Applied Mathematical Sciences, Vol. 20, 2026, no. 3, 131 - 148
HIKARI Ltd, www.m-hikari.com
https://doi.org/10.12988/ams.2026.919331

Price Prediction and Optimization of Energy Sharing in Renewable Energy Communities

Marta Biancardi

Department of Economics and Finance
University of Bari Aldo Moro
Largo Abbazia S. Scolastica, Bari (BA), Italy

Paola Catalano

Department of Social Sciences
University of Foggia
Via A. da Zara, 11, Foggia (FG), Italy

Giacinto Angelo Sgarro

Department of Social Sciences
University of Foggia
Via A. da Zara, 11, Foggia (FG), Italy

Antonio Vairo

Department of Social Sciences
University of Foggia
Via A. da Zara, 11, Foggia (FG), Italy

Abstract

The growth of Renewable Energy Communities (RECs) across Europe has increased the need for reliable tools to assess their technical performance and economic feasibility. In Italy, the implementation of the RED II Directive has made the design of RECs particularly relevant, especially under collective self-consumption schemes, where performance depends on the temporal alignment between photovoltaic generation and demand. This study proposes an integrated framework combining physics-based modeling and data-driven techniques to analyze and optimize energy sharing within RECs. Using real consumption profiles from ARERA and photovoltaic production simulated through PVGIS for a 3 kW system, multiple community configurations are evaluated to identify the structure that maximizes shared energy. Energy flows are modeled through a physically consistent formulation linking production, self-consumption, injected energy, and shared energy. After identifying the optimal REC configuration through combinatorial analysis, a machine learning approach is applied to model the relationship between energy variables and the economic value of shared energy, focusing on the best-performing configuration. The results show that the system is governed by a constrained relationship driven by the minimum between injected energy and demand. The superior performance of the linear model suggests that the problem is highly structured and largely predictable. Overall, the framework combines physical interpretability and predictive modeling, providing a practical and transparent tool for REC design and optimization.

Mathematics Subject Classification: 62M10, 68T05

Keywords: Renewable Energy Communities (RECs), Machine Learning, Incentive Scheme, Residential Consumer

1 Introduction

The global transition toward low-carbon energy systems has accelerated the development of decentralized energy paradigms, where Renewable Energy Communities (RECs) play a pivotal role. RECs enable collective generation, consumption, and sharing of renewable energy, fostering environmental sustainability, economic efficiency, and social inclusion. As highlighted by Belloni et al. [4], RECs represent a fundamental mechanism to promote distributed generation and optimize local energy exchanges, particularly when supported by appropriate planning tools and incentive schemes. At the European level, the regulatory framework established by Directive (EU) 2018/2001 (RED II) has formalized the role of citizens and local actors in energy markets. In Italy, this framework has been implemented through incentive schemes based on virtual self-consumption, where shared energy is defined as the minimum between injected and consumed energy within the community [2]. This regulatory design makes the temporal alignment between generation and demand a key performance driver, thus increasing the relevance of accurate forecasting and operational optimization.

A large body of literature has investigated the techno-economic optimization of RECs. Kachhad et al. [9] propose a Mixed Integer Non-Linear Programming approach for optimal sizing of photovoltaic and storage systems, showing how coordinated investment decisions improve profitability. Similarly, Barchi et al. [3] demonstrate that increasing the number of participants and integrating storage enhances both energy efficiency and economic returns in residential communities. More recent contributions emphasize the importance of uncertainty-aware operation. Nammouchi et al. [14] show that deterministic scheduling approaches may lead to infeasible or suboptimal outcomes when real-world variability in demand and renewable generation is considered. By introducing a robust optimization framework, the authors demonstrate that accounting for worst-case scenarios leads to more reliable and resilient REC operation, particularly in the Italian incentive context. Parallel research has explored economic and regulatory aspects. Magni et al. [13] show that incentive design significantly affects REC diffusion and regional equity. De Villena et al. [8] propose centralized allocation mechanisms to minimize system costs while ensuring fairness, whereas Awerkin et al. [2] highlight the importance of incentive-sharing rules using a game-theoretic framework. Another important research stream concerns local market design and peer-to-peer (P2P) energy trading. Alahmed and Tong [1] propose dynamic pricing schemes to align individual incentives with social welfare. Pereira et al. [15] introduce a multi-agent deep reinforcement learning framework for P2P energy trading, showing that adaptive agent-based strategies can significantly reduce energy costs and improve market efficiency. Beyond local optimization and trading, recent studies have started to investigate the system-level impact of RECs on electricity markets. Koltunov et al. [10] analyze the interaction between RECs and the Italian day-ahead market using a combined engineering and economic modeling approach. Their results indicate that REC deployment can simultaneously increase daytime supply, reduce grid exchanges through self-consumption, and potentially lower wholesale prices, highlighting both positive and negative systemic effects depending on seasonal conditions.

In parallel, the integration of machine learning techniques into energy systems has gained increasing attention. Liu et al. [11] demonstrate that combining feature selection with machine learning improves predictive performance in energy applications. Biancardi and Catalano [5] show that advanced models such as XGBoost outperform traditional approaches in electricity price forecasting, while emphasizing the need for interpretability in decision-support contexts. Despite these advances, the literature remains fragmented across three main dimensions: techno-economic optimization, market and regulatory analysis, and data-driven modeling. Moreover, existing machine learning applications often prioritize predictive accuracy over interpretability, limiting their practical adoption in REC contexts where transparency and stakeholder trust are essential. At the same time, robust optimization approaches and market-level analyses are rarely integrated with data-driven predictive frameworks.

This paper addresses these gaps by proposing a unified framework that integrates predictive modeling, comparative machine learning analysis, and Explainable Artificial Intelligence (XAI) techniques for RECs. By leveraging real-world data and incorporating interpretability tools such as SHapley Additive exPlanations (SHAP) values, the study provides both high predictive accuracy and a transparent understanding of the key drivers underlying shared energy dynamics.

1.1 Research Gap and Objectives

Despite the rapid growth of Renewable Energy Communities (RECs) in both policy and academic research, the existing literature remains fragmented across multiple analytical dimensions. A first stream of studies focuses on techno-economic optimization, primarily addressing system sizing and operational strategies through deterministic or optimization-based approaches. Kachhad et al. [9] propose a Mixed Integer Non-Linear Programming framework for optimal sizing of photovoltaic and storage systems, showing how coordinated investment decisions improve profitability. Similarly, Barchi et al. [3] demonstrate that increasing the number of participants and integrating storage enhances both energy efficiency and economic returns. While these contributions provide valuable insights into system design, they often rely on simplified assumptions and lack integration with real-world operational variability. A second stream of research investigates regulatory frameworks and economic mechanisms. Magni et al. [13] show that incentive design significantly affects REC diffusion and regional equity. De Villena et al. [8] propose centralized allocation mechanisms to minimize system costs while ensuring fairness among participants. In addition, Awerkin et al. [2] emphasize the importance of incentive-sharing rules through a game-theoretic framework. Although these studies provide important policy-oriented insights, they typically abstract from the dynamic and data-driven nature of energy flows, limiting their applicability for operational decision-making. More recent contributions explore decentralized and uncertainty-aware approaches. Pereira et al. [15] introduce a multi-agent reinforcement learning framework for peer-to-peer energy trading, demonstrating that adaptive strategies can significantly improve local market efficiency. In parallel, Nammouchi et al. [14] show that deterministic scheduling approaches may lead to infeasible solutions under uncertainty, and propose robust optimization techniques to improve system reliability. However, these approaches are often developed in isolation and do not explicitly address model interpretability or their integration into comprehensive decision-support frameworks.

At the system level, the impact of RECs on electricity markets remains underexplored. Koltunov et al. [10] analyze the interaction between RECs and the Italian day-ahead market, showing that REC deployment can influence both energy exchanges and price dynamics. Nevertheless, the connection between predictive modeling, operational behavior, and system-level outcomes remains limited. In parallel, machine learning techniques have been increasingly applied to energy forecasting problems. Liu et al. [11] demonstrate that combining feature selection with machine learning improves predictive accuracy in energy applications. Biancardi and Catalano [5] show that advanced models such as XGBoost outperform traditional approaches in electricity price forecasting. However, these models are often treated as black boxes, with limited attention to interpretability and transparency, which are crucial in community-based energy systems. Therefore, a clear research gap emerges from the lack of an integrated framework that simultaneously combines data-driven predictive modeling, interpretability of machine learning models, and application to Renewable Energy Communities within a realistic operational and regulatory context.

This paper aims to address this gap by developing a comprehensive analytical framework for RECs that integrates predictive modeling, comparative machine learning analysis, and XAI techniques. Specifically, the study pursues three main objectives. First, it evaluates and compares the performance of different machine learning models (including Linear Regression, Decision Trees, Random Forest, and XGBoost) in predicting key energy indicators within RECs. Second, it enhances model transparency through the application of interpretability tools such as SHAP values, enabling a deeper understanding of the main drivers of shared energy dynamics. Third, it bridges the gap between data-driven modeling and practical decision-making by providing actionable insights to support the design, optimization, and management of RECs within existing regulatory frameworks. By pursuing these objectives, the paper contributes to the literature by integrating traditionally separate research domains and promoting a more transparent and decision-oriented approach to REC analysis.

Roadmap of the paper. The remainder of the paper is organized as follows. Section 2 presents the methodological framework, including the construction of household demand archetypes, the modeling of photovoltaic generation and energy flows, the combinatorial generation of Renewable Energy Community configurations, and the machine learning and explainability tools adopted in the analysis. Section 3 reports the empirical results, first discussing the identification of the optimal REC configurations and then presenting the predictive performance of the selected machine learning models together with the corresponding interpretability analysis. Finally, Section 4 summarizes the main findings, discusses the limitations of the study, and outlines possible directions for future research.

2 Methodology

This section presents the methodological framework adopted to model and analyze energy sharing dynamics within Renewable Energy Communities (RECs). The approach integrates data-driven modeling, regulatory constraints, and machine learning techniques to provide a comprehensive representation of energy flows and economic incentives. The starting point of the analysis is the construction of representative electricity demand profiles. The baseline data were obtained from the standard residential load profile provided by ARERA (Autorità di Regolazione per Energia Reti e Ambiente), which reflects the average hourly consumption behavior of residential users in Italy. The baseline load profile was used to generate four distinct household archetypes, identified as follows: A#1 Retired couple, A#2 Family with two young commuter workers, A#3 Family with school-aged children and working parents, and A#4 Single commuter worker. These archetypes capture heterogeneous consumption behaviors driven by occupancy patterns and daily routines. The differentiation process was achieved by temporally redistributing the baseline energy demand, while preserving the same average hourly consumption. The redistribution was implemented using a Shapley value-based approach, which allows for a consistent allocation of variations in consumption by accounting for the marginal contribution of each user type. This method ensures fairness and coherence in the generation of synthetic load profiles. Furthermore, the temporal structure of consumption was adjusted according to the Italian time-of-use tariff scheme (fasce orarie F1, F2, F3), defined by ARERA. This adjustment ensures that the generated profiles realistically reflect behavioral responses to price signals.

Photovoltaic (PV) generation data were obtained from PVGIS (Photovoltaic Geographical Information System), an official tool developed by the European Commission for estimating solar energy production. The analysis considers a residential photovoltaic system with an installed capacity of 3 kW located in Naples. Standard installation parameters were adopted, including typical tilt angle and south-facing orientation, ensuring consistency with common residential configurations. Hourly production data over a full year were extracted to match the temporal resolution of the consumption profiles. The economic component of the model is based on the Italian electricity market and regulatory framework. The electricity purchase price was derived from the single national hourly price (PUN), which represents the reference wholesale electricity price in Italy. To approximate the actual cost faced by end users, additional components were included, such as distribution losses, system charges, and taxes. These elements ensure that the modeled electricity price reflects real market conditions. The selling price of electricity was estimated as a fraction of the PUN, reflecting common practices in the literature and market mechanisms, where prosumers receive a reduced price for injected energy due to transaction costs and aggregation processes. The incentive for shared energy was modeled according to the Italian regulatory framework defined by the Ministerial Decree of 7 December 2023 (DM CACER), implemented by the Gestore dei Servizi Energetici (GSE). This scheme introduces a tariff composed of a fixed component (approximately 0.08 €/kWh) and a variable component linked to market conditions. The total economic value associated with shared energy is therefore determined by the combination of avoided costs and regulatory incentives.

The behavior of the energy community is modeled through a deterministic simulation of energy flows at hourly resolution. For each time step, photovoltaic production, prosumer consumption, and aggregated demand of the remaining users are computed. Autoconsumption is defined as the minimum between photovoltaic production and prosumer demand. The energy injected into the grid corresponds to the residual production. The energy shared within the community is calculated as the minimum between injected energy and aggregated demand, reflecting the physical constraint that only the portion of energy simultaneously available and demanded can be shared. This formulation introduces a nonlinear relationship governed by a minimum operator, which represents the core mechanism underlying energy sharing in RECs. The residual energy not absorbed by the community is ultimately injected into the grid. All possible configurations of the community were simulated by varying the role of each user as prosumer or consumer. For each configuration, key performance indicators were computed, including total shared energy and residual injected energy.

2.1 Construction of REC configurations

The empirical analysis was developed in two sequential steps. First, all feasible Renewable Energy Community (REC) configurations were generated through a combinatorial procedure. Specifically, the four user profiles considered in the study (A#1, A#2, A#3, and A#4) were alternately assigned the role of prosumer, while all possible combinations of the remaining profiles were considered as consumers. For each configuration, hourly photovoltaic production was denoted as $E_{PV,t}$, while the prosumer's own demand was denoted as $C_{P,t}$. The prosumer's self-consumed energy was computed as:

$$SC_t = \min(E_{PV,t}, C_{P,t}) \tag{1}$$

The initial injected energy was then obtained as the residual surplus:

$$IE_t = E_{PV,t} - SC_t \tag{2}$$

For each consumer combination, the aggregated hourly demand was calculated as:

$$D_t = \sum_{j \in \mathcal{C}} C_{j,t} \tag{3}$$

where $\mathcal{C}$ denotes the set of consumers included in the given configuration. The shared energy, which represents the key performance variable of the REC under the Italian regulatory framework, was computed as the minimum between the injected surplus and the aggregated consumer demand:

$$SE_t = \min(IE_t, D_t) \tag{4}$$

Finally, the residual injected energy after sharing was defined as:

$$IE_t^{res} = \max(0; IE_t - SE_t) \tag{5}$$

For each configuration, the cumulative values of shared energy, residual injected energy, and self-consumed energy were then obtained by summing the corresponding hourly quantities over the full time horizon. These aggregated indicators were used to rank the feasible configurations and identify the best-performing REC structure.

In a second step, the machine learning dataset was constructed by selecting the optimal configuration identified in the combinatorial analysis. For this single best configuration, the same hourly variables were recomputed, namely photovoltaic production, self-consumption, injected energy, aggregated demand, and shared energy. These variables constitute the feature set used in the predictive phase. The target variable was defined as the economic value associated with shared energy. In operational terms, this quantity reflects the monetary benefit generated by shared energy and was computed as the product of hourly shared energy and the difference between the electricity purchase price and the selling price. The resulting machine learning dataset therefore consists of the following explanatory variables: photovoltaic production, self-consumption, injected energy, aggregated demand, and shared energy, while the target variable captures the economic value of shared energy under the selected REC configuration.

2.2 Machine Learning Models and Explainability Framework

To further investigate the relationship between energy variables and the economic incentive associated with shared energy, a supervised machine learning approach was adopted. The predictive analysis was carried out on the optimal REC configuration identified in the previous phase. In this setting, the target variable represents the monetary value of shared energy, while the input feature set includes photovoltaic production (E_{PV}), self-consumed energy (SC), injected energy (IE), total demand (D), and shared energy (SE). These variables collectively describe the physical and economic interactions within the selected Renewable Energy Community (REC). The modeling strategy aims to compare different levels of model complexity, ranging from linear approaches to advanced ensemble methods, to assess their ability to capture both linear and nonlinear relationships.

Linear Regression Linear Regression is used as a baseline model, assuming a linear relationship between the target variable y and the feature vector $\mathbf{x} = (x_1, x_2, \ldots, x_p)$. The model is defined as:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \epsilon_i \tag{6}$$

where β_0 is the intercept, β_j are the regression coefficients, and ϵ_i is the error term. The parameters are estimated by minimizing the Residual Sum of Squares (RSS):

$$\min_{\beta} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{7}$$

Despite its simplicity, Linear Regression provides a highly interpretable model and serves as a benchmark for evaluating more complex algorithms.

Decision Tree Regressor Decision Trees [16] partition the feature space into disjoint regions by recursively selecting splits that minimize prediction error. At each node, the algorithm selects the feature x_j and threshold t that minimize the Mean Squared Error (MSE):

$$\min_{j,t} \left[\sum_{x_i \in R_1(j,t)} (y_i - \bar{y}_{R_1})^2 + \sum_{x_i \in R_2(j,t)} (y_i - \bar{y}_{R_2})^2 \right] \tag{8}$$

where R_1 and R_2 represent the two regions generated by the split. Decision Trees are able to capture nonlinear relationships and interactions among variables, but they may suffer from overfitting and instability.

Random Forest Random Forest [6] is an ensemble learning method that builds multiple decision trees on bootstrap samples of the data and aggregates their predictions. The final prediction is obtained as:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^{B} T_b(x) \tag{9}$$

where $T_b(x)$ denotes the prediction of the b-th tree and B is the total number of trees. By introducing randomness in both data sampling and feature selection, Random Forest reduces variance and improves generalization, making it particularly effective for complex datasets.

XGBoost eXtreme Gradient Boosting (XGBoost) [7] is a gradient boosting algorithm that constructs models sequentially by minimizing a differentiable loss function. At iteration t, the prediction is updated as:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta f_t(x_i) \tag{10}$$

where f_t is the newly added tree and η is the learning rate. The objective function combines a loss term and a regularization component:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} L(y_i, \hat{y}_i^{(t)}) + \sum_{t=1}^{T} \Omega(f_t) \tag{11}$$

where $\Omega(f_t)$ penalizes model complexity. XGBoost is particularly effective in capturing nonlinear patterns and interactions, while maintaining computational efficiency.

Explainability: SHAP Values To ensure interpretability, SHapley Additive exPlanations (SHAP) [12] values are employed. SHAP is based on cooperative game theory and assigns to each feature a contribution to the prediction:

$$f(x) = \phi_0 + \sum_{j=1}^{p} \phi_j \tag{12}$$

where ϕ_j represents the marginal contribution of feature j and ϕ_0 is the expected model output. SHAP values provide a consistent and locally accurate explanation of model predictions, allowing the identification of the most influential variables.

3 Results

The empirical analysis was developed using the real-world dataset and the updated computational framework described in the previous section. The results are structured into two main parts: (i) identification of optimal Renewable Energy Community (REC) configurations through combinatorial analysis, and (ii) predictive modeling of the economic value associated with shared energy.

3.1 Optimal REC configurations

The combinatorial analysis evaluated all feasible configurations of the Renewable Energy Community (REC) by assigning, in turn, each user profile as prosumer and combining the remaining profiles as consumers. The four profiles considered in the analysis correspond to the household archetypes previously defined, namely: A#1 (Retired couple), A#2 (Family with two young commuter workers), A#3 (Family with school-aged children and working parents), and A#4 (Single commuter worker). For each configuration, hourly energy flows were computed by combining photovoltaic production, prosumer consumption, and aggregated demand of the selected consumers. In particular, the following quantities were derived: self-consumed energy, injected energy, aggregated demand, shared energy, and residual injected energy. The key performance indicator used to evaluate each configuration is the total shared energy over the full time horizon. The best-performing configurations are summarized in Table 1, which reports the top combinations ranked according to total shared energy. The results show that the most efficient configurations

Prosumer	Consumers	Shared Energy	Injected Energy	Self-Consumption
A#2	A#1, A#3, A#4	3444.51	363.81	4662.69
A#4	A#1, A#3, A#2	3441.44	363.81	4665.75
A#3	A#1, A#2, A#4	3382.05	426.27	4725.15
A#4	A#1, A#3	2965.18	840.07	4102.33
A#3	A#1, A#4	2905.78	901.12	4055.67

Table 1: Top-performing REC configurations in terms of shared energy. A#1–A#4 denote the four household archetypes considered in the analysis.

are those involving the maximum number of consumers. In particular, the best-performing configuration is obtained when A#2 (Family with two young commuter workers) acts as prosumer, while the remaining three profiles (A#1, A#3, and A#4) act as consumers. This configuration achieves the highest level of shared energy, with limited residual injected energy and a high level of self-consumption. Very similar results are observed when A#4 (Single commuter worker) is selected as prosumer, indicating that both profiles generate a balanced interaction between production and demand. A third highly competitive configuration is obtained when A#3 (Family with school-aged children and working parents) acts as prosumer, showing slightly lower shared energy but higher self-consumption levels. In contrast, configurations involving only two consumers exhibit significantly lower performance. In these cases, the mismatch between production and demand increases, leading to higher levels of residual injected energy and lower shared energy. This confirms that increasing the number of consumers improves the system's ability to absorb locally generated energy. Overall, these findings highlight that REC efficiency depends on both the identity of the prosumer and the size and composition of the consumer group. In particular, a larger and more diverse demand base enhances the temporal alignment between generation and consumption, thereby maximizing shared energy.

3.2 Prediction of the economic value of shared energy

Following the combinatorial optimization phase, a predictive modeling framework was applied to the best-performing REC configuration to estimate the economic value associated with shared energy. The target variable is defined as the monetary benefit derived from shared energy, computed as the product of shared energy and the difference between purchase and selling prices. The input features include photovoltaic production, self-consumed energy, injected energy, total demand, and shared energy. These variables capture the main physical and economic mechanisms governing the selected REC configuration. Four machine learning models were implemented: Linear Regression, Random Forest, XGBoost, and a Decision Tree model. The dataset was split chronologically into training (75%) and test (25%) sets to ensure a realistic forecasting scenario. To ensure a consistent and fair comparison across models, each algorithm was implemented using a standardized set of hyperparameters, selected to balance interpretability, generalization capability, and computational efficiency [17]. The Linear Regression model was configured with an intercept term enabled, allowing the model to account for baseline effects in the data while preserving the original feature structure. For tree-based methods, a common splitting criterion based on squared error minimization was adopted. The Decision Tree Regressor was implemented without imposing depth constraints, allowing the model to fully explore hierarchical partitions of the feature space. Default values were maintained for the minimum number of samples required for splitting and leaf creation, ensuring maximum flexibility in capturing local patterns. The Random Forest model was constructed as an ensemble of decision trees trained on bootstrap samples of the dataset. A fixed number of trees was used to ensure stability, while all available features were considered at each split. This configuration allows reducing variance and improving generalization while maintaining interpretability of the aggregated model behavior. Finally, the XGBoost model was implemented using a regression objective function based on squared error. The boosting structure was kept at its baseline configuration, with standard values for key parameters such as learning rate, tree depth, and number of estimators. This choice ensures consistency in the comparison and avoids overfitting through excessive tuning. The full set of hyperparameters used in the analysis is summarized in Table 2. The predictive performance of the models was evaluated using the coefficient of determination (R^2). The results reveal an unexpected but highly informative

Model	Hyperparameter	Value
Linear Regression	fit_intercept	True
	copy_X	True
Decision Tree	criterion	squared_error
	splitter	best
	max_depth	None
	min_samples_split	2
	min_samples_leaf	1
Random Forest	n_estimators	100
	criterion	squared_error
	bootstrap	True
	max_features	1.0
XGBoost	objective	reg:squarederror
	learning_rate	default
	max_depth	default
	n_estimators	default

Table 2: Hyperparameter configuration adopted for the machine learning models. Default values refer to standard library settings

outcome. Linear Regression achieves the best predictive performance, with an R^2 equal to **0.9176**. The Random Forest model reaches an R^2 of **0.8947**, followed by XGBoost with **0.8861**, while the Decision Tree model shows lower performance with an R^2 of **0.7822**. Table 3 summarizes the performance of the different algorithms. This result suggests that the relationship between

Model	R^2	MAE	MSE
Linear Regression	0.9176	0.0087	0.0004
Random Forest	0.8947	0.0085	0.0005
XGBoost	0.8861	0.0095	0.0005
Decision Tree	0.7822	0.0119	0.0010

Table 3: Predictive performance of the machine learning models in terms of R^2, Mean Absolute Error (MAE), and Mean Squared Error (MSE)

the input variables and the economic value of shared energy is largely linear and structurally well-defined. In particular, since the target variable is directly derived from shared energy and price differences, the underlying functional relationship can be effectively captured by a linear model without requiring more complex nonlinear approximations. More advanced models, such as Random Forest and XGBoost, while generally more flexible, do not provide additional predictive gains in this context. This behavior is consistent with situations in which the signal-to-noise ratio is high and the data-generating process is gov-

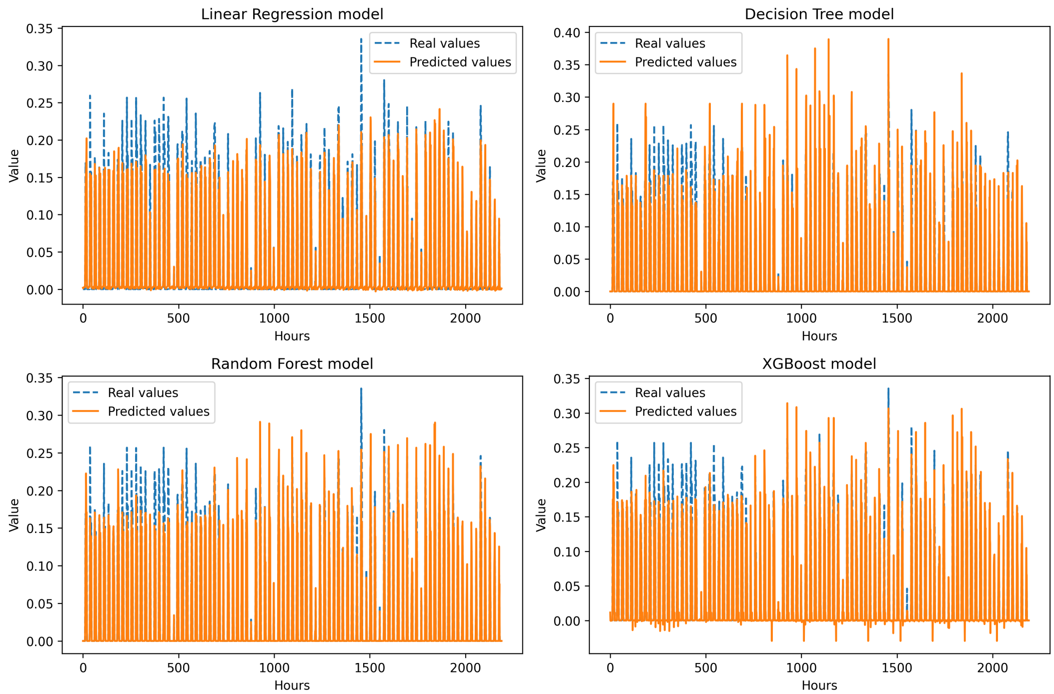

Figure 1: Comparison between observed and predicted values of the economic value associated with shared energy over the test set for the four models

erned by explicit deterministic relationships rather than complex interactions. Figure 1 provides a visual comparison between observed and predicted values for the different models. The results confirm that all models achieve a high level of accuracy, but the linear model exhibits the closest alignment with the actual values, highlighting its effectiveness in capturing the underlying structure of the problem. These findings indicate that, in the context of Renewable Energy Communities, model complexity should be carefully balanced against the intrinsic structure of the problem. When the economic mechanisms are directly linked to physical energy flows, simpler and more interpretable models may outperform more sophisticated approaches.

3.3 Model interpretability

To further investigate the drivers of the model predictions, SHapley Additive exPlanations (SHAP) values were employed. Figure 2 reports the SHAP summary plot, which ranks features according to their contribution to the predicted economic value of shared energy. The results clearly indicate that shared energy (SE) is the dominant explanatory variable. Higher values of SE are consistently associated with a strong positive contribution to the model output, confirming its central role in determining the economic value within the

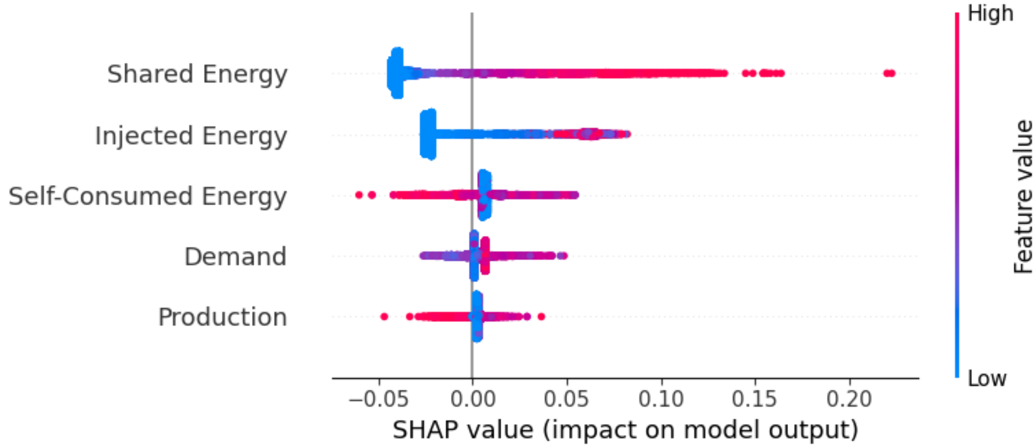

Figure 2: SHAP summary plot showing the contribution of each feature to the prediction of the economic value of shared energy. Features are ranked by importance

REC framework. This result is fully consistent with the regulatory structure, where incentives are directly linked to the amount of energy shared within the community. The variable injected energy (IE) shows a positive but less pronounced effect, suggesting that surplus energy contributes to the economic outcome only indirectly, through its potential to be shared. In contrast, self-consumed energy (SC) exhibits a more balanced impact, reflecting the trade-off between local consumption and the availability of surplus energy for sharing. The contribution of total demand (D) is moderate, indicating that higher demand supports energy sharing but does not independently determine the economic value. Finally, photovoltaic production (E_{PV}) has a relatively limited impact, highlighting that generation alone is not sufficient to drive system performance. Overall, the SHAP analysis confirms that the economic value of RECs is primarily driven by the effective alignment between production and demand, rather than by generation capacity alone. This result reinforces the interpretation of the system as being governed by a constrained supply–demand matching mechanism. This evidence is also consistent with the strong performance of the Linear Regression model, suggesting that the dominant contribution of shared energy imposes a relatively regular structure on the prediction problem.

4 Conclusions

This paper proposed a unified data-driven framework for the analysis of Renewable Energy Communities (RECs), integrating deterministic energy flow modeling, combinatorial optimization, comparative machine learning techniques, and Explainable Artificial Intelligence (XAI) tools. The results highlight how the temporal alignment between photovoltaic production and local demand represents the primary driver of shared energy, confirming the central role of community composition and load profiles, in line with previous studies [4].

From a methodological perspective, the comparative analysis of different machine learning models reveals that the relationship between the selected energy variables and the economic value of shared energy is highly structured and, in this context, largely linear. Contrary to common expectations in energy forecasting, the Linear Regression model achieves the highest predictive accuracy, outperforming more complex models such as Random Forest and XGBoost. This finding suggests that the underlying economic mechanism is governed by a well-defined functional relationship, rather than by highly nonlinear interactions. This result can be directly linked to the physical and regulatory formulation of the REC system. Since shared energy is determined by the minimum between injected energy and aggregated demand, the economic value emerges from a constrained supply–demand interaction. As a consequence, the system exhibits limited structural complexity and a strong deterministic component, making it particularly suitable for simple and interpretable models. The explainability analysis further supports this interpretation. SHAP values clearly identify shared energy as the dominant driver of the model output, while other variables, such as injected energy, demand, and self-consumption, play a secondary role. This confirms that REC performance is primarily driven by the effectiveness of the matching between local generation and demand, rather than by individual variables considered in isolation. From an application standpoint, the proposed framework demonstrates that improving REC performance requires enhancing the temporal alignment between production and consumption rather than increasing installed capacity alone. This highlights the importance of community composition and demand diversity as key design variables. Moreover, the effectiveness of simple and interpretable models represents a significant advantage in decision-support contexts, where transparency and stakeholder trust are essential.

Despite these contributions, several limitations remain. First, the analysis relies on deterministic energy profiles and does not explicitly account for uncertainty in photovoltaic production and consumption patterns, which may affect REC performance in real-world conditions. Second, the study considers a limited number of representative user profiles within a single geographical context, potentially restricting the generalizability of the results. Third, the economic modeling adopts simplified assumptions regarding energy prices and market interactions, which may differ from more complex real-world scenarios. Future research can extend this work in several directions. A first avenue concerns the integration of stochastic and robust modeling approaches to explicitly account for uncertainty in renewable generation and demand, improving the reliability of the framework. A second direction involves the inclusion of more advanced market mechanisms, such as peer-to-peer trading and dynamic pricing schemes, which may further enhance REC efficiency [15]. Additionally, extending the analysis to larger and more heterogeneous datasets would allow validating the framework under more realistic conditions and improving model generalization. Finally, future work may explore the integration of explainable artificial intelligence within optimization frameworks, enabling a tighter coupling between predictive accuracy, interpretability, and decision-making.

Overall, this study contributes to bridging the gap between data-driven modeling, economic analysis, and policy-relevant insights in Renewable Energy Communities, providing a flexible, interpretable, and practically applicable framework to support their design and operation.

References

[1] A. S. Alahmed and L. Tong. "Achieving Social Optimality for Energy Communities via Dynamic NEM Pricing". In: *IEEE* (2023).

[2] A. Awerkin, P. Falbo, and T. Vargiolu. "Optimal Investment and Fair Sharing Rules for Incentives in Virtual Renewable Energy Communities". In: *arXiv preprint* (2025).

[3] G. Barchi et al. "Residential Renewable Energy Community: a Techno-Economic Analysis of the Italian Approach". In: *Preprint* (2022).

[4] Elisa Belloni, Davide Fioriti, and Davide Poli. "Optimal design of renewable energy communities (RECs) in Italy: Influence of composition, market signals, buildings, location, and incentives". In: *Electric Power Systems Research* 235 (2024), p. 110895. DOI: 10.1016/j.epsr.2024.110895.

[5] M. Biancardi and P. Catalano. "Forecasting Italian Electricity Prices for Smarter Energy Communities". In: *Applied Mathematical Sciences* 19.8 (2025), pp. 297–310. DOI: 10.12988/ams.2025.919277.

[6] L. Breiman. "Random Forests". In: *Machine Learning* 45 (2001), pp. 5–32.

[7] T. Chen and C. Guestrin. "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. San Francisco, California, USA: Association for Computing Machinery, 2016, 785–794. ISBN: 9781450342322. DOI: 10.1145/2939672.2939785.

[8] Miguel Manuel De Villena et al. “Financial optimization of renewable energy communities through optimal allocation of locally generated electricity”. In: *IEEE Access* 10 (2022), pp. 77571–77587. DOI: 10.1109/ACCESS.2022.3191804.

[9] V. Kachhad et al. “A Techno-Economic Modelling and Component Sizing in Renewable Energy Communities”. In: *Preprint* (2023).

[10] Maksym Koltunov et al. “The Impact of Renewable Energy Communities in the Italian Day-Ahead Electricity Market: A Scenario Analysis”. In: *arXiv preprint* (2025).

[11] X. Liu et al. “Investigating the performance of machine learning models combined with different feature selection methods to estimate building energy consumption”. In: *Energy and Buildings* 273 (2022), p. 112408. DOI: 10.1016/j.enbuild.2022.112408.

[12] Scott M. Lundberg and Su-In Lee. “A unified approach to interpreting model predictions”. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS’17 (2017), 4768–4777.

[13] G. U. Magni, D. Bricca, and S. Familiari. “Economic incentives for Renewable Energy Communities: a scenario analysis”. In: *Journal of Sustainable Development of Energy, Water and Environment Systems* 13.2 (2025). DOI: 10.13044/j.sdewes.d13.0589.

[14] Amal Nammouchi et al. “Robust Operation of Energy Communities in the Italian Incentive System”. In: *IEEE Conference Proceedings* (2023).

[15] Helder Pereira, Luis Gomes, and Zita Vale. “Peer-to-peer energy trading optimization in energy communities using multi-agent deep reinforcement learning”. In: *Energy Informatics* 5.Suppl 4 (2022), p. 44. DOI: 10.1186/s42162-022-00235-2.

[16] J.R. Quinlan. “Induction of decision trees”. In: *Machine Learning* 1 (1986), pp. 81–106. DOI: 10.1007/BF00116251.

[17] Antonio Vairo, Luca Grilli, and Kamile Taujanskaite. “Cross-Country Generalization Bias in Agricultural Production: Do Machine Learning Models Learn the Same Way?” In: *Applied Mathematical Sciences* 20.2 (2026), pp. 69–77.

Received: March 21, 2026; Published: April 8, 2026