Panel (a) ranks the top 15 variables contributing to the prediction of project delay, while Panel (b) displays those most relevant for predicting project target failure. (P) stands for programmer, while (B) stands for beneficiary. The variable importance is measured by the Mean Decrease in Gini impurity, which indicates how strongly a variable helps partition the data into more homogeneous groups by improving the quality of splits within the decision trees.

Figure 2 presents the top 15 predictors of project underperformance as identified with the Random Forest.[11] For both outcomes, together to municipal-level features accounting for the local contextual conditions, the list of the top 15 predictors includes many project-level features, confirming how choices made within the CP implementation architecture are key predictors of the policy outcomes. Particularly relevant is the presence of variable classifying projects in the Regional Operational Programme (ROP)/National Operational Programme (NOP), the variable =1 when the project is activated through a tender, the share of private co-funding, different variables accounting for the type of the project's programmer and the type/industry of the beneficiary. Other relevant characteristics of the projects that emerge as top predictors for delays and target failures are the project amount of funding (absolute and

[11] The detailed description of each predictor is available in Appendix A.

per capita), the expected length of the project, its nature (i.e., being a subsidy/incentive) and the main fund financing the project when they are the Youth European Initiative-YEI or the European Social Fund-ESF. In predicting policy outcomes, these CP implementation architecture features are as key as some municipal-level variables accounting for population density, the age of the mayor, the average absences in the Public Administration (PA), the turnover in the PA, and being located in specific regions/macroarea (i.e., in Friuli-Venezia Giulia or LDRs).

These findings confirm that policy outcomes (and policy underperformance) are driven by the policy implementation architecture. In particular, most of the features are subject to choices that are taken within the implementation step of the policy: they can be easily reconsidered without the need for any specific reforms, new intervention tools nor resources integration. In this sense, this analysis provides an evidence-based that can guide a reconsideration of specific dimensions of the CP implementation.

Furthermore, it is worth noting how delays and target failures turn to offer two very distinct stories of underperformance, as the two outcomes respond to different set of policy implementation and socio-economic features. Therefore, it is confirmed that considering both outcomes is key for a comprehensive analysis of how the implementation architecture drives CP underperformance in Italy.

The following paragraph will discuss the results from the SHAP values analysis, i.e., directions and magnitudes of the features' influence on the predicted outcomes.

4.3 How the Drivers of Underperformance Work

A graphical representation of SHAP value reveals how individual features contribute to the predicted probability of project delays or target failure.[12] To guide the interpretation of the results, it is useful to recall that positive (negative) SHAP values indicate that the feature increases (reduces) the predicted probability of the project's delay or target failure.[13]

[12] For the sake of space we did not included the graphs for all the features that we discussed but they are available upon request.

[13] The relationship between continuous features and project underperformance is analysed using SHAP dependence plots, where the x-axis reports the observed feature value and the y-axis the corresponding SHAP value, capturing the feature's marginal contribution to the predicted probability of underperformance. The SHAP dependence plots will help uss asses both the direction and non-linearity of feature effects. For binary features, we use bar plots reporting the mean absolute SHAP value associated with the positive category. Confidence intervals are reported in the bar plots to assess the strength and robustness of SHAP contributions.

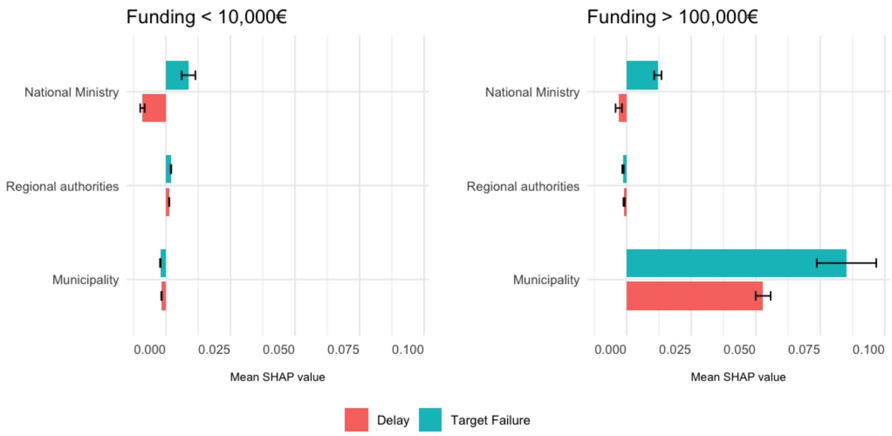
The SHAP graph in Figure 3 shows how projects belonging to the EU NOPs are associated with lower risks of delay and target failures, likely reflecting stronger central coordination and more streamlined implementation chains—a pattern that partly characterize also centrally managed Programmes of the Nationally-based CP. By contrast, projects belonging to the EU ROPs exhibit a higher probability of delay, reflecting the additional coordination and administrative burdens arising from interactions across institutional layers, and the presence of weak regional administrations. This finding is in line with evidence discussed in Section 2. Among projects belonging to Regional Programs, the EU ROPs ones are associated to lower risk of delays and target failures than the ones belonging to Regional Programs belonging to the Nationally-based CP. Also in this case, there is coherence with respect to what is concluded by recent causal evidence, according to which EU rules compensate administrative weaknesses, which, for some regions, might be pronounced (Celli et al., 2025).

Figure 3: SHAP Value Dependence Plots for Type of Programme

Note: This figure illustrates the relationship between the Programme under which each project is implemented and the predicted probabilities of project delay and target failure, as inferred from SHAP values. The height of each bar corresponds to the mean SHAP value for projects within the respective Programme, and the 95% confidence intervals indicate the uncertainty around these mean estimates.

Figure 4 shows how the risk of delays and target failures diverges clearly across types of programmers. Municipalities emerge as the most fragile programmers, being associated with the highest risks of both delays and target failure. As programmers, Regional Authorities tend to keep underperformance risk almost contained, with a negative, although limited, contribution to the probability of target failures. The role of National Ministries is mixed: while their projects are associated with a lower probability of delay, they also have a positive likelihood of target failures. What is interesting the most is that the choice of the best programmers should be made accounting for the scale of the project: while municipalities should be exempted from large projects (above €100,000), they can be the ideal programmer for small ones (below €10,000): in this case they reduce both delays and target failures, whereas Regional Authorities are associated to positive risks for both outcomes and while keeping the expected project's timeline Ministries are at the risk of missing project targets.

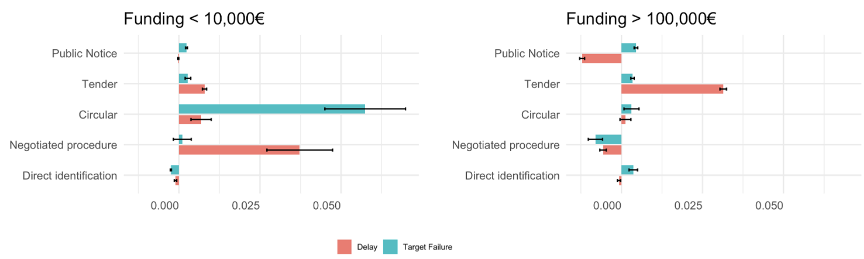
Figure 4: SHAP Value Dependence Plots for Type of Programmer



Note: This figure illustrates the relationship between the programmer under which each project is implemented and the predicted probabilities of project delay and target failure, as inferred from SHAP values, separately for projects with total funding below €10,000 (left panel) and above €100,000 (right panel). The height of each bar corresponds to the mean SHAP value for projects within the respective programmer, and the 95% confidence intervals indicate the uncertainty around these mean estimates.

Figure 5 focuses on the activation procedures (i.e., the legal instrument with which the project is generated within the Programme). At the moment, CP offers numerous options, with consequent significant administrative burdens. However, most of them (e.g., tenders and circulars) are useless (or detrimental) since strongly associated with underperformance risks. The best options to reduce both delays and target failures are direct identification and negotiated procedures. The former are particularly effective in the case of small-scale interventions; the latter for large-scale ones, when a closed coordination with potential beneficiaries is needed.

Figure 5: SHAP Value Dependence Plots for Activation Procedure



Note: This figure illustrates the relationship between the activation procedure under which each project is implemented and the predicted probabilities of project delay and target failure, as inferred from SHAP values. The height of each bar corresponds to the mean SHAP value for projects within the respective activation procedure, and the 95% confidence intervals indicate the uncertainty around these mean estimates. The left panel refers to projects with total funding below €10,000, while the right panel refers to projects with total funding above €100,000.

Moving further down the implementation chain, Figure 6 focuses on the role of the types of beneficiary. In the public sector, municipalities (and educational bodies at least in terms of target failures) struggle significantly, while Regions and Ministries in this role are capable of minimizing risk in both outcomes. In the private sector, the highest risks of underperformance are associated with different actors, all with a collective nature (Consortia and Associations/Foundations).[14] While legally unitary, these multi-actor structures can face internal complexity arising from coordination requirements. By contrast, when projects are assigned to individuals or to partnerships (companies), they are not at risk. Despite the central role played in CP, some types of actors turn out to be inadequate to act as direct beneficiaries of projects. Structural, organisational or capacity constraints might prevent them from benefiting from the projects they have been endorsed. In order to allow these actors to be protagonists of the CP, different project configurations should be considered for their involvement (e.g., actors endowed with sufficient resources and decision autonomy

[14] Consortia are formal collaborations among legally independent entities—typically firms—established to jointly implement projects within a shared legal or contractual framework.

could be identified as responsible for the delivery of projects targeting municipalities needs, instead of subordinate the municipalities' participation to the policy to an intake of administrative burdens that often overcome their actual resources).

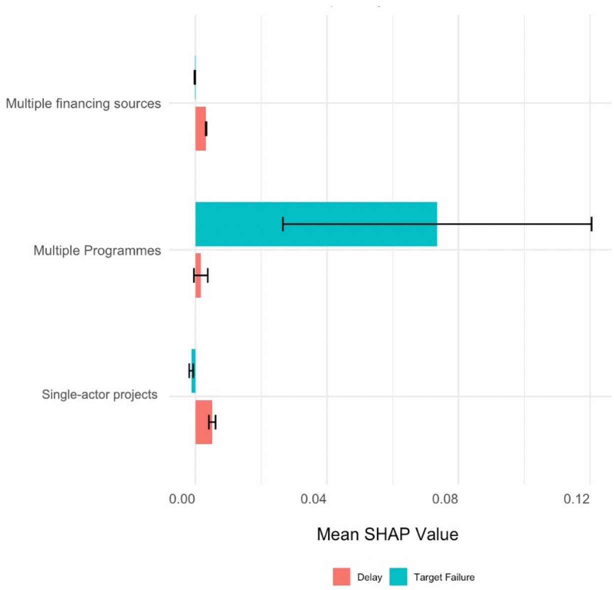
Figure 6: SHAP Value Dependence Plots for Type of Beneficiary

Note: This figure illustrates the relationship between the type of beneficiary under which each project is implemented and the predicted probabilities of project delay and target failure, as inferred from SHAP values. The height of each bar corresponds to the mean SHAP value for projects within the respective beneficiary type, and the 95% confidence intervals indicate the uncertainty around these mean estimates.

Irrespectively of the dynamics discussed up to now, Figure 7 shows that a project funded by multiple funds or in more than one Programme faces a higher risk of underperformance, reflecting additional administrative layers and coordination burdens across funding streams. Higher risk of delays is registered also at the opposite end of the spectrum, for projects where the same actor works as programmer and as beneficiary (we refer to these projects as "single-actor").[15] Overall, these patterns indicate that both excessive dispersion and concentration can undermine performance, also highlighting a trade-off between the reduction of coordination costs and the need for external accountability, typical of "single-actor" projects.

[15]This case applies to 9.57% of projects.

Figure 7: SHAP Value Dependence Plots for Multiple/Single Financing/Programmes/Actors



Note: This figure illustrates the relationship between project complexity features—namely multiple financing sources, participation in multiple programmes, and single- versus multi-actor implementation—and the predicted probabilities of project delay and target failure, as inferred from SHAP values. The height of each bar corresponds to the mean SHAP value for projects within the respective category, and the 95% confidence intervals indicate the uncertainty around these mean estimates.

The hypothesis of the paper, i.e., that implementation architecture features play a significant role in explaining how CP suffers from underperformance, is largely confirmed by the evidence presented so far. The analysis also shows how these features combine with conditioning factors attributable to the local context where the project is implemented: when projects are located in municipalities with a higher share of tertiary-educated personnel, characterized by a lower absenteeism rate and low turnover, they perform better (especially in terms of delays).[16] Finally, if a project is located in a LDR, it faces a lower risk of delay.[17]

To conclude the presentation of the results, we briefly summarise the evidence coming from the SHAP values associated to the features ranking among the top 15 predictors (Fig.2) belonging to implementation dimensions that have been not directly discussed so far: the expected length of the project is associated negatively with delays and positively to target failures; projects located in Friuli Venezia Giulia are associated with higher risks of both outcomes. The underperformance risk increases as it does the magnitude of the project in terms of total and per capita public funding but the relation is stronger for medium-sized projects, while small and large projects are associated to reduced risks;[18] a similar relation applies with population density, with projects located in averagely populated areas being at higher risk than others. Projects with private co-financing are at higher risk of delays, but only if the private resources are lower than the 60% of the total funding, otherwise the role of co-financing for delays (and target failure) is null. If the beneficiary belongs to the education sector, the project risks higher delays and target failures. Finally, if it belongs to the PA, the risk of target failures is lower; the same happens to projects classified as subsidy/incentive and to projects funded by the YEI, whereas the opposite applies for projects funded by the ESF.[19]

As anticipated in Section 3, we computed SHAP values by splitting the projects according to their magnitude (below €10,000 and above €100,000) and location (Northern-Southern regions). While the first sample split delivered interesting results

[16] The age of mayor was also among the top 15 predictors but the relation with the outcomes plotted in the SHAP graph is almost flat.

[17] The role of this feature in terms of target failure is not discussed because it does not compare among the top 15 predictors of this outcome.

[18] The reasons that can support a relative good performance of very small and very large projects respectively to the medium-sized ones can be different for the two extremes of the distribution: while smaller projects benefit from organisational simplicity, the highly funded ones tend to receive closer monitoring and strategic attention.

[19] The role of these features in terms of delays is not discussed because they do not compare among the top 15 predictors of this outcome.

(as discussed, for instance, for the cases of programmers and activation procedures), the results appeared substantially neutral to the North-South split. The CP implementation architecture seems to work with similar mechanisms across more or less developed regions.

The evidence obtained can be used to identify actionable guidelines for configuring successful projects. In order to provide concrete examples in this direction, Appendix D presents the case of two successful projects (completed on time and achieving their targets) belonging to the subset of projects for which the model predicts near-zero probabilities of delay and target failure: waterfall plots will illustrate how each feature contributes to the predicted outcomes.

5 Robustness checks

We implement a set of robustness checks, exploring the stability of our results under alternative specifications and definitions. In all cases, we obtain results that are consistent with the baseline.

To start, by following recent calls for the use of more flexible performance metrics (Molica, 2025), we consider a more conservative definition of both delay and target failure, classifying projects as delayed only when relative delays exceed the sample mean, and as unsuccessful only when at least one indicator underperforms its target by more than the indicator-specific average shortfall. Overall, results remain largely stable, with a stronger penalisation of projects in ROPs and when the Programmer is a Regional Authority, pointing to a higher incidence of severe underperformance in these cases. ROPs projects are further penalized also in the case of the second robustness test performed, which extends the analysis by adding Programme-specific Output Indicators to the Common Output Indicators used in the baseline (see Appendix B).

Then, we perform a series of checks that challenge methodological choices, obtaining results always consistent with the baseline. First, we check whether results are driven by extreme region-specific patterns by excluding the best- and worst-performing regions.[20] Second, we exclude the possibility that high correlations among some features could bias the computation of SHAP values. We compute statistically significant correlations between all features, dropping those that most frequently appear in the set of correlated pairs (value higher than 60%) and re-estimating the

[20] For the prediction of delay, we remove Friuli-Venezia Giulia (worst) and Puglia (best), while for target failure we exclude Friuli-Venezia Giulia (worst) and Marche (best).

models without these features. While the baseline results remain unchanged, predictive performance improves, suggesting that the removal of correlated features reduces noise in the data and that the model is capable of handling correlation among predictors. We, however, retain the original feature set, since some of the excluded variables carry relevant policy content. Finally, we repeat the analysis using regional rather than municipal sampling. This approach increases the independence between the training and testing sets, as entire regions are excluded from model training, thereby reducing the risk that region-specific characteristics or spatially correlated factors drive the estimated out-of-sample performance. In each iteration, fifteen regions are used for training, while five are consistently held out for testing. Results remain largely unchanged with respect to the baseline, both in terms of predictor rankings and SHAP values.

6 Conclusions

This paper carried out a comprehensive assessment of the CP implementation architecture and of its role in terms of projects' underperformance.

By leveraging project-level data on the implementation of the 2014-2020 CP in Italy, and thanks to an ML approach based on the use of a Random Forest model and SHAP values, it identifies how CP implementation architecture features predict projects' delays and target failures.

While the existing literature markedly focused on how policy impacts depend on the socio-economic and institutional conditions of the context where the policy is implemented, we focus on how projects' underperformance is driven by governance choices along the whole chain of the CP implementation architecture. The type of the Programme (EU vs. Nationally-based), its level of decentralization (National vs. Regional), the type and the industry of the project's programmer and beneficiary, the fund, nature and theme of the project, the legal procedure with which it is activated, the presence of private co-financing, of multiple financing sources/programs, the composition of the actors involved, the financial magnitude and the expected length, are some of the around 100 project-level features identified to build up the CP implementation architecture, together to a series of variables capturing characteristics of the project location at the municipal level (e.g., population density and type of municipality; P.A. skills, turnover, absenteeism; mayor characteristics, regional factors).

These features have been studied as predictors of projects' delays and target failures by our ML model, which revealed not only which are the most predictive features but also how each of them drives positively or negatively projects' outcomes.

The results show that underperformance is not random. Successful policy configurations capable of enhancing projects' performance can be identified and pursued *within* policy practice with existing tools and resources.

To embrace a virtuous change in this direction, CP should i) reinforce its EU anchoring and the role of the Central Administration of the Member States along the implementation chain, especially as managing authority of the Programmes; ii) exploit the consolidated know-how and apparatus of the Regions prioritizing their role as projects' programmers; iii) equip Municipalities with adequate support and solutions to making them able to benefiting from the policy without being directly responsible of large projects that burden them excessively; iv) narrow the spectrum of legal procedures leveraging the use of direct identifications for small projects and of negotiated procedures for large projects; v) find a balance between an excessive concentration of managing roles, associated to higher underperformance likely due to lack of enough accountability and excessive complexity that can arise when projects involve multiple programmes and/or funding sources. Finally, implementation remains more challenging in certain territorial contexts: *ceteris paribus*, projects implemented in municipalities with low-skilled PA, high absenteeism rate and low turnover display a significantly higher risk of underperformance. In this perspective, the socio-economic context should be considered *upstream*, with a place-based approach supporting a policy shift towards accessible, easily practicable and successfully-proven policy configurations rather than bottom-up participation per se: without targeted correctives the latter risks to translate into underperformance, with consequent waste of resources and failures in achieving CP objectives in the area who needed the most.

Far from being conclusive, the evidence provided in this paper can be intended as an evidence-based actionable guidance for the reconsideration of the implementation tools and options currently characterizing the CP implementation architecture, to shift implementation towards successful policy configurations. Embracing a performance-driven implementation paradigm is vital to turn ambitious strategies into timely, tangible results and make CP capable of tackling today's pressing challenges.

References

Bachtrögler-Unger, J., 2024. The Role of Administrative Capacity for an Effective Implementation of EU Cohesion Policy. Discussion Paper 24-067. ZEW – Centre for European Economic Research. doi:10.2139/ssrn.5011594.

Barca, F., 2009. An Agenda for a Reformed Cohesion Policy: A Place-Based Approach to Meeting European Union Challenges and Expectations. Independent Report. European Commission. Prepared at the request of Danuta Hübner, Commissioner for Regional Policy.

Breiman, L., 2001. Random forests. Machine learning 45, 5–32. doi:10.1023/A:1010933404324.

Caravaggio, N., Resce, G., Vaquero-Piñeiro, C., 2025. Predicting policy funding allocation with machine learning. Socio-Economic Planning Sciences 98, 102175. URL: 10.1016/j.seps.2025.102175, doi:10.1016/j.seps.2025.102175.

Carrieri, V., Blasio, G., Ferrara, A.R., Nisticò, R., 2026. Targeting and effectiveness of location-based policies. Journal of Regional Science 0, 1–17. doi:10.1111/jors.70054.

Celli, V., Crescenzi, R., de Blasio, G., Giua, M., 2025. Governance and the Implementation of the EU Cohesion Policy. Working Paper 50. Department of Geography and Environment, London School of Economics and Political Science. URL: https://researchonline.lse.ac.uk/id/eprint/127393.

Cerqua, A., Giannantoni, C., Zampollo, F., Mazziotta, M., 2025. The municipal administration quality index: the italian case. Social Indicators Research , 1–34doi:10.1007/s11205-024-03511-8.

Chernozhukov, V., Escanciano, J.C., Ichimura, H., Newey, W.K., Robins, J.M., 2022. Locally robust semiparametric estimation. Econometrica 90, 1501–1535. doi:10.3982/ECTA16294.

Coco, G., Monturano, G., Resce, G., 2025. Predicting Delays in Cohesion Infrastructure Projects. Economics and Statistics Discussion Paper 099/25. Università degli Studi del Molise, Dipartimento di Economia. URL: http://web.unimol.it/progetti/repec/mol/ecsdps/ESDP25099.pdf.

Coppola, G., Destefanis, S., Marinuzzi, G., Tortorella, W., 2020. European union and nationally based cohesion policies in the italian regions. Regional Studies 54, 83–94. doi:10.1080/00343404.2018.1447099.

Crescenzi, R., de Blasio, G., Giua, M., 2020. Cohesion policy incentives for collaborative industrial research: evaluation of a smart specialisation forerunner programme. Journal of Regional Science 60, 736–770. doi:10.1080/00343404.2018.1502422.

Crescenzi, R., Giua, M., 2016. The eu cohesion policy in context: Does a bottom-up approach work in all regions? Environment and Planning A 48, 2340–2357. doi:10.1177/0308518X16658291.

Crescenzi, R., Giua, M., 2024. The economic impacts of cohesion policy, in: Dotti, Musiałkowska, D.G.H., Walczyk (Eds.), EU Cohesion Policy: A Multidisciplinary Approach. Edward Elgar Publishing. Multidisciplinary Movements in Research, Political Science and Public Policy. doi:10.4337/9781802209402.

Crescenzi, R., Giua, M., Sonzogno, G.V., 2021. Mind the covid-19 crisis: An evidence-based implementation of next generation eu. Journal of Policy Modeling 43, 278–297. doi:10.1016/j.jpolmod.2021.03.002.

Cristofoletti, E., Gabriele, R., Giua, M., 2023. Gaining in impacts by leveraging the policy mix: Evidence from the european cohesion policy in more developed regions. Journal of Regional Science doi:10.1111/jors.12666.

Del Monte, A., De Iudicibus, A., Moccia, S., Pennacchio, L., 2022. Speed of spending and government decentralization: Evidence from italy. Regional Studies 56, 2133–2146. doi:10.1080/00343404.2022.2045010.

Del Monte, A., De Iudicibus, A., Moccia, S., Pennacchio, L., 2025. Administrative capacity and speed of public spending. Regional Studies 59. doi:10.1080/00343404.2025.2552865.

Di Stefano, R., Resce, G., 2025. The determinants of missed funding: Predicting the paradox of increased need and reduced allocation. Journal of Economic Behavior & Organization 231, 106910. doi:10.1016/j.jebo.2025.106910.

Foray, D., David, P.A., Hall, B.H., 2011. Smart specialisation: from academic idea to political instrument, the surprising career of a concept and the difficulties involved in its implementation. MTEI Working Paper .

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I., 2020. From local explanations to global understanding with explainable ai for trees. Nature Machine Intelligence 2, 56–67. doi:10.1038/s42256-019-0138-9.

Molica, F., 2025. Reassessing cohesion policy through the lens of the new eu industrial policy. JCMS: Journal of Common Market Studies 63, 302–319.

OECD, 2020. Strengthening Governance of EU Funds under Cohesion Policy: Administrative Capacity Building. Technical Report. Organisation for Economic Co-operation and Development.

OECD, 2025. Building Beneficiary Capacity under EU Cohesion Policy: A Framework for Action for Managing Authorities. Technical Report. OECD Multi-level Governance Studies. doi:10.1787/c2efa2e2-en.

Polverari, L., 2023. How to Better Support Administrative Capacity to Improve the Effectiveness of Cohesion Policy. Expert paper. European Commission. Final report, Contract No. 2023CE16BAT131.

Santos, A.M., Conte, A., Molica, F., 2025. Financial absorption of cohesion policy funds: How do programmes and territorial characteristics influence the pace of spending? JCMS: Journal of Common Market Studies 63, 227–245. doi:10.1111/jcms.13640.

Appendix A List of Features

Table A.1: List of features

Feature	Description	Mean/Freq
Per capita funding	Per capita public funding (population from 2019) for project i in municipality j	*7.68* *(103.62)*
Public funding	Total public funding (finanziamento totale pubblico netto) for project i in municipality j[21]	*68,996.37* *(1,961,850)*
Project lenght	Total number of expected days before completing project i in municipality j	*488.37* *(544.55)*
Big Project	Dummy equal to 1 if the project i is classified as a *Big project*	86 (0.017%)
Nature: Purchase of Goods	Dummy equal to 1 if the CUP nature of project i is *Purchase of Goods*	50,483 (9.75%)
Nature: Services	Dummy equal to 1 if the CUP nature of project i is *Purchase or realization of services*	174,074 (33.6%)
Nature: Subsidies and incentives	Dummy equal to 1 if the CUP nature of project i is *Contributions to units other than production units* or *Incentives to production units*	267,604 (51.7%)
Nature: Infrastructures	Dummy equal to 1 if the CUP nature of project i is *Public Works*	25,434 (4.91%)
Theme: Businesses	Dummy equal to 1 if the CUP theme of project i is *Business*	112,651 (21.76%)
Theme: Digital	Dummy equal to 1 if the CUP theme of project i is *Digitalization*	46,854 (9.05%)
Theme: Education	Dummy equal to 1 if the CUP theme of project i is *Education*	84,420 (16.31%)

Continued on next page

[21] If the project involves multiple municipalities, the total public funding is apportioned across them using 2019 population as weights.

Table A.1 – continued from previous page

Feature	**Description**	**Mean/Freq**
Theme: Energy	Dummy equal to 1 if the CUP theme of project i is *Energy*	6,351 (1.23%)
Theme: Environment	Dummy equal to 1 if the CUP theme of project i is *Environment*	6,344 (1.23%)
Theme: Governance	Dummy equal to 1 if the CUP theme of project i is *Governance*	3,656 (0.71%)
Theme: Inclusion	Dummy equal to 1 if the CUP theme of project i is *Social inclusion*	29,303 (5.66%)
Theme: Labour	Dummy equal to 1 if the CUP theme of project i is *Labour*	206,016 (39.80%)
Theme: R&D	Dummy equal to 1 if the CUP theme of project i is *Research and Development*	12,569 (2.43%)
Theme: Tourism	Dummy equal to 1 if the CUP theme of project i is *Tourism*	5,013 (0.97%)
Theme: Transport Systems	Dummy equal to 1 if the CUP theme of project i is *Transport Systems*	4,418 (0.85%)
t cohort	Dummy equal to 1 if project i started in year $t = 2014 - 2024$	
(P) entity: National Ministry	Dummy equal to 1 if the Programmer of project i is a National Ministry (Class 2.2 - Istat classification)	104,587 (20.621%)
(P) entity: Regional authorities	Dummy equal to 1 if the Programmer of project i is a Region (Class 2.4.1 - Istat classification)	335,478 (64.81%)
(P) entity: Province	Dummy equal to 1 if the Programmer of project i is a Province (Class 2.4.2 - Istat classification)	4,176 (0.81%)
(P) entity: Metropolitan city	Dummy equal to 1 if the Programmer of project i is a Metropolitan City (Class 2.4.6 - Istat classification)	143 (0.03%)

Continued on next page

Table A.1 – continued from previous page

Feature	Description	Mean/Freq
(P) entity: Municipality	Dummy equal to 1 if the Programmer of project i is a Municipality (Class 2.4.3 - Istat classification)	56,560 (10.93%)
(P) entity: Public economic entity	Dummy equal to 1 if the Programmer of project i is a Public Economic Entity (Class 1.6 - Istat classification)	5,942 (1.15%)
(P) entity: Public non-economic entity	Dummy equal to 1 if the Programmer of project i is a Public Non-Economic Entity (Class 2.7 - Istat classification)	1,568 (0.30%)
(P) entity: Hospitals and Healthcare	Dummy equal to 1 if the Programmer of project i is a Public Health Entity (Class 2.5 - Istat classification)	92 (0.02%)
(P) entity: Education body	Dummy equal to 1 if the Programmer of project i is an Education Body (Class 2.6 - Istat classification)	1,623 (0.31%)
(P) entity: Private	Dummy equal to 1 if the Programmer of project i is private (Classes 1.1-1.5, 1.7-1.9 - Istat classification)	6,568 (1.27%)
(P) entity: Others	Dummy equal to 1 if the Programmer of project i falls in "Others" (Class 9.9 - Istat classification)	858 (0.17%)
(P) Industry: Production	Dummy equal to 1 if the Programmer of project i is in the Production industry (Nace Rev. 2 Sect. A-E)	2,231 (0.43%)
(P) Industry: Construction	Dummy equal to 1 if the Programmer of project i is in the Construction industry (Nace Rev. 2 Sect. F)	502 (0.10%)
(P) Industry: Services	Dummy equal to 1 if the Programmer of project i is in the Services sector (Nace Rev. 2 Sect. G-J, L-N, R-T)	4,112 (0.79%)
(P) Industry: Finance and Insurance company	Dummy equal to 1 if the Programmer of project i is in the Finance and Insurance sector (Nace Rev. 2 Sect. K)	770 (0.15%)

Continued on next page

Table A.1 – continued from previous page

Feature	Description	Mean/Freq
(P) Industry: Public Administration	Dummy equal to 1 if the Programmer of project i is in Public Administration (Nace Rev. 2 Sect. O)	505,814 (97.72%)
(P) Industry: Education	Dummy equal to 1 if the Programmer of project i is in the Education sector (Nace Rev. 2 Sect. P)	3,278 (0.63%)
(P) Industry: Health	Dummy equal to 1 if the Programmer of project i is in the Health sector (Nace Rev. 2 Sect. Q)	361 (0.07%)
(P) Industry: Others	Dummy equal to 1 if the Programmer of project i is in other sectors (Nace Rev. 2 Sect. U)	527 (0.10%)
(B) entity: National Ministry	Dummy equal to 1 if the Beneficiary of project i is a National Ministry (Class 2.2 - Istat classification)	5,273 (1.02%)
(B) entity: Regional authorities	Dummy equal to 1 if the Beneficiary of project i is a Regional Authority (Class 2.4.10 - Istat classification)	135,663 (26.21%)
(B) entity: Metropolitan city	Dummy equal to 1 if the Beneficiary of project i is a Metropolitan city (Class 2.4.60 - Istat classification)	156 (0.03%)
(B) entity: Province	Dummy equal to 1 if the Beneficiary of project i is a Province (Class 2.4.20 - Istat classification)	2,829 (0.55%)
(B) entity: Municipality	Dummy equal to 1 if the Beneficiary of project i is a Municipality (Class 2.4.30 - Istat classification)	30,085 (5.81%)
(B) entity: Union of municipalities	Dummy equal to 1 if the Beneficiary of project i is a Island/mountain community or a union of municipalities (Class 2.4.30/2.4.50 - Istat classification)	1,004 (0.19%)

Continued on next page

Table A.1 – continued from previous page

Feature	**Description**	**Mean/Freq**
(B) entity: Public economic entity	Dummy equal to 1 if the Beneficiary of project i is a Public Economic Entity (Class 1.6 - Istat classification)	2,910 (0.56%)
(B) entity: Public non-economic entity	Dummy equal to 1 if the Beneficiary of project i is a Public non-economic Entity (Class 2.7 - Istat classification)	2,850 (0.55%)
(B) entity: Hospitals and Healthcare	Dummy equal to 1 if the Beneficiary of project i belongs to the National Health Service (NHS) (Class 2.5 - Istat classification)	579 (0.11%)
(B) entity: Education body	Dummy equal to 1 if the Beneficiary of project i is a School or University (Class 2.6 - Istat classification)	96,301 (18.61%)
(B) entity: Entrepreneur	Dummy equal to 1 if the Beneficiary of project i is an Individual entrepreneur, freelancer or self-employed worker (Class 1.1 - Istat classification)	63,022 (12.18%)
(B) entity: Partnership	Dummy equal to 1 if the Beneficiary of project i is a Partnership (Class 1.2 - Istat classification)	25,106 (4.85%)
(B) entity: Limited company	Dummy equal to 1 if the Beneficiary of project i is a Limited Company (Class 1.3 - Istat classification)	85,947 (16.61%)
(B) entity: Cooperative	Dummy equal to 1 if the Beneficiary of project i is a Cooperative (Class 1.4 - Istat classification)	10,638 (2.06%)
(B) entity: Consortium	Dummy equal to 1 if the Beneficiary of project i is a Consortium or other forms of cooperation between companies (Class 1.5 - Istat classification)	19,842 (3.83%)

Continued on next page

Table A.1 – continued from previous page

Feature	Description	Mean/Freq
(B) entity: Associations and Foundations	Dummy equal to 1 if the Beneficiary of project i is an Association, Foundation, Ecclesiastical Institution, Mutual Aid Society or a Committee (Class 1.7/1.8/1.9 - Istat classification)	25,297 (4.89%)
(B) entity: Natural Person	Dummy equal to 1 if the Beneficiary of project i is a Natural Person	10,093 (1.95%)
(B) industry: Agriculture and Natural Resources	Dummy equal to 1 if the Beneficiary of project i operates in Agriculture and Natural Resources industry (Nace Rev.2 Sect. A-B)	2,039 (0.39%)
(B) industry: Business and Finance	Dummy equal to 1 if the Beneficiary of project i operates in Business Services and/or Financial Services (Nace Rev.2 Sect. K-N)	27,494 (5.31%)
(B) industry: Construction	Dummy equal to 1 if the Beneficiary of project i operates in Construction Industries (Nace Rev.2 Sect. F)	11,259 (2.18%)
(B) industry: Education	Dummy equal to 1 if the Beneficiary of project i operates in Education Industry (Nace Rev.2 Sect. P)	151,852 (29.34%)
(B) industry: Health	Dummy equal to 1 if the Beneficiary of project i operates in Health Industry (Nace Rev.2 Sect. Q)	5,239 (1.01%)
(B) industry: ICT	Dummy equal to 1 if the Beneficiary of project i operates in ICT sector (Nace Rev.2 Sect. J)	9,209 (1.78%)
(B) industry: Manufacture and Energy	Dummy equal to 1 if the Beneficiary of project i operates in Manufacturing or Energy Sector (Nace Rev.2 Sect. C-D)	31,175 (6.02%)

Continued on next page

Table A.1 – continued from previous page

Feature	**Description**	**Mean/Freq**
(B) industry: Public Administration	Dummy equal to 1 if the Beneficiary of project i operates in the Public Administration sector (Nace Rev.2 Sect. O)	175,085 (33.83%)
(B) industry: Tourism	Dummy equal to 1 if the Beneficiary of project i operates in Tourism Industries (Nace Rev.2 Sect. I)	26,331 (5.09%)
(B) industry: Trade and Transports	Dummy equal to 1 if the Beneficiary of project i operates in Wholesale and retail trade or Transportation and storage (Nace Rev.2 Sect. G-H)	24,452 (4.72%)
(B) industry: Water Waste management and remediation act.	Dummy equal to 1 if the Beneficiary of project i operates in Water supply, waste management and remediation activities (Nace Rev.2 Sect. G-H)	1,025 (0.20%)
(B) industry: Natural Person	Dummy equal to 1 if the Beneficiary of project i is a Natural person	52,435 (10.13%)
EU: National Operational Programme (NOP)	Dummy equal to 1 if the project i belongs to a NOP	209,947 (40.56%)
EU: Regional Operational Programme (ROP)	Dummy equal to 1 if the project i belongs to a ROP	217,347 (41.99%)
Italian National Programme	Dummy equal to 1 if the project i belongs to an Italian National Programme	64,158 (12.40%)
Italian Regional Programme	Dummy equal to 1 if the project i belongs to an Italian Regional Programme	24,729 (4.78%)

Continued on next page

Table A.1 – continued from previous page

Feature	**Description**	**Mean/Freq**
Others	Dummy equal to 1 if the project i belongs to other programs such as Institutional Development Contracts (CIS) or Development and Cohesion Plan (PSC)	665 (0.13%)
Fund: ERDF	Dummy equal to 1 if the main fund of project i is ERDF, EMFF, or EAFRD	132,786 (25.65%)
Fund: ESF	Dummy equal to 1 if project i is funded by ESF (European Social Fund)	181,866 (35.14%)
Fund: YEI	Dummy equal to 1 if project i is funded by YEI (Youth Employment Initiative)	112,642 (21.76%)
Fund: FSC	Dummy equal to 1 if project i is funded by is FSC (Italy's Development and Cohesion Fund)	35,402 (6.84%)
Fund: PAC	Dummy equal to 1 if project i is funded by PAC (Action and Cohesion Plan)	10,898 (2.11%)
Fund: SNAI	Dummy equal to 1 if project i is funded by SNAI (National Strategy for Internal Areas)	43,141 (8.33%)
Fund: Interreg	Dummy equal to 1 if project i is funded by European Territorial Cooperation	111 (0.02%)
Fund: Multiple	Dummy equal to 1 project i is funded by multiple funds	749 (0.14%)
Direct identification	Dummy equal to 1 if project i is activated by direct identification	86,648 (16.74%)
Negotiated procedure	Dummy equal to 1 if project i is activated by Negotiated procedure	8,939 (1.73%)
Tender	Dummy equal to 1 project i is activated by a tender	131,366 (25.38%)
Public notice	Dummy equal to 1 if project i is activated by a public notice	288,053 (55.65%)

Continued on next page

Table A.1 – continued from previous page

Feature	Description	Mean/Freq
Circular	Dummy equal to 1 if project i is activated by a circular	2,589 (0.50%)
Share of private funding	Continuous variable (0-1), ratio of private to total funding for project i	*0.06 (0.19)*
Multiple financing sources	Dummy equal to 1 if project i is financed by more than one financing source	376,192 (72.7%)
Multiple activation procedures	Dummy equal to 1 if project i entails more than one activation procedure	122 (0.02%)
Multiple Programmes	Dummy equal to 1 if project i belongs to more than one Programme	749 (0.14%)
Implemented in multiple locations	Dummy equal to 1 if project i is implemented in more than one location	31,219 (6.03%)
Single-Actor project	Dummy equal to 1 if all the subjects involved in project i coincide (Programmer, Activator, Realizer, Beneficiary)	49,518 (9.57%)
Population density	Population density in municipality j computed as the ratio between the 2019 population and the municipality surface area	$961.2/km^2$ ($1460.3/km^2$)
Less Developed Region	Dummy equal to 1 if the project i is implemented in municipality j belonging to a Less Developed Region	198,078 (38.3%)
Inner Area	Dummy equal to 1 if the project i is implemented in municipality j classified as Inner Area	159,012 (30.72%)
Municipality type: Metropolitan City	Dummy equal to 1 if municipality j is a Metropolitan City	63,298 (12.23%)
Municipality type: City	Dummy equal to 1 if municipality j is classified as a City under ISTAT's Degree of Urbanization classification	113,374 (21.90%)

Continued on next page

Table A.1 – continued from previous page

Feature	Description	Mean/Freq
Municipality type: Small city and suburban areas	Dummy equal to 1 if municipality j is classified as a Small City or Suburban Area under ISTAT's Degree of Urbanization classification	222,319 (42.495%)
Municipality type: Rural area	Dummy equal to 1 if municipality j is classified as a Rural Area under ISTAT's Degree of Urbanization classification	118,604 (22.91%)
Implemented in $Region_k$	Dummy equal to 1 if the project i is implemented in a municipality j belonging to the Italian NUTS-2 region k	
Age of Mayor	Continuous variable reporting the age of the mayor of the municipality j at the start date t	*52.51* *(10.24)*
Education of Mayor: tertiary	Dummy variable equal to 1 if the mayor of the municipality j has a tertiary education at the start date t	341,325 (65.9%)
Avg. Education of Municipal PA: Tertiary	Dummy variable equal to 1 if, on average, PA employees in municipality j have a tertiary education (from Cerqua et al. (2025); average over 2014-2020)	181,730 (35.11%)
Average absences in the PA	Average number of absences by employees in municipality j (from Cerqua et al. (2025); average over 2014-2020)	*30.97* *(3.23)*
Turnover	Continuous variable (range 0-1) indicating turnover rate in municipality j (from Cerqua et al. (2025); average over 2014-2020)	*0.12* *(0.056)*

Note: For continuous variables, the last column reports the original feature mean, with the standard deviation in parentheses. For dummy variables, it shows the number of instances, with the corresponding percentage in parentheses.

Appendix B The Project-level Output Indicators

Establishing a common set of Output Indicators is essential for the effective monitoring and evaluation of EU Programmes. These indicators offer a standardized framework for measuring tangible results across diverse policy interventions, thereby ensuring alignment with both programme-specific goals and the EU's broader strategic priorities.

As set out in Regulation (EU) No. 1303/2013 of the European Parliament and of the Council of 17 December 2013,

> "Each priority shall set out indicators and corresponding targets expressed in qualitative or quantitative terms, in accordance with the Fund-specific rules, in order to assess progress in programme implementation aimed at achievement of objectives as the basis for monitoring, evaluation and review of performance."[22],

Accordingly, under the 2014–2020 National Monitoring System, the so-called Protocollo Unico di Colloquio (PUC) established the obligation to associate at least one Common Output Indicator (COI) with each project.[23]

In addition to these COIs (defined in a shared grid and valid across all Programmes), some projects are also associated with one (or more) Programme Indicator (PI). PIs are valid only within the context of the specific Programme and governed by individual fund regulations.

For all indicators—except a few explicitly described in the PUC—it is necessary to report both the ex-ante planned value (target) and, once measurable (typically during implementation or at completion), the achieved value at the date of the data release.

Table B.1: Total number of indicators by type

	COIs	**PIs**	**Total**
Total	1,761,536	340,172	2,101,708
Share	0.84	0.16	1.00

[22] European Parliament and Council of the European Union (2013), Regulation (EU) 30 no. 1301/2013 of 17 December 2013

[23] Ragioneria Generale dello Stato, *Monitoraggio unitario progetti protocollo unico di colloquio.*, 2020, version 2.2. Manual/documentation.

As shown in Table B.1, our dataset includes over 2 million Indicator records, each corresponding to an individual Indicator linked to a project in our sample. The number of COIs is significantly higher than PIs, representing more than 84% of all indicator records.

To assess performance, we define "target failure" as a project in which at least one associated Indicator falls short of its planned value at completion. While PIs are available for some projects, their influence on the overall classification is marginal. As shown in Table B.2, only 1% of observations change classification when PIs are taken into account.

This reinforces the robustness of our results to their exclusion and supports the choice to rely solely on COIs in the main analysis[24].

Table B.2: Distribution of Target success and failure by Indicator type

	Observations with COIs only	Observations with COIs and PIs	Total obs.
Target success	247,270 (0.48)	151,275 (0.29)	398,545 (0.77)
Target failure - COIs	*87,081 (0.17)*	*22,259 (0.04)*	*109,193 (0.21)*
Target failure - PIs	-	*4,835 (0.01)*	*4,835 (0.01)*
Target failure - Both	-	*5,022 (0.01)*	*5,022 (0.01)*
Total	334,351 (0.65)	183,244 (0.35)	517,595 (1)

To ensure comparability across projects and Programmes, we focus on COIs, which are systematically reported for all projects. Furthermore, in order to ensure homogeneity across data also exclude a small number of Indicators that are difficult to interpret or do not provide meaningful information[25].

[24]We conducted a Robustness check (see Section 5), proving that the inclusion of Programme-specific indicators does not alter our results.

[25]We excluded indicators 220–222, 2807, 2810, 326–327, 3891–3892, 418, 903, 915, and IO 3.1,

Table B.3: Distribution of projects by number of associated COIs

Number of COIs	Number of projects	Percentage
1	95,169	18.39
2	234,432	45.29
3	38,685	7.47
4	60,573	11.70
5	14,620	2.82
6	14,475	2.80
7	7,927	1.53
8	18,398	3.55
9	3,813	0.74
10	5,310	1.03
More than 10	24,193	4.67

Note: The table reports the number of projects by the number of associated COIs. Each unit represents a project i implemented in municipality j.. The maximum number of COIs associated with a single project is 40.

The type of COI associated to a project strongly depend on the nature of the project itself. Distinctive COIs range from physical milestones (e.g., surface area or road length) to social outcomes (e.g., number of unemployed participants), or financial outputs (e.g., volume of support granted). Accordingly, units of measure vary widely, including the number of people, square meters, kilometers, tonnes of CO, or euros, thus reflecting the diversity of intervention goals.

Nevertheless, our analysis abstracts from the specific content or unit of the indicators. What matters is whether a project fails the target for at least one of its COIs. This binary classification into target failure 0/1 allows us to evaluate performance consistently across heterogeneous interventions.

Table B.4: TOP 10 most frequent COIs by Project Nature

Project Nature	**COI Code**	**COI Description**	**Unit of Measure**
(1)	(2)	(3)	(4)

as they all measure the number of projects linked to a single project. These indicators are not informative for our purposes, as counting projects does not provide meaningful insight into project success or effectiveness.

Purchase of Goods	456	Equipped laboratories	number
Purchase of Goods	135	Capacity of infrastructure for childcare or supported education	number
Purchase of Goods	CV31M	Participants supported by actions to counter the effects of the COVID-19 pandemic (number) – Male	number
Purchase of Goods	458	Male students enrolled in pre-primary, primary, and secondary schools	number
Purchase of Goods	CV31F	Participants supported by actions to counter the effects of the COVID-19 pandemic (number) – Female	number
Purchase of Goods	457	Female students enrolled in pre-primary, primary, and secondary schools	number
Purchase of Goods	132	Decrease in annual primary energy consumption of public buildings	KWh/ year
Purchase of Goods	794	Units of goods purchased	number
Purchase of Goods	920	Development of applications and information systems	number
Purchase of Goods	CV33	Entities supported in the fight against COVID-19 (number)	number
Services	206M	People under 25 years old (male)	people
Services	206F	People under 25 years old (female)	people
Services	201F	Unemployed, including long-term unemployed (female)	people
Services	201M	Unemployed, including long-term unemployed (male)	people
Services	457	Female students enrolled in pre-primary, primary, and secondary schools	number
Services	458	Male students enrolled in pre-primary, primary, and secondary schools	number
Services	203M	Inactive people (male)	people
Services	203F	Inactive people (female)	people

Services	796	Beneficiaries	number
Services	205F	Workers, including self-employed (female)	people
Public Works	791	Surface area affected by intervention	square meters
Public Works	424	Number of public buildings or facilities affected by intervention	number
Public Works	798	Man-days worked	number
Public Works	114	Total length of roads rebuilt or renewed	km
Public Works	132	Decrease in annual primary energy consumption of public buildings	KWh/ year
Public Works	919	Lighting points	number
Public Works	134	Estimated annual reduction in greenhouse gas emissions	tonnes of CO_2 equiv.
Public Works	138	Open spaces created or restored in urban areas	square meters
Public Works	800	Beneficiaries / Beneficiary population	number
Public Works	468	Number of public buildings with improved energy consumption classification	number
Subsidies and Incentives	101	Number of businesses receiving support	number
Subsidies and Incentives	206M	People under 25 years old (male)	people
Subsidies and Incentives	206F	People under 25 years old (female)	people
Subsidies and Incentives	102	Number of businesses receiving grants (non-repayable)	number
Subsidies and Incentives	CV22	SMEs supported with working capital grants in response to the COVID-19 emergency (number)	number
Subsidies and Incentives	CV20	Direct support to SMEs for working capital (grants) in response to the COVID-19 emergency (total public cost)	euro

Subsidies and Incentives	201F	Unemployed, including long-term unemployed (female)	people
Subsidies and Incentives	201M	Unemployed, including long-term unemployed (male)	people
Subsidies and Incentives	796	Beneficiaries	number
Subsidies and Incentives	223	Number of micro, small, and medium enterprises financed (including cooperatives and social economy enterprises)	number

Note: The Table reports the 10 most frequent Indicators by project Nature. Column (1) reports the project nature, Column (2) contains the Indicator code, Column (3) provides a description of the indicator, while Column (4) indicates the unit of measure used by the Indicator.

Appendix C Supplementary ML materials

Table C1 reports a comprehensive comparison of prediction accuracies across alternative classification algorithms. We use standard logistic regression as a benchmark model and compare its performance with LASSO-regularised logistic regression (Friedman et al. 2023), Support Vector Machines (Cortes and Vapnik, 1995), Gradient Boosting (Friedman, 2001), and Random Forest (Breiman, 2001). The results clearly indicate that non-linear models—specifically Gradient Boosting and Random Forest—consistently outperform linear specifications such as logistic regression and its LASSO-regularized variant across all evaluation metrics. Notably, while the linear models yield Precision-Recall Area Under the Curve (PR AUC) values below 0.55, both Random Forest and Gradient Boosting exceed 0.84. This highlights their substantially greater ability to identify underperforming projects, particularly in the presence of a pronounced class imbalance. These models also deliver superior performance in terms of Receiving Operating Characteristics (ROC AUC), balanced accuracy, and specificity, suggesting more accurate classification and fewer false positives. Among the non-linear models, Random Forest attains the highest specificity (0.89) and a PR AUC of 0.841, which is especially relevant in our context where false alarms are costly. These findings underscore the advantages of flexible algorithms capable of capturing complex interactions and non-linearities in the data.

Table C1: Prediction accuracies

Model	**Accuracy**	**Specificity**	**Balanced Accuracy**	**ROC AUC**	**PR AUC**	**Share of positives**	**N. obs**
LOGIT	0.7596542	0.8266529	0.7931535	0.8693787	0.5485823	0.1256109	138,125
LOGIT LASSO	0.75423631	0.8318443	0.7930403	0.8695435	0.5481843	0.1256109	138,125
SVM	0.8538329	0.91710205	0.885467	0.9512389	0.800026	0.1256109	138,125
Gradient Boosting	0.9268012	0.8830056	0.9049034	0.9684796	0.8506445	0.1256109	138,125
Random Forest	0.9167957	0.8919885	0.9043921	0.9629817	0.841726	0.1256109	138,125

Note: We report standard measures of prediction accuracies (by column) for different methods we train (by row).

In particular, the Random Forest model, selected as baseline, performs strongly across all standard evaluation metrics for both outcomes (see Table 1 in the main text). For target failure, it achieves a balanced accuracy of 0.88, a ROC AUC of 0.94, and a PR AUC of approximately 0.77 -substantially higher than linear benchmarks and indicative of a robust ability to detect unsuccessful implementations, which are relatively frequent (about 21% of the sample). Similarly, the model predicting project delay achieves a balanced accuracy of 0.90, a ROC AUC of 0.96, and a PR AUC of roughly 0.84, demonstrating strong predictive power despite the lower prevalence of this class (around 13%). Specificity is above 0.89 in both cases, suggesting that the model effectively avoids false positives.

References

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189–1232.

Friedman, J., Hastie, T., Tibshirani, R., Narasimhan, B., Tay, K., Simon, N., Qian, J., & Yang, J. (2023). *Lasso and elastic-net regularized generalized linear models.* Astrophysics Source Code Library, ascl:2308.

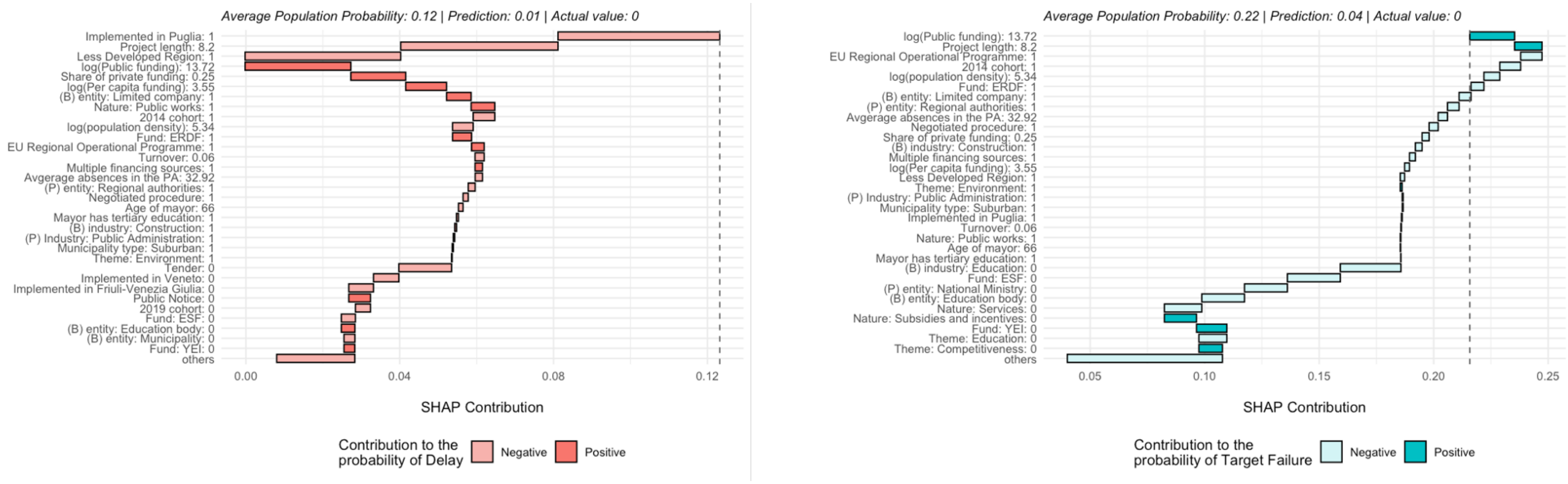
Appendix D Successful Projects' Configurations: Prediction and Fact Checking

The first project financed a large infrastructural intervention located in Apulia for the upgrade of a wastewater treatment plant to comply with health and safety standards. A low predicted delay risk (panel a) is driven, among others, by the expected (long) duration, by its location in an LDR, and the fact that it is activated through a negotiated procedure. These features offset the positive contribution to delay risk associated with the presence of a certain private co-financing, the type of the beneficiary, the financial magnitude of the project, the fact that it is an infrastructure and other aspects as plotted in Figure D1, (panel a). Panel (b) of Figure D1 shows the SHAP values for the features predicting target failures.

The second project supports the development of a web-based system for managing customer data. Panel (b) of Figure D2 shows that the risk of target failure was strongly reduced by the project classification as a subsidy or incentive, by its very small scale, by the fact that it originates within the European (vs. Nationally-based) CP, the type of the programmer (Regional Authorities) and of the beneficiary (Limited Company). Although belonging to the digital theme and incorporating a positive but limited private co-financing, the overall configuration of characteristics clearly favored success, yielding an almost zero predicted probability of target failure that corresponds to the actual outcome of the project. Similar considerations can be made around panel (a) of Figure D2 which shows the SHAP values for the features predicting delays.

Figure D1: Waterfall plots - Project 1

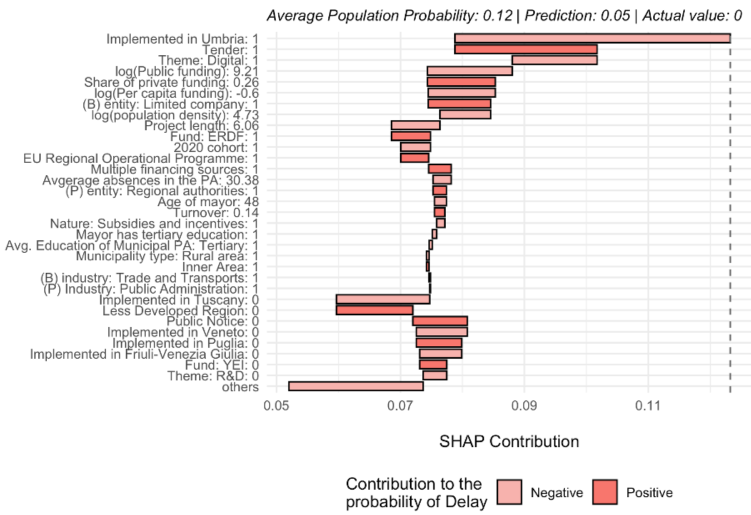
(a) Delay (b) Target Failure



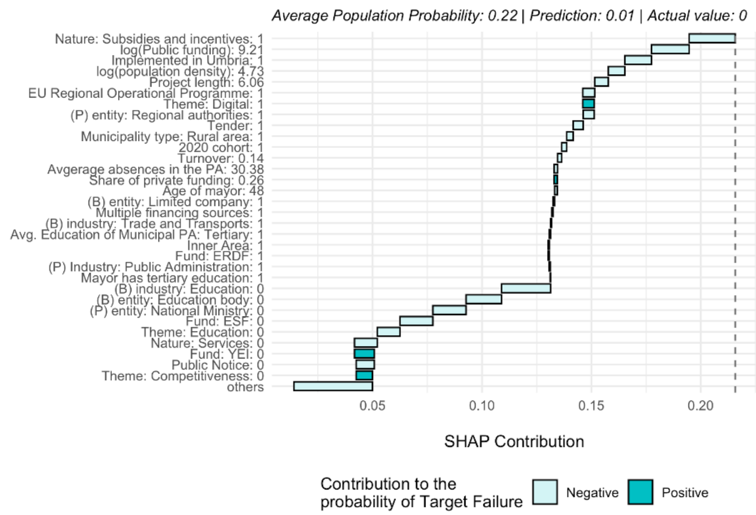
Note: These waterfall plots show how a project's feature (with positive value if present, 0 value if absent) pushes the prediction toward a corresponding outcome. The order of features reflects the absolute value of their SHAP contributions, with existing characteristics listed first and absent characteristics second. Please note that the absent characteristics beyond the top 10 are aggregated in the bar labeled "Other." The left panel corresponds to the prediction of delay, and the right panel corresponds to the prediction of target failure.

Figure D2: Waterfall plots - Project 2

(a) Delay



(b) Target Failure



Note: These waterfall plots show how a project's feature (with positive value if present, 0 value if absent) pushes the prediction toward a corresponding outcome. The order of features reflects the absolute value of their SHAP contributions, with existing characteristics listed first and absent characteristics second. Please note that the absent characteristics beyond the top 10 are aggregated in the bar labeled "Other." The left panel corresponds to the prediction of delay, and the right panel corresponds to the prediction of target failure.