

Forecasting Energy Performance Certificates (EPC) ratings in Italy using machine learning techniques



Francesco Braggiotti | Datasinc
Nicola Chiarini | Datasinc
Giulio Dondi | Datasinc
Luciano Lavecchia | Bank of Italy
Valeria Lionetti | Bank of Italy
Juri Marcucci | Bank of Italy
Riccardo Russo | Bank of Italy

Keywords: energy performance certificates, EPC, Italy, buildings, transition risk, machine learning, random forest classifier, random forest

JEL codes: Q47, Q54, Q58, C45, C53, C55

Abstract

EU member states have committed to reaching carbon neutrality by 2050. Since building-related activities are responsible for nearly 25% of the EU's greenhouse gas emissions, it is crucial to reduce emissions in this sector. However, policymakers need to be careful when acting on this asset class, as buildings are a major part of household wealth and bank assets. To make informed decisions, it is important to have accurate data on buildings' energy efficiency, such as the energy class reported in Energy Performance Certificates (EPCs). However, accurately assessing the energy efficiency of buildings remains a challenge, in Italy and elsewhere, given the limited availability of comprehensive data.

In this study, we developed a machine learning model to predict the energy class of Italian buildings using publicly available data. The model was trained on a specific geographic area in Italy and was able to correctly predict the energy class 37% of the time. However, if a margin of error of one class was allowed, the accuracy increased to 74%, improving upon standard techniques such as logistic regression. The results also raised concerns about the potential underreporting of buildings with the worst energy efficiency in the official EPC registry.

*The views expressed are those of the authors and do not necessarily reflect those of the institutions to which they belong. This paper is the result of the collaboration between the privately owned company Datasinc and the Bank of Italy, within the first Call for Proposals by the Bank of Italy's Milano Hub. The data utilized in this study were collected by Datasinc, a company acting under its sole responsibility. They were graciously provided to researchers at the Bank of Italy in an aggregated format. Neither the Bank of Italy nor its employees were involved in the data collection process or contributed to the creation of the dataset in any way. The policy brief is based on Braggiotti et al. (2024). "[Predicting buildings' EPC in Italy: a machine learning based approach](#)" Bank of Italy Occasional Papers, No. 850.

Introduction

With the European Union's commitment to achieving carbon neutrality by 2050, improving energy efficiency in buildings has become crucial. Buildings contribute significantly to greenhouse gas emissions (almost a quarter of all GHG emissions in the EU), and energy retrofitting is seen as a key strategy for reducing these emissions. In addition, energy consumption and energy prices impact the value of buildings. Changes in regulatory standards on the energy efficiency of buildings (i.e. the new Energy Performance of Building Directive in Europe) could lead to a deterioration in the value of those less energy efficient. And, accordingly, the bank's secured portfolio could be affected by a loss of value. In light of these potential transition risks, banks and supervisors consider the energy efficiency level (and certificate) as crucial data in the risk assessment process of credit granted by immovable property. However, accurately assessing the energy efficiency of buildings remains a challenge, given the limited availability of comprehensive data. In a [recent study](#) we address this gap in the context of Italy, by developing a machine learning model to predict the energy performance certificates (EPCs) of buildings using publicly accessible data¹.

The study utilized an open dataset comprising approximately 700,000 residential entries from Lombardy (and circa 130,000 from Piedmont as robustness check), the most affluent and populated Italian region, encompassing a wide range of building characteristics, such as age, size, value and floor, and geographical/climatic variables, such as altitude and average temperature. The data on the energy efficiency of the buildings (energy class) were obtained from the regional EPC registry, while building location, size and value were extracted from a revised, proprietary, version, of the Italian cadastre. This comprehensive public registry documents property ownership, boundaries, and physical characteristics across the country.

Leveraging Machine Learning for Accurate Energy Performance Predictions

The study employed a Random forest classifier (RFC) model, a type of ensemble learning method that builds multiple decision trees and merges their outcomes to improve predictive accuracy. This model was chosen for its ability to capture complex, potential non-linear relationships between the building characteristics and their energy performance. The RFC model proved effective, achieving an accuracy of 37% in predicting the exact energy class of buildings, which improved to 74% when a one-class margin (i.e. a class C might be in reality either a B or a D) of error was allowed. More importantly, these results are higher than what can be obtained with standard techniques such as logistic regression.

The analysis revealed that surface area and market value were among the most significant predictors of energy performance. These insights underline the model's ability to identify which characteristics most strongly influence a building's energy efficiency, providing valuable guidance for future energy policy and retrofitting efforts.

Revealing the Gaps: What Our Model Tells Us About Italy's Energy Efficiency

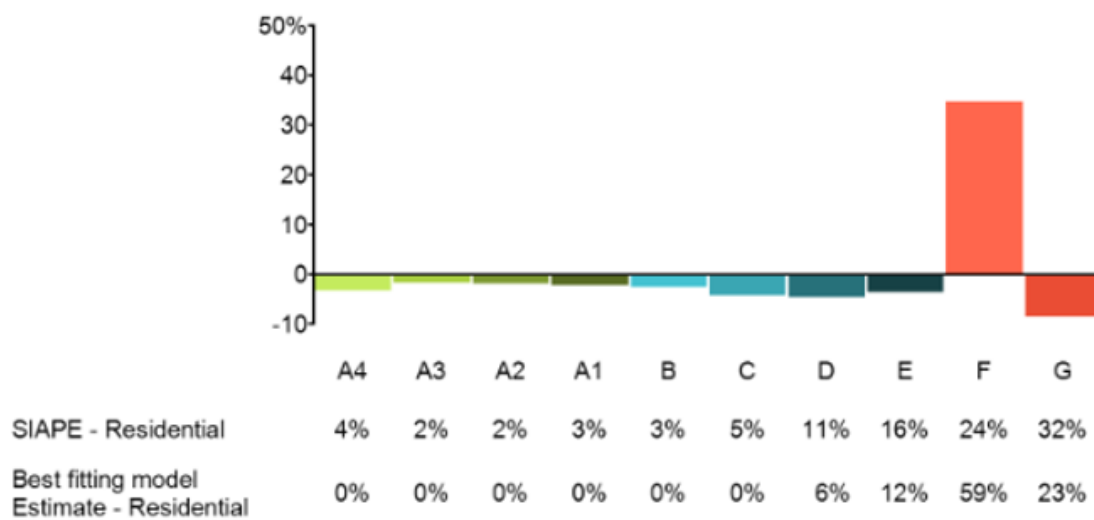
The model output uncovered discrepancies when compared with official EPC records. In particular, it identified a much higher proportion (82% vs 56%; fig.1) of buildings in the least efficient class, F. This suggests that official data might underrepresent less efficient buildings, probably due to biases in the data collection process. These findings underscore the potential underestimation of energy inefficiency in Italian buildings, which could have serious implications for energy policy design and implementation.

¹ EPC in Italy ranges from F (the worst energy-efficient class) to A4 (the most efficient). EPC is compulsory when selling, renting, or renovating any house. This might introduce a bias in the national cadaster (SIAPE).

A Wake-Up Call for Energy Policy: Confronting the Realities of Building Efficiency

This study hints at a potentially significant underestimation of energy inefficiency within Italy's building stock, suggesting that the problem may be more widespread than current official records indicate. While the model is not perfectly accurate, it offers valuable insights that underscore the need for more aggressive and targeted energy policies. These could include incentivizing comprehensive energy retrofits, revising regulations for both new and existing buildings, and expanding financial support for upgrading older, less efficient properties. As the EU advances towards stricter energy standards, this more realistic picture of building efficiency can guide the development of policies that effectively address the true scale of the challenge, paving the way for a more sustainable future.

Figure 1. Deviations in predictions vs. SIAPE data for national residential properties



About the authors

Francesco Braggiotti, CEO at Datasinc, with focus on Sales and Finance since 2020. He holds an MBA degree from Columbia University (New York), and a BA degree from the Università Cattolica del Sacro Cuore (Milano).

Nicola Chiarini is a Co-founder and President at Datasinc. Over 15 years of experience as Head of R&D, Data Architect, CRO, CTO, and Data Scientist in the financial services sector, leading technological development and advanced analytics projects. Master's degree in electronic engineering (2002), master's in finance (2003) from the University of Brescia, and a Master in Deep Learning from Bergamo Innovation District (2019).

Giulio Dondi, developer at DATASINC since 2022. Giulio holds a bachelor's degree in Physics from the University of Modena and Reggio Emilia with a thesis in Computational Astrophysics. Awarded the Erasmus study grant at Imperial College, London in 2017/2018. Awarded the master's degree in Theoretical Physics in 2019 at the University of Padova with a thesis on computational methods for high-energy particle physics. Awarded the Master in High-Performance Computing (MHPC) at SISSA, Trieste in 2021 with a thesis on computer-visualization over cadastral maps. Winner of the annual best thesis award. .

Luciano Lavecchia is an economist at the Climate Change and Sustainability Hub of the Bank of Italy. He is a member of the G7 Climate Change Mitigation Working Group (CCMWG), the Expert Network on Research, and the Task Force on Nature of the Network for Greening the Financial System (NGFS). From 2015 to 2018, he was seconded to the Technical Secretariat of the Italian Ministry of Economic Development. He is a fellow of the Istituto Bruno Leoni (IBL) and a co-founder of the Italian Observatory on Energy Poverty (OIPE). His work focuses on energy, energy poverty, and the sustainable data gap. He holds a degree in Economics and Social Sciences (DES) from Bocconi University in Milan and a Ph.D. in Economic Analysis from the University of Palermo.

Valeria Lionetti is a policy expert in climate risks and transition finance within banking regulation and supervision practices at the Climate and Sustainability Hub of the Bank of Italy. She is currently focused on addressing data gaps in energy-efficient buildings, green mortgages, and financial institutions' transition plans. She is also a member of the Supervision Workstream of the Network for Greening the Financial System (NGFS). Previously, she worked in the Bank of Italy's Supervision Directorate, where, from 2019 to 2022, she was involved in policy debates and analyses related to including climate and environmental risks in the bank prudential framework. She contributed to the ECB working group that developed supervisory expectations on climate-related and environmental risks, participated in the EBA Sustainable Finance Network, and was involved in the Workstream on Pillar 1 of the BCBS Task Force on climate-related risks.

Juri Marcucci holds a Ph.D. in Economics from UCSD and a doctoral degree from the Sant'Anna School of Advanced Studies in Pisa. He is an Advisor at the Bank of Italy's DG for Economics, Statistics, and Research and an adjunct professor at the Sapienza University of Rome. He is the organizer of the Italian Summer School of Econometrics for the Italian Econometric Association (SIDE) and co-organizer of the international webinar on Applied Machine Learning, Economics, and Data Science (AMLEDS). His work has focused on applying AI, Machine Learning, Natural Language Processing, and Big Data to macroeconomic forecasting and it appeared in the Journal of Econometrics, International Journal of Forecasting, and other peer-reviewed international journals. He has been a guest editor for the Journal of Econometrics, the International Journal of Forecasting, and Econometrics.

Riccardo Russo is a Data Scientist at the DG for Economics, Statistics, and Research of the Bank of Italy. His research interests focus on environmental, and climate change-related factors that affect economics and human activity.

Find more SUERF Policy Notes and Briefs at www.suerf.org/publications/suerf-policy-notes-and-briefs/



SUERF is a network association of central bankers and regulators, academics, and practitioners in the financial sector. The focus of the association is on the analysis, discussion and understanding of financial markets and institutions, the monetary economy, the conduct of regulation, supervision and monetary policy.

SUERF Policy Briefs (SPBs) serve to promote SUERF Members' economic views and research findings as well as economic policy-oriented analyses. They address topical issues and propose solutions to current economic and financial challenges. SPBs serve to increase the international visibility of SUERF Members' analyses and research.

The views expressed are those of the author (s) and not necessarily those of the institution(s) the author(s) is/are affiliated with.

Editorial Board

Ernest Gnan
David T. Llewellyn
Donato Masciandaro
Natacha Valla

SUERF Secretariat

c/o OeNB
Otto-Wagner-Platz 3
A-1090 Vienna, Austria
Phone: +43-1-40420-7206
www.suerf.org • suerf@oenb.at