The Determinants of Missed Funding

A Machine Learning Approach

DOCUMENTO DI VALUTAZIONE N. 17

DOCUMENTO DI VALUTAZIONE

Ufficio Valutazione Impatto

Impact Assessment Office

Senato della Repubblica

Questo *Documento di valutazione* è a cura di

ROBERTA DI STEFANO, Sapienza University of Rome, Rome, Italy

GIULIANO RESCE, University of Molise, Campobasso, Italy

I dati sono aggiornati al febbraio 2024

CODICI JEL: H5, H7, I3, J1, R5

PAROLE CHIAVE: COMPETITIVE FUNDING; NEXT GENERATION EU; ITALIAN NATIONAL RECOVERY AND RESILIENCE PLAN; PNRR; PREDICTIVE MODELING; MACHINE LEARNING

The Determinants of Missed Funding: A Machine Learning Approach

Luglio 2024

This research investigates how local governments overlook competitive funding opportunities utilizing machine learning and analyzing data from open calls within the Italian National Recovery and Resilience Plan (NRRP) financed by the European Next Generation EU. The focus is on predicting which local governments may face challenges in utilizing available funding, specifically examining the allocation of funds for Italian childcare services. The results demonstrate that it is possible to make out-of-sample predictions of municipalities that are likely to abstain from invitations, also identifying key determinants. Population-related factors play a pivotal role in predicting inertia, alongside other service-demand-related elements, particularly in regions with limited services. The study emphasizes the importance of local institutional quality and individual attributes of policy-makers. The adverse effects on participation resulting from factors that justify fund allocation may place regions with higher investment needs at a competitive disadvantage. Anticipating potential non-participants in calls can aid in achieving policy targets and optimizing the allocation of funds across various local governments.

Questa ricerca studia come le amministrazioni locali trascurino le opportunità di finanziamento competitivo utilizzando il machine learning e analizzando i dati dei bandi aperti nell'ambito del Piano Nazionale di Ripresa e Resilienza (PNR) italiano finanziato dall'Unione Europea Next Generation. L'attenzione si concentra sulla previsione di quali amministrazioni locali possono incontrare difficoltà nell'utilizzo dei fondi disponibili, in particolare esaminando l'assegnazione dei fondi per i servizi di assistenza all'infanzia in Italia. I risultati dimostrano che è possibile fare previsioni fuori dal campione sui comuni che probabilmente si asterranno dagli inviti, identificando anche i fattori determinanti. I fattori legati alla popolazione giocano un ruolo fondamentale nel prevedere l'inerzia, insieme ad altri elementi legati alla domanda di servizi, in particolare nelle regioni con servizi limitati. Lo studio sottolinea l'importanza della qualità istituzionale locale e delle caratteristiche individuali dei responsabili politici. Gli effetti negativi sulla partecipazione derivanti da fattori che giustificano l'allocazione dei fondi possono porre le regioni con maggiori esigenze di investimento in una posizione di svantaggio competitivo. Anticipare i potenziali non partecipanti ai bandi può aiutare a raggiungere gli obiettivi politici e a ottimizzare l'allocazione dei fondi tra le varie amministrazioni locali.

Sommario

1. Introduction

Addressing regional disparities and fostering sustainable development have proven to be enduring challenges, prompting various institutions, including the European Union, to develop comprehensive frameworks for a range of policies over the years (Farole, Rodr íguez-Pose, & Storper, 2011; Iammarino, Rodriguez-Pose, & Storper, 2019; Krugman, 1991). Within this context, there is a notable **emphasis on place-based policies directed towards under-developed areas** (Neumark & Simpson, 2015). The persistent struggle with regional disparities and the continuous discourse on the efficacy of these policies highlight the necessity for a nuanced comprehension of the factors shaping the distribution and utilization of funds.

From an empirical perspective, consensus regarding the efficacy of place-based policies remains elusive, given the divergence in outcomes when considering different timeframes, territorial levels, and the utilization of various econometric methodologies (Becker, Egger, & Von Ehrlich, 2018; Mohl & Hagen, 2010). Notably, recent times have witnessed a heightened **focus on econometric research**, driven by advancements in techniques and improved data accessibility. A meta-analysis by Dall'Erba and Fang (2017) illustrates that, overall, the development funds exhibit a generally positive and statistically significant impact on economic growth. Nevertheless, this impact tends to be relatively modest, particularly when evaluated in relation to the resources allocated. Focusing on the Italian case, in which place-based policies have been implemented since the post-World War II era, evaluation analyses yield scant empirical support for the notion that they can foster growth in the designated areas (Barone & de Blasio, 2023). Bronzini and De Blasio (2006) suggest that the initial effect of the funding is offset by a subsequent decline in the following years. However, Cerqua and Pellegrini (2014)'s analysis, which considers a more compelling evaluation strategy, reveals only a modest decline in the aftermath of the subsidy program. Regarding EU Funds, while some positive influence is observed (Giua, 2017), it lacks durability (Barone & de Blasio, 2023). In the case of other programs, the impacts are confined to specific micro areas within underdeveloped regions, but these are offset by adverse effects in adjacent regions, an occurrence often referred to as spatial displacement (Andini & De Blasio, 2016).

The recent economic literature has shown that **a significant aspect to consider is the role of local institutions**, which play a crucial part in the multilevel governance characterizing place- based policies. Referring to government quality indicators, it has been shown that institutional quality shows a wide heterogeneity among regions (Charron, Dijkstra, & Lapuente, 2014). Given that a substantial portion of place-based initiatives are managed at the local government level, the quality of these local government institutions becomes of pivotal importance. To exemplify this, Becker, Egger, and Von Ehrlich (2013), in their examination of European regions, demonstrate that institutional quality is a significant factor influencing the effectiveness of structural funds. **Administrative capacity is also considered the dominant explanation of the deficiency in EU funds absorption** (Incaltarau, Pascariu, & Surubaru, 2020; Milio, 2007; Surubaru, 2017).

Furthermore, **local politicians wield a great influence in molding the distribution of financial resources**. They can attract public funds even by selecting bureaucrats responsible for managing the allocation process as shown by Buscemi and Romani (2022) with the Italian example of the *Cassa per il Mezzogiorno*. It has been also shown that **many institutions** currently responsible for administering development funds, particularly local governments, **lack the motivation to use these funds effectively** (D'Amico, 2021). The argument posits that for development funds to stimulate growth, they should actively transform the local economy, including bringing back talented individuals who have migrated elsewhere. However, those who have left no longer have a say in local elections, leading local governors to focus on representing the interests of the remaining residents. This bias toward the current population encourages investments that maintain the status quo and bolster declining sectors rather than fostering innovation and development (D'Amico, 2021).

In the literature referenced above, an aspect that has yet to be explored is the **potential resistance of local potential beneficiaries to engage in competitive funding opportunities**, which is typically how funds are allocated. In practice, since many development funds are distributed through open calls, if the local government, organizations, enterprises, or citizens choose not to participate in such calls, this leads to their exclusion from the competition, subsequently resulting in a missed opportunity to secure funding and advance their development initiatives. Within the EU policy framework, **this is one of the reasons why several governments encountered difficulties in spending their allocated funds** (European Parliament, 2011). Such suboptimal execution performance has been partially attributed to factors associated with administrative capacity, as discussed by Cunico, Aivazidou, and Mollona (2022). This capacity is, in part, linked to the economic development levels of regions (Dincecco, 2017). Consequently, the limited allocation of funds is likely to have a more significant impact in territorial contexts where the imperative to invest is pronounced. This creates a paradox wherein the need for increased investment coexists with a reduction in resource allocation. In this context, empirical data indicates that regional policies frequently fall short in delivering results in the most underprivileged areas owing to their comparatively restricted planning capacities (Crescenzi & Giua, 2016; Kline & Moretti, 2014; Neumark & Simpson, 2015). In the case of specific funds aiming to promote development in underprivileged areas, there is a potential for the unintended consequence of achieving the opposite result. This is particularly significant as **regions with more pronounced needs may inadvertently miss out on opportunities, thereby worsening existing disparities**.

This paper investigates this aspect by leveraging one of the open calls of the Italian National Recovery and Resilience Plan (NRRP), which is the tool that outlines the objectives, reforms, and investments that Italy intends to implement thanks to the use of European Next Generation EU funds[1] **We use a machine learning (ML) model to predict local governments that fail to seize funding opportunities**, even when such funding is needed**, with a specific focus on the allocation of funds for childcare services** within the Italian NRRP. Due to the heterogeneity in participation of local governing bodies (UPB - Parliamentary Budget Office, 2022), this serves as **a case study on how and why potential beneficiaries, local administrations in our case, do not take advantage of available funding opportunities**. Identifying in

[1] https://next-generation-eu.europa.eu/index_en; https://www.italiadomani.gov.it/content/sogei-ng/it/en/home.html

advance potential no-participants in the calls can serve as a valuable tool for policymakers to push towards the policy target, and effectively allocate resources within the framework of the policy.

Among the scheduled initiatives, the **NPRR allocates 4.6 billion Euros for the development of nursery schools and the expansion of school infrastructure**. The goal is to enhance educational opportunities across the entire country by renovating existing nursery schools and con- structing new ones. This effort aims not only to expand the availability and enhance the quality of these services but also to assist families in balancing their personal and professional lives, promoting gender equality and women's employment, and stimulating an increase in the birth rate. During the implementation phase, the distribution of NPRR funding designated for nursery schools has been structured using specific calls for proposals *"Bandi asili nido"*. These calls define the criteria for distributing resources to municipalities in a competitive manner, requiring them to submit projects for consideration. In a nation facing challenges related to declining birth rates and women's unemployment, the *"Bandi asili nido"* program becomes a pivotal element within the context of the Italian NRRP. Extensive research demonstrates that childcare for preschool-aged children has a beneficial effect on maternal employment, especially among specific subgroups such as single mothers or those residing in economically disadvantaged areas (Baker, Gruber, & Milligan, 2008; Carta & Rizzica, 2018; Cascio, 2009; Fitzpatrick, 2010; Goux & Maurin, 2010; Havnes & Mogstad, 2011; Lefebvre & Merrigan, 2008; Nollenberger & Rodr íguez-Planas, 2015). Furthermore, in 2022, during the European Council meeting in Barcelona, it was mandated that Member states should strive to provide childcare services for at least 33 percent of children under 3 years old (*Barcelona European Council 15-16 March 2002. Presidency conclusions* 2002). However, the majority of municipalities fall significantly short of achieving this level of coverage. Among these, **over 3,400 municipalities with a serious shortage of nursery schools (coverage rate below 11 percent) did not participate in the *"Bandi asili nido"* calls** (UPB - Parliamentary Budget Office, 2022). This paper explores whether it is possible to predict which municipalities are likely not to apply to the call, even when they need to, and what are the main determinants of this inertia.

From an empirical perspective, we follow the recent economic literature arguing that targeting policy problems, such as predicting local governments that will miss out on funding chances, does not require ex-post correlation or causal inference solutions, but that new developments in the field of ML are of more use (Einav & Levin, 2014; Kleinberg, Ludwig, Mullainathan, & Obermeyer, 2015). ML techniques are gaining momentum for solving problems connected to poverty targeting (Jean et al., 2016), the effectiveness of public programs and spending (Andini, Ciani, de Blasio, D'Ignazio, & Salvestrini, 2018), and to identify corruption and political connections (de Blasio, D'Ignazio, & Letta, 2022; Mazrekaj, Titl, & Schiltz, 2023). Focusing on the Italian context, recent works have leveraged the potential of ML to predict the bankruptcy of local governments (Antulov-Fantulin, Lagravinese, & Resce, 2021), vaccine hesitancy in municipalities (Carrieri, Lagravinese, & Resce, 2021), to estimate local mortality and local inequality during the COVID-19 pandemic (Cerqua, Di Stefano, Letta, & Miccoli, 2021; Cerqua & Letta, 2022), to predict Geographical Indications areas (Resce & Vaquero-Pin˜eiro, 2022), and to predict dropout from higher education (Delogu, Lagravinese, Paolini, & Resce, 2024).

Predicting local governments that are likely to miss out on funding opportunities could become a relevant tool for intervening in the inertia of local authorities that need services but do not compete. Our results show that **it is possible to predict which local governments will not apply to the calls**, and that territorial socioeconomic features matter. In particular, **the population size and population density appear to be crucial for predicting the inertia of local governments**. Other important factors are connected to the demand for the services such as female occupation rate and birth rate. Further analysis highlights the role of the local institutional quality, the income and education level of the resident population, and the individual characteristics of the policymakers such as age, education, and gender. The outcomes of this study offer valuable insights and policy recommendations for improving the allocation of funds to diverse local governments with varying needs.

The remainder of the paper is structured as follows: Section 2 presents the Institutional framework, Section 3 is dedicated to the data and the methods, Section 4 shows the results and Section 5 concludes.

2. Institutional framework

The Italian Recovery and Resilience Plan provides additional and extraordinary financial resources to accomplish three main aims: to address the economic and social repercussions of the pandemic crisis; to drive a comprehensive ecological transition; and to tackle territorial disparities, gender inequality, weak productivity growth and a low rate of investment in human and physical capital[2]. Moreover, the NRPP planned to mitigate territorial disparities by ensuring that at least 40 percent of the resources are allocated to the Southern regions (Italian *Mezzogiorno*), the so-called "*quota-Sud*"[3]. The NRRP allocates most of the resources among territories through their participation in special calls for proposals that establish criteria for allocating resources in favour of participants on a competitive basis through submitting projects.

One of the important investments in the NRRP concerned resources provided for nursery schools and early childhood education and care services, the so-called *Bandi asili nido*. Research indicates that maternal labor supply measures boost mothers' engagement in the workforce and reduce the reservation wage for the unemployed, thereby enhancing their chances of securing employment (Carta & Rizzica, 2018). Promoting the quality and ample accessibility of childhood infrastructure contributes to increased female participation, thereby diminishing gender disparities in the labor market and fostering the accumulation of human capital for sustained long-term growth. This gained prominence as a significant policy consideration at the European level during the meeting of the European Council in Barcelona in 2002. The council emphasized the need for Member States to eliminate obstacles to female labor force participation and aim to offer childcare services to a minimum of 90 percent of children aged 3 to mandatory school age, as well as at least 33 percent of children under 3 years old, by the year 2010. Then, after the Covid-19 emergency, both objectives were updated[4].

In this context, Italy integrated the 33 percent objective into national legislation (Dlgs 65/2017), underlying the need to reduce territorial imbalances in the offer of early childhood services. In fact, Italy shows significant territorial inequalities in the availability and quality of childcare services, as depicted in Figure 1a where southern and more peripheral areas suffer a lack of adequate childcare services and are far from achieving the objective set at 33 percent of the number of nursery places for 100 children aged 0-2, hereafter LEP, i.e., essential levels of performance and services. In particular, the national mean of municipal LEP is around 16 nursery places for 100 children aged 0-2, while in the southern area it is around 9 and in the rest of Italy around 19[5]. Variations exist among regions, with none, on the average of municipal

[2] NRPP plans the investment and reforms in six thematic areas, the so-called missions, i.e., digitization, ecological transition, sustainable infrastructure, education and research, inclusion and cohesion, and health.

[3] Southern Italy, also known as *Meridione* or *Mezzogiorno* comprises the administrative regions that correspond to Abruzzo, Apulia, Basilicata, Calabria, Campania, Molise, Sicily, and Sardinia.

[4] A resolution of the EU Council of February 2021, has raised the objective of 90 percent in the 3-5 year age group to 96 percent, and the target of 33 percent for the under 3 age group to 45 percent, as part of the education targets to be achieved by 2030.

[5] If we sum the supply of places both in nursery schools and in supplementary early childhood services, the average municipal coverage is around 21 places for 100 children aged 0-2, 14 for the south and 25 for the centre and north. Note that, as reported by UPB - Parliamentary Budget Office (2022), the sum of available places divided by the total number of

LEP, meeting the 33 percent threshold. Additionally, when examining provinces (NUTS3 level), it becomes evident that only municipalities in eight provinces (one in Lombardy, one in Friuli-Venezia Giulia, two in Emilia-Romagna, and four in Tuscany) surpass the specified threshold on average[6].

In this framework, the NRRP resources are a unique opportunity to fill territorial gaps. Through Investment 1.1 (Plan for nurseries and nursery schools and early childhood education and care services) of Component 1 (Strengthening the offer of education services: from nurseries to universities) of Mission 4 (Education and Research), Italy assigns around 3.7 billion Euro for the creation of new spaces in early childhood education and care services through two distinct procedures. The first, as mentioned, pertains to ongoing projects totaling 0.7 billion Euros and was initiated with the DPCM (Decree of the President of the Council of Ministers) on December 30, 2020. The second, concerning the NRRP funds of 3 billion Euros,[7] was initiated through a public notice on December 2, 2021 (protocol number 48047) by the Ministry of Education. Both the DPCM in December 2020 and the public notice in December 2021 provide for access through special calls for proposals open to all Italian municipalities, individually or associated with other municipalities. To address economic and social imbalances, the former aims to allocate 60 percent of the funds for structures in disadvantaged areas and urban suburbs, identified by the social and material vulnerability index (IVSM) calculated by Istat. The latter, financed with NRRP resources, envisaged two constraints: the previously mentioned *"quota- Sud"* and considering the coverage rate at the regional level. However, different from the call of DPCM in December 2020, in which the requests were better than expected, for the NRRP one there was a poor response, particularly from southern municipalities. Subsequently, there was a further call only destined for the Mezzogiorno regions, with a priority for municipalities located in Basilicata, Molise, and Sicily.

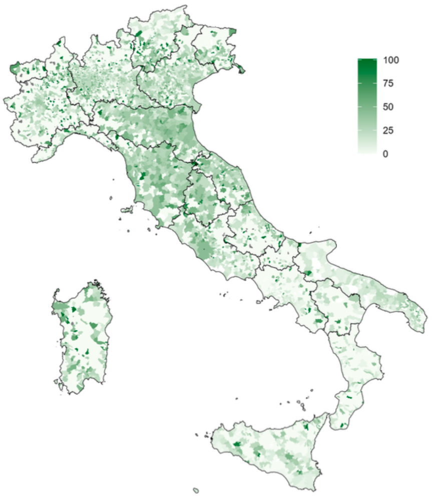
This framework aims to favour and protect the autonomy of municipalities through competitive calls for proposals and to integrate, in a second stage, the allocation one, in which criteria exist to guarantee the NPRR goals. However, around 60 percent of municipalities with a LEP less than or equal to 33 percent did not participate in the calls for proposals. Figure 1b, shows the municipalities with a LEP less than or equal to 33 percent which did not participate in either of the two calls for proposals.

children 0-2 in Italy is 26.9, the lower level of the figures reported here (the average of municipal LEP) depends on the fact that many (small) municipalities have small or zero LEP

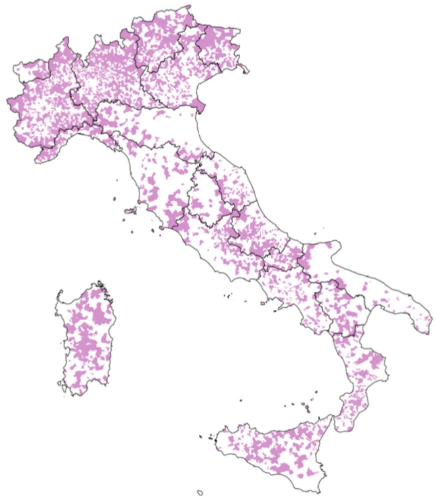
[6] If we sum the supply of places both in nursery schools and in supplementary early childhood services, we find 17 provinces with an average of municipal LEP more than 33 places for children aged 0-2.

[7] 2.4 billion for nursery schools, i.e., 0-2 years and 0.6 for schools of childhood, i.e., 3-6 years

Figure 1. Territorial distribution of nursery places and target municipalities



(a) Nursery places for 100 children aged 0-2 years in Italian municipalities in 2020



(b) Municipalities with a LEP less or equal to 33 percent that did not participate in the ”Bandi asilo” call

3. Data and Methods

3.1 Data

Our sample includes 7774 municipalities. We perform two different predictions, one for South- ern regions and one for the rest of the Country (Centre-North). The former includes 2486 municipalities while the latter includes 5288. Two primary issues affected this decision: firstly, the NPRR imposes an allocation constraint known as the abovementioned *"quota-Sud"* which mandates that 40 per cent of the funding must be directed toward the southern regions; secondly, the reopening of the *"Bandi asili nido"* call exclusively for southern regions (see section 2). These contribute to a scenario where, in practice, two distinct competitions have emerged in different parts of the country.

Our dependent variable is the non-application to at least one call *Bandi asilo* that is equal to 1 if the municipality did not apply for the part of the call referred to the nursery school for children 0-2 years for any call, and 0 otherwise. We identify five categories of predictors that could potentially affect non-application to the call and, consequently the inertia of local government: structural (St), socioeconomic (SE), demand (D), institutions (IQ), and politics (Pol). All features are taken in the period before the call opening.

The structural features include population size, population 0-2, population density, and the Compactness Index (Percent width of the largest built area polygon, it is indicated in the literature with LCPI - Largest Class Patch Index) four dimensions that synthesize the structure, the dynamics, and the capacity or likely difficulty of public administrations of project management. The socio-economic features consist of income and the proportion of the population with a degree as drivers of the development of territories but are also strictly correlated to the demand features of territories. To directly consider the demand features, we consider the proportion of the population 20-49, the proportion of the population over 70, the birth rate, the proportion of female occupation, the occupational gap, the proportion of families with 3 components or more, the mobility index (ratio between the sum of flows in and out of the municipality for work reasons and the employed population of the municipality), and the nursery schools cover- age rate both for the municipality and an average of the local labour system (SLL) in which the municipality insists. These factors have a dual role because they could drive call participation given that they identify the territories' need to meet an already existing demand on one hand, or the possibility of placing confidence in a particular territory and boosting its demand on the other hand. The institutional features that can be connected to local government inertia in leveraging funding opportunities include the proportion of municipal employees with a degree or more, the technical offices expenditure, the solidity of the balance sheet (administration surplus or deficit in relation to current revenues), and the provincial institutional quality index (Nifo & Vecchione, 2014) with its five components, i.e., Regulatory quality, Government Effectiveness, Rule of law, Corruption, Voice and accountability. To conclude, the political features include the human capital of policymakers and their attitudes to understanding decision-making and answering territorial needs. Politicians who are better equipped to navigate the intricacies of policymaking, financial management, and resource allocation, are more adept at identifying available funding sources, developing competitive grant proposals, and effectively managing funds once secured. The factors we include are the average age of both the municipal council and the assessors, the percentage of females of both the municipal council and the assessors, the age, gender, and education of the mayor, the seniority of the mayor (years in charge), and the political party position.

Table 1 reports the descriptive statistics for each predictor in the two subsamples.

3.2 Method

The prediction task is formulated as follows: for each municipality i at the year t, based on the set of lagged features $\{St, SE, D, IQ, Pol\}_{i,t-1}$, find the function $f(.)$(machine learning model) that predicts non-application to at least one call *Bandi asilo* ($NoApp_{i,t}$):

$$\{St, SE, D, IQ, Pol\}_{i,t-1} \xrightarrow{f(.)} NoApp_{i,t}.$$

Following the predominant approach in the literature using ML models in social science, we randomly divide the database as 80 percent for training and 20 percent for the out-of-sample testing set. Because various data splits can yield varying outcomes, we conducted ten distinct random data splits and calculated the average model performance across these iterations to enhance the stability of our results.

We use four different ML predicting algorithms plus a more ”classical” logistic regression model:

- Elastic Net (EN): a regression statistical method that performs features selection and regularization with a mix of L1 (LASSO-type) and L2 (ridge-type) penalization to reduce overfitting and increase prediction accuracy and interpretability (Tibshirani, 1996; Zou & Hastie, 2005);
- Random Forest (RF): a family of randomized tree-based classifier decision trees that uses different random subsets of the features at each split in the tree (Breiman, 2001);
- Gradient Boosting Machines (GBM): an ensemble method that works in an iterative way where at each stage new learner tries to correct the pseudo-residual of its predecessors (Friedman, 2001);
- Neural Network (NN): a model that uses a set of connected input/output units in which each connection has an associated weight and learns by adjusting the weights to predict the correct class label of the given input (Ripley, Venables, & Ripley, 2016).

Table 1. Descriptive statistics

	South		Centre-North	
	Mean	sd	Mean	sd
Non-application	0.45	0.50	0.67	0.47
Structural (St)				
Population	7,726.10	28,206.94	7,324.91	47,998.32
Population 0-2	290.49	805.25	299.75	527.26
Population density	166.86	645.45	148.31	973.14
Compactness Index	69.74	23.40	67.20	22.90
Socioeconomic (SE)				
Income	19,905.60	2,419.61	23,778.64	2,926.52
Share of population with degree or more	0.04	0.02	0.04	0.02
Demand (D)				
Female occupation	0.39	0.07	0.57	0.07
Share of families more 3 components	0.37	0.08	0.35	0.07
Cover of nursery schools	9.22	25.52	18.70	36.29
Birth rate	6.49	2.62	6.26	2.75
Share of population 20-49	0.36	0.03	0.35	0.03
Population over 70	0.19	0.05	0.19	0.05
Avg cover of nursery schools in SLL	9.20	8.36	18.73	10.36
Occupational gender gap	0.37	0.10	0.25	0.10
Mobility index	0.76	0.07	0.82	0.07
Institutional (IQ)				
Corruption	0.59	0.16	0.91	0.06
Government effectiveness	0.29	0.13	0.47	0.14
Regulatory quality	0.29	0.18	0.57	0.13
Rule of law	0.33	0.18	0.72	0.16
Voice and accountability	0.39	0.17	0.65	0.1
Institutional Quality Index	0.34	0.14	0.74	0.12
Education of officials	0.77	0.18	0.82	0.16
Technical offices expenditure	326,450.10	4,617,997.98	305,795.11	1787503.4
Solidity of the balance sheet	0.83	0.54	0.56	0.51
Politics (Pol)				
Mayor's Age	53.46	10.41	54.55	11.22
Mayor's Education	0.63	0.48	0.40	0.49
Female Mayor	0.1	0.29	0.17	0.38
Average age of the council	47.28	4.14	49.95	4.68
Share of females in the council	0.31	0.13	0.34	0.12
Average age of assessors	0.31	0.13	0.34	0.12
Share of females among assessors	47.28	4.14	49.95	4.68
Seniority of the Mayor	1.86	1.39	1.81	1.04
Political Party				
N	2,486		5,288	

The hyper-parameter optimization is only done on the training set using a repeated (10 times) five-fold cross-validation[8]. The performance of non-application classification prediction

[8] Il models have been implemented using R software trained with the optimisation algorithms available through the caret package (Kuhn, 2021)

is assessed by analyzing the Receiver Operating Characteristics curve (ROC) (Fawcett, 2006) on the test set. In our binary classification problem, the positive class is defined as the municipality that does not apply to the call, and the negative class is the municipality that applies to the call. The ROC curve shows the classifier's diagnostic ability by plotting the true positive rate (TPR) on the y-axis against the false positive rate (FPR) on the x-axis, since its discrimination threshold is varied (Antulov-Fantulin et al., 2021). When the classification task is completely unpredictable, the ROC curve is the diagonal line with an Area Under the Curve (AUC) of 0.5; a perfect classifier, instead, has AUC equal to 1.0, overall, the higher the AUC, the more predictive the model.

4. Results

In this section, we present the results of the model predicting local governments that miss out on funding chances. The focus will be on two main aspects: the predictability of our dependent variable (Section 4.1) and the features' importance of independent variables used for predictions (Section 4.2).

4.1 The Prediction Task

This section shows the out-of-the-sample performance of the models after the hyper-parameter optimization performed on the training set[9]. The AUC in Figure 2 shows that the RF models outperform all the other ML models, as well as the Logistic regression both in the Southern and Centre-North regions. The results of the DeLong, DeLong, and Clarke-Pearson (1988)'s statistical test in Table 2 suggest that, for both the South and Centre-North cases, the difference between the ROC curve of the RF model (the top-performing one) and the remaining models is statistically meaningful (p-value < 0.05) solely in the case of the NN model for the South area and in the case of EN, NN and Logit models for the Centre-North area. Consequently, it implies that the performance of GBM, EL, and Logit are substantially in line with that of the RF model for the South and only the performance of GBM is substantially on par with that of the RF model for the Centre-North.

The satisfactory performance of all the algorithms presented here is proven by the high level of accuracy, statistically higher than the no information rate, and the high level of all the other performance measures reported in Table 3. In terms of Cohen's Kappa, Table 3 shows that all models (except the Elastic Net) have values higher than 0.3 and close to 0.4 in the case of RF both in the Southern and Centre-North regions, figures which are in the range between 'fair' and 'moderate' strength of agreement about the prediction reliability (Altman, 1990; Landis & Koch, 1977). In terms of the P-value of [Accuracy > Accuracy Null (No Information Rate)], all models in any area have an accuracy statistically higher than the no information rate.

[9] In the case of Southern regions the hyperparameters best tune for EN is alpha = 1 and lambda = 0.00267395; for Random Forest is mtry (Number of variables available for splitting at each tree node) = 18; for Gradient Boosting Machine is n.trees (total number of trees to fit) = 100, interaction.depth (number of splits in each tree) = 2, shrinkage (learning rate) = 0.1, and n.minobsinnode (minimum number of observations in terminal nodes) = 10; and for Neural Network is size (number of units in the hidden layer) = 3, decay (parameter for weight decay) = 0.1. In the case of Centre-Northern regions the hyperparameters best tune for EN is alpha = 1 and lambda = 0.0001880039; for Random Forest is mtry = 2; for Gradient Boosting Machine is n.trees = 100, interaction.depth = 2, shrinkage = 0.1, and n.minobsinnode = 10; and for Neural Network is size = 5, decay = 0.1.

Figure 2. ROC curves for four ML models and Logistic regression

(a) South (b) Centre-North

Models trained on 80 percent of observations and tested on the remaining 20 percent. The ROC curves are calculated on the test set (20 percent of the data, the same for each algorithm).
Average (out of 10 splittings) AUC South: EN = 0.734, RF = 0.75, GBM = 0.748, NN = 0.674, Logistic = 0.732.
Average (out of 10 splittings) AUC Centre-North: EN = 0.745, RF = 0.767, GBM = 0.764, NN = 0.703, Logistic = 0.746.

Table 2. Results of DeLong's test for ROC curves

	South		Centre-North	
	Z	p-value	Z	p-value
RF vs GBM	0.191	0.606	0.383	0.493
RF vs EN	1.159	0.283	2.379	0.031
RF vs NN	3.851	0.010	5.625	0.013
RF vs logistic	1.222	0.278	2.263	0.044

Averages calculated across the repetitions for 10 different random splittings.

These findings indicate the feasibility of predicting municipalities that will not participate in the calls. Both for Southern regions and Centre-North areas, the best performer model in terms of accuracy is RF, while other algorithms show slightly lower performance. This aligns with previous empirical applications, confirming that the tree-based models are the more competitive methods for structured binary tasks, especially for municipality classifications (Antulov- Fantulin et al., 2021; Carrieri et al., 2021; Resce & Vaquero-Pin˜eiro, 2022). Furthermore, in terms of accuracy, Table 3 also shows that the Logistic regression exhibits good performances overall. This result is in line with studies that have found that, in some binary cases, ML methods do not outperform simple logistic regression (Christodoulou et al., 2019). While achieving similar performance to logistic regression, other tested algorithms bring added benefits such as increased flexibility, the ability to handle complex data relationships, and adaptability to diverse patterns. These non-parametric algorithms can uncover intricate patterns that logistic regression might struggle to capture. Indeed, as these characteristics are material to the problem at hand, they are manifested in slightly better performance (see Table 3, showing that RF has higher accuracy both in southern and northern municipalities and Figure 2, showing that also in terms of AUC RF outperforms all other algorithms). Overall, the results in this section clearly show that the inertia of local government in being engaged in cohesion calls can be predicted with almost all the ML models available in the literature.

Table 3. Models' performances

	South					Centre-North				
	GBM	RF	EN	NN	Logit	GBM	RF	EN	NN	Logit
Accuracy	0.681	0.691	0.652	0.666	0.667	0.739	0.744	0.716	0.736	0.738
Kappa	0.350	0.370	0.287	0.323	0.326	0.353	0.357	0.296	0.321	0.324
AccuracyLower	0.638	0.648	0.608	0.622	0.624	0.711	0.716	0.688	0.709	0.710
AccuracyUpper	0.722	0.732	0.694	0.707	0.709	0.766	0.770	0.743	0.763	0.764
AccuracyNull	0.550	0.550	0.550	0.550	0.550	0.667	0.667	0.667	0.667	0.667
AccuracyPValue	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.029	0.000	0.000
McnemarPValue	0.243	0.247	0.194	0.463	0.422	0.000	0.000	0.047	0.000	0.000
Sensitivity	0.591	0.605	0.557	0.616	0.619	0.897	0.910	0.876	0.929	0.930
Specificity	0.756	0.762	0.727	0.706	0.707	0.423	0.411	0.399	0.351	0.354
Pos Pred Value	0.664	0.674	0.640	0.631	0.633	0.757	0.756	0.749	0.741	0.742
Neg Pred Value	0.693	0.703	0.676	0.693	0.694	0.674	0.696	0.653	0.712	0.715
Precision	0.664	0.674	0.640	0.631	0.633	0.757	0.756	0.749	0.741	0.742
Recall	0.591	0.605	0.557	0.616	0.619	0.897	0.910	0.876	0.929	0.930
F1	0.625	0.637	0.58	0.623	0.625	0.821	0.825	0.803	0.825	0.825
Prevalence	0.449	0.449	0.449	0.449	0.449	0.667	0.667	0.667	0.667	0.667
Detection Rate	0.265	0.272	0.251	0.277	0.278	0.598	0.607	0.584	0.620	0.620
Detection Prevalence	0.4	0.403	0.401	0.439	0.439	0.791	0.804	0.785	0.836	0.835
Balanced Accuracy	0.673	0.683	0.642	0.661	0.663	0.660	0.660	0.638	0.640	0.642

Figures are estimated on the confusion matrix, which shows a cross-tabulation of the observed and predicted classes, generating the predicted classes based on the typical 50 percent cutoff for the probabilities (Kuhn, 2021). Averages were calculated across the repetitions for 10 different random splittings.

4.2 The Determinants of Inertia

This section presents an elaboration of the feature importance in the prediction of inertia of local governments. In this task, we implemented the RF, which is the model with the higher area under the ROC curves (see Figure 2). To maximize the utilization of available data for the analysis presented in this section, we retrained the RF model on the entire dataset using the hyperparameters fine-tuned through ten repetitions of five-fold cross-validation on the training set[10]. The focus of this section will be on two main aspects: the *Feature Importance* (Section 4.2.1) and the *Partial Dependence* (Section 4.2.2).

4.2.1 The Feature Importance

Results of the features' importance are shown in Figure 3. In this case, there are some differences between the Southern and the Centre-North areas (the correlation between the features' importance is 0.88).

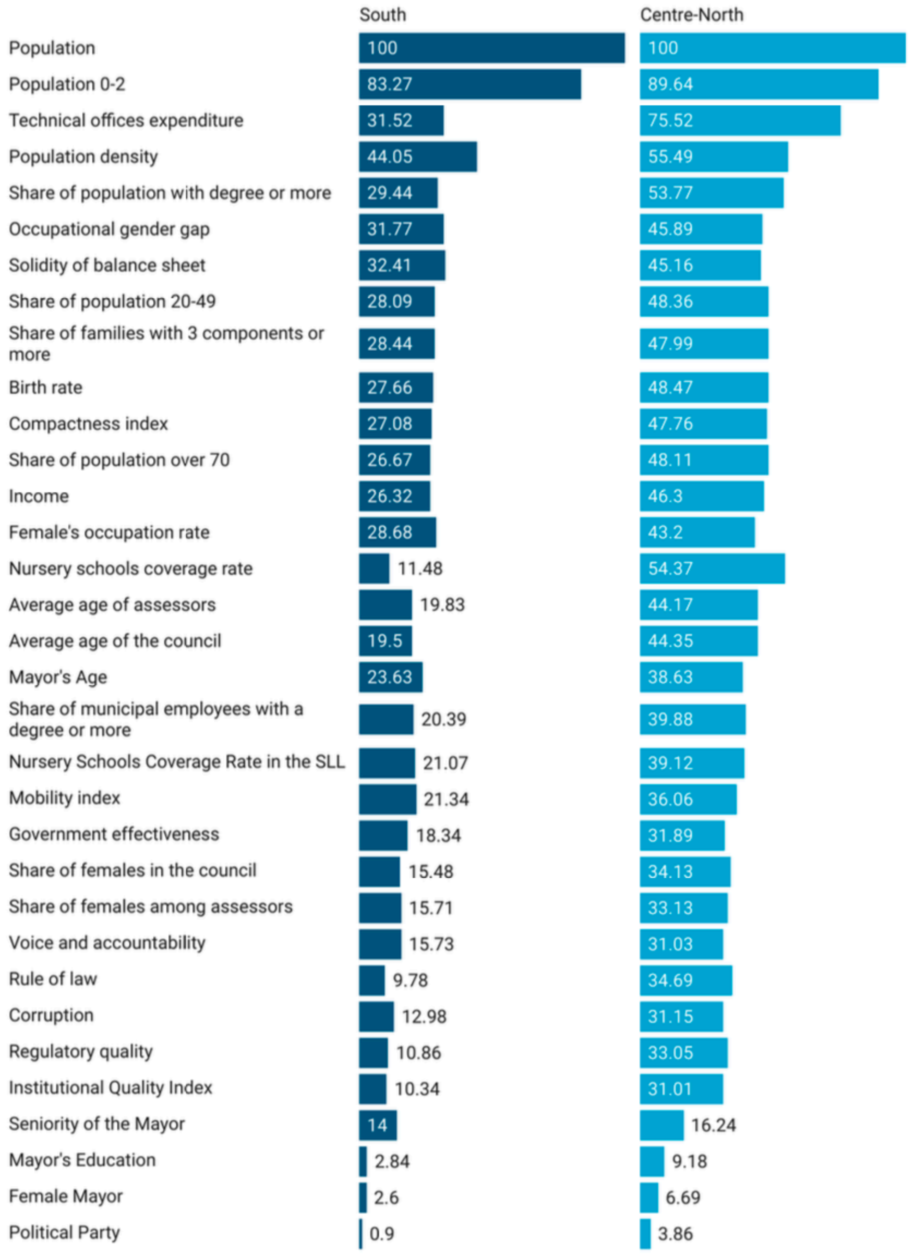
The most important factor both in Centre-North and in the South is the population size, meaning that the dimension of the municipality plays an important role in its inertia to compete

[10] Hyperparameters Best Tune are reported in footnote 9. The RF regression has been implemented in R, using the randomForest package (Liaw & Wiener, 2002).

for cohesion calls for funding. The second important factor for both samples is the population 0-2. On this point, it is possible that many municipalities that have low coverage are not particularly interested in obtaining funds to open a nursery because they have very few children under 3 years (this will be further investigated in section 4.2.2 by the Figure 4). The third important fac- tor on average is technical office expenditure, which is directly connected with the budget of the office in charge of developing projects. Then, it follows another population-related feature, i.e., the population density, which is strongly connected to urbanization, an indicator that has been recognized as crucial for the effectiveness of the cohesion policy (Albanese, Carrieri, Ferrara, & Speziali, 2023). In particular, the predominant role of population size and density in both Southern and Centre-North municipalities is in line with studies showing a different effect of the cohesion policy in urban and rural areas, as highlighted by Gagliardi and Percoco (2017) and Albanese, de Blasio, and Locatelli (2021). Another important structural factor in both Southern and Northern municipalities is the compactness index, which measures how prevalent the urban center is within a municipality. This may be partially connected to the fact that municipalities with fewer prevalent urban areas are more likely to require more than one nursery school.

An important category of factors explaining the inertia of local governments is related to the demand, i.e., the occupational gender gap, the female occupation, the mobility index, the birth rate, the proportion of the population over 70, the cover of nursery schools in the municipality and the average of nursery schools in the local labor system, the proportion of the population 20-49, and the proportion of families with 3 components or more. Demand factors are around 28% and 30% of importance for southern and centre-north municipalities, respectively. If these factors have a sign that is positively correlated with the call participation, i.e., the municipality participates because there is a need to satisfy the demand coming from citizens, these results give rise to some concern about the appropriateness of using cohesion (extraordinary) funds to cover something that needs to be covered by standard public expenditure.

Figure 3. Feature importance with the best model (RF)



Random Forest model trained on the whole database (bars sorted by average importance between South and Centre-North). The importance is estimated as the total decrease in node impurities from splitting on the variable, averaged over all trees. For classification, the node impurity is measured by the Gini index (Liaw & Wiener, 2002). Features ranked on the average importance between South and Centre-North.

Other significant factors explaining the inertia in both Southern and Centre-Northern municipalities are connected to the institutional quality as measured by Nifo and Vecchione (2014), by the education level of municipal employment, by technical offices expenditure, and by solidity of the balance sheet. In particular, the Institutional Quality Index is important as well as all the components of the index (Corruption, Government effectiveness, Regulatory quality, Rule of law, and Voice and accountability). The importance of Institutional quality strongly confirms the idea of Becker et al. (2013) on the role of local institutions in cohesion policy effectiveness. Furthermore, these results align with papers highlighting the role of administrative capacity in explaining the deficiency in EU funds absorption (Incaltarau et al., 2020; Milio, 2007; Surubaru, 2017). This evidence is also supported by the important role played by individual characteristics of local policymakers as the proportion of females in the council, the age of the mayor, and the average age of the council. Regarding age, Alesina, Cassidy, and Troiano (2019) noted the tendency of younger politicians to behave strategically, increasing spending and obtaining more transfers from higher levels of government, and these factors can somehow affect our outcome variable. Regarding gender, it has been shown that women in politics are usually more concerned about peoples' well-being, show higher cooperation and team working skills, and are less likely to engage in corruption, compared to their male counterparts (Chattopadhyay & Duflo, 2004; Herna ndez-Nicola ś, Mart ín-Ugedo, & M ínguez-Vera, 2018). Consequently, female political participation may affect the policies implemented and the applications to a call that can support gender equality and assist families in balancing their personal and professional lives (Funk & Gathmann, 2015). Moreover, our result on the important role of local policymakers in the cohesion policy is in line with the recent literature that introduced political economy elements into the cohesion debate (Buscemi & Romani, 2022; D'Amico, 2021).

Finally, important factors for predicting inertia in both Southern and Center-Northern municipalities are directly connected to socioeconomic development such as the level of education and the level of income. Overall, these features' importance confirms what has been shown by the growing body of literature highlighting how human capital development and the quality of local institutions may undermine or enhance the effectiveness of European funds (Aiello, Reverberi, & Brasili, 2019; Becker et al., 2013; Rodr íguez-Pose & Garcilazo, 2015). Our result corroborates previous findings by showing that these factors influence the inertia of local governments to participate in cohesion calls.

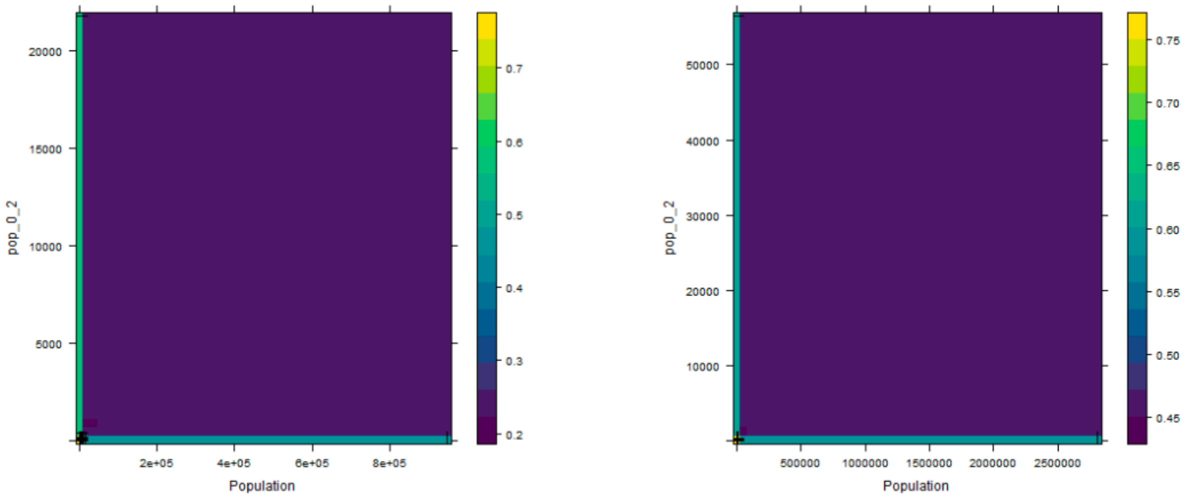
4.2.2 The Partial Dependence

Figure 3 highlights the importance of the features but does not provide insight into the sign of the relationship between the features and the inertia of local government. In this section, we deepen the sign of the association between features and the outcome of interest leveraging the partial dependence (i.e., marginal effect) plots. By relying on a partial dependence plot (PDP) we could graphically disentangle this relationship without the necessity of a previous mathematical model in the functional relationship also identifying possible non-linearity[11]. In this exercise, we only consider the most important factors explaining the inertia of local government as shown in Figure 3.

11 In this section we report PDPs with two input features of interest, able to show the interactions among the two features. For example, the two-variable PDP in Figure 4 shows the dependence of the probability of not participating in the call on joint values of population and population 0-2. We can see an interaction between the two features: with low population and low population 0-2 there is a higher probability of not participating in the call (more green to yellow). On the contrary, higher levels of both features are associated with a lower probability of not participating in the call (more purple)

Figure 4. Partial Dependence Plots for Population and Population 0-2 years

RFR



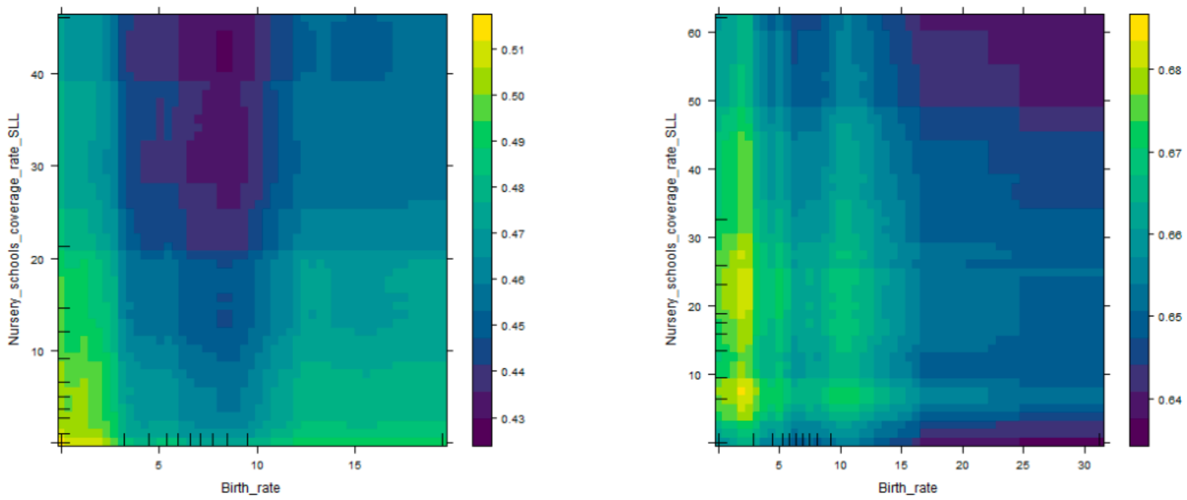
(a) South (b) Centre-North

RF model trained on the whole database. The Partial Dependence Plots are developed in R by the pdp package (Brandon, 2017).

Figure 4 shows the partial plots for Population and Population 0-2 years old. They are both the main drivers able to predict the probability of not participating in the call. Smaller municipalities are more likely to not participate, regardless of the Population 0-2. This may be partially connected to the lack of administrative capacity of smaller local governments, unable to plan in the medium to long term. The influence of the population aged 0-2 years closely aligns with the overall population impact. This could be attributed, in part, to the tendency of municipalities with a low number of children to not perceive a demand for nursery services, resulting in fewer applications to the call. It appears logical that factors linked to population matter, as childcare in sparsely populated areas, may inherently face economic challenges. However, this is the very rationale behind public intervention efforts aimed at encouraging a rise in the birth rate. In this regard, the negative impact on participation, stemming from factors like low population density and birth rate, which justify fund allocation, puts municipalities with greater investment requirements at a competitive disadvantage.

Figure 5 shows the partial plots for the birth rate and the nursery school coverage rate in the local labour system, which are two factors directly connected to the demand for nursery services. In both areas, the higher probability of not participating in the call is concentrated in municipalities with a lower birth rate. It appears that demand primarily influences applications in all municipalities although with some differentiation: in the southern municipalities, the nursery school coverage rate in the local labour system is more important while in northern municipalities the main driver is the birth rate. This could be partly connected to the heterogeneity in the availability of services and cultural differentiation between the two areas.

Figure 5. Partial Dependence Plots for Birth Rate and Nursery Schools Coverage Rate in the Local Labour System



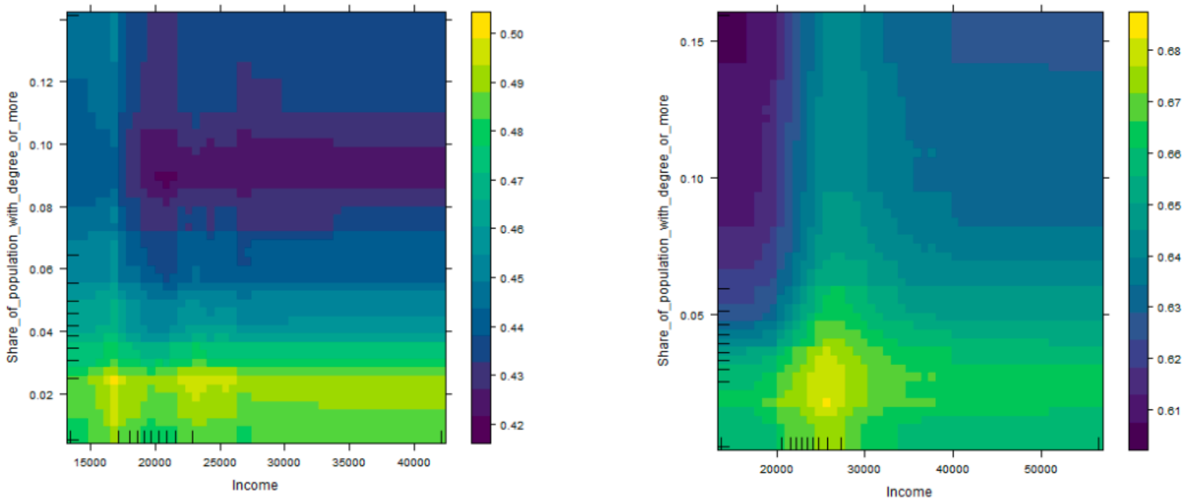
(a) South (b) Centre-North

RF model trained on the whole database. The Partial Dependence Plots are developed in R by the pdp package (Brandon, 2017).

However, it is crucial to note that, particularly in southern municipalities, the probability of applying is lower in municipalities in which there is lower local coverage of nursery schools in the local labour system. Overall, it is more likely that a municipality does not apply if the birth rate and the coverage of nursery schools in the local labour system are low: a clear case of increasing need correlating with missed opportunity. Furthermore, the positive role of demand-related factors, like the birth rate, raises apprehensions regarding the suitability of allocating extraordinary funds to address issues that should ideally be addressed through regular public expenditure. This prompts consideration of the appropriateness of using specialized funds to address needs that are typically within the purview of standard government budgetary allocations.

Figure 6 shows the PDPs for the average income and proportion of the population with a degree or higher in the municipality. The effect of the proportion of the population with a degree or higher is quite linearly negative on the inertia of local government both in Southern and North- ern municipalities. Municipalities with highly educated people are less likely to lose financing opportunities, mainly in the southern municipalities. The influence of income on the behaviour of local governments exhibits a distinctive pattern, and this pattern varies significantly between the Northern and Southern regions. Rather than following a straightforward linear relationship, the impact of income on the inertia of local government takes on a quite positive direction in the South and a curvilinear shape, resembling an inverted U in the North. Different factors may contribute to these results. Of course, income level is a proxy for development, but also high-income people may be less likely in need of childcare services. Overall, these nuanced variations in the impact of income on local government inertia underscore the complexity of regional dynamics and emphasize the importance of considering non-linear algorithms in policy analysis.

Figure 6. Partial Dependence Plots for Income and Share of Population with a degree or more

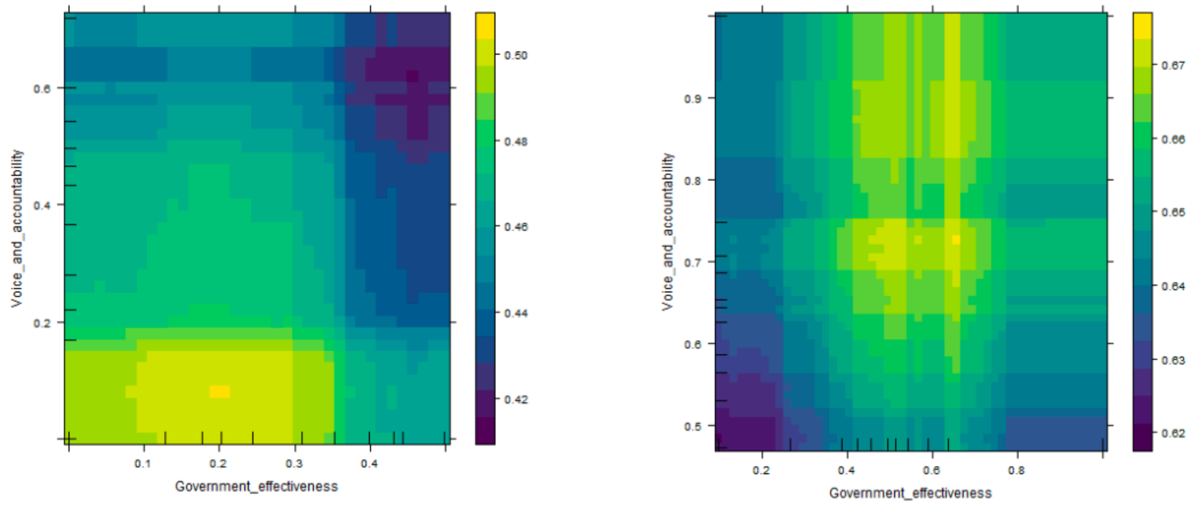


(a) South (b) Centre-North

RF model trained on the whole database. The Partial Dependence Plots are developed in R by the pdp package (Brandon, 2017).

In both areas of the country, the institutional quality indicators as measured by Nifo and Vecchione (2014) appear to be among the most important predictors of local government inertia in leveraging funding opportunities. In Figure 7 we show the PDP for the first two Institutional Quality Index components that are most important in average (Figure 3): Government Effectiveness and Voice and Accountability. In southern regions, higher Government Effectiveness and higher Voice and Accountability are associated with a lower probability of not participating in the call. This is quite expected and totally in line with the literature highlighting that a lower quality of the local institution seems to be associated with a lower effect of the cohesion policy (Becker et al., 2013). However, this association does not hold in the northern municipalities. This may depend on the fact that the institutional quality is, on average, higher in northern regions (this can be also noted by comparing the axes of the two panels in Figure 7), and the fact the institutional quality has lower heterogeneity in northern regions. To a certain degree, it appears that the beneficial impact of strong local institutions is somewhat diminished when these institutions are already well-established and of high quality. In other words, the correlation between the quality of local institutions and their positive outcomes may not be as pronounced or influential in municipalities where institutions are already functioning at a high standard.

Figure 7. Partial Dependence Plots for Government Effectiveness and Voice and Accountability



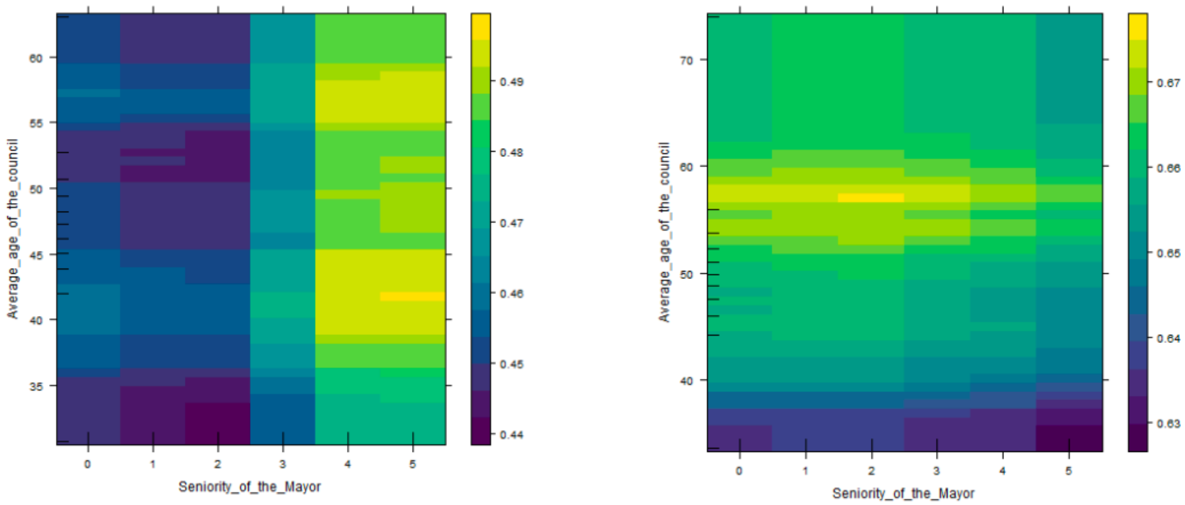
(a) South (b) Centre-North

RF model trained on the whole database. The Partial Dependence Plots are developed in R by the pdp package (Brandon, 2017).

Figure 8 shows the PDP for two important political features: the seniority of the mayor (years in charge)[12] and the average age of the council. Regarding average age, in both areas, younger politicians (particularly those under 40) are less likely to miss the funding opportunity in the Northern area. This observation could be attributed, at least in part, to the political aspirations of these politicians, who may exhibit a greater degree of concern for their career trajectory, as demonstrated in Alesina et al. (2019). Regarding the year in which they are in charge, there are some differences between the two areas. In the southern regions, a sort of U-shaped pattern emerges, indicating that politicians are more inclined to engage in initiatives during the second year of their tenure. This suggests that they refrained from participation at the outset, possibly due to inexperience, and abstained, more significantly, from involvement when they were nearing the end of their term, likely because they anticipated that the investment wouldn't reach completion before the upcoming election. Within the northern municipalities, we observe a decline in inertia as the length of a politician's service term increases. This pattern could be indicative of a lower level of opportunistic political behaviour among local politicians in this context.

[12] Article 51 of *Testo Unico sull'ordinamento degli enti locali* establishes the term of office for the mayor at five years, mirroring the duration of the municipal council.

Figure 8. Partial Dependence Plots for Seniority of the Mayor and Average Age of the Council

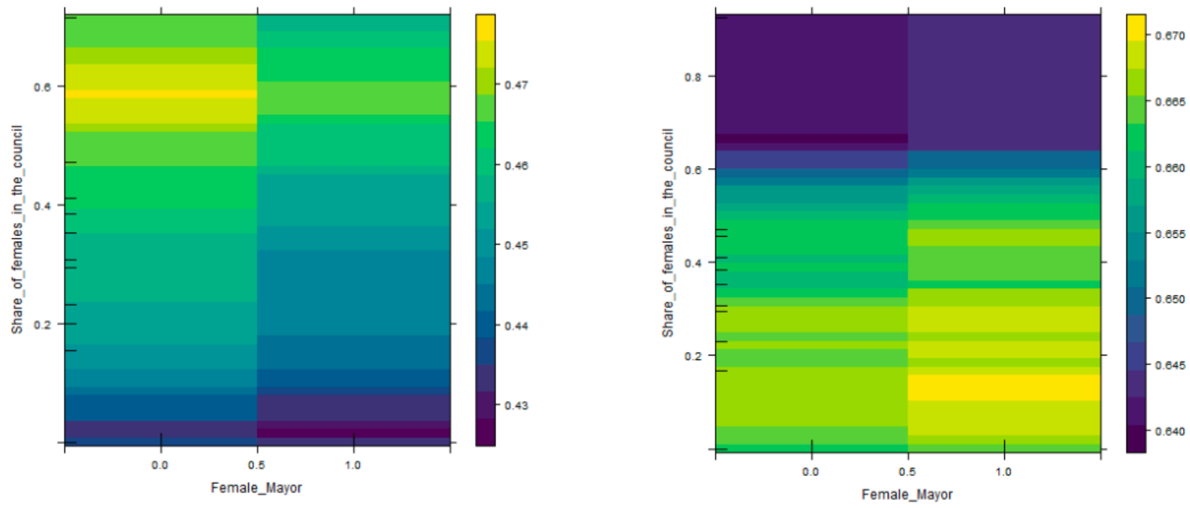


(a) South (b) Centre-North

RF model trained on the whole database. The Partial Dependence Plots are developed in R by the pdp package (Brandon, 2017).

Figure 9 shows the impact of gender in municipal governance by considering two important characteristics: the gender of the mayor and the percentage of females in the municipal council. The first factor is differentiated between Southern and Centre-Northern municipalities. In the south, female mayors are less likely to lose financing opportunities, and this is in line with the literature showing that women in politics are usually more concerned about peoples' well-being, in particular with services that can support gender equality (Chattopadhyay & Duflo, 2004; Funk & Gathmann, 2015; Herna ́ndez-Nicola ́s et al., 2018). This positive association is not confirmed in the Centre-Northern area of the Country, whilst impacting the percentage of females in the council, even more than in the southern municipalities. This is likely because having a greater representation of women on a council can mean that women's perspectives and experiences are more widely considered when making policy decisions rather than a unique leadership that, in structures characterized by high masculinity, may be more limited in influencing specific policies.

Figure 9. Partial Dependence Plots for Female Mayor and Share of Females in the City Council

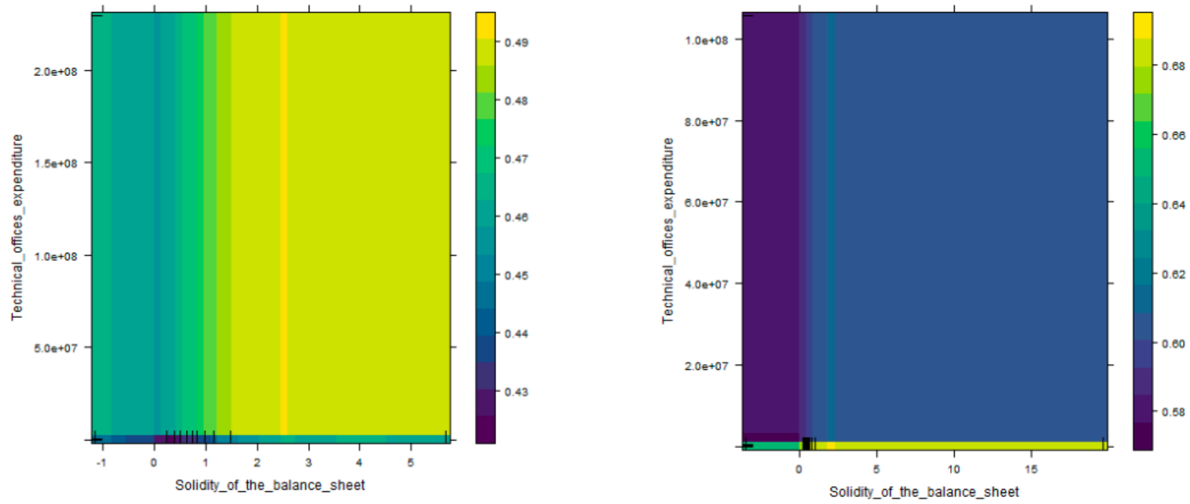


(a) South (b) Centre-North

RF model trained on the whole database. The Partial Dependence Plots are developed in R by the pdp package (Brandon, 2017).

Finally, in Figure 10 we show the technical office expenditure in relation to the solid of the balance sheet. These plots reveal that in the Centre-North there is a negative correlation between the expenses for technical offices and the inertia of local administration. Conversely, in the Southern municipalities, the inertia is not a problem of resources: high levels of resources spent in technical offices are not compensated for by greater efficiency. Another important factor is the solidity of the balance sheet, measured by the municipal administration's surplus or deficit in relation to current revenues. We use this indicator because mayors may have concerns about the current management of the nursery school after the structure is established by the NPRR. Results reveal that this is not the case; overall, higher inertia is observed in municipalities with a larger surplus.

Figure 10. Partial Dependence Plots for Solidity of the balance sheet and Technical Office Expenditure



(a) South (b) Centre-North

RF model trained on the whole database. The Partial Dependence Plots are developed in R by the pdp package (Brandon, 2017).

5. Conclusions

This paper delves into the intricacies of predicting which local governments, despite their need, are likely to miss out on funding opportunities. In this regard, open calls within the Italian National Recovery and Resilience Plan (NRRP), financed by the European Next Generation EU funds, are used as a case study. By leveraging machine learning techniques, the study sheds light on the determinants of inertia among potential beneficiaries, particularly focusing on the allocation of funds for childcare services.

The results indicate a strong capability to accurately predict which local government entities are less inclined to participate in funding initiatives. Furthermore, the empirical findings underscore the relevance of territorial structural features, with population size and population 0-2 emerging as crucial factors in predicting the inertia of local governments. Territories characterized by a limited number of children and a potential boost in birth rates through the presence of nursery schools are more prone to non-participation. This poses a substantial missed opportunity for local development, especially given that the investment required no co-financing. Additionally, factors tied to the demand for services, such as female occupation rate and birth rate, play a significant role. Moreover, the results highlight the multifaceted nature of the issue, emphasizing the influence of local institutional quality, income, education levels of the resident population, and individual characteristics of policymakers.

This analysis provides valuable insights for policymakers and practitioners, offering a nuanced understanding of the dynamics that contribute to the under-utilization of available funding. Identifying in advance potential no-participants in calls can serve as a valuable tool for policymakers to push towards the policy target, and effectively allocate resources within the framework of the policies. In particular, **the model can serve as an early warning system by flagging areas or groups that are at higher risk of missing funding opportunities.** This allows funding agencies to proactively reach out, provide additional support, or modify the application process to make it more accessible. In the existing policy framework, these predictions can be employed to target direct assistance, aiding the local government in promptly formulating appropriate plans, programs, and projects.

The research contributes to the broader discourse on place-based policies and the challenges associated with ensuring effective utilization of funds aimed at addressing regional underdevelopment. By exploring the under-participation of local authorities in competitive funding opportunities, **the study provides a foundation for developing targeted interventions to over- come inertia and enhance the impact of policy initiatives**. Moreover, the application of machine learning in predicting funding allocation challenges opens new avenues for research and policy development in the field of public programs and spending.

As nations strive to implement adequate policy plans and address pressing societal issues, the findings of this study offer actionable insights for optimizing the allocation of funds to diverse local governments with varying needs. In this regard, in many cases, more fragile territories that have more need to invest are not able to compete, thus missing opportunities. By **understanding the factors influencing inertia, policymakers can tailor interventions to**

encourage participation where it is most needed, ultimately fostering more effective and equitable development outcomes.

Ultimately, **the findings from this study can contribute to customizing funding calls to better address the distinct needs and challenges of targeted areas**, thereby minimizing the chance of overlooking opportunities. Recognizing the pivotal factors influencing the likelihood of missed funding provides valuable insights for refining and optimizing criteria and eligibility requirements in upcoming funding calls, ensuring a closer alignment with the intended recipients' needs. The persistence of underdevelopment in certain regions is intricately connected to local institutions, emphasizing the critical importance of policies capable of enduring and thriving despite the local context.

We acknowledge that, in 2024, the Italian Ministry of Education, in collaboration with the Italian Ministry of Economics and Finance, issued a new call for kindergarten applications. A subsequent phase of this study will analyze the applications submitted in response to this call.

References

Aiello, V., Reverberi, P. M., & Brasili, C. (2019). Regional diversity in experiences of cohesion policy: The cases of Emilia-Romagna and Calabria. *Papers in Regional Science*, *98*(6), 2275–2293.

Albanese, G., Carrieri, V., Ferrara, A., & Speziali, M. M. (2023). Looking for a star: A comprehensive evaluation of the EU cohesion policy. *Available at SSRN 4458080*.

Albanese, G., de Blasio, G., & Locatelli, A. (2021). Does EU regional policy promote local growth? evidence from the Italian Mezzogiorno. *Papers in Regional Science*, *100*(2), 327–348.

Alesina, A., Cassidy, T., & Troiano, U. (2019). Old and young politicians. *Economica*, *86*(344), 689–727.

Altman, D. G. (1990). *Practical statistics for medical research*. CRC press.

Andini, M., Ciani, E., de Blasio, G., D'Ignazio, A., & Salvestrini, V. (2018). Targeting with machine learning: An application to a tax rebate program in Italy. *Journal of Economic Behavior & Organization*, *156*, 86–102.

Andini, M., & De Blasio, G. (2016). Local development that money cannot buy: Italy's contratti di programma. *Journal of Economic Geography*, *16*(2), 365–393.

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

Baker, M., Gruber, J., & Milligan, K. (2008). Universal childcare, maternal labor supply, and family well-being. *Journal of political Economy*, *116*(4), 709–745.

Barcelona European Council 15-16 March 2002. Presidency conclusions. (2002). Retrieved from http://aei.pitt.edu/43345/ (Sections are bookmarked.)

Barone, G., & de Blasio, G. (2023). Place-based policies in the Italian case, part 1: A lot of money for little or no growth. *VoxEU. org*, *31*.

Becker, S. O., Egger, P. H., & Von Ehrlich, M. (2013). Absorptive capacity and the growth and investment effects of regional transfers: A regression discontinuity design with heterogeneous treatment effects. *American Economic Journal: Economic Policy*, *5*(4), 29–77.

Becker, S. O., Egger, P. H., & Von Ehrlich, M. (2018). Effects of EU regional policy: 1989-2013. *Regional Science and Urban Economics*, *69*, 143–152.

Brandon, G. (2017). pdp: An R package for constructing partial dependence plots. *The R Journal*, *9*(1), 421–436. Retrieved from https://journal.r-project.org/archive/2017/RJ-2017-016/index.html

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

Bronzini, R., & De Blasio, G. (2006). Evaluating the impact of investment incentives: The case of Italy's law 488/1992. *Journal of urban Economics*, *60*(2), 327–349.

Buscemi, T., & Romani, G. (2022). The political economy of regional development: Evidence from the Cassa per il Mezzogiorno. *Available at SSRN 4308599*.

Carrieri, V., Lagravinese, R., & Resce, G. (2021). Predicting vaccine hesitancy from area-level indicators: A machine learning approach. *Health Economics*, *30*(12), 3248–3256.

Carta, F., & Rizzica, L. (2018). Early kindergarten, maternal labor supply and children's outcomes: Evidence from Italy. *Journal of Public Economics*, *158*, 79-102. DOI: https://doi.org/10.1016/j.jpubeco.2017.12.012

Cascio, E. U. (2009). Maternal labor supply and the introduction of kindergartens into American public schools. *Journal of Human resources*, *44*(1), 140–170.

Cerqua, A., Di Stefano, R., Letta, M., & Miccoli, S. (2021). Local mortality estimates during the Covid-19 pandemic in Italy. *Journal of Population Economics*, *34*(4), 1189–1217.

Cerqua, A., & Letta, M. (2022). Local inequalities of the Covid-19 crisis. *Regional science and urban economics*, *92*, 103752.

Cerqua, A., & Pellegrini, G. (2014). Do subsidies to private capital boost firms' growth? a multiple regression discontinuity design approach. *Journal of Public Economics*, *109*, 114–126.

Charron, N., Dijkstra, L., & Lapuente, V. (2014). Regional governance matters: Quality of government within European Union member States. *Regional studies*, *48*(1), 68–90.

Chattopadhyay, R., & Duflo, E. (2004). Women as policy makers: Evidence from a randomized policy experiment in India. *Econometrica*, *72*(5), 1409–1443.

Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., & Van Calster, B. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of clinical epidemiology*, *110*, 12–22.

Crescenzi, R., & Giua, M. (2016). The EU cohesion policy in context: Does a bottom-up approach work in all regions? *Environment and Planning A: Economy and Space*, *48*(11), 2340–2357.

Cunico, G., Aivazidou, E., & Mollona, E. (2022). Decision-making traps behind low regional absorption of cohesion policy funds. *European Policy Analysis*, *8*(4), 439–466.

Dall'Erba, S., & Fang, F. (2017). Meta-analysis of the impact of European Union structural funds on regional growth. *Regional Studies*, *51*(6), 822–832.

D'Amico, L. (2021). Place based policies with local voting: Lessons from the EU cohesion policy. *Available at SSRN 3866098*.

de Blasio, G., D'Ignazio, A., & Letta, M. (2022). Gotham city. predicting 'corrupted 'municipalities with machine learning. *Technological Forecasting and Social Change*, *184*, 122016.

Delogu, M., Lagravinese, R., Paolini, D., & Resce, G. (2024). Predicting dropout from higher education: Evidence from Italy. *Economic Modelling*, *130*, 106583.

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 837–845.

Dincecco, M. (2017). *State capacity and economic development: Present and past*. Cambridge University Press.

Einav, L., & Levin, J. (2014). Economics in the age of big data. *Science*, *346*(6210), 1243089.

European Parliament. (2011). Report on absorption of structural and cohesion funds: Lessons learnt for the future cohesion policy of the EU (2010/2305 (ini)). *European Parliament*.

Farole, T., Rodrıguez-Pose, A., & Storper, M. (2011). Cohesion policy in the European Union: Growth, geography, institutions. *JCMS: Journal of Common Market Studies*, *49*(5), 1089–1111.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern recognition letters*, *27*(8), 861–874.

Fitzpatrick, M. D. (2010). Preschoolers enrolled and mothers at work? the effects of universal prekindergarten. *Journal of Labor Economics*, *28*(1), 51–85.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189–1232.

Funk, P., & Gathmann, C. (2015). Gender gaps in policy making: Evidence from direct democracy in Switzerland. *Economic Policy*, *30*(81), 141–181.

Gagliardi, L., & Percoco, M. (2017). The impact of european cohesion policy in urban and rural regions. *Regional Studies*, *51*(6), 857–868. 24

Giua, M. (2017). Spatial discontinuity for the impact assessment of the EU regional policy: The case of Italian objective 1 regions. *Journal of Regional Science*, *57*(1), 109–131.

Goux, D., & Maurin, E. (2010). Public school availability for two-year olds and mothers' labour supply. *Labour Economics*, *17*(6), 951–962.

Havnes, T., & Mogstad, M. (2011). Money for nothing? universal childcare and maternal employment. *Journal of Public Economics*, *95*(11-12), 1455–1465.

Herna ́ndez-Nicola ́s, C. M., Mart ́ın-Ugedo, J. F., & M ́ınguez-Vera, A. (2018). Women mayors and management of Spanish councils: An empirical analysis. *Feminist Economics*, *24*(1), 168–191.

Iammarino, S., Rodriguez-Pose, A., & Storper, M. (2019). Regional inequality in Europe: evidence, theory and policy implications. *Journal of economic geography*, *19*(2), 273– 298.

Incaltarau, C., Pascariu, G. C., & Surubaru, N.-C. (2020). Evaluating the determinants of EU funds absorption across old and new member states–the role of administrative capacity and political governance. *JCMS: Journal of Common Market Studies*, *58*(4), 941–961.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, *353*(6301), 790–794.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015). Prediction policy problems. *American Economic Review*, *105*(5), 491–495.

Kline, P., & Moretti, E. (2014). People, places, and public policy: Some simple welfare economics of local economic development programs. *Annu. Rev. Econ.*, *6*(1), 629–662.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of political economy*, *99*(3), 483–499.

Kuhn, M. (2021). CARET: Classification and Regression Training [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=caret (R package version 6.0-90).

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159–174.

Lefebvre, P., & Merrigan, P. (2008). Child-care policy and the labor supply of mothers with young children: A natural experiment from Canada. *Journal of Labor Economics*, *26*(3), 519–548.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomforest. *R News*, *2*(3), 18-22. Retrieved from https://CRAN.R-project.org/doc/Rnews/

Mazrekaj, D., Titl, V., & Schiltz, F. (2023). Identifying politically connected firms: a machine learning approach. *Oxford Bulletin of Economics and Statistics*.

Milio, S. (2007). Can administrative capacity explain differences in regional performances? evidence from structural funds implementation in Southern Italy. *Regional studies*, *41*(4), 429–442.

Mohl, P., & Hagen, T. (2010). Do EU structural funds promote regional growth? new evidence from various panel data approaches. *Regional Science and Urban Economics*, *40*(5), 353–365.

Neumark, D., & Simpson, H. (2015). Place-based policies. In *Handbook of regional and urban economics* (Vol. 5, pp. 1197–1287). Elsevier.

Nifo, A., & Vecchione, G. (2014). Do institutions play a role in skilled migration? the case of Italy. *Regional Studies*, *48*(10), 1628–1649.

Nollenberger, N., & Rodr íguez-Planas, N. (2015). Full-time universal childcare in a context of low maternal employment: Quasi-experimental evidence from Spain. *Labour Economics*, *36*, 124–136.

Resce, G., & Vaquero-Pin˜eiro, C. (2022). Predicting agri-food quality across space: A machine learning model for the acknowledgment of geographical indications. *Food Policy*, *112*, 102345.

Ripley, B., Venables, W., & Ripley, M. (2016). *Package 'nnet'. r package version, 7 (3–12), 700*. R, Vienna.

Rodr íguez-Pose, A., & Garcilazo, E. (2015). Quality of government and the returns of in- vestment: Examining the impact of cohesion expenditure in European regions. *Regional Studies*, *49*(8), 1274–1290.

Surubaru, N.-C. (2017). Administrative capacity or quality of political governance? EU cohesion policy in the new Europe, 2007–13. *Regional Studies*, *51*(6), 844–856.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267–288.

UPB - Parliamentary Budget Office. (2022). Piano asili nido e scuole dell'infanzia: prime evidenze dall'analisi delle graduatorie. *Focus tematico n. 9/2022 novembre 2022. Rome*. Retrieved from https://www.upbilancio.it/focus-tematico-n-9-25-novembre-2022/https://www.upbilancio.it/focus-tematico-n-9-25-novembre-2022/

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301–320.

SENATO DELLA REPUBBLICA

UFFICIO VALUTAZIONE DI IMPATTO

IMPACT ASSESSMENT OFFICE

www.senato.it/ufficiovalutazioneimpatto