Anticipating Delays in Cohesion Infrastruture Projects by Machine Learning

Giuseppe Coco[1], Gianluca Monturano[2], Giuliano Resce[3]
[1]Department of Economic - University of Bari
[2]Department of Economics - University of Modena and Reggio Emilia
[3]Department of Economics - University of Molise

Regional fragilities in incorporating the benefits of cohesion policies are partially due to difficulties in allocating the resources programmed. The efficiency of the allocation mechanism depends on several factors: project-related features, territorial characteristics, and institutional features such as coordination among public authorities. This paper proposes a machine learning model for predicting lags in cohesion project execution. Lags in policymaking are measured on cohesion projects monitored by opencoesione.gov.it. We measure execution times in different phases: (i) planning, (ii) execution, and (iii) conclusion. Results show that potential lags can be predicted and that institutional factors matter.

The detailed exploration of "Anticipating Delays in Cohesion Infrastructure Projects by Machine Learning" spans a comprehensive analysis of the challenges and opportunities presented by the implementation of cohesion policy projects. .

The research begins by setting the stage for the investigation, noting the significant investment in cohesion policies by the European Union aimed at reducing regional disparities and promoting sustainable development. Despite these efforts, the effective translation of programmed resources into real-world services and infrastructures has been fraught with delays and inefficiencies. The study identifies the need for improved tools to anticipate and manage these delays, suggesting machine learning (ML) as a promising solution.

The theoretical underpinnings of the study are rooted in the literature on project management, public administration, and machine learning. It discusses the evolution of Early Warning Models (EWM) and their application in various domains, emphasizing the shift towards data-driven approaches that leverage ML for predictive analysis. The research proposes that understanding the factors contributing to project delays in cohesion policy implementations can benefit significantly from ML techniques, which can process complex datasets to identify patterns and predictors of delays.

The methodology section provides a thorough description of the data collection process, drawing from the comprehensive dataset available on opencoesione.it, which includes detailed information on projects funded under the 2007 and 2013 cohesion policy cycles. The study defines the criteria for project selection, focusing on projects that are completed and pertain to public works, resulting in a dataset that encompasses a wide array of variables, from project-specific details to broader socioeconomic and institutional factors.

Using a binary classification approach, the research applies several ML algorithms, including Elastic Net, Random Forest, Gradient Boosting, and Neural Networks, to predict project delays. The selection of these algorithms is justified by their diverse strengths and capabilities in handling different aspects of the data, from linear relationships to complex, non-linear interactions.

The analysis reveals that the Random Forest algorithm outperforms others in predicting project delays, attributed to its ability to manage the high-dimensional and heterogeneous nature of the dataset. The research meticulously analyzes the importance of various predictors, uncovering that both project-specific and contextual factors play significant roles in determining delays. Factors such as the size of the project, the initial estimated completion time, and the socio-economic context of the implementing region are among the key determinants.

The discussion section reflects on the implications of these findings, considering the potential of ML models to serve as early warning systems for policymakers and project managers. It highlights the capacity of such models to not only predict delays but also offer insights into the underlying causes, thereby enabling more informed decision-making and strategic planning. The research also contemplates the broader applicability of these models beyond the scope of cohesion policy projects, suggesting their relevance for a wide range of public sector projects.

Concluding, the study underscores the significant contribution of ML to enhancing the predictive capacity in the realm of public project management. It calls for further research into refining these models, exploring their integration into the policy-making and project management processes, and extending their application to other domains. The research advocates for a more nuanced understanding of the interaction between project-specific characteristics, administrative capacities, and territorial contexts in shaping project outcomes.

This study its potential to revolutionize the management of cohesion policy projects through the application of machine learning. It not only contributes to the academic discourse on project management and public administration but also offers practical insights for policymakers, administrators, and practitioners tasked with the implementation of complex, large-scale projects.