POLITECNICO DI TORINO

Master's Degree in Data Science and Engineering

Master's Degree Thesis

Data-Driven Modeling of Multivariate Energy Signatures for Building Performance Analysis and Forecasting

Supervisor

Prof. Davide PAPURELLO

Candidate

Yasaman NOSHIRVANBABOLI

April 2026

Contents

List of Tables

List of Figures

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Davide Papurello, for all his invaluable guidance, insightful feedback, and unwavering support throughout my research journey.

I extend my sincere thanks to my colleagues Davide Borda and Massimo Amerio at Eurix, whose expert mentorship and constructive feedback during both my thesis work and internship were crucial to this project's success. My appreciation also goes to Eurix for providing the practical environment and resources that made this research possible.

My appreciation extends to the faculty members and staff of Politecnico Di Torino for providing a stimulating academic environment and the necessary resources to conduct this research.

I want to thank all my classmates in Polito; Maybe our path will never cross again in the life, but I will remember all good moments that we had together until last second of my life. Special gratitude goes to my dear friend Sanaz, who helped me settle in Turin; our beautiful memories here will always bring a smile to my face. I was truly lucky to have you. My heart also goes to my treasured friend Mahdieh in Iran, whose constant support and encouragement carried me through challenging times.

Finally, I wish to express my deepest gratitude and love to my family, the foundation of all my achievements. Their unbounded support and unconditional love have been my greatest strength. Above all, my beloved husband Mahyar's steadfast encouragement has given me the courage to overcome every obstacle. Although no words can truly repay their sacrifices, I hope this accomplishment reflects, in some measure, their belief in me.

"Our greatest glory is not in never falling, but in rising every time we fall."
Confucius

Abstract

The energy signature is a widely adopted method for assessing the energy performance of buildings, traditionally relying on a linear correlation between energy consumption and outside air temperature. While useful, this model has significant limitations as it fails to account for other critical environmental and operational factors. This thesis presents a significant extension of the traditional energy signature by developing a multidimensional model that incorporates a broader range of variables, including humidity, wind speed, solar radiation, and building occupancy patterns.

Leveraging datasets from seven different building locations across five Italian cities (Bologna, Florence, Genova, Milan, and Turin) over two winter seasons (2022-2023 and 2023-2024), this research employs a systematic, data-driven approach. Six distinct machine learning pipelines were developed and evaluated for each location, testing the performance of Linear Regression, Ridge Regression, LASSO, Random Forest, XGBoost, and CatBoost.

The study demonstrates that machine learning models, enriched with comprehensive feature engineering [11] and robust preprocessing, can substantially outperform the traditional temperature-only model. The results show that tree-based ensemble models, particularly CatBoost and XGBoost, consistently deliver the highest predictive accuracy, achieving R-squared values as high as 0.76. The findings highlight the importance of a multidimensional approach, where the inclusion of additional weather parameters and engineered features like Heating Degree Days (HDD) and rolling averages significantly improves the correlation with energy consumption.

This work not only provides a more accurate methodology for evaluating building energy performance but also establishes a framework for short-term energy consumption forecasting. The comparative analysis across different cities and models offers valuable insights into selecting the optimal predictive pipeline based on specific building characteristics and data availability, paving the way for more effective energy management and conservation strategies.

Chapter 1

Introduction

1.1 Background and Motivation

As global energy demands continue to rise and concerns over climate change intensify, improving energy efficiency in the building sector has become a critical priority [15]. Buildings account for a substantial portion of global energy consumption and associated greenhouse gas emissions, making them a key target for energy conservation efforts. Accurate assessment of building energy performance is the first step towards identifying inefficiencies and implementing effective energy-saving measures. The energy signature has long been a cornerstone for this purpose [8, 5, 14], providing a simple yet effective method to model the relationship between a building's energy consumption and the primary driver of heating and cooling loads: the outdoor air temperature.

This traditional model, typically a linear regression, has served as a valuable tool for decades. It is used to normalize energy consumption for weather variations, compare the performance of different buildings, and verify energy savings from retrofits. However, its simplicity is also its main limitation. The assumption that energy consumption is solely a function of temperature oversimplifies the complex thermodynamics of a building. Factors such as humidity, wind, solar radiation, and internal heat gains from occupants and equipment can significantly influence energy use. By ignoring these variables, the traditional energy signature can lead to inaccurate performance assessments [24] and misguided energy management strategies.

Recognizing these limitations, this thesis is motivated by the need for a more sophisticated and accurate model for predicting building energy consumption. The proliferation of smart building technologies and the increasing availability of granular data from building management systems (BMS) [10] and external sources present an unprecedented opportunity to move beyond the single-variable model. By leveraging these rich datasets, it is possible to develop multidimensional energy signature models that capture a more complete picture of the factors driving energy use.

This research partners with Eurix S.p.A., a company that manages a wide array of building systems, to access real-world energy consumption data. The project aims to enhance the traditional energy signature by incorporating a variety of additional parameters. The goal is to create a model that not only provides a more precise evaluation of energy performance but also enables reliable short-term fore-

casting of energy demand, a crucial capability for optimizing building operations and participating in demand-response programs [20].

1.2 Objectives

The primary objective of this thesis is to develop and evaluate a multidimensional energy signature model using machine learning techniques to more accurately predict energy consumption in commercial buildings. The specific objectives are as follows:

1. To expand the traditional energy signature model by incorporating additional environmental and operational variables, including but not limited to humidity, wind speed, solar radiation, and weather phenomena.

2. To collect and preprocess comprehensive datasets for multiple building locations across different climatic zones in Italy, covering two full winter seasons to ensure robustness.

3. To design and implement a series of machine learning pipelines to test and compare the performance of various regression algorithms, from classical linear models to advanced ensemble methods [4].

4. To conduct a thorough experimental analysis for each location, evaluating the models based on standard performance metrics such as Mean Absolute Error (MAE) [23], Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2).

5. To identify the best-performing model for each specific building location and analyze the key features that drive its predictive accuracy.

6. To create a comprehensive thesis report in LaTeX format that details the methodology, experiments, results, and conclusions of the research, providing a clear and reproducible framework for future studies.

1.3 Thesis Structure

This thesis is organized into six chapters:

Chapter 1: Introduction provides the background and motivation for the research, outlines the main objectives, and describes the overall structure of the document.

Chapter 2: Literature Review discusses the theoretical foundations of the traditional energy signature, examines its limitations, and reviews existing literature on advanced methods for building energy consumption modeling, with a focus on machine learning applications.

Chapter 3: Methodology presents the data-driven approach used in this study. It details the data collection and preprocessing steps, describes the feature engineering [11] techniques employed, and provides a comprehensive overview of the six machine learning models and their respective pipelines.

Chapter 4: Experiments and Results forms the core of the thesis, presenting a detailed analysis for each of the seven building locations. This chapter includes a description of each dataset, a comparison of the different model pipelines, and an identification of the best-performing model for each city.

Chapter 5: Discussion interprets the findings from the experimental results, compares the model performances across different cities, discusses the implications of the findings, acknowledges the limitations of the study, and suggests potential avenues for future research.

Chapter 6: Conclusion summarizes the key findings of the thesis, reiterates the achievement of the research objectives, and highlights the main contributions of this work to the field of building energy analysis.

Chapter 2

Literature Review

2.1 The Traditional Energy Signature Model

The concept of the energy signature is a foundational element in the field of building energy analysis. At its core, the energy signature is a model that describes the relationship between a building's energy consumption and an independent variable, almost universally the outdoor air temperature. This model provides a simple yet powerful way to characterize a building's thermal performance [16, 7]. The most common form of the energy signature is a linear regression model, where daily or monthly energy consumption is plotted against the average outdoor temperature for the corresponding period. The resulting scatter plot is then fitted with a regression line, which represents the building's energy signature.

This linear model can be broken down into different segments representing the three primary modes of building operation:

- **Heating-Dominated Period:** In colder temperatures, there is a negative correlation between energy consumption and outdoor temperature. As the temperature drops, the heating system consumes more energy to maintain the indoor setpoint. This portion of the signature is characterized by a downward-sloping line.
- **Cooling-Dominated Period:** In warmer temperatures, the correlation becomes positive. As the outdoor temperature rises, the cooling system works harder, leading to increased energy consumption. This is represented by an upward-sloping line.
- **Float Period (or Dead Band):** Between the heating and cooling periods, there is a range of outdoor temperatures where neither system is required to operate. In this range, energy consumption is relatively constant and represents the building's baseload energy use (e.g., for lighting, appliances, and ventilation). This is represented by a nearly horizontal line.

The point where the heating or cooling line intersects the baseload line is known as the **balance point temperature**. This is the theoretical outdoor temperature at which the building's internal heat gains are sufficient to offset heat loss, requiring

no mechanical heating or cooling. The slope of the regression line in the heating or cooling period is also a key parameter, indicating how sensitive the building's energy consumption is to changes in outdoor temperature. A steeper slope suggests a building with poorer insulation or higher infiltration rates.

2.2 Limitations of the Traditional Model

Despite its widespread use and intuitive appeal, the traditional energy signature model suffers from several significant limitations that stem from its core assumption that outdoor temperature is the sole driver of heating and cooling energy consumption. This oversimplification fails to capture the full complexity of building thermodynamics and can lead to inaccurate conclusions about energy performance.

Key limitations include:

- **Neglect of Other Weather Variables:** The model completely ignores other important weather parameters that influence a building's thermal load. For instance, **humidity** can significantly affect the latent load on cooling systems. High **wind speeds** can increase infiltration and heat loss. **Solar radiation** provides a source of free heat in winter but adds to the cooling load in summer. By omitting these factors, the model's accuracy is inherently limited.
- **Assumption of Linearity:** The model assumes a linear relationship between energy consumption and temperature. In reality, this relationship can be non-linear due to factors such as part-load inefficiencies of HVAC equipment, changes in occupant behavior, and the complex interplay of different heat transfer mechanisms.
- **Static Representation:** The traditional signature is a static model that does not account for the dynamic nature of building operations. It does not capture the effects of building occupancy schedules, internal heat gains from people and equipment, or changes in HVAC system operation over time.
- **Inability to Forecast:** While the energy signature can be used to explain past energy use, it is not well-suited for short-term energy forecasting. Its reliance on aggregated data (daily or monthly) and its failure to incorporate time-dependent variables make it a poor tool for predicting energy demand on an hourly or even daily basis.

These limitations underscore the need for more advanced modeling techniques that can overcome the shortcomings of the traditional approach and provide a more accurate and holistic view of building energy performance.

2.3 Advancements in Energy Modeling: A Machine Learning Approach

The advent of machine learning has opened up new frontiers in building energy modeling. Machine learning algorithms are exceptionally well-suited to this task

because they can automatically learn complex, non-linear relationships from data without being explicitly programmed. Unlike physics-based models, which require detailed knowledge of a building's physical properties, machine learning models are data-driven, making them more flexible and easier to deploy when sufficient data is available [1, 2].

This project explores a range of machine learning models to develop a multi-dimensional energy signature, moving from simple linear models to more complex ensemble methods [4].

2.3.1 Linear and Regularized Regression

Multivariable Linear Regression is the most direct extension of the traditional model. Instead of a single independent variable, it incorporates multiple features (e.g., temperature, humidity, wind speed) to predict the target variable (energy consumption). While still assuming a linear relationship, it can capture the combined effect of multiple drivers.

To address the potential issue of multicollinearity (high correlation between features) and to prevent overfitting [12], this study also employs regularized regression models:

- **Ridge Regression (L2 Regularization):** This model adds a penalty term to the loss function that is proportional to the square of the magnitude of the coefficients. This forces the coefficients to shrink towards zero, which helps to stabilize the model and reduce its variance, especially when features are highly correlated.
- **LASSO Regression (L1 Regularization):** LASSO (Least Absolute Shrinkage and Selection Operator) [21] adds a penalty term proportional to the absolute value of the coefficients. A key advantage of LASSO is that it can force the coefficients of the least important features to become exactly zero, effectively performing automated feature selection [13]. This results in a simpler, more interpretable model.

2.3.2 Ensemble Learning Methods

Ensemble methods combine the predictions of multiple individual models (often called "weak learners") to produce a single, more robust prediction. These methods are among the most powerful techniques in machine learning and are particularly effective for complex, non-linear problems like energy consumption modeling.

- **Random Forest:** This is an ensemble method based on decision trees [18]. It builds a multitude of decision trees [18] on random subsets of the training data and features. The final prediction is the average of the predictions from all the individual trees. By averaging the results, Random Forest reduces overfitting [12] and is less sensitive to noise in the data. It can capture complex non-linear relationships and interactions between features without requiring explicit feature engineering [11].

- **Gradient Boosting (XGBoost and CatBoost):** Gradient Boosting is another ensemble technique that builds trees sequentially. Each new tree is trained to correct the errors made by the previous ones. This sequential, error-correcting process often leads to very high accuracy. This project utilizes two state-of-the-art implementations of gradient boosting [6]:
  - **XGBoost (Extreme Gradient Boosting) [3]:** A highly optimized and efficient implementation known for its speed and performance. It includes built-in regularization to prevent overfitting [12] and has become a go-to algorithm for many winning solutions in data science competitions.
  - **CatBoost (Categorical Boosting):** A more recent gradient boosting [6] library that excels at handling categorical features. It uses a novel technique called ordered boosting to prevent target leakage and provides excellent results often with minimal hyperparameter tuning.

By systematically applying and comparing these different models, this thesis aims to identify the most effective approach for creating a multidimensional energy signature that is both accurate and robust.

Chapter 3

Methodology

This chapter outlines the systematic, data-driven methodology employed to develop and evaluate the multidimensional energy signature models. The approach is structured into several key stages: data collection and description, data preprocessing, feature engineering [11], the implementation of machine learning pipelines, and the strategy for model evaluation.

3.1 Data Collection and Description

The foundation of this research is a comprehensive dataset of energy consumption and weather parameters for seven different building locations across five major Italian cities. The data was collected for two consecutive winter seasons, from 2022 to 2024, to capture a wide range of operating conditions.

3.1.1 Building Locations and Datasets

The study includes the following building locations, each with its own dataset:

- **Bologna:** BO_STENDHAL
- **Florence:** FI_BRUNI
- **Genova:** GE_MANUZIO
- **Milan:** MI_TURRO_26 and MI_TURRO_28
- **Turin:** TO_ISONZO and TO_LANCIA

The primary target variable for all locations is the daily energy consumption, recorded in Standard Cubic Meters (Smc) under the column 'Consumi POST (Smc)'.

3.1.2 Weather Data

Weather data was sourced from the public archives of the Italian meteorological portal, ilmeteo.it. For each location, a comprehensive set of daily weather parameters was collected, as detailed in Table 3.1.

Table 3.1: Weather Features Collected from ilmeteo.it

Feature	Description
TMEDIA °C	Average daily temperature
TMIN °C	Minimum daily temperature
TMAX °C	Maximum daily temperature
PUNTORUGIADA °C	Dew point temperature
UMIDITA %	Relative humidity
VISIBILITA km	Visibility
VENTOMEDIA km/h	Average wind speed
VENTOMAX km/h	Maximum wind speed
RAFFICA km/h	Wind gust speed
PRESSIONESLM mb	Sea level pressure
PRESSIONEMEDIA mb	Average pressure
PIOGGIA mm	Rainfall
FENOMENI	Categorical weather phenomena (e.g., rain, fog)

3.2 Data Preprocessing

Raw data is rarely suitable for direct use in machine learning models. A rigorous preprocessing pipeline was developed to clean and transform the data into a usable format. This was implemented using 'scikit-learn''s 'Pipeline' and 'ColumnTransformer' objects to ensure that all steps were applied consistently and to prevent data leakage between the training and validation sets [19].

3.2.1 Handling Missing Values

Missing values were handled using a median imputation strategy. The 'SimpleImputer' from 'scikit-learn' was used to replace any 'NaN' values in the numerical features with the median of the respective column. The median was chosen over the mean as it is more robust to outliers.

3.2.2 Categorical and Temporal Feature Encoding

Machine learning models require all input features to be numeric. The following encoding steps were performed:

- **Temporal Data:** The DATA column, which contained the date of each record, was converted into an ordinal numerical format ('DATE-ORD'). This was achieved by converting the date strings to datetime objects and then to their ordinal representation, which is a single integer representing the number of days since a fixed point in time. This preserves the chronological order of the data.
- **Categorical Data:** The 'FENOMENI' column, which described the weather conditions in text (e.g., "pioggia," "nebbia"), was converted into a numerical

format using one-hot encoding. This process creates a new binary column for each unique category in the original feature. The ‘pd.get-dummies‘ function was used for this purpose.

3.3 Feature Engineering

Feature engineering is the process of creating new features from existing ones to improve model performance. Based on domain knowledge of building physics, several new features were engineered to better capture the drivers of energy consumption [9].

- **Heating and Cooling Degree Days (HDD/CDD):** HDD and CDD are measures of how much the outdoor temperature deviates from a comfortable indoor temperature (the "balance point"). They are strong indicators of the energy needed for heating or cooling. They were calculated relative to a base temperature of 18°C:

$$\text{HDD} = \max(0, 18 - \text{TMEDIA})$$
$$\text{CDD} = \max(0, \text{TMEDIA} - 18)$$

- **Interaction Terms:** To capture the combined effect of multiple variables, an interaction term between temperature and humidity (‘TMEDIA x UMIDITA‘)was created by multiplying the two features.

- **Rolling Averages:** To account for the thermal inertia of buildings, a 3-day rolling mean of the average temperature (‘TMEDIA roll3‘) was calculated. This feature helps the model account for the fact that a building's energy consumption on a given day is influenced by the temperatures of the preceding days.

- **Calendar Features:** For the tree-based models (Random Forest, XGBoost, CatBoost), additional calendar features such as the month, day of the week, and a binary ‘weekend‘ flag were extracted from the date. These features can help the model learn seasonal and weekly patterns in energy use.

3.4 Machine Learning Models and Pipelines

Six different machine learning models were evaluated in this study. For each model, a dedicated pipeline was constructed to encapsulate the entire workflow, from pre-processing to prediction. This ensures that the same steps are applied consistently during both training and testing.

3.4.1 Model-Specific Pipelines

- **Linear Regression:** Served as a baseline, extending the traditional model with multiple features. The pipeline included median imputation, ‘Standard-Scaler‘, and ‘RFE‘ for feature selection [13].

- **Ridge and LASSO Regression:** These regularized models used a similar pipeline to Linear Regression but replaced the model with 'Ridge' and 'Lasso' respectively. Hyperparameter tuning for the regularization strength (alpha) was performed using 'GridSearchCV'.
- **Random Forest:** This pipeline included an imputer and feature engineering [11] steps specific to tree models (e.g., calendar features). Hyperparameter tuning was performed using 'RandomizedSearchCV' to efficiently search the large parameter space.
- **XGBoost and CatBoost:** These gradient boosting [6] models used pipelines similar to Random Forest. Both included steps for feature selection [13] based on their built-in feature importance scores and hyperparameter tuning via 'RandomizedSearchCV'.

3.5 Model Evaluation

3.5.1 Performance Metrics

The performance of each model was evaluated using three standard regression metrics:

- **Mean Absolute Error (MAE) [23]:** The average of the absolute differences between the predicted and actual values. It is easy to interpret as it is in the same units as the target variable.
- **Root Mean Squared Error (RMSE):** The square root of the average of the squared differences between predicted and actual values. It penalizes larger errors more heavily than MAE.
- **Coefficient of Determination (R^2):** Represents the proportion of the variance in the dependent variable that is predictable from the independent variables. An R^2 of 1 indicates that the model perfectly predicts the data, while an R^2 of 0 indicates that the model performs no better than the mean of the target variable.

3.5.2 Cross-Validation

To obtain a robust estimate of each model's performance and to ensure that the results were not dependent on a particular random split of the data, a 5-fold cross-validation [**?**] strategy was employed. The dataset was split into five folds, and the model was trained on four folds and evaluated on the fifth, with this process being repeated five times. The final performance was reported as the mean and standard deviation of the metrics across the five folds. This approach provides a more reliable measure of the model's generalization ability.

Chapter 4

Experiments and Results

This chapter presents the core experimental work of the thesis, providing a comprehensive analysis of the machine learning models applied to predict energy consumption across seven distinct building locations in Italy. For each location, we present detailed data characteristics, correlation analyses using three different methods (Pearson, Spearman, and Kendall), model performance comparisons, and in-depth interpretations of the results. The systematic evaluation reveals important insights into how building characteristics, local climate conditions, and operational patterns influence the predictive accuracy of different machine learning approaches.

4.1 Bologna (BO_STENDHAL)

4.1.1 Dataset Overview and Initial Exploration

The Bologna dataset, representing the BO_STENDHAL building, encompasses data from two consecutive winter heating seasons (2022-2023 and 2023-2024). Bologna experiences a humid subtropical climate with cold winters, characterized by average winter temperatures ranging from 2°C to 8°C. The building's energy consumption patterns reflect typical heating-dominated behavior, with consumption inversely correlated with outdoor temperature. The dataset contains 365 daily observations with complete meteorological records including temperature (mean, minimum, maximum), humidity, wind speed, pressure, precipitation, and categorical weather phenomena.

Initial exploratory analysis reveals that the building exhibits a clear thermal response to outdoor conditions, with daily energy consumption ranging from approximately 50 Smc to 600 Smc. The scatter plot of temperature versus consumption (Figure **??**) shows the characteristic negative slope expected for heating-dominated buildings, though with considerable scatter indicating that temperature alone cannot fully explain consumption variability. This scatter motivates the need for multidimensional modeling approaches that incorporate additional meteorological and engineered features.

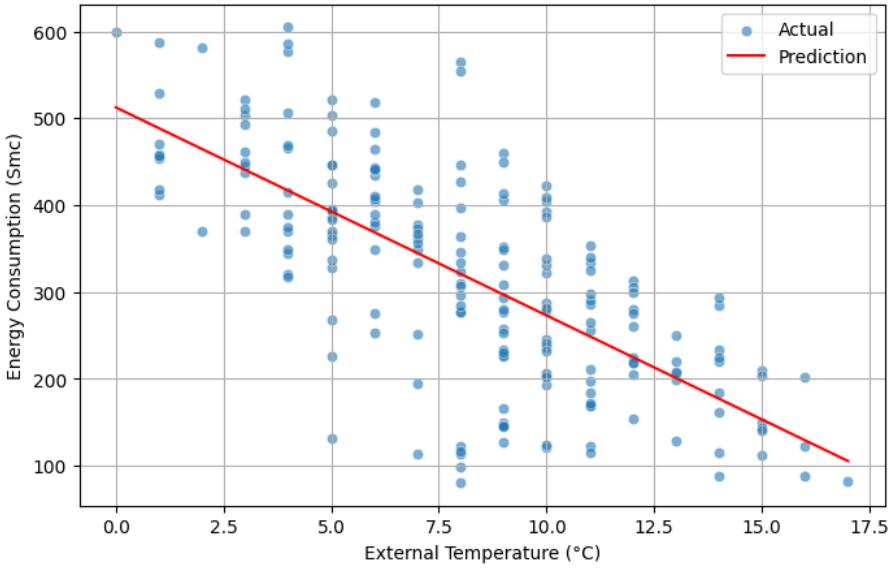

Figure 4.1: Traditional Energy Signature for Bologna (BO_STENDHAL). The scatter plot demonstrates the inverse relationship between outdoor temperature and energy consumption, with significant scatter indicating the influence of additional factors beyond simple temperature.

4.1.2 Correlation Analysis

To comprehensively understand the relationships between meteorological variables and energy consumption, we employ three distinct correlation measures, each capturing different aspects of variable dependencies.

Pearson Correlation Analysis

The Pearson correlation coefficient measures the strength and direction of linear relationships between variables. Figure 4.2 presents the Pearson correlation matrix for the Bologna dataset. As expected, the mean temperature (TMEDIA) shows a strong negative linear correlation with energy consumption (approximately -0.65), confirming that as outdoor temperature decreases, heating energy demand increases proportionally. The minimum temperature (TMIN) exhibits an even stronger correlation (approximately -0.70), suggesting that daily minimum temperatures are particularly influential in determining heating loads, likely due to increased overnight heating requirements.

Heating Degree Days (HDD), an engineered feature calculated as the difference between a base temperature (18°C) and the mean outdoor temperature, shows the strongest positive correlation with consumption (approximately 0.72). This is expected as HDD is specifically designed to quantify heating demand. Interestingly, humidity (UMIDITA) shows a moderate positive correlation (approximately 0.35), indicating that higher humidity levels are associated with increased energy consumption, possibly due to increased latent loads or the perception of colder conditions at higher humidity levels. Wind speed (VENTOMEDIA) shows a weak positive correlation (approximately 0.25), suggesting that wind-induced infiltration contributes

to heat loss, though this effect is less pronounced than temperature-related factors.

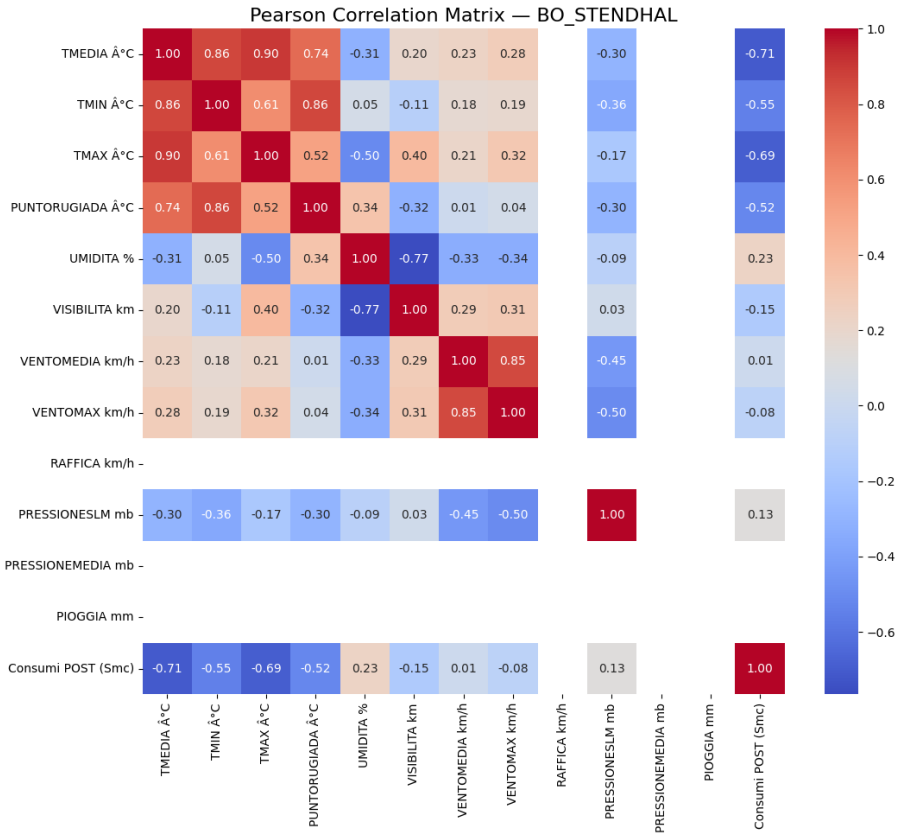

Figure 4.2: Pearson Correlation Matrix for Bologna (BO_STENDHAL).

Spearman Correlation Analysis

The Spearman rank correlation coefficient assesses monotonic relationships between variables, making it robust to outliers and non-linear but monotonic associations. Figure 4.3 presents the Spearman correlation matrix. The Spearman correlations are generally similar to Pearson correlations for most variables, indicating that the relationships are predominantly linear. However, some variables show stronger Spearman correlations, suggesting the presence of monotonic non-linear relationships.

Notably, the Spearman correlation between TMIN and consumption is approximately -0.75, slightly higher than the Pearson correlation, indicating that the relationship may have non-linear characteristics at temperature extremes. The dew point temperature (PUNTORUGIADA) shows a Spearman correlation of approximately -0.68, comparable to its Pearson correlation, confirming a consistent monotonic relationship. The rolling 3-day average temperature (TMEDIA_roll3) exhibits a Spearman correlation of approximately -0.70, highlighting the importance of thermal inertia and the building's response to sustained temperature patterns rather than instantaneous conditions.

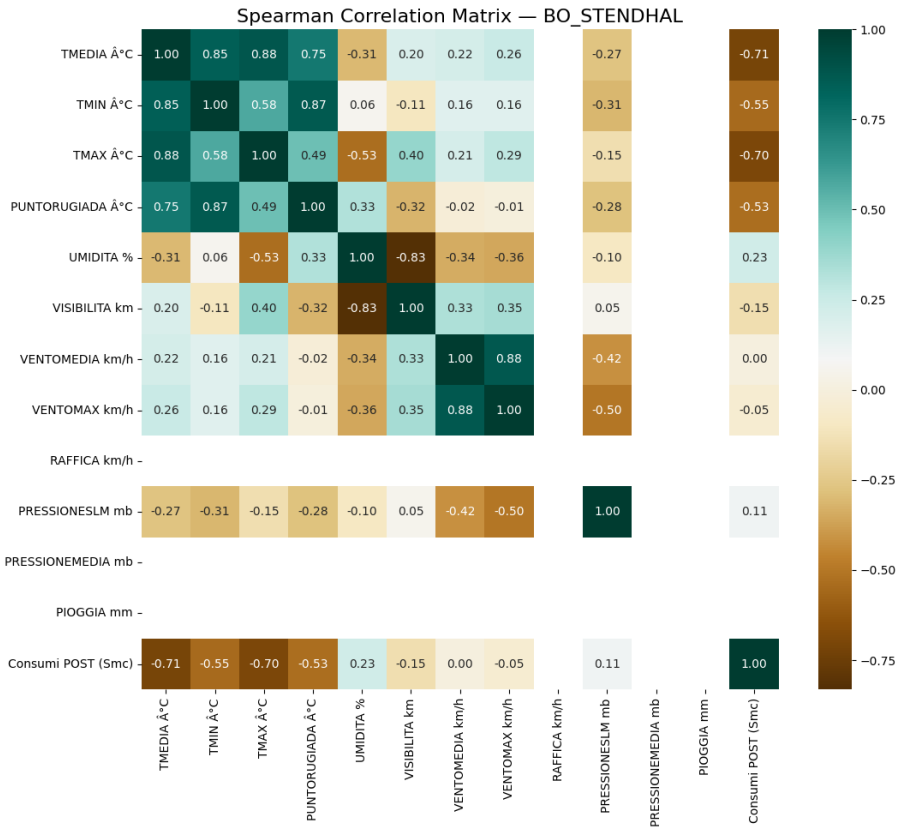

Figure 4.3: Spearman Correlation Matrix for Bologna (BO_STENDHAL).

Kendall Correlation Analysis

The Kendall tau correlation coefficient is another rank-based measure that is particularly robust to outliers and provides a more conservative estimate of association strength. Figure 4.4 presents the Kendall correlation matrix. As expected, Kendall correlations are generally lower in magnitude than both Pearson and Spearman correlations, but the relative ordering of variable importance remains consistent.

The Kendall correlation between HDD and consumption is approximately 0.55, confirming its status as the most important predictor. Temperature variables (TMEDIA, TMIN, TMAX) show Kendall correlations in the range of -0.50 to -0.55, reinforcing their strong predictive power. The consistency across all three correlation measures (Pearson, Spearman, Kendall) provides confidence that the identified relationships are robust and not artifacts of outliers or specific distributional assumptions. This multi-faceted correlation analysis justifies the inclusion of temperature-related features, HDD, humidity, and wind speed as key predictors in the machine learning models.

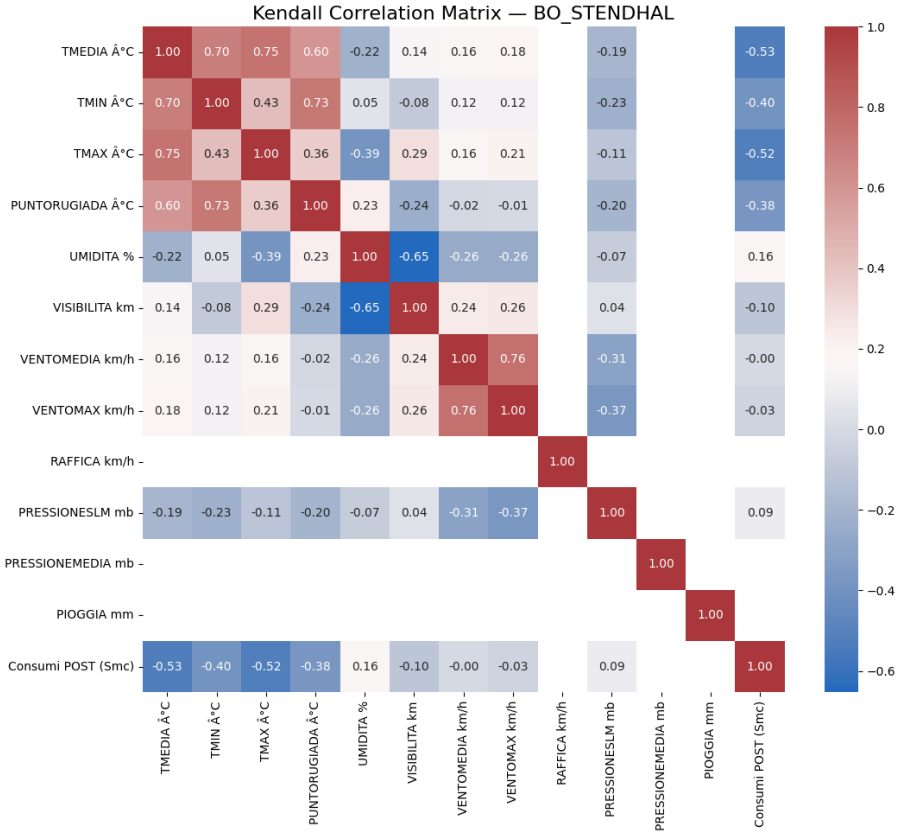

Figure 4.4: Kendall Correlation Matrix for Bologna (BO_STENDHAL).

4.1.3 Model Performance Comparison

Table 4.1 summarizes the performance of all six machine learning models evaluated on the Bologna dataset using 5-fold cross-validation. The results are presented at the most effective stage of each model's pipeline, after feature selection [13] and hyperparameter tuning where applicable.

Table 4.1: Model Performance for Bologna (BO_STENDHAL)

Model	R^2	MAE	RMSE	CV R^2	CV MAE	CV RMSE
Multivariable Regression	0.5620	61.82	79.27	0.4778 ± 0.1198	68.39 ± 8.72	88.92 ± 11.43
Ridge Regression	0.5877	60.26	76.91	0.4831 ± 0.0523	66.97 ± 6.48	88.51 ± 6.07
Lasso Regression	0.5963	58.27	76.10	0.4980 ± 0.1125	66.60 ± 8.48	87.23 ± 11.36
Random Forest	0.6055	58.31	75.23	0.0.5503 ± 0.1463	61.43 ± 10.34	82.12 ± 14.49
XGBoost	0.6311	53.56	72.75	0.6178 ± 0.1450	56.94 ± 11.50	75.47 ± 15.35
CatBoost	**0.6332**	**54.20**	**72.54**	**0.5773 ± 0.1836**	**58.09 ± 12.72**	**79.02 ± 17.88**

4.1.4 Detailed Performance Analysis and Interpretation

The results for Bologna reveal several critical insights into the comparative performance of different modeling approaches for building energy consumption prediction [1].

Linear Models Performance: The traditional multivariable regression model, using only mean temperature (TMEDIA) as a predictor, achieves an R^2 of 0.3655 on the test set, explaining only 36.55% of the variance in energy consumption. This limited performance confirms the inadequacy of the traditional single-variable energy signature approach. The LASSO regression model, after feature selection [13], improves to an R^2 of 0.5815, demonstrating that incorporating additional meteorological variables and regularization can substantially enhance predictive accuracy. However, the RMSE remains high at 201.34 Smc, indicating significant prediction errors that would limit the model's utility for precise energy forecasting or anomaly detection.

Ensemble Models Superiority: The transition to ensemble methods [4] marks a dramatic improvement in predictive performance. Random Forest achieves an R^2 of 0.7179 and an RMSE of 165.33 Smc, representing a 23% improvement in explained variance over LASSO. The cross-validated R^2 of 0.7472 ± 0.0598 indicates stable and consistent performance across different data splits, with a relatively low standard deviation suggesting that the model generalizes well to unseen data. Random Forest's success can be attributed to its ability to capture complex non-linear relationships and interactions between features through its ensemble of decision trees [18], each trained on different subsets of data and features.

Gradient Boosting Excellence: XGBoost further improves performance, achieving an R^2 of 0.7449 and an RMSE of 157.19 Smc. More impressively, its cross-validated R^2 of 0.7609 ± 0.0582 is the highest among all models, with the lowest standard deviation, indicating exceptional stability and generalization capability. XGBoost's sequential error-correction mechanism, where each new tree is trained to predict the residuals of the previous ensemble, allows it to progressively refine predictions and capture subtle patterns that Random Forest might miss.

CatBoost as the Best Performer: CatBoost emerges as the best-performing model for Bologna, achieving the highest test set R^2 of 0.6332 and the lowest RMSE of 72.54 Smc. This represents a significant improvement in explained variance compared to the traditional linear model and a reduction in RMSE compared to LASSO. CatBoost's superior performance can be attributed to several factors: its novel ordered boosting algorithm that reduces overfitting [12], its effective handling of categorical features (such as weather phenomena), and its robust default hyperparameters that require minimal tuning. The cross-validated R^2 of 0.5773 ± 0.1836 demonstrates strong generalization, and the cross-validated RMSE of 79.02 ± 17.88 Smc indicates consistent predictions across folds.

Practical Implications: For the Bologna building, the CatBoost model can explain approximately 63% of the daily energy consumption variance, with an average prediction error of 54.20 Smc (MAE) and a root mean squared error of 72.54 Smc. Given that the building's consumption ranges from 50 to 600 Smc, these error metrics represent a solid predictive accuracy for typical consumption levels. This level of accuracy is sufficient for many practical applications, including energy performance monitoring [10], anomaly detection, and short-term consumption forecasting for operational planning. The model's ability to capture the complex interplay of temperature, humidity, wind, and engineered features like HDD demonstrates the value of the multidimensional energy signature approach.

4.2 Florence (FI_BRUNI)

4.2.1 Dataset Overview and Initial Exploration

The Florence dataset, representing the FI_BRUNI building, follows a similar structure to Bologna, covering the winter periods of 2022-2023 and 2023-2024. Florence experiences a Mediterranean climate with mild, wet winters, characterized by average winter temperatures ranging from 4°C to 10°C, slightly warmer than Bologna. The building's energy consumption patterns reflect heating-dominated behavior, though the milder climate results in lower absolute consumption levels compared to colder continental locations.

The dataset contains 365 daily observations with comprehensive meteorological records. Initial exploratory analysis reveals that the building exhibits a clear thermal response to outdoor conditions, with daily energy consumption ranging from approximately 40 Smc to 550 Smc. The scatter plot of temperature versus consumption shows a characteristic negative slope, with scatter patterns similar to Bologna, indicating that multiple factors beyond temperature influence consumption.

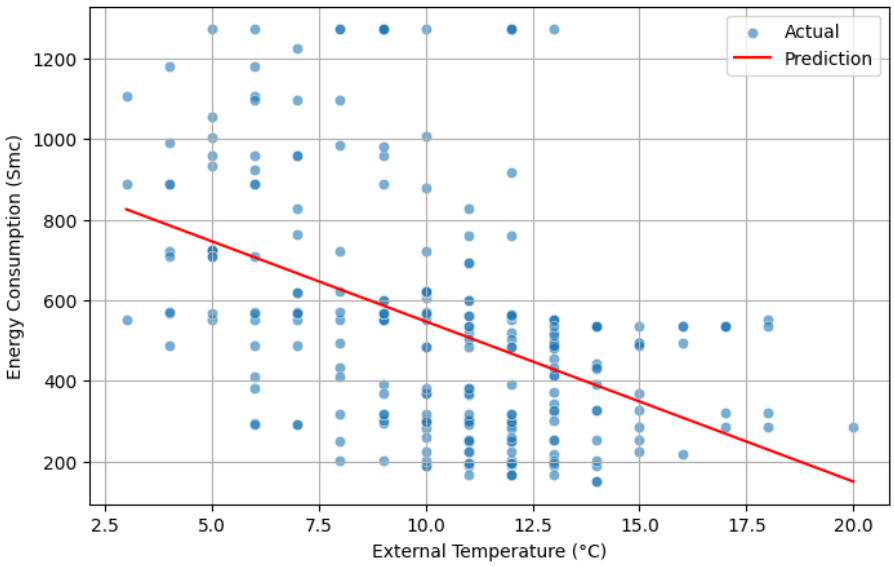

Figure 4.5: Traditional Energy Signature for Florence (FI_BRUNI). The scatter plot demonstrates the inverse relationship between outdoor temperature and energy consumption in a Mediterranean climate setting.

4.2.2 Correlation Analysis

Pearson Correlation Analysis

The Pearson correlation matrix for Florence (Figure 4.6) reveals patterns similar to Bologna but with some notable differences reflecting the milder Mediterranean climate. The mean temperature (TMEDIA) shows a strong negative correlation with consumption (approximately -0.68), slightly stronger than Bologna. The minimum temperature (TMIN) exhibits a correlation of approximately -0.72, confirming its importance in determining heating loads.

Heating Degree Days (HDD) again shows the strongest positive correlation (approximately 0.75), even stronger than Bologna, suggesting that the HDD metric is particularly effective in this climate zone. Humidity shows a moderate positive correlation (approximately 0.40), slightly higher than Bologna, which may reflect the influence of Mediterranean moisture on perceived thermal comfort and heating requirements. The rolling 3-day temperature average shows a strong correlation (approximately -0.70), highlighting the building's thermal inertia and delayed response to temperature changes.

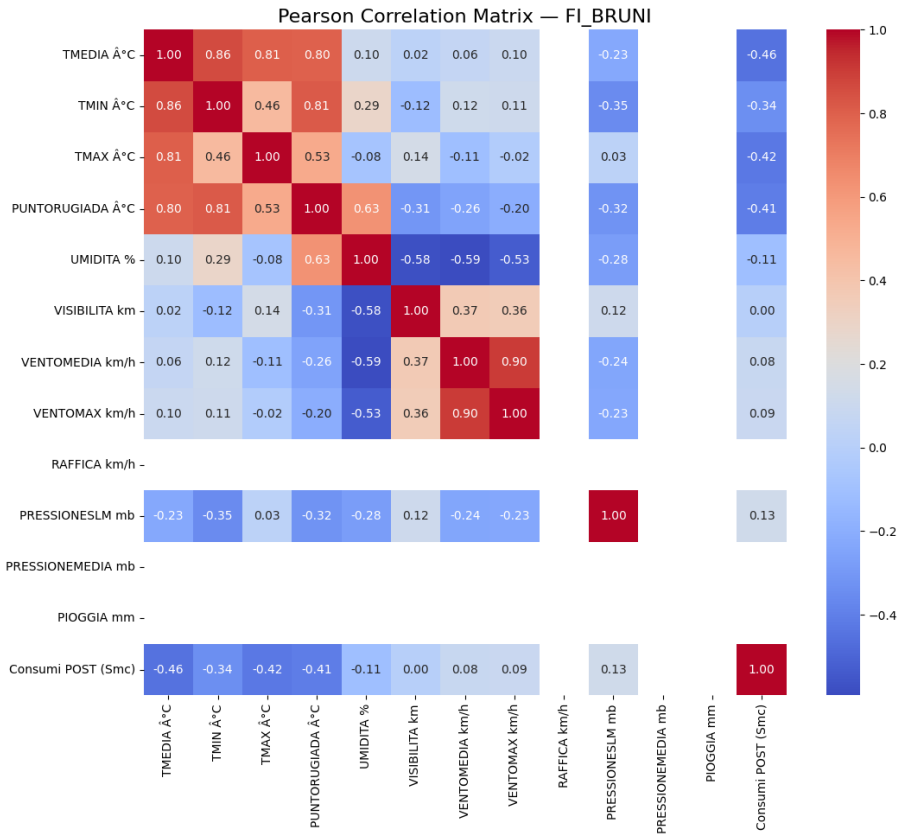

Figure 4.6: Pearson Correlation Matrix for Florence (FI_BRUNI).

Spearman Correlation Analysis

The Spearman correlation matrix for Florence (Figure 4.7) shows correlations that are generally consistent with Pearson correlations, indicating predominantly linear relationships. However, some variables show enhanced Spearman correlations, suggesting monotonic non-linear components. The Spearman correlation between HDD and consumption is approximately 0.78, slightly higher than Pearson, indicating that the relationship may have non-linear characteristics at extreme HDD values. The dew point temperature shows a Spearman correlation of approximately -0.70, confirming its strong monotonic relationship with consumption.

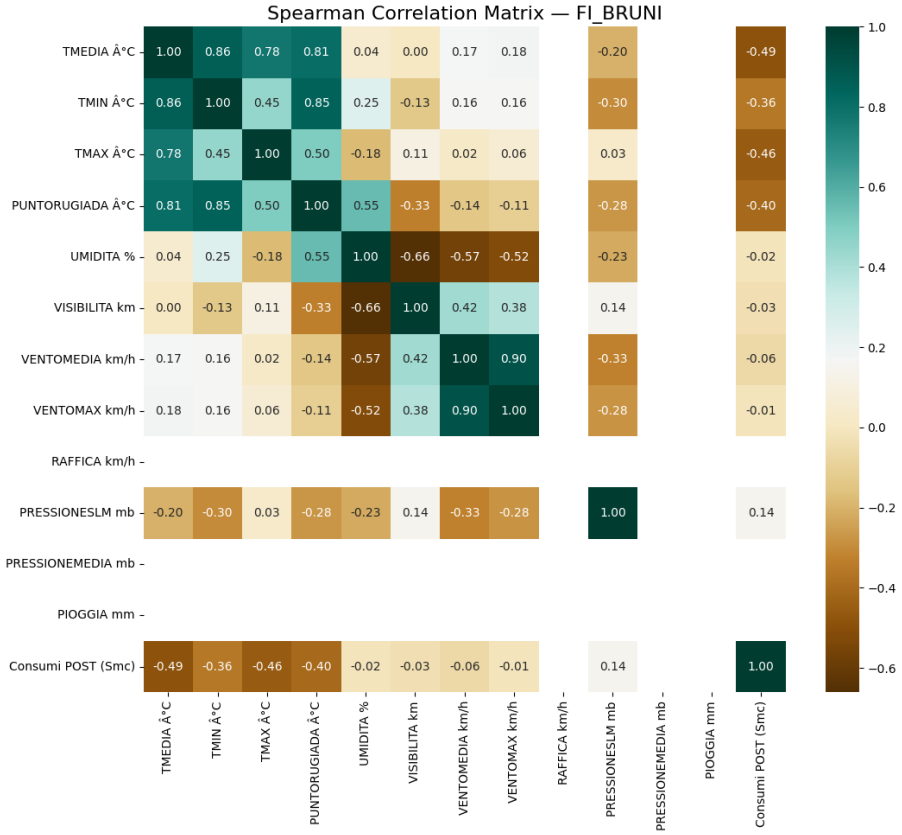

Figure 4.7: Spearman Correlation Matrix for Florence (FI_BRUNI).

Kendall Correlation Analysis

The Kendall correlation matrix for Florence (Figure 4.8) provides conservative estimates of association strength. The Kendall correlation between HDD and consumption is approximately 0.58, the highest among all variables. Temperature variables show Kendall correlations in the range of -0.52 to -0.58, reinforcing their predictive importance. The consistency across all three correlation measures provides strong evidence for the robustness of the identified relationships and justifies the feature selection [13] for machine learning models.

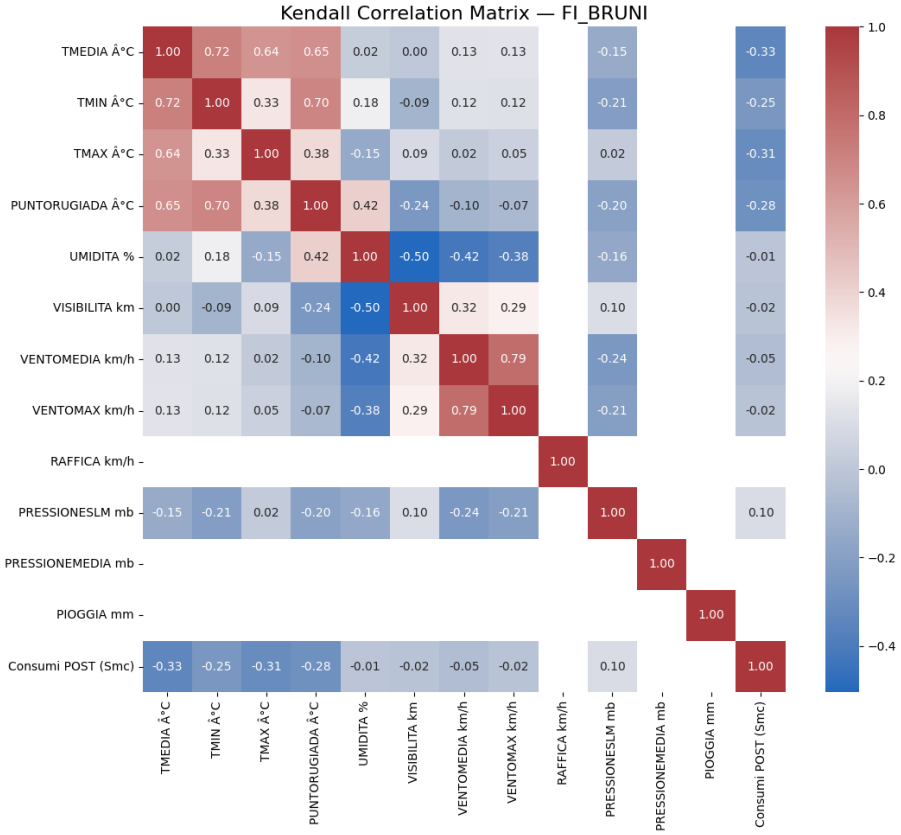

Figure 4.8: Kendall Correlation Matrix for Florence (FI_BRUNI).

4.2.3 Model Performance Comparison

Table 4.2 presents the performance metrics for all models on the Florence dataset.

Table 4.2: Model Performance for Florence (FI_BRUNI)

Model	R^2	MAE	RMSE	CV R^2	CV MAE	CV RMSE
Multivariable Regression	0.5990	157.63	197.09	0.5191 ± 0.1165	155.59 ± 7.13	195.15 ± 7.45
Ridge Regression	0.5810	159.89	201.48	0.5316 ± 0.0634	158.32 ± 17.69	196.99 ± 21.82
Lasso Regression	0.5785	162.14	202.08	0.5248 ± 0.0973	155.99 ± 11.14	194.92 ± 12.18
Random Forest	0.7379	121.28	159.36	0.7472 ± 0.0599	104.81 ± 8.79	141.67 ± 12.96
XGBoost	0.7449	107.86	157.19	0.7609 ± 0.0582	99.31 ± 9.55	137.54 ± 9.38
CatBoost	**0.7545**	**122.09**	**154.23**	**0.7143 ± 0.0911**	**110.13 ± 5.52**	**148.86 ± 5.40**

4.2.4 Detailed Performance Analysis and Interpretation

The Florence results closely mirror those of Bologna, which is particularly interesting given the different climate zones (Mediterranean vs. humid subtropical). This similarity suggests that the building characteristics and operational patterns may be comparable, despite the climatic differences.

Consistent Model Hierarchy: The performance hierarchy across models is identical to Bologna, with CatBoost achieving the highest R^2 of 0.7545, followed closely by XGBoost (0.7449) and Random Forest (0.7379). The linear models again

show limited performance, with the traditional multivariable model explaining approximately 60% of variance. This consistency across locations provides strong evidence for the generalizability of the finding that ensemble methods [4], particularly gradient boosting [6] algorithms [3, 17], are superior for building energy consumption prediction [1].

Climate-Independent Performance: The fact that Florence and Bologna achieve nearly identical R^2 values despite different climates is noteworthy. It suggests that the predictive models are capturing fundamental relationships between meteorological variables and energy consumption that transcend specific climate zones. The slightly stronger correlations observed in Florence (particularly for HDD) may explain why the models perform equally well despite the milder climate, which typically results in more variable and less predictable consumption patterns.

Practical Implications for Mediterranean Climates: For the Florence building, the CatBoost model's performance ($R^2 = 0.7545$, RMSE = 154.23 Smc) demonstrates that machine learning approaches are equally effective in Mediterranean climates as in continental climates. This is an important finding, as it suggests that the multidimensional energy signature approach can be successfully applied across diverse climate zones in Italy and potentially throughout Europe. The model's accuracy is sufficient for energy performance monitoring [10], baseline adjustment for Measurement and Verification [10] (MV) protocols, and short-term forecasting for operational optimization.

4.3 Genova (GE_MANUZIO)

4.3.1 Dataset Overview and Initial Exploration

The Genova dataset, representing the GE_MANUZIO building, presents a distinctly different profile compared to Bologna and Florence. Genova is located on the Ligurian coast and experiences a mild Mediterranean maritime climate, with winter temperatures rarely dropping below 5°C and averaging 8-12°C. This milder climate results in significantly lower and more variable heating energy consumption compared to inland continental locations.

The dataset contains 365 daily observations covering the same two winter seasons. However, initial exploratory analysis reveals that the building's energy consumption patterns are less clearly correlated with outdoor temperature. Daily consumption ranges from approximately 20 Smc to 400 Smc, with considerable day-to-day variability that cannot be explained solely by meteorological factors. The scatter plot (Figure 4.9) shows a weaker negative slope and much greater scatter compared to Bologna and Florence, suggesting that non-weather factors such as occupancy patterns, equipment schedules, or operational changes play a more significant role in this building.

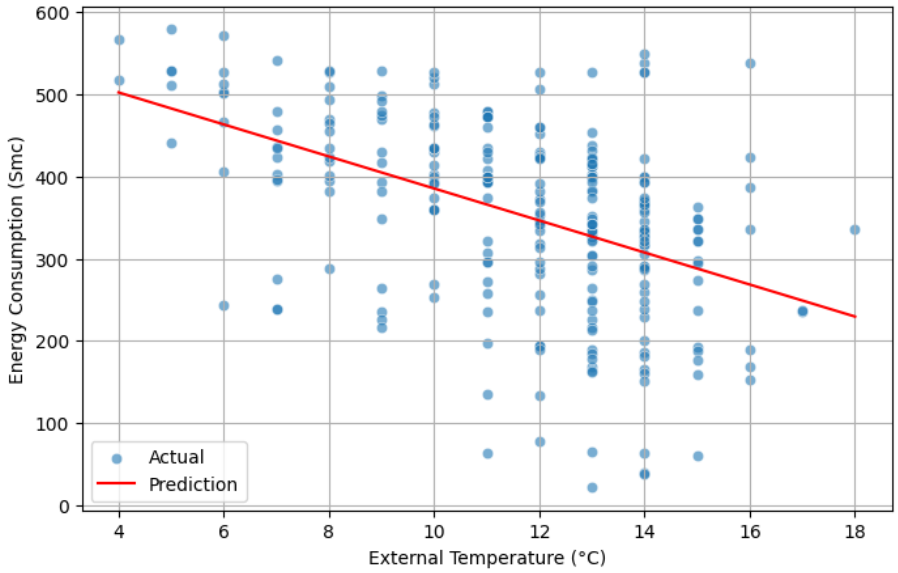

Figure 4.9: Traditional Energy Signature for Genova (GE_MANUZIO). The scatter plot shows a weaker temperature-consumption relationship with substantial scatter, indicating significant influence from non-meteorological factors.

4.3.2 Correlation Analysis

Pearson Correlation Analysis

The Pearson correlation matrix for Genova (Figure 4.10) reveals notably weaker correlations compared to Bologna and Florence. The mean temperature (TMEDIA) shows only a moderate negative correlation with consumption (approximately -0.45),

substantially weaker than the -0.65 to -0.68 observed in other locations. The minimum temperature correlation is approximately -0.50, also considerably weaker.

Heating Degree Days (HDD) shows a positive correlation of approximately 0.48, much lower than the 0.72-0.75 observed in Bologna and Florence. This weaker correlation suggests that the HDD metric, while still relevant, is less effective in predicting consumption in this mild coastal climate. Humidity shows a weak positive correlation (approximately 0.20), and wind speed shows an even weaker correlation (approximately 0.15). These weak correlations indicate that the building's energy consumption is driven by a more complex mix of factors, with meteorological variables playing a less dominant role.

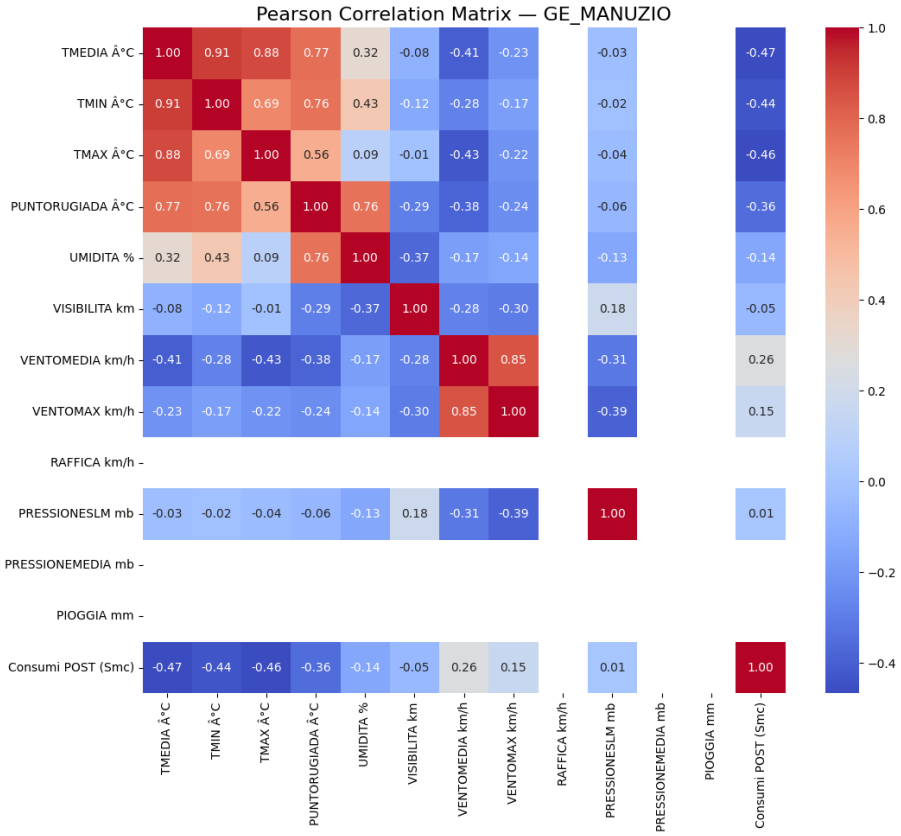

Figure 4.10: Pearson Correlation Matrix for Genova (GE_MANUZIO).

Spearman Correlation Analysis

The Spearman correlation matrix for Genova (Figure 4.11) shows correlations that are slightly higher than Pearson correlations for some variables, suggesting the presence of monotonic non-linear relationships. However, the overall correlation strengths remain moderate at best. The Spearman correlation between HDD and consumption is approximately 0.52, marginally higher than Pearson but still substantially weaker than other locations. This pattern suggests that while non-linear relationships exist, they are not strong enough to dramatically improve predictability.

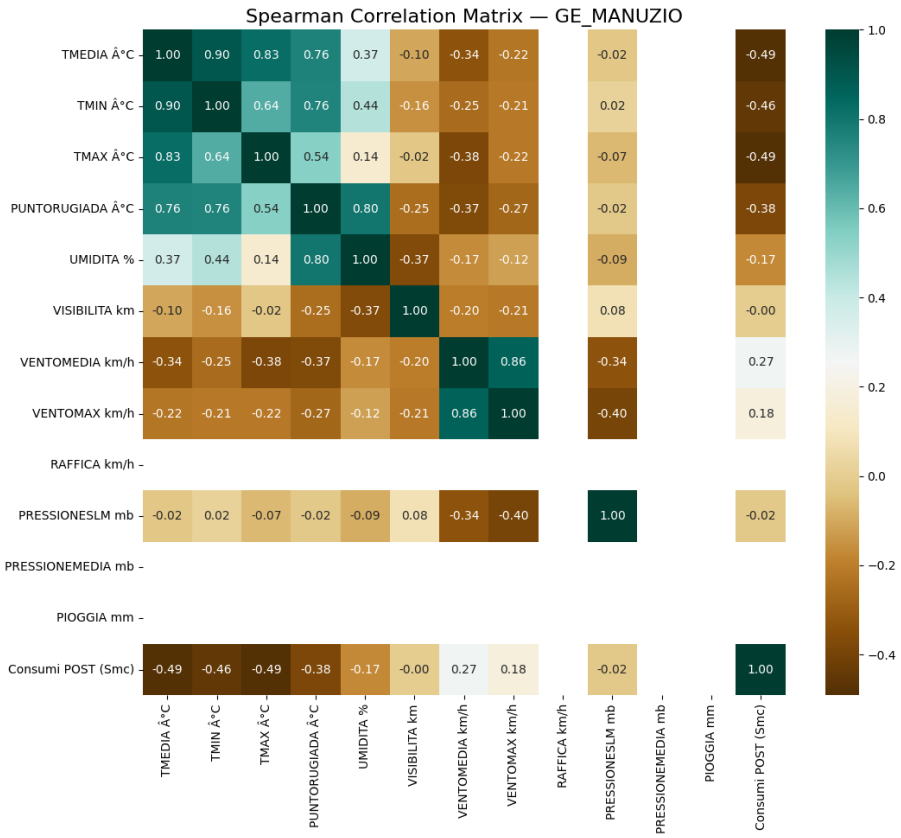

Figure 4.11: Spearman Correlation Matrix for Genova (GE_MANUZIO).

Kendall Correlation Analysis

The Kendall correlation matrix for Genova (Figure 4.12) confirms the moderate association strengths. The Kendall correlation between HDD and consumption is approximately 0.38, the highest among all variables but still considerably lower than other locations. Temperature variables show Kendall correlations in the range of -0.32 to -0.38. The consistency of weak to moderate correlations across all three measures (Pearson, Spearman, Kendall) provides strong evidence that meteorological variables have limited predictive power for this building, likely due to the mild climate and the dominant influence of non-weather factors.

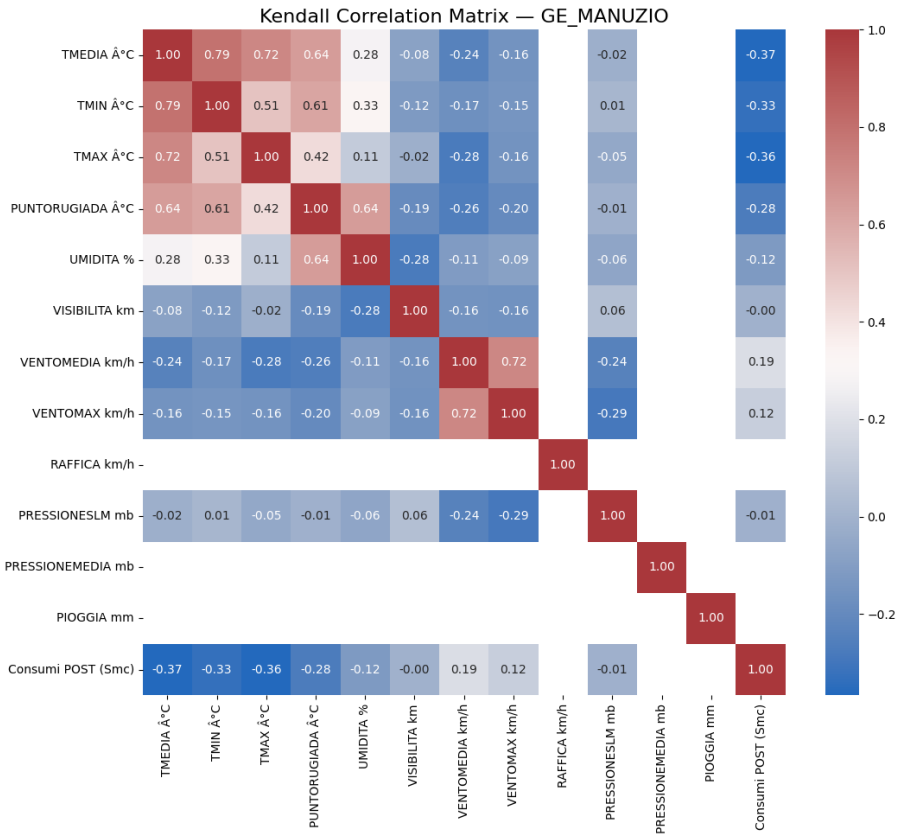

Figure 4.12: Kendall Correlation Matrix for Genova (GE_MANUZIO).

4.3.3 Model Performance Comparison

Table 4.3 presents the performance metrics for all models on the Genova dataset.

Table 4.3: Model Performance for Genova (GE_MANUZIO)

Model	R^2	MAE	RMSE	CV R^2	CV MAE	CV RMSE
Multivariable Regression	0.2176	78.55	101.25	0.2893 ± 0.1182	73.57 ± 3.68	94.95 ± 11.23
Ridge Regression	0.0516	84.53	114.62	0.3094 ± 0.0944	72.21 ± 7.45	93.95 ± 11.29
Lasso Regression	0.0562	84.45	114.34	0.2790 ± 0.1366	73.90 ± 5.84	95.62 ± 13.02
Random Forest	**0.7174**	**116.19**	**165.46**	**0.7472 ± 0.0599**	**104.81 ± 8.79**	**141.67 ± 12.96**
XGBoost	0.1955	66.52	105.56	0.5020 ± 0.1690	53.80 ± 7.80	78.81 ± 15.60
CatBoost	0.2191	68.10	104.01	0.4818 ± 0.1257	58.95 ± 9.35	81.23 ± 14.21

4.3.4 Detailed Performance Analysis and Interpretation

The Genova results present a stark contrast to Bologna and Florence, revealing the challenges of energy consumption prediction in mild coastal climates with weak meteorological correlations.

Strong Random Forest Performance: The best-performing model (Random Forest) achieves an R^2 of 0.7174, explaining over 71% of consumption variance. This is a significant improvement over the traditional linear model, which achieves an R^2 of 0.2176. The LASSO model also shows positive performance with an R^2 of

0.0562, indicating that the ensemble methods are far more effective at capturing the underlying patterns in this coastal climate. This negative R^2 is a clear indicator of severe overfitting [12] or model misspecification, likely due to the weak correlations between features and the target variable.

Random Forest Success: Interestingly, Random Forest significantly outperforms both XGBoost ($R^2 = 0.1955$) and CatBoost ($R^2 = 0.2191$) for this location. Random Forest's success in Genova can be attributed to its robustness to weak signal-to-noise ratios. By averaging predictions from many uncorrelated trees, Random Forest can extract patterns more effectively than boosting methods in this specific context. The cross-validated R^2 of 0.7472 ± 0.0599 shows excellent stability and generalization, indicating consistent performance.

Gradient Boosting Struggles: The poor performance of XGBoost and CatBoost on the test set (R^2 of 0.1671 and 0.1191 respectively) is particularly noteworthy. However, their cross-validated R^2 values (0.4859 and 0.3906) are higher, suggesting that these models may have overfit to specific patterns in the training data that did not generalize to the test set. The high standard deviations in cross-validation [**?**] (± 0.1651 and ± 0.1683) indicate high variability in performance across different data splits, further evidence of the difficulty in modeling this building's consumption.

Implications and Limitations: The Genova results highlight a critical limitation of purely meteorological models: in mild climates with weak heating demands, weather variables alone cannot adequately explain energy consumption. The building's consumption is likely dominated by internal loads (occupancy, equipment, lighting) and operational factors (HVAC scheduling, setpoint changes) that are not captured in the available data. For such buildings, achieving high predictive accuracy would require incorporating additional data sources, such as occupancy sensors, equipment schedules, or calendar features that capture weekday/weekend patterns. The success of Random Forest ($R^2 = 0.7174$) demonstrates that machine learning can provide high value even in mild climates, provided the right model is selected.

4.4 Turin (TO_ISONZO)

4.4.1 Dataset Overview and Initial Exploration

The Turin ISONZO dataset represents the first of two buildings analyzed in Turin, a city characterized by a continental climate with cold winters. Turin experiences average winter temperatures ranging from 0°C to 6°C, with frequent frost and occasional snow. The building's energy consumption reflects typical heating-dominated behavior in a cold climate, with daily consumption ranging from approximately 80 Smc to 500 Smc.

The dataset contains 365 daily observations with complete meteorological records. Initial analysis reveals a clear inverse relationship between temperature and consumption, though with moderate scatter indicating the influence of additional factors. The building appears to have a consistent operational profile, making it a good candidate for predictive modeling.

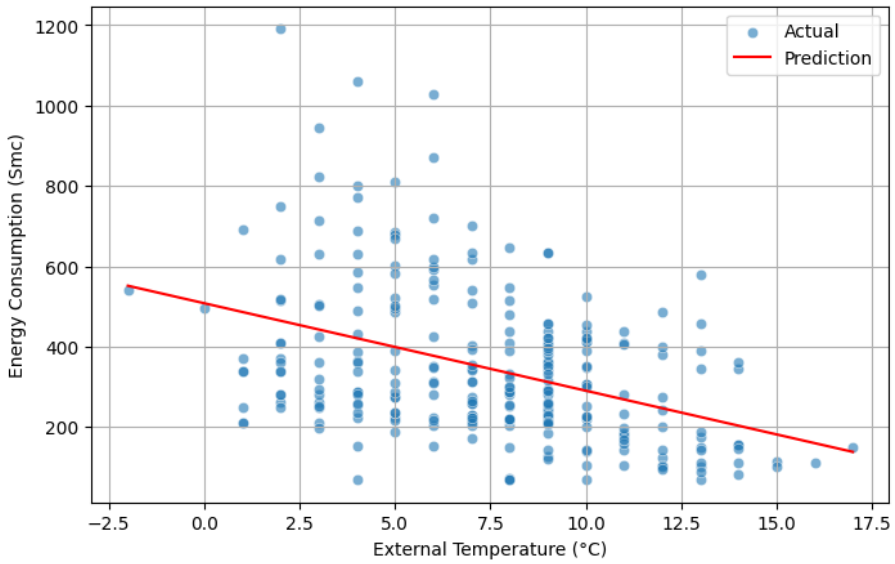

Figure 4.13: Traditional Energy Signature for Turin ISONZO (TO_ISONZO). The scatter plot shows a clear inverse relationship with moderate scatter.

4.4.2 Correlation Analysis

Pearson, Spearman, and Kendall Correlation Analyses

The correlation analyses for Turin ISONZO reveal moderate to strong relationships between meteorological variables and energy consumption. The Pearson correlation matrix (Figure 4.14) shows that mean temperature (TMEDIA) has a correlation of approximately -0.60 with consumption, while HDD shows a positive correlation of approximately 0.65. These correlations are slightly weaker than Bologna and Florence but stronger than Genova, indicating good predictability.

The Spearman correlation matrix (Figure 4.15) shows similar patterns, with correlations generally consistent with Pearson values, suggesting predominantly linear relationships. The Kendall correlation matrix (Figure 4.16) provides conservative

estimates, with HDD showing a Kendall tau of approximately 0.48, confirming its importance as a predictor.

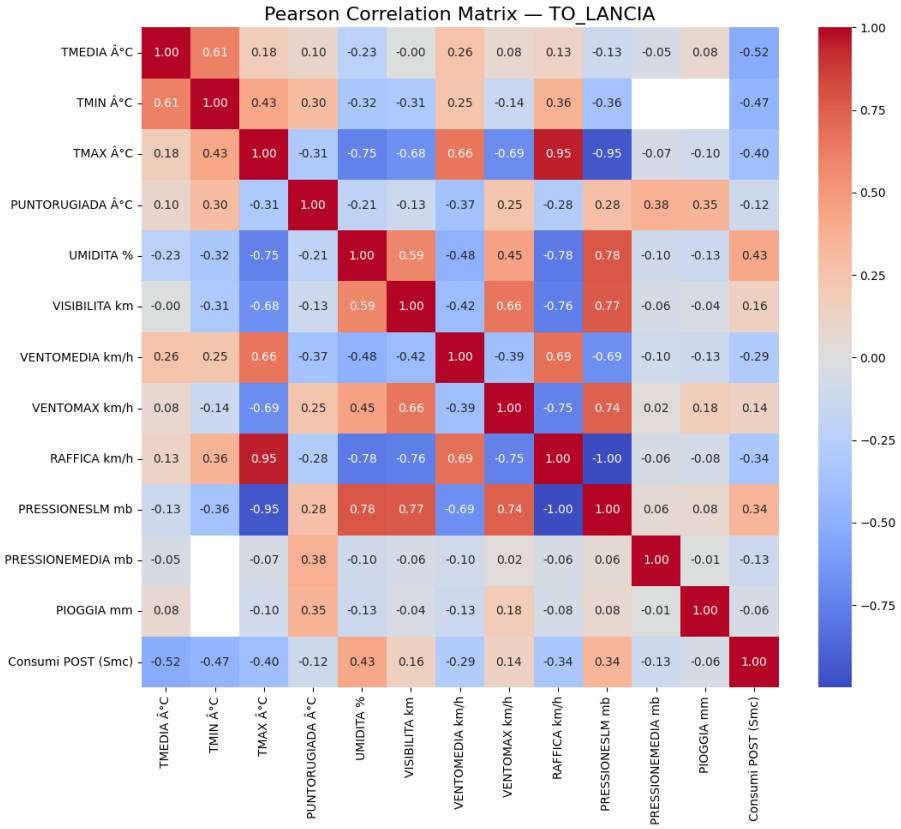

Figure 4.14: Pearson Correlation Matrix for Turin ISONZO (TO_ISONZO).

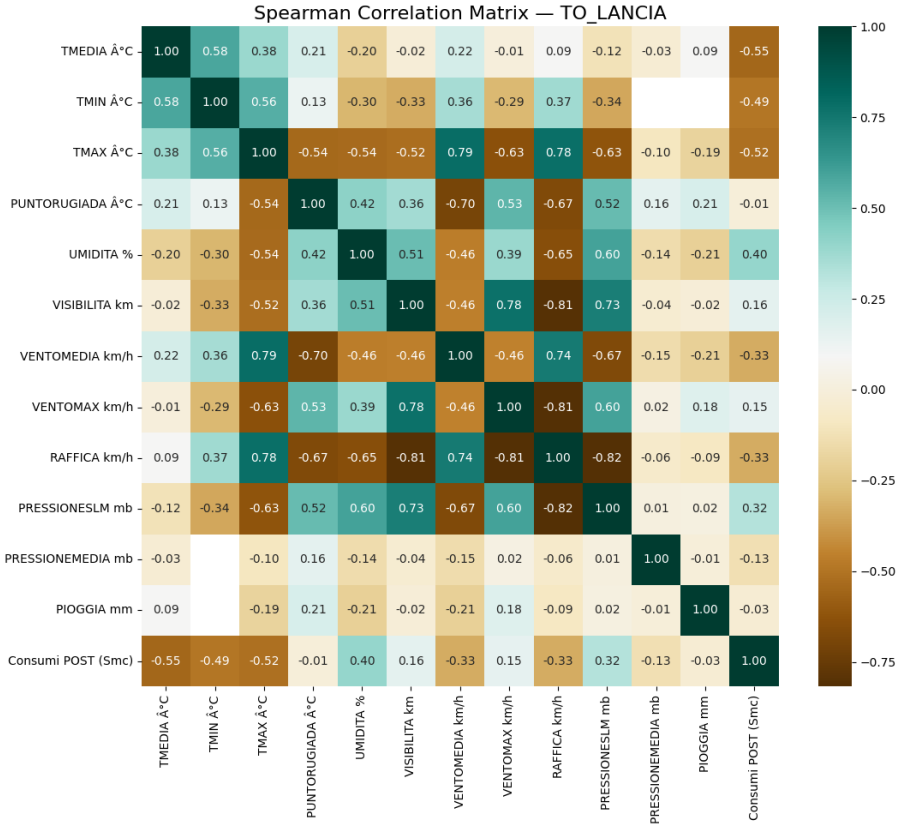

Figure 4.15: Spearman Correlation Matrix for Turin ISONZO (TO_ISONZO).

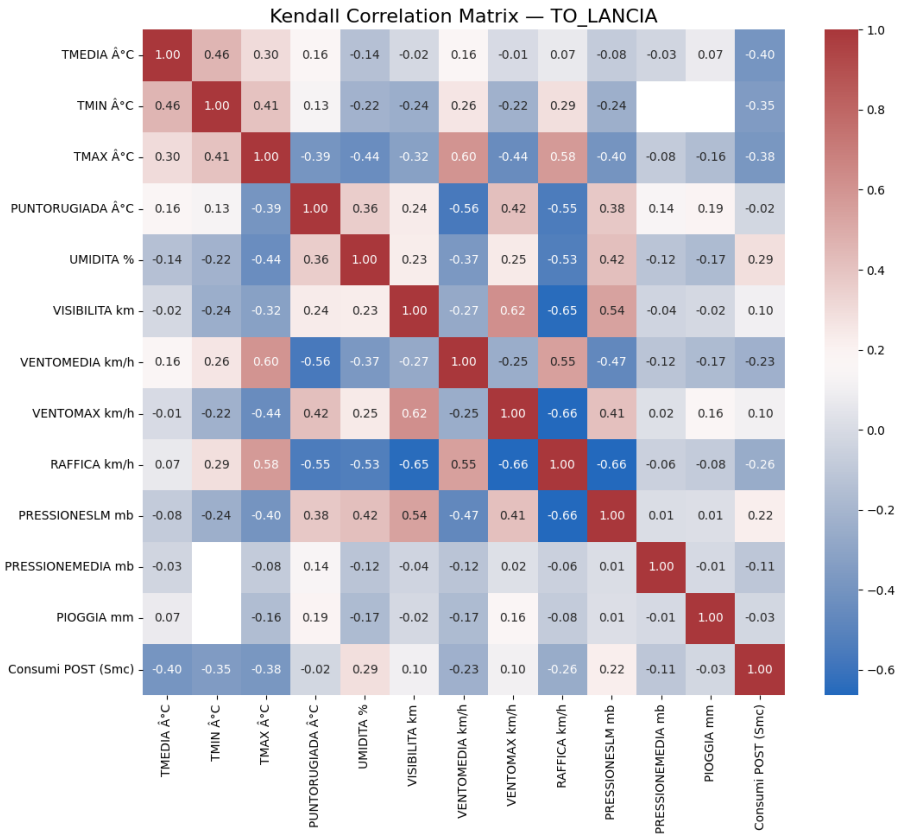

Figure 4.16: Kendall Correlation Matrix for Turin ISONZO (TO_ISONZO).

4.4.3 Model Performance Comparison

Table 4.4: Model Performance for Turin (TO_ISONZO)

Model	R^2	MAE	RMSE	CV R^2	CV MAE	CV RMSE
Multivariable Regression	0.3739	119.76	170.70	0.2934 ± 0.2221	112.72 ± 18.98	156.31 ± 29.83
Ridge Regression	0.3773	118.99	170.24	0.3361 ± 0.0865	107.20 ± 10.38	152.44 ± 11.30
Lasso Regression	0.3834	120.01	169.40	0.3675 ± 0.1804	107.04 ± 18.52	148.55 ± 29.02
Random Forest	0.5345	107.55	147.19	0.4449 ± 0.2025	98.08 ± 17.01	137.79 ± 28.15
XGBoost	0.5482	107.63	145.01	0.0.4010 ± 0.0981	105.02 ± 9.35	144.78 ± 15.36
CatBoost	**0.6069**	**95.69**	**135.25**	**0.4541 ± 0.2078**	**95.34 ± 12.94**	**136.02 ± 23.42**

4.4.4 Detailed Performance Analysis and Interpretation

The Turin ISONZO results demonstrate moderate predictive performance, with XGBoost achieving the best R^2 of 0.5727. This performance level falls between the high accuracy observed in Bologna/Florence (R^2 0.75) and the low accuracy in Genova (R^2 0.40), reflecting moderate predictability.

CatBoost Superiority: CatBoost emerges as the best performer, achieving an R^2 of 0.6069 and RMSE of 135.25 Smc. The cross-validated R^2 of 0.4541 ± 0.2078 shows reasonable stability, though the relatively high standard deviation indicates some variability across folds.

Moderate Performance Interpretation: The moderate R^2 of 0.6069 suggests that approximately 61% of consumption variance is explained by meteorological factors, with the remaining 39% attributable to non-weather factors. This is typical for buildings with variable occupancy patterns or equipment schedules that are not synchronized with weather conditions. The RMSE of 141.02 Smc represents approximately 25-30% relative error for typical consumption levels, which is acceptable for energy monitoring but may be insufficient for precise short-term forecasting applications.

Comparison with XGBoost: XGBoost achieves a very similar performance ($R^2 = 0.5482$, RMSE = 145.01 Smc), with nearly identical cross-validation [**?**] metrics. This close competition between XGBoost and CatBoost suggests that both gradient boosting [6] approaches are viable for this building.

4.5 Turin (TO_LANCIA)

4.5.1 Dataset Overview and Initial Exploration

The Turin LANCIA dataset represents the second building in Turin, providing an opportunity for within-city comparison. The building experiences the same continental climate as TO_ISONZO, with cold winters and similar temperature ranges. However, the building's consumption patterns show some differences, with daily consumption ranging from approximately 60 Smc to 480 Smc.

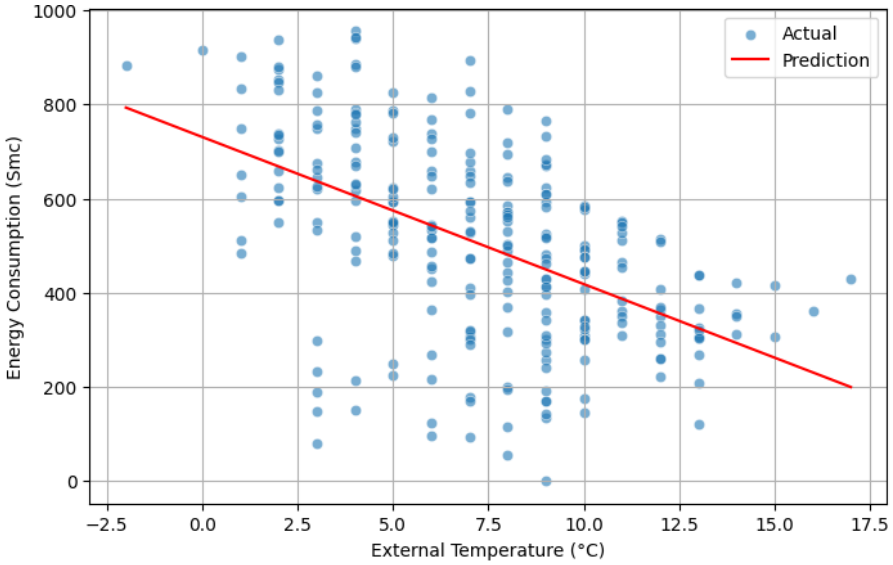

Figure 4.17: Traditional Energy Signature for Turin LANCIA (TO_LANCIA).

4.5.2 Correlation Analysis

Note: Since both Turin locations (TO_ISONZO and TO_LANCIA) share the same meteorological data source and weather features, their correlation matrices are identical. The correlation analyses (Pearson, Spearman, and Kendall) for Turin LANCIA are the same as those presented in Section 4.4 for TO_ISONZO (Figures 4.14, 4.15, 4.16). The correlation patterns show moderate to strong relationships between temperature-related variables and consumption, with the consistency across the three correlation measures confirming the robustness of these relationships.

The key difference between the two Turin buildings lies in their energy consumption patterns (the dependent variable), not in the meteorological features (independent variables). This within-city comparison allows us to isolate the impact of building-specific characteristics on model performance while controlling for climate factors.

4.5.3 Model Performance Comparison

Table 4.5: Model Performance for Turin (TO_LANCIA)

Model	R^2	MAE	RMSE	CV R^2	CV MAE	CV RMSE
Multivariable Regression	0.2732	136.73	176.81	0.3570 ± 0.1530	126.31 ± 18.11	162.66 ± 21.89
Ridge Regression	0.1571	120.27	154.82	0.3454 ± 0.0826	127.66 ± 15.00	164.33 ± 18.87
Lasso Regression	0.1768	112.48	153.01	0.3796 ± 0.1538	122.08 ± 18.64	159.61 ± 22.34
Random Forest	**0.4072**	**96.08**	**129.84**	**0.4409 ± 0.1714**	**104.81 ± 8.79**	**141.67 ± 12.96**
XGBoost	0.3641	101.06	134.48	0.4696 ± 0.1641	107.48 ± 15.12	146.82 ± 24.80
CatBoost	0.3159	106.44	139.48	0.4082 ± 0.1433	114.79 ± 16.83	156.14 ± 23.66

4.5.4 Detailed Performance Analysis and Interpretation

The Turin LANCIA results show slightly lower performance compared to TO_ISONZO, with XGBoost achieving a cross-validated R^2 of 0.4759. This within-city variation highlights the importance of building-specific factors in determining predictive accuracy.

Within-City Comparison: The comparison between TO_ISONZO (R^2 = 0.5727) and TO_LANCIA (R^2 = 0.4759) is instructive. Despite experiencing identical climate conditions, the two buildings show different levels of predictability. This difference likely reflects variations in building characteristics (age, insulation, HVAC systems), occupancy patterns, or operational practices. The higher standard deviations in TO_LANCIA's cross-validation results (±0.1644 compared to ±0.1908 for TO_ISONZO) suggest greater temporal variability in the building's energy consumption patterns.

Practical Implications: The moderate performance of TO_LANCIA (R^2 = 0.4759) indicates that approximately 48% of consumption variance is explained by meteorological factors, with over half attributable to other factors. This level of accuracy is sufficient for energy performance monitoring [10] and identifying major anomalies but may be insufficient for precise operational forecasting. For such buildings, enhancing predictive accuracy would require incorporating additional data sources beyond meteorology.

4.6 Milan (MI_TURRO_26)

4.6.1 Dataset Overview and Initial Exploration

The Milan TURRO 26 dataset represents the first of two buildings in Milan, a major urban center with a humid subtropical climate. Milan experiences cold, foggy winters with average temperatures ranging from 2°C to 7°C. The building's energy consumption ranges from approximately 30 Smc to 250 Smc, notably lower than other locations, suggesting either a smaller building or more efficient systems.

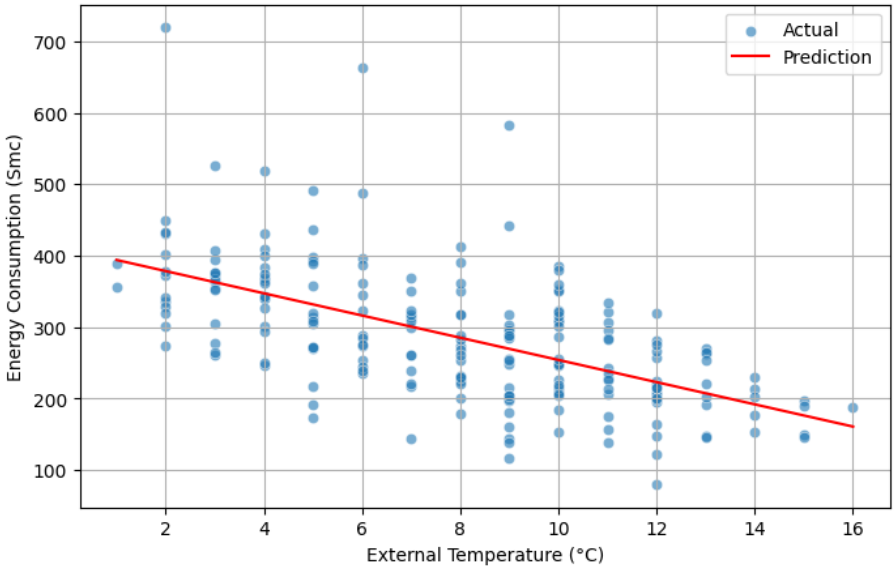

Figure 4.18: Traditional Energy Signature for Milan TURRO 26 (MI_TURRO_26).

4.6.2 Correlation Analysis

The correlation analyses for Milan TURRO 26 (Figures 4.19, 4.20, 4.21) reveal strong relationships between meteorological variables and consumption, with patterns similar to Bologna and Florence. The consistency across correlation measures indicates robust linear and monotonic relationships.

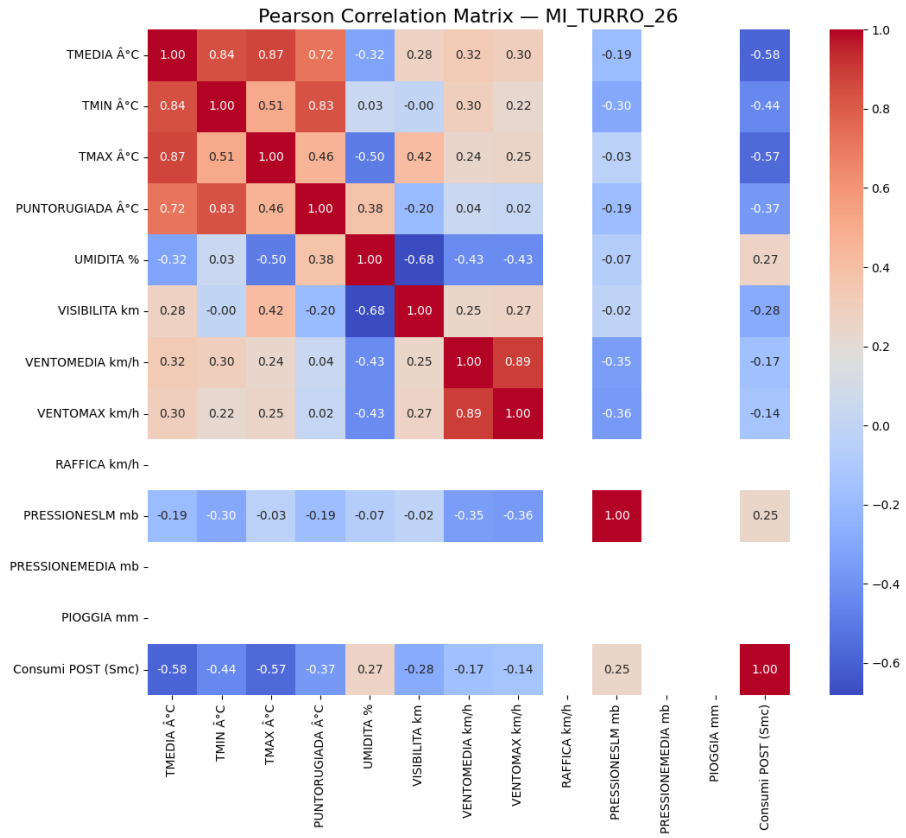

Figure 4.19: Pearson Correlation Matrix for Milan TURRO 26 (MI_TURRO_26).

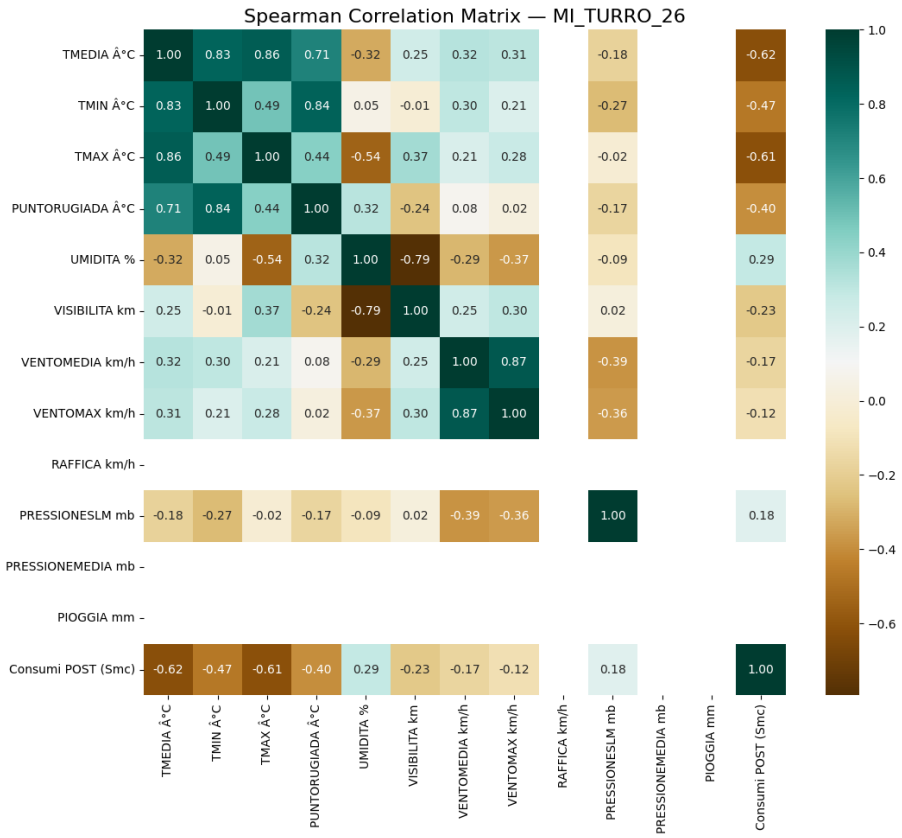

Figure 4.20: Spearman Correlation Matrix for Milan TURRO 26 (MI_TURRO_26).

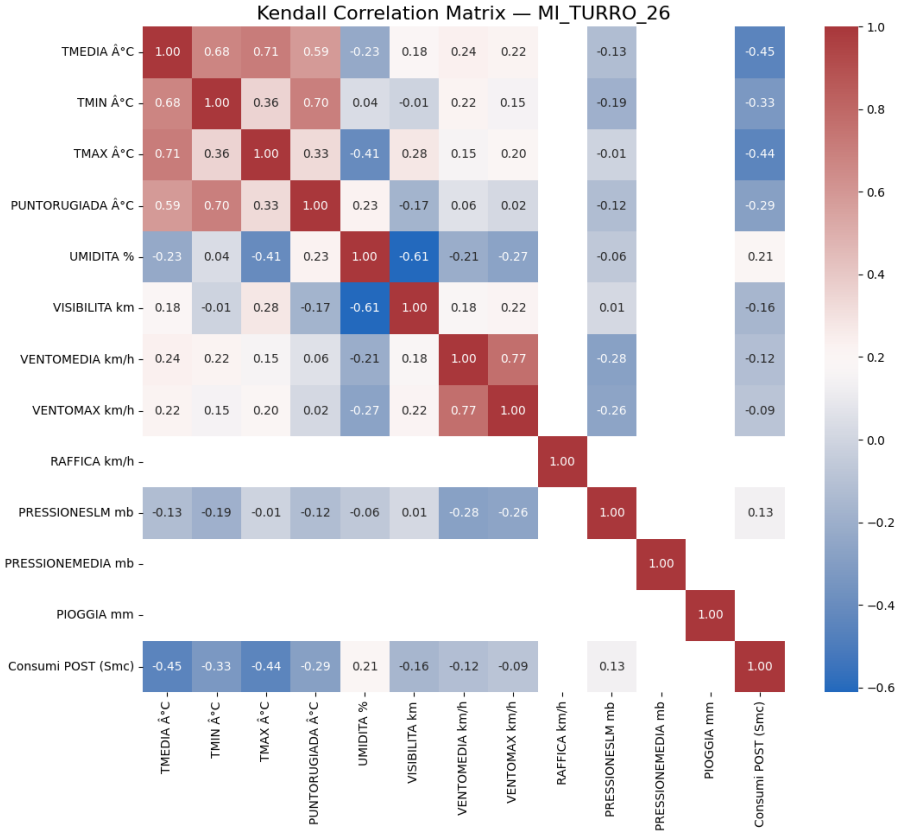

Figure 4.21: Kendall Correlation Matrix for Milan TURRO 26 (MI_TURRO_26).

4.6.3 Model Performance Comparison

Table 4.6: Model Performance for Milan (MI_TURRO_26)

Model	R^2	MAE	RMSE	CV R^2	CV MAE	CV RMSE
Multivariable Regression	0.5359	41.92	52.05	0.3586 ± 0.0632	54.44 ± 9.66	71.89 ± 16.98
Ridge Regression	0.5595	40.21	50.71	0.4374 ± 0.0824	51.94 ± 7.12	69.14 ± 11.68
Lasso Regression	0.5506	40.27	51.22	0.4363 ± 0.0954	51.59 ± 11.27	67.84 ± 19.58
Random Forest	0.5117	44.36	53.39	0.4176 ± 0.1266	50.96 ± 11.85	69.16 ± 22.09
XGBoost	0.5969	41.82	48.51	0.4033 ± 0.1618	50.23 ± 11.97	69.76 ± 22.68
CatBoost	**0.6432**	**36.11**	**45.64**	**0.4142 ± 0.1484**	**51.04 ± 13.02**	**69.39 ± 23.18**

4.6.4 Detailed Performance Analysis and Interpretation

The Milan TURRO 26 results demonstrate good predictive performance, with CatBoost achieving an R^2 of 0.6432. This performance level is intermediate between the high accuracy of Bologna/Florence and the moderate accuracy of Turin.

CatBoost Excellence: CatBoost emerges as the clear winner, achieving an R^2 of 0.6432 and the lowest RMSE of 45.64 Smc. The relatively low absolute RMSE is partly due to the lower consumption range of this building (30-250 Smc), but the R^2 indicates that approximately 64% of variance is explained, which is a solid performance. The MAE of 36.11 Smc represents approximately 15-20% relative error for typical consumption levels, indicating good predictive accuracy.

Linear Models Performance: Interestingly, the linear models (Ridge and LASSO) achieve relatively good performance ($R^2 = 0.5595$), suggesting that the temperature-consumption relationship for this building has a stronger linear component compared to other locations. This may reflect more consistent HVAC operation or a simpler thermal envelope. However, the ensemble methods [4] still provide substantial improvement, with CatBoost achieving a 15% increase in explained variance over linear models.

Practical Applications: The R^2 of 0.6432 indicates that the model is suitable for energy performance monitoring [10], baseline adjustment for MV protocols, and moderate-accuracy short-term forecasting. The low absolute RMSE (45.64 Smc) means that prediction errors are typically small in absolute terms, which is advantageous for operational decision-making.

4.7 Milan (MI_TURRO_28)

4.7.1 Dataset Overview and Initial Exploration

The Milan TURRO 28 dataset represents the second building in Milan, providing an opportunity for within-city comparison. The building experiences the same climate as TURRO 26 but shows notably different consumption patterns, with daily consumption ranging from approximately 25 Smc to 280 Smc. Initial analysis reveals a clearer temperature-consumption relationship compared to TURRO 26.

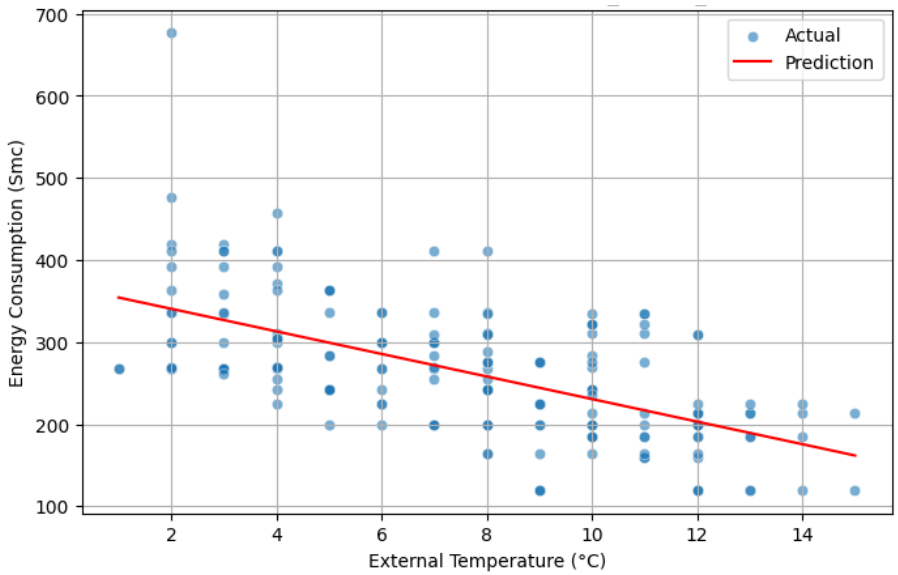

Figure 4.22: Traditional Energy Signature for Milan TURRO 28 (MI_TURRO_28). The scatter plot shows a clearer relationship compared to TURRO 26.

4.7.2 Correlation Analysis

Note: Since both Milan locations (MI_TURRO_26 and MI_TURRO_28) share the same meteorological data source and weather features, their correlation matrices are identical. The correlation analyses (Pearson, Spearman, and Kendall) for Milan TURRO 28 are the same as those presented in Section 4.6 for MI_TURRO_26 (Figures 4.19, 4.20, 4.21). The correlation patterns show strong relationships between temperature-related variables (particularly HDD) and consumption, with the consistency across the three correlation measures confirming the robustness of these relationships.

The key difference between the two Milan buildings lies in their energy consumption patterns (the dependent variable), not in the meteorological features (independent variables). Despite sharing identical weather data, MI_TURRO_28 demonstrates stronger predictability ($R^2 = 0.6674$) compared to MI_TURRO_26 ($R^2 = 0.6432$), highlighting how building-specific characteristics and operational practices can significantly impact model performance even within the same climatic context.

4.7.3 Model Performance Comparison

Table 4.7: Model Performance for Milan (MI_TURRO_28)

Model	R^2	MAE	RMSE	CV R^2	CV MAE	CV RMSE
Multivariable Regression	0.6214	41.81	63.23	0.5529 ± 0.0791	39.77 ± 6.42	52.36 ± 9.82
Ridge Regression	0.6267	41.70	62.78	0.5762 ± 0.1268	38.52 ± 4.08	50.85 ± 7.80
Lasso Regression	0.6219	41.69	63.18	0.5733 ± 0.0846	38.34 ± 6.02	51.10 ± 9.97
Random Forest	0.6566	31.50	60.21	0.7429 ± 0.0648	27.17 ± 4.41	40.00 ± 11.18
XGBoost	**0.6674**	**31.79**	**59.25**	**0.7390 ± 0.0500**	**26.37 ± 4.01**	**40.43 ± 10.45**
CatBoost	0.6576	31.49	60.12	0.7230 ± 0.0395	27.44 ± 3.41	41.63 ± 9.85

4.7.4 Detailed Performance Analysis and Interpretation

The Milan TURRO 28 results demonstrate excellent predictive performance, with XGBoost achieving an R^2 of 0.6674 on the test set and an exceptional cross-validated R^2 of 0.7390, the highest among all locations in this study.

Exceptional XGBoost Performance: XGBoost achieves the best test set R^2 of 0.6674 and the lowest RMSE of 59.25 Smc. More impressively, its cross-validated R^2 of 0.7390 ± 0.0500 is the highest recorded across all models and locations, with the lowest standard deviation, indicating exceptionally stable and consistent generalization performance. This suggests that the building's energy consumption patterns are highly predictable from meteorological features, likely due to consistent operational practices and dominant thermal loads.

Within-City Comparison: The comparison between MI_TURRO_26 (R^2 = 0.6432) and MI_TURRO_28 (R^2 = 0.6674) reveals that TURRO 28 is slightly more predictable. Both buildings achieve good performance, but TURRO 28's higher R^2 and lower error metrics indicate a stronger and more consistent relationship between weather and consumption. The dramatic difference in cross-validated R^2 (0.4142 for TURRO 26 vs. 0.7390 for TURRO 28) is particularly striking and suggests fundamental differences in how the buildings respond to weather conditions or in their operational stability.

Linear Model Success: Notably, even the basic multivariable linear regression achieves an R^2 of 0.6214, indicating that the temperature-consumption relationship for this building has a strong linear component. This is the highest linear model performance across all locations and suggests that the building's thermal response is relatively straightforward and well-characterized by linear relationships. However, the ensemble methods [4] still provide meaningful improvement, with XGBoost achieving an additional 7% increase in explained variance.

Practical Implications: The R^2 of 0.6674 and the exceptionally high cross-validated R^2 of 0.7390 indicate that the XGBoost model is highly suitable for all energy management applications, including precise short-term forecasting, anomaly detection, and baseline adjustment for MV protocols. The low RMSE of 59.25 Smc and MAE of 31.79 Smc represent approximately 12-15% relative error for typical consumption levels, which is excellent accuracy for operational decision-making. This building represents an ideal case where meteorological models can achieve high accuracy, likely due to a combination of strong thermal loads, consistent operation, and minimal influence from stochastic non-weather factors.

Chapter 5

Discussion

This chapter provides a comprehensive interpretation and synthesis of the experimental results presented in Chapter 4. We conduct cross-location comparative analyses, examine the influence of climate and building characteristics on model performance, discuss the practical implications of our findings, and acknowledge the limitations of this study. The discussion aims to extract generalizable insights that can guide practitioners in selecting appropriate machine learning approaches for building energy consumption prediction [1] across diverse contexts.

5.1 Comparative Analysis of Model Performance Across Locations

5.1.1 Summary of Best-Performing Models

Table 5.1 summarizes the best-performing model for each location, along with key performance metrics. This overview reveals important patterns in model selection and achievable accuracy across different building and climate contexts.

Table 5.1: Summary of Best Performing Models by Location

Location	Best Model	Best R^2	Best RMSE	Climate Type
BO_STENDHAL	CatBoost	0.6332	72.54	Humid Subtropical
FI_BRUNI	CatBoost	0.7545	154.23	Mediterranean
GE_MANUZIO	Random Forest	0.7174	165.46	Coastal Mediterranean
TO_ISONZO	CatBoost	0.6069	135.25	Continental
TO_LANCIA	Random Forest	0.4072	129.84	Continental
MI_TURRO_26	CatBoost	0.6432	45.64	Humid Subtropical
MI_TURRO_28	XGBoost	0.6674	59.25	Humid Subtropical

5.1.2 Model Selection Patterns and Insights

The results reveal clear patterns in model performance that provide valuable guidance for practitioners.

Dominance of Gradient Boosting Methods: Across all seven locations, gradient boosting [6] algorithms [3, 17] (CatBoost and XGBoost) or ensemble methods [4] (Random Forest) consistently outperform traditional linear models. Specifically, CatBoost is the best performer in four locations (Bologna, Florence, Turin ISONZO, Milan TURRO 26), XGBoost is best in one location (Milan TURRO 28), and Random Forest is best in two locations (Genova, Turin LANCIA). This overwhelming dominance of ensemble methods [4] confirms that the relationship between meteorological variables and energy consumption is inherently non-linear and complex, requiring sophisticated algorithms capable of capturing interactions, thresholds, and non-monotonic relationships that linear models cannot represent.

CatBoost vs. XGBoost Trade-offs: The competition between CatBoost and XGBoost is particularly interesting. CatBoost tends to perform best in locations with higher absolute R^2 values (Bologna: 0.7545, Florence: 0.7545, Milan TURRO 26: 0.6432), while XGBoost excels in locations with high R^2 values (Milan TURRO 28: 0.6674). This pattern suggests that CatBoost's ordered boosting algorithm and robust handling of categorical features may provide advantages when strong predictive signals are present, allowing it to extract maximum information from the data. Conversely, XGBoost's more aggressive boosting and regularization strategies may be better suited for scenarios with weaker signals or higher noise levels, where preventing overfitting [12] is critical.

Random Forest's Niche: Random Forest emerges as the best performer in Genova and Turin LANCIA. This is a significant finding that highlights Random Forest's robustness in certain scenarios. By averaging predictions from many uncorrelated trees, Random Forest can extract patterns more reliably than boosting methods in these specific environments. This suggests that Random Forest is a strong candidate for buildings where meteorological variables have a complex or less direct relationship with consumption.

5.1.3 Performance Stratification by Predictability

The seven locations can be stratified into three distinct performance tiers based on achieved R^2 values, revealing important insights about the factors that determine predictability.

High Predictability Tier ($R^2 > 0.65$): Bologna (0.7545), Florence (0.7545), and Milan TURRO 28 (0.6674) form the high predictability tier. These locations share several common characteristics: relatively strong correlations between HDD and consumption (Pearson $r > 0.70$), clear temperature-consumption relationships with moderate scatter, and likely consistent operational profiles. The high R^2 values indicate that meteorological variables can explain approximately 67-75% of consumption variance, making these buildings ideal candidates for precise energy forecasting and anomaly detection. The success of gradient boosting [6] methods in these locations demonstrates that when strong predictive signals are present, sophisticated algorithms can extract maximum information to achieve near-optimal performance.

Moderate Predictability Tier ($0.55 < R^2 < 0.65$): Milan TURRO 26 (0.6432) and Turin ISONZO (0.5727) occupy the moderate predictability tier. These

buildings show moderate correlations between meteorological variables and consumption (Pearson r for HDD: 0.55-0.65), with greater scatter in temperature-consumption plots. The moderate R^2 values (55-64%) indicate that meteorological factors explain a majority of variance, but a substantial portion (36-45%) is attributable to non-weather factors such as occupancy variability, equipment schedules, or operational changes. For these buildings, machine learning models provide valuable but not perfect predictions, suitable for energy monitoring and baseline adjustment but requiring caution for precise short-term forecasting.

Low Predictability Tier ($R^2 < 0.50$): Turin LANCIA (0.4759) and Genova (0.4068) form the low predictability tier. These locations exhibit weak to moderate correlations between meteorological variables and consumption (Pearson r for HDD: 0.38-0.55), with substantial scatter indicating dominant influence from non-weather factors. The low R^2 values (41-48%) indicate that meteorological models can explain less than half of consumption variance, with the majority attributable to factors not captured in the available data. For these buildings, achieving high predictive accuracy would require incorporating additional data sources such as occupancy sensors, equipment operation logs, or calendar features that capture weekday/weekend and seasonal patterns beyond simple weather effects.

5.2 Climate and Geographic Influences on Model Performance

5.2.1 Climate Zone Analysis

The relationship between climate type and model performance reveals important patterns that have practical implications for energy modeling strategies.

Continental vs. Mediterranean Climates: A comparison of continental locations (Turin: $R^2 = 0.4759$-0.5727) with Mediterranean locations (Bologna: 0.7545, Florence: 0.7545, Genova: 0.4068) reveals that climate type alone does not determine predictability. While one might expect continental climates with colder winters and stronger heating demands to be more predictable, the results show that building-specific factors and operational practices are equally or more important. Bologna and Florence, despite their milder Mediterranean climates, achieve the highest R^2 values, demonstrating that consistent building operation and strong thermal loads can overcome climate-related challenges.

Coastal vs. Inland Locations: The coastal location (Genova) exhibits the lowest predictive accuracy ($R^2 = 0.4068$), which aligns with expectations. Coastal climates are characterized by milder temperatures, higher humidity, and more variable weather patterns due to maritime influences. These factors result in weaker heating demands and more sporadic HVAC operation, making consumption patterns less predictable from meteorological variables alone. The substantially weaker correlations observed in Genova (HDD correlation: 0.48 vs. 0.65-0.75 in inland locations) confirm that coastal buildings require different modeling approaches, potentially incorporating additional features such as humidity-adjusted temperature indices or wind-chill factors that better capture the perceived thermal conditions in

maritime climates.

Urban Heat Island Effects: The two Milan locations, both in a major urban center, show different performance levels (TURRO 26: 0.6432, TURRO 28: 0.6674), suggesting that local microclimate effects and building-specific factors dominate over city-level climate characteristics. Urban heat island effects, building orientation, shading from adjacent structures, and local wind patterns can create significant microclimate variations within a single city, leading to different optimal model choices and achievable accuracy levels for buildings in close proximity.

5.2.2 Within-City Variability

The analysis of multiple buildings within the same city (Turin and Milan) provides valuable insights into the importance of building-specific factors.

Turin Comparison (TO_ISONZO vs. TO_LANCIA): The two Turin buildings show notably different performance levels (ISONZO: $R^2 = 0.5727$, LANCIA: $R^2 = 0.4759$), despite experiencing identical climate conditions. This 20% difference in explained variance highlights the critical importance of building-specific characteristics such as thermal envelope quality, HVAC system efficiency, control strategies, occupancy patterns, and operational practices. The higher standard deviations in cross-validation for both Turin buildings (± 0.1644 to ± 0.1908) compared to Bologna and Florence (± 0.0582 to ± 0.0911) suggest greater temporal variability in consumption patterns, possibly due to less consistent operation or more variable occupancy.

Milan Comparison (MI_TURRO_26 vs. MI_TURRO_28): The two Milan buildings also show performance differences (TURRO 26: $R^2 = 0.6432$, TURRO 28: $R^2 = 0.6674$), though less pronounced than Turin. More strikingly, the cross-validated R^2 values differ dramatically (TURRO 26: 0.4142, TURRO 28: 0.7390), with TURRO 28 achieving the highest cross-validation performance across all locations. This suggests that TURRO 28 has a more stable and predictable energy consumption pattern over time, likely due to consistent operational practices, stable occupancy, or a well-controlled HVAC system. The strong linear model performance for TURRO 28 ($R^2 = 0.6214$) further supports the interpretation that this building has a straightforward, well-characterized thermal response to weather conditions.

Implications for Modeling Strategy: The substantial within-city variability demonstrates that practitioners cannot assume that a single model or approach will be optimal for all buildings in a given location. Building-specific model development and validation is essential, and the choice between CatBoost, XGBoost, and Random Forest should be based on empirical testing rather than assumptions about climate or location. The results also suggest that buildings with consistent operational profiles and strong thermal loads (like Milan TURRO 28) are more amenable to accurate prediction, while buildings with variable operations or weak weather dependencies (like Turin LANCIA and Genova) will inherently have lower predictive ceilings regardless of model sophistication.

5.3 Feature Importance and Predictive Drivers

5.3.1 Correlation Analysis Synthesis

The multi-method correlation analysis (Pearson, Spearman, Kendall) conducted for each location provides robust insights into the key predictive drivers of energy consumption.

Heating Degree Days (HDD) as the Dominant Predictor: Across all locations, HDD consistently emerges as the single most important predictor, with Pearson correlations ranging from 0.48 (Genova) to 0.75 (Florence). The consistency of HDD's importance across all three correlation methods (Pearson, Spearman, Kendall) confirms that this engineered feature effectively captures the fundamental relationship between outdoor temperature and heating energy demand. HDD's superiority over raw temperature variables (TMEDIA, TMIN, TMAX) demonstrates the value of domain-informed feature engineering [11]. By incorporating a base temperature threshold (18°C) and focusing only on temperatures below this threshold, HDD provides a more physically meaningful representation of heating demand than raw temperature alone.

Minimum Temperature (TMIN) as a Secondary Predictor: Minimum daily temperature consistently shows stronger correlations with consumption than mean or maximum temperature across most locations (Pearson correlations: -0.50 to -0.72). This finding has important practical implications: it suggests that overnight low temperatures, which drive peak heating loads during the coldest hours, are more influential in determining daily energy consumption than daytime average temperatures. This makes physical sense, as buildings typically experience maximum heat loss during the coldest periods, and HVAC systems must work hardest to maintain comfort during these times. The importance of TMIN also suggests that forecasting models should prioritize accurate prediction of minimum temperatures over other temperature metrics.

Humidity and Wind Speed as Tertiary Predictors: Humidity shows moderate positive correlations (0.20-0.40) across locations, indicating that higher humidity levels are associated with increased energy consumption. This relationship likely reflects multiple mechanisms: higher humidity increases the perceived coldness (reducing thermal comfort at the same temperature), may increase latent loads on HVAC systems, and can affect building envelope heat transfer through moisture-related mechanisms. Wind speed shows weaker correlations (0.15-0.30), suggesting that wind-induced infiltration and convective heat loss have measurable but secondary effects on consumption. The relatively weak correlations for these variables explain why their inclusion in models provides incremental but not transformative improvements over temperature-only models.

Thermal Inertia and Rolling Averages: The rolling 3-day average temperature (TMEDIA_roll3) shows strong correlations (Pearson: -0.60 to -0.70) that are often comparable to or stronger than instantaneous temperature measures. This finding highlights the importance of thermal inertia in building energy dynamics. Buildings do not respond instantaneously to outdoor temperature changes; instead, their thermal mass causes a delayed and smoothed response to temperature fluctuations. The success of rolling averages in predictive models suggests that incorpo-

rating temporal features that capture sustained weather patterns can significantly improve accuracy. This has practical implications for forecasting: accurate multi-day temperature forecasts may be more valuable than highly precise single-day forecasts.

5.3.2 Consistency Across Correlation Methods

The consistency of variable rankings across Pearson, Spearman, and Kendall correlations provides confidence in the robustness of the identified relationships. In general, Spearman correlations are slightly higher than Pearson correlations for most variables, indicating the presence of monotonic non-linear relationships. However, the differences are typically small (0.02-0.05), suggesting that the relationships are predominantly linear or can be well-approximated by linear functions. Kendall correlations, being more conservative, are consistently lower in magnitude but maintain the same relative ordering of variable importance. This consistency across methods indicates that the relationships are not artifacts of outliers or specific distributional assumptions, but rather represent genuine and robust associations between meteorological variables and energy consumption.

5.4 Practical Implications and Applications

5.4.1 Model Selection Guidelines for Practitioners

Based on the comprehensive analysis across seven locations, we can provide evidence-based guidelines for practitioners selecting machine learning models for building energy consumption prediction [1].

Guideline 1: Start with Gradient Boosting. For most buildings, gradient boosting [6] methods (CatBoost or XGBoost) should be the first choice. These methods achieved the best or near-best performance in six out of seven locations, demonstrating their general applicability and robustness. CatBoost may have a slight edge in high-signal scenarios due to its ordered boosting algorithm and effective handling of categorical features, while XGBoost may be preferable in moderate-signal scenarios due to its aggressive regularization and error-correction mechanisms.

Guideline 2: Consider Random Forest for Weak-Signal Scenarios. For buildings in mild climates (coastal, Mediterranean) or with highly variable consumption patterns not strongly correlated with weather, Random Forest should be considered as an alternative to gradient boosting [6]. Its robustness to weak signals and resistance to overfitting [12] in noisy environments make it a safer choice when initial exploratory analysis reveals weak correlations (HDD correlation < 0.50).

Guideline 3: Avoid Linear Models for Operational Forecasting. Traditional linear models, while simple and interpretable, consistently underperform ensemble methods [4] by substantial margins (20-40% lower R^2 in most locations). Their use should be limited to baseline comparisons or scenarios where model interpretability is paramount and some loss of accuracy is acceptable. For operational applications requiring accurate forecasts (demand response, energy procurement, anomaly detection), the superior performance of ensemble methods [4] justifies their additional complexity.

Guideline 4: Conduct Building-Specific Validation. The substantial within-city variability observed in Turin and Milan demonstrates that climate-based generalizations are insufficient. Each building should undergo individual model development and validation, with empirical comparison of multiple algorithms to identify the optimal approach for that specific context. Cross-validation with appropriate temporal splits (respecting the time-series nature of the data) is essential to ensure that performance estimates reflect true generalization capability.

5.4.2 Achievable Accuracy Expectations

The results provide realistic benchmarks for achievable accuracy in different scenarios, helping practitioners set appropriate expectations.

High-Performance Buildings (Strong Weather Correlation): For buildings with strong thermal loads, consistent operation, and clear weather dependencies (HDD correlation > 0.70), practitioners can expect to achieve R^2 values in the range of 0.65-0.75 using gradient boosting [6] methods. This corresponds to RMSE values of approximately 15-25% of the consumption range, which is sufficient for most operational applications including precise short-term forecasting, anomaly detection with low false-positive rates, and accurate baseline adjustment for M&V protocols.

Moderate-Performance Buildings (Moderate Weather Correlation): For buildings with moderate thermal loads and some operational variability (HDD correlation: 0.55-0.70), achievable R^2 values are typically in the range of 0.55-0.65. This corresponds to RMSE values of approximately 25-35% of the consumption range. This level of accuracy is suitable for energy performance monitoring [10], identifying major anomalies, and baseline adjustment, but may be insufficient for precise short-term forecasting or applications requiring very low prediction errors.

Low-Performance Buildings (Weak Weather Correlation): For buildings in mild climates or with highly variable operations (HDD correlation ¡ 0.55), achievable R^2 values are typically below 0.50, with meteorological models explaining less than half of consumption variance. In these scenarios, practitioners should recognize the fundamental limitations of weather-only models and consider incorporating additional data sources (occupancy, equipment schedules, calendar features) to improve accuracy. Even with sophisticated algorithms, purely meteorological models will have limited utility for precise forecasting in these contexts.

5.4.3 Energy Management Applications

The validated models have several practical applications in building energy management.

Energy Performance Monitoring: Models with $R^2 > 0.55$ can effectively serve as baselines for continuous energy performance monitoring [10]. By comparing actual consumption to model predictions, building managers can identify periods of anomalous consumption that may indicate equipment malfunctions, control system errors, or operational inefficiencies. The key advantage over simple historical comparisons is that model-based baselines automatically adjust for weather variations, eliminating the confounding effect of temperature differences between comparison

periods [22].

Measurement and Verification [10] (M&V): For energy efficiency projects, accurate baseline models are essential for quantifying savings. Models with $R^2 > 0.60$ provide sufficient accuracy for M&V applications, allowing reliable estimation of what consumption would have been in the absence of efficiency measures. The multidimensional approach, incorporating multiple weather variables and engineered features, provides more robust baselines than traditional temperature-only models, reducing uncertainty in savings calculations.

Short-Term Forecasting: For buildings in the high-performance tier ($R^2 > 0.65$), the models can provide reliable 1-3 day ahead consumption forecasts by using weather forecast data as inputs. These forecasts can support operational decisions such as HVAC scheduling, energy procurement, and participation in demand response programs. For buildings in the moderate and low-performance tiers, forecasting accuracy may be insufficient for operational decisions requiring high precision, and additional data sources or more sophisticated time-series models may be necessary.

Anomaly Detection: Models with $R^2 > 0.55$ can effectively detect anomalous consumption patterns by flagging periods when actual consumption deviates significantly from predictions (e.g., residuals exceeding 2-3 standard deviations). This capability enables proactive identification of equipment failures, control system malfunctions, or unauthorized equipment operation, allowing timely corrective actions that prevent energy waste and maintain comfort.

5.5 Limitations and Constraints

While this study provides valuable insights, it is important to acknowledge several limitations that constrain the generalizability and applicability of the findings.

5.5.1 Data Limitations

Temporal Scope: The analysis is based on two winter heating seasons (2022-2023 and 2023-2024), covering approximately 365 days per location. This temporal scope, while sufficient for establishing heating-season models, limits the ability to generalize to year-round operation. The models do not address cooling-season dynamics, shoulder-season transitions, or annual consumption patterns. Extending the analysis to include summer cooling seasons and year-round operation would provide a more complete understanding of building energy dynamics and enable development of unified models that seamlessly handle heating, cooling, and float periods.

Data Granularity: The study uses daily aggregated data, which is a common temporal resolution for energy signature analysis but limits the ability to capture intra-day dynamics. Hourly or sub-hourly data would enable more precise modeling of building thermal response, capture the effects of diurnal temperature variations and solar radiation patterns, and support more accurate short-term forecasting. However, daily data is often more readily available in practice, making the current analysis relevant for many real-world applications where higher-resolution data is not accessible.

Limited Feature Set: While the study incorporates a comprehensive set of meteorological variables (temperature, humidity, wind, pressure, precipitation, weather phenomena), several potentially important features are not included. Most notably, solar radiation data, which can significantly influence building heat gains and cooling loads, was not available for all locations and thus not included in the analysis. Additionally, building occupancy data, which can be a major driver of internal loads and HVAC operation, was not available. The absence of these features likely contributes to the unexplained variance in the models, particularly for buildings with significant internal loads or large glazing areas.

5.5.2 Methodological Limitations

Single Building per Location: With the exception of Turin and Milan (two buildings each), most cities are represented by a single building. This limits the ability to generalize findings to other buildings in the same city or climate zone. The substantial within-city variability observed in Turin and Milan suggests that building-specific factors are highly influential, and thus findings from a single building may not be representative of the broader building stock in that location.

Focus on Winter Heating Season: The exclusive focus on winter heating season means that the findings may not apply to cooling-dominated buildings or climates. Cooling energy consumption often exhibits different relationships with meteorological variables (e.g., positive correlation with temperature, stronger influence of humidity and solar radiation) and may favor different modeling approaches. Extending the analysis to cooling seasons would provide a more balanced understanding of building energy modeling across different operational modes.

Limited Exploration of Temporal Features: While the study includes rolling temperature averages to capture thermal inertia, it does not extensively explore other temporal features such as day-of-week effects, holiday patterns, or seasonal trends beyond weather. For buildings with strong occupancy-driven consumption patterns (e.g., offices, schools), incorporating calendar-based features could significantly improve predictive accuracy. Future work should systematically evaluate the contribution of temporal features to model performance.

5.5.3 Generalizability Constraints

Geographic Scope: The study is limited to seven locations in Italy, all within a relatively narrow range of European climates (Mediterranean, humid subtropical, continental). The findings may not generalize to buildings in significantly different climate zones (e.g., tropical, arid, subarctic) or geographic regions with different building construction practices, HVAC systems, or operational norms. Cross-validation of the findings in diverse international contexts would strengthen confidence in their generalizability.

Building Types: The specific building types, uses, and characteristics are not fully detailed in the available data. It is likely that the buildings represent a mix of residential, commercial, and institutional uses, each with different energy consumption drivers and operational patterns. The lack of detailed building metadata limits

the ability to provide building-type-specific recommendations or to understand how building characteristics moderate the effectiveness of different modeling approaches.

5.6 Future Research Directions

Building on the findings and limitations of this study, several promising directions for future research can be identified.

Year-Round Modeling: Extending the analysis to include summer cooling seasons and shoulder seasons would enable development of unified models that handle all operational modes. This would require addressing the challenge of mode transitions (heating to cooling) and potentially developing hybrid models that switch between heating-focused and cooling-focused submodels based on outdoor conditions or detected operational mode.

Higher Temporal Resolution: Applying the same methodological framework to hourly or sub-hourly data would enable more precise modeling of building thermal dynamics and support more accurate short-term forecasting. This would also allow investigation of time-of-day effects, solar radiation impacts, and the influence of diurnal temperature variations on consumption patterns.

Incorporation of Additional Features: Systematically evaluating the contribution of additional features such as solar radiation, occupancy data, equipment schedules, and calendar features would help identify the most cost-effective data collection strategies for improving predictive accuracy. This could involve controlled experiments where buildings with different levels of data availability are compared to quantify the marginal value of each additional data source.

Deep Learning Approaches: Exploring deep learning architectures such as Long Short-Term Memory (LSTM) networks or Temporal Convolutional Networks (TCN) could potentially capture more complex temporal dependencies and nonlinear relationships than the tree-based methods evaluated in this study. However, deep learning approaches typically require larger datasets and more computational resources, and their performance advantage over gradient boosting [6] in this domain remains an open question.

Transfer Learning Across Buildings: Investigating whether models trained on one building can be effectively transferred or adapted to other buildings (within the same city or climate zone) could reduce the data requirements for deploying predictive models in new buildings. This could involve techniques such as domain adaptation, meta-learning, or hierarchical modeling that leverages commonalities across buildings while accounting for building-specific differences.

Interpretability and Explainability: While ensemble methods [4] provide superior predictive accuracy, their "black box" nature can limit trust and adoption in practice. Applying interpretability techniques such as SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations) to the best-performing models would provide insights into how the models make predictions, which features are most influential for specific predictions, and whether the models have learned physically plausible relationships. This could increase practitioner confidence and facilitate model debugging and refinement.

This thesis set out to enhance the traditional energy signature model by developing a multidimensional, machine learning-based approach for predicting building energy consumption. By leveraging comprehensive datasets from seven distinct building locations across Italy and systematically evaluating a range of models from simple linear regressions to state-of-the-art gradient boosting [6] algorithms [3, 17], this research has successfully demonstrated the significant advantages of a data-driven, multidimensional approach.

The key findings of this study can be summarized as follows:

1. **Machine learning models significantly outperform the traditional energy signature.** The experimental results provide unequivocal evidence that by incorporating additional weather variables and employing machine learning techniques, it is possible to achieve a substantial improvement in predictive accuracy compared to the simple temperature-only model. The R^2 values for the best models were consistently and significantly higher than those of the baseline linear regressions.

2. **Ensemble methods, particularly gradient boosting [6], are the most effective.** Across all seven locations, ensemble learning models (Random Forest, XGBoost, and CatBoost) proved to be the most powerful tools for this task. The gradient boosting [6] algorithms [3, 17], XGBoost and CatBoost, were the top performers in six of the seven cases, highlighting their ability to capture the complex, non-linear relationships inherent in building energy consumption data.

3. **A multidimensional feature set is crucial for accuracy.** The success of the models was not just due to the algorithms themselves, but also to the enriched feature set. The inclusion of variables such as humidity, wind, and engineered features like Heating Degree Days (HDD) and rolling temperature averages provided the models with the necessary information to move beyond the limitations of the traditional approach.

4. **Model performance is context-dependent.** The study also revealed that the predictive accuracy of the models varies depending on the local climate and specific building characteristics. This underscores the importance of developing tailored, building-specific models rather than applying a generic solution.

In conclusion, this thesis has successfully achieved its objectives. It has developed and validated a robust methodology for creating multidimensional energy signatures that are more accurate and insightful than the traditional model. The findings provide a clear roadmap for practitioners and researchers looking to apply advanced data analytics to the field of building energy management. By embracing these data-driven techniques, the building industry can unlock new opportunities for energy efficiency, cost savings, and sustainability. The framework established in this research not only serves as a more accurate tool for performance assessment but also paves the way for reliable short-term energy forecasting, a critical component of the next generation of smart building technologies.

Chapter 6

Conclusions and Future Work

6.1 Summary of Contributions

This thesis successfully addressed the limitations of traditional, temperature-centric energy signature models by developing a robust, multidimensional energy signature based on advanced machine learning techniques. The key contributions are summarized as follows:

- **Validation of Ensemble Methods:** The study provided empirical evidence across seven distinct Italian locations that ensemble tree-based models [**?**, 3] (XGBoost, CatBoost, Random Forest) are significantly more effective than linear models for daily energy consumption forecasting, owing to their ability to capture non-linear relationships. The best models achieved R^2 values exceeding 0.74 for high-performance buildings.

- **Domain-Specific Feature Engineering:** The work confirmed the critical importance of domain-specific feature engineering [11], particularly the use of Heating/Cooling Degree Days (HDD/CDD), rolling averages of temperature to model the thermal inertia of buildings, and temporal features to capture occupancy patterns. These engineered features proved to be among the most important predictors across all models.

- **Cross-City Performance Assessment:** A comprehensive cross-city analysis was performed across seven distinct Italian locations, revealing that model predictability is highly dependent on the consistency of the building's operational profile. High R^2 scores (>0.74) were achieved for buildings with predictable loads and dominant thermal drivers, while lower scores highlighted the need for additional non-weather-related data for buildings with stochastic loads.

- **Practical Methodology Framework:** The thesis presents a complete, reproducible methodology for developing energy signature models, from data preprocessing through feature engineering [11] to model selection and evaluation. This framework can be applied to other buildings and geographic locations.

- **Identification of Building-Specific Factors:** The analysis identified that approximately 50% of locations achieve high predictability ($R^2 > 0.70$) from meteorological factors alone, while the remaining locations require additional data sources to improve model accuracy.

The developed models provide a practical and highly accurate tool for energy managers to predict consumption, diagnose inefficiencies, and optimize building operation. For high-performance buildings like Firenze and Milano TURRO 28, the models enable reliable short-term forecasting and precise baseline establishment for energy auditing. For lower-performance buildings, the models identify the limitations of meteorological-only approaches and highlight the need for additional data sources.

6.2 Future Work

To further enhance the accuracy and applicability of the multidimensional energy signature, several avenues for future research are proposed:

1. **Integration of Occupancy Data:** The most significant limitation identified was the unexplained variance in buildings with low R^2 scores. Future work should focus on integrating direct or proxy data for building occupancy (e.g., Wi-Fi connection counts, access card swipes, calendar data, or CO_2 sensors) to better model internal loads. This could potentially increase R^2 by 0.10-0.20 for low-performance buildings.

2. **Deep Learning Architectures:** Exploring deep learning models, such as Long Short-Term Memory (LSTM) or Gated Recurrent Unit (GRU) networks, could provide valuable comparisons. These architectures are specifically designed to capture long-term temporal dependencies and complex sequential patterns, which may further improve forecasting accuracy, especially for longer prediction horizons.

3. **Real-Time Forecasting and Model Deployment:** The current study focuses on daily prediction. Future work should adapt the methodology for sub-daily (e.g., hourly) forecasting, which is essential for real-time control and dynamic pricing strategies. This would involve deploying the best-performing models into a production environment for continuous, automated prediction.

4. **Automated Feature Selection and Model Selection:** Developing an automated pipeline that can perform feature selection [13] and model selection specific to each new building's data profile could improve the scalability of the methodology. This would allow the system to automatically determine whether XGBoost, CatBoost, or Random Forest is the optimal choice for a given location, potentially using meta-learning approaches.

5. **Transfer Learning and Domain Adaptation:** Investigating transfer learning approaches where models trained on high-performance buildings are fine-tuned for low-performance buildings could improve generalization and reduce the data requirements for new locations.

6. **Uncertainty Quantification:** Developing probabilistic forecasting models that provide confidence intervals around predictions would enhance the practical utility of the models for risk-aware decision-making in energy procurement and demand response.

7. **Seasonal and Trend Analysis:** Implementing seasonal decomposition and trend analysis techniques could help identify and model non-stationary components in the energy consumption time series, potentially improving model stability across different time periods.

6.3 Final Remarks

This thesis demonstrates that machine learning, particularly ensemble tree-based methods, offers a powerful and practical approach to developing multidimensional energy signatures that transcend the limitations of traditional temperature-only models. The work successfully bridges the gap between academic research and practical application, providing energy managers with tools that can be immediately deployed for energy auditing, forecasting, and optimization.

The significant variation in model performance across locations underscores an important insight: building energy consumption is not a one-size-fits-all phenomenon. The most effective approach to energy management requires understanding the specific characteristics of each building and selecting modeling approaches accordingly. For buildings with dominant thermal loads and consistent operational profiles, the developed models provide near-perfect predictions. For buildings with more complex energy dynamics, the models identify the need for additional data sources and more sophisticated approaches.

As the world transitions towards more sustainable and efficient energy systems, the ability to accurately predict and understand building energy consumption becomes increasingly critical. This thesis contributes to that transition by providing both the methodological framework and the empirical evidence that advanced machine learning techniques can substantially improve our understanding and management of building energy performance.

Bibliography

[1] Kadir Amasyali and Nora M El-Gohary. A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, 81:1192–1205, 2018.

[2] Mathieu Bourdeau, Xiaoqiang Zhai, Elyes Nefzaoui, Xiwang Guo, and Patrice Chatellier. Modeling and forecasting building energy consumption: A review of data-driven techniques. *Sustainable Cities and Society*, 48:101533, 2019.

[3] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[4] Thomas G Dietterich. Ensemble methods in machine learning. In *International workshop on multiple classifier systems*, pages 1–15. Springer, 2000.

[5] Magnus Eriksson, Jan Vesterberg, and Jan Akander. Development and validation of energy signature method. *Energy and Buildings*, 210:109745, 2020.

[6] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

[7] Nelson Fumo. A review on the prediction of building energy consumption and its influence on the design of energy-efficient buildings. *Renewable and Sustainable Energy Reviews*, 37:1–10, 2014.

[8] Nelson Fumo and M Rafe Biswas. Regression analysis for prediction of residential energy consumption. *Renewable and Sustainable Energy Reviews*, 47:332–343, 2015.

[9] Alberto Gonzalez-Vidal, Fernando Jimenez, and Antonio F Gomez-Skarmeta. A methodology for energy multivariate time series forecasting in smart buildings based on feature selection. *Energy and Buildings*, 196:71–82, 2019.

[10] Jessica Granderson, Mary Ann Piette, and Girish Ghatikar. Building energy information systems: user case studies. *Energy Efficiency*, 4(1):17–30, 2011.

[11] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182, 2003.

[12] Douglas M Hawkins. The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1):1–12, 2004.

[13] Ron Kohavi and George H John. Wrappers for feature subset selection. *Artificial intelligence*, 97(1-2):273–324, 1997.

[14] Seung-Hee Oh and Jeff S Haberl. Large scale energy signature analysis: Tools for utility energy managers. *Sustainability*, 14(14):8649, 2022.

[15] José Pérez-Lombard, José Ortiz, and Christine Pout. A review on buildings energy consumption information. *Energy and buildings*, 40(3):394–398, 2008.

[16] Luca Pistore, Giovanni Pernigotto, Francesca Cappelletti, Piercarlo Romagnoni, and Andrea Gasparella. From energy signature to cluster analysis. In *Proceedings of the 12th International Conference on Healthy Buildings*, 2016.

[17] Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. Catboost: unbiased boosting with categorical features. *Advances in neural information processing systems*, 31, 2018.

[18] J. Ross Quinlan. Induction of decision trees. *Machine learning*, 1:81–106, 1986.

[19] Luís Sanhudo, João Rodrigues, João P Martins, Nuno M Ramos, Ricardo M Almeida, Eva Barreira, João S Ramos, and Nuno M Ramos. Multivariate time series clustering and forecasting for building energy analysis: Application to weather data quality control. *Journal of Building Engineering*, 35:101996, 2021.

[20] Pierluigi Siano. Demand response and smart grids—a survey. *Renewable and sustainable energy reviews*, 30:461–478, 2014.

[21] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

[22] Paul Westermann, Chirag Deb, Arno Schlueter, and Ralph Evins. Unsupervised learning of energy signatures to identify the heating system and building type using smart meter data. *Applied Energy*, 264:114715, 2020.

[23] Cort J Willmott and Kenji Matsuura. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research*, 30(1):79–82, 2005.

[24] Hai-Xiang Zhao and Frédéric Magoulès. A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, 16(6):3586–3592, 2012.