



**Politecnico  
di Torino**

## **Politecnico di Torino**

Corso di Laurea Magistrale in Ingegneria Gestionale  
A.a. 2022/2023  
Sessione di Laurea Ottobre 2023

# **Forecasting dei consumi di un magazzino refrigerato per l'inserimento in una CER**

Relatori:  
Prof. Giulio MANGANO  
Prof. Paolo LAZZERONI

Candidato:  
Mattia COLTRO



# Sommario

La questione energetica e soprattutto la tematica della transizione energetica si sta affermando come una dei più importanti in tutto il mondo. A sostegno di ciò le politiche nazionali e internazionali stanno promuovendo piani volti a raggiungere obiettivi di riduzione delle emissioni, di indipendenza energetica e di condivisione dell'energia. In questo elaborato viene esposto il percorso normativo europeo che ha portato fino alle direttive RED II e EMD II, e conseguentemente al loro recepimento nell'ordinamento italiano. Grazie a tale normativa lo strumento che viene promosso è quello delle Comunità Energetiche e in particolar modo quelle configurazioni che inglobano le Fonti di Energia Rinnovabile. Il presente elaborato ha l'obiettivo di andare ad esaminare i consumi elettrici del gruppo frigo afferente ad una cella frigo di un magazzino della filiera alimentare per poi effettuarne la previsione tramite un modello di forecasting. Tale modello come caratteristica peculiare tenta di considerare, per effettuare la previsione di consumo, feature di carattere gestionale operativo. La finalità futura è quella di usufruire di queste informazioni di previsione sul consumo nell'intento di inserire la struttura del magazzino analizzata in una configurazione CER. La tesi espone la problematica ambientale sotto la lente del settore della Food & Cold Supply Chain. In tale sezione viene fornita una panoramica del settore per meglio comprendere le ricadute dei consumi elettrici del un magazzino refrigerato oggetto di studio. Evidenziati gli effetti ambientali ed economici della filiera del Food&Berverage e della catena del freddo, vengono presentate le principali categorie e tecniche di Machine Learning per introdurre l'argomento, dato che il modello previsionale si avvale di semplici tecniche di Machine Learning. Nella parte finale dell'elaborato viene eseguita l'analisi della struttura e dei dati raccolti per la costruzione del modello. A seguito dell'analisi di regressione i dati raccolti non presentano sufficiente significatività per la costruzione del modello di previsione dei consumi della cella frigo. Tale previsione successivamente è stata effettuata con un modello di forecasting ARIMA. Il caso studio punta ad evidenziare le criticità del modello previsionale che vuole considerare le variabili gestionali operative, per poi poter integrare tali informazioni nell'inserimento della struttura analizzata in una configurazione CER.

# Ringraziamenti

Presentando questo elaborato come conclusione del mio percorso universitario ci tengo a ringraziare alcune persone in particolare che mi hanno accompagnato e sostenuto in questo cammino.

Ringrazio innanzitutto i miei genitori e mio fratello che sono sempre stati presenti nei momenti di gioia dopo gli esami superati, ma ancor più presenti nei momenti di faticosa preparazione sempre contraddistinti dalla contemporaneità dei dispendiosi impegni sportivi.

A questo riguardo una menzione va fatta anche a tutti i compagni di squadra che ho incontrato lungo il percorso e soprattutto a quelli che oggi posso chiamare amici.

Un ringraziamento particolare è dovuto ai miei Abici (Davide, Giuditta, Rabe e in special modo Federica e Elena) che spesso sono riusciti a tenermi a galla nei momenti di sconforto e difficoltà incontrati in questi anni, mi hanno sempre ricordato l'importanza di condividere pensieri, paure, gioie e soprattutto di saper festeggiare anche quando il mio pensiero era focalizzato solamente sugli impegni sportivi e universitari.

Il mio grazie più grande infine va a Bianca che, oltre a supportarmi e sopportarmi, è stata sempre in grado di tenermi coi piedi per terra, concedendomi un equilibrio emotivo per me fondamentale in tutto questo periodo.

*"Never too high, never too low"*





# Indice

<b>Indice delle Figure</b>	VI
<b>1 Comunità Energetiche</b>	1
1.1 Quadro europeo . . . . .	2
1.1.1 Normativa europea: direttive RED II e EMD II . . . . .	6
1.1.2 Normativa italiana . . . . .	8
1.2 Fonti di energia rinnovabile e fotovoltaico . . . . .	13
1.3 Vantaggi CE: benefici economici, sociali, ambientali . . . . .	16
1.4 Integrazione del forecasting nelle CER . . . . .	24
<b>2 Food Supply Chain &amp; Cold Supply Chain</b>	30
2.1 Spreco di cibo e di energia . . . . .	31
2.2 Normativa settore alimentare . . . . .	38
<b>3 Intelligenza Artificiale</b>	41
3.1 Introduzione AI e al Machine Learning . . . . .	41
3.2 Serie temporali . . . . .	49
3.2.1 Modelli autoregressivi . . . . .	52
3.3 Python tools . . . . .	54
<b>4 Caso studio</b>	59
4.1 Magazzino logistico refrigerato . . . . .	59
4.1.1 Gruppo frigo . . . . .	65
4.1.2 Misurazioni e dati . . . . .	70
4.2 Modello di forecasting . . . . .	77
4.2.1 Modello ARIMA . . . . .	85
<b>5 Conclusioni</b>	98
<b>Bibliografia</b>	103

# Indice delle Figure

1.1	Andamento delle emissioni di gas a effetto serra, UE-28, 1990-2017 <sup>1</sup>	2
1.2	Quota dei consumi finali lordi di energia coperta da FER [12]	14
1.3	Evoluzione della potenza e della numerosità degli impianti Fotovoltaici [13]	16
1.4	Andamento cumulato della potenza per AC/CER (al 31/12/22) [kW]; Distribuzione degli impianti fotovoltaici in AC/CER per taglia dell'impianto	16
1.5	Composizione del prezzo dell'energia per consumatore domestico tipo	17
1.6	Ricadute energetiche, ambientali e sociali con lo sviluppo delle Comunità energetiche in Italia <sup>2</sup>	20
2.1	Cibo perso nel mondo per mancanza di adeguata refrigerazione (2017)	32
2.2	Rappresentazione schematica di una Cold Chain [45]	35
3.1	Schematica rappresentazione dei metodi AI-based	42
3.2	Classificazione dei metodi di Machine Learning[72]	44
3.3	Rappresentazione grafica dei metodi di Classificazione, Regressione e Clustering <sup>3</sup>	45
3.4	Schema Rete Neurale semplificato <sup>4</sup>	46
3.5	Schematizzazione del funzionamento di un singolo neurone <sup>5</sup>	46
3.6	Schema di un algoritmo di Deep Learning che utilizza una Convolutional neural network <sup>6</sup>	47
3.7	Rappresentazione delle componenti di una serie temporale <sup>7</sup>	51
4.1	Planimetria del magazzino (scala 1:200): piano terra in alto, primo piano in basso	61
4.2	Planimetria del piano terra (scala 1:200)	62
4.3	Planimetria delle celle frigo a anticella (scala 1:200)	62
4.4	Esempio di scaffalatura portapallet e UPC	63
4.5	Dimensioni e struttura magazzino automatico	63
4.6	Condotti e scambiatori di condizionamento nel magazzino 2	65

4.7	Ciclo termodinamico del fluido refrigerante, Schema di un sistema di raffreddamento basato su un chiller a compressione e una torre di raffreddamento . . . . .	69
4.8	Dispositivo di misurazione con i collegamenti a tensione e corrente .	71
4.9	Schema collegamenti dispositivo di misurazione . . . . .	72
4.10	Consumi elettrici e temperature esterne, andamento sovrapposto . .	79
4.11	Consumo elettrico medio giornaliero con evidenza dei giorni non lavorativi . . . . .	80
4.12	Correlogramma con indicato l'indice di correlazione di Pearson . . .	82
4.13	Valori della prima regressione con tutte le variabili considerate . . .	84
4.14	Analisi dei residui sulla regressione con tutte le variabili . . . . .	84
4.15	Valori della regressione "pulita" . . . . .	85
4.16	Analisi dei residui sulla regressione "pulita" . . . . .	85
4.17	Parte di dati stazionaria utilizzata per creare il modello, con separazione train-test set . . . . .	87
4.18	Grafico di Autocorrelazione dei dati di training . . . . .	89
4.19	Grafico di Autocorrelazione Parziale dei dati di training . . . . .	89
4.20	Stepwise search per trovare il miglior ordine del modello ARMA . .	90
4.21	Scomposizione della serie temporale . . . . .	91
4.22	Grafico ACF componente stagionale, Grafico PACF componente stagionale . . . . .	92
4.23	Misure di performance dei modelli analizzati . . . . .	95
4.24	Rappresentazione dei valori del modello ARIMAX(4,0,3) per confronto training-fitting e test-forecast . . . . .	96



# Capitolo 1

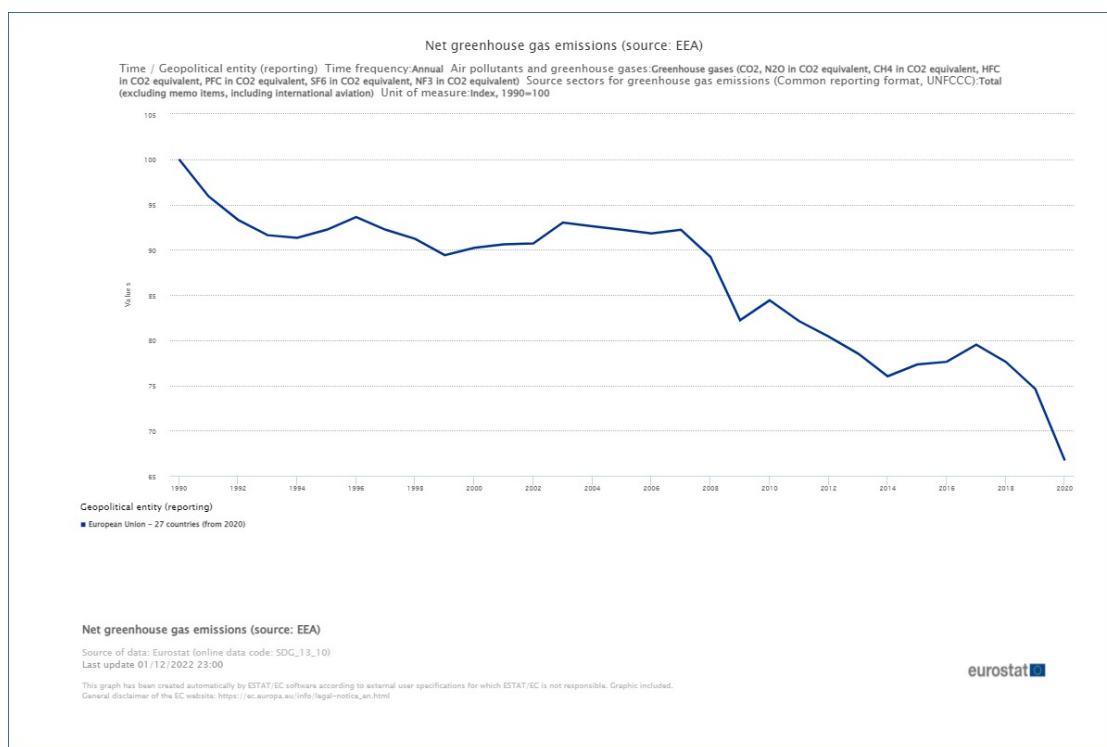
## Comunità Energetiche

Uno degli argomenti più caldi dell'attualità è sicuramente quello riguardante il cambiamento climatico e tutto ciò che ne consegue. Ovviamente il discorso non è circoscritto all'attualità ma ricade su tutti gli aspetti che ci circondano: ambientale, sociale, economico, scientifico, politico. La situazione odierna che ci si presenta riguardo la materia energetica è strettamente legata alle numerose iniziative che le istituzioni stanno portando avanti per creare un mondo più sostenibile a livello ambientale e sociale. Il processo di transizione ecologica ed energetica punta ad obiettivi come la riduzione di gas serra, l'indipendenza energetica, la condivisione dell'energia e tutti i benefici sociali ed economici derivanti. Uno degli strumenti che è stato individuato per realizzare gli ambiziosi piani energetici è la diffusione delle comunità energetiche. Tale configurazione, soprattutto nella forma di Comunità Energetica Rinnovabile (CER) è indicata come uno dei mezzi per aiutare a democratizzare, decarbonizzare e decentralizzare il settore energetico in tutta Europa. L'uso razionale dell'energia, la diffusione combinata di efficienza energetica negli usi finali, le fonti rinnovabili ed i prodotti e servizi a basso impatto energetico e ambientale dovranno essere i protagonisti della transizione che è in atto. Bisognerebbe pertanto ripensare all'uso dell'energia in modo olistico. Quando si affrontano i concetti di transizione energetica e di economia moderna efficiente sotto il profilo delle risorse non si può ignorare una visione di insieme, essa è difatti indispensabile per sfruttare il potenziale dei benefici non energetici connessi all'uso razionale dell'energia. Fra i benefici non energetici si possono considerare i miglioramenti della qualità della vita, la sicurezza, l'incremento di valore degli asset, gli impatti positivi sulla produttività e sull'occupazione. La trasformazione dei sistemi economici in sistemi "sostenibili" in senso ambientale impone tuttavia una profonda riflessione sulle modalità di sviluppo dei modelli scelti per attuare i progetti. L'ingresso di nuovi operatori, la liberalizzazione dei prezzi e l'affermazione delle nuove tecnologie FER hanno infatti gradualmente modificato il principio che l'economia di scala dovesse essere il driver dominante degli investimenti nella

generazione elettrica e tali impianti di più piccola taglia o specializzati per operare in determinati segmenti di mercato, la taglia decisamente contenuta degli impianti FER hanno dato equilibrio economico a mix produttivi più articolati dando una spinta verso la generazione distribuita. È esattamente in questo contesto che si inseriscono le comunità energetiche, le quali grazie al continuo miglioramento tecnologico danno la possibilità di sfruttare la condivisione dell'energia e tutti i vantaggi da essa derivanti.

## 1.1 Quadro europeo

Da lungo tempo le istituzioni hanno messo in campo numerosi piani, e nell'attuazione degli stessi alcuni obiettivi sono già stati raggiunti come la riduzione delle emissioni di CO<sub>2</sub> entro il 2020, e l'andamento sembra dare ragione alle strategie escogitate (Figura 1.1). L'Unione Europea (UE) porta avanti da tempo una



**Figura 1.1:** Andamento delle emissioni di gas a effetto serra, UE-28, 1990-2017 <sup>1</sup>

<sup>1</sup>Indice 1990=100, da dati Eurostat giugno 2019

politica che verte verso la decarbonizzazione e la transizione ecologica ed energetica. Fin dal 1995 è stato intrapreso un percorso dove le istituzioni comunitarie hanno posto l'accento sull'aver una politica energetica comune a tutti gli Stati membri, definendo obiettivi a lungo termine in materia energetica e sulla necessità di armonizzarli alle finalità generali della politica economica, improntata sull'integrazione, liberalizzazione e deregolamentazione dei mercati [1]. Nel tempo si sono progressivamente intensificati gli sforzi per combattere il cambiamento climatico diventando argomento sempre più centrale all'interno dell'agenda europea. L'importanza di mobilitarsi ha spinto l'Unione Europea ad acquisire un ruolo sempre più da leader nel contesto mondiale verso la lotta al cambiamento climatico. Nel 2007 viene formalizzato il primo documento programmatico denominato "Pacchetto Clima-Energia - Obiettivo: 20-20-20", che varava una serie di leggi atte a garantire il raggiungimento entro il 2020 di tre grossi obiettivi:

- diminuzione del 20% delle emissioni di gas serra rispetto ai livelli del 1990;
- 20% del fabbisogno energetico ricavato da fonti di energia rinnovabile;
- miglioramento del 20% dell'efficienza energetica .

Nel 2015 con l'accordo di Parigi l'UE si impegna a compiere ulteriori passi in avanti fissando obiettivi ancora più ambiziosi tramite il "Quadro Clima-Energia 2030" spingendosi verso una riduzione delle emissioni di CO<sub>2</sub> al 40%, con quota di energie rinnovabili ad almeno il 32% e miglioramento dell'efficienza energetica al 32,5%. La programmazione europea si è spinta ancora più in là del 2030, fino ad arrivare all'obiettivo cardine ovvero la neutralità carbonica entro il 2050. Per il raggiungimento degli obiettivi preposti per il 2030 e il 2050 l'UE ha strutturato un solido quadro normativo. Il "Clean Energy for all Europeans Package" (CEP), presentato dalla Commissione Europea nel 2016, comprende diverse misure legislative in cinque fondamentali settori della politica energetica europea:

- Sicurezza nell'approvvigionamento energetico;
- Mercato interno dell'energia;
- Efficienza energetica;
- Decarbonizzazione dell'economia;
- Ricerca, innovazione e competitività.

Tramite un ulteriore documento il "Green Deal", presentato l'11 dicembre 2019 dalla Commissione Europea, si definiscono le strategie di crescita mirate alla transizione che porterà l'UE verso una società giusta e prospera, dotata di un'economia moderna, efficiente sotto il profilo delle risorse e competitiva, che nel 2050 non



genererà emissioni nette di gas a effetto serra e in cui la crescita economica sarà dissociata dall'uso delle risorse. Appena un anno dopo, nel dicembre 2020, è stato approvato il pacchetto "Fit For 55%" un nuovo obiettivo vincolante, di riduzione interna netta delle emissioni di gas a effetto serra pari ad almeno il 55% (rispetto ai livelli del 1990) entro il 2030, un incremento considerevolmente ambizioso rispetto all'obiettivo concordato nel 2014.

Le misure portate avanti con le varie normative hanno il proposito di andare ad impattare non solo le grandi imprese, anzi sono rivolte maggiormente verso i singoli cittadini e piccole comunità, e alla loro consapevolezza energetica. Migliorando l'efficienza energetica e proponendo nuove configurazioni anche rispetto alla struttura del mercato viene esercitato un rafforzamento dei diritti dei consumatori, portando i cittadini ad avere una reale influenza sulla propria impronta energetica attraverso contatori intelligenti, con il controllo delle bollette domestiche, investendo direttamente nella produzione autonoma di energia rinnovabile, con possibilità di condivisione della stessa e di immetterla in rete. Nonostante l'UE vanti una ricerca di livello mondiale e industrie forti, nonostante un terzo di tutte le pubblicazioni scientifiche mondiali abbiano padri europeisti, c'è ancora difficoltà nel trasformare la leadership nella scienza in leadership nell'imprenditoria in questo ambito. È per questo che l'UE emana periodicamente direttive, finanziamenti e documenti riguardo alla pianificazione sulle politiche e strategie energetiche.

Nel CEP rientrano come disposizioni di maggiore rilevanza la "Direttiva sulla promozione dell'uso di energia da fonti rinnovabili" (RED II) e la "Direttiva sulle regole comuni per il mercato interno dell'elettricità" (EMD II). Come per qualsiasi direttiva europea, anche in questo caso è demandato il compito ai singoli governi nazionali degli Stati membri di tradurle in legge, in modo da omologare l'ordinamento di riferimento e allinearle con gli sviluppi comunitari.

È proprio nella RED II che viene finalmente precisato il ruolo e definita l'entità di Comunità Energetica Rinnovabile con la delineazione delle finalità a cui punta questo tipo di configurazione. Viene sottolineato che la partecipazione alle Comunità Energetiche Rinnovabili deve essere aperta a tutti i consumatori, compresi quelli appartenenti a famiglie a basso reddito o vulnerabili ponendo l'accento sul contrasto alla povertà energetica. Come specificato nella definizione, la dicitura è quella di CER e quindi al centro del discorso è posizionato tutto quel campo che comprende la tecnologia di generazione dell'energia rinnovabile. Non è così stringente invece il concetto di CEC (Comunità Energetica di Cittadini), esposto con la Direttiva 2019/944 (EMD II), considerato un soggetto giuridico quasi alla stregua delle CER, dove i membri partecipano alla generazione (da fonti rinnovabili ma anche da altre tecnologie), alla distribuzione, alla fornitura, al consumo, all'aggregazione, allo stoccaggio dell'energia, ai servizi di efficienza energetica, o a servizi di ricarica per veicoli elettrici o fornire altri servizi energetici ai suoi soci.

Proseguendo nel percorso delineato dalle direttive europee è stato collocato

un altro tassello fondamentale per il raggiungimento degli ambiziosi obiettivi in materia energetica. Il Mise ha pubblicato il 21 gennaio 2020 il Piano Nazionale Integrato per l'Energia e il Clima (PNIEC), testo col quale vengono recepite le novità contenute nel Decreto Legge sul Clima e quelle sugli investimenti per il Green New Deal previste nella Legge di Bilancio 2020. Il PNIEC ha come colonne portanti del processo di transizione verso la decarbonizzazione:

- la graduale eliminazione del carbone nella produzione di energia elettrica fino ad una sua totale interruzione nel 2025;
- la diffusione e l'integrazione delle energie rinnovabili (incentivando le comunità energetiche);
- la riduzione al minimo degli impatti ambientali (con priorità assegnata ad esempio agli impianti fotovoltaici sugli edifici per contenere l'uso del suolo, pompe di calore per evitare emissioni di particolato, biocarburanti avanzati per l'uso di residui e rifiuti, la rivalorizzazione delle infrastrutture, ecc.).

Il Piano si struttura in cinque linee d'intervento, che si svilupperanno in maniera integrata: dalla decarbonizzazione all'efficienza e sicurezza energetica, passando attraverso lo sviluppo del mercato interno dell'energia, della ricerca, dell'innovazione e della competitività. In linea con gli obiettivi europei si possono evidenziare alcuni dei numeri oggetto del piano (con orizzonte temporale 2021-2030): riduzione dei gas serra del 33%, per l'efficienza energetica si stima una riduzione dei consumi di energia primaria del 43%, quota di energia da fonti rinnovabili nei consumi finali lordi di energia pari al 30%. I vari settori industriali contribuiranno in modo diverso al conseguimento di quest'ultimo obiettivo, nel segmento elettrico si raggiungerà una quota prevista di fonti rinnovabili del 55,4%, nei trasporti il 21,6% e nel termico il 33% [2].

A seguito della situazione pandemica (Covid-19) in cui si è ritrovato tutto il mondo, l'Unione Europea ha deciso di combattere la crisi e affrontare la ripartenza con un progetto chiamato Next Generation EU, che prevede l'aiuto degli Stati membri tramite vari tipi di investimenti. Questo strumento temporaneo di ripresa e rilancio economico europeo ha l'obiettivo principale di mitigare l'impatto economico e sociale generato dalla crisi sanitaria. Tale progetto inserisce all'interno del bilancio europeo 2021-2027 una somma pari a 806,9 miliardi di euro (di cui 723,8 mld disposti per la ripresa e resilienza, ripartiti tra 385,8 mld di prestiti e 338 mld di sovvenzioni) messi a disposizione degli Stati membri che hanno presentato dei piani nazionali di resistenza e resilienza nei quali viene illustrato l'impiego di tali risorse.

Per l'Italia, il Piano Nazionale di Ripresa e Resilienza (PNRR) è il documento apposito predisposto dal governo per illustrare alla Commissione Europea come il nostro Paese intende investire i fondi che verranno forniti nell'ambito del programma Next generation Eu; la richiesta ammonta a 122,6 miliardi di euro (quota più

consistente tra i Paesi richiedenti). Col Decreto Legge del 31 maggio 2021 il Pnrr delinea anche le riforme che il governo intende adottare per modernizzare il Paese, riforme che costituiscono una *conditio sine qua non* per ottenere i finanziamenti. Tali riforme si distinguono in 4 tipologie: orizzontali o di contesto (misure d'interesse generale), abilitanti (interventi funzionali a garantire l'attuazione del piano), settoriali (riferite a singole missioni o comunque ad ambiti specifici), concorrenti (non strettamente collegate con l'attuazione del piano ma comunque necessarie per la modernizzazione del Paese es. riforma sistema fiscale e ammortizzatori sociali). Seguendo le linee guida emanate dalla Commissione Europea (tre assi principali: digitalizzazione e innovazione, transizione ecologica e inclusione sociale) il PNRR raggruppa i progetti di investimento in 16 componenti, a loro volta raggruppate in 6 missioni:

1. Digitalizzazione, innovazione, competitività, cultura e turismo;
2. Rivoluzione verde e transizione ecologica;
3. Infrastrutture per una mobilità sostenibile;
4. Istruzione e ricerca;
5. Coesione e inclusione;
6. Salute.

La Missione 2 è quella che interessa l'argomento di questa tesi, ed è coperta da 71,81 miliardi (nonostante l'Italia abbia assegnato poco più del minimo richiesto dal regolamento europeo 37% alla transizione ecologica), e per la messa in campo della stessa sono stati stanziati a fondo perduto 2,2 miliardi (concessi dalla UE) solamente per la promozione delle CER. L'atto attuativo ad inizio 2023 permetterà di generare 2.000 MW di nuova capacità di generazione elettrica in configurazione distribuita da parte di comunità di energie rinnovabili e autoconsumatori. L'obiettivo è puntare soprattutto sul fotovoltaico: supponendo 1.250 kWh medi di generazione annua di ogni impianto, le comunità energetiche di nuova creazione possono permettere di risparmiare 2.500 GWh di energia, ovvero 1,5 milioni di tonnellate di emissioni di anidride carbonica. Tutto questo mettendo in campo la produzione e la generazione nei territori di tutta la nazione [3, 4].

### **1.1.1 Normativa europea: direttive RED II e EMD II**

La Renewable Energy Directive, Direttiva UE 2018/2001 e la Energy Market Directive, Direttiva UE 2019/944 vanno a delineare il contesto legislativo su cui si basano le comunità energetiche. L'oggetto della direttiva RED II è indirizzato a stabilire un quadro comune per la promozione dell'energia da fonti rinnovabili,

mentre la direttiva EMD II riguarda le norme comuni per il mercato interno dell'energia elettrica. All'interno di questi documenti vi sono dettate anche le norme relative al sostegno finanziario, all'autoconsumo, alle garanzie di origine e alle procedure amministrative, di informazione e formazione.

Riferendoci in particolare alla direttiva RED II, nell'articolo 2, dove vengono riportate tutte le definizioni del testo normativo, vi è proprio quella di "comunità di energia rinnovabile" ivi delineata come: "soggetto giuridico basato sulla partecipazione aperta e volontaria, autonomo e effettivamente controllato da azionisti o membri che sono situati nelle vicinanze degli impianti di produzione di energia da fonti rinnovabili che appartengono e sono sviluppati dal soggetto giuridico in questione" [5]. L'obiettivo principale è quindi fornire benefici ambientali, economici o sociali a livello di comunità ai suoi azionisti o membri o alle aree locali in cui opera, piuttosto che profitti finanziari. Gli azionisti che la normativa prevede come membri sono rappresentate da persone fisiche, PMI o autorità locali, comprese le amministrazioni comunali.

Nell'articolo 22 [5] viene fatta una limitazione importante, ovvero che se il membro è un'impresa la loro partecipazione non può costituire l'attività commerciale o professionale principale. Inoltre viene chiarito ed assicurato che i clienti finali, in particolare quelli domestici manterranno i loro diritti e doveri in qualità di clienti finali senza essere soggetti a condizioni o procedure ingiustificate o discriminatorie. I membri di tali comunità hanno il diritto di produrre, consumare, immagazzinare e vendere l'energia rinnovabile anche tramite contratti di compravendita e possono accedere a tutti i mercati dell'energia elettrica appropriati, direttamente o mediante aggregazione, in modo non discriminatorio. Il diritto principale è quello dello scambio dell'energia elettrica rinnovabile prodotta all'interno della stessa comunità, la quale si configura sia come produttrice che come consumatrice. L'utilizzo dell'energia autoprodotta dalla comunità è prioritario per l'autoconsumo istantaneo e tramite condivisione con i membri della comunità, mentre l'energia eventualmente eccedentaria può essere accumulata e venduta.

La direttiva, all'articolo 19 [5], prevede una garanzia di origine per dimostrare ai clienti finali la quota o la quantità di energia da fonti rinnovabili nel mix energetico di un fornitore di energia e nell'energia fornita ai consumatori. In base a contratti conclusi con riferimento al consumo di energia prodotta da fonti rinnovabili, gli Stati membri assicurano che l'origine dell'energia da fonti rinnovabili sia garantita come tale ai sensi della presente direttiva, in base a criteri obiettivi, trasparenti e non discriminatori. Il rilascio della garanzia di origine può essere subordinato a un limite minimo di potenza corrispondente ad una quantità standard di 1 MWh. Tale garanzia deve indicare requisiti minimi come:

- la fonte energetica utilizzata per produrre l'energia e le date di inizio e di fine della produzione;

- se riguarda: energia elettrica, gas incluso l'idrogeno, riscaldamento o raffrescamento;
- la denominazione, l'ubicazione, il tipo e la potenza dell'impianto nel quale l'energia è stata prodotta;
- la data e il paese di rilascio, e il numero identificativo unico.

Viene definita un'altra tipologia di Comunità Energetica nella direttiva europea EMD II 2019/944, quella di "*Comunità Energetica dei Cittadini*" (CEC), la stessa viene ripresa nel Decreto Legislativo del 8 novembre 2021 numero 210 che attua la direttiva. La definizione nel D.Lgs 210/2021 (entrata in vigore 26/12/2021) è ripresa puntualmente dalla direttiva UE che indica tale configurazione come: *un soggetto di diritto, con o senza personalità giuridica fondato sulla partecipazione volontaria e aperta e controllato da membri o soci che ha lo scopo principale di offrire ai suoi membri o soci o al territorio in cui opera benefici ambientali, economici o sociali a livello di comunità anziché perseguire profitti finanziari* [6, 7]. Viene precisato che i membri o soci (che possono essere persone fisiche, piccole imprese, autorità locali, ivi incluse le amministrazioni comunali, gli enti di ricerca e formazione, gli enti del terzo settore e di protezione ambientale, gli enti religiosi, nonché le amministrazioni locali) delle comunità energetiche dei cittadini regolano i loro rapporti tramite un contratto di diritto privato, individuando un soggetto responsabile, ivi inclusi la Comunità stessa, un membro o socio di essa o un soggetto terzo. In accordo con le CER, una CEC può partecipare alla generazione, anche da fonti rinnovabili, alla distribuzione, alla fornitura, al consumo, all'aggregazione, allo stoccaggio dell'energia, ai servizi di efficienza energetica, o a servizi di ricarica per veicoli elettrici o fornire altri servizi energetici ai suoi membri o soci. All'interno di detta comunità i cittadini hanno il diritto di organizzare la condivisione dell'energia prodotta dalle unità di produzione controllate e nella disponibilità (non per forza di proprietà, discorso identico per gli impianti di stoccaggio) dalla comunità stessa, essendo pienamente responsabile della ripartizione di tale energia condivisa. Anche in questo caso la comunità ha la possibilità di distribuire l'energia condivisa, attività che avviene tramite la rete di distribuzione esistente [6, 7].

### 1.1.2 Normativa italiana

In Italia il recepimento delle direttive europee in materia di promozione dell'uso dell'energia da fonti rinnovabili è avvenuto tramite l'approvazione del Decreto Mil-leproroghe 162/2019, in particolare nell'articolo 42-bis, definitivamente convertito con la Legge n.8/2020 del 28 febbraio 2020. Successivamente sono stati emanati i relativi provvedimenti attuativi:

- dell’Autorità di regolamentazione per energia, rete e ambiente (ARERA) con la delibera 318/2020/R/eel che si occupa delle componenti tariffarie;
- del Ministero dello Sviluppo Economico (MiSE) con il Decreto Ministeriale del 16 settembre 2020 relativo alla regolazione degli incentivi per l’energia condivisa.

È da sottolineare il fatto che quanto definito dal Decreto Milleproroghe non è definitivo, ma introduce una fase sperimentale. Infatti, il monitoraggio delle nuove configurazioni è funzionale all’acquisizione di elementi utili all’attuazione delle disposizioni europee in materia di autoconsumo di cui alle citate direttive 2018/2001 e 2019/944 rispettivamente del Parlamento e del Consiglio Europeo. La definizione dell’entità giuridica di autoconsumo collettivo da fonti rinnovabili viene ripresa senza sostanziali modifiche proprio come quella di "Comunità di Energia Rinnovabile" dalla direttiva RED II di cui sopra. Vi sono però particolari condizioni a cui devono sottostare tali entità giuridiche costituite per la per la realizzazione di comunità energetiche, tra le più significative [8]:

- gli impianti di produzione ammessi alla configurazione devono essere entrati in esercizio in data successiva al 1 marzo 2020 e avere una potenza nominale non superiore ai 200 kW, essi devono essere detenuti dalla CER, ciò significa che ne hanno la piena disponibilità anche nel caso in cui non ne siano i proprietari e possono essere gestiti anche da un produttore esterno;
- gli impianti di produzione e tutti i punti di consumo facenti parte della CER devono essere sottesi alla medesima cabina di trasformazione MT/BT (secondaria);
- la condivisione deve avvenire attraverso la rete elettrica già esistente, ovvero con un approccio “virtuale”<sup>2</sup>; infatti, il decreto non prevede né la creazione di nuovi tratti di rete né la cessione di parte della rete pubblica esistente per uso privato della comunità.

Il D.Lgs 199/2021 (entrato in vigore il 15/12/2021), attuando la direttiva europea RED II, ha reso meno stringenti i requisiti di dimensionamento degli impianti rinnovabili stabilendo i seguenti criteri:

- Gli impianti di produzione di energia elettrica da fonti rinnovabili devono avere una potenza complessiva non superiore a 1 MW ed essere connessi alla rete

---

<sup>2</sup>Con autoconsumo “virtuale” (detto anche “commerciale” o “su perimetro esteso”) ci si riferisce ad uno schema di utenza che prevede l’utilizzo della rete pubblica per lo scambio di energia tra unità di generazione e di consumo

elettrica attraverso la stessa cabina primaria (corrispondente territorialmente a circa 3-4 Comuni, oppure 2-3 quartieri di una grande città);

- possono aderire alla comunità energetica anche impianti a fonti rinnovabili già esistenti alla data di entrata in vigore del D.Lgs 199/2021, purché in misura non superiore al 30% della potenza complessiva che fa capo alla comunità.

Quest'ultimo passaggio sta permettendo lo sviluppo della normativa riguardante la materia delle comunità energetiche e quindi consentirne una diffusione su larga scala, sicuramente maggiore di quella ad oggi esistente [9].

A corollario della normativa emanata dal governo vengono pubblicati vari documenti da diversi enti e agenzie per chiarire tutti gli aspetti tecnici ed economici che vanno a completare il quadro referente all'ambito delle comunità energetiche e del loro sviluppo. I principali enti incaricati nella pubblicazione di questi documenti sono l'ARERA (Autorità di Regolazione per Energia Reti e Ambiente) e il MiSE (Ministero dello Sviluppo Economico), tramite le delibere e i decreti si può specificare e capire meglio la terminologia riguardante l'argomento delle comunità energetiche.

**Prosumer** Nel rapporto produzione - consumo di energia, il protagonista della rivoluzione energetica in atto è un nuovo soggetto denominato *Prosumer* (crasi tra i termini inglesi *producer* e *consumer*), la cui figura è appunto al centro della direttiva europea RED II del 2018, in cui viene precisato che la crescente incidenza dell'autoconsumo di energia rinnovabile ha reso necessaria l'introduzione dell'autoconsumatore di energia rinnovabile e, tra questi, quelli che agiscono collettivamente proprio tramite le comunità energetiche. Diventa così realtà il diritto all'energia pulita "a km zero". Come precedentemente detto, insito nel concetto di prosumer si colloca quello di autoconsumo. L'autoconsumo indica quindi l'utilizzo diretto da parte dei prosumers di tutta o parte dell'energia prodotta dagli impianti che possiedono individualmente o attraverso una comunità energetica. La configurazione che l'UE sta cercando di far diffondere ha il suo fondamento nella condivisione dei sistemi di produzione e della stessa energia derivante. Quindi nella gerarchia d'uso dell'energia rinnovabile prodotta dai membri di una CE vi è l'autoconsumo in primis e successivamente la messa a disposizione dell'energia in surplus agli altri membri mettendo l'accento così sul vero senso di comunità. Pertanto viene definito "*autoconsumatore di energia rinnovabile*" il cliente finale che, operando in propri siti situati entro confini definiti o, se consentito da uno Stato membro, in altri siti, produce energia elettrica rinnovabile per il proprio consumo e può immagazzinare o vendere energia elettrica rinnovabile autoprodotta purché, per un autoconsumatore di energia rinnovabile diverso dai nuclei familiari, tali attività non costituiscano l'attività commerciale o professionale principale [5]. Il tentativo di tutta questa regolamentazione è quindi di portare a compimento la nozione di cliente attivo del mercato dell'energia. Il "*cliente attivo*", anche detto "energy prosumer", è definito

come: *un cliente finale o un gruppo di clienti finali consorziati che consuma o conserva l'energia elettrica prodotta nei propri locali situati all'interno di un'area delimitata o, se consentito da uno Stato membro, in altri locali, oppure vende l'energia elettrica autoprodotta o partecipa a meccanismi di flessibilità o di efficienza energetica, purché tali attività non costituiscano la principale attività commerciale o professionale* [7]. Sono compresi in questa dicitura i singoli cittadini, famiglie, organizzazioni non commerciali, enti pubblici e Pmi che non solo consumano energia, ma partecipano attivamente al mercato dell'energia, individualmente o collettivamente, anche attraverso una comunità energetica. La partecipazione attiva al mercato può consistere nella produzione di energia rinnovabile, nel miglioramento dell'efficienza energetica e/o nella gestione del sistema energetico con l'integrazione della produzione discontinua dell'energia da fonti rinnovabili con le esigenze della domanda, l'aggregazione, lo stoccaggio. In questa ottica si ritiene che i clienti attivi debbano avere facile accesso alla rete e ai mercati dell'energia a condizioni eque e che gli Stati debbano favorire la partecipazione di clienti attivi nei vari mercati con agevolazioni e incentivi atti a promuoverla [10].

Con la delibera del 27 dicembre 2022 da parte dell'ARERA è stato chiarito come la molteplicità delle diverse tipologie di configurazioni dovessero essere trattate. È stata quindi identificata la categoria di *configurazione per l'autoconsumo diffuso* [11] valida per le seguenti tipologie:

- gruppo di autoconsumatori di energia rinnovabile che agiscono collettivamente;
- gruppo di clienti attivi che agiscono collettivamente;
- comunità energetica rinnovabile o comunità di energia rinnovabile;
- comunità energetica dei cittadini;
- autoconsumatore individuale di energia rinnovabile “a distanza” con linea diretta;
- autoconsumatore individuale di energia rinnovabile “a distanza” che utilizza la rete di distribuzione;
- cliente attivo “a distanza” che utilizza la rete di distribuzione.

L'energia prodotta deve essere destinata a soddisfare i fabbisogni dei membri della comunità, ma se talvolta o in certe ore della giornata ci sono quantità eccedenti queste possono anche essere vendute a terzi (creando ricchezza per i soci). È evidente peraltro che questa non deve essere una attività primaria (come specificato dalla normativa) finalizzata alla produzione per la vendita a terzi perché in questo caso, si porrebbe in concorrenza con quella di società con specifiche finalità commerciali e quindi non sarebbe compatibile con il ruolo di una CE, cui vengono accordate anche agevolazioni fiscali per poter fornire servizi ai soci a condizioni più favorevoli.



**Energia condivisa** Parlando di autoconsumo diffuso, si va ad integrare nella concezione di comunità energetica proprio la diffusione dell'energia prodotta dagli attori appartenenti a una delle configurazioni precedentemente elencate, così facendo il discorso si sposta sulle modalità di scambio di tale energia. Viene in aiuto pertanto la definizione di *energia condivisa* che "è, in ogni ora e per l'insieme dei punti di connessione ubicati nella stessa zona di mercato che rilevano ai fini di una configurazione per l'autoconsumo diffuso, il minimo tra l'energia elettrica immessa ai fini della condivisione e l'energia elettrica prelevata ai fini della condivisione" [11]. Si potrebbe pensare che in luogo dello scambio diretto di energia tra condomini una rete ad hoc sia efficace, mentre invece è stato pensato uno scambio dell'energia di tipo virtuale: tale modello consente di realizzare nuove configurazioni di autoconsumo senza il bisogno di nuova installazione o subentro nelle reti fisiche. Lo scambio virtuale consente infatti la gestione a livello commerciale dei flussi di energia, garantendo ai singoli utenti delle unità abitative un elevato livello di qualità del servizio, che tra l'altro permette ai singoli utenti di continuare ad esercitare i loro diritti come consumatori, inclusi i diritti di scegliere il proprio fornitore e di esercitare lo "switching" del fornitore. Le nuove tecnologie, quali gli smart meter elettrici di seconda generazione e i sistemi peer-to-peer, consentiranno agli aggregatori il coinvolgimento diretto del cliente finale come prosumer, avvalendosi dei distributori locali (DSO) quali soggetti neutrali in grado di abilitare e facilitare la transizione energetica. Tali scelte sono state fatte in un'ottica di efficientamento del sistema energetico per ridurre le perdite di rete e tutti i costi che in relazione ad esse devono essere sopportati dagli utenti finali.

**Rete distribuita** L'importanza della sintesi normativa racchiusa nel CEP riguarda l'attenzione che viene messa dall'istituzione europea sulle nuove configurazioni in grado di dar vita ad una vera e propria *generazione distribuita dell'energia*. Al giorno d'oggi viene considerato come un grave deficit per l'edificazione di un mercato libero europeo dell'energia la limitata partecipazione dei piccoli consumatori in tale mercato. Impostando così la normativa il legislatore considera le configurazioni residenziali di produzione e stoccaggio di energia elettrica il nuovo strumento chiave per raggiungere il prima possibile l'obiettivo della completa transizione energetica, aumentando considerevolmente le quote di energia elettrica provenienti da fonti rinnovabili. Contestualmente avverrebbe una diminuzione delle importazioni energetiche da paesi non UE, favorendo l'indipendenza ma senza massicci investimenti di massa nel settore industriale e di produzione dell'energia. I benefici della generazione distribuita riguardano da un lato la riduzione delle perdite di trasporto sulle reti elettriche grazie all'autoconsumo, dall'altro la distribuzione dei rischi della generazione su un ampio numero di impianti che sfruttano fonti rinnovabili, anche se vanno valutate varie problematiche legate al sistema di backup. Un sistema policentrico caratterizzato dalla diffusione di sistemi energetici locali è

inoltre caratterizzato da due tipologie di implicazioni: quelle di tipo generale/sistemiche rispetto alla produzione energetica (tra cui: risparmio energetico e costi di produzione, le performance ambientali e di riduzione delle emissioni inquinanti, la gestione delle infrastrutture, i costi della rete nazionale e le esternalità negative), e le implicazioni rispetto ai singoli utenti o a coalizioni di utenti (riduzione dei consumi e miglioramento della qualità ed affidabilità della fornitura energetica; riduzione della vulnerabilità energetica di alcune fasce di popolazione) [10].

## 1.2 Fonti di energia rinnovabile e fotovoltaico

In questa trattazione si vuole presentare una panoramica su tutto ciò che comprende il concetto di Comunità di Energia Rinnovabile. Precedentemente è stata trattata la parte normativa con le annesse definizioni mentre in questa sezione si vuole approfondire il ruolo della generazione di energia rinnovabile. Partendo dalla Direttiva europea RED II e spostandoci verso i vari piani nazionali come il PNIEC si evince una marcata pressione riguardo all'aumento delle quote di consumi dovute ad energia rinnovabile. In termini di decarbonizzazione l'obiettivo è il raggiungimento della quota del 55% di generazione energetica da fonti rinnovabili. Nonostante negli ultimi anni le installazioni siano incrementate, le quote di consumi relative alle FER (Fonti di Energia Rinnovabile) non presentano lo stesso andamento. Vengono presentati di seguito alcune rilevazioni statistiche<sup>3</sup> (estratte da rapporti del GSE) che mettono in evidenza la situazione delle FER in Italia.

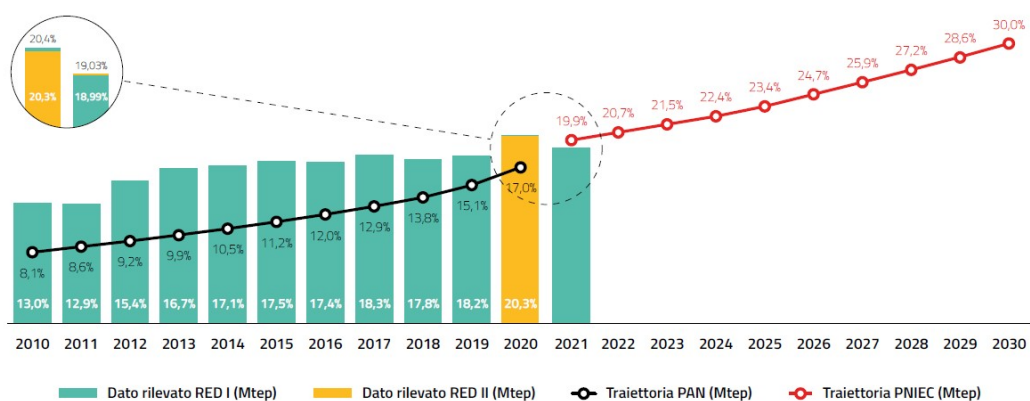
Il consumo finale lordo di energia da fonti rinnovabili rilevato in Italia nel 2021 ammonta a 22,9 Mtep<sup>4</sup>, equivalenti a circa 960.000 TJ (266,7 TWh). Il 48,7% dei consumi si concentra nel settore Termico (11,2 Mtep) ed è associato principalmente agli impieghi di biomassa solida per riscaldamento (legna da ardere, pellet) e all'ampia diffusione di apparecchi a pompa di calore. Molto rilevante è anche il ruolo delle FER nel settore Elettrico (10,2 Mtep, per un'incidenza del 44,5% sul totale dei consumi di energia da FER); in questo caso, oltre alla tradizionale fonte idrica (4,17 Mtep, dato normalizzato), assumono un ruolo significativo tutte le altre fonti rinnovabili: solare (2,15 Mtep), eolica (1,75 Mtep, dato normalizzato), bioenergie (1,63 Mtep), e geotermica (0,51 Mtep) [12]. Nel 2021 la quota dei consumi finali lordi di energia coperta da fonti rinnovabili (Figura 1.2) risulta

---

<sup>3</sup>Ai fine del monitoraggio dei target UE vi è la necessità di calcolare i valori secondo la metodologia della Direttiva europea RED II creando una piccola discrepanza in fase di confronto tra i valori relativi agli anni precedenti calcolati secondo la metodologia precedente (relativa alla RED I).

<sup>4</sup>unità di misura che indica le tonnellate equivalenti di petrolio. Conversione 1ktep = 11.63 GWh

pari al 19,03%, in diminuzione rispetto al dato 2020 pari a 20,3% ed inferiore alla traiettoria prevista dal PNIEC per il 2021 (19,9%). Su questa dinamica si notano gli effetti della pandemia: a fronte di una crescita di consumi di energia da FER relativamente contenuta (+3,9%), i consumi energetici complessivi del Paese sono infatti cresciuti ad un ritmo più che doppio rispetto al 2020 (+10,6%). Per quanto riguarda il contributo dei settori specifici il GSE ne individua tre: Elettrico, Termico, Trasporti. L'unico settore in linea con le previsioni PNIEC è quello dei Trasporti (10% rispetto al 9.9% previsto dal PNIEC) mentre gli indicatori-obiettivo relativi al settore Elettrico (36% rispetto al 37.5%) e al settore Termico (19.7% rispetto al 22.1%) mostrano valori inferiori alle previsioni del PNIEC: in entrambi i casi, infatti, nel 2021 la quota dei consumi complessivi coperti da FER risulta inferiore alle previsioni [12].



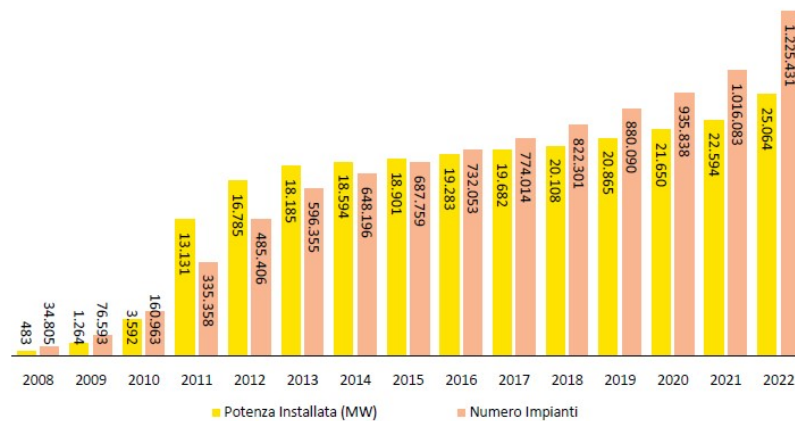
**Figura 1.2:** Quota dei consumi finali lordi di energia coperta da FER [12]

Per quanto concerne le CER e più specificatamente dei parametri economici valutati per la loro diffusione si evidenzia l'importanza dell'energia elettrica condivisa che è poi quella oggetto di valorizzazione economica e incentivazione. Ecco perché è utile dare uno sguardo più approfondito al settore Elettrico. A fine 2021, per il settore Elettrico la potenza efficiente lorda dei circa 1.030.000 impianti a fonti rinnovabili installati in Italia è pari a 58,0 GW; l'incremento rispetto al 2020 (+2,5%) è legato principalmente alle nuove installazioni di impianti fotovoltaici (+944 MW) ed eolici (+383 MW). La produzione lorda di energia elettrica da FER nel 2021 è pari a 116,3 TWh, in leggera diminuzione rispetto al 2020 (-0,5%); essa rappresenta il 40,2% della produzione complessiva nazionale. La produzione elettrica calcolata applicando i criteri delle direttive europee sulle energie rinnovabili risulta invece in lieve aumento (+0,3%); in questo caso essa rappresenta il 36,0% del consumo interno lordo di energia elettrica.

Considerando la ripartizione per tipo di fonte, la fonte energetica rinnovabile che nel 2021 garantisce il principale contributo alla produzione complessiva di energia elettrica da FER si conferma quella idroelettrica (39,0% del totale); seguono solare (21,5%), eolica (18,0%), bioenergie (16,4%) e geotermica (5,1%) [12]. Nonostante la quota parziale di energia solare che interviene sul settore Elettrico non sia la preponderante, per lo sviluppo delle CER quella solare si configura come la fonte di energia rinnovabile più adatta su cui puntare, soprattutto in aree urbane e industriali dove è difficile installare impianti eolici o idroelettrici. Per diverse ragioni il fotovoltaico è una delle tecnologie di generazione distribuita più significative. Tra queste: la maturità della tecnologia che permette di trasformare la radiazione solare in energia elettrica, seppur continui a essere attivo il settore di ricerca e lo sviluppo per migliorarla; il suo basso costo (se confrontato con gli altri sistemi di produzione da fonti rinnovabili); l'abbondanza della fonte; la facilità di installazione negli edifici e la sua scalabilità. Secondo numerosi scenari che prospettano il raggiungimento di produzione di energia elettrica 100% rinnovabile, il fotovoltaico rappresenta la soluzione ad oggi più economicamente conveniente per incrementare la produzione di energia pulita [10].

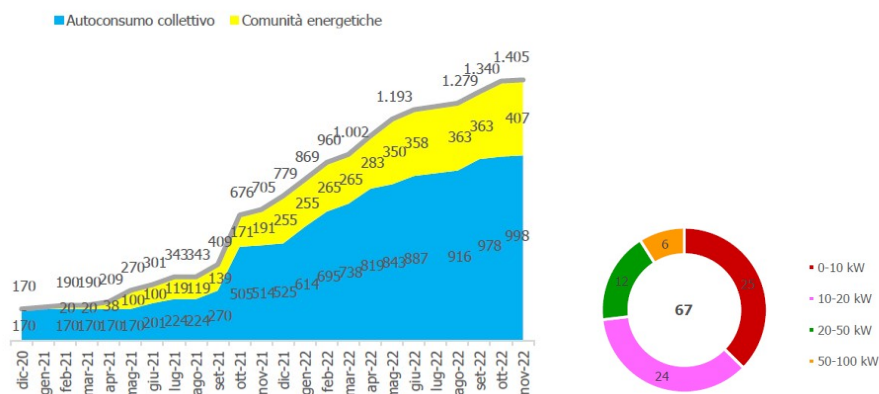
La tecnologia fotovoltaica ha visto un grande sviluppo dal 2008 ad oggi. Al 31 dicembre 2022 risultano installati in Italia 1.225.431 impianti fotovoltaici, per una potenza complessiva pari a 25.064 MW (+21% di impianti installati annualmente, +11% di potenza, +12% di produzione). Gli impianti di potenza inferiore o uguale a 20 kW costituiscono il 93% del totale in termini di numerosità e il 26% in termini di potenza (Figura[1.3]); la taglia media degli impianti è poco superiore a 20 kW [13]. Il numero degli impianti entrati in esercizio nel corso del 2022 è significativamente più elevato rispetto all'analogo dato rilevato nel 2021 (+161%); risulta ampiamente positiva anche la variazione della potenza installata (+165%). Il trend del fotovoltaico è in linea con la traiettoria PNIEC, il quale prevede al 2030 una potenza fotovoltaica installata pari a 51 GW. Importante è anche il dato decrescente della quota dei sistemi fotovoltaici collocati non a terra, che fino al 2013 hanno visto un incremento quasi alla pari, ma che oggi vede il doppio della potenza installata nei sistemi non a terra. Alla fine del 2022 la potenza fotovoltaica installata a terra ammonta a 8.403 MW (+4,6% rispetto al 2021), pari al 34% del dato complessivo nazionale; i 16.661 MW di potenza installata non a terra (+14,6% rispetto al 2021) rappresentano il restante 66% del totale nazionale[13].

Dai dati del GSE al 31/12/2022 risultano 46 configurazioni di Autoconsumo Collettivo e 21 Comunità Energetiche Rinnovabili per un totale di 67 configurazioni e in termini di potenza risultano 1,4 MW, tutti da fonte fotovoltaica. Complessivamente delle 67 configurazioni in esercizio, quasi il 75% sono relative a impianti fino a 20 kW di potenza e il numero medio di clienti finali facenti parte delle configurazioni è crescente con la taglia dell'impianto. Alla data di rilevazione risultano 501 clienti finali connessi a configurazioni AC/CER, di cui oltre il 70%



**Figura 1.3:** Evoluzione della potenza e della numerosità degli impianti Fotovoltaici [13]

relativi ad Autoconsumo Collettivo [14].



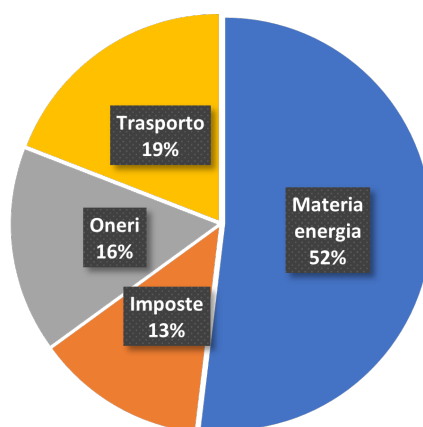
**Figura 1.4:** Andamento cumulato della potenza per AC/CER (al 31/12/22) [kW]; Distribuzione degli impianti fotovoltaici in AC/CER per taglia dell'impianto

### 1.3 Vantaggi CE: benefici economici, sociali, ambientali

Alla base dello sviluppo delle configurazioni di comunità energetiche, come esPLICITATO nella correlata normativa, vi è l'obiettivo di concretizzare benefici di tipo ambientale, economico, sociale per diverse figure beneficiarie. Ovviamente perché

ci sia uno sviluppo e una diffusione sostenibile delle CE il primo beneficio che deve essere realizzabile è quello economico. Attualmente si parla molto del peso della bolletta elettrica e se il consumatore finale non dovrà affrontare i costi di distribuzione, trasmissione e trasporto dell'energia elettrica (che rappresentano circa il 35%) la stessa bolletta graverà in misura minore. Secondo le stime di ARERA [10], per un consumatore domestico tipo<sup>5</sup>, la composizione percentuale del prezzo dell'energia elettrica è data da:

- spesa per la materia energia. Rappresenta circa il 51,9% ed è composta da approvvigionamento dell'energia (43,6% e commercializzazione al dettaglio (8,3%);
- imposte per circa il 13,1%;
- spesa per oneri di sistema che equivalgono a circa il 15,9% della bolletta;
- spesa per il trasporto e la gestione del contatore. Rappresenta circa il 19,1%, al suo interno si trova la trasmissione, che è l'attività di Terna e pesa per circa il 3,5% sulla bolletta elettrica.



**Figura 1.5:** Composizione del prezzo dell'energia per consumatore domestico tipo

I piani europei e nazionali sono costruiti per dare sostegno ai cittadini in questa transizione. Il PNRR prevede l'allocazione di 2,2 miliardi di euro per le iniziative di CER e di autoconsumo collettivo costituite nei Comuni con popolazione inferiore a 5000 abitanti. L'incentivazione del decentramento dell'energia e il tentativo

---

<sup>5</sup>Condizioni economiche di fornitura per una famiglia con 3 kW di potenza impegnata e 2.700 kWh di consumo annuo

di coinvolgere anche le piccole comunità permetterà sicuramente una maggior diffusione delle comunità energetiche, il tutto è aiutato dall'ampliamento delle restrizioni alle CER con potenza massima installabile fino a 1MW e maggiore area di condivisione con la possibilità di collegare i sistemi di generazione FER alla cabina di scambio primaria. Il legislatore ha previsto una serie di misure, volte ad incentivare l'appartenenza ad una comunità energetica, che vanno a favorire in particolare l'energia condivisa, ovvero la quantità di energia prodotta dagli impianti a energia rinnovabile, ceduta alla rete e che contemporaneamente viene utilizzata da qualcuna delle utenze appartenenti alla CER. Le entrate economiche sono costituite sia dal corrispettivo per l'energia elettrica ceduta alla rete (corrispondente al prezzo zonale orario) che dall'incentivo ventennale in tariffa per l'energia condivisa. Enti come il MiSE (Ministero dello Sviluppo Economico), il MASE (Ministero dell'Ambiente e della Sicurezza Energetica), il GSE (Gestore Servizi Energetici), l'ARERA sono incaricati di redigere la documentazione tecnica ed economica per attuare i decreti del governo in materia di energia rinnovabile ed energia condivisa, in modo da chiarificare in maniera inequivocabile le modalità di partecipazione alle configurazioni e le partite economiche relative ad ognuna di esse. Il 4 agosto 2020 l'ARERA con la delibera 318/2020/R/eel [15] aveva fornito le prime disposizioni riguardo la regolazione delle partite economiche sull'energia elettrica condivisa, queste sono state aggiornate con l'introduzione del concetto di *Autoconsumo Diffuso* nella delibera 727/2022/R/eeL del 27 dicembre 2022. Con il TIAD (Testo Integrato Autoconsumo Diffuso) l'ARERA conferma il modello regolatorio virtuale. In quest'ultimo documento viene indicato intanto che l'autoconsumo diffuso è incompatibile con il regime di scambio sul posto, e che tutti gli adempimenti sono in capo al GSE. Il GSE è deputato al calcolo della quantità di energia elettrica condivisa oraria e mensile per le configurazioni di autoconsumo diffuso. Inoltre ai fini della determinazione del contributo per la valorizzazione dell'energia autoconsumata ( $C_{ACV}$ ), calcola per ciascuna configurazione per l'autoconsumo diffuso, la quantità di energia elettrica autoconsumata oraria e mensile ( $E_{ACV}$ ), quest'ultima pari alla somma delle quantità di energia elettrica autoconsumata oraria nelle ore del mese. Nel caso di comunità energetiche rinnovabili, di comunità energetiche dei cittadini, di autoconsumatori individuali di energia rinnovabile "a distanza" che utilizzano la rete di distribuzione e di clienti attivi "a distanza" che utilizzano la rete di distribuzione, il GSE calcola su base mensile, per ciascuna configurazione, il contributo per la valorizzazione dell'energia elettrica autoconsumata ( $C_{ACV}$ ), espresso in euro (€), pari al prodotto tra l'energia elettrica autoconsumata  $E_{ACV}$  e il corrispettivo unitario di autoconsumo forfetario mensile  $CU_{Afa},m$ . Quest'ultimo è pari alla parte unitaria variabile, espressa in c€/kWh, della tariffa di trasmissione ( $TRAS_E$ ) definita per le utenze in bassa tensione.

$$C_{ACV} = CU_{Afa},m * E_{ACV} \quad (1.1)$$

Viene riconosciuto alla comunità energetica il contributo per l'energia elettrica autoconsumata ( $C_{ACV}$ ) e il contributo per l'energia elettrica oggetto di incentivazione ( $C_{ACI}$ ). Esso, su base mensile, è pari alla somma de:

- il termine  $C_{ACV}$  calcolato secondo la formula (1.1);
- il prodotto tra la tariffa incentivante definita dal MASE ai sensi del decreto ministeriale del 16 settembre 2020 e la quantità di energia elettrica a cui essa è riferita.

Nel suddetto decreto ministeriale [16] all'articolo 3 viene definita la tariffa incentivante, concessa per un periodo di 20 anni, riconosciuta per la quota di energia condivisa sotto forma di tariffa premio pari a:









- 100 €/MWh nel caso in cui l'impianto di produzione faccia parte di una configurazione di autoconsumo collettivo, ovvero gruppi di autoconsumatori di energia rinnovabile che agiscono collettivamente o gruppi di clienti attivi che agiscono collettivamente;
- 110 €/MWh nel caso in cui l'impianto faccia parte di una comunità energetica, rinnovabile o di cittadini.

Secondo una consultazione pubblica del MASE [17] vengono fatte alcune integrazioni alla disciplina di regolamentazione degli incentivi come:

- si ipotizza di prevedere un contingente complessivo sull'intero periodo posto (20 anni) pari a 5 GW, al raggiungimento del quale il decreto non sarebbe più applicabile, salva una successiva disposizione o un aumento della potenza messa a contingente;
- per gli impianti fotovoltaici la tariffa verrebbe corretta per tenere conto dei diversi livelli di insolazione secondo un fattore di correzione di: +4 €/MWh per le Regioni del Centro (Lazio, Marche, Toscana, Umbria, Abruzzo) e +10 €/MWh per le Regioni del Nord (Emilia-Romagna, Friuli-Venezia Giulia, Liguria, Lombardia, Piemonte, Trentino-Alto Adige, Valle d'Aosta, Veneto);
- nel caso in cui la quota di energia condivisa fosse pari o superiore al 70% dell'energia prodotta, la quota residua di energia potrebbe essere liberamente venduta dal produttore;
- nel caso, invece, in cui la quota di energia condivisa fosse inferiore al predetto limite del 70%, sull'energia elettrica eccedentaria venduta sarebbe previsto un tetto di prezzo pari a 80 €/MWh.



Un rapporto dedicato alle comunità energetiche, pubblicato nel 2020 dall'Energy & Strategy Group del Politecnico di Milano [18] ha mostrato (Figura 1.6) le stime sulle ricadute energetiche, ambientali e sociali di tali configurazioni. Tali stime sono state calcolate secondo l'ipotesi di tre scenari con diversi gradi di ambizione (basso, medio, alto), incentrati sullo sviluppo delle CE e associati alle condizioni di recepimento delle direttive europee RED II e EMD II. L'istituzione di una CER

		SCENARIO BASSO	SCENARIO MEDIO	SCENARIO ALTO
	Potenza di fotovoltaico installata	100 MW, pari al 1,5% dell'obiettivo PNIEC per il 2025	3.600 MW, pari al 55% dell'obiettivo PNIEC per il 2025	5.400 MW, pari al 80% dell'obiettivo PNIEC per il 2025
	Energia prodotta da rinnovabili (cumulato)	2,3 TWh	84 TWh	126 TWh
	Energia prodotta da rinnovabili (annuale - 2025)	120 GWh/anno, pari allo 0,3% dell'obiettivo PNIEC per il 2025	4.400 Wh/anno, pari al 12% dell'obiettivo PNIEC per il 2025	6.600 GWh/anno, pari al 18% dell'obiettivo PNIEC per il 2025
		SCENARIO BASSO	SCENARIO MEDIO	SCENARIO ALTO
	Riduzione perdite di rete	2 GWh, pari a circa 100 mila €	74 GWh, pari a circa 4 mln €	98 GWh, pari a circa 5 mln €
	Riduzione costi di distribuzione e di trasmissione	circa 14,5 mln €	circa 540 mln €	circa 720 mln €
	Riduzione emissioni CO <sub>2</sub> (cumulato)	0,8 mln ton: circa 24 mln € se CO <sub>2</sub> valorizzata a 30 €/ton	25,6 mln ton: circa 750 mln € se CO <sub>2</sub> valorizzata a 30 €/ton	39 mln ton: circa 1.170 mln € se CO <sub>2</sub> valorizzata a 30 €/ton
	Riduzione emissioni CO <sub>2</sub> (annuale 2025)	Trascurabile	1,3 mln ton	2,1 mln ton
	Posti di lavoro diretti	100 unità	6.500 unità	10.500 unità

**Figura 1.6:** Ricadute energetiche, ambientali e sociali con lo sviluppo delle Comunità energetiche in Italia<sup>6</sup>

prevede un'oculata pianificazione dell'investimento per effettivamente ottenerne un ritorno economico. Attualmente è possibile pianificare e valutare iniziative per la costituzione di CER tramite diversi tipi di strumenti disponibili sul mercato da utilizzare nella fase di avvio. Tra questi si cita il RECON (Renewable Energy Community ecONomic calculator<sup>7</sup>) uno strumento sviluppato da ENEA, finalizzato a supportare valutazioni preliminari di tipo energetico, economico e finanziario delle configurazioni di comunità energetica rinnovabile (CER) o di autoconsumatori

<sup>6</sup>le stime sono calcolate rispetto agli incentivi divulgati limitatamente al 2020

<sup>7</sup><https://recon.smartenergycommunity.enea.it>

di energia rinnovabile che agiscono collettivamente. Tramite RECON è possibile calcolare la resa energetica, l'autoconsumo, la condivisione dell'energia, l'impatto ambientale (riduzione delle emissioni di CO<sub>2</sub>), i flussi di cassa attualizzati e i principali indicatori finanziari (VAN, TIR, payback time, ecc.) considerando diverse forme di finanziamento dell'investimento, le detrazioni fiscali e gli incentivi introdotti dalla recente normativa.

**Benefici sociali** Viene rimarcato in maniera considerevole nella normativa come le comunità energetiche sono soggetti atti a fornire "*benefici ambientali, economici o sociali*". Una delle problematiche sociali che queste configurazioni hanno la possibilità di attenuare è quella della povertà energetica, tematica che non va sottovalutata. A fine 2021 l'Osservatorio Italiano sulla Povertà Energetica (OIPE) ha aggiornato la stima dell'incidenza del fenomeno (povertà energetica) nel 2020, segnalando come esso coinvolgesse l'8% dei nuclei familiari italiani, per la precisione 2,1 milioni di famiglie, per questo è diventato un problema non più trascurabile vista l'incidenza sulla popolazione nazionale. Per *povertà energetica (PE)* si intende "*la difficoltà ad acquistare un paniere minimo di beni e servizi energetici oppure come la condizione per cui l'accesso ai servizi energetici implica una distrazione di risorse (in termini di spesa o di reddito) superiore a quanto socialmente accettabile*" (misura convenzionale usata da molti Paesi: si considera PE quando la spesa energetica supera il 10% della spesa totale) [2]. L'attenzione alla povertà energetica, la cui stima è utilizzata dal Governo Italiano nella Strategia Energetica Nazionale (SEN) del 2017 e nel Piano nazionale integrato energia e clima (PNIEC) del 2019, è a maggior ragione elevata in questa fase segnata dal rischio latente della crisi energetica e dal caro-bollette, dovute principalmente a crisi economica post Covid e alla situazione bellica tra Russia e Ucraina [19].

Per far sì che le comunità energetiche possano giocare un ruolo chiave nel processo di decarbonizzazione, e tutto ciò che ne consegue, occorre un sistema di norme trasparenti che sia in grado non solo di fissare obiettivi precisi e concreti ma altresì di coordinare i diversi organismi incaricati e di intervenire nel processo autorizzativo e di gestione, così da non ostacolarne lo sviluppo ma anzi favorirlo. Nel promuovere la crescita del mercato dell'energia da FER e da altre forme di generazione che possano contribuire alla creazione/riuso/risparmio di energia, è necessario tener conto del loro impatto positivo non solo sullo sviluppo a livello regionale e locale ma anche sulla coesione sociale e sulla creazione di posti di lavoro, in particolare per quanto riguarda le Pmi e i produttori indipendenti di energia [10].

A supporto della diffusione delle comunità energetiche e dello sviluppo della generazione distribuita, un team del Politecnico di Milano ha redatto un rapporto sulle Smart grid [20] in cui ha classificato i benefici derivanti dalle CE come segue:

1. Benefici per le utenze energetiche:

- Migliore qualità dell'energia e affidabilità.
- Maggiore accettabilità delle tecnologie rinnovabili da parte dei cittadini tramite la partecipazione attiva della popolazione ai vari progetti comunitari (le CE consentono una maggior presa di coscienza e accettazione delle tecnologie rinnovabili, incrementandone le installazioni).
- Ottimizzazione della spesa energetica:
  - Copertura del fabbisogno delle utenze con costo minimo;
  - Dispacciamento in tempo reale delle risorse locali e massimizzazione autoconsumo;
  - Partecipazione attiva alla gestione del sistema.
- Economia di scala: partecipando in assetto comunitario agli investimenti, il costo pro-capite è minore, inoltre generalmente impianti di taglia maggiori hanno un'efficienza maggiore.

2. Benefici per il sistema elettrico:

- Maggiore stabilità e affidabilità del sistema poiché è minore la quota di energia derivante da fonti rinnovabili messa in rete, la maggior parte di essa viene autoconsumata all'interno delle comunità.
- Riduzione delle perdite di trasmissione, distribuzione e conversione se la rete interna delle LEC (Local Energy Community) ha lunghezza ridotta.
- Riduzione della dipendenza energetica a livello nazionale perché si fa maggior ricorso alle tecnologie rinnovabili del territorio.
- Migliore e minore utilizzo delle infrastrutture di trasmissione e dispacciamento dell'energia. Essendo l'energia prodotta localmente all'interno delle comunità le infrastrutture della rete pubblica sono meno utilizzate.
- Le comunità energetiche sono delle aggregazioni e in quanto tali potranno partecipare ai servizi di rete.

3. Benefici per il territorio:

- Minori inquinanti immessi in ambiente.
- Promozione dell'immagine Green del territorio.
- Miglior uso del territorio (occupazione del territorio).
- Maggior tasso di occupazione: le nuove installazioni producono posti di lavoro a lungo termine, le tecnologie vanno infatti installate e mantenute nel tempo.

#### 4. Benefici per il consumatore:

- I benefici di scala (dovuti alla maggiore dimensione degli investimenti) e i benefici riguardanti le sinergie che si riescono ad ottenere dall'unione di più utenze energetiche. I benefici di scala si hanno non solo sull'investimento iniziale, in riferimento al fatto che, gli impianti di maggiore dimensione hanno costi unitari inferiori) ma anche sul funzionamento degli asset in riferimento al fatto che, gli impianti di taglia maggiore a parità di altre condizioni hanno una maggiore efficienza.
- Lo sfruttamento locale di energia: l'aggregazione di più utenze energetiche può abilitare la possibilità di sfruttare a livello locale l'energia che altrimenti andrebbe persa (o ceduta in rete ad un valore inferiore).
- La riduzione della potenza contrattuale impiegata: fa riferimento alla riduzione della cosiddetta "quota potenza" della bolletta energetica.
- La diminuzione del numero di punti di prelievo – POD: la riduzione della cosiddetta "quota fissa" della bolletta energetica.

Il potenziale di diffusione delle CE in Italia varia da un minimo di 25.000 ad un massimo di quasi 100.000 nei diversi scenari analizzati, cui corrisponde un volume d'investimenti complessivo dell'ordine dei 50-150 mld €[20]. Le ricadute sistemiche associate alla diffusione delle comunità energetiche sono dunque piuttosto rilevanti ed i costi sostenuti dal sistema elettrico (su cui ha impatto la diffusione delle CE) potrebbero ridursi di circa il 10-20% [20].

Come indicato precedentemente ad oggi sono 67 le configurazioni di AC/CER nel Paese. Possiamo notare un aumento degli investimenti nel settore delle rinnovabili elettriche nel 2022 rispetto al 2021 (dai 2 mld€ del 2021 ai 3,8 mld€ del 2022) [14], trainati dal fotovoltaico e dall'eolico. Questo trend è sicuramente favorevole allo sviluppo delle CE. L'impatto sociale che possono avere le CE influenza anche la sfera occupazionale, che viene quantificata in termini di ULA (Unità di Lavoro per Anno), essa tiene conto di tre componenti: occupazione diretta, indiretta e indotta. Secondo le previsioni del PNIEC, nel settore della produzione di energia elettrica in termini di ULA, gli occupati cresceranno da 37.710 unità nel 2017 a 50.683 nel 2030: con un saldo positivo (+34% circa) pari a 12.973 ULA, rispettivamente +1.086 da idroelettrico, +5.101 da eolico, +9.197 da solare e +94 da geotermico [2]. Le ricadute occupazionali temporanee dirette e indirette (legate alla costruzione e installazione di nuovi impianti) nel 2022 si stimano a circa 25 mila ULA, mentre gli occupati permanenti diretti e indiretti (legati alla gestione e manutenzione degli impianti esistenti) nel 2022 si stimano pari a circa 35 mila ULA permanenti [14].

## 1.4 Integrazione del forecasting nelle CER

Le Comunità Energetiche Rinnovabili come presentato nelle sezioni precedenti hanno l'obiettivo di ridurre degli sprechi di energia tramite la loro nozione fondante di condivisione dell'energia green; inoltre grazie alla direzione presa dalla normativa queste configurazioni possiedono anche un notevole potenziale di risparmio economico per gli appartenenti alle CER. Per fare in modo che si sfrutti al meglio tale potenziale è importante seguire l'onda della tecnologia e della ricerca. Gli studi su questo argomento sono di ampio raggio ma sicuramente i più interessanti e completi sono quelli che considerano più prospettive insieme soprattutto quella tecnica e quella economica.

Nel campo della gestione efficiente dell'energia vi sono moltissimi esempi in letteratura di studi e meccanismi che hanno l'obiettivo di stimare la domanda energetica di lungo o breve periodo, poiché i consumi di energia sono spesso dinamici e vi è la necessità di strumenti accurati che ne possano prevedere i valori. La necessità deriva dal continuo incremento della popolazione e della sempre maggiore richiesta di energia (in questa sede ci si limita a considerare l'energia elettrica). Per risolvere o migliorare il problema della gestione dell'energia dalla generazione alla trasmissione e distribuzione è stato dimostrato che le Smart Grid [21] hanno un grosso potenziale soprattutto in collaborazione con tecniche di forecasting che permettono di avere previsioni sui consumi futuri. Tramite algoritmi di Machine Learning c'è la possibilità di gestire meglio la generazione di energia e far sì che il bilanciamento della rete elettrica avvenga in modo più rapido. Diversi studi hanno analizzato le metodologie di forecasting come [22] che mette a confronto diverse tecniche di ML come Logistic Regression (LR), Support Vector Machines (SVM), Decision Tree Classifier (DTC) per verificare la loro accuratezza e il loro errore sul forecast riguardo al problema del Short-Term Load Forecast (STLF).

Le iniziative relative alle Smart Grid in tutto il mondo stanno trasformando la rete in un sistema più robusto, dinamico e aperto che porterà i consumatori e i loro dispositivi direttamente nel regno della gestione della rete stessa. Sono state proposte molte strategie di pianificazione Demand Side Management (DSM) per le reti intelligenti, ma per implementare tali metodi di pianificazione è fondamentale la conoscenza della quantità di domanda di energia a livello domestico. Ciò richiede una previsione del carico a breve termine per le case e, in alcuni casi, anche per i dispositivi. Nel lavoro [23] vengono proposte due metodologie per il STLF attraverso cui l'accuratezza della previsione dei carichi domestici migliora del 50%. Le tecniche di forecasting utilizzate sono quelle di Multi Linear Regression (MLR) e Artificial Neural Network (ANN), che aiutano di conseguenza anche sul lato della Demand Response per rendere più stabile la rete a cui sono collegati e che servono grazie alla modulazione del proprio consumo, proprio quest'ultima è facilitata conoscendo le previsioni di consumo dettate degli algoritmi di forecasting.

Un altro lavoro [24] propone un modello basato sul Demand Response (DR) per gestire le risorse di una comunità energetica. Tale modello si avvale di un algoritmo di forecasting, basato su una rete neurale, per prevedere il consumo di energia, la generazione di fotovoltaico e il profilo di flessibilità del giorno seguente. Dopo aver previsto i consumi energetici e considerati tutti i vicoli del modello vengono identificati i possibili eventi di Demand Response. Tale analisi permette di individuare periodi corrispondenti a surplus di produzione, che verranno considerati a supporto dell'eventuale evento di DR. La comunità considerata è composta da 250 costruzioni residenziali di cui 200 hanno installato dei pannelli fotovoltaici e per ogni residenza si ipotizza la presenza di due tipi di veicoli elettrici con differenti capacità e tempi di ricarica. Questo modello cerca di sfruttare la flessibilità energetica dei membri per bilanciare le risorse energetiche della comunità. Dopo aver identificato i possibili eventi di DR attraverso un algoritmo di forecasting, utilizzando un algoritmo di clustering vengono selezionati i membri della comunità che saranno invitati a partecipare all'evento per abbassare il picco di produzione fotovoltaica nella rete e distribuire il surplus. I risultati ottenuti testando questo modello con i dati della comunità energetica dimostrano l'efficienza del modello, consentendo di risparmiare 13,66 € e di ridurre 32,34 kg di emissioni di CO<sub>2</sub> per un solo evento di DR.

Nel lavoro [25] viene studiato il valore economico del fotovoltaico distribuito per un sistema di comunità energetica rinnovabile rispetto alle condizioni specifiche del Paese, il confronto è fatto tra otto Paesi europei. Per arrivare a proporre le stime economiche lo studio presenta la modellazione della domanda elettrica, che in questo caso è stata effettuata tramite un tool di simulazione basato sul Load Profile Generator (LPG). La valutazione dei profili di domanda elettrica potrebbe essere integrata con tecniche di forecasting (sempre in presenza di una quantità di dati storici sufficiente) per rendere più precisa la valorizzazione dell'energia condivisa. I risultati mostrano che le comunità di energia rinnovabile hanno il potenziale per ridurre i costi dell'elettricità grazie agli investimenti comunitari e all'aggregazione del carico.

Diversi in letteratura sono gli studi che analizzano configurazioni come le comunità energetiche, o simili dove vi è presenza di condivisione di energia, sotto il profilo del modello di business e valutando le relazioni che intercorrono tra i membri utilizzando gli strumenti della teoria dei giochi. Ad esempio la pubblicazione [26] propone di trattare la comunità energetica come un gioco di coalizione (campo specifico della teoria dei giochi), in cui i giocatori sono i finanziatori e i consumatori membri della configurazione. Seguendo questo approccio è possibile determinare una regola di allocazione equa che dia un adeguato profitto agli stakeholder, riflettendo il loro reale contributo sui ricavi. Utilizzando questo modello di coalizione viene presentato in un altro studio [27], un business model basato sulla regolamentazione italiana pertinente alle CER e agli incentivi a loro dedicati

sull'energia condivisa e autoconsumata. La struttura della CER utilizzata in tale lavoro è definita in modo che sia necessario considerare il contributo di quattro componenti: produzione di energia dall'unità fotovoltaica, consumo elettrico per i carichi, comportamento dell'unità di accumulo e scambio di potenza con l'energia elettrica di rete. In tale approccio, la produzione fotovoltaica e i carichi elettrici sono considerati noti, ma se così non fosse anche in tal caso le metodologie di forecasting potrebbero essere implementate per rendere più affidabile e real-time il modello utilizzando dati storici per prevedere i consumi e la produzione di energia elettrica futura.

Nello studio [28] viene analizzata una comunità residenziale di prosumer, che tramite un sistema definito di P2P (peer to peer, ovvero condivisione tra pari) si scambia energia solare generata tramite pannelli fotovoltaici. Anche se il modello P2P non è afferente alle modalità di comunità energetica, perché lo scambio avviene su collegamenti privati, il modello mostra l'efficiente coesistenza di energia condivisa tramite l'utilizzo di metodologie di forecasting dei consumi. Il modello [28] include il fornitore di energia e il gestore della rete elettrica, diversi nuclei familiari e i loro collegamenti. Nel modello sono considerati sistemi di accumulo, contatori smart meter e tutti gli elettrodomestici che definiscono i carichi della rete elettrica. In questo modello è importante il ruolo del contatore intelligente utilizzato per registrare il consumo energetico. Esso comunica e fornisce informazioni al fornitore di energia e al gestore della rete, allo stesso tempo comunica con il sistema di gestione dell'energia domestica (home management system, HMS). Supponiamo che tutti gli apparecchi siano collegati a HMS tramite connessioni cablate o wireless. Quando la generazione supera la domanda, la batteria verrà utilizzata per risparmiare energia in eccesso. La comunità di prosumer viene modellata in questo studio come un set di utenze domestiche ognuna delle quali possiede un set di carichi elettrici. Tale studio si propone di creare un forecasting del consumo energetico del giorno successivo con l'obiettivo di minimizzare l'utilizzo di energia non rinnovabile da parte della comunità di prosumer P2P. Per fare questa ottimizzazione viene inoltre programmata la quantità di energia caricata o scaricata dalla batteria e la quantità di energia condivisa con altre famiglie in modo tale da ridurre lo squilibrio tra la produzione solare totale e la domanda di energia. Per affrontare il problema i ricercatori hanno utilizzato un modello di previsione dell'energia basato sullo strumento di ML chiamato Long Short-Term Memory (LSTM). Il modello ottenuto è stato poi confrontato con le prestazioni di un altro modello di previsione ARIMA del quale supera le prestazioni secondo i parametri di RMSE e MSE, raggiungendo una maggiore precisione e prestazioni migliori anche sul problema di time-consuming.

Uno dei problemi più difficili da affrontare rimane sempre riuscire a conciliare la produzione e il consumo di energia soprattutto nella prospettiva di una comunità energetica. Per aumentare il consumo di energia rinnovabile locale un modo è

quello di dotare le comunità energetiche di carichi elettrici controllati e batterie [29], inoltre per gestire questi dispositivi numerose strategie basate sul forecasting sono presenti [30]. Anche nello studio [31] vengono esaminate diverse tecniche di forecasting in un ambiente di comunità energetica e valutate le loro performance. In questo caso le tecniche prese in considerazione sono: regressione multipla, K-Nearest neighbours, XGBoost. Sono state valutate misure di performance come RMSE e MASE. Le previsioni della domanda energetica vengono utilizzate per gestire le risorse della comunità, quindi è importante convalidare la loro utilità sia dal punto di vista della qualità di previsione che del valore economico. Questo studio esamina la qualità delle previsioni e il valore economico derivato nel contesto di una comunità energetica reale non influenzata da specifici regimi normativi e suggerisce parametri per misurarli entrambi. La simulazione del caso studio dimostra che previsioni di qualità significativamente maggiore (ad esempio, in termini di MASE) si traducono in un valore leggermente più alto in termini di KPI per la comunità energetica (load-matching indicators [32]).

Un altro caso studio interessante [33] riguarda un quartiere residenziale che punta a diventare una comunità energetica rinnovabile nella campagna di Firenze. Anche qui la comunità è costituita sia da soli consumer che anche da prosumer. In questo paper viene analizzato il caso sotto il punto di vista tecnico del calcolo del consumo, poi supportato da quello socio-economico che ne determina la sostenibilità per i membri della comunità. Il modello delineato prevede di investigare come in una CER tramite l'utilizzo di Battery Management Systems (BMS) è possibile massimizzare l'autoconsumo collettivo. Presentando diversi scenari viene mostrato come con l'utilizzo di un BMS standard viene massimizzato l'autoconsumo del singolo prosumer mentre con un BMS smart è possibile agire ottimizzando quello collettivo portando maggior vantaggio all'intera comunità. Come energia rinnovabile anche in questo caso viene usata quella solare tramite la tecnologia fotovoltaica, la meno intrusiva come tipo di impianto. Per effettuare le stime ultime del modello è necessario il profilo di carico, ma purtroppo i misuratori installati nelle abitazioni sotto analisi sono di vecchia generazione e quindi i profili di carico non sono disponibili. Diventa così necessario generare i profili di carico usando una tecnica di forecasting. È stato sottoposto ai partecipanti della comunità un sondaggio sulle loro abitudini circa l'uso degli elettrodomestici, tali informazioni sono poi state utilizzate come input in un programma di simulazione bottom-up [34] che permetteva di simulare i carichi. Inoltre è stato utilizzato il forecast dei consumi e della produzione di energia fotovoltaica per creare uno scenario ottimale da considerare come benchmark per valutare gli altri scenari possibili e realizzabili. I risultati mostrano che installare sistemi di batterie decentralizzati in una CER e gestirli con un BMS standard è un problema perché l'autoconsumo collettivo diminuisce e questo limita il potenziale della CER per l'indipendenza energetica dalla rete a media tensione. Per risolvere il problema è necessario gestire le batterie



in base al monitoraggio dei dati in tempo reale (tramite BMS smart) o, meglio ancora, in base al forecasting (come fatto nello scenario ottimale).

Uno dei principali punti di interesse per le comunità energetiche è la previsione efficace della produzione di elettricità da fonti energetiche rinnovabili (FER), poiché è vitale per bilanciare l'offerta e la domanda di elettricità e, di conseguenza, per programmare e analizzare le reti di distribuzione e garantire una maggiore indipendenza della comunità [35]. In questo studio [36] il problema del forecasting viene spostato sulla produzione del fotovoltaico e più precisamente di una Solar Power Plant, la cui precisa previsione della produzione è fondamentale per bilanciare domanda e offerta e per programmare il funzionamento delle reti di distribuzione nel contesto di smart city e comunità energetiche. Il problema rilevante studiato in questa casistica è la situazione in cui vi è scarsità di dati o perché gli impianti sono stati installati di recente o per la mancanza di smart meter. Nel caso studio viene portata avanti l'analisi utilizzando come metodo di forecasting un approccio ML, il LSTM e utilizzando tre differenti strategie di Transfert Learning (TL) per ovviare al problema di insufficienza dei dati. L'obiettivo dello studio è proprio il confronto sull'accuratezza del forecasting con e senza l'approccio TL e con tutte e 3 le strategie di TL i modelli hanno performato meglio rispetto all'approccio non-TL.

Un innovativo algoritmo per il controllo della gestione energetica di una comunità è presentato in questo paper [37]. Il modello è in grado di coinvolgere i clienti nello scambio di energia sfruttando il loro potenziale di flessibilità energetica. L'innovazione principale si basa su un sistema di controllo basato su una matrice in cui la strategia considera simultaneamente le priorità individuali e quelle della comunità. Nel lavoro gli autori si avvalgono di un sistema di previsione basato sulla tecnica RNN, la quale esegue una previsione real-time sulla domanda energetica delle abitazioni considerate. Come in altri casi descritti in precedenza la previsione di domanda dell'energia viene fatta usando un ambiente di sviluppo basato sul Deep Learning ovvero un algoritmo di Long-Short Term Memory. Il caso studio a cui è applicato il modello prevede delle unità domestiche equipaggiate con veicoli elettrici, BESS (Battery Energy Storage System) e un impianto fotovoltaico integrato sul tetto. L'innovativo controllo della gestione dell'energia interpone il tradizionale contratto individuale con il fornitore di energia, per: affidarsi ad algoritmi predittivi riguardanti la domanda di energia e la produzione rinnovabile, massimizzare l'autoconsumo di energia rinnovabile all'interno della comunità, partecipare con eventi DR a risposta immediata, garantire un risparmio economico. I risultati numerici ottenuti come risultato dell'analisi del caso di studio sono incoraggianti e hanno evidenziato una riduzione della domanda di energia durante gli eventi di DR circa del 21%, mentre l'interazione con la rete elettrica diminuisce di circa il 15%, con notevoli benefici per la stabilità della rete.

In altri lavori come [38] sono state studiate tecniche di forecasting di lungo periodo esplorando l'applicazione dei metodi di Recurrent Neural Network (RNN)

basati sempre sul Long-Short Term Memory (LSTM) per il forecast dei consumi elettrici, la cui previsione è una componente chiave per la gestione di lungo termine delle reti elettriche. Possiamo riportare anche analisi su metodi di forecasting eseguiti su casi di strutture commerciali come un magazzino [39], dove diverse metodologie di Deep Learning (RNN, LSTM, Gated Recurrent Unit GRU) e ML (SVM, Random Forest, Extreme Gradient Boosting XGBoost) vengono messe a confronto per effettuare la previsione del consumo energetico della struttura per il giorno seguente.

I casi presentati sono solo alcuni di quelli presenti in letteratura sull'argomento delle Comunità Energetiche Rinnovabili, ma si può percepire da questi esempi che lo studio di queste configurazioni come la loro sostenibilità economica trova grande sostegno nelle metodologie di forecasting. Le considerazioni sugli studi sotto il punto di vista dei business model delle CER mettono in luce la natura che queste configurazioni hanno di un gioco di coalizione dove la funzione di profitto da massimizzare è quella della comunità a cui afferiscono i singoli membri. Per far sì che le CER abbiano maggiori sviluppi ovviamente i membri devono prevedere nel lungo periodo un NPV positivo ed è per questo che la normativa ha inserito gli incentivi sull'energia condivisa. Tutti questi processi di stima sono decisamente più efficienti se vengono svolti con l'ausilio delle tecniche di forecasting che siano sul lato del calcolo della domanda energetica o della produzione (spesso fotovoltaica). È necessario che per ogni caso studio vengano studiate le variabili a disposizione e si scelga la metodologia più adeguata in base ai dati di cui si è in possesso e dell'accuratezza della previsione che si necessita per il singolo caso.

## Capitolo 2

# Food Supply Chain & Cold Supply Chain

Per comprendere meglio il contesto in cui è delineato il caso studio oggetto di questa tesi, si approfondirà in questo capitolo l'argomento della Cold Supply Chain (CSC). Il caso studio va ad analizzare i consumi elettrici di un magazzino refrigerato. La proprietà refrigerante dell'hub logistico si rende necessaria per il tipo di filiera che l'azienda va a servire. Si sta parlando della filiera alimentare, a cui appartengono la quasi totalità dei clienti serviti dalla struttura presa in esame. Quella alimentare è sicuramente una filiera molto complessa nella sua gestione, presentando caratteristiche, criticità e peculiarità uniche. Tale filiera, partendo dalle materie prime e dalla produzione degli alimenti fino alla loro consegna ("from farm to fork"), richiede un'attenzione particolare sotto tanti aspetti diversi come igiene e sicurezza, tradotti in processi necessari e obbligatori di tracciabilità, trasparenza e coerenza normativa lungo tutta la supply chain. Questo settore viene identificato come Food Supply Chain e ingloba diversi attori, identificabili nelle numerose aziende operanti con differenti obiettivi e posizionate a differenti stadi che attraverso i processi propri della supply chain vanno così a creare questo network.

In generale, gli alimenti che necessitano di temperature controllate per il loro adeguato mantenimento si suddividono in quattro categorie secondo la temperatura di stoccaggio:

- Surgelati (frozen) mantenuti a  $-18^{\circ}\text{C}$  o meno;
- Refrigerati (cold-chilled) mantenuti tra gli  $0^{\circ}\text{C}$  e  $1^{\circ}\text{C}$ ;
- Mediamente refrigerati (medium-chilled) a  $5^{\circ}\text{C}$ ;
- Esotici refrigerati (exotic-chilled) tra i  $10^{\circ}\text{C}$  e i  $15^{\circ}\text{C}$ ;

Questi livelli di temperature sono stati stabiliti per adattarsi a differenti tipologie di prodotti alimentari [40, 41, 42].

La maggior parte dei prodotti di questa filiera è sottoposta a processi che mirano a rallentare la deteriorazione di cibi congelati o rinfrescati, per garantirne un'adeguata qualità al consumatore finale. Si identifica quindi con il nome di Cold Supply Chain, talvolta Cool Supply Chain, quella filiera produttiva e logistica che interessa prodotti principalmente agroalimentari e deperibili che necessitano di essere mantenuti a temperature controllate. Più precisamente possiamo definire una Cold Chain come un sistema ininterrotto di trasporti e stoccaggi a temperature controllate di prodotti refrigerati tra il produttore a monte e il consumatore finale, progettato per mantenere la qualità e la sicurezza di questi prodotti [43, 42, 44]. Le criticità nella filiera food sono numerose e presenti in ogni step, se ne possono citare alcune: la gestione dell'inventario che deve considerare la data di scadenza, la possibilità di contaminazioni, le temperature di conservazione, e i livelli di scorte che devono essere monitorati per ridurre i rischi di esaurimento o approvvigionamento eccessivo. Diventa per tanto necessario che gli attori attuino procedure idonee di monitoraggio, controllo, e di rilevazione delle temperature durante le attività di stoccaggio e spostamento nel rispetto degli standard imposti dalla legge [45, 46, 47, 48].

## **2.1 Spreco di cibo e di energia**

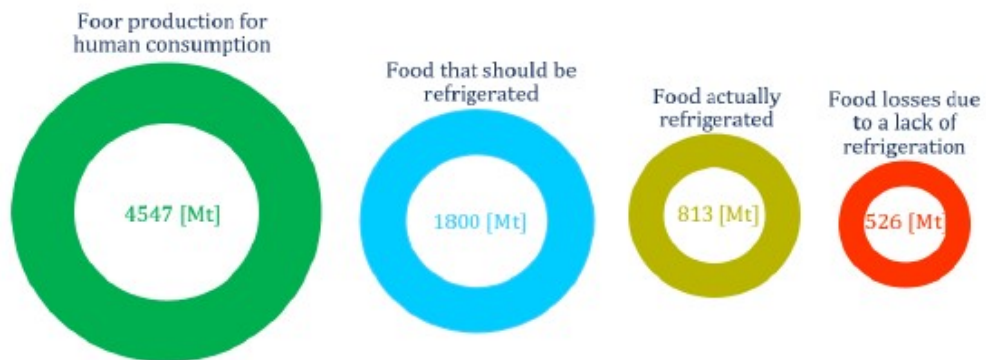
Lo spreco di cibo è una delle conseguenze negative nella gestione della Cold Supply Chain e tale fenomeno porta con sé problemi ambientali, etici e economici. Nel mondo più di un terzo del cibo (1.3 milioni di tonnellate, FAO 2011 [49]) viene perso o sprecato prima di raggiungere il consumatore finale, ciò rappresenta uno spreco di lavoro, acqua, energia, terreno e ulteriori fattori produttivi. La perdita di cibo avviene quando vi è una diminuzione di quantità e qualità degli alimenti attraverso i vari step della Food Supply Chain (FSC) che nella maggior parte dei casi sono dovute a una incorretta refrigerazione lungo la catena. Inoltre le stime fanno emergere come la carbon footprint totale del cibo sprecato genera circa 4.4 GtCO<sub>2</sub> equivalente. Facendo un paragone se queste emissioni fossero create da un Paese sarebbe il terzo più inquinante al mondo per emissioni [49].

La FSC è inoltre responsabile di circa un terzo delle emissioni globali di idrofluorocarburi (HFCs): queste sono un gruppo di sostanze chimiche industriali principalmente usate per la refrigerazione e il raffreddamento, sono dei gas serra molto potenti avendo un impatto sul riscaldamento globale di circa 3790 volte superiore rispetto all'anidride carbonica. Esempi di idrofluorocarburi sono tra i più utilizzati R134a, R1234ze, R12, R32 e numerosi miscele. Queste sostanze sono assoggette ad un parametro: il Global Warming Potential (GWP), che è una misura relativa di quanto calore intrappola nell'atmosfera una determinata massa di gas ad effetto serra. Ad oggi gli HFCs rappresentano circa 1% dei gas serra

totali ma stanno crescendo con un tasso del 10-15% annuo, dato non positivo [50, 51]. Si evince che ridurre lo spreco di cibo diventa un imperativo morale, economico e ambientale per contribuire al raggiungimento di uno sviluppo sostenibile, anche dettato dalle percentuali in crescita degli HFCs nell'atmosfera, le cui perdite potrebbero ridursi con una più accurata gestione della CSC.

Molte cause conducono allo spreco di alimenti, basti pensare a livelli di qualità inaccettabili, alimenti scartati inutilmente dai venditori retail o dai consumatori, ma gli stessi articoli deperibili vengono gettati via per la mancata consegna nei tempi previsti. In questi casi e in molti altri non citati, la causa si riscontra in diversi nodi della catena, ma sempre riconducibile alla difficoltà di mantenere una temperatura adatta e costante per i prodotti trattati, senza appunto interrompere la catena del freddo.

Secondo un documento dell'International Institute of Refrigeration (IIR) nel 2017 il 12% del cibo prodotto nel mondo è stato scartato per insufficienze nella catena del freddo. Si deve considerare che per compensare tali perdite viene messa in atto una sovrapproduzione, che porta ad innumerevoli conseguenze, basti pensare a quelle sui consumi e sulle emissioni. È stata calcolata globalmente una perdita di cibo dovuta a mancanza di giusta refrigerazione pari a 526 milioni di tonnellate (Figura 2.1). Un modello proposto dal IIR mostra come una Cold Chain migliorata (maggiormente estesa portando in tutti i Paesi del mondo il livello di performance presente nei Paesi sviluppati in termini di quantità di strumentazione refrigerante per abitante e efficienza energetica della stessa) porti a livelli di spreco del cibo molto minori riducendo tale perdita del 55% e con riduzioni significative delle emissioni di CO<sub>2</sub> equivalente fino al 47% [44].



**Figura 2.1:** Cibo perso nel mondo per mancanza di adeguata refrigerazione (2017)

La sicurezza e la qualità degli alimenti nella Food Cold Chain possono essere compromesse quando gli alimenti subiscono cambiamenti imprevisti di temperatura

e umidità o anche nell'abuso della regolazione di queste proprietà, che in ultima istanza possono portare all'incremento dei livelli di cibi sprecati. Si evidenzia quindi l'importanza nella gestione delle temperature evitandone l'abuso anche tramite i recenti miglioramenti della tecnologia di supporto. L'abuso di temperatura è definito come "una deviazione inaccettabile dalla temperatura ottimale o dal regime di temperatura ottimale rispetto ad un determinato prodotto alimentare per un certo periodo di tempo, tendo conto della temperatura dell'ambiente esterno e del tipo di attività a cui è esposto il prodotto (carico, scarico, spostamenti, ecc.)" [42]. L'identificazione dei luoghi o delle fasi nella Cold Chain dove l'abuso delle temperature avviene è cruciale per un intervento efficace [52]. Questa problematica si riscontra in tutte le fasi della Cold Chain e non è limitata a nessun prodotto particolare, principalmente si verifica nelle attività di trasporto, stoccaggio (specialmente dei venditori retail e nell'esposizione dagli stessi degli alimenti), e nella refrigerazione domestica [45, 53]. In letteratura si riscontrano diversi studi che rivolgono il focus su questo argomento e la maggior parte analizzano il contesto europeo. Si possono citare casi come:

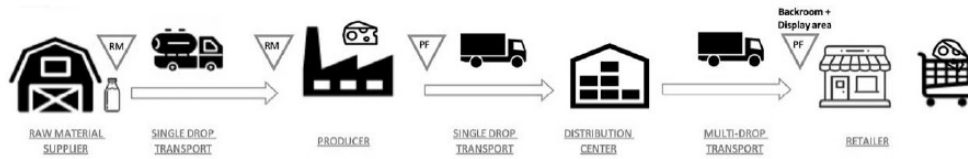
- Francia per la Food Chain di carne preconfezionata, prodotti pronti da mangiare o da cuocere e yogurt il 13.6% delle temperature nell'intera catena del freddo sono più alte di quelle raccomandate (4-6°C) [54];
- Islanda per lombi di merluzzo e filetto di eglefino rispettivamente il 17% e il 36.1% del tempo totale di due trasporti aerei la temperatura è più alta di 5°C [55];
- Finlandia per pesce, carne e prodotti pronti da mangiare il 50% delle temperature è maggiore di 1°C in un range di tempo che va tra i 249 ai 781 minuti [56];

e molti altri. Come si nota dai casi citati l'abuso della temperatura è riferito al periodo di tempo nel quale gli alimenti sono fuori dal loro range ottimale di temperatura; diventa quindi anche il tempo un fattore cruciale ed è per questo che il problema viene considerato in letteratura congiuntamente come "time-temperature abuse". Viene riscontrata inoltre una significativa discrepanza tra la temperatura effettiva dei prodotti e quella indicata dalla strumentazione di refrigerazione, arrivando fino a 6°C di differenza [42]. È stato provato che abbassando la temperatura da 7°C a 4°C per gli alimenti che necessitano di basse temperature si può estendere la conservazione e ridurre lo spreco di cibo, che per il solo Regno Unito è stato stimato un valore di 283.8 milioni di GBP (sterlina britannica) con un impatto di riduzione dell'energia valorizzato a 80.9 milioni di GBP [57]. Diventano quindi fondamentali le procedure di monitoraggio, tracciabilità e misurazione delle temperature degli alimenti lungo tutta la catena. Come supporto real-time sono utilizzabili diverse tecnologie come Radio Frequency Identification (RFID), Wireless Sensor Networks

(WSN), Time Temperature Integrators (TTIs), con anche proposte di architetture basate sull' Internet of Things (IoT) per la raccolta dati in tempo reale ad ogni punto della Food Cold Chain [58]. È stato sviluppato un sistema di monitoraggio delle temperature all'interno del framework del progetto europeo FRISBEE [59] per migliorare l'efficienza e la sostenibilità del sistema valutando simultaneamente la qualità degli alimenti a basse temperature, il consumo di energia e l'impatto della Food Cold Chain sul riscaldamento globale. Molti sono quindi i fattori che influenzano l'abuso di temperature come le pratiche degli operatori, la progettazione della strumentazione di refrigerazione, il posizionamento e l'impacchettamento degli alimenti nei container di stoccaggio. Si evidenzia anche la necessità da parte delle autorità di regolamentare tramite linee guida che considerino l'utilizzo di energia, il rischio economico, l'impatto sociale e l'ambientale il tutto con l'obiettivo di mantenere l'integrità degli alimenti nella food cold chain [42].

Il settore del Food&Beverage è considerato uno dei più grandi consumatori di energia e causa il 67% delle emissioni di gas serra del settore agro-alimentare [60]. Il maggiore impatto deriva dal consumo di elettricità, che nel mondo è stimato al 17%, al fine di refrigerare o per il condizionamento dell'aria [59]. Lunghe Supply Chain sfruttano le economie di scala che generalmente sono più efficienti, ma allo stesso tempo condizionano negativamente e con un grado maggiore l'ambiente, a causa degli stoccaggi prolungati e dei tempi di distribuzione che avvengono su scala globale [61]. Il mantenimento della qualità dei prodotti e il consumo di energia nei vari stadi della Supply Chain provoca un gigantesco trade-off: una maggiore qualità richiede che la conservazione degli alimenti deperibili venga fatta a temperature più basse e quindi con carichi di refrigerazione maggiori, allo stesso tempo il consumo energetico aumenta e con esso i conseguenti costi e con essi le emissioni di gas serra relative. In letteratura sono già stati investigati metodi di gestione della catena del freddo, ma prevalentemente è stato fatto dalla prospettiva della singola azienda che in via principale considera come ottimo il minimizzare la funzione di costo totale, in questo modo il focus sui costi porta ad una sostenibilità economica ma non necessariamente include la dimensione sociale e ambientale fortemente influenzate dal trade-off tra energia e qualità [45]. È possibile avere una visione olistica della Cold Supply Chain (Figura 2.2) considerando un modello che valuta le performance energetiche di questo tipo di catena. In questo tipo di modello di Marchi-Zanoni viene considerata la Supply Chain nel settore Food&Beverage e vengono valutate le perdite sulla qualità dei prodotti dovute ad un'inappropriata refrigerazione. L'alta applicabilità del modello in vari contesti di prodotto è data dal fatto che non vengono considerati i consumi relativi alla produzione (specifici per ogni tipologia di prodotto) ma solo sulle attività logistiche. I dati utilizzati sono stati ricavati dai vari attori della catena: magazzino del fornitore di materie prime, la materia prima e i magazzini di prodotti finiti presso il produttore, il magazzino del centro di distribuzione e il magazzino del retailer e l'area espositiva presso il rivenditore, inoltre sono

considerati tutti i trasporti fatti tra questi stage. Nel dettaglio il modello analizza



**Figura 2.2:** Rappresentazione schematica di una Cold Chain [45]

tutti i vettori energetici che contribuiscono allo scopo di refrigerare o rinfrescare: elettricità, gas naturale, diesel. Questi contributi sono stati normalizzati in modo da avere un'unica unità di misura per il consumo energetico. L'energia equivalente è espressa in kWh/kg, rispetto al prodotto analizzato, utilizzando per ogni tipo di carico un fattore di conversione dal SEC (Specific Energy Consumption) di ogni attore della catena. In diversi casi studio si nota che gli stage a monte relativi al produttore sono quelli con la SEC maggiore. In particolare si riporta il caso della supply chain del formaggio che mostra come il 71.48% dei SEC sono relativi ai consumi elettrici dei sistemi di refrigerazione mentre il restante 29% è relativo ai consumi di carburante per i veicoli. Nella frammentazione della supply chain si annovera l'incidenza dei SEC del produttore con il 33% (0.179 kWh/kg) e quella dei trasporti tra produttore e centro di distribuzione pari al 19% (0.107 kWh/kg) dei SEC totali. Per quanto riguarda l'incidenza puntuale del centro di distribuzione, considerato come magazzino intermedio di smistamento dei prodotti, occupa una quota relativamente bassa pari al 13% (0.070 kWh/kg) ma comunque non trascurabile. L'altra variabile indagata è la perdita di qualità dei prodotti, che implementata nel modello porta la SEC totale da 0.55 kWh/kg a 1.079 kWh/kg, ciò è dovuto alle risorse spese per produrre beni che verranno scartati a causa di inadeguate proprietà organolettiche e nutritive dovute alla mal conservazione. Viene riscontrato che il 43% di queste perdite avviene proprio nello stage iniziale della catena, per inefficienza nella refrigerazione delle materie prime e del prodotto finito nei magazzini del produttore [45].

Tramite il modello sono stati individuati alcuni parametri critici nella Cold Chain quali: le temperature di stoccaggio nei vari magazzini, le dimensioni dei lotti nei trasporti e i tempi di stoccaggio. Applicando le misure di efficienza energetica (EEMs, Energy Efficiency Measures) negli stadi della supply chain dove l'incidenza della SEC è maggiore, permette di ottenere migliori risultati, senza eliminare però il trade-off tra l'aumento delle temperature di stoccaggio e la diminuzione del consumo energetico con conseguente maggiore degrado qualitativo dei prodotti. Nel caso studio è stata trovata una soluzione di ottimo: abbassare la temperatura del magazzino materie prime e aumentare quella del prodotto finito (per evitare abusi di temperatura); tali misure hanno portato ad una diminuzione del consumo



totale (SEC che considera anche la degradazione dei prodotti) del 41% [45]. Attraverso il modello precedentemente illustrato è stata indagata la supply chain del pesce, mettendo a confronto la gestione della materia prima mantenuta fredda (chilled) oppure congelando il prodotto. In linea con il caso precedente il 71.4% dei SEC sono relativi ai consumi elettrici dei sistemi di refrigerazione e il 28.6% è attinente al consumo di gasolio per lo spostamento dei veicoli. Per quanto riguarda la rilevanza puntuale nelle varie fasi si nota che il prodotto freddo ha la più alta SEC presso il fornitore di materia prima, coprendo circa il 38% dei consumi totali, e a seguire il trasporto singolo al centro di distribuzione con un peso del 23%; mentre per il prodotto surgelato gli stage con la più alta SEC sono il magazzino e le vetrine dei retailer rispettivamente con il 23% e il 13% dei SEC totali e a seguire il trasporto singolo dal produttore al centro di distribuzione col 15%. Anche in questo studio viene considerato il contributo energetico della perdita di qualità principalmente riscontrata nei siti del retailer (circa 10%) e nel centro di distribuzione (4%) che fa aumentare la SEC totale da 0.669 kWh/kg a 0.781 kWh/kg; mentre grazie alla soluzione con minori temperature la qualità persa nel prodotto surgelato è trascurabile rimanendo 1.051 kWh/kg. Si ripresenta il trade-off tra mantenimento della qualità (conseguente minor spreco) e consumo di energia; possono quindi essere implementate EEMs che nello scenario del prodotto freddo si concentrano sulla riduzione dei tempi di stoccaggio a livello di materia prima con un più rapido ciclo di rifornimento, mentre nello scenario del prodotto surgelato il focus è da porre sulla diminuzione della potenza necessaria per fornire il carico di refrigerazione o ridurre i tempi di refrigerazione. In entrambi gli scenari le misure attuabili possono essere una maggiore sincronizzazione e coordinazione nell'intera catena e ad esempio tramite l'aggiustamento delle temperature di raffreddamento per evitare abusi [62].

Le EEMs in una Cold Chain possono interessare miglioramenti tecnici sulla strumentazione già in utilizzo, pratiche di manutenzione o progettazione e gestione dell'operatività come ad esempio l'aggiustamento delle temperature e il loro controllo. Le soluzioni tecnologiche solitamente sono più promettenti in termini di risparmio energetico (tra il 15% e il 40% di riduzione dei consumi) ma allo stesso tempo più costose, mentre miglioramenti nelle pratiche di manutenzione e operative hanno un costo quasi trascurabile e al contempo permettono un significativo risparmio (circa 15%) [61]. Le principali misure sono raggruppate in base all'area d'intervento:

- tecnologie ausiliarie (sistema di luci a LED, motori e sistemi di ventilazione più efficienti, ecc.);
- edifici (isolamento delle pareti, magazzini con comparti separati, isolamento del tetto e sostituzione delle finestre);
- dipendenti (addestramento del personale, accrescimento della consapevolezza, responsabilità e gestione attiva);

- generazione e recupero dell'energia (recupero del calore disperso, sistemi di stoccaggio dell'energia termica e/o elettrica);
- simbiosi industriale (cooperazione su problematiche di interesse comune, condivisione delle infrastrutture, ecc.);
- manutenzione (minimizzazione delle perdite di aria compressa, ottimizzazione dei sistemi di distribuzione del raffreddamento, pulizia regolare dei condensatori ed evaporatori, ecc.);
- management (revisione energetica, sfruttamento dell'energia ottimale);
- monitoraggio e controllo (sistemi di monitoraggio real-time, uso di sistemi di controllo automatici);
- sistemi di refrigerazione (raffreddamento libero, cicli frigoriferi alternativi es. a due stadi);
- trasporti (isolamento migliore attraverso barriere d'aria, ottimizzazione dei viaggi, Portable Refrigerated Units - PRUs).

Nella gestione delle temperature sui mezzi di trasporto la problematica principale è quella delle aperture delle porte che diventa più rilevante quando si tratta dei trasporti multi-drop dal centro di distribuzione ai retailer, per questo si necessitano misure come le PRUs o le barriere d'aria. Le misure di efficienza energetica applicabili ai magazzini sono principalmente: isolamento delle pareti, sistema di illuminazione a LED, manutenzione, monitoraggio e controllo, recupero del calore disperso, raffreddamento libero (free cooling), sistemi di refrigerazione a due stadi. Ognuna di queste misure ha grande impatto energetico ma diversa valutazione sull'investimento iniziale, perciò va valutata la fattibilità economica di ogni misura. Tuttavia se viene considerato un tempo di circa 20 anni per la maggior parte di queste misure di efficienza il risparmio cumulato è maggiore dell'investimento totale portano ad un valore attuale netto positivo [61].

Di pari passo con le EEMs vanno i così detti benefici non energetici NEBs (Non-Energy Benefits) che possono essere descritti come miglioramenti dovuti a tecnologie tecnicamente efficienti che aggiungono ulteriore beneficio ai processi produttivi. Si possono includere questi miglioramenti in aree come la riduzione di rifiuti ed emissioni, miglioramento di operatività e manutenzione, aumento della produzione, della qualità dei prodotti e dell'ambiente di lavoro, immagine pubblica e morale dei lavoratori [63, 64].

## 2.2 Normativa settore alimentare

Nelle numerose ricerche che riguardano la supply chain alimentare e le sue numerose criticità, una grande responsabilità viene addossata alla normativa, la quale deve porre le basi per le *good practices* da adottare in materia di salvaguardia della qualità degli alimenti e quindi del consumatore finale. Nel settore del Food&Beverage la normativa di riferimento principale è siglata **HACCP** codificata per la prima volta con la Direttiva 43/93/CEE e successivamente recepita in Italia con il Decreto Legislativo 155/97. Il percorso normativo ha visto in seguito abrogato quest'ultimo D.Lgs e sostituito da:

- Regolamento CE 178/2002: che ha introdotto l'obbligo del controllo su tutta la filiera alimentare e ha attribuito maggiore responsabilità al produttore quale garante della salubrità del proprio prodotto [65];
- Regolamento CE 852/2004: che contiene le norme generali in materia di igiene dei prodotti alimentari rivolte agli OSA (operatori del settore alimentare), elencando i loro obblighi relativi alla tutela del consumatore [66];
- Regolamento CE 853/2004: che indica le norme di igiene relative agli alimenti di origine animale, anch'esse rivolte agli OSA [67].

HACCP è l'acronimo di Hazard Analysis and Critical Control Points, ovvero un sistema di analisi dei rischi e punti di controllo critico, risulta come un insieme di procedure, mirate a garantire la salubrità degli alimenti, basate sulla prevenzione anziché sull'analisi del prodotto finito. Il concetto a monte di questo sistema è impuntato sulla tutela della salute pubblica piuttosto che sulla semplice soddisfazione del cliente, in particolare si sottolinea la definizione di salubrità [68] nel campo della qualità alimentare, dove per *alimenti salubri* si considerano gli alimenti idonei al consumo umano dal punto di vista igienico. Viene inoltre identificato dalla normativa [68] l'attore principale su cui cade l'incombenza di far rispettare tale sistema ed è il responsabile dell'industria alimentare (Art.2 comma e) definito come: il titolare dell'industria alimentare ovvero il responsabile specificatamente delegato. In questo frangente per *industria alimentare* si intende (Art.2 comma b) ogni soggetto pubblico o privato, con o senza fini di lucro, che esercita una o più delle seguenti attività: la preparazione, la trasformazione, la fabbricazione, il confezionamento, il deposito, il trasporto, la distribuzione, la manipolazione, la vendita o la fornitura, compresa la somministrazione, di prodotti alimentari. Il protocollo HACCP è sistematico ed ha basi scientifiche, la sua finalità è di individuare ed analizzare pericoli e mettere a punto sistemi adatti per il loro controllo e regolazione. Si fonda sul monitoraggio dei punti di lavorazione degli alimenti in cui si prospetta un pericolo di contaminazione sia di natura biologica che chimica o

fisica. Queste attività sono distribuite su tutta la filiera del processo di produzione e distribuzione alimentare, ovvero lungo tutta la Food Supply Chain. Si possono individuare 7 principi alla base del sistema HACCP [69, 66]:

1. Individuazione dei pericoli e analisi dei rischi: identificare ogni pericolo che deve essere prevenuto, eliminato o ridotto a livelli accettabili;
2. Individuazione dei punti critici di controllo (CCP): identificare i punti critici di controllo nella fase o nelle fasi in cui il controllo stesso si rivela essenziale per prevenire o eliminare un rischio o per ridurlo a livelli accettabili;
3. Definizione dei limiti critici: stabilire, nei punti critici di controllo, i limiti critici che differenziano l'accettabilità e l'inaccettabilità ai fini della prevenzione, eliminazione o riduzione dei rischi identificati;
4. Definizione delle procedure di monitoraggio: stabilire ed applicare procedure di sorveglianza efficaci nei punti critici di controllo;
5. Definizione e pianificazione delle azioni correttive: stabilire le azioni correttive da intraprendere nel caso in cui dalla sorveglianza risulti che un determinato punto critico non è sotto controllo;
6. Definizione delle procedure di verifica: stabilire le procedure, da applicare regolarmente, per verificare l'effettivo funzionamento delle misure precedentemente definite;
7. Definizione delle procedure di registrazione: predisporre documenti e registrazioni adeguati alla natura e alle dimensioni dell'impresa alimentare al fine di dimostrare l'effettiva applicazione delle misure di cui sopra.

In ogni azienda appartenente a questo settore è necessario, nonché obbligatorio, che sia presente una figura responsabile dell'elaborazione e della gestione della procedura, concretizzandola nel piano di autocontrollo alimentare (o anche manuale di autocontrollo Haccp). Tale individuo deve aver ricevuto un'adeguata formazione per l'applicazione dei principi del sistema HACCP. Per essere precisi il metodo HACCP si focalizza solamente sulle problematiche di sicurezza sanitaria dei prodotti alimentari e non sulla qualità del prodotto, anche se la maggioranza dei controlli di qualità e dei sistemi di garanzia sono basati sui principi HACCP [47].

Per soddisfare le legittime richieste di sicurezza e garanzia alimentare da parte dei consumatori, è stata concepita la norma e la relativa certificazione ISO 22000, anch'essa intesa a superare l'approccio retroattivo del controllo di qualità ed è maggiormente improntata sull'approccio preventivo. "ISO 22000 Food Safety Management Systems – Requirements for any organization in the food chain" è lo standard fondamentale per i sistemi di gestione della sicurezza del settore

agroalimentare, applicabile appunto a tutte le aziende che in tale filiera operano in modo diretto o indiretto. Il manuale dello standard ISO 22000 si basa su principi fondamentali [48] riconosciuti a livello internazionale:

- comunicazione interattiva: si tratta di un fattore innovativo e fondamentale della gestione dei rischi, che definisce un flusso di informazioni strutturate sia verso l'interno sia verso l'esterno dell'azienda, per garantire un controllo efficace dei fattori di rischio;
- gestione di sistema: permette il controllo di tutte le interazioni tra gli elementi che fanno parte del sistema, per garantire l'efficienza e l'efficacia del sistema stesso;
- prerequisiti: è l'adozione degli schemi GMP (Good Manufacturing Practice), GHP (Good Hygiene Practice), GAP (Good Agricultural Practice), di programmi e procedure di manutenzione per attrezzature ed edifici e di programmi di disinfestazione;
- principi HACCP.

La norma conferma inoltre la validità dei principi del Codex Alimentarius (raccolta di standard, linee guida e codici di condotta, redatto dalla Codex Alimentarius Commission per un programma di collaborazione tra WHO e FAO) [46] per lo sviluppo del sistema HACCP e crea un elemento di congiunzione tra i sistemi di gestione della qualità implementati in accordo alla ISO 9001 e i sistemi di autocontrollo aziendale sviluppati ai sensi del D.Lgs 193/2007 (ex D. Lgs. 155/97) e del regolamento comunitario Reg. 852/04 in vigore dal 1° gennaio 2006.

# Capitolo 3

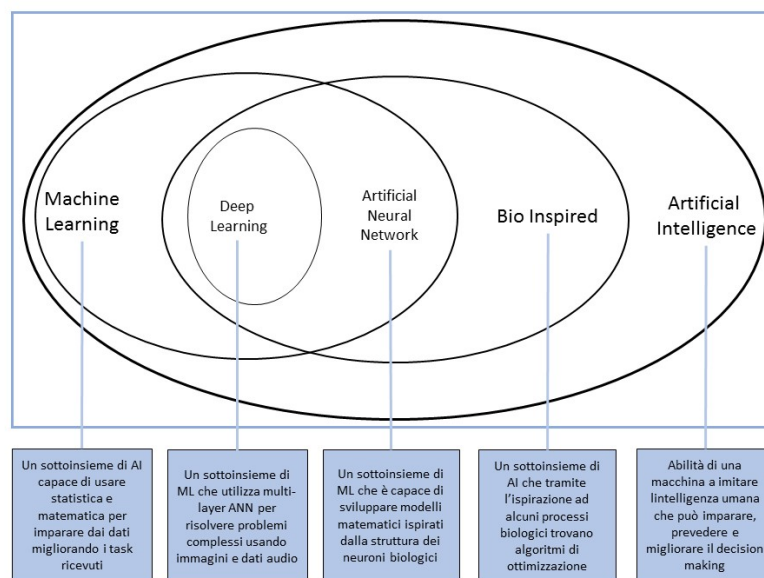
## Intelligenza Artificiale

### 3.1 Introduzione AI e al Machine Learning

L'Intelligenza Artificiale (AI) ha rivoluzionato il modo in cui oggi affrontiamo i problemi complessi nelle diverse aree dell'industria, dell'economia e della ricerca scientifica. Tra le branche più promettenti dell'AI si trovano il Machine Learning (ML) e il Deep Learning (DL), che hanno dimostrato un potenziale significativo nel trattare una vasta gamma di problemi. L'Intelligenza Artificiale (AI) si riferisce alla simulazione dell'intelligenza umana in macchine programmate per pensare e agire come esseri umani. Implica lo sviluppo di algoritmi e programmi per computer in grado di eseguire attività che in genere richiedono le capacità dell'intelligenza umana come il ragionamento, l'apprendimento, la percezione sensoriale (percezione visiva, il riconoscimento vocale, la traduzione linguistica), il processo decisionale e la risoluzione di problemi complessi. Più precisamente per "intelligenza" si intende l'abilità di imparare e risolvere problemi. L'obiettivo dell'AI quindi è quello di far sì che i computer diventino intelligenti in modo che possano acquisire tutte o alcune delle capacità precedentemente associate all'intelligenza umana. Riprendendo la semantica del termine "intelligenza", questa è la capacità di acquisire e applicare la conoscenza, che è a sua volta l'informazione acquisita attraverso l'esperienza e a sua volta l'esperienza è il sapere che si acquisisce attraverso l'esposizione (formazione). L'analisi condotta consente di poter definire l'intelligenza artificiale come una copia di qualcosa di naturale (cioè gli esseri umani) che è in grado di acquisire e applicare le informazioni apprese attraverso l'esposizione [70]. I maggiori campi di utilizzo sono per ora nei settori militare, healthcare e nel computing, ma stanno prendendo piede in molti altri come: retail, finanza, manifatturiero, trasporti, sicurezza, customer service, marketing e nell'educazione.

Si delinea molto vasto il campo dell'intelligenza artificiale, ma dandone una rappresentazione semplificata si descrivono successivamente le sue principali aree.

Possiamo considerare il Machine Learning (ML) come un sottoinsieme di metodologie di AI, in cui una macchina viene addestrata ad eseguire un compito a partire da esempi ovvero dai dati che gli vengono forniti. Un'altra branca è quella dell'intelligenza artificiale Bio-Ispirata, una serie di tecniche che si ispirano al funzionamento degli esseri viventi (degli esempi possono essere gli algoritmi genetici o algoritmi di ottimizzazione ispirati al funzionamento di una colonia di formiche o al funzionamento del sistema immunitario) ovvero ispirate alla biologia. Tramite l'intersezione di questi due insiemi possiamo rappresentare quelle tecniche che prendono il nome di Reti Neurali, proprio perché si basano sul funzionamento dei neuroni. Come sottoinsieme delle Artificial Neural Network(ANN) ci sono le tecniche di Deep Learning, che attraverso a un numero di layer sempre maggiori e quindi più profondi riescono a rappresentare e risolvere problemi sempre più astratti. La rappresentazione secondo un diagramma di Venn può essere più chiara (Figura 3.1).



**Figura 3.1:** Schematica rappresentazione dei metodi AI-based

**Machine Learning** Il Machine Learning come detto in precedenza è una branca dell'Intelligenza Artificiale che si occupa dello sviluppo di algoritmi e modelli che consentono ai computer di apprendere da dati senza essere esplicitamente programmati; cioè permette ai sistemi di acquisire conoscenza e migliorare le prestazioni nel tempo attraverso l'analisi di un insieme di dati di addestramento. Nel 1959 Arthur Samuel, un americano leader nel campo del computer gaming e

intelligenza artificiale ha coniato il termine "Machine Learning", definendolo come: "Il campo di studio che concede ai computer l'abilità di imparare senza essere esplicitamente programmati". Anche se non esiste una definizione universale di Machine Learning ve ne sono diverse che permettono di comprendere a pieno il concetto portato avanti da tale disciplina [71]:

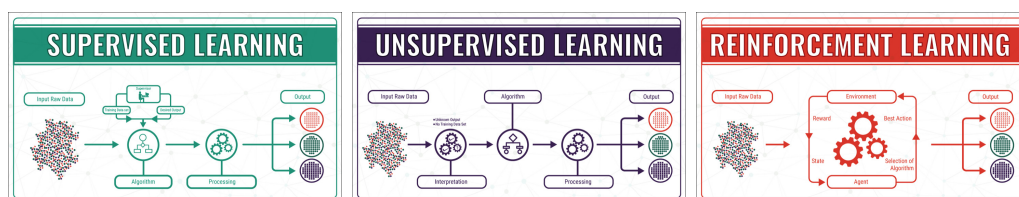
- Il Machine Learning è un modo di programmare i computer per ottimizzare un criterio di performance usando esempi di dati o passate esperienze. Si crea un modello definito da certi parametri, e l'addestramento è l'esecuzione di un programma informatico per ottimizzare i parametri del modello usando i dati di training. Il modello può essere predittivo per fornire previsioni sul futuro, o descritto per ottenere conoscenza dai dati. Il campo di studio del ML si preoccupa di creare programmi informatici che automaticamente migliorano con quella che si può definire esperienza.
- Il Machine Learning è un sottocampo dell'intelligenza artificiale che fa riferimento allo sviluppo di algoritmi e modelli statistici permettendo ai computer di migliorare le loro performance sui compiti assegnati tramite l'esperienza. Questi algoritmi e modelli sono progettati in modo addestrarsi tramite i dati e fare previsioni o prendere decisioni senza specifiche istruzioni.

Ci sono diversi tipi di approcci di Machine Learning, e si differenziano in base alla natura del segnale o risposta di addestramento disponibile al sistema che deve essere addestrato; i principali sono:

- **Supervised learning:** questo tipo di apprendimento è per la maggior parte utilizzato per problemi di classificazione e regressione. Consiste nel compito di Machine Learning di apprendere una funzione che mappa un input verso un output basandosi su degli esempi che sono coppie di input-output. Questi esempi sono forniti al computer sotto forma di dati etichettati, generalmente dei dataset formati da numerose righe (ognuna rappresentante un esempio). Il modello impara a generalizzare da questi dati per creare delle previsioni su dati nuovi e mai visti.
- **Unsupervised learning:** è un tipo di Machine Learning dove l'algoritmo (o modello) impara a riconoscere pattern e schemi nei dati che gli vengono somministrati senza che l'addestramento sia stato effettuato con esempi/dati etichettati. Lo scopo dell'addestramento non supervisionato è quello di scoprire la struttura o la distribuzione sottostante ai dati di input. Ci sono anche situazioni ibride dove per un problema si hanno un grosso numero di input ma solamente alcuni sono etichettati, quindi in questo caso possiamo parlare di semi-supervised learning. Tale tipologia è principalmente usata in problemi di clustering e di riduzione della dimensionalità dei dati.



- **Reinforcement learning:** questo tipo di apprendimento è una classe di metodi di Machine Learning in cui un agente impara a interagire con un ambiente compiendo azioni e ricevendo feedback sotto forma di ricompense o penalità basate sulle azioni intraprese. L'obiettivo dell'apprendimento rinforzato è di imparare un metodo, che è una mappatura dagli stati alle azioni, che massimizza la ricompensa cumulativa prevista nel tempo. Esempi di applicazione di questa metodologia sono la robotica o i veicoli a guida autonoma.



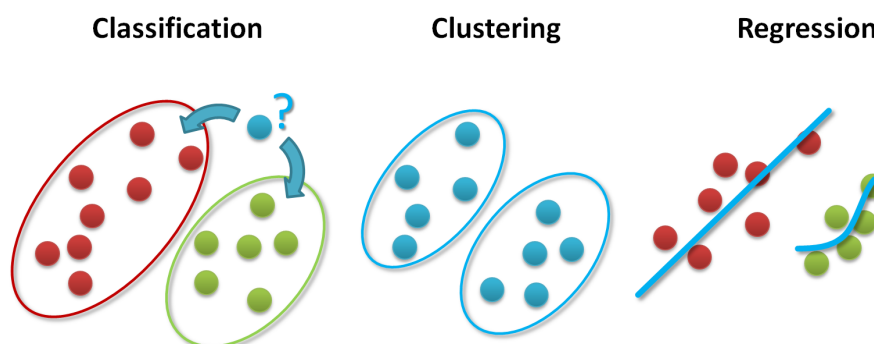
**Figura 3.2:** Classificazione dei metodi di Machine Learning[72]

Oltre alle tipologie di metodi dei sistemi di Machine Learning vi è una categorizzazione basata sul tipo di output che viene richiesto di produrre al sistema di ML e quindi il risultato desiderato [71]. Tra i principali possiamo citare:

- **Classificazione:** quando gli input vengono divisi in due o più classi, l'apprendimento consiste nel produrre un modello che assegnati degli input nuovi riesce a classificarli in quelle classi definite nella fase di training. Un esempio diffuso per spiegare questa tipologia è la selezione fatta nelle email quando in automatico le mail spam vengono incasellate nell'etichetta apposita.
- **Regressione:** in questo caso l'output richiesto al sistema è di natura continua, sia sulle variabili di input che di output. Può funzionare su modelli semplici o molteplici, avendo più di una feature, e con modelli lineari e non lineari. Verrà spiegato nel dettaglio successivamente.
- **Clustering:** quando un set di input deve essere diviso in gruppi, ma a differenza della classificazione i gruppi non sono noti a priori dal sistema, ecco perché è un problema di unsupervised learning. Il sistema deve imparare a riconoscere i dati con simili attributi e raggrupparli assieme.

Per concludere la descrizione del mondo dall'intelligenza artificiale, diamo una generica e non esaustiva spiegazione di cosa sia il Deep Learning. Il DL è una sottocategoria del Machine Learning che si basa su reti neurali. Queste ultime sono

<sup>1</sup><http://www.big-data.tips/machine-learning-methods>

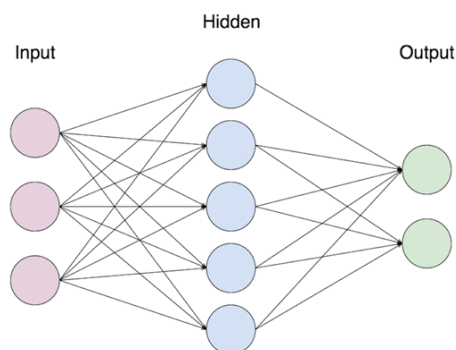


**Figura 3.3:** Rappresentazione grafica dei metodi di Classificazione, Regressione e Clustering<sup>1</sup>

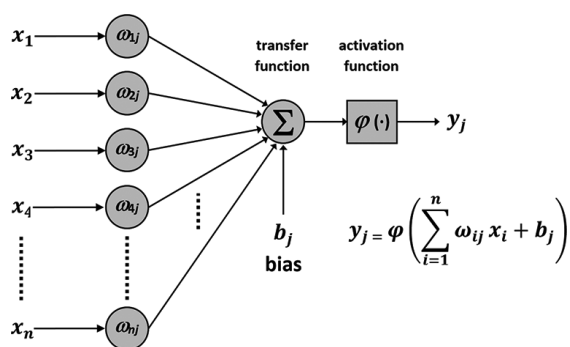
modelli computazionali ispirati dal cervello umano e composti da numerosi strati di neuroni artificiali interconnessi. Il concetto chiave del Deep Learning è l'uso di reti neurali con più strati nascosti, che consentono di estrarre automaticamente rappresentazioni complesse dai dati di input. Una Artificial Neural Network (ANN) ha come elementi principali strati, neuroni e connessioni. L'intento è quello di imitare il cervello umano che è estremamente complesso e che può essere approssimato a una vasta rete neurale comprendente quasi 100 miliardi di neuroni, ognuno dei quali è collegato ad altre migliaia di neuroni tramite le sinapsi, ciò porta a milioni di miliardi di collegamenti. Allo stesso modo le ANN sono costituite dai layer (strati) composti da un certo numero di neuroni. I layer sono di differenti tipi in base all'applicazione che se ne deve fare, ed ogni neurone è connesso a tutti i neuroni del layer adiacente. I dati sono forniti all'input layer, poi uno o più layer sono connessi all'input layer e questi sono detti "hidden layer", ed infine vi è l'output layer che può essere composto da uno o più neuroni (Figura 3.4). I neuroni interagiscono con gli altri attraverso un'unità detta Linear Threshold Unit composta da una funzione di "trasferimento" e una di "attivazione". Ad ogni neurone di ogni strato viene assegnato un peso che in base al valore assegnato da una funzione di ottimizzazione, ovvero la "funzione di attivazione", permette l'attivazione o meno di quel neurone indicando l'importanza della connessione e di conseguenza delle feature coinvolte. La "funzione di trasferimento" misura la somma di ogni neurone moltiplicato per il suo peso ( $w$ ) con l'aggiunta di un coefficiente, detto "bias" ( $b$ ), utile per la traslazione della funzione di attivazione [73]. Si propone una rappresentazione schematica in Figura 3.5, che presenta anche la formulazione di ciò precedentemente enunciato.

<sup>2</sup><https://laptrinhx.com/titanic-prediction-with-artificial-neural-network-in-r-3087367370/>

<sup>3</sup><https://www.intechopen.com/chapters/53590>



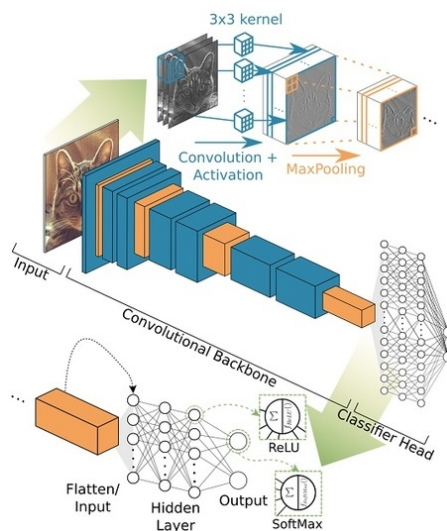
**Figura 3.4:** Schema Rete Neurale semplificato<sup>2</sup>



**Figura 3.5:** Schematizzazione del funzionamento di un singolo neurone<sup>3</sup>

La procedura tramite cui una rete neurale è in grado di fornire output dagli input, mediante pesi e bias, è chiamata "forward propagation" perché l'informazione viaggia dall'input agli strati di output del modello. Questo processo alla fine genera un costo scalare  $C(W)$ . Per far sì che si minimizzi questo costo è stato creato l'algoritmo di "backpropagation", che consente il ritorno dell'informazione di costo scalare ottenuta, calcolando la derivata della funzione di perdita rispetto ai parametri della ANN [74]. Nel Deep Learning viene quindi applicata la teoria delle reti neurali che possono essere di vario tipo come: Feedforward neural network, Convolutional neural network o Recurrent neural network. Ciò che sta alla base del DL è la numerosità di layer delle ANN che si vanno a inserire tra input e output layer per riuscire ad elaborare problemi sempre più astratti e complessi; una rappresentazione del riconoscimento di un'immagine è proposta in Figura 3.6. Il Deep Learning necessita di una numerosità di dati etichettati (supervised learning) molto elevata e una grossa potenza di calcolo da parte dell'hardware, ma allo stesso tempo può raggiungere risultati eccellenti in molte applicazioni. Il DL è particolarmente adatto per l'elaborazione di dati complessi, come immagini, testi o suoni e tale tecnologia ha ottenuto risultati notevoli in diversi ambiti come:

il riconoscimento delle immagini, la traduzione automatica, la diagnosi medica, i veicoli a guida autonoma, il rilevamento di frodi nell'ambito finanziario e nel trading automatico.



**Figura 3.6:** Schema di un algoritmo di Deep Learning che utilizza una Convolutional neural network <sup>4</sup>

In conclusione, l'Intelligenza Artificiale, il Machine Learning e il Deep Learning sono discipline in continua evoluzione che offrono un enorme potenziale per risolvere problemi complessi, compresi quelli legati alle serie temporali. Nel prossimo capitolo, esamineremo le sfide specifiche associate al forecasting sulle serie temporali e le metodologie tradizionali utilizzate per affrontarle.

**Data science** La Data Science è la scienza che si occupa di analizzare i dati e da questi estrarre insights e previsioni, utilizzando un approccio “umano”. Questa disciplina può essere vista come l'insieme di regole e processi necessari alla trasformazione di ipotesi e dati in previsioni concretamente utilizzabili [75]. Il data scientist utilizza tecniche statistiche per produrre analisi, grafici, regressioni, previsioni e tutto quello che la sua abilità riesce a ricavare dai dati a sua disposizione. In questo ambito vengono utilizzati anche algoritmi di Machine Learning per creare modelli dei dati che si cerca di analizzare. Proprio nel campo della Data Science è diffuso lo standard di processo CRISP-DM (Cross Industry Standard Process for Data Mining) le cui fasi sono: Business Understanding, Data Understanding,

<sup>4</sup><https://www.mdpi.com/2072-4292/12/10/1667#>

Data Preparation, Modeling, Evaluation, Deployment [75]. Parallelamente per un progetto di Machine Learning possiamo distinguere alcune fasi:

- *Analisi e preparazione dei dati*: in questa fase si analizza il dataset che abbiamo a disposizione e lo si prepara in modo da “adattarlo” all’algoritmo che vogliamo utilizzare. Alcuni algoritmi di machine learning richiedono che i dati siano organizzati in una maniera ben definita (ad esempio tutti dati numerici, dati normalizzati, ecc). L’approccio da avere in questa fase è quello di considerare sempre l’algoritmo che si andrà ad utilizzare: i dati, cioè, dovranno essere preparati secondo l’esigenza dell’algoritmo.
- *Apprendimento*: l’algoritmo apprende dai dati presenti nel dataset di apprendimento (training set) per elaborare un modello previsionale.
- *Valutazione*: si decide se il modello previsionale creato dall’algoritmo è accurato o meno. All’algoritmo viene fornito il test set e, se i risultati su quest’ultimo rientrano nei limiti di una percentuale di errore stabilita a priori, l’algoritmo viene promosso, altrimenti si ritorna al primo step.
- *Previsione*: il modello è sufficientemente accurato e può essere utilizzato su dati reali per creare delle predizioni sulle feature interessate e utilizzarli come fattori decisionali.

L’attività di preparazione dei dati a sua volta viene tipicamente suddivisa in tre step principali:

- *Data Collection*: in questa fase si effettua la vera e propria raccolta delle informazioni. Se è ben chiaro il problema da affrontare, stendiamo un elenco delle informazioni di cui abbiamo bisogno per modellarlo, dopodiché andiamo a caccia di queste ultime, prelevandole dalle varie fonti e mettendole insieme prima di iniziare ad elaborarle.
- *Data Preprocessing*: abbiamo a disposizione il nostro set di dati grezzo che ora va ripulito e bilanciato.
- *Feature Engineering* (data transformation): in questa fase iniziamo il lavoro di trasformazione del nostro dataset. Il risultato sarà un dataset diverso da quello originale, meno facile da interpretare ma più adatto al modello che andremo ad applicare.

**Regressione** Si parla di problemi di regressione quando la classe di output è numerica e le feature a nostra disposizione sono numeriche ed entrambe di natura continua. L’algoritmo più semplice da considerare in questa classe di metodi è la Regressione Lineare. Essa consiste in una interpolazione, appunto

lineare, ovvero creando una relazione lineare tra le variabili in input che si riduce nell'identificare una retta che se abbastanza precisa attraversa la nuvola di punti a nostra disposizione come dataset. La regressione lineare può essere semplice (eq. 3.1), ovvero dipendente da una sola variabile o più comunemente multipla (eq. 3.2). Il modello è quindi rappresentato da una funzione che restituisce la misura di  $y$  quando gli vengono fornite le  $n$  variabili dipendenti. La funzione presenta una costante detta anche bias ( $\beta_0$ ) e ad ogni variabile dipendente è assegnato un coefficiente.

$$y = \beta_0 + \beta_1 x_1 \quad (3.1)$$

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (3.2)$$

Quando la relazione tra variabili di input e output non è lineare si può ricorrere alla regressione polinomiale, dove le  $n$  variabili dipendenti possono assumere gradi anche maggiori del primo ma la formulazione è comunque simile alla precedente, dove ogni variabile in ogni grado è moltiplicata per un coefficiente sempre con l'aggiunta della costante (eq. 3.3).

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p^p \quad (3.3)$$

## 3.2 Serie temporali

Le serie temporali rappresentano un tipo di dato particolarmente rilevante in molti campi, come l'economia, la finanza, l'energia, la meteorologia e molti altri, poiché incorpora la componente temporale che permette di studiare l'evoluzione di un fenomeno nel corso del tempo. L'analisi delle serie temporali è un'utile tecnica in tutte le problematiche dove la variabile tempo ricopre un ruolo cardine nel definire la variabile dipendente o target che interessa indagare, e successivamente riuscire a prevedere. La definizione di *serie temporale* può essere data come segue: sequenza di osservazioni in un dato periodo, cioè la sequenza di valori che una data misura assume lungo l'arco temporale del nostro dataset [75]. Si tratta di un processo stocastico rispetto alla variabile aleatoria  $X_t$  con  $t$  che indica l'istante di tempo in cui la variabile viene misurata. L'obiettivo di questa tesi è quello di effettuare un forecasting o una previsione del dato osservato per capire quale valore assumerà in futuro. Le serie temporali possono essere suddivise in due grandi famiglie:

- deterministiche: appartengono a questo gruppo le serie in cui i valori possono essere esattamente determinati sulla base dei valori precedenti;
- stocastiche: appartengono a questo gruppo le serie i cui valori possono essere determinati solo parzialmente in base ai valori precedenti.

L'obiettivo di questa tesi è esplorare alcune delle metodologie per realizzare un modello di forecasting sulle serie temporali, utilizzando un approccio basato sul Machine Learning. Il forecasting sulle serie temporali consiste nel prevedere i valori futuri di una serie temporale sulla base dei dati storici osservati. Questo processo riveste un'importante ruolo in molte applicazioni pratiche; nel caso in questione la previsione è eseguita rispetto ai consumi energetici di un magazzino a temperatura controllata, o meglio della cella frigo situata all'interno del magazzino. Nella maggior parte dei casi avremo a che fare con serie stocastiche, di conseguenza sarà impossibile prevedere con certezza il valore senza assumersi una responsabilità di errore. Generalmente, infatti, possiamo indicare il valore assunto da una serie temporale come la somma di due componenti:  $f(t)$  che esprime l'evoluzione temporale del fenomeno e  $w(t)$  che è la componente randomica rappresentante tutta la serie di circostanze che non riusciamo a considerare per descrivere  $y(t)$

$$y(t) = f(t) + w(t) \quad (3.4)$$

Assumendo la parte residua di  $y(t)$  non sia spiegabile con la sola funzione  $f(t)$  la assimiliamo ad errori accidentali, che equivale ad introdurre una variabile stocastica chiamata *white noise* (variabile casuale indipendente e identicamente distribuita, di media zero e varianza costante). Generalmente nell'analisi delle serie temporali per sviluppare un modello adeguato viene fatta un'analisi delle componenti (Figura 3.7) della stessa serie che possono essere:

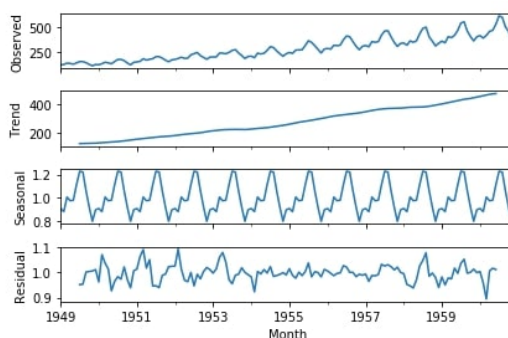
- **Trend:** tendenza generale del fenomeno considerato (es. crescente, decrescente);
- **Stagionalità:** parte relativa ai movimenti che il fenomeno assume nel periodo considerato. Generalmente dovuta a fenomeni ciclici quali clima, fattori sociali, ecc;
- **Residui:** parte più randomica della relazione tra la variabile osservata e le relative componenti.

L'analisi e la previsione delle serie temporali presenta alcune peculiarità rispetto ad altri tipi di dato, alcune delle principali sono:

- Rumore e variabilità: le serie temporali possono essere influenzate da rumore o fluttuazioni casuali che possono rendere difficile l'identificazione dei modelli sottostanti. È importante separare il segnale di interesse dal rumore per ottenere previsioni accurate.

---

<sup>5</sup><https://coderzcolumn.com/tutorials/data-science/how-to-remove-trend-and-seasonality-from-time-series-data-using-python-pandas>



**Figura 3.7:** Rappresentazione delle componenti di una serie temporale<sup>5</sup>

- Non stazionarietà: molti fenomeni nel tempo mostrano cambiamenti nel tempo, come trend o variazioni stagionali. Questa non stazionarietà può influire sulla capacità di generalizzazione dei modelli e richiede l'adeguamento delle metodologie di analisi.
- Gestione dei dati mancanti: le serie temporali spesso presentano dati mancanti o valori mancanti, che possono derivare da problemi di rilevamento, errori di registrazione o altre ragioni. È importante sviluppare strategie per affrontare i dati mancanti e gestirli in modo appropriato per l'analisi e la previsione.
- Autocorrelazione: le serie temporali possono presentare autocorrelazione, ovvero la dipendenza dei valori osservati nel tempo. L'autocorrelazione può essere positiva (i valori successivi sono correlati positivamente) o negativa (i valori successivi sono correlati negativamente). La gestione dell'autocorrelazione è cruciale per la costruzione di modelli accurati.

Esistono diverse metodologie per l'analisi e il forecasting delle serie temporali. Le metodologie possono essere divise in due categorie principali: quelle tradizionali e quelle basate su Machine Learning e Deep Learning. Le metodologie tradizionali per l'analisi delle serie temporali includono tecniche statistiche, modelli di decomposizione e metodi di smoothing. Alcuni esempi di queste metodologie sono:

- Tecniche statistiche: queste tecniche si basano sull'analisi statistica delle serie temporali, utilizzando modelli come ARIMA (Autoregressive Integrated Moving Average) e SARIMA (Seasonal ARIMA). Questi modelli incorporano la dipendenza dai dati osservati passati e l'effetto di stagionalità per fare previsioni sulle serie temporali.



- Modelli di decomposizione: questi modelli suddividono una serie temporale nelle sue componenti di trend, stagionalità e componente residua, permettendone l'analisi in modo separato con la possibilità di riscontrare l'ordine di stagionalità con le appropriate tecniche.
- Metodi di smoothing: questi metodi, come la media mobile esponenziale (Exponential Moving Average, EMA) e il filtraggio di Kalman (Kalman Filtering), sono utilizzati per ridurre il rumore e identificare i pattern sottostanti nelle serie temporali.

Le metodologie basate su Machine Learning e Deep Learning sono diventate sempre più popolari per l'analisi e il forecasting per questa tipologia di dato. Questi approcci possono gestire in modo più flessibile la complessità dei dati temporali e catturare modelli non lineari. Alcuni esempi di metodologie basate su ML e DL per le serie temporali sono:

- Reti Neurali Artificiali (ANN): le reti neurali artificiali possono essere utilizzate per modellare e prevedere le serie temporali. Le reti neurali feedforward, come il Multilayer Perceptron (MLP), possono essere utilizzate per l'apprendimento supervisionato, mentre le reti neurali ricorrenti (Recurrent Neural Networks, RNN) sono adatte per modellare la dipendenza temporale.
- Long Short-Term Memory (LSTM): le LSTM sono un tipo di RNN che possono catturare le dipendenze a lungo termine nelle serie temporali. Sono particolarmente efficaci nella modellazione delle sequenze temporali complesse con pattern a lungo termine.
- Reti Neurali Convoluzionali (CNN): le CNN, che sono state inizialmente sviluppate per l'elaborazione delle immagini, possono anche essere applicate all'analisi delle serie temporali. Possono essere utilizzate per estrarre caratteristiche rilevanti dalle sequenze temporali e per la previsione.
- Metodi di ensemble: gli approcci di ensemble combinano diversi modelli per migliorare le prestazioni di previsione. Ad esempio, il metodo di media di ensemble può combinare previsioni da diversi modelli per ottenere un'accuratezza migliore.

### 3.2.1 Modelli autoregressivi

Quando la componente di trend della serie temporale non presenta linearità e quindi è costante, siamo in presenza di una serie di dati che si definiscono *stazionari*. Entrando in gioco la stazionarietà della serie temporale possiamo nell'analisi usufruire di modelli autoregressivi, che sono i modelli più semplici per descrivere

e simulare l'andamento di una serie temporale. La teoria alla base di tali modelli è che il valore  $y$  della serie storica al tempo  $t$ , è dato dalla somma di una parte che dipende dai valori precedenti della stessa variabile nel tempo e una parte di novità  $u(t)$ , chiamata *white noise* (ovvero come una gaussiana con deviazione standard  $\sigma$ ). Si può definire un modello  $AR(p)$  di ordine  $p$  (eq. 3.5), dove l'ordine indica gli istanti di tempo precedenti presi in considerazione per stimare il valore al tempo attuale  $t$ . Il parametro  $\phi_p$  costituisce il coefficiente di regressione lineare relativo ad ogni ordine di tempo, e a seconda del valore assunto distinguiamo tre casi:

- $\phi = 1$  si dice che il processo è random walk.
- $|\phi| > 1$  si dice che il processo è esplosivo.
- $|\phi| < 1$  si dice che il processo è stazionario.

$$y(t) = \phi_1 y(t-1) + \phi_2 y(t-2) + \dots + \phi_p y(t-p) + u(t) \quad (3.5)$$

Un altro modello utile per l'analisi delle serie temporali è quello della *media mobile (moving average MA)*. A differenza del modello AR che effettua la regressione sui valori passati della serie, il modello MA (eq. 3.6) descrive la serie storica utilizzando gli errori passati come variabile. Il modello Moving Average si occupa di gestire l'errore di regressione definito come combinazione lineare degli errori che si sono verificati in passato, nella formula generica che descrive un modello  $MA(q)$  di ordine  $q$  sono presenti:

- $\mu$ : media della serie;
- $\alpha(t-i)$ : errore al periodo passato  $t-i$ ;
- $u(t)$ : termine randomico di white noise;
- $q$ : numero di periodi precedenti presi in considerazione per stimare il tempo attuale.

$$y(t) = \mu + \omega_1 \alpha(t-1) + \omega_2 \alpha(t-2) + \dots + \omega_q \alpha(t-q) + u(t) \quad (3.6)$$

Combinando i due modelli precedenti si ottiene un modello più completo il modello  $ARMA$  (*autoregressive moving average*) che viene definito in termini generali  $ARMA(p,q)$ , considerando i due ordini discussi prima  $p$  e  $q$  in modo separato, con la seguente formulazione:

$$y(t) = \sum_{i=1}^p \phi_i y(t-i) + \sum_{i=1}^q \omega_i \alpha(t-i) + u(t) \quad (3.7)$$

É possibile creare una generalizzazione maggiore utilizzando il modello *ARIMA* (*autoregressive integrated moving average*) nel caso in cui i dati abbiano una tendenza non stazionaria. La parte *integrata* indica che i valori dei dati sono stati sostituiti con la differenza tra i loro valori attuali e quelli precedenti, secondo l'operatore *lag*  $L$ :

$$L(y(t)) = y(t - 1)$$

che applicato alla equazione 3.7 restituisce il modello generale *ARIMA*( $p, d, q$ ) (eq. 3.8) con  $d$  che indica l'ordine di integrazione utilizzato per eliminare la stazionarietà<sup>6</sup>:

$$\left(1 - \sum_{i=1}^p \phi_i L(i)\right) (1 - L_d) y(t) = \left(1 + \sum_{i=1}^q \omega_i L(i)\right) u(t) \quad (3.8)$$

É importante notare che la scelta della metodologia per l'analisi e il successivo forecasting di una serie temporale dipende dalle caratteristiche specifiche della serie stessa, dalla disponibilità dei dati e dagli obiettivi di previsione. Nel prossimo capitolo, esamineremo più in dettaglio le metodologie tradizionali e quelle basate su Machine Learning e Deep Learning per il forecasting delle serie temporali.

### 3.3 Python tools

Nel vasto mondo dell'analisi dei dati, Python è diventato uno strumento fondamentale per i data scientist, gli analisti e gli ingegneri. Grazie alla sua semplicità, versatilità e ricchezza di librerie, Python si è guadagnato un posto di rilievo come uno dei linguaggi di programmazione più popolari per l'analisi e la manipolazione dei dati. Python è un linguaggio di programmazione ad alto livello, interpretato e object-oriented, che offre una sintassi semplice e leggibile. Una delle principali ragioni per cui Python è così popolare nell'analisi dei dati è la sua vasta gamma di librerie specializzate, che offrono funzionalità avanzate per la manipolazione, la visualizzazione e l'elaborazione dei dati. Per la costruzione del modello di Machine Learning presentato in questo elaborato ci si è serviti, oltre che delle strutture base del linguaggio Python [76], proprio di quelle librerie e metodi specifici di supporto per la materia ML.

Le principali librerie a supporto del modello riguardante il caso studio sono: Pandas, Numpy, Scikit-learn, Matplotlib e Statsmodels. Ognuna di queste viene di seguito brevemente descritta per introdurne le caratteristiche e funzionalità.

---

<sup>6</sup>assunto  $p=1-d$

**Pandas** Pandas è una potente libreria open source per la manipolazione e l'analisi dei dati in Python. Offre strutture dati e strumenti di analisi facili da usare e altamente efficienti, che rendono la gestione dei dati un compito semplice e intuitivo. Una delle strutture dati principali offerte da Pandas è il *DataFrame*, struttura essenzialmente tabulare bidimensionale di dati, simile a un foglio di calcolo, con assi etichettati (righe e colonne). La particolarità di questa struttura è che accetta solo dati di tipo numerico (ad esempio non stringhe). I DataFrame consentono di organizzare, filtrare, modificare e analizzare facilmente i dati in modo flessibile. Pandas offre anche molte funzioni utili per l'elaborazione dei dati, come l'aggregazione, la trasformazione e il riordino. Queste operazioni sono spesso necessarie durante la fase di preparazione dei dati. Inoltre, Pandas supporta l'importazione e l'esportazione di dati da e verso una varietà di formati, inclusi CSV, Excel, SQL e molti altri. L'analisi dei dati e il ML sono strettamente correlati, e Python, insieme alla libreria Pandas, offre una solida base per l'applicazione di algoritmi di ML. Grazie alla facilità di utilizzo di Pandas per l'importazione, la pulizia e la manipolazione dei dati, è possibile preparare i dati in modo efficiente per l'addestramento di modelli di ML [77].

**Numpy** NumPy è una libreria open-source in Python ampiamente utilizzata per la manipolazione efficiente di array multidimensionali e per l'esecuzione di operazioni matematiche su di essi. È uno strumento fondamentale nell'ecosistema di calcolo scientifico e di dati in Python ed è spesso utilizzato in combinazione con altre librerie come SciPy, Pandas e Matplotlib per analisi dati, calcoli scientifici e visualizzazioni. L'array è sicuramente la struttura più utilizzata nella libreria Numpy, e questa è definita come una tabella di elementi (solitamente numeri), tutti dello stesso tipo, indicizzati da una tupla di interi positivi. In Numpy, il numero di dimensioni dell'array è chiamato rango (*rank*) dell'array. Una tupla di numeri interi che fornisce la dimensione dell'array lungo ciascuna dimensione è nota come forma (*shape*) dell'array [78]. Alcune delle principali funzionalità di NumPy sono:

- Array multidimensionali: gli ndarray, sono un tipo di dato array multidimensionale omogeneo. Questi array possono avere qualsiasi numero di dimensioni e sono ideali per rappresentare dati tabulari, immagini, suoni ecc;
- Operazioni e funzioni matematiche avanzate: è fornita un'ampia gamma di funzioni per eseguire operazioni matematiche su array, come addizione, sottrazione, moltiplicazione, divisione, potenza e funzioni matematiche avanzate come funzioni trigonometriche, logaritmi, esponenziali, funzioni statistiche, ecc. Queste operazioni sono solitamente eseguite elemento per elemento;

- Trasposizione e manipolazione degli array: è possibile manipolare gli array, ad esempio cambiando le loro dimensioni, facendo la trasposizione, estrazione di sottoinsiemi di dati, concatenazione di array, ecc;
- Broadcasting: è consentita l'esecuzione di operazioni su array con forme diverse in modo automatico ed efficiente, estendendo automaticamente gli array più piccoli per abbinare le dimensioni dell'array più grande;
- Algebra lineare: è inclusa la funzionalità per eseguire operazioni di algebra lineare come inversione di matrici, calcolo di determinanti, risoluzione di sistemi di equazioni lineari, sviluppo in serie di Taylor, ecc;
- Generazione di numeri casuali: è presente un generatore di numeri casuali che consente di creare sequenze di numeri casuali con diverse distribuzioni;
- Efficienza computazionale: gli array NumPy sono implementati in C, il che li rende molto efficienti dal punto di vista computazionale. Le operazioni su array NumPy sono notevolmente più veloci rispetto all'utilizzo di liste Python native.

In sintesi, NumPy è una libreria essenziale per chiunque lavori con dati scientifici, ingegneristici o numerici in Python.

**Scikit-learn** In Python, l'analisi dei dati e la costruzione di modelli di Machine Learning vanno spesso di pari passo. Python offre una vasta gamma di librerie specializzate per il ML, una di queste è proprio Scikit-learn. È possibile utilizzare la libreria per selezionare il modello appropriato per il problema che si vuole risolvere. Scikit-learn offre un'ampia varietà di algoritmi di Machine Learning, tra i più noti possiamo citare: Linear Regression, Logistic Regression, Decision Trees, Support Vector Machines, Neural Network e molto altro. Nella libreria troviamo funzioni che ci semplificano tutti i passaggi da eseguire per creare un modello di ML, come ad esempio l'eliminazione dei NaN (not a number) o la divisione del dataset in train e test per i processi di addestramento e validazione e moltissimi altri. Da sottolineare sono anche le funzioni implementate nella libreria che permettono il calcolo delle metriche di errore dei modelli. Dopo aver addestrato un modello per validarlo è necessario fare il confronto tra i dati reali e i dati creati dal modello e in base al livello di errore se ne valutano le performance. Ci sono diverse metriche che permettono di valutare i diversi modelli di ML e Scikit-learn presenta sicuramente i principali: accuracy score, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R2 score. Quelli presentati sono solamente alcune delle metriche che offre la libreria [79]. Nel caso studio Scikit-learn è stata utilizzata per il calcolo delle metriche di errore e delle funzioni build-in al suo interno per tale scopo.

**Matplotlib** Matplotlib è una straordinaria libreria di visualizzazione dati in Python. È una libreria di visualizzazione dati multiplatforma costruita su array NumPy e progettata per funzionare con lo stack SciPy più ampio, e offre una vasta gamma di funzionalità per creare grafici e visualizzazioni di dati di alta qualità in modo flessibile e personalizzabile. Uno dei maggiori vantaggi della visualizzazione è che consente l'accesso visivo a enormi quantità di dati in immagini facilmente gestibili. Matplotlib permette di produrre diverse tipologie di grafici come quelli a linee, a barre, a dispersione, istogrammi ecc. [80]. In definitiva, Matplotlib è una libreria fondamentale per la creazione di grafici e visualizzazioni dati in Python. La sua flessibilità, personalizzazione e varietà di opzioni di visualizzazione lo rendono uno strumento ideale per rappresentare in modo efficace dati complessi in una vasta gamma di contesti, dall'analisi dei dati scientifici alla creazione di grafici per la comunicazione e la presentazione.

**Statsmodels** Statsmodels è una libreria open-source in Python utilizzata principalmente per la modellazione statistica e l'analisi dei dati. Questa libreria fornisce un ampio set di strumenti per eseguire analisi statistica, stimare modelli, eseguire test ipotetici e condurre diverse tecniche di regressione. Ecco una descrizione più dettagliata delle sue funzionalità e aree di utilizzo:

- **Analisi statistica:** offre una vasta gamma di strumenti statistici per esplorare i dati e condurre analisi descrittive, come calcolo di medie, deviazioni standard, percentili, correlazioni e altro ancora;
- **Regressione lineare:** è possibile eseguire analisi di regressione lineare, che consente di modellare le relazioni tra variabili indipendenti e dipendenti. Statsmodels offre funzionalità avanzate per valutare l'adeguatezza del modello e ottenere stime dei parametri del modello;
- **Regressione logistica:** La libreria supporta anche la regressione logistica, che è utilizzata per modellare variabili binarie o categoriche dipendenti;
- **Modelli di serie temporali:** Statsmodels offre una vasta gamma di strumenti per l'analisi delle serie temporali, compreso il supporto per la modellazione ARIMA (AutoRegressive Integrated Moving Average) e modelli SARIMA (Seasonal ARIMA);
- **Analisi di varianza (ANOVA):** è possibile eseguire analisi di varianza per confrontare le medie di più gruppi e determinare se ci sono differenze significative tra di essi;
- **Test statistici e inferenza:** La libreria fornisce una serie di test statistici per eseguire test di ipotesi, test di normalità, test di omoschedasticità e altro ancora;

- Grafici di supporto: vi sono presenti strumenti per la creazione di grafici come diagrammi di dispersione, grafici di residui e grafici Q-Q plot per valutare i risultati dei modelli;
- Integrazione con altre librerie: Statsmodels può essere utilizzato in combinazione con altre librerie Python come NumPy e Pandas per la manipolazione dei dati e con Matplotlib per la visualizzazione dei risultati;
- Documentazione completa: La libreria è ben documentata con tutorial, esempi e una guida dettagliata che aiutano gli utenti a comprenderne l'uso e le funzionalità [81].

In particolare per la creazione del modello di forecasting oggetto del presente elaborato si è sfruttato principalmente la parte della libreria relativa ai modelli autoregressivi ARIMA e SARIMA, rispettivamente tramite l'utilizzo delle classi *sm.tsa.arima.model* e *sm.tsa.SARIMAX*. Nelle varie analisi per decretare l'ordine del modello ci si è avvalsi del test d'ipotesi ADF (Augmented Dickey-Fuller) e dei grafici di autocorrelazione (ACF) e autocorrelazione parziale (PACF) forniti dalla libreria. Inoltre è stata di aiuto anche la funzione *auto\_arima*, fornita però dalla libreria *pmdarima*, ed è stata utilizzata per la selezione automatica dei migliori parametri per il modello ARIMA. Da evidenziare anche l'utilizzo del metodo *get\_forecast* per la creazione dei valori nell'intervallo che si è deciso di prevedere.

# Capitolo 4

## Caso studio

### 4.1 Magazzino logistico refrigerato

Il caso studio che verrà trattato in questa tesi riguarda l'analisi dei consumi elettrici e delle variabili gestionali relative ad un third part provider nel settore Food&Beverage. L'attività aziendale presa in considerazione è una realtà che lavora nel campo dei trasporti e della logistica. Oltre a gestire il trasporto nazionale offre servizi di deposito, logistica integrata, operazioni doganali e fiscali e anche lavorazioni. Le parti più incisive del business sono quelle relative al trasporto e al deposito, infatti l'azienda è dotata di una flotta con un ampio parco mezzi per ogni esigenza dei clienti (autotreni, bilici, motrici, furgoni). Lavorando nel settore del Food&Beverage l'azienda si è dotata di mezzi refrigerati, multi-temperatura e isothermici per soddisfare i requisiti necessari a fornire una corretta continuità lungo la Food Supply Chain soprattutto quando interviene la Cold Chain. Per quanto riguarda il servizio di deposito l'azienda possiede circa 10 mila mq di magazzino di cui 5 mila mq a temperatura controllata. Le temperature delle celle sono costantemente monitorate tramite sensori collegati ad un sistema di controllo che mette in funzione il raffreddamento superate le soglie di temperatura dei locali. Tutta la struttura è dotata della strumentazione idonea al rispetto della normativa HACCP e anche il personale applica tutte le best practice relative a tale normativa, inoltre è da sottolineare come i prodotti gestiti per i vari committenti sono sempre confezionati (mai gestito il prodotto "nudo").

Per quanto riguarda il supporto dato alle operations dal sistema informativo aziendale è sicuramente completo ma al tempo stesso frammentato. È presente un sistema ERP (Enterprise Resource Planning), software gestionale che si occupa della gestione dei dati di tutte le funzioni aziendali comprendendo quelli del magazzino ai fini commerciali. Per la gestione dei dati riguardanti il trasporto o il deposito della merce vengono utilizzati due sistemi separati ma con la possibilità di comunicare tra loro: TMS (Transport Management System) e WMS (Warehouse Management

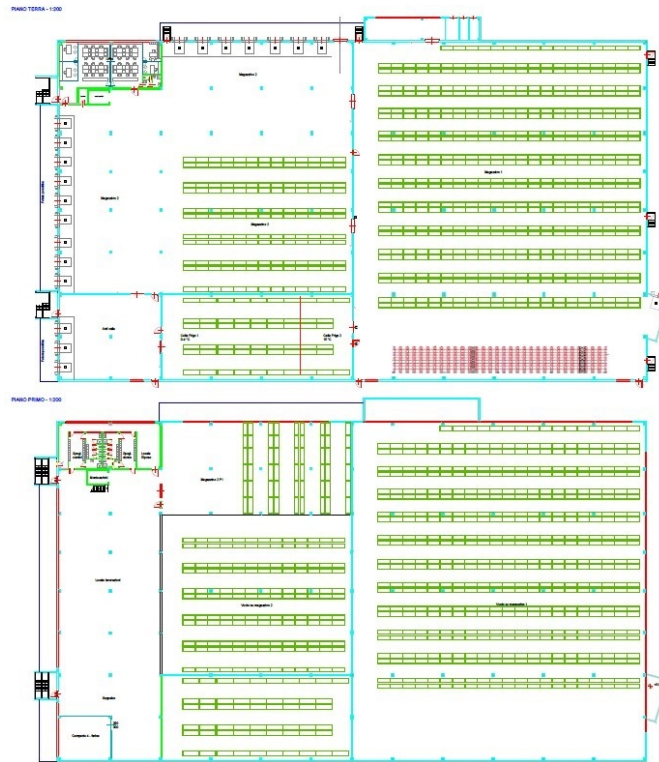


System). L'obiettivo primario di un magazzino è gestire la movimentazione e lo stoccaggio delle merci nel modo più efficiente possibile. Un WMS è un software progettato per aiutare a ridurre i costi attraverso processi di magazzino efficaci. Si rivolge alle aziende che hanno bisogno di ricevere e spedire merci, pur mantenendo un utilizzo ottimale dello spazio e conoscendo in modo specifico dove sono immagazzinate tutte le merci in un dato momento. Questi software possono essere anche integrati nel sistema ERP dell'azienda insieme a scanner e stampanti di codici a barre, apparecchiature di etichettatura. Il WMS permette l'acquisizione dei dati in tempo reale e l'automazione del magazzino. In questo modo le attività di magazzino comuni possono essere tutte ottimizzate per risparmiare tempo e per creare maggiori profitti [82]. Uno dei principali compiti del WMS è assicurare l'evasione dei prodotti secondo i criteri desiderati dall'azienda, solitamente i criteri più utilizzati sono il FIFO (First In First Out) o il LIFO (Last In First Out). Grazie alla capacità del software di gestire e tenere traccia delle informazioni del lotto o del pezzo singolo è possibile adempiere in modo semplice all'obbligo che hanno gli attori della Food Supply Chain riguardo la tracciabilità dei prodotti alimentari. Come spiegato nel capitolo 2 in questo settore la data di scadenza è di fondamentale importanza. Ecco perché nel nostro caso parlando di un magazzino refrigerato, in cui vengono stoccati prodotti alimentari, la politica di evasione è principalmente quella FEFO (First Expiration First Out), proprio grazie al WMS è possibile una gestione efficiente anche di questo requisito. È tramite tutti questi sistemi che avremo la possibilità di accedere ai dati gestionali che saranno utili per costruire il modello previsionale presentato di seguito.

Come possiamo osservare in Figura 4.1 viene riportata la disposizione degli ambienti all'interno dell'intero capannone e vengono indicate le temperature alle quali sono soggette. In particolare la struttura è composta da due piani: il piano terra dove sono locati i magazzini e le celle a temperatura controllata e il primo piano con un locale lavorazioni e una parte di magazzino.

La distinzione dei locali è la seguente:

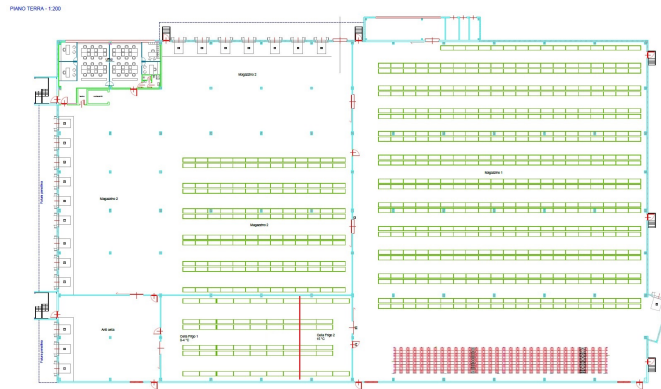
- **Cella frigo 1:** situata al piano terra (600 mq). Temperatura impostata a circa  $+2/3^{\circ}\text{C}$  (range normativo  $0 - +4^{\circ}\text{C}$ );
- **Cella frigo 2:** situata al piano terra (200 mq). Temperatura impostata a circa  $+16^{\circ}\text{C}$  (è stata inserita in un secondo tempo e spilla aria fredda dalla cella 1 per il mantenimento dei  $+18$  in estate);
- **Anticella:** situata al piano terra (500 mq). Temperatura impostata a circa  $+8/10^{\circ}\text{C}$  (ma utilizzata soltanto in determinate fasce orarie);
- **Locale lavorazioni:** situato al primo piano. Temperatura impostata intorno ai  $+24^{\circ}\text{C}$ . Il raffreddamento in questo locale è utilizzato esclusivamente nel funzionamento estivo o eventualmente durante i cicli di deumidificazione;



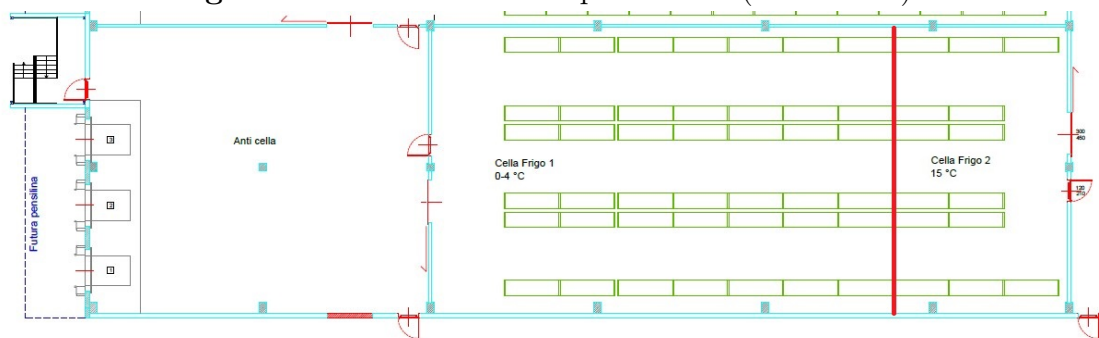
**Figura 4.1:** Planimetria del magazzino (scala 1:200): piano terra in alto, primo piano in basso

- **Magazzino 1:** situato al piano terra (5000 mq). Non prevede una temperatura controllata quindi si definisce a temperatura ambiente;
- **Magazzino 2:** situato al piano terra (3500 mq). Temperatura impostata a +18/20°C. Il raffreddamento è utilizzato esclusivamente nel funzionamento estivo.

Per quanto riguarda l'organizzazione interna dello stoccaggio nei diversi locali della struttura è presente un unico tipo di scaffalatura che è quella di tipologia portapallet (Figura 4.4). Queste tipologie di scaffalature sono le migliori per organizzare la merce gestita su europallet per l'efficienza dimensionale; infatti le dimensioni degli EPAL sono 80x120 cm e una scaffalatura tipo portapallet ha una larghezza di 240 cm per contenere esattamente 3 EPAL. Tale scaffalatura non viene sfruttata in modo efficace quando deve gestire i pallet con standard USA che hanno dimensioni 100x120 cm; con questo standard vi è la possibilità di stoccare sullo scaffale solo 2 pallet e si vanno a perdere 40 cm di larghezza che diventano inutilizzati.



**Figura 4.2:** Planimetria del piano terra (scala 1:200)



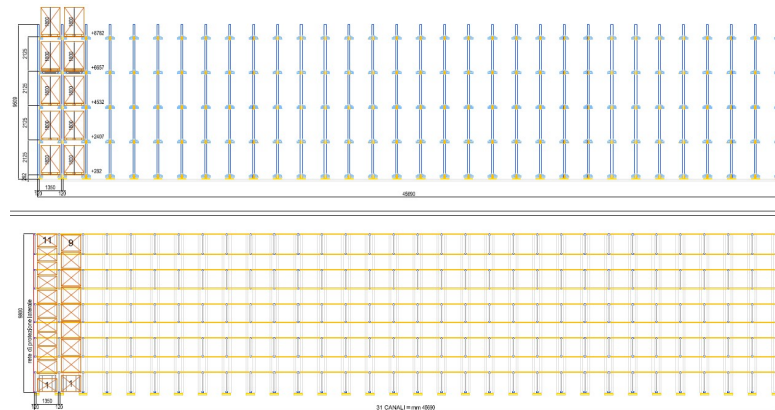
**Figura 4.3:** Planimetria delle celle frigo a anticella (scala 1:200)

Come si può notare dalla planimetria del piano terra (Figura 4.2), nella parte inferiore destra della pianta, che ancora comprende il Magazzino 1, è illustrata una scaffalatura rossa rappresentante la parte di magazzino automatizzata. Questo è un sistema integrato per un impianto di scaffalatura multi-profondità. Il sistema di stoccaggio compatto con UPC (Under Pallet Carrier) è composto da un carrier (o shuttle Figura 4.4), un carrello elevatore e i moduli di scaffalatura a canale. L'UPC è particolarmente efficace nel caso di ripetuto svuotamento/riempimento nello stesso canale. Per il trasferimento in un canale, il carrello di servizio preleva l'UPC sollevandolo con le proprie forche. Lì, transita in modo autonomo al di sotto dei pallet stoccati senza essere collegato con il carrello di servizio. Dopo il deposito del primo pallet sulla piattaforma UPC nel canale e dopo aver premuto il pulsante di avvio, lo shuttle effettua automaticamente tutti i movimenti di traslazione e di sollevamento. Grazie a sensori, in grado di riconoscere la posizione dei pallet stoccati, non c'è pericolo di collisione nel canale durante le fasi di stoccaggio e di prelievo. Un'altra delle caratteristiche di efficienza di questo sistema è la gestione ottimizzata sui diversi tipi di pallet (sia i 80x120 cm che i 100x120 cm) per una massima adattabilità alle varie tipologie di impiego; il carrier è munito di sensori

per il riconoscimento delle dimensioni dei pallet. In ogni canale vi è la possibilità di stoccare 11 europallet o 9 pallet USA (Figura 4.5). Il dettaglio dei locali di



**Figura 4.4:** Esempio di scaffalatura portapallet e UPC



**Figura 4.5:** Dimensioni e struttura magazzino automatico

stoccaggio con la relativa organizzazione delle scaffalature è la seguente:

- Magazzino 1: 23 scaffalature con 8 livelli ognuna, 12 corridoi intra-scaffale;
- Magazzino automatico (in magazzino 1): 31 canali con 5 livelli ognuno, un solo corridoio frontale alla struttura per l'accesso dei carrelli al carico e scarico;
- Magazzino 2: 11 scaffalature con 8 livelli ognuna, 6 corridoi intra-scaffale;
- Cella Frigo 1: 6 scaffalature con 5 livelli ognuna, 4 corridoi intra-scaffale;
- Cella Frigo 2: 6 scaffalature con 5 livelli ognuna, 4 corridoi intra-scaffale;
- Anticella: funzionalità di carico e scarico non adibita a stoccaggio in scaffalatura.

Lo studio effettuato sulla struttura appena descritta ha come obiettivo di andare prima ad analizzare e in seguito a prevedere i consumi elettrici della *Cella frigo 1*. Il raffreddamento di questi locali è demandato ad una macchina frigorifera di grandi dimensioni che viene definita *Gruppo Frigo*. Perciò per lo sviluppo del modello è stato necessario reperire i dati di consumo elettrico (kWh) e potenza attiva (kW) del gruppo frigo per tutto il periodo di studio.

**Flussi termici** Il caso studio prevede l'analisi dei consumi relativi alla cella frigo 1 (0-4°C) e all'anticella, per tutto il periodo preso in considerazione dall'analisi che si vuole effettuare; diventa quindi importante riuscire ad isolare il solo consumo attinente a. Durante le varie ispezioni eseguite nella struttura è stata studiata una caratteristica importante del sistema climatizzato di raffreddamento. Tale caratteristica è rappresentata dai flussi di raffreddamento che sono imputati totalmente all'unico gruppo frigo presente nella struttura; questo vuol dire che tutti i locali che vengono in qualche modo raffrescati o refrigerati vanno a richiedere potenza frigorifera all'unico gruppo frigo di cui si sono misurati e collezionati i dati di consumo. Tuttavia non sono disponibili informazioni su come è distribuita in ogni locale la domanda di raffrescamento e refrigerazione. Questo a causa della mancanza di una strumentazione di misura dedicata. Nella fattispecie si dovrebbero installare dei misuratori che determinano la potenza termica fornita, attraverso la rilevazione della temperatura di mandata, di ritorno e della portata dell'acqua refrigerata all'interno degli scambiatori che raffrescano i locali indicati. Ovviamente tale strumentazione non è necessaria in un tale impianto, ma per conoscere la parzializzazione dei consumi del gruppo frigo rispetto a ogni locale e di conseguenza per rispettare i vincoli dell'analisi del caso studio sarebbe stata molto utile. In particolare i locali direttamente condizionati sono fondamentalmente quattro: la cella frigo 1, la cella frigo 2, l'anticella e il magazzino 2. Il magazzino 2 e la cella frigo 2 vengono raffrescate solamente nel periodo estivo, ma tale periodo non è codificato tramite delle date precise bensì è suscettibile alla temperatura esterna, per cui quando quest'ultima ha un sostanziale incremento il regime "estivo" (mediamente nel mese di maggio) entra in azione. Più precisamente per il raffrescamento la cella frigo 2 spilla aria fredda dalla cella frigo 1 tramite una ventola, mentre il magazzino 2 è fornito di condotti di aerazione a soffitto i quali raffrescano il voluminoso ambiente tramite due scambiatori (Figura 4.6) a cui si dovrebbero applicare gli strumenti di misura precedentemente descritti. All'interno della struttura si riscontra che i principali elementi di dissipazione dei flussi termici sono le baie di carico, che permettono lo scambio termico diretto con l'ambiente esterno, importante soprattutto quando entra in gioco il periodo estivo e il delta di temperatura tra interno ed esterno è di 10/15°C. L'anticella presenta 3 baie di carico mentre il magazzino 2 è dotato di 16 baie di carico e per ciò questi ultimi locali rispetto agli altri sono maggiormente penalizzati per la dissipazione



**Figura 4.6:** Condotti e scambiatori di condizionamento nel magazzino 2

dei flussi termici di aria raffrescata e refrigerata, in maniera più marcata lo è il magazzino 2 che presenta una volumetria decisamente maggiore. Inoltre è da sottolineare come tra il magazzino 1 (formalmente a temperatura ambiente) e il magazzino 2 (raffrescato a 18/20°C) ci sono 2 portelloni scorrevoli che vengono azionati tramite sensori di movimento per agevolare gli spostamenti di merce con i carrelli. Questa comunicazione tra i due locali a differenti temperature complica ulteriormente il quadro dei flussi termici perché aprendosi innumerevoli volte durante le ore lavorative questi portelloni diventano un elemento di dissipazione dei flussi termici di raffrescamento. Non vi è stata una rilevazione strumentale ma con la sola percezione sensoriale durante i sopralluoghi si può affermare che anche il magazzino 1 sia in un certo modo condizionato e non a temperatura ambiente proprio per questo effetto di dissipazione. Per ridurre tale effetto dissipativo si potrebbe pensare di installare dei portelloni taglia aria sia dalle baie di carico che nei punti comunicanti tra i magazzini 1 e 2 in modo che gli scambi termici con l'esterno e tra i locali si riducano in modo considerevole. Per quanto riguarda le celle frigo 1 e 2 sono fornite di pannelli termoisolanti e non hanno sbocchi diretti con l'ambiente esterno. Il calcolo dei consumi inerenti alla sola cella frigo 1 diventa ancora più difficoltosa per la presenza della ventola che mette in comunicazione i flussi termici con la cella frigo 2, che dovrebbero essere misurati per avere la giusta parzializzazione dei consumi.

#### 4.1.1 Gruppo frigo

Quando si tratta di raffreddare edifici o impianti industriali generalmente ci si avvale di strumentazione quali i Chiller, che si configurano come i dispositivi tra i più efficienti per lo smaltimento del calore e con la possibilità di raggiungere temperature anche molto basse del fluido in uscita. Possiamo parlare indistintamente di Chiller, Refrigeratore o Gruppo Frigorifero (o frigo) per indicare questa tipologia di macchina

termodinamica. In base al meccanismo di raffreddamento possiamo identificare due tipologie di refrigeratori [83]:

- *a compressione*: viene utilizzato il lavoro meccanico del compressore che comprimendo il fluido refrigerante permette lo scambio termico nel ciclo all'interno della macchina. I componenti principali di questo di questo sistema sono: il compressore, il condensatore, la valvola di laminazione e l'evaporatore [84];
- *ad assorbimento*: sfrutta un principio fisico-chimico utilizzando come refrigerante una miscela di acqua e sale. L'evaporazione dell'acqua provoca la rimozione del calore dal fluido di processo. L'acqua stessa viene poi assorbita dalla soluzione di sale. Successivamente, la soluzione di sale viene riscaldata per liberare l'acqua evaporata, che poi viene condensata e riutilizzata nel ciclo di raffreddamento.

Nei vari sistemi possiamo incontrare due tipi di condensatore:

- raffreddato ad aria: la circolazione del fluido refrigerante in tubi alettati permette l'estrazione del calore che viene trasferito all'aria circolante all'interno del condensatore, che diventata calda semplicemente viene aspirata da una ventola e dissipata nell'ambiente;
- raffreddato ad acqua: il mezzo di dissipazione è la fonte idrica, che poi viene inviata ad una tecnologia di raffreddamento (come una torre di raffreddamento, esempio della Figura 4.7 a solo scopo illustrativo che non rappresenta la tecnologia del caso studio).

In questa sede ci interessa fornire una panoramica di come avvenga in questo tipo di dispositivo il ciclo frigorifero a compressione per capire dove viene generata la richiesta di consumo elettrico del gruppo frigo. La Figura 4.7 mostra in modo schematico il funzionamento di un chiller a compressione ed è rappresentato il ciclo termodinamico completo.

In particolare la macchina di cui tratteremo i consumi è un impianto frigorifero per raffreddamento a fluido indiretto miscela incongelabile acqua antigelo a vaso aperto funzionante a refrigerante propano (R290). Per ridurre i punti critici ed eliminare i fermi macchina è stato scelto di costruire una macchina con 2 circuiti frigoriferi completamente distinti e indipendenti in modo da garantire almeno la metà della potenza in caso di guasto a qualsiasi componente di uno dei 2 circuiti. L'impianto si divide in 2 sezioni fisicamente distinte:

- sezione 1: formata da 2 circuiti separati e viene definita lato primario (circuito del fluido refrigerante) perché è la sezione che contiene il gas refrigerante e svolge il ciclo frigorifero;

- sezione 2: formata da un unico circuito e viene definita lato secondario (circuito del fluido di processo o da raffreddare) perché è la sezione che contiene soltanto una soluzione di acqua e antigelo svolgendo la funzione di fluido indiretto per il trasporto del "freddo" alle utenze.

**Modalità di funzionamento lato primario** Tale insieme a pressione si configura come un impianto frigorifero ad espansione diretta utilizzato per il raffreddamento di miscela incongelabile. All'interno del compressore, il fluido refrigerante in fase di vapore freddo a bassa pressione viene compresso ad alta pressione e alta temperatura (minimo  $+60^{\circ}\text{C}$ , massima  $+110^{\circ}\text{C}$ ) portandolo alla temperatura di saturazione per permetterne la condensazione. Da qui, in fase totalmente gassosa ed in alta pressione viene inviato al condensatore, dove per scambio termico con aria ambiente rilascia il suo calore latente di condensazione e viene totalmente liquefatto. Il fluido in fase liquida ad alta pressione ed alta temperatura ( $+30/+50^{\circ}\text{C}$ ) defluisce ed entra nello scambiatore aspirazione/liquida. Quest'ultimo viene utilizzato per effettuare uno scambio termico atto a sottoraffreddare il fluido liquido ad alta temperatura in uscita dal condensatore per renderlo più stabile e allo stesso tempo surriscaldare il vapore freddo in uscita dall'evaporatore per garantire l'assenza di tracce di liquido in aspirazione al compressore. Attraverso le tubazioni il fluido in fase di liquido sottoraffreddato viene fatto passare prima all'interno di un filtro disidratatore e di uno scambiatore economizzatore, per depurare sottoraffreddare ulteriormente e stabilizzare il fluido liquido. Il fluido in fase di liquido sottoraffreddato ad alta pressione attraversa la valvola solenoide e viene mandato alla valvola di laminazione che regolando la sua apertura in base alla temperatura rilevata a valle dell'evaporatore lascia fluire la giusta quantità di fluido verso l'evaporatore operante in bassa pressione. Nell'evaporatore il fluido evaporando assorbe dal fluido secondario (miscela acqua/glicole) il calore latente di evaporazione consentendo di raggiungere la temperatura voluta alle utenze. Il fluido ora in fase di vapore freddo in bassa pressione viene mandato allo scambiatore aspirazione/liquida. A questo punto il fluido in fase di vapore freddo surriscaldato viene ricondotto al compressore per iniziare nuovamente il suo ciclo.

**Modalità di funzionamento lato secondario** Il lato secondario dell'impianto ha un funzionamento molto più semplice del lato primario sopra descritto in quanto ha la sola funzione di trasportare il fluido secondario (di processo) gelido dall'impianto frigorifero verso le utenze posizionate all'interno dello stabilimento, questo permette come descritto nelle pagine precedenti di ridurre il fattore di pericolosità dell'impianto frigorifero (rischio incendio). Il lato secondario è composto essenzialmente da:

- una vasca d'accumulo, per l'accumulo della miscela acqua/antigelo;

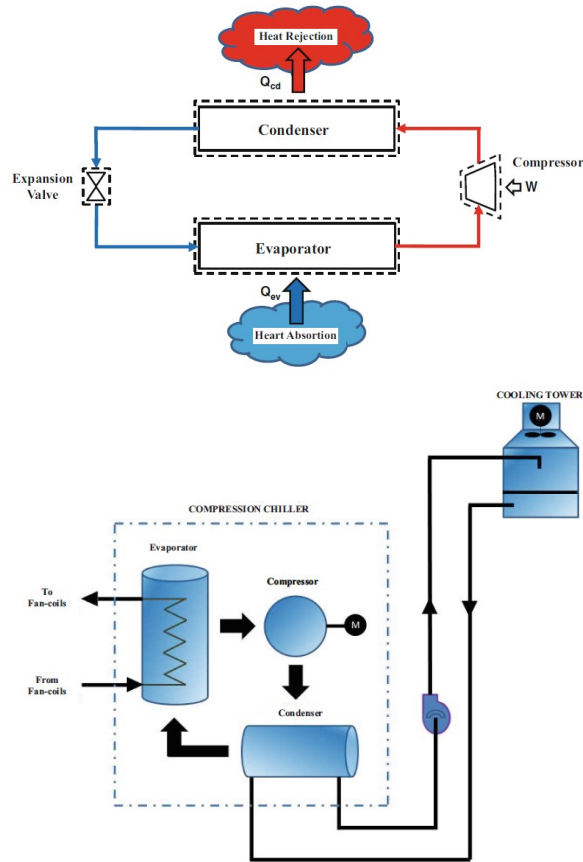


- 2 pompe centrifughe denominate pompe evaporatore;
- 2 pompe centrifughe denominate pompe utenze;
- serie di valvole e filtri;
- tubazioni di vario diametro.

Le 2 pompe evaporatore aspirano miscela acqua/antigelo dalla vasca d'accumulo che viene pompata all'interno di 2 tubazioni in acciaio inox, qui 2 filtri a Y provvedono a pulire la miscela acqua/antigelo da eventuali sporchie prima di entrare nei 2 evaporatori. All'interno degli evaporatori, l'evaporazione del gas refrigerante del lato primario, raffredda la miscela acqua/antigelo che ritorna poi nella vasca d'accumulo tramite un'unica tubazione di grosso diametro. La miscela acqua/antigelo gelida viene poi aspirata dalle 2 pompe utenze e pompata all'interno di un collettore e successivamente in una tubazione in acciaio inox collegata all'impianto idraulico dello stabilimento. La miscela acqua/antigelo una volta uscita dalle utenze ad una temperatura più calda torna alla vasca d'accumulo.

La giusta attenzione va posta sulla scelta del fluido refrigerante da utilizzare all'interno di un ciclo frigorifero per motivi sia di sicurezza dell'impianto che per questioni ambientali e climatiche. Un refrigerante spesso utilizzato nei chiller è il gas R410A (una miscela di due gas R32 e R125) energeticamente efficiente ma anche molto sicuro in caso di perdite, in quanto non tossico e non infiammabile; spesso utilizzato è anche il refrigerante R513A, che ha un GWP (Global Warming Potential) più contenuto pur mantenendo buone performance. Entrambi questi fluidi refrigeranti sono gas fluorurati, anche detti idrofluorocarburi (HFC), e se vengono emessi nell'atmosfera contribuiscono all'alterazione delle condizioni climatiche. Come detto in precedenza nel capitolo sulla Cold Chain, l'incidenza degli HFC attualmente non è alta rispetto al resto dei gas serra, ma il trend fortemente crescente richiede misure di controllo adeguate sulle emissioni di questi gas. È per questo che le organizzazioni che sfruttano macchinari con al loro interno queste sostanze necessitano di figure professionali dedicate.

Nel caso studio la soluzione adottata per il fluido refrigerante è il gas refrigerante propano R290 che ha alcuni vantaggi: è un gas naturale con GWP quasi a zero, è un gas puro (e non una miscela di più componenti come tutti i gas refrigeranti sintetici), possiede eccellenti proprietà termodinamiche che permettono ottimi rendimenti del macchinario. Di contro essendo un gas infiammabile necessita di alcuni accorgimenti come: l'installazione del sistema esclusivamente all'esterno dello stabilimento e all'aria aperta (in modo da non creare atmosfera esplosiva in caso di perdita del gas refrigerante) e limitare la carica del circuito frigorifero per diminuire la pericolosità.



**Figura 4.7:** Ciclo termodinamico del fluido refrigerante, Schema di un sistema di raffreddamento basato su un chiller a compressione e una torre di raffreddamento

**COP** Per calcolare la resa di un dispositivo come una macchina frigorifera non viene utilizzato l'indicatore di rendimento, ma bensì il Coefficiente di Performance o Prestazione (COP), il quale è il risultato del rapporto tra il calore sottratto a bassa temperatura  $Q_{evap}$  (o potremmo dire la quantità di freddo prodotta e scambiata nell'evaporatore) e il lavoro prodotto da tutti i componenti del gruppo frigorifero  $L$  [85].

$$COP = \frac{Q_{evap}}{L}$$

Un'ulteriore misura che viene presa solitamente in considerazione è il SCOP (Seasonal COP) che determina l'efficienza stagionale durante l'intera stagione di raffreddamento, ottenuta dividendo la domanda di raffreddamento per i consumi elettrici annuali. Nell'applicazione in questione quest'ultima misura non è affatto utile in quanto avendo un uso prevalente nel periodo estivo il SCOP non è di fatto

rappresentativo poiché il massimo utilizzo del chiller avviene durante i periodi in cui il COP è più basso a causa della temperatura esterna (mesi estivi). Durante i periodi in cui la temperatura esterna consentirebbe un COP migliore la struttura non ha di fatto dispersioni importanti e rimane attivo il raffreddamento soltanto della cella frigo 1 che, essendo coibentata con pannelli di spessore 180/200 cm e chiusa all'interno di un magazzino a +20°C, ha un consumo piuttosto basso in relazione alla potenza del chiller.

Il dispositivo frigorifero analizzato è stato realizzato da un'impresa su misura per la struttura in questione e non essendo una macchina di serie o dedicata a servizio di climatizzazione non ha rese certificate, quindi non è possibile fornire il COP teorico dei compressori ad una determinata condizione di lavoro. La macchina frigorifera presa in considerazione nel caso studio prepara una soluzione di acqua glicolata (fluido da raffreddare, di processo) a -5°C in tutti i periodi dell'anno, in quanto questa temperatura è necessaria alla Cella Frigo 1 (quella con la temperatura più bassa 2/3°C) che è l'unica utenza attiva 365 giorni all'anno 24 ore su 24. È possibile ricavare il COP definendo le condizioni di lavoro, e per darne una rappresentazione delle performance nel periodo in cui il chiller è sottoposto al maggiore sforzo prendiamo in esame il lavoro estivo. Considerando che le condizioni reali sono normalmente molto più favorevoli in quanto la condensazione flottante e la parzializzazione della potenza dei compressori variano in modo sensibile il rendimento energetico, in particolar modo con un compressore parzializzato con inverter a pari temperatura esterna di 32°C si ottengono COP molto più alti di quanto sotto dichiarato. Le condizioni per il calcolo sono:

- compressori con economizzatore attivo;
- temperatura esterna +32°C;
- condensazione a +43°C ( $\Delta T$  condensatore 4 K);
- evaporazione a -11,5°C con acqua glicolata a -6°C;
- potenza termica sviluppata da ciascun compressore 100,5 Kw.

Date queste condizioni i valori di performance sono:

$$COP_{\text{dei soli compressori}} = 2,12$$

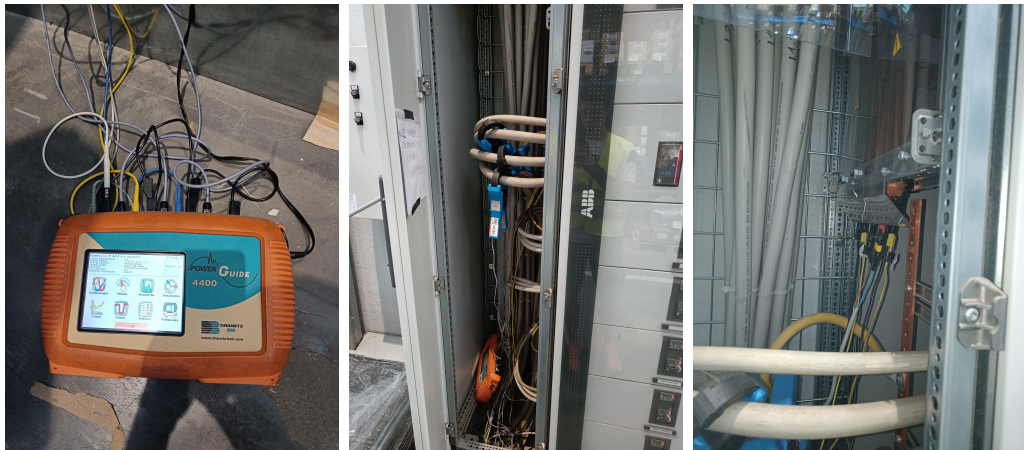
$$COP_{\text{globale Chiller}} = 1,85$$

#### 4.1.2 Misurazioni e dati

Essendo l'intento di questa tesi proporre un modello previsionale che crei delle previsioni sul consumo del gruppo frigo è stato necessario individuare quali potessero

essere le variabili significative da considerare per poi andare a testare la loro significatività grazie ai metodi di Machine Learning. Siamo quindi in presenza di un modello **data-driven**, dunque la materia prima per poter costruire il modello vero e proprio sono i **dati**.

**Dati di consumo elettrico** Avendo per prima cosa investigato e analizzato la struttura è stato deciso di procedere con la rilevazione del primo parametro necessario per la sua stima futura: il consumo del gruppo frigo. La rilevazione dei dati di consumo è stata fatta con l'utilizzo di uno strumento collegato sulla linea di alimentazione elettrica della macchina frigorifera, in particolare si tratta di un Dranetz PowerGuide® 4400 provvisto di otto canali. La misura è stata fatta direttamente sul quadro elettrico generale da cui parte la linea che alimenta il solo gruppo frigo (Figura 4.8). Avendo a che fare con un sistema trifase con



**Figura 4.8:** Dispositivo di misurazione con i collegamenti a tensione e corrente

collegamento a stella è stato necessario misurare per ogni fase la tensione di fase e la corrente di linea. Per il rilevamento delle correnti di linea sono state collegate al dispositivo 3 pinze amperometriche e inserendo tra le ganasce di ogni pinza un unico conduttore di fase presente nel quadro; questo procedimento è stato ripetuto per ciascuna delle 3 fasi. Per quanto riguarda le tensioni di fase sono stati collegati al dispositivo per ogni fase un cavo che pinzasse il conduttore della fase e un cavo che pinzasse il neutro (per entrambi i cavi ci siamo serviti di morsetti a coccodrillo). Il collegamento in modo schematico è rappresentato in Figura 4.9.

Tramite questi collegamenti il dispositivo, per tutto il periodo in cui è rimasto collegato, ha memorizzato per ogni fase i valori minimi, medi e massimi di tensione ( $E$ ), corrente ( $I$ ) e potenza attiva ( $P$ ). Inoltre lo strumento proprio per calcolare la potenza attiva ha la capacità di memorizzare la forma d'onda di ogni fase che le permette di derivare l'angolo di sfasamento ( $\varphi$ ) tra la tensione di fase e la corrente

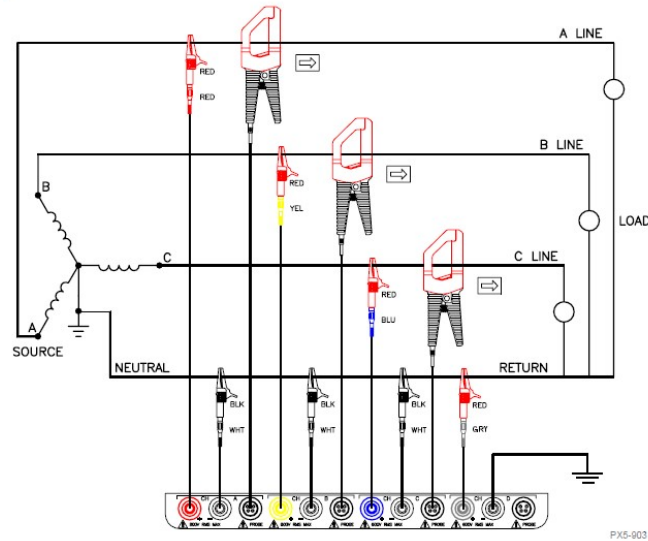


Figura 4.9: Schema collegamenti dispositivo di misurazione

di linea. Da questi valori si è scelto di considerare la potenza media assorbita da ogni fase che sommate danno in quell'istante la potenza attiva assorbita dal gruppo frigo.

$$P_{\text{Media, Tot}} = P_{\text{Media, A}} + P_{\text{Media, B}} + P_{\text{Media, C}}$$

Il campionamento delle misurazioni è stato impostato con un intervallo di 15 minuti, per cui ogni quarto d'ora lo strumento memorizzava le informazioni. Per tramutare le potenze elettriche (kW) in consumi energetici (kWh) si è moltiplicata la potenza per un fattore di scala pari a 0.25, equivalente ad  $\frac{1}{4}$  la frazione di ora che viene campionata dal dispositivo. A causa del reperimento dei dati di temperatura su base oraria e per avere nel modello dati omogenei si è deciso di avere un frazionamento orario anche dei dati di consumo, perciò per questi dati è stata fatta per ogni ora una media delle quattro rilevazioni effettuate nell'ora stessa per avere il consumo medio in quella fascia oraria.

**Dati di temperatura** Dal momento che il caso studio si prefigge di analizzare e prevedere i consumi di una macchina frigorifera, una variabile che non può essere tralasciata è quella della temperatura. Nonostante la struttura delle celle frigo sia ben coibentata gli scambi termici con l'ambiente esterno avvengono, in maniera decisamente più marcata nel periodo estivo. I flussi di calore che possiamo riscontrare sono banalmente quando si apre la cella frigo per effettuare delle operazioni, in questi casi l'aria ad una temperatura maggiore entra creando l'innalzamento

della temperatura della cella stessa. Ovviamente maggiore è la temperatura esterna maggiore è l'incidenza sul consumo del chiller a cui verrà richiesto di produrre un maggiore lavoro termico. Per questi motivi una delle variabili da considerare è la temperatura esterna. Per ottenere i dati di temperatura esterna non è stato possibile inserire dei rilevatori in prima persona per avere la misura nella posizione geografica esatta dello stabilimento. Tuttavia è stato possibile reperire dati di temperatura attendibili tramite l'ARPA (Agenzia Regionale per la Protezione Ambientale) della regione in cui è situato lo stabilimento. La struttura è localizzata in una posizione intermedia lungo la linea che unisce due stazioni metereologiche supervisionate dall'ARPA. Questo ha permesso di considerare e richiedere i dati da entrambe le stazioni per il periodo in esame. Valutando che le discrepanze tra i valori di temperatura rilevati dalle due stazioni non fossero considerevoli, si è scelto di utilizzare come dato nel modello la media delle temperature rilevate dalle due stazioni.

La grandezza fisica che è maggiormente influenzata dalla temperatura esterna è proprio la temperatura di ogni comparto dello stabilimento. È proprio il rilevamento di queste temperature tramite sensori interni agli ambienti specifici che mette in azione il raffreddamento dei locali alle temperature prestabilite. Perciò anche le temperature interne della cella frigorifera e degli altri ambienti a temperatura controllata sono state acquisite ed analizzate. Si ipotizza che il dato sulla cella frigo possa essere il più significativo essendo il consumo relativo a questo ambiente il più dispendioso per il chiller e diventare nel modello una variabile significativa per la previsione del consumo, visto che se la temperatura interna aumenta, con un andamento rilevato dai dati, ci si può aspettare che all'istante successivo o all'ennesimo istante successivo il consumo aumenterà e viceversa. Questo tipo di dato è memorizzato dal sistema gestionale interno e quindi ci è stato fornito direttamente dall'azienda. I file che ci sono stati messi a disposizione dall'azienda avevano una suddivisione giornaliera, ovvero un file per giorno, e una cadenza di rilevazione delle misure riportate al loro interno di 30 minuti. È stato necessario quindi avvalersi di un codice Python per rendere automatica la lettura dei vari file e stanziare tutti i dati contenuti al loro interno in un unico dataset che è stato esportato sotto forma di file *xlsx*. I file inoltre riportavano le misure di temperatura e umidità di tutti i locali a temperatura controllata, si è dovuto quindi fare un filtraggio per estrapolare solo le temperature della cella frigo e dell'ante-cella, unici locali d'interesse per lo studio della tesi.

Per le temperature interne i dati sono stati riportati aggregati all'ora, in modo da avere una struttura omogenea con tutti i dataset utilizzati per il modello. I file originali riportavano un frazionamento delle rilevazioni ogni 30 minuti e sono quindi state fatte le medie per le due misure di ogni ora. Per quanto riguarda le temperature esterne i dati ricevuti dall'ARPA avevano già un campionamento orario.

Una delle motivazioni dello studio riguarda proprio la necessità che ha avuto l'azienda di modificare la temperatura del locale anticella. A causa di una richiesta di gestione ordini maggiore da parte dei clienti, la cella frigo 1 presentava una capacità di stoccaggio insufficiente ed è per questo che dal primo di marzo la temperatura dell'anticella (rilevata anche dai file esaminati) è stata abbassata da 8/10°C a 3/4°C utilizzando il locale come stoccaggio momentaneo. Questa inefficienza ha portato sicuramente un incremento nel consumo del gruppo frigo, soprattutto per l'errato utilizzo dell'anticella che dovrebbe servire come intermezzo per evitare lo scambio termico diretto tra la cella frigo e l'esterno.

**Dati gestionali operativi** La peculiarità di questo studio è proprio la previsione del consumo elettrico prodotto per raffreddare gli ambienti a temperatura controllata, in particolare della cella frigo, ma considerando non solo variabili fisiche o prettamente correlate con l'efficienza tecnica della struttura e della strumentazione. Ciò che si vuole studiare è la correlazione tra i consumi elettrici del gruppo frigo e l'operatività del magazzino. Con operatività si intende tutte quelle operazioni che coinvolgono in modo fisico l'apertura (ingresso, uscita) della cella la movimentazione di materiale all'interno e del personale. Tramite il WMS dell'azienda, che registra tutte le informazioni relative all'operatività, è stato possibile reperire un numero elevato di queste informazioni che possono essere utili nel modello. Va detto che con un elevato numero di variabili è compito in sede di validazione del modello valutare se la significatività di ogni variabile presa in considerazione è ad un livello abbastanza alto oppure è preferibile scartare quelle statisticamente non significative.

In prima istanza l'azienda ci ha reso disponibili dei file giornalieri estratti dal sistema di gestione del magazzino (WMS), contenenti al loro interno una media di circa 4-5 mila record per ogni file giornaliero. Questi dati si riferiscono tutti a movimenti ed operazioni che vengono effettuate strettamente nella cella frigo e nell'ante-cella, locali di interesse dello studio in questione. I file sono stati inviati in formato xlsx con numerose colonne riportanti informazioni non solamente di carattere fisico ma anche logico. Infatti questi file sono compilati automaticamente dal sistema e registrano in ogni istante tutte le operazioni fisiche o logiche (informatiche) che si svolgono. Per quanto riguarda le operazioni logiche registrate dal sistema intendiamo quelle operazioni che non rispondono ad un'azione appunto fisica in sito ma solamente azioni di tipo contabile informatico e fiscale. Tutte le operazioni in questi file sono classificate e possiedono delle caratteristiche esclusive in modo che tali operazioni siano identificabili in modo univoco. Tra tutti i parametri che sono stati presi in considerazione quelli più importanti e di interesse per lo studio di questa tesi sono principalmente:

- *referenza*: è il codice univoco relativo ad un ordine di prelievo o di stoccaggio,

viene identificato come il codice dell'ordine;

- *numero bolla*: è il codice identificativo della bolla (documento di trasporto<sup>1</sup>) per quel dato ordine;
- *quantità*: questa dimensione viene espressa attraverso due parametri: i pezzi e i colli. Vengono presi in considerazione entrambi perché non è esclusa la movimentazione di pezzi singoli (per singola operazione) per determinati tipi di prodotti e quindi la corrispondenza collo-numero pezzi non è assicurata;
- *codice identificativo vano*: è il codice che indica dove una determinata operazione viene eseguita per i pezzi/colli interessati e quindi la loro ubicazione. Il codice è formato da quattro blocchi di cifre separati dal carattere "-". Gli ultimi due blocchi sono quelli che ci interessano per individuare la posizione in cui viene eseguita l'operazione. Il penultimo blocco di due cifre indica la campata o colonna (dimensione orizzontale) dell'ordine, mentre l'ultimo blocco di due cifre indica l'altezza in cui è posizionato il collo/pezzo oggetto dell'operazione (esempio: 133-0A-24-01, quest'ordine è posizionato alla campata 24 al livello 1).

Tramite l'incrocio dei due parametri di numero bolla e referenza si è riusciti a mettere in relazione l'ordine con la spedizione ad esso associata e collegare tutte le caratteristiche associate a questi due elementi. Mentre il codice identificativo del vano è stato tenuto in conto come proxy per il calcolo degli spostamenti orizzontali e verticali eseguiti per una data operazione.

I file originali includevano un'elevata numerosità di informazioni che non sarebbero state utili in sede di elaborazione del modello, alcuni esempi sono: descrizione dell'articolo, produttore, data di scadenza, numero documento, referenza documento, riferimento ERP esterno, magazzino logico, commessa, sotto commessa, peso variabile, prezzo al kg, barcode esterno, barcode master, committente, anagrafica, mittente, locazione mittente, destinatario, locazione destinatario. Dopo aver eseguito l'analisi dei file originali pervenuti e riconosciute le caratteristiche opportune relativamente alle informazioni necessarie al caso studio, è stato fondamentale pulire il file filtrando solamente le informazioni utili al fine di costruire il modello ed avere una struttura dati più ordinata, comprensibile e chiara per la lettura

---

<sup>1</sup>Documento emesso prima della consegna o della spedizione della merce con l'indicazione degli elementi principali dell'operazione, che si riassumono nei seguenti dati: il numero progressivo, la data, le generalità del cedente, del cessionario e del soggetto eventualmente incaricato del trasporto, la quantità (espressa in cifre) dei beni trasportati, la descrizione dei beni trasportati con indicazione di natura e qualità degli stessi. Viene emesso in due copie: una deve essere trattenuta e conservata dall'emittente, l'altra deve essere consegnata al cessionario in accompagnamento della merce ceduta e trasportata [86].



umana in prima battuta ma soprattutto per la lettura dei dati da parte del modello di machine learning. Proprio questa fase di preparazione e manipolazione dei dati è stata eseguita tramite l'ausilio di un codice Python che ha reso possibile automatizzare la lettura dei numerosi file giornalieri presi in considerazione e farne l'elaborazione per estrarne i dati utili e ordinati. Grazie alla struttura DataFrame messa a disposizione dalla libreria Pandas è stato possibile salvare i dati sotto il formato tabulare (righe, colonne) e farne poi l'esportazione su file xlsx. Il file "pulito" è stato organizzato in modo da far combaciare le informazioni con gli altri tipi di dato (temperature e consumi) ed avere tutti i dataset uniformati secondo la stessa struttura, per questo motivo i dati sono stati aggregati all'ora. Ogni record del dataset indica un'ora e la corrispondente data del giorno in questione e per ognuna delle caratteristiche, che in seguito verranno descritte, sono stati calcolati gli aggregati per ognuna di esse in quel lasso di tempo orario. Avendo i dati salvati con timestamp con precisione al secondo per ogni attività, è stata calcolata la somma nell'ora presa in considerazione di ogni parametro interessato, in modo da poter avere l'andamento orario (in linea con gli altri dataset) di ogni attività. In definitiva nel dataset finale sono state individuate ed estratte 5 tipologie di attività:

- INB: inbound. Attività di scarico della merce in entrata dai mezzi di trasporto e posizionata in ante-cella.
- STO: storage. Attività di stoccaggio della merce dopo che è stata scaricata nell'ante-cella e viene quindi posizionata negli scaffali dentro la cella frigo.
- REL: relocation. Attività di ri-allocazione della merce da una posizione in uno scaffale ad un'altra, attività eseguita internamente alla cella.
- PIC: picking. Attività di prelievo della merce dagli scaffali e posizionata nell'ante-cella.
- OUT: outbound. Attività di carico della merce in uscita dall'ante-cella sui mezzi di trasporto.

Alle attività nel dataset finale sono associati dei parametri che daranno la possibilità al modello di valutare i movimenti che interessano le aree dell'ante-cella e della cella frigo. Per ogni attività si riscontrano i valori dei pezzi (pieces), dei colli (packs) e del numero di ordini (orders) effettuati nella fascia oraria per una determinata attività. Per quanto riguarda le attività che richiedono un accesso allo scaffale, come lo storage la relocation e il picking, si hanno tre parametri supplementari calcolati grazie all'informazione di posizione dei pezzi/colli movimentati ricavata dal codice identificativo del vano presente nel file originale:

- LEV: level. Indica il numero di spostamenti verticali.

- COL: column. Indica il numero di spostamenti orizzontali.
- AIS: aisle. Indica il numero di corsie presenti nella cella frigo, facendo partire il conteggio dal muro a sinistra dell'ingresso della stessa cella.

Per le attività di storage e picking gli spostamenti sono calcolati come somma dei livelli e colonne perché fanno riferimento ad una posizione iniziale (per lo storage) o finale (per il picking) di riferimento, che potremmo considerare come livello 0 e colonna 0 siccome vengono prelevati o posizionati nell'ante-cella (livello terreno). Differente è il calcolo degli spostamenti per l'attività di relocation che tiene conto della differenza delle posizioni in cui erano i pezzi/colli da movimentare e la nuova posizione in cui vengono stoccati.

Ottenuto il file finale pulito e ordinato è stato possibile riconoscere alcune pratiche ripetute riguardanti la gestione delle attività:

- tutte le attività vengono effettuate in un range temporale che va dalle ore 7 alle ore 18, indicando le ore lavorative (da aggiornare quando si avranno tutti i dati);
- le attività di picking e outbound si concentrano nella seconda parte della giornata, generalmente dalle 13 in poi.

Riconoscendo che le attività operative di magazzino vengono svolte solamente in orario diurno, si è deciso di aggiungere in questo dataset una colonna booleana che indicasse l'orario operativo (ovvero lavorativo per le attività di magazzino) in modo che nel creare il modello l'incidenza delle variabili logistiche venga presa in considerazione unicamente quando la variabile booleana è impostata su True(1). La scelta è stata presa in conseguenza del fatto che se questo controllo non viene inserito il modello leggerebbe tutte le variabili logistiche a zero durante le ore non operative e nel calcolo dei coefficienti ci sarebbe uno sbilanciamento per l'incidenza di tali variabili sulla previsione. In sostanza le variazioni nel consumo durante le ore non operative non vengono inficiate dalle variabili logistiche (che sarebbero tutte a zero), ma solamente dalle altre variabili precedentemente descritte.

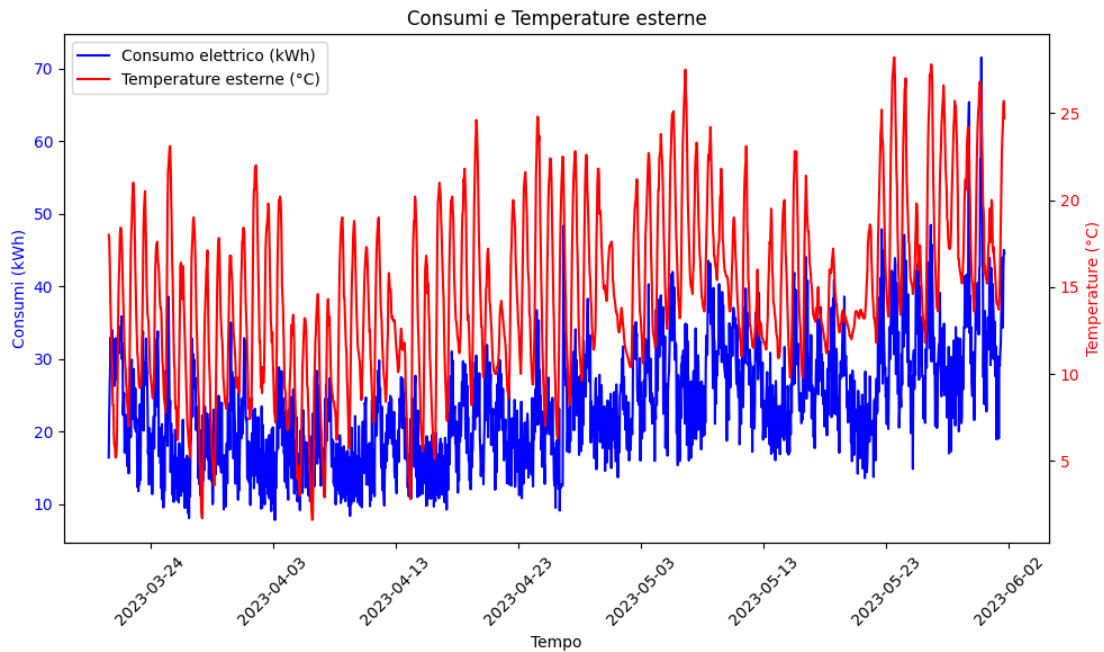
## 4.2 Modello di forecasting

In questa parte viene descritto il processo per arrivare a strutturare un modello di Machine Learning tramite il linguaggio di programmazione Python che permetta di prevedere i consumi della cella frigo del magazzino in questione. Precedentemente sono state descritte le tipologie di dati che sono stati raccolti per provare a determinare le variabili significative da considerare nel modello. Prima di elaborare il modello vero e proprio è necessario effettuare un'analisi dei dati raccolti per

verificare alcune assunzioni necessarie nella costruzione di un modello di forecasting. Siccome in letteratura non sono presenti studi che valutano il consumo di una struttura di questo genere prendendo come variabili indipendenti anche le variabili gestionali che interessano gli spostamenti di merce effettuati nella cella frigo e nell'anticella, si è deciso di puntare ad un modello di analisi e forecasting semplice come prima proposta in questo ambito. L'obiettivo diventa creare una regressione lineare, che vista la presenza di una molteplicità di variabili indipendenti, sarà multipla; quindi l'intento sarà di definire un modello previsionale utilizzando una Regressione Lineare Multipla. Come anticipato, prima di progettare il modello di ML per il forecasting è stata effettuata un'analisi di regressione. In un'analisi di regressione generalmente si compiono i seguenti passi [87]:

- Studio statistico dei dati;
- Creazione della regressione;
- Validazione dell'analisi e delle assunzioni (analisi dei residui);
- Trarre conclusioni sulla significatività dei predittori sull'effetto che hanno sulla risposta;
- Costruire il modello di regressione lineare;
- Rivedere le assunzioni di linearità nel modello e effettuare delle trasformazioni se necessario;
- Ottimizzare il modello per determinare i livelli ottimi dei predittori;
- Verificare i risultati.

Per quanto riguarda l'analisi di regressione ci si è avvalsi del software statistico MINITAB®, che permette di effettuare questi tipi di analisi in modo efficace e veloce grazie alle funzionalità build-in presenti in esso. Per prima cosa è stata analizzata quella che è la risposta del nostro modello ovvero i consumi elettrici del gruppo frigo. Si è notato come l'andamento dei consumi verso il mese di maggio presentasse un trend in crescita, dovuto principalmente all'innalzamento delle temperature. La correlazione tra le variabili temperatura esterna e consumo elettrico è decisamente marcata come si può constatare anche dal grafico in Figura 4.10, dove l'andamento dei consumi segue l'andamento della temperatura. Si porta all'attenzione anche l'andamento del consumo medio giornaliero nel quale si riesce ad osservare come nei giorni non lavorativi (nel grafico in Figura 4.11 evidenziati) i consumi siano sempre minori rispetto ai giorni di lavorativi che li precedono. È stato calcolato che il consumo medio giornaliero nei giorni lavorativi è di 25.17 kWh mentre il consumo medio giornaliero nei giorni non lavorativi è pari a 19.63 kWh. Questo

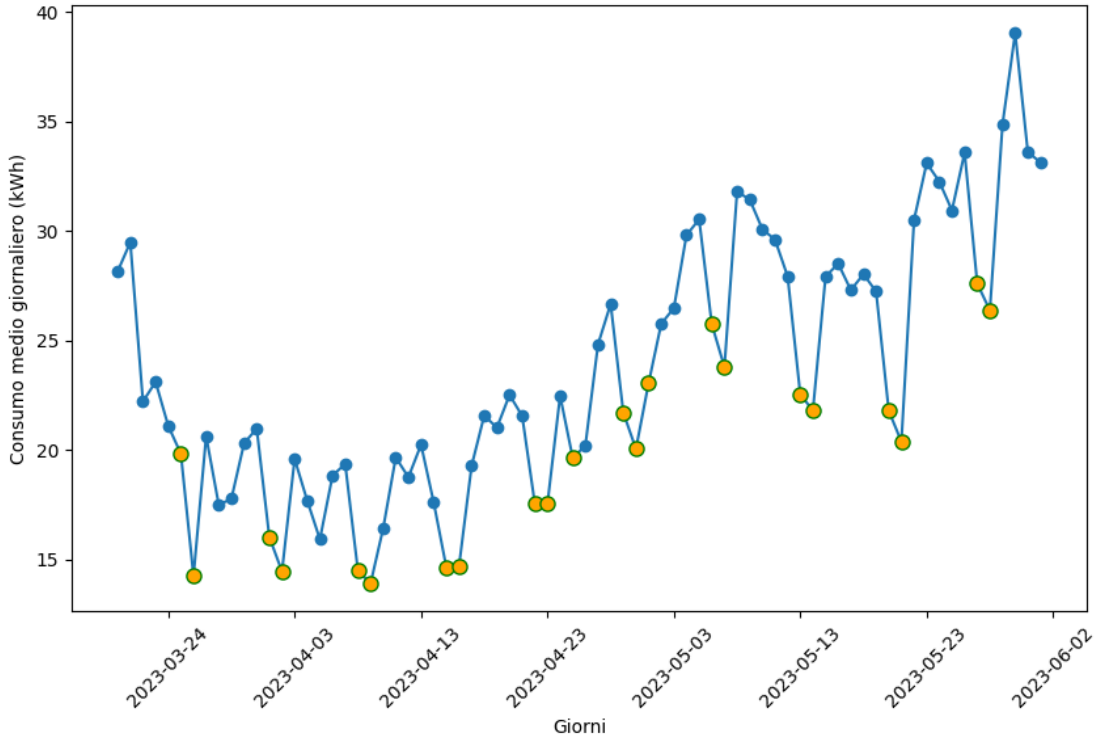


**Figura 4.10:** Consumi elettrici e temperature esterne, andamento sovrapposto

dato è sostenuto anche dai dati gestionali che ci sono pervenuti visto che non vi è presenza di attività in queste date, a supporto del fatto che l'attività di carico e scarico all'interno del magazzino è minima.

Ritornando al trend dei consumi, esso evidenzia la non stazionarietà della serie temporale presa in considerazione, perciò diventa necessario o considerare solamente la parte stazionaria oppure eseguire delle trasformazioni, dette integrazioni, per rendere stazionaria tutta la serie. A questo problema si aggiunge quello della ripartizione dei consumi sui diversi locali climatizzati della struttura, che non è possibile effettuare per mancanza dell'adeguata strumentazione, come esplicitato nel paragrafo riguardante i flussi termici. In definitiva il problema della ripartizione dei consumi si presenta quando entra a regime il "periodo estivo", dove non solo la cella frigo e l'anticella sono climatizzate ma anche gli altri locali concorrono ai consumi del gruppo frigo. Non avendo specifiche informazioni su quando il periodo di gestione estivo entra in esercizio, è stato deciso di considerare solamente i dati in cui la serie temporale risulta stazionaria, ipotizzando che il periodo estivo cominci con l'innalzamento dei consumi, corrispondente a quello delle temperature esterne.

Per determinare il periodo di stazionarietà della serie temporale è stato utilizzato un test d'ipotesi conosciuto come ADF test (Augmented Dickey-Fuller test), che prevede come ipotesi nulla la presenza di una radice unitaria nella serie storica rappresentativa della presenza di un trend in essa, l'ipotesi alternativa è data dalla



**Figura 4.11:** Consumo elettrico medio giornaliero con evidenza dei giorni non lavorativi

non presenza della radice unitaria. In particolare il test è realizzato tramite una funzione di regressione che include una costante. La formulazione del test è [88]:

$$\begin{cases} H_0 : \gamma = 0 \\ H_1 : \gamma < 0 \end{cases}$$

$$\Delta Y_t = \gamma Y_{t-1} + \mu + \sum_{i=1}^p \delta_i \Delta Y_{t-i} + \varepsilon_t$$

Il test prevede il calcolo della statistica

$$t_{\delta_i} = \frac{\hat{\delta}_i}{se(\hat{\delta}_i)} \quad \text{e se } t_{\delta_i} < DF_{critical}$$

allora  $H_0$  è rigettata e la serie è stazionaria. Il valore di  $DF_{critical}$  è derivato dalla distribuzione Dickey–Fuller.

Per il caso in questione è stato effettuato il test in modo iterativo, tramite l'utilizzo di un codice Python, ampliando la serie analizzata considerando una

settimana aggiuntiva ad ogni iterazione finché il test non è risultato negativo. Grazie a questa analisi si è deciso di prendere come periodo d'esame dal 20 marzo (inizio rilevazioni sui consumi) fino al 28 aprile (data ultima per avere una settimana completa di dati stazionari), così facendo si considerano settimane intere senza dover troncare i dati e avendo un dataset più bilanciato. Questa procedura viene applicata per essere in linea con i dati gestionali a disposizione non dovendoli troncare.

Dopo aver fatto alcune considerazioni sui dati presi in analisi tramite MINITAB è stata creata la regressione ricevendo in output l'analisi completa su significatività dei coefficienti e analisi dei residui. In un'analisi di regressione lineare multipla per creare un modello affidabile è necessario che si rispettino alcune assunzioni chiave [89]:

- non vi deve essere multicollinearità tra le variabili indipendenti;
- ci deve essere una relazione lineare tra le variabili indipendenti e quella dipendente;
- deve essere rispettata l'omoschedasticità;
- i residui devono essere normalmente distribuiti.

La prima condizione che è stata studiata è quella della multicollinearità. Per **multicollinearità** si intende il fenomeno statistico che si verifica quando due o più variabili indipendenti in un modello di regressione sono altamente correlate tra di loro. Tale fenomeno può creare difficoltà nel determinare accuratamente il singolo effetto che ogni variabile indipendente ha sulla variabile indipendente. La stima dei coefficienti, con la presenza di multicollinearità, può risultare instabile e non affidabile rendendo più difficile trarre conclusioni significative dal modello. Le stime dei coefficienti effettuate in presenza di multicollinearità essendo instabili presentano una maggiore varianza, producendo intervalli di confidenza ampi e maggior variabilità nella previsione della variabile dipendente del modello [90]. La presenza di questo fenomeno in un dataset può comparire per diversi motivi: nuove variabili create che sono dipendenti da altre variabili, informazioni ridondanti nel modello che possono portare a variabili altamente correlate, esperimento progettato male con un gran numero di dati osservati e con l'impossibilità di manipolarli, anche l'uso di variabili dummy può causare multicollinearità. L'ottimizzazione dei coefficienti in una regressione lineare viene fatta con il metodo degli OLS (Ordinary Least Squares, metodo dei minimi quadrati) e l'assenza di collinearità perfetta è proprio una delle assunzioni di tale metodo, quindi necessariamente da escludere. Per studiare la multicollinearità in MINITAB ci si è serviti di un diagramma che mostra quanto sono correlate le variabili indipendenti tra di loro, un correlogramma. Come si può vedere in Figura 4.12 ogni variabile indipendente

è incrociata con le altre e per ogni intersezione è calcolato l'indice di correlazione di Pearson, che solitamente porta a considerare la correlazione tra le due variabili alta se il valore dell'indice supera 0.8 [91, 92] e di conseguenza crea nel modello l'effetto di multicollinearità prima descritto. Come si può notare dal diagramma sono molte le coppie di variabili che presentano alta correlazione, zone facilmente riconoscibili per la legenda grafica che mostra in rosso acceso queste intersezioni.

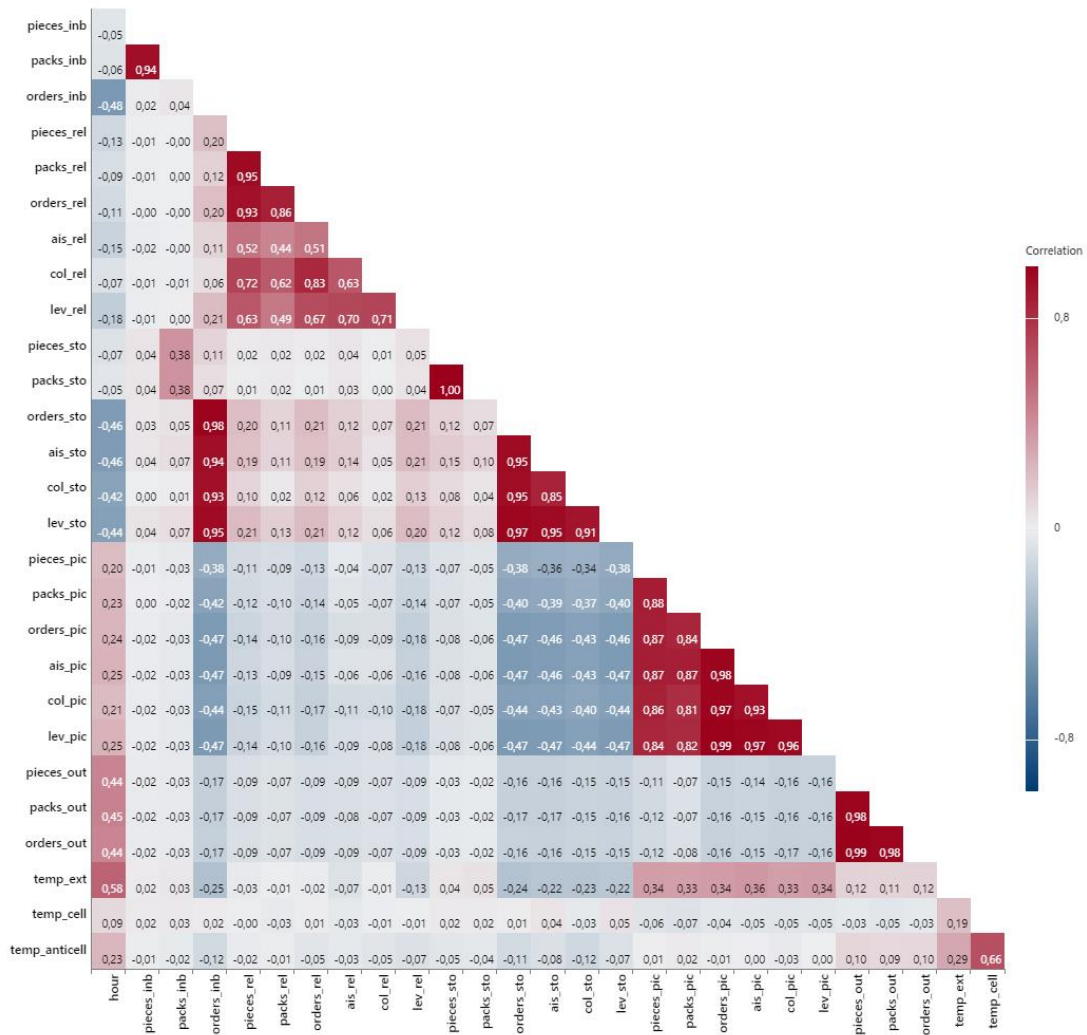


Figura 4.12: Correlogramma con indicato l'indice di correlazione di Pearson

Per evitare o anche solo ridurre questo fenomeno la prassi è eliminare dal modello le variabili maggiormente correlate fra loro, ma in questa fase la problematica si sposta sul criterio col quale vengono scelte le variabili correlate da scartare? Intanto si può partire dallo scartare quelle variabili che abbiano informazioni meno

importanti, ma questa è solamente una valutazione qualitativa e soggettiva di chi analizza; in generale è un processo iterativo per cui si eliminano alcune variabili e si controllano i livelli di performance della regressione andando avanti per tentativi e seguendo se possibile l'esperienza dell'analista. Tuttavia è presente un indice che ci permette di avere una valutazione quantitativa di quelle variabili che maggiormente sono correlate con tutte le altre nella regressione. Tale misura è il VIF (Variance Inflation Factor) e determina la forza della correlazione tra le variabili indipendenti di una regressione. Il punteggio VIF di una variabile indipendente rappresenta quanto bene viene spiegata la variabile da altre variabili indipendenti. Questa misura ha come limite inferiore 1 e non ha limite superiore, per un  $VIF = 1$  non vi è correlazione tra la variabile indipendente e le altre. Se il VIF eccede il valore 5 o 10 (in base al caso che si sta studiando) si considera alto il livello di multicollinearità tra la variabile interessata e le altre [90].

Quindi grazie all'analisi di correlazione con il correlogramma si ha già idea che coppie di variabili quali "packs" e "pieces" per tutte le operazioni (stoccaggio, prelievo, rilocazione, scarico e in uscita) sono altamente correlate. In particolare tra tutte le variabili del picking si riscontrano alti valori di correlazione. La forte correlazione tra pezzi e colli delle varie attività comunque non stupisce, poiché rientra in uno dei punti precedentemente elencati che possono far nascere multicollinearità, ovvero variabili che presentano informazioni parzialmente ridondanti. Per avere una migliore immagine del valore di correlazione tra le variabili si è eseguita una prima regressione considerando tutte le variabili e producendo i risultati in Figura 4.13. Come ci si poteva aspettare i valori dell'indicatore VIF di correlazione per molte feature sono a livelli decisamente troppo alti. Tre sono le variabili che si deve cominciare ad escludere "pieces\_inb", packs\_inb e "packs\_sto" con rispettivamente valori VIF pari a 79929.88, 93338.50, 14416.44. Si è deciso di considerare, visti i livelli generalmente alti, di considerare come livello di correlazione accettabile la soglia per il  $VIF \leq 10$ . Da questa premessa sono state svolte numerose regressioni che iterativamente venivano "pulite" delle variabili con i VIF più alti per mantenere tutte le variabili del modello sotto il livello di soglia prefissato. La prima regressione che ha rispettato i parametri è quella presentata in Figura 4.15, in cui si può vedere come tutte le variabili rimaste soddisfano il vincolo  $VIF \leq 10$ .

Arrivati a questa riduzione e trovando le feature utilizzabili per il modello il focus si è spostato sul trovare le variabili significative nel modello. Per ricavare questa informazione si deve ricercare il p-value, una statistica che indica proprio il livello di significatività della variabile all'interno del modello sviluppato. Tale statistica è presente nel prospetto della regressione in Figura 4.15. Generalmente viene considerata significativa la variabile che ha un  $p - value \leq 0.05$  [93, 94]. Nel nostro caso solamente le feature rappresentanti le temperature esterne e delle celle rientrano al di sotto del 5%. La significatività delle temperature delle celle è stata verificata in quasi tutte le iterazioni, mentre per le temperature esterne è



Coefficients					
Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	15.08	3.06	4.92	0.000	
hour	0.327	0.142	2.31	0.022	2.54
pieces_inb	-0.000272	0.000230	-1.18	0.239	79929.88
packs_inb	0.00271	0.00230	1.18	0.240	93338.50
orders_inb	0.2219	0.0992	2.24	0.026	43.92
pieces_rel	-0.00516	0.00485	-1.06	0.288	25.16
packs_rel	0.00096	0.00411	0.23	0.816	14.97
orders_rel	0.1916	0.0965	1.99	0.048	15.11
ais_rel	0.0917	0.0460	1.99	0.047	2.53
col_rel	-0.01059	0.00728	-1.46	0.146	5.28
lev_rel	-0.0537	0.0518	-1.04	0.301	3.32
pieces_sto	0.000333	0.000572	0.58	0.560	845.74
packs_sto	-0.00277	0.00237	-1.17	0.243	14416.44
orders_sto	-0.060	0.157	-0.38	0.702	106.33
ais_sto	-0.0203	0.0216	-0.94	0.348	15.87
col_sto	-0.00970	0.00400	-2.42	0.016	14.04
lev_sto	0.0299	0.0325	0.92	0.358	26.04
pieces_pic	0.000388	0.000994	0.39	0.697	7.09
packs_pic	0.00207	0.00263	0.79	0.432	5.95
orders_pic	-0.0127	0.0546	-0.23	0.817	116.23
ais_pic	0.0069	0.0136	0.51	0.613	43.71
col_pic	-0.00290	0.00199	-1.46	0.146	30.78
lev_pic	0.0173	0.0154	1.13	0.261	50.11
pieces_out	-0.00001	0.00107	-0.01	0.991	126.47
packs_out	-0.00002	0.00100	-0.02	0.988	30.73
orders_out	0.0015	0.0124	0.12	0.905	94.63
temp_ext	0.3914	0.0902	4.34	0.000	1.95
temp_cell	1.48	1.00	1.48	0.140	2.04
temp_anticell	-1.622	0.749	-2.17	0.031	2.11

Figura 4.13: Valori della prima regressione con tutte le variabili considerate

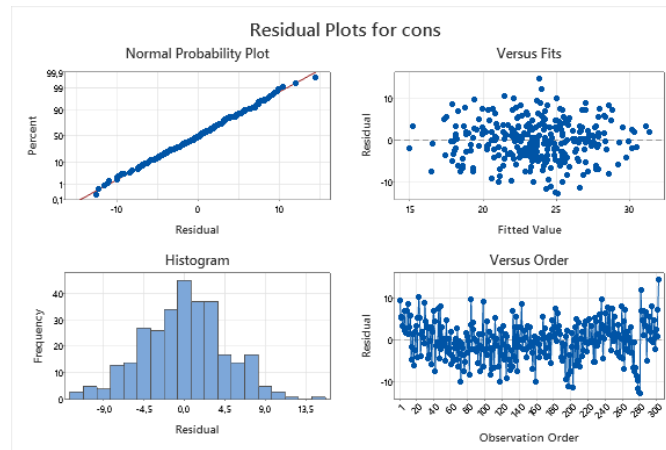


Figura 4.14: Analisi dei residui sulla regressione con tutte le variabili

stata verificata sempre, a dimostrazione del forte legame tra consumo elettrico del gruppo frigo e temperatura esterna.

Per quanto riguarda le altre assunzioni della regressione possiamo osservare dall'analisi dei residui in Figura 4.14 che in Figura 4.16, che viene rispettata la normalità dei residui (dimostrata sia dall'istogramma che dal probability plot). Anche l'omoschedasticità, ovvero la distribuzione casuale dei residui senza tendenze, è rispettata e anche questa proprietà è riscontrabile dai grafici a destra delle Figure 4.14 4.16. Nonostante l'analisi dei residui abbia permesso di riscontrare risultati positivi e l'omoschedasticità è stata provata il modello non può essere considerato affidabile. L'obiettivo era quello di creare un modello che tenesse conto delle variabili gestionali operative nella previsione dei consumi, ovvero cercare di prevedere il consumo futuro basandosi anche sulle informazioni dei movimenti

Coefficients					
Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	15.08	3.06	4.92	0.000	
hour	0.327	0.142	2.31	0.022	2.54
pieces_inb	-0.000272	0.000230	-1.18	0.239	79929.88
packs_inb	0.00271	0.00230	1.18	0.240	93338.50
orders_inb	0.2219	0.0992	2.24	0.026	43.92
pieces_rel	-0.00516	0.00485	-1.06	0.288	25.16
packs_rel	0.00096	0.00411	0.23	0.816	14.97
orders_rel	0.1916	0.0965	1.99	0.048	15.11
ais_rel	0.0917	0.0460	1.99	0.047	2.53
col_rel	-0.01059	0.00728	-1.46	0.146	5.28
lev_rel	-0.0537	0.0518	-1.04	0.301	3.32
pieces_sto	0.000333	0.000572	0.58	0.560	845.74
packs_sto	-0.00277	0.00237	-1.17	0.243	14416.44
orders_sto	-0.060	0.157	-0.38	0.702	106.33
ais_sto	-0.0203	0.0216	-0.94	0.348	15.87
col_sto	-0.00970	0.00400	-2.42	0.016	14.04
lev_sto	0.0299	0.0325	0.92	0.358	26.04
pieces_pic	0.000388	0.000994	0.39	0.697	7.09
packs_pic	0.00207	0.00263	0.79	0.432	5.95
orders_pic	-0.0127	0.0546	-0.23	0.817	116.23
ais_pic	0.0069	0.0136	0.51	0.613	43.71
col_pic	-0.00290	0.00199	-1.46	0.146	30.78
lev_pic	0.0173	0.0154	1.13	0.261	50.11
pieces_out	-0.00001	0.00107	-0.01	0.991	126.47
packs_out	-0.00002	0.00100	-0.02	0.988	30.73
orders_out	0.0015	0.0124	0.12	0.905	94.63
temp_ext	0.3914	0.0902	4.34	0.000	1.95
temp_cell	1.48	1.00	1.48	0.140	2.04
temp_anticell	-1.622	0.749	-2.17	0.031	2.11

Figura 4.15: Valori della regressione "pulita"

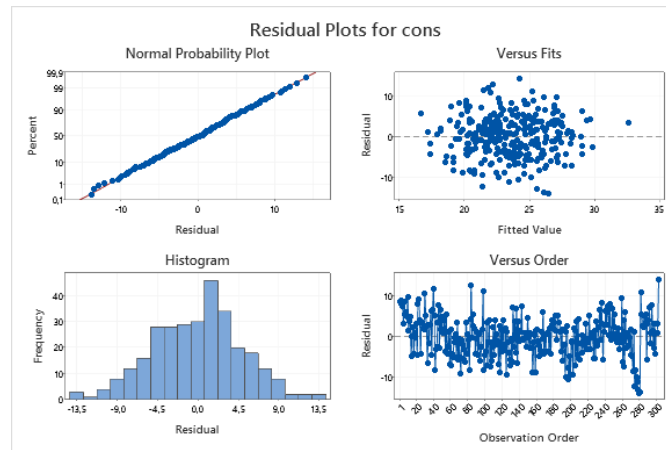


Figura 4.16: Analisi dei residui sulla regressione "pulita"

effettuati nella cella frigo e nell'anticella. Purtroppo l'analisi, anche dopo aver escluso la variabili che introducevano multicollinearità nel modello, ha mostrato tramite i p-value che la significatività delle variabili restanti non è sufficiente. Sicuramente l'alta correlazione presente fra molte delle informazioni del dataset ha impedito di sfruttare i dati raccolti.

## 4.2.1 Modello ARIMA

Non potendo strutturare in modo affidabile un modello di regressione lineare multipla, si è virato su un modello previsionale che tenesse conto della serie temporale relativa ai consumi del gruppo frigo con i dati raccolti. Questo modello mantiene sempre l'obiettivo di prevedere i consumi elettrici della cella frigo e

dell'anticella, ma solamente dall'analisi dei consumi passati senza l'inserimento nel modello di ulteriori variabili esplicative. Vedremo in seguito che, in realtà, verrà usata la variabile "temperatura esterna" per provare a migliorare il modello, siccome tra le variabili di consumo e di temperatura si è riscontrata una discreta dipendenza nell'andamento soprattutto della prima rispetto alla seconda. Si tratta a tutti gli effetti di un modello autoregressivo di cui precedentemente sono state descritte le caratteristiche nel capitolo 3.2.1.

Anche in questo caso, essendo una regressione, è necessario considerare solamente la parte stazionaria della serie temporale per due motivi:

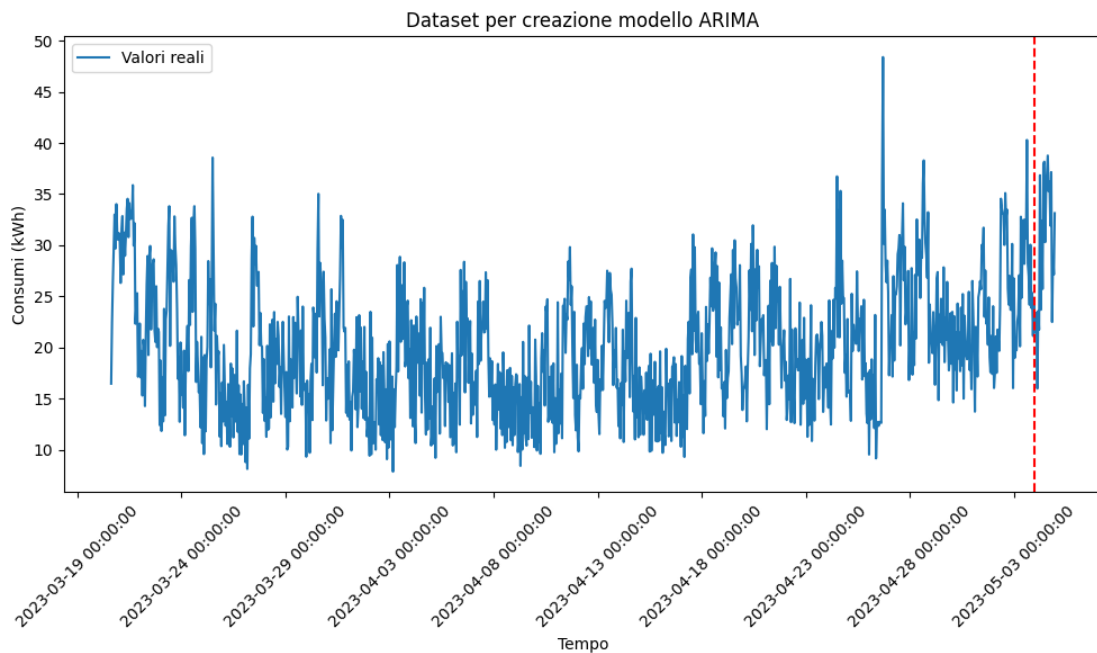
1. se ci fosse una componente di trend si dovrebbero applicare delle trasformazioni dette integrazioni;
2. considerando i flussi termici della struttura subentra il problema di parzializzazione dei consumi quando entra in gioco la gestione del "periodo estivo". Non avendo la possibilità di stimare la ripartizione dei consumi in tale periodo per singolo locale condizionato, presumiamo che tale periodo inizi con l'innalzamento dei consumi (di pari passo con quello delle temperature) come analizzato precedentemente.

Esposte queste premesse è stato ulteriormente eseguito il test ADF per valutare fino a che punto la serie temporale potesse considerarsi stazionaria considerando singoli giorni e non settimane intere. Il punto di rottura della stazionarietà è stato individuato nell'elemento del dataset corrispondente all'orario 7:00 del giorno 5 maggio 2023. Con il nuovo modello autoregressivo ci poniamo l'obiettivo di creare una previsione dei valori di consumo per il giorno successivo. Siccome i dati campionati sono orari le previsioni di forecast saranno sui 24 punti successivi. In questo modo per poter considerare dei blocchi di dati che permettano al modello di apprendere in maniera più completa è stato scelto di considerare come dati solo quelli fino alle ore 23:00 del 4 maggio, in modo da avere giornate complete (ovvero dalle ore 00:00 alle 23:00) tutte campionate. Quindi il dataset che andremo ad utilizzare per la creazione del modello va dal 20 marzo 2023 al 4 maggio 2023.

Per la creazione di un modello di Machine Learning generalmente la strategia usata è quella di dividere il dataset in due parti:

- il **training-set**: è la parte di dati più corposa ed è utilizzata per addestrare il modello, nel nostro caso sono i dati che il modello utilizza per stimare i coefficienti della regressione;
- il **test-set**: è la parte di dati che viene utilizzata per verificare la bontà del modello, serve per mettere in relazione le previsioni calcolate dal modello con i dati reali acquisiti in fase di raccolta.

Solitamente le proporzioni usate per la divisione train-test sono dell'ordine 70-30% o 80-20%. Nel caso studio in questione, a causa del numero esiguo di elementi nel dataset, è stata scelta una divisione che avvantaggiasse la fase di addestramento lasciando molte più istanze nel training-set. Il dataset stazionario utilizzato per la costruzione del modello presenta un numero di istanze pari a 1089, e considerando che il forecast viene fatto su 24 istanze è stato deciso di prendere quella come dimensione del test-set; pertanto i due dataset hanno un numero di istanze pari a 1065 per il training-set e 24 per il test-set (Figura 4.17).



**Figura 4.17:** Parte di dati stazionaria utilizzata per creare il modello, con separazione train-test set

Dopo aver definito le dimensioni dei vari dataset, il passo successivo è stato quello di scegliere i parametri del modello ARIMA. Tale modello che è da intendersi come Auto Regressive Integrated Moving Average è caratterizzato proprio da 3 parametri che ne compongono l'ordine:

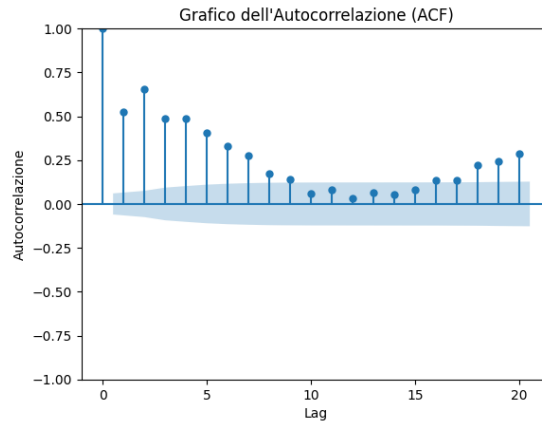
- **p**: i lag temporali a ritroso che vengono considerati per calcolare i coefficienti sulla variabile d'interesse (in questo caso il consumo del gruppo frigo);
- **d**: i gradi di differenziazione utilizzati per avere una serie temporale stazionaria;
- **q**: i lag temporali a ritroso che vengono considerati per stimare i coefficienti degli errori nella media mobile.

Come già spiegato per le scelte fatte in precedenza il parametro  $d$  è pari a 0, proprio perché si è deciso di considerare solamente la parte stazionaria della serie temporale, quindi mancando la parte di integrazione si può parlare di modello ARMA e non ARIMA. Per quanto riguarda i parametri  $p$  e  $q$  sono stati utilizzati diversi metodi per definirli. In primis sono stati sviluppati i grafici rappresentati nelle Figure 4.19, 4.18. Essi sono i grafici di "autocorrelazione" e "autocorrelazione parziale" definiti grazie alle funzioni ACF (Auto Correlation Function) e PACF (Partial Auto Correlation Function) presenti nel pacchetto *statsmodels*. Questi tipi di grafici in letteratura [95] sono utilizzati per comprendere quanti lag è necessario considerare in un modello ARMA e quindi definirne i parametri. Per andare ad individuare il valore di  $p$ , ovvero i lag della parte autoregressiva, si deve osservare il grafico PACF. Dalla Figura 4.19 si nota come per i primi 3 lag il valore di autocorrelazione è molto alto, tendente ad 1, mentre per il lag 4 seppur ridotto è ancora fuori dall'intervallo di confidenza, raffigurato dalla fascia azzurro chiaro. Generalmente si sceglie come valore di  $p$  l'ultimo lag fuori dall'intervallo, che in questo caso è il quarto. Per quanto riguarda la Figura 4.18 tramite la sua analisi permette di attribuire il numero di lag relativi alla parte di moving average del modello. Si nota in questo grafico un numero maggiore di lag con valori al di fuori dell'intervallo di confidenza, ma è possibile distinguere che dal lag numero 3 in poi vi è un trend decrescente verso lo 0. È possibile in questo caso considerare come lag per la parte MA il valore 3.

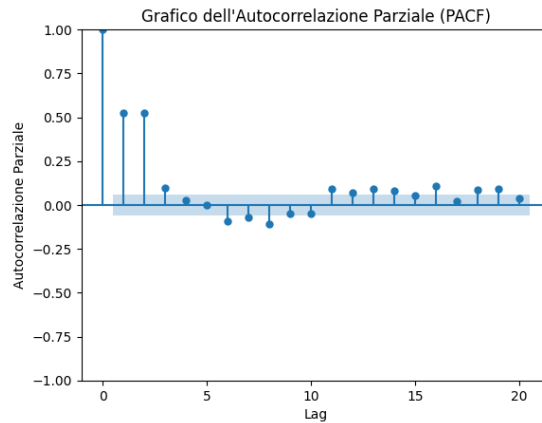
Per avere una conferma dell'analisi grafica effettuata sulle Figure 4.19, 4.18 è stata sfruttata la funzione del pacchetto *pmдарima.arima* che possiede al proprio interno il metodo *auto\_arima*. Questa funzione permette, in ambiente Python, di trovare il miglior modello ARIMA restituendo così i parametri di tale modello. Il Best Model viene scelto in base a quale metodo di calcolo si sceglie per l'ottimizzazione. I criteri di validazione utilizzabili sono 4: AIC, BIC, HQIC, OOB. Il più comune, utilizzato anche in questa sede, è il criterio AIC (Akaike's Information Criterion). È un metodo per la validazione e il confronto fra modelli statistici, il quale fornisce una misura della qualità della stima di un modello statistico tenendo conto sia della bontà di adattamento che della complessità del modello. Il punteggio è calcolato:

$$AIC = 2K - 2\ln(L) \quad AIC = 2\ln \frac{e^k}{L}$$

dove  $L$  è la stima della massima verosimiglianza. Il criterio di ottimizzazione prevede di scegliere il modello con il punteggio AIC più basso. Utilizzando la formula riscritta, si può vedere come il punteggio AIC del modello aumenterà in proporzione alla crescita del valore del numeratore, che contiene il numero di parametri nel modello  $k$  (cioè una misura della complessità del modello). Al contrario il punteggio AIC diminuirà in proporzione alla crescita del denominatore che contiene il logaritmo della massima verosimiglianza (MLE, Maximum Likelihood



**Figura 4.18:** Grafico di Autocorrelazione dei dati di training



**Figura 4.19:** Grafico di Autocorrelazione Parziale dei dati di training

Estimation) del modello, che è una misura della bontà di adattamento del modello [96].

Questo metodo è particolarmente utile per trovare gli ordini  $p$  e  $q$  del modello ma non  $d$ , siccome nel nostro caso  $d$  è già assegnato ed è pari a zero, l'AIC score è un buon modo di valutare i modelli con i vari parametri [97]. Il Best Model (Figura 4.20) è scelto dalla funzione *auto\_arima* come quello che minimizza il punteggio AIC. Dopo l'esecuzione del codice viene confermata l'analisi dei grafici PACF e ACF restituendo come miglior modello ARIMA(4,0,3). Data anche la conferma dalla stepwise search possiamo definire l'ordine del modello, ovvero ARIMA(4,0,3).

Analizzando l'andamento del consumo elettrico della cella frigo si è deciso di analizzare anche la parte stagionale della serie temporale, per verificare se potesse migliorare la modellazione e restituire in fase di forecasting un errore minore. In questo caso, andando a considerare anche la stagionalità dei dati, il modello non

```

Performing stepwise search to minimize aic
ARIMA(0,0,0)(0,0,0)[0] : AIC=9490.229, Time=0.03 sec
ARIMA(1,0,0)(0,0,0)[0] : AIC=6837.897, Time=0.11 sec
ARIMA(0,0,1)(0,0,0)[0] : AIC=8615.762, Time=0.16 sec
ARIMA(2,0,0)(0,0,0)[0] : AIC=inf, Time=0.12 sec
ARIMA(1,0,1)(0,0,0)[0] : AIC=6371.527, Time=0.18 sec
ARIMA(2,0,1)(0,0,0)[0] : AIC=6278.184, Time=0.20 sec
ARIMA(3,0,1)(0,0,0)[0] : AIC=6279.373, Time=0.39 sec
ARIMA(2,0,2)(0,0,0)[0] : AIC=6279.157, Time=0.30 sec
ARIMA(1,0,2)(0,0,0)[0] : AIC=6297.902, Time=0.25 sec
ARIMA(3,0,0)(0,0,0)[0] : AIC=inf, Time=0.18 sec
ARIMA(3,0,2)(0,0,0)[0] : AIC=6281.432, Time=0.64 sec
ARIMA(2,0,1)(0,0,0)[0] intercept : AIC=6226.346, Time=0.42 sec
ARIMA(1,0,1)(0,0,0)[0] intercept : AIC=6331.116, Time=0.55 sec
ARIMA(2,0,0)(0,0,0)[0] intercept : AIC=6234.670, Time=0.28 sec
ARIMA(3,0,1)(0,0,0)[0] intercept : AIC=6228.339, Time=0.77 sec
ARIMA(2,0,2)(0,0,0)[0] intercept : AIC=6228.333, Time=0.65 sec
ARIMA(1,0,0)(0,0,0)[0] intercept : AIC=6570.582, Time=0.13 sec
ARIMA(1,0,2)(0,0,0)[0] intercept : AIC=6240.781, Time=0.64 sec
ARIMA(3,0,0)(0,0,0)[0] intercept : AIC=6226.873, Time=0.47 sec
ARIMA(3,0,2)(0,0,0)[0] intercept : AIC=6215.793, Time=1.80 sec
ARIMA(4,0,2)(0,0,0)[0] intercept : AIC=6205.755, Time=2.44 sec
ARIMA(4,0,1)(0,0,0)[0] intercept : AIC=6230.125, Time=0.57 sec
ARIMA(5,0,2)(0,0,0)[0] intercept : AIC=6225.696, Time=2.57 sec
ARIMA(4,0,3)(0,0,0)[0] intercept : AIC=6193.160, Time=2.34 sec
ARIMA(3,0,3)(0,0,0)[0] intercept : AIC=6203.605, Time=2.14 sec
ARIMA(5,0,3)(0,0,0)[0] intercept : AIC=6204.024, Time=2.83 sec
ARIMA(4,0,4)(0,0,0)[0] intercept : AIC=6202.302, Time=3.22 sec
ARIMA(3,0,4)(0,0,0)[0] intercept : AIC=6193.769, Time=2.61 sec
ARIMA(5,0,4)(0,0,0)[0] intercept : AIC=6200.157, Time=2.93 sec
ARIMA(4,0,3)(0,0,0)[0] : AIC=6283.457, Time=0.71 sec

Best model: ARIMA(4,0,3)(0,0,0)[0] intercept
Total fit time: 30.663 seconds

```

Figura 4.20: Stepwise search per trovare il miglior ordine del modello ARMA

avrà più la denominazione ARIMA ma diventa qualificato come SARIMA (Seasonal AutoRegressive Integrated Moving Average). Includendo la stagionalità nel modello, tale caratteristica deve essere esaminata e anch'essa presenta un suo ordine con i relativi parametri da definire. Un modello SARIMA generico presenta i seguenti parametri:

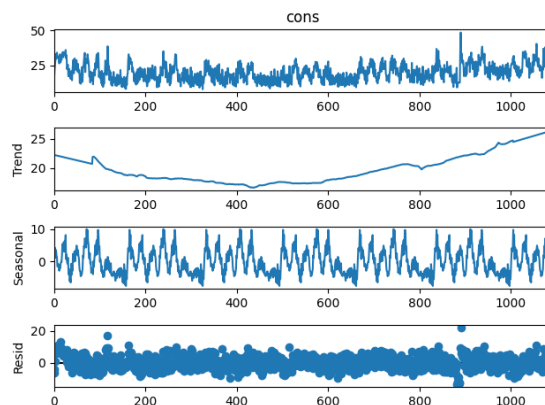
- ordine (p,d,q), che si riferisce all'ordine delle serie temporali. Questo ordine è utilizzato anche nel modello ARIMA (che non considera la stagionalità);
- ordine stagionale (P,D,Q,s), che si riferisce all'ordine della componente stagionale della serie storica.

Per una migliore comprensione di seguito sono commentati i vari termini del "seasonal order" [98]:

- **P** (Seasonal Auto Regressive order): è il numero di ritardi stagionali nell'autoregressione. In altre parole, indica quanti passi indietro nel tempo si stanno considerando come input stagionale. Ad esempio, se  $P$  è 1, il modello considererà il valore dell'osservazione stagionale al passo precedente;
- **D** (Seasonal Integrated order): rappresenta il numero di volte che è necessario differenziare la parte stagionale della serie temporale per renderla stazionaria. In altre parole, quanti passaggi di differenziazione stagionale sono necessari per stabilizzare la varianza o la tendenza stagionale;
- **Q** (Seasonal Moving Average order): è il numero di ritardi stagionali nella media mobile. Indica quanti valori passati dell'errore di previsione stagionale vengono utilizzati per prevedere il valore attuale;
- **s** (Seasonal Periodicity): rappresenta la lunghezza del periodo stagionale nella serie temporale, la periodicità. Ad esempio, se la serie fosse mensile e presentasse una stagionalità annuale, allora  $s$  sarebbe 12. Nel caso studio in questione la serie temporale è su circa due mesi con dati orari, perciò la periodicità  $s$  sarà pari a 24.

La specifica dell'ordine stagionale è importante nel modello SARIMA perché tiene conto della natura ciclica o stagionale dei dati. Identificare correttamente gli ordini stagionali è cruciale per ottenere previsioni più accurate e coerenti con la struttura dei dati.

Per effettuare l'analisi di stagionalità come primo step la serie temporale presa in esame, sempre con l'aiuto della libreria *statsmodels*, è stata decomposta (Figura 4.21) nelle sue componenti di *trend*, *stagionalità*, *residui* [98]. Effettuata la scompo-



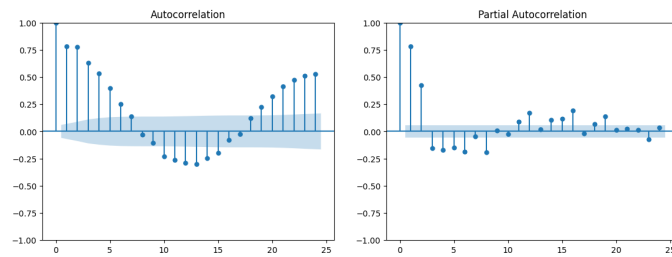
**Figura 4.21:** Scomposizione della serie temporale

sizione il primo parametro che si va ad analizzare è quello dell'integrazione, ovvero



$D$ . Come per il parametro  $d$  del normale ordine ARIMA, anche di questo ne viene incrementato il valore se la serie non risulta stazionaria, e quindi necessita della differenziazione per essere resa tale. In questa analisi però la serie temporale presa in considerazione è solo la componente stagionale della decomposizione, e di essa viene proprio testata, come in precedenza, la stazionarietà tramite il test d'ipotesi ADF. Come risultato il test ha rigettato l'ipotesi nulla  $H_0$  di "non stazionarietà" con un  $p$ -value pari a 0.0034, ciò sta a significare che la componente stagionale è stazionaria e non necessita di ulteriori trasformazioni, traducendosi nel valore del parametro  $D = 0$ .

Proseguendo nell'analisi dell'ordine di stagionalità per i parametri  $P$  e  $Q$  generalmente si procede allo studio dei grafici PACF (per il parametro  $P$ ) e ACF (per il parametro  $Q$ ) della sola parte stagionale, in modo da rilevare i lag significativi per le componenti stagionali di AR e MA. Purtroppo l'analisi grafica (Figura 4.22)



**Figura 4.22:** Grafico ACF componente stagionale, Grafico PACF componente stagionale

non permette una chiara distinzione dei parametri stagionali  $P$  e  $Q$ . Per avere una valutazione più attendibile nella scelta dell'ordine stagionale ci si è serviti di un codice che sostanzialmente replica la funzionalità della precedente funzione `auto_arima`, andando a calcolare l'AIC score di vari modelli, ma sta volta iterando solamente sui parametri stagionali  $P$  e  $Q$  e mantenendo fissi tutti gli altri ai valori già decretati. È stato inoltre calcolato per ognuno dei modelli l'errore RMSE per metterli in relazione su due misure di performance. Il risultato dell'analisi è stato che il Best Model per entrambe le misure è stato:

$$SARIMA(4,0,3)(1,0,2,24) \quad \text{con} \quad AIC = 6158.37867 \quad \text{e} \quad RMSE = 4.342875$$

In definitiva i parametri stagionali scelti sono:

$$P = 1, D = 0, Q = 2, s = 24.$$

Terminata la ricerca dei parametri del nostro modello il passo successivo è addestrare tutti i modelli e di ognuno calcolarne le misure di performance, che per un modello di regressione equivale a computare gli errori tra i valori reali e i valori di previsione prodotti dal modello.

Prima di proporre l'analisi delle performance dei modelli vi è da fare una precisazione su quali modelli verranno esaminati. Si è deciso per completezza di considerare anche un modello che comprendesse la variabile esogena *temperatura esterna* visto che l'andamento del consumo elettrico del gruppo frigo è altamente relazionato a quello della variabile temperatura per il periodo di osservazione. Esaminato il caso studio questa è la variabile, di cui possediamo i dati, che maggiormente può aiutare a spiegare la varianza della nostra variabile target, e di conseguenza potrebbe nel modello di forecasting produrre una previsione più accurata.

Pertanto i modelli che si andranno ad esaminare sono i seguenti:

- ARIMA(4,0,3)
- ARIMAX(4,0,3)<sup>2</sup>
- SARIMA(4,0,3)(1,0,2,24)

e per ognuno dei modelli verranno mostrate le performance sia dell'addestramento (fitting) sia del forecast dei 24 valori.

Le misure di performance scelte per confrontare i risultati dei modelli sono quelli più comuni utilizzate in questi contesti come R<sup>2</sup> score, MSE (Mean Square Error), RMSE (Root Mean Square Error), MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error). Di seguito è proposta una breve descrizione matematica degli indicatori [99]. Nella formulazione seguente, X<sub>i</sub> è il valore i-esimo previsto e l'elemento Y<sub>i</sub> è il valore i-esimo effettivo (reale). Vengono definite due costanti: la media dei valori reali

$$\bar{Y} = \frac{1}{m} \sum_{i=1}^m Y_i$$

e la media della somma totale dei quadrati (Mean Squares Total)

$$MST = \frac{1}{m} \sum_{i=1}^m (Y_i - \bar{Y})^2$$

Il coefficiente di determinazione o meglio conosciuto come R<sup>2</sup> può essere interpretato come la porzione di varianza della variabile dipendente che è prevedibile dalle variabili indipendenti. Ha un range di valori che va dal limite inferiore pari a  $-\infty$  al limite superiore +1. Il valore di questo indicatore pari a 1 viene raggiunto quando l'adattamento del modello è perfetto e la varianza della variabile dipendente è completamente spiegata dalle variabili indipendenti. Sebbene non ci sia un valore

---

<sup>2</sup>La "X" sta a significare che il modello prende in esame anche la variabile esogena

numerico che limita inferiormente la misura, il valore 0 corrisponde ad un pessimo adattamento del modello e non si perde informazione se si considera solamente l'intervallo (0,1) poiché tutti i valori negativi indicano un adattamento peggiore rispetto alla linea del valor medio.

$$R^2 = 1 - \frac{\sum_{i=1}^m (X_i - Y_i)^2}{\sum_{i=1}^m (\bar{Y} - Y_i)^2}$$

Al contrario della valutazione di  $R^2$ , l'output di MSE, RMSE, MAE e MAPE abbraccia l'intero ramo positivo dei valori reali, con il limite inferiore a 0 che implica un adattamento perfetto, e valori che crescono progressivamente e all'infinito per i modelli con prestazioni peggiori. Per definizione, questi valori dipendono fortemente dagli intervalli delle variabili descrittive, rendendoli incomparabili sia tra loro che all'interno della stessa metrica: un dato valore di output per una metrica non ha alcuna relazione interpretabile con un valore simile per una misura diversa, e anche lo stesso valore per la stessa metrica può riflettere prestazioni del modello profondamente diverse per due attività distinte.

$$MSE = \frac{1}{m} \sum_{i=1}^m (X_i - Y_i)^2$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (X_i - Y_i)^2}$$

Poiché  $R^2 = 1 - \frac{MSE}{MST}$  e poiché MST è fissato per i dati che si analizzano,  $R^2$  è monotonicamente correlato a MSE (una relazione monotona negativa), il che implica che un ordinamento di modelli di regressione basati su  $R^2$  sarà identico (anche se in ordine inverso) a un ordinamento di modelli basati su MSE o RMSE.

$$MAE = \frac{1}{m} \sum_{i=1}^m |X_i - Y_i|$$

$$MAPE = \frac{1}{m} \sum_{i=1}^m \left| \frac{Y_i - X_i}{Y_i} \right|$$

MAPE è un'altra metrica di prestazione per i modelli di regressione, con un'interpretazione molto intuitiva in termini di errore relativo: per la sua definizione, il suo utilizzo è consigliato in compiti in cui è più importante essere sensibili alle variazioni relative più che a quelle assolute [99].

Gli indicatori MSE, RMSE, MAE e MAPE sono estremamente dipendenti dai range delle variabili descrittive e avendo un range di valori che va da 0 a  $+\infty$  diventa difficile prenderli in considerazione come metriche per una valutazione assoluta su modelli differenti. Questa assunzione invece può essere fatta per quanto

riguarda la misura dell' $R^2$  che ha le proprietà per permettere di paragonare tra loro due modelli differenti tramite il suo valore. La valutazione sui modelli proposti in questo elaborato comunque tiene in considerazione tutte queste misure perché in tutti i modelli i range delle variabili descrittive non cambiano.

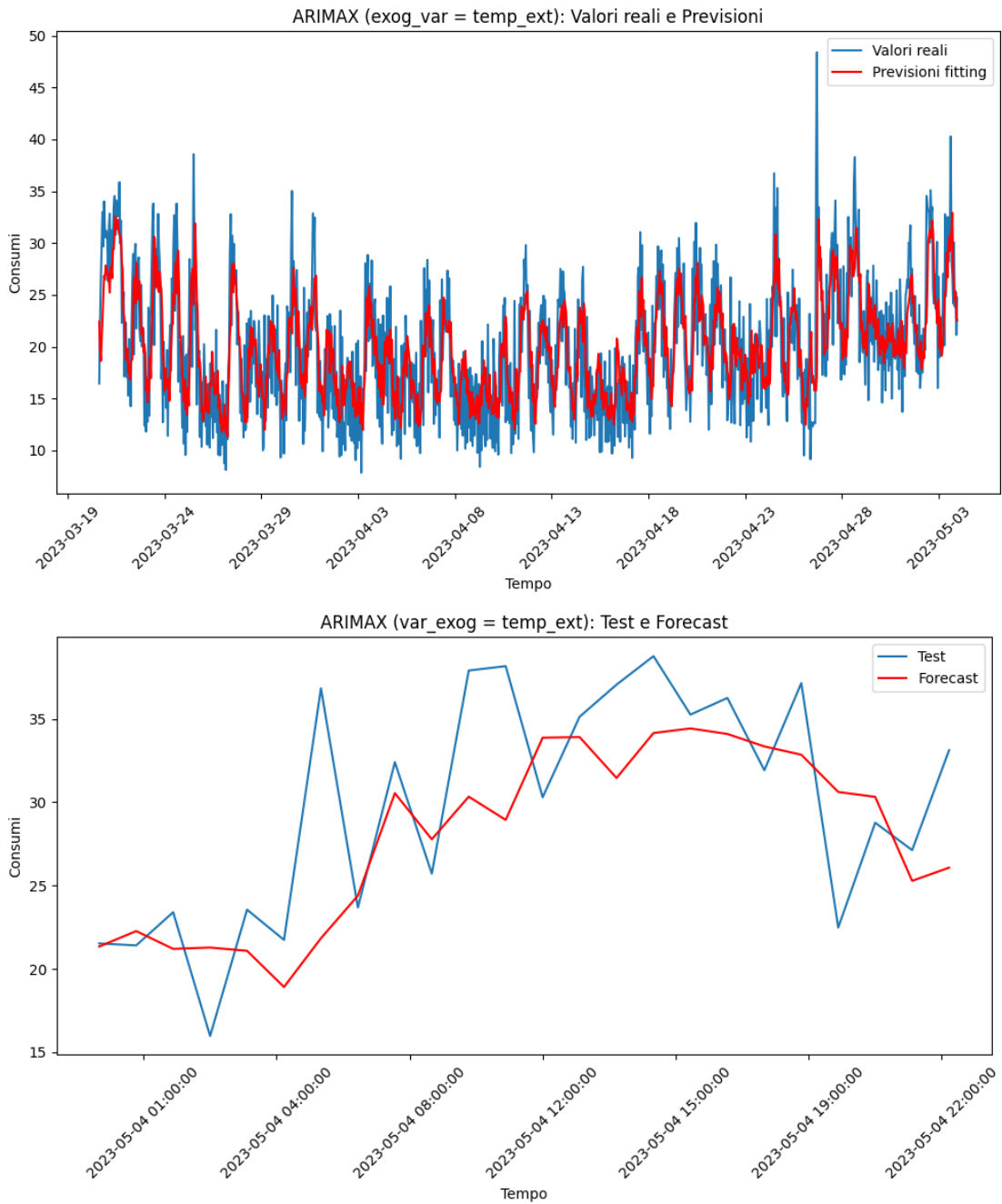
Dopo aver eseguito il codice Python che ha permesso, tramite tutte le librerie, funzioni e metodi esposti nel capitolo 3.3, di creare i modelli di forecast con i vari parametri precedentemente trovati e tutti i relativi grafici per eseguire in conclusione l'analisi delle performance. Sono stati integrati nel codice i calcoli per l'output di tali performance presentate in Figura 4.23 La tabella dei risultati

Modello		MSE	RMSE	MAE	MAPE	R2
Valori di addestramento	ARIMA(4,0,3)	19.97892	4.46978	3.57142	19.95174	0.47926
	ARIMAX(4,0,3)	18.51599	4.30302	3.38393	18.87882	0.51739
	SARIMA(4,0,3)(1,0,2,24)	18.62644	4.31584	3.39931	19.06043	0.48833
Valori forecast	ARIMA(4,0,3)	31.02022	5.56958	4.12016	24.72854	0.31366
	ARIMAX(4,0,3)	26.55976	5.15362	3.85612	25.23609	0.41235
	SARIMA(4,0,3)(1,0,2,24)	26.58069	5.15565	3.92868	24.76859	0.41189

**Figura 4.23:** Misure di performance dei modelli analizzati

mostra che per ogni modello sono state effettuate le valutazioni sui dati di training e su quelli di test: per i dati di addestramento le performance sono state calcolate tra i valori reali dei consumi nel training-set e i fitted-values calcolati dal modello, mentre per i 24 valori di cui è stata fatta la previsione le performance sono state calcolate tra i valori nel test-set e i valori di forecast creati tramite il modello. come si può notare sono evidenziati i valori migliori per ogni misura di performance e, a parte per il MAPE relativo alla parte di forecast, si riscontra che il modello ARIMAX(4,0,3) è quello che presenta decisamente i migliori risultati sia per il training che per il test set. Di seguito viene illustrato il confronto grafico tra i valori di training-fitting e i valori di test-forecast (Figura 4.24)

I risultati appena presentati mostrano come la correlazione tra la temperatura e il consumo elettrico della cella frigo sia importante nel forecast dei consumi, e in misura maggiore della considerazione di stagionalità nel modello permette di effettuare misure più precise. A valle di queste valutazioni è da sottolineare che nonostante la maggiore precisione, anche nel modello più performante riscontriamo misure di errore relativamente alte. Come anticipato precedentemente per i valori di MSE, RMSE, MAE, MAPE non possiamo fare valutazioni assolute rispetto alla bontà del modello, ma per i valori di  $R^2$  score si. In particolare si riscontra un valore ridotto per tale indicatore pari a 0.51739 per l'addestramento e 0.41235 per il forecast.



**Figura 4.24:** Rappresentazione dei valori del modello ARIMAX(4,0,3) per confronto training-fitting e test-forecast

Il modello così definito riesce a realizzare una previsione che spiega solamente il 41% della varianza del consumo elettrico della cella frigo, e come si può comprendere il

livello di accuratezza è basso per una misura di questa portata. Sicuramente i dati a disposizione per la costruzione del modello sono abbastanza esigui e questo è uno dei motivi per il basso livello di accuratezza. Per quanto riguarda la stagionalità anche in questo caso la dimensione del dataset influisce parecchio, perché con un più ampio intervallo di osservazione la stagionalità può portare ad una accuratezza maggiore catturando migliori trend e ripetitività lungo l'arco temporale esaminato. Nella gamma dei modelli autoregressivi è poi possibile creare l'integrazione di entrambe le componenti, considerando sia la stagionalità che la variabile esogena in un modello chiamato SARIMAX che però deve essere supportato da una quantità di rilevazioni necessariamente maggiore di quella presente in questo caso studio.

La conclusione che scaturisce dall'analisi effettuata e dai risultati proposti è che ci sia la necessità di considerare altre variabili per avere una previsione accurata del consumo elettrico della cella frigo, poiché la sola temperatura esterna e i consumi precedenti non sono sufficienti. Nell'analisi dei flussi termici del caso studio si è appurato che la possibilità di avere la parzializzazione dei consumi elettrici relativi al condizionamento dei singoli locali è particolarmente rilevante principalmente per riuscire a considerare un periodo di raccolta dati più esteso, vista la possibilità di includere i mesi estivi. La struttura della cella frigo e dell'anticella, grazie alla coibentazione, rende chiaro che la dissipazione del calore avviene per mezzo dello sbocco delle baie di carico verso l'ambiente esterno e la loro apertura dipende dalle attività operative svolte all'interno di questi locali. È perciò fondamentale riuscire a manipolare, in modo più efficiente di come fatto in questa sede, i dati gestionali operativi che descrivono le attività che vengono svolte nei suddetti locali per incrementare la qualità del forecast (banalmente per avere un  $R^2$  score maggiore), e ricavare delle feature significative da inserire nel modello previsionale.

Si ricorda che l'obiettivo di realizzare una previsione affidabile è sorto per poter, in un secondo momento, utilizzare questa informazione in un ottica di inserimento della struttura logistica in una configurazione di Comunità Energetica Rinnovabile. Grazie ad un forecast accurato sarebbe possibile attivare azioni di Demand Response e accoppiare in modo efficace un eventuale produzione fotovoltaica, ottimizzando sia richiesta che produzione con le esigenze della comunità. Tutto ciò però richiede la premessa fondamentale di una previsione accurata. In studi successivi a questo sarà necessario non solo avere la possibilità di separare i consumi dei diversi locali per avere un periodo di osservazione più ampio, ma si può considerare di utilizzare anche modelli di Machine Learning più complessi (come ad esempio le Artificial Neural Network) che possano elaborare meglio le informazioni gestionali estrapolabili dal sistema WMS dell'azienda.

# Capitolo 5

## Conclusioni

La presente tesi si è posta l'intento di proporre un modello di forecasting dei consumi elettrici relativi al raffreddamento della cella frigo di un magazzino logistico refrigerato. Lo studio ha come scopo originale quello di inserire come variabili indipendenti nel modello previsionale informazioni di carattere gestionale operativo riguardanti la cella frigo del magazzino. Il caso studio è stato affrontato con l'idea che la creazione di un modello di forecasting possa permettere, in un'ottica di inserimento in una Comunità Energetica Rinnovabile, un'ottimizzazione tra consumo e produzione energetica.

Nella prima parte è stata affrontato l'argomento delle Comunità Energetiche sotto il profilo normativo europeo e nazionale. Si è potuto riscontrare come vi sia una forte spinta delle istituzioni verso le configurazioni di condivisione dell'energia atte a promuovere la decentralizzazione della produzione energetica. Questa diffusione si porta dietro numerosi vantaggi economici come la riduzione dei costi dell'energia (soprattutto per trasporto e distribuzione) e profitti dalla condivisione della stessa, ma anche risvolti sociali positivi come la diminuzione della povertà energetica (ad oggi coinvolge l'8% delle famiglie italiane) e un maggior tasso di occupazione. I benefici derivanti dalla diffusione delle CER riguardano anche per il sistema elettrico diminuendo le perdite di rete e aumentando stabilità e affidabilità e portano ad una maggiore accettazione delle tecnologie rinnovabili da parte dei cittadini.

La tecnologia diventa alleata di questi obiettivi perché tramite l'utilizzo di Intelligenza Artificiale e tecniche di Machine Learning è possibile creare dei modelli di forecasting che permettano di avere informazioni sui consumi energetici futuri. Grazie a questi tipi di modelli vi è la possibilità di associare le informazioni di consumo con quelle di produzione dell'energia nelle Comunità Energetiche in modo da ottimizzare le esigenze dei membri della comunità. Tale potrebbe essere lo scopo ultimo di analisi future del caso studio, che grazie ad un modello di previsione dei consumi relativo alla cella frigo del magazzino (sfruttando le informazioni gestionali) renderebbe possibile abbinarne la produzione fotovoltaica e in un contesto CER

si potrebbero progettare eventi di Demand Response o comunque si potrebbe ottimizzare la produzione di energia a servizio dei consumi del magazzino e quando le necessità energetiche della struttura non sono elevate (surplus di produzione) dirottare queste risorse verso la comunità o in sistemi di accumulo a disposizione della stessa.

Andando ad analizzare la filiera in cui opera il magazzino del caso studio, ovvero quella alimentare, si è potuto riscontrare una notevole incisività dei consumi dovuti alla refrigerazione lungo tutta quella che viene definita Cold Supply Chain. L'incisività di questi consumi porta non solamente a problematiche di tipo economico per gli agenti del settore, ma va anche a colpire tematiche quali il riscaldamento climatico, lo spreco di alimenti (nel 2017 il 12% del cibo prodotto nel mondo è stato scartato per insufficienze nella catena del freddo) e in generale di risorse quali anche l'energia. Si può evidenziare come in una Cold Supply Chain circa il 70% dei SEC sono relativi a consumi elettrici dei sistemi di refrigerazione e per una struttura come un centro di distribuzione l'incidenza può essere del 10-20% di tali consumi.

Per la creazione del modello di forecasting a supporto del caso studio ci si è avvalsi di una metodologia semplice di Machine Learning come la Regressione Lineare Multipla, visto il carattere innovativo dello studio. È stata comunque presentata un'introduzione sulla materia Intelligenza Artificiale e sono state descritte le principali tecniche di Machine Learning e di analisi delle serie temporali.

Infine è stato descritto il caso studio partendo dall'inquadramento dell'azienda per avere un'idea del servizio e delle attività svolte e poi delineando i caratteri significativi che possono incidere sul modello di forecasting. In tale ottica è stata fornita una dettagliata descrizione dei locali afferenti alla struttura del magazzino e soprattutto dei flussi termici relativi al raffreddamento dei diversi locali. Il condizionamento di questi ultimi è generato da un gruppo frigorifero (chiller) unico di cui si è brevemente riportato il funzionamento. Dopo aver illustrato la metodologia di raccolta dei dati e quali tipologie di dati sarebbero stati considerati nel modello, si è proseguito nell'analisi di regressione. Proprio per la definizione di un modello di Regressione Lineare Multipla è necessario che vengano rispettate alcune assunzioni fondamentali come: assenza di multicollinearità tra le variabili indipendenti, relazione lineare tra variabili indipendenti e quella dipendente, presenza di omoschedasticità e distribuzione normale dei residui. Partendo dall'analisi di multicollinearità e servendosi di strumenti grafici come il correlogramma e analitici come il calcolo del VIF si sono andate ad escludere tutte le variabili che maggiormente erano correlate tra loro e avrebbero reso difficile la stima dei regressori. Si è così arrivati ad avere un pool di variabili che rispettassero il vincolo del  $VIF < 10$ , la relazione lineare, l'omoschedasticità e la distribuzione normale dei residui. Nonostante ciò non è stato possibile costruire il modello a causa della non significatività delle variabili gestionali, poiché tutte quelle sopravvissute



all'analisi di multicollinearità presentavano dei p-value decisamente maggiori di 0.05, il che le rende non significative al fine di rappresentare le feature in un modello di forecasting. L'unica variabile che in tutte le regressioni tentate è sempre risultata significativa è la temperatura esterna, che come anticipato nell'analisi dei dati ha una forte relazione con i consumi di elettricità della cella frigo. Si è deciso quindi di proporre un modello autoregressivo ARIMA per provare a prevedere i consumi della cella frigo del magazzino. Per costruire il modello è stata studiata la stazionarietà dei dati di consumo elettrico e i grafici ACF e PACF per definire i vari ordini di autoregressione, integrazione e media mobile. Da tale analisi e grazie al supporto di un codice Python è stato individuato il modello ARIMA(4,0,3). Per provare a ricavare una previsione più accurata sono stati integrati due modelli ulteriori che tenevano conto uno della variabile esogena "temperatura esterna" e uno della stagionalità. È stata effettuata quindi l'analisi sulla stagionalità della serie temporale presa in considerazione che ha portato alla definizione dei rispettivi ordini. Sono stati eseguiti i codici Python dei tre modelli di forecast ARIMA(4,0,3), ARIMAX(4,0,3) e SARIMA(4,0,3)(1,0,2,24) che hanno permesso il calcolo della previsione dei consumi elettrici per il giorno seguente, ovvero 24 istanti di tempo in avanti. Di questi modelli sono state calcolate le misure di performance identificate negli indicatori  $R^2$ , MSE, RMSE, MAE, MAPE per poterli confrontare tra loro e darne una valutazione di bontà predittiva. Il modello che ha mostrato le performance migliori è ARIMAX(4,0,3) a conferma della stretta relazione tra consumi e temperatura esterna nel caso studio. Nonostante questo risultato il miglior modello sia nella fase di training che in quella di forecast, spicca all'occhio il valore del  $R^2$  score (0.51739 per l'addestramento, 0.41235 per il forecast). Questa misura è significativa della limitata bontà della previsione spiegando solamente il 41% della varianza dei dati di consumo, valore troppo basso in un'applicazione concreta come potrebbe essere quella dell'implementazione di una CER.

Lo studio effettuato in questo elaborato ha portato alla luce le difficoltà nella costruzione di un modello di forecasting dei consumi di una cella frigo in un magazzino refrigerato che vuole considerare le variabili gestionali. Il modello che viene proposto mostra tutte le sue lacune in una poco accurata previsione rispecchiata dal valore del  $R^2$  score e risaltando la necessità di considerare le variabili gestionali. Nonostante lo studio non sia stato in grado di proporre il modello originariamente pensato, ha sicuramente evidenziato le problematiche da risolvere in futuri lavori per effettuare analisi di maggior rilievo.

Essendo quello analizzato un modello data-driven, si è riscontrata in particolare la necessità, di una più grande quantità di dati che si traduce in un maggiore arco temporale di rilevazione per diverse motivazioni: sia per avere l'opportunità di considerare i trend di stagionalità presenti in serie temporali più estese, sia per avere un campionamento che osservi i consumi in tutti i periodi dell'anno anche per più anni. Per poter estendere il rilevamento dei dati è necessario riuscire ad avere la

parzializzazione dei consumi di raffreddamento per ogni singolo locale, problematica risolvibile solamente con l'installazione di appositi strumenti di misurazione della potenza termica.

Una considerazione da far risaltare riguarda la forte correlazione tra la temperatura esterna e i consumi, questa viene ricondotta alla presenza delle baie di carico presenti nell'anticella e nel magazzino 2. Infatti anche se si avesse la possibilità di avere le misurazioni parzializzate dei locali il problema della forte dissipazione di calore sia verso l'esterno tramite le baie sia quella passiva tra il magazzino 2 e il magazzino 1 non sarebbe risolto. Nell'ottica di inserimento della struttura in una CER, il primo passo da fare è sicuramente quello di ridurre queste dissipazioni rendendo più efficienti i consumi per il raffreddamento dei locali. A questo proposito si potrebbero sfruttare alcune soluzioni come l'installazione di porte taglia aria sia dalle baie sia tra i due magazzini, ciò contribuirebbe alla riduzione delle dissipazioni. In seguito a questi efficientamenti si potrebbe eventualmente pensare di intervenire, post analisi, sullo scheduling logistico tramite il forecasting dei consumi elettrici per il raffreddamento in modo da poterli ottimizzare. In un ancor successivo step sarebbe dunque plausibile arrivare ad integrare il forecasting con la produzione di energia (esempio fotovoltaico) e sfruttare i risultati a servizio di una Comunità Energetica.

Osservando il dataset creato per la Regressione Lineare Multipla (consumi, temperature, dati gestionali) emerge una problematica, ovvero la presenza di numerosi zeri per alcune attività gestionali in determinate fasce orarie della giornata. In particolare si riscontra che le operazioni di inbound e stoccaggio vengono effettuate nella mattina fino circa alle ore 13, mentre le operazioni di picking e outbound sono prevalentemente eseguite nel pomeriggio dopo le ore 15. Essendo il dataset composto da colonne che rappresentano le attività operative e da record corrispondenti ad un'ora di un determinato giorno, è naturale che in ogni riga si presentino tanti campi con valori pari a zero, perché con tale organizzazione operativa le diverse attività sono svolte in determinate fasce orarie (se sono ore mattutine le colonne corrispondenti a OUT e PIC saranno nulle e nelle ore pomeridiane le colonne INB e STO saranno nulle). La presenza di numerosi campi nulli in un modello semplice come quello di una regressione può creare problemi nella stima dei coefficienti e nella loro significatività. Una possibile soluzione potrebbe essere quella di creare modelli previsionali diversi per fasce orarie che includono gruppi di attività eseguite in quel lasso di tempo, in modo da ovviare alla presenza di numerosi zeri nei campi del dataset somministrati al programma di Machine Learning.

Un'altra tematica da sottoporre in studi futuri è la possibilità di impiegare tecniche di Machine Learning più complesse per riuscire a spiegare in modo più completo le relazioni tra le variabili gestionali e il consumo elettrico per il raffreddamento della cella frigo. Sicuramente l'utilizzo di tecniche della famiglia ANN (Artificial Neural Network), possedendo una potenza descrittiva e una complessità maggiori,

può migliorare il livello di performance di un modello di forecasting come quello che si è cercato di proporre in questo elaborato.

# Bibliografia

- [1] Enrico Giarmanà. «Il diritto a produrre energia e el nascenti figure dei prosumers e delle renewable energy communities». In: *ReteAmbiente* (2020) (cit. a p. 3).
- [2] Ministero dello Sviluppo Economico. «Piano Nazionale Integrato per l’Energia e il Clima». In: <https://www.mise.gov.it/> (2019) (cit. alle pp. 5, 21, 23).
- [3] openpolis. «Decreto legge 31 maggio 2021, n. 77 Governance del Piano nazionale di ripresa e resilienza e prime misure di rafforzamento delle strutture amministrative e di accelerazione e snellimento delle procedure». In: *Gazzetta ufficiale della repubblica italiana* (2021) (cit. a p. 6).
- [4] openpolis. «Cos’è il Pnrr, piano nazionale ripresa e resilienza». In: (2021). URL: <https://www.openpolis.it/parole/cose-il-pnrr-piano-nazionale-ripresa-e-resilienza/> (cit. a p. 6).
- [5] «Direttiva (UE) 2018/2001 del Parlamento europeo e del Consiglio dell’11 dicembre 2018 sulla promozione dell’uso dell’energia da fonti rinnovabili». In: *OJ L 328/82* (2018) (cit. alle pp. 7, 10).
- [6] *Decreto legislativo 8 novembre 2021, n. 210 Attuazione della direttiva UE 2019/944, del Parlamento europeo e del Consiglio, del 5 giugno 2019, relativa a norme comuni per il mercato interno dell’energia elettrica e che modifica la direttiva 2012/27/UE, nonché recante disposizioni per l’adeguamento della normativa nazionale alle disposizioni del regolamento UE 943/2019 sul mercato interno dell’energia elettrica e del regolamento UE 941/2019 sulla preparazione ai rischi nel settore dell’energia elettrica e che abroga la direttiva 2005/89/CE. (21G00233)*. 2021 (cit. a p. 8).
- [7] *Direttiva (UE) 2019/944 del Parlamento europeo e del Consiglio del 5 giugno 2019 relativa a norme comuni per il mercato interno dell’energia elettrica*. 2019 (cit. alle pp. 8, 11).

- [8] *Decreto legge 30 dicembre 2019, n. 162 coordinato con la Legge di conversione 28 febbraio 2020, n. 8 recante: «Disposizioni urgenti in materia di proroga di termini legislativi, di organizzazione delle pubbliche amministrazioni, nonché di innovazione tecnologica.»* 2019 (cit. a p. 9).
- [9] *Decreto legislativo 8 novembre 2021, n. 199 Attuazione della direttiva (UE) 2018/2001 del Parlamento europeo e del Consiglio.* 2021 (cit. a p. 10).
- [10] AIEE Federmanager. «Il ruolo delle comunità energetiche nel processo di transizione verso la decarbonizzazione». In: (2021) (cit. alle pp. 11, 13, 15, 17, 21).
- [11] Autorità di Regolazione per l'Energia Reti e Ambiente. *Deliberazione 27 dicembre 2022, 727/2022/R/EEL di ARERA. Testo Integrato Autoconsumo Diffuso (TIAD).* 2022. URL: <https://www.arera.it/> (cit. alle pp. 11, 12).
- [12] GSE. «Rapporto statistico 2021 - Energia da fonti rinnovabili in Italia». In: (2023). URL: <https://www.gse.it/> (cit. alle pp. 13–15).
- [13] GSE. «Rapporto statistico Solare Fotovoltaico 2022». In: (2023). URL: <https://www.gse.it/> (cit. alle pp. 15, 16).
- [14] GSE. «Energia e Clima in Italia - rapporto trimestrale (Q4/2022)». In: (2023). URL: <https://www.gse.it/> (cit. alle pp. 16, 23).
- [15] Autorità di Regolazione per l'Energia Reti e Ambiente. *Deliberazione 4 agosto 2020, 318/2020/R/EEL, Regolazione delle partite economiche relative all'energia elettrica condivisa da un gruppo di autoconsumatori di energia rinnovabile che agiscono collettivamente in edifici e condomini oppure condivisa in una comunità di energia rinnovabile.* 2020. URL: <https://www.arera.it/> (cit. a p. 18).
- [16] *Decreto 16 settembre 2020 Individuazione della tariffa incentivante per la remunerazione degli impianti a fonti rinnovabili inseriti nelle configurazioni sperimentali di autoconsumo collettivo e comunità energetiche rinnovabili, in attuazione dell'articolo 42-bis, comma 9, del decreto-legge n. 162/2019, convertito dalla legge n. 8/2020.* 2020 (cit. a p. 19).
- [17] Dipartimento energia Ministero dell'ambiente e della sicurezza energetica. *Consultazione pubblica: Attuazione della disciplina per la regolamentazione degli incentivi per la condivisione dell'energia di cui all'articolo 8 del decreto legislativo 8 novembre 2021, n.199 (Comunità energetiche e sistemi di autoconsumo – impianti di potenza fino a 1 MW).* 2022. URL: <https://www.mase.gov.it/notizie/comunita-energetiche-mase-al-la-consultazione-pubblica-sul-decreto-di-erogazione-incentivi> (cit. a p. 19).

- 
- [18] Energy & Strategy Group. «Le ricadute energetiche, ambientali e sociali con lo sviluppo delle Comunità energetiche in Italia». In: (2020). URL: <https://www.energystrategy.it/> (cit. a p. 20).
- [19] Raffaele Miniaci e Paola Valbonesi Ivan Faiella Luciano Lavecchia. «La povertà energetica in Italia, Secondo rapporto dell'Osservatorio Italiano sulla Povertà Energetica (OIPE)». In: <https://oipeosservatorio.it/publicazioni> (2020) (cit. a p. 21).
- [20] Davide Chiaroni, Federico Frattini, Simone Franzo et al. «Smart Grid Report. Le prospettive di sviluppo delle Energy Community in Italia». In: (2014) (cit. alle pp. 21, 23).
- [21] Vehbi C Gungor, Dilan Sahin, Taskin Kocak, Salih Ergut, Concettina Buccella, Carlo Cecati e Gerhard P Hancke. «Smart grid technologies: Communication technologies and standards». In: *IEEE transactions on Industrial informatics* 7.4 (2011), pp. 529–539 (cit. a p. 24).
- [22] Thamer Alquthami, Muhammad Zulfiqar, Muhammad Kamran, Ahmad H Milyani e Muhammad Babar Rasheed. «A performance comparison of machine learning algorithms for load forecasting in smart grid». In: *IEEE Access* 10 (2022), pp. 48419–48433 (cit. a p. 24).
- [23] Fahad Javed, Naveed Arshad, Fredrik Wallin, Iana Vassileva e Erik Dahlquist. «Forecasting for demand response in smart grids: An analysis on use of anthropologic and structural data and short term multiple loads forecasting». In: *Applied Energy* 96 (2012), pp. 150–160 (cit. a p. 24).
- [24] Ruben Barreto, Luis Gomes e Zita Vale. «Demand Response-based Model for an Energy Community Considering Members and Electric Vehicles Participation». In: *2023 IEEE International Conference on Environment and Electrical Engineering and 2023 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I<sup>2</sup>CPS Europe)*. IEEE. 2023, pp. 1–6 (cit. a p. 25).
- [25] Johannes Radl, Andreas Fleischhacker, Frida Huglen Revheim, Georg Lettner e Hans Auer. «Comparison of profitability of PV electricity sharing in renewable energy communities in selected European countries». In: *Energies* 13.19 (2020), p. 5007 (cit. a p. 25).
- [26] Matteo Moncecchi, Stefano Meneghello e Marco Merlo. «A game theoretic approach for energy sharing in the italian renewable energy communities». In: *Applied Sciences* 10.22 (2020), p. 8166 (cit. a p. 25).
- [27] A Cielo, P Margiaria, P Lazzeroni, I Mariuzzo e M Repetto. «Renewable Energy Communities business models under the 2020 Italian regulation». In: *Journal of Cleaner Production* 316 (2021), p. 128217 (cit. a p. 25).

- [28] Luyao Zou, Md Shirajum Munir, Kitae Kim e Choong Seon Hong. «Day-ahead energy sharing schedule for the P2P prosumer community using LSTM and swarm intelligence». In: *2020 International Conference on Information Networking (ICOIN)*. IEEE. 2020, pp. 396–401 (cit. a p. 26).
- [29] Action Plan et al. «Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions». In: *European Commission* (2011) (cit. a p. 27).
- [30] Zhiping Cheng, Dongqiang Jia, Zhongwen Li, Shuai Xu e Jikai Si. «Multi-time-scale energy management for microgrid using expected-scenario-oriented stochastic optimization». In: *Sustainable Energy, Grids and Networks* 30 (2022), p. 100670 (cit. a p. 27).
- [31] Dominik Putz, Michael Gumhalter e Hans Auer. «The true value of a forecast: Assessing the impact of accuracy on local energy communities». In: *Sustainable Energy, Grids and Networks* 33 (2023), p. 100983 (cit. a p. 27).
- [32] Girolama Airò Farulla, Giovanni Tumminia, Francesco Sergi, Davide Aloisio, Maurizio Cellura, Vincenzo Antonucci e Marco Ferraro. «A review of key performance indicators for building flexibility quantification to support the clean energy transition». In: *Energies* 14.18 (2021), p. 5676 (cit. a p. 27).
- [33] Mattia Pasqui, Alex Felice, Maarten Messagie, Thierry Coosemans, Tommaso Tiozzo Bastianello, Duccio Baldi, Pietro Lubello e Carlo Carcasci. «A new smart batteries management for Renewable Energy Communities». In: *Sustainable Energy, Grids and Networks* 34 (2023), p. 101043 (cit. a p. 27).
- [34] Francesco Lombardi, Sergio Balderrama, Sylvain Quoilin e Emanuela Colombo. «Generating high-resolution multi-energy load profiles for remote areas with an open-source stochastic model». In: *Energy* 177 (2019), pp. 433–444 (cit. a p. 27).
- [35] Alejandro Pena-Bello, David Parra, Mario Herberz, Verena Tiefenbeck, Martin K Patel e Ulf JJ Hahnel. «Integration of prosumer peer-to-peer trading decisions into energy community modelling». In: *Nature Energy* 7.1 (2022), pp. 74–82 (cit. a p. 28).
- [36] Elissaios Sarmas, Nikos Dimitropoulos, Vangelis Marinakis, Zoi Mylona e Haris Doukas. «Transfer learning strategies for solar power forecasting under data scarcity». In: *Scientific Reports* 12.1 (2022), p. 14643 (cit. a p. 28).
- [37] Andrea Petrucci, Giovanni Barone, Annamaria Buonomano e Andreas Athienitis. «Modelling of a multi-stage energy management control routine for energy demand forecasting, flexibility, and optimization of smart communities using a Recurrent Neural Network». In: *Energy Conversion and Management* 268 (2022), p. 115995 (cit. a p. 28).

- [38] Aya Abdelsalam Ismail, Timothy Wood e Héctor Corrada Bravo. «Improving long-horizon forecasts with expectation-biased LSTM networks». In: *arXiv preprint arXiv:1804.06776* (2018) (cit. a p. 28).
- [39] Andrea Maria NC Ribeiro, Pedro Rafael X do Carmo, Patricia Takako Endo, Pierangelo Rosati e Theo Lynn. «Short-and Very Short-Term Firm-Level Load Forecasting for Warehouses: A Comparison of Machine Learning and Deep Learning Models». In: *Energies* 15.3 (2022), p. 750 (cit. a p. 29).
- [40] John Fernie e Leigh Sparks. *Logistics and retail management: insights into current practice and trends from leading experts*. Kogan Page Publishers, 2004 (cit. a p. 30).
- [41] Kerstin Gustafsson, Gunilla Jönson, David Smith e Leigh Sparks. *Retailing logistics and fresh food packaging: managing change in the supply chain*. Kogan Page Publishers, 2006 (cit. a p. 30).
- [42] Nodali Ndraha, Hsin-I Hsiao, Jelena Vlajic, Ming-Feng Yang e Hong-Ting Victor Lin. «Time-temperature abuse in the food cold chain: Review of issues, challenges, and recommendations». In: *Food Control* 89 (2018), pp. 12–21 (cit. alle pp. 30, 31, 33, 34).
- [43] Roberto Montanari. «Cold chain tracking: a managerial perspective». In: *Trends in Food Science & Technology* 19.8 (2008), pp. 425–431 (cit. a p. 31).
- [44] J Sarr, JL Dupont e J Guilpart. «The Carbon Footprint of the Cold Chain, 7th Informatory Note on Refrigeration and Food». In: *International Institute of Refrigeration: Paris, France* (2021) (cit. alle pp. 31, 32).
- [45] Beatrice Marchi e Simone Zanoni. «Cold Chain Energy Analysis for Sustainable Food and Beverage Supply». In: *Sustainability* 14.18 (2022), p. 11137 (cit. alle pp. 31, 33–36).
- [46] Codex Alimentarius Commission. *Codex Alimentarius, International Food Standards*. 2023. URL: <https://www.fao.org/fao-who-codexalimentarius> (cit. alle pp. 31, 40).
- [47] Chinaza Godswill Awuchi. «HACCP, quality, and food safety management in food and agricultural systems». In: *Cogent Food & Agriculture* 9.1 (2023), p. 2176280 (cit. alle pp. 31, 39).
- [48] DNV. *Lo standard ISO 22000*. 2023. URL: <https://www.dnv.it/services/lo-standard-iso-22000-5173> (cit. alle pp. 31, 40).
- [49] Jenny Gustavsson, Christel Cederberg, Ulf Sonesson, Robert Van Otterdijk e Alexandre Meybeck. *Global food losses and food waste*. 2011 (cit. a p. 31).
- [50] CCAC. *Climate & Clean Air Coalition*. 2023. URL: <https://www.ccacoalition.org/fr/sleps/hydrofluorocarbons-hfcs> (cit. a p. 32).



- [51] AA Kefalidou. «Sustainable Energy Solutions to ‘Cold Chain’ Food Supply Issues». In: *Brief for GSDR–2016 Update*. (2022). URL: [https://sustainabledevelopment.un.org/content/documents/968624%5C\\_Kefalidou%5C\\_Sustainable%5C%%20energy%5C%%20solutions%5C%%20to-cold%5C%%20chain-food%5C%%20supply%5C%%20issues.pdf](https://sustainabledevelopment.un.org/content/documents/968624%5C_Kefalidou%5C_Sustainable%5C%%20energy%5C%%20solutions%5C%%20to-cold%5C%%20chain-food%5C%%20supply%5C%%20issues.pdf) (cit. a p. 32).
- [52] Seth-Oscar Tromp, René Haijema, Hajo Rijgersberg e Jack GAJ van der Vorst. «A systematic approach to preventing chilled-food waste at the retail outlet». In: *International Journal of Production Economics* 182 (2016), pp. 508–518 (cit. a p. 33).
- [53] Onrawee Laguerre, Steven Duret, Hong-Minh Hoang e Denis Flick. «Using simplified models of cold chain equipment to assess the influence of operating conditions and equipment design on cold chain performance». In: *International journal of refrigeration* 47 (2014), pp. 120–133 (cit. a p. 33).
- [54] Evelyne Derens, B Palagos e J Guilpart. «The cold chain of chilled products under supervision in France». In: (2006), pp. 823–823 (cit. a p. 33).
- [55] Nga Thi Tuyet Mai, BJÖRN MARGEIRSSON, Sveinn Margeirsson, Sigurdur Gretar Bogason, SJÖFN SIGURGÍSLADÓTTIR e Sigurjon Arason. «Temperature mapping of fresh fish supply chains—air and sea transport». In: *Journal of food process engineering* 35.4 (2012), pp. 622–656 (cit. a p. 33).
- [56] J Lundén, V Vanhanen, K Kotilainen e K Hemminki. «Retail food stores’ internet-based own-check databank records and health officers’ on-site inspection results for cleanliness and food holding temperatures reveal inconsistencies». In: *Food Control* 35.1 (2014), pp. 79–84 (cit. a p. 33).
- [57] T Brown, NA Hipps, S Easteal, A Parry e JA Evans. «Reducing domestic food waste by lowering home refrigerator temperatures». In: *International journal of refrigeration* 40 (2014), pp. 246–253 (cit. a p. 33).
- [58] Chih-Wen Shih e Chih-Hsuan Wang. «Integrating wireless sensor networks with statistical quality control to develop a cold chain system in food industries». In: *Computer Standards & Interfaces* 45 (2016), pp. 62–78 (cit. a p. 34).
- [59] Sunny George Gwanpua et al. «The FRISBEE tool, a software for optimising the trade-off between food quality, energy use, and global warming impact of cold chains». In: *Journal of Food Engineering* 148 (2015), pp. 2–12 (cit. a p. 34).
- [60] Fabio Monforti-Ferrario et al. *Energy use in the EU food sector: State of play and opportunities for improvement*. Publications Office Luxembourg, 2015 (cit. a p. 34).

- [61] Beatrice Marchi, Laura Bettoni e Simone Zanoni. «Assessment of Energy Efficiency Measures in Food Cold Supply Chains: A Dairy Industry Case Study». In: *Energies* 15.19 (2022), p. 6901 (cit. alle pp. 34, 36, 37).
- [62] Beatrice Marchi e Simone Zanoni. «Energy efficiency in cold supply chains of the food and beverage sector». In: *Transportation Research Procedia* 67 (2022), pp. 56–62 (cit. a p. 36).
- [63] Ernst Worrell, John A Laitner, Michael Ruth e Hodayah Finman. «Productivity benefits of industrial energy efficiency measures». In: *Energy* 28.11 (2003), pp. 1081–1098 (cit. a p. 37).
- [64] Lisa Neusel e Simon Hirzel. «Energy efficiency in cold supply chains of the food Sector: An exploration of conditions and perceptions». In: *Cleaner Logistics and Supply Chain* 5 (2022), p. 100082 (cit. a p. 37).
- [65] *Regolamento (CE) n. 178/2002 del Parlamento europeo e del Consiglio, del 28 gennaio 2002, che stabilisce i principi e i requisiti generali della legislazione alimentare, istituisce l’Autorità europea per la sicurezza alimentare e fissa procedure nel campo della sicurezza alimentare.* 2002 (cit. a p. 38).
- [66] *Regolamento (CE) n. 852/2004 del Parlamento europeo e del Consiglio, del 29 aprile 2004, sull’igiene dei prodotti alimentari.* 2004 (cit. alle pp. 38, 39).
- [67] *Rettifica del regolamento (CE) n. 853/2004 del Parlamento europeo e del Consiglio, del 29 aprile 2004, che stabilisce norme specifiche in materia di igiene per gli alimenti di origine animale.* 2004 (cit. a p. 38).
- [68] *Decreto legislativo 26 maggio 1997, n. 155 Attuazione delle direttive 93/43/CEE e 96/3/CE concernenti l’igiene dei prodotti alimentari.* 1997 (cit. a p. 38).
- [69] Wikipedia. *HACCP — Wikipedia, L’enciclopedia libera.* [Online; in data 21-aprile-2023]. 2023. URL: [//it.wikipedia.org/w/index.php?title=HACCP&oldid=132995410](https://it.wikipedia.org/w/index.php?title=HACCP&oldid=132995410) (cit. a p. 39).
- [70] Geeks for Geeks. *Artificial Intelligence | An Introduction.* 2023. URL: <https://www.geeksforgeeks.org/artificial-intelligence-an-introduction/> (cit. a p. 41).
- [71] Geeks for Geeks. *An introduction to Machine Learning.* 2023. URL: <https://www.geeksforgeeks.org/introduction-machine-learning/> (cit. alle pp. 43, 44).
- [72] Ronald van Loon. *A Machine Learning Explained: Understanding Supervised, Unsupervised, and Reinforcement Learning.* 2018. URL: <https://www.linkedin.com/pulse/machine-learning-explained-understanding-supervised-ronald-van-loon/> (cit. a p. 44).
- [73] Francois Chollet. *Deep learning with Python.* Simon e Schuster, 2021 (cit. a p. 45).

- 
- [74] Ian Goodfellow, Yoshua Bengio e Aaron Courville. *Deep learning*. MIT press, 2016 (cit. a p. 46).
- [75] Di Nuzzo Michele. *Data Science e Machine Learning: Dai dati alla conoscenza*. Independently Published, 2021 (cit. alle pp. 47–49).
- [76] Python Software Foundation. *Python Beginner’s Guide*. 2023. URL: <https://wiki.python.org/moin/BeginnersGuide> (cit. a p. 54).
- [77] Inc. NumFOCUS. *Pandas User Guide*. 2023. URL: [https://pandas.pydata.org/docs/user\\_guide/index.html#user-guide](https://pandas.pydata.org/docs/user_guide/index.html#user-guide) (cit. a p. 55).
- [78] NumPy Developers. *NumPy user guide*. 2022. URL: <https://numpy.org/doc/1.25/user/index.html#user> (cit. a p. 55).
- [79] scikit-learn developers (BSD License). *Scikit-learn Machine Learning in Python*. 2023. URL: <https://scikit-learn.org/stable/> (cit. a p. 56).
- [80] John Hunter, Darren Dale, Eric Firing, Michael Droettboom e Matplotlib development team. *Using Matplotlib*. 2023. URL: <https://matplotlib.org/stable/users/index.html> (cit. a p. 57).
- [81] Josef Perktold, Skipper Seabold, Jonathan Taylor e statsmodels-developers. *Statsmodels - statistical models, hypothesis tests, and data exploration*. 2023. URL: <https://www.statsmodels.org/stable/index.html> (cit. a p. 58).
- [82] Anas M Atieh, Hazem Kaylani, Yousef Al-Abdallat, Abeer Qaderi, Luma Ghoul, Lina Jaradat e Iman Hdairis. «Performance improvement of inventory management system processes by an automated warehouse management system». In: *Procedia Cirp* 41 (2016), pp. 568–572 (cit. a p. 60).
- [83] Frigofluid cooling systems. *Chiller a compressione: come funzionano*. 2023. URL: <https://www.frigofluid.it/risorse/articoli-tecnici/chiller-a-compressione-come-funzionano/> (cit. a p. 66).
- [84] Nadia Nedjah, Luiza de Macedo Mourelle e Marcelo Silveira Dantas Lizarazu. «Mathematical modeling of chiller-based refrigeration systems for energy efficiency optimization». In: *Computational Science and Its Applications-ICCSA 2022 Workshops: Malaga, Spain, July 4-7, 2022, Proceedings, Part I*. Springer. 2022, pp. 273–288 (cit. a p. 66).
- [85] Matteo Corti. «Modelli semplificati delle proprietà di fluidi refrigeranti per la simulazione rapida di cicli frigoriferi a compressione di vapore». In: (2012) (cit. a p. 69).
- [86] Luca Massalongo. *la scheda di trasporto* (cit. a p. 75).

- [87] Donald P Lynch, William S Duff e JE Fleischmann. «A guideline for using experimental design mathematical modeling in welding of coil lead frames». In: *Proceedings: Electrical Insulation Conference and Electrical Manufacturing and Coil Winding Conference (Cat. No. 01CH37264)*. IEEE. 2001, pp. 133–140 (cit. a p. 78).
- [88] Paolo Mattana. *Il test Dickey-Fuller*. 2019. URL: [https://www.performancetrading.it/Documents/PmEconometria/PmEq\\_Dickey\\_Fuller.htm](https://www.performancetrading.it/Documents/PmEconometria/PmEq_Dickey_Fuller.htm) (cit. a p. 80).
- [89] Zoumana Keita. *Multiple Linear Regression in R: Tutorial With Examples. A complete overview to understanding multiple linear regressions in R through examples*. 2022. URL: <https://www.datacamp.com/tutorial/multiple-linear-regression-r-tutorial> (cit. a p. 81).
- [90] Aniruddha Bhandari. *Multicollinearity | Causes, Effects and Detection Using VIF*. 2023. URL: <https://www.analyticsvidhya.com/blog/2020/03/what-is-multicollinearity/> (cit. alle pp. 81, 83).
- [91] Mavuto M Mukaka. «A guide to appropriate use of correlation coefficient in medical research». In: *Malawi medical journal* 24.3 (2012), pp. 69–71 (cit. a p. 82).
- [92] Patrick Schober, Christa Boer e Lothar A Schwarte. «Correlation coefficients: appropriate use and interpretation». In: *Anesthesia & analgesia* 126.5 (2018), pp. 1763–1768 (cit. a p. 82).
- [93] Sara Dadipoor, Azin Alavi, Zainab Kader, Shokrollah Mohseni, Hadi Eshaghi Sani Kakhaki e Nahid Shahabi. «Predictive power of PEN-3 cultural model in cervical cancer screening among women: a cross-sectional study in South of Iran». In: *BMC cancer* 23.1 (2023), pp. 1–12 (cit. a p. 83).
- [94] Rui Li, Dongdong Ye, Qiukun Zhang, Jianfei Xu e Jiabao Pan. «Nondestructive Evaluation of Thermal Barrier Coatings’ Porosity Based on Terahertz Multi-Feature Fusion and a Machine Learning Approach». In: *Applied Sciences* 13.15 (2023), p. 8988 (cit. a p. 83).
- [95] Prabhat Pathak. *How to Create an ARIMA Model for Time Series Forecasting in Python*. 2020. URL: <https://www.analyticsvidhya.com/blog/2020/10/how-to-create-an-arima-model-for-time-series-forecasting-in-python/> (cit. a p. 88).
- [96] Hirotogu Akaike. «Information theory and an extension of the maximum likelihood principle». In: *Selected papers of hirotugu akaike*. Springer, 1998, pp. 199–213 (cit. a p. 89).
- [97] George Athanasopoulos Rob J Hyndman. *Forecasting: principles and practice*. 2022. URL: <https://otexts.com/fppit/> (cit. a p. 89).

- [98] Angelica Lo Duca. *Understanding the Seasonal Order of the SARIMA Model*. 2021. URL: <https://towardsdatascience.com/understanding-the-seasonal-order-of-the-sarima-model-ebef613e40fa> (cit. alle pp. 90, 91).
- [99] Davide Chicco, Matthijs J Warrens e Giuseppe Jurman. «The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation». In: *PeerJ Computer Science* 7 (2021), e623 (cit. alle pp. 93, 94).