POLITECNICO DI TORINO

Corso di Laurea Magistrale in
Ingegneria Energetica e Nucleare

Tesi di Laurea Magistrale

Sviluppo di algoritmi di machine learning per la previsione dei consumi di una centrale frigorifera industriale

Relatore

Prof. Marco Badami

Candidato

Marco Mozzi

Novembre 2022

Sviluppo di algoritmi di machine learning per la previsione dei consumi di una centrale frigorifera industriale

ABSTRACT

Negli ultimi anni si sta prendendo sempre più consapevolezza della criticità e dell'impatto del cambiamento climatico sul nostro pianeta. In questo contesto la gestione dell'energia e la "transizione energetica" hanno ormai assunto centralità nel dibattito pubblico e istituzionale. Al fine di garantire una transizione energetica efficace, è necessario affiancare all'aumento di produzione di energia pulita, una gestione efficiente dell'energia a disposizione. Nel presente elaborato sono riportati i risultati di uno studio condotto su un impianto di refrigerazione composto da tre gruppi frigo. Il lavoro è stato sviluppato a partire dall'analisi impiantistica e dell'architettura di monitoraggio energetico, fino ad arrivare alla modellazione e previsione dei suoi consumi di energia elettrica. Attraverso l'implementazione di algoritmi predittivi statistici e di machine learning, capaci di individuare pattern e correlazioni puntuali a partire da grandi moli di dati, sono stati sviluppati diversi modelli, caratterizzati da tecniche e finalità differenti. Per la realizzazione di questi, è stato necessario analizzare e studiare le variabili fisiche e temporali proprie del sistema. Sfruttando questo tipo di strumenti, è stato possibile predire secondo diverse prospettive temporali i consumi del sistema in esame, oltre che stimare le quote di consumo allocabili alle diverse utenze servite dal sistema frigorigeno. Infine, si è andati ad esplorare le opportunità applicative e le potenzialità a livello industriale derivanti da una analisi e conoscenza profonda di questi tipi di sistemi.

Development of machine learning algorithms for consumption forecasting of an industrial refrigeration plant

ABSTRACT

In recent years there has been increasing awareness of the criticality and impact of climate change on our planet. In this context, energy management and the concept "energy transition" have now taken centrality in the public and institutional debate. In order to ensure an effective energy transition, increasing clean energy production must be accompanied by efficient management of available energy. This paper reports the results of a study conducted on a refrigeration plant consisting of three refrigeration units. The work was developed starting from the plant and energy monitoring architecture analysis to the achievement of the modeling and prediction of its power consumption. Through the implementation of predictive statistical and machine learning algorithms, capable of detecting patterns and correlations from large amount of data, several models were developed, characterized by different techniques and purposes. For the realization of these, it was necessary to analyze and study the physical and temporal variables peculiar to the system. Taking advantage of these types of tools, it was possible to predict according to different temporal perspectives the consumption of the system under consideration, as well as to estimate the consumption shares allocable to the different utilities served by the refrigeration system. In conclusion, it has been carried out the evaluation of the application opportunities and potential at the industrial level arising from a deep analysis and knowledge of these types of systems.

Indice

1. Il quadro energetico ambientale 1
1.1 L'accordo di Parigi e il Green Deal europeo 2
1.2 L'impatto della refrigerazione e del condizionamento 5
2. Descrizione del sito industriale 7
2.1 Il processo produttivo 7
2.2 I consumi di stabilimento 9
2.2.1 I consumi elettrici 9
2.2.2 I consumi di combustibile 11
2.2.3 I consumi globali 12
2.2.4 I consumi delle UTA di stabilimento e dei gruppi frigo 14
3. Acquisizione dei dati e pre-processing 18
3.1 L'architettura di monitoraggio 18
3.2 Analisi dei dati di consumo elettrico dei gruppi frigo 20
3.3 Pre-processing 26
3.3.1 Outliers e Outlier Detection 26
3.3.2 Gestione dei missing values 27
3.4 Metriche Prestazionali 28
4. Modelli 30
4.1 Modelli di regressione lineare 30
4.1.1 Modello di regressione lineare semplice 31
4.1.2 Modello di regressione lineare multipla 35
4.1.3 Modello di regressione lineare per previsione dei consumi UTA 37
4.2 Decision Tree 40
4.2.1 Funzionamento algoritmo Decision Tree 40
4.2.2 Applicazione modello Decision Tree 41
4.3 Random Forest 47
4.3.1 Funzionamento algoritmo Random Forest 47
4.3.2 Modello annuale – Random Forest 48
4.3.3 Modelli dei consumi di UTA e macchinari – Random Forest 50
4.3.4 Ricapitolazione modelli – Random Forest 57
4.4 Reti Neurali Artificiali 59
4.4.1 Funzionamento algoritmo ANN 59
4.4.2 Preparazione dati 61
4.4.3 Modello annuale – ANN 64

4.4.4 Modelli dei consumi di UTA e macchinari – ANN....70
4.4.5 Modello 24h – ANN....73
4.4.6 Ricapitolazione modelli – ANN....76
5. Opportunità applicative dei modelli implementati....77
5.1 Riassunto delle performance dei modelli....77
5.2 Benchmarking interno: 2021-2022....78
5.3 Possibili implementazioni per l'ottimizzazione....83
6. Conclusioni....85
Glossario....86
Bibliografia....87

Indice delle Figure

Figura 1 - Emissioni di GHG (Greenhouse Gases) totali, 1990-2021, EGR 2022 1
Figura 2 - Emissioni totali di CO_2 per energia da combustione e processi industriali e la loro variazione annuale [2] 1
Figura 3 - Accordo di Parigi, COP21, 12 dicembre 2015 2
Figura 4 - Green Deal europeo, 11 dicembre 2019 [6].....3
Figura 5 - Illustrazione del Green Deal europeo, dalla comunicazione della Commissione Europea COM(2019) 640 3
Figura 6 - Industria 4.0 [8].....4
Figura 7 - Consumo mondiale di energia per il raffrescamento di locali interni negli edifici [9] 5
Figura 8 - Consumo energetico finale per il raffrescamento di locali per regione e numero di unità operanti nel Net Zero Scenario, 2000-2030 5
Figura 9 - Emissioni di CO_2 e intensità di emissione per il condizionamento dell'aria nel Net Zero Scenario, 2000-2030 [11] 6
Figura 10 - Schema del processo produttivo.....8
Figura 11 - Andamento mensile dei consumi elettrici nel periodo gennaio 2016 - giugno 2019 9
Figura 12 - Curva di assorbimento di potenza elettrica nell'anno 2018 9
Figura 13 - Curva di durata della potenza elettrica dell'anno 2018.....10
Figura 14 - Heatmap prelievi elettrici del 2018 10
Figura 15 - Andamento mensile dei consumi di gas naturale e kerosene nel periodo Gennaio 2016 - Giugno 2019.....11
Figura 16 - Curva di prelievo orario della potenza di combustibile dell'anno 2018 11
Figura 17 - Ripartizione dei consumi globali del 2018 da bolletta (dx) e in energia primaria (sx).....12
Figura 18 - Suddivisione dei consumi elettrici del 2018 12
Figura 19 - Suddivisione dei consumi elettrici del 2018 per area funzionale.....13
Figura 20 - Profili di carico delle UTA nelle prime due settimane di settembre 2021 15
Figura 21 - Profili di temperatura delle UTA 1 e 2 e della temperatura esterna nelle prime due settimane di settembre 2021.....15
Figura 22 - Profilo di carico e di temperatura della UTA1 e locale 1, prime due settimane di settembre 2021 16
Figura 23 - Profili di carico dei gruppi frigo, prime due settimane di settembre 2021 16
Figura 24 - Schema logico della domanda aziendale di liquido refrigerante 17
Figura 25 - Schema logico dell'interfaccia tra software e sistema di campo 19
Figura 26 - Profili di temperatura esterna e potenza media oraria dei gruppi frigo nel periodo 2020-2022 20
Figura 27 - Distinzione dei profili di carico frigo in funzionamento "estivo" e "invernale".....21
Figura 28 - Heatmaps rappresentanti i consumi orari frigo e la temperatura media oraria esterna del mese di ottobre.....21
Figura 29 - Heatmaps rappresentanti i consumi orari frigo e la temperatura media oraria esterna del mese di luglio.....22
Figura 30 - Confronto dei profili settimanali di consumo dei gruppi frigo per tipo di funzionamento.....23

Figura 31 - Boxplot dei consumi frigo orari per giorno della settimana in funzionamento invernale....24
Figura 32 - Boxplot dei consumi frigo orari per giorno della settimana in funzionamento estivo....24
Figura 33 - Boxplot orari giornalieri tipici per i due tipi di funzionamento dei gruppi frigo...25
Figura 34 - Schema esemplificativo outlier puntuale....26
Figura 35 - Schema esemplificativo sequenza anomala....26
Figura 36 - Schema esemplificativo di interpolazione lineare per la gestione dei missing values....27
Figura 37 - Parallelismo tra profili di temperatura esterna e potenza media oraria associata ai gruppi frigo....30
Figura 38 - Scatter plot della potenza media oraria associata ai gruppi frigo in funzione della temperatura oraria nel periodo estate 2021....31
Figura 39 - Rappresentazione range orario giornaliero di riferimento per modello di regressione lineare....32
Figura 40 - Scatter plot e rette di regressione considerando i valori compresi nel range orario di riferimento....33
Figura 41 - Schema logico rappresentante variabili di input e variabili di output nel modello34
Figura 42 - Rappresentazione del rapporto tra variabili climatiche sfasate e potenza oraria ...34
Figura 43 - Rappresentazione logica delle variabili di input e output al modello di regressione lineare multipla....35
Figura 44 - Rappresentazione 3D della regressione lineare multipla....36
Figura 45 - Schema logico ripartizione dei consumi frigo....37
Figura 46 - Centroide rappresentante il profilo tipo dei consumi frigo invernali....37
Figura 47 - Rappresentazione della quota di consumo frigo allocabile ai macchinari in funzionamento estivo....38
Figura 48 - Schema semplificativo della struttura di un CART....40
Figura 49 - Schema logico rappresentante le variabili di input al modello CART....42
Figura 50 - Rappresentazione grafica del modello CART generato....43
Figura 51 - Confronto tra profili reali e predetti per il periodo di training e di testing....44
Figura 52 - Confronto tra profilo reale e profilo predetto dal modello CART di una settimana invernale (sx) ed estiva (dx)....44
Figura 53 - Schema logico rappresentante le variabili di input al modello Random Forest 48
Figura 54 - Rappresentazione grafica di cinque generici CART costituenti il modello Random Forest....49
Figura 55 - Confronto tra profilo reale e predetto dal modello Random Forest della 3° e 4° settimana di settembre 2021....50
Figura 56 - Procedimento di creazione di un modello predittivo per le UTA....52
Figura 57 - Schema logico delle variabili di input al modello di previsione dei consumi frigo dei macchinari....52
Figura 58 - Cumulata dei consumi frigorigeni allocabili a macchinari e UTA di un generico periodo di funzionamento estivo....53
Figura 59 - Stima della quota di consumo allocabile alle UTA in un generico periodo del funzionamento estivo....54
Figura 60 - Procedimento di testing complessivo dei due modelli....55
Figura 61 - Consumo frigorigeno reale (linea scura) e l'aggregato dei consumi (aree colorate), Random Forest....56

Figura 62 - Profili di consumo allocabili alle UTA per gruppi frigo e ventilatori.....57
Figura 63 - Rappresentazione layers di una rete neurale.....59
Figura 64 - Input e Output di un generico neurone artificiale 60
Figura 65 - Esempio di One-hot encoding con la categoria "giorno della settimana" 63
Figura 66 - Esempio di Binary encoding con la categoria "giorno della settimana" 63
Figura 67 - Encoding basato sui profili tipici di consumo.....65
Figura 68 - Encoding relativo al tipo di funzionamento, estivo vs invernale.....66
Figura 69 - Rappresentazione delle variabili di input al modello ANN.....66
Figura 70 - Andamento della metrica loss al passare delle epoche, per i dati di training e testing.....67
Figura 71 - Profili reali e predetti del periodo di test 68
Figura 72 - Profili reali e predetti del modello ANN estivo nel periodo di testing.....69
Figura 73 - Profili reali e predetti del modello ANN invernale nel periodo di testing 70
Figura 74 - Profilo aggregato composto dalla stima dei consumi dei macchinari e del consumo delle UTA.....71
Figura 75 - Consumo frigorigeno reale (linea scura) e l'aggregato dei consumi (aree colorate), ANN.....72
Figura 76 - Schema logico delle variabili di input al modello di forecasting 24 h ANN.....74
Figura 77 - Profili reali e predetti dal modello di previsione 24h ANN nel periodo di testing 75
Figura 78 - Procedimento di benchmarking interno.....79
Figura 79 - Stima delle variabili autoregressive con l'approccio "recursive multi-step forecasting" 80
Figura 80 - Testing dell'applicazione del "recursive multi-step forecasting" 81
Figura 81 - Rappresentazione dei profili a confronto e del risparmio stimato 81
Figura 82 - Zoom-in sui profili di benchmark 82
Figura 83 - Andamento del risparmio economico nel mese di luglio 2022.....82
Figura 84 - Possibile procedimento di conversione dei consumi elettrici in domanda frigorigena.....83
Figura 85 - Creazione di un modello per la stima della domanda frigorigena 84

Indice delle Tabelle

Tabella 1 - Consumi annuali di gasolio e benzina del 2018 12
Tabella 2 - Suddivisione dei consumi elettrici 2018 13
Tabella 3 - Caratteristiche generali delle UTA di stabilimento 2018 14
Tabella 4 - Dati di targa gruppi frigo 17
Tabella 5 - Metriche prestazionali per predizione dell'energia frigo totale considerando i diversi driver esogeni 35
Tabella 6 - Metriche prestazionali regressione multipla 36
Tabella 7 - Metriche prestazionali per predizione dell'energia frigo UTA considerando i diversi driver esogeni 38
Tabella 8 - Metriche prestazionali del modello CART 45
Tabella 9 - Metriche prestazionali del modello annuale Random Forest 49
Tabella 10 - Metriche prestazionali modello predittivo consumi frigorigeni macchinari 53
Tabella 11 - Metriche prestazionali previsione consumi frigorigeni UTA 54
Tabella 12 - Metriche prestazionali complessive dei due modelli, Random Forest 56
Tabella 13 - Riassunto modelli Random Forest 57
Tabella 14 - Metriche prestazionali modello ANN 67
Tabella 15 - Metriche prestazionali modello ANN estivo 69
Tabella 16 - Metriche prestazionali modello ANN invernale 70
Tabella 17 - Metriche prestazionali modello ANN macchinari 71
Tabella 18 - Metriche prestazionali modello ANN UTA 72
Tabella 19 - Metriche prestazionali complessive dei due modelli, ANN 73
Tabella 20 - Metriche prestazionali modello di forecasting 24h ANN 75
Tabella 21 - Riassunto modelli ANN 76
Tabella 22 - Confronto delle metriche prestazionali tra i modelli implementati 77
Tabella 23 - Metriche prestazionali sul periodo di testing dell'applicazione del recursive multi-step forecasting 81
Tabella 24 - Risparmi stimati attraverso il procedimento di benchmarking interno 83

1. Il quadro energetico ambientale

La transizione energetica e la sostenibilità sono ormai temi fondamentali e attuali nel mondo d'oggi. Il riscaldamento globale e i suoi effetti stanno raggiungendo livelli sempre più critici: le emissioni di gas effetto serra non sono mai state così alte come nell'ultimo decennio [1].

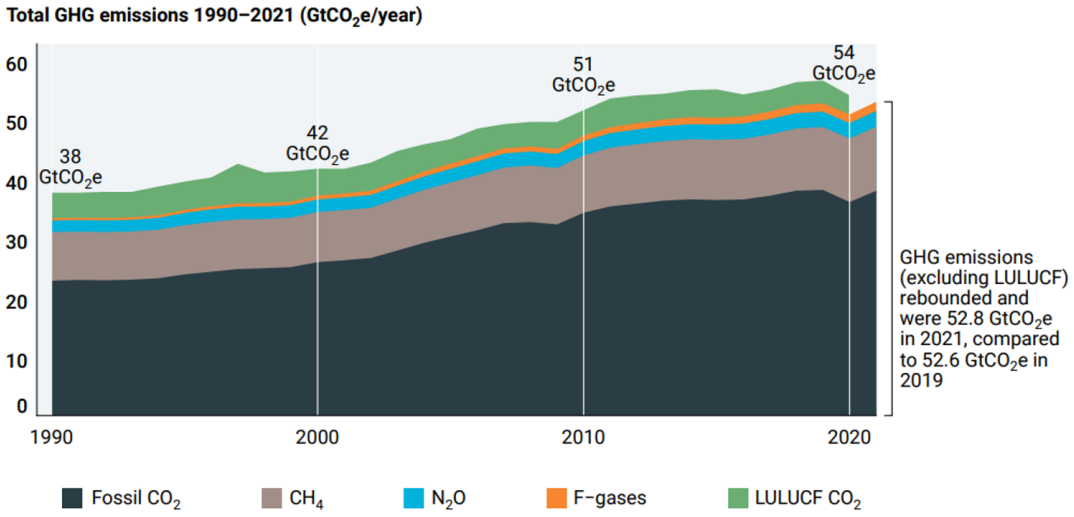

Figura 1 - Emissioni di GHG (Greenhouse Gases) totali, 1990-2021, EGR 2022

Il grafico messo a disposizione sul report annuale EGR dall'UNEP (*United Nations Environment Programme*), evidenzia come il consumo massivo di fonti fossili sia responsabile della maggior parte delle emissioni di GHG.

A livello industriale, si può osservare l'andamento delle emissioni e l'incremento di anno in anno assegnabile a tale settore [2]. L'incremento che si ha avuto tra 2020 e 2021 è il più alto mai registrato, come si evidenzia in Figura 2.

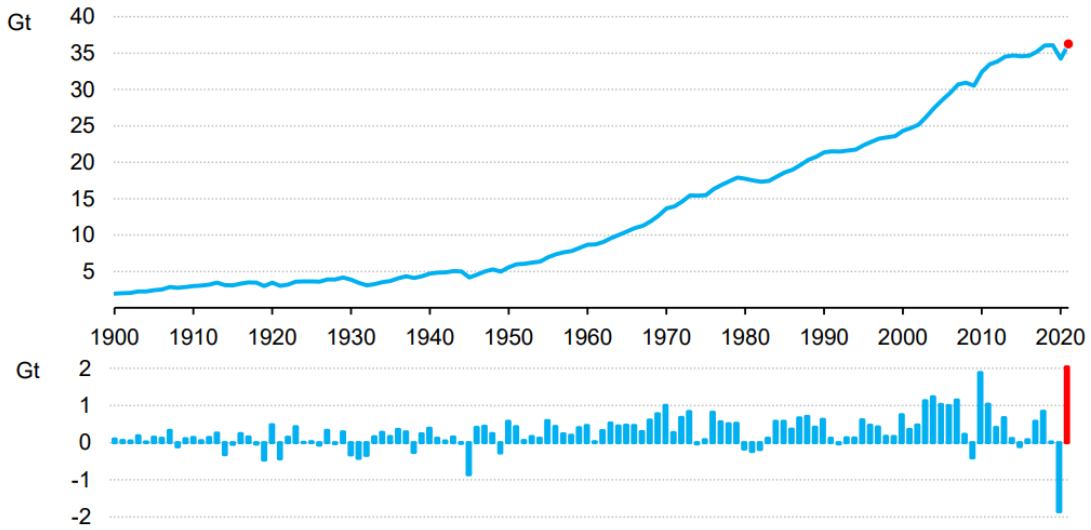

Figura 2 - Emissioni totali di CO_2 per energia da combustione e processi industriali e la loro variazione annuale [2]

Vi sono fenomeni fisico-climatici che possiedono una certa inerzia, e risultano ormai inarrestabili anche se si smettesse di produrre CO_2 (e gli altri gas effetto serra) all'istante. La responsabilità maggiore è proprio dell'uso ingente di combustibili fossili (carbone, petrolio, gas naturale, ecc.), che ha reso l'anidride carbonica il predominante tra i gas effetto serra sulla Terra.

Lo scioglimento dei ghiacciai, l'innalzamento del livello del mare, i forti cambiamenti climatici, le ondate di calore, la siccità, sono solo una parte delle conseguenze che in questo momento stanno danneggiando l'equilibrio nostro e del nostro pianeta. Fondamentale risulta quindi essere la decarbonizzazione dei vari settori, in modo da arginare i danni e garantire un futuro migliore.

Gli studi scientifici hanno permesso di aumentare la consapevolezza in questo tema, permettendo alle varie istituzioni di mettersi in moto per sopperire a una situazione sempre più critica.

1.1 L'accordo di Parigi e il Green Deal europeo

Di cruciale importanza per il nostro pianeta è stata l'emanazione dell'Accordo di Parigi (Parigi, COP21 [3]): in data 12 dicembre 2015 i 195 Stati membri dell'UNFCCC *(United Nations Framework Convention on Climate Change*, Convenzione Quadro delle Nazioni Unite sui cambiamenti climatici) hanno adottato il primo documento giuridicamente vincolante concernente il cambiamento climatico. Questo documento, determinato dalla comune intesa degli Stati membri, è un passo rivoluzionario verso la neutralità climatica e una realtà a impatto zero.

Figura 3 - Accordo di Parigi, COP21, 12 dicembre 2015

L'obiettivo comune posto è l'ottenimento di un aumento limitato della temperatura media globale, con soglia minima di 2°C oltre i livelli preindustriali, e se possibile un contenimento di questa crescita anche sotto a 1.5°C [4]. Insieme a questo, nell'articolo 4 del documento, si preme tanto sul concetto di mitigazione e sul monitoraggio di questo.

L'accordo di Parigi non specifica però i mezzi e le procedure per il raggiungimento di tali obiettivi. Per quanto riguarda i paesi appartenenti all'Unione Europea, si è giunti a un insieme di direttive che vertono proprio al recepimento di tali linee guida.

I paesi facenti parte dell'Unione Europea sostengono risolutamente l'accordo di Parigi, in modo da contenere in modo sicuro gli effetti del riscaldamento globale. Tutti i suoi membri infatti sono determinati nell'attuazione e il raggiungimento degli obiettivi di questo.

Per il raggiungimento di questi traguardi, è fondamentale la creazione di un piano accurato e comune, che permetta non solo dei progressi ma anche una misurazione e monitoraggio di questi, e che vi sia una comunicazione e condivisione degli obiettivi. A questo scopo, l'Unione Europea, l'11 dicembre del 2019 ha adottato il Green Deal europeo [5].

Figura 4 - Green Deal europeo, 11 dicembre 2019 [6]

Gli Stati membri dell'Unione Europea si sono impegnati nel raggiungimento della neutralità climatica entro il 2050, rispettando le responsabilità fatte proprie con l'accordo di Parigi [7].

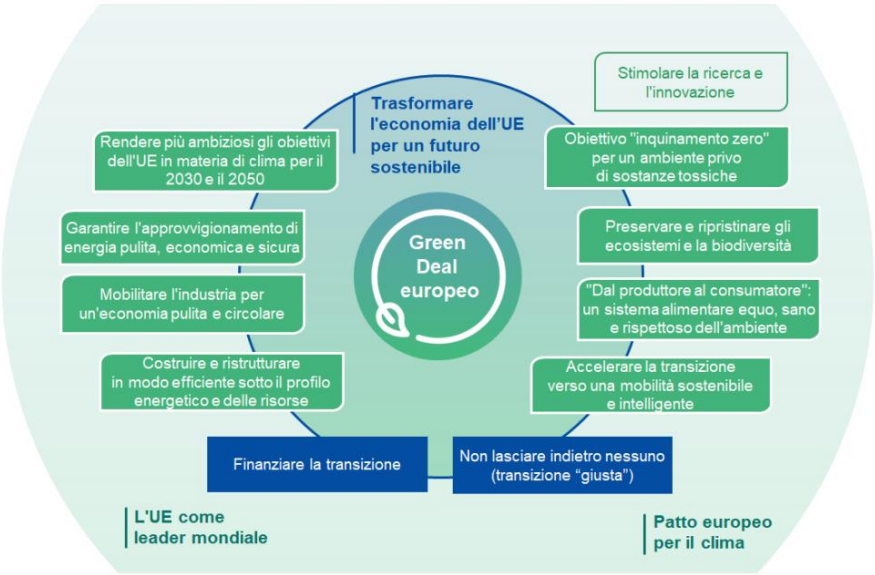

Figura 5 - Illustrazione del Green Deal europeo, dalla comunicazione della Commissione Europea COM(2019) 640

Parte delle strategie della commissione è la seguente: far sì che il settore industriale UE sia alla guida della transizione verde e digitale, aumentando la sua competitività su scala mondiale.

Per il raggiungimento in un'economia più sostenibile, digitale, resiliente, serve che start-up, PMI (piccole e medie imprese) e grandi imprese, non solo stiano al passo col cambiamento,

ma che ne siano esse stesse il motore. Un ruolo cruciale lo hanno quindi le innovazioni di carattere verde e digitale.

Concetti fondamentali [8] introdotti nell'Accordo di Parigi e nelle traiettorie del Green Deal sono:

- Adattamento
- Mitigazione

Per adattamento si intende l'insieme di tutte quelle azioni che vertono alla previsione e alla minimizzazione dell'impatto degli effetti e dei danni relativi al cambiamento climatico. I cambiamenti climatici possono però essere visti anche come opportunità da cui trarre possibili vantaggi. Le misure di adattamento possono avere sia carattere infrastrutturale, come la costruzione di protezioni per l'innalzamento dei mari, che di carattere comportamentale, attraverso l'educazione e una rivoluzione delle abitudini che permettano un accoglimento dei cambiamenti in atto. L'adattamento è fondamentale, perché vi sono fenomeni climatici che possiedono ormai una certa inerzia, e non sono quindi arrestabili attraverso alcuna azione umana. L'unica soluzione è trovare un modo per conviverci, attraverso un processo di adeguamento agli effetti attuali e futuri del cambiamento climatico.

Per mitigazione invece si intendono tutte quelle azioni il cui obiettivo è la riduzione dell'impatto dei cambiamenti climatici dovuti al cambiamento climatico e quindi alle emissioni di gas effetto serra. Esempi di mitigazione sono l'incremento di capacità per quanto riguarda le fonti di energia rinnovabile, l'implementazione di sistemi di trasporto ad emissioni ridotte (se non nulle), la creazione di sistemi di stoccaggio (naturale, come foreste, o artificiale, con sistemi appositi), l'efficientamento di sistemi energetici e la valorizzazione dell'energia a disposizione. In poche parole, il concetto di mitigazione racchiude tutti gli interventi che permettono una riduzione delle emissioni di gas serra, rafforzando parallelamente i pozzi di assorbimento di queste.

Un punto fondamentale per la transizione energetica, necessaria per il raggiungimento della neutralità climatica, è quindi anche l'implementazione di nuove strategie di gestione dell'energia. L'opportunità di un mondo sempre più interconnesso e della possibilità di scambio di quantità sempre più grandi e varie di dati permette l'implementazione di algoritmi che permettano una descrizione e un'ottimizzazione dei sistemi energetici, sia per il lato produzione che per il lato consumo.

Figura 6 - Industria 4.0 [8]

Nel presente elaborato, un punto cardine è proprio la creazione di modelli che simulino in modo accurato l'andamento dei consumi di un'utenza industriale, con una seguente esplorazione delle possibilità applicative derivanti da questi.

Queste implementazioni evidenziano l'importanza della digitalizzazione e della gestione intelligente delle proprie risorse: non è importante solo implementare sistemi sostenibili e puliti, ma anche gestire in modo efficace le risorse e gli strumenti a disposizione, in modo da minimizzare il più possibile gli sprechi di diversa natura.

1.2 L'impatto della refrigerazione e del condizionamento

I consumi relativi al condizionamento e raffrescamento dei locali è più che triplicato negli ultimi vent'anni, e anche la sua quota rispetto ai consumi totali di un edificio è in continuo aumento [9].

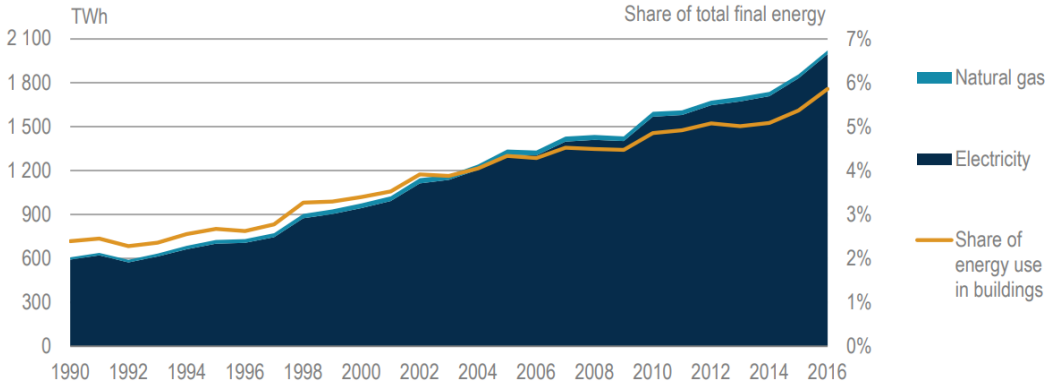

Figura 7 - Consumo mondiale di energia per il raffrescamento di locali interni negli edifici [9]

Analizzando le proiezioni al 2030, messe a disposizione in un report dell'IEA [11], si capisce che questo trend, pur considerando il Net Zero Scenario, non è previsto che si stabilizzi.

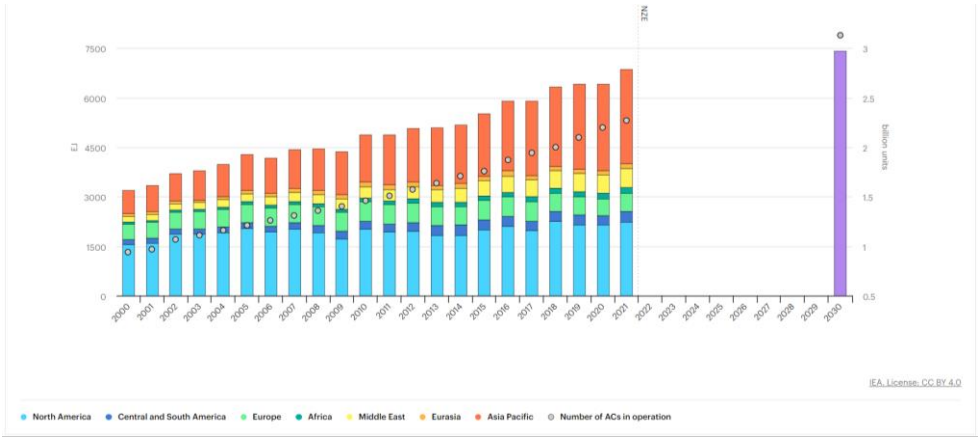

Figura 8 - Consumo energetico finale per il raffrescamento di locali per regione e numero di unità operanti nel Net Zero Scenario, 2000-2030

È quindi evidente che i consumi associati ad una domanda di refrigerazione, non solo impattino in modo consistente, ma sono anche in aumento.

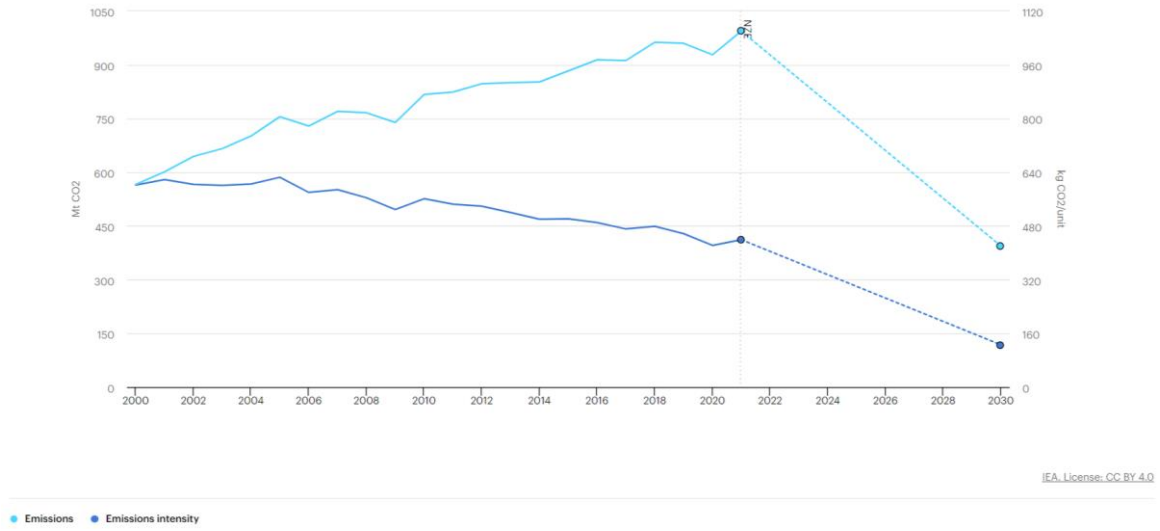

Figura 9 - Emissioni di CO_2 e intensità di emissione per il condizionamento dell'aria nel Net Zero Scenario, 2000-2030 [11]

Per raggiungere gli obiettivi del Green Deal e dell'accordo di Parigi, è quindi necessario anche un incremento delle performance di gruppi frigo e di sistemi di condizionamento dell'aria, per l'ottenimento di un rateo di emissioni contenuto (Figura 9) [11].

L'efficienza dei sistemi di condizionamento sta comunque migliorando costantemente negli ultimi anni, ma considerando uno scenario senza un passaggio alle BAT (Best Available Techniques), la domanda di elettricità per il raffrescamento degli ambienti negli edifici potrebbe aumentare del 40% a livello globale nel 2030. Per raggiungere lo scenario Net Zero, le efficienze medie dei nuovi sistemi dovrebbero aumentare fino al 50% entro il 2030 [11].

Lo studio eseguito nel presente elaborato, verte su un'analisi approfondita dell'andamento dei consumi di un gruppo frigo industriale, per prevederne e simularne le dinamiche. Questo tipo di analisi permette non solo di capire e conoscere il sistema in esame, ma anche valutare le possibili opportunità applicative nel campo del monitoraggio dei consumi e dell'ottimizzazione di questi.

In linea con la filosofia della mitigazione promulgata nei documenti precedentemente citati, l'obiettivo è estrarre la conoscenza necessaria per poi permettere lo sviluppo di algoritmi di ottimizzazione e poter valutare l'impatto che i miglioramenti hanno sui sistemi.

2. Descrizione del sito industriale

Il sito in esame di cui sono stati analizzati ed elaborati i dati, è un corpo di fabbrica, ospitante diverse fasi del processo produttivo e degli uffici. All'interno dell'edificio principale sono presenti locali adibiti a centrale termica, sala compressori e condizionamento di locali e macchinari. Le informazioni relative al caso studio sono state estratte dal documento della diagnosi aziendale del sito. In questo documento sono descritti i processi del ciclo produttivo, i consumi dei vari vettori energetici, il layout aziendale e tante altre informazioni. Molte informazioni, tabelle e grafici sono state estratte dall'ultima diagnosi energetica effettuata sull'azienda in esame.

2.1 Il processo produttivo

All'interno dell'edificio principale è possibile distinguere diverse zone e aree funzionali, sulla base delle diverse lavorazioni che vengono svolte in esse. I reparti distinguibili sono rispettivamente:

- Il reparto di ispezione
- Il reparto di ispezione
- Il reparto saldatura
- Il reparto coating
- Il reparto sabbiatura
- Il reparto elettroerosione
- Il reparto foratura laser
- Il reparto trattamento termico
- Il reparto collaudo
- I servizi ausiliari (sala compressori, torri evaporative e filtri depolveratori)
- I servizi generali (illuminazione, gruppi frigoriferi, uffici, servizi generali, UTA)

Tutti questi reparti hanno la funzione di apportare lavorazioni di precisione su palette di turbine a gas, secondo disegni specifici, volte alla realizzazione del componente finito o alla riparazione di componenti usurati.

Perciò, i processi principali in cui è suddividibile l'attività dello stabilimento sono due:

- Lavorazione di "Nuovo"
- Lavorazione di "Riparazione"

Nella lavorazione di pezzi nuovi, l'azienda riceve palette grezze prodotte da altre aziende, su cui si svolgono principalmente operazioni di foratura (tramite elettroerosione o laser) e di rivestimento, coinvolgendo prevalentemente i seguenti reparti:

- Reparto di collaudo, per verificare in più momenti del ciclo produttivo la presenza o meno di anomalie
- Reparto di sabbiatura, in cui vene effettuata una pulizia preliminare dei pezzi prima di mandarli al reparto di rivestimento
- Reparto di rivestimento, dove i pezzi ricevono due tipi di rivestimento, metallico o ceramico, mediante plasma spray o mediante HVOF
- Reparto trattamento termico, in cui si eliminano le cricche attraverso un processo di ricristallizzazione dei grani, per poi effettuare i processi di diffusione e invecchiamento
- Reparto foratura, dove avvengono per l'appunto realizzate forature mediante elettroerosione o mediante il taglio laser

Per quanto riguarda invece le operazioni di riparazione e services non esiste un ciclo di lavoro standard, in quante le operazioni svolte sono funzione del pezzo in esame e dell'esperienza e sensibilità dell'operatore. Nonostante non vi sia una procedura standardizzata, il ciclo prevede sempre inizialmente una verifica ispettiva all'ingresso in stabilimento, dopodiché è mandato ad un'altra azienda in cui si svolge lo stripping (eliminazione dello strato ceramico iniziale). Rientrato il pezzo in azienda è poi possibile attuare ispezioni con liquidi fluorescenti per mettere in evidenza eventuali cricche presenti sul materiale. Successivamente, in seguito all'esito dell'ispezione, sono svolte lavorazioni di saldatura e molatura manuale, trattamenti termici, sabbiatura, foratura e rivestimento.

Si possono quindi riassumere i due processi principali nello schema di seguito.

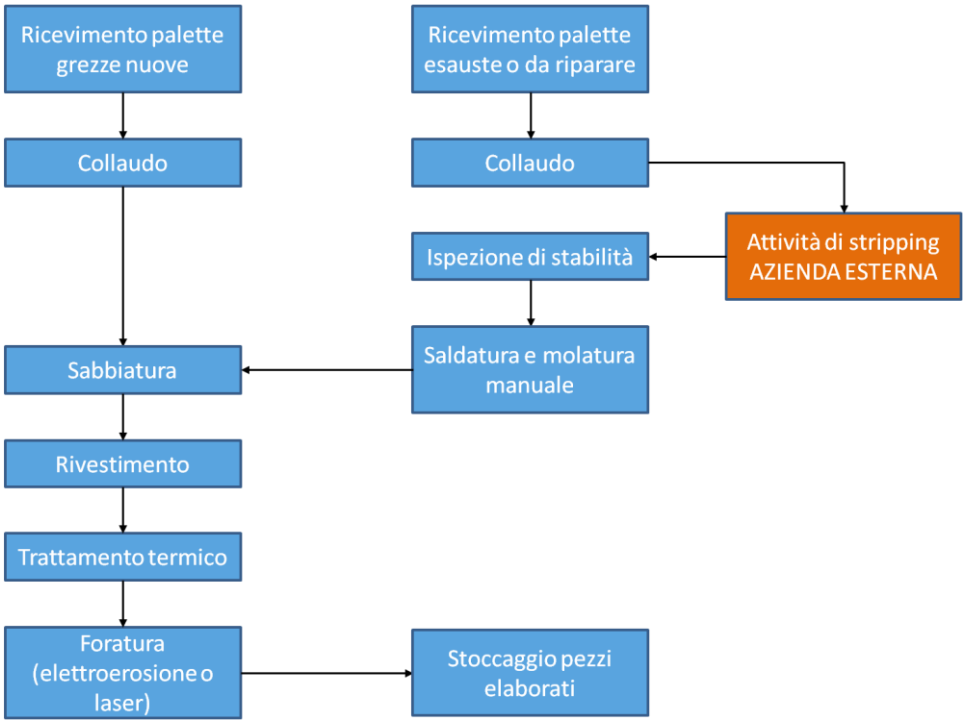

Figura 10 - Schema del processo produttivo

2.2 I consumi di stabilimento

2.2.1 I consumi elettrici

L'intero stabilimento è alimentato da una stazione di trasformazione primaria, che riceve dalla rete energia in media tensione per immetterla e distribuirla in bassa tensione.

Dai dati estraibili dalle bollette, è possibile avere un'idea degli assorbimenti elettrici dell'azienda nel periodo tra gennaio 2016 e giugno 2019, riportati graficamente evidenziando le quote di energia consumata nelle rispettive fasce mensilmente.

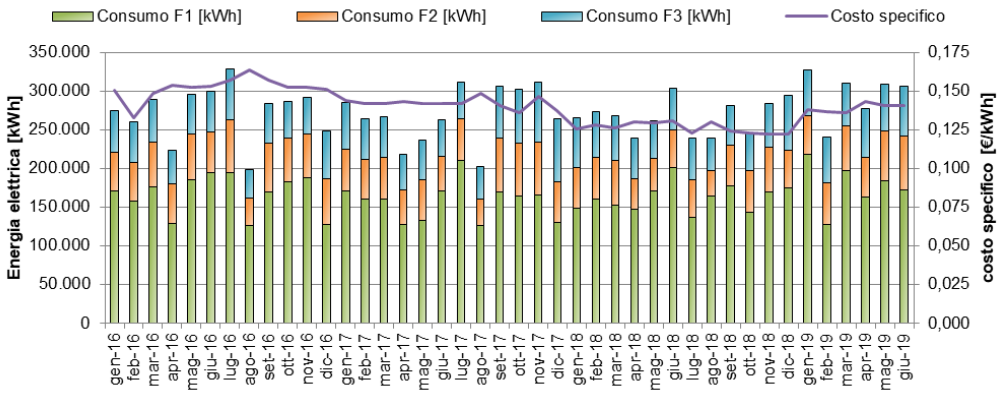

Figura 11 - Andamento mensile dei consumi elettrici nel periodo gennaio 2016 - giugno 2019

Dalla Figura 11 si osserva un consumo mensile prevalentemente compreso tra i 200.000 e i 350.000 kWh. Insieme ai consumi suddivisi nelle tre fasce, è inoltre riportato l'andamento del consumo specifico in €/kWh dell'energia elettrica.

Analizzando in modo specifico la curva di assorbimento orario dell'anno 2018 (in Figura 12), si denota la periodicità settimanale dell'andamento dei consumi, che mediamente avviene in un range di potenza non superiore ai 1500 kW. Salta all'occhio, inoltre, la riduzione di attività in corrispondenza delle ferie estive e natalizie.

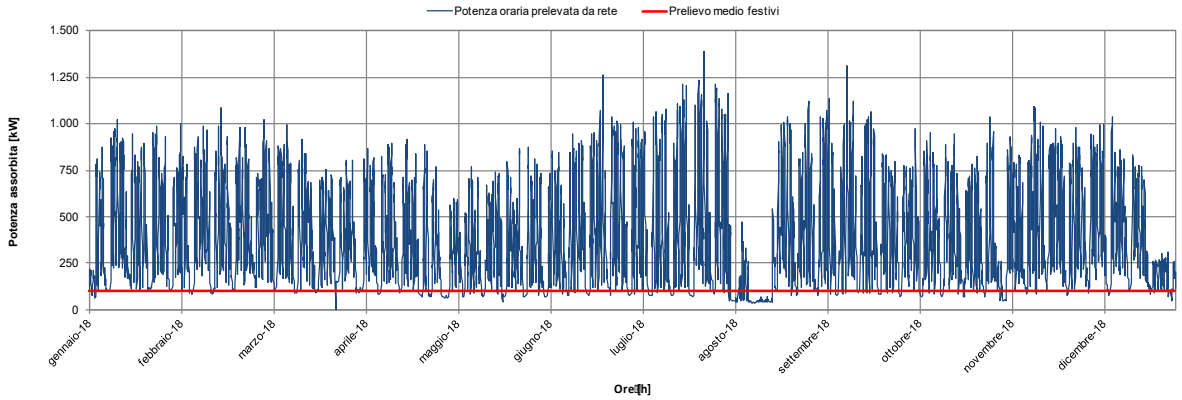

Figura 12 - Curva di assorbimento di potenza elettrica nell'anno 2018

Da questi dati è possibile effettuare una cumulata dei prelievi orari, e analizzare il livello di potenza in funzione delle ore di funzionamento (Figura 13).

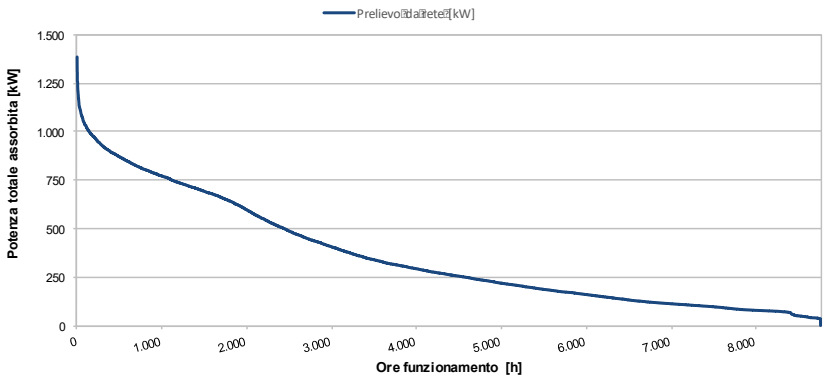

Figura 13 - Curva di durata della potenza elettrica dell'anno 2018

Per una rappresentazione più fruibile si può visualizzare il *carpet plot* (o *heatmap*) dei prelievi, in cui l'intensità cromatica permette di distinguere i periodi di maggior assorbimento (colori caldi) da quelli più bassi (colori freddi). L'alternarsi dei giorni di colorazione completamente fredda permette di identificare i giorni di attività (lavorativi) da quelli di non attività (week-end). Inoltre, si nota che il turno lavorativo avviene tra mattina e pomeriggio, con i consumi maggiori nella parte centrale della giornata. Stagionalmente invece, se non si considera la pausa di agosto, il periodo estivo risulta quello livelli di potenza più alti, probabilmente dovuto ai carichi frigorigeni di condizionamento.

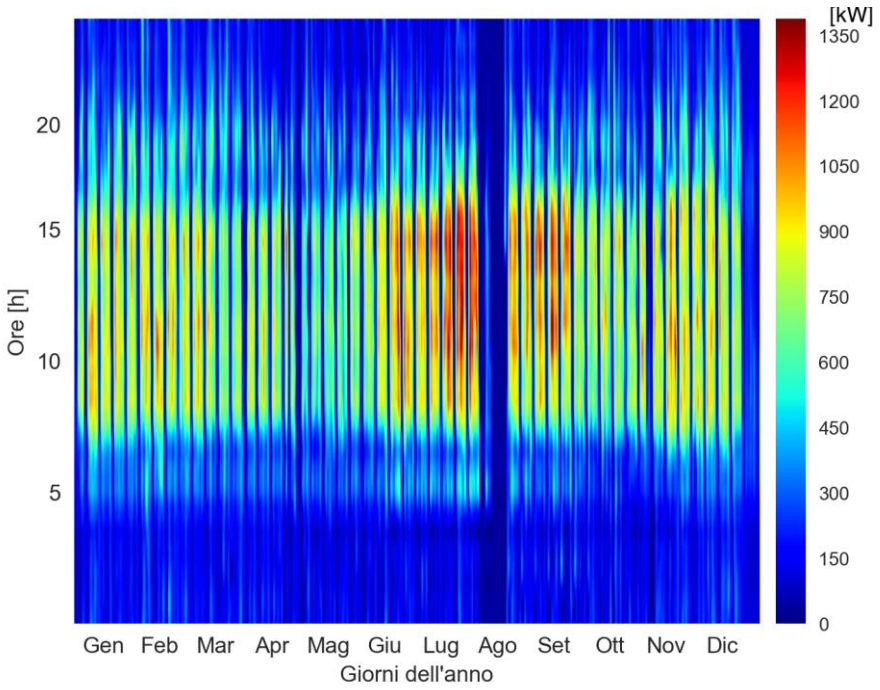

Figura 14 - Heatmap prelievi elettrici del 2018

2.2.2 I consumi di combustibile

I combustibili consumati all'interno dello stabilimento sono i seguenti:

- Gas naturale
- Kerosene
- Benzina
- Gasolio

I consumi più impattanti a livello aziendale risultano quello del gas naturale e del kerosene, di cui si rappresentano i consumi mensili e costo specifico nella Figura 15.

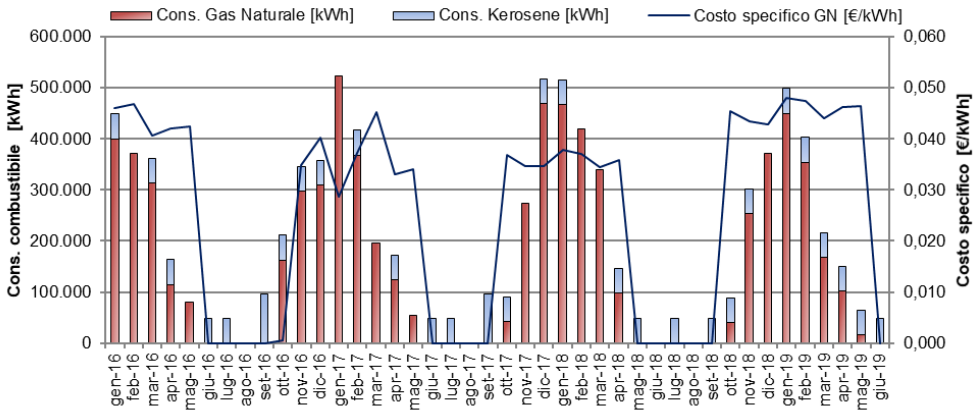

Figura 15 - Andamento mensile dei consumi di gas naturale e kerosene nel periodo Gennaio 2016 - Giugno 2019

Dall'andamento dei consumi nel periodo in esame, si può apprezzare una periodicità di questi. Data la stagionalità del trend, si può ipotizzare che questi combustibili siano prevalentemente sfruttati ai fini di riscaldamento, come si può vedere anche dalla curva di prelievo orario dell'anno 2018 (Figura 16).

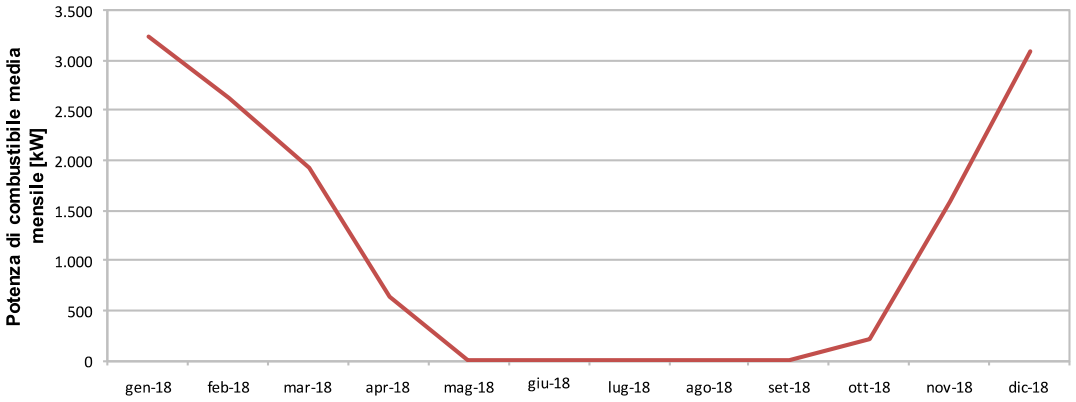

Figura 16 - Curva di prelievo orario della potenza di combustibile dell'anno 2018

Dai bilanci di combustibile è infatti confermato che il consumo di gas naturale sia principalmente sfruttato per il riscaldamento (servizi generali), mentre il kerosene è usato principalmente per il coating (attività principali).

Per quanto riguarda invece benzina e gasolio, il loro consumo è imputabile all'uso dei mezzi di trasporto di proprietà dell'azienda. Nella Tabella 1 sono riportati i consumi dell'anno 2018 sia

in litri che in kWh (considerando rispettivamente un PCI pari a 10 kWh/l per il gasolio e 9.05 kWh/l per la benzina).

Tabella 1 - Consumi annuali di gasolio e benzina del 2018

	Gasolio		Benzina	
	[lt]	[kWh]	[lt]	[kWh]
TOT - 18	2862	28620	106	960

2.2.3 I consumi globali

Dai dati di bolletta, è possibile riportare gli share di consumo in funzione dei MWh e di tep (tonnellate equivalenti di petrolio, unità di misura dell'energia primaria). Il consumo complessivo di stabilimento risulta essere di 789 tep, in cui i consumi di gasolio e benzina (rispettivamente 0.08 e 2 tep) risultano trascurabili.

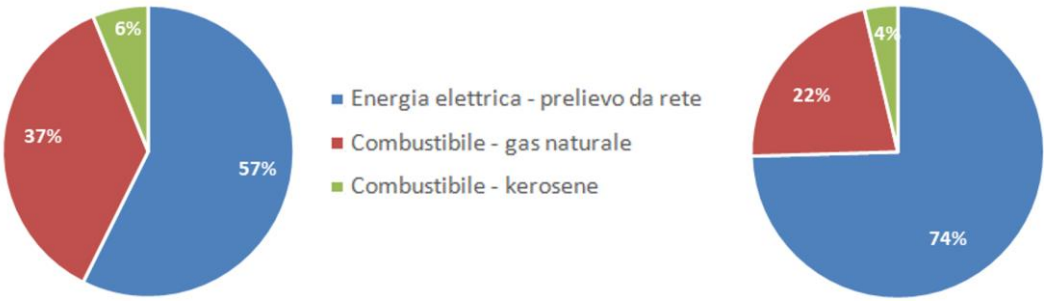

Figura 17 - Ripartizione dei consumi globali del 2018 da bolletta (dx) e in energia primaria (sx)

È evidenziata fin da subito la prevalenza dei consumi elettrici sia in termini di energia primaria (74%) che in termini energetici complessivi (57%).

Da un'analisi più accurata dei bilanci elettrici, è possibile ripartire i rispettivi consumi in funzione delle utenze ed in funzione delle aree funzionali.

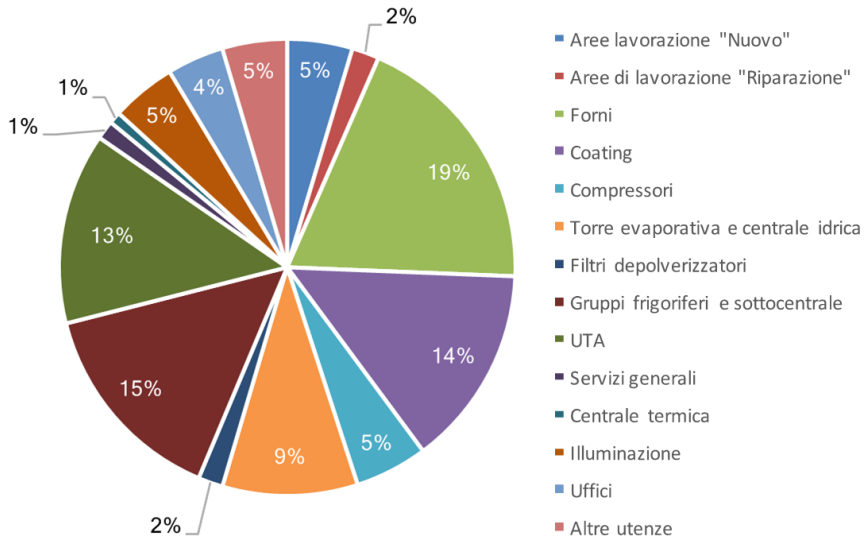

Figura 18 - Suddivisione dei consumi elettrici del 2018

Sulla base di questi dati si possono definire le utenze di maggior interesse (forni 19%, gruppi frigoriferi e sottocentrale 15%, UTA di stabilimento 14%) e le aree funzionali che più consumano (attività principali e servizi generali).

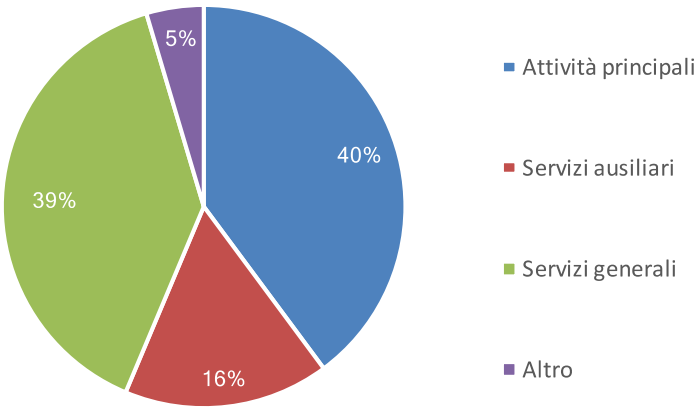

Figura 19 - Suddivisione dei consumi elettrici del 2018 per area funzionale

Si riportano i consumi del 2018 suddivisi per reparto e area funzionale per una valutazione quantitativa e qualitativa dell'impatto di ognuno di questi sul totale.

Tabella 2 - Suddivisione dei consumi elettrici 2018

Consumi 2018				
Reparto	**Totale**		**Area funzionale**	**Area funzionale**
	[kWh]	**[%]**	**[%]**	
Aree lavorazione "Nuovo"	145.054	5.0%	40%	Attività principali
Aree di lavorazione "Riparazione"	60.870	2.0%		
Forni	597.860	19.0%		
Coating	446.642	14.0%		
Compressori	161.340	5.0%	16%	Servizi ausiliari
Torre evaporativa e centrale idrica	300.715	10.0%		
Filtri depolverizzatori	54.963	2.0%		
Gruppi frigoriferi e pompaggi	461.676	15.0%	39%	Servizi generali
UTA	425.576	14.0%		
Servizi generali	41.689	1.0%		
Centrale termica	26.283	1.0%		
Illuminazione	144.137	5.0%		
Uffici	126.565	4.0%		
Altro	143.989	4.6%		
TOTALE	**3.137.359**	**100%**		

2.2.4 I consumi delle UTA di stabilimento e dei gruppi frigo

Nella seguente sottosezione si concentra il focus sui consumi elettrici delle UTA e gruppi frigoriferi, che insieme rappresentano rispettivamente il 14% e il 15% (Tabella 2) dei consumi elettrici dell'impianto nell'anno 2018 (quasi un terzo dei consumi totali). Innanzitutto, si riporta una scheda tecnica riassuntiva delle UTA presenti nell'impianto.

Tabella 3 - Caratteristiche generali delle UTA di stabilimento 2018

UTA	Marca	Portata di mandata [m^3/h]	Portata di ripresa [m^3/h]	Potenza nominale [kW]	Ore di funzionamento [h/anno]	Consumo utenza [kWh]	Misuratore	Consumo totale [kWh]
UTA 1	Wolf Italia srl	68.000	68.000	45		54.802	Misurato Mis8 (M31)	
UTA 2	Wolf Italia srl	41.500	41.500	30		71.355	Misurato Mis8 (M46)	
UTA 3	Wolf Italia srl	90.000	90.000	75		118.287	Misurato Mis8 (M47)	
UTA 4	Wolf Italia srl	54.000	54.000	37	**3731**	40.519	Misurato Mis8 (M32)	**425576**
UTA 5	Wolf Italia srl	90.000	90.000	75		54.787	Misurato Mis8 (M48)	
UTA 6	Wolf Italia srl	15.000	15.000	11		35.811	Misurato Mis8 (M33)	
UTA 7	Wolf Italia srl	14.500	14.500	11		50.015	Misurato Mis8 (M14)	

Le sette UTA, hanno la funzione di condizionamento dell'aria sia dei locali industriali che degli uffici.

L'azienda monitora ogni UTA con dei multimetri, in modo da registrare i consumi elettrici di queste singolarmente. Questi consumi sono principalmente imputabili al funzionamento del ventilatore (e quindi del motore elettrico), il quale deve dare la prevalenza necessaria a vincere le perdite delle varie sezioni dell'unità di trattamento aria e servire i locali da condizionare. Inoltre, per verificare l'azione di condizionamento apportata da queste unità di trattamento aria, sono stati posizionati dei sensori di temperatura sul ritorno in ogni reparto/locale condizionato. La temperatura dell'aria di ripresa di un locale, nell'ipotesi di un modesto miscelamento, può essere considerata coincidente con la temperatura del locale stesso. In questo modo si possono monitorare gli andamenti delle temperature nei locali condizionati.

Altri misuratori di interesse sono quelli che monitorano i gruppi frigoriferi: vi sono tre unità frigorifere, e per ognuna di queste è presente un multimetro che registra l'assorbimento elettrico corrispettivo. Questi gruppi servono sia le UTA (per il processo di raffreddamento e/o deumidificazione dell'aria) che i macchinari (per refrigerazione di processo, raffreddamento di componenti, ecc.).

Per un'analisi qualitativa dell'andamento degli assorbimenti delle UTA, si riporta la Figura 20, in cui sono rappresentati i valori di potenza media oraria dei ventilatori delle unità, insieme al profilo aggregato dei consumi complessivi di questi. Si considerano unicamente le prime due settimane di settembre 2021, dato che non si hanno vere e proprie variazioni dei profili settimanali.

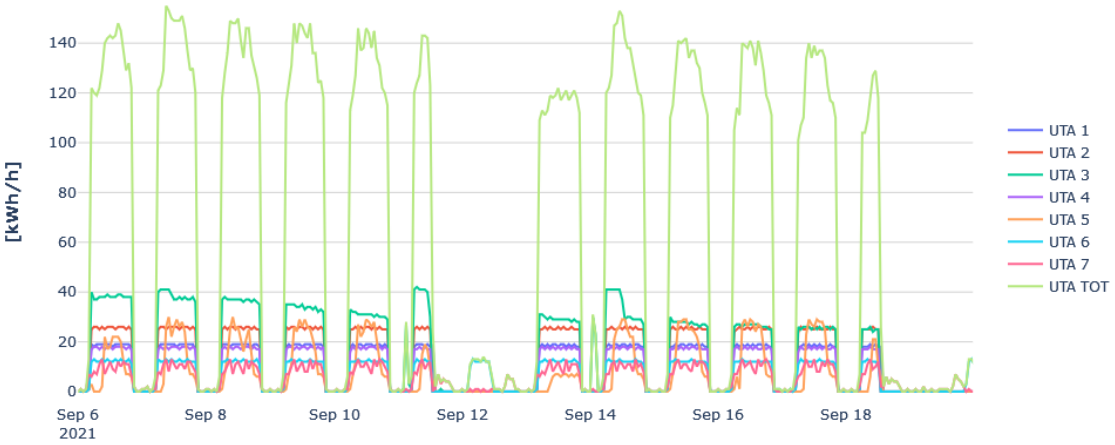

Figura 20 - Profili di carico delle UTA nelle prime due settimane di settembre 2021

Analizzando i profili di carico, e verificando poi gli orari, si può affermare che la settimana lavorativa (nell'anno 2021) è di sei giorni:

- Dal lunedì al venerdì l'orario di funzionamento è tra le 5:00 (talvolta con una possibile preaccensione o comunque piccolo assorbimento tra le 4:00 e le 5:00) e le 20:00. Questo per tutte le UTA eccetto quella che serve il locale 7 (UTA7) che anche nel periodo tra le 21:00 e le 4:00 presenta sempre un livello basso o nullo di consumo. In generale si può affermare che il turno lavorativo tipo sia 5:00 - 20:00.
- Il sabato invece possiamo parlare di un mezzo turno lavorativo, in quanto gli orari di funzionamento sono tra le 5:00 e le 12:00 (per la UTA1 fino alle 13:00 e per le UTA7 sembra sempre si mantenga un livello di potenza minimo tra le 12:00 e le 5:00 del lunedì successivo).

Si possono vedere comunque situazioni in cui i consumi di certe UTA non seguono un andamento in linea con i turni lavorativi: questi scostamenti potrebbero essere dovuti alle lavorazioni per l'attività di riparazione e services, le quali non essendo standardizzate (dipende molto dai pezzi da riparare e dagli operatori), non è detto seguano i trend di consumo attesi.

Da un'analisi dei profili di temperatura relativi ai locali climatizzati e le condizioni esterne, si può intuire la logica di gestione degli ambienti interni.

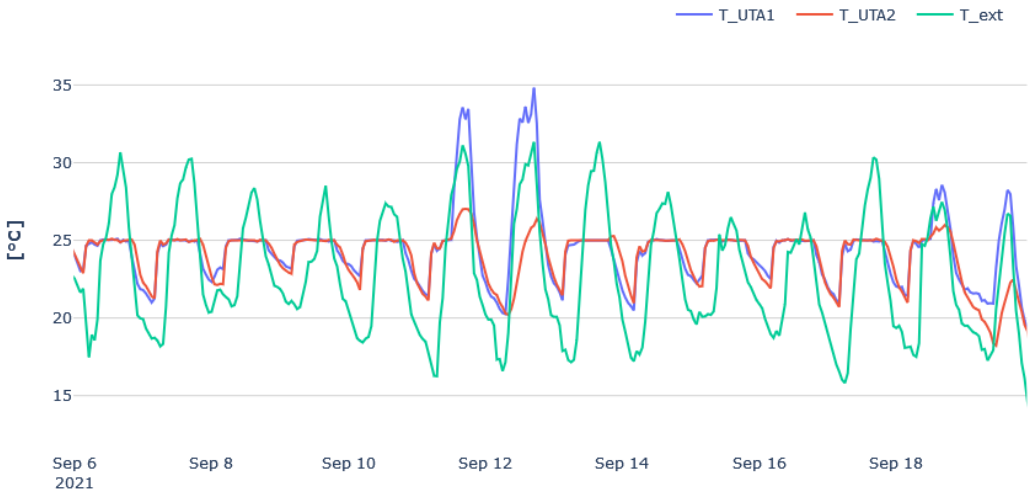

Figura 21 - Profili di temperatura delle UTA 1 e 2 e della temperatura esterna nelle prime due settimane di settembre 2021

Si sono riportati unicamente i profili di temperatura del locale 1 e 2 perché ritenuti rappresentativi anche delle altre unità. Dall'andamento dei profili di temperatura e potenza oraria, si può ipotizzare che sia impostata una temperatura di set-point (per i locali in esame risulta essere di circa 25°C). Nel fine settimana invece, si nota come la temperatura locale tenda a seguire l'andamento della temperatura esterna, dato che non è apportato alcun tipo di controllo. Riportando in uno stesso grafico gli andamenti qualitativi di un'UTA e la temperatura del locale che serve, si sottolinea questo tipo di logica di controllo, come si può vedere in Figura 22.

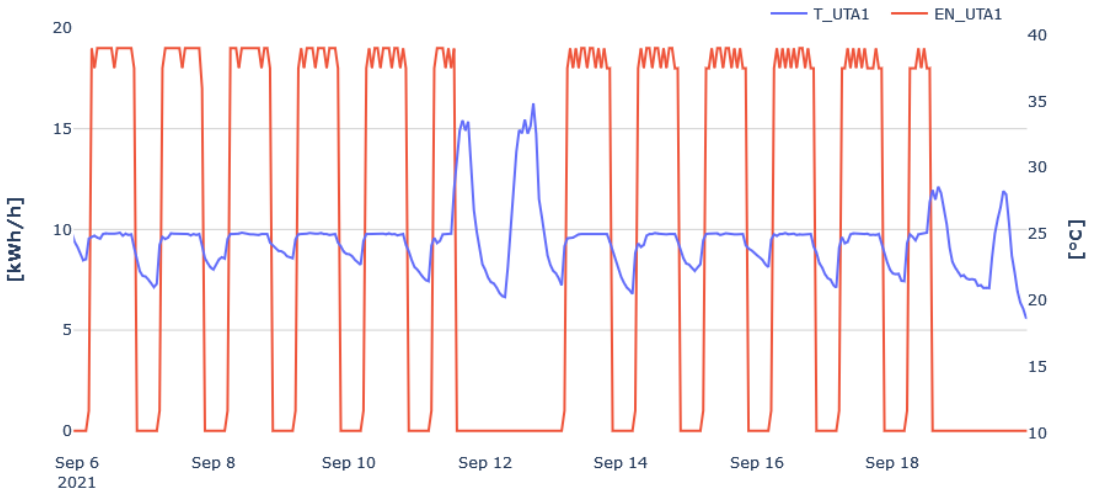

Figura 22 - Profilo di carico e di temperatura della UTA1 e locale 1, prime due settimane di settembre 2021

Temperatura interna e carico elettrico sono quasi in "fase" in orario lavorativo, e risulta anche facile distinguere le giornate dei turni lavorativi completi (lunedì-venerdì), da quelle con mezzi turni (sabato) dai festivi (domenica).

Se invece si analizzano i consumi dei gruppi frigo (Figura 23), si nota un andamento quasi completamente in linea con i consumi delle UTA, eccetto la domenica a mezzogiorno, in cui si verifica periodicamente un picco dei consumi da parte di tutti e tre, probabilmente dovuto a qualche operazione di routine.

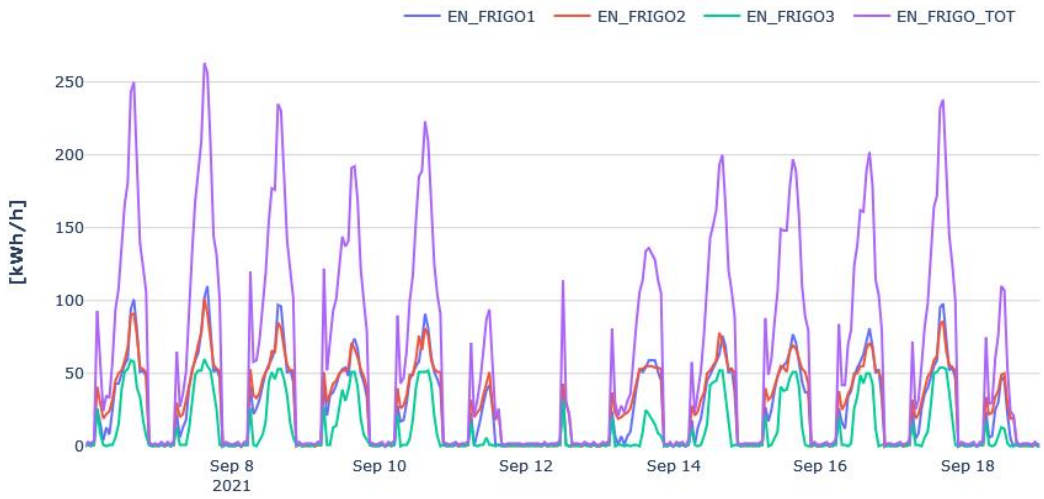

Figura 23 - Profili di carico dei gruppi frigo, prime due settimane di settembre 2021

I dati di targa dei gruppi frigo sono invece riassumibili nella tabella seguente.

Tabella 4 - Dati di targa gruppi frigo

Unità Frigorifera	Marca	Modello	Potenza elettrica nominale [kW]	Potenza frigo [kW]	Ore di funzionamento [h/anno]	Consumo utenza [kWh]	Misuratore	Consumo totale [kWh]
Gruppo frigo 1	YORK	YCIV 930 HE LS	282	773	338	95247	Misurato M9 (M55)	
Gruppo frigo 2	YORK	YCIV 930 HE LS	282	773	374	105420	Misurato M9 (M56)	**287530**
Gruppo frigo 3	YORK	YCIV 930 HE LS	282	773	308	86863	Misurato M9 (M57)	

Si presenta di seguito uno schema logico rappresentante i vettori energetici che interessano la parte frigorigena aziendale.

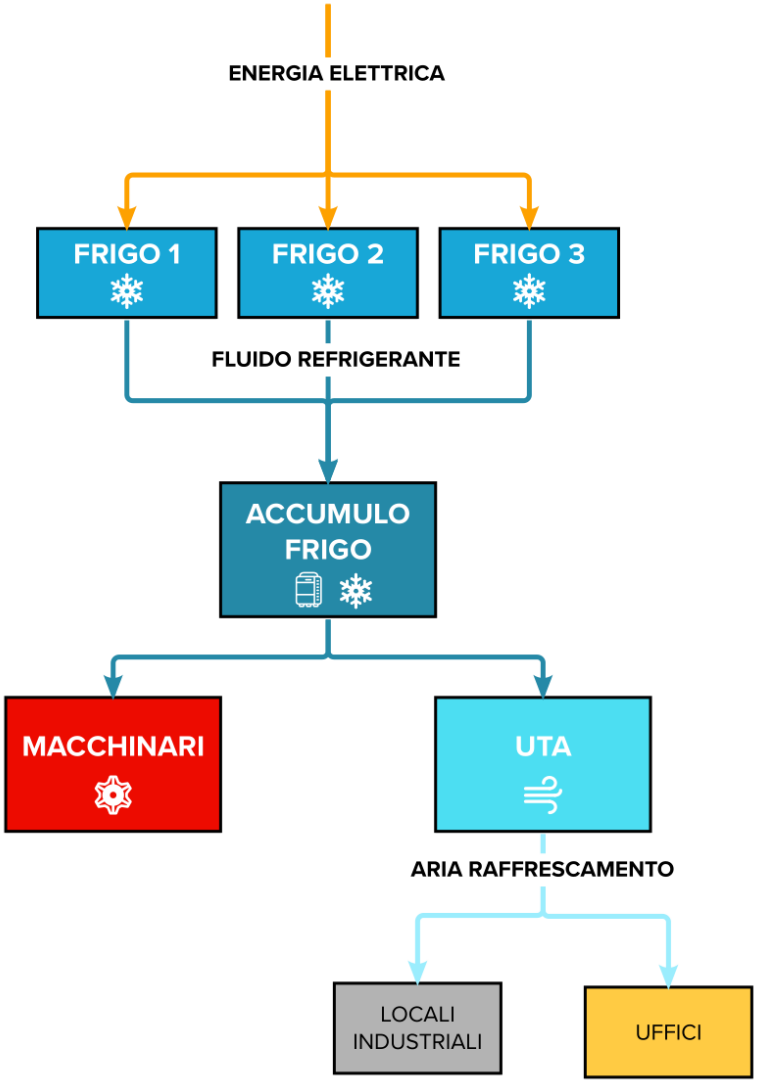

Figura 24 - Schema logico della domanda aziendale di liquido refrigerante

3. Acquisizione dei dati e pre-processing

3.1 L'architettura di monitoraggio

Il mondo in cui viviamo sta diventando sempre più data-centric: attraverso l'acquisizione, lo scambio e l'elaborazione dei dati, è possibile ricavare informazioni interessanti e utili per la gestione e la comprensione di fenomeni a livello sociale, economico e scientifico.

Questo spostamento di attenzione sul mondo dei big-data porta, anche a livello energetico, grandi vantaggi per la gestione e la manutenzione dei sistemi che si possono ritrovare su scala residenziale, terziaria e industriale.

L'acquisizione di questa grossa mole di dati avviene tramite i sistemi di monitoraggio, molteplici per natura e funzionalità richieste, ma assimilabili da un punto di vista architetturale. Questi, infatti, sono formati da tre livelli fondamentali:

- **Livello di campo**: costituito dai dispositivi di misura dotati di interfaccia di comunicazione installati in sito, che interagiscono direttamente con quest'ultimo e lo mettono in relazione con il livello successivo. Il compito principale di questo livello è quindi quello di acquisire le grandezze fisiche e trasformarle in segnali riconoscibili e interpretabili dal livello di comunicazione.
- **Livello di collegamento**: permette la comunicazione tra il livello di campo ed il livello di gestione. Esso può essere composto da reti wireless (WiFi, GPRS) oppure sistemi cablati, in cui i segnali sono trasmessi via cavo (cavi Modbus, cavi Ethernet), in base alle distanze e alle esigenze.
- **Livello di Analisi e Gestione**: In quest'ultimo livello, il dato proveniente dal campo e letto attraverso gli appositi driver di comunicazione, viene poi analizzato a seconda delle specifiche esigenze. Questo viene fatto grazie all'utilizzo di algoritmi e software che permettono l'implementazione di strategie di gestione volte all'efficientamento.

Come detto in precedenza i software che compongono il livello di Analisi e Gestione possono essere molteplici da un punto di vista della potenzialità di calcolo ma esistono due soluzioni principali in termini di architettura informatica:

- Soluzione **on Premise**, ovvero il software si trova in un server interno alla rete del sito dove avviene la raccolta dei dati.
- Soluzione **in Cloud**, ovvero il software si trova installato su un server collocato in datacenter di terze parti. In questo caso il dato non resterà più circoscritto alla rete aziendale ma, attraverso il livello di collegamento e appositi protocolli di comunicazione sicuri, dovrà poter raggiungere la rete esterna.

L'azienda in esame presenta un'architettura on Premise che sfrutta due protocolli di comunicazione, uno per il trasferimento dei dati a livello di campo Modbus RTU e l'altro per il trasferimento dei dati dal livello di capo al livello di Analisi e Gestione Modbus TCP/IP [11]:

Il protocollo Modbus RTU è principalmente usato per l'acquisizione di tutte le informazioni energetiche, come le misure elettriche alle utenze. Il Modbus RTU rappresenta uno dei mezzi

di comunicazione più sfruttati a livello industriale: consente lo scambio dati tra controllori logici programmabili e elementi software, attraverso linee seriali. L'architettura è di tipo master/slave, e grazie alle sue caratteristiche versatili è spesso implementata per i sistemi SCADA nel mondo dell'industria. Un'altra caratteristica di questo protocollo è la codifica binaria e il robusto controllo sugli errori di comunicazione (errori CRC, quando si ha un danneggiamento del pacchetto dati e si necessita una ritrasmissione).

Il Modbus TCP è invece semplicemente un'estensione del protocollo che permette il trasporto dati anche su reti TCP/IP, con numerosi vantaggi gestionali, per quanto riguarda la possibilità di scambio sincronizzato di più informazioni. Questa versione di Modbus consente la comunicazione su reti Internet/Intranet. Inoltre, attraverso dei byte identificativi, è possibile distinguere i diversi slave, mantenendo un gateway con un singolo indirizzo IP. Questo tipo di interfaccia permette di gestire diversi aspetti applicativi a livello più alto.

Si rappresenta uno schema logico dell'architettura e degli elementi di interesse nel monitoraggio del gruppo frigo industriale.

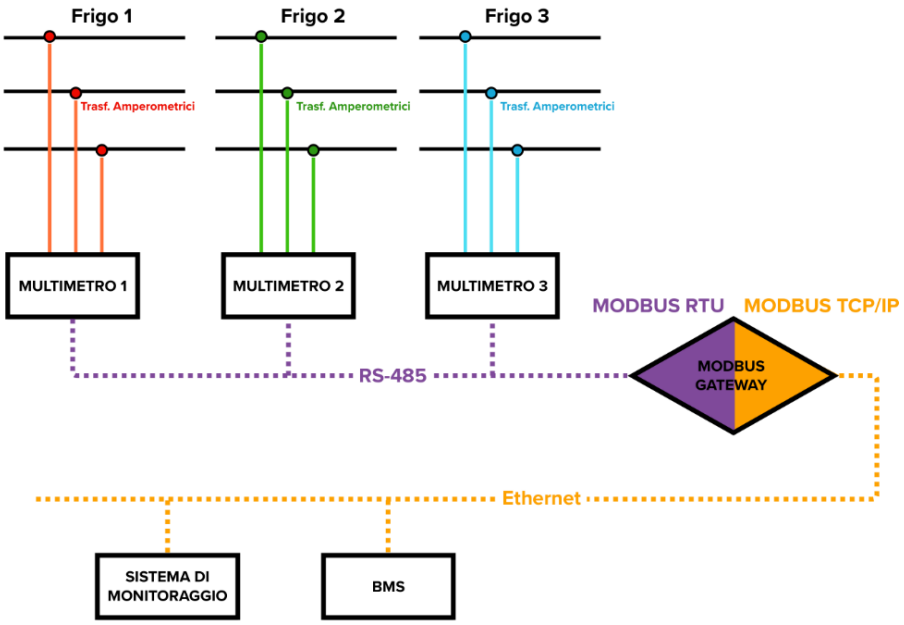

Figura 25 - Schema logico dell'interfaccia tra software e sistema di campo

Per ogni utenza trifase monitorata, per la valutazione dei relativi consumi di energia elettrica, si hanno installati tre trasformatori amperometrici per fase. Questi sono connessi su un multimetro, il quale permette di computare e stimare varie grandezze elettriche di interesse come il cos(φ), l'energia consumata, ecc.

Questi multimetri sono tutti fisicamente cablati tra loro attraverso un bus (Modbus RTU), che permette la comunicazione seriale con un gateway. Il gateway presenta al suo interno un driver di lettura dedicato per la traduzione e codifica dei segnali ricevuti e da inviare, permettendo poi il passaggio a rete Ethernet. Sulla rete Ethernet si interfacciano il sistema di monitoraggio e il BMS (Building Management System).

Attraverso questo sistema, non solo sono a disposizione le informazioni di consumo elettrico delle varie utenze industriale, ma anche i dati del BMS relativi alla gestione dell'edificio risultano fruibili.

3.2 Analisi dei dati di consumo elettrico dei gruppi frigo

Parte integrante del seguente studio è l'analisi critica dei consumi di energia relativi ai gruppi frigo dell'azienda. I concetti base su cui si è inizialmente indagato sono i seguenti:

- L'utilizzo per cui è prodotto il fluido refrigerante.
- Il riconoscimento di possibili stagionalità, sequenze periodiche.
- La definizione dei fattori influenzanti, definiti anche come *drivers*.
- La comprensione del significato fisico delle relazioni tra le componenti e le variabili del sistema in esame.

Come già accennato nel capitolo precedente, i 3 gruppi frigo servono un accumulo termico dedicato, il quale è messo a disposizione di:

- Macchinari in cui è necessario fluido refrigerante di processo (trattamenti termici come tempra, ecc.)
- Sistemi ausiliari per il raffreddamento di elementi meccanici, per evitarne il deterioramento dati i grandi stress termici a cui sono sottoposti certi componenti.
- Unità di Trattamento Aria, messe a servizio sia degli uffici che dei locali industriali

I dati a disposizione erano quindi i consumi elettrici associati ai tre gruppi frigo nel periodo tra gennaio 2020 e fine luglio 2022.

Per poter analizzare e capire il funzionamento del sistema e l'andamento della domanda di potenza frigorifera si è inizialmente visualizzato su scala macroscopica il profilo di potenza media oraria.

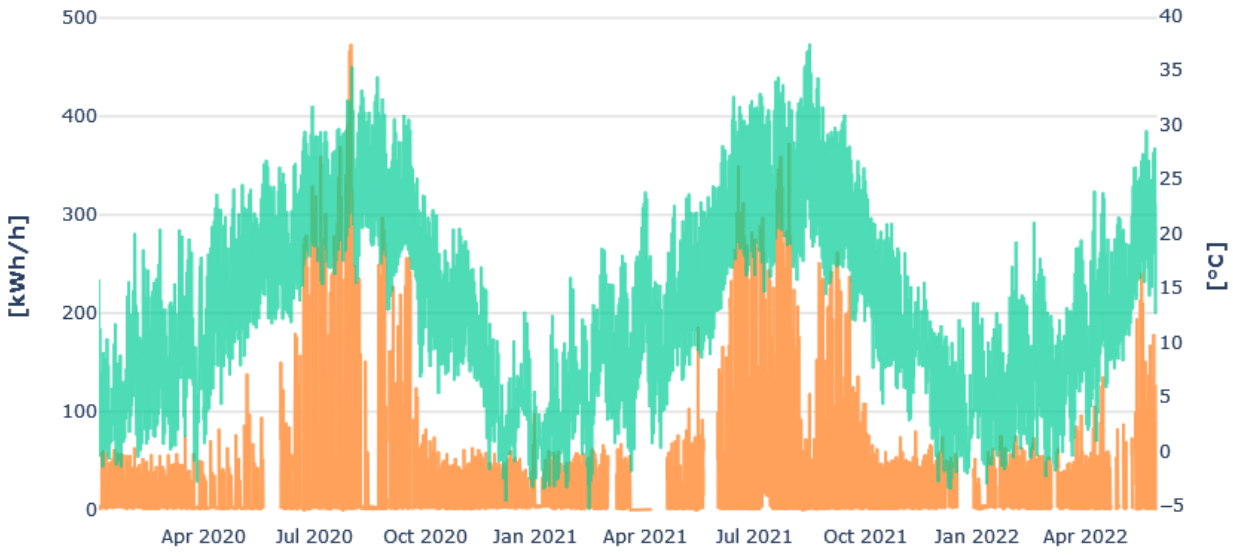

Figura 26 - Profili di temperatura esterna e potenza media oraria dei gruppi frigo nel periodo 2020-2022

In Figura 26 è possibile apprezzare la stagionalità del funzionamento dei gruppi frigo. Come si può vedere si ha un incremento dei consumi nel periodo estivo, con un possibile legame all'andamento della temperatura media esterna del sito in esame.

Da un'analisi preliminare macroscopica si è scelto di classificare il consumo dei gruppi frigo durante l'anno solare in due categorie:

- Funzionamento "estivo": comprende il periodo dal 15 di maggio al 15 settembre
- Funzionamento "invernale": comprende il periodo dal 1° gennaio al 30 aprile e dal 1°ottobre al 31 dicembre

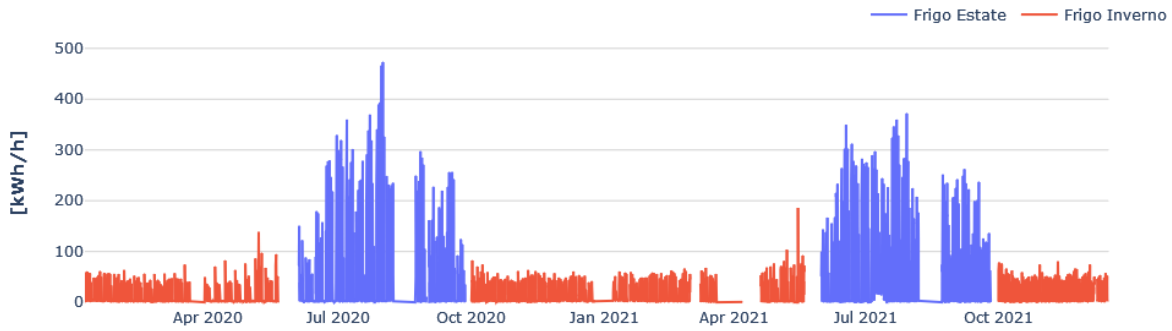

Figura 27 - Distinzione dei profili di carico frigo in funzionamento "estivo" e "invernale"

Le due settimane di inframezzo tra le due modalità di funzionamento rispecchiano pattern di consumo non tipici e altamente variabili, e quindi si è deciso di escluderli dall'analisi in quanto non propriamente rappresentativi e significativi per questa.

Per avere un quadro più completo dell'intensità e della stagionalità dei consumi si è fatto uso delle mappe di calore (o heatmaps), che permettono di avere una visione qualitativa dei consumi nelle 24 ore giornaliere (asse delle ascisse) per un certo periodo in esame. Per l'analisi qualitativa dei profili si sono quindi considerati due mesi di riferimento, uno relativo al funzionamento "invernale" (ottobre 2021) e uno relativo al funzionamento "estivo" (luglio 2021), rappresentati nelle seguenti figure (Figura 28).

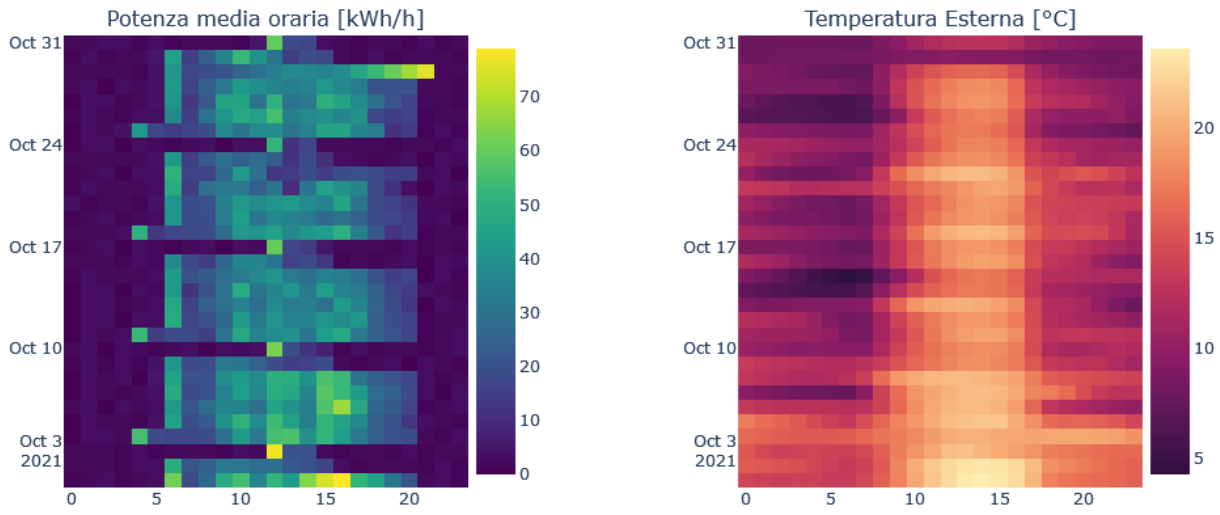

Figura 28 - Heatmaps rappresentanti i consumi orari frigo e la temperatura media oraria esterna del mese di ottobre

La visualizzazione attraverso heatmap permette di cogliere diversi aspetti dell'andamento dei consumi.

È possibile, infatti, distinguere l'alternarsi delle settimane, attraverso la periodica presenza dei week-end caratterizzati da una gradazione di colori più fredda. Inoltre, si distingue una preaccensione anticipata del sistema alle 4:00 AM, probabilmente per una messa a regime dei macchinari, che invece si verifica alle 6:00 AM i restanti giorni lavorativi. Nella parte centrale della giornata si hanno i consumi più elevati, con una leggera attenuazione tra le 13:00 e le 14:00 probabilmente legata alla pausa pranzo. È riportata inoltre un'heatmap relativa alla temperatura esterna, che permette di visualizzare non solo la periodicità giornaliera di questa, ma anche il passaggio a temperature sempre più rigide, apprezzabile dal passaggio a gradazioni più scure verso fine ottobre. È possibile, inoltre, cogliere un'anomalia nei consumi il giorno 29 ottobre, che potrebbe essere legato o a un errore di funzionamento del sistema, o a un errore di campionamento dei sistemi di misura o anche semplicemente a una lavorazione di tipo riparativo, e quindi non legata al processo produttivo periodico tipico.

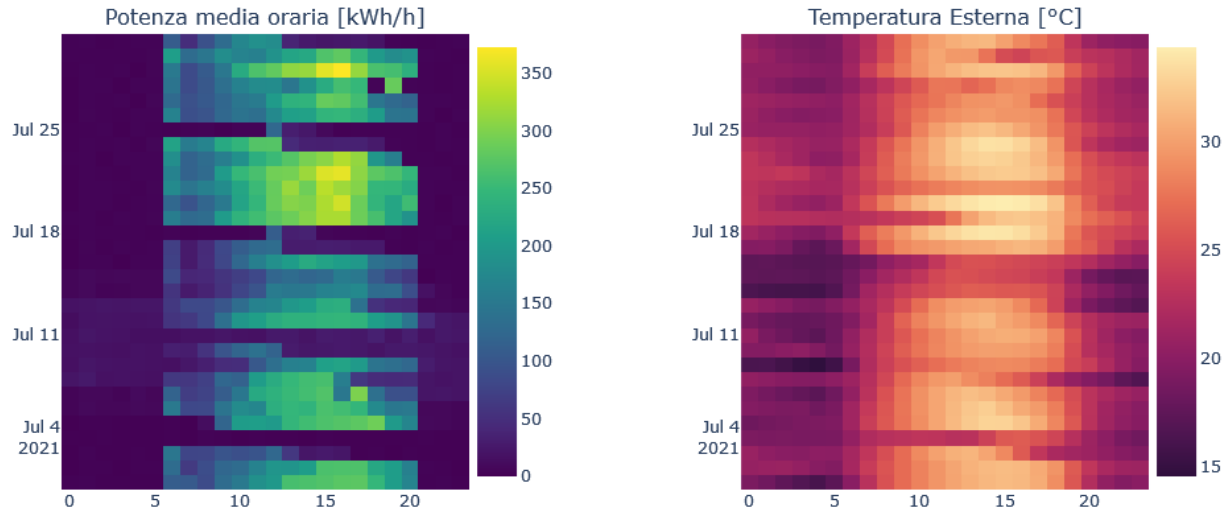

Figura 29 - Heatmaps rappresentanti i consumi orari frigo e la temperatura media oraria esterna del mese di luglio

Analogamente si possono analizzare i consumi e le temperature del mese di luglio 2021. Come si può notare i consumi risultano molto più elevati, e si ha anche un cambio nella schedulazione settimanale. I consumi più elevati sono legati alla modifica delle impostazioni di funzionamento delle unità di trattamento aria di stabilimento, che hanno il compito di mantenere una temperatura accettabile all'interno di locali industriali e uffici.

Per poi capire anche a livello quantitativo l'andamento dei profili, si sono scelte delle settimane tipiche per le rispettive categorie di funzionamento, rappresentate nella seguente figura.

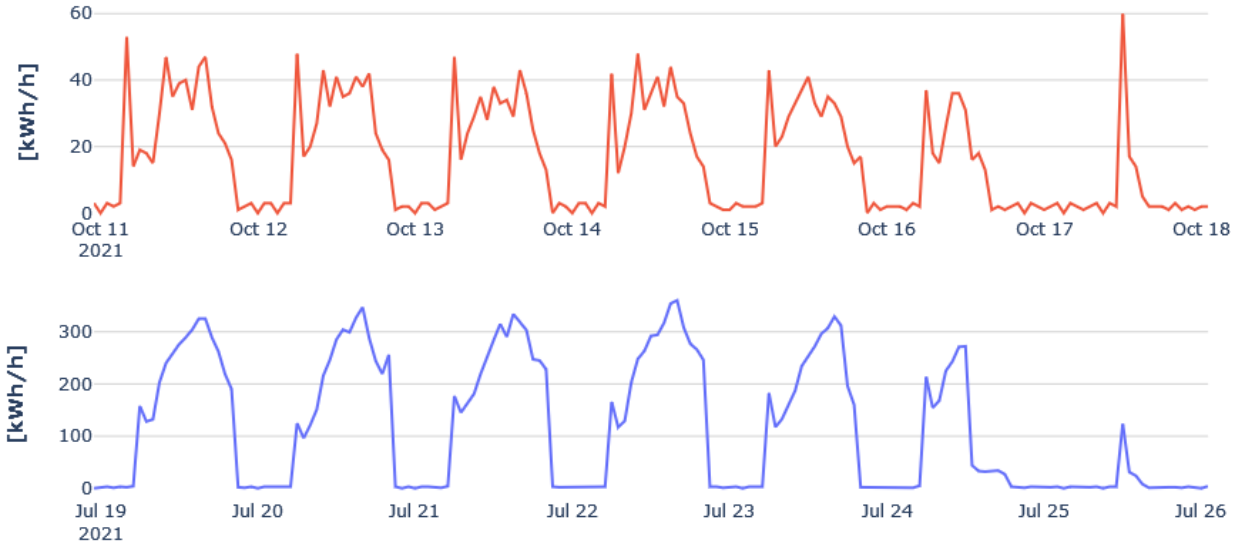

Figura 30 - Confronto dei profili settimanali di consumo dei gruppi frigo per tipo di funzionamento

Come si poteva già intendere dalla lettura dell'heatmap, la tipizzazione dei profili può essere suddivisa nel seguente modo:

- **Lunedì**: caratterizzato da un profilo relativamente conforme a quelli degli altri giorni della settimana lavorativa (Lun-Ven) se non per la preaccensione più intensa e anticipata nel caso invernale (alle 4:00 AM d'inverno mentre alle 6:00 AM d'estate).
- **Giornata Lavorativa**: comprende i giorni dal martedì al venerdì, con i trend di consumo più caratteristici del sistema. La preaccensione in questo caso si verifica sempre alle 6:00 AM.
- **Sabato**: caratterizzato da un mezzo turno lavorativo, con preaccensione uguale alla giornata lavorativa classica ma con l'attenuamento e poi spegnimento delle utenze frigo tra le 13:00 e 18:00
- **Domenica**: caratterizzata da un consumo di routine che ha picco alle 12:00 sia nel caso estivo che in quello invernale

Questo tipo di etichettatura permette di raggruppare in modo consapevole i pattern simili di consumo per poi capire meglio le relazioni con le variabili fisiche e aziendali.

Per cogliere la presenza o meno di differenze di consumo nei giorni settimanali, e valutare se vi sia o meno una periodicità-stagionalità di settimana in settimana, è possibile sfruttare i diagrammi a scatola (o *boxplots*). Raggruppando in diversi modi i dati sulla base del giorno della settimana e sull'ora del giorno, è possibile visualizzare i profili tipici dei giorni della settimana.

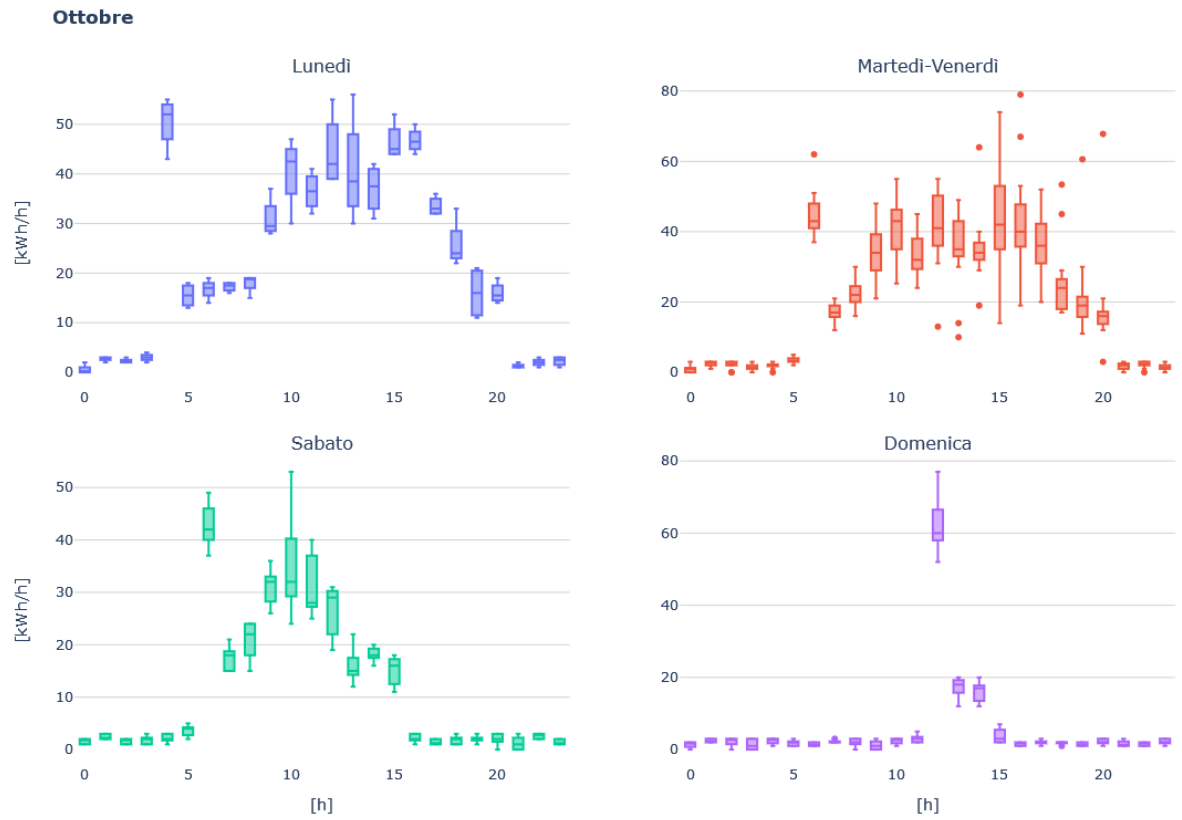

Figura 31 - Boxplot dei consumi frigo orari per giorno della settimana in funzionamento invernale

Figura 32 - Boxplot dei consumi frigo orari per giorno della settimana in funzionamento estivo

Le linee verticali più spesse (o “scatole”) rappresentano il range interquartile: un range esteso può significare un’ampia dispersione dei valori.

Attraverso la seguente rappresentazione (Figura 31 e Figura 32) si notano situazioni ricorrenti: il sabato si ha mezzo turno lavorativo, la domenica è presente un consumo legato a una qualche operazione di routine aziendale, mentre il resto dei giorni si può assumere che il profilo sia pressocchè simile (eccetto il lunedì che presenta sempre un consumo, probabilmente per sopperire allo spegnimento prolungato dell’intera domenica).

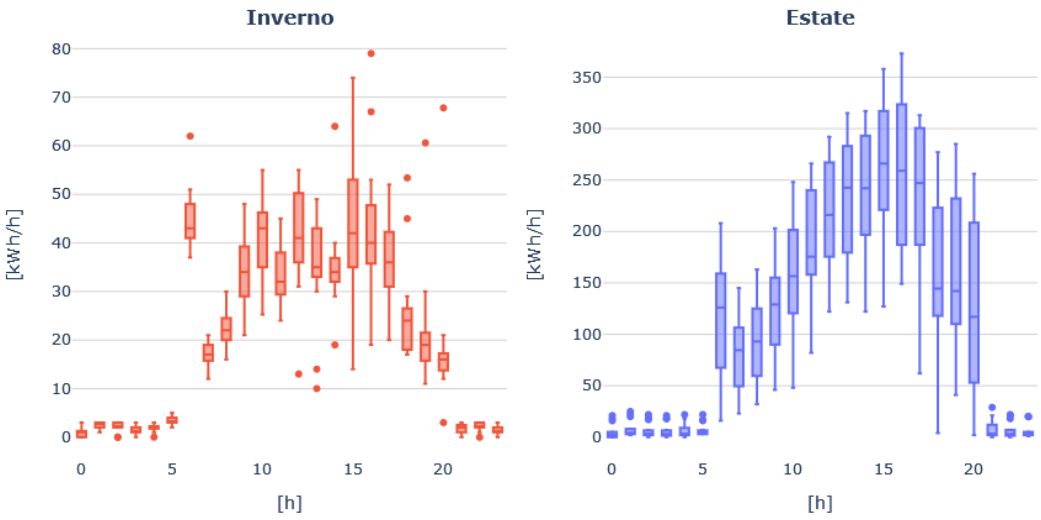

Figura 33 - Boxplot orari giornalieri tipici per i due tipi di funzionamento dei gruppi frigo

La linee sottili verticali rappresentano i valori di massimo e minimo all’interno di un range definito sulla base dell’intervallo interquartile. La linea in mezzo al box rappresenta invece la mediana, mentre i puntini singoli rappresentano gli outlier statistici (e quindi al di fuori del range prima definito).

Questi tipi di rappresentazionni permettono di capire in modo preliminare la presenza di turni standard e l’influenza delle schedulazioni industriali all’interno di un’azienda. Si è verificato successivamente che non vi fossero cambiamenti significativi nelle schedule e nei trend di consumo anche per gli altri mesi costituenti il periodi di funzionamento estivo e invernale.

Una volta capito a livello generale il profilo di consumo tipico, è possibile addentrarsi nella modellazione di questo.

3.3 Pre-processing

Prima di creare un qualunque modello però è di cruciale importanza tenere sempre a mente il seguente concetto:

"Garbage in, garbage out"

Ovvero che se si introducono dati di scarsa qualità e non rappresentativi per la la creazione o l'utilizzo di un modello, ciò che ne esce sarà altrettanto di bassa qualità e insensato.

Per evitare questo, è necessario visualizzare i dati in modo critico per identificare elementi atipici o non significativi, dovuti a errori di monitoraggio, trasmissione di informazioni o vere e proprie anomalie del sistema che potrebbero traviare la modellazione.

Questo tipo di procedimento è una delle parti più lunghe e impegnative nella creazione di un modello ed è composto principalmente da due sotto-procedimenti: l'outlier detection e la sostituzione dei missing values [11].

3.3.1 Outliers e Outlier Detection

In questa fase si va proprio a identificare ed eliminare tutte le anomalie all'interno dei dataset disponibili. Gli outliers (o anomalie) possono essere divisi nelle seguenti categorie:

- Outliers puntuali: sono valori puntuali solitamente fuori scala, e quindi identificabili con strumenti statistici. Possono anche essere dei valori realistici se visti singolarmente ma insensati se contestualizzati nelle condizioni in cui si trovano. Si possono ricondurre spesso a errori momentanei delle architetture di monitoraggio, degli strumenti di misura o degli stessi macchinari.

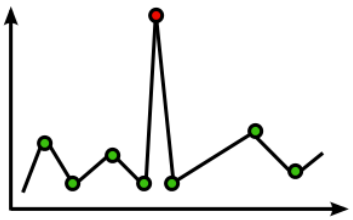

Figura 34 - Schema esemplificativo outlier puntuale

- Sequenze anomale: sono delle vere e proprie sequenze di dati che non rispecchiano il normale funzionamento del sistema in esame. Possono essere dovuti anche in questo caso al mal funzionamento dei sistemi di monitoraggio, del sistema stesso, o alla presenza di festività/particolari eventi (test degli impianti, cambio di macchinari, ecc.).

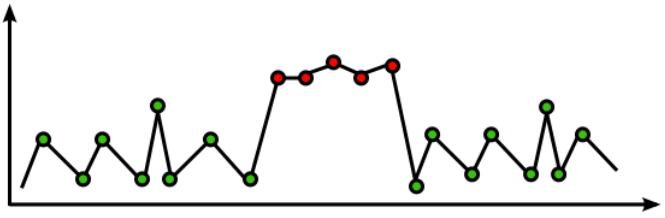

Figura 35 - Schema esemplificativo sequenza anomala

Nel dataset disponibile erano distinguibili da subito outliers di tipo puntuale, che riportavano dei consumi estremamente elevati fuori scala. Il dataset è stato da subito ripulito da tutti questi, affidandosi principalmente al metodo statistico boxplot (eliminazione dei valori al di fuori di certi valori soglia determinati statisticamente dal range interquartile).

Per quanto riguarda le sequenze anomale, sono da tenere da conto questi aspetti:

- il consumo frigo è legato ai macchinari in periodo invernale, e a UTA e macchinari in periodo estivo.
- l'azienda in esame non attua unicamente operazioni di routine ma anche riparazioni su richiesta, e ciò comporta una variabile non prevedibile nei consumi frigo.
- i dati a monitorati comprendono ovviamente anche i periodi di festività, di manutenzione, di mal funzionamento.

Si è quindi andati a ripulire il dataset dai giorni festivi (periodo natalizio e vacanze di Ferragosto), eliminando anche giornate con profili particolarmente irregolari.

3.3.2 Gestione dei missing values

Successivamente è stato necessario sostituire ai valori mancanti: questi possono verificarsi in caso di momentanei errori di monitoraggio/misura o anche solo per la precedente pulizia del dataset dalle tutte le inconsistenze rilevate.

I periodi con dati estremamente discontinui sono stati direttamente eliminati (sopra le 8 ore di informazioni mancanti). Per quanto riguarda invece la gestione dei periodi al di sotto delle 6 ore, considerando che il turno lavorativo standard duri tra le 14 e le 15 ore, si è valutato che la miglior metodologia di sostituzione fosse l'interpolazione lineare. Altri tipi di sostituzione come quello facente uso della media mobile (*moving avarage*) o della costante globale (*global constant*) sarebbero risultati relativamente approssimativi. Un'altra opzione valida sarebbe stato l'uso della sostituzione con costante locale (*local constant*), attraverso l'ausilio di una *lookup-table*: data però la presenza di una forte dipendenza dalla temperatura (in periodo estivo) più che dalla periodicità settimanale vera e propria, e data anche la possibilità di lavorazioni di riparazione non programmate, si è pensato che questo tipo di sostituzione non costituisse per il caso in esame la migliore scelta.

Attraverso l'interpolazione lineare si riesce invece a mantenere per brevi periodi mancanza dati una buona approssimazione del profilo di consumo. La scelta delle 6 ore di "vuoto" come soglia massima per l'interpolazione è frutto di un compromesso tra la necessità di mantenere dei dati attendibili e realistici e comunque la valorizzazione dei dati disponibili, senza che si vada a perdere una giornata intera di dati validi per 2 o 3 ore di assenza di informazioni.

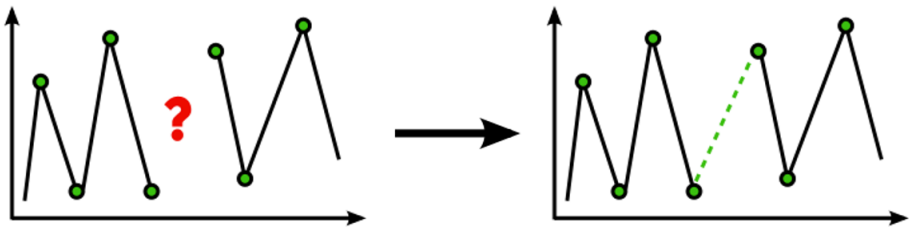

Figura 36 - Schema esemplificativo di interpolazione lineare per la gestione dei missing values

3.4 Metriche Prestazionali

Per misurare le prestazioni predittive di un modello, oltre alla semplice analisi qualitativa dei profili, è necessario determinare delle metriche di riferimento per poter confrontare oggettivamente le performance di questo.

Per valutare le prestazioni del seguente modello si sono sfruttate le seguenti metriche:

- **R^2** – Coefficiente di determinazione:

$$R^2 = 1 - \frac{RSS}{TSS} = \frac{ESS}{TSS}$$

Dove:

$$RSS = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$ESS = TSS - RSS = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

Come si può intendere dalla natura dei termini, il coefficiente di determinazione permette di stimare quanto il modello riesce a spiegare la forza della relazione tra variabile dipendente e variabili indipendenti. Valori prossimi a 1 (o -1) sono indice di bontà del modello mentre valori prossimi a 0 indicano un basso grado di correlazione.

- **MAE** – Errore assoluto medio (*Mean Absolute Error*):

$$MAE = \frac{1}{n} \cdot \sum_{i=1}^{n} |y_i - \hat{y}_i| = \frac{1}{n} \cdot \sum_{i=1}^{n} |e_i|$$

Il MAE permette di verificare l'errore medio evitando che errori negativi e positivi si compensino a vicenda. Il vantaggio di tale metrica è che ha la stessa unità di misura della variabile dipendente e quindi interpretabile a livello fisico.

- **RMSE** – Radice dell'errore quadratico medio (*Root Mean Square Error*):

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (\hat{y}_i - y_i)^2}$$

Anche l'RMSE, come il MAE, ha la caratteristica di avere la stessa unità di misura della variabile da predire (qualità che l'MSE non possiede). Questo tipo di metrica penalizza gli errori con discostamenti maggiori proprio grazie al quadrato del binomio della differenza tra valore predetto e valore attuale. Sarà quindi sempre maggiore (o uguale) al MAE.

- **CV(RMSE)** – Coefficiente di variazione della radice dell'errore quadratico medio:

$$CV(RMSE) = \frac{RMSE}{\bar{y}} \cdot 100$$

Questa metrica è frutto della normalizzazione dell'RMSE, e a differenza delle altre metriche, presenta il vantaggio di essere un valore relativo (percentuale). Per la sua natura, risulta essere ovviamente molto influenzato dalla media dei valori della variabile indipendente. Dalla letteratura si può affermare che valori di CV(RMSE) prossimi o inferiori al 25% sono indice di bontà del modello

Spesso è anche considerato come metrica di riferimento signficativa il Mean Absolute Percentage Error (MAPE), che però presenta come termine al denominatore l'i-esimo valore reale delle osservazioni. Ciò rende questa metrica molto sensibile alla presenza di valori vicini allo 0 e di conseguenza non risulta rappresentativa per il caso in esame, in cui si hanno frequentemente valori nulli nei periodi al di fuori del classico turno di lavoro.

Queste metriche prestazionali appena introdotte saranno fondamentali nella valutazione e verifica dei modelli creati.

4. Modelli

4.1 Modelli di regressione lineare

Per la creazione del modello di regressione lineare, si è inizialmente andati ad analizzare gli andamenti tra possibili driver e la variabile dipendente da prevedere, ovvero il consumo frigo totale.

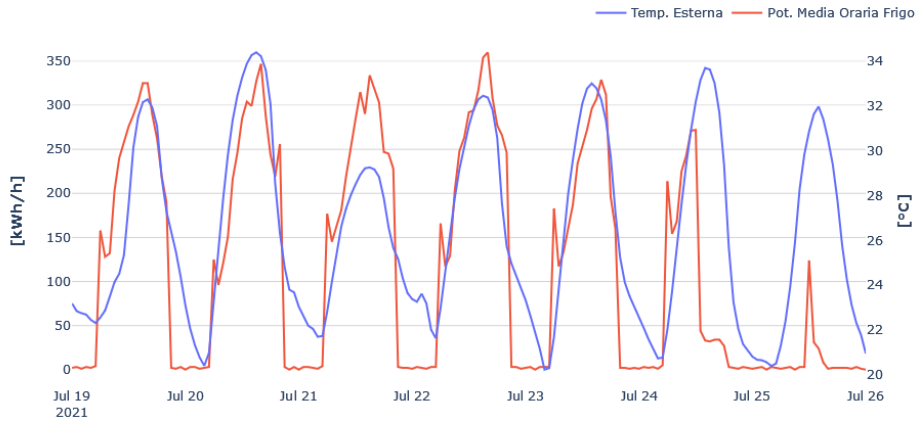

Figura 37 - Parallelismo tra profili di temperatura esterna e potenza media oraria associata ai gruppi frigo

Come si può vedere in Figura 37, considerando una settimana estiva di riferimento, si può notare una correlazione tra i profili di temperatura e i profili di consumo dei gruppi frigo.

Le variabili climatiche esogene a disposizione sono state ottenute grazie al "POWER Data Access Viewer" (*Prediction Of Worldwide Energy Resources*) messo a disposizione dalla NASA. Attraverso questo servizio gratuito è possibile accedere agli archivi di dati climatici di diversa natura e su diverse scale temporali, computati attraverso misurazioni, stime ed elaborazioni delle informazioni ricevute dai sistemi dei satelliti. Nota la geolocalizzazione del sito in esame, si sono quindi estratti i dati orari delle variabili ritenute più influenti per i consumi frigorigeni, ovvero:

- *Temperature at 2 Meters*, cioè la temperatura [°C] oraria a 2 metri dal suolo.
- *All Sky Surface Shortwave Downward Irradiance*, che descrive l'energia per unità di superficie [Wh/m^2] trasmessa dal Sole, considerando anche la presenza o meno di nuvolosità (al contrario delle variabili "*Clear Sky*"). Questa grandezza tiene conto delle radiazioni ad Onda Corta, le quali si può assumere abbiano il maggiore contenuto energetico e quindi il maggiore impatto per il caso in esame.

In ambito energetico si fa spesso riferimento al concetto di firma energetica: mettendo in relazione la temperatura esterna con variabili come la potenza media giornaliera di riscaldamento (per l'inverno) o raffrescamento (per l'estate) è possibile definire un modello lineare descrittivo dei rispettivi sistemi. L'approccio usato nei prossimi sotto-capitoli è molto simile a quello appena esplicitato.

4.1.1 Modello di regressione lineare semplice

Inizialmente, si è per l'appunto creato un modello che mettesse in relazione la potenza media oraria elettrica consumata dai gruppi frigo con la temperatura media esterna oraria riferita a quell'ora stessa.

La retta retta di regressione, fa sì che sia minimizzata la sommatoria degli scarti quadratici tra essa e i punti stessi: noto il valore della variabile indipendente (temperatura esterna) è possibile stimare la variabile dipendente (energia elettrica cosumata dai gruppi frigorigeni).

La formula della retta e la computazione dei suoi coefficienti sono presentati qui di seguito:

$$y = b + m \cdot x$$

Dove:

$$m = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$

$$b = \frac{\sum y - m\sum x}{n}$$

Da cui quindi:

$$P_h = b + m \cdot T_h$$

Dove P_h [kWh/h] rappresenta la potenza media oraria associata al consumo elettrico totale dei gruppi frigo, mentre T_h [°C] rappresenta la temperatura media oraria esterna del sito.

Il primo step è stato quello di rappresentare la totalità dei valori estivi del 2021 per iniziare a fare le prime assunzioni nella creazione del modello.

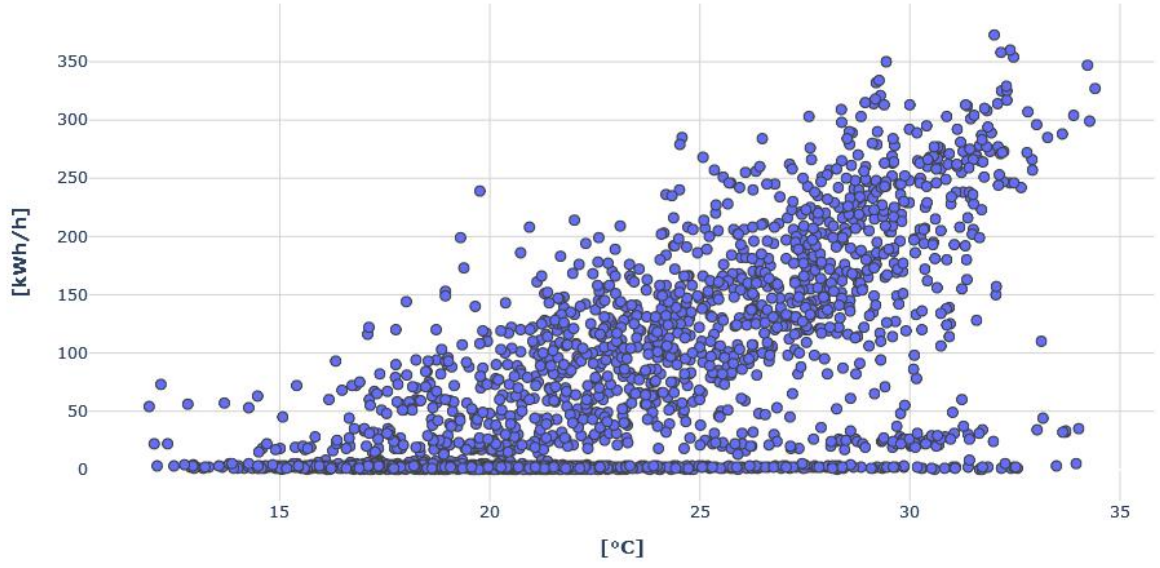

Figura 38 - Scatter plot della potenza media oraria associata ai gruppi frigo in funzione della temperatura oraria nel periodo estate 2021

Come si può vedere la dispersione dei valori è molto elevata, ma comunque si nota fin da subito una certa correlazione tra energia elettrica e temperatura esterna. I valori di monitoraggio relativi ai periodi al di fuori del turno di lavoro non risultano effettivamente correlati alla variabile indipendente (temperatura media esterna).

Per questo motivo si è analizzato andati a considerare unicamente i valori che si ipotizza fossero realmente correlati con la temperatura esterna: i periodi al di fuori del turno di lavoro e i periodi di preaccensione sono stati esclusi nella creazione del modello. L'orario considerato è quindi stato quello tra le 8:00 AM e le 18:00 PM, come si può visionare nella seguente figura, non considerando l'ultima parte del turno di lavoro poichè, valutando i risultati ottenuti da analisi preliminari, si è potuto concludere che le UTA dalle 18 poi venissero spente (lasciando in operazione i macchinari).

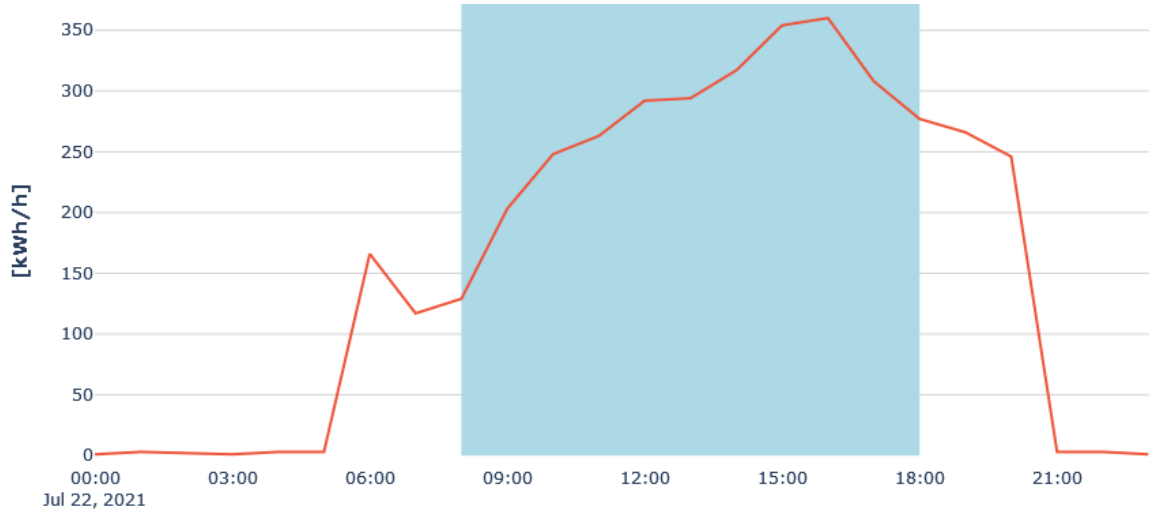

Figura 39 - Rappresentazione range orario giornaliero di riferimento per modello di regressione lineare

Questo tipo di assunzione ha permesso una pulizia del dataset da dati non utili alla correlazione, in modo da creare il modello con informazioni da esso fruibili.

Ripulito il dataset, è stato possibile verificare l'effetto della scelta intoducendo nella rappresentazione la retta di regressione e valutando le performance di questo primo tentativo di modello, non solo sfruttando come driver la temperatura esterna (chiamata anche T_ext [°C] per praticità) ma anche l'irradianza solare (abbreviata invece come I_sol [Wh/m^2]).

Figura 40 - Scatter plot e rette di regressione considerando i valori compresi nel range orario di riferimento

Come si può cogliere dalla rappresentazione, temperatura e potenza media sembra che abbiano una forte correlazione, mentre per quanto riguarda l'irradianza ciò non risulta vero.

Per la valutazione della bontà della retta di regressione è spesso usato in letteratura il parametro R^2, il quale indica elevata correlazione se vicino a 1 (pendenza positiva) o -1 (pendenza negativa).

Il valore ottenuto nel seguente caso è pari a 0.61 (con come driver la temperatura esterna), che risulta accettabile ma non eccellente.

La causa di questa alta dispersione dei dati potrebbe essere imputabile a:

- la presenza di una domanda variabile frigorigena dei macchinari, legata più alle lavorazioni programmate che alla temperatura esterna.
- l'inerzia termica delle componenti fisiche del sistema (dipendente dai materiali e dalla capacità termica di questi) potrebbe interferire e ritardare il carico sui gruppi frigo.

Entrambe questi problemi sono stati analizzati criticamente, e per entrambi questi si è andati a definire delle possibili soluzioni.

Per valutare l'impatto dell'inerzia termica, si sono tenuti in considerazione le seguenti elaborazioni delle variabili climatiche:

- il singolo valore di temperatura/irradianza, shiftato di n-ore precedenti.
- la SMA, Simple Moving Average, ovvero la media degli ultimi valori di temperatura sempre di n-ore precedenti

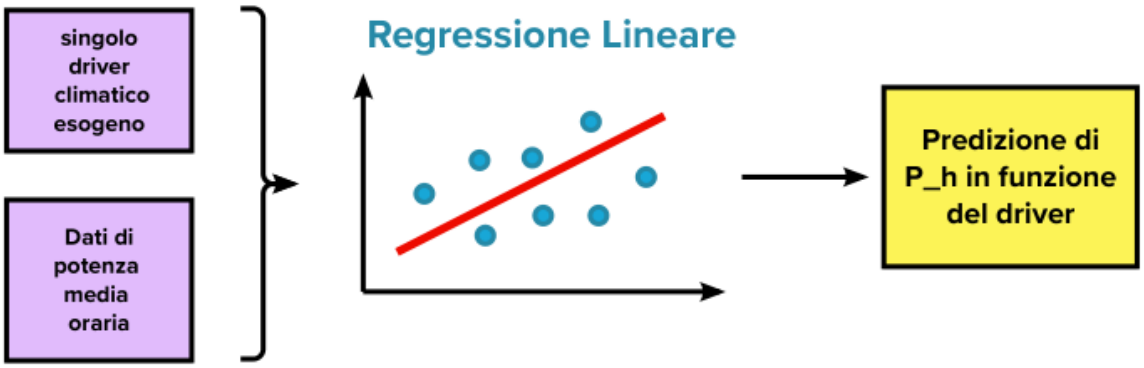

Figura 41 - Schema logico rappresentante variabili di input e variabili di output nel modello

Sfruttando questo approccio, si riportano in Figura 42 parte dei risultati ottenuti:

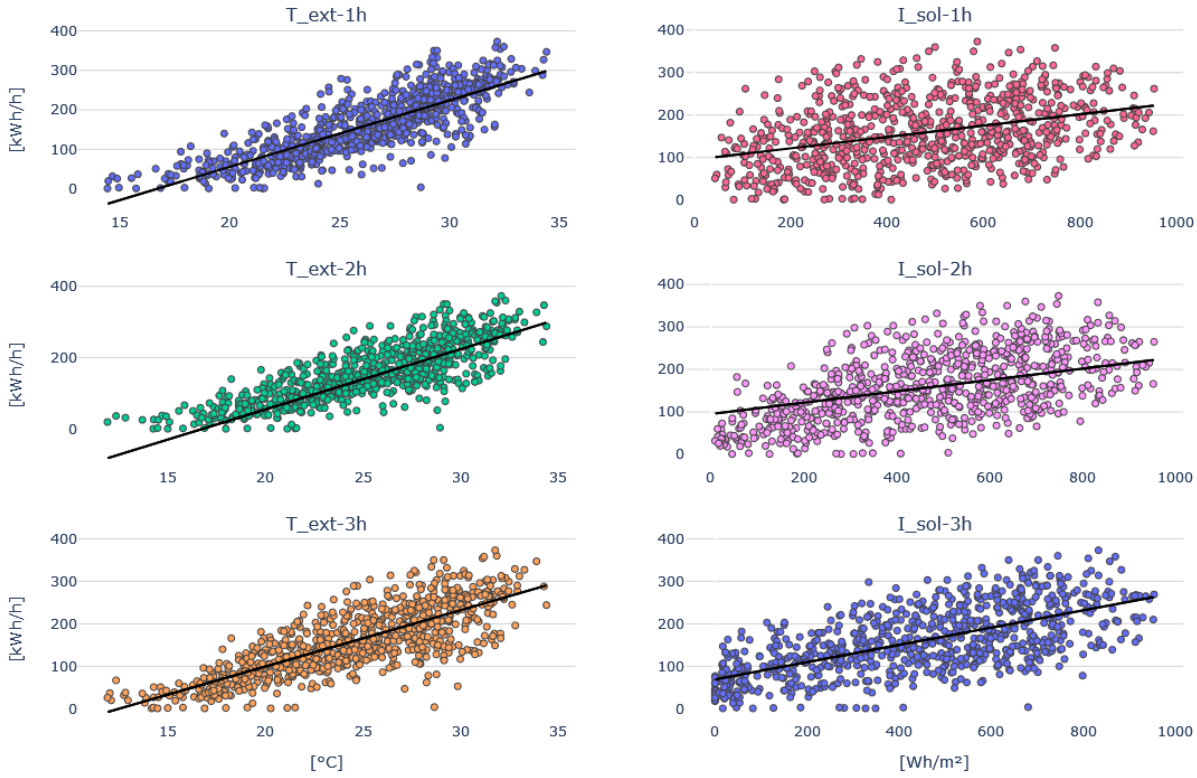

Figura 42 - Rappresentazione del rapporto tra variabili climatiche sfasate e potenza oraria

Si può apprezzare immediatamente una minore dispersione dei dati, estremamente evidente soprattutto per quanto riguarda l'irradiazione solare.

Per valutare però a livello quantitativo quale sia il driver più rappresentativo si è andati a valutare le metriche prestazionali dei modelli fino a 6 ore di sfasamento delle variabili, i cui risultati sono presenti in Tabella 5.

Tabella 5 - Metriche prestazionali per predizione dell'energia frigo totale considerando i diversi driver esogeni

Driver	R2 [-]	RMSE [kWh/h]	MAE [kWh/h]	CV(RMSE) [%]
T_ext	0.61	49.07	39.02	30.96
T_ext-1h	0.67	45.59	35.53	28.76
T_ext-2h	0.65	46.49	36.20	29.33
T_ext-3h	0.62	48.78	37.99	30.78
T_ext-4h	0.57	51.99	40.77	32.80
T_ext-5h	0.50	55.75	44.32	35.17
T_ext-6h	0.43	59.58	48.00	37.59
I_sol	0.01	78.59	64.93	49.59
I_sol-1h	0.13	73.47	60.73	46.35
I_sol-2h	0.34	64.26	52.29	40.54
I_sol-3h	0.43	59.37	48.19	37.46
I_sol-4h	0.42	59.92	49.26	37.80
I_sol-5h	0.35	63.48	52.49	40.05
I_sol-6h	0.27	67.49	55.83	42.58
SMA_T-2h	0.66	46.19	36.13	29.14
SMA_T-3h	0.68	44.84	34.81	28.29
SMA_T-4h	0.68	44.51	34.45	28.08
SMA_T-5h	0.68	44.91	34.70	28.34
SMA_T-6h	0.66	45.79	35.29	28.89

Per quanto riguarda la temperatura, i driver più rappresentativi risultano essere la temperatura media esterna dell'ora precedente e la media mobile comprendente le ultime 3 ore. Per quanto riguarda l'irradianza invece, si considera come più significativa la variabile I_sol-3h, ovvero l'irradianza solare relativa a 3 ore prima.

4.1.2 Modello di regressione lineare multipla

Si è poi valutata la possibilità di una regressione lineare multipla, il cui principio è analogo e la cui formula è la seguente:

$$P_h = \beta_0 + \beta_1 \cdot T_h + \beta_2 \cdot I_{sol,h}$$

In questo caso però si è direttamente andati a considerare le due variabili climatiche indipendenti più rappresentative e semplici, ovvero la T_ext-1h e la I_sol-3h.

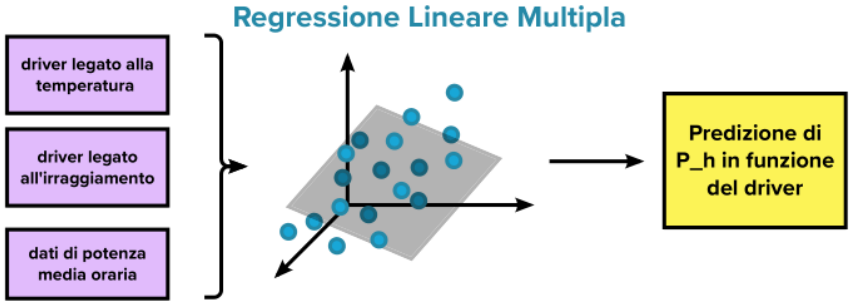

Figura 43 - Rappresentazione logica delle variabili di input e output al modello di regressione lineare multipla

Da tenere da conto è la relazione di multicollinearità (ovvero alta correlazione) tra le due variabili dipendenti, temperatura e irradianza, che potrebbe portare a un modello ridondante, senza alcun miglioramento.

Riportando infatti graficamente le variabili, è possibile apprezzare la correlazione tra queste, notando però comunque un'alta dispersione dei dati.

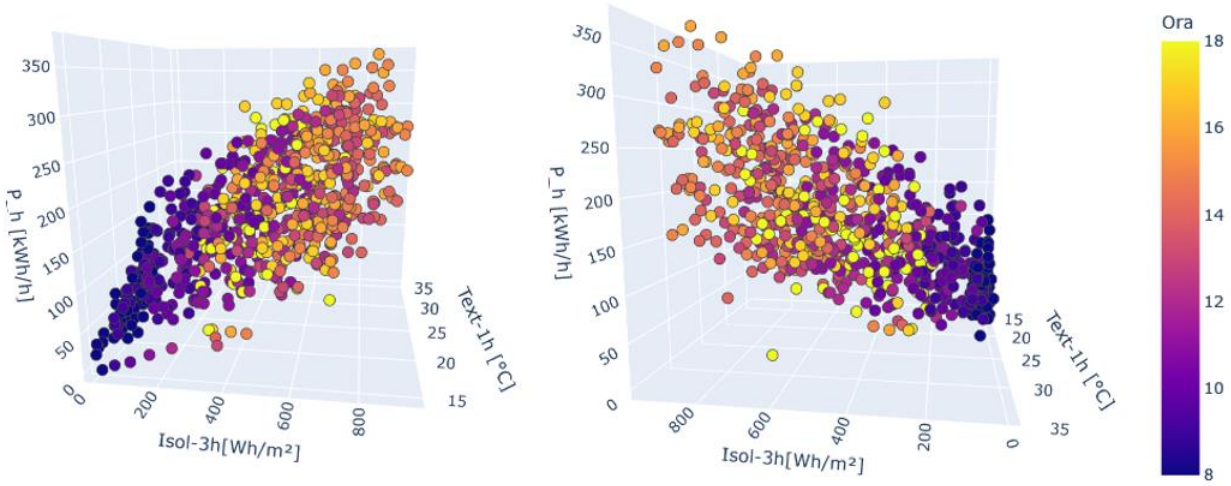

Figura 44 - Rappresentazione 3D della regressione lineare multipla

Per la rappresentazione 3D si è scelto di rappresentare attraverso la scala cromatica l'ora del giorno associata ai consumi. Come si può vedere, tendenzialmente i consumi più elevati si hanno quando la gradazione risulta tendente all'arancione, e quidni tra le 15:00 e le 17:00. Non si ha però una disposizione complanare dei punti, e ciò significa una inter-correlazione debole tra variabili indipendenti e variabile dipendente. Questo si rispecchia anche nei valori delle metriche prestazionali, le quali non presentano particolari miglioramenti rispetto alle semplici regressioni precedentente viste.

Tabella 6 - Metriche prestazionali regressione multipla

Regressione Multipla	
R2 [-]	0.67
MAE [kWh/h]	35.32
RMSE [kWh/h]	45.09
CV(RMSE) [%]	28.45

4.1.3 Modello di regressione lineare per previsione dei consumi UTA

Attraverso un'analisi critica delle caratteristiche del sistema e del caso studio in esame, si è determinata una metodologia di allocazione ipotizzando che la domanda frigo legata alle lavorazioni dei macchinari, sia totalmente indipendente dalla temperatura esterna

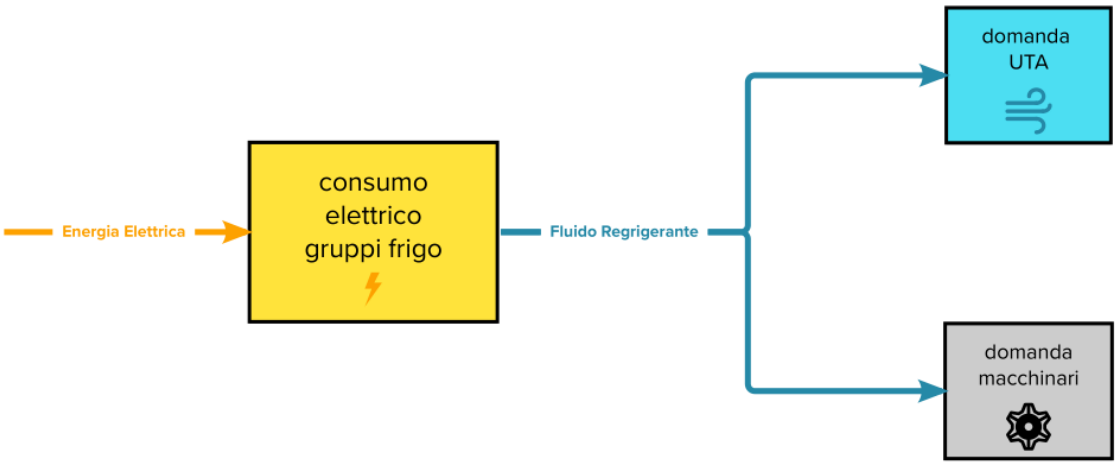

Figura 45 - Schema logico ripartizione dei consumi frigo

Inizialmente si è infatti andati a verificare se la temperatura fosse un driver valido nel periodo invernale. Si è potuto affermare che non vi fosse forte correlazione tra temperatura e consumi frigo (l'R^2 risulta infatti sotto il valore di 0.005), dato che le UTA in inverno non fanno uso di fluido refrigerante e che il funzionamento dei macchinari non è apparentemente influenzato dalle condizioni esterne.

Si è quindi usato un processo semplificato di determinazione del profilo tipico invernale, che possiamo definire come centroide rappresentativo, ipotizzando che questo possa descrivere la domanda frigorigena allocabile unicamente al funzionamento dei macchinari. Questo profilo tipico può essere visto come uno zoccolo duro anche per il periodo estivo.

In questo modo è stato possibile ottenere, dalla differenza tra consumo totale e consumo dei macchinari, il valore di potenza media oraria riferibile alle UTA.

Si rapprenta di seguito il profilo medio ottenuto dall'anailisi dei dati invernali.

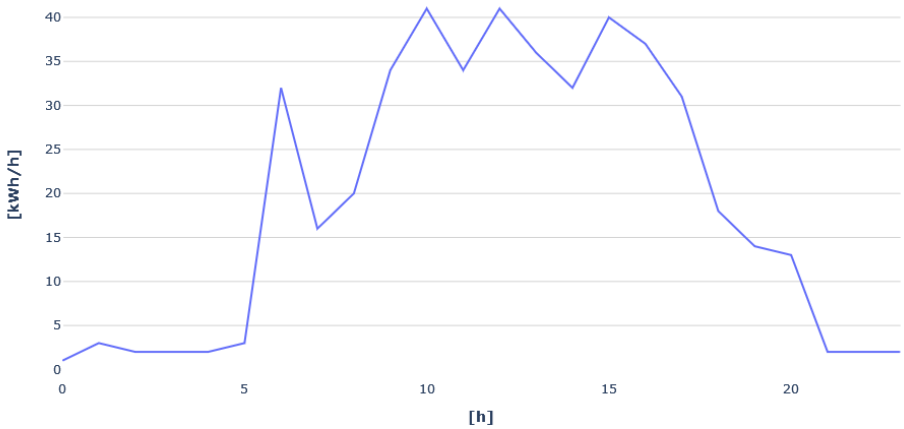

Figura 46 - Centroide rappresentante il profilo tipo dei consumi frigo invernali

Conseguentemente si è attuata tra i consumi totali giornalieri e questo profilo standard, in modo da ottenere una stima dei consumi frigo delle UTA.

Si rappresenta qui di seguito a titolo esemplificativo la profilazione ottenuta:

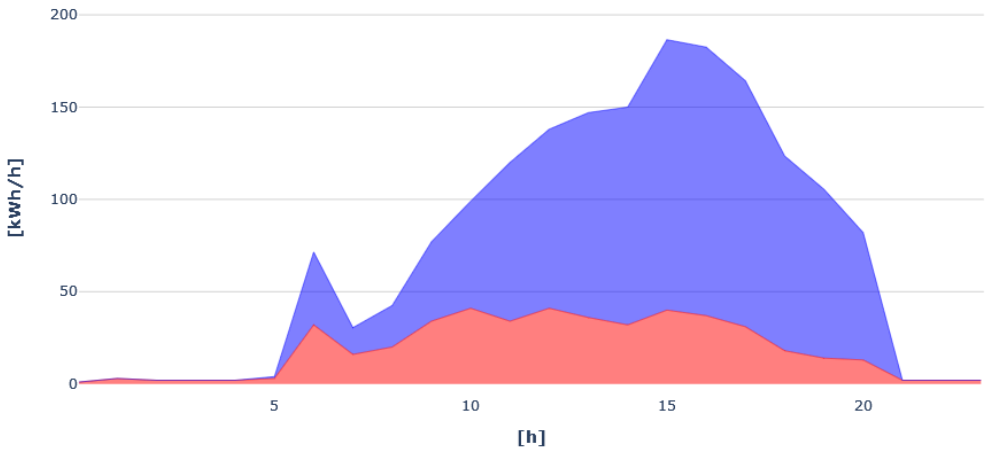

Figura 47 - Rappresentazione della quota di consumo frigo allocabile ai macchinari in funzionamento estivo

Analogamente a quanto fatto per quanto riguarda i consumi frigo totali, si è quindi andati a valutare le prestazioni dei modelli di regressione lineare semplice al variare del driver climatico. Le metriche prestazionali computate sono riportate in Tabella 7.

Tabella 7 - Metriche prestazionali per predizione dell'energia frigo UTA considerando i diversi driver esogeni

Driver	R2 [-]	RMSE [kWh/h]	MAE [kWh/h]	CV(RMSE) [%]
T_ext	0.59	49.21	39.28	39.09
T_ext-1h	0.66	44.76	34.97	35.55
T_ext-2h	0.66	44.74	35.03	35.53
T_ext-3h	0.64	46.18	36.28	36.68
T_ext-4h	0.60	48.52	38.11	38.54
T_ext-5h	0.54	51.65	41.08	41.02
T_ext-6h	0.48	55.21	44.47	43.85
I_sol	0.00	76.41	63.42	60.69
I_sol-1h	0.10	72.60	60.36	57.66
I_sol-2h	0.30	64.02	52.40	50.85
I_sol-3h	0.42	58.40	47.46	46.38
I_sol-4h	0.43	57.72	47.42	45.84
I_sol-5h	0.38	60.36	50.07	47.94
I_sol-6h	0.30	63.92	53.02	50.77
SMA_T-2h	0.64	45.95	35.99	36.49
SMA_T-3h	0.67	44.12	34.37	35.04
SMA_T-4h	0.68	43.25	33.69	34.35
SMA_T-5h	0.68	43.07	33.50	34.21
SMA_T-6h	0.68	43.38	33.73	34.45

Come si può vedere, risultano complessivamente più efficaci le medie mobili della temperatura nella descrizione di questo tipo di consumo, non solo per quanto riguarda l'R^2 ma anche per il CV(RMSE).

La creazione di modelli che sfruttino la regressione lineare per la previsione dei consumi ha il vantaggio di essere molto semplice e interpretabile a livello fisico, soprattutto nel caso sia importante un confronto tra diverse condizioni di funzionamento o modifiche del proprio sistema (la pendenza della retta può essere interpretabile come funzione dell'efficienza del sistema, e la sua intersezione allo stesso tempo può essere correlata ai carichi endogeni o altre caratteristiche del sistema, come le coperture di questo).

Gli svantaggi principali dei modelli di regressione lineare sono però l'accuratezza limitata e l'incapacità di modellare e fare proprie le variabili temporali.

4.2 Decision Tree

4.2.1 Funzionamento algoritmo Decision Tree

Un altro modello considerato per la caratterizzazione e predizione dei consumi frigo è il Classification And Regression Tree (CART), un algoritmo facente parte della categoria Decision Tree, contraddistinto dalla tipica rappresentazione ad albero.

Il funzionamento può essere descritto brevemente nel seguente modo:

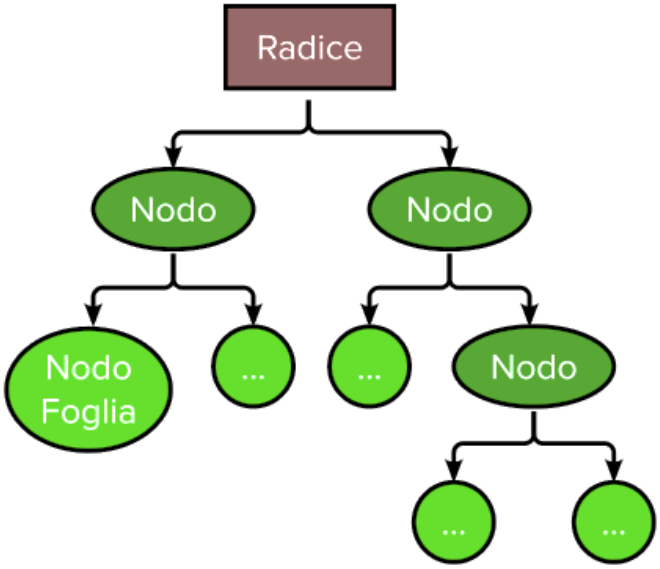

Figura 48 - Schema semplificativo della struttura di un CART

L'intero insieme delle osservazioni (ovvero i valori che si vogliono predirre) può essere localizzato all'interno del primo nodo, detto nodo radice. Ad ognuno di questi valori sono associate una serie di etichette, di tipo categorico (giorno della settimana, ora del giorno, ecc.) o di tipo numerico (temperatura esterna attuale, valore di potenza media oraria dell'ora precedente, ecc.). Questo set iniziale di dati è splittato in modo binario ad ogni nodo, sulla base del valore soglia di una certa etichetta (ad esempio in un nodo generico, si potrebbe avere una divisione dei del dataset nel presente nodo in base a se la temperatura esterna dell'ora prima sia superiore o inferiore a 25 °C). Le variabili decisionali (etichette) e i rispettivi valori soglia sono funzione degli iperparametri settati, permettendo o meno una divisione del dataset in altri due dataset, caratterizzati da una maggiore intra-similarità (o "purezza") rispetto al nodo precedente. Sulla base quindi degli iperparametri scelti e dei valori a questi assegnati, si raggiunge un numero finito di nodi foglia, nella quale sono presenti osservazioni tra loro simili, caratterizzate dallo stesso "percorso" definito dalle differenti condizioni poste sulle variabili decisionali (etichette).

Il modello riesce quindi a definire quale sia la variabile decisionale e il valore soglia associato ad essa che permetta lo split più puro ad ogni nodo fino ad arrivare a un nodo foglia finale. Nel caso in esame, trattandosi di un problema di regressione, l'output di ogni nodo è di tipo numerico, e risulta essere il valore più rappresentativo (e generalmente più vicino) ai valori delle osservazioni di questo.

Gli iperparametri più comunemente usati sono i seguenti [14]:

- **Criterio** (*criterion*), è il parametro su cui si basa la purezza dello split. Il più comune risulta essere l'errore quadratico medio.
- **Profondità massima** (*max_depth*), è per l'appunto la profondità massima raggiungibile dall'albero, contando come 0 il nodo radice.
- **Numero minimo di campioni per nodi interni** (*min_samples_split*), impedisce la creazione di un ulteriore nodo di split (fermandosi quindi a un nodo foglia) nel caso la divisione portasse ad un numero di campioni inferiore al valore settato del suddetto iperparametro.
- **Numero minimo di campioni per nodi foglia** (*min_samples_leaf*), simile all'iperparametro precedente, definisce un numero minimo di entità per nodo foglia.
- **Diminuzione minima di impurità** (*min_impurity_decrease*), permette o meno il verificarsi di uno split sulla base del miglioramento della purezza complessiva che si otterrebbe nel caso della seguente divisione.

È importante settare in modo opportuno i seguenti iperparametri per evitare alberi troppo o troppo poco profondi. Un albero esageratamente profondo, porterebbe ad un modello non abbastanza generico, e quindi molto efficace se applicato ai dati sfruttati per la sua creazione, ma poco accurato se applicato a dati esterni (e quindi poco utile nel caso di forecasting). Questo tipo di problema, in cui si ottengono alte performance con i dati sfruttati per la creazione del modello, ma basse prestazioni con dati esterni, è chiamato in letteratura sovradattamento o *overfitting*. Al contrario, con un modello non abbastanza profondo, si rischia di non descrivere sufficientemente in dettaglio il sistema in esame, perdendo quindi possibili informazioni caratterizzanti di questo.

4.2.2 Applicazione modello Decision Tree

Prima di eseguire la calibrazione e messa appunto dei seguenti iperparametri (anche detta "tuning") è necessario scegliere quali variabili sia sensato fornire al modello. Il vantaggio dell'algoritmo dei CART è che la variabile decisionale e il suo valore soglia sono scelti "automaticamente" sulla base della purezza dello split, ma comunque ciò non toglie che una pessima scelta dei dati di input possa danneggiare la modellazione del sistema.

Oltre alle variabili climatiche esogene e le combinazioni di queste sfruttate nel modello del sottocapitolo precedente (T_ext, I_sol e SMA), si è valutato che risultasse impattante introdurre altri due tipi di variabile:

- **Variabili autoregressive**, ovvero i valori di potenza media oraria dei consumi riferita a 1, 2, … , n-ore precedenti.

- **Variabili categoriche temporali**, ovvero il giorno della settimana, l'ora del giorno, il mese e settimana dell'anno, che consentono una modellazione in grado di cogliere possibili schedulazioni tipiche dell'azienda.

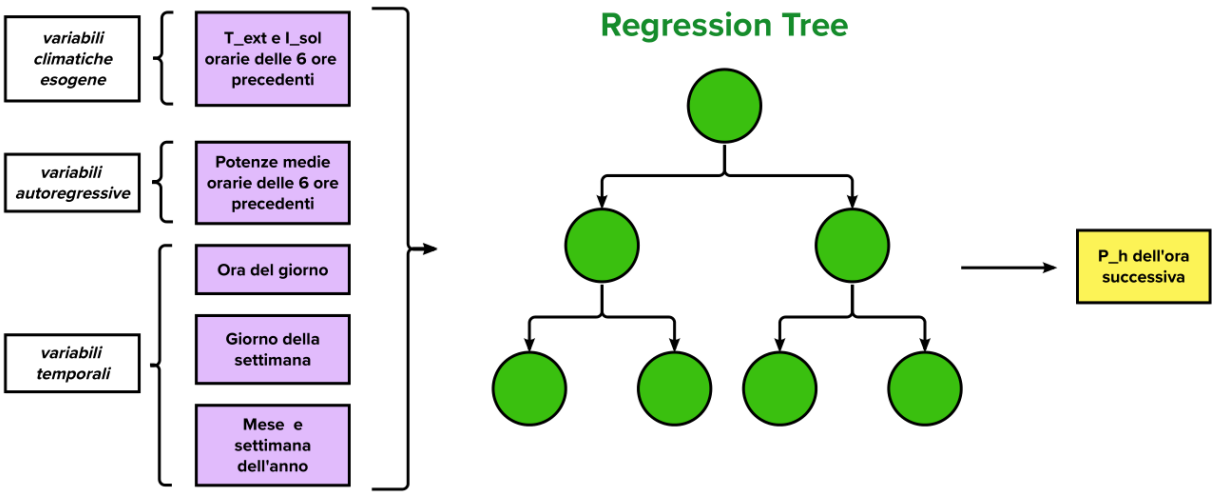

Figura 49 - Schema logico rappresentante le variabili di input al modello CART

Anche in questo caso, si è preso sempre come riferimento unicamente l'anno 2021. Si è però provato a sfruttare il vantaggio delle variabili temporali, senza una vera e propria distinzione tra periodo di funzionamento estivo e periodo di funzionamento invernale.

Per valutare l'efficacia del modello, e anche per poter verificare la presenza o meno di overfitting nel modello, si è scelto di dividere il dataset in due parti:

- Dataset di Training: comprendente circa l'80% dei dati, è il dateset su cui avviene il fitting e quindi su cui il modello si "allena". Proprio perché rappresenta i dati di allenamento è importante la scelta delle informazioni da fornire: se non si danno le informazioni giuste si rischia che il modello non riesca a cogliere certi pattern di funzionamento.

- Dataset di Testing: è composto da 8 settimane (una di gennaio, una di febbraio, 2 di giugno, due di settembre, una di novembre e infine una di dicembre), tutte caratterizzate da diversi livelli medi di consumo in modo da avere un confronto valido. In questo modo si ha a disposizione un dataset sufficentemente vario per valutare le prestazioni del modello nelle le varie modalità di funzionamento.

Successivamente si è attuato il processo di tuning. Sfruttando come leve gli iperparametri, si è andata a modificare la profondità dell'albero, la purezza degli split e il numero minimo di elementi in ogni nodo foglia, per raggiungere delle prestazioni accattabili nel modello.

L’albero ottenuto è riportato qui di seguito.

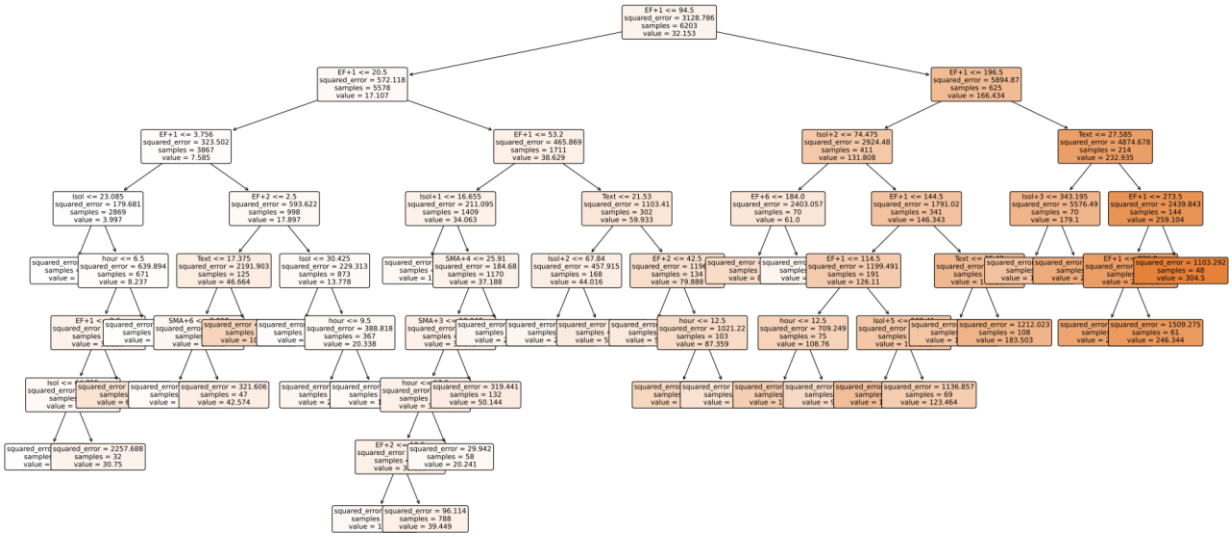

Figura 50 - Rappresentazione grafica del modello CART generato

Si può notare come nei primi split la variabile più ricorrente sia la “EF+1”, ovvero la potenza media oraria riferita all’ora precedente. Ciò significa che le variabili autoregressive risultano essere molto impattanti nella descrizione del sistema.

Gli iperparametri settati sono:

- Come *criterion*, è stato scelto l’errore quadratico medio, in quanto consigliato spesso dalla natura e avente anche il vantaggio di non essere sensibile alla presenza di valori prossimi allo 0 (caratterisitca cruciale per il caso in esame, dove spesso si hanno valori di potenza media oraria estremamente bassi se non nulli).
- La *max_depth* impostata pari a 10, per ottenere comunque valori abbastanza rappresentativi senza però ricadere nell’overfitting.
- Un *min_impurity_decrease*, settato a 2, poiché con valori al di sopra di questo si effettuavano degli “*stopping*” degli split preventivi che non giovavano alle capaictà predittive dell’algoritmo, mentre al di sotto di questo erano permessi split che non migliaravano sensibilmente la purezza dei nodi foglia .
- Un *min_samples_leaf*, impostato pari a 30, in quanto si è valutato che porlo inferiore avrebbe portato a nodi foglia eccessivamente specifici (e quindi poco rappresentativi).

In totale i nodi foglia ottenuti risultano essere 32. Ciò comporta che l’intero anno di consumi del sistema può essere descritto unicamente da 32 valori di potenza media, che approssimano al meglio, entro i limiti del modello, la potenza media oraria dell’ora successiva (da prevedere).

Si rappresentano qui i profili di consumi orari previsti e reali per il periodo di training e per quello di test.

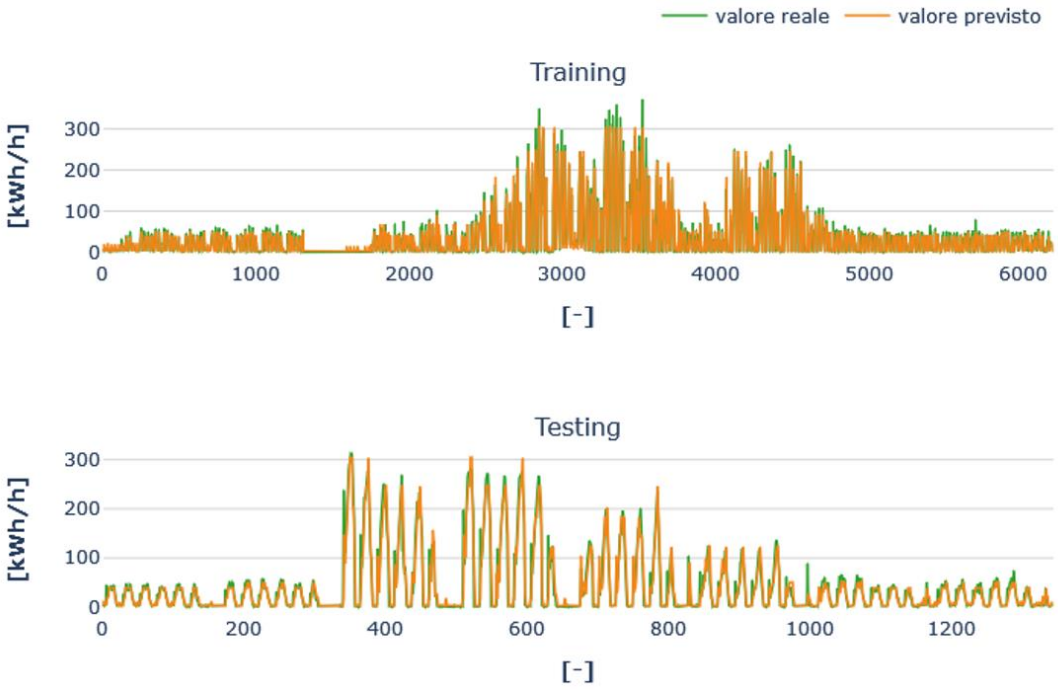

Figura 51 - Confronto tra profili reali e predetti per il periodo di training e di testing

Si può apprezzare visivamente da subito la bontà del modello ottenuto, ma per una migliore analisi si riportano anche due settimane generiche del periodo di testing.

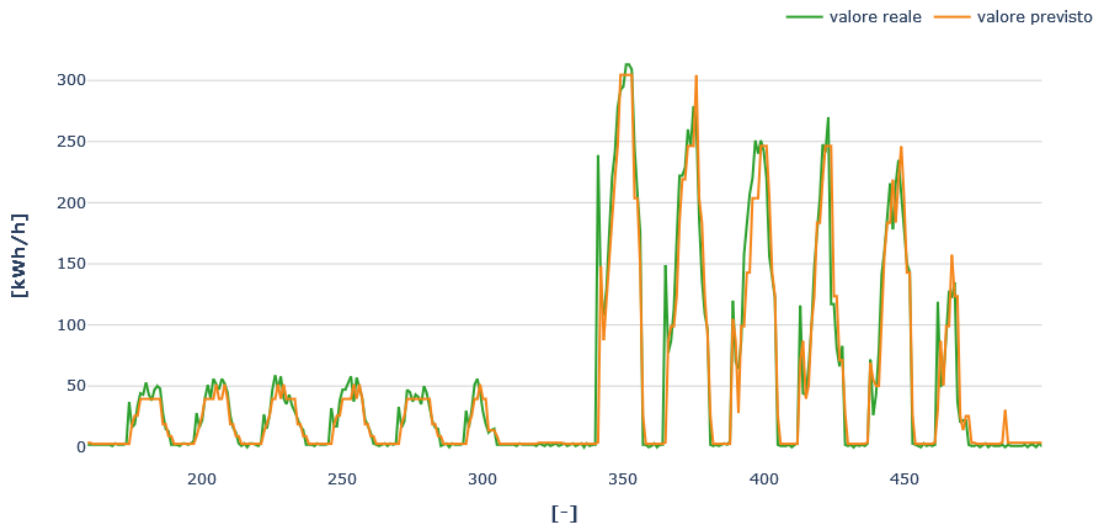

Figura 52 - Confronto tra profilo reale e profilo predetto dal modello CART di una settimana invernale (sx) ed estiva (dx)

La presenza delle variabili temporali porta i seguenti vantaggi, evidenti in Figura 52:

- Modellazione sufficientemente accurata dei consumi legati alla preaccensione, che rappresentava invece un grande limite per quanto riguarda i modelli di regressione lineare classici.
- Riconoscimento del giorno della settimana grazie all'etichetta dedicata, che quindi permette al modello di distinguere anche sabati e domeniche.
- Riconoscimento di una schedule giornaliera, anch'esso possibile grazie alla variabile "ora del giorno"
- Riconoscimento del tipo di funzionamento (estivo o invernale) in cui ricade grazie alle variabili "mese dell'anno" e "settimana dell'anno".

Come si può notare però il modello presenta ancora dei limiti nel forecasting delle ore di punta, ma riesce comunque a cogliere in modo efficace molte altre dinamicità dei consumi.

I risultati ottenuti dal seguente modello sono i seguenti:

Tabella 8 - Metriche prestazionali del modello CART

Metrica	Training	Testing
R2 [-]	0.90	0.91
MAE [kWh/h]	8.18	9.98
RMSE [kWh/h]	17.33	19.83
CV(RMSE) [%]	53.90	46.00

Si notano dei valori di RMSE e MAE relativamente bassi, se confrontati con gli ordini di grandezza dei profili di potenza media oraria, sintomo di buone prestazioni da parte del modello. Anche il valore di R^2 risulta essere relativamente vicino a 1 sia nel periodo di training che quello di testing

Salta però all'occhio un valore abbastanza elevato per quanto riguarda il CV(RMSE). Il coefficiente di variazione della radice dell'errore quadratico medio permette di osservare e quantificare a livello relativo/percentuale gli scostamenti tra i valori reali e quelli previsti. Dato che in questo caso, il periodo considerato è l'intero anno, comprensivo anche delle ore fuori i turni di lavoro, il valore medio dei campioni di potenza media reali ottenuto per la normalizzazione dell'RMSE (ovvero per la computazione del CV(RMSE)) risulta inferiore rispetto a quello ottenuto nel modello di regressione lineare. Di conseguenza ciò fa sì che non siano ovviamente confrontabili tra di loro.

Ciò nonostante, il modello ottenuto risulta essere molto più potente rispetto alla regressione lineare, in quanto riesce a riconoscere i pattern di funzionamento di tipo stagionale (estate-inverno), settimanale (giorni lavorativi-weekend) e giornaliero (preaccensione-ora di picco). Questo permette di prevedere in modo plausibile, con un timestep di 1 ora, il consumo frigo futuro. È però più complesso rispetto alla regressione, e richiede anche più informazioni come input.

Per quanto riguarda le criticità riguardanti la disponibilità delle variabili di input:

- Le variabili temporali sono facili da reperire e anche da introdurre nel modello (non vi è un particolare pre-processing dei dati).
- Le variabili autoregressive anch'esse fanno parte dello storico del sistema, e quindi con un sistema di monitoraggio appropriato si potrebbero reperire facilmente anche queste.
- Le variabili climatiche esogene sono disponibili su differenti piattaforme meteo con accurattezze elevate: i valori storici sono registrati, mentre quelli a distanza di 1h hanno incertezze molto basse (e inoltre nel modello in esame è usata unicamente la T_ext come variabile esogena definibile come non storica, che risulta sicuramente più facile da reperire/stimere rispetto alla I_sol).

Questo modello quindi risulta avere un vantaggio prestazionale molto più elevato, nonostante effettivamente risulti più complesso di una semplice regressione lineare. Il vantaggio di questi modelli è la capacità di integrare variabili decisionali categoriche e di mantenere comunque, se l'albero non è troppo esteso, una certa interpretabilità logico-fisica. Da quali siano le prime "etichette" usate per gli split, si può capire l'importanza delle variabili introdotte. Inoltre la logica binaria, e la scelta delle condizioni risultano spesso comprensibili: se ho una temperatura più alta di una certa soglia e mi trovo in una giornata lavorativa, è probabile che il mio consumo sarà alto.

4.3 Random Forest

4.3.1 Funzionamento algoritmo Random Forest

Il modello Random Forest, come si può intendere dal nome, è rappresentabile proprio come una foresta di alberi decisionali. È frutto del concetto di "*Ensemble Learning*", in cui uno dei metodi chiave sfruttati è quello di usare un gruppo di modelli semplici/deboli in modo da creare un modello più potente dal loro agglomeramento. In questo caso quindi, la foresta casuale è proprio data dall'unione di tanti alberi di regressione.

Il vantaggio di avere tanti CART come riferimento, permette un processo di calibrazione degli iperparametri meno accurato e la possibilità di avere alberi più profondi, e quindi più precisi. D'altra parte, si ha una perdita del significato logico-fisico e quindi l'interpretabilità del modello.

Per ottenere alberi più profondi, senza rischiare di ricadere nel fenomeno dell'overfitting, è possibile allenare ogni albero non più con l'intero set di dati, ma solo con una parte casuale di questo: ogni albero quindi si allenerà con dati diversi, permettendo di mantenere una certa genericità nel modello. Questo ci permette di arrivare a profondità estremamente più elevate senza aver paura di essere troppo specifici, proprio per questa natura randomica nella fase di training degli alberi.

Diventa però cruciale la scelta di un nuovo iperparametro, ovvero la grandezza del campione di dati con cui ogni albero dovrà allenarsi: campioni piccoli rispetto all'intero dataset a disposizione permettono di evitare l'overfitting, a discapito però di una riduzione delle prestazioni.

Inoltre, come si è visto nel capitolo precedente, risulta evidente che vi siano variabili più incidenti rispetto ad altre e che quindi l'algoritmo CART porterebbe alla creazione di alberi correlati, simili tra di loro, e quindi ridondanti, nonostante la componente randomica nella scelta del dataset di training. Per ovviare a questo problema, nei modelli Random Forest, ad ogni nodo dell'i-esimo CART vengono considerate come "candidate" per lo split solo una parte di variabili decisionali/etichette, prese in modo casuale: questo permette di "forzare" la creazione di alberi differenti tra di loro e quindi ottenere performance complessive migliori.

Una volta definita la foresta, le predizioni finali di questa saranno la media degli output di ogni albero.

Come variabili di input, sulla base dell'esperienza col CART e di una sperimentazione sui dati, si sono scelte le stesse che erano state impostate nel modello Decision Tree, eccetto la I_sol (e derivate), in quanto non particolarmente influenti.

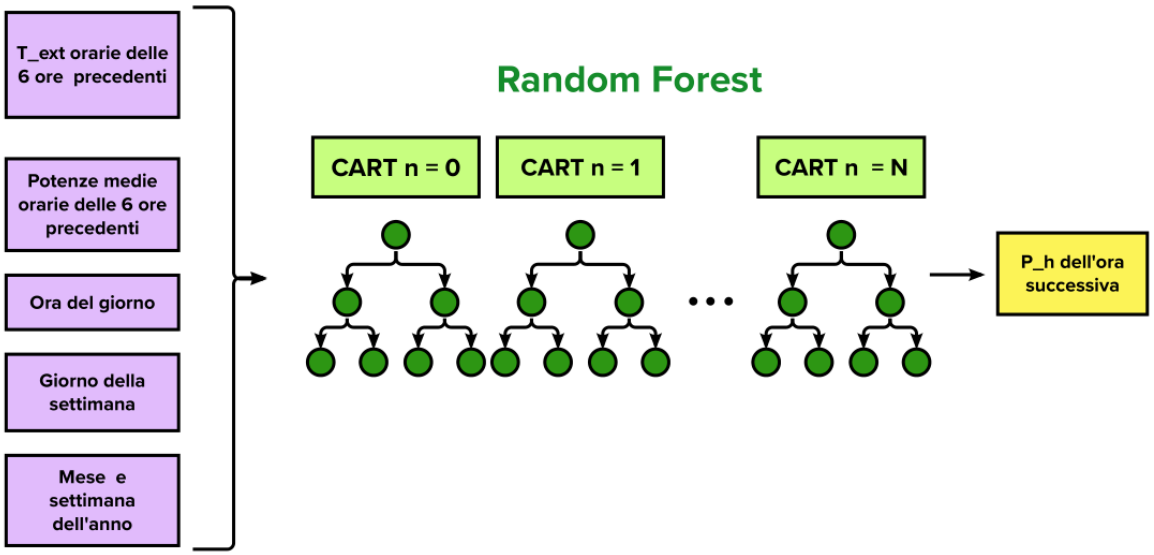

Figura 53 - Schema logico rappresentante le variabili di input al modello Random Forest

Per la definizione del modello, i nuovi iperparametri introdotti sono stati i seguenti [15]:

- Il **massimo di caratteristiche** (*max_features*), rappresenta il numero di variabili decisionali candidate (in modo randomico) ad ogni nodo per lo splitting. Come precedentemente accennato, questo permette di evitare la creazione di alberi correlati evitando la predominanza di variabili con capacità predittiva più alta.

- La **dimensione dei sotto-campioni** per il training (*bootstrap* e *max_samples*), permette di scegliere la dimensione dei dataset di training casuali su cui si alleneranno i singoli alberi.

4.3.2 Modello annuale – Random Forest

L'obiettivo del modello descritto nella seguente sezione è lo stesso di quello del CART: prevedere il consumo dell'ora successiva nell'intero funzionamento annuale dei gruppi frigo. I dataset di training e di testing sono stati lasciati invariati, in modo da avere un confronto più veritiero con il modello CART semplice.

Anche in questo caso si è andati ad attuare un processo di calibrazione degli iperparametri, in modo da ottenere la maggiore accuratezza possibile senza ricadere nel sovradattamento.

Dopo varie iterazioni, si è definita quella che risulta essere la miglior impostazione degli iperparametri:

- Il numero di stimatori (*n_estimators*), ovvero il numero totale di alberi, si è settato pari a 100 in quanto superando questa soglia non vi erano sensibili miglioramenti prestazionali.
- Una profondità massima degli alberi (*max_depth*), impostata anch'essa pari a 100. Questo valore così elevato permette di ottenere maggiore profondità, e quindi valori più precisi, senza overfitting grazie alla natura randomica dell'algoritmo.

- Una diminuzione minima di impurità (*min_impurity_decrease*), settata pari a 0.0005, con motivazione analoga al punto precedente.
- Un numero minimo di campioni per nodi foglia (*min_samples_leaf*), impostato pari a 5, in modo da comunque lasciare un limite alla specificità senza però creare *stopping* prematuri.
- La dimensione dei sotto-campioni (*max_samples*), impostata pari all'90% del dataset originario. Questo permette di mantenere una genericità nel modello, senza però nuocere alle performance del modello
- Il massimo di caratteristiche (*max_features*), si è settato pari a 11, ovvero il 50% delle variabili decisionali totali. Anche questa restrizione ha permesso di limitare sensibilmente il fenomeno di overfitting.

Come si può da subito notare, i singoli alberi presentano restrizioni di gran lunga meno severe del modello CART semplice. Questo, unito alle restrizioni imposte dagli iperparametri *max_samples* e *max_features*, permettono la creazione di un modello predittivo più efficace.

Si rappresentano di seguito cinque degli alberi facenti parte della Random Forest a titolo esemplificativo.

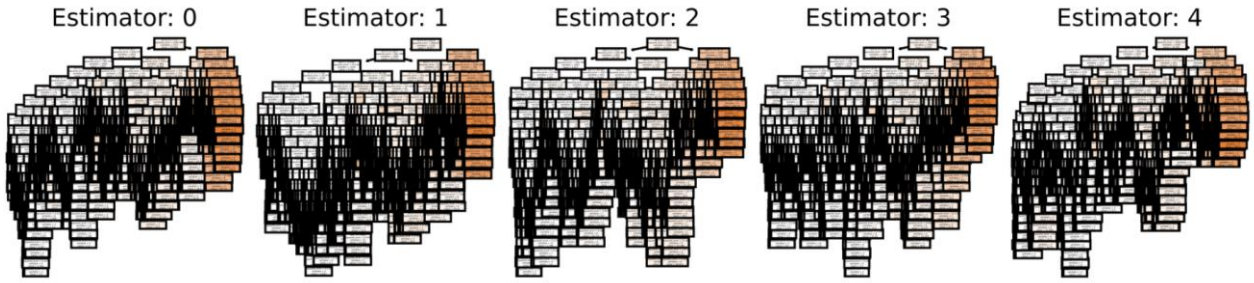

Figura 54 - Rappresentazione grafica di cinque generici CART costituenti il modello Random Forest

Di seguito si riportano in Tabella 9 le metriche prestazionali rappresentanti le performance del modello.

Tabella 9 - Metriche prestazionali del modello annuale Random Forest

Metrica	Training	Testing
R2 [-]	0.97	0.94
MAE [kWh/h]	4.26	6.98
RMSE [kWh/h]	10.42	15.65
CV(RMSE) [%]	32.40	36.31

Come si può vedere dalle differenze prestazionali tra training e testing, il modello risulta leggermente overfittato, il che rispecchia comunque una delle caratteristiche e maggiori criticità dell'algoritmo Random Forest. La scelta degli iperparametri però è stata fatta in modo da

minimizzare comunque questa differenza tra le metriche di training e testing, senza però andare a diminuire le performance sui dati di quest'ultimo.

Nonostante questo tipo di modellazione risulti più pesante a livello computazionale (si è passati da 32 a 5998 nodi foglia totali), e l'interpretabilità logico-fisica venga un po' a meno, sono evidenti i miglioramenti prestazionali.

È riportata di seguito, come fatto per il modello precedente, un confronto tra profilo reale e profilo previsto di due settimane generiche del periodo di testing.

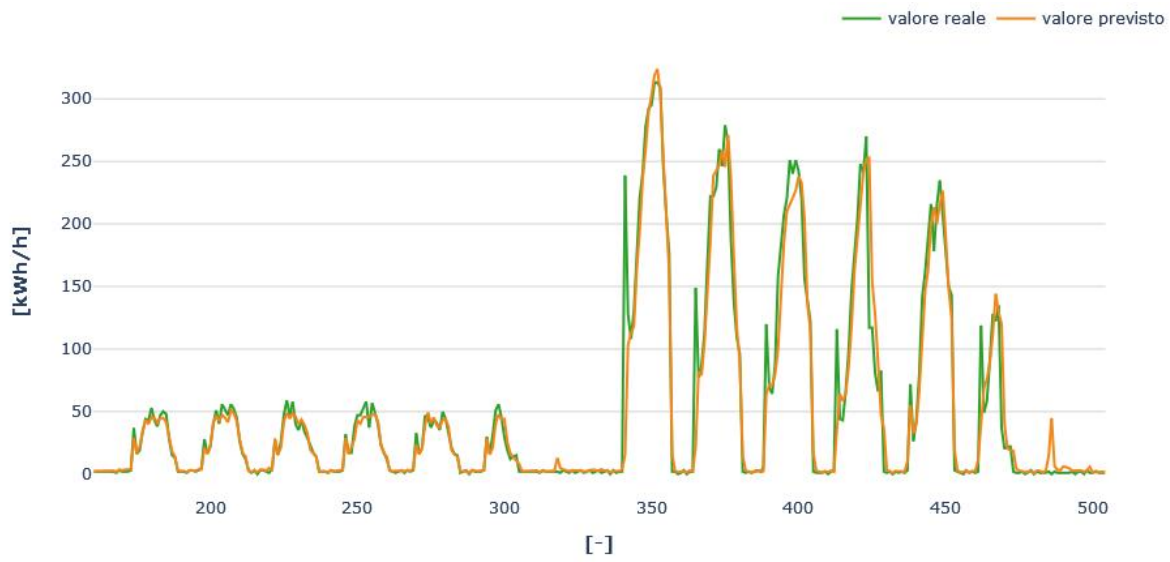

Figura 55 - Confronto tra profilo reale e predetto dal modello Random Forest della 3° e 4° settimana di settembre 2021

In Figura 55 si sono rappresentate le stesse settimane analizzate col modello CART (Figura 52), per permettere un confronto critico tra i due modelli: a differenza del CART, in questo caso i profili risultano meno discretizzati e più continui, seguendo in modo più accurato il reale andamento della potenza media.

4.3.3 Modelli dei consumi di UTA e macchinari – Random Forest

Data la validità e potenza di questo algoritmo, si è poi provato inoltre a sfruttarlo per il discernimento di carico frigorigeno legato alle UTA e carico frigorigeno legato ai macchinari.

Le ipotesi alla base di questo procedimento, molto forti, sono:

- Nel periodo di funzionamento invernale, si può assumere che il carico frigorigeno sia unicamente legato ai macchinari (e non alle UTA).

- Non vi è variazione nella logica di funzionamento dei macchinari (e quindi una sorta di stagionalità) tra il periodo di funzionamento estivo e quello invernale.

La scaletta finale è descrivibile nei seguenti passi:

1) Creare un modello che simuli e predica il comportamento dei macchinari: per farlo ci si può basare sui dati di funzionamento invernale
2) Proiettare questo modello sul periodo estivo, in modo da avere quindi un sotto profilo che descriva il consumo frigorigeno allocabile unicamente ai macchinari anche quando le UTA sono in modalità di raffrescamento.
3) Dalla differenza tra il profilo reale di consumo frigorigeno totale e quello previsto dei macchinari si estrae il profilo di consumo allocabile alle UTA.
4) Sfruttando i dati disgiunti relativi ai consumi frigorigeni delle UTA, è possibile creare un altro modello che quindi permetta di prevedere unicamente il consumo attribuito alle UTA.

Seguendo questo procedimento, il risultato è la possibilità di prevedere i consumi frigorigeni allocabili a macchinari e UTA in modo distinto, e quindi sulla base di questo monitorare in modo più consapevole i propri consumi, valutando l'efficacia dell'implementazione di nuove strategie gestionali, la presenza di anomalie, il risparmio dovuto a nuovi macchinari e tante altre opportunità.

La criticità maggiore di questa fase è la scelta delle variabili di input da dare al modello. Come si è visto nei modelli precedenti le variabili autoregressive risultano molto incidenti. Se però l'obiettivo è quello di proiettare il modello di funzionamento dei macchinari sul periodo estivo, se si usassero delle variabili di tipo autoregressivo il modello sarebbe influenzato da dei consumi precedenti che non rispecchiano il consumo isolato dei macchinari, ma quello aggregato di UTA e macchinari.

Di conseguenza le variabili di input al modello devono essere puramente di tipo esogeno. Le variabili esogene di tipo temporale risulteranno quindi le più impattanti, dato che permettono di definire in modo accurato i pattern di consumo stagionali. Al contrario, le variabili esogene di natura climatica non dovrebbero influire nel modello se si pensa ad esempio alla bassa correlazione tra potenza media e temperatura nel periodo invernale (dimostrato nel capitolo sui modelli di regressione lineare). Ciò non toglie che il modello possa definire una serie di condizioni di carattere olistico, che permettano di cogliere dinamicità non ovvie e non strettamente legate al significato fisico. Il modello potrebbe infatti cogliere relazioni tra le condizioni climatiche e l'EER dei gruppi frigo che a livello macroscopico risultano impercettibili.

Per l'intero percorso, ai modelli sono state tenute "nascoste" due settimane di dati (l'ultima di giugno e una di fine agosto) caratterizzate da diversi livelli di consumo, che saranno usate come testing finale per la valutazione delle performance complessive di questi.

Per una migliore comprensione dei passaggi logici, si mette a disposizione uno schema del procedimento attuato in Figura 56.

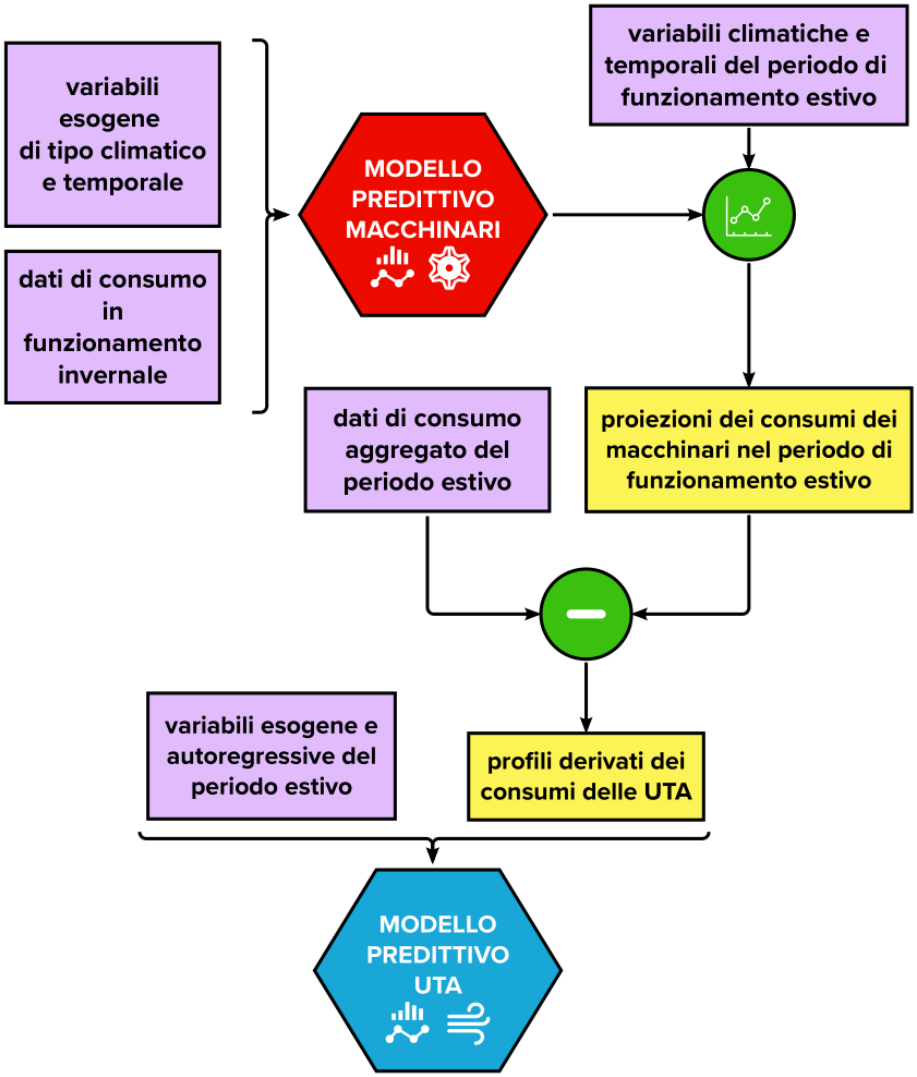

Figura 56 - Procedimento di creazione di un modello predittivo per le UTA

Attraverso questo procedimento è quindi possibile avere una stima previsionale dei consumi delle UTA. Di seguito si riporta invece lo schema rappresentante la natura delle variabili di input al modello predittivo dei consumi dei macchinari.

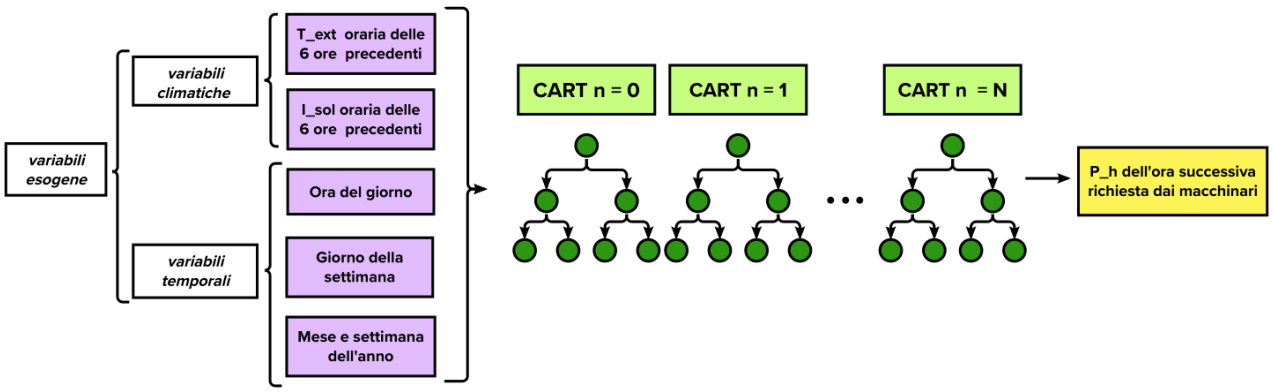

Figura 57 - Schema logico delle variabili di input al modello di previsione dei consumi frigo dei macchinari

Come precedentemente affermato, non ci si aspetta che le variabili climatiche spieghino in modo eccellente i consumi frigo, ma nonostante questo forte limite, il modello riesce a ottenere delle metriche prestazionali accettabili (malgrado non si riesca ad evitare completamente il problema di sovradattamento) come si può vedere in Tabella 10.

Tabella 10 - Metriche prestazionali modello predittivo consumi frigorigeni macchinari

Metrica	Training	Testing
R2 [-]	0.92	0.89
MAE [kWh/h]	2.79	3.62
RMSE [kWh/h]	4.86	6.08
CV(RMSE) [%]	27.02	32.48

Anche in questo caso vi è stato un processo di calibrazione degli iperparametri per limitare l'overfitting e mantenere un buon livello di accuratezza.

Dopo aver creato il modello, si è andati a proiettare le variabili esogene estive sul modello appena generato, ottenendo quindi una stima dei profili di consumo dei macchinari nel periodo estivo.

Si riportano graficamente i consumi dei macchinari stimati per il periodo estivo, a confronto con i consumi totali. La differenza, riportata con colorazione blu, rappresenta proprio quella che si può assumere sia la quota relativa alle UTA.

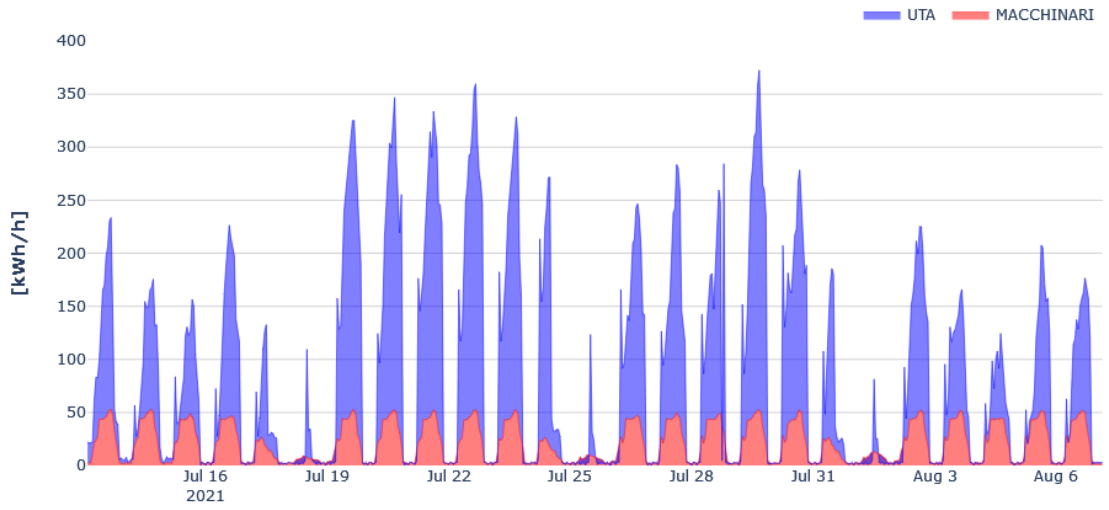

Figura 58 - Cumulata dei consumi frigorigeni allocabili a macchinari e UTA di un generico periodo di funzionamento estivo

Come si può vedere, la domenica il consumo stimato non è accurato, e presenta sempre lo stesso pattern di errore. Questo potrebbe significare diverse cose, come ad esempio un cambio nella

gestione delle lavorazioni o una variazione nella schedulazione aziendale. Tenendo conto comunque di questa variazione, il modello di previsione del carico frigorigeno associato alle UTA non considererà all'interno del proprio dataset di training le domeniche.

Si riporta di seguito la stima della quota assegnabile alle UTA.

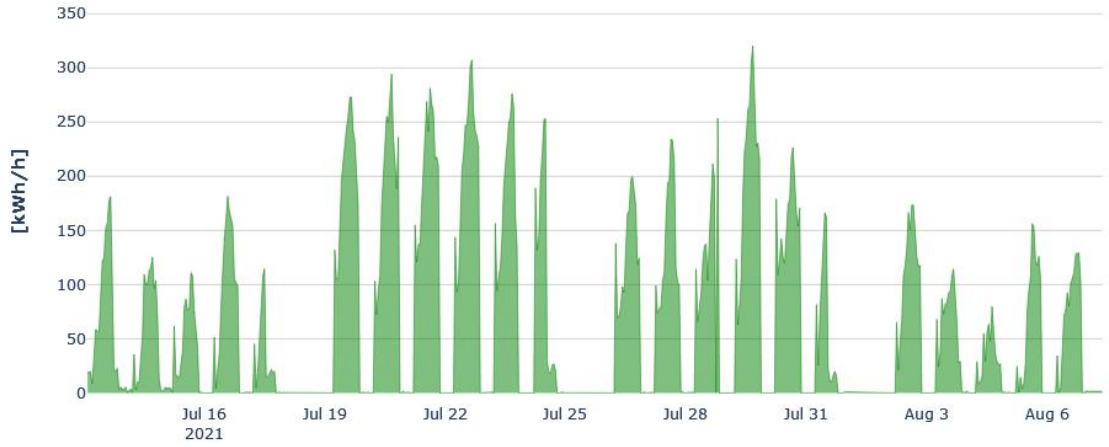

Figura 59 - Stima della quota di consumo allocabile alle UTA in un generico periodo del funzionamento estivo

Grazie alla disponibilità di queste stime è possibile quindi creare un nuovo modello che vada a simulare/prevedere l'andamento dei consumi frigorigeni relativi alle UTA.

Il modello in esame può far uso di tutte le variabili di input precedentemente sfruttate per il modello Random Forest di predizione oraria creato per l'anno 2021 (Figura 53).

Effettuando sempre il tuning degli iperparametri, si è riusciti ad ottenere un modello valido, le cui metriche prestazionali sono riportate di seguito in Tabella 11.

Tabella 11 - Metriche prestazionali previsione consumi frigorigeni UTA

Metrica	Training	Testing
R2 [-]	0.93	0.94
MAE [kWh/h]	10.58	11.21
RMSE [kWh/h]	20.78	18.91
CV(RMSE) [%]	31.56	28.99

Il modello però ottenuto è frutto di ipotesi ed assunzioni molto forti. Un limite forte della metodologia usata è proprio l'impossibilità di valutare direttamente se le assunzioni fatte sono considerabili come attendibili.

Per verificare la validità complessiva dei due modelli, si sono sfruttate le due settimane "nascoste", non considerate nel procedimento di creazioni dei modelli (quello predittivo dei consumi UTA e quello predittivo dei consumi dei macchinari) proprio a questo scopo.

Proiettando le variabili climatiche, temporali, e autoregressive (solo per il modello UTA quest'ultime ovviamente) di questo periodo sui due modelli, è stato possibile definire delle stime dei singoli profili di carico di UTA, e dei macchinari. Per eseguire un check dell'efficacia predittiva, si sono poi aggregati questi profili, ottenendo una stima del profilo di consumo totale dei gruppi frigo. Confrontando questo profilo fittizio, frutto della somma delle due stime, con il profilo di carico reale del periodo in esame, è stato possibile computare le metriche prestazionali e valutare le performance complessive dei modelli.

Si rappresenta di seguito in Figura 60 il procedimento di testing attuato.

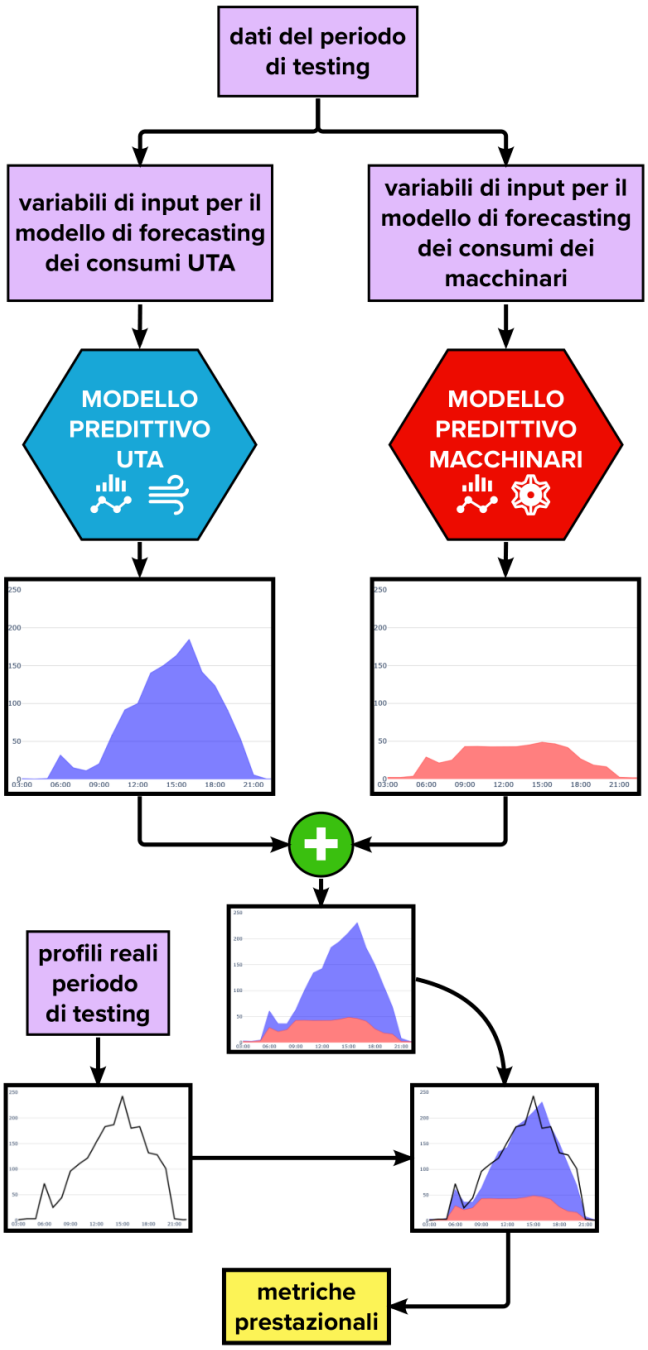

Figura 60 - Procedimento di testing complessivo dei due modelli

Le metriche prestazionali ottenute sono quindi riportate in Tabella 12.

Tabella 12 - Metriche prestazionali complessive dei due modelli, Random Forest

Metrica	Testing
R2 [-]	0.89
MAE [kWh/h]	17.73
RMSE [kWh/h]	28.33
CV(RMSE) [%]	27.20

Come si può vedere, i valori risultano completamente soddisfacenti:

- Il valore di R^2 è più che prossimo a 1.
- Il valore dell'errore assoluto medio è relativamente basso se si considera che la media dei valori di consumo nell'orario di funzionamento (tra le 4:00 AM e le 22:00) è di circa 100 kWh/h.
- Stesso discorso per la radice dell'errore quadratico medio, che come si sa tende però a dare più peso a discostamenti più alti.
- Il CV(RMSE) risulta non troppo distante dal valore 25%, al di sotto del quale si può definire come efficace un modello.

Nella seguente figura, si riporta un confronto tra i consumi reali e i consumi

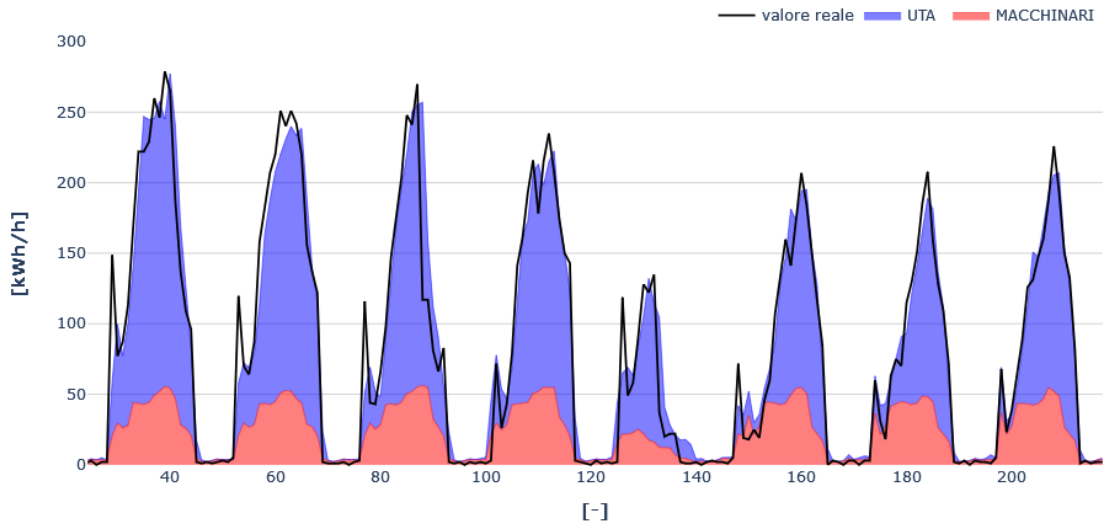

Figura 61 - Consumo frigorigeno reale (linea scura) e l'aggregato dei consumi (aree colorate), Random Forest

Un ulteriore check che si può attuare per valutare se il modello generato risulta veritiero e valido, è confrontare le schedulazioni di funzionamento dei profili fittizi ottenuti (riguardanti il carico frigo delle UTA) con i profili di consumo dei ventilatori di esse. Questa seconda entità è stata computata sommando elemento per elemento i relativi consumi orari dei ventilatori delle sette UTA, ottenendo quindi un profilo unico.

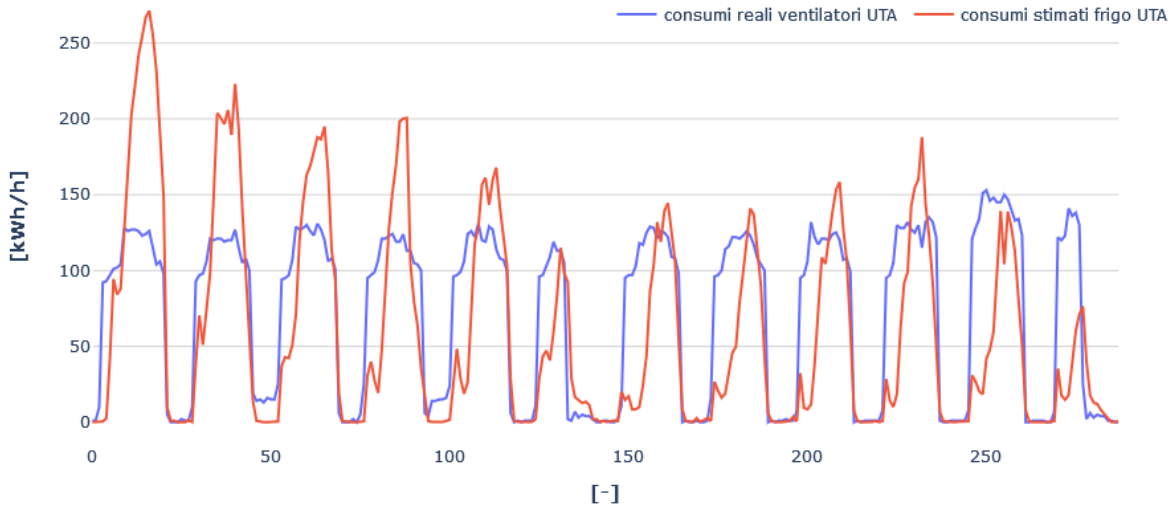

Figura 62 - Profili di consumo allocabili alle UTA per gruppi frigo e ventilatori

Come si può vedere dalla rappresentazione grafica (Figura 62), eccetto qualche deviazione occasionale, i profili risultano praticamente in fase per quanto riguarda gli orari di accensione/spegnimento. Ciò può essere visto come una ulteriore conferma delle assunzioni e ipotesi alla base della creazione dei modelli.

4.3.4 Ricapitolazione modelli – Random Forest

I modelli generati sono riassunti qui di seguito.

Tabella 13 - Riassunto modelli Random Forest

MODELLO	output	input	timestep	periodo validità
MODELLO PREDITTIVO GRUPPI FRIGO	Potenza media oraria dei gruppi frigo	Variabili di tipo climatico, temporale e autoregressivo	1 h	Intero periodo annuale
MODELLO PREDITTIVO MACCHINARI	Potenza media oraria dei gruppi frigo allocabile ai macchinari	Variabili esogene di tipo climatico, temporale	1 h	Periodo di funzionamento estivo
MODELLO PREDITTIVO UTA	Potenza media oraria dei gruppi frigo allocabile alle UTA	Variabili di tipo climatico, temporale e autoregressivo	1 h	Periodo di funzionamento estivo

Questi tre modelli permettono quindi di prevedere e descrivere i consumi orari dei gruppi frigo e delle domande che li caratterizzano. In questo modo si possono attuare nuove metodologie di ottimizzazione anche in funzione delle diverse domande di energia frigorigena.

È possibile, inoltre, valutare l'impatto di deviazioni dal funzionamento atteso, attraverso un benchmarking interno con due obiettivi:

- Verifica dell'efficacia di un intervento di retrofit o altre strategie/azioni di miglioramento del sistema.
- Verifica di errori di funzionamento nel sistema se si riscontrano profili anomali rispetto a quelli predetti dal modello.

Nel caso di attuazione di nuove strategie di gestione, singolarmente sulle UTA o sui macchinari, è possibile verificarne l'efficacia. Ovviamente per una corretta analisi e accertamento del loro impatto, i miglioramenti di gestione e della componentistica del sistema andrebbero fatti in periodi separati: se si attuassero delle modifiche contemporaneamente al sistema UTA e al sistema macchinari, non si riuscirebbe a capire come assegnare le quote di risparmio.

4.4 Reti Neurali Artificiali

4.4.1 Funzionamento algoritmo ANN

Un altro modello adatto al forecasting sono le reti neurali artificiali (*Artificial Neural Network*, in breve ANN). Sono considerate come un modello black box: i parametri interni al modello non risultano effettivamente interpretabili. Questo tipo di modello, quindi, ha come grande difetto la completa perdita dell'interpretabilità fisica, ma ha come punto di forza proprio quello di cogliere pattern e caratteristiche dei dati non ovvie e non comprensibili all'occhio umano.

In questa sezione si va a descrivere in breve il funzionamento delle reti neurali artificiali. I tipi di rete implementati al giorno d'oggi sono tanti. Di seguito, si va a descrivere la tipologia più semplice, che è anche stata la prima a essere implementata: la *Feedforward Neural Network* (FNN). Le FNN non presentano cicli interni tra i nodi della rete e hanno un'unica direzione di propagazione delle informazioni.

I nodi della rete rappresentano i neuroni, che comunicano tra di loro attraverso un passaggio di informazioni. La struttura delle ANN può essere semplificata nei seguenti elementi:

- ***Input layer***, rappresenta il primo strato della rete, in cui si hanno tanti neuroni quante sono le variabili di input
- ***Hidden layer***, rappresenta uno strato intermedio tra input e output. Ve ne possono essere più di uno o anche nessuno. Nelle reti classiche, questo tipo di layer è di tipo "*dense*", ovvero un singolo neurone di questo strato è interconnesso a tutti i neuroni dello strato precedente e di quello successivo.
- ***Output layer***, rappresenta lo strato di output, in cui il numero di neuroni corrisponde proprio quello delle variabili di output desiderate.

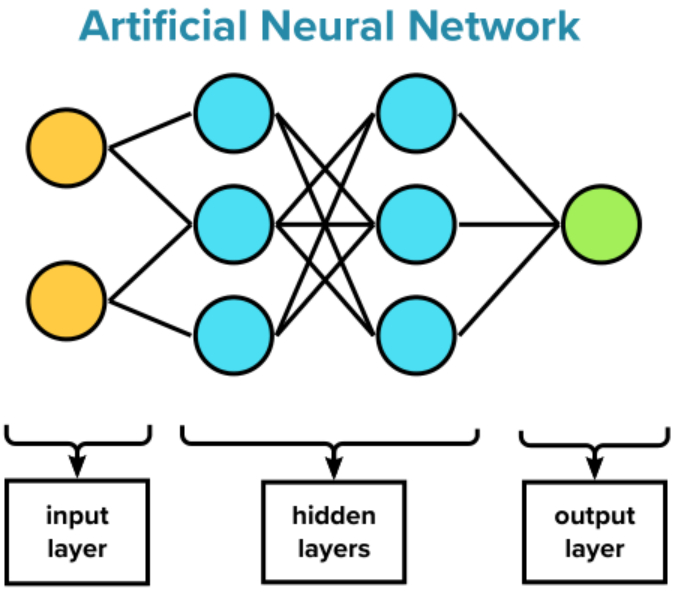

Figura 63 - Rappresentazione layers di una rete neurale

Per spiegare in modo efficace e intuitivo l'interazione di un neurone, è utile sfruttare l'analogia con le regressioni lineari multiple [16], di cui si riporta la generica formula qua di seguito.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_n x_{in}$$

Come sappiamo, l'output è rappresentato dalla y_i, le x_{in} rappresentano le n-variabili indipendenti di input, il β_0 rappresenta il termine costante, mentre gli altri β_n sono gli n-coefficienti (pendenze) associati ad ogni variabile indipendente.

Analogamente, per i neuroni si ha la seguente struttura.

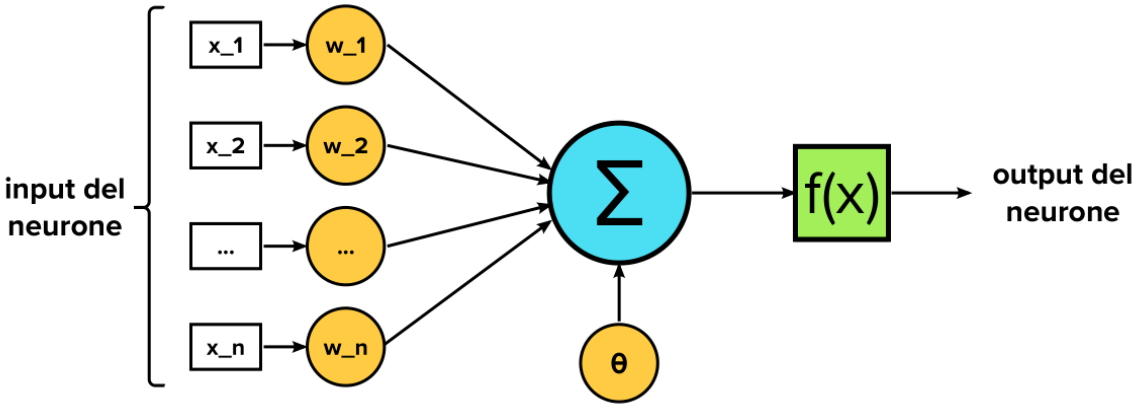

Figura 64 - Input e Output di un generico neurone artificiale

In questo caso l'elemento θ è analogo al termine costante, e i coefficienti β_i sono interpretati come pesi w_i. Viene attuata una somma di tutti gli input moltiplicati per i propri pesi e il valore di "*bias*" (ovvero θ). L'output di questa somma passerà per una funzione di attivazione f(x) che restituirà l'output effettivo del neurone. La funzione di attivazione f(x) permette la trasmissione di informazioni o meno, sulla base del superamento di un certo valore soglia (un esempio è la funzione ReLU). Anche questa può essere selezionata in funzione del tipo di problema.

Nel caso di layer di tipo dense, quindi, ogni neurone riceve come input un valore x_i per ogni neurone presente nel layer precedente.

Il processo di apprendimento delle reti neurali sta proprio nel procedimento di stima di questi pesi e bias, per minimizzare la differenza tra il valore predetto e quello reale. Per valutare le prestazioni di questo, è definita una funzione costo (*cost function*) che può essere di diversa tipologia in base al problema. L'obiettivo primario è quindi minimizzare questa funzione. Ogni volta che viene fatta "vedere" l'interezza del dataset al modello, si può dire che è passata un'epoca, e quindi si può calcolare la somma di tutti gli output della funzione costo. Attraverso il metodo del "gradiente discendente", "gradiente discendente stocastico" o altri metodi, si stima in che direzione si dovranno "spostare" complessivamente i pesi w_i, per l'iterazione/epoca successiva per avvicinarsi al minimo della funzione costo. Al fine di capire e associare la responsabilità del valore della funzione costo ad ogni peso, si attua un processo di *backward propagation*, in cui, attraverso un articolato algoritmo matematico, avviene una ridefinizione dei pesi dei rami di input per minimizzare l'errore, partendo dall'output layer verso quello di input. Dopo questo procedimento si hanno tutti i pesi aggiornati pronti per una nuova iterazione.

Ovviamente nella prima epoca i pesi sono computati in modo randomico (saranno tutti valori simili tra di loro e diversi dal valore nullo).

Gli iperparametri più comuni in letteratura [17] sono riportati di seguito, e possono essere calibrati con diverse strategie in base alla complessità del problema:

- Numero di hidden layers: per problemi più complessi, si può andare ad esplorare le performance del modello aggiungendo layer e quindi aumentando la "profondità" della rete (*depth*).
- Numero di neuroni associato ad ognuno di questi strati: anche in questo caso, in base alla complessità e alle caratteristiche del problema, possono servire strati più o meno "ampi" (*breadth*). Come si può capire dalla complessità del modello, le prestazioni cambiano anche in base all'ordine degli hidden layers.
- Numero di epoche: il numero di epoche, e quindi di reiterazioni e aggiustamento dei pesi, determina la specificità ed efficacia del modello (anche se non è propriamente una caratteristica della rete, ma del suo allenamento). Un numero troppo elevato di epoche può però portare al fenomeno di overfitting, e quindi poca genericità del modello se sfruttato per dati esterni dal dataset di training.

Per "combattere" il problema di overfitting, si può anche introdurre il concetto di *dropout*: questo consiste nella "potatura" randomica di una certa parte di rami della rete tra due strati, generando quindi nuove architetture fittizie (tutte simili a quella originale) ad ogni epoca, proprio per stimolare il modello a rimanere sufficientemente generico.

Le FNN in particolare non hanno la capacità di mantenere memoria degli input passati. Questo tipo di qualità può però essere implementato introducendo in modo "forzato" delle variabili autoregressive associate alle ore precedenti.

4.4.2 Preparazione dati

Le ANN sono uno strumento molto forte, che però necessita una attenta pre-elaborazione dei dati di input. Infatti, oltre ovviamente alla outliers detection ed eliminazione di questi, è necessario riformulare i dati in modo che il modello non interpreti in modo errato le variabili di input. Procedimenti fondamentali sono:

- La normalizzazione per le variabili di tipo regressivo.
- L'encoding delle variabili categoriche.

La normalizzazione dei dati permette che le variabili regressive di input, vengano considerate e interpretate in modo imparziale dal modello. Infatti, le variabili possono essere di diversa natura e quindi avere scale differenti (una temperatura esterna di 35°C e una potenza media oraria di 35 kWh/h hanno stesso valore assoluto ma descrivono anche rispettivamente diversi livelli di intensità del proprio parametro), ed è quindi necessario normalizzarle per far sì che non si dia maggior peso a una certa variabile unicamente perché il suo valore assoluto è maggiore, dato che le reti neurali non hanno interpretabilità fisica.

Per quanto riguarda invece l'encoding, questo è necessario per permettere al modello di considerare variabili categoriche nella sua computazione. Un esempio si può fare con i giorni della settimana: se si assegnassero dei numeri dall'1 al 7 per distinguerli, si rischierebbe che la domenica (rappresentata dal numero 7) potrebbe essere interpretata come un dato di maggiore intensità e quindi importanza, quando ciò non è assolutamente vero. Si sono quindi introdotti diversi tipi di codifica per evitare questo tipo di problemi.

Attraverso questi due tipi di procedimento, tutte le variabili vengono tradotte in numeri compresi tra 0 e 1, garantendo una migliore interpretabilità alle reti neurali artificiali.

4.4.2.1 Normalizzazione

Prima di attuare una qualunque normalizzazione, è necessario considerare il significato fisico della grandezza in esame. Le normalizzazioni più comuni in letteratura per grandezze fisiche sono:

- Massimo, si attua semplicemente dividendo per il valore massimo della rispettiva variabile l'intero insieme di dati della variabile stessa.

$$z_i = \frac{x_i}{\max(x)}$$

  Ha come criticità la gestione dei numeri negativi.

- Massimo-minimo, è simile ma permette di riportare in scala in modo efficace anche in presenza di numeri negativi.

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

- Z-score, fa sì che i dati siano normalizzati in modo che la media di questi sia 0 e che la loro deviazione standard sia 1.

$$z_i = \frac{x_i - \bar{x}}{stdev(x)}$$

Nel caso in esame le variabili regressive sono principalmente di due tipi:

- Energia elettrica associata ai gruppi frigo, che può essere semplicemente scalata con una normalizzazione tramite massimo.
- Temperatura esterna, in cui il meno non rappresenta un vero e proprio significato fisico, e quindi il metodo più adatto risulta il metodo massimo-minimo.

4.4.2.2 Encoding

In questo caso si sono attuate diversi tipi di encoding.

- ***One-hot encoding***, crea una nuova variabile (e quindi una nuova colonna di input) per ogni elemento categorico. Per ogni riga (e quindi dato di input) i valori relativi agli elementi categorici saranno 0 eccetto per quello che rappresenta la categoria di effettiva appartenenza che sarà pari a 1 ("hot"). Si riporta di seguito un esempio veloce con i giorni della settimana.

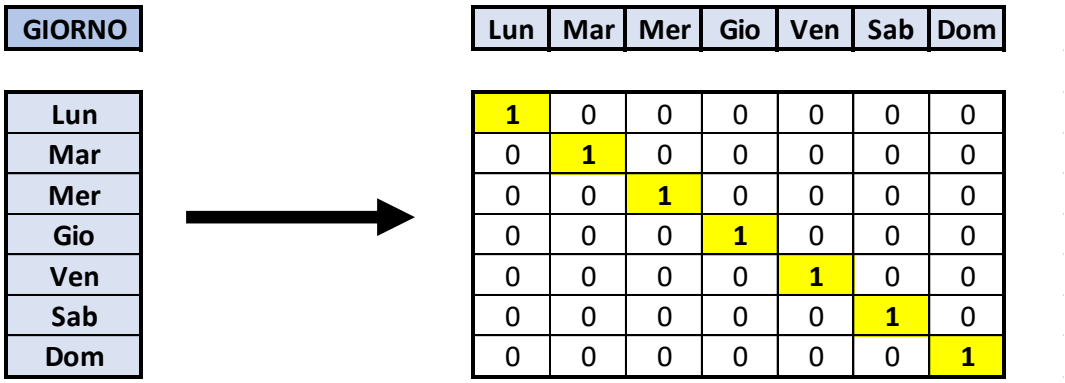

Figura 65 - Esempio di One-hot encoding con la categoria "giorno della settimana"

- ***Binary encoding***, è molto simile al precedente, ma sfrutta la codifica binaria per la conversione in variabili categoriche. Il modello infatti riesce a cogliere questo tipo di informazione senza grandi problemi. Il vantaggio rispetto al caso precedente è la riduzione del numero di colonne e quindi di variabili di input da dare al modello. Anche in questo caso si fornisce un esempio analogo per una migliore comprensione.

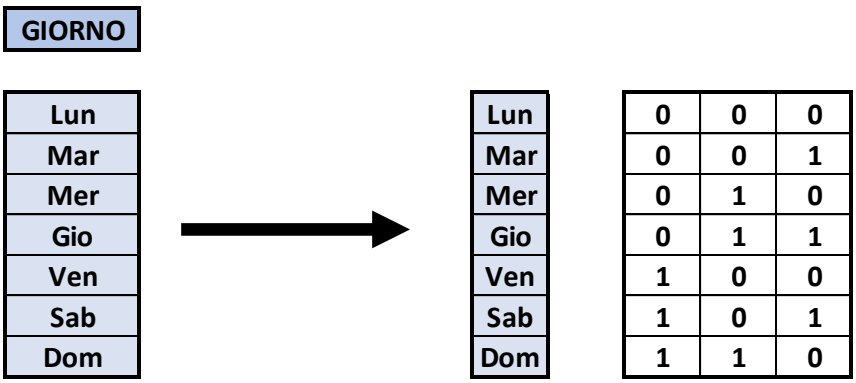

Figura 66 - Esempio di Binary encoding con la categoria "giorno della settimana"

Questo tipo di codifica permette di ridurre le variabili di input e quindi alleggerire il peso computazionale, migliorando le prestazioni del modello stesso.
Semplicemente sfruttando il Binary encoding invece del One-hot, considerando la variabile "Ora del giorno", si passa da 24 colonne a 5 colonne ($2^5 = 32 > 24$).

Si è sperimentata l'efficacia dei due metodi di codifica sulle diverse variabili categoriche nella creazione del modello.

4.4.2.3 Preparazione dati del caso studio

Come detto precedentemente, di grande importanza è la preparazione dei dati di input. La tipologia dei dati è sempre la stessa:

- Informazioni climatiche esogene, in questo caso solo la temperatura esterna.
- Informazioni autoregressive, ovvero i valori delle ore passate di consumo.
- Informazioni temporali categoriche, ovvero il giorno della settimana, l'ora del giorno e il mese dell'anno.

Per quanto riguarda la temperatura, si è usata una normalizzazione Min-Max usando i dati del periodo 2021-2022. Si sono considerati questi due anni perché, nel caso si volessero proiettare modelli di un anno sull'altro o viceversa, è opportuno avere i dati tutti nel range 0-1, dato che i modelli ANN tendono ad avere difficoltà in caso di estrapolazione.

Analogamente, per l'energia elettrica associata ai gruppi frigo, si è normalizzato sul massimo dei due anni solari 2021-2022.

Relativamente alle variabili temporali, si sono sfruttati i due tipi di codifica in funzione del problema. Da evidenziare è la differenza a livello di numero di variabili che si introducono nei due casi: si passa da 43 variabili con l'One-hot encoding (24 ore, 7 giorni della settimana, 12 mesi) a 12 con il Binary encoding (5 per descrivere le ore, 3 per i giorni della settimana e 4 per i mesi).

4.4.3 Modello annuale – ANN

Il primo modello creato con le reti neurali è stato un modello che, che con un singolo valore di output, sulla base delle variabili di input, andasse a stimare il valore di potenza media oraria associata ai gruppi frigo dell'ora successiva (come il modello Random Forest annuale). Per questo tipo di modello si sono sfruttate variabili autoregressive, variabili climatiche esogene a diversi time-step, e variabili di tipo temporale. Per ottenere ottimi risultati, nei modelli black-box, è molto importante offrire come input variabili significative, che non siano mal interpretabili dall'algoritmo. Dare variabili totalmente scorrelate come input non farebbe che impegnare del potenziale computazionale in modo inutile.

Di conseguenza anche la scelta degli n-shift temporali delle variabili autoregressive e climatiche risulta importante: se sono le 12:00, sapere quale fosse il consumo delle 4:00 AM del mattino stesso risulta meno utile, rispetto invece a sapere i consumi delle ultime sei ore.

Di conseguenza si è andati a usare le seguenti leve a disposizione per ottenere le migliori performance nella fase di testing:

- Variazione degli input relativi ai valori passati di consumo dei gruppi frigo, in un range tra le 8 e le 3 ore precedenti a quello da prevedere.
- Variazione dei valori passati di temperatura media oraria esterna.
- Variazione analoga al punto precedente, ma considerando le temperature future, ipotizzando di avere a disposizione un tool o un servizio efficiente di forecasting per le variabili climatiche.
- Variazione delle variabili temporali di input, come il mese dell'anno, il giorno della settimana e l'ora del giorno, variando anche il tipo di codifica usato.

Insieme a questo procedimento, composto da varie iterazioni, si è attuata la calibrazione degli iperparametri nel seguente modo:

- Variando la profondità (*depth*, si attua aggiungendo o togliendo strati agli *hidden layers*) e l'ampiezza (*breadth*, si attua aumentando o diminuendo il numero di neuroni per ogni strato facente parte degli *hidden layers*) della rete. Ovviamente, se le prestazioni ottenute dal modello sono basse sia per il training che per il testing, è necessario aumentare la complessità della rete aumentandone i collegamenti.
- Aumentando o diminuendo le epoche, in quanto, come spiegato precedentemente, ad ogni epoca si ha un aggiornamento dei pesi e coefficienti che compongono la rete, in modo da minimizzare la funzione costo. Questo tipo di minimizzazione è però ovviamente fatta sui dati di training: ciò significa che un numero troppo elevato di epoche, porterebbe sì a prestazioni sempre più buone sui dati di allenamento, a discapito però delle capacità di generalizzazione del modello. Ciò comporterebbe quindi a prestazioni eccellenti sui dati di training e pessime su quello di testing, rendendo il modello inutile per applicazioni di forecasting. Allo stesso tempo, troppe poche epoche non permetterebbero la creazione di un modello sufficientemente specifico e accurato. È necessario, come avviene anche per il settaggio degli altri iperparametri, un compromesso tra generalità e accuratezza.
- Inserendo e variando il valore dei *dropout*, e quindi inserendo la possibilità di perdere una percentuale dei dati (la percentuale dei dati è fissata ma i "rami" eliminati sono scelti in modo randomico) nel passaggio di informazioni tra uno strato ed un altro. Questo tipo di calibrazione si attua quando è necessario aumentare la complessità di una rete per aumentare l'accuratezza del modello, con come effetto collaterale l'overfitting. Per limitare questo fenomeno indesiderato, si attua questa perdita parziale forzata di dati ad ogni iterazione in modo che la rete non si sovradatti sui dati di allenamento. È come se ad ogni epoca si allenasse una rete neurale leggermente diversa, permettendo una generalizzazione del modello (concettualmente molto simile al passaggio logico che c'è tra CART e Random Forest).

Per migliorare la comprensione del sistema in esame, si è deciso di usare il tipo di etichettatura categorica dei dati che si era presentato nelle Figura 31 e Figura 32 (lunedì, martedì-venerdì, sabato, domenica). Per renderle variabili leggibili al modello si è usata sia la codifica One-hot encoding che il Binary encoding, in modo da poter poi scegliere la migliore tra le due opzioni.

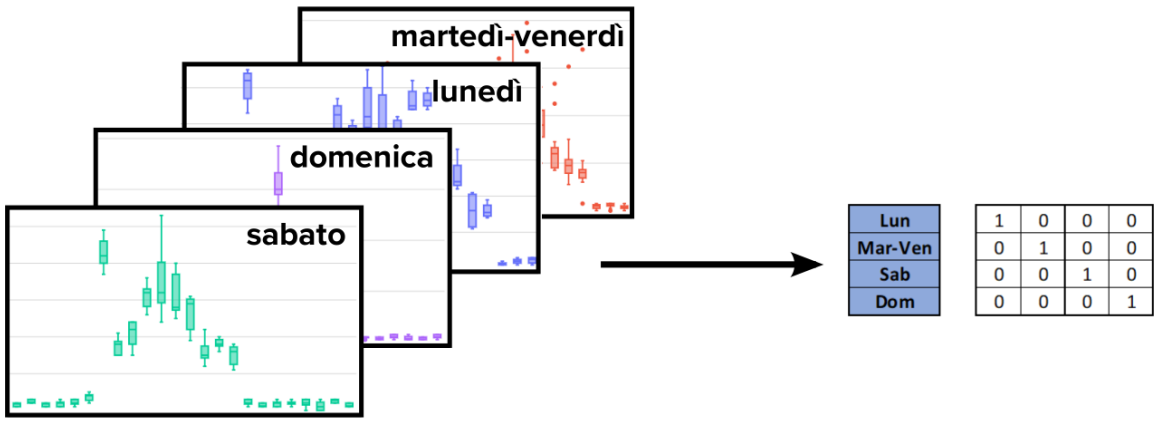

Figura 67 - Encoding basato sui profili tipici di consumo

Inoltre, per passare l'informazione del tipo di funzionamento (invernale o estivo) e quindi se in quel periodo le UTA siano accese o meno, si è creata una nuova variabile di input denominata "tipo funzionamento" che presenta il valore 1 quando le unità di trattamento aria risultano accese (tra maggio e settembre) e un valore nullo per il resto del tempo.

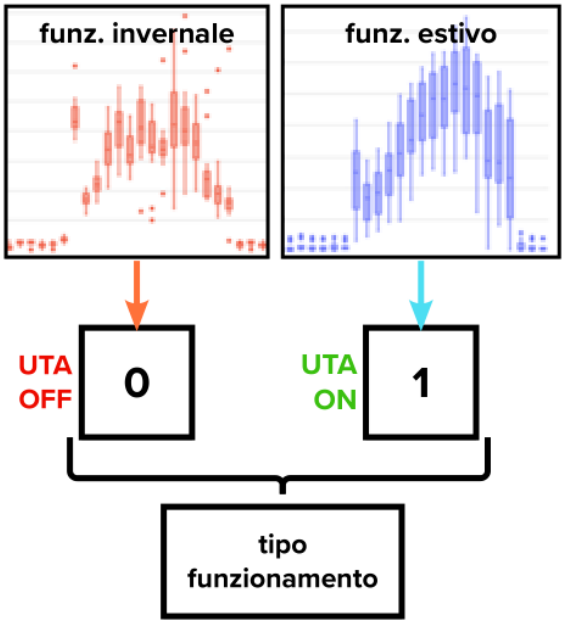

Figura 68 - Encoding relativo al tipo di funzionamento, estivo vs invernale

Queste nuove informazioni legate ai tipi di funzionamento e schedulazione forniscono informazioni da subito più chiare al modello per il riconoscimento di pattern.

Dopo varie iterazioni, con calibrazione degli iperparametri e variazione delle possibili combinazioni delle variabili di input, si riporta di seguito lo schema logico definitivo riguardante le variabili di input al modello.

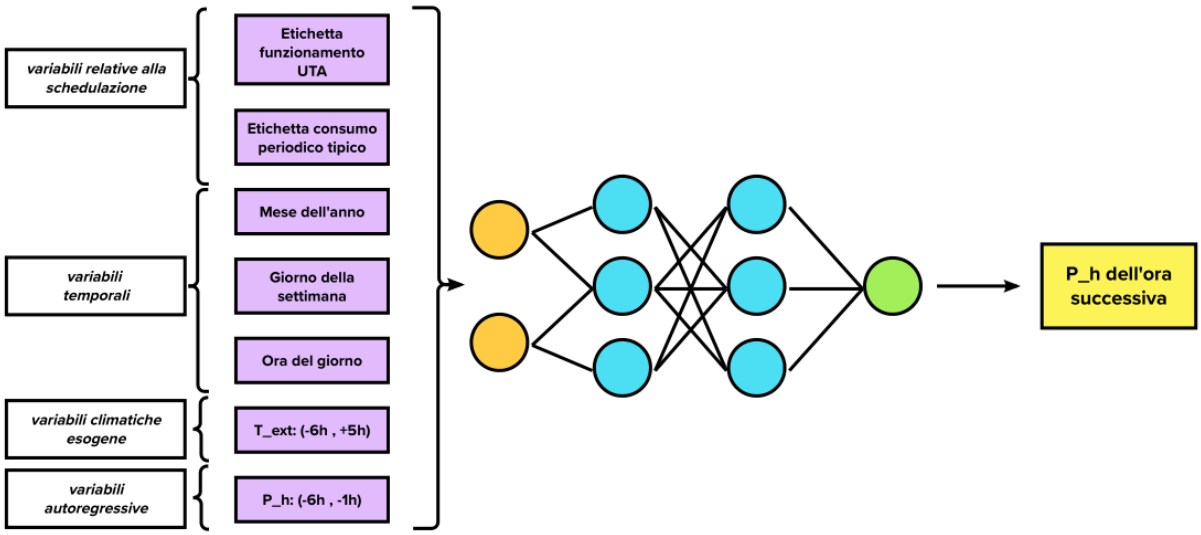

Figura 69 - Rappresentazione delle variabili di input al modello ANN

Come si può vedere dalla Figura 69, le variabili sfruttate sono anche in questo caso di diversa natura, con diversi range temporali per quanto riguarda le temperature e i consumi shiftati.

Per questo modello, la codifica utilizzata per le variabili categoriche (temporali e relative alla schedulazione) è sempre stata l'One-hot encoding, poiché risultava più efficace a livello prestazionale.

L'impostazione degli iperparametri si riassume nel seguente modo:

- 100 epoche di allenamento, per avere una buona accuratezza senza cadere nel fenomeno di sovradattamento.
- 3 hidden layers, nel seguente ordine e con le seguenti caratteristiche:
    - 128 neuroni + Dropout 20%
    - 256 neuroni + Dropout 50%
    - 64 neuroni

Come si può vedere, un ruolo fondamentale lo ha il Dropout, che ad ogni iterazione fa sì che vi sia una perdita del 50% dei dati in modo da mantenere un buon grado di generalità all'interno del modello.

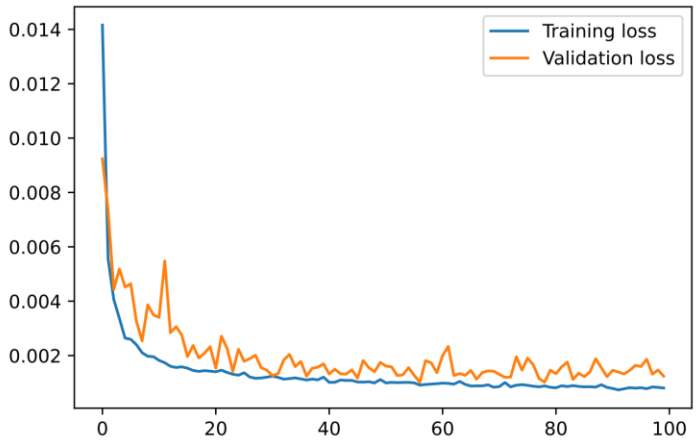

Figura 70 - Andamento della metrica loss al passare delle epoche, per i dati di training e testing

Come si può notare, col passare delle epoche la rete neurale artificiale tende ad avere accuratezza sempre più elevata fino a stabilizzarsi su un valore fisso di funzione costo. Si nota che comunque il dataset di training tenda ad avere valori leggermente migliori, proprio perché il modello si sta fittando sui dati di quel periodo.

Le metriche prestazionali ottenute sono rappresentate nella tabella qui di seguito

Tabella 14 - Metriche prestazionali modello ANN

Metrica	Training	Testing
R2 [-]	0.97	0.95
MAE [kWh/h]	4.81	6.76
RMSE [kWh/h]	9.74	14.31
CV(RMSE) [%]	28.58	32.72

Le performance del modello risultano soddisfacenti, dato che il modello creato permette la previsione dei consumi orari di tutto l'anno.

Dalla Figura 71, si può comunque cogliere la bontà predittiva del modello, nonostante vi siano dei discostamenti leggeri nel periodo invernale, nelle giornate domenicali, e nei periodi di picco.

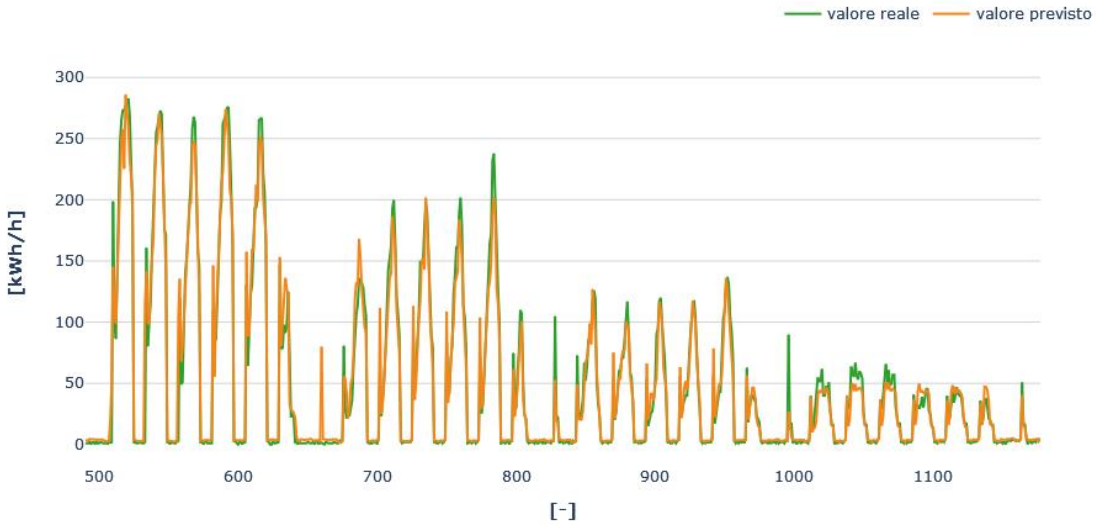

Figura 71 - Profili reali e predetti del periodo di test

Come si può vedere, i risultati ottenuti sono pressoché comparabili col modello Random Forest (Figura 55) creato per la modellazione di consumi di tutto l'anno 2021. Il fatto che anche con modelli più complessi non si riesca ad ottenere un ulteriore incremento dell'accuratezza e quindi un miglioramento delle metriche prestazionali, potrebbe essere dovuto alla presenza di attività di riparazione: questo tipo di attività non avviene con alcun tipo di periodicità, e dipende molto dal tipo di lavorazioni.

Per sopperire a questo tipo di problema, una possibilità sarebbe creare una nuova variabile di input, in modo da considerare anche le seguenti circostanze: si potrebbero sfruttare variabili presenti all'interno del registro di attività (come "ore lavorate", "lavorazione svolta", ecc.) in modo da definire quale tra queste meglio riporta l'informazione del tipo di lavorazione e l'intensità di questa. Una volta raggiunto una mole sufficiente di dati, si potrebbe implementare all'interno del modello questo aspetto, e sfruttarlo per il forecasting usando anche come input le informazioni relative alle future lavorazioni (note). Questo tipo di procedure e informazioni permetterebbe anche un enorme vantaggio per la creazione di un modello per il forecasting della domanda frigo relativa ai macchinari, senza l'uso di variabili autoregressive.

Successivamente si sono creati dei modelli disgiunti, uno per il funzionamento estivo e uno per il funzionamento invernale, per valutare se vi fosse una variazione di prestazioni per l'aspetto predittivo.

4.4.3.1 Modello Funzionamento Estivo – ANN

Per creare il seguente modello si è andati a ridurre le variabili che risultavano superflue, permettendo al contempo stesso una riduzione di complessità del modello. Si sono perciò sfruttate le seguenti variabili:

- T_ext(-6h , +6h), ovvero la temperatura attuale, delle 6 ore precedenti, e delle future 6 ore.
- P_h(-1h , -6h), cioè i valori di consumo delle ultime 6 ore.
- La variabile temporale "ora del giorno", con codifica One-hot encoding.
- La variabile categorica basata sui profili tipici (Lun, Mar-Ven, Sab, Dom), sempre con codifica One-hot encoding.

Non si sono considerate le variabili "giorno della settimana", "tipo funzionamento" e "mese dell'anno" perché non fornivano informazioni aggiuntive utili nella formazione del modello.

I risultati ottenuti dalle metriche prestazionali sono rappresentati nella seguente tabella.

Tabella 15 - Metriche prestazionali modello ANN estivo

Metrica	Training	Testing
R2 [-]	0.96	0.92
MAE [kWh/h]	9.25	11.99
RMSE [kWh/h]	17.20	24.47
CV(RMSE) [%]	24.78	31.98

Il modello non sembra avere particolari vantaggi rispetto a quello annuale precedentemente presentato, anzi, nonostante varie iterazioni e cambi di variabili, tende a sottostimare. Questo limite potrebbe essere dovuto alla scarsità di dati estivi, che quindi permettono di raggiungere solo un certo livello di accuratezza, come si può vedere dai profili relativi al periodo di test nella seguente figura.

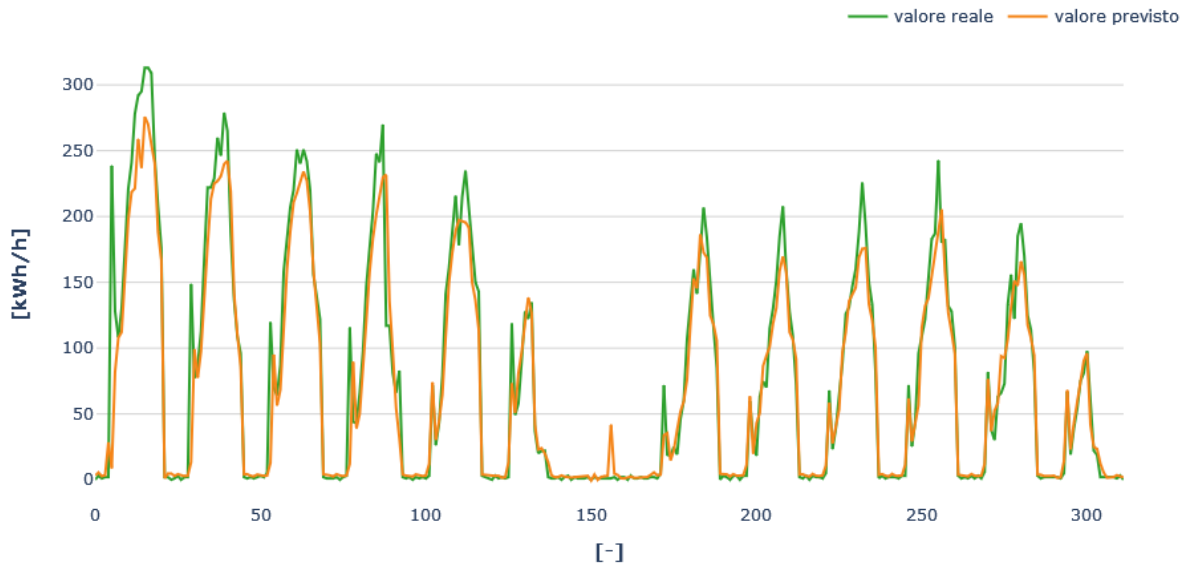

Figura 72 - Profili reali e predetti del modello ANN estivo nel periodo di testing

4.4.3.2 Modello Funzionamento Invernale – ANN

Analogamente per quanto si è fatto nel caso estivo, si andati a creare un modello per la previsione oraria dei consumi frigo nel periodo invernale. Anche in questo caso, determinante è stata la scelta delle variabili di input: la combinazione di variabili di input più efficaci è stata la stessa che si è usata nel caso estivo. Si riportano le metriche prestazionali ottenute.

Tabella 16 - Metriche prestazionali modello ANN invernale

Metrica	Training	Testing
R2 [-]	0.93	0.92
MAE [kWh/h]	2.97	3.43
RMSE [kWh/h]	4.90	5.32
CV(RMSE) [%]	24.19	25.03

Come si può intendere dalla valutazione complessiva di metriche e profili ottenuti, i di MAE e RMSE risultano inferiori anche perché i livelli di potenza raggiunti sono relativamente minori rispetto al caso estivo. Ne consegue che anche la media e quindi il CV(RMSE) abbia valori inferiori, che toccano il 25%.

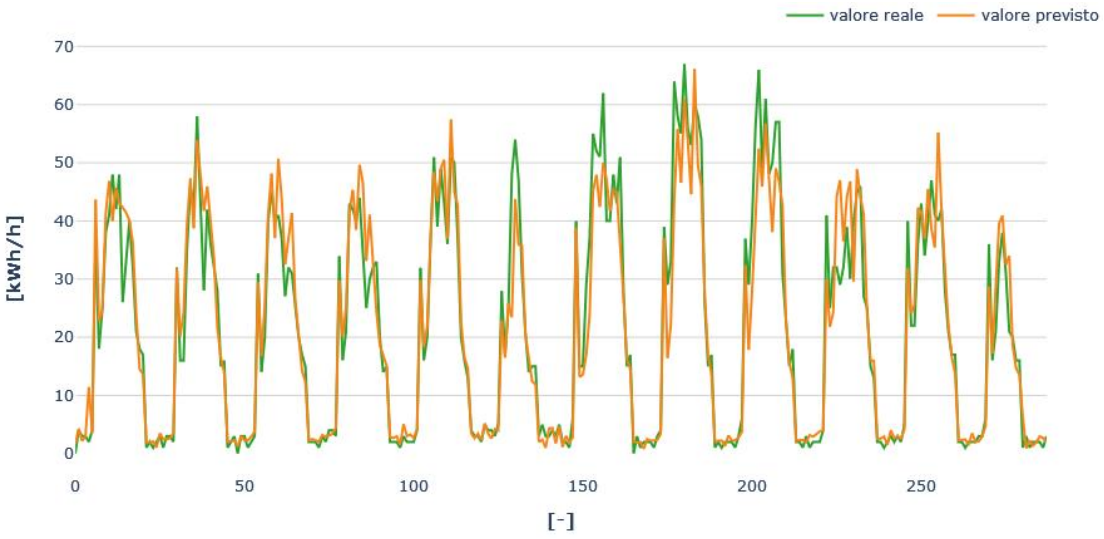

Figura 73 - Profili reali e predetti del modello ANN invernale nel periodo di testing

Il modello riesce a seguire in modo abbastanza accurato le dinamiche del consumo frigo legato ai macchinari, nonostante la correlazione delle variabili climatiche (temperatura esterna) sia meno d'aiuto in questo caso.

4.4.4 Modelli dei consumi di UTA e macchinari – ANN

Successivamente, analogamente alla metodologia e al procedimento usato con il modello Random Forest, si sono creati i modelli di previsione per UTA e macchinari in modo da poter allocare la domanda di energia elettrica dei gruppi frigo alle rispettive utenze.

A differenza del modello Random Forest, in cui l'unica criticità era la scelta delle variabili, si aggiunge un'altra difficoltà: la normalizzazione dei dati sui consumi. Dato che nel

procedimento di creazione dei modelli (schematicamente riportato a Figura 56) si attuano operazioni aritmetiche sui profili reali e previsti, si sono definiti esattamente gli stessi criteri e valori di normalizzazione, per rendere più semplice e sicura la preparazione dati.

Seguendo quindi il medesimo iter, si è creato il modello di previsione dei consumi allocabili ai macchinari sfruttando unicamente variabili esogene. Le variabili sfruttate sono state:

- T_ext(-6h , +6h), ovvero la temperatura attuale, delle 6 ore precedenti, e delle 6 future.
- La variabile temporale "ora del giorno", con codifica One-hot encoding.
- La variabile categorica basata sui profili tipici (Lun, Mar-Ven, Sab, Dom), sempre con codifica One-hot encoding.

Tabella 17 - Metriche prestazionali modello ANN macchinari

Metrica	Training	Testing
R2 [-]	0.85	0.89
MAE [kWh/h]	4.41	4.00
RMSE [kWh/h]	7.25	5.85
CV(RMSE) [%]	35.58	29.19

Come ci si aspettava, le performance del tale modello non sono perfette, ma si possono comunque ritenere accettabili. Il fatto che il modello performi meglio sul dataset di testing si suppone sia dovuto a settimane anomale nel periodo di testing, in cui effettivamente possono essersi verificate lavorazioni fuori programma di riparazione.

Il passaggio successivo è quello di proiettare le variabili estive sul modello dei macchinari, per valutare quale quota di consumo sia allocabile ai macchinari e quale alle UTA. I profili disaccoppiati ottenuti sono riportati qui di seguito.

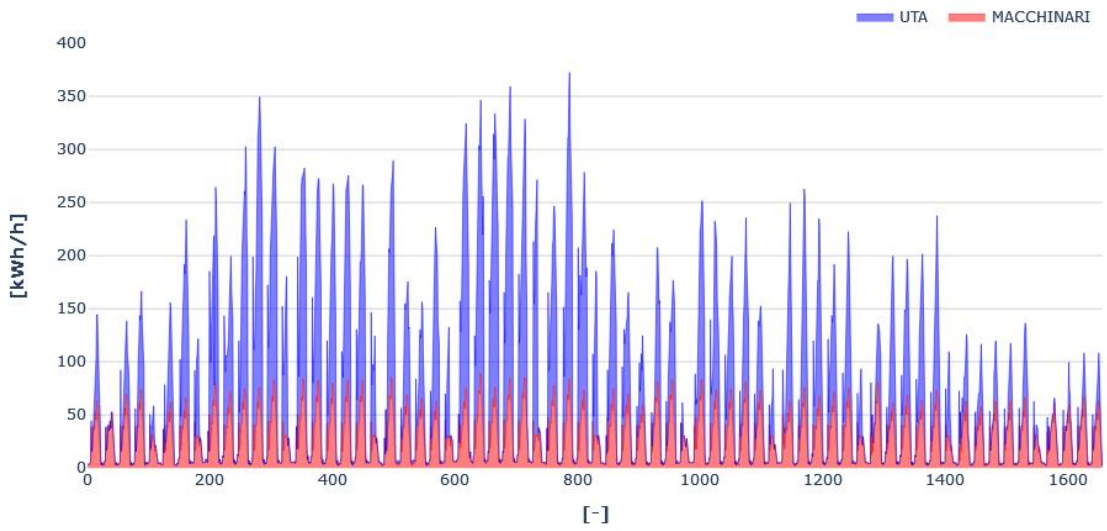

Figura 74 - Profilo aggregato composto dalla stima dei consumi dei macchinari e del consumo delle UTA

Da questo, si è poi estratta la stima dei consumi UTA, su cui si è poi creato il modello di previsione per la quota ad esse allocabile.

Tabella 18 - Metriche prestazionali modello ANN UTA

Metrica	Training	Testing
R2 [-]	0.94	0.94
MAE [kWh/h]	9.31	10.08
RMSE [kWh/h]	17.25	16.50
CV(RMSE) [%]	34.60	28.34

Anche in questo caso il modello creato risulta sufficientemente attendibile, come si può vedere dalla tabella delle metriche prestazionali. Per un controllo finale delle prestazioni complessive dei due modelli, anche in questo caso si è andati ad attuare il procedimento di testing complessivo.

Infatti, per avere un confronto valido, si sono usati esattamente gli stessi dati di testing e training che si erano usati per l'algoritmo Random Forest. Anche in questo caso si sono quindi "nascoste" fin dall'inizio due settimane significative per il testing complessivo.

Proiettando le variabili del periodo di test sui due modelli, come è stato fatto anche con il modello Random Forest, si è poi potuto verificare la bontà dei modelli implementati (il procedimento di testing è completamente analogo a quello rappresentato a Figura 60). Si rappresentano di seguito i profili aggregati delle stime delle quote, affiancate dai valori reali

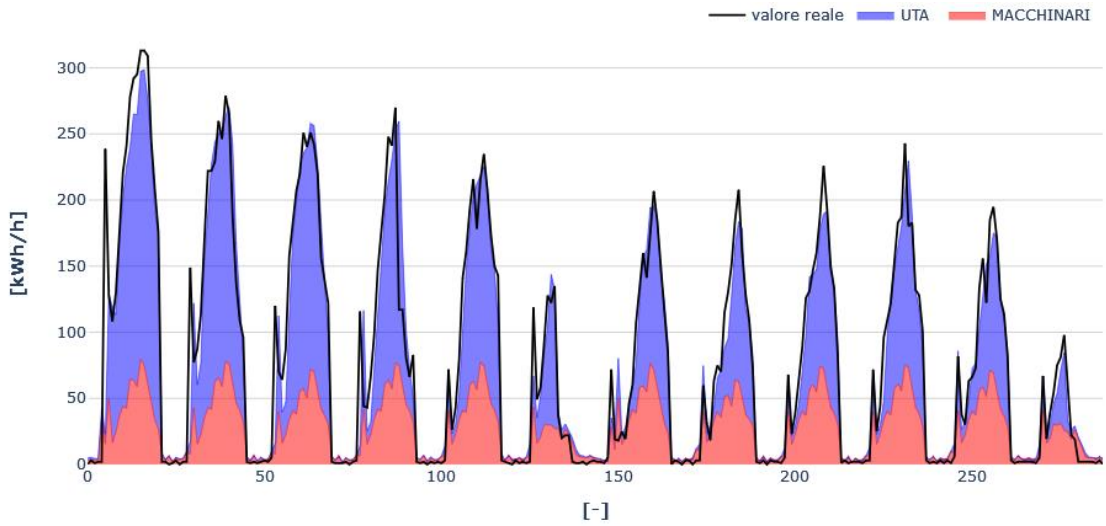

Figura 75 - Consumo frigorigeno reale (linea scura) e l'aggregato dei consumi (aree colorate), ANN

Confrontando i seguenti profili con quelli che si erano ottenuti con il modello Random Forest (Figura 61), si nota che le reti neurali tendono stimare livelli di potenza media più alta nelle ore di picco, rispetto ai profili più piatti che si ottengono con l'altro algoritmo. Questo perché uno dei punti critici delle reti neurali è l'estrapolazione dei dati: se si danno come input al modello dei valori al di fuori del range tipico che si ha nel dataset di training, è possibile che le reti neurali diano in output dei valori insensati. Le temperature raggiunte a ottobre e maggio però non si distanziavano eccessivamente in valore assoluto da quelle del periodo estivo. Le reti avrebbero quindi potuto cogliere un pattern basato sulla relazione delle temperature sfasate, su determinate dinamiche di consumo, o anche solo una relazione tra l'EER e le condizioni climatiche (temperature esterne maggiori diminuiscono l'efficienza dei gruppi frigo e ne

aumentano i consumi a parità di effetto utile). Il problema degli algoritmi black-box è proprio l'interpretabilità fisica: non è possibile capire appieno quali parametri portano il modello a computare un certo output. È perciò necessario avere ben presente le dinamiche e i funzionamenti logico-fisici dei sistemi in esame per poter interpretare al meglio i risultati ottenuti.

Le performance del modello risultano comunque valide, come si può vedere dai valori delle metriche prestazionali nella tabella riportata di seguito.

Tabella 19 - Metriche prestazionali complessive dei due modelli, ANN

Metrica	Testing
R2 [-]	0.91
MAE [kWh/h]	13.72
RMSE [kWh/h]	25.59
CV(RMSE) [%]	30.90

Questo tipo di approccio, basato sul metodo allocativo, permette di stimare in modo distinto le quote delle utenze, e quindi prevedere, e allo stesso tempo controllare, i consumi dei gruppi frigo. Un grande limite di questo procedimento è la mancanza di driver strettamente correlati alla produzione e alle lavorazioni: l'inserimento di variabili più caratterizzanti per quanto riguarda il funzionamento dei macchinari, permetterebbe una previsione più sicura e accurata dei consumi frigo a loro allocabili.

4.4.5 Modello 24h – ANN

Un altro uso dei modelli che sfruttano le reti neurali artificiali, è la possibilità di avere più di un output contemporaneamente. Questo, se l'obiettivo è quello della previsione dei consumi, permette di prevedere più time-step futuri. Nel presente caso, si è andati a creare un modello che permettesse di prevedere i consumi delle 24 ore seguenti. Questo tipo di modello necessita un maggiore numero di informazioni e un livello più alto di complessità computazionale, data anche la sfida più difficoltosa.

Inoltre, per l'interpretazione dei dati, a differenza del caso precedente, è risultata molto più efficace la codifica Binary encoding per quanto riguarda le variabili categoriche. Questo perché si riesce a passare in modo ugualmente efficace questo tipo di informazioni, riducendo il numero di variabili in ingresso relative e ottimizzando quindi le prestazioni computazionali del modello.

Anche in questo caso è stata necessaria una attenta scelta delle variabili e una accurata calibrazione degli iperparametri.

Si rappresenta di seguito lo schema logico relativo alle variabili in entrata al modello.

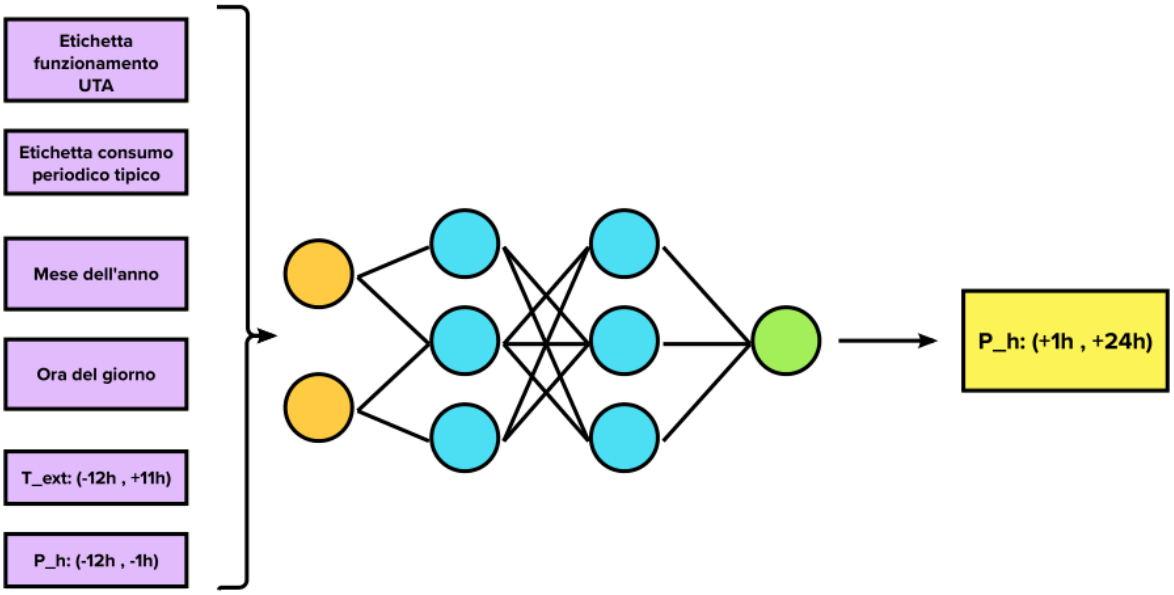

Figura 76 - Schema logico delle variabili di input al modello di forecasting 24 h ANN

In questo caso, per ottimizzare e migliorare le performance, è stato necessario implementare una rete più profonda e quindi più complessa.

Il modello è riassumibile attraverso i seguenti iperparametri:

- 120 epoche di allenamento, dato che analizzando dopo varie iterazioni, si è valutato che questo permettesse un buon compromesso tra accuratezza e generalità dell'allenamento.

- 6 hidden layers, nel seguente ordine e con le seguenti caratteristiche:
    - 256 neuroni + Dropout 50%
    - 128 neuroni + Dropout 30%
    - 64 neuroni
    - 128 neuroni + Dropout 20%
    - 64 neuroni

  In questo caso, inoltre, dato il differente obiettivo del modello, si hanno 24 neuroni nel layer di output (la previsione è sulle 24 ore successive) invece che uno.

Una volta quindi definiti sia gli iperparametri che le variabili, è stato possibile fittare il modello e verificare le sue performance sul dataset di testing.

Il periodo di testing considerato è sempre lo stesso scelto per i modelli annuali, comprendente diversi tipi di funzionamento in diversi periodi dell'anno, come si può vedere in Figura 77.

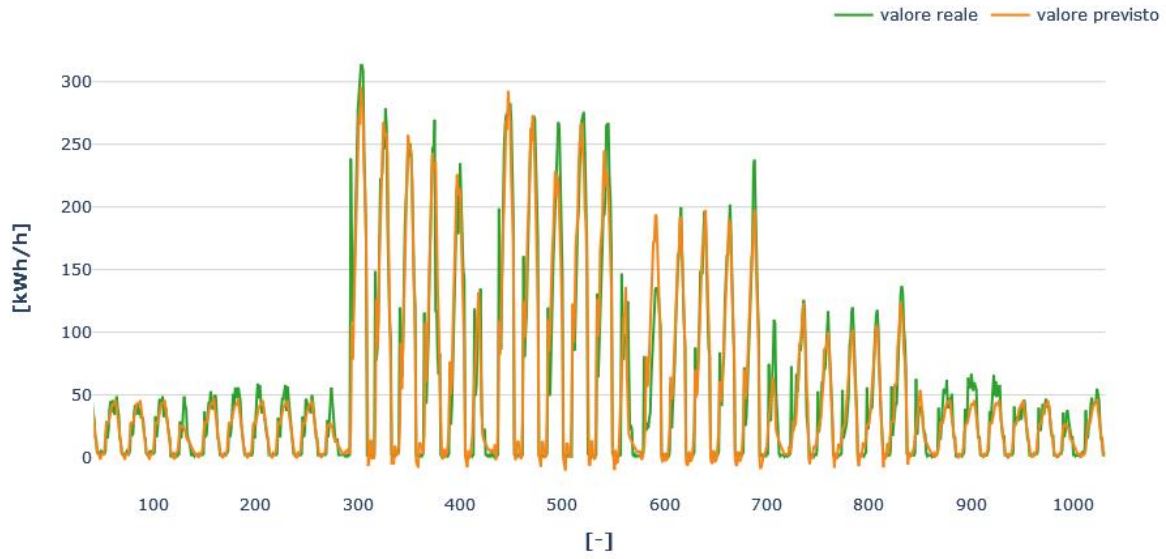

Figura 77 - Profili reali e predetti dal modello di previsione 24h ANN nel periodo di testing

Eccetto qualche discostamento occasionale nelle ore di punta e qualche valore leggermente sotto lo 0, si può notare come effettivamente sia accurato il modello di forecasting a multi-step temporale.

Per valutare in termini numerici le sue performance, ci si affida però sempre alle metriche prestazionali (Tabella 20).

Tabella 20 - Metriche prestazionali modello di forecasting 24h ANN

Metrica	Training	Testing
R2 [-]	0.93	0.93
MAE [kWh/h]	9.69	10.67
RMSE [kWh/h]	16.29	18.52
CV(RMSE) [%]	40.97	37.36

I risultati ottenuti sono soddisfacenti, dato che le prestazioni ottenute sono comparabili con quelle ottenute con l'algoritmo Random Forest di previsione a singolo step. Il presente modello riesce quindi a prevedere il profilo dei consumi delle future 24 ore (con dati orari), tutto l'anno, in qualunque tipo di funzionamento.

A discapito di una leggera perdita di accuratezza, con questo tipo di modello previsionale si possono quindi stimare i consumi di un'intera giornata futura, permettendo la creazione di strategie di ottimizzazione basate su un maggiore range temporale.

4.4.6 Ricapitolazione modelli – ANN

Anche alla fine di questo capitolo si riporta un riassunto dei modelli creati e le loro caratteristiche.

Tabella 21 - Riassunto modelli ANN

MODELLO	output	input	timestep	periodo validità
	potenza media oraria dei gruppi frigo	variabili: • autoregressive • climatiche • temporali • schedulari	1h	Intero periodo annuale
	potenza media oraria dei gruppi frigo	variabili: • autoregressive • climatiche • temporali • schedulari	1h	Periodo di funzionamento estivo
	potenza media oraria dei gruppi frigo	variabili: • autoregressive • climatiche • temporali • schedulari	1h	Periodo di funzionamento invernale
	potenza media oraria dei macchinari	variabili: • climatiche • temporali • schedulari	1h	Periodo di funzionamento estivo
	potenza media oraria delle UTA	variabili: • autoregressive • climatiche • temporali • schedulari	1h	Periodo di funzionamento estivo
	potenza media oraria dei gruppi frigo	variabili: • autoregressive • climatiche • temporali • schedulari	24h	Intero periodo annuale

5. Opportunità applicative dei modelli implementati

5.1 Riassunto delle performance dei modelli

Per avere una panoramica generale delle prestazioni dei modelli annuali di forecasting, si riporta di seguito una tabella riassuntiva di questi.

Tabella 22 - Confronto delle metriche prestazionali tra i modelli implementati

MODELLO	R2 [-]	MAE [kWh/h]	RMSE [kWh/h]	CV(RMSE) [%]
Regressione lineare	0.68	34.45	44.51	-
Decision Tree	0.91	9.98	19.83	46.00
Random Forest	0.94	6.98	15.65	36.31
Reti Neurali Artificiali	0.95	6.76	14.31	32.72

Per quanto riguarda il modello di regressione lineare, si è scelto il modello più robusto, ovvero quello che sfrutta come driver esogeno la media mobile delle ultime 4 ore (SMA-4h). Il CV(RMSE) dovrebbe essere pari a 28.08% ma non risulta confrontabile con quello degli altri modelli ovviamente (la media su cui si adimensionalizza il RMSE è diversa rispetto agli altri).

Tutti i modelli creati possono performare durante l'intero anno eccetto proprio il modello di regressione lineare semplice, il quale risulta utile unicamente durante il periodo di funzionamento delle UTA, in cui il carico frigorigeno è maggiormente influenzato dalle condizioni climatiche (temperatura esterna e carico solare).

Per ognuno di questi modelli, vi sono pro e contro:

- **Regressione Lineare**:
  - **PRO**: Modello estremamente semplice e facilmente interpretabile dal punto di vista fisico: si può associare un significato tecnico-fisico non solo ai parametri caratterizzanti ma anche a variazioni di questi in caso di modifiche del sistema.
  - **CONTRO**: Accuratezza relativamente bassa, e poco versatile in quanto può essere usato solo nel periodo estivo e nell'orario in cui le UTA risultano in normale funzionamento.
- **Decision Tree**:
  - **PRO**: Modello relativamente semplice e interpretabile dal punto fisico, decide lui le migliori variabili e soglie nella fase di training. Performa durante l'intero anno con un'accuratezza del tutto accettabile (con unicamente 32 valori di output riesce ad approssimare/stimare i consumi dell'intero anno).
  - **CONTRO**: Necessita una attenta calibrazione degli iperparametri per evitare il fenomeno di sovradattamento, e proprio per la natura dell'algoritmo è necessario un trade-off tra precisione dell'output e le capacità di generalizzazione. Ciò comporta dei profili troppo discretizzati e accurati.

- **Random Forest**:
    - **PRO**: Molto accurato, non necessita una calibrazione degli iperparametri meticolosa quanto il Decision Tree. L'*ensemble learning* permette di aumentare la robustezza e precisione delle previsioni rispetto ad un singolo albero decisionale.
    - **CONTRO**: La natura binaria dell'algoritmo permette di stimare sempre dei valori interni al range disponibile dai dati di allenamento. Questo fa sì che non si abbiano mai valori irrealistici, ma non permette neanche di estrapolare informazioni. Inoltre, proprio per come funziona l'algoritmo, si ha una perdita di interpretabilità fisica.

- **Artificial Neural Network**:
    - **PRO**: Permette di cogliere pattern e dinamiche non banali, raggiungendo accuratezze molto elevate. Grazie alla sua natura, riconosce interazioni tra variabili anche di carattere olistico, non spiegabili con gli altri modelli.
    - **CONTRO**: Si perde completamente l'interpretabilità fisica dei parametri, oltre alla necessità di una preparazione dei dati accurata e attenta. Nonostante riesca a cogliere correlazioni non lineari, in caso di condizioni e variabili di input al di fuori del range su cui si allenato, le capacità di estrapolazione non risultano robuste: i valori che si ottengono potrebbero non avere né senso fisico, né senso logico.

Il modello più performante risulta quello basato sulle reti neurali artificiali, che potrebbe essere ulteriormente migliorato attraverso l'introduzione di nuove variabili di input che descrivano la domanda frigorigena da parte dei macchinari. Si valuta però come modello più resiliente e praticamente allo stesso modo prestante quello creato sull'algoritmo Random Forest. La scelta di partire da algoritmi più semplici, per arrivare allo sviluppo di modelli di maggiore complessità, è stata maturata in funzione del principio del "*rasoio di Occam*": tra più soluzioni equivalenti a un problema, si preferisce la più semplice. Ragionando in funzione di questo pensiero, non risulta quindi scontato scegliere un modello leggermente meno accurato, se questo appare più semplice, resiliente e interpretabile.

5.2 Benchmarking interno: 2021-2022

Un esempio pratico di applicazione dei modelli creati è quello della valutazione dei risparmi dopo un intervento di retrofit. Infatti, nel periodo estivo 2022 il sistema che regolava le UTA è stato modificato. Questa variazione della logica di funzionamento aveva come obiettivo il risparmio energetico attraverso una variazione degli orari di accensione/spegnimento degli impianti di condizionamento. Ciò non ha influito solamente sul consumo associato ai ventilatori, ma anche sulla domanda frigorifera richiesta ai chiller.

Per valutare il risparmio energetico ottenuto da questa nuova strategia di gestione, è necessario in primis creare un modello che simuli in modo accurato il sistema in condizioni di pre-retrofit. In questo caso quindi il modello dovrà essere basato sui dati dell'estate 2021.

Una volta creato il modello di baseline, ovvero il modello che mi permette di simulare il sistema con cui confrontarsi, è necessario proiettare le condizioni di post-retrofit su questo. Ciò che si ottiene da tale procedura sono i profili di consumo che si avrebbero se non si fossero apportate modifiche/interventi di retrofit al sistema.

Questo tipo di procedimento è definito in letteratura come benchmarking interno, ovvero un confronto di un sistema con sé stesso con fine ultimo la valutazione dell'impatto di cambiamenti o semplicemente la valutazione di possibili variazioni/guasti nel sistema.

È riportato di seguito uno schema logico rappresentante il procedimento di benchmarking interno attuato.

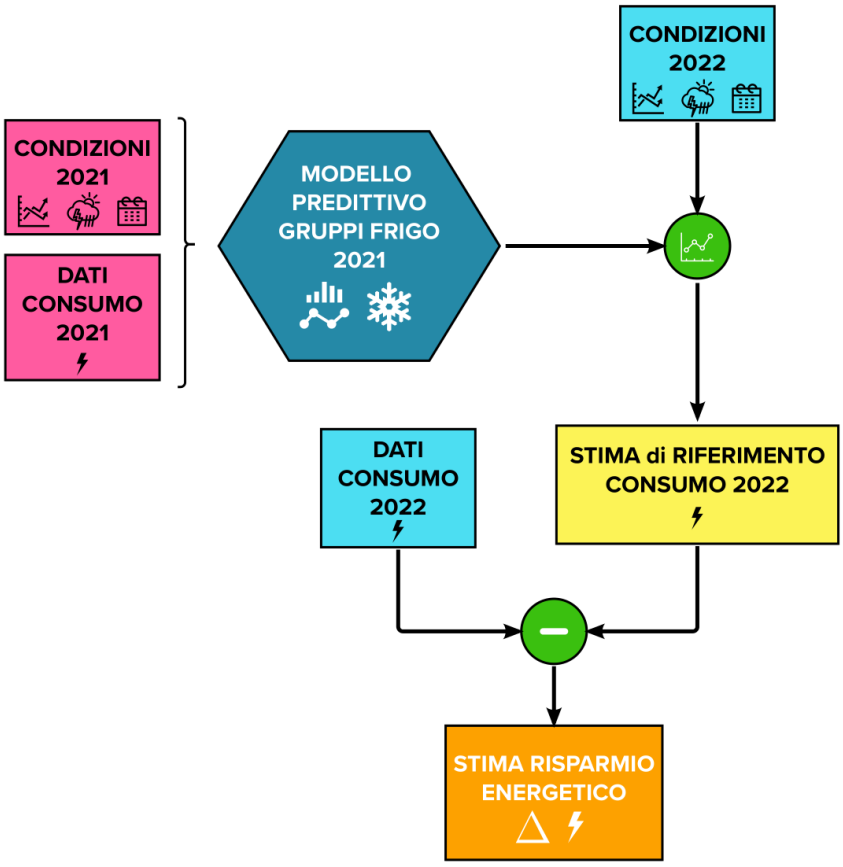

Figura 78 - Procedimento di benchmarking interno

Una criticità di questo procedimento è la gestione delle variabili autoregressive: queste sono infatti rappresentative della dinamicità del sistema da modellare. Dato che il sistema è intrinsecamente cambiato, vista l'impostazione di nuove schedule e logiche di funzionamento, anche le sue dinamiche di consumo non risultano più le stesse.

Per sopperire a questo tipo di problema ci sono due alternative:

- Creare un nuovo modello che sfrutti variabili puramente esogene.
- Sfruttare un approccio concettualmente simile al *recursive multi-step forecasting*.

La prima alternativa risulterebbe la più sicura se il sistema fosse sufficientemente caratterizzabile attraverso le variabili esogene a disposizione. Purtroppo, però, dopo vari tentativi di creazione di un modello composto solo da variabili climatiche, temporali, e relative

alla schedulazione, ci si è resi conto che molto spesso, soprattutto durante il funzionamento estivo, il modello tendeva a sottostimare i consumi e risultando effettivamente poco accurato.

La seconda alternativa invece permette di sfruttare sempre le variabili di natura autoregressiva, e quindi cogliere meglio le dinamiche tipiche del sistema. Questo è possibile grazie al continuo aggiornamento delle nuove variabili regressive di input: partendo da delle variabili regressive iniziali reali, dopo sfruttate come input consumo dello step temporale successivo, si prende questa ultima stima come valore di riferimento per la predizione successiva. Si genera quindi una finestra mobile (in rosso in Figura 79) che ad ogni predizione fa come propria variabile di input autoregressiva la stima appena ottenuta. Per una migliore comprensione di tale procedimento, si riporta qui di seguito uno schema logico di questo.

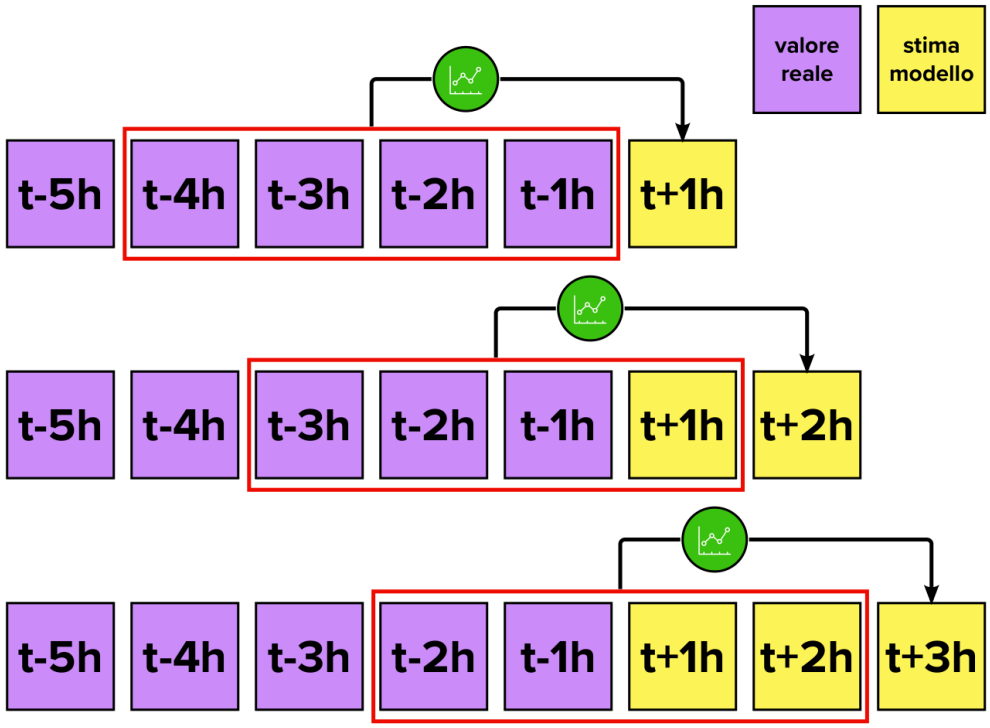

Figura 79 - Stima delle variabili autoregressive con l'approccio "recursive multi-step forecasting"

Come si può vedere, sempre più variabili stimate entrano a far parte della finestra mobile, rappresentante delle variabili di input autoregressive. Questo fa sì che dopo poche iterazioni il sistema non abbia più in ingresso valori reali ma solo valori fittizi stimati. Il vantaggio di questa procedura è che quindi non si usano le variabili autoregressive reali, che travierebbero il modello, ma allo stesso tempo si rischia di accumulare incertezze ad ogni step temporale e quindi ottenere dei profili incorretti.

Si è quindi andati a verificare come questo tipo di metodologia performasse sul dataset di testing precedentemente usato per testare il modello di forecasting dei consumi frigo relativo al 2021. Applicando quindi tale strategia, ciò che si ottiene sono i seguenti profili.

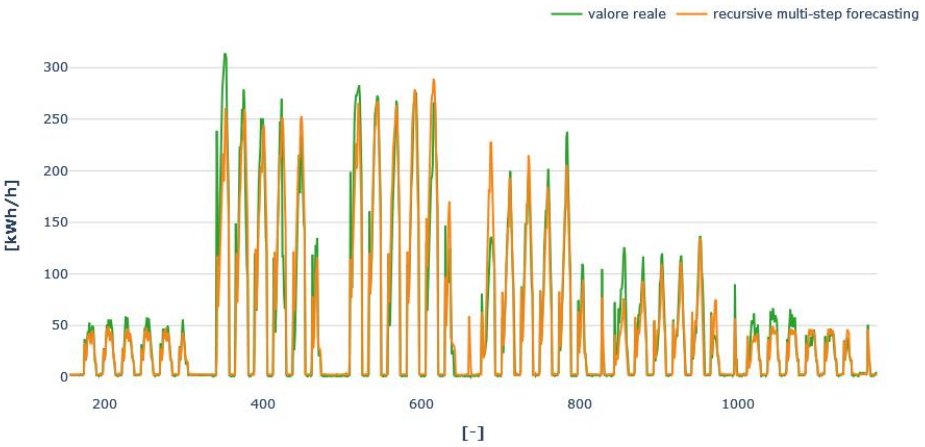

Figura 80 - Testing dell'applicazione del "recursive multi-step forecasting"

Come si può vedere, l'incertezza e le performance risultano inferiori a quelle ottenibili con un approccio classico (Tabella 23), ma comunque si può notare una buona approssimazione dei profili, considerando anche che a parte i primi quattro valori di consumo, tutti i successivi sono frutto di stime di valori autoregressivi.

Tabella 23 - Metriche prestazionali sul periodo di testing dell'applicazione del recursive multi-step forecasting

Metrica	Testing
R2 [-]	0.91
MAE [kWh/h]	8.89
RMSE [kWh/h]	19.25
CV(RMSE) [%]	44.00

Considerando quindi come sufficientemente valido il modello, si sono successivamente proiettate le condizioni del mese di luglio 2022 su questo (climatiche e temporali reali, mentre quelle autoregressive stimate), seguendo come preannunciato l'approccio sfruttato per il *recursive multi-step forecasting*. Si analizza il risultato ottenuto per i giorni di lavoro (dal lunedì al venerdì).

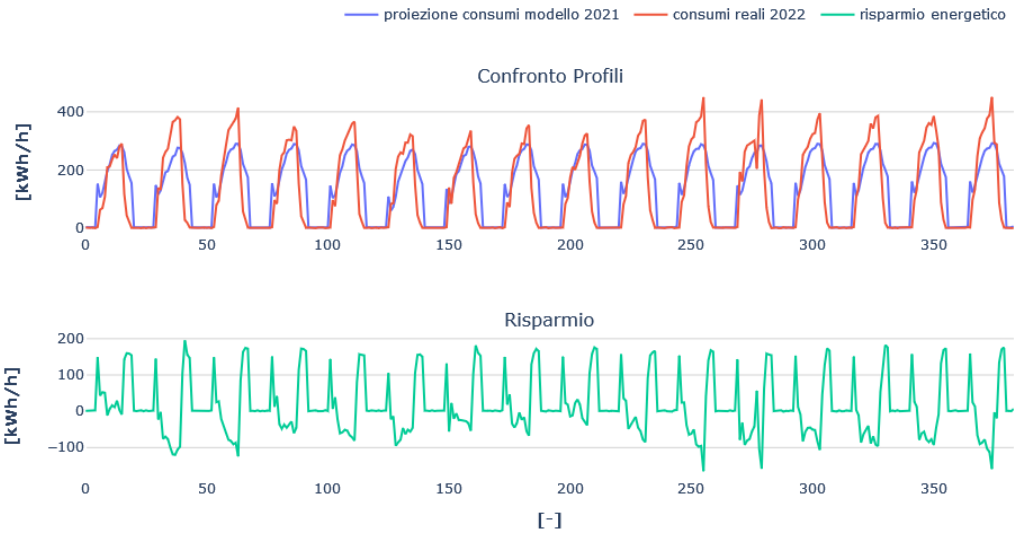

Figura 81 - Rappresentazione dei profili a confronto e del risparmio stimato

Si riporta un ingrandimento dei profili per capire al meglio la nuova logica di funzionamento rispetto a quella precedentemente attuata.

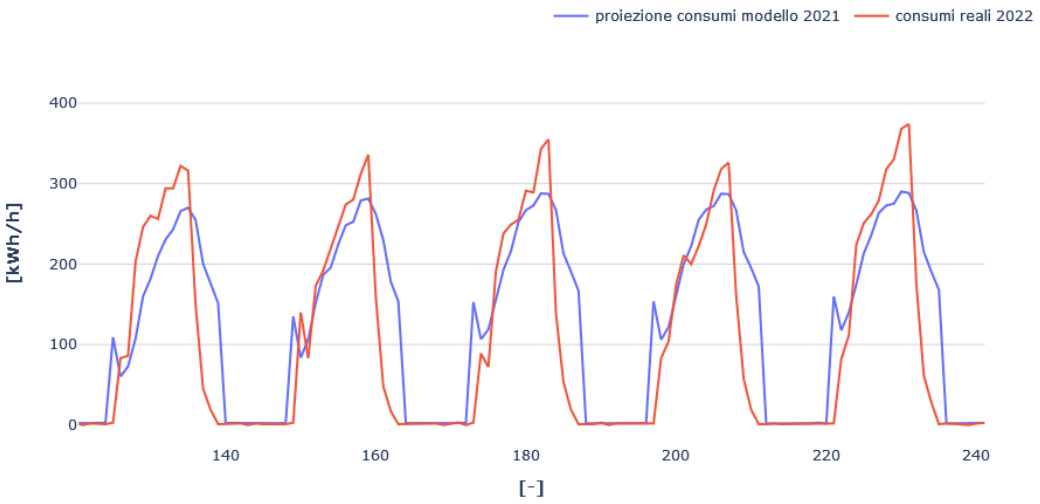

Figura 82 - Zoom-in sui profili di benchmark

Come si può vedere, la nuova schedulazione prevede una accensione dei gruppi frigo posticipata, accompagnata da uno spegnimento anticipato. Questo fa sì che, per soddisfare la domanda di carico frigorigeno, vi sia un livello di potenza di picco superiore nelle ore di punta. Per capire se la nuova strategia funziona, è necessario valutare se il risparmio che si ottiene a inizio e fine turno di lavoro riesce a prevalere sul sovraconsumo che si ha nell'ora di punta.

Per valutare il risparmio economico, si è andati a integrare la differenza dei profili (sottraendo profilo reale da quello stimato) ipotizzando un costo dell'energia elettrica pari a 33 c€/kWh [18].

Figura 83 - Andamento del risparmio economico nel mese di luglio 2022

Dall'analisi, si è potuto valutare l'impatto dell'intervento di retrofit dal punto di vista energetico, economico, e ambientale (considerando che mediamente in Italia si ha un'emissione di 0.3 kg di CO_2 equivalente al kWh elettrico prodotto [19]).

Tabella 24 - Risparmi stimati attraverso il procedimento di benchmarking interno

Risparmio	Luglio 2022 [-]	Risparmio Medio Giornaliero [-/gg]
Energetico [kWh]	4707	294
Economico [€]	1541	96
Ambientale [kg_{CO2eq}]	1469	92

Percentualmente, nei 16 giorni rappresentativi di luglio, il risparmio è stato circa del 9% (considerando che il consumo totale stimato dal modello corrisponde a 50770 kWh).

L'implementazione di modelli accurati e validi per valutare l'impatto di interventi di retrofit è fondamentale per una giusta stima dei risparmi. Banalmente, se si installasse un nuovo gruppo frigo di migliori prestazioni, ma l'estate a venire fosse più calda rispetto alla precedente, ne consegue che anche se i consumi risulteranno più alti, ciò non vuol dire che l'intervento sia stato un fallimento: il modello su cui proiettare le condizioni post-retrofit serve proprio a questo.

5.3 Possibili implementazioni per l'ottimizzazione

Nella seguente sezione si riportano degli spunti volti all'ottimizzazione del sistema.

Uno dei primi passi per avere anche una maggiore interpretabilità fisica e quindi più possibilità di implementazione di algoritmi volti all'ottimizzazione dei consumi frigo, è la traduzione dei consumi elettrici frigo in una stima della domanda frigorigena. Infatti, il consumo elettrico è funzione di:

- Temperatura esterna (che influenza l'EER).
- Il PLR (*Part Load Ratio*) dei gruppi frigo durante il funzionamento.
- La logica di regolazione dei gruppi frigo.

Dai dati storici si potrebbero estrarre i profili relativi alla domanda frigorigena, sfruttando i dati di targa dei gruppi frigo e correlazioni semi-empiriche (l'EER è principalmente funzione della T_ext e del PLR).

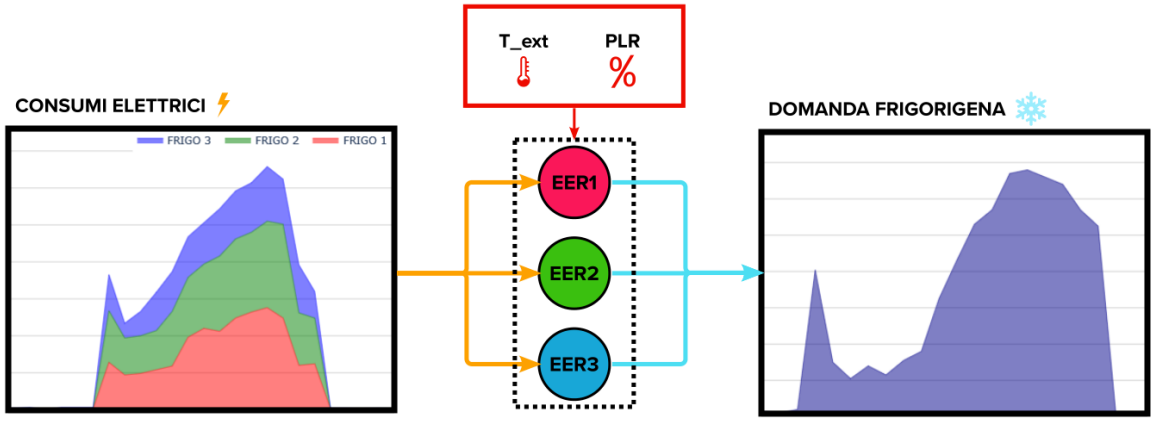

Figura 84 - Possibile procedimento di conversione dei consumi elettrici in domanda frigorigena

Una volta estratte tali informazioni, sarebbe possibile creare un nuovo modello di forecasting, non più dei consumi elettrici, ma della domanda frigorigena stessa. Per farlo in modo accurato, si potrebbero inserire nuove variabili, legate anche al funzionamento dei macchinari che necessitano fluido refrigerante, ed evitare di fare uso di variabili autoregressive.

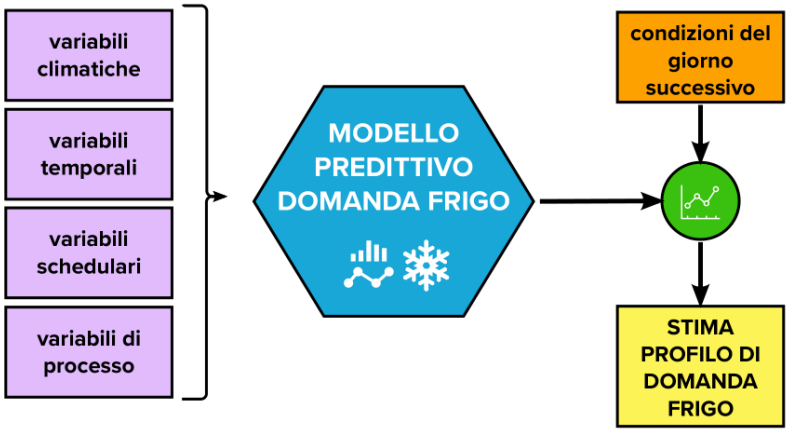

Figura 85 - Creazione di un modello per la stima della domanda frigorigena

I driver che influenzano la domanda frigorigena si immagina saranno molto simili a quelli sfruttati per i modelli sui consumi elettrici, dato che il procedimento in Figura 84 non varia in modo sensibile la natura del profilo.

Una volta stimata la domanda frigorigena, è possibile provare a ottimizzare i consumi elettrici dei gruppi frigo, andando ad applicare possibili strategie sia sui funzionamenti di UTA e macchinari, lato domanda, sia sulla logica di funzionamento dei gruppi frigo, lato produzione. Per questo tipo di operazioni potrebbe risultare di importanza rilevante anche il metodo allocativo sfruttato nella creazione dei modelli di previsione dei consumi delle UTA e dei macchinari.

Mantenendo il focus unicamente sui gruppi frigo, si potrebbero attuare delle strategie che prendano in considerazione sia la temperatura esterna, che le performance dei gruppi frigo stessi, in modo da far sì che questi funzionino possibilmente sempre al massimo delle performance, coordinando in modo intelligente la loro accensione e regolazione.

Ad esempio, si potrebbero schedulare le attività produttive che richiedono fluido refrigerante (di processo o anche solo per gli ausiliari) in funzione della temperatura esterna e della previsione del carico relativo alle UTA: andando ad appiattire la domanda frigorifera, si potrebbe evitare l'accensione di un gruppo frigo a PLR ridotti (di norma, se non ci si ritrova in condizioni critiche, bastano due unità per soddisfare il carico termico) e quindi permettere a quelli già in operazione di essere sempre in condizioni prossime al nominale (concetto molto affine alla tematiche di Demand Side Management e Demand Response). Per rendere possibili questi dislocamenti della domanda frigorifera si potrebbe fare affidamento sulla capacità dell'accumulo termico.

Per implementare nuovi algoritmi di ottimizzazione, sarà perciò fondamentale l'introduzione di nuova componentistica di campo, in grado ad esempio di rilevare le temperature relative al fluido refrigerante all'interno dell'accumulo termico, o altre grandezze relative ai macchinari serviti dai gruppi frigo.

6. Conclusioni

Oggetto del presente elaborato è stato lo studio di modelli di forecasting per la predizione dei consumi elettrici di una centrale frigorifera industriale, a partire da dati reali di un'azienda.

Si è partiti da semplici modelli statistici fino ad arrivare allo sviluppo di algoritmi di machine learning di maggiore complessità, consentendo il raggiungimento di un'accuratezza sempre più elevata. Per fare questo, si è tenuto conto della fisica del problema, della logica aziendale, e dei driver influenzanti e descrittivi del sistema, che hanno permesso l'estrazione e l'utilizzo di informazioni caratterizzanti della centrale.

Lo studio approfondito dei consumi frigoriferi e della loro modellazione ha permesso una distinzione degli stessi in due componenti principali: una componente legata ai fabbisogni energetici dei macchinari e una componente relativa al fabbisogno delle UTA. Questo non solo permette di prevedere con un passo temporale di un'ora il consumo dell'unità di refrigerazione industriale, ma anche di distinguere le quote che lo compongono, creando opportunità per ulteriori implementazioni e aumentando la consapevolezza dell'uso dell'energia. L'utilizzo combinato di metodo allocativo e algoritmi di machine learning ha permesso di evidenziare la rilevanza delle variabili indipendenti e della loro influenza sui sistemi analizzati: si ritiene quindi che tale approccio possa rappresentare uno strumento valido nell'analisi dei consumi energetici di sistemi complessi e meritevole di futuri approfondimenti.

Avvalendosi di differenti tipi di encoding delle informazioni categoriche è stato inoltre possibile ottimizzare il numero di variabili di input ai modelli analizzati. Infatti, per il modello di previsione giornaliero, di orizzonte temporale pari a 24 ore e risoluzione oraria, la codifica binaria delle informazioni temporali ha consentito di raggiungere prestazioni predittive totalmente soddisfacenti. Questa capacità predittiva anche su una più larga scala temporale, apre diverse opportunità di miglioramento sia per l'aspetto di controllo dei sistemi industriali, sia per la gestione stessa dell'azienda.

Nell'ultima parte dell'elaborato si è riportata un'analisi di benchmarking interno, confrontando il funzionamento dei gruppi frigo nell'estate 2022 rispetto a quello nell'estate 2021. Durante questa fase, è emersa non solo l'elevata influenza delle variabili autoregressive, ma anche la criticità a loro associata. Ciò ha evidenziato l'importanza della metodologia sia nella creazione di un modello che nella sua applicazione. Lo sviluppo di tale modello e l'implementazione di questa metodologia, hanno consentito la valutazione dell'impatto di una modifica nella schedulazione delle UTA sui consumi frigoriferi, altrimenti difficile da stimare. L'analisi svolta ha permesso di mettere in risalto il potenziale che questi modelli hanno, sia nel forecasting, che nel monitoraggio e supervisione dei sistemi energetici.

L'insieme delle conoscenze estratte, delle metodologie sviluppate e degli strumenti acquisiti, permette l'implementazione di nuove strategie di gestione e automazione del sistema frigorifero. Gli sviluppi futuri sono quindi volti a un'innovazione dell'architettura di monitoraggio, improntata verso lo sviluppo di algoritmi di ottimizzazione che fruiscano delle informazioni derivate dai modelli di forecasting.

Glossario

ANN	*Artificial Neural Network*, reti neurali artificiali
CART	*Classification And Regression Tree*, albero decisionale
CV(RMSE)	Coefficiente di Variazione del RMSE
EER	*Energy Efficiency Ratio*, rapporto di efficienza energetica
EF	Consumo di energia elettrica relativo ai gruppi frigo [kWh/h]
I_sol	Irradianza solare [Wh/m^2]
MAE	*Mean Absolute Error*, errore assoluto medio
PLR	*Part Load Ratio*, rapporto di carico parziale dei gruppi frigo
P_h	Potenza media oraria [kWh/h]
RMSE	*Root-Mean-Square Error*, radice dell'errore quadratico medio
R2	Coefficiente di determinazione
SMA	*Simple Moving Average*, media mobile semplice
T_ext	Temperatura media esterna oraria [°C]

Bibliografia

[1] UNEP, Emission Gap Report 2022.

[2] IEA, Global Energy Review: CO2 Emissions in 2021.

[3] UNFCCC, Accordo di Parigi.

[4] MATTM, Report sull'Accordo di Parigi.

[5] Commissione Europea, Il Green Deal Europeo.

[6] Prosperous Network, Green New Deal: tra rivoluzione verde e lacune giuridiche. https://prosperousnetwork.com/2020/11/green-new-deal/.

[7] Consiglio dell'Unione Europea, Green Deal europeo. https://www.consilium.europa.eu/it/policies/green-deal/ .

[8] EEA, Definizioni di Adattamento e Mitigazione. https://www.eea.europa.eu/it/help/domande-frequenti/qual-e-la-differenza-tra.

[9] ICT, Smart-Factory. https://icxt.di.unito.it/laboratori/smart-factory-and-materials-lab/ .

[10] IEA, The Future of Cooling.

[11] IEA, Report on the final energy consumption for space cooling. https://www.iea.org/data-and-statistics/charts/final-energy-consumption-for-space-cooling-by-region-and-number-of-space-cooling-equipment-units-in-operation-in-the-net zero-scenario-2000-2030.

[12] Electronic Team Publishing, Modbus RTU communication guide. https://www.virtual-serial-port.org/articles/modbus-rtu-guide/#:~:text=in%20related%20articles-,What%20does%20Modbus%20RTU%20stand%20for%3F,%2C%20RS485%2C%20anc %20RS422%20protocols.

[13] Capozzoli Alfonso, materiale didattico del corso "Gestione energetica e automazione negli edifici".

[14] Scikit-Learn Library, Decision Tree Regressor. https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeRegressor.html

[15] Scikit-Learn, Random Forest Regressor. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html.

[16] Bhattiprolu Sreenivas, Materiale didattico open source

[17] Keras, The Sequential model. https://keras.io/guides/sequential_model/.

[18] Selectra Italia, Costo dell'energia elettrica per le aziende. https://luce-gas.it/business/offerte/costo-kwh-aziende.

[19] Qualenergia, CO2 e mix elettrici degli stati UE. https://www.qualenergia.it/articoli/co2-quanto-sono-sporchi-i-mix-elettrici-dei-diversi-stati-ue/.