

# POLITECNICO DI TORINO

Corso di Laurea Magistrale in Ingegneria Energetica e  
Nucleare

Anno Accademico 2022/2023



**Politecnico  
di Torino**

Tesi di Laurea Magistrale

## **Profilazione di utenze residenziali attraverso tecniche di machine learning nel contesto delle Comunità Energetiche Rinnovabili**

Relatore:  
Prof. Andrea Lanzini

Candidato:  
Lorenzo Giannuzzo

Correlatore:  
Dr. Francesco Demetrio Minuto



*Ai miei genitori,  
da sempre,  
i miei più grandi  
sostenitori*

# Abstract

Le recenti evoluzioni del quadro normativo europeo ed italiano, sul tema energia, hanno manifestato una forte spinta sul tema delle Comunità Energetiche Rinnovabili. La loro diffusione sul territorio rappresenta un'opportunità per accelerare il processo di de-carbonizzazione e di evoluzione del sistema elettrico nazionale, passando da una configurazione centralizzata di difficile gestione ad una segmentata che permette un più agevole bilanciamento della rete. Tuttavia, la costituzione delle Comunità di Energia Rinnovabile è un processo complesso e richiede competenze multidisciplinari che spaziano dal settore economico ed energetico a quello giuridico e amministrativo. La realizzazione di un progetto di successo, nel medio-lungo termine, è strettamente legata al corretto bilanciamento tra la produzione degli impianti e il consumo delle utenze all'interno della comunità, su base oraria.

Allo stato attuale, le infrastrutture di misura del carico elettrico, soprattutto in ambito residenziale, non permettono la rilevazione del dato orario e ciò non consente di valutare con precisione il loro contributo in termini di condivisione dell'energia. Il presente lavoro mira a profilare, con risoluzione oraria, le utenze residenziali che desiderano calarsi in questo contesto attraverso tecniche non intrusive di ricostruzione del carico. Lo scopo è quello di identificare il profilo tipologico del carico del singolo utente, attraverso tecniche di *machine learning*, a partire dai dati mensili, facilmente reperibili dalle bollette elettriche, e di estrapolare un profilo orario secondo una metodologia *data-driven*, con l'obiettivo di stimare i principali parametri energetici che determinano la valorizzazione e l'incentivazione della condivisione dell'energia all'interno di Comunità Energetiche Rinnovabili.

Il processo di ricostruzione del profilo orario, a partire dai dati di consumo mensili, è composto principalmente da tre fasi: la prima consiste nell'individuazione dei principali *pattern* di consumo delle utenze residenziali, attraverso un *clustering* di tipo *k-Means*, preceduto da un ampio *pre-processing* dei dati reali grezzi; la seconda riguarda l'implementazione di *Random Forest*, ovvero un modello supervisionato di *machine learning* in grado di classificare ed individuare la tipologia di consumo del singolo utente sulla base dei consumi mensili; la terza riguarda la generazione del profilo orario attraverso un processo di *re-scaling* sulla base delle caratteristiche del consumo precedentemente individuato tramite il classificatore.

L'efficacia della metodologia proposta è valutata sulla base dell'errore relativo sul calcolo dell'energia condivisa, sia al livello aggregato che di singolo utente. I risultati ottenuti dimostrano che è possibile ricostruire il profilo orario di utenze residenziali a partire dai consumi mensili con un errore medio del 21%, al livello di singolo utenze, e del 0.56% al livello aggregato, ovvero al livello di comunità.

# Abstract

The recent developments in the European and Italian regulatory framework on energy have highlighted a strong push towards the theme of Renewable Energy Communities. Their diffusion in the territory represents an opportunity to accelerate the decarbonization process and the evolution of the national electrical system, moving from a centralized configuration that is difficult to manage to a segmented one that allows for easier network balancing. However, the establishment of Renewable Energy Communities is a complex process that requires multidisciplinary expertise ranging from the economic and energy sectors to the legal and administrative fields. The successful implementation of a project in the medium to long term is intricately linked to the correct balance between the production of the plants and the consumption of the users within the community, on an hourly basis.

At present, the electric load measurement infrastructure, especially in the residential sector, does not allow for hourly data detection, which does not allow for a precise assessment of their contribution in terms of energy sharing. This study aims to profile residential users who wish to be involved in this context with hourly resolution through non-intrusive load reconstruction techniques. The aim is to identify the typological load profile of each user through machine learning techniques, based on monthly data that can easily be obtained from electricity bills, and to extrapolate an hourly profile according to a data-driven methodology, with the objective of estimating the main energy parameters that determine the valorization and incentivization of energy sharing within Renewable Energy Communities.

The process of reconstructing the hourly profile based on monthly consumption data mainly consists of three phases: the first involves identifying the main consumption patterns of residential users through k-means clustering, preceded by extensive preprocessing of the raw data; the second concerns the implementation of Random Forest, a supervised machine learning model capable of classifying and identifying the typological consumption of each user based on monthly consumption; while the third involves generating the hourly profile through a rescaling process based on the characteristics of the previously identified typological consumption. The actual effectiveness of the proposed methodology is evaluated based on the distance between the generated hourly profile and the actual profile, as well as the relative error in calculating shared energy, both at the aggregated and individual user levels.

The effectiveness of the proposed methodology is assessed based on the relative error in the calculation of shared energy, both at the aggregate and individual user level. The results obtained demonstrate that it is possible to reconstruct the hourly profile of residential users starting from monthly consumption with an average error of 21% at the individual user level, and 0.56% at the aggregate level, i.e., at the community level.





# INDICE

<b>LISTA DELLE FIGURE .....</b>	<b>5</b>
<b>LISTA DELLE TABELLE .....</b>	<b>7</b>
<b>NOMENCLATURA.....</b>	<b>8</b>
<b>INTRODUZIONE .....</b>	<b>11</b>
<b>1. COMUNITÀ ENERGETICHE IN ITALIA ED EUROPA .....</b>	<b>14</b>
1.1 LA NORMATIVA EUROPEA SULLE COMUNITÀ ENERGETICHE .....	18
1.2 LA NORMATIVA ITALIANA SULLE COMUNITÀ ENERGETICHE .....	22
1.3 LE COMUNITÀ ENERGETICHE IN EUROPA E NEL MONDO.....	31
1.4 LE COMUNITÀ ENERGETICHE IN ITALIA.....	37
<b>2 ANALISI DATI E STATO DELL'ARTE .....</b>	<b>44</b>
2.1 ANALISI DATI .....	44
2.1.1 Pre-processing.....	45
2.2.1 Clustering gerarchico.....	53
2.2.2 K-means clustering .....	57
2.2.3 Latent Profile Analysis .....	59
2.2.4 Time Series Clustering (DTW) .....	60
2.3 OVIERVIEW SUI PRINCIPALI MODELLI DI CLASSIFICAZIONE .....	62
2.3.1 Alberi decisionali .....	64
2.3.3 Support Vector Machines.....	70
2.3.4 Naive Bayesian Classification.....	71
2.4 LITTERATURE REVIEW .....	72
<b>3 STRUTTURA DEL DATABASE UTILIZZATO .....</b>	<b>78</b>
<b>4 METODOLOGIA.....</b>	<b>82</b>
4.1 Pre-processing.....	85
4.2 Clustering dei profili medi mensili.....	90
4.3 Implementazione del modello di classificazione .....	93
4.4 Ricostruzione del profilo orario.....	99

<b>5 RISULTATI OTTENUTI .....</b>	<b>102</b>
5.1 PRE-PROCESSING.....	102
5.2 CLUSTERING .....	111
5.3 CLASSIFICAZIONE .....	113
5.4 RICOSTRUZIONE DEL PROFILO ORARIO .....	118
<b>6 CONCLUSIONI .....</b>	<b>124</b>
<b>APPENDICE A.....</b>	<b>126</b>
<b>BIBLIOGRAFIA.....</b>	<b>129</b>
<b>RINGRAZIAMENTI .....</b>	<b>135</b>

## LISTA DELLE FIGURE

---

<b>FIGURA 1</b> EMISSIONI DI GAS AD EFFETTO SERRA IN EUROPA PER TIPOLOGIA DI SETTORE [15].....	14
<b>FIGURA 2</b> SCHEMATIZZAZIONE STRUTTURA PACCHETTO "FIT FOR 55" [12]. .....	16
<b>FIGURA 3</b> SCHEMA DI UNA CONFIGURAZIONE DI 'AUTOCONSUMO COLLETTIVO. ....	19
<b>FIGURA 4</b> SCHEMA DI UNA COMUNITÀ ENERGETICA. ....	20
<b>FIGURA 5</b> CONFIGURAZIONE BASATA SU UN PIANO REGOLATORIO FISICO [18].....	24
<b>FIGURA 6</b> CONFIGURAZIONE BASATA SU UN PIANO REGOLATORIO VIRTUALE [18]. ....	24
<b>FIGURA 7</b> SVILUPPO DELLE CER E DELLE CEC IN EUROPA [25]. .....	32
<b>FIGURA 8</b> NUMERO DI CER NEL 2022 PER 9 PAESI EUROPEI [6]. ....	33
<b>FIGURA 9</b> MAPPATURA DELLE MIGLIORI ESPERIENZE SULLA CONDIVISIONE DELL'ENERGIA [19].....	40
<b>FIGURA 10</b> SOSTITUZIONE DI MISSING VALUES TRAMINE IL METODO GLOBAL CONSTANT [51]. ....	46
<b>FIGURA 11</b> SOSTITUZIONE DI MISSING VALUES TRAMITE IL METODO LOCAL CONSTANT [51]. ....	47
<b>FIGURA 12</b> SOSTITUZIONE DI MISSING VALUES TRAMITE MOVING AVERAGE SUBSTITUTION [51]. ....	48
<b>FIGURA 13</b> SOSTITUZIONE DI MISSING VALUES TRAMITE INTERPOLAZIONE LINEARE [51].....	48
<b>FIGURA 14</b> RAPPRESENTAZIONE GRAFICA DEL DECISION RANGE. ....	50
<b>FIGURA 15</b> RAPPRESENTAZIONE GRAFICA DEL PAA [51]. ....	51
<b>FIGURA 16</b> ESEMPIO DI UN DENDOGRAMMA [50].....	55
<b>FIGURA 17</b> ILLUSTRAZIONE GRAFICA DELL'ELBOW METHOD DOVE K = NUMERO DI CLUSTERS. ....	56
<b>FIGURA 18</b> ESEMPIO DI K-MEANS CLUSTERING. ....	58
<b>FIGURA 19</b> EFFETTO DELL'UTILIZZO DI DUE NUMERI DI CLUSTER DIFFERENT NEL K-MEANS CLUSTERING.....	59
<b>FIGURA 20</b> STRUTTURA DI UN ALBERO DECISIONALE.....	66
<b>FIGURA 21</b> DATI GREZZI DI UN DATASET GENERICO. ....	70
<b>FIGURA 22</b> ESEMPIO DI CLASSIFICAZIONE TRAMITE KNN. ....	70

<b>FIGURA 23</b> METODOLOGIA PER LA STIMA DEL PROFILO ORARIO. ....	84
<b>FIGURA 24</b> SCHEMA LOGICO DELE PRE-PROCESSING ADOTTATO. ....	85
<b>FIGURA 25</b> DIVISIONE DEL DATASER PER L'INDIVIDUAZIONE DEGLI OUTLIER GIORNALIERI. ....	87
<b>FIGURA 26</b> IDENTIFICAZIONE DEI CLUSTER INFREQUENTI E POCO RAPPRESENTATIVI.....	88
<b>FIGURA 27</b> IDENTIFICAZIONE PROFILI ANOMALI SU SCALA MENSILE. ....	91
<b>FIGURA 28</b> ELIMINAZIONE DEI PROFILI ANOMALI GIORNALIERI PER LE DOMENICHE AUTUNNALI. ....	104
<b>FIGURA 29</b> ELIMINAZIONE DEI PROFILI ANOMALI GIORNALIERI PER I MARTEDÌ PRIMAVERILI. ....	105
<b>FIGURA 30</b> PROFILI GIORNALIERI ANOMALI PER LA STAGIONE AUTUNNALE. ....	106
<b>FIGURA 31</b> PROFILI GIORNALIERI ANOMALI PER LA STAGIONE INVERNALE. ....	106
<b>FIGURA 32</b> PROFILI GIORNALIERI ANOMALI PER LA STAGIONE ESTIVA.....	107
<b>FIGURA 33</b> PROFILI GIORNALIERI ANOMALI PER LA STAGIONE PRIMAVERILE. ....	108
<b>FIGURA 34</b> RISULTATI DEL PRE-PROCESSING SVILUPPATO. ....	110
<b>FIGURA 35</b> VARIAZIONE DEL DBI INDEX AL VARIARE DEL NUMERO DI CLUSTER. ....	111
<b>FIGURA 36</b> PATTERN TIPOLOGICI DI CONSUMO MENSILE INDIVIDUATI.....	112
<b>FIGURA 37</b> IDENTIFICAZIONE DEI PATTERN DI CONSUMO ANOMALI DEI PROFILI MEDI MENSILI. ....	113
<b>FIGURA 38</b> IMPORTANZA DELLE VARIABILI DI CLASSIFICAZIONE UTILIZZATE.....	117
<b>FIGURA 39</b> CONFRONTO TRA PROFILO REALE E SIMULATO - 1.....	119
<b>FIGURA 40</b> CONFRONTO TRA PROFILO REALE E SIMULATO - 2.....	119
<b>FIGURA 41</b> CONFRONTO TRA PROFILO REALE E SIMULATO - 3.....	120
<b>FIGURA 42</b> CONFRONTO TRA PROFILO REALE E SIMULATO - 4.....	120
<b>FIGURA 43</b> CONSUMO AGGREGATO REALE E SIMULATO A CONFRONTO - 1 .....	121
<b>FIGURA 44</b> CONSUMO AGGREGATO REALE E SIMULATO A CONFRONTO - 2 .....	122
<b>FIGURA 45</b> CONSUMO AGGREGATO REALE E SIMULATO A CONFRONTO - 3 .....	122
<b>FIGURA 46</b> CONSUMO AGGREGATO REALEE E SIMULATO A CONFRONTO - 4 .....	123

# LISTA DELLE TABELLE

---

<b>TABELLA 1.</b> SINTESI DEGLI OBIETTIVI CLIMATICI DEL 2020 E 2030 PER EUROPA E ITALIA .....	17
<b>TABELLA 2</b> PRINCIPALI INIZIATIVE DI CER E AUTOCONSUMO COLLETTIVO [19]. .....	40
<b>TABELLA 3</b> UTILIZZO DEI DATI IN AMBITO ENERGETICO IN FUNZIONE DELLA SCALA TEMPORALE. ....	52
<b>TABELLA 4</b> PRINCIPALI CARATTERISTICHE DELLE METODOLOGIE DI CLUTERING DESCRITTE .....	61
<b>TABELLA 5</b> STRUTTURA DEL DATABASE NELLA SUA FORMA GREZZA.....	79
<b>TABELLA 6</b> POPOLAZIONE DEL DATABASE PER ANNO DI ATTIVITÀ. ....	80
<b>TABELLA 7</b> CARATTERISTICHE DEL DATASET RIFERITE AL 2012. ....	81
<b>TABELLA 8</b> FASCE DI CONSUMO DI LONDON POWER PER LE UTENZE RESIDENZIALI LONDINESI [42]. ....	82
<b>TABELLA 9</b> PARAMETRI AGGIUNTIVI, NECESSARI PER IL PROCESSO DI CLUSTERING DI TIPO K-MEANS. ....	89
<b>TABELLA 10</b> PARAMETRI DEL CLUSTERING DI TIPO K-MEANS SU SCALA MENSILE. ....	91
<b>TABELLA 11</b> VARIABILI UTILIZZATE PER ALLENARE IL MODELLO COMUNI A PISCITELLI ET AL. ....	94
<b>TABELLA 12</b> ULTERIORI VARIABILI UTILIZZATE NEL PROCESSO DI APPRENDIMENTO DEL MODELLO. ....	95
<b>TABELLA 13</b> CARATTERISTICHE DEL DATASET DI TRAIN E TEST. ....	96
<b>TABELLA 14</b> PARAMETRI AGGIUNTIVI NECESSARI PER LA DEFINIZIONE DEL MODELLO. ....	98
<b>TABELLA 15</b> VALORI MANCANTI E OUTLIERS PUNTUALI TROVATI.....	102
<b>TABELLA 16</b> DIVISIONE STAGIONALE DEL DATASET. ....	103
<b>TABELLA 17</b> RISULTATI DEL PROCESSO DI IDENTIFICAZIONE DEGLI OUTLIERS SU SCALA GIORNALIERA. ....	109
<b>TABELLA 18</b> VARIABILI UTILIZZATE PER ALLENARE IL CLASSIFICATORE. ....	114
<b>TABELLA 19</b> CARATTERISTICHE DEL RANDOM FOREST MIGLIORE OTTENUTO DAL “TUNING”. ....	115
<b>TABELLA 20</b> PRECISIONE DEL MODELLO NEL PROCESSO DI CLASSIFICAZIONE. ....	116
<b>TABELLA 21</b> MATRICE DI CONFUSIONE DEL TESTING SET.....	116
<b>TABELLA 23</b> ERRORE RELATIVO AL LIVELLO AGGREGATO PER L'E.C.....	123

## NOMENCLATURA

---

<b>CE</b>	Comunità Energetiche
<b>CER</b>	Comunità Energetiche Rinnovabili
<b>REC</b>	Renewable Energy Communities
<b>ENEA</b>	Energia Nucleare Energie Alternative
<b>COP</b>	Conferenza delle Parti
<b>UNFCCC</b>	Convegno delle nazioni unite sui Cambiamenti Climatici
<b>UE</b>	Unione Europea
<b>CEP</b>	Clean Energy Package
<b>RED</b>	Renewable Energy Directive
<b>ETS</b>	Enti del Terzo Settore
<b>ESCo</b>	Energy Service Company
<b>CEC</b>	Comunità Energetiche dei Cittadini
<b>PMI</b>	Piccole e Medie Imprese
<b>MT</b>	Media Tensione
<b>BT</b>	Bassa Tensione
<b>MiSE</b>	Ministero dello Sviluppo Economico
<b>GSE</b>	Gestore dei Servizi Energetici
<b>ARERA</b>	Autorità di Regolazione per Energia Reti e Ambiente
<b>SMA</b>	Simple Moving Average
<b>PAA</b>	Piecewise Aggregate Approximation
<b>DTW</b>	Dynamic Time Warping Time Series Clustering
<b>LPA</b>	Latent Profile Analysis

<b>SSE</b>	Sum of Square Errors
<b>KNN</b>	K-Nearest Neighbors
<b>NBC</b>	Naive Bayesian Classification
<b>CEC</b>	Comunità Energetiche dei Cittadini
<b>DER</b>	Distributed Energy Resources
<b>CAM</b>	Customer Adoption Modedistribute
<b>SOM</b>	Self-Organizing Map
<b>DBI</b>	Davies-Bouldin Index
<b>WSSSE</b>	Within Set Sum of Square Errors
<b>MIA</b>	Mean Adequancy Index
<b>CDI</b>	Clustering Disperios Indicator
<b>LCL</b>	Low Carbon London
<b>EC</b>	Energia Condivisa
<b>dToU</b>	Dynamic Time of Use





# Introduzione

Il raggiungimento della neutralità climatica entro il 2050 rappresenta uno dei principali obiettivi delle politiche europee. A tale scopo, l'Unione Europea ha presentato nel dicembre 2019 il Green Deal, una nuova strategia a lungo termine finalizzata a creare un'economia moderna, competitiva e climaticamente neutra. Il documento enfatizza l'importanza delle fonti rinnovabili di energia e sottolinea il ruolo fondamentale dei consumatori, che diventano sempre più attivi e rilevanti nella transizione ecologica. Secondo stime di Energia Nucleare ed Energie Alternative (ENEA), entro il 2050 circa 264 milioni di cittadini dell'Unione Europea si uniranno al mercato dell'energia come *prosumer*, generando fino al 45% dell'energia elettrica rinnovabile complessiva del sistema.

In questo contesto, le Comunità Energetiche Rinnovabili (CER) rappresentano un'importante novità introdotta in Italia con il recepimento della direttiva europea REDII (Direttiva UE 2018/2001). Le CER sono organizzazioni locali che riuniscono cittadini, imprese, istituzioni e altri enti locali per sviluppare progetti di produzione, condivisione e utilizzo locale dell'energia rinnovabile. Esse si basano fondamentalmente su tecnologie come i pannelli solari, le turbine eoliche e le centrali idroelettriche, e possono includere la produzione di energia elettrica, il riscaldamento e il raffreddamento degli edifici e la ricarica dei veicoli elettrici. Le CER si presentano come potenziali catalizzatori del processi di transizione energetica, e si basano sulla partecipazione e la condivisione locale delle risorse energetiche.

Gli obiettivi principali delle CER sono di promuovere l'efficienza energetica, l'uso di fonti rinnovabili e la riduzione delle emissioni di gas serra, ma anche di favorire la creazione di comunità locali più sostenibili. Inoltre, le CER consentono ai cittadini e agli enti attivi di partecipare direttamente al mercato energetico come *prosumers*, promuovendo l'autoconsumo e diventando produttori e proprietari di impianti alimentati da fonti rinnovabili. In questo modo, esse generano valore economico e sociale sul territorio e rappresentano un potenziale di innovazione per la decentralizzazione del mercato energetico e per la necessaria transizione ecologica.

Le CER, quindi, rappresentano un nuovo soggetto per i mercati energetici che può svolgere diverse attività finalizzate a fornire benefici a livello di comunità ai suoi azionisti o membri e sono un esempio concreto di come la transizione energetica possa diventare una realtà partecipativa e inclusiva, in cui i cittadini diventano protagonisti attivi nella gestione del proprio territorio e del proprio futuro energetico.

In letteratura sono numerosi gli studi riguardanti il dimensionamento ottimizzato delle CER [8, 9, 14]. Tuttavia, molti di questi studi si focalizzano sulla ricerca di una configurazione ottimizzata dal punto di vista tecnico, tralasciando aspetti importanti come la stabilità a lungo termine dell'aggregato, la corretta distribuzione dei benefici economici generati dalla CER, le possibili barriere che ostacolano l'ingresso in comunità e le problematiche che derivano dall'uscita dell'aggregato da parte degli utenti nel corso del tempo [3]. In aggiunta, nella maggior parte dei casi, gli studi si basano su dati di consumo reali con risoluzione oraria o sub-oraria. L'utilizzo di dati con un'elevata risoluzione temporale è fondamentale per il corretto bilanciamento tra produzione e consumo, aspetto cardine per la buona riuscita di un progetto di CER, sia dal punto di vista energetico che economico.

Tuttavia, dati con risoluzione oraria o sub-oraria, in un contesto reale, non sono spesso disponibili, soprattutto nell'ambito delle utenze residenziali e in paesi, come l'Italia, dove non è ancora stato implementato un sistema di misurazione di nuova generazione. La scarsa reperibilità dei dati orari non permette, in prima istanza, il corretto bilanciamento tra produzione e consumo al livello orario e spesso si è costretti a delle approssimazioni che generalmente introducono un errore notevole nel calcolo dei principali parametri energetici ed economici.

Il presente lavoro di tesi ha come obiettivo la presentazione di una metodologia volta a stimare il profilo orario di utenze residenziali a partire dai consumi mensili in fasce, facilmente reperibili dalle bollette elettriche, attraverso tecniche non intrusive. La metodologia proposta si propone di colmare le lacune esistenti nella disponibilità di dati orari nell'ambito residenziale, una problematica che affligge inevitabilmente anche il settore delle Comunità

Energetiche Rinnovabili, dal momento che la valutazione dei principali parametri energetici che generano valore economico, come l'energia condivisa, è condotta a livello orario.

La capacità di stimare con precisione il consumo orario di un utente sulla base dei soli dati di consumo mensile, attraverso tecniche non intrusive, consente di effettuare un calcolo più accurato dell'energia condivisa e di individuare, in fase di simulazione preliminare, la configurazione ottimale che garantisce il bilanciamento opportuno tra produzione e consumo, assicurando la stabilità economica della CER. La metodologia proposta previene la realizzazione di progetti di CER sulla carta economicamente sostenibili, ma instabili nel breve e medio periodo, a causa di una valutazione imprecisa dei consumi dei propri membri e, di conseguenza, dell'energia condivisa generata, nonché di parte degli introiti economici.

Nel Capitolo 1 si ripercorre l'iter normativo che ha visto nascere le CER in Italia e nel mondo, attraverso un rapido excursus delle principali tappe legislative. Inoltre si effettua una breve analisi descrittiva della struttura delle CER e della loro diffusione nel territorio italiano ed europeo, riportando alcune esperienze di Comunità Energetiche già ben consolidate.

Nel Capitolo 2 si descrive l'utilità dell'analisi dati in contesti energetici e si descrivono le principali tecniche di *clustering* e di classificazione supervisionata, delle quali si farà ampio utilizzo nei capitoli successivi. Nel Capitolo 3 si descrive la struttura del *database* utilizzato e le sue principali caratteristiche, definendo la quantità e la qualità dei dati utilizzati.

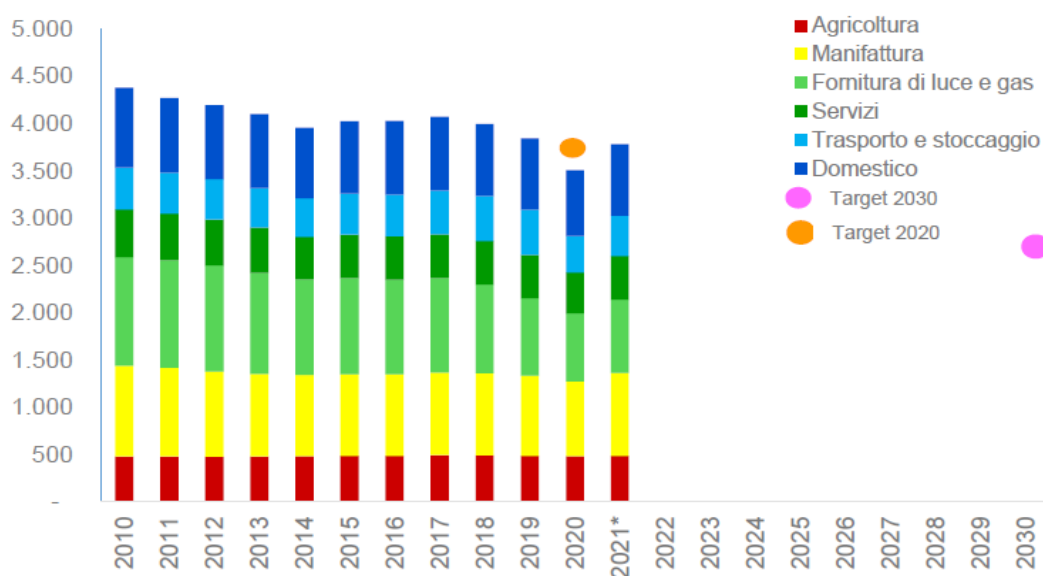
Successivamente, nei Capitoli 4, 5 e 6 viene descritta nel dettaglio la metodologia proposta e le tecniche utilizzate, giustificando le scelte effettuate e esponendo i principali risultati ottenuti, in termini di efficacia della metodologia implementata, flessibilità e scalabilità del modello.

# 1. Comunità Energetiche in Italia ed Europa

Nel dicembre 2015, i governi, dopo un processo di trattative che è durato circa due decenni, hanno adottato il primo accordo universale per contrastare i cambiamenti climatici. L'adozione dell'accordo è avvenuta in occasione della ventunesima conferenza delle parti (COP 21) della Convenzione quadro delle Nazioni Unite sui cambiamenti climatici (UNFCCC) tenutasi a Parigi. L'accordo di Parigi [24] ha come obiettivo principale quello di mantenere l'aumento della temperatura media globale ben al di sotto dei 2°C e, se possibile, a 1,5°C rispetto ai livelli preindustriali. A tal fine, le parti si sono impegnate a stabilizzare quanto prima possibile le emissioni di gas serra a livello mondiale e a raggiungere l'azzeramento delle emissioni nette nella seconda metà del secolo.

Nel 2019, l'Unione Europea (UE) aveva già ridotto le sue emissioni di gas serra del 24% rispetto ai livelli del 1990, dimostrando di essere sulla buona strada per raggiungere l'obiettivo stabilito nel protocollo di Kyoto di ridurre le emissioni del 20% entro il 2020. Nonostante la ripresa post Covid delle emissioni targata 2021 [15], la tendenza a lungo termine delle emissioni di gas a effetto serra dell'UE mostra una graduale riduzione, come rappresentato in **Figura 1**.

## Andamento emissioni gas ad effetto serra in Europa [Mt CO<sub>2</sub>]



**Figura 1** Emissioni di gas ad effetto serra in Europa per tipologia di settore [15].

Nel dicembre del 2019, la Commissione Europea ha presentato il *Green Deal* [4], un pacchetto di misure volte a stabilire obiettivi più ambiziosi per la riduzione delle emissioni di gas serra entro il 2030, e per la de-carbonizzazione dell'economia dell'UE entro il 2050, in linea con gli accordi stabiliti dall'Accordo di Parigi. Tale iniziativa rappresenta un supporto alla trasformazione dell'UE in una società equa e prospera, con un'economia moderna e competitiva. Il *Green Deal* sottolinea la necessità di un approccio olistico e intersettoriale, in cui tutti i settori strategici pertinenti contribuiscano all'obiettivo finale in termini di clima [13]. Le misure proposte variano dai tagli ambiziosi alle emissioni, agli investimenti in attività di ricerca e innovazione all'avanguardia, con lo scopo di preservare l'ambiente naturale dell'Europa.

Il *Green Deal* è sostenuto da investimenti in tecnologie verdi, soluzioni sostenibili e nuove imprese, ed è concepito come una nuova strategia di crescita in grado di trasformare l'UE in un'economia sostenibile e competitiva. Tra le iniziative incluse nel *Green Deal*, al fine di favorire un percorso equilibrato verso la neutralità dell'UE in termini di emissione di CO<sub>2</sub> entro il 2050, nell'aprile 2021 la Commissione ha convenuto di innalzare dal 40% al 55% il precedente obiettivo di riduzione delle emissioni di gas serra entro il 2030 attraverso l'accordo europeo *Fit for 55* [12]. Tale pacchetto consiste in una serie di proposte volte a rivedere la legislazione in materia di clima, energia e trasporti e a mettere in atto nuove iniziative legislative per allineare la legislazione dell'UE rispetto ai propri obiettivi climatici (**Figura 2**). Esso comprende tra le varie proposte:

- una revisione della direttiva sulla promozione delle energie rinnovabili;
- una rifusione della direttiva sull'efficienza energetica;
- una revisione della direttiva sulla tassazione dei prodotti energetici;
- una revisione della direttiva sulla realizzazione di un'infrastruttura per combustibili alternativi;
- un fondo sociale per il clima;
- una revisione della direttiva sulla presentazione energetica nell'edilizia;

- la riduzione delle emissioni di metano nel settore dell'energia;
- una revisione del terzo pacchetto "Energia" sul gas.

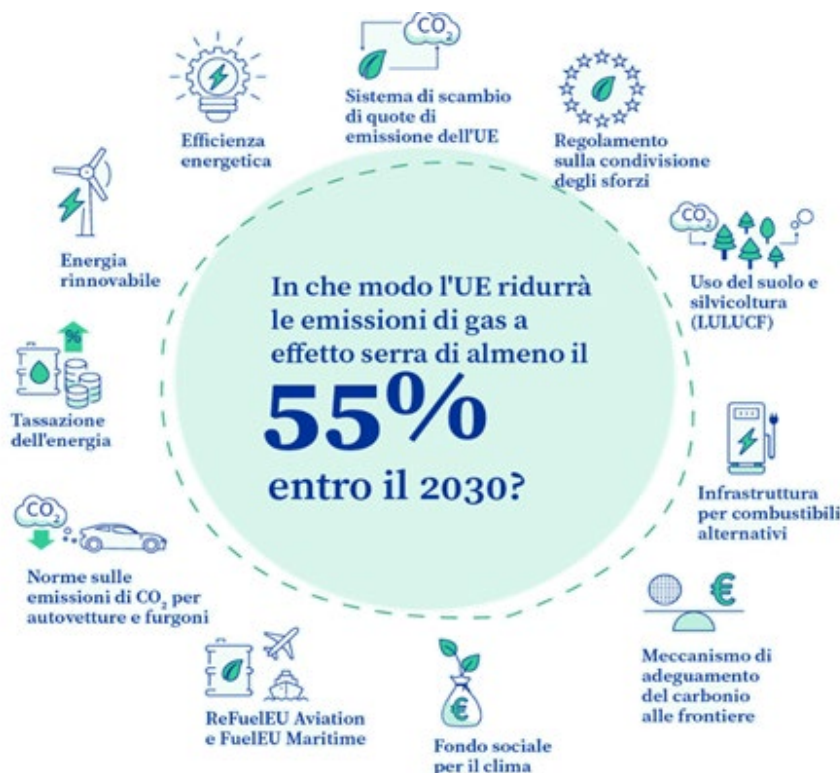


Figura 2 Schematizzazione struttura pacchetto "Fit for 55" [12].

In prospettiva del raggiungimento degli obiettivi fissati per il 2030, l'Unione Europea fornisce un solido quadro normativo in materia di efficienza energetica e utilizzo di fonti rinnovabili tramite il *Clean Energy for all Europeans Package* (CEP), che comprende diverse misure legislative in cinque fondamentali settori della politica europea: sicurezza nell'approvvigionamento energetico, mercato interno dell'energia, efficienza energetica, decarbonizzazione dell'economia, ricerca, innovazione e competitività [25, 29]. Le misure in esso contenute non riguardano solo le grandi imprese, ma coinvolgono anche i cittadini e le loro attività.

Secondo il CEP, tramite una migliore efficienza del mercato e il rafforzamento dei diritti dei consumatori, i cittadini acquisiscono un reale potere decisionale sulla propria impronta energetica, sia attraverso l'impiego di *smart meter* di seconda generazione, che attraverso l'investimento nella produzione autonoma tramite fonti di energia rinnovabile. All'interno di questo pacchetto si

inseriscono la direttiva sulla promozione dell'uso di energia da fonti rinnovabili (REDII) e la direttiva sulle regole comuni per il mercato interno dell'energia elettrica (IEM), che definiscono nuove modalità di partecipazione dei cittadini alla transizione energetica, in particolare le Comunità Energetiche.

Sulla base del CEP sono state emanate le direttive europee vigenti e sono stati redatti i Piani di Azione Nazionale per l'Energia e il Clima che definiscono gli obiettivi da raggiungere in Europa e in Italia entro il 2030, schematizzati in **Tabella 1**.

**Tabella 1.** Sintesi degli obiettivi climatici del 2020 e 2030 per Europa e Italia

	OBIETTIVI 2020		OBIETTIVI 2030	
	EU	ITA	EU	ITA
<b>ENERGIE RINNOVABILI</b>	.	.	.	.
<i>Quota Rinnovabile dei consumi finali lordi</i>	20%	17%	32%	30%
<i>Quota Rinnovabile dei consumi finali lordi dei trasporti</i>	10%	10%	14%	22%
<i>Quota Rinnovabile dei consumi finali lordi termici</i>			+1,3% anno	+1,3% anno
<b>EFFICIENZA ENERGETICA</b>	.	.	.	..
<i>Riduzione consumi primari rispetto allo scenario</i>	-20%	-24%	-32,50%	-43%
<i>Riduzione consumi finali da politiche attive</i>	-1,5% anno	-1,5% anno	-0,8% anno	-0,8% anno
<b>EMISSIONI DI GAS SERRA</b>	.	.	.	.
<i>Riduzione GHG (2005) nei settori ETS</i>	-21%		-43%	
<i>Riduzione GHG (2005) nei settori non ETS</i>	-10%	-13%	-30%	-33%
<i>Riduzione GHG totali (1990)</i>	-20%		-40%	

Entro il prossimo decennio, l'Unione Europea mira a raggiungere una quota del 32% di fonti rinnovabili nei consumi finali lordi, mentre l'Italia ha stabilito un obiettivo del 30%, con particolare attenzione al settore dei trasporti, dove l'obiettivo di quota di fonti rinnovabili è del 22%, superiore a quello dell'Unione Europea. Gli obiettivi di efficienza energetica prevedono una riduzione dei consumi primari dell'energia del 32,5% per l'UE e del 43% per l'Italia, con una riduzione dei consumi dovuta a politiche attive pari allo 0,8% annuo. Per quanto riguarda le emissioni di gas serra, si prevede una riduzione del 43% nel settore ETS a livello europeo e del 33% nei settori non-ETS in Italia entro il 2030.

Questo capitolo fornisce una panoramica dell'evoluzione del quadro normativo sulle Comunità Energetiche in Europa e in Italia, dalle prime fasi fino ai più recenti sviluppi legislativi. Inoltre, vengono presentati e descritti i principali progetti a livello europeo e nazionale che hanno avviato il processo di diffusione delle Comunità Energetiche.

## **1.1 La normativa europea sulle Comunità Energetiche**

Nel panorama normativo europeo, rivestono particolare rilevanza due direttive che attribuiscono centralità agli utenti finali nel processo di transizione energetica [6]. La prima è la Direttiva (UE) 2001/2018 del Parlamento europeo e del Consiglio, emanata l'11 dicembre 2018, che promuove l'utilizzo di fonti energetiche rinnovabili (REDII). La seconda è la Direttiva (UE) 944/2019 del Parlamento europeo e del Consiglio, datata 5 giugno 2019, che stabilisce norme comuni per il mercato interno dell'energia elettrica (IEM).

La RED II nasce con lo scopo di sviluppare dell'energia prodotta da fonti rinnovabili sul territorio europeo, favorendo la partecipazione attiva dei cittadini e, più in generale, dei clienti finali. Essa introduce dei modelli di partecipazione a complessità crescente, definendo e normando l'autoconsumo singolo, l'autoconsumo collettivo e le Comunità Energetiche Rinnovabili con lo scopo di aumentare la partecipazione nei processi di produzione e consumo di energia.

Essa definisce un autoconsumatore di energia rinnovabile come un cliente finale che:

- operando in propri siti situati entro confini definiti o, se consentito da uno Stato membro, in altri siti, produce energia elettrica rinnovabile per il proprio consumo;
- può immagazzinare o vendere energia elettrica rinnovabile per il proprio consumo e può immagazzinare o vendere energia elettrica rinnovabile purché, per un autoconsumatore di energia rinnovabile diverso dai nuclei familiari, tali attività non costituiscano l'attività commerciale o professionale principale.



Definisce inoltre un autoconsumatore di energia rinnovabile che agisce collettivamente come:

- un gruppo di almeno due autoconsumatori di energia rinnovabile che agiscono collettivamente e si trovano nello stesso edificio o condominio.

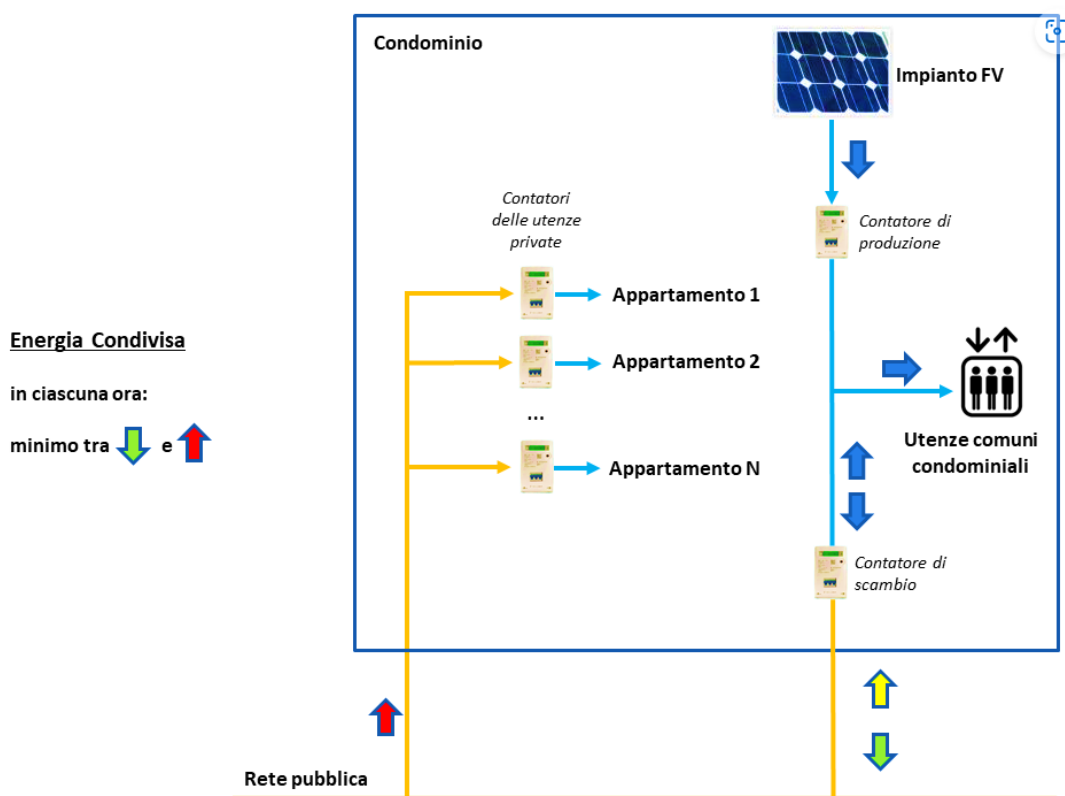


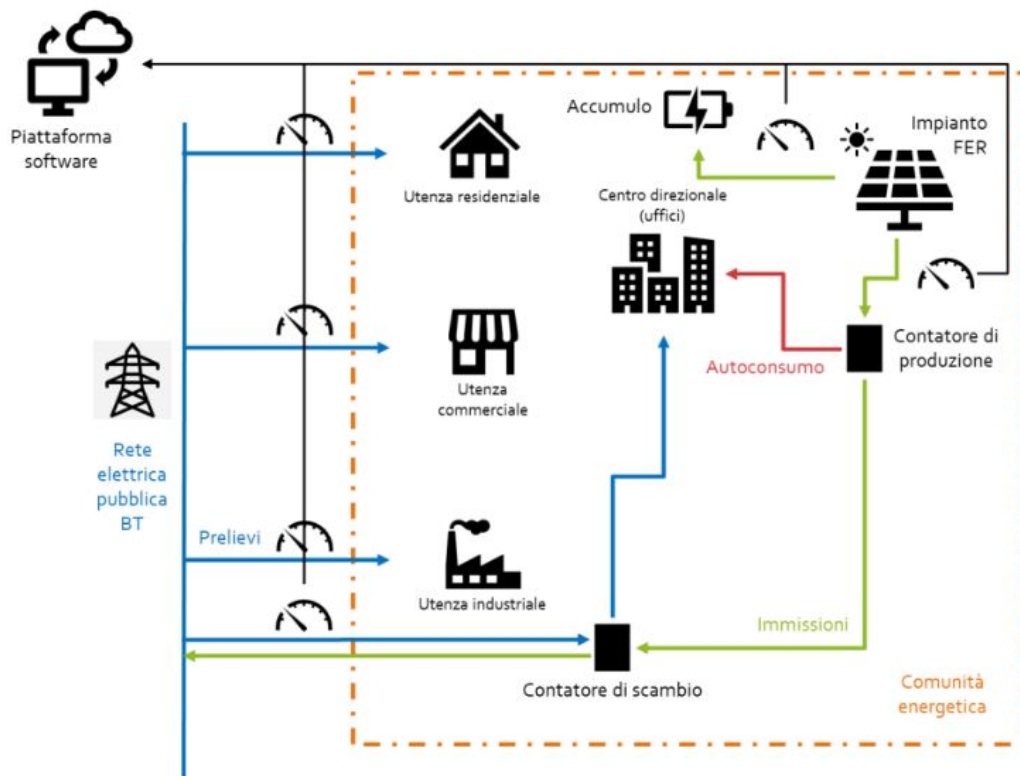
Figura 3 Schema di una configurazione di 'autoconsumo collettivo'.

La RED II introduce poi le Comunità Energetiche definendole come un soggetto giuridico che:

- si basa sulla partecipazione aperta e volontaria;
- è autonomo ed effettivamente controllato da azionisti o dai membri che sono situati nelle vicinanze degli impianti di produzione di energia da fonti rinnovabili che appartengono e sono sviluppati dal soggetto giuridico in questione;
- i cui azionisti o membri sono persone fisiche, PMI o autorità locali,

comprese le amministrazioni comunali;

- il cui obiettivo principale è fornire benefici ambientali, economici o sociali a livello di comunità ai suoi azionisti o membri o alle aree locali in cui opera, piuttosto che generare profitti finanziari.



**Figura 4** Schema di una comunità energetica.

Le CER rappresentano quindi un soggetto nuovo per i mercati energetici, più attento alle caratteristiche locali e con una finalità diversa da quella del lucro. Sono enti basati su una governance democratica dove le decisioni sono prese dai membri della comunità in modo indipendente e autonomo, che possono collaborare con altri attori del mercato senza che questi debbano diventare necessariamente membri delle comunità. Per garantire il carattere no profit delle Comunità Energetiche non è ammessa la partecipazione, in qualità di membri della comunità, alle aziende del settore energetico (fornitori ed ESCO) che possono, invece, prestare servizi di forniture ed infrastruttura.

È evidente che tanto maggiore è la complessità delle configurazioni assunte e dei servizi erogati, tanto più grandi e diversificate saranno le competenze in

campo energetico che dovranno essere introdotte dai soci o membri [6]. Per questa ragione, la RED II evidenzia inoltre come sia necessaria una collaborazione tra i cittadini e i soggetti che già oggi operano in campo energetico, come ad esempio le ESCo, i distributori locali o gli altri operatori industriali. Tali osservazioni sono condivise anche nel panorama scientifico da numerosi studi di settore [3, 8, 14].

Analogamente a quanto riportato nella direttiva RED II, anche la IEM introduce alcune figure a complessità crescente per favorire la partecipazione degli utenti finali e non, al mercato dell'energia elettrica. La direttiva definisce il cliente attivo, o del cliente attivo che opera in modo collettivo come:

- un cliente finale o un gruppo di clienti finali consorziati che consuma o conserva l'energia elettrica prodotta nei propri locali situati all'interno di un'area delimitata o, se consentito da uno Stato membro, in altri locali, oppure vende l'energia elettrica autoprodotta o partecipa a meccanismi di flessibilità o efficienza energetica, purché tali attività non costituiscano la principale attività commerciale o professionale.

La direttiva definisce inoltre le Comunità Energetiche dei Cittadini (CEC) come un soggetto giuridico che:

- è fondato sulla partecipazione volontaria e aperta;
- è effettivamente controllato da membri o soci che sono persone fisiche, autorità locali comprese le amministrazioni comunali o PMI;
- ha lo scopo principale di offrire ai suoi membri o soci o al territorio in cui opera benefici ambientali, economici o sociali a livello di comunità anziché generare profitti finanziari;
- può partecipare alla generazione, anche da fonti rinnovabili, alla distribuzione, al consumo, all'aggregazione, allo stoccaggio dell'energia, ai servizi di efficientamento energetico o a servizi di ricarica per veicoli elettrici, o fornire altri servizi energetici ai suoi membri o soci.

Rispetto alle CER, vengono meglio dettagliati gli ambiti in cui le CEC possono operare, pur mantenendo il fine principale di rispondere alle esigenze dei propri membri. Inoltre, le CEC non prevedono i principi di autonomia e prossimità caratteristici delle CER e può solo gestire l'elettricità, che sia essa prodotta da fonte rinnovabile o fossile.

## **1.2 La normativa italiana sulle Comunità Energetiche**

Nel periodo compreso tra la fine del 2019 e l'inizio del 2020, in Italia è stato avviato un processo di recepimento parziale della direttiva RED II attraverso la Legge 8/2020, con l'obiettivo di sperimentare gli effetti, le ricadute e le possibili criticità connesse all'introduzione delle Comunità Energetiche (CE) e delle configurazioni di autoconsumo collettivo nel contesto nazionale.

Con tale legge sono state definite le modalità e le condizioni per l'attivazione dell'autoconsumo collettivo da fonti rinnovabili e la realizzazione di comunità di energia rinnovabile. Ciò ha avviato la sperimentazione di un quadro regolatorio che mira a consentire agli utenti finali e ai produttori di energia di associarsi al fine di condividere l'energia elettrica prodotta localmente da nuovi impianti alimentati da fonti rinnovabili.

In questa fase sperimentale, sono stati introdotti alcuni vincoli tecnici relativi agli impianti di produzione e al perimetro di operazione delle CER:

- gli impianti detenuti all'interno delle configurazioni di autoconsumo collettivo e delle CER devono essere entrati in esercizio dopo il 1 marzo 2020 ed entro i 60 giorni successivi alla data di entrata in vigore del decreto di recepimento della direttiva UE 2018/2001;
- la massima potenza installabile riferita al singolo impianto non può essere superiore a 200 kW;
- gli impianti e i consumatori, che siano essi membri, soci o soggetti

partecipanti alla configurazione, devono sottostare alla medesima cabina di trasformazione MT/BT nel caso delle CER, e devono afferire al medesimo edificio nel caso di schemi di autoconsumo collettivo.

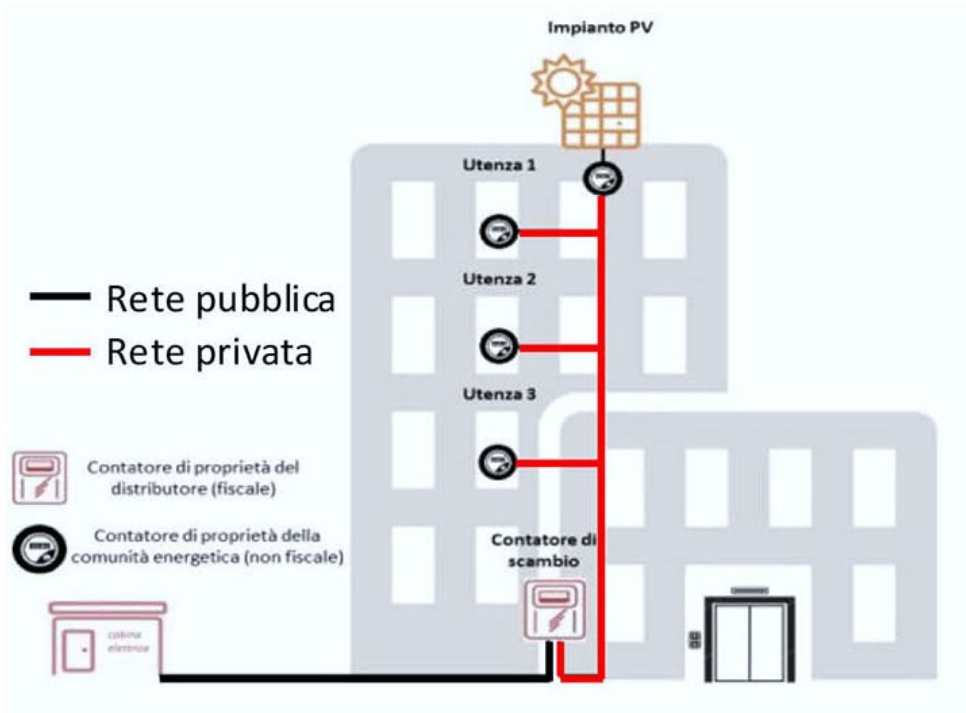
Così come esplicitamente espresso nella direttiva, in questa fase sperimentale è stata garantita la partecipazione a enti commerciali e industriali che si occupano di produzione e gestione dell'energia da fonti rinnovabili in qualità di soggetti terzi. Nel corso del 2020, ARERA e il Ministero per lo Sviluppo Economico (MiSE), hanno rispettivamente definito:

- il modello regolatorio e le componenti tariffarie applicabili ai membri delle CER e ai partecipanti agli schemi di autoconsumo collettivo;
- gli incentivi riconoscibili ad entrambe le configurazioni.

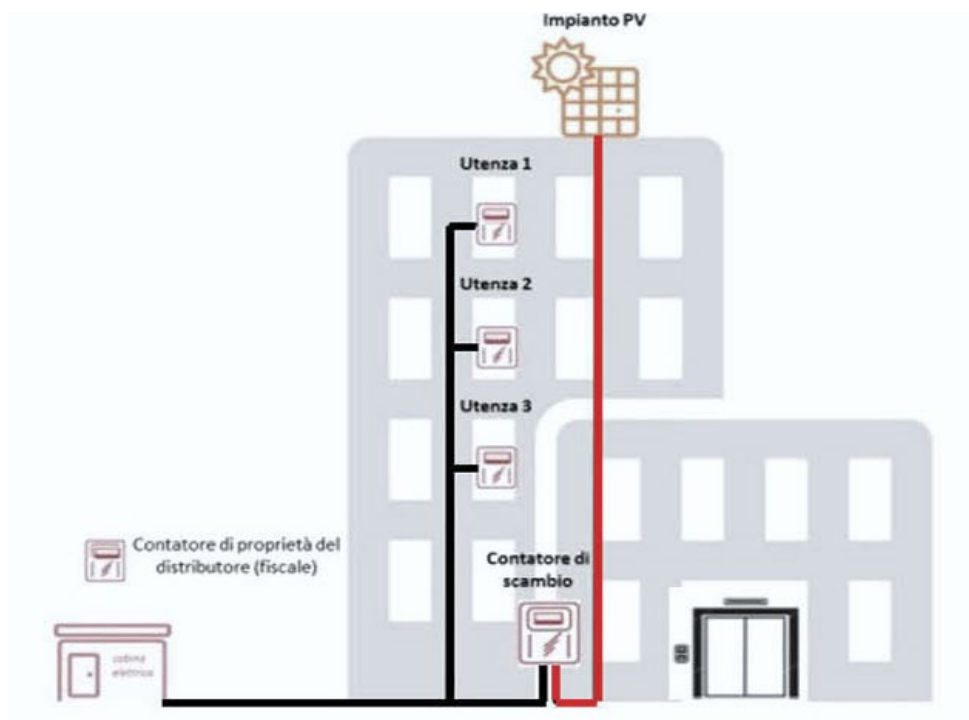
In materia di regolazione ARERA tramite la Delibera 318/2020 ha deciso di ricorrere ad un modello di tipo virtuale per la gestione delle partite economiche.

Tale modello permette di:

- evitare la realizzazione di nuove reti private generalmente molto costose;
- mantenere separata l'evidenza dei benefici associati all'autoconsumo (che non dipendono da fonti, tipologia di reti e/o assetti societari) e agli incentivi espliciti (che, in quanto tali, possono essere opportunamente calibrati in funzione delle fonti e/o delle tecnologie).



**Figura 5** Configurazione basata su un piano regolatorio fisico [18].



**Figura 6** Configurazione basata su un piano regolatorio virtuale [18].

La principale differenza tra una configurazione virtuale e una fisica (**Figura 5 e 6**) sta nel fatto che nella prima viene coinvolta direttamente la rete elettrica pubblica mentre nella seconda viene utilizzata principalmente quella privata e, là dove non ve ne fosse una adeguata, è necessario realizzarne una nuova con delle

caratteristiche conformi alla configurazione scelta.

Il modello regolatorio virtuale prevede che il Gestore dei Servizi Energetici (GSE) eroghi il “servizio di valorizzazione e incentivazione dell’energia condivisa” per tramite del Referente delle configurazioni. Viene riconosciuta alle CER la restituzione di alcune componenti di rete che ammontano a circa 8 €/MWh rispetto all’energia prodotta dagli impianti all’interno delle configurazioni e consumata nel medesimo arco orario dai propri membri. Tale restituzione può salire fino a circa 10 €/MWh per gli schemi di autoconsumo collettivo. Alle componenti di regolazione va aggiunta la componente incentivante di tipo “feed-in premium” definita dal MiSE attraverso il Decreto Ministeriale del 16 settembre 2020 che corrisponde a:

- 100 €/MWh per l’energia condivisa dai partecipanti agli schemi di autoconsumo collettivo;
- 110 €/MWh per l’energia condivisa dai membri delle CER.

Per la corretta gestione del meccanismo regolatorio virtuale precedentemente descritto, il GSE è chiamato ad espletare i seguenti adempimenti:

- predisporre e trasmettere, per verifica positiva da parte del Direttore della Direzione Mercati Energia all’Ingrosso e Sostenibilità Ambientale dell’ARERA, lo schema di istanza, lo schema di contratto e le Regole tecniche, contenenti tra le altre cose, i criteri puntuali di calcolo eventualmente necessari, le modalità di comunicazione al Referente delle configurazioni che beneficiano del servizio di valorizzazione ed incentivazione dell’energia condivisa e le modalità di profilazione dei dati di misura con le annesse modalità di utilizzo;
- fornire servizi di assistenza territoriale alle Pubbliche Amministrazioni;

- predisporre un apposito portale informatico interoperabile con il sistema GAUDI, ai fini dell'accesso al servizio di valorizzazione e incentivazione dell'energia elettrica condivisa, nonché ai fini della gestione tecnica ed economica del medesimo servizio.

Nel 22 dicembre 2020 sono state pubblicate sul sito del GSE le Regole tecniche per l'accreditamento alle CER e agli schemi di autoconsumo collettivo [17], aggiornate poi ad aprile 2021 a seguito dell'evoluzione del quadro normativo e regolatorio di riferimento. I primi mesi del 2021 sono stati caratterizzati dall'attivazione e dalla candidatura delle prime comunità energetiche e dei primi schemi di autoconsumo collettivo [6]. Conformemente agli obiettivi stabiliti per la fase sperimentale, tali iniziative hanno permesso di identificare gli elementi critici da modificare in vista del recepimento complessivo della RED II e della IEM. A partire dal mese di agosto 2021 sono circolate le prime bozze di recepimento definitivo delle due direttive europee, recepite poi dai decreti legislativi 199 (REDII) e 210 (IEM) nel novembre 2021.

### **DLGS 199/2021 (REDII)**

#### *Partecipazione e controllo:*

- La CER è un soggetto di diritto autonomo e l'esercizio dei poteri di controllo fa capo esclusivamente a persone fisiche, piccole e medie imprese (no attività commerciale/industriale principale), enti territoriali e autorità locali, ivi incluse le amministrazioni comunali, gli enti di ricerca e formazione, del terzo settore e di protezione ambientale nonché le amministrazioni locali, che sono situate nel territorio degli stessi Comuni in cui sono ubicati gli impianti per la condivisione di energia;
- la partecipazione alle CER è aperta a tutti i consumatori, compresi quelli appartenenti a famiglie a basso reddito o vulnerabili, fermo restando che l'esercizio dei poteri di controllo è detenuto dai soggetti aventi le caratteristiche di cui sopra.

#### *Obiettivi:*



- La CER ha l'obiettivo principale di fornire benefici ambientali, economici o sociali a livello di comunità ai suoi soci o membri o alle aree locali in cui opera la comunità e non quello di realizzare profitti finanziari.

*Ambiti di attività:*

- La CER può produrre altre forme di energia da fonti rinnovabili finalizzate all'utilizzo da parte dei membri, può promuovere interventi integrati di domotica ed efficienza energetica, nonché offrire servizi di ricarica dei veicoli elettrici ai propri membri e assumere un ruolo di società di vendita al dettaglio e può offrire servizi ancillari e di flessibilità.

*Condivisione dell'energia:*

- In una comunità di energia rinnovabile o in un gruppo di autoconsumatori di energia rinnovabile che agiscono collettivamente, l'energia condivisa è pari al minimo, in ciascun periodo orario, tra l'energia elettrica prodotta e immessa in rete dagli impianti a fonti rinnovabili e l'energia elettrica prelevata dall'insieme dei clienti finali associati situati nella stessa zona di mercato;
- ai fini del calcolo dell'energia condivisa rileva solo la produzione di energia rinnovabile degli impianti che risultano nella disponibilità e sotto il controllo della comunità.

*Ricorso all'energia prodotta da impianti esistenti:*

- Gli impianti a fonti rinnovabili per la produzione di energia elettrica realizzati dalla comunità sono entrati in esercizio dopo la data di entrata in vigore del presente decreto legislativo, fermo restando la possibilità di adesione per impianti esistenti, sempre di produzione di energia elettrica rinnovabile, per una misura comunque non superiore al 30% della potenza complessiva che fa capo alla comunità.

*Regolazione:*

- Nei casi in cui gli impianti di produzione e i punti di prelievo siano connessi alla porzione di rete di distribuzione sottesa alla stessa cabina

primaria, ARERA individua, anche in via forfettaria, il valore delle componenti tariffarie disciplinate in via regolatoria, nonché di quelle connesse al costo della materia prima energia, che non risultano tecnicamente applicabili all'energia condivisa, in quanto energia istantaneamente autoconsumata sulla porzione di rete.

*Incentivi:*

- Possono accedere all'incentivo gli impianti a fonti rinnovabili di potenza non superiore a 1 MW che entrano in esercizio in data successiva a quella di entrata in vigore del presente decreto;
- è erogato solo in riferimento alla quota di energia condivisa da impianti e utenze di consumo connesse sotto la stessa cabina primaria;
- nei casi in cui la condivisione è effettuata sfruttando la rete pubblica di distribuzione, è previsto un unico conguaglio, composto dalla restituzione delle componenti di cui all'articolo 32, comma 3, lettera a), compresa la quota di energia condivisa, e dall'incentivo di cui al presente articolo.

**DLGS 210/2021 (IEM)**

*Partecipazione e controllo:*

- La CEC è un soggetto di diritto, con o senza personalità giuridica fondato sulla partecipazione volontaria e aperta, controllato da membri o soci che siano persone fisiche, autorità locali, ivi incluse le amministrazioni comunali, gli enti di ricerca e formazione, del terzo settore e di protezione ambientale nonché le amministrazioni locali.

*Obiettivi:*

- La CEC ha lo scopo principale di offrire ai suoi membri o soci o al territorio in cui opera benefici ambientali, economici o sociali a livello di comunità anziché perseguire profitti finanziari.

*Ambiti di attività:*

- La CEC può partecipare alla generazione, alla distribuzione, alla fornitura, al consumo, all'aggregazione, allo stoccaggio dell'energia, ai servizi di efficienza energetica, o a servizi di ricarica per veicoli elettrici o fornire altri servizi energetici ai suoi membri o soci.

*Condivisione dell'energia:*

- L'energia è condivisa nell'ambito della porzione della rete di distribuzione sottesa alla stessa zona di mercato;
- l'energia condivisa è pari, in ciascun periodo orario, al valore minimo tra quello dell'energia elettrica prodotta e immessa in rete dagli impianti e quello dell'energia prelevata dall'insieme dei clienti associati.

*Regolazione:*

- ARERA determina, anche in via forfettaria, l'ammontare delle componenti tariffarie regolate che non devono essere applicate all'energia condivisa nell'ambito della porzione di rete di distribuzione sottesa alla stessa cabina primaria e istantaneamente autoconsumata, in quanto corrispondenti a costi evitati per il sistema.

Attraverso i due decreti legislativi che sanciscono il definitivo recepimento delle direttive europee vengono introdotte delle novità sostanziali per le CER, mentre per gli schemi di autoconsumo collettivo non vengono definiti cambiamenti notevoli. Le principali novità riguardanti le CER sono rappresentate dalla possibilità di essere incentivati sull'energia condivisa attraverso impianti con potenza fino a 1 MW con l'estensione del perimetro di incentivazione fino alla cabina primaria.

Nel 27 dicembre 2022 è stata emanata la Delibera ARERA 727/2022/R/EEL che si esprime in merito alla precedente sperimentazione delle configurazioni di autoconsumo collettivo e CER approvando il modello regolatorio virtuale in quanto:

- consente ad ogni soggetto partecipante di modificare le proprie scelte con facilità, sia in relazione alla configurazione per l'autoconsumo, sia per effetto delle proprie scelte di approvvigionamento dell'energia elettrica, poiché non ha comportato l'esigenza di nuove connessioni o di nuovi collegamenti elettrici;
- garantisce trasparenza e flessibilità per tutti coloro che intendono prendervi parte: ogni cliente e produttore continua ad avere la propria autonomia e continua a mantenere i propri diritti e doveri;
- valorizza l'autoconsumo in modo esplicito in funzione della miglior stima possibile dei benefici indotti dall'autoconsumo medesimo nel sistema elettrico; allo stesso modo, pone le basi per attribuire un incentivo esplicito per promuovere determinate fonti o tecnologie. Il modello regolatorio virtuale consente inoltre di separare la quantità di energia elettrica oggetto di valorizzazione per l'autoconsumo dalla quantità di energia elettrica oggetto di incentivazione;
- garantisce flessibilità ai gruppi di autoconsumatori di energia rinnovabile che agiscono collettivamente e alle comunità energetiche rinnovabili, che possono organizzarsi come ritengono più opportuno. Anche la remunerazione, erogata dal GSE al referente, può essere liberamente ripartita tra i membri, sulla base di contratti di diritto privato.

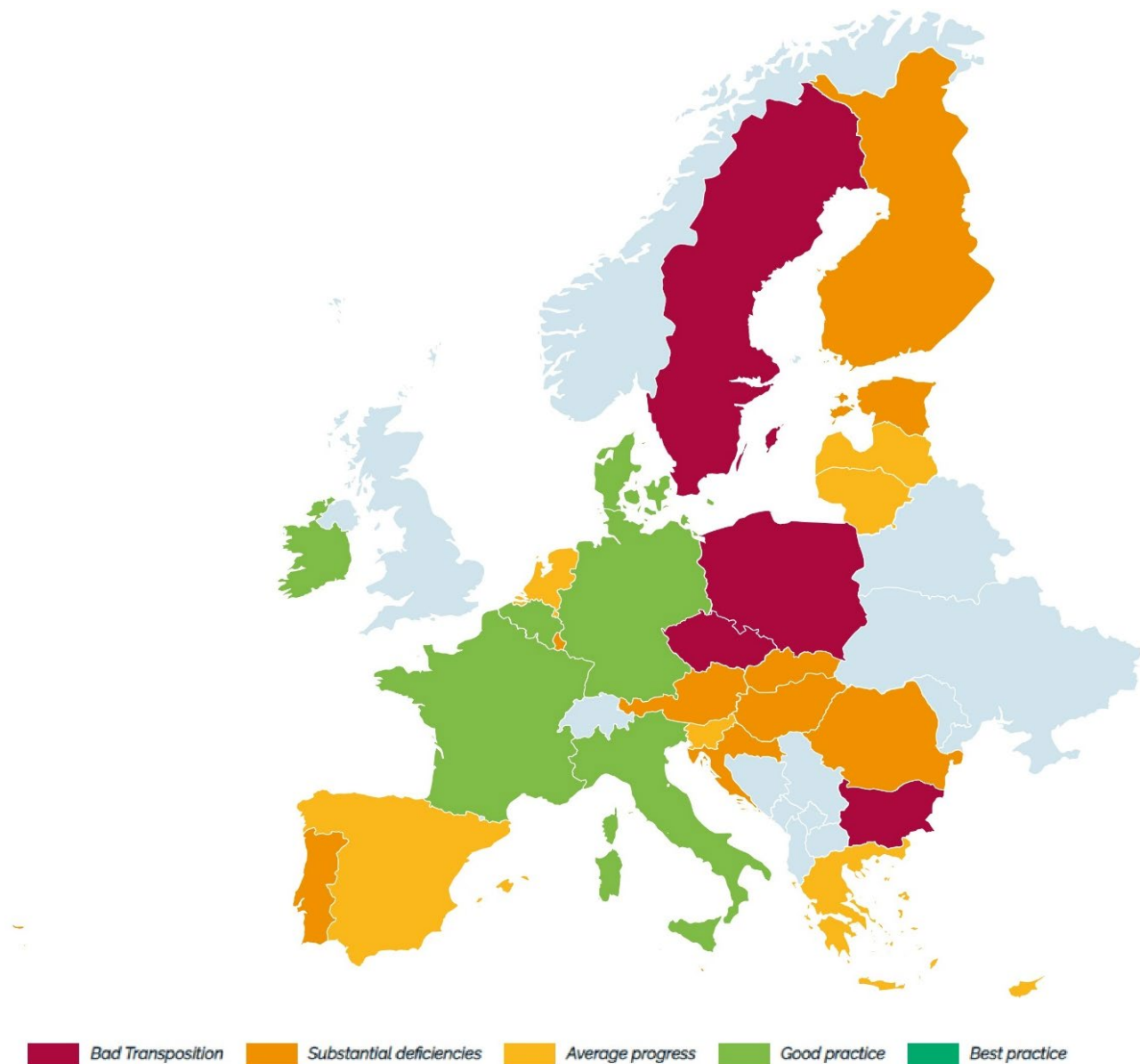
Per questi motivi, in attuazione del modello regolatorio virtuale si definisce che:

- ogni cliente finale e ogni produttore acquista e vende, rispettivamente, la propria energia elettrica prelevata e la propria energia immessa;
- il GSE riconosce al referente la valorizzazione dell'autoconsumo ai sensi della deliberazione 318/2020/R/EEL;

- il GSE riconosce al referente l'incentivo previsto dal decreto ministeriale 16 settembre 2020;
- il referente ripartisce gli importi ricevuti dal GSE tra i membri del gruppo di autoconsumatori di energia rinnovabile, secondo le modalità autonomamente definite sulla base di contratti di diritto privato.

### **1.3 Le Comunità Energetiche in Europa e nel mondo**

La nascita di una comunità energetica prevede l'aggregazione di un certo numero di produttori e consumatori disposti a condividere impianti di produzione di energia elettrica prodotta da fonte rinnovabile [16]. Allo stato attuale, in molti paesi, sono presenti varie iniziative di autoconsumo collettivo e di comunità energetiche. Il progetto *European Federation of citizen energy cooperatives* ha tracciato i movimenti dei singoli paesi europei nell'ambito delle iniziative e del recepimento delle normative europee riguardo le comunità energetiche rinnovabili e le comunità energetiche dei cittadini (**Figura 7**) [25].



**Figura 7** Sviluppo delle CER e delle CEC in Europa [25].

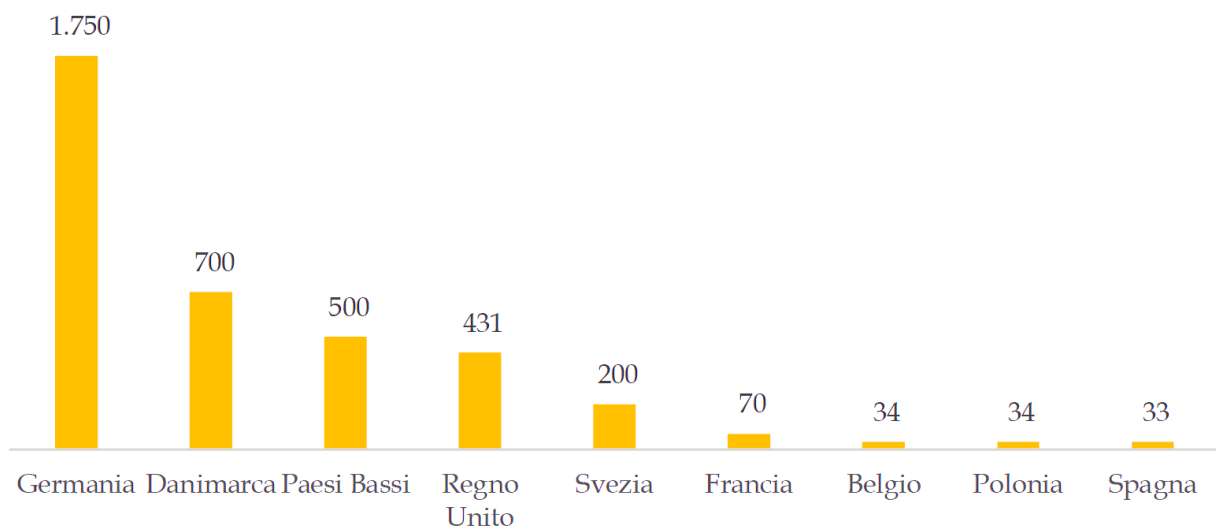
La valutazione del progresso dei singoli paesi è stata fatta tenendo conto:

- di come i criteri e i concetti espressi nelle normative europee sono state integrate dalla legislazione nazionale;
- del grado di dettaglio con il quale sono state recepite le norme;
- del ruolo che le CEC e le CER assumono nel contesto nazionale;
- di come i concetti di governance cooperativa vengono integrati nel panorama legislativo della nazione;

- della predisposizione di specifiche strutture giuridiche;
- del grado di partecipazione dei cittadini;
- della presenza o meno di una o più autorità designate al monitoraggio delle CER e delle CEC.

La **Figura 7** mette in risalto i paesi più avanti nello sviluppo delle tematiche relative alle CER e alle CEC che sono: Italia, Germania, Francia, Belgio, Irlanda e Danimarca. Presentano un minor avanzamento paesi come Spagna, Slovenia, Cipro, Grecia, Lituania e Olanda mentre i paesi che hanno introdotto sostanziali differenze rispetto alle normative europee sono: Portogallo, Romania, Slovacchia, Lussemburgo, Malta, Finlandia, Estonia, Croazia, Austria.

In **Figura 8** vengono rappresentati il numero di comunità energetiche nel 2022 per 9 dei paesi europei elencanti in precedenza.



**Figura 8** Numero di CER nel 2022 per 9 Paesi europei [6].

Si effettua un breve focus sulla Germania, paese con il numero di più alto di CER nel territorio. La proprietà di impianti energetici da parte di piccole comunità locali è un fenomeno diffuso da tempo in Germania, fin dal 2017, quando è stato introdotto il Renewable Energy Sources Act (EGG) introducendo il concetto di Citizens' Energy Company. Esso consiste in un gruppo di almeno dieci persone fisiche localizzate nella stessa città o nello stesso distretto, in cui è situato almeno

un impianto di produzione da fonte eolica. La definizione di questa configurazione è stata poi rivisitata dal primo gennaio del 2023, diventando di fatto un gruppo di almeno cinquanta persone che vivono in un'area con lo stesso codice postale ed entro un raggio di 50 km, alimentato da impianti fotovoltaici o eolici, di proprietà. Tali configurazioni, non sono sempre definibili come Comunità di Energia Rinnovabile, secondo la definizione della RED II. Non solo per via della semplice definizione ma anche per la definizione dei diritti, degli obblighi e delle azioni consentite all'interno di queste configurazioni.

In generale, l'elemento che accomuna le varie iniziative tra le CER realizzate in Europa e nel mondo è la dimensione locale dei sistemi di produzione e di scambio di energia da fonti rinnovabili.

Di seguito, vengono riportati alcuni esempi di progetti avviati e conclusi tra il 1999 e il 2016.

### **Esempi di Comunità Energetiche Extra UE**

#### ***Grupo Creluz, Rio Grande do Sul, Brasile – 1999***

*Territorio, scala:* Rurale, piccola scala.

*Tecnologia:* 6 piccoli impianti idroelettrici.

1\

*Collaborazioni:* 20 mila membri della cooperativa, impresa sociale locale.

*Altri fattori;* Fornire gratuitamente elettricità a 600 famiglie povere. Il gruppo possiede e gestisce allo stato attuale 6 impianti idroelettrici collocati lungo alcuni corsi fluviali, che garantiscono circa 4 MW di fornitura, pari al 27% dell'energia elettrica richiesta da 80.000 persone. L'elettricità viene poi venduta ai residenti della zona, soci della cooperativa. I membri delle cooperative che generano l'elettricità sono anche i membri che la consumano.

#### ***The Brooklyn Microgrid (BMG), New York (USA) – 2016***

*Territorio, scala:* Piccola scala.

*Tecnologia:* Principalmente FV residenziale.

*Collaborazioni:* BMG è stata fondata come benefit corporation dalla società L03 Energy nel tentativo di modificare il modello tradizionale di rete energetica



presente incorporando il concetto di una rete energetica comunali dove i cittadini di Brooklyn, sia essi di natura residenziale che commerciale, possono acquistare e vendere energia rinnovabile generata localmente.

*Altri fattori:* I partecipanti accedono al mercato locale dell'energia attraverso l'applicazione mobile di BMG. Nell'app, le persone possono scegliere di acquistare o vendere energia solare locale, energia rinnovabile e/o energia dalla rete. Possono impostare il loro budget di spesa giornaliero per l'acquisto di energia sul mercato. Una volta installati dei contatori "intelligenti" targati BMG, che raccoglie e registra i dati energetici per l'utilizzo interno dei mercati dell'energia, i professionisti possono scegliere se vendere l'energia solare in eccesso al mercato o continuare a utilizzare l'energia di rete.

## **Esempi di Comunità Energetiche in Europa**

### ***Bioenergy vilalge Jühnde, Germania – 2004***

*Territorio, scala:* Rurale, scala regionale.

*Tecnologia:* Impianto di cogenerazione da 700 kW a biogas per produrre elettricità che viene fornita alla rete pubblica. Una caldaia a cippato da 550 kW che viene utilizzata in inverno per fornire il riscaldamento che circola intorno alla rete locale.

*Collaborazioni:* Cooperativa locale e Università di Göttingen.

*Altri fattori:* La comunità energetica produce il 70% di calore e il doppio dell'energia elettrica richiesta. L'impianto bioenergetico è di proprietà locale e collettiva degli abitanti del posto. I residenti possono acquistare azioni della società cooperativa che possiede l'impianto (attualmente quasi il 75% degli abitanti del paese sono membri di questa società). Una volta acquistate le azioni e diventati soci, i cittadini possono acquistare il riscaldamento e l'elettricità dall'azienda. Ciò comporta che gli stessi consumatori di energia diventano anche produttori.

### ***Localised Energy Systems – Community Energy Generation, Aggregation and***

### ***Demand Shaping (les-cegads), Regno Unito – 2015***

*Territorio, scala:* Cluster di villaggi dell'Oxfordshire, Shrivenham, Watchfield e Longcot.

*Tecnologia:* Pannelli FV; 45 kW di picco, 60 kWh di accumulo termico in 6 abitazioni e 2 kWh di batterie agli ioni di litio per 9 altre abitazioni.

*Collaborazioni:* 48 famiglie partecipanti sono state reclutate per questa iniziativa.

*Altri fattori:* Il successo del processo CEGADS ha portato una prima implementazione di natura puramente commerciale del concetto di comunità nella piccola cittadina di Bethesda, nel nord del paese.

### ***ACCESS Project Isole Mull e Iona, Scozia – 2015***

*Territorio, scala:* Locale.

*Tecnologia:* Ideazione e creazione di un sistema in cui la quantità di energia elettrica che viene esportata sulla rete di distribuzione e/o di trasmissione è limitata dall'accensione di una "domanda" corrispondente sottoforma di riscaldamento ad accumulo elettrico, cilindri di acqua calda o caldaie elettriche a flusso nelle abitazioni.

*Collaborazioni:* Finanziato dal governo scozzese attraverso il local Energy Challenge Fund del 2015. Il costo totale del progetto ammonta a circa due milioni e mezzo di sterline, con una sovvenzione di 1.8 milioni da parte del governo, 0.3 milioni da parte dell'Electricity Network Innovation Allowance (NIA) di Ofgem tramite SSE Networks e circa 0.4 milioni di sterline da parte dei partner del progetto.

*Altri fattori;* Dimostra come il carico aggregato può essere utilizzato per controllare la quantità di energia importata e, se necessario, esportata dall'isola.

### ***Tower Power, Scozia – 2016***

*Territorio, scala:* Locale.

*Tecnologia:* Aggregazione della domanda di energia delle comunità locali e

abbinamento alle tecnologie rinnovabili locali (principalmente FV residenziale). Ha anche l'obiettivo di installare riscaldatori ad accumulato.

*Collaborazioni:* È finanziato dal fondo Local Energy Challenge del governo scozzese e collabora con Energy Local, Our Power, TMA, The city of Edinburgh Council e Glasgow City Council.

## **1.4 Le Comunità Energetiche in Italia**

In Italia si registrano le prime esperienze di CER a partire dai primi anni 2000 [6]. Solo negli ultimi anni però hanno assunto un nuovo modello volto ad incrementare la partecipazione degli utenti ad assumere un nuovo modello volto ad aumentare la partecipazione degli utenti sulla base dei principi di autoconsumo e sostenibilità ambientale.

Per la prima volta, le CER vengono menzionate nel 2017 attraverso la Strategia Energetica Italiana del 2017 e, successivamente, dal Piano Nazionale Energia e Clima nel 2018. Nello stesso anno poi, la Regione Piemonte ha approvato una legge sulle CER, risultando dunque la prima iniziativa legislativa esplicitamente dedicata al settore delle comunità energetiche.

Inizialmente, le CER erano piccole realtà locali, ma negli ultimi anni queste realtà si sono ampliate, allargando la platea di membri e progetti, beneficiando di economie di scala e diventando più efficienti nel fornire i servizi. Nel 2020 è infatti stata definita in maniera più accurata la legislazione applicabile a queste realtà che rende possibile, dal punto di vista formale, la costituzione di comunità energetiche rinnovabili nel territorio italiano.

Allo stato attuale, infatti, le CER realizzate in Italia seguono le norme definite nell'ambito dell'art. 42-bis del D.L. n. 162/2019 che recepisce in via transitoria la Direttiva RED II.

Secondo il recente report di Legambiente datato 2022, sono presenti sul territorio italiano ben 35 realtà effettivamente operative tra CER e configurazioni di autoconsumo collettivo, 41 in fase di progetto e 24 in fase di avviamento. Di queste 59 sono quelle di nuova realizzazione, censite tra giugno 2021 e maggio

2022, una crescita importante che evidenzia l'altro interesse che sta muovendo amministrazioni comunali, condomini, cittadini e imprese del terzo settore [19].

Tra queste, 39 sono comunità energetiche rinnovabili e 20 sono invece configurazioni di autoconsumo collettivo. Le iniziative che si stanno sviluppando partono dal basso coinvolgendo centinaia di famiglie, decine di comuni e imprese.

Lo sviluppo di queste nuove realtà apre a opportunità di innovazione come la CER Nuove Energie Alpine, la prima comunità energetica ad aver superato la criticità del vincolo della cabina primaria o quella di Ventotene, pensata per soddisfare i bisogni e valorizzare la potenzialità dell'isola, o ancora come la Comunità Energetica Critaro, in Calabria, insieme a quelle siciliane di Messina, Sortino e Blufi, che hanno posto l'accento sui benefici sociali per le fasce di popolazione che vivono in condizioni di disagio socioeconomico [19].

Tante esperienze che coinvolgono 55 Comuni, da San Daniele, prima CER operativa del progetto di area vasta in Friuli-Venezia Giulia, a Ragusa, dove l'amministrazione comunale ha deciso di intraprendere un vasto progetto di promozione di comunità energetiche sul territorio; da Basiglio, prima CER nell'area metropolitana di Milano, fino ad Ussaramanna e Villanovaforru, in Sardegna, e oltre 20 imprese coinvolte direttamente e non nella realizzazione di comunità energetiche, come ad esempio il Gruppo Amaranto e ad AMARES.

Come dimostrano le numerose esperienze riportate da Legambiente e dagli studi di settore [6, 19, 20, 21], le CER non rappresentano solamente una opportunità di produzione energetica e di riduzione dei costi in bolletta, ma un'occasione di assumere il ruolo di protagonisti, nel proprio territorio, del cambiamento necessario al miglioramento delle proprie condizioni di vita.

Non a caso, numerose amministrazioni locali stanno sviluppando comunità energetiche con lo scopo di combattere la povertà energetica portando allo stesso tempo innovazione e opportunità di sviluppo locale. Proprio in questa direzione viaggiano le CER di Napoli Est e di Ferla che hanno ispirato la realizzazione della Rete delle Comunità Energetiche Rinnovabili e Solidali, un soggetto informale che si pone l'obiettivo di costruire un'alleanza dal basso per la lotta alla povertà

energetica, portano queste realtà in contesti con notevoli criticità ambientali ma soprattutto socioeconomiche, per la costruzione di processi di partecipazione e innovazione sociale capaci di innescare un profondo cambiamento nella mentalità del territorio, nell'ottica di una maggior giustizia ambientale e sociale.

A questa rete di CER al momento hanno aderito 48 soggetti diversi tra amministrazioni locali, associazioni, enti e CER con l'obiettivo di mettere insieme le competenze e le conoscenze da esportare poi nei singoli territori contribuendo così alla lotta contro le disuguaglianze e alla transizione energetica.

In **Figura 9** è presentata una mappatura delle migliori esperienze registrate ad oggi sulla condivisione dell'energia (CER e autoconsumo collettivo).



Figura 9 Mappatura delle migliori esperienze sulla condivisione dell'energia [19].

In **Tabella 2** sono elencate le principali iniziative di CER e autoconsumo collettivo sul territorio italiano.

**Tabella 2** Principali iniziative di CER e autoconsumo collettivo [19].

### Comunità Energetiche

- 
- 01 CER Nuove Energie Alpine
  - 02 Comunità Energetica Quartiere Tannino

- 03 Comunità Energetica Rinnovabile Basiglio A
- 04 Comunità Energetica "Solisca"
- 05 "Monticello Green Hill"
- 06 La "CER Comunità Collinare del Friuli – San Daniele 1"
- 07 Energia Verde Connessa
- 08 Comunità Energetica Rinnovabile Gallese
- 09 "Verso il Futuro", la Comunità Energetica Rinnovabile del Lazio Meridionale
- 10 Comunità energetica di Ventotene
- 11 "CE Rossini", la comunità energetica di Montelabbate
- 12 AMARES, la comunità energetica di Ripalimosani
- 13 La comunità energetica di Foiano di Val fortore
- 14 Comunità Energetica Rinnovabile e Solidale "Critaro"
- 15 Blue Green Energy
- 16 Comunità Energetica Rinnovabile e Solidale di Messina
- 17 Le comunità Energetiche di Ragusa
- 18 Comunità Energetica Rinnovabile e Solidale di Sortino
- 19 GECO Green Energy Community
- 20 CommOn Light
- 21 Comunità energetica Biddanoa E' Forru
- 22 Comunità energetica di Ussaramanna
- 23 Miracer
- 24 Amendolara punta sulle rinnovabili condivise
- 25 La comunità energetica d'impresе a Bari
- 26 Associazione Comunità Energetica di Fondo Saccà – E.T.S.

### **Territori in movimento**

---

- 01 Aosta verso le comunità energetiche
- 02 La comunità energetica di Rudiano
- 03 Il progetto pilota di Ampezzo
- 04 La comunità energetica di Treviso
- 05 Roma verso le comunità energetiche
- 06 La comunità energetica di Gagliano Aterno

- 07 Il Molise dell'energia collettiva
- 08 La comunità energetica rinnovabile e solidale di Piaggine
- 09 San Giovanni a Piro puta sulle comunità energetiche
- 10 La Basilicata energetica e rinnovabile
- 11 La comunità energetica di Tiriolo
- 12 La comunità energetica per le periferie di Reggio Calabria
- 13 La comunità energetica allo Zen di Palermo
- 14 Il progetto REACT

### **Autoconsumo Collettivo**

---

- 01 Gli Autoconsumatori Collettivi del Progetto Energheia – Autoconsumo collettivo
- 02 Il Residence Cicogna
- 03 RE(Y) Venezia – Retail Efficiency Venezia

Tra queste, vengono riportare e descritte le prime comunità energetiche costituite sul territorio italiano.

### ***Comunità energetica e solidale di Napoli est – Prima comunità energetica del Sud Italia***

*Comune di realizzazione:* Comune di Napoli.

*Fonti rinnovabili:* Solare fotovoltaico 53 kW.

*Promotore:* Legambiente, Fondazione Famiglia di Maria,  
Fondazione Con il Sud.

*Finanziamenti:* Fondi di Fondazione Con il Sud.

Rivoluzione energetica che nasce dalla periferia est di Napoli. Precisamente nel quartiere di San Giovanni a Teduccio, è stata avviata la prima comunità energetica rinnovabile e solidale del nostro Paese. Il progetto ha richiesto un investimento di circa 100.000 euro, finanziato da Fondazione con il Sud e promosso da Legambiente e dal territorio, a partire dalla Fondazione Famiglia di Maria e da 40 famiglie locali con disagi sociali. Il percorso intrapreso ha anche lo scopo di sensibilizzare le famiglie verso una maggior consapevolezza dei temi



energetici, al fine di efficientare i benefici generati dalla comunità.

La comunità dispone di un impianto fotovoltaico di 53 kW realizzato sulla copertura della Fondazione Famiglia di Maria, in grado di produrre circa 65mila kWh/anno di energia elettrica [19]. La stima del risparmio reale, in termini di minor energia elettrica consumata dagli aderenti alla CER è circa pari a 300.000 euro in 25 anni.

### ***Comunità energetica “energy city hall” – Prima comunità energetica del Nord Italia***

*Comune di realizzazione:* Comune di Magliano Alpi.

*Fonti rinnovabili:* Solare fotovoltaico 20 kW.

*Promotore:* Comune di Magliano Alpi e 5 privati cittadini.

*Finanziamenti:* Fondi comunali.

Il Comune di Magliano Alpi ha voluto rendere atto concreto la sua adesione al Manifesto delle Comunità Energetiche per una centralità attiva del Cittadino nel nuovo mercato dell’energia, promosso dall’Energy Center del Politecnico di Torino realizzando la prima comunità energetica del nord Italia. Essa è alimentata da un impianto fotovoltaico da 20 kW, installato su una struttura comunale.

L’impianto ha la finalità di garantire l’autosufficienza della struttura, di una biblioteca, delle scuole comunali e di una palestra e condividere energia in surplus con cinque famiglie, oltre che alimentare una colonnina di ricarica per le auto elettriche gratuita per i membri della CER. In questa configurazione, il comune ha inoltre acquistato dei contatori intelligenti che, collegati ai POD dei membri della comunità, monitorato e rilevano i consumi delle singole utenze ed effettuando l’analisi dei flussi energetici tramite la piattaforma di gestione Energy4Com.

## 2 Analisi dati e stato dell'arte

### 2.1 Analisi dati

L'analisi dei dati è un processo in cui i dati vengono raccolti, elaborati, esplorati e interpretati al fine di trarre informazioni utili e guidare decisioni basate sui dati. L'analisi dei dati è diventata sempre più importante in un'ampia gamma di settori, in particolare in quelli in cui i dati hanno un ruolo centrale, come ad esempio nel settore energetico. L'analisi dei dati per i processi *data driven* si riferisce all'utilizzo dei dati come base per prendere decisioni e guidare le operazioni in senso lato. In questo approccio, i dati sono al centro del processo e vengono utilizzati per identificare i modelli, le tendenze e le opportunità di miglioramento di processi generici.

Il processo di analisi dei dati per sistemi *data driven* inizia con la raccolta dei dati da diverse fonti, come sensori, strumenti di monitoraggio, sistemi di raccolta dati o piattaforme *web*. Una volta raccolti i dati, vengono elaborati e trasformati in formati utili per l'analisi, come tabelle, grafici e visualizzazioni. Successivamente, i dati vengono esplorati per identificare i modelli e le tendenze, utilizzando diverse tecniche, tra cui le tecniche di *data mining* e analisi statistica. L'analisi dei dati può essere divisa in diversi passaggi, a seconda del tipo di dati e degli obiettivi dell'analisi. Alcune delle fasi comuni includono l'analisi descrittiva, l'analisi esplorativa, l'analisi predittiva e l'analisi prescrittiva.

L'analisi descrittiva si focalizza sull'esplorazione dei dati per comprendere i pattern e le tendenze presenti nei dati. Questo tipo di analisi può aiutare a identificare eventuali problemi o inefficienze nei processi e suggerire soluzioni.

L'analisi esplorativa si concentra sull'identificazione dei *pattern*, delle correlazioni e delle relazioni tra i dati. Questo tipo di analisi può aiutare a individuare nuove opportunità e migliorare la comprensione di uno specifico fenomeno.

L'analisi predittiva si concentra sull'utilizzo dei dati storici per fare

previsioni sul futuro. Questo tipo di analisi può aiutare a prevedere la domanda futura, identificare le tendenze di un determinato fenomeno e identificarne le caratteristiche.

Infine, l'analisi prescrittiva, si concentra sull'utilizzo dei dati per suggerire decisioni basate sui dati. Questo tipo di analisi può aiutare a identificare le opportunità di ottimizzazione, sviluppare nuove strategie e migliorare l'efficienza di processi generici.

### 2.1.1 Pre-processing

Il *pre-processing* racchiude le fasi principali dell'analisi dati per i processi *data-driven*. Esso consiste nell'elaborazione e nella ricostruzione dei dati grezzi e serve a migliorare le performance di processi come il *clustering*, la creazione di modelli di *machine learning* e non solo. Il *pre-processing* è composto principalmente da due fasi:

- *Data Preparation*;
- *Data Characterization*.

Il *data preparation* è un passaggio di vitale importanza per processi di *data analytics* e spesso consiste nella parte più corposa dell'analisi. Esso è articolato in tre passaggi:

1. *Data cleaning*;
2. *Data transformation*;
3. *Data reduction*.

Il processo di *data cleaning* consiste nell'individuazione e rimozione degli *outliers*, nella gestione e sostituzione dei *missing values* e nella rimozione e sostituzione dei valori negativi. Si definiscono *outliers* i valori che non hanno una corrispondenza statistica con gli altri dati registrati e sono spesso caratterizzati dall'essere fuori scala. Essi possono essere sia valori puntuali, ovvero registrazioni anomale del dato al massimo grado di risoluzione della scala (*outliers* puntuali) o possono essere elementi molto dissimili dai restanti, come ad esempio forme del

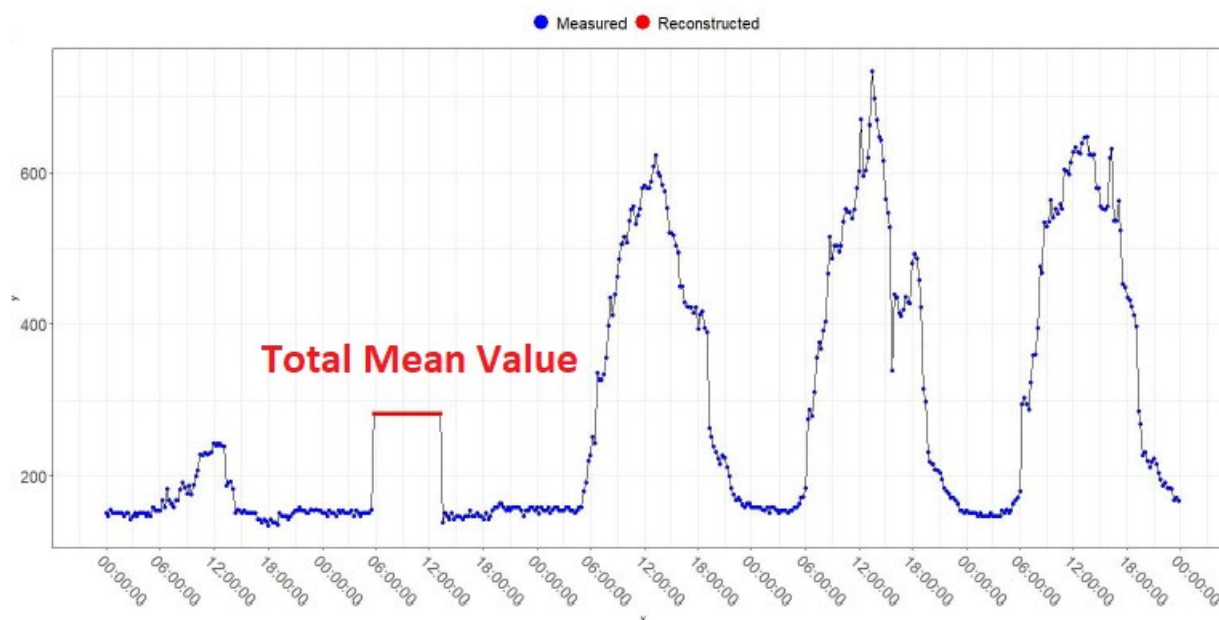
profilo di carico anomale o profili di carico su scala temporale differente caratterizzati da un consumo notevolmente diverso.

L'individuazione e la sostituzione dei valori mancanti rappresentano una fase cruciale del processo di pulizia dei dati. Tale procedura può essere effettuata in molteplici modi, a seconda della natura e delle caratteristiche del *database* di origine.

Partendo da esso, i metodi comunemente più utilizzati in letteratura sono:

- *global constant*;
- *local constant*;
- *moving average*;
- *linear interpolation*.

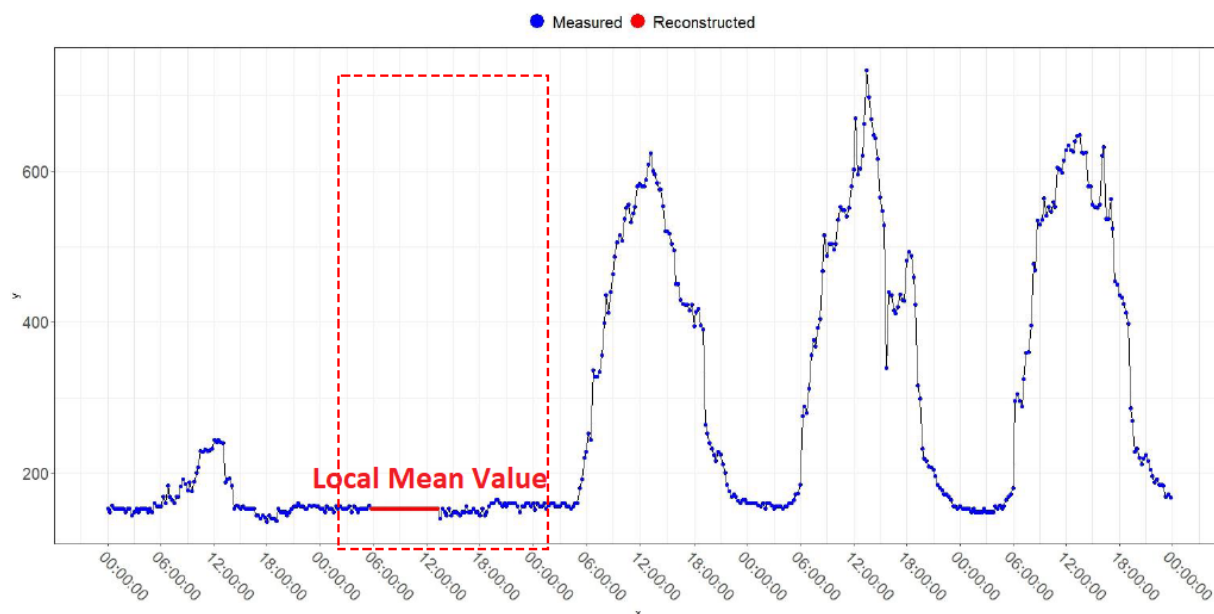
Il *global constant method* consiste nella sostituzione dei *missing values* con un solo valore, come ad esempio il valore medio o la mediana dei dati di partenza (**Figura 10**).



**Figura 10** Sostituzione di *missing values* tramite il metodo *global constant* [51].

Il *local constant method* consiste nella sostituzione dei *missing values* con un

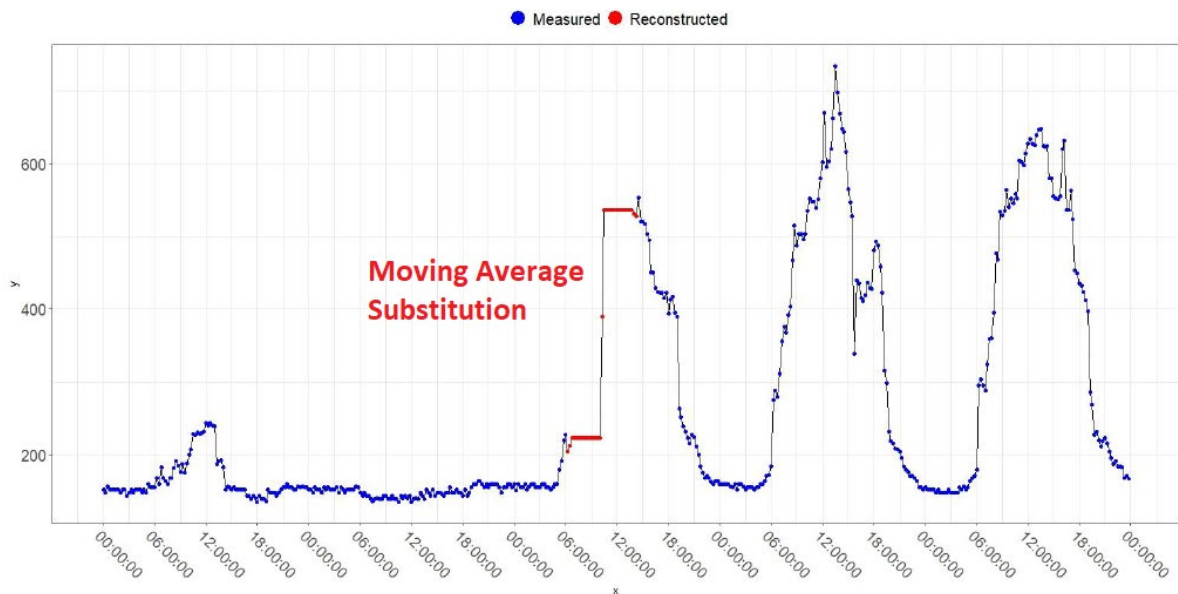
valore medio calcolato localmente, in prossimità dei dati mancanti (**Figura 11**).



**Figura 11** Sostituzione di missing values tramite il metodo local constant [51].

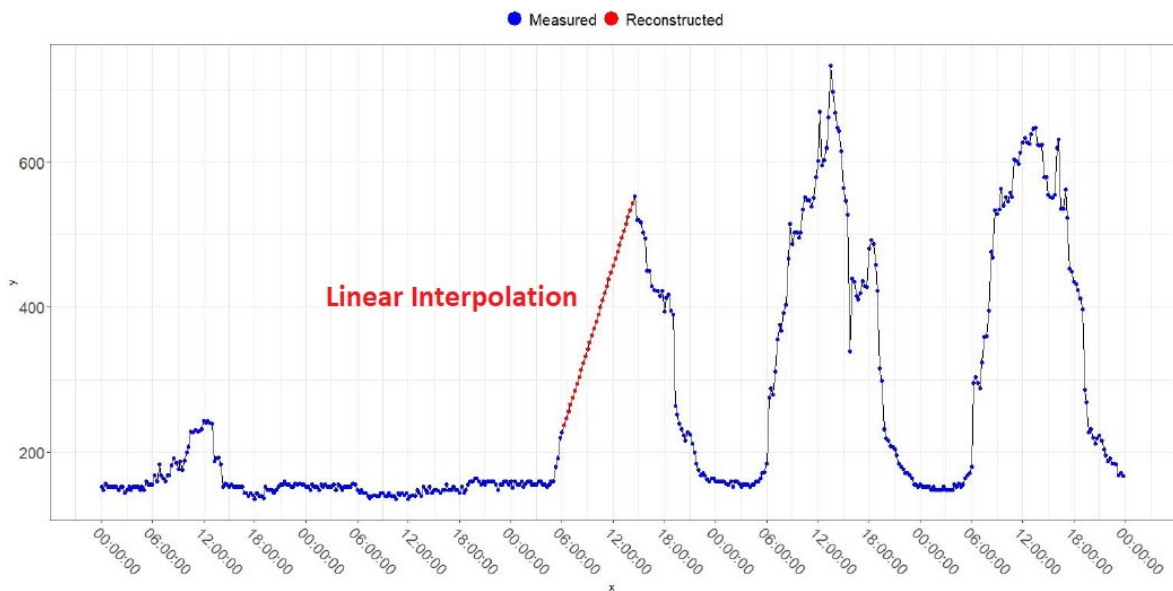
Il *moving average method*, noto anche come *simple moving average* (SMA) sostituisce i valori mancanti con la media non pesata delle  $n$  precedenti o successive misurazioni presenti (**Equazione 1**). Il valore con la quale effettuare la sostituzione viene quindi calcolato utilizzando le  $n$  misurazioni successive o precedenti ai dati mancanti. Inoltre, la sostituzione può essere considerata accettabile per numero massimo  $m$  di valori da ricostruire, al di sopra del quale viene iterato nuovamente il calcolo utilizzando i nuovi valori più vicini ai valori mancanti restanti, come rappresentato in **Figura 12**.

$$SMA = \frac{1}{N} \sum_{i=1}^N y_{-i} \quad (1)$$



**Figura 12** Sostituzione di missing values tramite Moving Average Substitution [51].

Il *linear interpolation method* consiste nella sostituzione dei *missing values* con un'interpolazione lineare tra le due misurazioni valide che delimitano il valore mancante (**Figura 13**).



**Figura 13** Sostituzione di missing values tramite interpolazione lineare [51].

In ogni caso, prima di procedere con l'effettiva sostituzione dei valori mancanti, è necessario accertarsi della presenza di periodi di mancanza del dato prolungati. La presenza di periodi estesi di dati necessita un trattamento diverso. Infatti, le tecniche comunemente utilizzate per la ricostruzione dei valori

mancanti, come l'interpolazione lineare, possono commettere un errore sull'approssimazione del profilo reale, nel caso di intervalli di assenza del dato molto estesi.

In questi casi, è opportuno optare per una sostituzione del profilo effettuata, ad esempio, mediante la media oraria dei giorni del medesimo mese, o con il profilo medio mensile.

Successivamente, per l'identificazione degli *outliers*, uno dei metodi più utilizzati in letteratura è l'"interquantile method". Esso rappresenta una delle metodologie più semplici ed efficaci utilizzate in ambito scientifico e richiede il calcolo di tre parametri fondamentali:

- la mediana, ovvero il valore centrale della distribuzione dei dati campionata;
- Q1, definito come primo interquartile, ovvero il valore al di sotto della quale sono presenti il 25% dei dati di partenza;
- Q3, definito come terzo interquartile, ovvero il valore al di sotto del quale risiedono il 75% dei dati di partenza.

La distanza tra Q3 e Q1 è definita *Inter-Quartile Range* (IQR), e tramite essa viene definito un nuovo intervallo, noto come *decision range*, i cui elementi al di fuori del quale vengono identificati come *outliers*.

Il *decision range* è definito secondo quanto rappresentato dalla **Figura 14**.

$$\text{Lower Bound} = Q_1 - 1.5 * IQR \quad (2)$$

$$\text{Upper Bound} = Q_3 - 1.5 * IQR \quad (3)$$

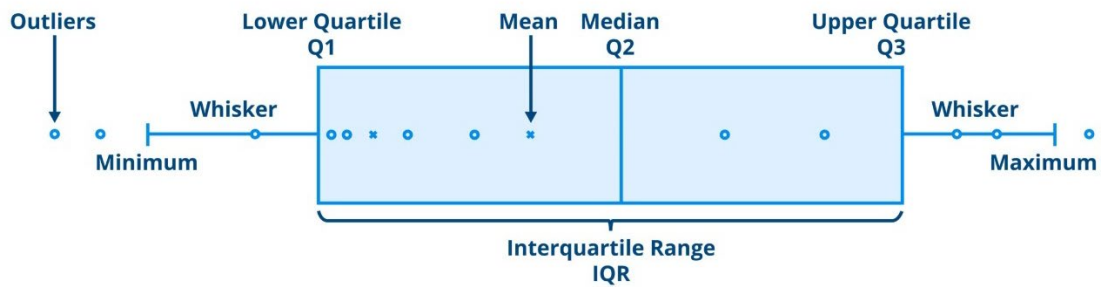


Figura 14 Rappresentazione grafica del decision range.

Una volta identificati gli *outliers* puntuali, essi possono essere sostituiti con le medesime tecniche di sostituzione discusse per i *missing values*.

Al *Data Cleaning* segue il processo di *Data Transformation*, che si articola nei seguenti passaggi:

- *Data Scaling*;
- *Data Type Transformation*.

Il *Data Scaling* (noto anche come normalizzazione) consiste nella normalizzazione dei dati presenti all'interno del dataset di partenza. Questo tipo di operazione è spesso utilizzata nello studio di serie storiche in ambito energetico per escludere l'impatto che alcuni parametri hanno sulla valutazione dei profili di carico delle utenze. Le principali tecniche di normalizzazione utilizzate sono:

- normalizzazione sul massimo:

$$x_i = \frac{x_i}{\max(x)} \quad (4)$$

- normalizzazione min-max:

$$x_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (5)$$

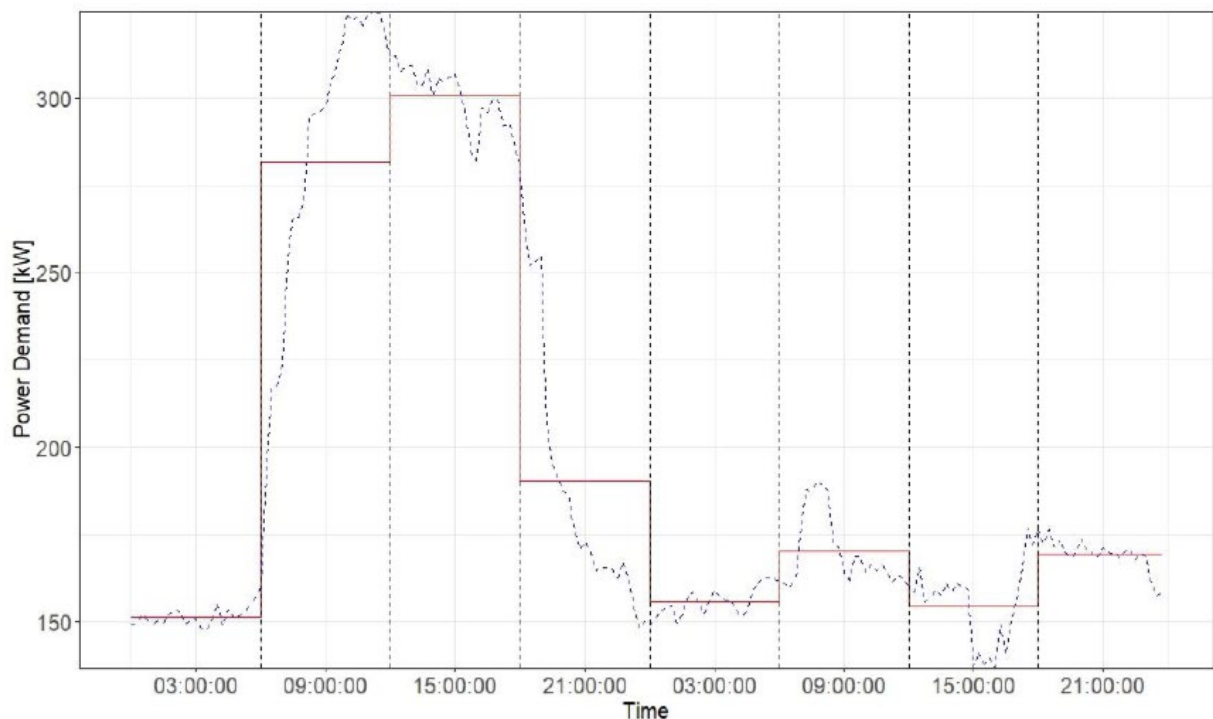
- normalizzazione "standard":



$$x_i = \frac{x_i - \text{mean}(x)}{\text{sd}(x)} \quad (6)$$

A tale processo segue il *Data type transformation*. Esso non è sempre necessario nel *pre-processing* ma può essere utile per organizzare il *dataset* in modo ottimizzato. Uno degli obiettivi più frequenti di questo passaggio è quello di trasformare dati numerici in variabili categoriche (come ad esempio consumo alto, medio, basso).

A conclusione del processo di *pre-processing*, si procede con il *Data Reduction*. Esso ha lo scopo di migliorare l'efficienza degli algoritmi, come ad esempio gli algoritmi di clustering e di classificazione, riducendo il costo computazionale attraverso il passaggio da una risoluzione dei dati più fine ad una più grossolana. Una delle tecniche utilizzate nel *Data Reduction* è la *Piecewise Aggregate Approximation* (PAA), che consiste nella riduzione del numero dei dati all'interno di una determinata finestra temporale con il valore medio (**Figura 15**).



**Figura 15** Rappresentazione grafica del PAA [51].

In **Tabella 3** vengono sintetizzate le scale temporali più utilizzate, per ambito di

utilizzo, nei processi *data-driven*.

**Tabella 3** Utilizzo dei dati in ambito energetico in funzione della scala temporale.

<i>Risoluzione sub-oraria ( &lt; 1 hour )</i>	<i>Risoluzione oraria ( 1 hour )</i>	<i>Risoluzione giornaliera ( 1 day )</i>	<i>Risoluzione mensile ( 1 month )</i>	<i>Risoluzione annua ( 1 year )</i>
<i>Short Term Forecasting</i>	<i>Short Term Forecasting</i>  <i>Load Profile Characterization</i>  <i>Benchmarking</i>	<i>Medium Term Forecasting</i>	<i>Medium-Long Term Forecasting</i>	<i>Scenario Analysis</i>

## 2.2 Overview delle principali tecniche di clustering applicate a serie storiche

In ambito accademico, vengono utilizzate diverse tecniche di *clustering* volte a individuare i *pattern* tipici del consumo energetico da parte di utenze generiche. La letteratura sottolinea come il *driver* principale per un corretto *clustering* di serie temporali energetiche sia rappresentato dal *daily shape*, ovvero l'andamento medio annuo su scala giornaliera del consumo energetico.

Tuttavia, alcuni studi dimostrano che per una suddivisione più accurata delle diverse tipologie di carico, le tradizionali tecniche di *clustering* basate sul *daily shape* possono essere integrate con opportune segmentazioni temporali in grado di evidenziare i diversi comportamenti delle utenze in finestre temporali specifiche [50]. Degli esempi di segmentazione temporale utilizzata a supporto di algoritmi di *clustering* sono la suddivisione per stagioni, per giorni della settimana

o per i diversi mesi dell'anno. Tale approccio è noto in letteratura come *Nested approach* e generalmente migliora le performance dei processi di *clustering* [50].

Le tecniche più utilizzate per il *clustering* di serie temporali nel contesto preso in considerazione sono il cosiddetto *Hierarchical clustering* e il *K-means clustering*. Tuttavia, possono essere utilizzate anche tecniche come il *Dynamic Time Warping Time Series clustering* e il *Latent Profile Analysis*. Ciascuna di queste tecniche presenta vantaggi e svantaggi e sono di tipo euristico, ovvero richiedono la normalizzazione dei dati grezzi di partenza.

Di seguito, viene fornita una panoramica delle principali caratteristiche delle metodologie di *clustering*, in modo da comprendere meglio le scelte effettuate nella metodologia proposta per l'analisi dei profili delle utenze in esame.

### 2.2.1 Clustering gerarchico

Il *clustering gerarchico*, noto anche come *Hierarchical clustering*, rappresenta una delle tecniche più famose nel campo dell'analisi dei dati. Esso si può suddividere in due tipologie, ovvero agglomerativo o divisivo. Il *clustering gerarchico agglomerativo* adotta un approccio *bottom-up*, in cui si parte dai singoli elementi, considerati tutti come *cluster* distinti e si procede con l'aggregazione degli elementi più vicini, fino ad ottenere un solo ed unico *cluster*, coincidente con l'insieme di partenza. Tale approccio viene definito gerarchico in quanto l'aggregazione avviene gradualmente.

Una volta che un elemento viene incluso in un *cluster*, vi rimane all'interno fino alla fine del processo. In modo opposto, il *clustering gerarchico divisivo* segue un approccio in cui tutti gli elementi sono inizialmente raggruppati nello stesso *cluster* e successivamente vengono suddivisi passo dopo passo in base alla loro distanza, fino al raggiungimento di un numero di *cluster* pari al numero degli elementi analizzati.

La tipologia di *clustering gerarchico* più utilizzata nella distinzione dei profili tipologici di utenze generiche è quella di tipo agglomerativo, che si articola nei seguenti passaggi:

1. calcolo della matrice delle distanze;
2. costituzione dei cluster contenenti un singolo elemento;
3. unione dei cluster più "vicini";
4. calcolo della matrice delle distanze con i nuovi *cluster* appena formati
5. Iterazione dello step 3 e 4 fino a quando non si ottiene un singolo *cluster*.

Nel calcolo della matrice delle distanze possono essere utilizzati diversi criteri:

*Distanza Euclidea*

$$D(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (7)$$

*Distanza di Mahalanobis*

$$D^2(x, y) = (X - Y)^T S^{-1} (X - Y) \quad (8)$$

dove S è la matrice di covarianza delle variabili e  $S^{-1}$  è la sua matrice inversa.

*Distanza di Minkowsky*

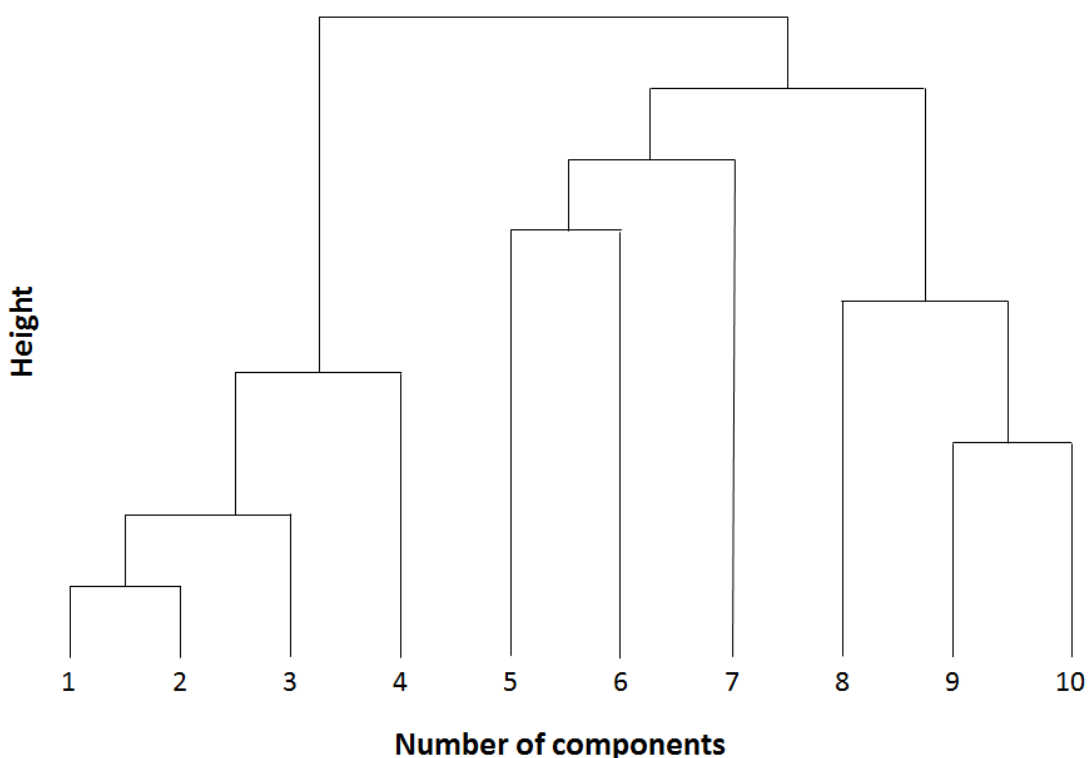
$$D(x, y) = \sum_{i=1}^n (\text{abs}(x_i - y_i)^w)^{1/w} \quad (9)$$

Per il *clustering* gerarchico di tipo agglomerativo è necessario inoltre definire, oltre al criterio del calcolo della distanza, il metodo di agglomerazione. I più utilizzati sono i seguenti:

- *Simple linkage* – distanza tra i due *cluster* più vicini;
- *Complete linkage* – distanza tra i due *cluster* più lontani;
- *Average linkage* – media tra le distanze tra tutti i casi di due *cluster*;
- *Centroid linkage* – distanza tra i centroidi dei *cluster*;

- *Ward linkage* – la minimizzazione dell'errore quadratico medio minimo che si ottiene dall'unione di due *cluster*.

Tra i metodi di agglomerazione elencati, il *Ward linkage* è noto in letteratura per essere il meno sensibile agli *outliers*. È consigliabile di utilizzare la metodologia di agglomerazione *Ward* insieme la distanza euclidea, per il calcolo della matrice delle distanze, nella condizione in cui la deviazione standard in ogni cluster è simile [50].



**Figura 16** Esempio di un dendrogramma [50].

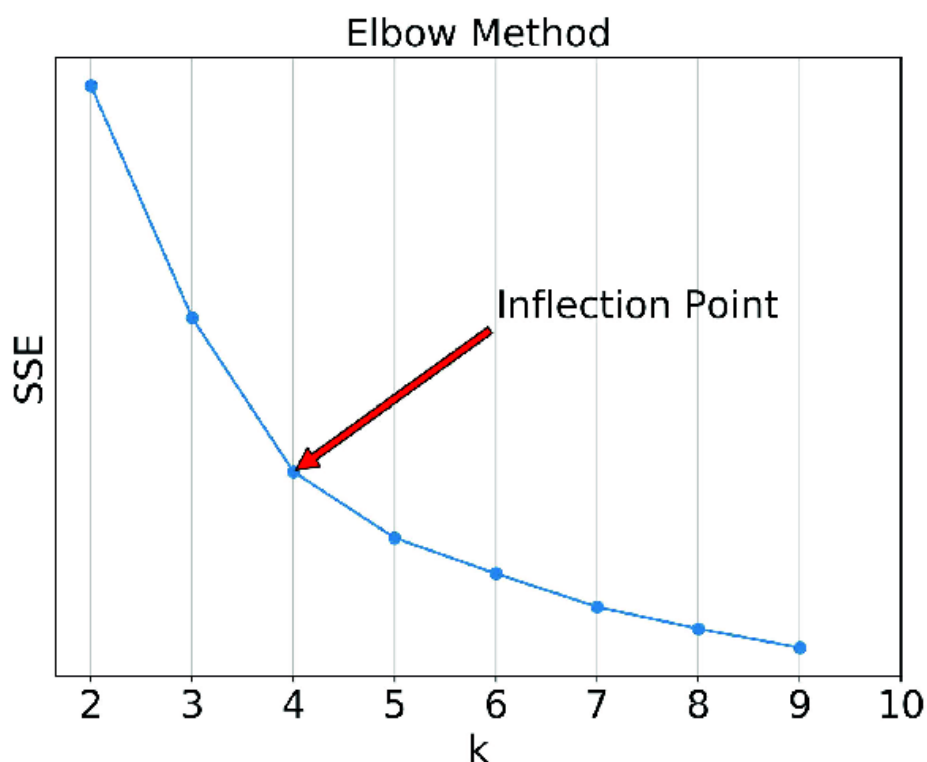
Dopo aver individuato la metodologia di agglomerazione ed aver effettuato il calcolo della matrice delle distanze, è necessario individuare il numero ottimale di *cluster* in cui si decide di dividere il *dataset*. È una delle parti più cruciali dei processi di *clustering* e non esiste una metodologia universale in grado di definire un numero ottimale da utilizzare. L'identificazione del numero ottimale di *cluster* in un *clustering* gerarchico può essere fatta in diversi modi, tra cui:

- *Elbow method* utilizzando la somma degli errori quadratici (SSE), calcolata secondo la **Equazione 10**;

- Utilizzo del dendrogramma ottenuto come risultato dal clustering gerarchico.

Il primo consiste nel trovare il numero di cluster tale per cui un incremento unitario di tale valore non è in grado di provocare un forte aumento della somma degli errori quadratici medi. Tale valore corrisponde graficamente al “ginocchio” della curva rappresentata in **Figura 17**.

$$SSE = \sum_{i=1}^K \sum_{x \in C_i} dist^2(m_i, x) \quad (10)$$



**Figura 17** Illustrazione grafica dell'Elbow Method dove  $k$  = numero di clusters.

Il secondo consiste nell'utilizzo del dendrogramma, ovvero dell'albero ramificato frutto dal processo di *clustering*. Per determinare il numero di *cluster* ottimale a partire dal dendrogramma di un clustering gerarchico agglomerativo, è possibile visualizzare il dendrogramma e individuare il punto in cui la distanza tra i *cluster* inizia a diminuire in modo meno significativo rispetto alle iterazioni precedenti, formando una sorta di restringimento nel grafico. Questo punto indica

il numero ottimale di cluster da selezionare per l'analisi dei dati.

In alternativa ai due, si può utilizzare il “metodo silhouette”, che calcola un coefficiente di similarità tra i dati all'interno di un cluster e tra i dati di un *cluster* e quelli degli altri *cluster*. Il coefficiente di *silhouette* assume valori tra -1 e 1, dove valori più elevati indicano una maggiore coesione dei dati all'interno del *cluster* e una maggiore separazione dai dati degli altri *cluster*. Il numero di *cluster* ottimale è quello che massimizza il valore medio del coefficiente di *silhouette*.

Il *clustering* gerarchico presenta alcuni svantaggi, tra cui il costo computazionale elevato. Infatti, nel primo passaggio del processo, viene generata una matrice delle distanze che può essere molto grande in base alla mole dei dati di partenza. Inoltre, il *clustering* gerarchico è generalmente sensibile agli *outliers* e ai cosiddetti “not convex clusters”. I *not convex clusters* sono raggruppamenti di dati che presentano forme irregolari e non omogenee, che possono essere difficili da gestire con le tecniche di *clustering* tradizionali.

D'altra parte, il *clustering* gerarchico presenta anche alcuni vantaggi, come la capacità di fornire una soluzione univoca e stabile, indipendentemente dal processo iterativo adottato. Inoltre, la flessibilità del metodo permette di utilizzare diverse definizioni di distanza e di metodi di agglomerazione, rendendolo adatto a molteplici tipologie di dati e strutture.

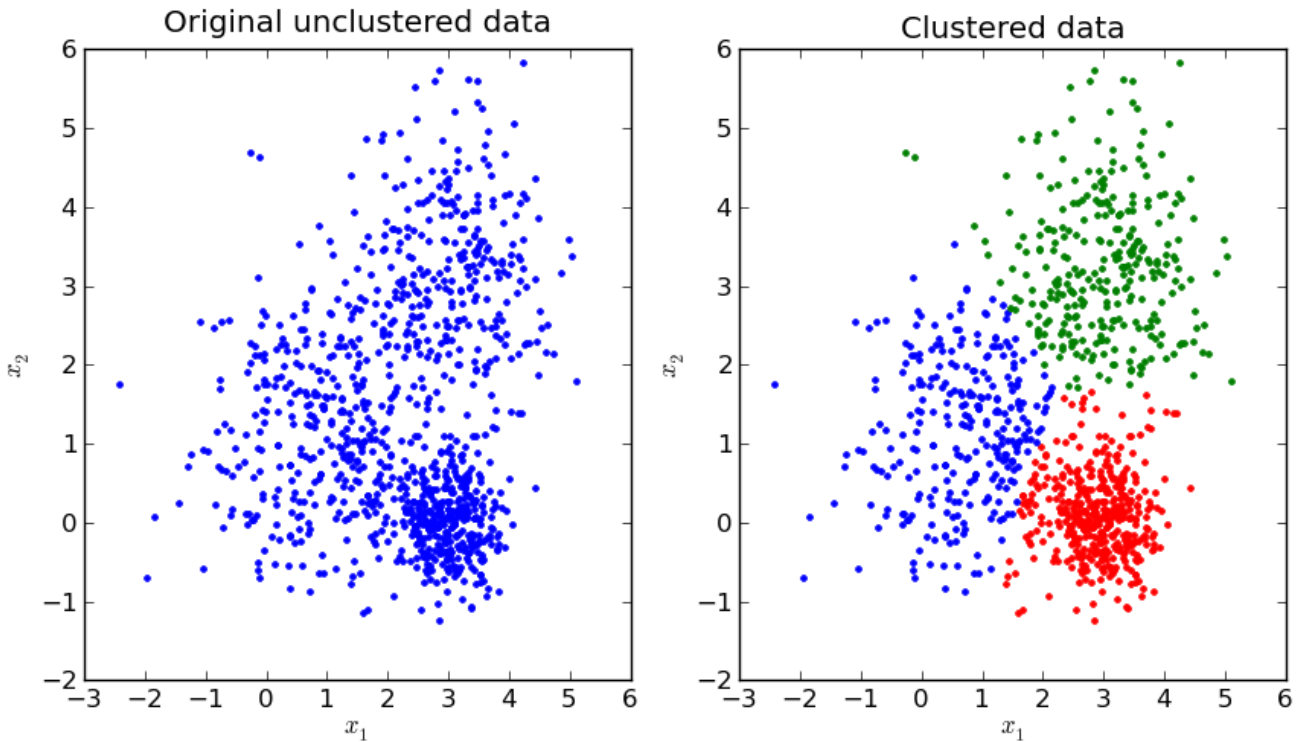
### 2.2.2 K-means clustering

Il *K-means* clustering rappresenta una delle tecniche maggiormente impiegate in campo analitico. Si caratterizza per la sua velocità ed efficienza, ma presenta alcuni svantaggi. L'algoritmo essendo di tipo partitivo, genera dei *cluster* non sovrapposti, ovvero raggruppamenti dove ogni elemento viene associato ad un solo e unico *cluster* (**Figura 18**).

Il processo si articola nei seguenti passaggi:

1. definizione di  $n$  elementi di partenza come centroidi dei *cluster* iniziali;
2. assegnazione dei *cluster* in base al concetto di distanza dai centroidi

- selezionati;
3. ricalcolo dei nuovi centroidi dei cluster appena generati;
  4. iterazione del secondo e del terzo punto fino al raggiungimento di una stabilità del sistema.

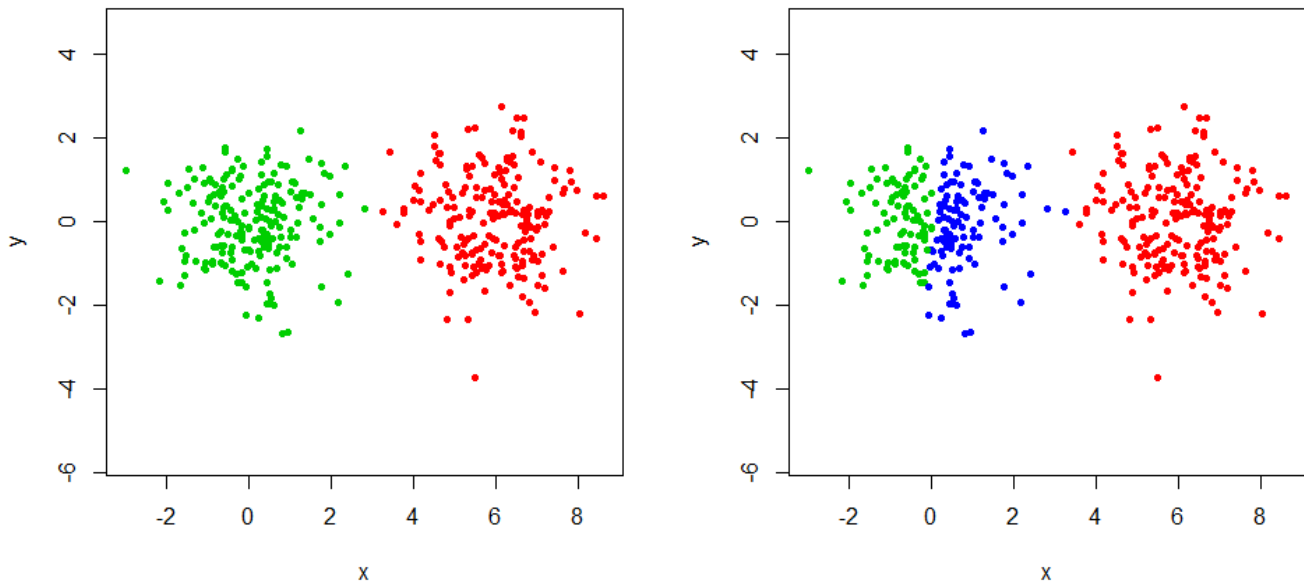


**Figura 18** Esempio di *K-means clustering*.

La definizione dei centroidi dei *cluster* è il punto cruciale del *K-means clustering*. La criticità di questo passaggio è legata alla scelta dei primi centroidi nella fase iniziale, poiché la scelta di centroidi differenti può portare a risultati diversi.

La velocità di esecuzione di questa tecnica è uno dei suoi vantaggi principali soprattutto rispetto al *clustering* gerarchico, inoltre può gestire un numero di dati di gran lunga più elevato. In aggiunta, nel *K-means clustering* è necessario definire il numero di cluster prima di intraprendere il processo e ciò può essere fatto, ad esempio, utilizzando un *cluster* gerarchico.





**Figura 19** Effetto dell'utilizzo di due numeri di cluster different nel *k-Means clustering*.

### 2.2.3 Latent Profile Analysis

Uno dei principali svantaggi del *clustering* gerarchico e del *K-means* è l'utilizzo di un approccio euristico non fondato su modelli formali. I così detti "model based clustering algorithms" propongono un approccio alternativo al processo di segmentazione dei dati, dove la scelta del numero di cluster è strettamente associata alla scelta del modello.

Il *Latent Profile Analysis* (LPA) è una tecnica statistica che può essere utilizzata per l'analisi dei dati energetici al fine di identificare gruppi omogenei di utenti o di consumi energetici. In questo ambito, l'obiettivo principale dell'analisi è quello di identificare i profili di consumo energetico degli utenti, ovvero le caratteristiche distintive dei loro consumi energetici.

Il primo passo del metodo LPA consiste nella definizione delle variabili che si desidera includere nell'analisi. Nell'ambito energetico, queste variabili possono includere il consumo di energia elettrica, il consumo di gas, le temperature esterne, l'orario di utilizzo dei dispositivi elettronici e altre variabili correlate, a seconda del contesto di utilizzo.

Successivamente, si procede alla stima dei modelli LPA, attraverso l'utilizzo di un *software* statistico apposito. Durante la stima del modello, si cerca di

identificare il numero ottimale di *cluster*, ovvero il numero di gruppi omogenei di utenti che meglio spiega la variabilità dei dati. Questo viene solitamente effettuato utilizzando una combinazione di criteri oggettivi (ad esempio, il criterio di informazione di Akaike o il criterio di Bayes) e di giudizio soggettivo, in cui l'analista cerca di trovare il giusto compromesso tra la complessità del modello e la sua capacità di rappresentazione dei dati [51].

Una volta identificato il numero ottimale di *cluster*, si procede alla caratterizzazione dei profili di consumo energetico degli utenti all'interno di ciascun *cluster*. Questo può essere fatto utilizzando una serie di tecniche di analisi multivariata, come l'analisi delle componenti principali, l'analisi discriminante o l'analisi delle varianze [51].

#### 2.2.4 Time Series Clustering (DTW)

Tutte le metodologie descritte in precedenza non tengono conto che i dati sono distribuiti lungo una serie temporale. I profili di carico, infatti, contengono precise informazioni riguardo la posizione temporale dei dati registrati e tali informazioni possono essere utilizzate per raggruppare serie temporali simili nello stesso cluster. Il *Time Series Clustering* tiene conto di queste peculiarità e può essere condotto in tre modi diversi:

- approccio "model based";
- approccio "feature based";
- approccio "raw-data-based".

Uno degli sviluppi più comuni del "raw-data-based times series clustering" consiste nell'utilizzo di una speciale formula per il calcolo della distanza. Viene utilizzata, infatti, una funzione denominata "Dynamic time Warping distance" (DTW) per il calcolo della matrice delle distanze. Tale processo è fortemente legato alla forma dei profili di carico ma richiede un elevato costo computazionale e fa fatica a gestire e riconoscere leggere variazioni all'interno della serie temporale [49]. Alcuni studi [46] dimostrano che l'utilizzo del DTW è una buona scelta nel momento in cui si desidera inquadrare efficacemente profili molto simili mentre non è di rilevante importanza individuare i carichi molto diversi tra loro.

## 2.2.5 Confronto tra le diverse metodologie di clustering

Un recente studio condotto dalla Commissione Europea sul clustering di profili di carico di natura residenziale [50] ha sintetizzato le principali caratteristiche delle metodologie di clustering descritte in precedenza (**Tabella 4**) e ha sperimentato l'utilizzo di esse su un vasto numero di serie temporali provenienti da diversi paesi [50].

**Tabella 4** Principali caratteristiche delle metodologie di clustering descritte

	<i>Clustering Gerarchico</i>	<i>K-means</i>	<i>Latent Profile Analysis</i>	<i>Time series approach (DTW)</i>
<i>Tipologia di dati richiesta</i>	Dati "grezzi" normalizzati	Dati "grezzi" normalizzati	Dati "grezzi" normalizzati	Dati "grezzi" normalizzati
<i>Dimensione dei dati</i>	Quasi illimitata	Quasi illimitata	Molto limitata	Limitata
<i>Livello di misurazione dei dati</i>	Continuo	Continuo	Continuo	Continuo
<i>Distribuzione dei dati</i>	Nessuna assunzione	Nessuna assunzione	Multivariata-normale	Nessuna assunzione
<i>Tempo di esecuzione</i>	Lento	Veloce	Lento	Molto lento
<i>Sensibilità agli outliers</i>	Elevata	Elevata	Bassa	Elevata
<i>Determinazione del numero di cluster</i>	Nessun metodo predefinito (SSE, dendogramma, interpretazione libera, etc.)	Nessun metodo predefinito (SSE, interpretazione libera, etc.)	BIC	Nessun metodo predefinito (SSE, interpretazione libera, etc.)
<i>Altri svantaggi</i>		Non restituisce una soluzione univoca	Deve essere parametrizzato Ha difficoltà a performare bene con delle	Deve essere parametrizzato Distingue con difficoltà profili non simili

Dalla tabella sopra riportata emerge che tutte le tecniche necessitano di dati normalizzati e solo il *Latent Profile* necessita di una distribuzione normalizzata di variabili continue e di un numero di dati di input molto limitato. Nel caso di un numero elevatissimo di dati solo il *K-means* rappresenta una soluzione viabile mentre le altre tecniche di *clustering* lavorano bene con una quantità più ridotta di dati.

Inoltre, lo studio dimostra che, tra le precedenti metodologie analizzate, quella che meglio si presta al clustering di profili residenziali, con un numero contenuto di dati, è il *clustering* gerarchico. Ciò è dovuto principalmente al fatto che i risultati ottenuti dal *clustering* gerarchico sono quelli più facilmente interpretabili ed utilizzabili.

L'interpretabilità dei risultati ottenuti è uno degli aspetti principali per la valutazione di un processo di *clustering*, è infatti importante capire come i *cluster* sono stati ottenuti, cosa contengono e cosa rappresentano gli elementi al proprio interno in modo da poter risalire alle ragioni sociali ed economiche alla base della struttura delle serie temporali in essi contenute.

## 2.3 Overview sui principali modelli di classificazione

L'obiettivo dei modelli di classificazione è quello di individuare un insieme di regole, sulla base di dati noti, che permette di predire la classe di un nuovo dato inizialmente ignoto. Il processo di classificazione fa parte di una vasta categoria di algoritmi definita *Data Mining*, nota anche come *discovery process*. Esso è un processo logico utilizzato per estrarre informazioni chiave da un set di dati in modo da poter individuare delle regole utili allo scopo dell'analisi. In generale, gli algoritmi di *Data Mining* sono suddivisi nei seguenti passaggi [22]:

- *exploration*;
- *pattern identification*;

- *deployment*.

La fase dell'*exploration* consiste nella rifinitura dei dati di partenza e nell'individuazione delle variabili più importanti per la creazione del modello in funzione dell'obiettivo dell'analisi. Successivamente, la fase del *pattern identification* ha lo scopo di identificare le regole più efficaci che generano, in termini di precisione del modello, la miglior classificazione dei dati ignoti. Infine, il *deployment* ha il compito di gestire e modificare i risultati ottenuti dal modello in base all'output desiderato.

In generale, gli algoritmi di **Data Mining** seguono tre diversi approcci in fase di apprendimento: approccio supervisionato, non supervisionato e semi-supervisionato. Nell'approccio supervisionato l'algoritmo è allenato su un set di dati ben definito di cui si conoscono le caratteristiche chiave utili al processo di estrazione di conoscenza. Le variabili sulla quale è allenato il modello possono essere numeriche o categoriche.

Dall'altro lato, i processi non supervisionati lavorano con un set di dati di cui le caratteristiche chiave sono ignote. In questo caso, l'algoritmo ha egli stesso il compito di individuare quali sono le variabili chiave che permettono di estrapolare le regole necessarie ai fini dell'estrazione di conoscenza facendo uso di processi di caratterizzazione, raggruppamento e *clustering*. Infine, i processi semi-supervisionati vengono utilizzati quando si ha a disposizione un piccolo set di dati noto e caratterizzato, insieme ad un vasto dataset di dati ignoti di cui si desidera effettuare la caratterizzazione.

I modelli di classificazione fanno parte del sottoinsieme delle tecniche supervisionate dove, per ogni singolo elemento in fase di apprendimento, sono note la classe e le variabili chiave utili alla classificazione. Il set di dati di partenza viene diviso in due parti denominate *training set* e *testing set*.

Corrispondentemente, il processo di classificazione si divide in:

- *training* del modello;

- *testing* del modello.

In fase di *training* si costruisce il modello utilizzando la parte di dati precedentemente identificate come *training set*. In questa fase, il modello creato acquisisce la conoscenza necessaria a classificare i dati in base a come essi vengono classificati nel *training set*, cioè individua le relazioni tra le variabili degli elementi e la propria classe.

In fase di *testing*, dei nuovi elementi appartenenti al set di dati di partenza, vengono inseriti nel modello privati della propria classe, lasciando al modello il compito di classificarli nuovamente. Successivamente, viene valutata la precisione del modello in funzione della nuova classe predetta e della loro classe originariamente nota attraverso delle metriche ben definite.

Gli algoritmi di classificazione più comuni sono:

- Alberi decisionali (*Decision Tree*);
- *K-Nearest Neighbor*;
- *Support Vector Machines*;
- *Naive Bayesian Classification*;
- *Neural Networks*.

### 2.3.1 Alberi decisionali

Gli alberi decisionali sono una tecnica di apprendimento automatico utilizzata per la classificazione e la regressione di dati. In generale, un albero decisionale è costituito da un insieme di nodi collegati tra loro da archi, dove ogni nodo rappresenta una domanda o una condizione che permette di prendere una decisione.

La costruzione di un albero decisionale avviene attraverso un processo iterativo che prevede la scelta della variabile di input che migliora maggiormente la separazione delle classi o la riduzione dell'errore di regressione. Tale processo prevede di suddividere il *dataset* di partenza in sottoinsiemi sempre più omogenei fino ad arrivare alla creazione di un albero completo.

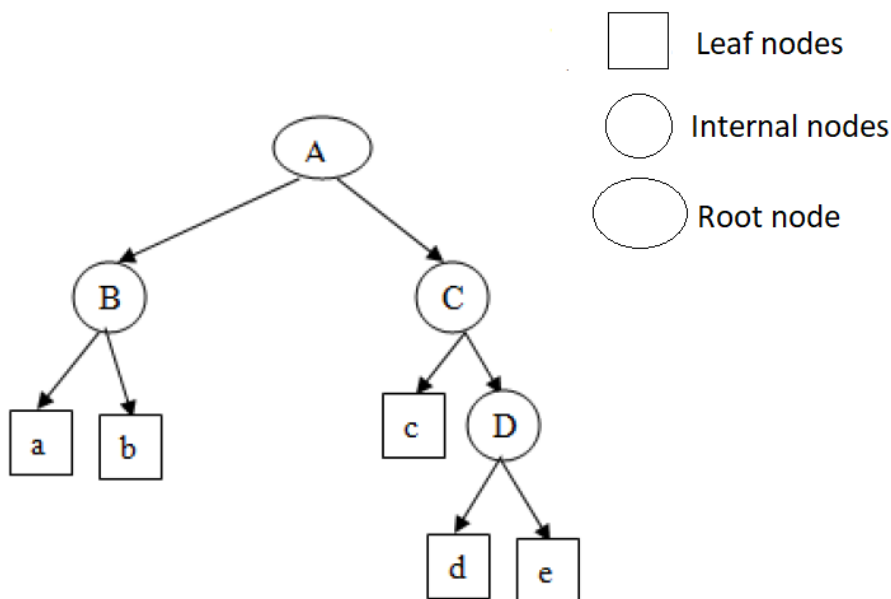
Gli alberi decisionali presentano diverse caratteristiche che li rendono una tecnica di apprendimento automatico molto utilizzata in vari settori:

- facilità di interpretazione: gli alberi decisionali sono facilmente interpretabili in quanto mostrano in maniera chiara e intuitiva le decisioni prese sulla base dei dati di input.
- scarsa dipendenza dai dati: gli alberi decisionali sono poco sensibili alle variazioni dei dati di input, a differenza di altre tecniche di apprendimento automatico.
- possibilità di gestire sia dati qualitativi che quantitativi: gli alberi decisionali possono gestire sia dati categorici che numerici.
- possibilità di gestire dataset di grandi dimensioni: gli alberi decisionali possono essere utilizzati su dataset di grandi dimensioni senza che ne venga compromessa l'efficienza.
- capacità di identificare interazioni tra variabili: gli alberi decisionali sono in grado di identificare le interazioni tra le variabili di input e la loro influenza sulla variabile di output.

Tuttavia, l'albero decisionale presenta anche alcuni svantaggi, come la tendenza all'*overfitting*, ovvero la costruzione di un albero eccessivamente complesso che rischia di adattarsi troppo al dataset di *training* e di non generalizzare bene sui dati di *test*. Inoltre, l'albero decisionale può essere sensibile alla presenza di dati rumorosi o *outliers*, che possono influire negativamente sulla qualità delle decisioni prese.

Più nel dettaglio, gli alberi decisionali consistono in una struttura ramificata

composta da nodi che si dividono in *root nodes*, ovvero nodi che non presentano alcun flusso in entrata, *internal o test nodes*, vale a dire nodi con flussi entranti ed uscenti, e *leaf nodes*, ovvero i nodi terminali (o anche detti decisionali) dell'albero. In un albero decisionale, ogni *internal node* è caratterizzato da almeno una divisione doppia del flusso in ingresso che segue la logica di una determinata funzione discreta applicata sulle variabili in ingresso (**Figura 20**).



**Figura 20** Struttura di un albero decisionale.

Per definire la classe di un elemento attraverso un albero decisionale, partendo dal *root node*, vengono percorsi in successione gli *internal nodes*, rappresentativi di regole estrapolate dal modello in fase di *training*, fino al raggiungimento di un *leaf node*. Una misura quantitativa dell'utilità di uno split all'interno degli *internal node* è rappresentata da un valore statico noto in letteratura come *information gain*, che misura quanto bene la regola utilizzata è in grado di scindere i dati all'interno del *training set* [22].

Esistono diverse tipologie di alberi decisionali utilizzate in ambito di data mining e machine learning. Di seguito ne elenchiamo alcune:

- alberi decisionali binari: questo tipo di albero decisionale presenta solo due possibili risultati per ogni nodo interno, generando così un albero a due rami. Ogni nodo interno rappresenta una condizione sul valore di un



attributo, mentre le foglie rappresentano le classi o le etichette delle istanze di dati. Questo tipo di albero è particolarmente utile per problemi di classificazione binaria.

- alberi decisionali multi-ramificati: al contrario degli alberi decisionali binari, questo tipo di albero presenta più di due rami per ogni nodo interno. Questo consente una maggior flessibilità nella modellizzazione di problemi di classificazione multi-classe, in cui le istanze di dati possono appartenere a più di due classi.
- alberi decisionali regolari: questo tipo di albero decisionale ha una struttura regolare, ovvero tutti i rami hanno la stessa lunghezza e tutti i nodi interni hanno lo stesso grado (ovvero lo stesso numero di rami). Questo tipo di albero è spesso utilizzato per problemi di classificazione in cui ogni attributo ha lo stesso peso nella definizione delle classi.
- alberi decisionali non regolari: questo tipo di albero decisionale ha una struttura irregolare, ovvero i rami possono avere lunghezze diverse e i nodi interni possono avere diversi gradi. Questo tipo di albero è particolarmente utile quando alcuni attributi hanno un peso maggiore rispetto ad altri nella definizione delle classi.
- alberi decisionali basati su regressione: questo tipo di albero decisionale viene utilizzato per problemi di regressione, ovvero per prevedere un valore numerico invece che una classe. In questo caso, le foglie dell'albero rappresentano i valori di output invece che le classi.
- alberi decisionali basati su insiemi: questo tipo di albero decisionale viene utilizzato in ambito di machine learning per la classificazione di istanze di dati che appartengono a più di una classe. Ogni foglia dell'albero rappresenta un insieme di classi, invece di una singola classe.

In generale, gli alberi decisionali sono molto flessibili e possono essere utilizzati per una vasta gamma di problemi di classificazione e regressione. Tuttavia, è importante scegliere il tipo di albero decisionale più adatto al problema specifico, in modo da ottenere i migliori risultati possibili.

### 2.3.2 K-Nearest Neighbor Classifiers (KNN)

Il *K-Nearest Neighbors* (KNN) è un algoritmo di apprendimento supervisionato utilizzato per la classificazione e la regressione. L'obiettivo dell'algoritmo è di classificare un nuovo punto dati in base alla sua vicinanza ai punti dati di addestramento, ovvero dati già classificati. In sostanza, l'algoritmo cerca i  $K$  punti dati più vicini al nuovo punto e determina la classe del nuovo punto in base alla maggioranza dei voti tra i  $K$  punti più vicini.

L'algoritmo KNN è considerato un metodo di apprendimento "pigro" perché non richiede una fase di addestramento vera e propria, ma piuttosto memorizza i dati di addestramento e li utilizza in fase di classificazione. L'algoritmo può essere utilizzato sia per problemi di classificazione che di regressione.

Il parametro  $K$  dell'algoritmo rappresenta il numero di vicini più prossimi da poter utilizzare per la classificazione o la regressione. Maggiore è il valore di  $K$ , maggiore sarà la complessità del modello e maggiore sarà la varianza dell'errore, mentre un valore di  $K$  troppo basso potrebbe portare a un'alta sensibilità al rumore nei dati di addestramento.

L'algoritmo KNN si basa sulla definizione di una funzione di distanza che misura la distanza tra due punti nello spazio delle caratteristiche. Le funzioni di distanza più comuni sono la distanza euclidea e la distanza di Manhattan. Inoltre, per migliorare le prestazioni dell'algoritmo, è possibile utilizzare tecniche di normalizzazione dei dati e di selezione delle caratteristiche.

Tra le principali caratteristiche dell'algoritmo KNN vi sono la semplicità, la flessibilità e l'adattabilità a diverse tipologie di dati. Tuttavia, l'algoritmo può

essere dal punto di vista computazionale e costoso in presenza di un grande numero di punti dati e la scelta del valore di  $K$  può essere un'operazione non banale. Inoltre, l'algoritmo può essere influenzato dalla presenza di dati rumorosi o *outliers*.

La classificazione tramite *K-Nearest Neighbor* è basata quindi sul concetto di analogia e vicinanza (**Figura 21 e 22**). Ogni elemento del *training set* viene riportato in un spazio  $n$ -dimensionale. Successivamente, quando viene inserito un nuovo elemento ignoto al modello, esso viene collocato allo stesso modo all'interno dello spazio  $n$ -dimensionale creato valutando la vicinanza agli elementi noti, determinandone quindi un campo di appartenenza. In base al campo di appartenenza assegnatoli, gli viene associata una classe. La vicinanza agli elementi noti dello spazio  $n$ -dimensionale può essere valutata, come detto in precedenza, tramite distanza euclidea (**Equazione 11**).

$$X = (x_1, x_2, \dots, x_n) \text{ e } Y = (y_1, y_2, \dots, y_n)$$

$$d(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (11)$$

In generale, questo processo di classificazione è più lento degli alberi decisionali in quanto è necessario molto tempo per il calcolo delle distanze [22]. Inoltre, a differenza della classificazione tramite albero decisionale, questo metodo attribuisce lo stesso peso a tutte le  $n$  variabili selezionate per classificare gli elementi.

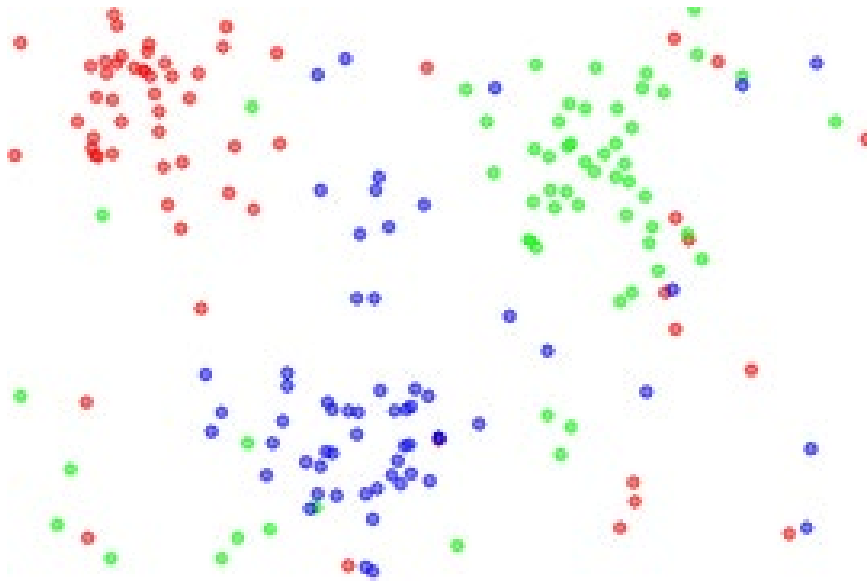


Figura 21 Dati grezzi di un dataset generico.

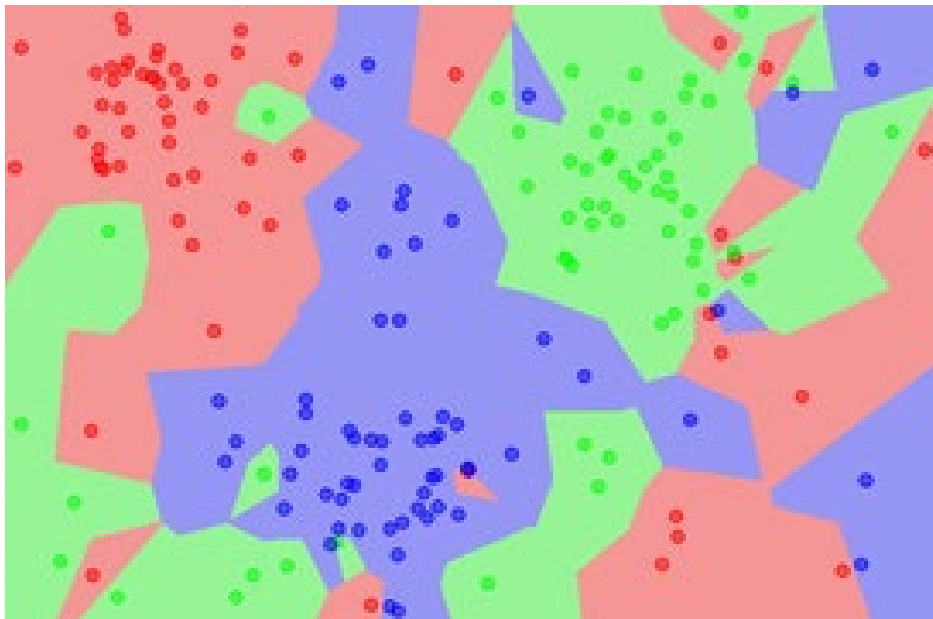


Figura 22 Esempio di classificazione tramite KNN.

### 2.3.3 Support Vector Machines

Il metodo *Support Vector Machine* (SVM) è stato introdotto per la prima volta da Vapnik [23] e si è dimostrato essere molto efficace per problemi regressivi, di classificazione e per il *pattern recognition*. È considerato un buon classificatore per via delle sue elevate performance senza l'aggiunta di "conoscenza" ulteriore da instillare al modello. Il *Support Vector Machines* è un algoritmo di apprendimento

supervisionato utilizzato principalmente per la classificazione binaria. L'obiettivo del SVM è trovare un iperpiano che separi i dati in due classi in modo ottimale.

L'ipotesi alla base del SVM è che i dati di input siano rappresentati in uno spazio di dimensione superiore (spesso definito come spazio delle *feature*), in cui è più facile separare i dati in due classi. L'ipotesi dell'iperpiano si traduce nel trovare il miglior piano possibile che separi i dati in modo che il margine tra i due insiemi di punti (uno per classe) sia massimizzato. Il margine rappresenta la distanza tra l'iperpiano e i punti di ogni classe più vicini all'iperpiano stesso.

Il metodo SVM inizia con la definizione della funzione di decisione, che determina a quale classe appartiene un nuovo punto dati in base alla sua posizione rispetto all'iperpiano trovato. L'obiettivo è trovare la funzione di decisione ottimale che minimizzi l'errore di classificazione sui dati di training e massimizzi il margine. Questa funzione di decisione può essere rappresentata come una combinazione lineare delle funzioni kernel, che trasformano i dati in uno spazio delle feature di dimensione superiore.

Il SVM ha diverse caratteristiche importanti, tra cui la capacità di generalizzare bene, anche con un numero relativamente piccolo di dati di training. Inoltre, l'uso di funzioni kernel consente al SVM di gestire dati che non sono linearmente separabili nello spazio originale. Tuttavia, l'addestramento del SVM può essere dal punto di vista computazionale costoso per grandi quantità di dati, e la scelta del kernel e dei suoi parametri può influire notevolmente sulla performance dell'algoritmo.

#### **2.3.4 Naive Bayesian Classification**

La *Naive Bayesian Classification* (NBC) è un algoritmo di classificazione supervisionata che utilizza il teorema di Bayes per predire l'appartenenza di un'osservazione ad una classe. L'algoritmo si basa sull'assunzione di indipendenza tra le variabili esplicative, ovvero che ogni variabile influisca indipendentemente sulla variabile di risposta.

L'idea alla base del NBC è di calcolare la probabilità condizionata

dell'appartenenza di un'osservazione ad una classe, dato il valore delle variabili esplicative. Questa probabilità condizionata viene calcolata per ogni classe e l'osservazione viene assegnata alla classe con la probabilità condizionata più alta.

Il calcolo delle probabilità condizionate avviene attraverso la stima delle probabilità a priori delle classi e delle distribuzioni di probabilità delle variabili esplicative condizionate alle classi. Nel caso delle variabili continue, si assume di solito una distribuzione normale, mentre per le variabili categoriche si utilizza la distribuzione multinomiale.

Una volta stimati i parametri del modello, il *Naive Bayesian Classification* è in grado di classificare nuove osservazioni in modo efficiente e accurato. L'algoritmo è particolarmente adatto per *dataset* di grandi dimensioni e con un elevato numero di variabili esplicative.

Tra le caratteristiche principali del *Naive Bayesian Classification*, vi sono la semplicità di implementazione, la velocità di calcolo e la capacità di gestire sia dati continui che categorici. Tuttavia, l'assunzione di indipendenza tra le variabili esplicative può portare ad una riduzione dell'accuratezza del modello in presenza di correlazioni tra le variabili.

## 2.4 Literature Review

L'argomento CER è molto dibattuto nel panorama scientifico. Pierpaolo Garavaso et al. analizzano nel loro studio una CER generica secondo il modello "energy hub", ovvero un modello con diversi sistemi di produzione dove l'energia viene prodotta e convertita in altre forme per soddisfare il fabbisogno energetico di altre strutture [8]. Lo studio presenta una metodologia basata sul *Standardized Matrix Modelling* con lo scopo di identificare le migliori strategie di gestione del sistema sotto un sistema incentivante pilotato da politiche energetiche ed economiche. I risultati mostrano come, attraverso la metodologia proposta, sia possibile ottenere un risparmio sui costi energetici pari al 28% rispetto ad un *benchmark scenario* da loro proposto.

Lo studio di Emi Minghui Gui et al. considera la potenziale importanza delle “Clean Energy Communities” (CEC) nella trasformazione dei regimi sociotecnici attuali che sostengono i nostri sistemi energetici centralizzati, verso un futuro più distribuito e decentralizzato. In questo lavoro viene spiegata la struttura centralizzata, distribuita e decentralizzata, integrata nei sistemi energetici e nei mercati dell’energia attuali, vengono analizzate le dinamiche a lungo termine e i possibili percorsi della de-centralizzazione del sistema energetico, e l’evoluzione delle comunità, utilizzando una struttura esplorativa, attingendo da teorie di transizione, studi sull’innovazione e concetti di reti sociali.

Ciò include l’analisi delle diverse forme di CEC, tra cui le centrali virtuali, il commercio *peer-to-peer*, le *microgrid* e i progetti energetici su scala comunitaria, emergenti in diverse giurisdizioni. Questa analisi suggerisce che i percorsi di transizione a bassa emissione di carbonio saranno vari, guidati da contesti sociali, tecnologici e organizzativi diversi, e plasmati dai processi di cambiamento istituzionale e dall’interazione con le amministrazioni.

Lo studio sostiene inoltre che gli imprenditori sociali e tecnologici e le aziende di servizi pubblici, operanti in un ambiente sempre più definito dai consumatori di energia, dai gruppi industriali e dagli attori politici, dovranno essere innovativi nella scelta di direzioni strategiche, decisioni di investimento, nelle politiche di acquisizione delle risorse energetiche per raggiungimento dei loro obiettivi in termini di transizione energetica [45].

Lo studio di A. Caramizaru et al. fornisce una panoramica delle attività, dell’organizzazione e delle implicazioni delle comunità energetiche come enti partecipanti ad un sistema energetico ampio e complesso [46]. Inoltre, si propone di identificare dei percorsi in funzione delle future implicazioni politiche e delle iniziative di ricerca. Inoltre, lo studio esplora i risultati analizzando una vasta gamma di attività, forme organizzative e benefici sociali delle comunità energetiche e il loro contributo all’espansione delle energie rinnovabili.

Gjorgievski et al. dimostrano nel loro lavoro di ricerca che le comunità energetiche consentono ai consumatori di perseguire congiuntamente i propri obiettivi economici, ambientali e sociali individuali e collettivi, contribuendo

contemporaneamente alla decarbonizzazione del sistema energetico. In considerazione dell'interesse crescente per questo settore, questo lavoro si propone di migliorare la comprensione delle disposizioni normative, dei progetti tecnici e degli impatti delle comunità energetiche. Tali argomenti vengono discussi in relazione ai diversi attori e ai loro ruoli.

In aggiunta, il documento rivede gli aspetti tecnici della progettazione dei vari sistemi energetici locali, tenendo conto degli obiettivi dei membri della comunità energetica e degli attori esterni. La letteratura da loro esaminata viene analizzata sul piano dei metodi, degli obiettivi di modellizzazione e dei vincoli utilizzati nel processo di progettazione. Infine, il loro lavoro di ricerca quantifica gli impatti economici, ambientali, tecnici e sociali delle comunità energetiche, rivede gli indicatori numerici utilizzati per quantificare questi impatti e fornisce una discussione critica dei risultati [47].

L'articolo di J. Jung et al propone una tecnica per la pianificazione e il design ottimale di sistemi ibridi di energia rinnovabile per applicazioni di *microgrids*. Viene utilizzato un *Distributed Energy Resources Customer Adoption Model* distribuite (DER-CAM) per determinare la dimensione e il tipo ottimali di risorse energetiche distribuite (DER) e i loro programmi operativi per un sistema di distribuzione tipo. Utilizzando i risultati del DER-CAM, nello studio viene effettuata una valutazione delle prestazioni elettriche del circuito di distribuzione. Vengono presentati anche i risultati delle analisi riguardanti i benefici economici dell'utilizzo delle configurazioni ottimali per i DER interni al sistema. In questo lavoro viene utilizzata la rete elettrica del campus del Brookhaven National Laboratory (BNL) per dimostrare l'efficacia del loro approccio.

I risultati da loro esposti mostrano che le analisi tecniche ed economiche condotte sui sistemi ibridi di energia rinnovabile sono essenziali per l'utilizzo efficiente delle risorse energetiche rinnovabili per le applicazioni di *microgrids*.

Lo studio di Volpato et al. affronta insieme tre aspetti principali che influenzano la convenienza di aggregazioni di consumatori e produttori come le CER: la complementarità tra la generazione e la domanda di diversi prosumer, il



criterio di allocazione dei costi operativi delle comunità energetiche e l'applicazione dei programmi di risposta alla domanda [48]. Il loro lavoro quantifica il peso relativo di questi aspetti utilizzando il *Mixed-Integer Linear Programming* per minimizzare i costi operativi delle comunità di energia rinnovabile e dei cittadini, dove i *prosumer* sono connessi alla rete come entità singole o separate. I risultati ottenuti permettono di identificare linee guida generali per il funzionamento economico ottimale delle comunità energetiche.

Anche il tema del *pattern recognition* in ambito residenziale e commerciale è stato ampiamente sviscerato nel panorama scientifico. Pérez-Chacón et al. propongono una metodologia di clustering per dataset di elevata dimensione attraverso quattro metriche di validazione differenti [26], ovvero DB-Dunn index, Davies-Bouldin Index, Within Set Sum of Square Errors (WSSSE) Index e il DB-Silhouette Index.

La valutazione del numero di cluster ottimale individuato da tutte e quattro le metriche porta all'identificazione del  $k$  migliore attraverso ad un'analisi in frequenza dei primi tre  $k$  ottimali definiti dagli indici precedentemente elencati. Inoltre, lo studio identifica il *k-Means* come l'algoritmo migliore per l'individuazione dei profili tipologici in presenza di un numero elevato di dati.

Maher Azaza et al. propongono una metodologia per l'identificazione dei *pattern* di consumo tipologici da dati registrati da *smart meter*, in relazione ad alcune metriche di consumo [27]. Inoltre, in questo studio viene valutata di due differenti algoritmi di *clustering*, ovvero il *Hierarchical Clustering* e il *Self-Organising Map* (SOM), concludendo che il primo algoritmo identifica dei profili di carico simili per forma, ma molto diversi per metriche di consumo (variabilità del carico e *responsibility factor*, definito come il rapporto tra il carico di un singolo componente nel periodo di carico massimo dell'aggregato, diviso per il picco massimo del medesimo componente in senso assoluto), mentre il secondo algoritmo risulta essere più efficace nel raggruppare utenze con consumi diversi, ma simili per metriche di consumo.

S. Ghosh et al. propongono l'utilizzo di differenti modelli di classificazione

per l'identificazione dei carichi elettrici in ambito residenziale [28]. Nel loro studio vengono valutate le performance di diversi algoritmi di classificazione supervisionati, in particolare vengono utilizzati modelli quali il *Logistic Regression Classifier*, *Decision Tree Classifier*, *Multi-Layer Perceptron Classifier*, *k-Nearest Neighbors Classifies* (KNN), *Gradient Boosting Classifier* e *Random Forest Classifier*.

Qiaochu Wang et al. utilizzano un *clustering* a due step per il *load pattern recognition* nel campo degli edifici combinando un *Dynamic Time Warping Clustering* (DTWC) ad un *K-means* [29]. Nel loro studio si dimostra che utilizzare un approccio di questo tipo permette di individuare in modo più preciso e agevole i profili di carico tipologici, partendo prima dalla forma dei profili, con il primo metodo, per poi passare alla distinzione per intensità di carico, con il secondo.

Nello studio di June Young Park et al., invece, vengono valutate tre metodologie di *clustering* per l'identificazione degli andamenti di carico tipologici [30]. Nello specifico vengono analizzati il *K-means clustering*, il *Bisecting k-Means Clustering* e il *Gaussian Mixture Model Based Clustering*. La scelta dell'algoritmo ottimale per il caso studio specifico è stata fatta valutando la similarità degli elementi in ogni *cluster* generato, la distanza tra i centroidi dei *cluster* generati e il Calinski-Harabaz Index. I risultati dimostrano che il *K-means* è, tra i tre, l'algoritmo che meglio identifica le diverse tipologie di consumo.

Analogamente, nello studio di M. Fang et al vengono valutati differenti algoritmi di *clustering* per il *load pattern identification* [32], tra cui il *K-means*, *Hierarchical Clustering*, *Gaussian Mixture Model*, *C-vine Copula Mixture Model* e *R-vine Mixture Model*, in combinazione con diverse tipologie di classificatori, tra cui il KNN, *Classification and Regression Tree*, *Naive Bayes Classifier* (NB) e *Random Forest Classifier*. Lo studio dimostra come la combinazione tra l'*R-vine Mixture Model* e il *Random Forest Classifier* performa meglio, in materia di classificazione, rispetto alle restanti combinazioni. Inoltre, per l'identificazione del numero *k* di *cluster* ottimale, si fa uso di quattro metriche diverse, che sono il Sum of Squared Error (SSE), il Davies-Boulding Index (DBI), il Mean Index Adequancy (MIA) e il Clustering Disperios Indicator (CDI).

Anche nello studio di Tureczek, A.M. et al. vengono utilizzate le stesse metriche di valutazione del numero ottimale di *cluster* nell'ambito del riconoscimento dei profili tipologici per le utenze residenziali, a partire da dati rilevati da *smart meter*, con l'aggiunta del Silhouette Index (SI), per poi utilizzare un *clustering* di tipo *K-means*. Così come in alcuni degli studi precedentemente elencati, anche nel lavoro di M. Batou et al. si fa uso di un *K-means clustering* ma in forma rivisitata, per il *load estimation* di infrastrutture locali.

A fronte della letteratura revisiona in tema CER e identificazione dei profili tipologici per utenze generiche, emerge che non sono presenti, in ambito scientifico, studi che ricostruiscono il profilo orario di utenze residenziali con il fine di valutare i principali output energetici delle comunità, su scala oraria, a partire dai soli valori di consumo mensile. Ciò permetterebbe di simulare le CER, in fase di studio preliminare, con una maggiore precisione, anche in assenza di dati orari, attraverso tecniche di ricostruzione del dato non intrusive.

Di fatti, la metodologia proposta nei capitoli successivi pone come obiettivo la ricostruzione dei profili orari, a partire dai consumi in fasce facilmente reperibili dalle bollette elettriche, volta al calcolo dell'energia condivisa all'interno di una CER, parametro fondamentale per il sostentamento economico dell'aggregato in contesti di valorizzazione ed incentivazione della condivisione dell'energia.

### 3 Struttura del Database utilizzato

Nel presente lavoro di tesi viene utilizzato il dataset pubblico del progetto Low Carbon London, guidato da UK Power Networks e sviluppato tra il novembre 2011 e febbraio 2014 [43]. Il database del progetto, reso pubblico per scopi di ricerca, consiste nell'insieme delle letture del consumo elettrico di 5567 utenti residenziali londinesi, espresso in kWh e misurato ogni 30 minuti, in formato CSV, per un totale di 167 milioni di righe per un peso di circa 10 GB.

Il progetto Low Carbon London (LCL) è stato un programma di ricerca del valore di £28 milioni che si è svolto dall'inizio del 2011 alla fine del 2014 e che è stato finanziato da consumatori energetici attraverso il fondo di rete a basso livello di carbonio di Ofgem. Il progetto è stato ideato per indagare l'impatto di una vasta gamma di tecnologie a basso livello di carbonio sulla rete di distribuzione dell'elettricità di Londra.

È in questo contesto che si è svolto il primo esperimento nel settore residenziale britannico, relativo alla tariffazione dinamica del costo dell'energia elettrica. Questo esperimento ha coinvolto molteplici famiglie nell'area di Londra, di cui 1.122 sono state sottoposte ad una tariffa del costo dell'energia elettrica sperimentale (dToU) che è entrata in vigore per l'intero anno 2013.

La sperimentazione è stata condotta da un partenariato di organizzazioni:

- UK Power Networks: la società che gestisce la rete di distribuzione dell'elettricità a Londra e il principale partner del programma;
- Imperial College London: progettazione dell'esperimento e analisi dei risultati;
- EDF Energy: fornitore di energia elettrica e supervisione della tariffa dToU;
- Siemens: implementazione del database e delle comunicazioni;
- Logica (ora CGI): implementazione del sistema di misurazione intelligente.

All'interno del dataset sono presenti due gruppi di utenti residenziali, il

primo è quello sottoposto a tariffe sperimentali del costo dell'energia elettrica nel 2013, mentre il secondo sottogruppo, formato da circa 4500 utenti, ha continuato a consumare l'energia elettrica abitualmente mantenendo il piano tariffario precedente.

Per gli scopi specifici del presente lavoro, sono stati utilizzati i dati relativi al secondo sotto-gruppo, ovvero agli utenti alla quale non è stato proposto l'utilizzo di un piano tariffario diverso da quello abituale.

Il *database* utilizzato è composto da una prima colonna, contenente il nome delle utenze in un formato standard, da una seconda colonna contenente un "tag" che identifica gli utenti sottoposti a piani tariffari sperimentali, una terza contenente data e ora delle rilevazioni eseguite ed un'ultima colonna contenenti i valori di consumo espressi in kWh (**Tabella 5**).

**Tabella 5** *Struttura del database nella sua forma grezza.*

<i>Utente</i>	<i>Tag</i>	<i>DateTime</i>	<i>Consumo [kWh]</i>
<i>MAC00002</i>	Std	2012-10-12 00:30:00	0.0002
<i>MAC00002</i>	Std	2012-10-12 01:00:00	0.0004
<i>MAC00002</i>	Std	2012-10-12 01:30:00	0.005
...	...	...	...

Come detto in precedenza, il file contiene le letture del consumo elettrico di utenze residenziali a partire dal 2011 fino al 2014. La popolazione di utenti presenti nel *database*, per i quattro anni di rilevazione dei dati è riportata in **Tabella 6**.

**Tabella 6** *Popolazione del database per anno di attività.*

	<b>2011</b>	<b>2012</b>	<b>2013</b>	<b>2014</b>
N. Utenti	354	4433	4411	4065
N. mesi con almeno una rilevazione per utente	2	12	12	2
% Utenti sul totale degli utenti coinvolti	7,9%	99,8%	99,3%	91,5%

Il 2011 è un anno molto povero dal punto di vista delle rilevazioni dei dati. Esso presenta un numero di utenti pari al 7,9% rispetto agli utenti totali complessivamente coinvolti nell'intera durata del progetto e un numero di mesi con almeno una rilevazione per utente pari a 2.

Il 2012, al contrario, è l'anno più ricco di rilevazioni. Esso contiene un numero di utenti pari al 99,8% rispetto al numero di totale di utenti e contiene rilevazioni dei per tutti i mesi dell'anno.

Analogamente, il 2013 presenta un numero di utenti leggermente inferiore a quello del 2012 ma lo stesso numero di mesi nei quali è stata effettuata almeno una rilevazione per tutti gli utenti interessati.

Il 2014, differentemente dalle altre sezioni del database, contiene un numero elevato di utenti (91,5% del totale) ma sono pochi i mesi nei quali è presente almeno una rilevazione per utente (2 mesi).

A fronte di ciò, la parte di database che meglio si presta alla metodologia descritta nei capitoli è la sezione corrispondente al 2012.

Più nel dettaglio, la sezione del database d'interesse è strutturata secondo quanto riportato in **Tabella 7**.

*Tabella 7 Caratteristiche del dataset riferite al 2012.*

<i>Mese</i>	1	2	3	4	5	6	7	8	9	10	11	12
<i>N. Utenti</i>	494	713	1028	1714	2945	3375	3792	3801	3997	4406	4418	4433
<i>N. Utenti con rilevazioni complete (*)</i>	346	483	0	1003	1610	1977	3259	3697	3710	3894	4281	3952
<i>% Utenti con rilevazioni complete</i>	70,0%	67,7%	0,0%	58,5%	54,7%	58,6%	85,9%	97,2%	92,8%	88,3%	96,9%	89,1%

(\*) Numero di utenti di cui è noto il consumo elettrico orario per tutte le ore del mese

I mesi in cui sono presenti più utenti, e quindi più dati, sono quelli relativi ai mesi di finali dell'anno, e gli utenti con rilevazioni complete, ovvero quegli utenti che sono campionati per tutte le ore del mese, sono compresi tra lo 0,0% (riferito al mese di marzo) e il 97,2% (riferito al mese di agosto).

Marzo è un mese particolarmente critico dal punto di vista della rilevazione del dato, perché non presenta utenze con rilevazioni complete. Questo comportamento è ricorrente anche per le sezioni *dataset* relative agli altri anni, pertanto, si assume che quel periodo possa coincidere con il periodo di manutenzione degli *smart meter* utilizzati.

Per questo motivo, dall'analisi svolta in seguito viene completamente escluso il mese di marzo.

## 4 Metodologia

Il presente lavoro di tesi ha come obiettivo la presentazione di una metodologia volta a stimare il profilo orario di utenze residenziali a partire dai consumi mensili in fasce, facilmente reperibili dalle bollette elettriche, attraverso tecniche non intrusive.

Essendo geograficamente localizzati in Inghilterra, più precisamente nella città di Londra, i consumi mensili vengono valutati secondo la suddivisione di consumo in fasce della compagnia London Power (LOND), schematizzata in **Tabella 8**, considerata valida per tutti i giorni dell'anno.

**Tabella 8** *Fasce di consumo di London Power per le utenze residenziali londinesi [42].*

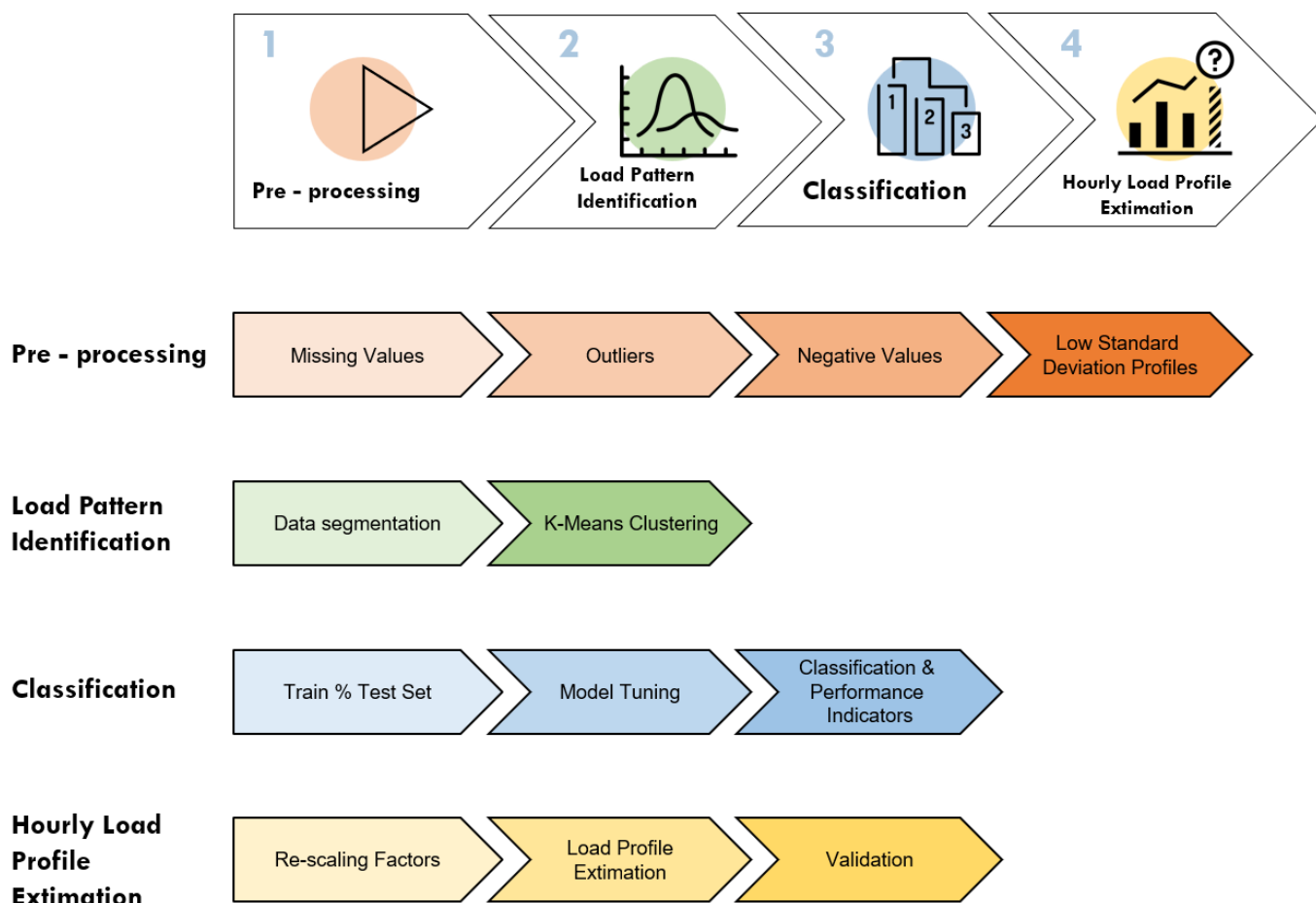
<i>Fascia di consumo</i>	<i>Intervallo temporale</i>
Red	11:00-14:00 & 16:00-19:00
Amber	07:00-11:00 & 14:00-16:00 & 19:00-23:00
Green	00:00-07:00 & 23:00-24:00

La metodologia proposta ha come inizio l'elaborazione dei dati selezionati in precedenza attraverso un ampio *pre-processing*, articolato su diverse scale temporali (oraria, giornaliera e mensile); prosegue con l'identificazione delle principali tipologie di consumo degli utenti attraverso un *clustering* di tipo *K-means*, classifica le utenze residenziali per tipologia di consumo attraverso i consumi mensili attraverso *Random Forest*, ovvero un modello supervisionato di machine learning; ed infine ricostruisce il profilo orario delle utenze analizzate per stimare l'energia condivisa al livello singolo ed aggregato (**Figura 23**).

Nel dettaglio, la metodologia si compone di:



- Pre-processing:
  - Individuazione e sostituzione dei valori mancanti (*missing values*) e dei valori negativi;
  - Individuazione e sostituzione degli *outliers* puntuali;
  - Identificazione ed eliminazione dei profili giornalieri infrequenti;
  - Identificazione ed eliminazione dei profili a basso consumo mensile e bassa deviazione standard;
  
- Load Pattern identification:
  - Divisione stagionale de dataset;
  - Scoposizione in giorni della settimana, sabati e domeniche;
  - Individuazione dei profili tipici tramite il *K-means clustering*;
  
- Classification:
  - Scelta del *dataset* di *train* e *test*;
  - Tuning del modello di classificazione;
  - Valutazione delle performance di classificazione;
  - Classificazione delle utenze per tipologia, a partire dai consumi mensili nelle diverse fasce orarie;
  
- Hourly Load Profile estimation:
  - Individuazione delle caratteristiche del consumo annuo, su scala mensile, tramite classificatore;
  - Stima del profilo annuo orario tramite un processo di *rescaling* di tipo *data-driven*.



**Figura 23** Metodologia per la stima del profilo orario.

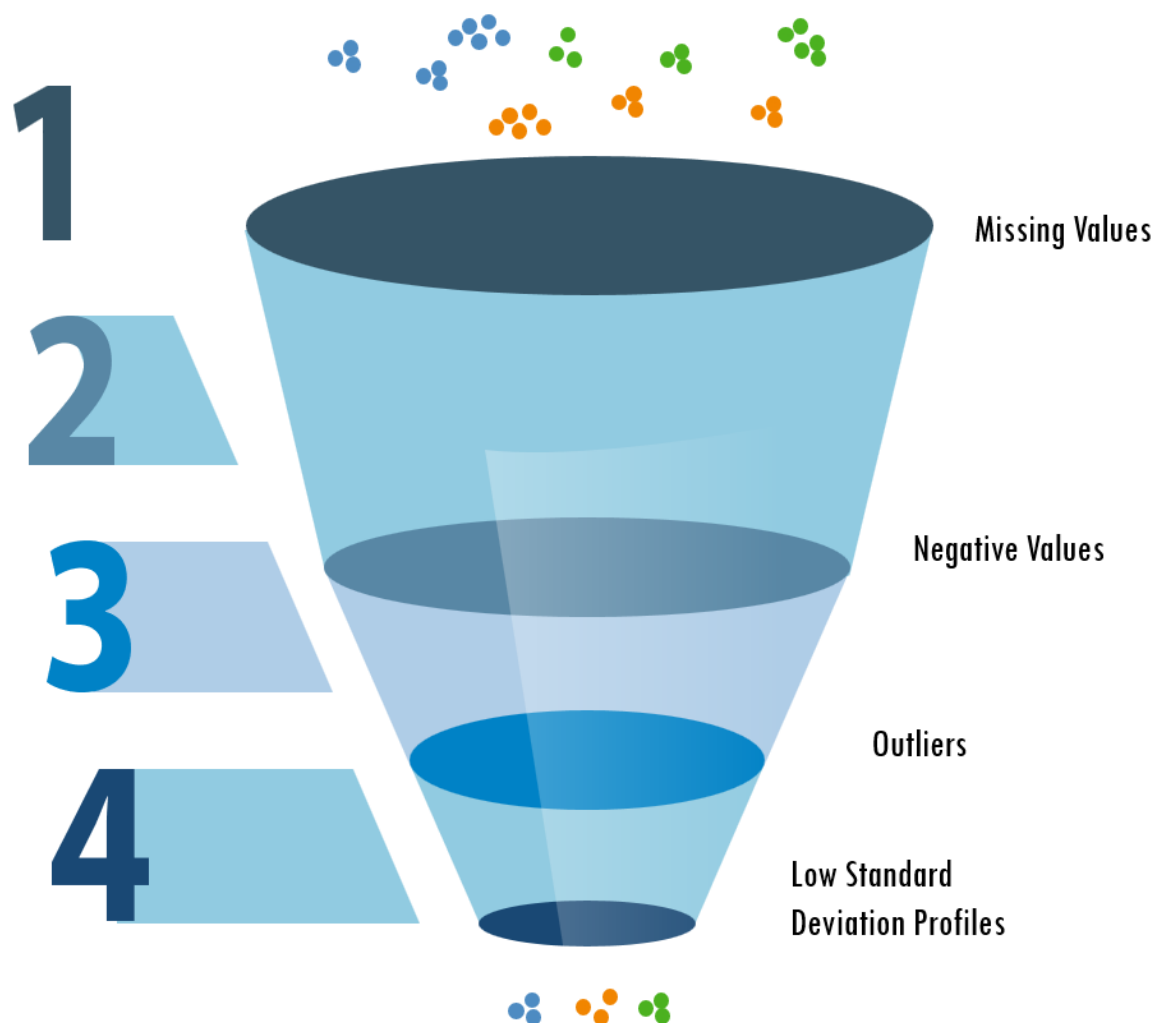
La metodologia proposta si propone di colmare una lacuna esistente nella disponibilità di dati orari nell'ambito residenziale, una problematica che affligge inevitabilmente anche il settore delle Comunità Energetiche Rinnovabili, dal momento che la valutazione dei principali parametri energetici che generano valore economico, come l'energia condivisa, è condotta a livello orario.

La capacità di stimare con precisione il consumo orario di un utente sulla base dei soli dati di consumo mensile, attraverso tecniche non intrusive, consente di effettuare un calcolo più accurato dell'energia condivisa e di individuare, in fase di simulazione preliminare, la configurazione ottimale che garantisce il bilanciamento opportuno tra produzione e consumo, assicurando la stabilità economica della CER. La metodologia proposta previene la realizzazione di progetti di CER sulla carta economicamente sostenibili, ma instabili nel breve e medio periodo, a causa di una valutazione imprecisa dei consumi dei propri membri e, di conseguenza, dell'energia condivisa generata, nonché di parte degli

introiti economici.

## 4.1 Pre-processing

Il *pre-processing* del *database* analizzato è articolato nei passaggi illustrati in **Figura 24**.



**Figura 24** Schema logico delle *pre-processing* adottato.

Il primo passaggio è quello dell'identificazione e ricostruzione dei dati mancanti (*missing values*). Essi vengono ricostruiti tramite un'interpolazione lineare, con un numero massimo di valori consecutivi pari a 3. I valori mancanti che non ricostruiti da questo passaggio vengono elaborati in modo differente, ovvero attraverso la ricostruzione del dato tramite il valore medio mensile.

Il secondo consiste nell'individuazione ed eliminazione dei valori negativi (*negative values*). Essi rappresentano generalmente una non corretta rilevazione del dato da parte delle apparecchiature utilizzate e una volta eliminati, vengono ricostruiti allo stesso modo dei valori mancanti.

Il terzo passaggio riguarda l'individuazione degli *outliers*. Essi vengono processati su diverse scale temporali:

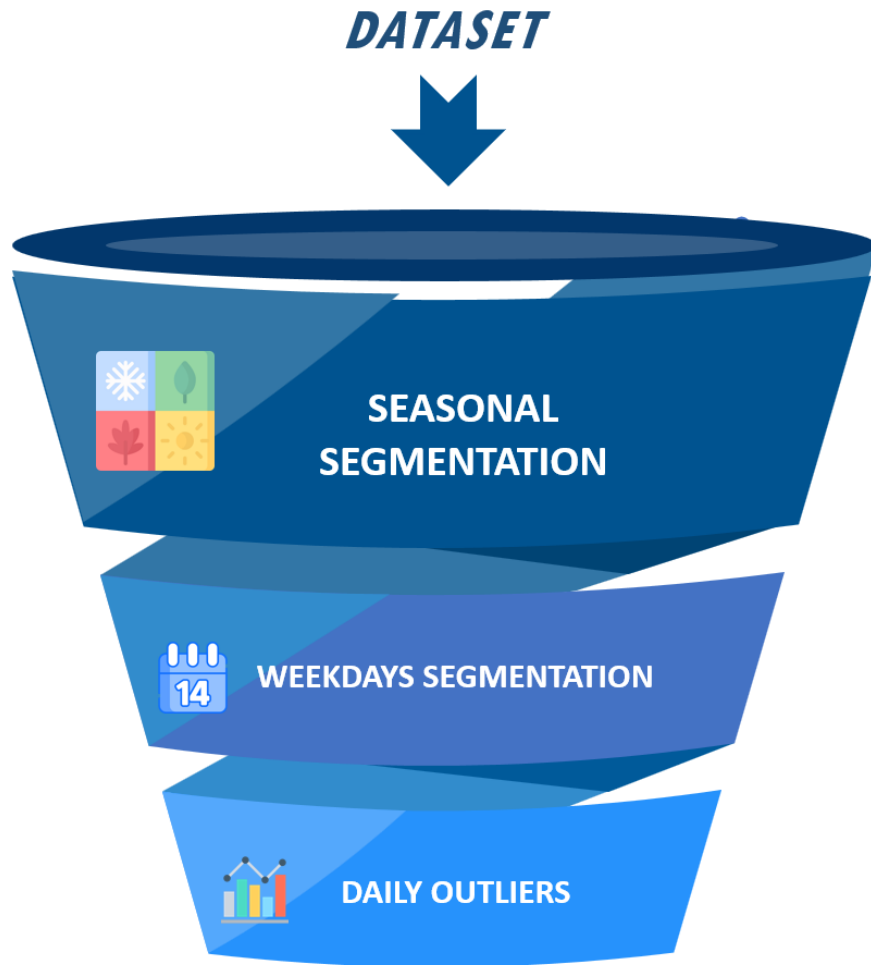
- Scala oraria - *outliers* puntuali (*hourly-scale outliers*);
- Scala giornaliera - profili di carico infrequenti e/o relativi ad un numero basso di utenze;
- Scala Mensile - profili a basso consumo.

Gli *outliers* puntuali al livello orario vengono identificati attraverso l' "interquartile method", in forma rivisitata.

Normalmente i percentili utilizzati per l'individuazione dei valori Q1 e Q3 sono il 25% e il 75%. Nel caso specifico, vista la struttura dei dati a disposizione, si opta per dei percentili diversi, pari a 10% e 90%. L'impiego di percentili diversi nella procedura previene l'effetto "colabrodo", ovvero previene l'eliminazione di un numero di dati eccessivamente elevato tale da generare degli intervalli di mancanza del dato molto estesi e molto frequenti, perdendo molte delle informazioni contenute nel *dataset*.

Gli *outliers* su scala giornaliera, ovvero i profili infrequenti e relativi ad un numero basso di utenti, vengono identificati utilizzando parte della metodologia sviluppata da Piscitelli et al. [39]. Si fa uso, quindi, di un *clustering* su scala giornaliera con il fine di identificare quei sotto-gruppi poco popolati e con pochi utenti al loro interno, attraverso l'utilizzo di valori soglia definiti in seguito.

Tale processo è applicato su diverse sottosezioni del *dataset*, che viene quindi segmentato per stagioni e giorni della settimana come rappresentato in **Figura 25**, utilizzando di fatto un approccio nidificato già proposto da diversi studi di settore [31, 39, 41].



**Figura 25** Divisione del dataset per l'individuazione degli outlier giornalieri.

Nello specifico, viene applicato un *k-Means clustering* tramite algoritmo "Lloyd" sui profili di carico normalizzati sul massimo giornaliero, secondo l' **Equazione 12**, e vengono considerati profili infrequenti e relativi ad un numero basso di utenti quei profili all'interno di *cluster* che contengono meno del 5% dei profili giornalieri analizzati o meno del 5% degli utenti complessivamente presenti nella sezione del *dataset* in questione (**Figura 26**).

$$\text{Consumo normalizzato}_h = \frac{\text{Consumo}_h}{\max(\text{Consumo}_{h,\text{giornaliero}})} \quad (12)$$

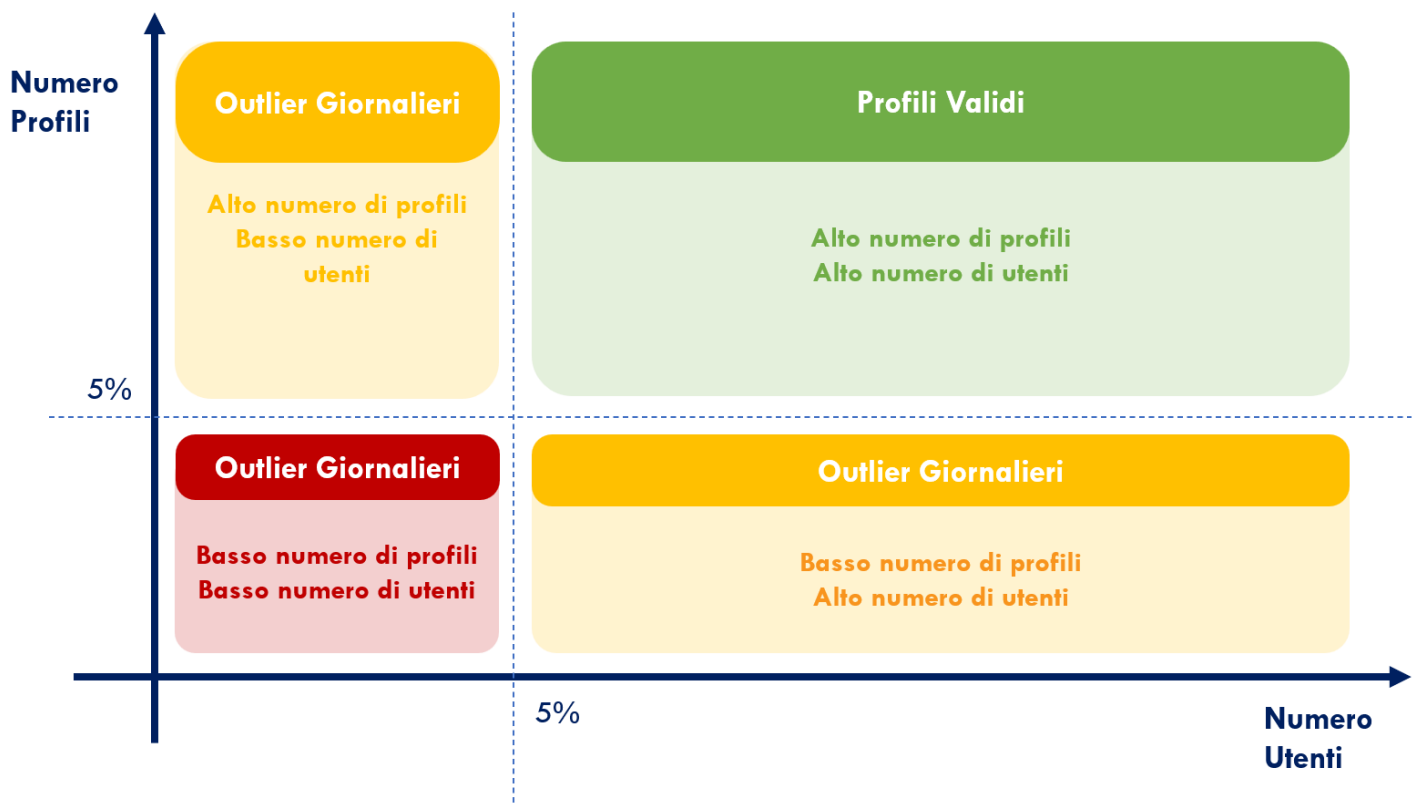


Figura 26 Identificazione dei cluster infrequenti e poco rappresentativi.

L'algoritmo "Lloyd" è utilizzato per dividere un insieme di dati in gruppi diversi in funzione della loro somiglianza. L'obiettivo dell'algoritmo è quello di minimizzare la somma delle distanze quadrate tra ogni punto e il centroide del *cluster* di appartenenza.

L'algoritmo "Lloyd" segue i seguenti passaggi:

- a. inizializzazione dei centroidi: si selezionano  $k$  punti casuali dall'insieme di dati come centroidi iniziali per i  $k$  *cluster*.
- b. assegnazione dei punti ai centroidi: ogni punto viene assegnato al cluster il cui centroide è il più vicino.
- c. aggiornamento dei centroidi: il centroide di ogni *cluster* viene aggiornato come la media di tutti i punti assegnati al *cluster*.
- d. Ripetizione dei passaggi b e c: i passaggi b e c vengono ripetuti finché non

si verifica una convergenza.

L'algoritmo è sensibile alla scelta iniziale dei centroidi, poiché essa influisce notevolmente sulla convergenza dell'algoritmo e sulla qualità del *clustering* effettuato.

I parametri ulteriori necessari alla definizione univoca del processo di *clustering* sono sintetizzati in **Tabella 9**.

**Tabella 9** Parametri aggiuntivi, necessari per il processo di *clustering* di tipo *k*-Means.

<i>Nome</i>	<i>Descrizione</i>	<i>Valore</i>
k	Numero di cluster	Definito tramite DBI Index
nitermax	Numero massimo di iterazioni	700
nstart	Set di centroidi presi casualmente	100

Il numero ottimale di *clusters* viene definito, per ogni sottosezione del *dataset*, tramite la valutazione dell'indice di Davies-Bouldin (DBI Index). L'indice di Davies-Bouldin è una metrica utilizzata per valutare la qualità di un algoritmo di *clustering* ed è definito come la media del rapporto tra la somma delle varianze all'interno di ogni *cluster* e la distanza tra il centroide di ogni *cluster* e quello dei *cluster* più simili [39]. Formalmente, per un insieme di  $n$  oggetti da suddividere in  $k$  *cluster*, il DBI Index è definito secondo l'**Equazione 13**.

$$DBI = \frac{1}{k} * \sum_{i \neq j} \frac{\max(R_i + R_j)}{d(C_i, C_j)} \quad (13)$$

dove  $i$  e  $j$  sono gli indici dei cluster,  $R_i$  è la varianza all'interno del *cluster*  $i$ ,  $C_i$  è il

centroide del *cluster*  $i$ , e  $d(C_i, C_j)$  è la distanza euclidea tra i centroidi dei *cluster*  $i$  e  $j$ . Il valore massimo viene calcolato su tutte le possibili coppie di *cluster*, esclusa la coppia con i valori di  $i$  e  $j$  coincidenti.

In sostanza, l'indice di Davies-Bouldin misura la somiglianza tra i *cluster* in base alla loro distribuzione interna e alla loro grado di separazione. Un valore basso dell'indice indica una buona suddivisione. Al contrario, un valore alto dell'indice significa che la suddivisione effettuata non è ottimale e che i *cluster* generati non sono facilmente distinguibili e che gli elementi al loro interno sono notevolmente diversi tra loro.

Successivamente, vengono processati gli *outliers* su scala mensile. Essi vengono identificati utilizzando il consumo elettrico medio delle utenze domestiche rilevato dal governo britannico nel 2020 [40]. Il database nazionale fornisce il valore medio di consumo elettrico annuo, pari a 3880 kWh. Il valore soglia utilizzato, che individua le utenze con un consumo basso, è calcolato secondo l'Equazione 14.

$$Electric\ Consumption\ threshold\ value = \frac{3880}{12} * 0,10 \quad (14)$$

Infine, il quarto ed ultimo passaggio consiste nell'individuazione ed eliminazione dei profili caratterizzati da una deviazione standard bassa, ovvero quei profili che presentano una forma del carico relativamente piatta.

## 4.2 Clustering dei profili medi mensili

Dopo aver eliminato gli *outliers* dal *dataset* di partenza, si procede al *clustering* dei profili medi mensili tramite un *k-Means clustering*. Anche in questo caso, la scelta del numero ottimale di *cluster* è stata fatta utilizzando il DBI Index.

Ripercorrendo la metodologia adottata per l'individuazione e l'eliminazione degli *outliers* su scala giornaliera, vengono eliminati i *cluster* contenenti un numero basso di profili e/o un numero basso di utenti al loro interno. A differenza del caso precedente, i valori soglia utilizzati sono pari al 10% per entrambe le



variabili (Figura 27).

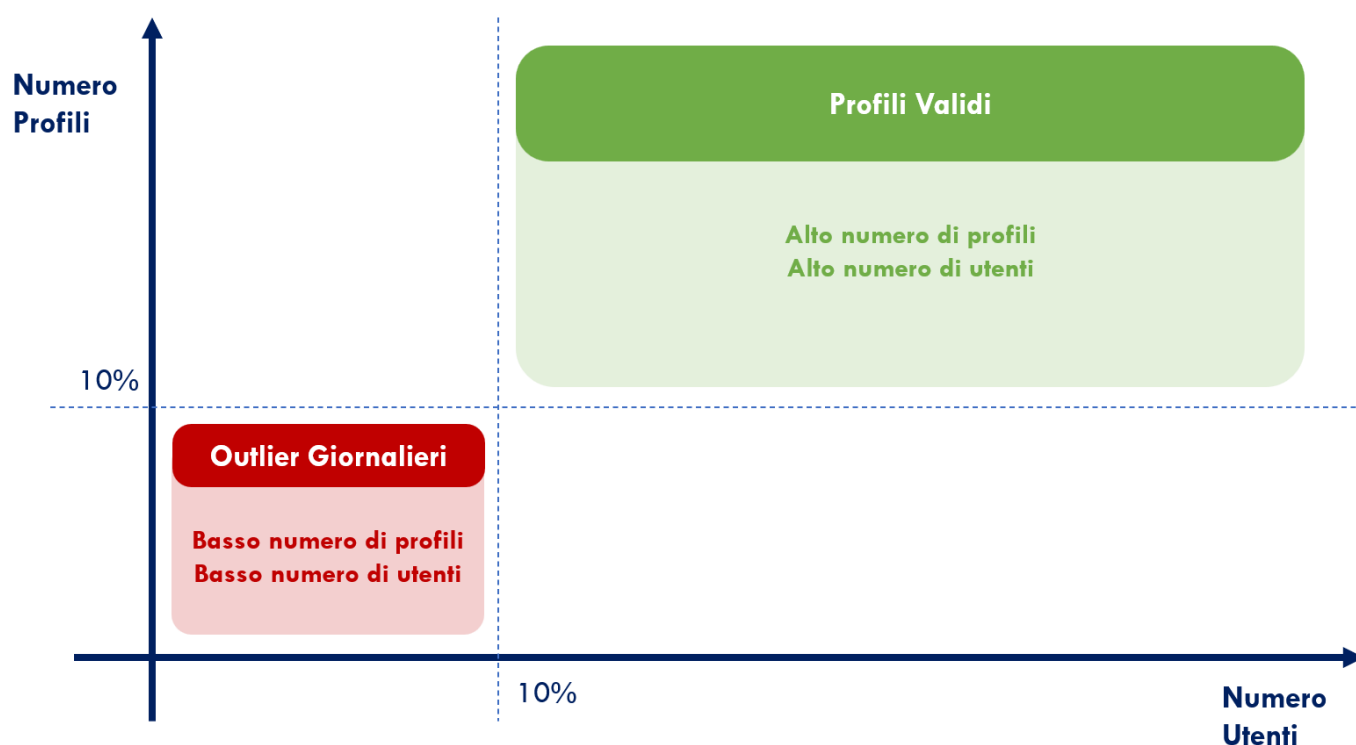


Figura 27 Identificazione profili anomali su scala mensile.

Trattandosi di un *K-means clustering* i parametri utilizzati in fase di implementazione dell'algoritmo sono sintetizzati in **Tabella 10**.

Tabella 10 Parametri del clustering di tipo *k-Means* su scala mensile.

Nome	Descrizione	Valore
k	Numero di cluster	Definito tramite DBI Index
nitermax	Numero massimo di iterazioni	700
nstart	Set di centroidi presi casualmente	400

A differenza del *clustering* sui profili giornalieri durante il processo di eliminazione degli *outliers*, viene utilizzato l'algoritmo Hartigan e Wong.

Esso è un algoritmo di *clustering* di tipo *k-Means* che è stato proposto come

alternativa all'algoritmo di *Lloyd* ed è in grado di fornire un clustering più robusto e stabile [52, 53].

Esso è composto dai seguenti passaggi:

- 1) Inizializzazione dei centroidi: come nell'algoritmo di *Lloyd*, si selezionano  $k$  punti casuali dall'insieme di dati come centroidi iniziali per i  $k$  cluster.
- 2) Assegnazione dei punti ai centroidi: ogni punto viene assegnato al *cluster* il cui centroide è il più vicino.
- 3) Calcolo della somma delle distanze quadrate: per ogni *cluster*, viene calcolata la somma delle distanze quadrate tra ogni punto assegnato al *cluster* e il centroide del *cluster*.
- 4) Tentativo di spostamento di un punto da un *cluster* all'altro: per ogni punto, viene valutata la possibilità di spostarlo da un *cluster* all'altro, e viene calcolata la nuova somma delle distanze quadrate per ogni *cluster* coinvolto nel tentativo di spostamento.
- 5) Confronto tra le somme delle distanze quadrate: se la somma delle distanze quadrate risulta inferiore dopo lo spostamento del punto, il punto viene effettivamente spostato al nuovo *cluster*, altrimenti viene mantenuto nel cluster di appartenenza attuale.
- 6) Aggiornamento dei centroidi: dopo ogni spostamento di punto, i centroidi di tutti i *cluster* vengono aggiornati.
- 7) Ripetizione dei passaggi 3-6: i passaggi 3-6 vengono ripetuti fino a quando non si verifica una convergenza.

A differenza dell'algoritmo di *Lloyd*, l'algoritmo di *Hartigan* e *Wong* prevede il tentativo di spostamento di ogni punto da un cluster all'altro, in modo da verificare se il nuovo assegnamento dei punti ai *cluster* possa portare a una

riduzione della somma delle distanze quadrate, ovvero a un *clustering* migliore [52, 53] Inoltre, l'algoritmo di *Hartigan e Wong* può essere eseguito più volte con diversi centroidi iniziali, in modo da migliorare la qualità del *clustering*.

L'algoritmo di *Hartigan e Wong* è più oneroso dal punto di vista computazionale dell'algoritmo di *Lloyd*, ma può fornire un *clustering* più stabile e robusto, soprattutto quando il numero di cluster  $k$  è elevato e i centroidi iniziali sono scelti in modo non ideale [52, 53]

Analogamente a quanto fatto per l'identificazione degli *outliers* su scala giornaliera, nel processo corrente viene normalizzato il consumo sul massimo mensile.

Al termine del processo di *clustering* vengono individuati  $n$  profili tipologici, caratteristici delle utenze residenziali analizzate. Le diverse tipologie di carico individuate vengono utilizzate in seguito per allenare un modello di classificazione supervisionato con il fine di creare un sistema in grado di assegnare una tipologia di consumo ad un utente in funzione dei propri consumi mensili.

### 4.3 Implementazione del modello di classificazione

Il processo di implementazione del modello di classificazione è complesso e si articola nei seguenti passaggi:

- scelta del modello di classificazione;
- scelta delle variabili con le quali allenare il modello;
- scelta del dataset di *train* e *test*;
- tuning del modello in fase di *train*;
- valutazione della precisione del modello in fase di *training* e *testing*.

La scelta del modello di classificazione è una delle fasi più importanti del processo. Secondo lo studio di Mengqiu Fan et al., il modello che meglio si lega ad un clustering di tipo *k-Means* per i processi di classificazione è il *Random Forest* [33]. Per questo motivo, si decide di implementare un *Random Forest* nell'ambiente

di sviluppo *R* tramite il pacchetto “RandomForest”.

Per allenare il modello in modo opportuno, è necessario individuare le variabili su scala mensile da utilizzare. Ripercorrendo la metodologia proposta da Piscitelli et al., tra le variabili impiegate nel processo di apprendimento [39] si utilizzano i valori riportati in **Tabella 11**.

**Tabella 11** Variabili utilizzate per allenare il modello comuni a Piscitelli et al.

<i>Nome</i>	<i>Descrizione</i>	<i>Unità di misura</i>
Amber/Total	Valore della fascia di consumo “Amber” sul consumo mensile totale	[-]
Red/Total	Valore della fascia di consumo “Red” sul consumo mensile totale	[-]
Green/Total	Valore della fascia di consumo “Green” sul consumo mensile totale	[-]
Amber/Red	Valore della fascia di consumo “Amber” sul valore della fascia di consumo “Red”	[-]
Green/Amber	Valore della fascia di consumo “Green” sul valore della fascia di consumo “Amber”	[-]
Green/Red	Valore della fascia di consumo “Green” sul valore della fascia di consumo “Red”	[-]

In aggiunta, si decide di utilizzare degli ulteriori parametri con lo scopo di migliorare la precisione del modello. Nello specifico vengono utilizzate anche le variabili riportate in **Tabella 12**.

**Tabella 12** Ulteriori variabili utilizzate nel processo di apprendimento del modello.

<i>Nome</i>	<i>Descrizione</i>	<i>Unità di misura</i>
Green/Total - corrected	Valore della fascia di consumo "Green" sul consumo mensile totale, corretto secondo il fattore di correzione $f$	[-]
Green/Red - corrected	Valore della fascia di consumo "Green" sul valore della fascia di consumo "Red" corretto, secondo il fattore di correzione $f$	[-]
Green/Amber – corrected	Valore della fascia di consumo "Green" sul valore della fascia di consumo "Amber", corretto secondo il fattore di correzione $f$	[-]
Amber/Red - corrected	Valore della fascia di consumo "Amber" sul valore della fascia di consumo "Red", corretto secondo il fattore di correzione $f$	[-]
Red/Total – corrected	Valore della fascia di consumo "Green" sul consumo mensile totale, corretto secondo il fattore di correzione $f$	[-]
Amber/Total – corrected	Valore della fascia di consumo "Amber" sul consumo mensile totale, corretto secondo il fattore di correzione $f$	[-]
Green	Valore della fascia di consumo "Green"	[kWh]
Amber	Valore della fascia di consumo "Amber"	[kWh]
Red	Valore della fascia di consumo "Red"	[kWh]
Total	Consumo mensile totale	[kWh]

Il fattore di correzione  $f$  è valutato per ogni singolo parametro, e corrisponde al

rapporto tra le ore mensili delle fasce orarie messe a confronto, come rappresentato nell'Equazione 15.

$$f = \frac{\text{Ore fascia}_x}{\text{Ore fascia}_y} \quad (15)$$

Nel caso in cui uno dei termini del rapporto sia il consumo totale mensile, viene utilizzato il numero di ore complessivamente presenti in quel mese.

Utilizzare il fattore di correzione permette di trasferire al modello le informazioni relative alla durata temporale delle fasce di consumo, informazione altrimenti non utilizzabile in maniera diretta.

Successivamente, vengono scelti il *dataset* di *train* e *test*, ovvero quelle sottosezioni del *dataset* utilizzate per allenare e testare il modello. Nello studio di Piscitelli et al. è stato utilizzato l'85% del *dataset* per la fase di *train*, e il restante 15% per il *test* [39]. Nello loro lavoro i dati sono stati scelti in maniera casuale con il vincolo di avere utenti diversi tra *dataset* di *train* e *test*.

Nel presente lavoro le caratteristiche del *dataset* di *train* e *test* sono schematizzate in **Tabella 13**.

**Tabella 13** Caratteristiche del *dataset* di *train* e *test*.

<i>Nome del dataset</i>	<i>% rispetto all'intero dataset</i>	<i>Descrizione</i>
Train	70%	Viene utilizzato il 70% del <i>dataset</i> globalmente utilizzato, selezionato in maniera casuale
Test	30%	Viene utilizzato il restante 30% del <i>dataset</i> globalmente utilizzato, selezionato in maniera casuale

Si decide di utilizzare una percentuale minore per il *dataset* di *training* rispetto alla metodologia proposta da Piscitelli et al. per evitare di incorrere nel fenomeno dell'*overfitting*, ovvero l'eccessivo adattamento del modello sui dati con il quale viene allenato, memorizzandoli invece di generalizzare le regole apprese per la previsione di nuovi dati.

A questo punto, dopo aver individuato il modello, le variabili con cui allenarlo e il *dataset* di *train* e *test*, è necessario stabilire gli iperparametri che determinano un modello ottimale in termini di precisione nel processo di classificazione.

In un *Random Forest*, gli iperparametri principali sono sostanzialmente tre:

- numero di alberi generati;
- numero di variabili utilizzate ad ogni *split*;
- elementi minimi nei nodi foglia.

Il numero di alberi generati in fase di implementazione del *Random Forest* è uno dei parametri più importanti. Nell'articolo di Leo Breiman, l'inventore del *Random Forest*, si discutono diversi aspetti dell'algoritmo, tra cui il numero di alberi da utilizzare. Breiman afferma che il numero di alberi in un *Random Forest* dipende dalla dimensione del *dataset*, dal tempo disponibile per la costruzione del modello e dal *trade off* tra accuratezza e velocità [54].

Breiman suggerisce inoltre che, in generale, il numero di alberi può essere scelto in modo arbitrario, ma afferma che, in pratica, un numero di alberi compreso tra qualche centinaio e qualche migliaio può fornire buoni risultati per la maggior parte dei problemi. In aggiunta, sottolinea inoltre che l'accuratezza del modello tende a migliorare all'aumentare del numero di alberi, ma che, dopo un certo punto, l'aumento del numero di alberi non porta ad alcun miglioramento significativo dell'accuratezza [54].

Il numero di variabili utilizzate ad ogni *split* indica il numero di elementi necessari alla creazione di una regola che determina una generica scissione binaria che effettua l'albero ad ogni passaggio. In generale, utilizzare un numero di variabili di *split* diverso nella creazione dell'albero può portare a performance

del modello significativamente differenti.

Infine, il numero di elementi minimi nei nodi foglia rappresenta il valore limite minimo di elementi presenti in un nodo terminale del modello. Tale valore è strettamente legato alla dimensione dell'albero generato. Per un numero di elementi minimo nei nodi foglia basso, si ottengono alberi molto estesi, contenenti molti split. Al contrario, per un numero alto di questo parametro, si ottengono alberi meno estesi e con un numero di split limitato.

Per trovare la combinazione degli iperparametri che ottengono la precisione migliore, si fa uso dei pacchetti "tuneRanger" e "mlf". Nello specifico si effettuano numerose iterazioni nelle quali vengono valutate le diverse combinazioni degli iperparametri, fino al raggiungimento di un numero di iterazioni massimo prestabilito.

Infine, per definire in maniera univoca il modello, è necessario definire una serie di altri parametri, sintetizzati in **Tabella 14**.

**Tabella 14** Parametri aggiuntivi necessari per la definizione del modello.

<i>Parametro</i>	<i>Valore</i>	<i>Descrizione</i>
importance	Impurity/impurity_corrected/permutation	Indica la metrica con la quale viene valutata l'importanza di ogni variabile a supporto del modello di classificazione. Per "impurity" e "impurity_corrected" si intende la valutazione del Gini index.
split rule	Gini/Extratrees/Hellinger	Regola per la definizione delle variabili migliori in uno split node



---

		Indica la metrica con la quale vengono valutati i modelli ad ogni iterazione.
measure	acc/auc/brier/ber/bac	<ul style="list-style-type: none"> <li>- acc = accuracy</li> <li>- auc = area under the curve</li> <li>- brier = Brier score</li> <li>- ber = balanced error rate</li> <li>- bac = balanced accuracy</li> </ul>

---

A valle della definizione di tutti i parametri, il processo di *tuning* del modello estrapola gli iperparametri che garantiscono una precisione maggiore, in termini di classificazione, ed essi vengono utilizzati per la creazione del modello definitivo.

#### 4.4 Ricostruzione del profilo orario

La ricostruzione del profilo orario consiste nel passaggio dai valori di consumo mensile e dalla tipologia di consumo identificata dal classificatore, agli dei valori di consumo orario simulati.

Tale operazione richiede un passaggio intermedio, ovvero la definizione e il calcolo del fattore di scala  $p$ , definito secondo la metodologia utilizzata da Piscitelli et al. e rielaborato secondo le **Equazioni 16 e 17** [39].

$$p = \frac{Q_{red}}{q} \quad (16)$$

$$q = T \left( \sum_{11}^{14} q_i + \sum_{16}^{19} q_i \right) \quad (17)$$

dove:

- $Q_{red}$  è il consumo mensile in fascia di consumo “Red”;
- $T$  = lunghezza dell’intervallo temporale della fascia di consumo “Red”;
- $q_i$  = consumo normalizzato nell’arco temporale “Red”, ovvero tra le 11:00 e le 14:00 e tra le 16:00 e le 19:00.

Per il fattore di scala si decide di utilizzare il valore di consumo in fascia “Red” come riferimento perché coinvolge generalmente i periodi di consumo più elevati per le utenze residenziali, e perché copre un arco temporale di 6 ore, corrispondente a circa il 50% delle ore in cui le utenze residenziali presentano tendenzialmente consumi elevati, ovvero tra le 08:00 e le 20:00.

In definitiva, il profilo orario si ottiene moltiplicando il fattore  $p$  per il valore normalizzato del profilo relativo alla tipologia di consumo individuata tramite il processo di clustering e classificazione, per il mese  $n$ -esimo. Il profilo così ottenuto rappresenta il consumo, al livello orario, per tutti i giorni del mese considerato (**Equazione 18**).

$$\text{Consumo orario}_h = \text{Consumo normalizzato}_h * p \quad (18)$$

L’effetto della ricostruzione del profilo orario viene valutato in termini di errore relativo sull’energia condivisa sia per singolo utente che al livello aggregato. L’energia condivisa viene calcolata secondo l’**Equazione 19**.

$$\text{Energia Condivisa}_x = \begin{cases} \text{Produzione}_x & \text{se } \text{Consumo}_x > \text{Produzione}_x \\ \text{Consumo}_x & \text{se } \text{Produzione}_x > \text{Consumo}_x \end{cases} \quad (19)$$

La produzione, ovvero l’energia prodotta dagli impianti presenti nella CER, viene simulata al livello orario tramite il tool PVGIS, nell’ipotesi di avere un impianto fotovoltaico di riferimento situato nella periferia di Londra.

Il tool PVIGS è costituito da un dispositivo di simulazione dell’irradiazione solare e della temperatura, che viene utilizzato per simulare le condizioni ambientali in cui l’impianto fotovoltaico opererà. Inoltre, il sistema comprende un inverter e un simulatore di rete elettrica, che consentono di verificare il

funzionamento dell'impianto nelle diverse condizioni di carico. La simulazione avviene attraverso l'utilizzo di algoritmi di calcolo avanzati, che tengono conto delle caratteristiche dell'impianto fotovoltaico, dell'inverter e della rete elettrica, nonché delle condizioni ambientali in cui l'impianto è installato.

Per la valutazione dell'errore relativo al livello di singolo utente, l'impianto fotovoltaico di riferimento utilizzato ha una taglia pari a 20 kW, mentre al livello aggregato il confronto viene fatto con un impianto di taglia superiore, pari a 200 kW.

## 5 Risultati ottenuti

In questo capitolo vengono discussi i risultati ottenuti dalla metodologia proposta. Il capitolo è diviso nelle seguenti sezioni:

- *Pre-processing* – risultati ottenuti dall’elaborazione dei dati grezzi;
- *Clustering* – profili tipologici ottenuti e caratteristiche;
- *Classificazione* – precisione e caratteristiche del modello;
- *Ricostruzione del profilo orario* – profili orari ottenute e metriche di valutazione.

### 5.1 Pre-processing

I risultati della metodologia proposta in riferimento al *pre-processing* dei dati grezzi riguardano la sostituzione ed eliminazione dei valori mancanti, l’identificazione, sostituzione ed eliminazione degli *outliers* e dei valori negativi. Come spiegato nel Capitolo 4, gli *outliers* vengono valutati al livello puntuale, giornaliero e mensile.

Vengono identificati e sostituiti tramite interpolazione lineare zero valori mancanti. Conseguentemente, non sono stati identificati archi temporali con valori mancanti consecutivi più ampi di 3 ore, pertanto non è stata necessaria una ricostruzione dei dati più approfondita. Non sono stati individuati valori negativi, mentre sono vengono identificati 24017 *outliers* puntuali, sostituiti tramite “interquartile method” (Tabella 15).

**Tabella 15** Valori mancanti e outliers puntuali trovati.

<i>Parametro</i>	<i>Valore</i>	<i>Percentuale sul totale</i>
Missing Values	0	0%

Valori negativi	0	0%
Outliers Puntuali	24017	1%

Gli *outliers* al livello giornaliero segue la divisione del dataset illustrata nel Capitolo 4. Il dataset di partenza viene quindi diviso per stagioni e per giorni della settimana. La divisione stagionale è fatta come quanto illustrato in **Tabella 16**.

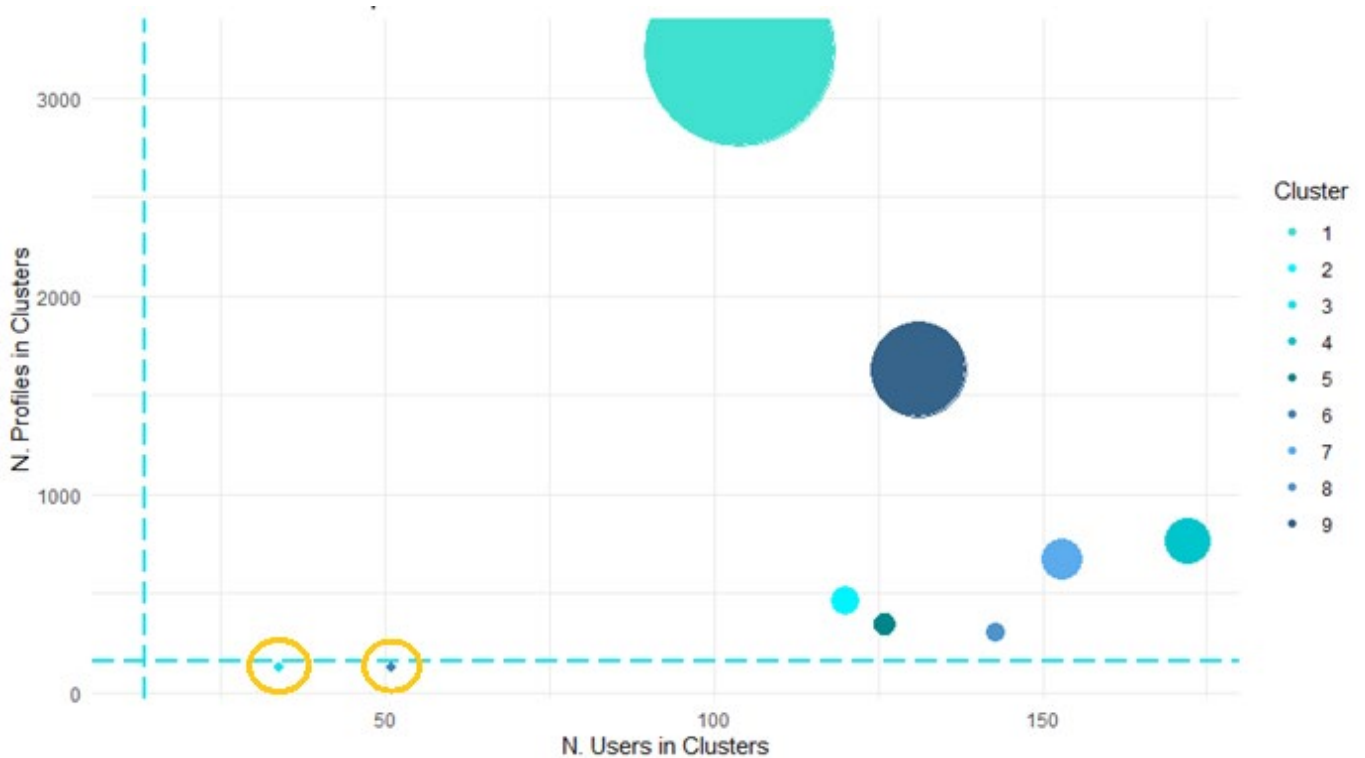
**Tabella 16** *Divisione stagionale del dataset.*

<i>Stagione</i>	<i>Mesi</i>
Estate	Giugno, Luglio, Agosto
Inverno	Dicembre, Gennaio, Febbraio
Primavera	Marzo, Aprile, Maggio
Autunno	Settembre, Ottobre, Novembre

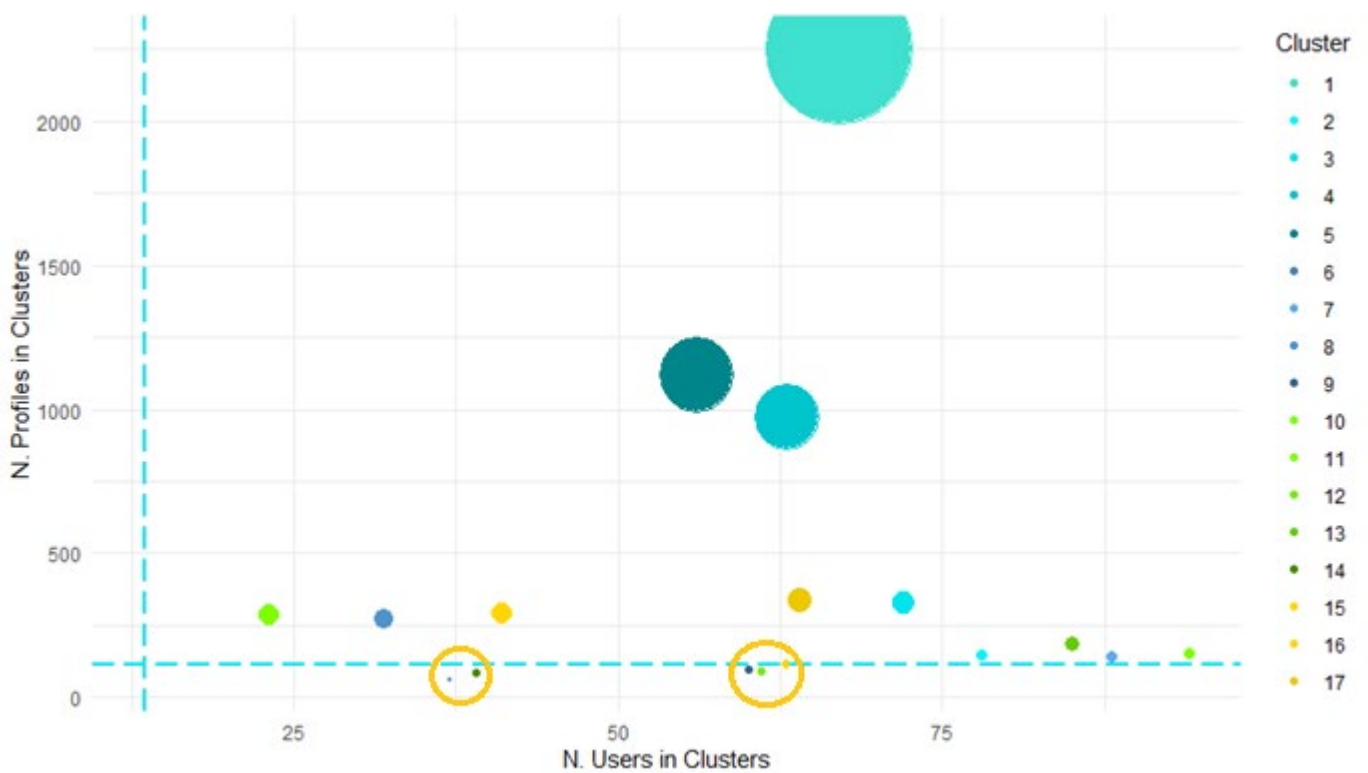
Attraverso un *clustering* di tipo *k-Means* e delle metriche di individuazione

dei profili giornalieri infrequenti e poco rappresentativi discusse nel Capitolo 4, vengono dunque identificati ed eliminati i cluster contenenti i profili di carico infrequenti e relativi ad un numero basso di utenti, su scala giornaliera.

Di seguito vengono illustrati alcuni dei passaggi più significativi dell'identificazione degli *outliers* giornalieri per le diverse stagioni per alcuni dei giorni della settimana analizzati (Figura 28 e 29).



**Figura 28** Eliminazione dei profili anomali giornalieri per le domeniche autunnali.



**Figura 29** Eliminazione dei profili anomali giornalieri per i martedì primaverili.

Ogni cerchio corrisponde ad un *cluster* e la dimensione dei cerchi indica la percentuale dei profili sul totale, all'interno del *cluster* stesso. Le figure sopra riportate mettono in risalto un aspetto fondamentale, ovvero la numerosità di cluster ottenuti dal DBI Index, che sottolinea la notevole diversità di consumo tra utenti anche se nello stesso giorno della settimana e nella stessa stagione.

Successivamente vengono riportati tutti i profili identificati come *outliers* dal processo di individuazione ed eliminazione su scala giornaliera, divisi per stagioni (**Figure 30, 31, 32 e 33**).

### Profili Anomali - Autunno

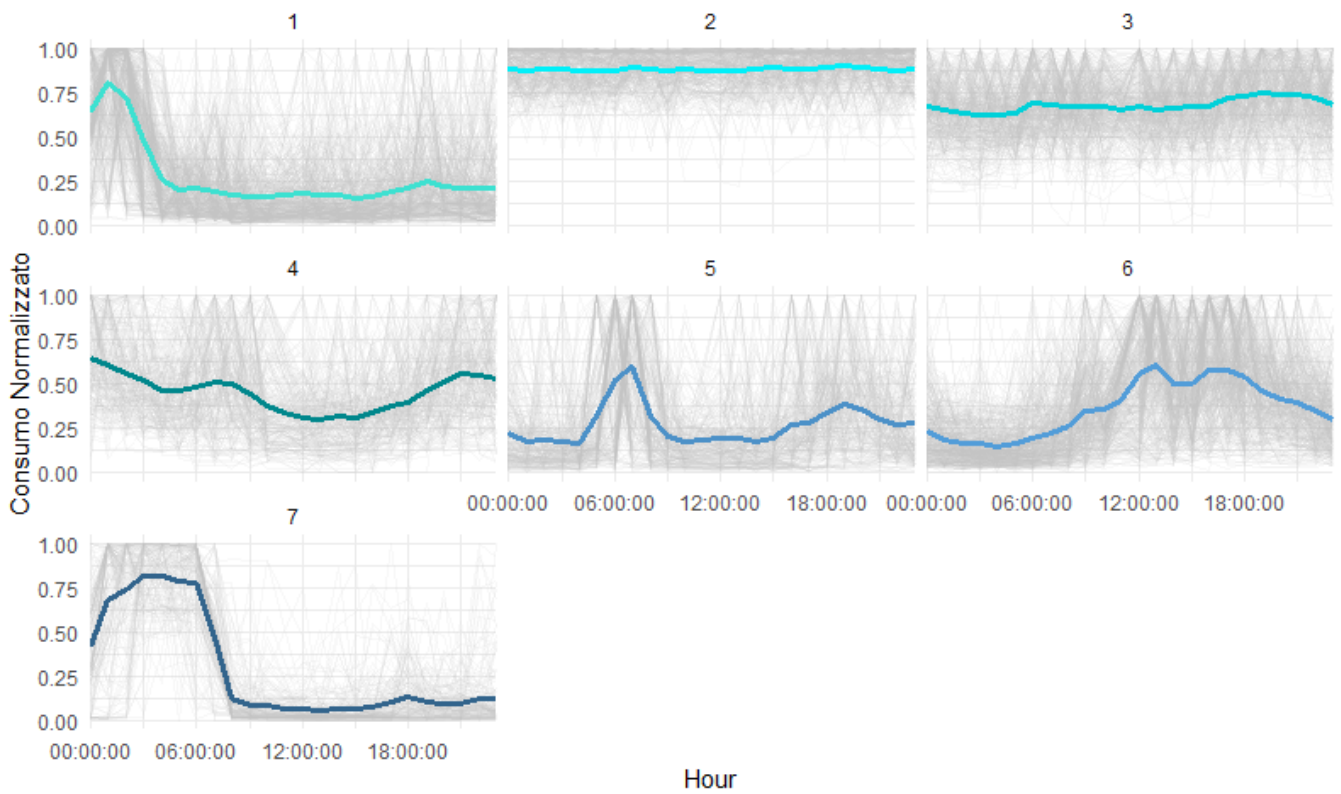



Figura 30  Profili giornalieri anomali per la stagione autunnale.

### Profili Anomali - Inverno

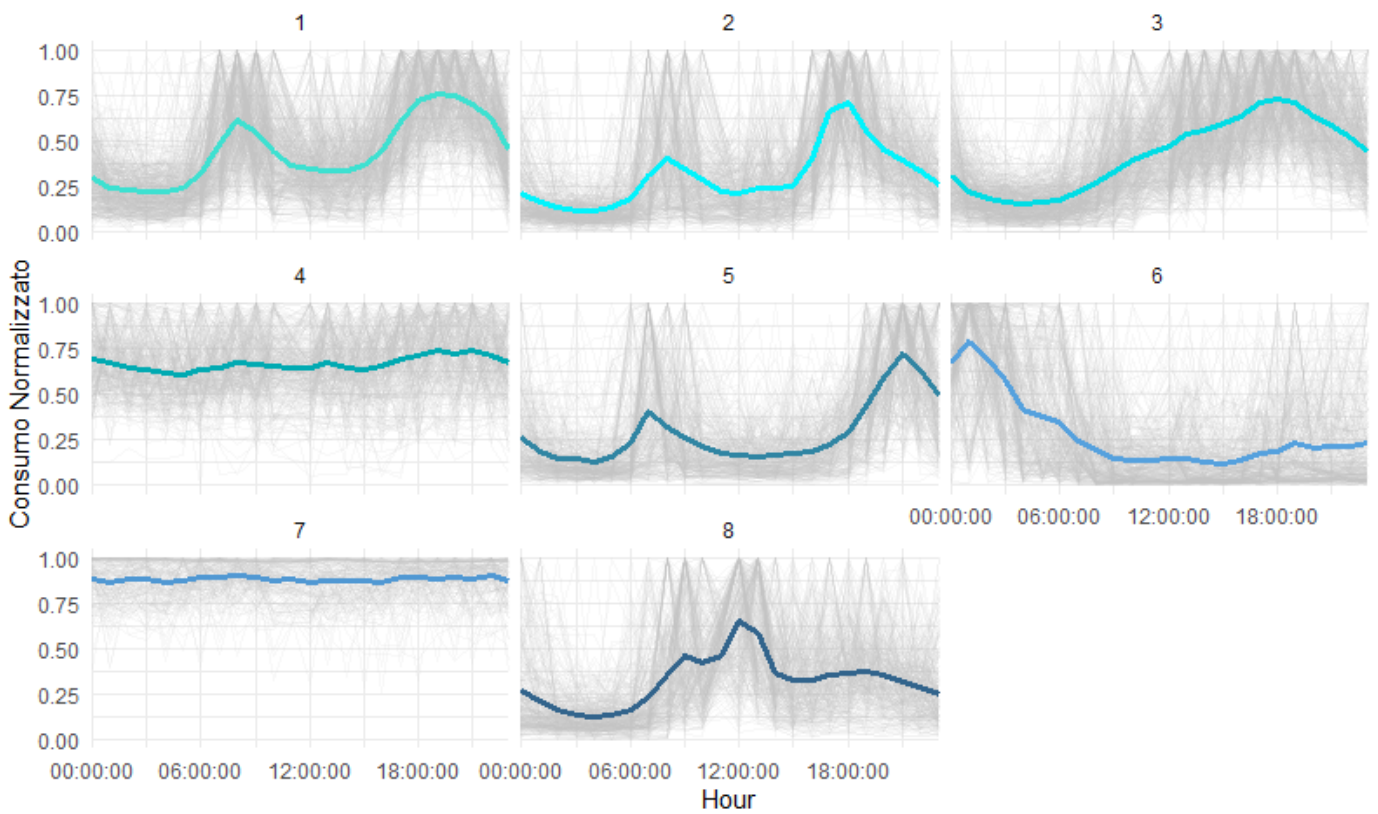
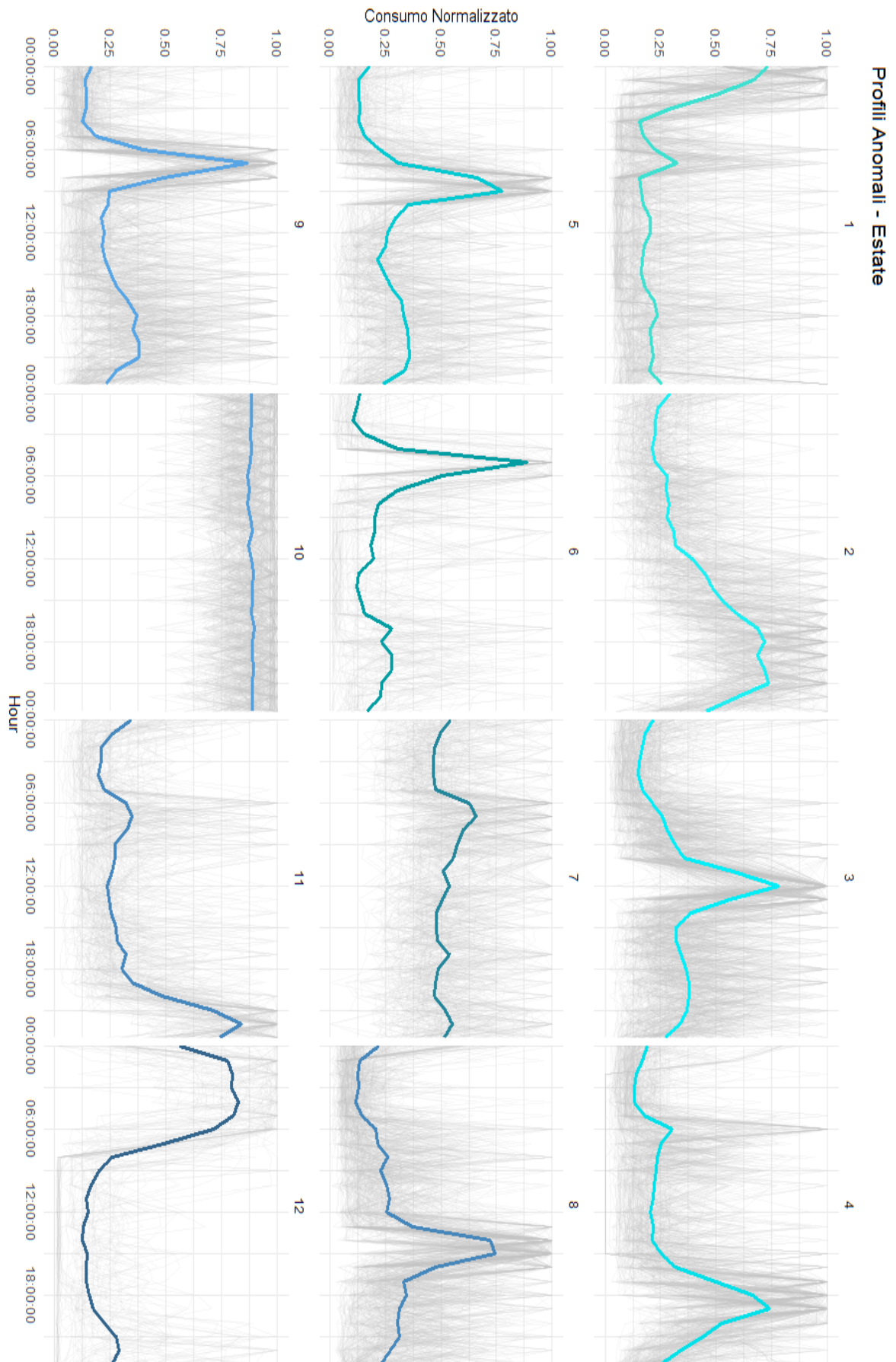
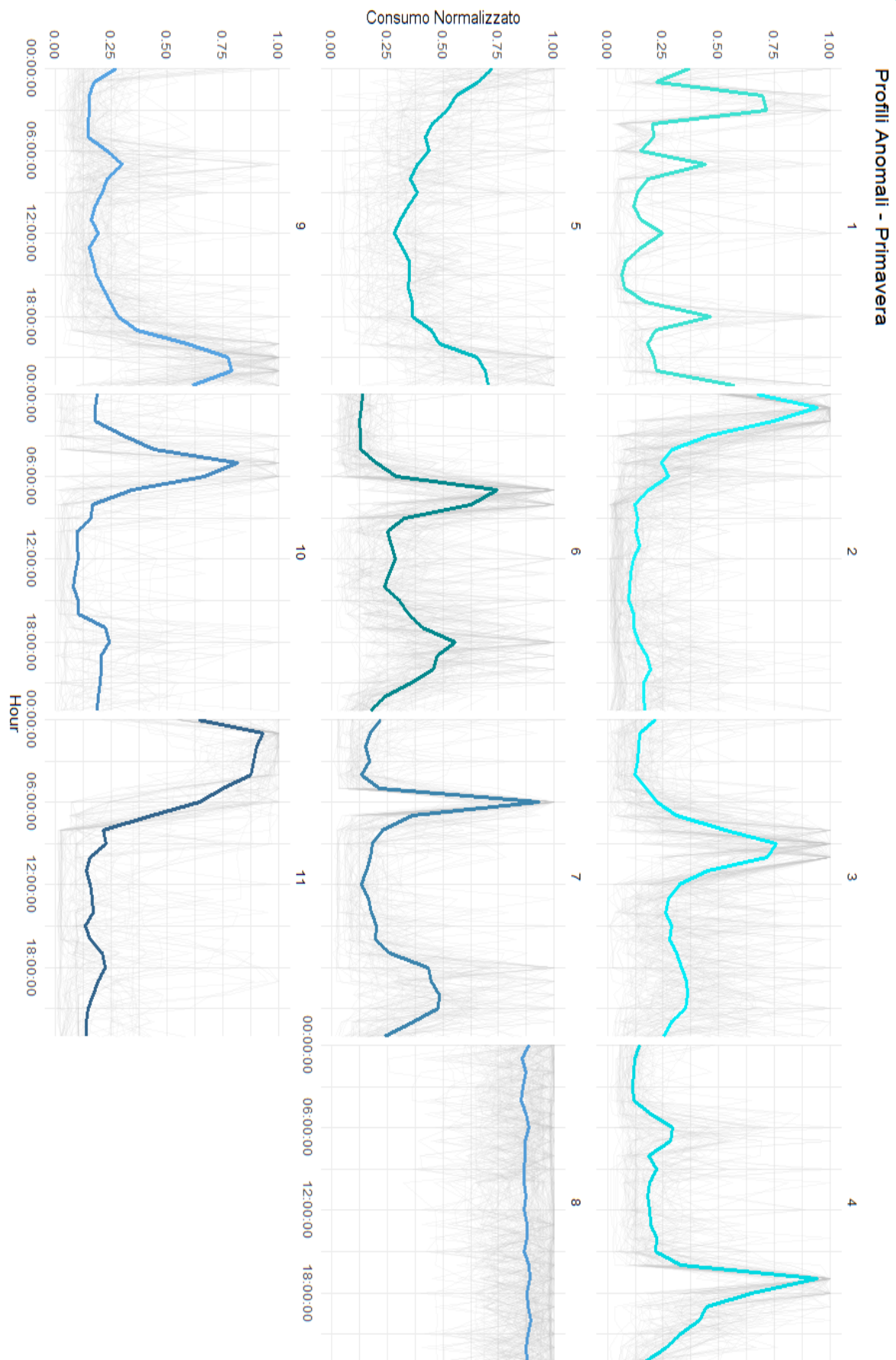


Figura 31 Profili giornalieri anomali per la stagione invernale.





**Figura 32** *Profili giornalieri anomali per la stagione estiva.*



**Figura 33** *Profili giornalieri anomali per la stagione primaverile.*

Come mostrato dai grafici, la maggior parte dei profili comuni tra le quattro stagioni sono principalmente profili con carichi principalmente concentrati nelle ore notturne, carichi costanti o carichi bassi caratterizzati da un forte picco localizzato in una o più ore specifiche.

Nello specifico, in autunno (**Figura 30**) vengono identificati due sottogruppi di profili di carico costante a diversa intensità, sei sottogruppi a carico basso ma con picchi localizzati e un sottogruppo con un'elevata intensità del carico in periodi notturni.

In estate (**Figura 32**) si individuano due sottogruppi con carico costante e sei con picchi localizzati e uno con carichi alti nelle ore notturne, similmente al periodo autunnale.

La primavera è la stagione con più sottogruppi anomali individuati. In particolare, sono presenti quindici sottogruppi con carico basso e picchi localizzati, un solo sottogruppo con carico costante mentre i restanti racchiudono profili con molteplici picchi spalmati lungo il pomeriggio e nelle ore notturne (**Figura 33**).

Il periodo invernale è molto simile, in termini di anomalie, al periodo autunnale ed estivo. Sono presenti anche qui due sottogruppi con carico costante ma a diversi livelli di intensità, nove sottogruppi con un picchi localizzati e un sottogruppo con carichi elevati nelle ore notturne (**Figura 31**).

In definitiva, vengono identificati ed considerati anomali 7223 profili su scala giornaliera, schematizzati in **Tabella 17**.

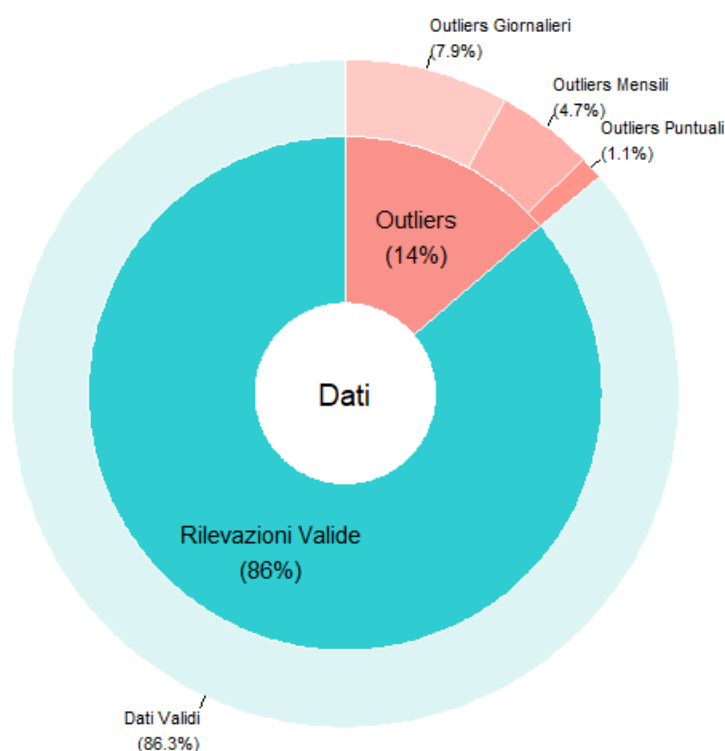
**Tabella 17** Risultati del processo di identificazione degli outliers su scala giornaliera.

<i>Stagione</i>	<i>Profili anomali</i>	<i>Numero di sottogruppi anomali (DBI Index)</i>
Estate	2850	9

Inverno	750	12
Primavera	1981	20
Autunno	1642	9

Gli *outliers* su scala mensile, invece, sono di meno rispetto ai precedenti e interessano 143 profili medi mensili, che vengono eliminati dal *dataset*. Pertanto, tutte le rilevazioni orarie del consumo elettrico relative ai mesi caratteristici di un basso consumo non vengono utilizzate ai fini dell'analisi.

Complessivamente, il *pre-processing* ha portato all'eliminazione di dati anomali per una percentuale pari al 13.7% rispetto al totale, scomposto nei termini rappresentati in **Figura 34**.



**Figura 34** Risultati del *pre-processing* sviluppato.

In sintesi, l'1.1% delle rilevazioni è stato identificato come *outlier* puntuale, il 7.9% come *outlier* giornaliero e il 4.7% come *outlier* su scala mensile. L'assenza di valori mancanti e di valori negativi lascia pensare ad una elaborazione preliminare del *dataset* originale prima che fosse reso pubblico.

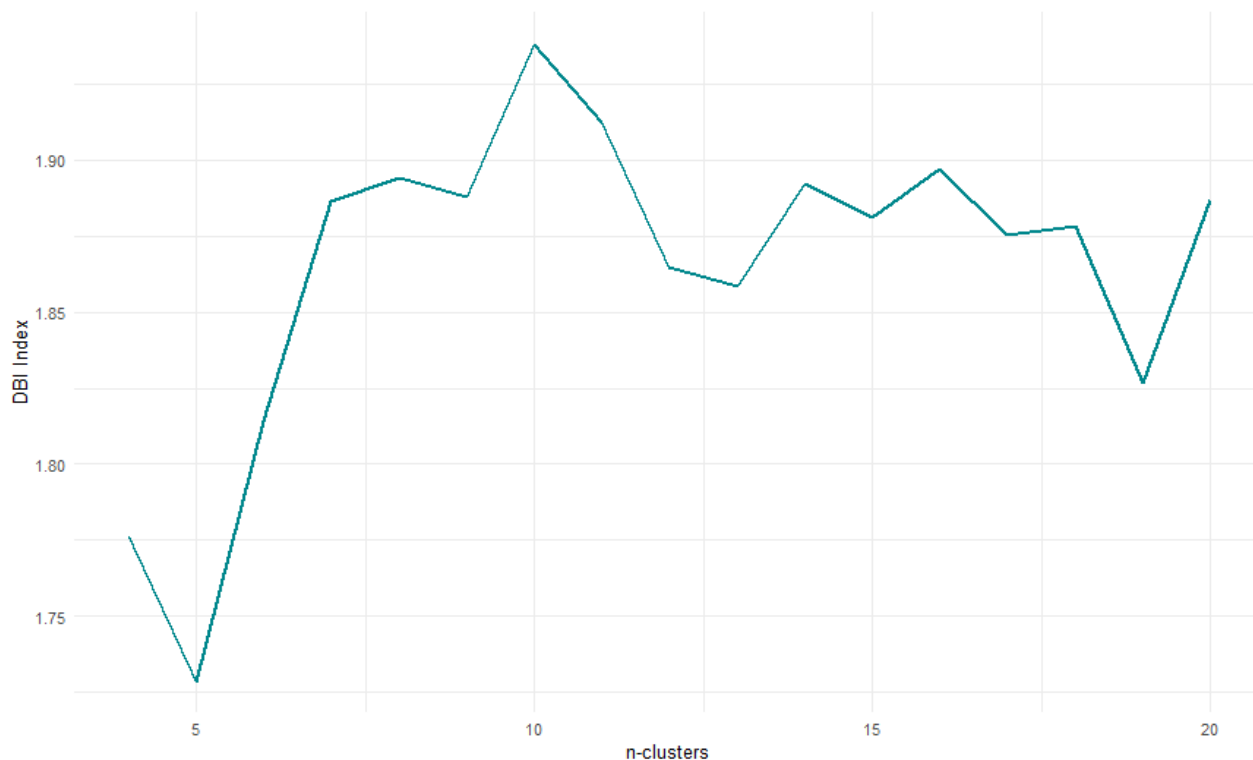
## 5.2 Clustering

Il *clustering* per le utenze residenziali, descritto nel Capitolo 4, restituisce i seguenti risultati, suddivisi nelle seguenti sezioni:

- valutazione del DBI Index per il numero ottimale di *cluster*  $k$ ;
- identificazione dei *pattern* di consumo tipologici;
- selezione dei *cluster* più popolati e più rappresentativi delle utenze residenziali.

### Valutazione del DBI Index

La variazione del DBI Index, per l'identificazione dei *pattern* tipologici di consumo mensile, al variare del numero del numero di cluster ( $k$  compreso tra 4 e 20) è rappresentata in **Figura 35**.

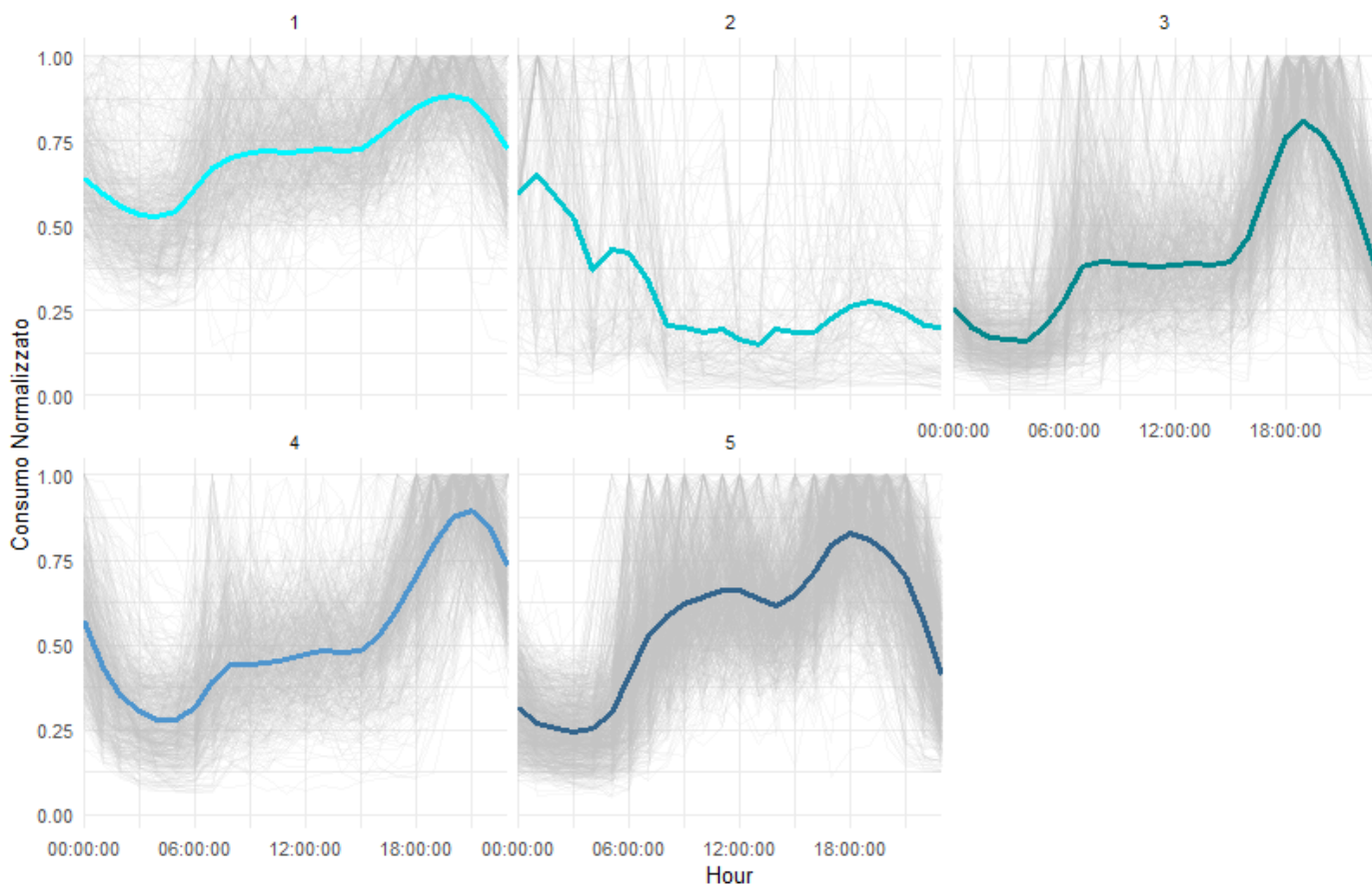


**Figura 35** Variazione del DBI Index al variare del numero di cluster.

Come si evince dal grafico, il numero ottimale di *cluster k* scelto è pari a 5.

### *Pattern tipologici di consumo*

I *pattern* tipologici di consumo individuati tramite il *k-Means* descritto nella metodologia restituisce i seguenti sottogruppi di consumo, rappresentati in **Figura 36**.



**Figura 36** *Pattern tipologici di consumo mensile individuati.*

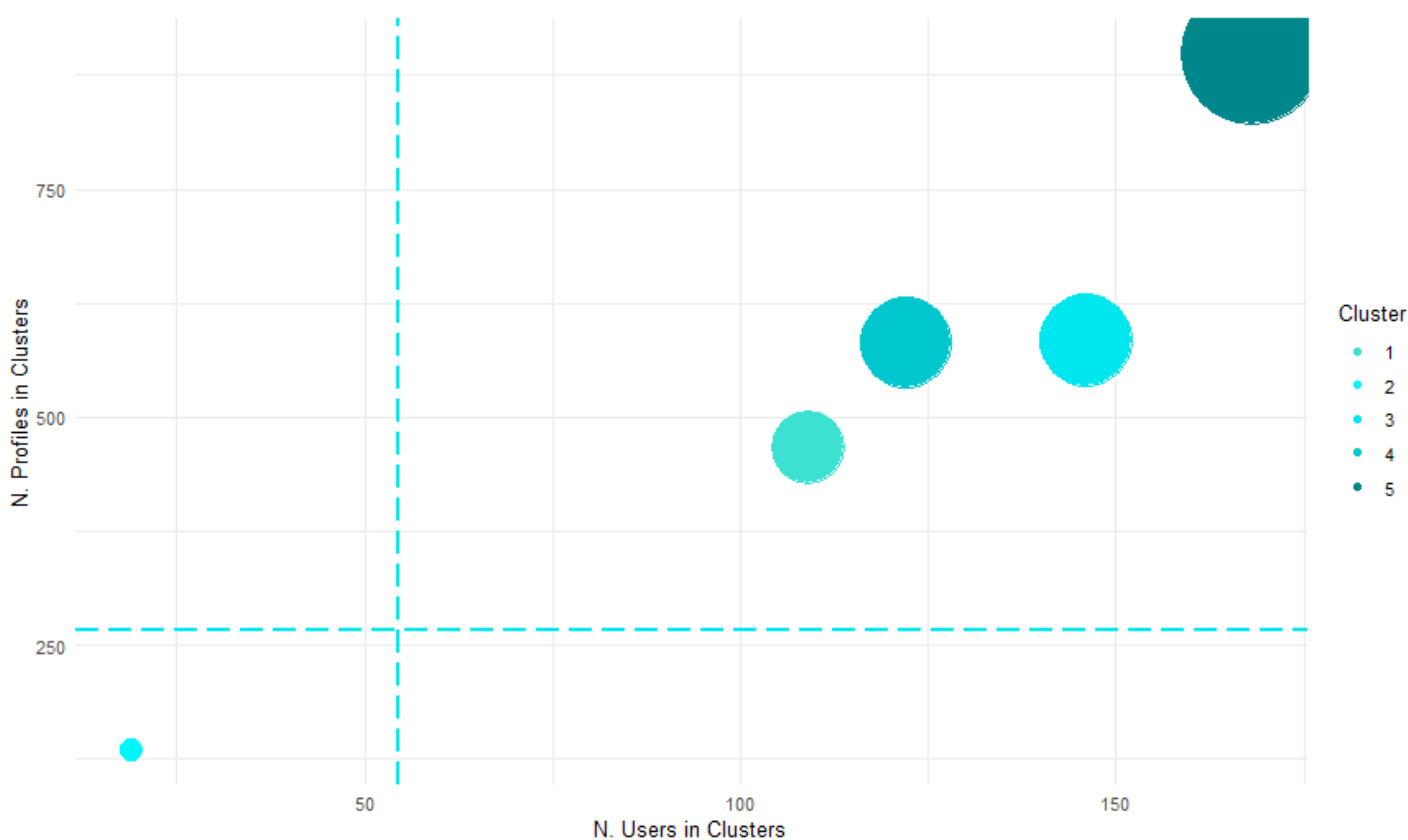
In definitiva, vengono identificate 5 diverse tipologie di consumo, nello specifico:

- *Tipologia 1* – consumo mediamente costante nel corso della giornata, con un abbassamento nelle ore serali ed un notevole picco tra le 21:00 e le 22:00;
- *Tipologia 2* – consumo mediamente basso con dei picchi di maggiore intensità nelle ore notturne;
- *Tipologia 3* – consumo basso nelle ore serali e con un picco a ridosso delle ore 18:00, con un carico costante dalle 06:00 alle 15:00;
- *Tipologia 4* – consumo mediamente elevato con un leggero abbassamento durante le ore notturne e un leggero rialzo dopo le

20:00;

- *Tipologia 5* – consumo moderato con un rialzo mattutino, seguito da un leggero abbassamento nelle ore di pranzo e un conseguente rialzo a ridosso delle ore 18:00.

Dopo aver individuato le diverse tipologie di consumo, si effettua un ulteriore controllo per escludere i sottogruppi poco popolati e con un numero basso di utenti al loro interno, così come eseguito per l'identificazione degli *outliers* su scala giornaliera (**Figura 37**).



**Figura 37** Identificazione dei pattern di consumo anomali dei profili medi mensili.

La metodologia evidenzia come la Tipologia 2 di consumo contiene al proprio interno effettivamente pochi profili e un numero basso di utenti. Per questo motivo, tale sottogruppo viene escluso dal continuo dell'analisi.

### 5.3 Classificazione

Il modello utilizzato è stato ottenuto, come descritto nel **Capitolo 4**, attraverso un processo iterativo volto ad individuare gli iperparametri che

generano una migliore precisione in fase di *training*.

Per riassumere, le variabili utilizzate per allenare il modello e le caratteristiche principali dello stesso vengono riassunte in **Tabella 18 e 19**.

**Tabella 18** Variabili utilizzate per allenare il classificatore.

<i>Nome</i>	<i>Descrizione</i>	<i>Unità di misura</i>
Green/Total - corrected	Valore della fascia di consumo "Green" sul consumo mensile totale, corretto secondo il fattore di correzione $f$	[-]
Green/Red - corrected	Valore della fascia di consumo "Green" sul valore della fascia di consumo "Red" corretto, secondo il fattore di correzione $f$	[-]
Green/Amber - corrected	Valore della fascia di consumo "Green" sul valore della fascia di consumo "Amber", corretto secondo il fattore di correzione $f$	[-]
Amber/Red - corrected	Valore della fascia di consumo "Amber" sul valore della fascia di consumo "Red", corretto secondo il fattore di correzione $f$	[-]
Red/Total - corrected	Valore della fascia di consumo "Green" sul consumo mensile totale, corretto secondo il fattore di correzione $f$	[-]
Amber/Total - corrected	Valore della fascia di consumo "Amber" sul consumo mensile totale, corretto secondo il fattore di correzione $f$	[-]
Green	Valore della fascia di consumo "Green"	[kWh]
Amber	Valore della fascia di consumo "Amber"	[kWh]
Red	Valore della fascia di consumo "Red"	[kWh]



Total	Consumo mensile totale	[kWh]
Amber/Total	Valore della fascia di consumo "Amber" sul consumo mensile totale	[-]
Red/Total	Valore della fascia di consumo "Red" sul consumo mensile totale	[-]
Green/Total	Valore della fascia di consumo "Green" sul consumo mensile totale	[-]
Amber/Red	Valore della fascia di consumo "Amber" sul valore della fascia di consumo "Red"	[-]
Green/Amber	Valore della fascia di consumo "Green" sul valore della fascia di consumo "Amber"	[-]
Green/Red	Valore della fascia di consumo "Green" sul valore della fascia di consumo "Red"	[-]
Amber/Total	Valore della fascia di consumo "Amber" sul consumo mensile totale	[-]
Red/Total	Valore della fascia di consumo "Red" sul consumo mensile totale	[-]
Green/Total	Valore della fascia di consumo "Green" sul consumo mensile totale	[-]
Amber/Red	Valore della fascia di consumo "Amber" sul valore della fascia di consumo "Red"	[-]

*Tabella 19 Caratteristiche del Random Forest migliore ottenuto dal "tuning".*

<i>Parametro</i>	<i>Valore</i>
------------------	---------------

Numero di alberi	4000
Variabili utilizzate ad ogni split	2
Numero minimo di elementi per nodo foglia	16
importance	Impurity
split rule	Gini
respect.unordered.factor	order
measure	acc

La precisione del modello è valutata in base al numero di classificazioni corrette effettuate, sia in fase di *testing* che in fase di *training* (**Tabella 20**).

**Tabella 20** Precisione del modello nel processo di classificazione.

<i>Parametro</i>	<i>Valore</i>
Precisione in <i>training</i>	69%
Precisione in <i>testing</i>	62%

Di seguito, viene riportata la matrice di confusione per il *testing set* (**Tabella 21**).

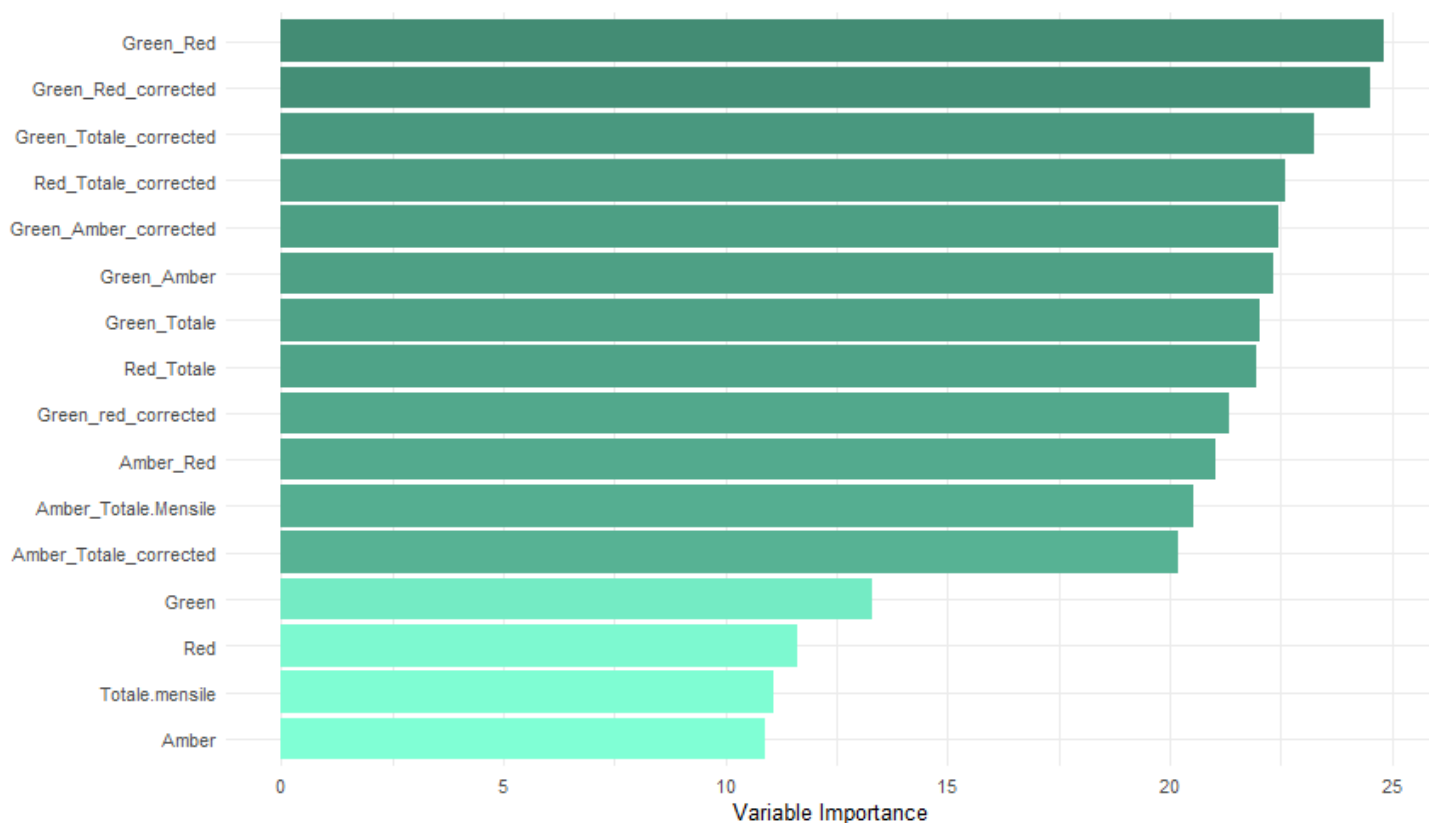
**Tabella 21** Matrice di confusione del *testing set*.

<i>Tipologia</i>	<b>1</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>Precisione</b>
<b>1</b>	62	0	13	19	<b>65.9%</b>

3	6	42	17	54	35.3%
4	20	7	72	17	62.0%
5	6	23	10	138	78.0%

Come si evince dalla matrice di confusione appena illustrata, il modello distingue molto bene la Tipologia 5 dalle altre, mentre ha una precisione peggiore per gli altri casi. In particolare, il modello fatica a distinguere la Tipologia 3 dalla 5.

Inoltre, è possibile valutare, attraverso il Gini Index, quale delle variabili è stata più utile al modello durante il processo di classificazione (**Figura 38**).



**Figura 38** *Importanza delle variabili di classificazione utilizzate.*

Come emerge dal grafico, le due variabili più importanti nel processo di classificazione sono il rapporto tra i consumi in fascia “Green” e “Red”, sia

corretto tramite il fattore  $f$  che nella sua forma più semplice. Ciò esprime il fatto che nel processo di classificazione è molto importante il rapporto tra il consumo rilevato nelle ore notturne e tra il consumo rilevato nelle ore centrali della mattina e del pomeriggio.

Successivamente, per ordine di importanza, emergono in successione tutte le restanti variabili corrette del fattore  $f$  introdotte, seguite dalle restanti variabili definite come rapporto. Ciò dimostra che tali variabili influenzano positivamente le scelte decisionali del modello, per i motivi discussi nel Capitolo 4.

Infine, le variabili meno importanti per il modello sono i consumi energetici grezzi, non rapportati a nessun'altra variabile. Tali parametri sono infatti privi di ogni confronto e non aggiungono delle informazioni aggiuntive, utili al modello.

## 5.4 Ricostruzione del profilo orario

Per gli utenti del *testing set*, ovvero 52 utenti residenziali scelti casualmente, si procede alla ricostruzione del profilo orario. Calcolando l'energia condivisa reale e quella simulata mettendo a confronto il profilo reale e quello elaborato dalla metodologia corrente con il profilo di produzione degli impianti fotovoltaici, è possibile valutare l'errore relativo commesso, al livello di singolo utente, sintetizzato in **Tabella A**, presente in **Appendice A**.

I risultati dimostrano che l'errore relativo medio riferito calcolo dell'energia condivisa, al livello di singolo utente è pari al 21.33%. Questo implica che, in generale, il profilo orario simulato approssima abbastanza bene, nell'anno, il profilo reale, ottenendo così un valore di energia condivisa molto simile a quello effettivo. In **Figura 39, 40, 41 e 42** vengono messi a confronto il profilo orario simulato e quello reale per alcuni mesi dell'anno, per diversi utenti del *testing set*.

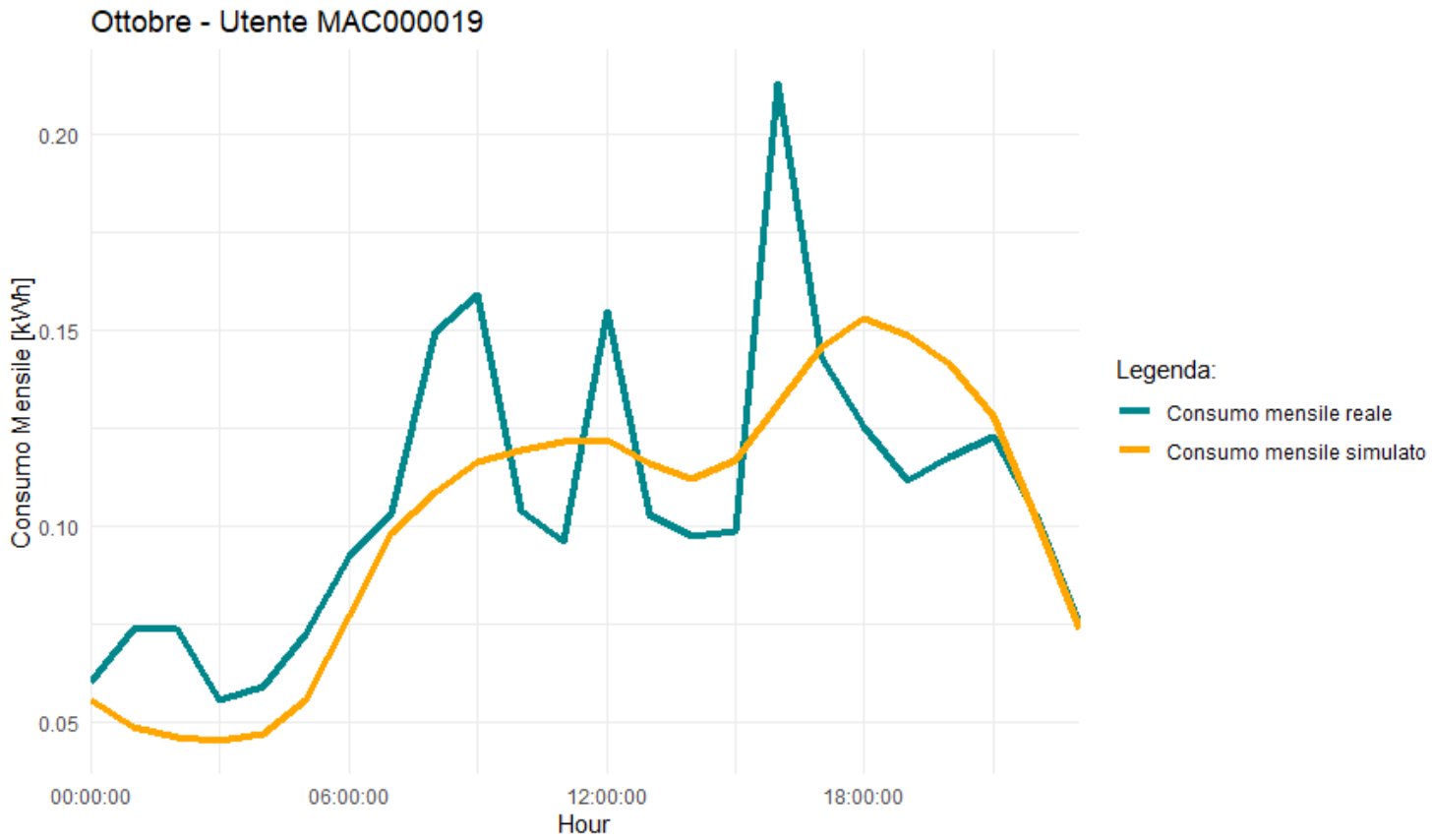


Figura 39 Confronto tra profilo reale e simulato - 1

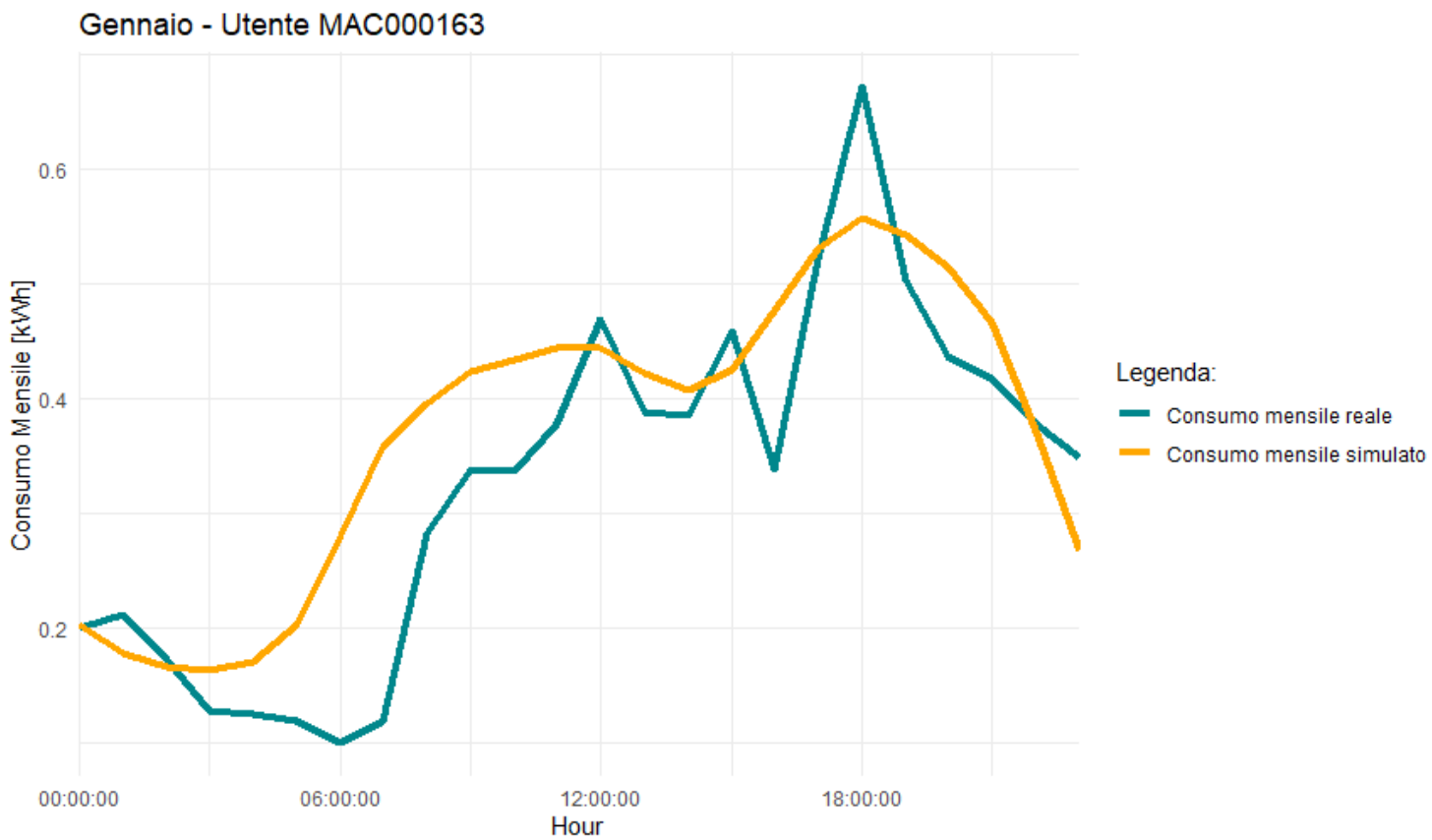


Figura 40 Confronto tra profilo reale e simulato - 2

Luglio - Utente MAC000097

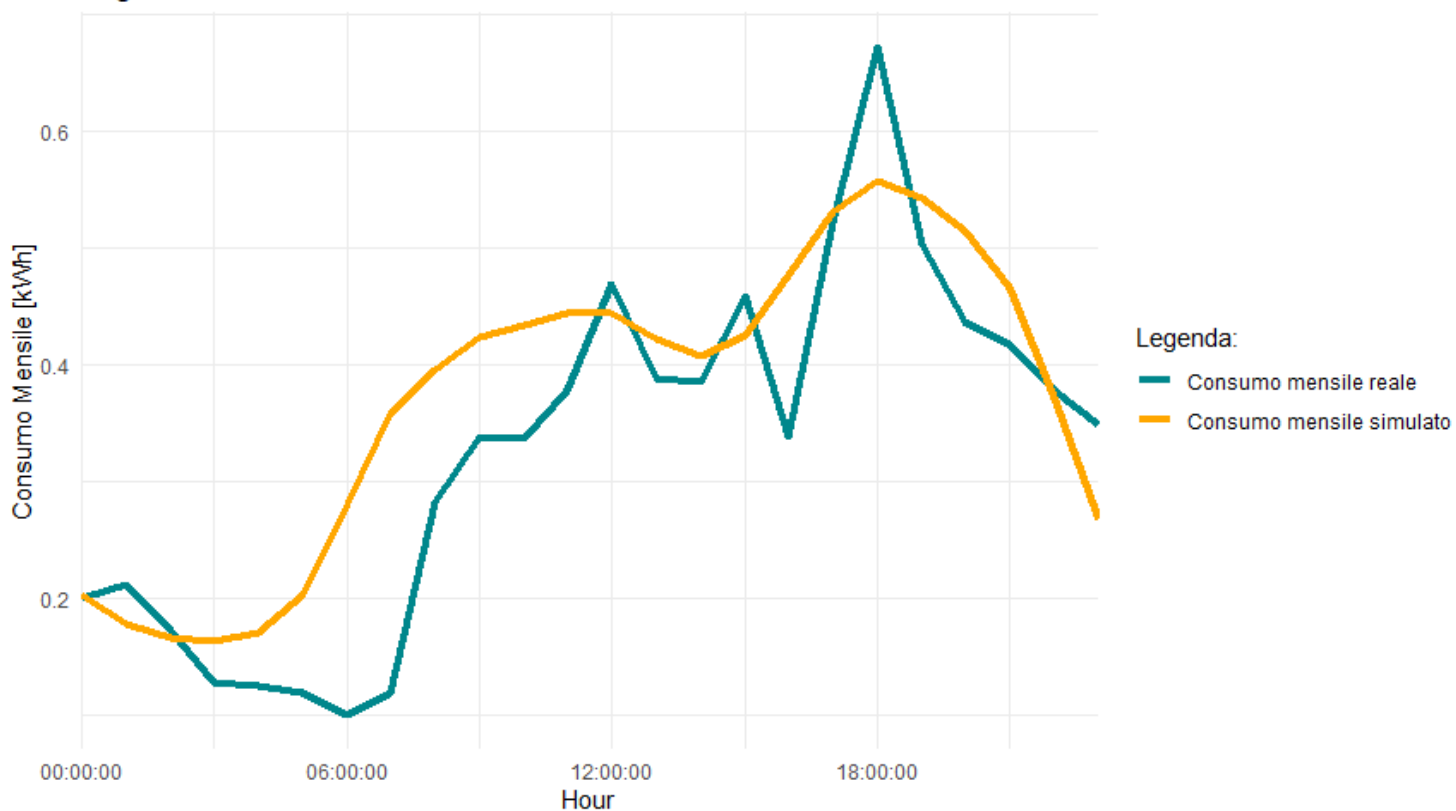


Figura 41 Confronto tra profilo reale e simulato - 3

Giugno - Utente MAC000055

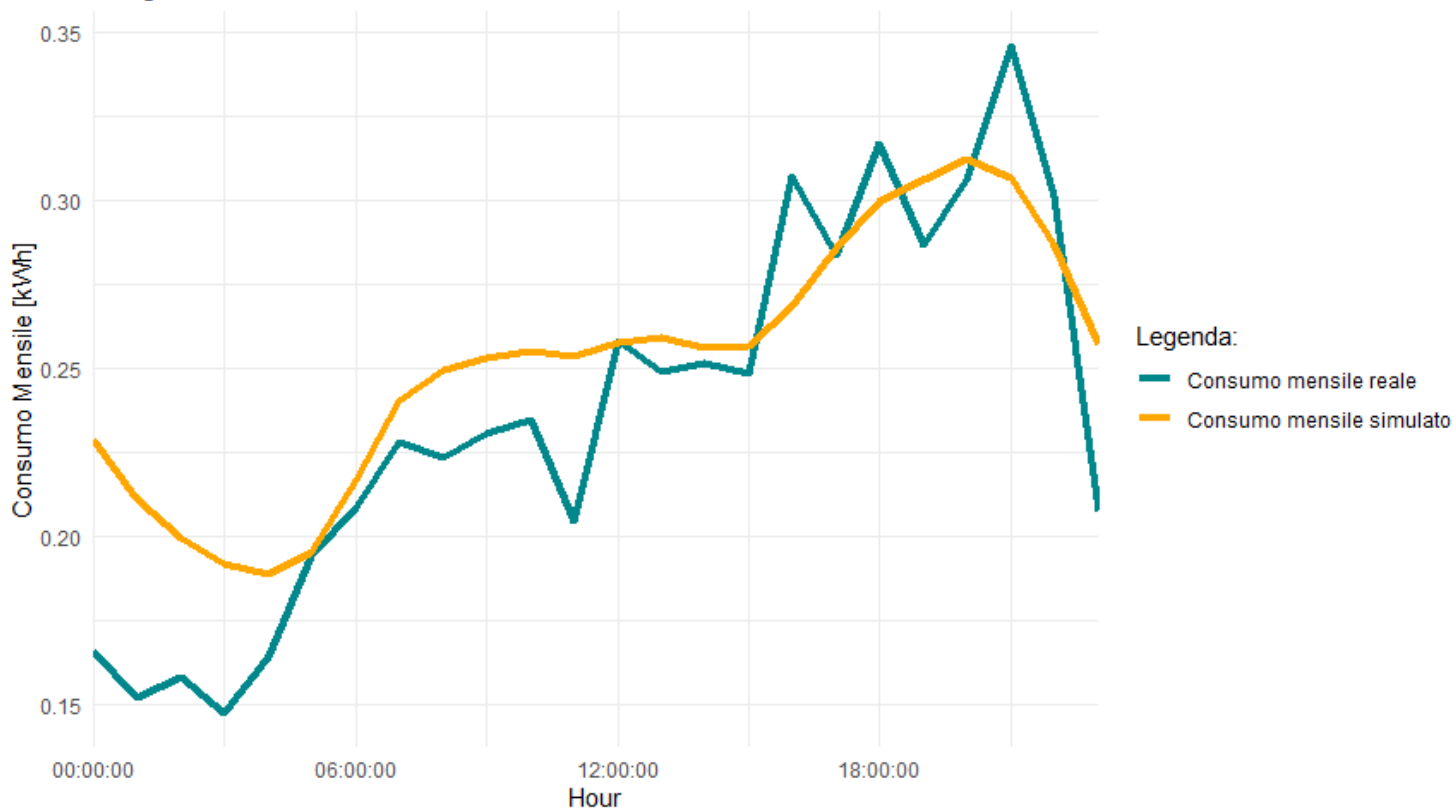
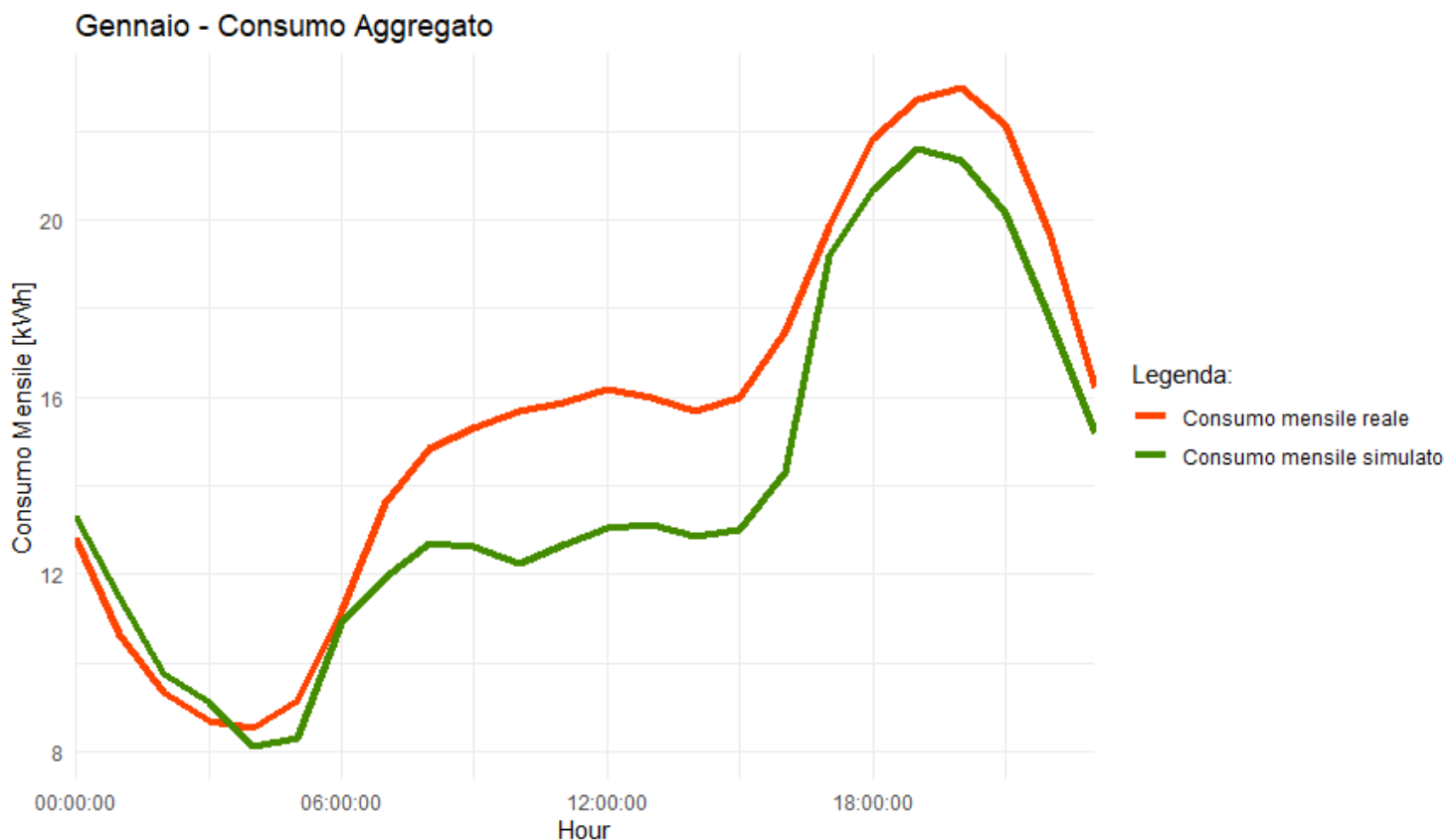


Figura 42 Confronto tra profilo reale e simulato - 4

Come si vede dai grafici sopra riportati, il profilo reale simulato è molto simile al profilo reale, su scala mensile, ma differisce per alcuni picchi isolati. Ciò è dovuto al comportamento stocastico degli utenti, ovvero alla loro imprevedibilità nei consumi, fenomeno preponderante su scala giornaliera. L'effetto del comportamento stocastico delle utenze residenziali, tende a ridursi man mano che si estende la scala temporale di osservazione, infatti, osservando i consumi medi di un utente, su scala annuale, i picchi relativi all'imprevedibilità e alla casualità dei consumi elettrici hanno un peso minore, rispetto a risoluzioni temporali più basse, come quella giornaliera o mensile.

La metodologia proposta perde la capacità di prevedere i picchi del singolo utente, tuttavia, essi vengono mediati al livello aggregato, come mostrato in **Figura 43, 44, 45 e 46**, dove vengono messi a confronto i profili medi mensili reali e simulati, al livello aggregato, per alcuni dei mesi analizzati.



*Figura 43 Consumo aggregato reale e simulato a confronto - 1*

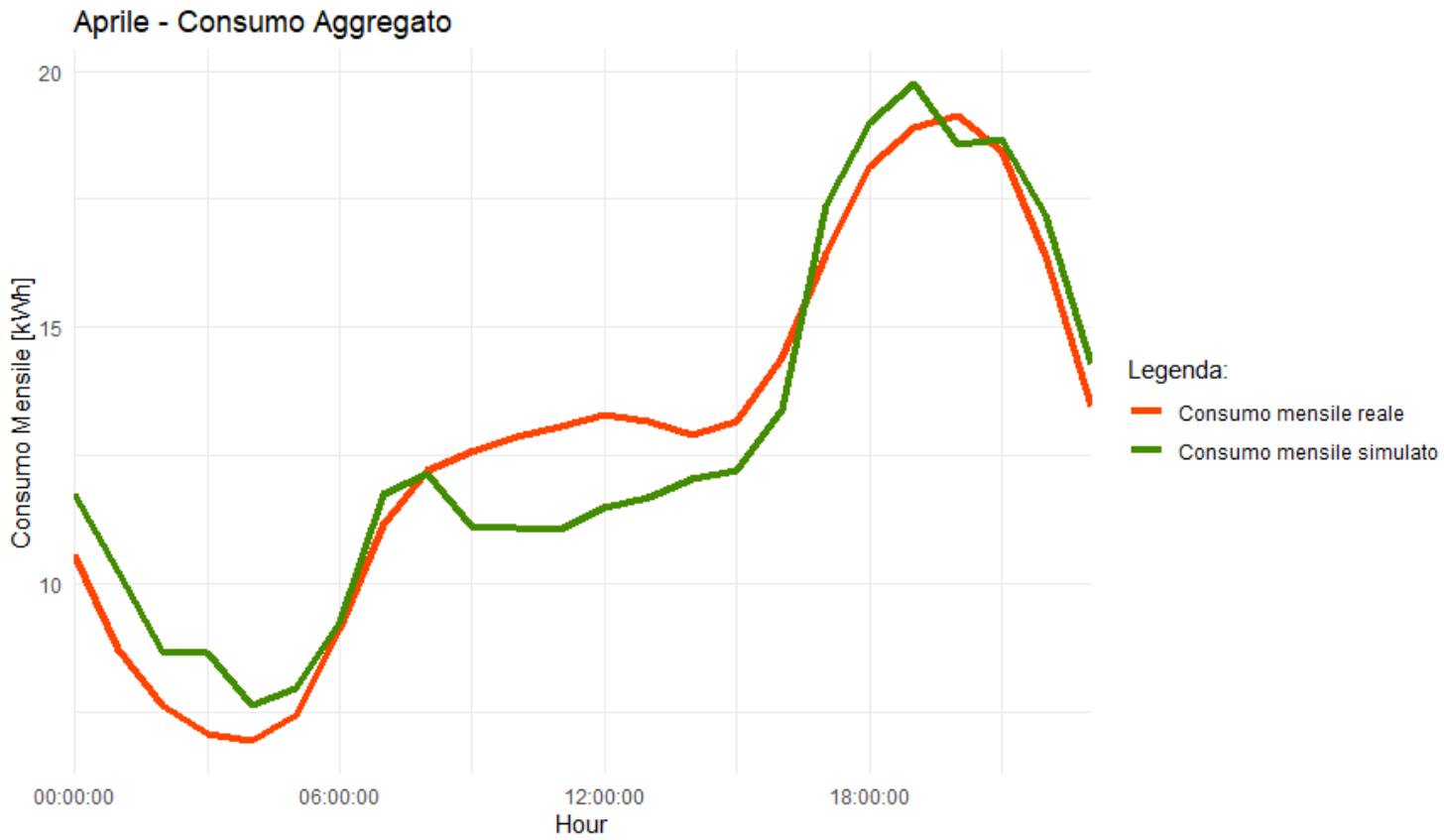


Figura 44 Consumo aggregato reale e simulato a confronto – 2

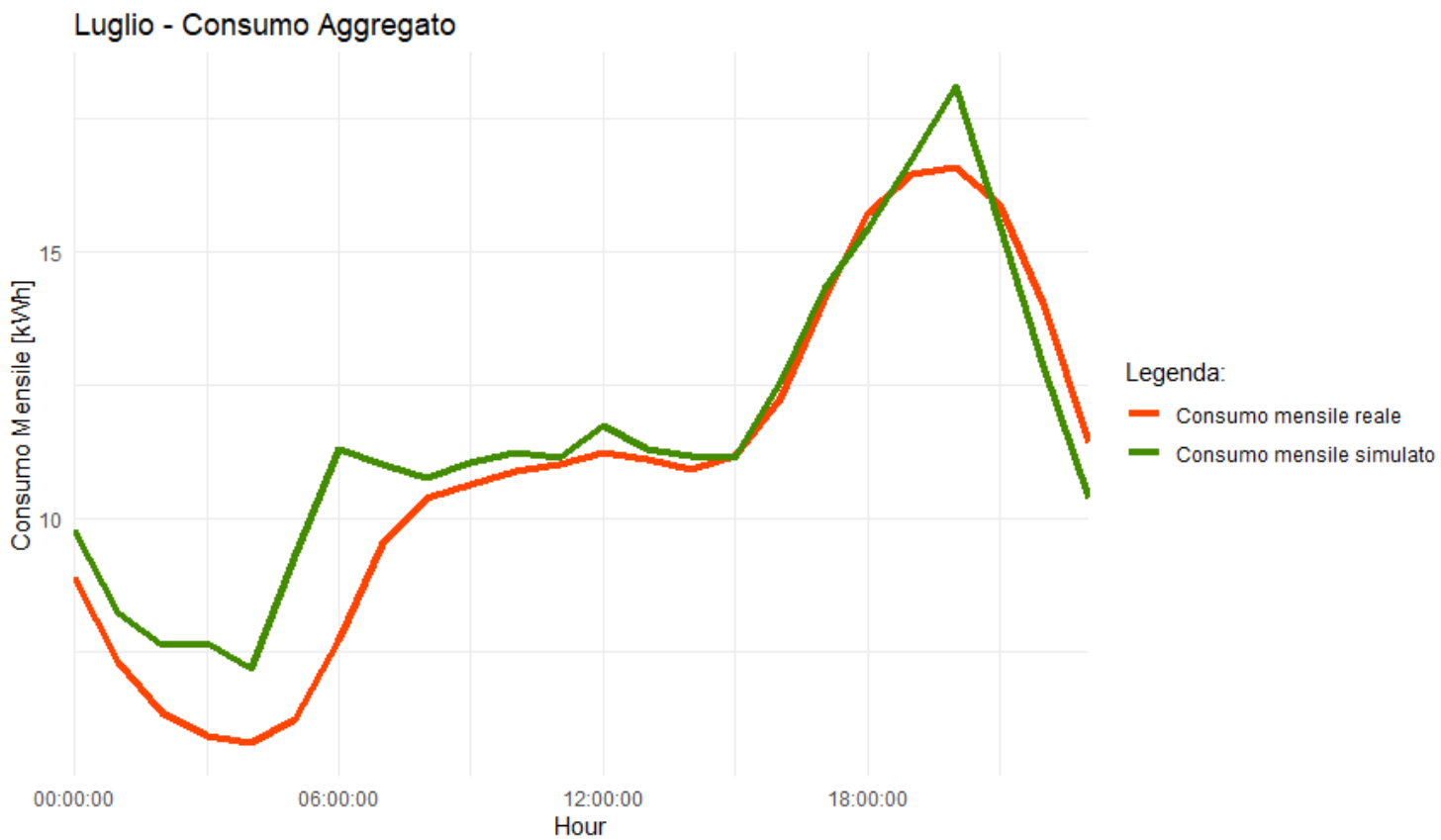


Figura 45 Consumo aggregato reale e simulato a confronto – 3



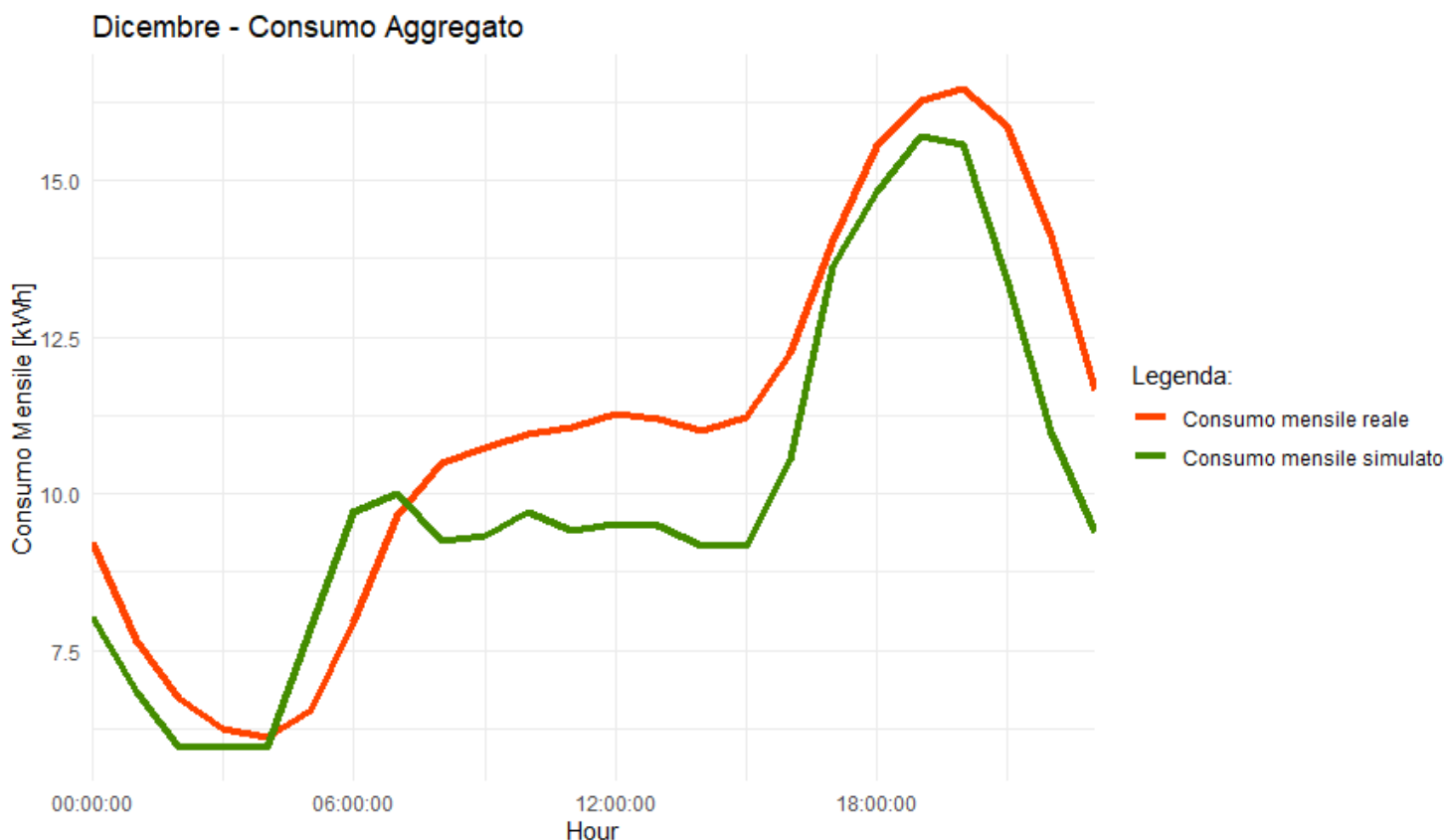


Figura 46 Consumo aggregato reale e simulato a confronto - 4

Al livello aggregato, considerando i 52 utenti del testing set come utenze residenziali afferenti alla stessa CER con le caratteristiche di produzione descritte nel Capitolo 4, si commette un errore sul calcolo dell'energia condivisa pari al **0.56%** (Tabella 23). Analizzando le utenze residenziali insieme, si commette un errore minore perché le imperfezioni del modello, al livello di singolo utente, assumono un peso minore man mano che la CER è più grande, in termini di utenti al suo interno e in termini di energia condivisa generata all'interno della CER stessa.

Tabella 22 Errore relativo al livello aggregato per l'E.C.

<i>Energia condivisa reale dell'aggregato</i>	<i>Energia condivisa simulata dell'aggregato</i>	<i>Errore relativo</i>
<i>[kWh]</i>	<i>[kWh]</i>	<i>[%]</i>
36179.09	35995.92	0.56

## 6 Conclusioni

L'obiettivo della metodologia proposta è quello di stimare il profilo orario di utenze residenziali a partire dai soli consumi mensili attraverso tecniche non intrusive. Tale scopo emerge dall'esigenza di far fronte alla mancanza di dati orari reali negli studi di prefattibilità delle CER, soprattutto nel contesto residenziale, dove il corretto bilanciamento, al livello orario, tra consumo e produzione determina ingenti ingressi economici in contesti di valorizzazione ed incentivazione dell'energia condivisa, necessari al sostentamento economico della comunità.

Ripercorrendo i passaggi principali della metodologia illustrata nel presente lavoro di tesi, si è effettuato, in prima istanza, un intenso *pre-processing*, atto ad eliminare le anomalie e gli elementi di disturbo che avrebbero potuto ridurre la precisione dei modelli utilizzati in seguito, riducendo di fatto il dataset utilizzato del 14%.

Successivamente, è stato fatto un clustering di tipo *K-means* individuando quattro *pattern* di consumo tipologici delle utenze residenziali, in seguito utilizzati per creare un modello supervisionato di *machine learning*, nello specifico un *Random Forest*, caratterizzato precisione del 69% in fase di *training* e del 62% in fase di *testing*, con lo scopo di implementare un modello in grado di identificare il *pattern* di consumo di un'utenza generica a partire dai soli consumi mensili.

Una volta implementato il modello, si è proceduto al *rescaling* dei profili normalizzati utilizzati per i processi di *clustering* e di classificazione precedenti. Nello specifico è stato individuato un fattore di scala  $p$  che consentisse il passaggio dai profili normalizzati, rappresentativi della sola forma del carico, a profili con un'intensità di carico vera e propria, di fatto ricostruendo un profilo che approssimasse quanto più possibile il profilo reale.

I profili così ottenuti, confrontati con un impianto fotovoltaico tipo, hanno generato mediamente un'energia condivisa, relativa alla singola utenza, con un errore relativo del 21% rispetto all'energia condivisa calcolata tramite i profili

reali. Al livello aggregato, e quindi al livello di CER, l'errore commesso sul calcolo dell'energia condivisa è notevolmente inferiore, ed è pari al 0.56%.

I motivi dell'abbassamento così drastico dell'errore passando dal singolo utente al livello aggregato è dovuto ad un fenomeno di compensazione dell'errore sulla stima dell'energia condivisa. Nello specifico, la metodologia implementata, al livello di singolo utente, in alcuni casi sottostima l'energia condivisa mentre in altri la sovrastima, generando quindi al livello aggregato un fenomeno di compensazione che porta complessivamente alla riduzione, nel caso specifico quasi all'annullamento, dell'errore.

In definitiva, la riduzione dell'errore, passando dal livello di singolo utente al livello aggregato, può essere tanto più significativa quanto più l'intensità dell'energia condivisa sovrastimata è pari all'energia condivisa sottostimata. Ciò dipende quindi dalle caratteristiche degli utenti che vengono processati.

Per evitare di ottenere un errore basso a causa di fenomeni di compensazione dell'errore, e non per l'effettiva efficacia del modello, alcuni dei possibili lavori futuri al presente progetto di tesi potrebbero riguardare l'aumento della precisione del classificatore utilizzato, ad esempio attraverso l'impiego di dati più completi e descrittivi delle utenze residenziali, come ad esempio dataset contenenti, oltre che al semplice valore di consumo, il numero di elementi in famiglia per utenza, le caratteristiche impiantistiche dello stabile o l'impiego lavorativo degli occupanti, aspetti che influenzano notevolmente le caratteristiche del carico elettrico. In aggiunta, potrebbe essere interessante valutare l'impiego di tecniche di *clustering* e di *pre-processing diverse*, valutandone l'impatto sulla precisione del modello.

## APPENDICE A

**Tabella A** *Errore relativo sull'energia condivisa simulata*

<i>Nome Utente</i>	<i>Energia Condivisa Reale [kWh]</i>	<i>Energia Condivisa Simulata [kWh]</i>	<i>Errore Relativo [%]</i>
MAC000019	571.65	511.07	10.59
MAC000020	592.98	493.18	16.83
MAC000026	1104.07	895.68	18.87
MAC000034	554.15	510.79	7.83
MAC000055	1008.38	1055.98	4.72
MAC000077	390.103	591.84	51.71
MAC000086	616.03	487.81	19.18
MAC000097	1666.05	2094.19	25.69
MAC000099	344.34	303.08	11.98
MAC000102	178.56	177.42	1.20
MAC000112	1047.85	1016.03	3.03
MAC000115	604.68	566.22	6.36
MAC000118	554.27	633.98	14.38
MAC000136	1495.06	1947.72	30.27
MAC000160	477.09	438.07	8.18
MAC000161	900.67	833.48	7.46
MAC000163	1087.52	1180.18	8.52
MAC000180	871.14	520.00	40.30
MAC000205	1021.99	799.33	21.78

MAC000206	916.93	1289.02	40.58
MAC000209	253.04	291.07	15.03
MAC000216	1984.67	2324.83	17.14
MAC000223	383.66	542.40	41.37
MAC000228	772.26	646.96	16.22
MAC000231	381.70	316.61	17.05
MAC000243	1046.91	969.37	7.40
MAC000251	835.65	689.93	17.43
MAC000253	774.04	1018.26	31.55
MAC000254	882.14	1143.49	29.63
MAC000255	1691.75	1879.16	11.08
MAC000273	364.09	717.63	97.1
MAC004460	889.09	715.55	19.52
MAC004472	778.93	890.07	14.26
MAC004480	1710.05	1844.83	7.88
MAC004496	248.31	206.95	16.65
MAC004499	755.77	619.45	18.04
MAC004515	243.80	223.83	8.19
MAC004518	484.51	728.69	50.39
MAC004524	225.43	317.86	41.00
MAC004528	551.59	695.02	26.00
MAC004537	857.41	571.15	33.38
MAC004539	443.74	556.63	25.44
MAC004547	805.27	634.56	21.19

MAC004550	574.00	398.55	30.56
MAC004555	484.72	316.35	34.73
MAC004569	461.74	513.31	11.16
MAC004571	1108.84	810.02	26.95
MAC004572	514.75	520.65	1.14
MAC004583	313.55	257.22	17.96
MAC004846	445.66	365.00	18.09
MAC004856	469.47	611.40	30.23
MAC004861	957.17	899.67	6.00
<hr/>			
<b>Errore Relativo Medio</b>	-	-	<b>21.33</b>
<b>[%]</b>			
<hr/>			

## Bibliografia

[1] Commissione Europea – Comunicazione della commissione al parlamento europeo, al consiglio, al comitato economico e sociale europeo, al comitato delle regioni e alla banca europea per gli investimenti. Energia pulita per tutti gli europei Bruxelles, 30.11.2016 COM (2016) 860 final

[2] Accordo di Parigi. 12 dicembre 2015. 21a Conferenza delle Parti della Convenzione quadro delle Nazioni Unite sui cambiamenti climatici (COP 21). FCCC/CP/2015/L.9/Rev.1

[3] Davide Fioriti, Antonio Frangioni, Davide Poli – Optimal sizing of energy communities with fair revenue sharing and exit clauses: Value, role and business model of aggregators and users Applied Energy, 1 October 2021, DOI ???

[4] Commissione Europea – Comunicazione al parlamento europeo, al consiglio, al comitato economico e sociale europeo e al comitato delle regioni. Il Green Deal Europeo Bruxelles, 11.12.2019 COM (2019) 640 final

[5] Sito A2A Life Company - <https://magazine.a2aenergia.eu/energia/comunita-energetiche-che-cosa-sono-e-dove-sono-italia>

[6] RSE – Le Comunità Energetiche in Italia, Orange Book 2022

[7] Direttiva 2018/2001/UE del Parlamento europeo e del Consiglio, del 11 dicembre 2018, sulla promozione dell'uso dell'energia da fonti rinnovabili. Journal ufficiale dell'Unione europea, L 328/82, 21.12.2018

[8] Garavaso, Pierpaolo, et al. "Optimal Planning and Operation of a Residential Energy Community under Shared Electricity Incentives." Energies 14.8 (2021): 2045

[9] Ma, Tengfei, et al. "Energy flow matrix modeling and optimal operation analysis of multi energy systems based on graph theory." Applied Thermal Engineering 146 (2019): 648-663

[10] Aura Caramizaru, Andreas Uihlein – Energy communities : an overview of energy and social innovation European Commission, JRC Science for policy report 2020

[11] Consiglio europeo – Conclusioni sul quadro 2030 per le politiche dell'energia e del clima Bruxelles, 23 ottobre 2014 (OR. En) SN 79/14

[12] Pacchetto Fit for 55. Commissione europea. 14 luglio 2021. Proposta di revisione del sistema di scambio delle quote di emissione dell'UE del pacchetto Fit for 55

[13] Green Deal europeo. Commissione europea. 2019. Proposta di legge sul clima europea del Green Deal europeo

[14] Francesco Demetrio Minuto, Andrea Lanzini, Lorenzo Giannuzzo, Romano Borchiellini – Digital platforms for Renewable Energy Communities projects: an overview IOP Science, 09 June 2022, DOI 10.1088/1755-1315/1106/1/012007

[15] GSE – Energia e Clima in Italia, rapporto trimestrale, Novembre 2022

[16] ARERA - Le Comunità Energetiche in Italia, una guida per orientare i cittadini nel nuovo mercato dell'energia

[17] GSE – Gruppi di autoconsumatori di energia rinnovabile che agiscono collettivamente e comunità di energia rinnovabile, Regole tecniche per l'accesso al servizio di valorizzazione e incentivazione dell'energia elettrica condivisa, Aprile 2022

[18] RSE Schemi di autoconsumo collettivo e comunità energetiche - <https://dossierse.it/17-2020-gli-schemi-di-autoconsumo-collettivo-e-le-comunita-dellenergia/>

[19] Rapporto Legambiente 2022 Comunità Rinnovabili – Ufficio Energia di Legambiente, Maggio 2022

[20] Rapporto Legambiente 2021 Comunità Rinnovabili – Ufficio Energia di



Legambiente, Maggio 2021

[21] A. Lanzini et al. – Guida alle Comunità Energetiche Rinnovabili a impatto sociale – Versione 1 del 13.04.2022, Fondazione Compagnia di San Paolo

[22] S. Neelamegam, Dr. E. Ramaraj – Classification algorithm in Data mining: Na Overview International Journal of P2P Network Trends and Technology (IJPTT) – Volume 4 Issue 8, 08 September 2013

[23] Christopher J.C. Burges – A tutorial on Support Vector Machines for Patter Recognition

[24] European Commission – Paris Climate Conference, Parigi, Dicembre 2015

[25] RESCOOP - <https://www.rescoop.eu/transposition-tracker>

[26] Pérez-Chacón, R.; Luna-Romera, J.M.; Troncoso, A.; Martínez-Álvarez, F.; Riquelme, J.C. Big Data Analytics for Discovering Electricity Consumption Patterns in Smart Cities. *Energies* 2018, 11, 683. <https://doi.org/10.3390/en11030683>

[27] Maher Azaza, Fredrik Wallin, Smart meter data clustering using consumption indicators: responsibility factor and consumption variability, *Energy Procedia*, Volume 142, 2017, Pages 2236-2242, ISSN 1876-6102, <https://doi.org/10.1016/j.egypro.2017.12.624>

[28] S. Ghosh, D. K. Panda, S. Das and D. Chatterjee, "Cross-Correlation Based Classification of Electrical Appliances for Non-Intrusive Load Monitoring," 2021 International Conference on Sustainable Energy and Future Electric Transportation (SEFET), Hyderabad, India, 2021, pp. 1-6, doi: 10.1109/SeFet48154.2021.9375687.

[29] Qiaochu Wang, Yan Ding, Xiangfei Kong, Zhe Tian, Linrui Xu, Qing He, Load pattern recognition based optimization method for energy flexibility in office buildings, *Energy*, Volume 254, Part C, 2022, 124475, ISSN 0360-5442, <https://doi.org/10.1016/j.energy.2022.124475>.

[30] June Young Park, Xiya Yang, Clayton Miller, Pandarasamy Arjunan, Zoltan Nagy, Apples or oranges? Identification of fundamental load shape profiles for benchmarking buildings using a large and diverse dataset, *Applied Energy*, Volume 236, 2019, Pages 1280-1295, ISSN 0306-2619, <https://doi.org/10.1016/j.apenergy.2018.12.025>.

[31] Tureczek, A.M.; Nielsen, P.S. Structured Literature Review of Electricity Consumption Classification Using Smart Meter Data, *Energies* 2017, 10, 584. <https://doi.org/10.3390/en10050584>.

[32] M. Fang et al., "Data-Driven Load Pattern Identification," 2021 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia), Chengdu, China, 2021, pp. 568-573, doi: 10.1109/ICPSAsia52756.2021.9621531.

[33] M. Fang et al., "Data-Driven Load Pattern Identification," 2021 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia), Chengdu, China, 2021, pp. 568-573, doi: 10.1109/ICPSAsia52756.2021.9621531.

[34] M. Batou, G. J. Tsekouras and A. Moronis, "Load estimation of a small autonomous island power system using k-means method," 2019 3rd European Conference on Electrical Engineering and Computer Science (EECS), Athens, Greece, 2019, pp. 60-65, doi: 10.1109/EECS49779.2019.00024.

[35] László Czétány, Viktória Vámos, Miklós Horváth, Zsuzsa Szalay, Adrián Mota-Babiloni, Zsófia Deme-Bélafi, Tamás Csoknyai, Development of electricity consumption profiles of residential buildings based on smart meter data clustering, *Energy and Buildings*, Volume 252, 2021, 111376, ISSN 0378-7788, <https://doi.org/10.1016/j.enbuild.2021.111376>.

[36] Sicheng Zhan, Zhaoru Liu, Adrian Chong, Da Yan, Building categorization revisited: A clustering-based approach to using smart meter data for building energy benchmarking, *Applied Energy*, Volume 269, 2020, 114920, ISSN 0306-2619, <https://doi.org/10.1016/j.apenergy.2020.114920>.

[37] Behzad Najafi, Monica Depalo, Fabio Rinaldi, Reza Arghandeh, Building characterization through smart meter data analytics: Determination of the most

influential temporal and importance-in-prediction based features, *Energy and Buildings*, Volume 234, 2021, 110671, ISSN 0378-7788, <https://doi.org/10.1016/j.enbuild.2020.110671>.

[38] Anna Marszal-Pomianowska, Per Heiselberg, Olena Kalyanova Larsen, Household electricity demand profiles – A high-resolution load model to facilitate modelling of energy flexible buildings, *Energy*, Volume 103, 2016, Pages 487-501, ISSN 0360-5442, <https://doi.org/10.1016/j.energy.2016.02.159>.

[39] Marco Savino Piscitelli, Silvio Brandi, Alfonso Capozzoli, Recognition and classification of typical load profiles in buildings with non-intrusive learning approach, *Applied Energy*, Volume 255, 2019, 113727, ISSN 0306-2619, <https://doi.org/10.1016/j.apenergy.2019.113727>.

[40] Energy consumption in the UK 2021 - <https://www.gov.uk/government/statistics/energy-consumption-in-the-uk-2021>

[41] L. Bottaccioli, S. Di Cataldo, A. Acquaviva and E. Patti, "Realistic Multi-Scale Modeling of Household Electricity Behaviors," in *IEEE Access*, vol. 7, pp. 2467-2489, 2019, doi: 10.1109/ACCESS.2018.2886201

[42] Electricity billing in the UK – Wikipedia - [https://en.wikipedia.org/wiki/Electricity\\_billing\\_in\\_the\\_UK](https://en.wikipedia.org/wiki/Electricity_billing_in_the_UK)

[43] Schofield, James & Tindemans, Simon & Carmichael, Richard & Woolf, Matt & Bilton, Mark & Strbac, G.. (2016). Low Carbon London project: Data from the dynamic time-of-use electricity pricing trial, 2013. 10.5255/UKDA-SN-7857-1

[44] Melović, Boban, and Dragana Ćirović. "Analysis of financial incentives as an instrument of renewable energy sources management in Montenegro." *E3S Web of Conferences*. Vol. 157. EDP Sciences, 2020.

[45] Emi Minghui Gui, Iain MacGill, Typology of future clean energy communities: An exploratory structure, opportunities, and challenges, *Energy Research & Social Science*, Volume 35, 2018, Pages 94-107, ISSN 2214-6296, <https://doi.org/10.1016/j.erss.2017.10.019>

- [46] Iglesias, Félix, and Wolfgang Kastner. "Analysis of similarity measures in times series clustering for the discovery of building energy patterns." *Energies* 6.2 (2013): 579-597
- [47] Vladimir Z. Gjorgievski, Snezana Cundeva, George E. Georghiou, Social arrangements, technical designs and impacts of energy communities: A review, *Renewable Energy*, Volume 169, 2021, Pages 1138-1156, ISSN 0960-1481, <https://doi.org/10.1016/j.renene.2021.01.078>
- [48] Gabriele Volpato, Gianluca Carraro, Marco Cont, Piero Danieli, Sergio Rech, Andrea Lazzaretto, General guidelines for the optimal economic aggregation of prosumers in energy communities, *Energy*, Volume 258, 2022, 124800, ISSN 0360-5442, <https://doi.org/10.1016/j.energy.2022.124800>
- [49] Li, Lei, and B. Aditya Prakash. "Time series clustering: Complex is simpler!." *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*. 2011
- [50] Natconsumers. <http://www.natconsumers.eu>. Accessed 28 Aug 2015
- [51] Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural equation modeling*, 14(4), 535-569. <https://doi.org/10.1080/10705510701575396>
- [52] Hartigan, J. A. and Wong, M. A. (1979). Algorithm AS 136: A K-means clustering algorithm. *Applied Statistics*, 28, 100--108. 10.2307/2346830
- [53] Lloyd, S. P. (1957, 1982). Least squares quantization in PCM. Technical Note, Bell Laboratories. Published in 1982 in *IEEE Transactions on Information Theory*, 28, 128—137
- [54] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.

## Ringraziamenti

Un ringraziamento va a tutte quelle persone che hanno dedicato il loro tempo, la loro pazienza e il loro supporto alla realizzazione di questo progetto di tesi.

Ringrazio il mio relatore, il Professore Andrea Lanzini, per avermi accolto come un figlio all'interno del gruppo di ricerca dell'Energy Center, e per avermi dato la possibilità di poter esprimere le mie potenzialità in un contesto nuovo, come quello delle Comunità Energetiche, sfruttando a pieno le conoscenze maturate durante il mio percorso di laurea, soprattutto nel campo del *data analytics*, la mia passione. Un sentito grazie per avermi insegnato a gestire i momenti più difficili e complicati con calma ed intelligenza, ma soprattutto con il sorriso, senso dell'humor, serietà e anche tanta ironia.

Ringrazio il mio co-relatore, il Dottor Francesco Demetrio Minuto, per avermi forgiato nel tempo, rimediando alle mie lacune e indicandomi la giusta via nei momenti bui, e per avermi insegnato la disciplina, la dedizione al lavoro e il rispetto per il lavoro altrui.

Ringrazio il Professor Romano Borchiellini, per l'opportunità di condurre il lavoro di tesi e la mia attività di ricerca presso l'Energy Center del Politecnico di Torino.

Ringrazio i miei genitori, i miei primissimi sostenitori, coloro che non hanno mai dubitato delle mie qualità e che hanno sempre cercando di indirizzarmi per la strada migliore. Un grazie di cuore per avermi insegnato la disciplina e l'educazione, e per avermi trasmesso la passione per il lavoro. Grazie per i numerosi confronti avuti nel tempo, che mi hanno reso forte e sicuro del percorso che ho scelto, e che mi hanno reso sempre più determinato a raggiungere i miei obiettivi, a qualunque costo. Grazie per i mille sacrifici fatti in tutti questi anni, dei vostri consigli e del supporto avuto, senza il quale non sarei arrivato a questo magnifico traguardo. Spero con tutto il cuore di aver riscattato la mia famiglia, ottenendo finalmente quella laurea da voi tanto desiderata ma mai conseguita, non per demerito ma per vicissitudini della vita.

Ringrazio mio fratello, compagno di stanza e di vita in questi cinque anni a Torino, per il sostegno e il supporto avuto per tutta la durata del mio percorso. Grazie per tutte le parole di conforto ricevute nei momenti bui, e per il sorriso e l'ironia che ci ha sempre accompagnato nei momenti più difficili.

Ringrazio i miei "mesci", Rocco, Eugenio e Gregorio, per gli incredibili momenti passanti insieme. Le risate, gli scherzi e le avventure vissute insieme saranno per sempre parte di me, e rappresentano quanto di più caro ho raccolto in questo meraviglioso percorso. Un grazie particolare a Rocco, fedele compagno di università, con il quale sono cresciuto insieme, sia come ingegnere che come uomo, e con il quale ho condiviso il campo di battaglia per tutta la durata del percorso di studi, vincendo e perdendo insieme, ed aiutandoci a vicenda nei momenti di difficoltà.

Ringrazio i miei "leccesi", Elena, Francesco, Davide M., Luca C., Davide F., i miei conterranei, la mia seconda famiglia. Grazie per aver portato gioia e allegria nelle serate grigie di Torino, per l'ironia e per l'affetto dimostrato da sempre. Grazie per aver riportato nella mia vita quel pizzico di Salento, che tanto mi era mancato in questi anni.

Ringrazio i miei compagni di università, Fabio, Alessandro, Davide, Nino, Agnese, Luca, Stefano D., Stefano M. Maddalena, Gabriele, Philippe, Ivan, Salvatore, Federico, Samuele e Simone per essermi stato vicino e per aver reso questa esperienza unica e irripetibile. Un gruppo coeso, capace di darsi forza e di stimolare le qualità dei singoli, agendo da squadra contro le sfide affrontate in questo cammino.

Ringrazio i miei amici più stretti, Alfreda, Gianluca, Simone, Roberto, Renato, Martina, Antonia, Gregorio, Andrea, Elisa, Alice, Cristina e Rebecca, che da sempre mi accompagnano e mi sostengono nei momenti felici e di difficoltà, e che sono sempre stati presenti, in ogni parte del mondo e in ogni contesto. Un grazie di cuore per i bellissimi momenti passati insieme, per i consigli e per le lezioni di vita, che mi hanno formato come persona e come uomo. Grazie per avermi sostenuto, anche a migliaia di chilometri di distanza, il nostro legame non conosce

la distanza. Sono un uomo veramente fortunato, ad avervi incontrato nella vita.

Un grazie super speciale all'Ufficio 7, ad Alessandro, Federico, Filippo, Sabatino, Antonio e Franco, fedeli compagni di lavoro, sempre presenti e disponibili in qualunque difficoltà lavorativa e non. Grazie per i momenti di formazione passati insieme, per l'ottimo ambiente creato e per avermi cresciuto come lavoratore e come persona, arricchendomi delle vostre esperienze e dei vostri trascorsi. Grazie per tutte le avventure e per i momenti di difficoltà superati come gruppo granitico e determinato. Un ringraziamento particolare a Sabatino ed Antonio, inizialmente sotto la mia guida, vere e proprie rivelazioni che hanno superato in tempo record i loro maestri, me compreso, dimostrando di essere persone estremamente valide e competenti, nonché ottimi amici.

Grazie ai miei nonni, i miei maestri di vita, coloro che mi hanno insegnato che alla fine, il duro lavoro ripaga sempre. Grazie a mio nonno Luigi G., per avermi insegnato l'autorevolezza e il senso di sacrificio, la generosità e la laboriosità. Grazie per avermi insegnato il rispetto dei ruoli e delle regole, e per aver continuato a proteggermi, anche da lassù. Grazie a mio nonno Luigi C., per avermi insegnato l'audacia, l'ironia e la serietà, per avermi forgiato nel carattere e nello spirito, e per avermi regalato i momenti più belli della mia vita. Grazie a mia nonna Anatolia Vittoria, per avermi insegnato il rispetto di se stessi, l'astuzia e la fiducia nei propri mezzi. Grazie per avermi accompagnato per mano, da semplice bambino che ero fino all'uomo che sono oggi, mi manchi veramente tanto. Grazie a mia nonna Michela, per avermi insegnato la gioia di vivere, l'espressività, la sincerità e l'arte della comunicazione, l'unica persona della mia vita con il quale basta solo uno sguardo, per dirci tutto. Grazie per avermi portato fino a qui, resistendo fino all'ultimo pur di vedermi raggiungere questo prestigioso traguardo, sarai sempre nei miei pensieri, qualunque cosa accada.

Grazie ai miei zii, cugini e parenti, che hanno accolto con entusiasmo l'ultimo arrivato della famiglia e lo hanno visto crescere, fino a diventare la persona che sono oggi.

Infine, grazie a me stesso, senza troppi vantì e senza troppe glorie. Sarà bello,

rileggendo queste ultime righe, ripensare a tutto quello che ho fatto, per arrivare fin qui.

*Lorenzo Giannuzzo*