

POLITECNICO DI TORINO  
CORSO DI LAUREA MAGISTRALE IN INGEGNERIA  
ELETTRICA



**Politecnico  
di Torino**

TESI DI LAUREA

**Analisi di dati sulle forniture elettriche:  
campionamento stratificato e procedure di clustering su  
utenze trifase e monofase**

Relatore:

Prof. Gianfranco Chicco

Candidato:

Daniele Bonansinga

*Marzo 2022*

# Sommario

<b>POLITECNICO DI TORINO .....</b>	<b>1</b>
<b>Introduzione.....</b>	<b>4</b>
<b>1. Acquisizione dati: contatori 1G e 2G .....</b>	<b>7</b>
<b>1.1. Concetti sulla misurazione dell'energia .....</b>	<b>7</b>
<b>1.2. Smart metering 1G.....</b>	<b>8</b>
<b>1.3. Smart metering 2G.....</b>	<b>12</b>
<b>2. Profili di carico.....</b>	<b>17</b>
<b>3. Approccio statistico per determinare i diversi livelli di clienti: campionamento stratificato.....</b>	<b>25</b>
<b>3.1. Teoria del campionamento stratificato.....</b>	<b>27</b>
<b>4. Applicazione del metodo.....</b>	<b>32</b>
<b>4.1. Caso studio: Dati trifase, utenti non residenziali .....</b>	<b>32</b>
<b>4.1.1. Campionamento stratificato avendo come variabile discriminante la potenza contrattuale degli utenti non residenziali trifase.....</b>	<b>35</b>
<b>4.1.2. Campionamento stratificato avendo come variabile discriminante l'energia consumata degli utenti non residenziali trifase .....</b>	<b>38</b>
<b>4.1.3. Ripartizione finale degli utenti.....</b>	<b>41</b>
<b>4.2. Caso studio: dati monofase, utenti residenziali .....</b>	<b>43</b>
<b>4.2.1. Ripartizione finale degli utenti.....</b>	<b>47</b>
<b>5. Metodi di cluster nell'era dei big data.....</b>	<b>48</b>
<b>5.1. Algoritmi di clustering.....</b>	<b>49</b>
<b>5.2. Caso studio: Clustering degli utenti Ireti .....</b>	<b>52</b>
<b>5.2.1. Caso studio 1: Clustering K-means dei dati secondo la potenza contrattuale per i dati trifase mensili .....</b>	<b>52</b>
<b>5.2.2. Caso studio 2: Clustering k-means dei dati secondo l'utilizzazione per i dati trifase mensili .....</b>	<b>56</b>
<b>5.2.3. Caso studio 3: Clustering K-means dei dati secondo la potenza e l'utilizzazione per i dati monofase residenziali.....</b>	<b>60</b>
<b>5.3. Metodo di Clustering K-means dei dati tenendo conto di più variabili.....</b>	<b>63</b>
<b>5.3.1. Caso studio 4: Clustering K-means bidimensionale tenendo conto di energia e potenza.....</b>	<b>63</b>

5.3.2. Caso studio 5: Clustering K-means bidimensionale tenendo conto di utilizzazione e potenza.....	64
5.3.3. Caso studio 5.1: Clustering K-means bidimensionale tenendo conto dell'utilizzazione normalizzata e della potenza.....	66
5.3.4. Caso studio 6: Clustering K-means bidimensionale tenendo conto della potenza e dell'energia facendo il clustering sull'utilizzazione.....	67
<b>6. Determinazione delle classi di clienti studiando le curve di carico.....</b>	<b>69</b>
6.1. Creazione delle micro-classi tramite procedure di clustering.....	74
6.2. Creazione delle micro-classi tramite il metodo <i>k-means</i> .....	76
6.3. Creazione delle micro-classi tramite il clustering gerarchico.....	80
6.3.1. Cluster delle curve di carico rappresentative mensili.....	81
6.3.2. Cluster delle curve di carico rappresentative annuali.....	85
<b>7. Analisi dei dati sul consumo: confronto 2019-2020.....</b>	<b>89</b>
<b>8. Conclusioni.....</b>	<b>93</b>
<b>9. SITOGRAFIA.....</b>	<b>94</b>
<b>10. BIBLIOGRAFIA.....</b>	<b>95</b>
<b>RINGRAZIAMENTI.....</b>	<b>96</b>

## Introduzione

La progressiva installazione degli smart meter di ultima generazione sugli utenti finali sta mettendo a disposizione un numero gigantesco di dati, poiché essi riescono a raccogliere informazioni in tempi uguali o inferiori al quarto d'ora. Questi dati devono essere acquisiti, salvati ed elaborati. Gestire il flusso continuo di milioni di dati non è semplice; infatti, sono necessari strumenti e metodologie adeguate a migliorare la corretta analisi dei dati poiché troppi dati disponibili difficili da elaborare portano a nessuna informazione utile. L'impossibilità di analizzare e salvare tutte le informazioni ricavate dai contatori porta a studiare delle metodologie per determinare il corretto numero di dati da analizzare cercando però di abbassare il più possibile la soglia di errore nella stima finale.

Le metodologie presentate nell'elaborato per l'analisi sulle curve di carico sono la teoria del campionamento stratificato e due principali procedure di clustering: k-means e clustering gerarchico. Grazie ad esse sarà possibile ridurre l'onere dei dati, da più di un milione a qualche migliaio, potendo così studiare al meglio le caratteristiche di consumo dei diversi utenti al fine di creare, grazie alle procedure di clustering, classi di utenti sulla base della similarità. In questa circostanza sarà possibile una miglior comprensione del sistema elettrico generale e come obiettivo finale si potranno studiare dei miglioramenti del servizio all'utente aumentando la sua proattività e la consapevolezza dei propri consumi permettendo così un perfezionamento dei piani tariffari portando ad un risparmio nei consumi. Inoltre, grazie alla conoscenza dei consumi degli utenti, si potrà avere un miglioramento del controllo della flessibilità da parte dell'ente distributore oltre che un miglioramento dell'efficienza operativa del sistema.

La curiosità su questo argomento è nata quando sono venuto a conoscenza che uno dei più grandi enti di distribuzione elettrica italiana ha presentato il piano per la messa in servizio degli smart meter di ultima generazione. Dopo aver appreso le enormi potenzialità e vantaggi dei contatori 2G mi sono posto una domanda: come riuscire a migliorare l'analisi dei consumi degli utenti finali?

Questo è stato il punto di partenza del mio elaborato. Dopo aver contattato il responsabile dell'area metering energia elettrica di Ireti S.p.A, ovvero, l'ingegner Renato Mulinacci, sono venuto in possesso dei dati delle forniture elettriche trifase e monofase,

degli anni 2019 e 2020 raccolti dagli smart meter installati, che hanno permesso di iniziare la mia analisi.

Avere a disposizione dati di questi due anni ha permesso, nell'ultimo capitolo, di fare anche un confronto sui consumi pre e post pandemia.

Ulteriori sviluppi della mia ricerca possono portare all'aggiornamento e al perfezionamento di alcune procedure sull'analisi di dati, le quali sono essenziali per rispondere alle necessità di un sistema in continua evoluzione come quello elettrico, nonché all'inclusione dello studio di nuovi carichi anche di tipo attivo, ad esempio la generazione distribuita e sistemi di accumulo, che stanno segnando la transazione energetica permettendo il corretto passaggio dell'elettrificazione degli utenti.



# 1. Acquisizione dati: contatori 1G e 2G

## 1.1. Concetti sulla misurazione dell'energia

Quando si effettuano misure elettriche tramite i contatori bisogna dividere le misure in quantità istantanee e quantità integrali.

Le quantità istantanee sono quantità non misurabili nel tempo, poiché istantanee, mentre le quantità integrali sono misurabili e il loro significato varia al trascorrere di un certo intervallo di tempo. Nelle applicazioni della misura dell'energia, il concetto rilevante è la precisione con cui viene misurata l'energia, preservando le quantità integrali. Il dato fondamentale per i nostri calcoli è la potenza media, la potenza in un certo intervallo di tempo, dunque l'energia. È importante specificare che dall'energia misurata non è possibile dedurre l'evoluzione della potenza media per intervalli di tempo inferiori a quello utilizzato per la misura. Per questa motivazione, è molto importante specificare l'intervallo del campionamento, poiché più è lungo l'intervallo del campionamento più vi saranno perdite delle informazioni relative alle veloci variazioni della potenza media.

Tutte le normative nazionali e internazionali fanno riferimento a un intervallo di campionamento, dove tutti i tempi di passo hanno la stessa durata che varia dal quarto d'ora all'ora, e dove il modello di potenza è rappresentato come potenza media in fasi temporali successive.

Da queste poche cose dette possiamo trovare un'importante conclusione, ovvero è importante ridurre al minimo il passo di campionamento per aver una misurazione più accurata ma ciò comporta un aumento del numero di dati inviati ai canali di comunicazione che può risultare superiore al numero di dati gestibili.

## 1.2. Smart metering 1G

Il sistema di acquisizione dati analizzato successivamente è quello utilizzato da IRETI S.p.A, esso è un sistema di vecchia generazione ma ancora molto utilizzato. Il sistema di *smart metering* 1G è costituito da una serie di apparecchiature hardware e software che riescono a raccogliere i dati di misura dei clienti collegati alla rete di distribuzione e di amministrare e coordinare da remoto gli apparati di campo.

### Architettura

I principali componenti che costituiscono il sistema di *smart metering* 1G sono:

- Sistemi di campo: insieme di dispositivi collocati sulla rete BT e presso le utenze finali. Possiamo distinguere tre principali dispositivi.
  1. *Contatori elettronici*. Essi sono fondamentali per la misurazione e per il trasferimento dei dati ai concentratori.
  2. *I concentratori*. I quali devono memorizzare i dati sulle utenze e mandarli al sistema centrale, contemporaneamente possono attribuire comandi dal sistema centrale ai contatori elettronici.
  3. *Modem e antenne*, fondamentali per la telegestione.
- Sistema centrale: composto da diversi dispositivi fondamentali (*operating system, server, computer, database*) che permettono il corretto funzionamento del software centrale di telegestione e la corretta operatività dei concentratori.
- Apparati mobili.

Il contatore 1G interagisce con i concentratori grazie ad una connessione in BT chiamata PLC - *power line communication*, mentre i concentratori, situati in ciascuna cabina MT/BT, trasmettono informazioni al sistema centrale tramite una rete di telecomunicazioni GSM/GPRS.

Come detto sopra tutti i sistemi sono bidirezionali, infatti, il sistema centrale può sia ricevere i dati dal concentratore che trasmettere dati e impartire comandi al misuratore, sempre passando dal concentratore.

L'attuale sistema 1G di IRETI è caratterizzato da un singolo canale per la trasmissione dei dati di misura fra misuratore e concentratore.

In Figura 1 è riportato uno schema semplificato dell'architettura del sistema.

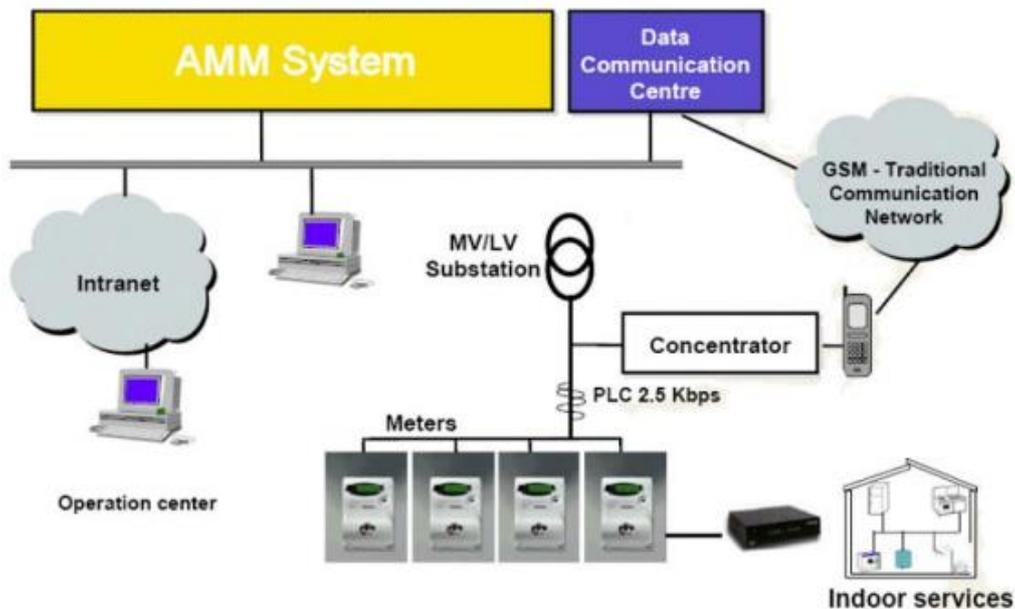


Figura 1 - Schema semplificato architettura smart metering 1G- ireti.it/documents

## Funzionalità e prestazioni

Di seguito verranno riportate alcune principali caratteristiche funzionali dei diversi componenti.

### Contatori 1G

I contatori 1G hanno la caratteristica di misurare, memorizzare e trasmettere le curve di carico per fasce orarie per il mese corrente e per il mese precedente, inoltre, possono lavorare in modalità remota permettendo di cambiare, diminuire o aumentare, la potenza utilizzabile. L'abilità di memorizzare i consumi di energia, sia attiva che reattiva, permette di dividere le tariffe per differenti fasce.

Il periodo di utilizzo e di funzionamento di questa tipologia di *smart meter* è pari o superiore a 15 anni.

## **Concentratori 1G**

Le funzionalità principale dei concentratori 1G è quella di raccogliere e immagazzinare diversi tipi di dati, ad esempio: dati di misura di tensione e corrente, dati sulle curve di carico, dati di allarme relativi a qualsiasi tipo di guasto o malfunzionamento e dunque dati riguardanti la *power quality* del sistema in generale.

## **Il sistema centrale 1G**

Il sistema centrale 1G permette di controllare e organizzare a livello centralizzato tutte le operazioni che riguardano nuove installazioni e attivazioni di *smart meter* e coordina tutti gli interventi necessari da fare in loco. L'aggiornamento della rete fisica e logica è di necessaria importanza per mappare i contatori ed eseguire le lavorazioni precedentemente elencate.

Il livello di prestazione del sistema 1G permette di evidenziare il consumo energetico mensile tramite telelettura per il 96% dei misuratori e nell'esecuzione da remoto, telegestione, degli ordini di lavoro relativi alla gestione utenza, nell'84% dei casi. Per effettuare l'acquisizione dei dati del contatore o effettuare alcune lavorazioni di gestione utenza nel caso in cui vi è l'impossibilità della telelettura o telegestione è necessario intervenire recandosi sul posto.

## **Criticità emerse del sistema di *smart metering* 1G**

Nell'arco degli ultimi anni si sono riscontrate diverse criticità che hanno evidenziato quanto questo sistema di misurazione sia obsoleto.

L'incessante aumento nel numero di dati è una delle maggiori problematiche di questi contatori poiché all'epoca della loro installazione essi erano limitati alla raccolta delle letture da remoto una volta al mese e all'esecuzione da remoto delle attività di gestione utenza, come da norma vigente. Inoltre, la mancanza di un canale di back-up tra contatori e concentratori favorisce la presenza di forti disturbi che causano attenuazioni

del segnale lungo la linea elettrica, per questo motivo sono state riscontrate diverse problematiche durante alcune attività di telelettura e telegestione. Un'altra criticità emersa negli ultimi anni è stata la mancanza di un ulteriore canale di comunicazione isolato, oltre al PLC, che permetta al cliente finale la conoscenza dei propri consumi.

Oltre alle criticità descritte precedentemente, nel corso degli ultimi anni, la frequenza dei guasti dei concentratori dovuta ad alcune componenti del sistema obsolete è aumentata notevolmente. Tutte queste problematiche hanno permesso lo sviluppo di una nuova generazione di contatori.

### 1.3. Smart metering 2G

Tutte le più grandi società di distribuzione, in Italia, stanno sostituendo gli *smart meter* 1G con i nuovi *smart meter* 2G, questo cambiamento avviene in base al decreto legislativo n.102/2014 in materia di efficienza energetica che recepisce la Direttiva europea 27/2012. I distributori di energia elettrica si sono dovuti adattare a dei nuovi sistemi di misurazione intelligenti che soddisfino i requisiti di questo decreto. Inoltre, i distributori sono obbligati da ARERA a garantire un nuovo e sicuro modello di contatori ai propri clienti ogni 15 anni. Ciò permetterà, oltre al rinnovo completo dei misuratori di campo, di superare le criticità dell'attuale sistema 1G evidenziate nel paragrafo precedente.

#### Architettura 2G

I principali componenti che costituiscono il sistema di *smart metering* 2G sono:

- *Il contatore elettronico*: Il quale riesce a campionare i consumi e ad inviare i dati ai concentratori.
- *Il concentratore*: inoltra i dati campionati dal contatore al sistema centrale
- *Il sistema centrale*: elabora e memorizza i dati di misura e li inoltra ad altri sistemi per il successivo invio al Sistema Informativo Integrato. Inoltre, il sistema centrale permette la gestione del sistema da remoto; le comunicazioni fra misuratore, concentratore e sistema centrale sono infatti bidirezionali.

La principale differenza tra il sistema di vecchia generazione e il sistema di nuova generazione è che il nuovo sistema disporrà di due canali indipendenti per la comunicazione tra contatore e concentratore.

- **Canale primario**: *Power Line Communication* (PLC). Grazie a questo canale sarà possibile la comunicazione bidirezionale fra misuratore e concentratore tramite la linea elettrica in BT.

- **Canale di back-up:** Radio Frequenza (RF) nella banda di frequenza 169 MHz, che utilizzerà la comunicazione via etere per la trasmissione dei dati bidirezionali tra contatore e concentratore in caso di indisponibilità del canale primario.

Il nuovo sistema 2G permetterà la comunicazione diretta fra il misuratore e gli eventuali dispositivi utente (DU), così facendo il cliente finale potrà avere accesso alle informazioni relative ai propri consumi. Nella figura 2 possiamo osservare un semplice schema di come è organizzata l'architettura del nuovo sistema 2G.

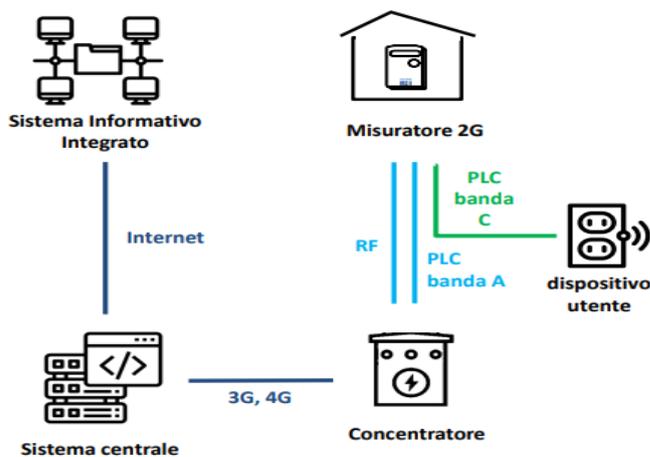


Figura 2 - Schema semplificato architettura smart metering 2G- ireti.it/documents

### Funzionalità e prestazioni

Il contatore 2G, essendo di nuova generazione, ha alcune funzionalità nuove e più efficienti rispetto al contatore di vecchia generazione. Il nuovo *smart meter 2G* supporta la programmazione di una struttura settimanale. Ciò significa che è possibile associare diversi piani tariffari a seconda della giornata e a seconda di un intervallo di tempo definito. Per ogni giorno è possibile associare fino a sei tariffe. Inoltre, è possibile definire una struttura giornaliera dedicata per i giorni festivi. Oltre a riuscire a memorizzare i valori più alti di potenza attiva e reattiva quart'oraria, il nuovo contatore memorizza le quantità di energia attiva e reattiva prelevata o immessa per ogni fascia tariffaria di un determinato periodo temporale. Avendo a disposizione una memoria il nuovo dispositivo riesce a salvare il valore massimo di potenza giornaliera e a

memorizzarlo per i successivi 38 giorni. Grazie al display di cui sono dotati, i nuovi contatori possono rendere visibili agli utenti le loro curve di carico con intervalli di campionamento controllabili, dunque, quart'orari o orari.

L'inedito *sistema centrale 2G* consentirà la telelettura e la telegestione di ciascuno dei diversi dispositivi di campo, garantendo maggiori livelli di prestazione rispetto alla vecchia generazione. Esso sarà responsabile della gestione e configurazione dei dispositivi installati sul campo, della raccolta dei dati di misura e dell'esecuzione delle richieste di gestione utenza.

Dopo ulteriori analisi e molteplici studi su questi nuovi dispositivi, si è potuto affermare che i livelli di prestazione attesi dagli *smart metering 2G* permetteranno disponibilità giornaliera al Sistema Informativo Integrato delle curve quart'orarie di energia con aggiornamento giornaliero secondo i seguenti livelli di prestazione:

- il 95% dei misuratori 2G entro 24 ore dalla mezzanotte del giorno di consumo.
- il 97% dei misuratori 2G entro 96 ore dalla mezzanotte del giorno di consumo

Principali caratteristiche e funzionalità degli *smart meter* 1G e 2G messe a confronto

Tabella 1: confronto smart meter 1G e 2G

1G	2G
<b>RACCOLTA DELLE LETTURE PER FASCIA ORARIA UNA VOLTA AL MESE</b>	<b>DISPONIBILITA' GIORNALIERA DELLE CURVE QUARTORARIE DELL'ENERGIA</b>
<b>TARIFFAZIONE SULLA BASE DI 4 FASCE ORARIE PREFISSATE</b>	<b>MAGGIOR NUMERO DI FASCE ORARIE (6 GIORNALIERE) CHE PERMETTONO DI SVILUPPARE SOLUZIONI FLESSIBILI PROFILATE, CARATTERIZZATE DALLA VARIANZA DI PREZZO TRA LE DIVERSE ORE DEL GIORNO</b>
<b>UN UNICO CANALE PER LA COMUNICAZIONE TRA CONCENTRATORI E SISTEMA CENTRALE ATTRAVERSO UNA RETE DI TELECOMUNICAZIONE GSM/GPRS</b>	<b>COMUNICAZIONE AD ALTA VELOCITA' 3G/4G SU RETE PUBBLICA TRA CONCENTRATORE E SISTEMA CENTRALE</b>
<b>ASSENZA DI UN CANALE DEDICATO ALLA COMUNICAZIONE LATO UTENTE</b>	<b>PRESENZA DI DUE CANALI INDIPENDENTI PER LA TRASMISSIONI DI DATI, CHE PERMETTE LA COMUNICAZIONE TRA IL CONTATORE ED EVENTUALE DISPOSITIVO DI PROPRIETA' DEL CLIENTE FINALE</b>

## **Benefici**

Molteplici sono i benefici del sistema *smart metering* 2G sia per gli utenti finali sia per il sistema elettrico nel suo complesso. Grazie al campionamento giornaliero dei dati di consumo, l'utente finale potrà essere a conoscenza dei consumi relativi al giorno precedente, dunque la bolletta sarà più comprensibile, permettendo al cliente di conoscere i propri comportamenti, e nel caso, l'utente potrà cambiarli per poter risparmiare accedendo a offerte innovative e personalizzate secondo le proprie esigenze e le proprie abitudini. Inoltre, le informazioni sulle interruzioni di servizio permetteranno di migliorare la pianificazione degli interventi sulla rete contribuendo nella tempestività della rivelazione del guasto e nel corretto ripristino della rete stessa riducendo al minimo i disservizi e garantendo il più possibile la continuità del servizio. La possibilità di avere numerose informazioni di tutti gli utenti a cadenza giornaliera permetterà la precisa allocazione delle perdite fisiche e dunque economiche agli operatori di mercato. Per concludere, le caratteristiche di telegestione degli *smart meter* 2G renderanno maggiormente efficiente l'esercizio della rete, riducendo e rendendo più agevoli e rapide alcune attività oggi svolte manualmente da personale tecnico operativo.

## 2. Profili di carico

Quando studiamo i profili di carico analizziamo l'evoluzione temporale della potenza, attiva e reattiva, ottenuta da misurazioni specifiche. La potenza è una variabile istantanea per cui le misurazioni dell'energia sono predefinite in intervalli temporali che possono essere di 15 minuti se utilizziamo i nuovi contatori 2G, altrimenti le misurazioni sono ad intervalli orari. L'importante è che le misurazioni dell'energia devono essere a intervalli regolari. Per costruire correttamente la curva dei profili di carico in ogni intervallo orario viene misurata la potenza media e viene memorizzata. Bisogna partire dalla misurazione dell'energia per ogni intervallo e poi dividendo quest'ultima con la durata dell'intervallo si calcola la potenza media, essa viene mantenuta costante per tutta la durata dell'intervallo temporale di riferimento. Riducendo l'intervallo di tempo, quindi prendono le misure con una minore frequenza di campionamento, si consente una migliore approssimazione della realtà quindi una miglior evoluzione della richiesta di potenza al carico aggregato.

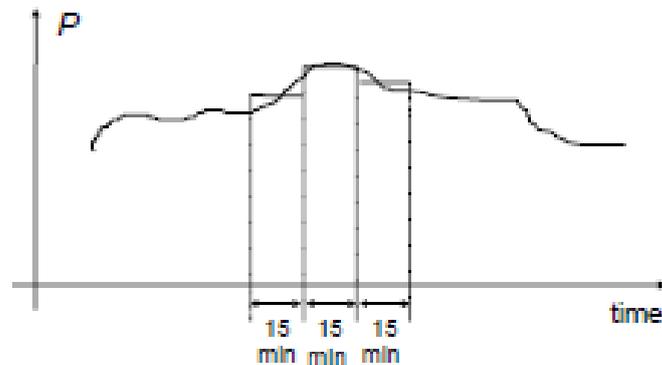


Figura 3: Esempio misurazione della potenza. Riferimento: materiale didattico dell'insegnamento "Distribuzione e utilizzazione dell'energia elettrica" (prof. Gianfranco Chicco, anno accademico 2020-2021)

Questo è il punto di partenza per costruire le LDC (*load duration curve*) che rappresentano, per ogni valore di potenza, la durata per la quale il valore di potenza è

stato raggiunto o superato. La potenza massima ha l'intervallo minimo di tempo mentre la potenza minima ha una durata pari all'intero periodo di tempo analizzato.

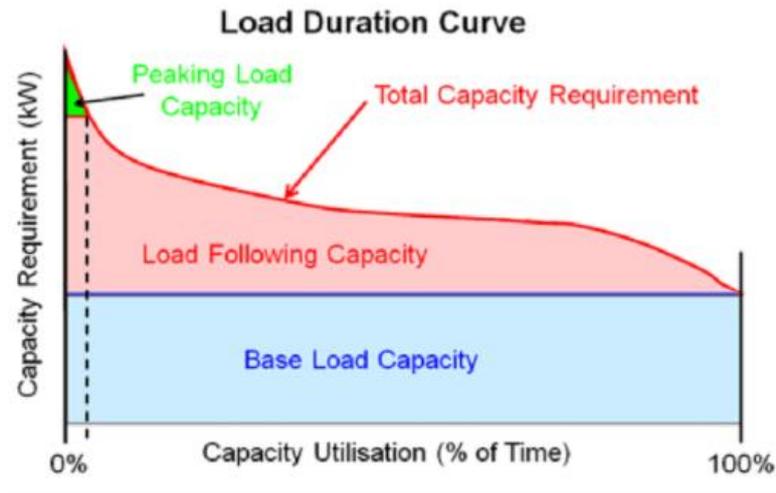


Figura 4: Esempio di LDC per un dato intervallo di tempo. Riferimento: materiale didattico dell'insegnamento "Distribuzione e utilizzazione dell'energia elettrica" (prof. Gianfranco Chicco, anno accademico 2020-2021)

### Categorizzazione degli utenti

Esistono diverse macrocategorie di utenti, esse sono contraddistinte dal diverso modo di utilizzare l'energia nell'arco della giornata, e soprattutto dall'energia usata. Le diverse macrocategorie sono:

- utenti residenziali
- utenti industriali
- utenti del settore terziario
- altre tipologie di utenti (illuminazione, trazione)

## Carico aggregato per utenti residenziali

Il sistema di distribuzione non alimenta il singolo utente residenziale, ma alimenta un aggregato di utenti residenziali fin alla cabina elettrica (prima della suddivisione capillare ai singoli utenti). Il carico aggregato è visto come unico carico che ingloba tutti i carichi della stessa categoria, esso è il carico da studiare quando si considerano gli utenti residenziali. Il consumo di un unico carico residenziale è molto volatile, esso dipende da numerose variabili: il numero di membri familiari, l'età, le attività che le persone svolgono e il loro stile di vita. Per questo bisogna unire un aggregato di carichi residenziali per avere una caratterizzazione dei consumi residenziali, essa è evidente con una analisi statistica basata sui vari aspetti che influenzano il consumo in una famiglia. Fortunatamente sotto la stessa sottostazione elettrica vi può essere un elevato numero di carichi residenziali, in questo modo l'analisi diventa molto più semplice.

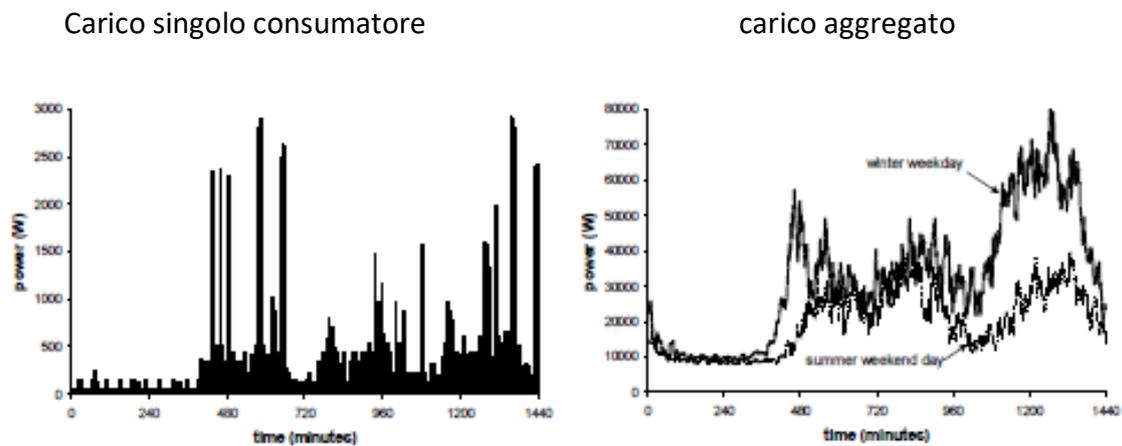
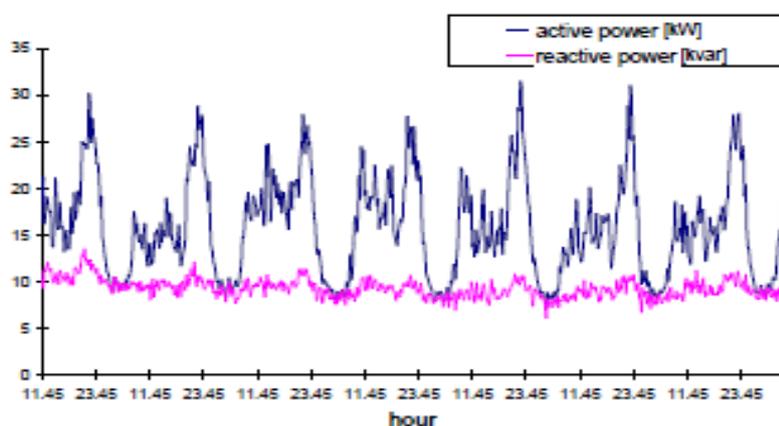


Figura 5: Esempio di potenza consumata da singolo carico (sx) e da carico aggregato (dx). Riferimento: materiale didattico dell'insegnamento "Distribuzione e utilizzazione dell'energia elettrica" (prof. Gianfranco Chicco, anno accademico 2020-2021)

Come si può notare, il carico singolo ha molte irregolarità che dipendono da aspetti non elettrici, citati precedentemente, e i picchi possono dipendere dall'uso di elettrodomestici ad alta potenza nei diversi momenti della giornata. Notiamo che la caratteristica

predominante per i singoli utenti residenziali è l'aver grandi picchi di consumo in tempi scarsamente prevedibili. Mentre se analizziamo l'aggregazione del carico residenziale possiamo notare un modello di consumo abbastanza regolare con aumento di carico nelle prime ore della mattina, una flessione del carico nelle ore centrali della giornata per poi riavere un picco di consumi serali. Se l'analisi del profilo di carico venisse fatta per diversi giorni, si potrebbe notare che la caratteristica del carico residenziale aggregato rimane quasi costante.



*Figura 6: Esempio di analisi del profilo di carico di un aggregato residenziale per una settimana.  
Riferimento: materiale didattico dell'insegnamento "Distribuzione e utilizzazione dell'energia elettrica"  
(prof. Gianfranco Chicco, anno accademico 2020-2021)*

L'intervallo di variazione e la precisione dei profili di carico dipendono molto dal numero di utenti analizzato. Se il numero di utenti è basso si avrà un'alta incertezza, mentre se il numero di utenti analizzati è alto l'incertezza si ridurrà sempre di più. Come possiamo vedere nella Figura 7, l'incertezza del profilo di carico per 20 utenti è molto maggiore rispetto all'incertezza dei profili di carico di 300 utenti.

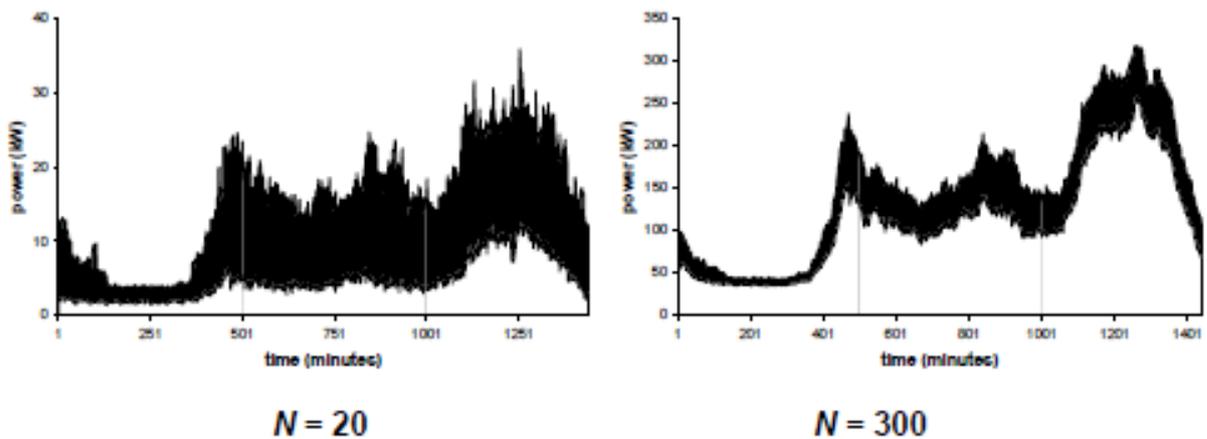


Figura 7: esempio di come varia l'incertezza nel caso se considerassi 20 utenti o 300. Riferimento: materiale didattico dell'insegnamento "Distribuzione e utilizzazione dell'energia elettrica" (prof. Gianfranco Chicco, anno accademico 2020-2021)

### Profili di carico per utenti industriali

Quando analizziamo i profili di carico per utenti industriali, possiamo notare alcune caratteristiche che li distinguono dal carico aggregato per utenti residenziali. I consumi sono prevalentemente concentrati durante le ore lavorative, quindi si hanno 2 picchi giornalieri che vanno dall'inizio della giornata lavorativa fino all'orario di pranzo e il secondo picco si ha nel pomeriggio. Durante la pausa pranzo si ha un crollo del consumo dovuto allo stop della produzione. Naturalmente durante la notte l'energia consumata si riduce notevolmente. Nella Figura 8 possiamo osservare il profilo di carico settimanale di un'industria con potenza di riferimento di 400 kW. Si notano molto evidentemente i due picchi giornalieri e il crollo dell'energia utilizzata durante la pausa pranzo, mentre la notte e nel fine settimana la potenza diminuisce poichè i macchinari dell'azienda sono spenti.

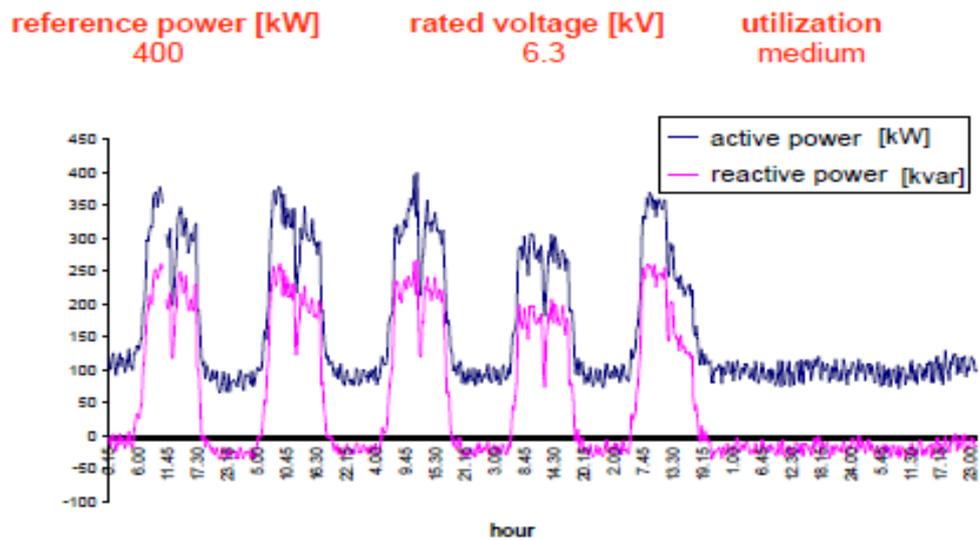


Figura 8: Esempio profilo di carico industriale osservato in una settimana. Riferimento: materiale didattico dell'insegnamento "Distribuzione e utilizzazione dell'energia elettrica" (prof. Gianfranco Chicco, anno accademico 2020-2021)

### Profili di carico per utenti terziari

Gli utenti terziari sono piccoli-medi utenti con il profilo di carico molto simile a quello industriale. I grandi carichi terziari, centri commerciali o grandi uffici, hanno un singolo picco durante il giorno dovuto al continuo lavoro ma la potenza non scende mai di tanto, neanche durante la notte. Questo perché vi sono alcuni servizi che sono sempre operativi, come ad esempio l'illuminamento e le celle frigorifere. Principalmente gli andamenti dei profili di carico del settore terziario variano molto tra di loro, proprio perché hanno differenti destinazioni d'uso.

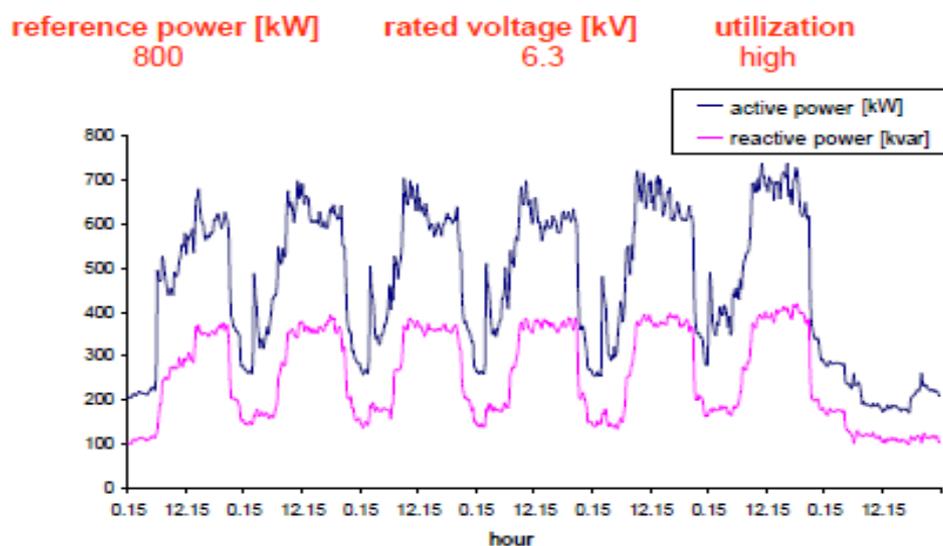


Figura 9: Esempio di andamento settimanale di un utente terziario con potenza di riferimento di 800 kW.  
Riferimento: materiale didattico dell'insegnamento "Distribuzione e utilizzazione dell'energia elettrica"  
(prof. Gianfranco Chicco, anno accademico 2020-2021)

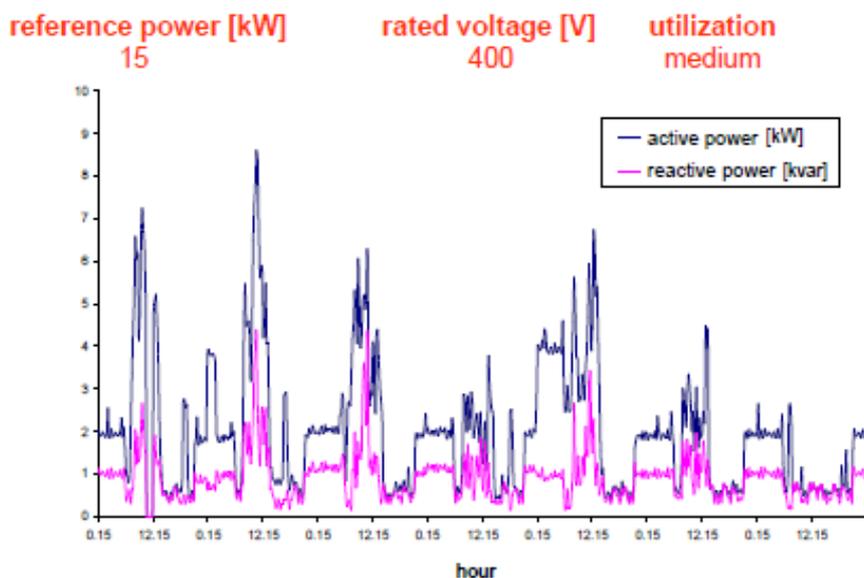


Figura 10: Esempio di andamento settimanale di un utente terziario con potenza di riferimento di 15 kW.  
Riferimento: materiale didattico dell'insegnamento "Distribuzione e utilizzazione dell'energia elettrica"  
(prof. Gianfranco Chicco, anno accademico 2020-2021)

In conclusione, possiamo dire che dopo l'introduzione del mercato competitivo dell'energia elettrica, i fornitori di energia analizzano i profili di carico e riescono a prevedere l'evoluzione dei consumi al livello della sottostazione AT/MT o MT/BT, queste informazioni consentono di identificare criticità e periodicità (settimanale, mensile o stagionale) delle oscillazioni dei consumi. Grazie a queste analisi, i fornitori di energia elettrica creano nuovi gradi di libertà per formulare nuove strutture tariffarie. Per questo motivo la conoscenza dell'evoluzione del carico elettrico è essenziale per la definizione delle tariffe variabili nel tempo. Inoltre, da analisi dettagliate effettuate su specifiche categorie di carico, si possono fare delle normalizzazioni. I profili di carico vengono estratti e poi sono normalizzati rispetto al picco della curva di carico, questo procedimento si fa per facilitarne l'utilizzo con diverse aggregazioni di carico.

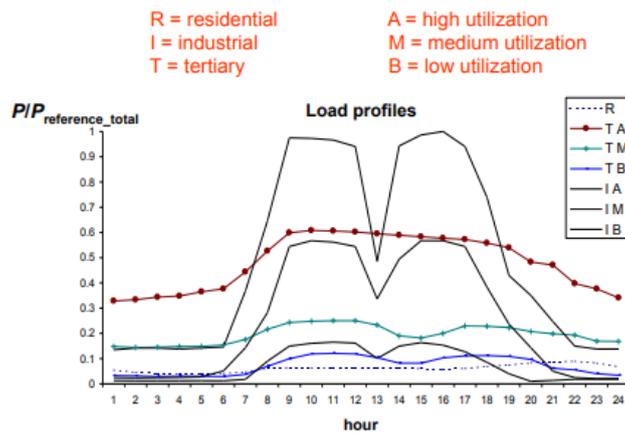


Figura 11: Esempio di normalizzazione di differenti profili di crico rispetto al loro picco di consumo. Riferimento: materiale didattico dell'insegnamento "Distribuzione e utilizzazione dell'energia elettrica" (prof. Gianfranco Chicco, anno accademico 2020-2021)

### 3. Approccio statistico per determinare i diversi livelli di clienti: campionamento stratificato

Il profilo di carico descrive la domanda di energia elettrica per un cliente o un gruppo di clienti, in un determinato periodo di tempo. Lo studio sui profili di carico dei clienti aggregati ha descritto per diversi anni il piano tariffario di molti clienti e con la liberazione del mercato elettrico ci si aspetta un incremento della pressione sulle utenze elettriche. In questa circostanza, una miglior comprensione delle caratteristiche del cliente elettrico permetterà di stabilire un'offerta tariffaria che possa aiutare le utenze elettriche nel risparmio dei consumi. In sostanza la buona e accurata previsione del profilo di carico è utile sia per affrontare una corretta progettazione della strategia di marketing sia per migliorare la pianificazione dell'espansione del sistema per ottenere un buon servizio che permetta il miglioramento dell'efficienza operativa del sistema. Grazie alle autorità di regolamentazione sono state coordinate le procedure che hanno permesso diverse soluzioni tenendo sostanzialmente conto:

- della tipologia di utenza (residenziale, industriale, commerciale)
- del livello di tensione di alimentazione
- del fattore di carico (*load factor*).

Una procedura generale per determinare la classe o tipologia di utenza del carico è mostrata nella Figura 12.

Il primo problema è la definizione della classe dei clienti. Normalmente la classe è definita dalla potenza contrattuale e dal tipo di attività, mentre dal punto di vista del mercato elettrico sarebbe più interessante raggruppare i clienti sulla base di diversi profili di carico con le seguenti caratteristiche:

- Ogni profilo di carico deve rappresentare un gruppo relativamente omogeneo di clienti
- Ogni profilo di carico deve essere nettamente diverso dagli altri
- I parametri per l'assegnazione di qualsiasi cliente a un profilo di carico dovrebbero essere facilmente leggibili

- Il numero dei diversi profili di carico deve essere basso

Molto importante è avere un ampio sondaggio tra i clienti per trovare possibili profili di carico diversi. Nella situazione reale le risorse di misurazione possono essere molto limitate. Questo documento considera una possibile strategia per allocare le risorse di misurazione tra tali classi, utilizzando le loro caratteristiche statistiche e la teoria del campionamento stratificato.



*Figura 12: Processo per rilevare i profili di carico*

## 3.1. Teoria del campionamento stratificato

Come accennato in precedenza, i clienti elettrici si dividono in differenti classi o livelli dipendenti dalla forma contrattuale che hanno. Quindi, l'approccio più rigoroso per determinare il profilo di carico di ciascuna classe di utenti sarebbe quello di effettuare misurazioni su ciascun membro della classe e quindi elaborare tutti i dati, ma visto che il numero di utenti è molto elevato tale approccio risulta essere poco pratico e troppo costoso. Siamo quindi costretti a utilizzare alcune tecniche statistiche significative per determinare il miglior campionamento di ogni classe di clienti. Nel caso in esame si è scelto di utilizzare la teoria del campionamento stratificato dove le principali caratteristiche sono riportate di seguito.

### Scelta della variabile discriminante

Una variabile  $Y$  è scelta come principale caratteristica discriminante, ovvero, la potenza attiva contrattuale o il consumo medio di energia in un determinato intervallo di tempo. Questa variabile caratteristica deve essere facile da valutare con i dati fatturati della società di servizi elettrici. Questa caratteristica può essere valutata cumulativamente su tutti i clienti  $y_{tot}$  e su tutti i livelli  $y_h$ . La teoria del campionamento stratificato campiona le classi per ottimizzare la stima di  $y_{tot}$  utilizzando le stime di  $y_h$  delle differenti classi.

### Stima della media e della varianza dello strato

Supponiamo che la popolazione totale di  $N$  clienti sia partizionata in diversi livelli  $H$  e per il livello  $H$ -esimo generico siano definite le seguenti notazioni:

- $N_h$  numero totale di livelli
- $n_h$  numero totale di livelli campionati
- $y_{hi}$  valore della caratteristica variabile all'esempio  $i$ -esimo
- $\mu_h$  rappresenta la media dei veri livelli
- $\sigma^2_h$  rappresenta i livelli di varianza reali (variazione standard)
- $\mu_h^{\wedge}$  media dello strato del campione
- $\sigma^{\wedge 2}_h$  varianza del livello del campione (variazione standard).

$$\mu = \frac{1}{N_h} \sum_{i=1}^N (y)$$

$$\sigma^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y - \mu)^2$$

$$\hat{\mu}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} (y_{hi})$$

$$\hat{\sigma}^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y - \mu)^2$$

### Stima della media e della varianza della popolazione

Considerando  $N$  uguale alla sommatoria di tutti i livelli da uno a  $h$  come il totale numero di unità e  $n$  il numero totale di campioni. La vera popolazione media  $\mu$  e la sua stima  $\mu^{\wedge}$  sono date dalle seguenti formulazioni:

$$\mu = \frac{1}{N} \sum_{h=1}^H (N_h * \mu_h) \quad \hat{\mu} = \frac{1}{N} \sum_{h=1}^H (N_h * \hat{\mu}_h)$$

La varianza associata a  $\mu^{\wedge}$  è:

$$\hat{\sigma}^2 = \frac{1}{N^2} \cdot \sum_{h=1}^H N_h^2 \frac{(N_h - n_h)}{(N_h - 1)} \cdot \frac{(\sigma_h)^2}{n_h} \quad (1)$$

Con l'ipotesi di una distribuzione normale, l'intervallo di confidenza della popolazione media  $\mu$  è stimato da  $\mu^{\wedge} = k * \sigma^{\wedge}$ .

Possiamo vedere i valori dell'intervallo di confidenza al differente valore di  $k$  nella Tabella 2.

Tabella 2: valori di  $k$  e rispettive probabilità di confidenza

	K1	K2	K3	K4	K5
Probabilità di confidenza	50%	68.2%	95.4%	99%	99.7%
k	0.67	1	2	2.58	3

## Dimensione del campione

Bisogna selezionare la miglior dimensione del campione  $n_h$  per  $h=1, 2...H$  dato  $C$  come costo totale

$$C = c + \sum c_h \cdot n_h$$

dove  $C_0$  sono i costi fissi sostenuti per l'esecuzione dell'indagine e  $c_h$  i costi associati ad una singola misurazione nell'indagine per il livello  $h$ . Questo problema di ottimizzazione viene risolto utilizzando il *Lagrangian multiplier method* dove i valori  $n_h$  per  $h=1, 2...H$  e il *Lagrangian multiplier*  $\lambda$  sono selezionati per ridurre al minimo la funzione obiettivo senza vincoli.

$$L = \hat{\sigma}^2 + \lambda \left( \sum_{h=1}^H (c_h \cdot n_h) - C + c \right)$$

Ponendo la derivata uguale a zero  $\delta L / \delta \lambda = 0$  con il ritorno al vincolo di uguaglianza (*equality constraint*) e per  $\delta L / \delta n_h = 0$  per  $h=1, 2... H$  e prendendo l'unica soluzione valida e sommandola per tutti i livelli e dividendola con la soluzione del singolo livello otteniamo che:

$$n_h = n \cdot \frac{N_h \cdot \sigma \sqrt{N_h \cdot \frac{1}{c_h(N_h - 1)}}}{\sum_{m=1}^H (N_m \sigma \cdot \sqrt{\frac{N_m}{c_m(N_m - 1)}})}$$

Approssimando che il costo di indagine dei singoli costi è uguale per tutti i livelli abbiamo che  $C=C_h$  per cui  $n_h$  diventa:

$$n_h = n \cdot \frac{N_h \cdot \sigma \sqrt{N_h \cdot \frac{1}{(N_h - 1)}}}{\sum_{m=1}^H (N_m \sigma \cdot \sqrt{\frac{N_m}{(N_m - 1)}})} \quad (2)$$

Per un dato livello si dovrebbe prendere un numero di campioni  $n_h$  più grande se:

- Se il livello è più grande
- Se il livello è variabile internamente

### Stima dell'intervallo di confidenza

Il processo di allocazione ottimale è la partizione di  $n$  nei vari  $n_h$  dove

$$n = \frac{C - c_1}{c'}$$

Una relazione esplicita  $n$  e la deviazione standard  $\sigma^{\wedge}$  è ottenuta sostituendo  $n_h$  data dall'equazione (2) nella equazione (1), trovando:

$$\hat{\sigma}^2 = \left[ \left( \sum_{h=1}^H N_h * \sigma \cdot \sqrt{\frac{N_{\pi}}{N_h - 1}} \right) - n \sum_{h=1}^H \frac{N_{\pi}^2 \sigma^2}{N_h - 1} \right] \frac{1}{N^2 n}$$

Per dare una rappresentazione significativa dei risultati possiamo affermare che l'intervallo di confidenza sulla stima sulla vera popolazione media dai campioni è delimitato da  $\mu^{\wedge} \pm k * \sigma^{\wedge} = \mu^{\wedge} \pm d$  è possibile quindi relazionare  $d = k \times \sigma^{\wedge}$  al numero complessivo delle unità in tutti i livelli dei campioni per qualsiasi valore del moltiplicatore  $k$  (Tabella 2).

### Riassunto del metodo

Il metodo del campionamento stratificato può essere riassunto in pochi semplici passaggi:

1. Definire la caratteristica variabile  $Y$  (la potenza contrattuale e se è possibile l'energia consumata in un dato periodo di tempo)
2. Definire il budget disponibile  $C$  e il numero  $n$  di misurazioni massime possibili
3. Considerare i livelli proposti e per ogni livello calcolare  $n_h$  (numero totale di livelli campionati) e la vera varianza del livello  $\sigma_h$
4. Calcolare il numero suggerito di campioni  $n_h$  per ogni strato usando l'eq. (2)
5. Verificare che l'intervallo di confidenza sia accettabile, ovvero se il budget sia ragionevole.

Una volta stabilito il sistema di indagine, i dati saranno raccolti eseguendo le misurazioni sul campo e verrà determinata la curva cumulativa dei profili di carico da seguire.

### **Vantaggi**

Questo tipo di campionamento si rivela particolarmente utile quando gli strati, all'interno della popolazione, sono di numerosità molto diversa. In una tale situazione di elevata variabilità del fenomeno, infatti, sarebbe necessario un campione molto ampio. Stratificando la popolazione, invece, è possibile ottenere una adeguata copertura degli strati meno numerosi anche con un campione di dimensioni ridotte, con un apprezzabile risparmio di tempi e costi di rilevazione.

Non è indispensabile, infatti, che il numero dei soggetti che compongono il campione all'interno di ciascuno strato sia proporzionale alla dimensione dello strato nella popolazione, né che gli strati siano tutti della stessa numerosità.

Tra i pregi legati all'utilizzo di un campionamento stratificato vi sono la migliore rappresentatività e uno sfruttamento delle informazioni disponibili sulla popolazione che altre tecniche di campionamento non hanno. La conseguenza pratica di tutto ciò è la capacità di generare stime più efficienti.

### **Limiti**

La principale limitazione del campionamento stratificato è che lo strato di tutte le unità di campionamento, rispetto ai fattori su cui è basata la stratificazione, deve essere noto prima di scegliere il campione in quanto la procedura di estrazione del campione è comunque probabilistica.

## 4. Applicazione del metodo

Il metodo di campionamento stratificato è stato applicato per impostare un'indagine sui profili di carico per i clienti della città di Torino. I dati sui consumi sono stati forniti dall'azienda Ireti S.P.A, facente parte del gruppo IREN. Essi sono stati inviati dal responsabile dell'*area metering* di Ireti.

I dati necessari per la realizzazione del metodo sono:

- numero di utenti totali residenziali (tipo monofase)
- numero di utenti totali non residenziali (tipo trifase)
- per ogni utente avere un'informazione di potenza contrattuale e di energia annuale
- avere alcune informazioni sulla energia utilizzata mensilmente

### 4.1. Caso studio: Dati trifase, utenti non residenziali

I dati relativi ai clienti trifase a uso non residenziale sono stati suddivisi per fornitura, potenza contrattuale, tipologia d'uso ed energia consumata. Questi dati sono stati forniti mensilmente e comprendevano tutti i mesi del 2020.

#### Primo passo: suddivisione delle classi

Una volta acquisiti i dati è stata fatta un'analisi per capire al meglio come dividere i dati nelle sottoclassi. I criteri scelti per la suddivisione sono i seguenti:

1. Sono stati individuati e filtrati i codici della tipologia d'uso più numerosi
  - OR (tipologia d'uso ordinario)
  - SA (tipologia d'uso ordinario ma con funzioni specifiche)
2. Sono state create dieci classi per ogni tipologia d'uso in base ai limiti per la potenza contrattuale

Nelle tabelle sottostanti, Tabella 1.1 e Tabella 1.2 possiamo vedere i dati degli utenti e la suddivisione delle diverse classi.

*Tabella 1.1: suddivisione delle classi per la tipologia d'uso ordinaria (OR)*

CLASSE	LIMITE PER POTENZA CONTRATTUALE (kW)	NUMERO	SOMMATORIA POT CONTRATTUALE (kW)	SOMMATORIA ENERGIA TOT (kWh)
ORT1	0-4.6	7901	22327	1323463
ORT2	4.6-6.6	6829	40559	2125193
ORT3	6.6-17	18142	220035	12834653
ORT4	17-27	5689	123912	8207760
ORT5	27-47	4117	141106	10151000
ORT6	47-67	1784	98120	7416353
ORT7	67-107	1240	104243	9037427
ORT8	107-147	290	36137	3647622
ORT9	147-187	212	33566	3647406
ORT10	187-4413	190	48714	4091705
TOTALE		46396	868722	62482582

*Tabella 1.2: suddivisione delle classi per la tipologia d'uso ordinaria ascensore (SA)*

CLASSE	LIMITE PER POTENZA CONTRATTUALE (kW)	NUMERO	SOMMATORIA POT CONTRATTUALE (kW)	SOMMATORIA ENERGIA TOT (kWh)
SAT1	0-4.6	1285	4675	312204
SAT2	4.6-6.6	6612	37765	1592260
SAT3	6.6-17	15433	178377	6249377
SAT4	17-27	2174	46358	1901503
SAT5	27-47	791	26942	1660509
SAT6	47-67	383	21072	1450105
SAT7	67-107	287	24463	1844998
SAT8	107-147	73	9007	720002

SAT9	147-187	49	7828	873597
SAT10	187-4413	35	8497	660069
TOTALE		27122	365105	17266973

Si riescono a suddividere 20 sottoclassi, ad esse verrà applicato il campionamento stratificato.

L'applicazione del campionamento stratificato è stata fatta utilizzando il programma Excel.

L'analisi verrà fatta suddividendo due casi:

- 1 Campionamento stratificato avendo come variabile discriminante la potenza contrattuale degli utenti non residenziali trifase
- 2 Campionamento stratificato avendo come variabile discriminante l'energia consumata degli utenti non residenziali trifase

#### 4.1.1. Campionamento stratificato avendo come variabile discriminante la potenza contrattuale degli utenti non residenziali trifase

##### Input ed elaborazione dei dati

Per procedere con il campionamento stratificato sono necessari alcune informazioni iniziali, come spiegato nel paragrafo precedente. Questi dati sono stati calcolati scegliendo come variabile discriminante la potenza contrattuale, per cui l'analisi del campionamento stratificato è stata fatta su un mese casuale (marzo), questo perché durante l'anno la potenza contrattuale degli utenti non cambia. Quando si considererà come variabile discriminante l'energia consumata bisognerà ragionare in modo differente.

I dati di input per ogni classe sono: il numero di utenti  $N_h$ , la media e la deviazione standard  $\sigma$ .

Tabella 1.3: esempio di campionamento stratificato degli utenti trifase del mese di marzo 2020

variabile: potenza contrattuale																				
INPUT																				
classe	ORT1	ORT2	ORT3	ORT4	ORT5	ORT6	ORT7	ORT8	ORT9	ORT10	SAT1	SAT2	SAT3	SAT4	SAT5	SAT6	SAT7	SAT8	SAT9	SAT10
numero di ute	7901	6829	18142	5689	4117	1784	1240	290	212	190	1285	6612	15433	2174	791	383	287	73	49	35
valore medio	2.82	5.939	12.128	21.78	34.27	55	84.06	124.61	158.33	256.39	3.63	5.711	11.55	21.32	34.06	55.01	85.23	123.29	159.77	242.77
deviazione st	0.945	0.24	2.53	2.39	5.23	5.23	11.69	9.96	11.17	88.06	0.6945	0.4529	2.43	2.22	5.53	5.14	11.91	10.03	11.4	85.68

Avendo come input questi dati si può iniziare la loro elaborazione. L'elaborazione dei dati consiste nel calcolare tutti i dati utili delle diverse classi necessari a trovare il numero di utenti da ripartire per la costruzione del grafico della variazione dell'intervallo di confidenza necessario per decidere il numero di utenti finale da analizzare.

Tabella 1.4: esempio di campionamento stratificato degli utenti trifase del mese di marzo 2020 tenendo conto delle prime tre classi del tipo d'uso OR.

ELABORAZIONE DATI			
numero totale utenti N	73516		
valore medio totale	16.77805456		
	ORT1	ORT2	ORT3
varianza	0.893025	0.0576	6.4009
valore totale	22280.82	40557.43	220026.2
dev stand % rispetto alla media	33.5106383	4.041084	20.86082
$N_h \cdot \sigma \cdot \text{radq}(N_h / (N_h - 1))$	7466.917545	1639.08	45900.53
$(N_h \cdot \sigma)^2 / (N_h - 1)$	7056.683663	393.408	116131.5

## Output

È stata creata la Tabella 1.4 nella quale, nella prima colonna, sono stati inseriti il numero di utenti  $n$ , decisi a priori, i quali rappresentano il numero totale in cui le diverse classi si devono ripartire; infatti, la sommatoria dei valori delle singole righe deve essere uguale al valore della prima colonna per ogni riga. Il primo numero scelto (5) è molto basso, ma questa analisi è servita per poi ricostruire meglio il grafico dell'intervallo di confidenza.

Tabella 1.5: Diverse ripartizioni del numero di utenti nelle diverse classi

numeri utenti	ORT1	ORT2	ORT3	ORT4	ORT5	ORT6	ORT7	ORT8	ORT9	ORT10	SAT1	SAT2	SAT3	SAT4	SAT5	SAT6	SAT7	SAT8	SAT9	SAT10
5	0	0	1	1	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0
10	1	0	2	1	1	1	1	0	0	1	0	0	2	0	0	0	0	0	0	0
20	1	0	5	1	2	1	2	0	0	2	0	0	4	1	1	0	0	0	0	0
30	1	0	7	2	3	2	2	0	0	3	0	0	6	1	1	0	1	0	0	1
50	2	0	12	3	5	3	4	1	1	4	0	1	10	1	1	1	1	0	0	1
70	3	1	17	5	8	3	5	1	1	6	0	1	12	2	2	1	1	0	0	1
100	4	1	23	7	11	5	7	1	1	9	0	2	19	2	2	1	2	0	0	2
130	5	1	30	9	14	6	10	2	2	11	1	2	25	3	3	1	2	0	0	2
160	6	1	37	11	18	8	12	2	2	14	1	2	31	4	4	2	3	1	0	2
200	8	2	47	14	22	10	15	3	2	17	1	3	38	5	4	2	3	1	1	3
240	9	2	56	17	26	11	18	4	3	21	1	4	46	6	5	2	4	1	1	4
280	11	2	65	19	31	13	21	4	3	24	1	4	53	7	6	3	5	1	1	4
320	12	3	75	22	35	15	24	5	4	27	1	5	61	8	7	3	6	1	1	5
370	14	3	86	26	41	18	27	5	4	32	2	6	71	9	8	4	6	1	1	6
400	15	3	94	28	44	19	30	6	5	34	2	6	76	10	9	4	7	2	1	6
450	17	4	105	31	49	21	33	7	5	38	2	7	86	11	10	5	8	2	1	7
500	19	4	117	35	55	24	37	7	5	43	2	8	96	12	11	5	9	2	1	8
550	21	5	129	38	60	26	41	8	7	47	3	8	105	14	12	6	10	2	2	9
600	23	5	140	42	66	29	44	9	7	51	3	9	115	15	13	6	10	2	2	9
650	25	5	152	45	71	31	48	10	8	56	3	10	124	16	14	7	11	2	2	10
700	27	6	164	48	77	33	52	10	8	60	3	11	134	17	16	7	12	3	2	11
800	30	7	187	55	88	38	59	12	10	68	4	12	153	20	18	8	14	3	2	12
900	34	8	210	62	99	43	66	13	11	77	4	14	172	22	20	9	16	3	3	14
1000	38	8	234	69	110	48	74	15	12	85	5	15	191	25	22	10	17	4	3	15

## Stima dell'intervallo di confidenza

Una volta calcolato  $n_h$  di ogni classe per ogni numero di utenti totali decisi a priori si è fatta una stima dell'intervallo di confidenza. Si è calcolato  $d\%$ . Esso è un valore percentuale che si trova moltiplicando la stima della vera deviazione standard per la probabilità di confidenza ( $k$ ), diviso il valor medio totale. I valori  $k$  scelti che rappresentano la probabilità di confidenza sono stati scelti come mostrato in Tabella 1.6.

Tabella 1.6: i diversi valori di probabilita di confidenza di K1,K2,K3,K4,K5

	K1	K2	K3	K4	K5
Probabilità di confidenza	50%	68.2%	95.4%	99%	99.7%
k	0.67	1	2	2.58	3

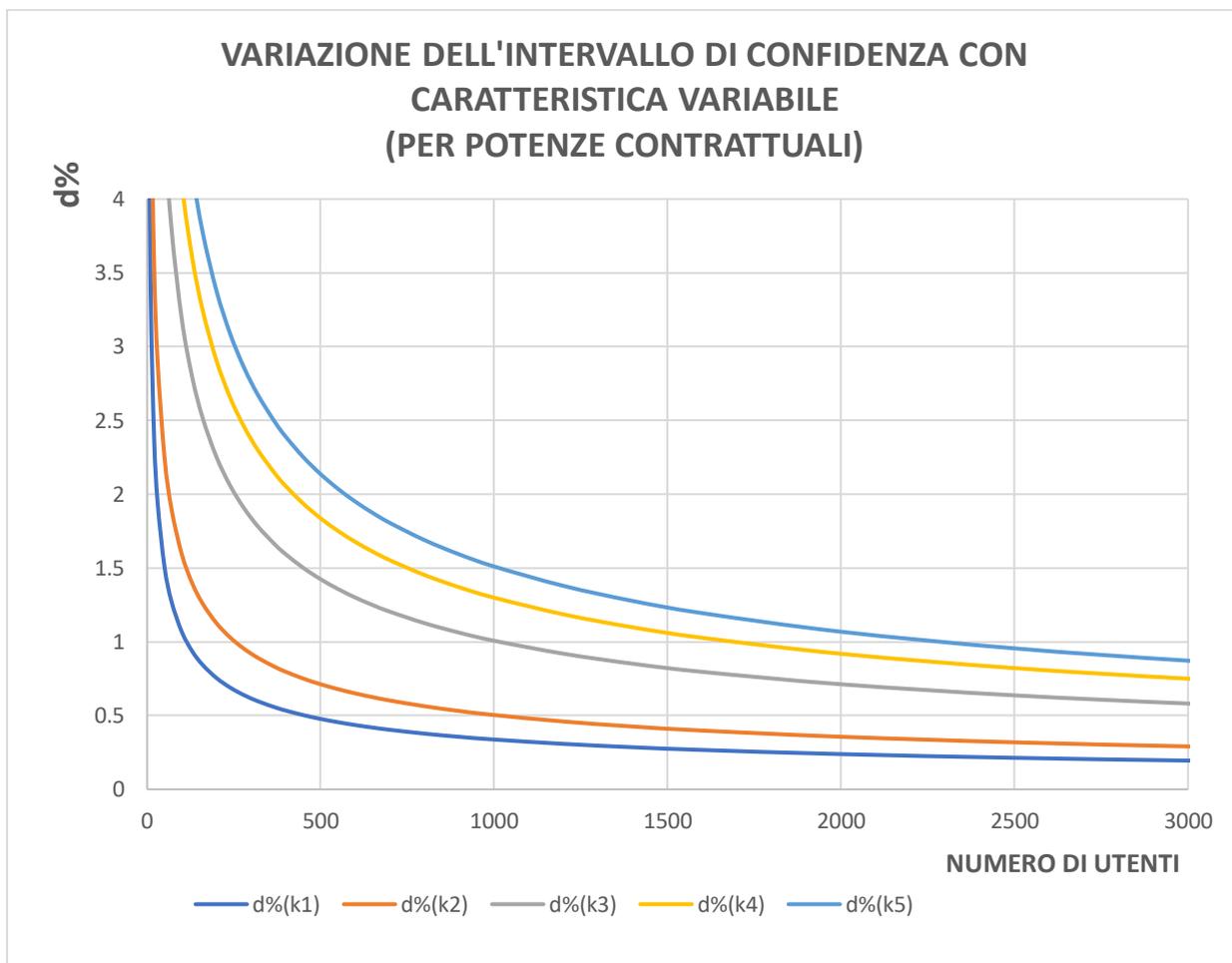


Grafico 1: variazione dell'intervallo di confidenza con caratteristica variabile per potenze contrattuali, esempio degli utenti trifase del mese di marzo 2020

#### 4.1.2. Campionamento stratificato avendo come variabile discriminante l'energia consumata degli utenti non residenziali trifase

In questo caso, l'analisi del campionamento stratificato è stata fatta scegliendo come variabile discriminante l'energia consumata mensilmente. Visto che mensilmente il consumo energetico cambia quest'analisi è stata fatta su tutti i mesi dell'anno e non su un'unica mensilità come nel caso precedente. Per coerenza i dati e le figure riportate di seguito saranno dello stesso mese del caso precedente (marzo 2020).

##### Input ed elaborazione dei dati

Il procedimento rimane immutato, ma cambiano i dati in input, poiché riferiti all'energia consumata. Come possiamo osservare nella Tabella 1.7, la deviazione standard è molto più alta rispetto al caso precedente poiché l'energia consumata dipende solamente dall'utente e non è un valore fisso come la potenza contrattuale.

*Tabella 1.7: input di campionamento stratificato degli utenti trifase del mese di marzo 2020 tenendo conto delle prime tre classi del tipo d'uso OR e come variabile l'energia consumata mensilmente*

variabile:	energia consumata		
INPUT			
classe	ORT1	ORT2	ORT3
numero di utenti Nh	7903.00	6829.00	18142.00
valore medio	167.00	311.00	707.00
deviazione standard	252.00	334.00	742.00

Adesso si può procedere con l'elaborazione dati.

Tabella 1.8: elaborazione dati di campionamento stratificato degli utenti trifase del mese di marzo 2020 tenendo conto delle prime tre classi del tipo d'uso OR e come variabile l'energia consumata mensilmente

ELABORAZIONE DATI			
numero totale utenti N	73519.00		
valore medio totale	1091.962105		
Classi	ORT1	ORT2	ORT3
varianza	63504	111556	550564
valore totale	1319801	2123819	12826394
dev stand % rispetto alla media	150.8982036	107.3954984	104.950495
$Nh \cdot \sigma \cdot \sqrt{Nh / (Nh - 1)}$	1991682.012	2281053.018	13461735.02
$(Nh \cdot \sigma)^2 / (Nh - 1)$	501935624	761927496.3	9988882682

## Output

Grazie ai dati elaborati possiamo ricreare la tabella dove verranno descritti i valori  $n_h$  suddivisi per ogni classe e per il numero di utenti totali da rappresentare.

Tabella 1.9: Diverse ripartizione del numero di utenti nelle diverse classi

numeri utenti	ORT1	ORT2	ORT3	ORT4	ORT5	ORT6	ORT7	ORT8	ORT9	ORT10	SAT1	SAT2	SAT3	SAT4	SAT5	SAT6	SAT7	SAT8	SAT9	SAT10
5	0	0	2	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
10	1	0	2	2	2	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0
20	1	1	3	2	3	2	2	1	1	1	0	0	2	0	0	0	1	0	0	0
30	1	1	5	3	4	3	3	1	1	1	0	0	3	1	1	0	1	0	0	0
50	1	1	8	5	7	5	6	2	2	2	0	1	4	1	1	1	1	0	1	0
70	2	2	12	7	9	6	8	3	3	3	0	1	6	2	1	1	2	1	1	1
100	2	3	17	9	13	9	11	4	4	5	0	2	9	2	2	2	2	1	1	1
130	3	4	22	12	17	12	15	5	6	6	1	2	11	3	3	2	3	1	1	1
160	4	5	27	15	21	15	18	7	7	7	1	3	14	4	3	3	4	1	2	1
200	5	6	33	19	26	18	23	8	9	9	1	3	18	5	4	3	4	2	2	2
240	6	7	40	23	31	22	28	10	11	11	1	4	21	6	5	4	5	2	2	2
280	7	8	47	26	36	26	32	12	13	13	1	5	25	6	6	4	6	2	3	3
320	8	9	54	30	42	29	37	13	14	15	1	5	28	7	7	5	7	2	3	3
370	9	10	62	35	48	34	43	15	17	17	2	6	32	9	8	6	8	3	4	3
400	10	11	67	38	52	37	46	17	18	18	2	6	35	9	8	6	9	3	4	4
450	11	13	75	42	59	41	52	19	20	21	2	7	39	10	9	7	10	3	5	4
500	12	14	84	47	65	46	57	21	22	23	2	8	44	12	10	8	11	4	5	4
550	14	16	92	52	72	50	63	23	25	25	2	9	48	13	11	9	12	4	6	5
600	15	17	100	56	78	55	69	25	27	28	3	10	53	14	12	9	13	5	6	5
650	16	18	109	61	85	59	75	27	29	30	3	11	57	15	13	10	14	5	7	6
700	17	20	117	66	91	64	80	29	31	32	3	11	61	16	14	11	16	5	7	6
800	20	23	134	75	104	73	92	33	36	37	3	13	70	18	17	13	18	6	8	7
900	22	26	151	84	117	82	103	37	40	42	4	15	79	21	19	14	20	7	9	8
1000	25	28	167	94	130	92	115	42	45	46	4	16	88	23	21	16	22	8	10	9
1200	30	34	201	113	156	110	138	50	54	55	5	19	105	28	25	19	27	9	12	11
1400	35	40	234	131	182	128	161	58	63	65	6	23	123	32	29	22	31	11	15	13
1600	40	45	268	150	208	146	184	66	71	74	7	26	140	37	33	25	36	12	17	14

## Stima dell'intervallo di confidenza

Avendo tutti i dati possiamo calcolarci la stima dell'intervallo di confidenza, ovvero  $d\%$  calcolato con gli stessi  $k$ , probabilità di confidenza, del caso precedente.

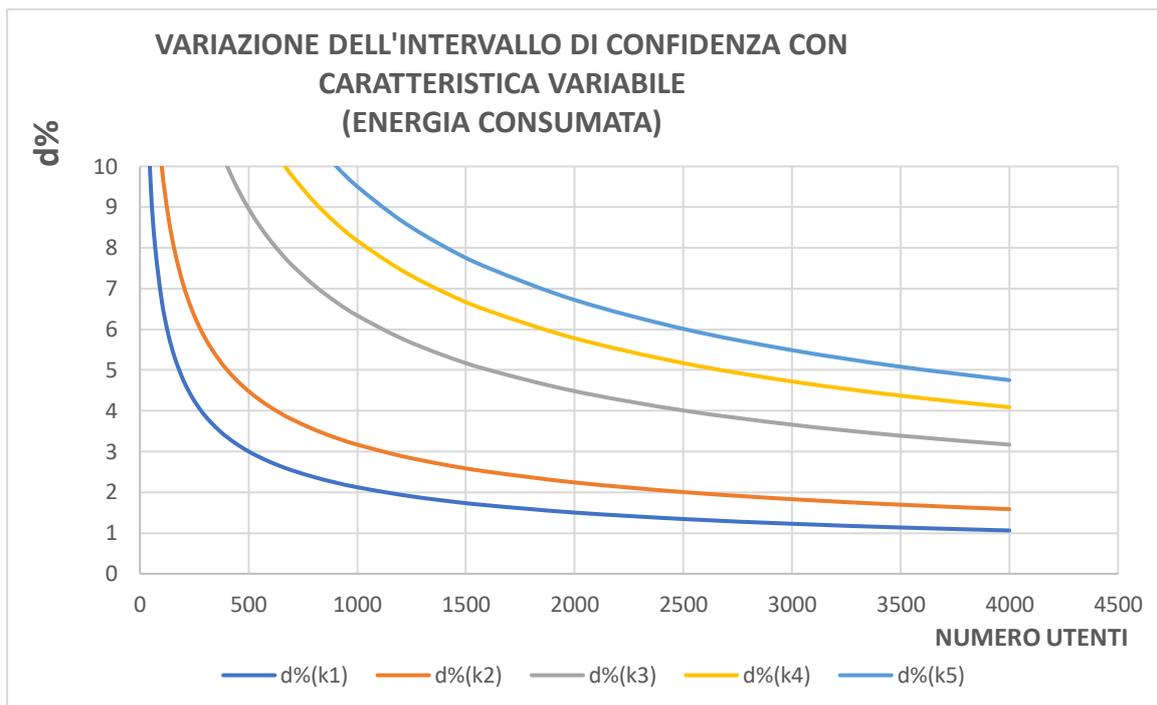


Grafico 2: variazione dell'intervallo di confidenza con caratteristica variabile per energia consumata, esempio degli utenti trifase del mese di marzo 2020

### Confronto tra i due casi

Come si può osservare graficamente il primo caso (grafico 1), ovvero quello avente come variabile discriminativa la potenza contrattuale, si comporta meglio rispetto al caso dell'energia consumata (grafico 2), difatti a 2500 utenti nel primo caso siamo sotto l' 1% della variazione dell'intervallo di confidenza anche con la probabilità di confidenza del 99.7% (caso con  $k=3$ ) mentre nel caso dell'energia consumata a 2500 utenti il valore della variazione dell'intervallo di confidenza è del 6% con la probabilità di confidenza del 99.7% ( $k=3$ ). Questo risultato è dato dalla diversa tipologia di dati. Le potenze contrattuali, essendo valori fissi e non modificabili, hanno una deviazione standard e dunque una varianza molto minore rispetto ai dati relativi all'energia consumata, la quale è molto variabile anche all'interno delle diverse classi, poiché dipende esclusivamente dal consumo dell'utente. Per questo motivo, la variazione dell'intervallo di confidenza è maggiore nel caso dell'energia consumata.

### 4.1.3. Ripartizione finale degli utenti

Analizzando i dati in uscita della stima dell'intervallo di confidenza, si è scelto di puntare a prendere 2500 utenti divisi come nel caso dell'energia consumata, questa scelta è stata presa in considerazione perché, rispetto al caso della potenza contrattuale, il caso dell'energia consumata ha i dati distribuiti più uniformemente sulle diverse classi per via della elevata deviazione standard presente in ciascuna classe. Come visto nel confronto tra i due casi, il numero di utenti 2500 è stato scelto poiché sia nel caso della potenza contrattuale sia nel caso dell'energia consumata si hanno buoni valori di intervallo di confidenza. Il numero puntato dunque è 2500 utenti, il numero per ogni classe deve essere distribuito come in Tabella 1.10.

Tabella 1.10: ripartizione ideale delle classi

utenti	ORT1	ORT2	ORT3	ORT4	ORT5	ORT6	ORT7	ORT8	ORT9	ORT10	SAT1	SAT2	SAT3	SAT4	SAT5	SAT6	SAT7	SAT8	SAT9	SAT10
2500	62	71	418	234	325	229	287	104	112	116	11	40	219	58	52	39	56	19	26	22

Per un miglior studio dei dati trifase è stato osservato quante forniture rimanevano allacciate tutti e 12 i mesi, poiché se si vuole fare l'analisi annuale sarebbe scorretto prendere degli utenti che non sono attivi tutti e 12 i mesi, ovvero che per alcuni mesi dell'anno hanno la fornitura staccata. Dunque, tramite un filtro si è notato che solo 6165 utenti si ripetevano nel corso dei 12 mesi. Tramite Matlab abbiamo trovato e suddiviso nelle diverse classi questi 6165 utenti.

Essi sono così distribuiti:

Tabella 1.11: ripartizione delle classi sul numero di utenti reali allacciati nei 12 mesi

utenti tot	ORT1	ORT2	ORT3	ORT4	ORT5	ORT6	ORT7	ORT8	ORT9	ORT10	SAT1	SAT2	SAT3	SAT4	SAT5	SAT6	SAT7	SAT8	SAT9	SAT10
6165	315	446	1759	670	495	241	102	23	12	14	6	73	1623	304	73	4	3	0	2	0

Come possiamo notare vi sono certe classi nella Tabella 1.11 che non hanno un numero sufficiente di utenti per coprire la suddivisione presa come obiettivo della Tabella 1.10. Bisogna, dunque, fare un ulteriore passaggio per aver il numero giusto di utenti da richiedere per poi analizzare i profili di carico.

Si confrontano le 3 tabelle successive e si prende il numero massimo di utenti per ogni classe.

Distribuzione dei 2500 utenti tenendo conto dell'energia consumata.

*Tabella1.12*

utenti	ORT1	ORT2	ORT3	ORT4	ORT5	ORT6	ORT7	ORT8	ORT9	ORT10	SAT1	SAT2	SAT3	SAT4	SAT5	SAT6	SAT7	SAT8	SAT9	SAT10
2500	62	71	418	234	325	229	287	104	112	116	11	40	219	58	52	39	56	19	26	22

Distribuzione dei 2500 utenti tenendo conto della potenza contrattuale.

*Tabella1.13*

UTENTI TOT	ORT1	ORT2	ORT3	ORT4	ORT5	ORT6	ORT7	ORT8	ORT9	ORT10	SAT1	SAT2	SAT3	SAT4	SAT5	SAT6	SAT7	SAT8	SAT9	SAT10
2500	41	66	399	255	316	231	281	114	114	143	10	49	194	59	52	45	57	22	27	23

Distribuzione dei 6156 utenti reali nei 12 mesi.

*Tabella1.14*

utenti tot	ORT1	ORT2	ORT3	ORT4	ORT5	ORT6	ORT7	ORT8	ORT9	ORT10	SAT1	SAT2	SAT3	SAT4	SAT5	SAT6	SAT7	SAT8	SAT9	SAT10
6165	315	446	1759	670	495	241	102	23	12	14	6	73	1623	304	73	4	3	0	2	0

Il numero finale di utenti da analizzare e la rispettiva suddivisione delle differenti classi diventa:

*Tabella1.15*

utenti tot	ORT1	ORT2	ORT3	ORT4	ORT5	ORT6	ORT7	ORT8	ORT9	ORT10	SAT1	SAT2	SAT3	SAT4	SAT5	SAT6	SAT7	SAT8	SAT9	SAT10
6842	315	446	1759	670	495	241	287	114	114	143	11	73	1623	304	73	45	57	22	27	23

Questo è il nostro punto di partenza per analizzare i profili di carico degli utenti trifase a uso non residenziale di ogni classe.

## 4.2. Caso studio: dati monofase, utenti residenziali

I dati inviati dal responsabile dell'area *metering* di Ireti S.p.A., per quanto riguarda i clienti monofase a uso residenziale, erano suddivisi per: fornitura, potenza contrattuale, tipologia d'uso ed energia consumata divisa nelle differenti fasce (F1, F2, F3). Questi dati sono annuali e descrivono il consumo degli utenti residenziali nell'anno 2020.

### Primo passo: suddivisione delle classi

Una volta acquisiti i dati, è stata fatta un'analisi per capire al meglio come dividere i dati nelle rispettive sottoclassi. I criteri scelti per la suddivisione sono gli stessi usati nel caso precedente.

Sono stati individuati e filtrati i codici della tipologia d'uso più numerosi:

- OR (tipologia d'uso ordinario)
- SA (tipologia d'uso ordinario ma con funzioni specifiche)

Sono state create cinque classi per ogni tipologia d'uso in base ai limiti della potenza contrattuale.

Nella Tabella 2.1 possiamo vedere i dati degli utenti e la suddivisione delle diverse classi.

Tabella 2.1: suddivisione delle classi per la tipologia d'uso ordinaria ascensore (OR)

LIMITE PER POTENZA CONTRATTUALE (kW)	CLASSE	NUMERO	SOMMATORIA POTENZA CONTRATTUALE (kW)	SOMMATORIA ENERGIA ANNUA (kWh)
0-2.3	ORM1	27846	36956	20367070
2.3-3.6	ORM2	544839	1635367	825253034
3.6-5.3	ORM3	34551	155437	89751770
5.3-6.6	ORM4	27663	165937	95366816
6.6-708	ORM5	1569	15308	10200001
TOTALE		636722	2009008	1040940269

Tabella 2.2: suddivisione delle classi per la tipologia d'uso ordinaria ascensore (SA)

LIMITE PER POTENZA CONTRATTUALE (kW)	classe	NUMERO N <sub>h</sub>	SOMMATORIA POTENZA CONTRATTUALE (kW)	SOMMATORIA ENERGIA ANNUA (kWh)
0-2.3	SAM1	254	338	310089
2.3-3.6	SAM2	744	2231	1076414
3.6-5.3	SAM3	167	753	459021
5.3-6.6	SAM4	286	1714	902485
6.6-708	SAM5	40	460	147511
TOTALE		1491	5497	2895520

Notiamo come la suddivisione delle classi è diversa dal caso precedente, qui si riescono a suddividere dieci sottoclassi e ad esse verrà applicato il campionamento stratificato.

L'analisi è stata fatta esattamente come nel caso precedente ovvero dividendo due casi:

- 1 Campionamento stratificato avendo come variabile discriminante la potenza contrattuale degli utenti residenziali monofase
- 2 Campionamento stratificato avendo come variabile discriminante l'energia consumata degli utenti residenziali monofase

### **Output e stima dell'intervallo di confidenza**

Andando ad analizzare direttamente l'output e la stima dell'intervallo di confidenza si evidenziano le principali caratteristiche di questa tipologia di utenti.

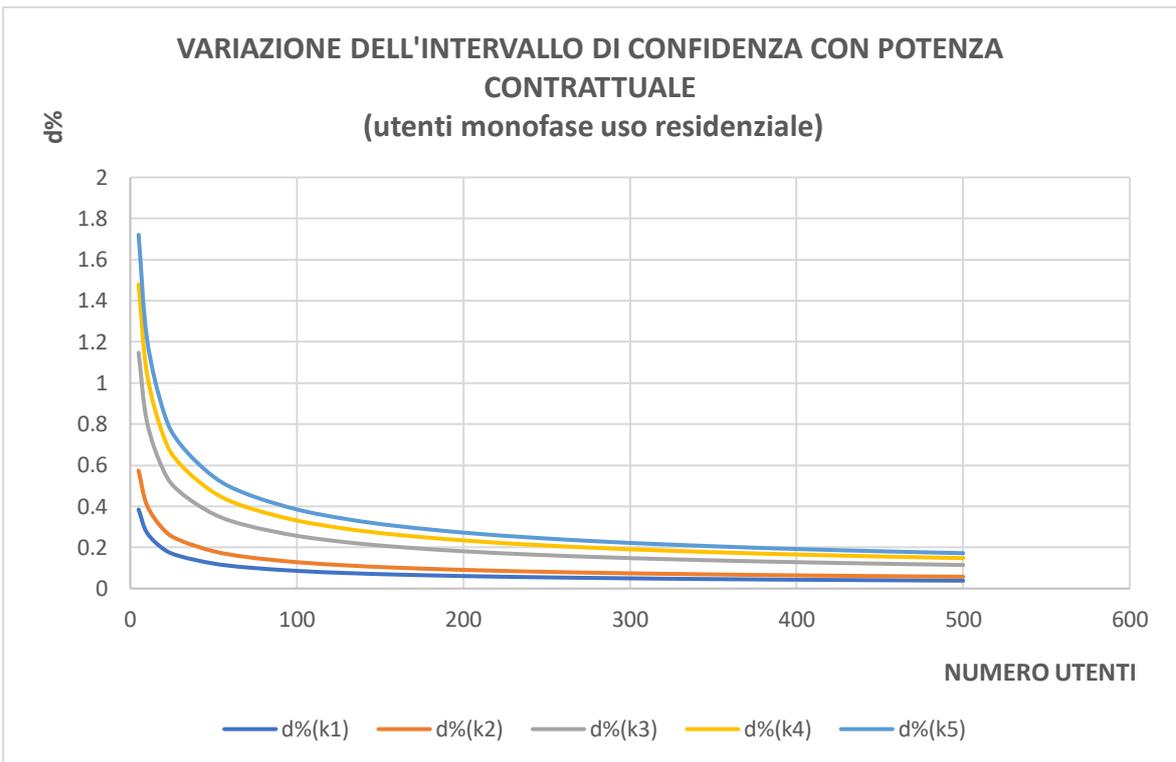


Grafico 3: variazione dell'intervallo di confidenza con caratteristica variabile potenza contrattuale, esempio degli utenti monofase anno 2020

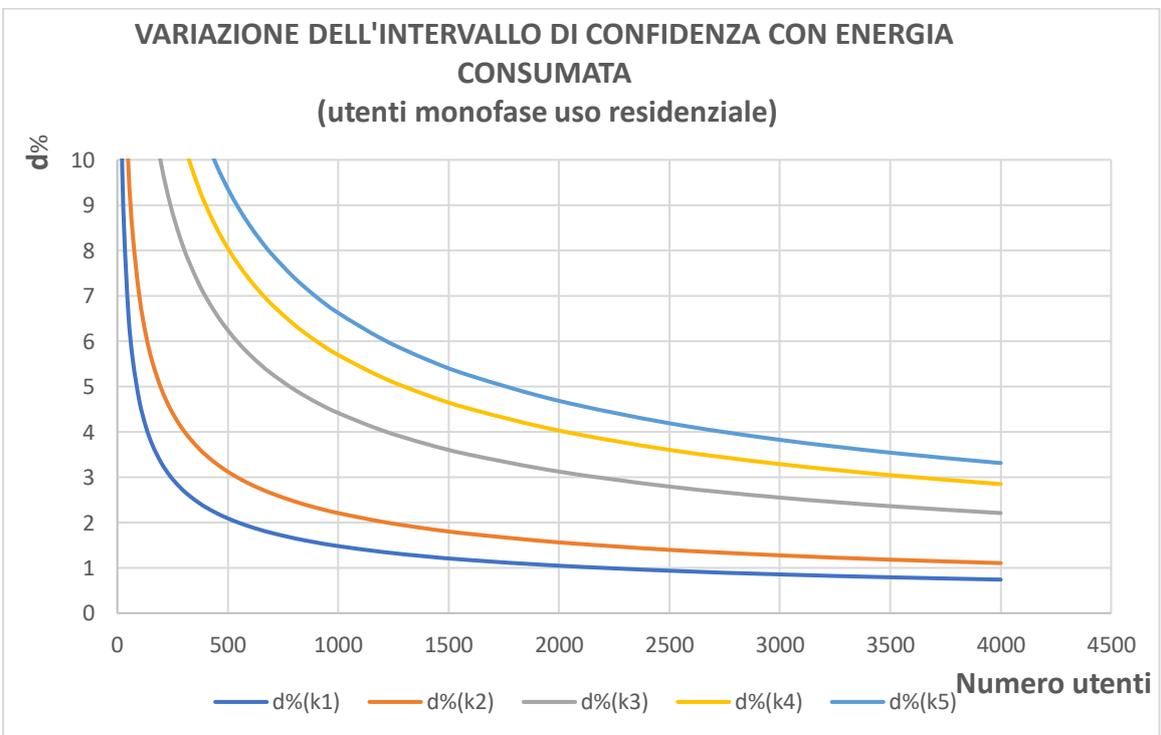


Grafico 4: variazione dell'intervallo di confidenza con caratteristica variabile per energia consumata, esempio degli utenti monofase anno 2020

Nel caso monofase, dove oltre il 90% dei dati sono utenti residenziali a potenza contrattuale 3 kW, si può osservare, nel grafico 3, che per 200 utenti siamo sotto 0.2% della variazione dell'intervallo di confidenza anche con la probabilità di confidenza del 99.7% ( $k=3$ ). Quindi possiamo fare un'accurata analisi prendendo solamente 200 utenti per il caso a potenza contrattuale. Nel caso dell'energia consumata (grafico 4), notiamo che per 3000 utenti il valore della variazione dell'intervallo di confidenza è del 4% con la probabilità di confidenza del 99.7%. Dunque, servirebbero molti più dati rispetto al primo caso per fare un'accurata analisi. Ciò è dovuto all'elevato numero di utenti che hanno potenze contrattuali uguali, esse essendo valori fissi e non modificabili hanno una deviazione standard e dunque una varianza, molto minore rispetto ai dati relativi all'energia consumata, la quale è molto variabile anche all'interno delle stesse classi poiché dipende esclusivamente dal consumo dell'utente.

#### 4.2.1. Ripartizione finale degli utenti

Analizzando i dati in uscita della stima dell'intervallo di confidenza, si è scelto di puntare a prendere 2500 utenti sia nel caso a potenza contrattuale che nel caso dell'energia consumata. Notando le distribuzioni molto diverse, si è pensato di confrontare le 2 tabelle e di prendere il numero massimo di utenti per ogni classe. In questo modo sappiamo che, come minimo, abbiamo preso il numero sufficiente di dati per tutte le classi anche se qualcuna sarà in eccesso.

Distribuzione dei 2500 utenti tenendo conto dell'energia consumata

Tabella2.1

UTENTI TOT	ORM1	ORM2	ORM3	ORM4	ORM5	SAM1	SAM2	SAM3	SAM4	SAM5
2500	219	1802	198	241	30	2	4	1	3	0

Distribuzione dei 2500 utenti tenendo conto della potenza contrattuale

Tabella2.2

UTENTI TOT	ORM1	ORM2	ORM3	ORM4	ORM5	SAM1	SAM2	SAM3	SAM4	SAM5
2500	1071	733	446	86	138	10	2	3	2	10

Il numero finale di utenti da analizzare e la rispettiva distribuzione nelle differenti classi diventa:

Tabella2.3

UTENTI TOT	ORM1	ORM2	ORM3	ORM4	ORM5	SAM1	SAM2	SAM3	SAM4	SAM5
3728	1071	1802	446	241	138	10	4	3	3	10

Questo è il nostro punto di partenza per analizzare i profili di carico degli utenti monofase a uso residenziale di ogni classe.

## 5. Metodi di cluster nell'era dei big data

Oggi è sempre più complicato acquisire, salvare ed elaborare l'enorme quantità di dati del sistema elettrico, il quale invia ai sistemi centrali milioni di informazioni al giorno. Troppi dati disponibili difficili da elaborare portano a nessuna informazione, per cui è molto importante escogitare dei metodi efficienti che ci permettano la corretta analisi dei dati per poter migliorare il funzionamento del sistema elettrico in generale che ci consente di comprendere i comportamenti degli utenti, permettendo sia di perfezionare i loro piani tariffari che per aver un miglior controllo della flessibilità.

Un metodo efficiente per analizzare le caratteristiche di un certo numero di dati è il *clustering* o analisi dei gruppi. Il *clustering* consiste in un insieme di metodi per unire differenti dati in classi omogenee dove i differenti gruppi di dati prenderanno il nome di cluster. I dati all'interno di un cluster presentano tra di loro delle similarità e presentano delle dissimilarità con i dati in altri cluster. I differenti algoritmi di cluster funzionano in maniera molto simile tra di loro: l'input è costituito da una serie di dati casuali, mentre in output abbiamo la suddivisione dei dati in diversi gruppi (o cluster) in cui i dati sono suddivisi per la loro similarità.

### Calcolo della similarità e dissimilarità nei cluster

I diversi algoritmi di cluster variano tra di loro per i differenti metodi di calcolo di similarità tra un dato e l'altro all'interno dello stesso gruppo e per i vari metodi di calcolo di dissimilarità tra gruppi differenti.

Esistono molteplici metodi di calcolare la similarità dei dati, i due principali sono:

Distanza di *Minkowski* 
$$d(x_i, x_j) = \left( \sum_{k=1}^p |x_{ki} - x_{kj}|^m \right)^{\frac{1}{m}}$$

Distanza euclidea  $d(x_i, x_j) = \sqrt{\sum_{k=1}^n (x_{ki} - x_{kj})^2} = |x_{ki} - x_{kj}|$

Come possiamo notare la distanza di *Minkowski* con  $m=2$  diventa una distanza *euclidea*.

La misura di dissimilarità tra due oggetti può essere definita dai disallineamenti delle caratteristiche corrispondenti, essa è nota come la distanza di Gower.

La distanza di Gower può essere espressa in questo modo:

$$d(x_i, x_j) = \sum_{k=1}^p \delta(x_{ki}, x_{kj}) \quad \text{dove } d(x_i, x_j) = \begin{cases} 0 & x_i = x_j \\ 1 & x_i \neq x_j \end{cases}$$

La dissimilarità totale può essere calcolata come:

$$d(X_i, X_j) = \sum_{k=1}^p \delta(x_{ki}, x_{kj}) + \sqrt{\sum_{k=p+1}^n (x_{ki} - x_{kj})^2}$$

## 5.1. Algoritmi di clustering

I differenti algoritmi di clustering si possono suddividere in due categorie principali:

- *Algoritmi gerarchici*: suddividono gli oggetti da analizzare in differenti gruppi rappresentabili in una struttura ad albero. Costruiscono una struttura gerarchica.
- *Algoritmi partizionali*: suddividono gli oggetti da analizzare in differenti cluster in modo da ridurre il più possibile la dispersione all'interno del cluster e aumentare il più possibile la dispersione tra i diversi cluster.

Analizziamo le principali caratteristiche di entrambi.

### Algoritmi partizionali

Di seguito sono analizzati i principali passaggi degli algoritmi di tipo partizionali, di fondamentale importanza per il nostro studio.

1. Si sceglie in anticipo il numero totale di gruppi, cluster, che si vogliono creare.
2. Si assegna a ciascuno dei cluster, in maniera casuale, i dati presi in esame.
3. Troviamo i centroidi di ogni gruppo con la seguente formula

$$M_p = (1/n_p) * \sum_{i=1}^{n_p} x_{kp}$$

dove  $M_p$  è il centroide del cluster  $p$ ,  $n_p$  è il numero dei dati nel gruppo  $p$ , e  $x_{kp}$  è l' $x$ -esimo dato del cluster  $p$ .

4. Si calcola la distanza degli elementi dal centroide e otteniamo un errore  $e_p$ .
5. La sommatoria di tutti gli errori dell'insieme dei gruppi dà l'errore totale  $e_{tot}$ . Esso è l'errore da minimizzare.
6. Si ripete il passaggio dal punto 2 riassegnando gli elementi in base al meno distante dal centroide.
7. Ripetendo i passaggi fino al punto 6 si minimizza l'errore totale  $e_{tot}$  fino al suo valore minimo preimpostato, oppure il processo va avanti finché i dati all'interno del cluster non raggiungono un equilibrio.

### ***K-means e K-medoids***

*K-means* e *k-medoids* sono due dei principali algoritmi partizionali. Entrambi, per funzionare, necessitano di conoscere il numero di cluster  $k$  già nel primo passaggio.

Il *K-means* determina i cluster con la distanza maggiore tra il valore medio e tutti gli elementi di un singolo cluster. L'algoritmo del *K-means* è sensibile al modo in cui sono posizionati i dati iniziali, per questo motivo l'errore potrebbe convergere a un minimo locale e non assoluto. Inoltre questo metodo è molto sensibile ai disturbi all'interno dei dati soprattutto se essi sono molto diversi (es. diversi ordini di grandezza come vedremo in seguito) rispetto a tutti gli altri elementi.

Il *k-medoids* determina i cluster con la distanza maggiore tra i centroidi in ogni cluster, non il valore medio come nel caso precedente. Questa variante gli concede di essere meno sensibile ai dati rumorosi e ai dati molto diversi tra loro.

### **Algoritmi gerarchici**

Non solo gli algoritmi di tipo partizionale verranno usati nel nostro studio, ma anche quelli gerarchici. Il clustering gerarchico produce raggruppamenti di oggetti, dati, seguendo un percorso gerarchico.

Due sono le strategie per il clustering gerarchico:

1. *Agglomerativo* o “bottom up”: il punto di partenza è l'inserimento di ciascun dato in un cluster differente e si procede quindi all'accorpamento graduale di cluster a due a due.
2. *Divisivo* o “top down”: vengono posizionati tutti i dati in un unico cluster che piano piano viene suddiviso in ulteriori micro-cluster.

In principio, tutti i dati disponibili non appartengono a nessun raggruppamento specifico. Successivamente, l'algoritmo procede iterativamente, in una delle due strategie elencate, aggregando gruppi di dati per similarità, se qualche dato non è simile a nessun altro dato verrà isolato e verrà chiamato outlier.

Quando si studia il clustering gerarchico, come nel caso degli algoritmi di tipo partizionale, un fattore molto importante è il calcolo della distanza o dissimilarità tra un dato rispetto agli altri. Come visto poche pagine prima, esistono numerose tecniche per il calcolo della distanza, spesso variano a seconda del tipo di dato studiato poiché alcune misurazioni sono appropriate solo per certe tipologie di analisi.

Di seguito sono elencate alcune metriche utilizzate negli algoritmi di clustering gerarchico:

- La distanza euclidea
- La distanza di Manhattan
- La distanza di Mahalanobis
- La distanza di Hamming

## 5.2. Caso studio: Clustering degli utenti Ireti

Il primo passo per l'applicazione del metodo del campionamento stratificato, vista nel primo capitolo, è stata la suddivisione dei dati, forniti dall'azienda Ireti S.P.A, in classi o gruppi. Il procedimento usato era molto semplice: sono stati identificati gli utenti con tipologia d'uso più numerosa, ovvero gli SA e gli OR e successivamente sono state create le classi con un'analisi empirica basata solamente sui limiti della potenza contrattuale.

Una volta analizzate le diverse tipologie di *clustering* si è pensato di provare a vedere quali fossero i risultati del metodo di clustering *K-means* variando però in ogni caso studio i criteri del cluster.

### 5.2.1. Caso studio 1: Clustering K-means dei dati secondo la potenza contrattuale per i dati trifase mensili

Come primo caso, si è presa in considerazione la stessa grandezza usata nell'applicazione del campionamento stratificato, ovvero la potenza contrattuale. Ciò è stato fatto per vedere le differenze nei risultati tra l'analisi empirica e l'algoritmo di *clustering k-means*.

Il procedimento di *clustering k-means* è stato elaborato su Matlab. Avendo i dati per dodici mesi il procedimento è stato ripetuto per ogni mese del 2020. E' stato riportato come esempio il clustering dei dati di gennaio 2020.

Tabella 3.1: suddivisione dei gruppi per la tipologia d'uso ordinaria (OR) utilizzando il k-means

GRUPPI	LIMITE PER POTENZA CONTRATTUALE (kW)	NUMERO GEN
ORT1	0-8	15033
ORT2	9-16	17841
ORT3	16.5-24	3697
ORT4	25-39	5014
ORT5	40-67.5	2885

ORT6	68-113	1269
ORT7	114-185	457
ORT8	187-300	167
ORT9	316-504	25
ORT10	1050	1
TOTALE		46389

Tabella 3.2: suddivisione dei gruppi per la tipologia d'uso ordinaria (SA) utilizzando il k-means

GRUPPI	LIMITE PER POTENZA CONTRATTUALE (kW)	NUMERO UTENTI GEN
SAT1	0-7	8198
SAT2	8-13.5	10107
SAT3	14-31	7657
SAT4	33-56.3	562
SAT5	56.5-84	286
SAT6	85-122	184
SAT7	123-167	68
SAT8	169-220	36
SAT9	243-420	11
SAT10	630	1
TOTALE		27110

Facendo un confronto con i risultati ottenuti con il metodo empirico (vedere Tabella 1.1 e Tabella 1.2) si può notare che nell'algoritmo *k-means* alcuni gruppi vengono uniti creando così classi molto numerose. Prendendo in considerazione il mese di gennaio vediamo che il gruppo ORT1 con il metodo *k-means* è composto da 15033 utenti, mentre la classe ORT1 con il metodo empirico ha 7901 utenti, ovviamente questo perché sono cambiati i limiti massimi e minimi della potenza contrattuale che definiscono le classi per cui cambierà il numero di utenti in tutti i gruppi, o classi, da un caso all'altro.

Un modo interessante per rappresentare i risultati è costruire delle cumulative dei diversi cluster, sia per il caso ORT sia per il caso SAT per ogni mese dell'anno.

Di seguito verranno rappresentati solo i dati di gennaio poiché il raggruppamento dei successivi mesi è pressoché identico.

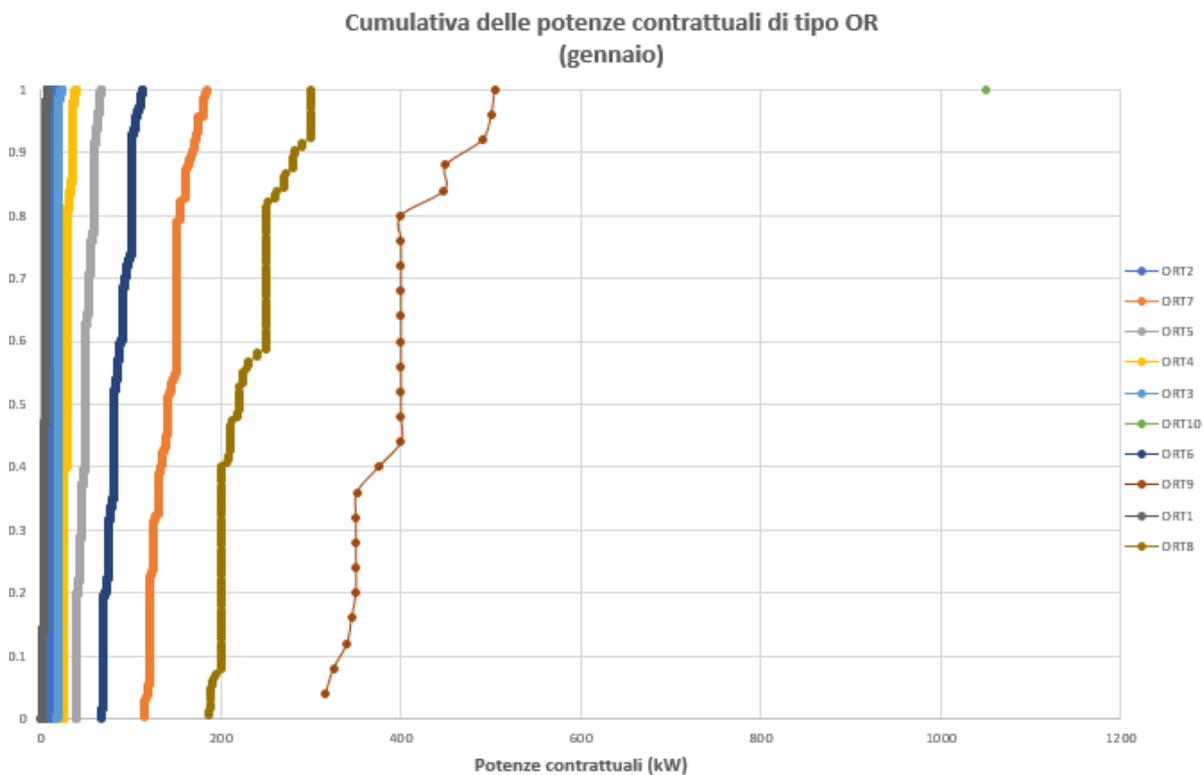


Figura 3.1: grafico della cumulativa delle potenze contrattuali di tipo OR

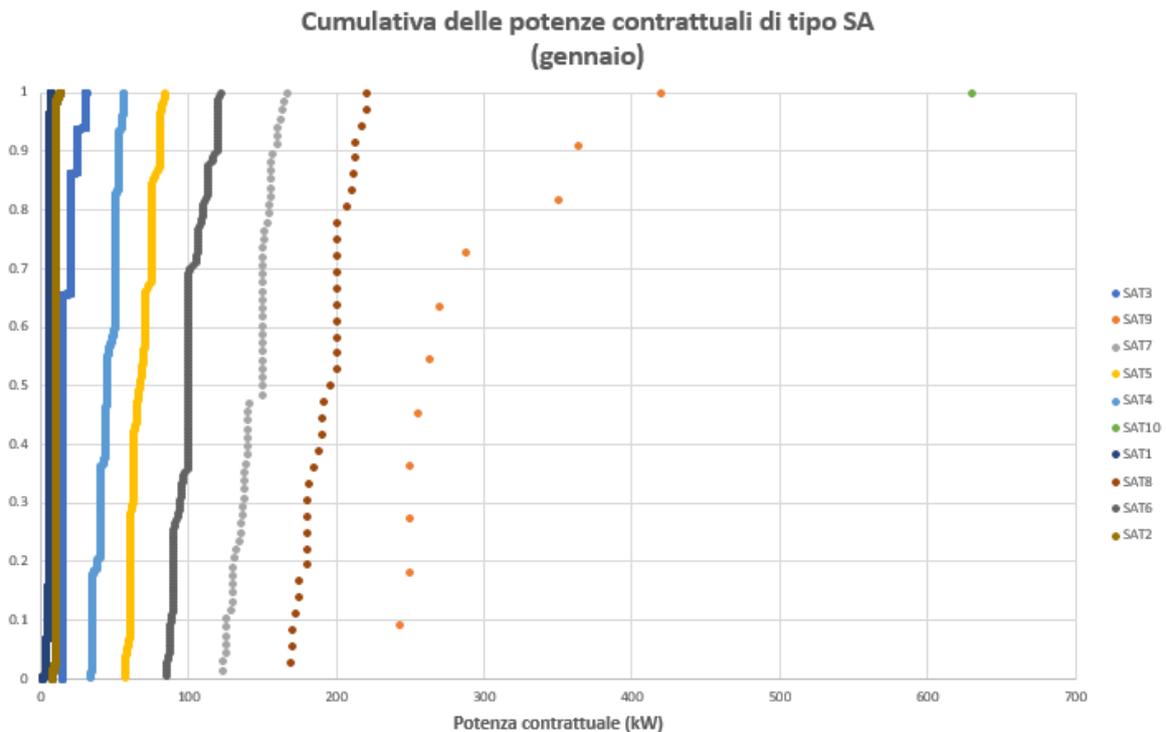


Figura 3.2: grafico della cumulativa delle potenze contrattuali di tipo SA

Come possiamo notare dai grafici e dalle tabelle numeriche (Tabella 3.1 e Tabella 3.2) i primi cluster (da ORT1 a ORT4 e da SAT1 a SAT4) sono i più numerosi, per cui l'algoritmo *k-means* li divide anche se hanno potenze contrattuali simili tra i diversi gruppi.

Si può anche notare come questo l'algoritmo di *clustering k-means* è sensibile ai dati molto diversi tra di loro, come si vede nelle tabelle 3.1 e 3.2 e nelle figure 3.1 e 3.2. Il gruppo ORT10 e il gruppo SAT10 sono rappresentati da un solo utente con potenza contrattuale molto maggiore rispetto a tutte le altre. Per queste loro diversità i due utenti sono stati assegnati a un gruppo isolato.

### 5.2.2. Caso studio 2: Clustering k-means dei dati secondo l'utilizzazione per i dati trifase mensili

In questo caso si è presa in considerazione l'utilizzazione di ogni singola utenza. L'utilizzazione è il rapporto tra l'energia elettrica consumata e la potenza massima prelevata in un periodo di tempo, nel nostro caso, di un mese. L'utilizzazione è una grandezza che viene misurata in ore.

$$U = \frac{E}{P} [h]$$

Il rapporto tra le ore di utilizzazione, della massima potenza, e le ore totali nel relativo periodo, un mese, rappresenta un indice della continuità del carico. Per cui questo dato fornisce un'idea sulla variabilità del carico nel periodo preso in esame.

Dopo aver calcolato l'utilizzazione mensile di ogni utente per tutti i dodici mesi, si è potuto utilizzare il metodo di *Clustering K-means* per creare dieci gruppi con caratteristiche di utilizzazione differente. Di seguito vediamo i risultati di gennaio, mese campione.

Tabella 3.3: suddivisione dei gruppi prendendo come grandezza l'utilizzazione

GRUPPI	Utilizzazione (h)	Numero utenti Gennaio
Gruppo1	0-25.7	19276
Gruppo2	25.7-51.8	20139
Gruppo3	51.8-83.4	13166
Gruppo4	83.4-121.8	8952
Gruppo5	121.8-168.3	5882
Gruppo6	168.3-228.1	3486
Gruppo7	228.3-314.6	1775
Gruppo8	314.6-459	673
Gruppo9	461-830.5	127
Gruppo10	851.7-1780	18
TOTALE		73491

Come visto nel caso precedente i risultati del clustering si possono osservare anche attraverso una cumulativa.

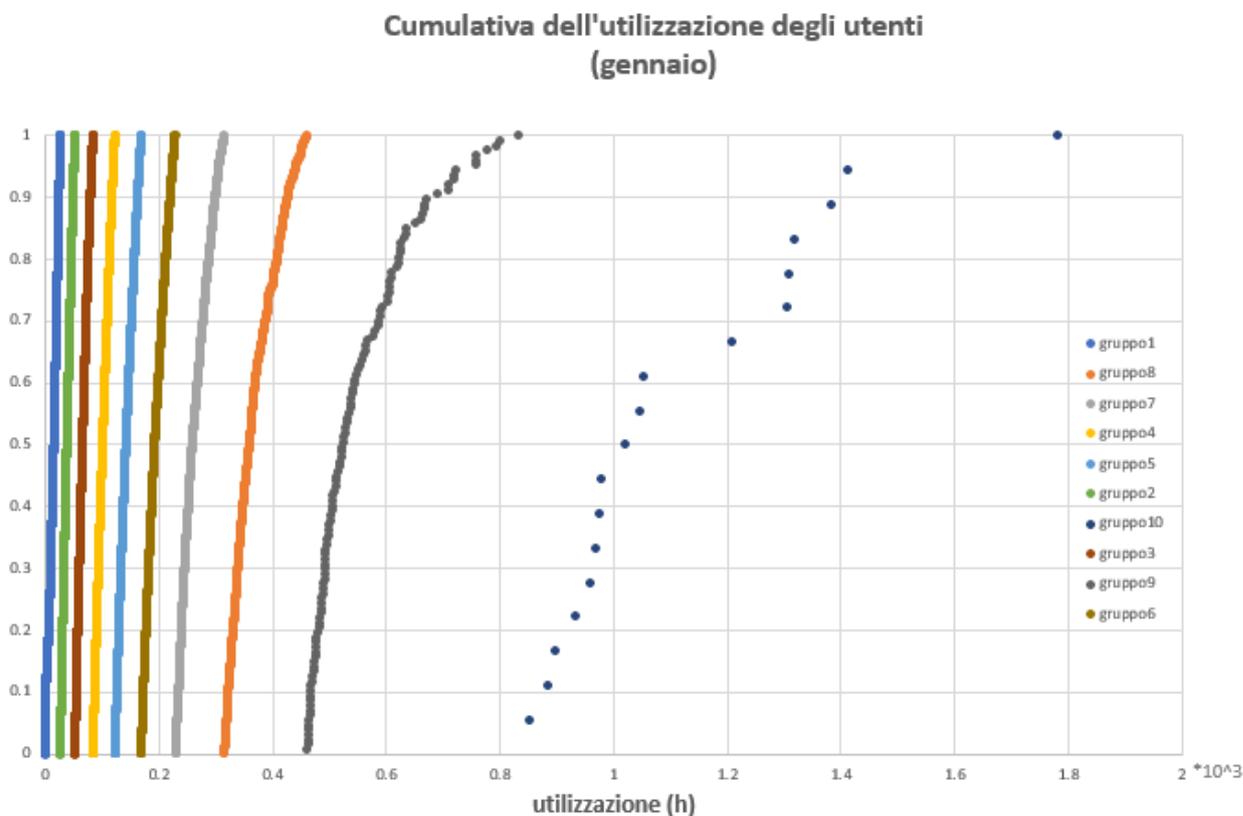


Figura 3.3: grafico della cumulativa dell'utilizzazione

Come già osservato in precedenza, con il metodo di *clustering k-means* i dati con ordine di grandezza molto diverso rispetto agli altri dati sono isolati e sono numericamente minori rispetto a quelli che hanno lo stesso ordine di grandezza tra loro. Come possiamo notare nella Tabella 3.3, i gruppi più numerosi sono i primi, ovvero quelli con utilizzazione simile.

### Caso studio 2.1: Clustering k-means dei dati secondo l'utilizzazione per i dati trifase mensili del mese di aprile 2020

Avendo analizzato l'utilizzazione in tutti i mesi dell'anno 2020 si è notata una diminuzione dell'utilizzazione, che coinvolge a grandi linee tutti gli utenti trifase, nel mese di aprile. Questa leggera diminuzione è dovuta alla chiusura della maggior parte delle attività commerciali e industriali durante il periodo del lockdown varato dal

governo a metà marzo del 2020 e finito a inizio maggio 2020. Avendo unicamente dati mensili, l'unico mese utile dove osservare qualche cambiamento riguardante il consumo è aprile.

*Tabella 3.3.1: suddivisione dei gruppi prendendo come grandezza l'utilizzazione nel mese di aprile 2020*

GRUPPI	Utilizzazione (h)	Numero utenti Aprile
Gruppo1	0-18.5	23505
Gruppo2	18.5-38.7	21942
Gruppo3	38.7-64.2	13265
Gruppo4	64.2-98.2	7437
Gruppo5	98.2-144.7	4021
Gruppo6	144.7-210.2	2127
Gruppo7	210.2-314.4	953
Gruppo8	315.6-534.7	262
Gruppo9	546.8-946.3	23
Gruppo10	1248.1-1786.2	3
TOTALE		73540

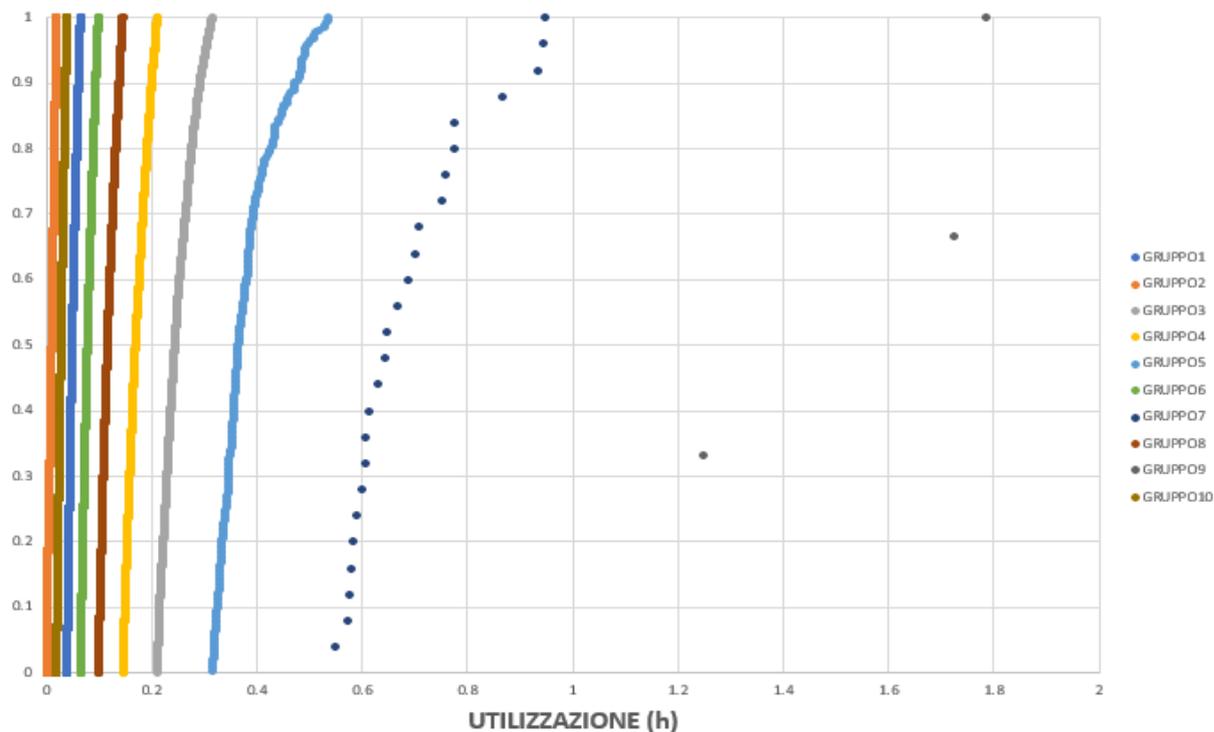


Figura 3.3.1: grafico della cumulativa dell'utilizzazione del mese di aprile 2020

Confrontando la Tabella 3.3 con la Tabella 3.3.1 e la Figura 3.3 con la Figura 3.3.1, si può notare che l'utilizzazione nelle differenti classi è diminuita. Difatti ogni gruppo del mese di aprile ha il limite massimo di utilizzazione minore rispetto a quello di gennaio. Ciò vuol dire che gli utenti hanno consumato meno nel mese di aprile rispetto a quello di gennaio. Come accennato prima, questo è strettamente collegato con le diverse chiusure di attività commerciali e industriali dovute al lockdown imposto dal governo.

### 5.2.3. Caso studio 3: Clustering K-means dei dati secondo la potenza e l'utilizzazione per i dati monofase residenziali

Cambiano i dati ma il ragionamento è lo stesso dei casi 1 e 2. Avendo solo dati residenziali annuali, in questo caso, non devo prendere nessun mese come mese campione.

Come numero di cluster si è preso lo stesso usato nell'applicazione dei dati monofase a uso residenziale del campionamento stratificato ovvero dieci, cinque per la tipologia d'uso ORM e cinque per SAM. Il procedimento di *clustering k-means* è stato elaborato su Matlab e di seguito sono riportati i differenti raggruppamenti.

Risultati del *clustering K-means* dei dati secondo la potenza contrattuale.

Tabella 3.4: suddivisione dei gruppi prendendo come grandezza la potenza contrattuale per i dati monofase ad uso residenziale.

GRUPPI	LIMITE PER POTENZA CONTRATTUALE (kW)	NUMERO UTENTI
ORM1	0.5-2	14662
ORM 2	2.2-3.5	478676
ORM 3	3.95-5	32007
ORM 4	5.3-8	26015
ORM 5	9-20	1344
SAM1	0.5-2	161
SAM2	3-3.5	531
SAM3	4-5	152
SAM4	5.5-8	250
SAM5	10-15	38
TOTALE		553836

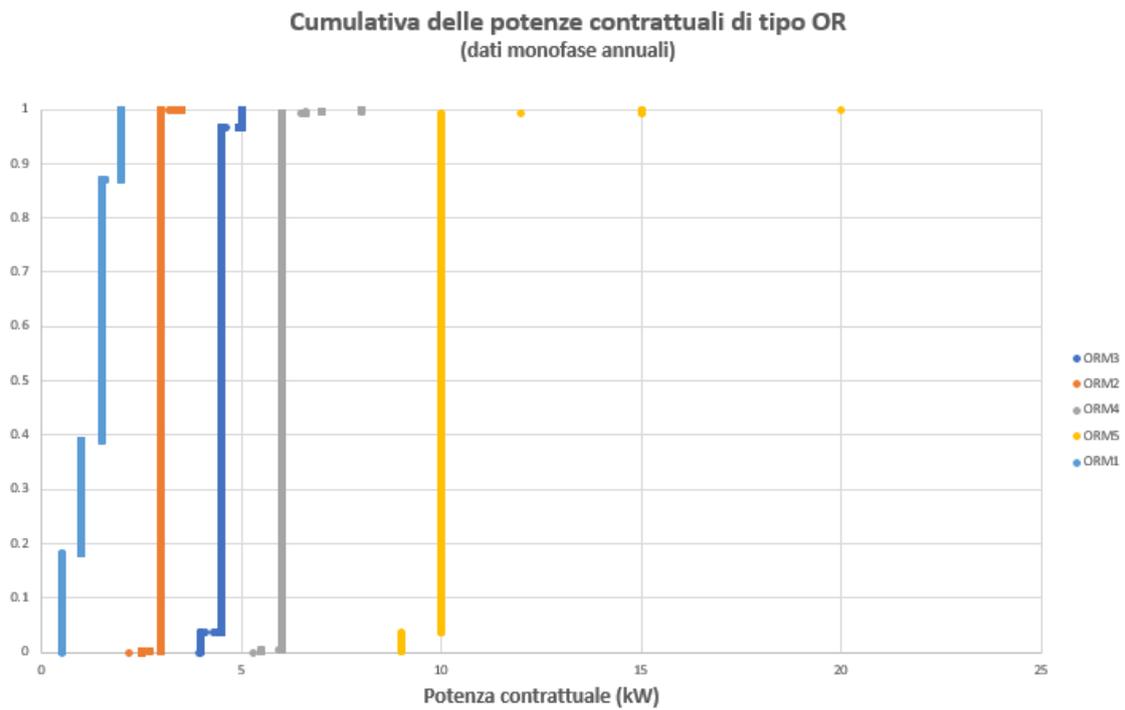


Figura 3.4: grafico della cumulativa delle potenze contrattuali di tipo OR dati residenziali monofase annuali

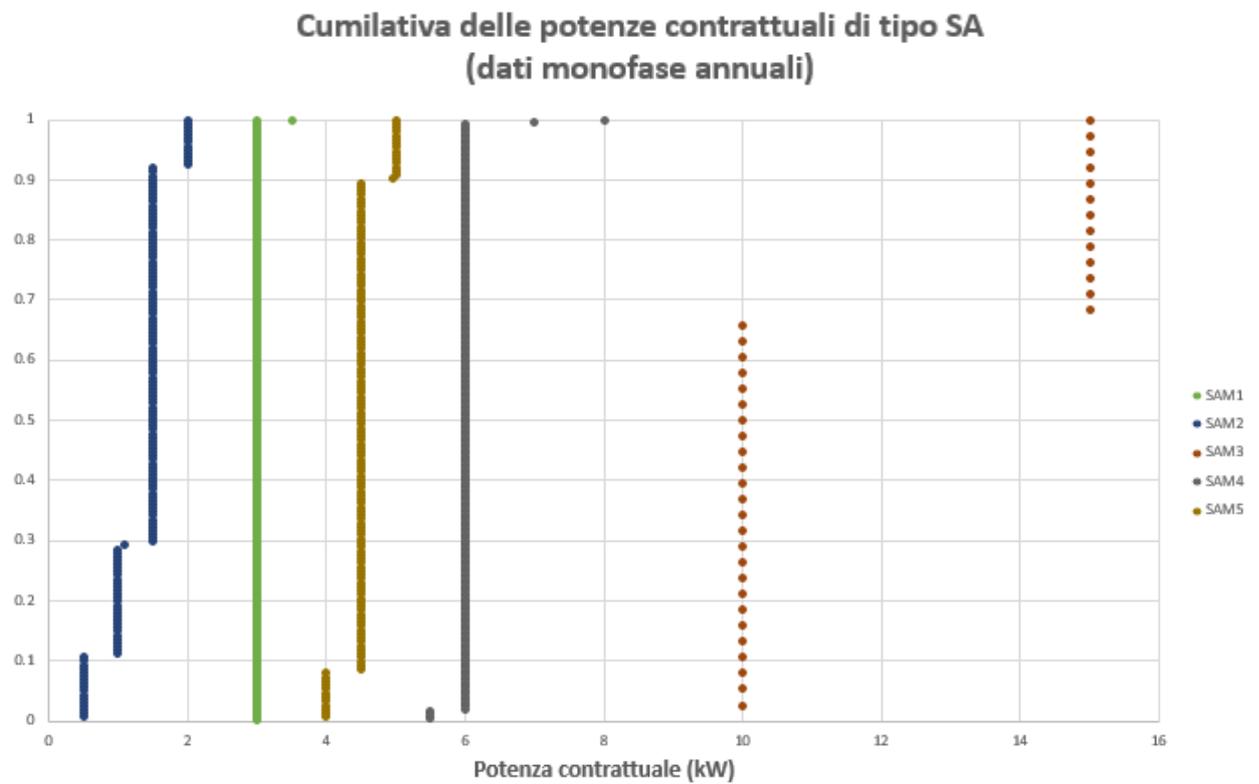


Figura 3.5: grafico della cumulativa delle potenze contrattuali di tipo SA dati residenziali monofase annuali

Risultati del *clustering K-means* dei dati secondo l'utilizzazione.

Tabella 3.5: suddivisione dei gruppi prendendo come grandezza l'utilizzazione per i dati monofase ad uso residenziale.

GRUPPI	Utilizzazione (h)	NUMERO UTENTI
Gruppo1	50.7-308.5	107471
Gruppo2	308.6-448.17	120323
Gruppo3	448.22-595.3	109757
Gruppo4	595.3-760.8	91041
Gruppo5	760.8-958.3	62530
Gruppo6	958.4-1216	36947
Gruppo7	1216.4-1600	17301
Gruppo8	1600.8-2261	6271
Gruppo9	2262-3588	1813
Gruppo10	3601-17850	382
TOTALE		73491

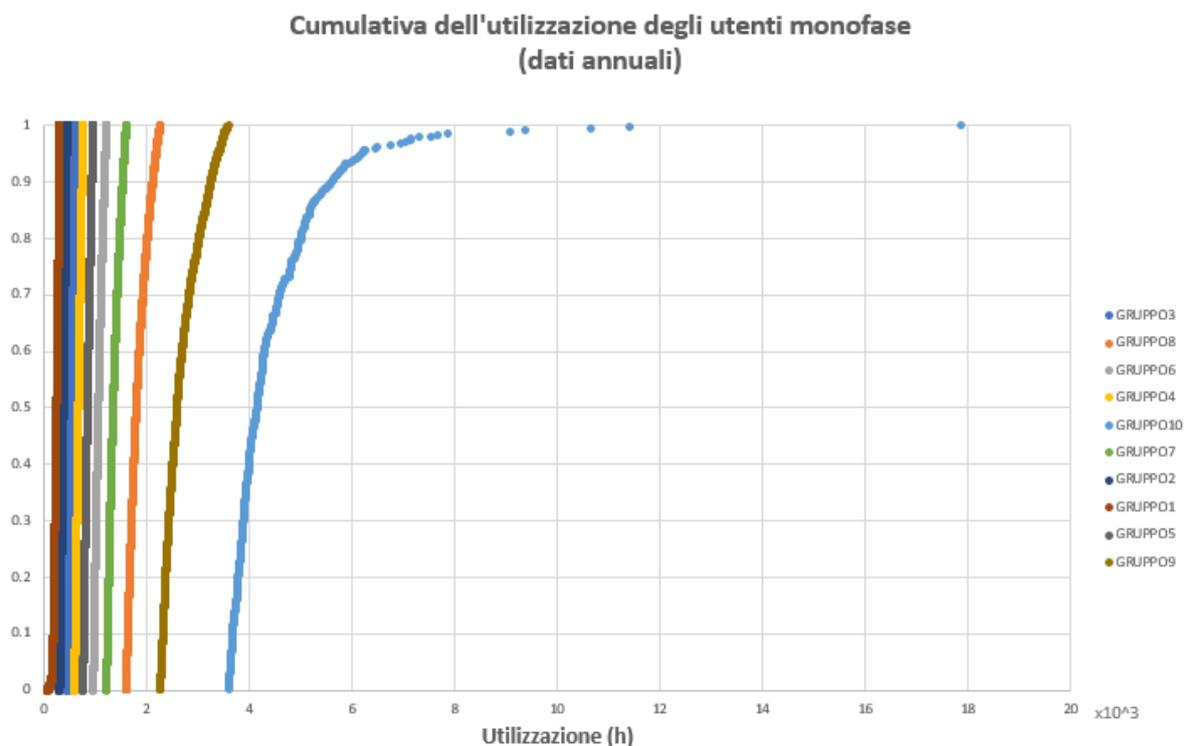


Figura 3.6: grafico della cumulativa dell'utilizzazione per gli utenti residenziali monofase

## **5.3. Metodo di Clustering K-means dei dati tenendo conto di più variabili**

Finora si sono analizzate le differenti tipologie di *clustering* considerando solo una variabile alla volta, ovvero la potenza contrattuale o l'utilizzazione. Di seguito sono riportati ulteriori casi studio di come creare i gruppi, sempre con il metodo di *clustering k-means*, tenendo conto di più variabili come potenza, energia e utilizzazione. Questo procedimento può essere chiamato clustering bidimensionale.

Studiando i differenti casi si riesce a comprendere come si comporta l'algoritmo di *clustering del k-means* a seconda dei dati forniti.

### **5.3.1. Caso studio 4: Clustering K-means bidimensionale tenendo conto di energia e potenza**

In questo primo caso si rielaborano i dati di gennaio in una chiave diversa rispetto al caso 1. Il clustering dei dati viene effettuato utilizzando due variabili differenti ma fissando il numero dei gruppi sempre a dieci. Sull'asse delle ascisse poniamo la potenza contrattuale, mentre sulle ordinate l'energia. I risultati di questo metodo possono essere visualizzati nella Figura 3.7.

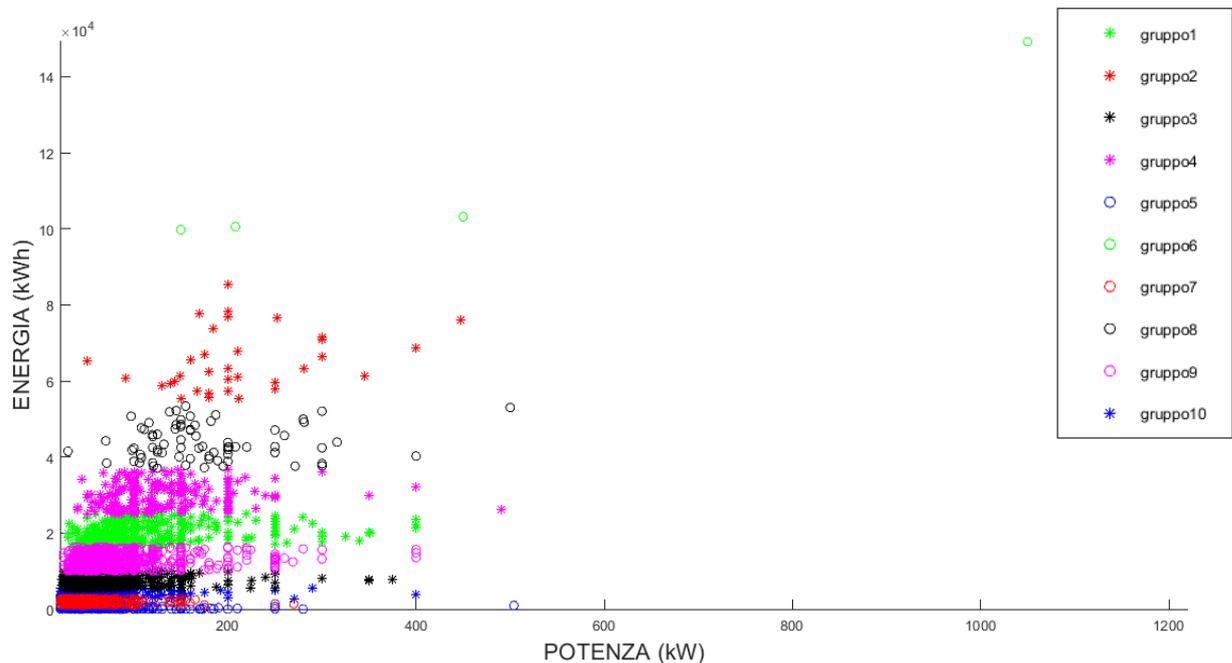


Figura 3.7: Grafico del clustering k-means tramite energia e potenza

Come possiamo osservare, il taglio dei cluster è orizzontale, ovvero all'interno dei gruppi si mantengono differenti valori di potenza, ma il gruppo cambia con il cambiare dell'energia, ciò vuol dire che è predominante la variabilità dell'energia rispetto a quella della potenza. Questo ha senso, perché le potenze sono valori fissi e non modificabili da contratto, mentre il dato relativo all'energia è molto variabile poiché dipende esclusivamente dalle caratteristiche di consumo dell'utente.

### 5.3.2. Caso studio 5: Clustering K-means bidimensionale tenendo conto di utilizzazione e potenza

In questo caso analizziamo un nuovo raggruppamento considerando come variabile per creare i gruppi l'utilizzazione. L'utilizzazione, come visto nel capitolo precedente, è il rapporto tra energia e potenza ciò vuol dire che le due grandezze analizzate in questo caso non sono indipendenti. Analizzando il risultato nel grafico sottostante possiamo vedere delle piccole differenze rispetto al caso 4.

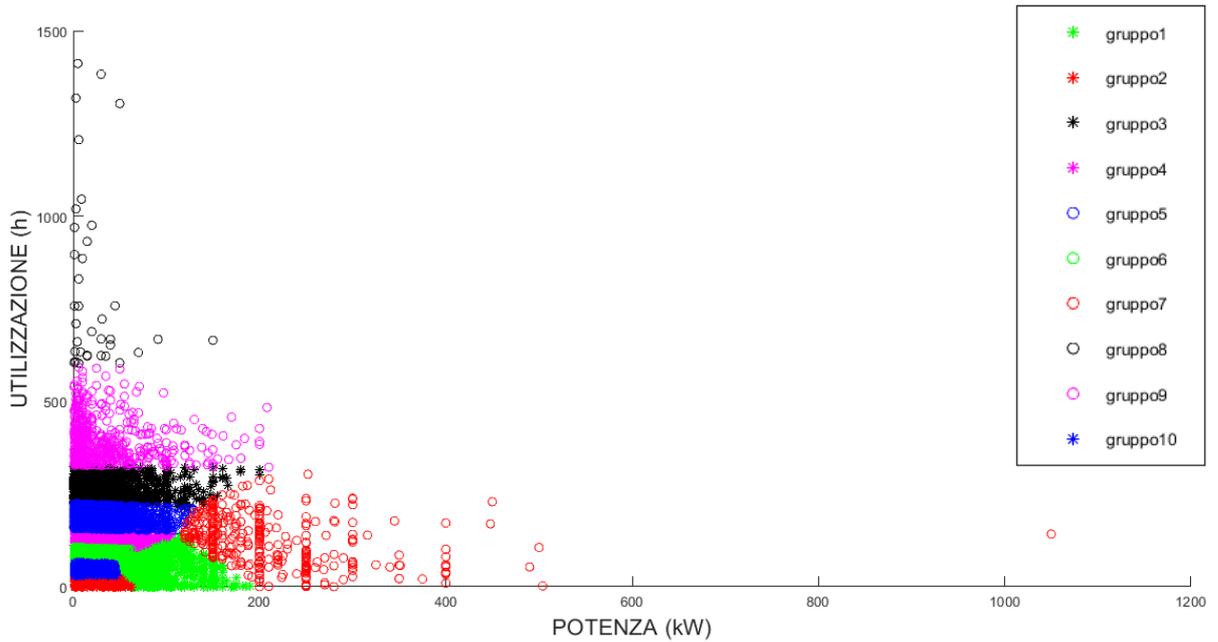


Figura 3.8: Grafico del clustering *k*-means dell'utilizzazione e della potenza

Confrontandolo con la Figura 3.7, possiamo notare che il taglio della maggior parte dei cluster è orizzontale proprio come nel caso 4, questo poiché l'utilizzazione ha una variabilità maggiore rispetto alla potenza. Non tutti i cluster però hanno il taglio orizzontale, ad esempio il gruppo 8 e il gruppo 2 hanno un taglio diagonale, ciò è dovuto al fatto che le due variabili poste sugli assi non sono indipendenti.

### 5.3.3. Caso studio 5.1: Clustering K-means bidimensionale tenendo conto dell'utilizzazione normalizzata e della potenza

Si vuole fare una prova per analizzare i risultati dell'algoritmo di *clustering k-means* utilizzando la normalizzazione dell'utilizzazione. La normalizzazione dell'utilizzazione è il rapporto tra le ore di utilizzazione con il numero di ore del mese preso in esame, nel nostro caso gennaio, per cui si è diviso per 744.

Nel grafico sottostante possiamo notare che normalizzando l'utilizzazione alcuni utenti superano il valore dell'unità, ciò significa che l'utente preso in esame sta consumando per più ore del mese il massimo della potenza consentita dal suo contratto, e ciò non è possibile.

Andando ad analizzare meglio i dati della potenza contrattuale delle poche utenze che superano il limite di utilizzazione pari a uno, si è potuto osservare un errore nella memorizzazione dei dati da parte dell'azienda Ireti poiché le forniture in esame in realtà hanno un'altra potenza contrattuale che è cambiata durante l'anno precedente, ma il database non è stato aggiornato adeguatamente. Motivo per il quale i clienti che nella Figura 3.9 superano il valore dell'unità in realtà hanno una potenza contrattuale più elevata per cui la loro reale utilizzazione è inferiore a uno.

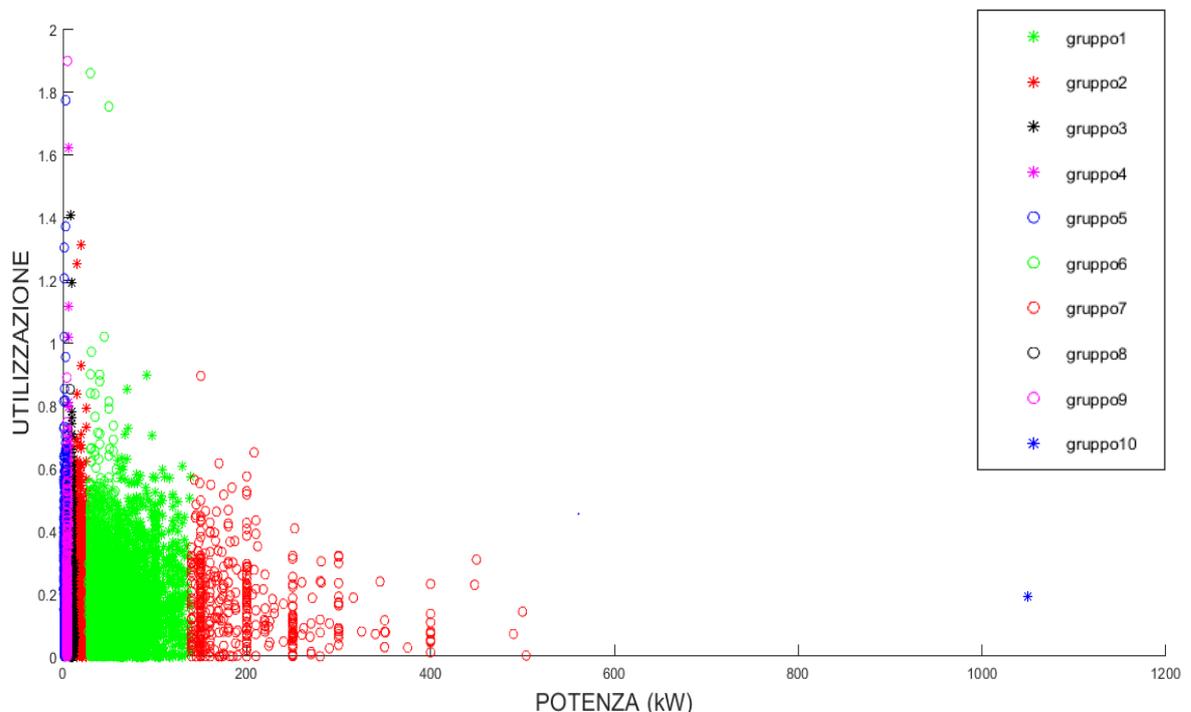


Figura 3.9: Grafico del clustering k-means tramite la normalizzazione dell'utilizzazione e potenza

Analizzando ulteriormente la Figura 3.9 e confrontandola con la Figura 3.8 del caso 5, possiamo accorgerci che le grandezze dei due grafici poste sugli assi sono le stesse, ma il risultato del clustering tra un caso e l'altro è completamente diverso. Si può osservare che nel caso 5 i valori dell'utilizzazione non erano normalizzati, dunque i risultati del cluster venivano formati seguendo la variabilità dell'utilizzazione, ovvero formando tagli orizzontali. Ora con i valori dell'utilizzazione normalizzati si ha l'ordine di grandezza dell'unità, per cui l'algoritmo *k-means* crea nuove tipologie di gruppi tenendo in considerazione la variabilità delle potenze che è predominante rispetto all'utilizzazione solo perché l'utilizzazione ha cambiato il suo ordine di grandezza passando da valori con due o tre zeri a valori sotto l'unità. Per questo motivo il taglio dei cluster è verticale.

Dopo aver notato questa importante caratteristica, si può fare un'importante considerazione sull'algoritmo di *clustering k-means*. Cambiando l'ordine di grandezza dei dati, l'algoritmo cambia il modo di raggruppare le classi, ciò è dovuto alle caratteristiche intrinseche dell'algoritmo. Questo esempio serve come dimostrazione per aumentare la nostra attenzione sull'importanza dei dati quando utilizziamo algoritmi di clustering.

#### **5.3.4. Caso studio 6: Clustering K-means bidimensionale tenendo conto della potenza e dell'energia facendo il clustering sull'utilizzazione**

Come ultimo caso studio si è pensato di creare un nuovo raggruppamento unendo tutte e tre le grandezze analizzate nei casi precedenti. Questo è possibile poiché, come già accennato, l'utilizzazione è dipendente sia dalla potenza che dall'energia. Sapendo ciò si è deciso di considerare sull'asse delle ascisse la potenza e su quello delle ordinate l'energia e di fare il clustering dei diversi gruppi sull'utilizzazione. Il risultato può essere intuibile. Le classi sono divise in linee diagonali poiché il dato su cui si è fatto il cluster dei gruppi (l'utilizzazione) è strettamente dipendente dagli assi (potenza ed energia).

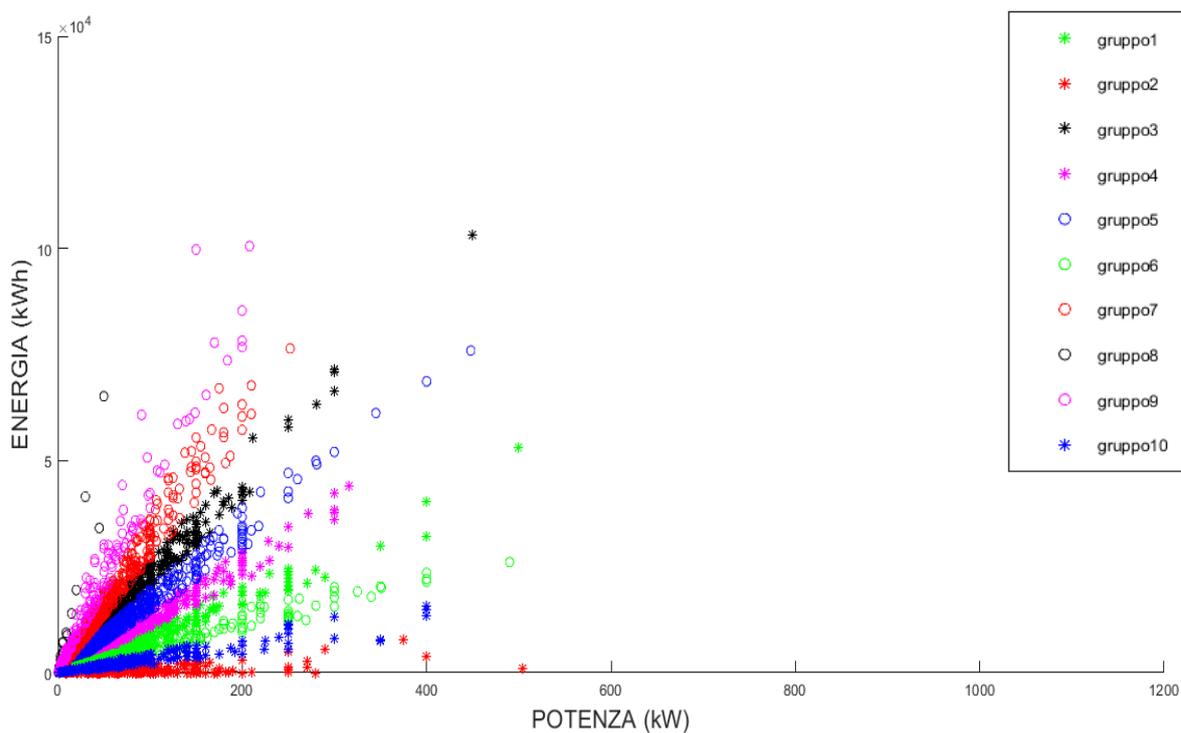


Figura 3.10: Grafico del clustering *k-means* sull'utilizzazione tenendo conto dell'energia e della potenza

### Osservazioni

Grazie all'algoritmo di *clustering k-means* sono state create numerose classi molto differenti dal ragionamento empirico riportato nel capitolo 4. La diversità della formazione delle classi nei differenti casi è data dalla tipologia e dalla dimensione della variabile considerata. Infatti, variando la grandezza presa in considerazione cambia il modo in cui l'algoritmo taglia i differenti cluster, creando in ogni caso studio gruppi con caratteristiche totalmente diverse gli uni dagli altri.

Questa analisi è utile alla creazione delle diverse macro-classi per una miglior comprensione delle caratteristiche dei clienti a seconda della variabile presa in esame che può essere la potenza contrattuale, l'utilizzazione o l'energia consumata.

## 6. Determinazione delle classi di clienti studiando le curve di carico

Alla fine del capitolo 4.1 e 4.2, grazie all'applicazione del metodo del campionamento stratificato, abbiamo trovato la ripartizione degli utenti finali, dunque, il numero minimo e sufficiente di forniture, in tutte le classi, necessari per continuare la classificazione dei consumatori rispetto alle macrocategorie predefinite.

La ripartizione ideale delle classi richiesta a Ireti per avere un'ottima stima dell'intervallo di confidenza è riportata in Tabella 1.15 e in Tabella 2.3, rispettivamente per le utenze trifase non residenziali e per le utenze monofase a uso residenziale. Purtroppo la realtà, come spesso accade, non è identica alla teoria, per cui le curve di carico delle forniture richieste a Ireti non è la stessa dell'obiettivo prefissato nei capitoli 4.1 e 4.2. Bisogna dunque cambiare obiettivo modificando il numero di utenti da analizzare e quindi la stima dell'intervallo di confidenza.

Una premessa che bisogna fare è che le curve di carico degli utenti con una potenza contrattuale al di sotto dei 50 kW non vengono salvate nel database di Ireti, per cui per le macro-classi più basse e tutti gli utenti monofase non è stato possibile effettuare l'analisi sulle curve di carico.

L'unica analisi possibile da effettuare è stata per gli utenti trifase a uso non residenziale con potenza contrattuale maggiore o uguale a 50 kW.

Un'ulteriore premessa da fare, prima di passare all'applicazione delle diverse procedure di clustering, è che Ireti ha dato a disposizione le curve di carico giornaliere di tutti i giorni dell'anno per ogni fornitura, ma noi non possiamo studiarle ed analizzarle tutte poiché l'onere di calcolo diventerebbe nuovamente troppo grande. Bisogna, dunque, andare ad analizzare le curve di carico dei giorni che possiamo definire coerenti che sono sostanzialmente i giorni feriali, escluse le festività che ovviamente possono cadere all'interno della settimana. A questo punto si va a prendere un giorno a metà del mese dei giorni risultanti coerenti e si prende la curva di carico di ogni utente di quel giorno. Prendiamo questo giorno come esempio rappresentativo del mese e lo facciamo per ogni utente.

Successivamente vediamo alcuni esempi delle curve di carico rappresentative di diversi utenti in diversi mesi di differenti macro-classi.



Grafico 6.1: curva di carico rappresentativa del mese di agosto dell'utente 200428720 della macro-classe ORT9



Grafico 6.2: curva di carico rappresentativa del mese di febbraio dell'utente 200380346 della macro-classe ORT8

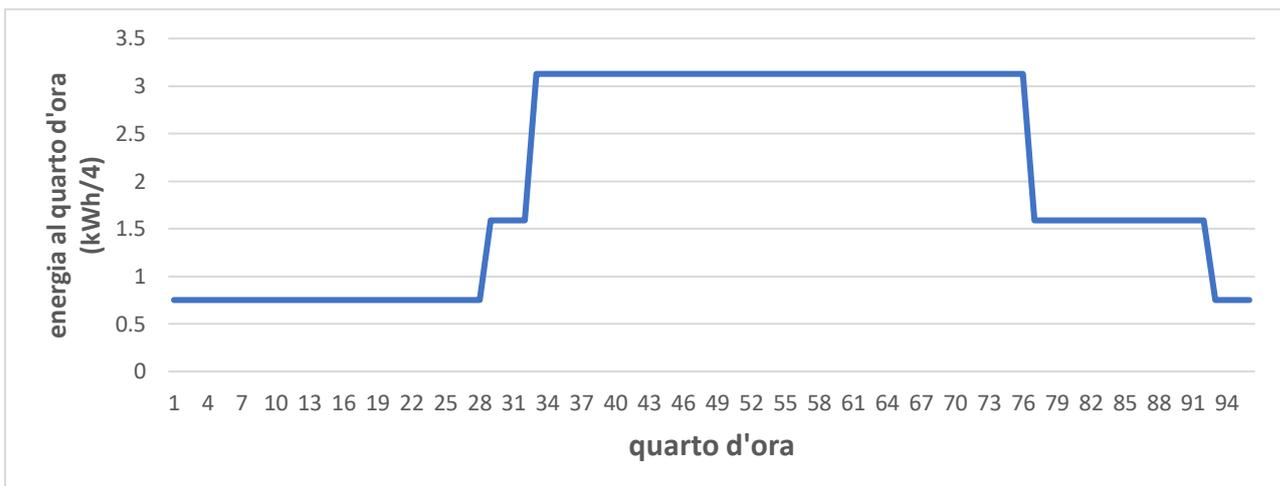


Grafico 6.3: curva di carico rappresentativa del mese di giugno dell'utente 101169055 della macro-classe SAT7

Come possiamo vedere nel grafico 6.3 la curva di carico dell'utente 101169055 rimane costante nel tempo per certi tratti, questo accade non solo all'utente in questione, ma a molti altri utenti. Una motivazione per cui ciò accade è che i misuratori hanno una risoluzione verticale bassissima a quattro livelli, nel senso che quando il contatore misura il consumo il risultato viene posizionato o sopra o al di sotto di uno dei quattro livelli preimpostati. Questa discretizzazione verticale fa sì che le letture nel tempo abbiano dei valori fissi e dunque per certi tratti costanti.

### Ripartizione reale delle macro-classi per utenti trifase a uso non residenziale

Le curve di carico dei clienti ricevute sono state analizzate e filtrate per togliere alcuni errori, ad esempio alcune forniture si ripetevano più volte nello stesso mese o alcuni dati erano incompleti. Inoltre, tramite un filtro si è notato che solo 795 utenti avevano le curve di carico che si ripetevano nel corso dei 12 mesi. Visto che la nostra analisi prevede lo studio delle curve di carico durante tutto l'anno, siamo obbligati a prendere e studiare le curve di carico di questi 795 utenti, eliminando dalle restanti forniture. Le curve di carico sono così ripartite nelle macro-classi predefinite

Tabella 6.1: Suddivisione reale trovata delle macro-classi

ORT6	ORT7	ORT8	ORT9	ORT10	SAT6	SAT7	SAT8	SAT9	SAT10
83	251	103	100	131	18	50	14	23	22

Ripartizione richiesta.

Tabella 1.15

utenti tot	ORT1	ORT2	ORT3	ORT4	ORT5	ORT6	ORT7	ORT8	ORT9	ORT10	SAT1	SAT2	SAT3	SAT4	SAT5	SAT6	SAT7	SAT8	SAT9	SAT10
6842	315	446	1759	670	495	241	287	114	114	143	11	73	1623	304	73	45	57	22	27	23

Bisogna confrontare i risultati ottenuti dalla Tabella 6.1 con l'output del procedimento del campionamento stratificato, prendendo in considerazione il caso come variabile discriminate l'energia consumata, Tabella 1.9, per i motivi spiegati precedentemente. Questo si fa poiché dobbiamo cambiare i valori di riferimento che ci eravamo prefissati variando la precisione decisa.

Per andare a identificare il livello di accuratezza del campionamento stratificato che possiamo garantire dobbiamo andare a confrontare la nuova ripartizione creata (Tabella 6.1) con il foglio di calcolo del campionamento stratificato (Tabella 1.9), cambiando il numero totale di utenti per andare a trovare il numero che può portare ad avere tutti i valori della Tabella 6.1 rispettati. Per fare ciò, dobbiamo farci guidare dal caso peggiore della nostra suddivisione reale. Nel nostro caso la classe peggiore è la classe ORT6, ovvero quella di cui disponiamo di pochi utenti rispetto a quanti davvero dovremmo averne.

Il confronto con la Tabella 1.9 indica che il numero di utenti totali dove abbiamo rispettato la grandezza di tutte le classi, compresa la peggiore, è 900.

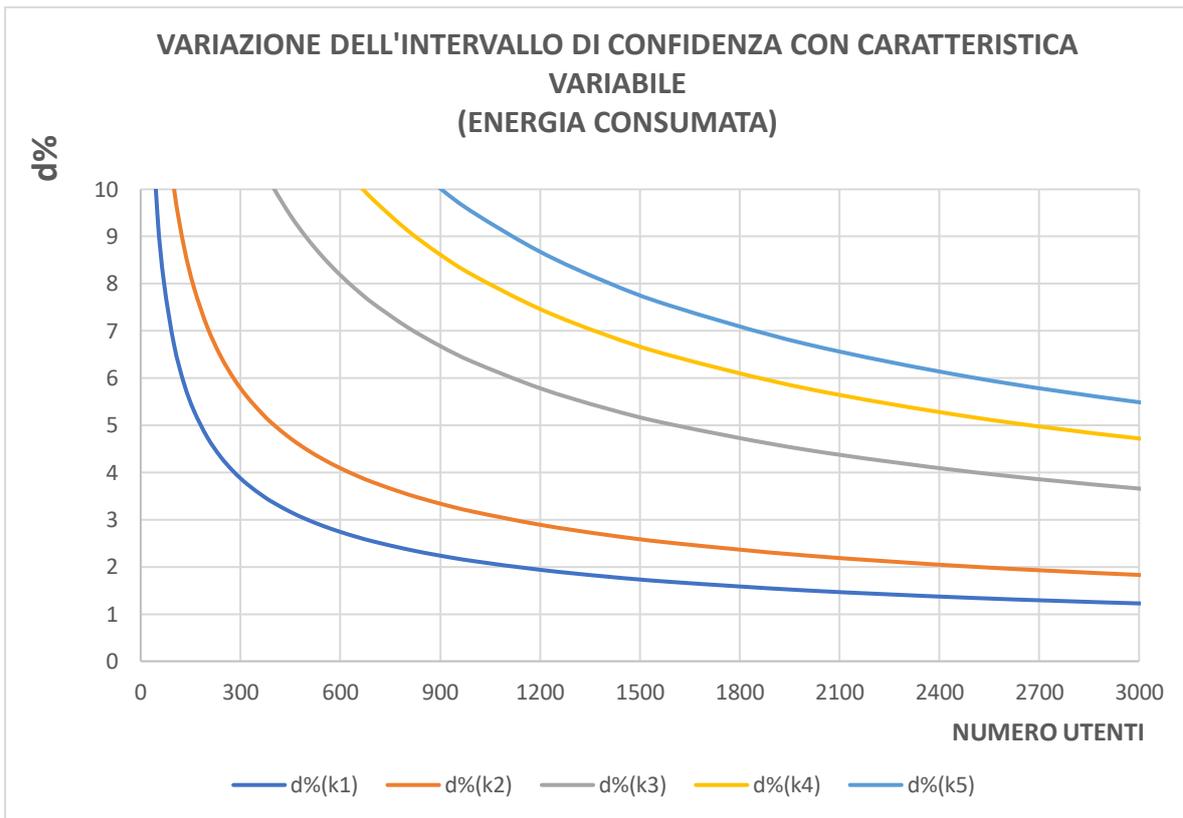
Essi sono così distribuiti:

Tabella 6.2

ORT6	ORT7	ORT8	ORT9	ORT10	SAT6	SAT7	SAT8	SAT9	SAT10
82	103	37	40	42	14	20	7	9	8

Il livello di confidenza garantito dal campionamento stratificato, con il numero totale di utenti pari a 900, oscilla tra il 2.23%, nel caso migliore, e per il 10%, nel caso peggiore, come si può vedere nel grafico sottostante.

Avendo un numero superiore di utenti in tutte le macro-classi analizzate, come si può vedere confrontando la Tabella 6.1 e la Tabella 6.2, possiamo sicuramente affermare che siamo molto lontani dal caso peggiore.



*Grafico 6.4: variazione dell'intervallo di confidenza con caratteristica variabile per energia consumata, esempio degli utenti trifase del mese di marzo 2020*

## 6.1. Creazione delle micro-classi tramite procedure di clustering

### Curve di carico mensili e annuali

Dopo aver fatto alcune premesse sulla modifica dell'intervallo di confidenza e sulla creazione delle curve di carico rappresentative mensili per ogni fornitura, si può passare alla creazione delle micro-classi tramite procedure di clustering partendo dalle macro-classi precedentemente create.

Per vedere se all'interno di ogni macro-classe emergono dei cluster, bisogna andare a suddividere le curve di carico rappresentative di ogni fornitura mese per mese. Lo stesso ragionamento verrà fatto andando a mettere in fila le curve di carico rappresentative mese per mese di ogni fornitura, creando la nuova curva di carico rappresentativa annuale come si può osservare nel Grafico 6.6.

Prendendo come esempio la macro-classe ORT6 possiamo vedere come si distribuiscono le curve di carico mensili di tutte le forniture senza utilizzare nessun procedimento di clustering.

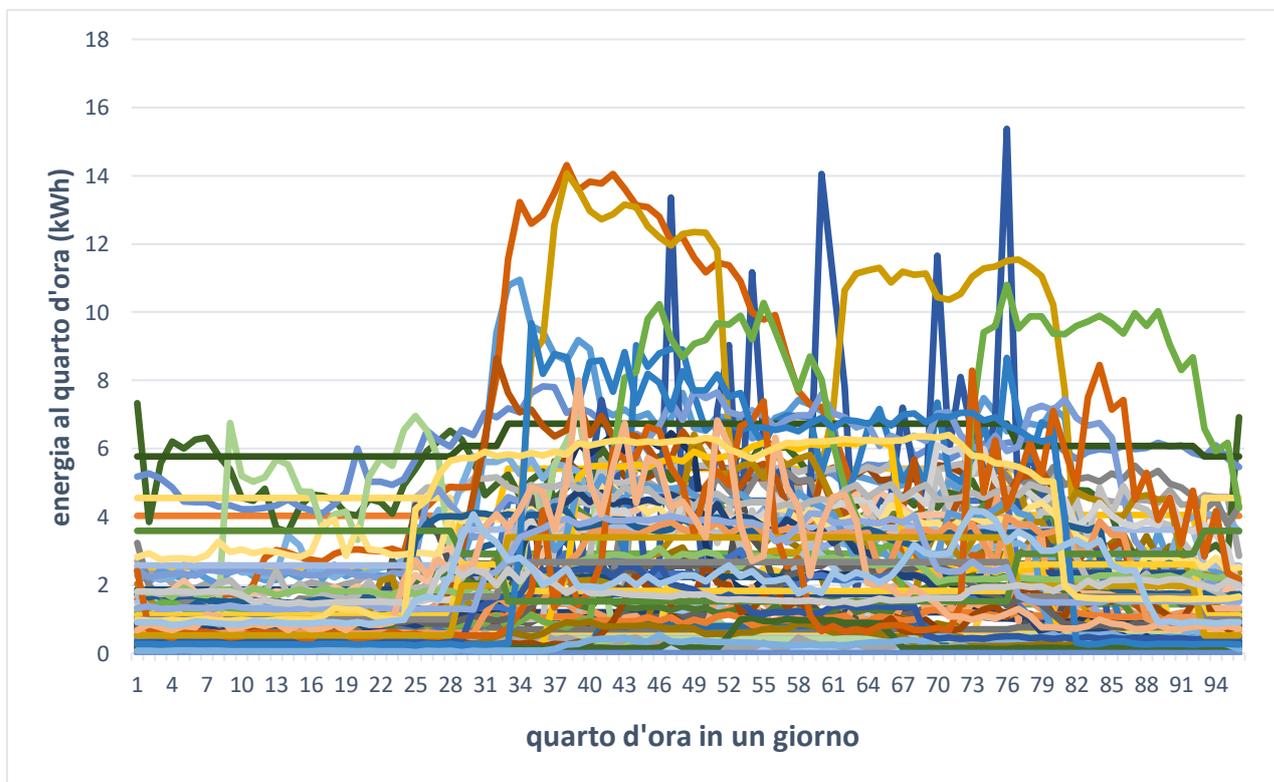


Grafico 6.5: curve di carico rappresentative di tutte le forniture della classe ORT6 nel mese di gennaio

Allo stesso modo, vediamo un esempio della distribuzione delle curve di carico annuali di tutte le forniture della macro-classe SAT6.

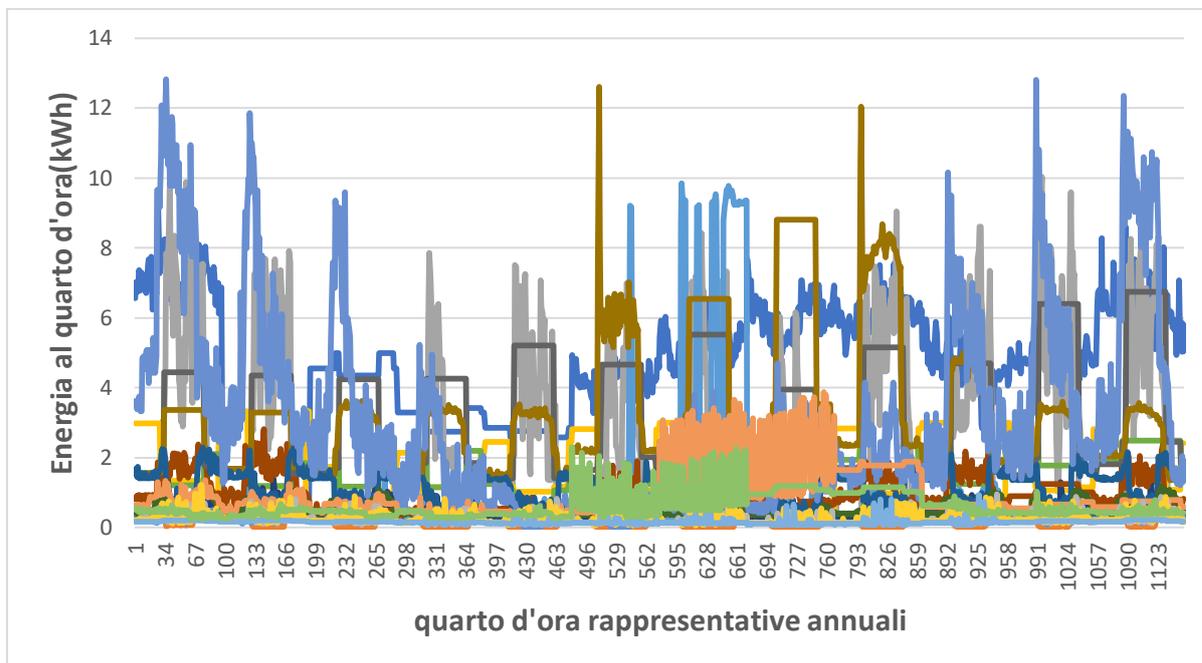


Grafico 6.6: curve di carico rappresentativa annuale di tutte le forniture della classe SAT6 nell'anno 2020

## 6.2. Creazione delle micro-classi tramite il metodo *k-means*

### Preparazione dati

Prima di parlare della procedura di clustering del *k-means*, bisogna preparare i dati da dare in ingresso all'algoritmo di clustering. I dati possono essere preparati in due modi differenti, ma il procedimento di clustering finale sarà lo stesso, solo che come input avremo due tipologie di dati differenti.

*Primo metodo:* creare per ogni fornitura una media annua della sua curva di carico. Si prendono tutte le curve di carico di ogni mese di un'unica fornitura e si fa la media. Il vettore finale sarà un vettore lungo 1x96 e rappresenterà la curva di carico media annua.

*Secondo metodo:* creare per ogni fornitura un vettore lungo 12x96 dove sulla prima riga vi sarà la curva di carico rappresentativa di gennaio e nella riga sottostante ci sarà la curva di carico di febbraio e così via. Il vettore risultante sarà dunque composto da tutte le curve di carico di ogni mese.

### Clustering *K-means* sulle forniture

La prima procedura di clustering utilizzata per le creazioni dei gruppi di utenti per ogni classe è il *k-means*, procedura utilizzata anche nel capitolo 5.

Per ogni mese e per ogni macro-classe si è fatto girare l'algoritmo del *k-means* su Matlab e come risultato si ha la suddivisione delle forniture in cluster mese per mese. Il problema è che ogni mese il cluster può cambiare. Nel senso che il cluster numero 1 nel mese di gennaio può non rimanere lo stesso nel mese di febbraio poiché esso può essere rappresentato, ad esempio, dal cluster numero 4.

La risoluzione di questo problema si ha andando a vedere quante volte una fornitura sta nello stesso cluster di un'altra fornitura nei dodici mesi. Questo è facilmente visualizzabile in una matrice quadrata di dimensioni numero\_forniture x numero\_forniture, dove su righe e colonne si hanno le forniture della macro-classe presa in esame e su ogni casella vi è scritto il numero di volte che due forniture stanno nello stesso cluster nei diversi mesi dell'anno, indipendentemente dal numero del cluster dato dal output del *k-means*.

Questa matrice è stata creata partendo dai risultati del cluster *k-means* mese per mese e contando quante volte le forniture stavano nello stesso cluster utilizzando il programma Matlab.

*Tabella 6.3: esempio di una piccola parte della matrice che conta quante volte le forniture stanno nello stesso cluster ogni mese della macro-classe ORT10*

	<b>200266484</b>	<b>200277607</b>	<b>200293589</b>	<b>200293735</b>	<b>200302984</b>	<b>200302988</b>	<b>200309074</b>
<b>200266484</b>	12	11	5	5	1	10	2
<b>200277607</b>	0	12	6	4	1	9	1
<b>200293589</b>	0	0	12	2	0	3	0
<b>200293735</b>	0	0	0	12	7	5	7
<b>200302984</b>	0	0	0	0	12	3	10
<b>200302988</b>	0	0	0	0	0	12	4
<b>200309074</b>	0	0	0	0	0	0	12

Come possiamo vedere nella Tabella 6.3, la matrice in questione è diagonale superiore e la diagonale ha ovviamente il valore di 12, poiché sulla diagonale si trovano le forniture identiche che staranno nello stesso cluster tutti i mesi dell'anno.

Adesso bisogna prendere le forniture che hanno più elementi uguali a dodici sulla riga. Di esse ci interessano le curve di carico rappresentative che ci serviranno per trovare il centroide. Con il termine centroide si intende la media delle curve di carico delle forniture scelte.

La creazione del centroide è effettuata per i due modi diversi visti precedentemente. Nel primo metodo viene fatta la media annuale delle curve di carico; quindi, come output avremo un vettore che contiene la curva di carico media lungo 1x96, mentre nel secondo metodo il centroide viene creato dalla media dei vettori lunghi 12x96.

Una volta trovati i centroidi, si calcolano le distanze con le altre curve di carico dal centroide, permettendo in seguito di creare i cluster di utenti con curve di carico per similarità.

## Problema nelle creazioni dei cluster

Durante lo studio delle curve di carico per la creazione dei cluster si è verificato un problema che è dato dalla natura intrinseca dell'algoritmo *k-means*. Una caratteristica fondamentale di questa procedura è quella di mettere insieme tutti i dati diversi sotto un unico cluster. Nel procedere con l'analisi, questa predominante caratteristica è venuta fuori. La maggior parte dei cluster che si ritrovano insieme 12-11-10 volte nell'arco dell'anno sono tutte curve di carico completamente diverse tra di loro, dunque non hanno nessuna similarità, per cui è impossibile la creazione di micro-classes secondo questa procedura.

Si riporta un esempio di un raggruppamento di curve secondo la procedura del *k-means*.

Le curve di carico indicate sono le medie annue di ogni fornitura. Per questo esempio è stato usato il primo metodo per la creazione dei dati.

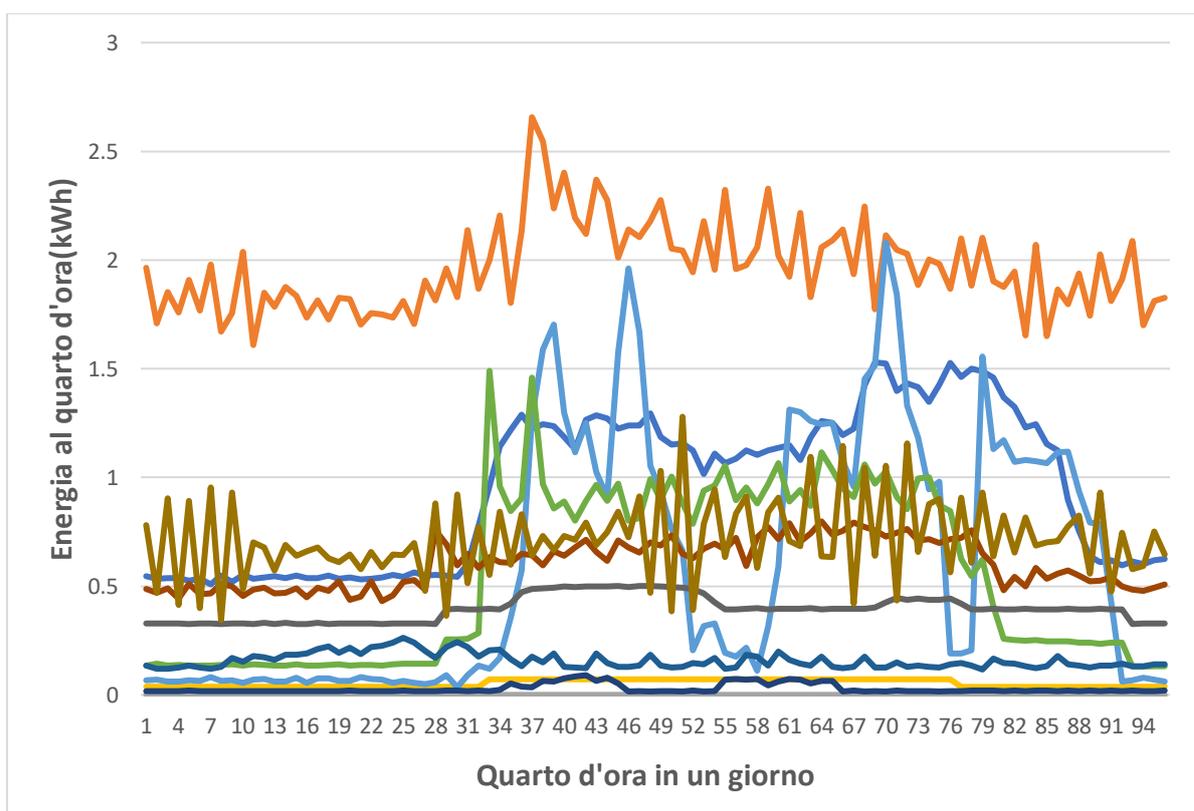


Grafico 6.7: esempio di curve di carico di diverse forniture della macro-classe ORT8 raggruppate sotto lo stesso cluster per 12 mesi l'anno

Come si può vedere, le forniture raggruppate sotto lo stesso cluster per tutti i mesi dell'anno hanno una distribuzione della curva di carico e una grandezza complessivamente diversa tra di loro. Possiamo affermare che l'algoritmo di clustering *k-means* non è adeguato a questo tipo di suddivisione.

## 6.3. Creazione delle micro-classi tramite il clustering gerarchico

Data l'impraticabilità di creare cluster di curve di carico attraverso l'algoritmo *k-means* si è pensato di utilizzare un'altra procedura di clustering; il clustering gerarchico. Il clustering gerarchico ha un algoritmo diverso da quello del *k-means* anche se essi si assomigliano su alcuni ragionamenti generali.

Un'unica premessa è da fare. Essendo, il numero dei cluster che si vogliono creare deciso a priori, si è scelto di assegnare il numero di dieci cluster per le macro-classi ORT, molto numerose, e il numero di cinque cluster per le macro-classi SAT, poco numerose. In questo modo per ogni caso analizzato si potranno vedere molto bene i vari raggruppamenti e i principali outlier. Con outlier si intendono le curve di carico anomale che non hanno nessuna caratteristica di similarità con le altre curve, dunque sono isolate.

### Preparazione dati

Per l'algoritmo del clustering gerarchico, la preparazione dei dati è molto più semplice rispetto a quella vista precedentemente. Bisogna prendere per ogni macro-classe tutte le curve di carico di ogni fornitura che si vuole analizzare e darle in ingresso all'algoritmo creato su Matlab.

La preparazione dei dati viene suddivisa in due differenti modi poiché si vogliono avere due differenti output.

*Prima suddivisione:* L'analisi viene fatta sulla curva di carico rappresentativa mensile di ogni macro-classe. In output si avranno i diversi cluster che potranno cambiare mese per mese a seconda del raggruppamento assegnato dall'algoritmo. Si avranno dodici output per ogni macro-classe, uno per ogni mese.

*Seconda suddivisione:* si fa un'analisi delle curve di carico rappresentative annuali delle forniture. Per creare ciò si mettono una dietro l'altra le curve di carico rappresentative di ogni fornitura mese per mese facente parte della stessa macro-classe. Si darà in pasto all'algoritmo una matrice lunga 1152 quarti d'ora, che corrispondono a 96x12 elementi, i primi corrispondono ai quarti d'ora della curva di carico rappresentativa di un mese, moltiplicato per 12, ovvero i mesi dell'anno. Vi sarà un solo output contenente le

suddivisioni nei cluster delle curve di carico rappresentative annuale; quindi, quali forniture hanno tra di loro una certa similarità per tutto l'anno.

### 6.3.1. Cluster delle curve di carico rappresentative mensili

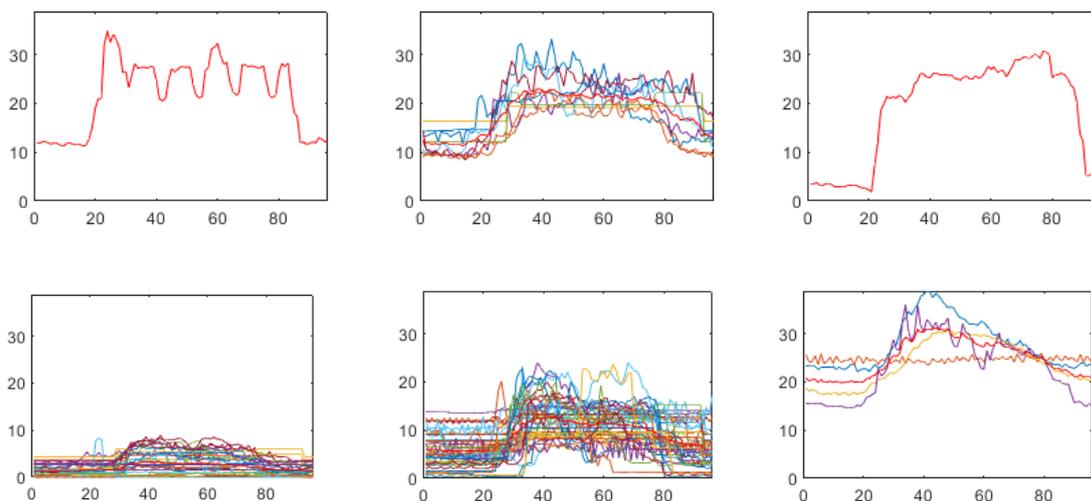
Di seguito vediamo alcuni output dei cluster per la prima suddivisione. Molteplici sono gli esempi riportati, poiché così si potranno vedere le principali differenze tra le forniture delle diverse macro- classi e come cambiano le curve e i raggruppamenti nei diversi mesi dell'anno.

Ogni grafico visionato successivamente avrà sull'asse delle ordinate l'energia al quarto d'ora e sull'asse delle ascisse il tempo scandito in 96 quarti d'ora.

#### Confronto tra stessa macro-classe

Vediamo cosa succede confrontando i risultati dell' algoritmo gerarchico ottenuti dalla stessa macro-classe, quindi identiche forniture, ma per mesi differenti.

Cluster delle forniture della classe ORT9 del mese gennaio.



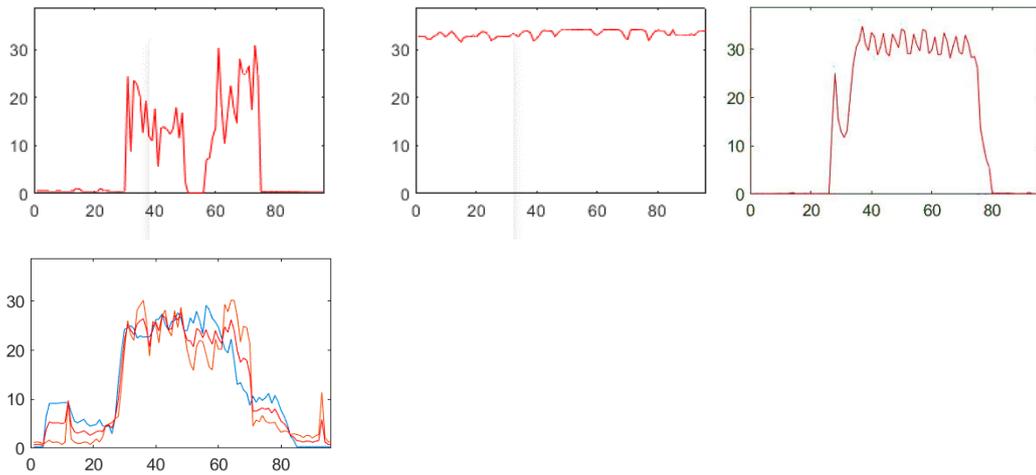


Figura 6.1: insieme di grafici che rappresentano le suddivisioni dei dieci cluster della macro-classe ORT9 del mese di gennaio

Cluster delle forniture della classe ORT9 del mese di giugno.

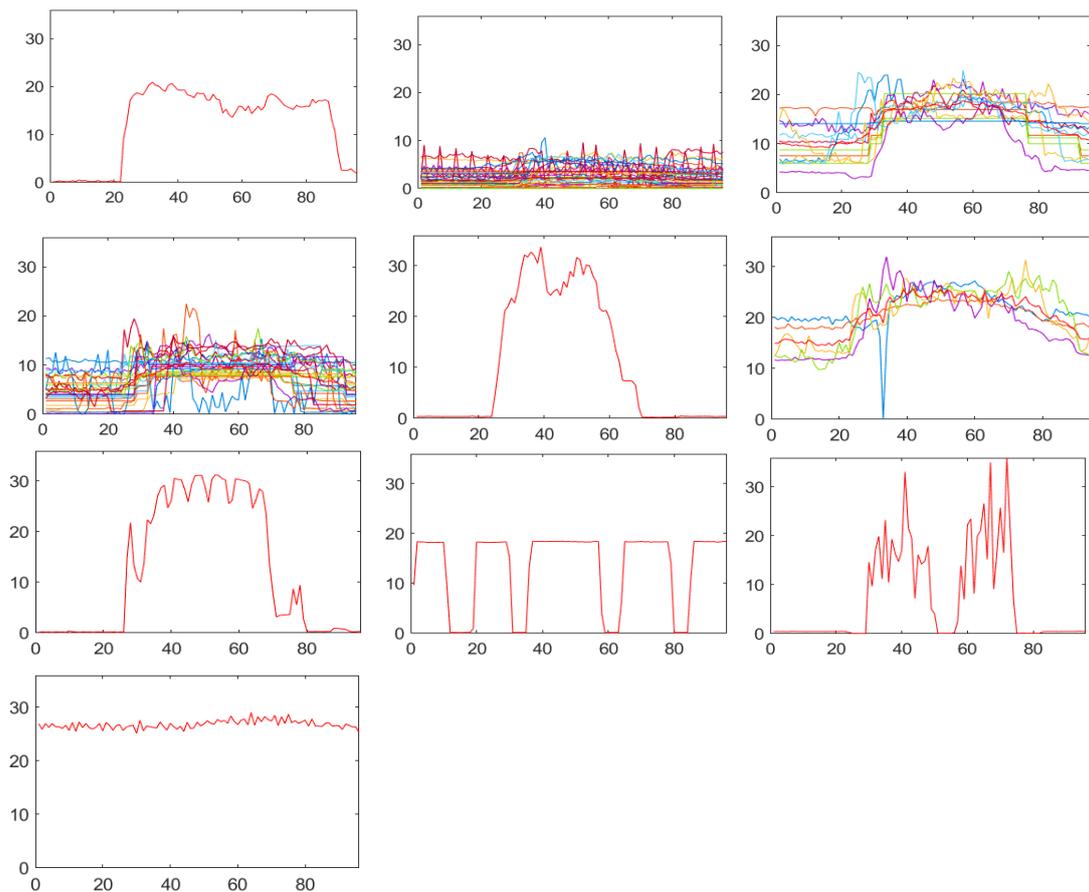


Figura 6.2: insieme di grafici che rappresentano le suddivisioni dei dieci cluster della macro-classe ORT9 del mese di giugno

Come si può vedere confrontando la Figura 6.1 e la Figura 6.2, quindi prendendo due mesi casuali dell'anno della stessa macro-classe, le diverse tipologie di cluster creati nei due mesi hanno caratteristiche molto diverse anche se le forniture sono identiche. Questo risultato è visibile studiando gli outlier e gli aggregati di carichi nei diversi cluster dei due diversi mesi. Infatti, per il mese di gennaio possiamo notare che il cluster gerarchico ha creato ben cinque micro-classi con un aggregato di carichi che hanno andamenti e grandezze molto simili. Lo stesso ragionamento viene fatto per il mese di giugno solo che in questo mese il cluster gerarchico ha creato un aggregato di carichi solo quattro volte.

### Confronto tra diverse macro-classi

Vediamo cosa succede confrontando i risultati dell'algoritmo gerarchico ottenuti da diverse macro-classi, quindi diverse forniture, ma nello stesso mese.

Cluster delle forniture della macro-classe SAT6 del mese di febbraio

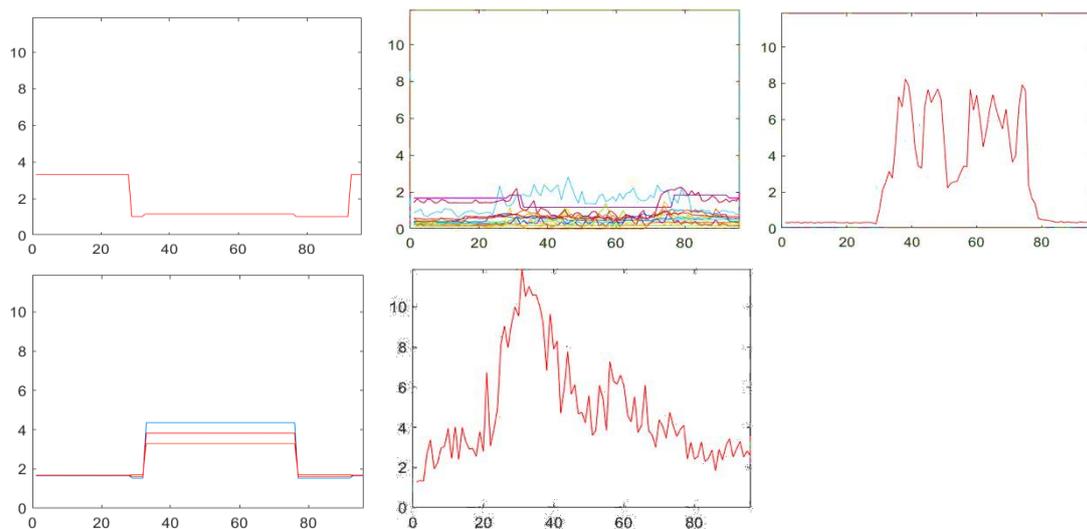


Figura 6.3: insieme di grafici che rappresentano le suddivisioni dei cinque cluster della macro-classe SAT6 del mese di febbraio.

## Cluster delle forniture della classe SAT9 del mese di febbraio

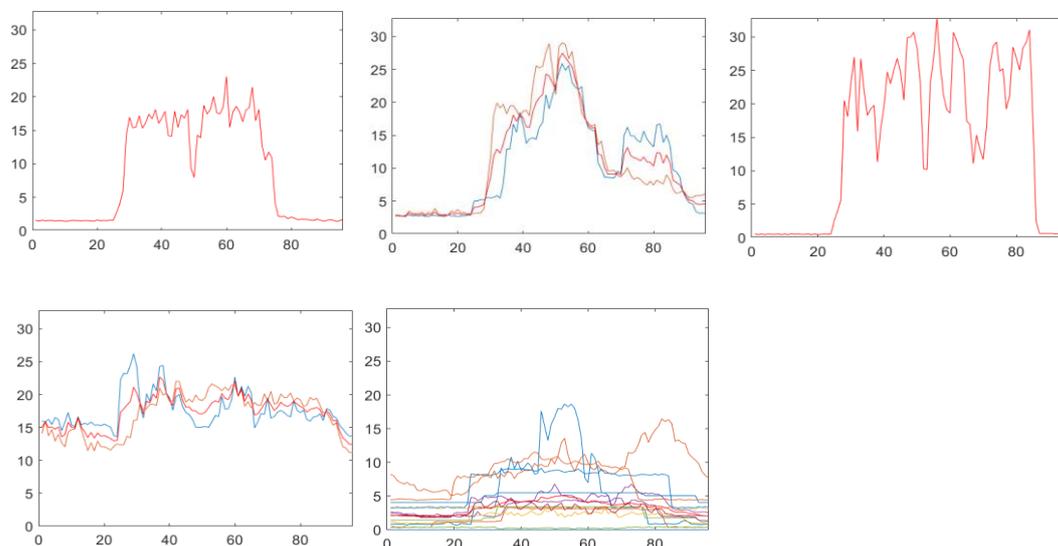


Figura 6.4: insieme di grafici che rappresentano le suddivisioni dei cinque cluster della macro-classe SAT10 del mese di febbraio

Come si può osservare dal confronto delle figure 6.3 e 6.4, la reale differenza sta nell'ordine di grandezza dell'energia. Difatti le differenti macro-classi, come scritto nel quarto capitolo, sono divise per potenza contrattuale. Questo vuol dire che le curve di carico delle forniture presenti nelle classi più alte, come ad esempio quella della Figura 6.4, avranno un'energia consumata maggiore rispetto alle forniture presenti nelle classi più basse, Figura 6.3.

### Osservazioni

Dall'analisi delle curve di carico mensile possiamo trarre alcune conclusioni importanti. Avendo tutti i cluster di tutti i mesi di tutte le macro-classi è possibile avere una miglior comprensione delle caratteristiche mensili delle utenze elettriche e del loro sviluppo mese per mese. Capire le principali caratteristiche e le principali evoluzioni dei carichi mensilmente potrà portare a ulteriori studi su diversi campi.

Per l'ente distributore sarà importante capire l'evoluzione dei carichi mese per mese di ogni classe creata al fine di aver un miglior controllo della flessibilità e della sua variazione mensile. Questa analisi inoltre potrà incentivare le autorità competenti al fine

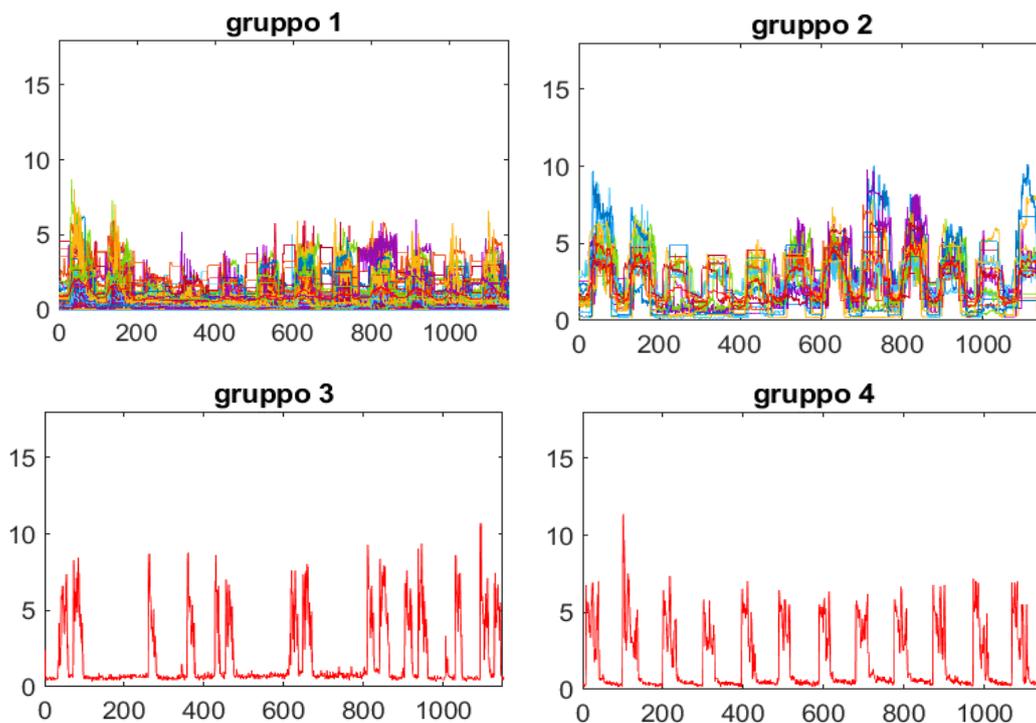
di creare tariffe dedicate per le differenti classi nei diversi mesi, o trimestri, in esame, permettendo così di studiare nuove strategie di business e innovativi servizi dedicati. Per il cliente finale sarà importante conoscere i propri consumi in ogni periodo dell'anno per evitare un eccessivo spreco nei consumi; dunque, potrà imparare a risparmiare dal punto di vista energetico ma anche economico.

### 6.3.2. Cluster delle curve di carico rappresentative annuali

Di seguito sono mostrati alcuni esempi di output del cluster gerarchico per la seconda suddivisione. Le curve di carico rappresentative annuali create saranno divise in diversi cluster una volta sola per ogni macro-classe. Anche in questo caso non si formeranno solo carichi aggregati con caratteristiche simili, ma ci saranno cluster con una sola fornitura, gli outlier.

Ogni grafico visionato successivamente avrà sull'asse delle ordinate l'energia al quarto d'ora e sull'asse delle ascisse il tempo scandito in 1152 quarti d'ora, ovvero 96 elementi della curva di carico rappresentativa di un mese, moltiplicato per 12, ovvero i mesi dell'anno.

Cluster delle forniture della macro-classe ORT6



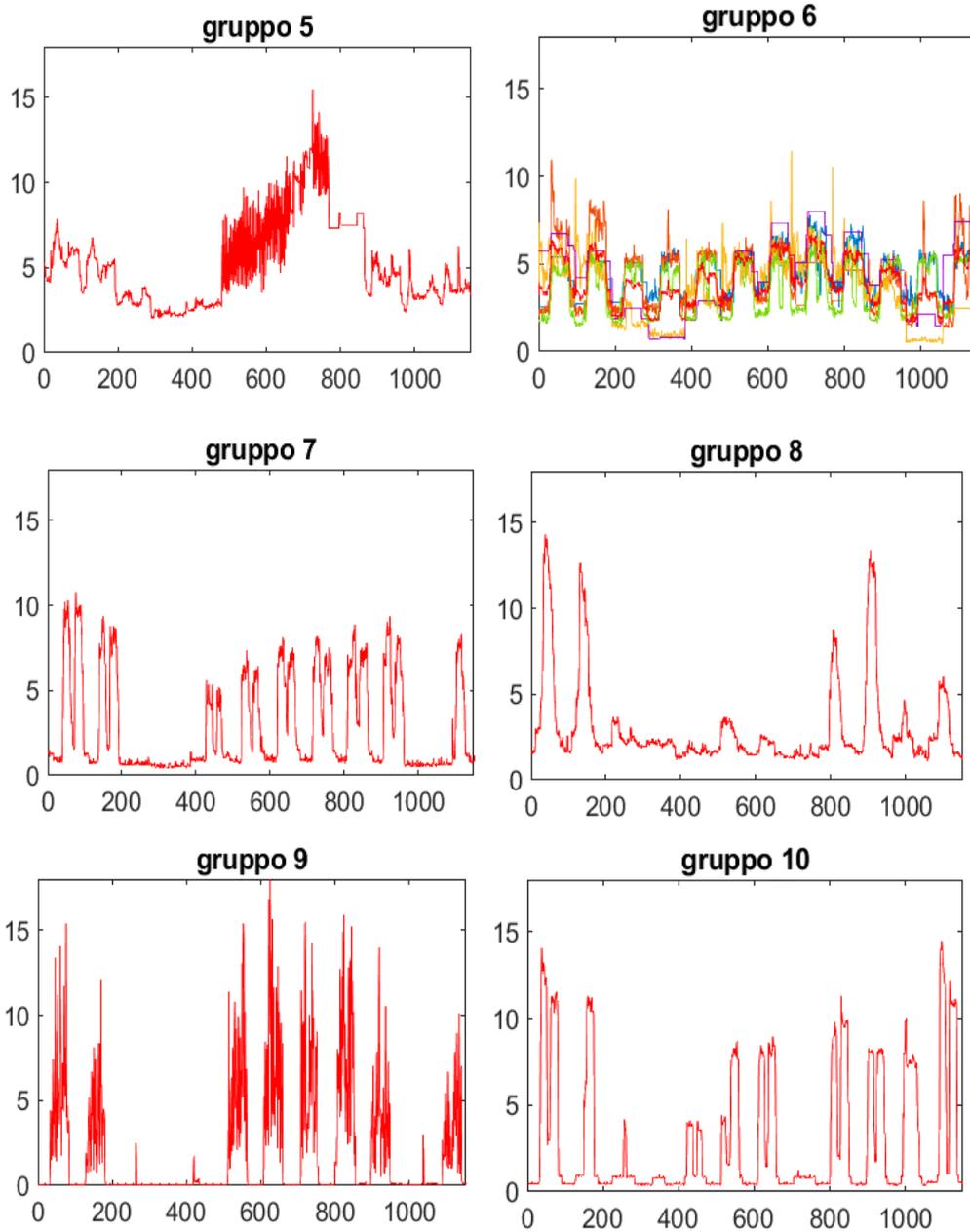
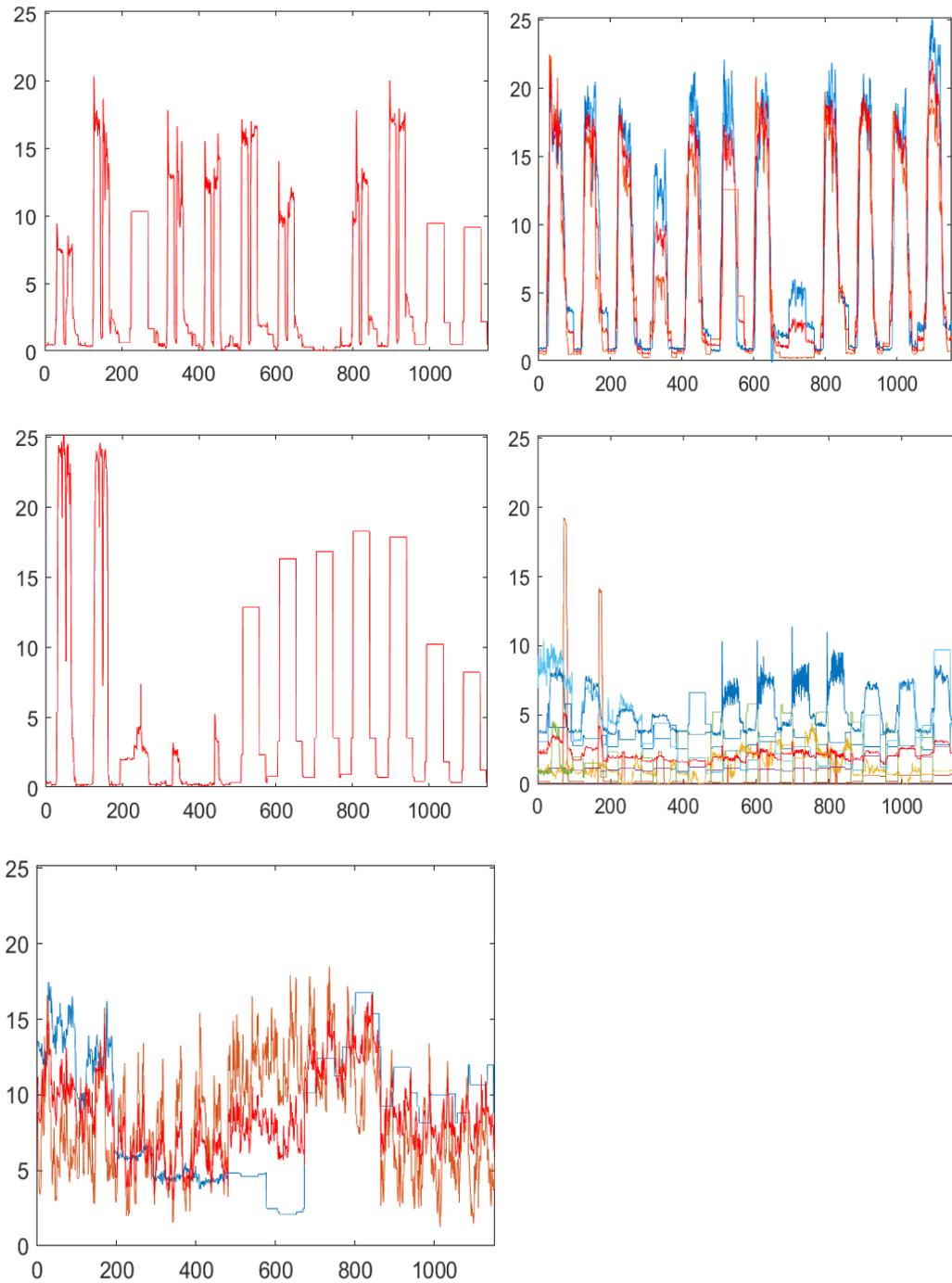


Figura 6.5: insieme di grafici che rappresentano le suddivisioni dei dieci cluster della macro-classe ORT6 con le curve di carico rappresentative annuali

## Cluster delle forniture della macro-classe SAT8



*Figura 6.6: insieme di grafici che rappresentano le suddivisioni dei cinque cluster della macro-classe SAT8 con le curve di carico rappresentative annuali*

Come si può osservare nei due esempi riportati precedentemente, Figura 6.5 e Figura 6.6, per le curve di carico annuali il clustering gerarchico crea più outlier rispetto a quelle dei carichi aggregati, nella Figura 6.5 infatti si può vedere che della macro-classe ORT6

solo 3 gruppi hanno un carico aggregato, mentre i restanti gruppi sono tutti outlier. Ciò è dato sia dalle caratteristiche intrinseche dell'algoritmo di clustering gerarchico e sia perché le curve di carico rappresentative delle forniture possono variare mese per mese, non permettendo una corretta similarità tutto l'anno che consente alla procedura di clustering di unirle sotto un unico gruppo.

### **Osservazioni**

Dall'analisi delle curve di carico annuale possiamo trarre alcune conclusioni importanti. Avendo tutti i cluster di tutte le macro-classi è possibile aver una miglior comprensione delle caratteristiche annuali delle utenze elettriche che unite alla conoscenza del loro sviluppo mese per mese studiate in precedenza, può portare ulteriori analisi in diversi campi al fine di avere una miglior comprensione delle abitudini elettriche degli utenti in generale.

Questo studio inoltre può servire a confrontare le caratteristiche delle curve di carico dei consumatori di anno in anno permettendo ulteriori analisi sui continui sviluppi del comportamento degli utenti finali.

## 7. Analisi dei dati sul consumo: confronto 2019-2020

Oltre ad avere a disposizione i dati relativi alle curve di carico, il rappresentante dell'area metering di Ireti S.p.A. e i suoi collaboratori, hanno gentilmente fornito i dati sul consumo degli utenti richiesti riguardante gli anni 2019 e 2020. Avendo a disposizione questi dati, si è pensato di fare una piccola analisi per vedere come sono cambiate le abitudini dei clienti nel corso dei due anni.

Questo studio è stato effettuato per mettere in evidenza le differenze di consumo durante il periodo del lock-down che ha bloccato molte attività commerciali e industriali tra marzo e maggio del 2020.

### Numero di utenze totali

Il primo grafico, Grafico 7.1, che andiamo a mostrare presenta un confronto tra il numero di utenti totali trifase a uso non residenziale nell'anno 2019 e 2020. Il confronto è fatto mese per mese.

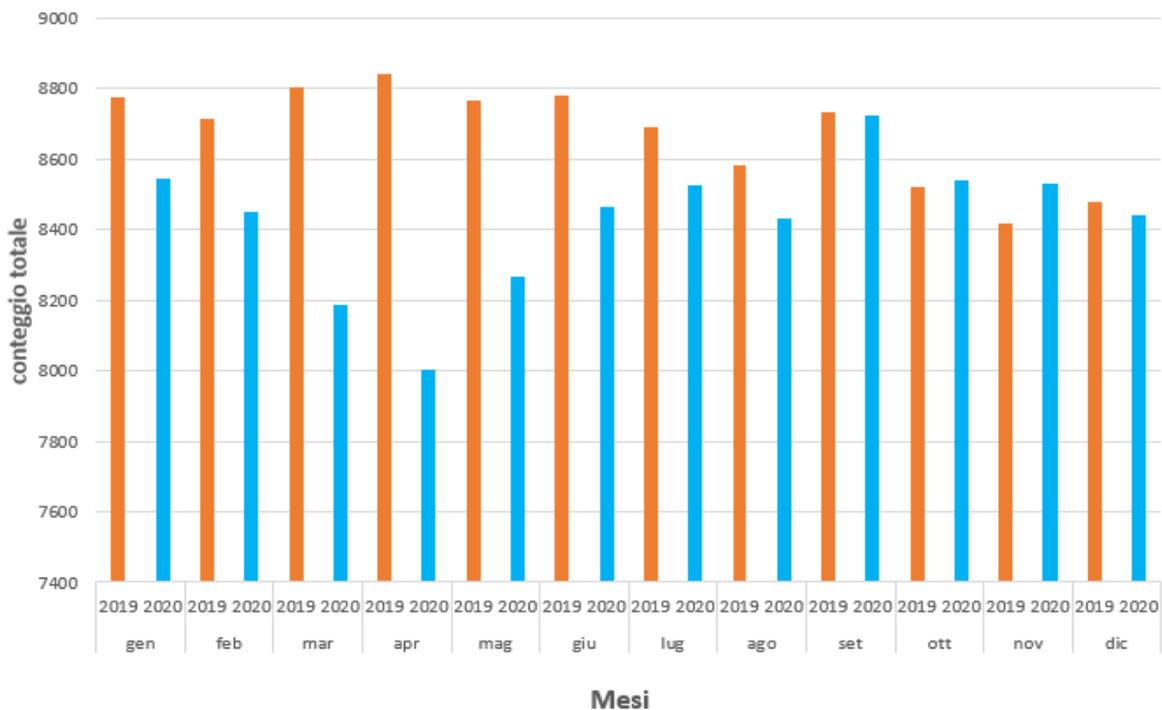


Grafico 7.1: Conteggio totale di utenze trifase per ogni mese dell'anno 2019- 2020

Come si può notare, nei mesi di lockdown, tra marzo e maggio 2020, il numero di utenti totali è diminuito, perdendo addirittura nel mese di aprile il 10% degli utenti rispetto all'anno precedente.

Questo è accaduto perché molte utenze hanno preferito staccarsi dall'ente di distribuzione visto che, in quel periodo, molte attività erano obbligate a rimanere chiuse e inoltre non si sapeva per quanto tempo quella situazione sarebbe continuata. Nei mesi successivi possiamo notare una leggera crescita del numero di utenze, fino ad arrivare a settembre, in cui il numero di utenti del 2020 ha raggiunto quello del 2019.

### Consumo totale degli utenti trifase

Disponendo dell'energia totale consumata da ogni utente per ogni mese dell'anno, 2020 e 2019, si è calcolata la sommatoria del consumo totale di tutti gli utenti per ogni mese, per andare a vedere la riduzione del consumo degli utenti trifase durante il periodo che va da marzo a maggio 2020.

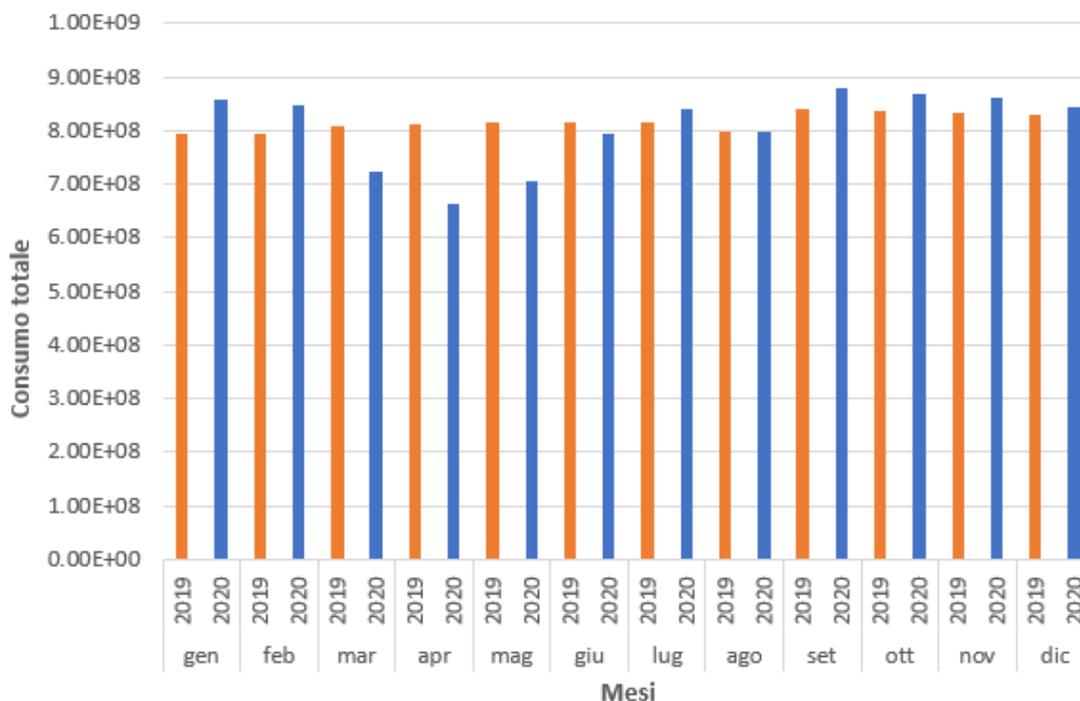


Grafico 7.2: consumo totale degli utenti trifase a uso non residenziale, confronto anno 2019 e 2020. Sull'asse delle ordinate troviamo la sommatoria del consumo di tutte le utenze in un mese (kWh)

Analizzando il Grafico 7.2, possiamo notare come i primi due mesi del 2020 avevano segnato un aumento dei consumi rispetto all'anno precedente mentre da marzo a maggio, come si poteva immaginare, il consumo totale degli utenti si è abbassato arrivando, nel mese di aprile, ad una riduzione di circa il 20% rispetto all'anno precedente. Anche in questo caso, il mese più critico è aprile del 2020 poiché, come si

osserva nel Grafico 7.1, il numero di utenti è diminuito di più in questo mese rispetto agli altri, e in generale la maggior parte delle attività commerciali e industriali sono rimaste chiuse per via del lockdown.

### Media di consumo delle utenze nei diversi mesi

Considerando che molte utenze nel periodo del lockdown sono state scollegate, la riduzione del consumo totale rispetto all'anno precedente è stata una diretta conseguenza. Si è pensato di fare un'analisi sulla media del consumo, così da tenere in considerazione la riduzione delle utenze nel periodo critico che va da marzo a maggio 2020.

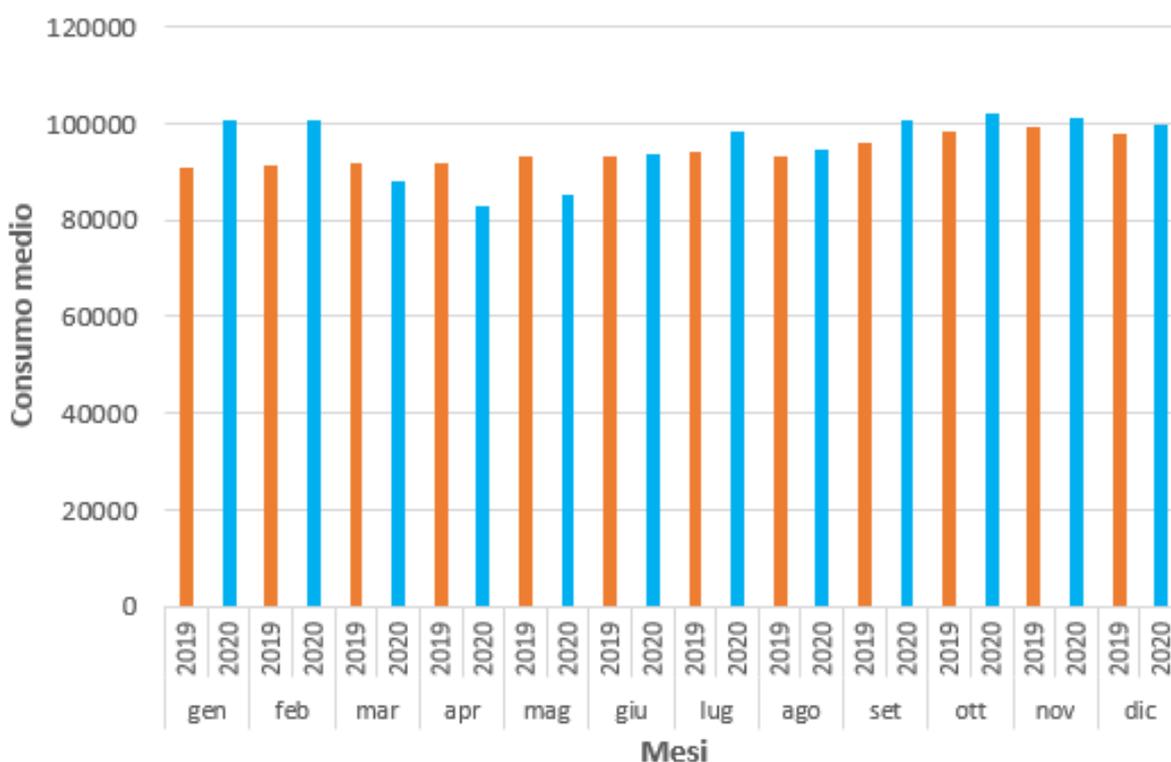


Grafico 7.3: media dei consumi degli utenti trifase a uso non residenziale, confronto anno 2019 e 2020. Sull'asse delle ordinate troviamo la media del consumo di tutte le utenze in un mese (kWh)

Dal Grafico 7.3 possiamo osservare come la media mensile degli utenti trifase durante i mesi del lockdown non sia diminuita di tanto rispetto all'anno precedente; quindi, possiamo concludere dicendo che la causa predominante della diminuzione del consumo durante i mesi critici è avvenuta poiché numerose utenze hanno preferito scollegarsi e rimanere senza alimentazione, risparmiando sul consumo e quindi avendo un risparmio economico. Mentre come si può osservare da quest'ultimo grafico, le attività che sono

rimaste collegate hanno consumato mediamente la stessa energia dell'anno precedente.

## 8. Conclusioni

Lo scopo di questo elaborato è stato quello di approfondire il tema dell'analisi dei dati e apprendere la sua importanza al fine di portare dei miglioramenti sul sistema elettrico generale; infatti, capire l'importanza e l'efficacia del corretto dispiegamento degli *smart meter* di ultima generazione sulla rete elettrica, porta ad un alto quantitativo di dati utili. Su queste informazioni sono state utilizzate differenti tecniche per l'analisi dei dati per ottenere diverse classificazioni dei clienti. I diversi cluster di clienti sono stati creati utilizzando le informazioni relative ai loro consumi come ad esempio la potenza contrattuale, l'energia consumata o l'utilizzazione ma anche dati relativi alle curve di carico quart'orarie.

Oltre a una facile visualizzazione delle caratteristiche principali dei diversi utenti nelle diverse classi, basata su alcune tipologie di dati sul consumo come, ad esempio, le curve di carico, questa analisi porta ulteriori benefici, che comprendono una miglior conoscenza dei profili di carico delle macro-classi di clienti e una miglior conoscenza delle curve di carico per periodi di tempo coerenti. Ciò determina l'abbandono del concetto tradizionale di classificazione dei clienti solo in base alla loro effettiva forma di consumo. Tenere sotto controllo il consumo degli utenti porta vantaggio alla società di distribuzione, poiché tener sotto controllo i consumi dei clienti può permettere la riduzione dei costi relativi all'esecuzione dei controlli antifrode, oltre che ad un miglioramento della flessibilità e dunque ad un'efficienza operativa del sistema. I risultati di questa analisi portano benefici anche al cliente finale, poiché portano ad una consapevolezza specifica sul consumo, permettendo così un risparmio economico oltre ad incentivare l'autorità nazionale nella definizione di tariffe dedicate per le differenti classi di clienti, permettendo di sviluppare nuove strategie di business e innovativi servizi dedicati.

Si augura che tutti i benefici portati da questa analisi, precedentemente elencati, possano essere utilizzati per migliorare l'esercizio della rete elettrica in generale e possano portare in futuro a ulteriori sviluppi di queste procedure di analisi dei dati, permettendo così di rispondere alle necessità di un sistema in continua evoluzione come quello elettrico, includendo nell'analisi ulteriori tipologie di carico quali veicoli elettrici e carichi attivi come la generazione distribuita, permettendo un ammodernamento del sistema elettrico che deve stare al passo con una grande sfida: la transazione energetica.

## 9. SITOGRAFIA

- Ireti: piano di messa in servizio del sistema di smart metering 2G (PMS2)  
<https://www.ireti.it/contatori-2g>
- Arera: scheda smart meter  
[https://arera.it/it/operatori//SmartMeter\\_scheda.htm](https://arera.it/it/operatori//SmartMeter_scheda.htm)
- e-distribuzione: qualità del contatore 2G  
<https://www.e-distribuzione.it/open-meter/la-qualita-del-contatore.html>
- materiale didattico dell'insegnamento "Distribuzione e utilizzazione dell'energia elettrica" (prof. Gianfranco Chicco, anno accademico 2020-2021)
- materiale didattico dell'insegnamento "Smart electricity systems" (prof. Gianfranco Chicco, anno accademico 2020-2021)
- Dataskills: tecniche di clustering  
<https://www.dataskills.it/tecniche-di-clustering/>

## 10. BIBLIOGRAFIA

- E.Bompard, E.Carpaneto, G.Chicco, R.Napoli, F.Piglione, P.Postolache, M.Scutariu, ***“Stratified sampling of the electricity customers for setting up a load profile survey”***, 2000.
- G. Chicco, D. Labate, A. Notaristefano, F. Piglione, ***“Unveil the Shape: Data Analytics for Extracting Knowledge from Smart Meters”***, Energia Elettrica Supplement Journal AEIT, numero 6 Vol. 96, pag 1-15, 2019.
- G. Chicco, D. Labate, P. Giubbini, M. Ettore, Exeura ***“Shape: a new business analitics web platform for getting insights on electrical load patterns”*** CIRED Workshop, paper 0354, pag. 1-5, 2014.

## RINGRAZIAMENTI

Mi è doveroso dedicare questo spazio del mio elaborato alle persone che hanno contribuito, con il loro supporto, alla realizzazione dello stesso.

In primis, un ringraziamento speciale al mio relatore Gianfranco Chicco per la sua immensa pazienza, per i suoi indispensabili consigli, per le conoscenze trasmesse durante tutto il percorso di stesura dell'elaborato.

Un ringraziamento particolare va al responsabile dell'area metering di Ireti, l'ingegner Renato Mulinacci, e a tutti i suoi collaboratori, i quali mi hanno seguito, con un'infinita disponibilità, in ogni step della realizzazione dell'elaborato.

Ringrazio infinitamente mia madre e mio fratello che mi hanno sempre sostenuto, appoggiando ogni mia decisione, fin dalla scelta del mio percorso di studi.

Infine, dedico questa tesi a me stesso, ai miei sacrifici e alla mia tenacia che mi hanno permesso di arrivare fin qui.