



UNIVERSITÀ POLITECNICA DELLE MARCHE

FACOLTÀ DI INGEGNERIA

Corso di Laurea triennale in Ingegneria Meccanica

**Sviluppo ed applicazione di una soluzione di coaching
personalizzato per la misura e l'ottimizzazione del comfort
nell'ambiente costruito**

Development and application of a personalized coaching solution for
measuring and optimizing comfort in the built environment

Relatore:

Prof. Revel Gian Marco

Correlatrici:

Dott.sa Cipollone Vittoria

Dott.sa Serroni Serena

Tesi di Laurea di:

Rossini Giacomo

A.A. 2023/2024

Abstract

Il lavoro descritto in questa tesi consiste nell'applicazione di modelli Machine Learning per la previsione della sensazione termica, espressa come sensazione termica (Thermal Sensation Vote), utilizzando parametri ambientali quali temperatura esterna, temperatura interna e umidità relativa interna, misurati in edifici adibiti ad utilizzo ufficio. L'obiettivo principale è sviluppare un approccio innovativo basato sull'applicazione di algoritmi di Machine Learning (ML), in grado di prevedere il feedback partendo da parametri ambientali misurati in ambienti reali: in questo modo è possibile superare i limiti evidenziati dalla letteratura nell'ambito dell'ottimizzazione della misura del comfort e della sua componente soggettiva, quali l'utilizzo di dati fisiologici, difficilmente registrabili, e la periodica necessità di richiedere feedback agli utenti. Sono stati analizzati i vari algoritmi di previsione presenti in letteratura, e ne sono state individuate due categorie da sviluppare, la classificazione e la regressione, da cui sono stati scelti gli algoritmi Support Vector (SV) e Random Forest (RF). Sono stati analizzati due dataset provenienti da campagne di misura in edifici reali, uno situato in Italia e l'altro in Finlandia, facenti parte del progetto Horizon Europe (HE), ai quali sono stati poi applicati i modelli di previsione del comfort termico. I risultati ottenuti dalla previsione evidenziano una buona accuratezza dei modelli di classificazione (79% di accuratezza per il SV e 97% di accuratezza per il RF), che si sono rivelati più adatti rispetto a quelli di regressione, i cui risultati migliori, ottenuti da RF, hanno registrato un errore assoluto medio di previsione di 0,05. Una volta valutati i modelli, per misurarne la validità è stato considerato necessario applicarli ad un nuovo dataset: è stato quindi stilato un protocollo di misura per l'acquisizione di dati in merito alla sensazione termica degli utenti, all'interno dell'ambiente di ricerca. Sono stati coinvolti dieci partecipanti, distribuiti in quattro diverse stanze, ai quali è stato inviato un link con cadenza oraria per aprire il sondaggio a cui rispondere. All'interno delle stanze sono stati installati dei sensori per monitorare i dati ambientali e confrontarli con i feedback. Inoltre, attraverso un dataset estratto da un database globale per la raccolta di dati ambientali e delle sensazioni termiche, è stata condotta un'indagine statistica sull'influenza dei dati personali dei soggetti, quali età, peso e altezza, nella percezione termica degli stessi.

Indice

1. Introduzione	5
1.1 Background	5
1.2 Scopo dello studio	7
2. Stato dell'arte	8
2.1 Il comfort termico.....	8
2.2 Termoregolazione e bilancio energetico	9
2.3 Modelli di misura del comfort termico.....	11
2.3.1 PMV e PPD.....	11
2.3.2 Modelli di comfort personalizzati.....	12
2.3.3 Thermal Sensation Vote.....	13
2.4 Coaching nell'ambiente costruito	14
2.5 Applicazione del Machine Learning per la misura del comfort	15
2.5.1 Limiti dello stato dell'arte	16
3. Materiali e metodi	17
3.1 Dataset.....	18
3.2 Analisi e processamento dati.....	19
3.2.1 Analisi dati caso studio 1 (Focchi)	19
3.2.2 Analisi dati caso studio 2 (FVH)	24
3.3 Sviluppo ed applicazione algoritmi ML	29
3.3.1 Predizione TSV con classificatori	29
3.3.2 Predizione TSV con regressori	33
3.4 Metriche di valutazione.....	33
3.5 Sviluppo ed applicazione di protocollo di misura per raccolta feedback utenti ...	36
3.5.1 Protocollo di misura	36
3.6 Analisi database ASHRAE e ricerca correlazioni	43

4.	Risultati	45
4.1	Metriche di valutazione dei modelli ML.....	45
4.2	Analisi di correlazione del database ASHRAE	53
5.	Discussione	58
6.	Conclusione	59

1. Introduzione

1.1 Background

Il comfort termico è la condizione mentale in cui il soggetto esprime soddisfazione in merito alla sensazione termica percepita all'interno dell'edificio in cui si trova [1]. Questo rappresenta da tempo un argomento di grande rilevanza nell'ambito dell'ingegneria meccanica, poiché studi recenti confermano che le persone trascorrono oltre l'80% del loro tempo in ambienti chiusi [2]. Garantire condizioni ottimali di benessere negli spazi interni è quindi fondamentale, soprattutto per preservare la salute e la produttività degli individui, con particolare attenzione ai contesti lavorativi.

La gestione delle condizioni termiche negli edifici è affidata ai sistemi HVAC (Heating, Ventilation, and Air Conditioning), che rappresentano una delle principali fonti di consumo energetico, utilizzando una quota rilevante dell'energia totale degli edifici. Pur essendo essenziali per garantire il comfort degli occupanti, il loro funzionamento può essere ottimizzato per ridurre gli sprechi energetici. Dalla letteratura emergono diverse strategie [3][4], basate su modelli adattivi per il comfort termico o l'ottimizzazione dei set-point di temperatura che possono diminuire i consumi senza compromettere il benessere interno. Inoltre, l'introduzione di tecnologie avanzate basate sull'intelligenza artificiale (IA), offre soluzioni innovative per migliorare l'efficienza energetica nell'ambiente costruito [5].

Il comfort termico è regolato da normative internazionali che definiscono standard di progettazione e gestione dei sistemi HVAC, nonché criteri per la misura e valutazione del benessere termico negli edifici. Tra le norme più importanti a livello globale vi sono la ISO 7730 [6] e gli standard ASHRAE [1]. La prima fornisce i metodi per la previsione del comfort termico globale e il grado di insoddisfazione termica in ambienti termici moderati, attraverso la definizione delle formule per il calcolo degli indici di Predicted Mean Vote (PMV) e Predicted Percentage of dissatisfied (PPD), che verranno illustrati nei capitoli successivi. L'obiettivo della seconda normativa consiste nel definire le combinazioni di parametri ambientali e personali che garantiscono un livello di comfort termico accettabile per la maggior parte degli individui coinvolti.

Per far sì che sia garantito uno stato accettabile di comfort, col tempo sono stati sviluppati dei metodi efficaci per la misura di quest'ultimo. Ad oggi, i modelli di misura del comfort termico si suddividono in due approcci: quello stazionario e quello adattivo [7]. Il primo si basa sul principio del bilancio energetico che regola gli scambi di calore del corpo umano; il secondo sfrutta alcune tecniche che mettono in relazione i fattori ambientali interni accettabili con i fattori ambientali e metereologici esterni.

Come già detto sopra, la normativa ASHRAE 55 [1] identifica il comfort termico come uno stato mentale del soggetto. In questi termini, la componente soggettiva ricopre un ruolo di grande rilevanza nella percezione delle condizioni ambientali da parte dell'individuo. La definizione stessa sottolinea l'importanza di considerare di fattori fisici, fisiologici, psicologici e di altra natura nel processo di formulazione del responso in merito al comfort termico [8]. Le condizioni ambientali necessarie per assicurare il comfort non sono quindi uguali per tutti, ma variano in base alle caratteristiche personali e alle percezioni individuali. Tutto ciò rende complesso garantire condizioni di comfort che soddisfino tutti gli occupanti di uno spazio [1].

Un metodo efficace per ottenere le condizioni ambientali che soddisfino la maggioranza degli utenti, è la periodica raccolta di feedback espressi dai soggetti coinvolti [8]. Una delle modalità più comuni per raccogliere il feedback consiste nel sottoporli a sondaggi sulla sensazione termica, utilizzando una scala a sette livelli (da -3, freddo, a +3, caldo, con 0 neutralità termica), definita dalla normativa ASHRAE Standard 55 [1], per misurare il comfort termico attraverso il Thermal Sensation Vote (TSV) [1]. Questo voto soggettivo permette di esprimere una valutazione del comfort percepito. L'utilizzo di sondaggi è inoltre utile al miglioramento dell'efficienza energetica dei sistemi HVAC, dato che impostare parametri ambientali con criterio contribuisce alla riduzione del consumo energetico. Tuttavia, richiedere feedback periodici può risultare invasivo dal punto di vista dell'occupante, soprattutto in ambito lavorativo, dove può interferire con la concentrazione e le attività [9].

Per superare questo limite, una possibile soluzione consiste nell'elaborare modelli di previsione del comfort termico [10], mirati a stimare il TSV a partire da dati ambientali, fisiologici e dai feedback raccolti in un certo arco temporale sia all'interno sia all'esterno

degli edifici. I modelli di previsione del comfort termico si basano sull'utilizzo di algoritmi di Machine Learning [10], una tecnologia avanzata nell'ambito dell'intelligenza artificiale che si occupa dello sviluppo di algoritmi e modelli statistici in grado di analizzare dati, identificare schemi e prendere decisioni o effettuare previsioni. Nel contesto specifico, tali algoritmi vengono impiegati per prevedere il feedback degli utenti in termini di TSV. Attraverso l'analisi di dati ambientali e fisiologici raccolti nel tempo, gli algoritmi di Machine Learning permettono di stimare con precisione i valori di TSV. Questi valori, a loro volta, sono associati a specifiche configurazioni dei parametri ambientali, i quali possono essere implementati nei sistemi HVAC per garantire condizioni termiche ottimali. In questo modo, è possibile non solo migliorare il benessere termico degli utenti, ma anche ottimizzare il consumo energetico degli edifici.

1.2 Scopo dello studio

Il presente studio si propone di approfondire le metodologie esistenti per la previsione del comfort termico, misurato attraverso il TSV, il quale corrisponde ad un indice di sensazione termica degli utenti. Partendo quindi dallo stato dell'arte ed i suoi limiti, l'obiettivo principale è sviluppare un modello innovativo di previsione della sensazione termica degli utenti, attraverso l'utilizzo di algoritmi di Machine Learning (ML). Dopo aver analizzato dei dataset reali contenenti dati relativi a parametri ambientali e TSV, raccolti attraverso campagne di acquisizione all'interno di due diversi edifici, situati in zone climatiche differenti, sono stati infatti scelti e applicati alcuni algoritmi di ML per la previsione del TSV, valutati poi attraverso le metriche che verranno di seguito illustrate. In questo lavoro è stato poi stilato un protocollo di misura per l'acquisizione del feedback degli utenti all'interno dell'ambiente in cui è stata condotta la ricerca, così da ottenere un nuovo dataset su cui verificare la validità dei modelli implementati. Infine, a partire da un dataset estratto dal database online "ASHRAE global thermal comfort database", è stata condotta un'analisi statistica dell'andamento del TSV fornito dagli utenti a confronto con i loro dati personali, quali età, peso e altezza, e con i dati ambientali di temperatura, per osservare l'influenza di questi parametri nella percezione termica e ricercare un'eventuale correlazione.

2. Stato dell'arte

2.1 Il comfort termico

Secondo la normativa ASHRAE 55, il comfort termico è definito come “quella condizione mentale che esprime soddisfazione in merito all'ambiente termico ed è stimata attraverso una valutazione soggettiva” [1]. Questa definizione sottolinea il ruolo centrale della componente soggettiva nella raccolta dei dati relativi al comfort termico, poiché la percezione della sensazione termica varia in modo unico e personale per ciascun individuo.

Negli ultimi decenni, tuttavia, lo sviluppo tecnologico, e in particolare i progressi nel campo dell'intelligenza artificiale, hanno permesso di superare alcune delle limitazioni legate alla soggettività. In questi termini, sono stati sviluppati modelli basati sul Machine Learning [11], capaci di analizzare e interpretare i dati relativi alla sensazione termica in termini oggettivi. Questi modelli permettono di elaborare informazioni utili per identificare i parametri ambientali ottimali, che possono essere utilizzati per regolare in modo efficace i sistemi HVAC. L'utilizzo di tali approcci rappresenta un passo avanti significativo nella gestione del comfort termico, consentendo di integrare le esigenze individuali con soluzioni tecniche efficienti e sostenibili.

Per la definizione del comfort termico, c'è bisogno di una serie di fattori ambientali e personali che influenzano la percezione dell'individuo. La misurazione dei parametri ambientali segue le indicazioni della norma UNI EN ISO 7726 [12], che indica i requisiti per gli strumenti di misura. La norma ASHRAE standard 55 [1] definisce invece le condizioni ambientali che corrispondono ad un comfort termico accettabile per la maggior parte degli utenti coinvolti.

Parametri ambientali

- Temperatura dell'aria: si riferisce alla temperatura dell'aria che circonda le persone all'interno della zona di interesse.
- Temperatura media radiante: è la temperatura media registrata sulle superfici circostanti all'area di misurazione, che comprende muri, finestre, soffitto e pavimento. Anche la radiazione termica influenza la sensazione termica percepita.
- Umidità relativa: definita come la quantità di vapore acqueo presente nell'aria divisa per la quantità massima che essa può contenere per unità di volume ad una certa temperatura. Rappresenta anch'essa un parametro fondamentale per il benessere dell'individuo, poiché influenza la qualità dell'aria respirata.
- Velocità dell'aria: corrisponde alla velocità dell'aria all'interno dello spazio preso in considerazione, ed è importante per ottenere un buon ricircolo dell'aria all'interno della stanza.

Parametri personali

- Metabolismo (*Met*): è un parametro che descrive il flusso energetico generato dal corpo umano dovuto all'attività metabolica, misurata in "met", la cui unità, ovvero 1 met, corrisponde all'energia prodotta per unità di superficie in media da una persona seduta a riposo, ovvero 58.2 W/m^2 . Una persona occupa in media una superficie di 1.8 m^2 . Queste informazioni si rifanno alla normativa ASHRAE [1].
- Isolamento termico dovuto all'abbigliamento (*Icl*): misura la resistenza al trasferimento di calore sensibile corrispondente ad un capo di abbigliamento, e viene espresso in unità "clo". Anche questa definizione proviene dalla normativa ASHRAE [3].

2.2 Termoregolazione e bilancio energetico

La sensazione termica varia nel tempo, poiché il corpo umano possiede la capacità di adattamento alle condizioni ambientali, di conseguenza ne risente anche il comfort percepito.

La variazione della sensazione termica è dovuta ai meccanismi di termoregolazione che il corpo attiva per reagire ai cambiamenti ambientali, risultando fondamentale per il mantenimento dell'omeostasi fisiologica sia in condizioni di riposo che in movimento [13]. Numerosi studi dimostrano che all'origine dei meccanismi di termoregolazione ci sia l'attività del sistema nervoso [14], in particolare dal sistema nervoso simpatico, che riceve segnali dai termorecettori [15] e provoca la risposta del sistema circolatorio [16].

Allo stesso tempo, anche il tasso metabolico viene regolato dal sistema termoregolatore [17]: si può dire allora che esso dovrebbe essere considerato una variabile dinamica, strettamente legata alle caratteristiche fisiologiche individuali, piuttosto che una costante [18]. Questi processi influenzano in maniera diretta la temperatura corporea e della pelle, e ciò ricorda ancora una volta l'importanza di un'analisi integrata dei parametri fisiologici per comprendere e prevedere la risposta del corpo umano alle variazioni termiche [19].

Il bilancio energetico [20] del corpo umano si basa sul primo principio della termodinamica, secondo cui l'energia può trasformarsi passando da una forma all'altra ma non può essere creata o distrutta. L'equazione $E_s = E_i - E_o$ rappresenta l'applicazione di questo principio ai flussi energetici corporei, dove E_s rappresenta la variazione dell'energia immagazzinata nel corpo, E_i è l'energia in ingresso mentre E_o quella in uscita. L'energia in ingresso (E_i) deriva principalmente dall'energia chimica ottenuta dal nutrimento del corpo attraverso l'introduzione di alimenti, mentre l'energia in uscita (E_o) comprende la perdita di calore per irraggiamento, conduzione e convezione, il lavoro svolto dal corpo e il calore latente disperso attraverso l'evaporazione. La variazione di energia del corpo (E_s) riflette le variazioni dei valori dei macronutrienti presenti nel corpo, dimostrando come il principio di conservazione dell'energia sia perfettamente applicabile ai processi fisiologici umani.

Le condizioni ambientali, come temperatura, umidità e velocità del vento, influenzano in maniera significativa il meccanismo della termoregolazione ed il bilancio energetico. In situazioni caratterizzate da alte temperature ed umidità elevate, il corpo potrebbe risentire di questi effetti diminuendo la propria capacità di cedere calore, provocando problemi di salute.

Al contrario, a basse temperature, il corpo deve impostare il proprio meccanismo di termoregolazione per mantenere la temperatura corporea, richiedendo una maggiore quantità di apporto energetico.

2.3 Modelli di misura del comfort termico

A partire dagli anni Settanta, il tema del comfort termico ha attirato crescente attenzione nella comunità scientifica, portando allo sviluppo di modelli specifici per la valutazione e la previsione del benessere termico. Questi modelli, basati sulla percezione termica e sui parametri ambientali, rappresentano un punto di riferimento fondamentale per comprendere e migliorare l'interazione tra le persone e il loro ambiente [21]. Nella sezione che segue verranno illustrati i principali modelli di misura del comfort termico, evidenziandone l'importanza all'interno della ricerca nel campo del benessere negli ambienti lavorativi.

2.3.1 PMV e PPD

Il Predicted Mean Vote (PMV) è definito, secondo la normativa UNI ISO 7730 [6] e ASHRAE standard 55 [1], come “l'indice che predice il valore medio dei voti di un ampio gruppo di persone in una scala di sette valori che misura la sensazione termica”.

+3	hot
+2	warm
+1	slightly warm
0	neutral
-1	slightly cool
-2	cool
-3	cold

Figura 1: scala a sette valori – ASHRAE standard 55 [1]

Questo modello di misura del comfort è stato sviluppato da Fanger nel 1970 [22], e rappresenta una formalizzazione della percezione soggettiva del comfort termico, sviluppata sulla base di dati empirici ottenuti da studi condotti su un ampio campione di individui. Questo indice consente di stimare la media dei feedback sul comfort termico espressi da un gruppo di persone. L'equazione per il calcolo del PMV si trova all'interno della normativa UNI EN ISO 7730 [6] e necessita della conoscenza di sei parametri, quali il tasso metabolico, l'isolamento termico dovuto all'abbigliamento, la temperatura dell'aria, la temperatura media radiante, la velocità dell'aria e l'umidità dell'aria.

Il Predicted Percentage of Dissatisfied (PPD) è definito dalla normativa ASHRAE come “un indice che stabilisce una predizione quantitativa della percentuale delle persone termicamente insoddisfatte determinata con il PMV” [1]. Secondo la normativa UNI EN ISO 7730 [6], le persone che sono insoddisfatte della loro sensazione termica sono quelle che, come risposta al feedback, forniranno valori corrispondenti a troppo caldo (hot), caldo (warm), freddo (cool) e troppo freddo (cold), nella scala a sette valori illustrata sopra. Anche in questo caso la norma 7730 fornisce la formula per il calcolo di questo indice, che dipende esclusivamente dal PMV.

$$PPD = 100 - 95 \cdot \exp(-0,033\,53 \cdot PMV^4 - 0,217\,9 \cdot PMV^2)$$

Figura 2: equazione del PPD secondo la normativa UNI EN ISO 7730 [6]

Dato che il PPD è dipendente dall'indice PMV, è possibile ottenere una relazione tra le due variabili che può essere mostrata attraverso il grafico presente nella Figura 3.

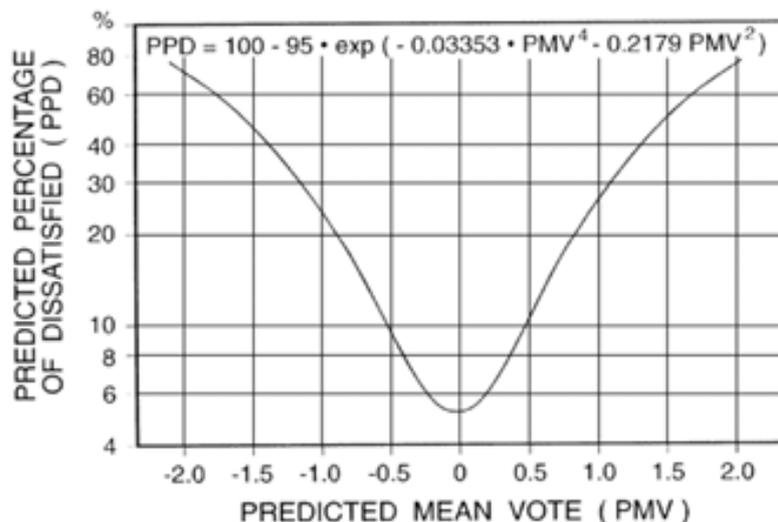


Figura 3: andamento del PPD in funzione del PMV, norma ASHRAE standard 55 [1]

2.3.2 Modelli di comfort personalizzati

Allo stato attuale degli studi, esistono due categorie di modelli per la misura del comfort termico: l'approccio allo stato stazionario e l'approccio adattivo [7]. L'esempio di maggior rilevanza del primo gruppo è quello di Fanger [22], il quale inizia raccogliendo dati sperimentali riferiti ad un gruppo di studenti universitari all'interno di una camera

climatica. Questo tipo di approccio sfrutta il principio del bilancio termico, ed è attualmente il modello di riferimento su cui si basa la misura del comfort termico degli utenti all'interno degli edifici. Fanger, durante la prova, ha fatto in modo che le variazioni di temperatura all'interno della camera fossero minime, mentre agli utenti coinvolti è stato chiesto di indossare un abbigliamento standardizzato. Considerata la difficoltà che spesso emerge nel raccogliere dati in merito ai parametri definiti da Fanger, studi successivi hanno sviluppato un modello semplificato del PMV, basato sulla conoscenza dei valori di temperatura dell'aria e umidità relativa: ne è un esempio quello ideato da Buratti et al. [23], che verrà applicato in questo lavoro. Rispetto ai modelli a stato stazionario, i modelli adattivi offrono un approccio alternativo basato su dati raccolti tramite studi sul campo. Questi modelli sfruttano le regressioni lineari, che mettono in correlazione temperatura operativa interna accettabile e temperatura esterna o altri fattori meteorologici. Tale approccio risulta particolarmente utile per gli edifici a ventilazione naturale, poiché tiene conto delle variazioni ambientali esterne e dell'adattamento degli occupanti [7].

2.3.3 Thermal Sensation Vote

Il Thermal Sensation Vote (TSV) è un indice che misura la sensazione termica basato sulla scala a sette valori dell'ASHRAE [1], rappresentata nella Figura 1, e viene raccolto attraverso un sondaggio [8]. Questo sondaggio viene somministrato agli occupanti di un edificio, i quali esprimono un giudizio sulla propria percezione di comfort termico. Il TSV, quindi, rappresenta un indice interamente soggettivo, poiché riflette la percezione termica individuale che, come discusso in precedenza, è influenzata dalla combinazione di fattori di tipo ambientale, fisiologico e psicologico [1], i quali contribuiscono a determinare la risposta di ogni individuo alle condizioni termiche dell'ambiente. Il sondaggio viene fatto solitamente definendo un numero preciso di partecipanti, a cui vengono richieste informazioni personali, principalmente in merito al genere, al peso, all'altezza e all'età. Viene inoltre scelto un luogo e un periodo specifici, oltre che un intervallo di tempo secondo cui richiedere il feedback degli utenti. Durante il periodo in questione, vengono registrati dati ambientali come temperatura, umidità, velocità dell'aria, correlabili alle percezioni termiche dei soggetti [9].

2.4 Coaching nell'ambiente costruito

Il concetto di Coaching [24] nell'ambito del benessere termico si riferisce ad un approccio mirato a guidare e supportare le persone a migliorare la propria condizione di comfort termico, attraverso suggerimenti in merito a parametri ambientali da impostare, ad esempio con la regolazione del termostato, o ad azioni da compiere, come aprire o chiudere la finestra, utilizzare un certo abbigliamento o adottare un certo stile di vita all'interno dell'edificio. Questo metodo sfrutta in primis la raccolta dei dati ambientali ed i feedback degli utenti che ne derivano, i quali per mezzo di tecniche sviluppate vengono analizzati, in modo da ottenere delle correlazioni tra i valori dei parametri dell'ambiente e la percezione termica, capire quali situazioni corrispondono a quelle ottimali, e fornire infine all'utente le indicazioni per raggiungerle. Un esempio significativo di applicazione del coaching per l'ottimizzazione del comfort è descritto nell'articolo [24], in cui viene presentato un metodo innovativo per la definizione della conversazione tra un coach virtuale e persone anziane. Questo approccio si basa sulla raccolta e sull'analisi di dati relativi ai parametri che determinano la qualità ambientale interna (Indoor Environmental Quality, IEQ), integrati con l'utilizzo di grafi di conoscenza (Knowledge Graphs, KG). Un altro esempio è presente all'interno dell'articolo [25], dove viene sviluppato un sistema di misurazione composto da un robot sociale ed alcuni sensori per valutare il comfort termico, così da ottenere un sistema intelligente che, attraverso il coaching, guidi il soggetto attraverso consigli per raggiungere il comfort termico.

Dunque, l'obiettivo principale del metodo di coaching è quello di porre l'utente al centro del processo, offrendo soluzioni tecnologiche che consentano di migliorare il proprio stato di comfort termico e, più in generale, il benessere complessivo nell'ambiente domestico. Questo approccio non solo promuove una maggiore consapevolezza delle condizioni ambientali, ma facilita anche l'adozione di strategie mirate per ottimizzare il comfort in modo semplice e accessibile.

2.5 Applicazione del Machine Learning per la misura del comfort

Un ulteriore passo in avanti nell'ambito del comfort termico è rappresentato dall'introduzione del machine learning (ML), una branca dell'intelligenza artificiale (IA) che consente al sistema di identificare schemi e fare previsioni in diversi ambiti dell'ingegneria. Il Machine Learning ha rappresentato una svolta nello sviluppo dei modelli personalizzati di misura del comfort. Questo sviluppo si è reso necessario poiché, come precedentemente discusso, il TSV è un indice soggettivo del comfort termico, variabile da persona a persona ed influenzato da molteplici fattori ambientali, fisiologici ma anche psicologici.

La previsione del TSV attraverso l'utilizzo di queste tecniche aiuta a rendere questo indice più oggettivo, il che può essere utile per stabilire le condizioni ottimali per cui si massimizza il comfort termico. Inoltre, la periodica richiesta di feedback agli utenti potrebbe risultare invasiva in ambiente lavorativo, poiché rispondere ad un sondaggio è oggetto di distrazione delle persone, che distolgono l'attenzione dal proprio compito, rischiando così di influenzarle psicologicamente ed alterare perciò la loro percezione termica. Un esempio di applicazione del machine learning nell'ambito del comfort termico si può riscontrare nell'articolo [10], dove i dati, provenienti dal database ASHRAE RP-884 Adaptive Model Project, una volta analizzati e processati attraverso tecniche di pulizia ed espansione del dataset per il bilanciamento dati, sono stati sottoposti ad un modello di classificazione chiamato Support Vector Classification (SVC) [10], il quale ha fornito come risultato una previsione del TSV. Questo dataset comprendeva sia parametri ambientali che fisiologici, oltre al feedback degli utenti espresso come TSV. Un altro esempio di Machine Learning per la previsione del comfort lo si ritrova all'interno dell'articolo [11]. In questo studio vengono messe a confronto molteplici categorie di tecniche di machine learning utilizzando dei dati ambientali e fisiologici raccolti prima da dispositivi avanzati e più costosi, poi da dispositivi più accessibili ma meno accurati, per ricercare il corretto bilanciamento tra costi e accuratezza della previsione. Un ulteriore esempio si trova all'interno dell'articolo [26], dove i dati fisiologici di ciascun soggetto ed i dati ambientali sono stati raccolti per mezzo di sensori, poi, attraverso il dataset, è stato pre-addestrato un modello ML, combinando algoritmi di Deep Learning (DL) [26] e machine learning. Dal modello addestrato, è stata applicata la tecnologia di ensemble

Transfer Learning (TL), utilizzata per superare i limiti di performance dei modelli di previsione singoli. Il machine learning è quindi diventato parte integrante della ricerca nell'ambito del comfort termico, poiché ha introdotto un nuovo modo di affrontare l'argomento, ottenendo notevoli risultati.

2.5.1 Limiti dello stato dell'arte

Nonostante i significativi progressi ottenuti con l'introduzione dell'ML nella previsione del comfort termico, permangono alcune criticità che questo studio si propone di minimizzare. Uno dei possibili limiti riguarda l'uso dei parametri fisiologici, i quali rappresentano un elemento fondamentale per la misura del comfort [27].

Tuttavia, la raccolta di questi dati non sempre si adatta alla disponibilità degli utenti, sia per l'assenza di strumenti adeguati sia perché molti dispositivi richiedono un'interazione diretta, come l'utilizzo di sensori indossabili, un aspetto che può essere percepito come invasivo [28]. Questo potrebbe alterare la percezione del comfort da parte del soggetto, compromettendo l'accuratezza dei dati. Inoltre, la richiesta periodica di feedback sulla sensazione termica risulta invasiva dal punto di vista dell'occupante, soprattutto in ambito lavorativo, dove può diminuire la concentrazione del soggetto [9].

Per ovviare a questa limitazione, in questo studio si è scelto di utilizzare esclusivamente parametri ambientali, come temperatura e umidità, al fine di sviluppare un modello di previsione meno intrusivo ma comunque efficace. Un'ulteriore criticità evidenziata dalla letteratura riguarda la necessità di disporre di dataset di ampie dimensioni. Questo requisito nasce dal fatto che la percezione termica varia significativamente da individuo a individuo [8], rendendo indispensabile una grande varietà di dati per garantire un adeguato addestramento del modello. Tuttavia, nei contesti di ricerca come quello di questo studio, l'obiettivo è sviluppare una soluzione personalizzata, basata su dati ambientali raccolti in casi studio reali. L'intento di questa ricerca è quindi quello di individuare un modello predittivo che sia applicabile a scenari quotidiani, affrontando i limiti discussi e garantendo un compromesso tra la semplicità di implementazione, la personalizzazione e l'accuratezza nella previsione del comfort termico.

3. Materiali e metodi

La metodologia sviluppata nel lavoro di tesi ambisce ad applicare algoritmi di ML per la misura del comfort termico in termini di TSV, partendo da parametri ambientali raccolti in edifici reali adibiti ad ufficio. Nello specifico, nella prima fase dello studio, sono stati analizzati i dati ambientali relativi a due edifici distinti situati in Europa. Il primo edificio è situato nel centro Italia, mentre il secondo edificio di uffici è situato in Finlandia. In primo luogo, sono stati selezionati i parametri utili per lo studio corrente, cioè la temperatura interna (°C), la temperatura esterna (°C), l'umidità relativa interna (%) e il TSV. Successivamente, i dati sono stati processati e visualizzati graficamente, ed è stata condotta un'analisi di correlazione tra i vari parametri a disposizione (Figura 4).

Terminata la fase di analisi, il lavoro è proseguito con l'applicazione di diversi algoritmi ML ai dataset ottenuti dall'analisi. La valutazione dei modelli è stata fatta attraverso metriche specifiche che verranno illustrate in seguito in un paragrafo dedicato.

Il passo successivo dello studio ha riguardato lo sviluppo di un protocollo di misura finalizzato alla raccolta di feedback da parte degli utenti, con particolare attenzione al loro voto di sensazione termica (TSV). Questo protocollo è stato progettato per raccogliere dati ambientali all'interno della struttura in cui si svolgeva l'attività lavorativa degli utenti e acquisire informazioni sul loro comfort termico e globale. L'obiettivo principale era la creazione di un nuovo dataset completo e strutturato, adatto a testare i modelli di machine learning selezionati e sviluppati nel corso dello studio.

Infine, è stata condotta un'indagine statistica su dati relativi alla sensazione termica umana forniti da un database online (database ASHRAE), per studiare le diverse correlazione tra parametri personali come età, peso, altezza e sensazione termica dei soggetti. Lo scopo di quest'ultima fase era valutare l'importanza dei dati personali rispetto alla percezione termica degli utenti.

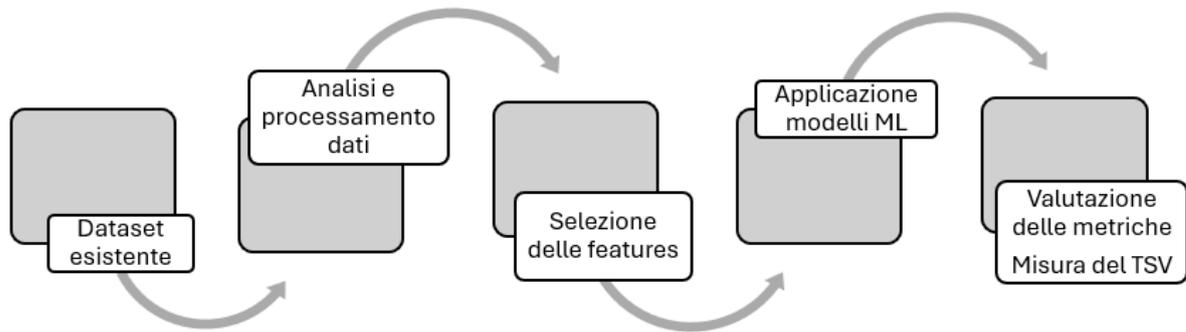


Figura 4: Schema riassuntivo della metodologia applicata.

3.1 Dataset

Come già anticipato, i dataset su cui si è basato il lavoro di analisi e applicazione dei modelli di previsione sono due, i quali, provenendo da acquisizioni in edifici di zone climatiche differenti, le loro caratteristiche e i dati misurati contengono informazioni molto diverse. Il primo contiene dati raccolti all'interno di un edificio situato in Italia, che viene denominato 'Focchi', dal nome dell'azienda di cui è sede, la quale fa parte del progetto 'Horizon Europe'.

Questo dataset, mostrato nella Figura 5, contiene informazioni in merito a temperatura esterna (°C), temperatura interna (°C), umidità relativa interna (%) e il feedback degli utenti raccolto come TSV.

date	outdoor T	indoor T	indoor RH	id	pilot	building	building_id	floor	room	TSV
2024-01-14 23:00:04	3.06	16.29	41.8	533	Focchi	Head Quarter	2	1	DR1	-1
2024-01-14 23:01:04	3.12	16.29	41.58	533	Focchi	Head Quarter	2	1	DR1	-1
2024-01-14 23:02:05	3.12	16.24	41.77	533	Focchi	Head Quarter	2	1	DR1	-1
2024-01-14 23:03:05	3.12	16.24	41.63	533	Focchi	Head Quarter	2	1	DR1	-1
2024-01-14 23:04:05	3.18	16.4	41.55	533	Focchi	Head Quarter	2	1	DR1	-1
2024-01-14 23:05:05	3.18	16.24	41.55	533	Focchi	Head Quarter	2	1	DR1	-1

Figura 5: Dataset 'Focchi'

Il secondo dataset (Figura 6) proviene da una campagna di raccolta Feedback svolta ad Helsinki, in Finlandia, in un edificio denominato 'FVH', anch'esso sede di un gruppo, il 'Forum Virium Helsinki', facente parte del progetto Horizon Europe.

Le informazioni disponibili in questo caso riguardano temperatura esterna (°C), temperatura interna (°C), umidità relativa interna (%), concentrazione CO₂ (parti per milione, ppm) e consumo energetico (kW).

date	building	room_x	floor_x	section_x	indoor T_y	ndoor RH_1	outT	value	TSV
2024-01-15 10:20:00	KYMP Building	333	4	B	20.833	13	-8.722	2.041	0
2024-01-15 10:40:00	KYMP Building	333	4	B	21.833	14	-8.889	2.041	0
2024-01-15 11:00:00	KYMP Building	333	4	B	22.222	14	-8.889	2.56833	0
2024-01-16 15:40:00	KYMP Building	333	4	B	21.333	7	-16.611	1.898	0
2024-01-16 16:00:00	KYMP Building	333	4	B	21.222	7	-16.778	1.66667	0
2024-01-16 16:20:00	KYMP Building	333	4	B	21.056	7	-17.056	1.66667	0

Figura 6: Dataset 'FVH'

3.2 Analisi e processamento dati

3.2.1 Analisi dati caso studio 1 (Focchi)

Vedremo in questo paragrafo il procedimento seguito nell'analisi e trattamento dei dati corrispondenti al dataset denominato Focchi, riferiti ad un edificio situato in Italia.

Come già anticipato, i parametri presenti a disposizione sono temperatura esterna, temperatura interna, umidità relativa interna e TSV.

In primo luogo, sono stati selezionati solo i dati d'interesse. Successivamente, sono stati estratti i dati corrispondenti ad una sola stanza dell'edificio, la stanza 'DR1', applicando un filtro alla colonna del dataset 'room'. È stato ottenuto un dataset (Figura 7) di 10535 samples dei parametri sopracitati, relativo al periodo di acquisizione che va dal 14 gennaio 2024 ore 23:00:04, al 2 febbraio 2024 ore 11:01:24, con una frequenza di campionamento di un minuto. È poi stata eseguita un'analisi approfondita per identificare outlier (valori fuori scala), duplicati e Nan (valori non registrati) all'interno del dataset, ma non sono state trovate anomalie.

date	outdoor T	indoor T	indoor RH	TSV
2024-01-14 23:00:04	3.06	16.29	41.8	-1
2024-01-14 23:01:04	3.12	16.29	41.58	-1
2024-01-14 23:02:05	3.12	16.24	41.77	-1
2024-01-14 23:03:05	3.12	16.24	41.63	-1
2024-01-14 23:04:05	3.18	16.4	41.55	-1
2024-01-14 23:05:05	3.18	16.24	41.55	-1

Figura 7: dataset 'Focchi' filtrato e analizzato

Nelle figure 8 e 9 vengono mostrati gli andamenti dei parametri ambientali e del TSV.

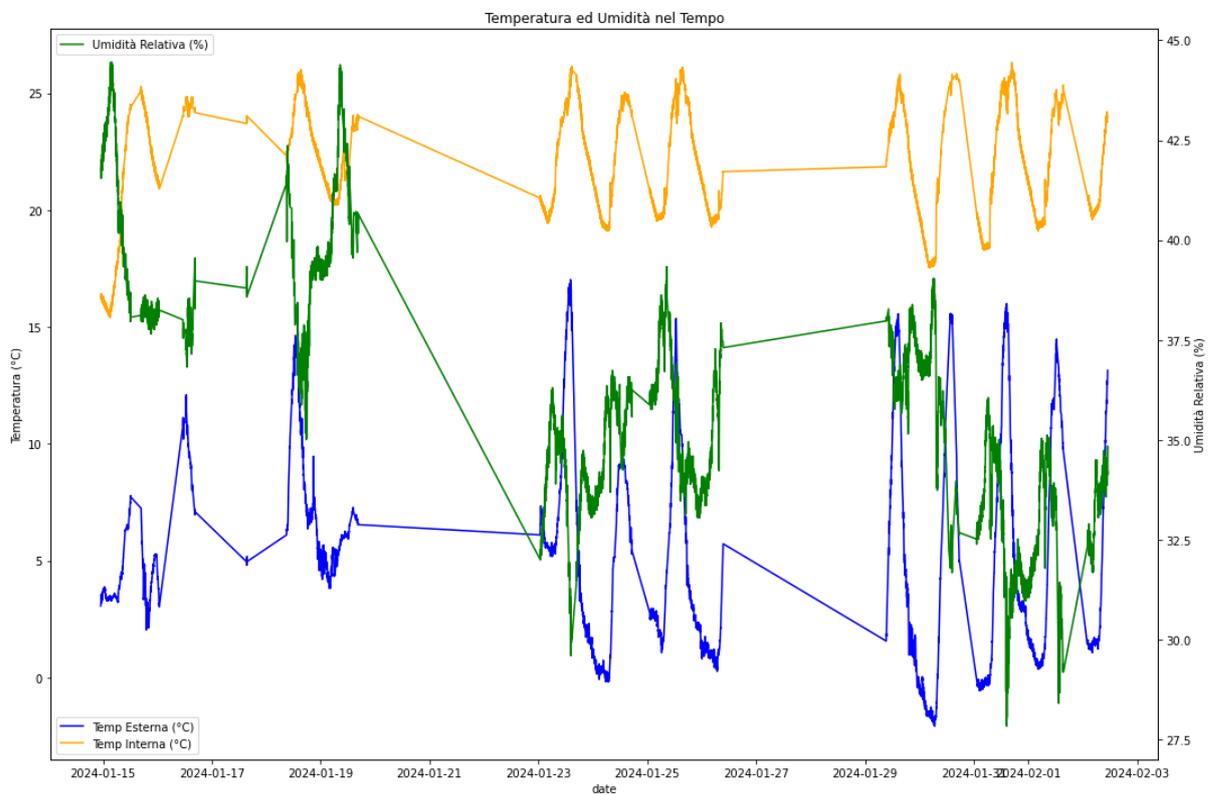


Figura 8: Andamento della temperatura interna, esterna e umidità relativa interna

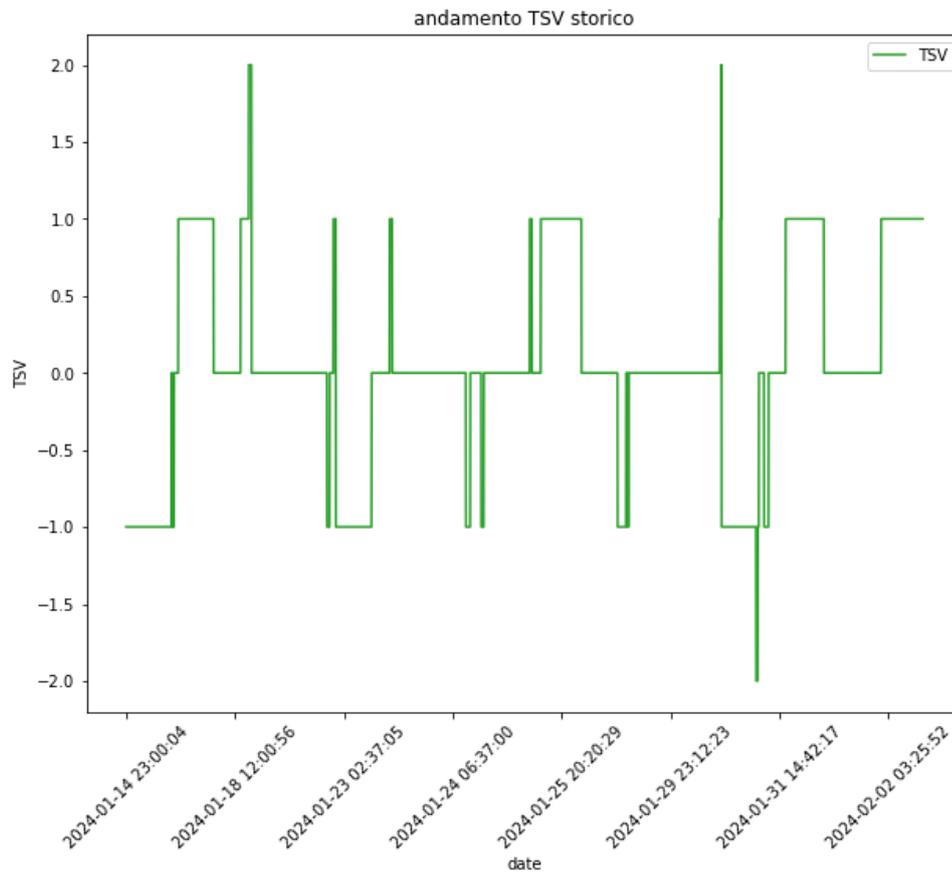


Figura 9: andamento del TSV

Dopodiché è stata fatta un'analisi di correlazioni tra i vari parametri per capire se ci fosse una relazione tra gli andamenti dei parametri ambientali e il feedback degli utenti. Per fare ciò è stata costruita una matrice di correlazione, illustrata in Figura 10.

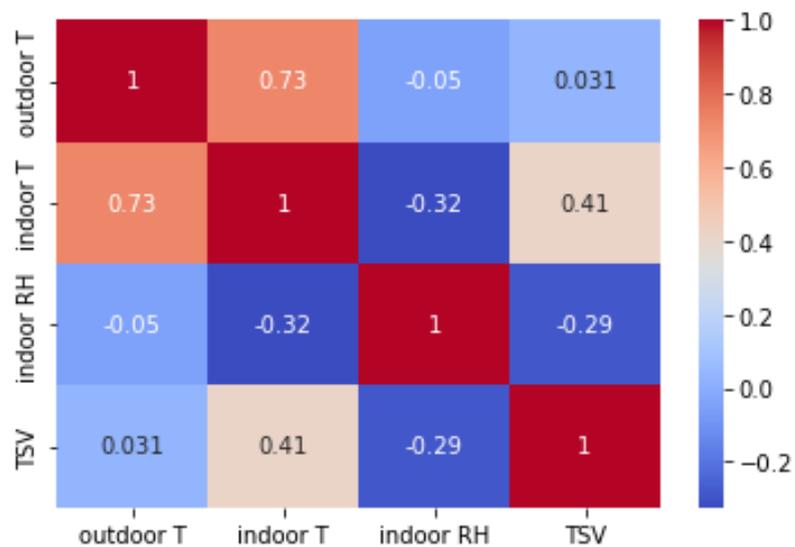


Figura 10: heatmap della matrice di correlazione tra i parametri ambientali e il TSV

Da come si può notare, la correlazione più alta per il TSV la si riscontra con la temperatura interna, mentre la correlazione più alta tra i parametri ambientali è quella tra temperatura interna ed esterna.

Per ottenere un ulteriore parametro da confrontare con il TSV, è stato applicato un modello di PMV semplificato (sPMV) secondo il metodo sviluppato da Buratti et al. [23], che utilizza unicamente i parametri di temperatura interna e umidità relativa interna per il calcolo del PMV. Viene di seguito mostrato l'andamento.

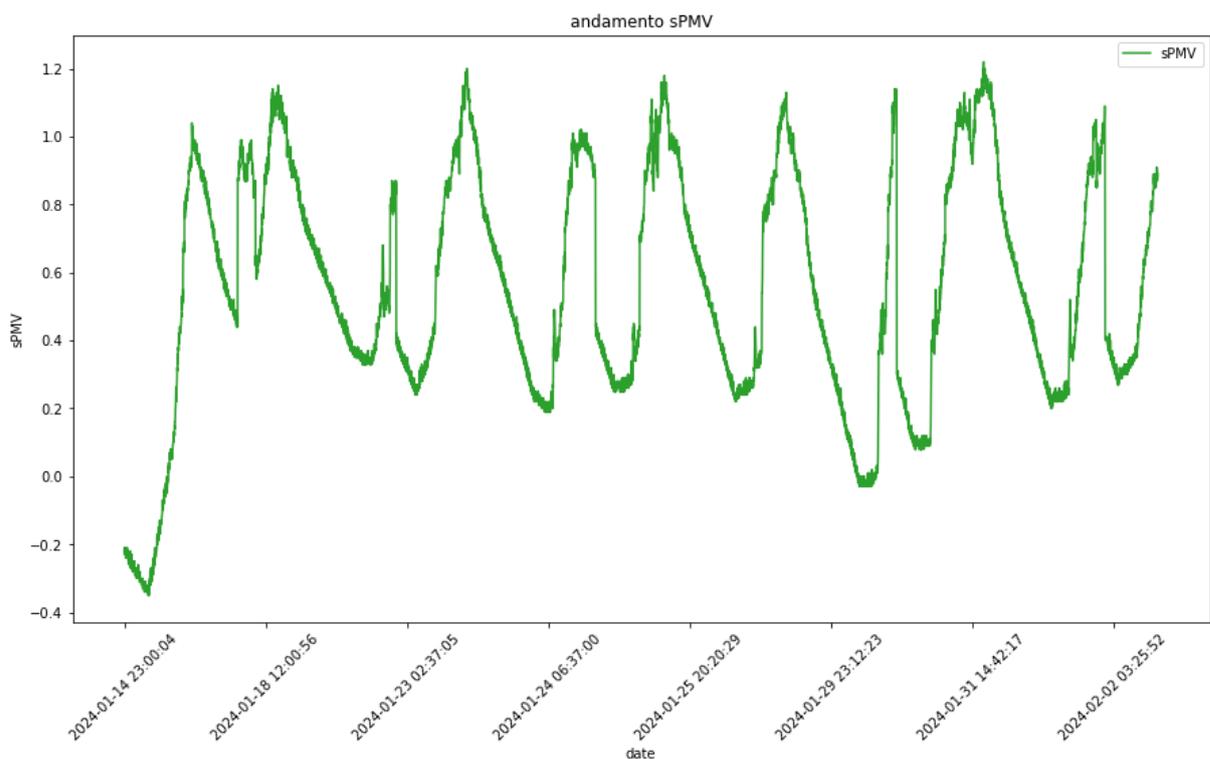


Figura 11: andamento del sPMV nel tempo

È stata quindi aggiunta la colonna contenente i valori sPMV al dataset (Figura 12), e successivamente è stata ricostruita la matrice di correlazione (Figura 13) ed è stato messo in evidenza l'andamento del sPMV a confronto con il TSV (Figura 14).

date	sPMV	ated n_sens	applied_model	outdoor T	indoor T	indoor RH	TSV
2024-01-14 23:00:04	-0.22	533	simplified pmv - w	3.06	16.29	41.8	-1
2024-01-14 23:01:04	-0.22	533	simplified pmv - w	3.12	16.29	41.58	-1
2024-01-14 23:02:05	-0.23	533	simplified pmv - w	3.12	16.24	41.77	-1
2024-01-14 23:03:05	-0.23	533	simplified pmv - w	3.12	16.24	41.63	-1
2024-01-14 23:04:05	-0.21	533	simplified pmv - w	3.18	16.4	41.55	-1
2024-01-14 23:05:05	-0.23	533	simplified pmv - w	3.18	16.24	41.55	-1

Figura 12: dataset aggiornato con colonna sPMV

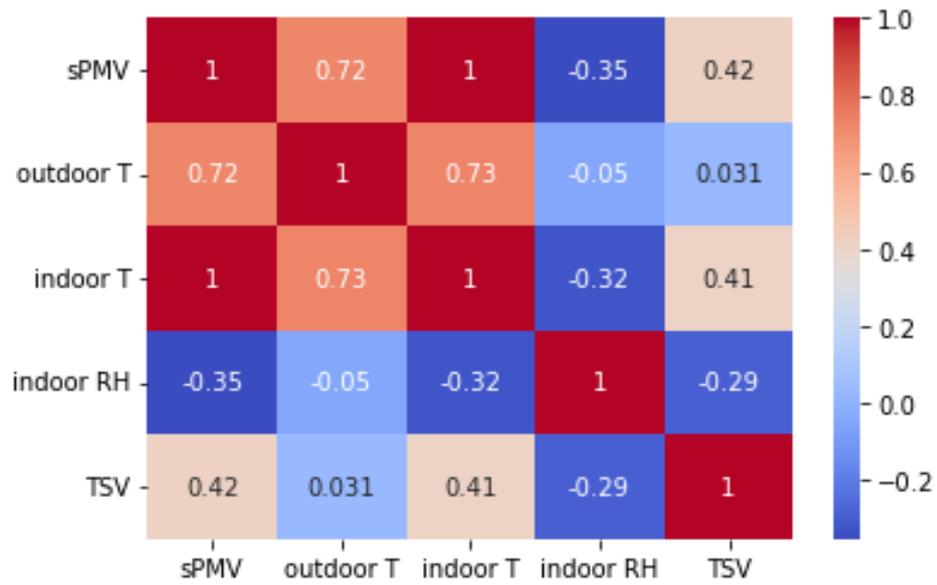


Figura 13: matrice di correlazione con sPMV

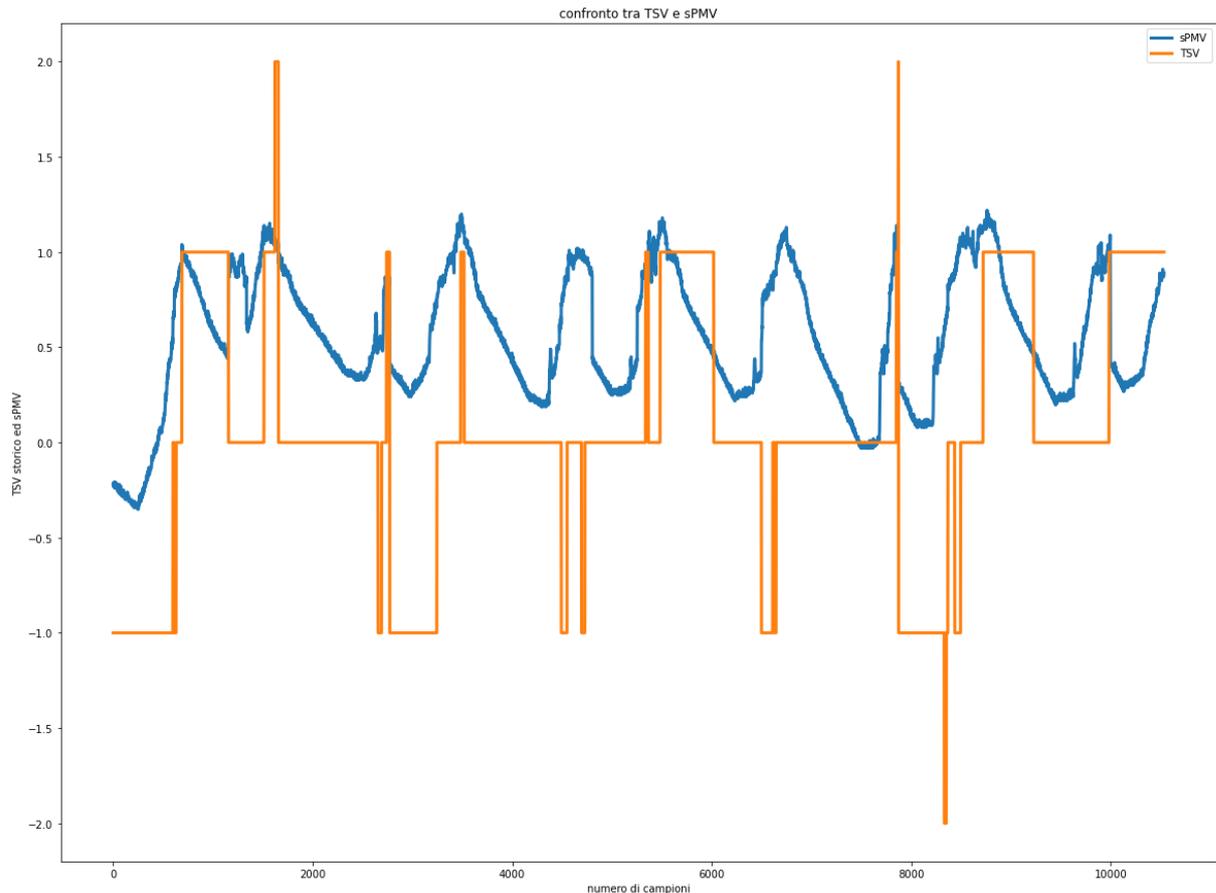


Figura 14: andamento di TSV e sPMV a confronto

3.2.2 Analisi dati caso studio 2 (FVH)

Il secondo dataset utilizzato per lo studio proviene da una raccolta di dati effettuata in un edificio situato in Finlandia, dal nome 'FVH'. Il dataset si riferisce ai dati raccolti nella stanza 333 del quarto piano dell'edificio, e i parametri misurati sono la temperatura interna, la temperatura esterna, l'umidità relativa interna, il consumo energetico e il comfort termico degli utenti, ottenuto con il TSV. Anche in questo caso si è proceduto con la selezione delle colonne di interesse, e la ricerca di eventuali anomalie, quali outlier, duplicati o valori non registrati. Si è visto che erano presenti alcuni duplicati, i quali sono stati rimossi, ottenendo un dataset di 192 valori contro i 271 iniziali. Il periodo di acquisizione va dal 15 gennaio 2024 ore 10:20 al 31 gennaio 2024 ore 14:50, con un intervallo di acquisizione di 20 minuti (Figura 15).

date	room_x	indoor T_y	ndoor RH_y	outT	value	TSV
2024-01-19 13:40:00	333	22.333	13	-10.222	1.81817	-1
2024-01-19 14:00:00	333	22.167	13	-10.167	1.67533	-1
2024-01-22 08:20:00	333	23.222	18	-0.944	2.336	0
2024-01-22 08:40:00	333	23.5	18	-0.944	2.336	-0.5
2024-01-22 09:00:00	333	23.444	18	-0.944	2.457	-0.5
2024-01-22 09:20:00	333	22.889	18	-1.111	2.457	-1

Figura 15: dataset FVH senza duplicati

Come si può notare dalla figura 15, il dataset contiene dei buchi temporali di alcuni giorni, il che rende disomogenea l'intera misurazione, cosa che può portare a delle difficoltà nel momento in cui si va ad applicare il modello di previsione del TSV. Si è deciso quindi di adottare un metodo di interpolazione per i dati mancanti, utilizzando il tipo di interpolazione denominata 'time linear interpolation'. Questa tecnica si basa sull'interpolazione lineare dei dati in base alla distanza temporale, che consente di riempire i valori mancanti in una serie temporale, e viene applicata tramite il linguaggio di programmazione Python.

Alla fine di questo processo il numero di sample era 389, ed è stato applicato un metodo di ricampionamento per ottenere il dataset orario attraverso il calcolo delle medie orarie dei parametri. Nelle figure seguenti (16, 17, 18) vengono illustrati gli andamenti dei parametri interpolati e la matrice di correlazione.

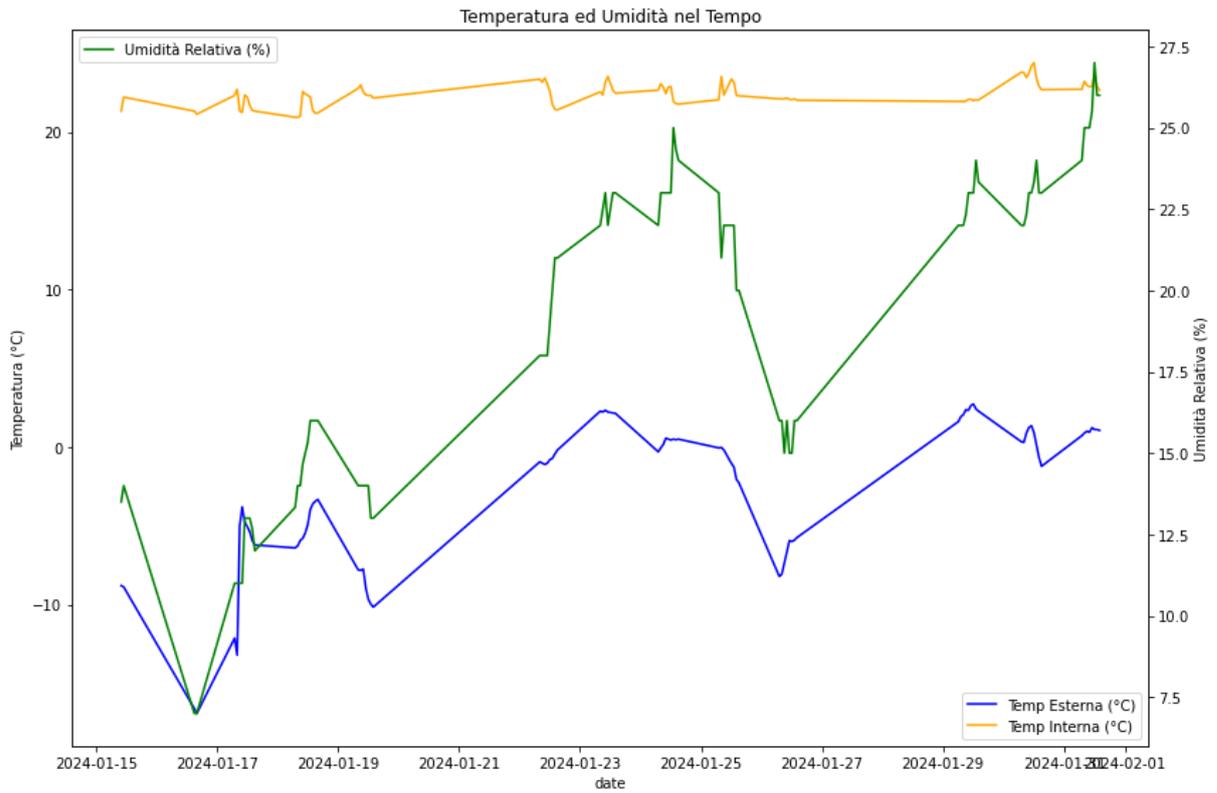


Figura 16: andamento parametri ambientali dopo l'interpolazione

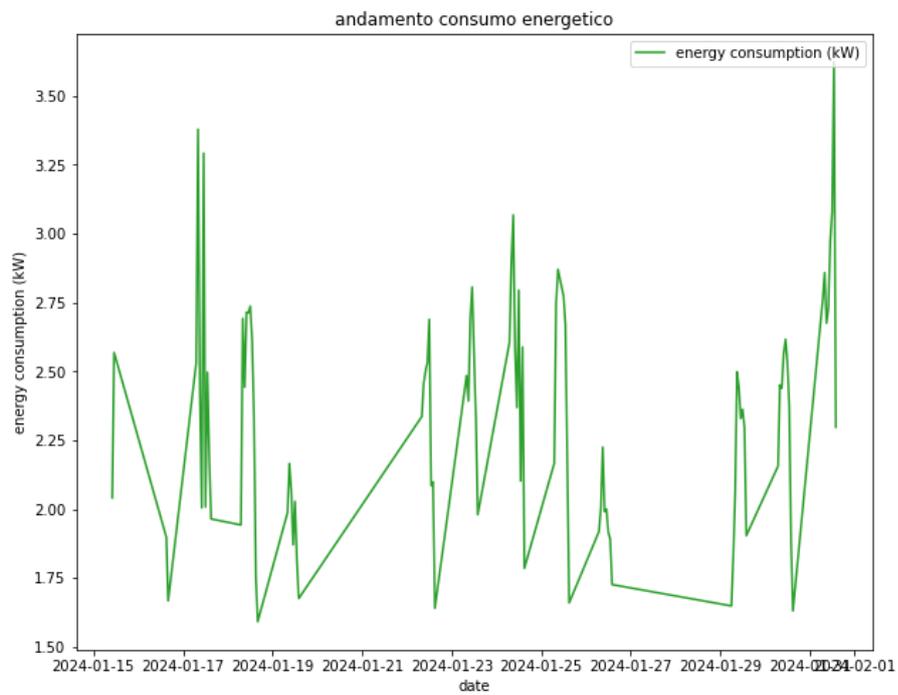


Figura 17: andamento consumo energetico in kW dopo l'interpolazione

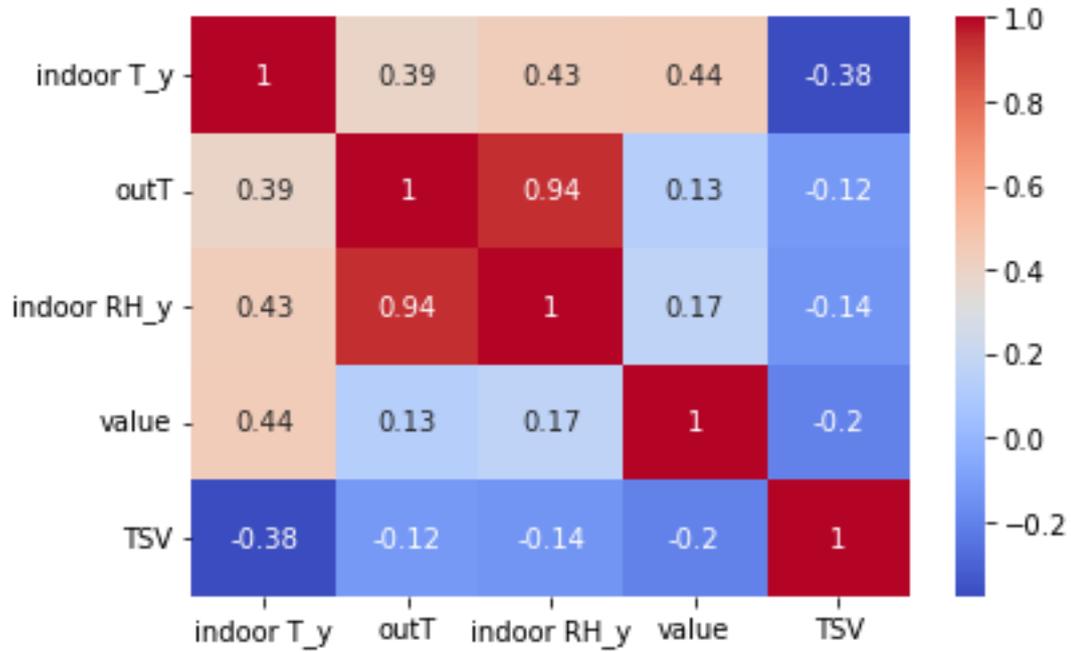


Figura 18: matrice di correlazione del dataset interpolato

Dalla matrice di correlazione in figura 18 si nota come, rispetto al primo dataset, il TSV presenti dei valori di correlazione minori con i parametri ambientali.

Come già fatto per il primo dataset, anche in questo caso è stato calcolato il sPMV attraverso il modello semplificato proposto da Buratti [23]. Di seguito le figure corrispondenti a questi passaggi dell'analisi.

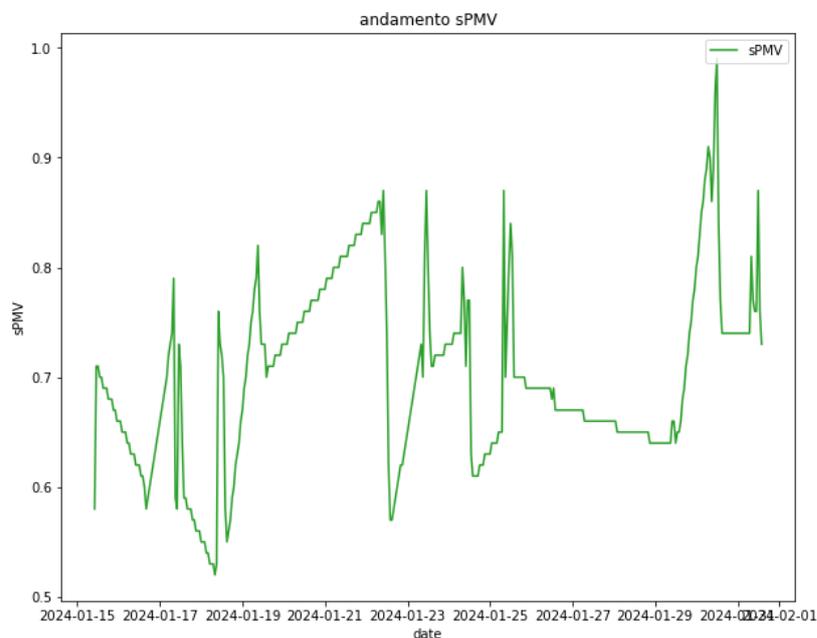


Figura 19: andamento sPMV

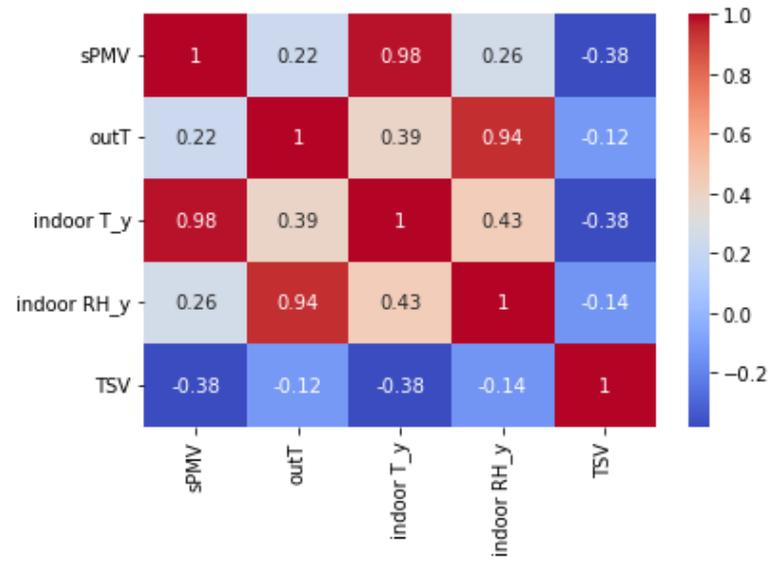


Figura 20: matrice di correlazione con sPMV

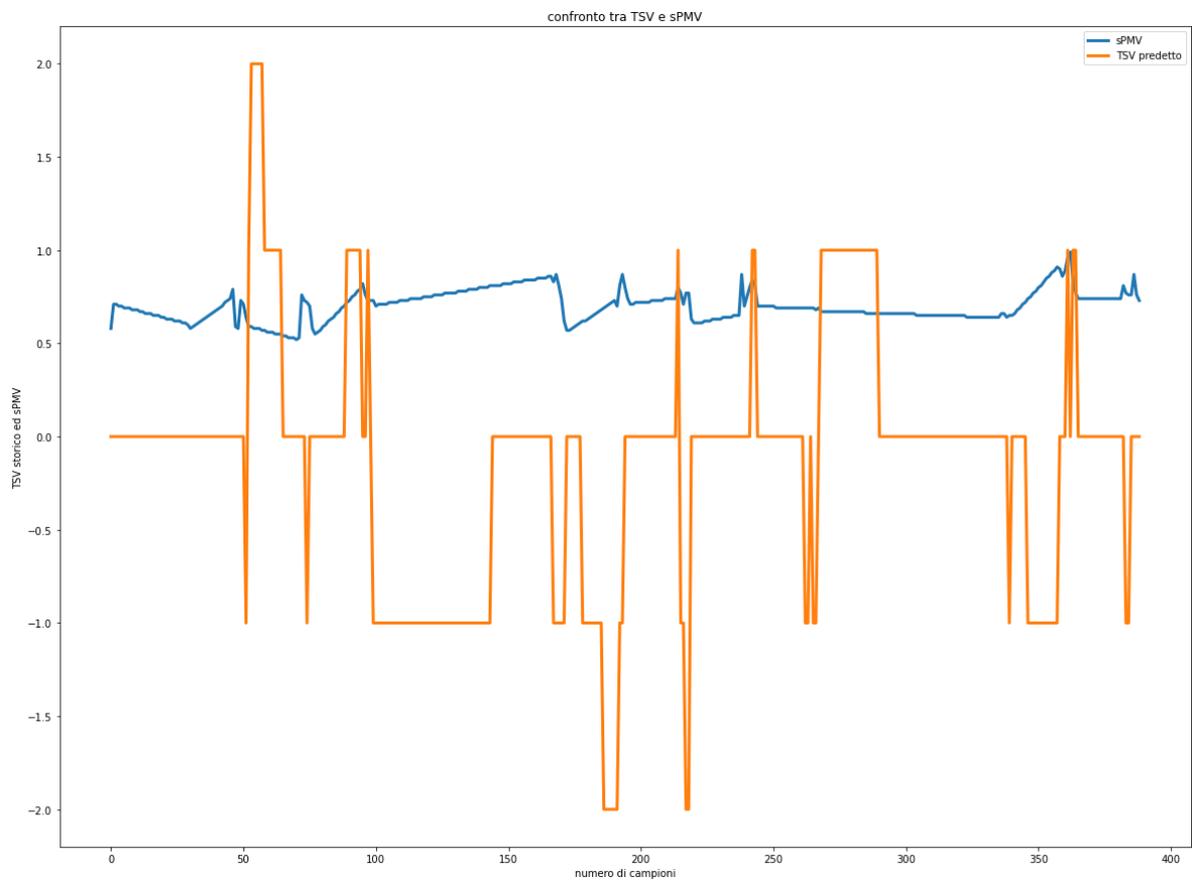


Figura 21: andamento del TSV e del sPMV a confronto

3.3 Sviluppo ed applicazione algoritmi ML

Terminata la fase dell'analisi dati ed ottenuti i dataset processati, si è passati all'applicazione di alcuni algoritmi di machine learning per la previsione del TSV. Sono stati messi a confronto diversi modelli di previsione per valutare quale fosse il più adatto al caso di studio. Nei seguenti paragrafi verranno analizzati caso per caso.

3.3.1 Predizione TSV con classificatori

Il primo tipo di metodo di machine learning utilizzato è la classificazione. Sono stati applicati due modelli diversi di classificazione ad entrambi i dataset, effettuando molteplici prove variando di volta in volta sia il numero di variabili utilizzate (i parametri ambientali), sia i parametri intrinseci, che verranno di seguito illustrati. Si è iniziato con l'algoritmo di classificazione Support Vector Classifier (SVC) [10], che è una variante della Support Vector Machine (SVM), un algoritmo di apprendimento supervisionato che migliora il meccanismo di machine learning rendendo minimo il rischio strutturato.

SVC è un modello per problemi a due classi, una positiva e una negativa, il cui obiettivo di apprendimento è massimizzare il margine tra le classi e trasformare questo obiettivo in una soluzione di programmazione quadratica convessa [10]. Il dataset utilizzato viene diviso in due categorie di variabili, x e y . x rappresenta le 'features' di input del modello, che nel caso di studio corrente corrispondono ai parametri ambientali; y viene denominata 'label', la variabile da prevedere, e corrisponde alla sensazione termica (TSV). Il principio di base della SVC è utilizzare un "iperpiano" come confine decisionale (equazione 1) per separare le due classi. La distanza del punto dal piano di ogni campione deve essere uguale o superiore a 1 (equazione 2).

$$w^T x + b = 0 \quad (1)$$

$$y_i (w^T x_i + b) \geq 1 \quad (2)$$

I parametri w e b sono dei vettori, il primo perpendicolare al piano, il secondo lo intercetta. Questo confine costruisce in realtà due iperpiani paralleli, che rappresentano i limiti di un intervallo utilizzato per distinguere la classificazione dei campioni.

I campioni situati al di sopra del limite superiore fanno parte della classe positiva (equazione 3), quelli al di sotto del limite inferiore appartengono a quella negativa (equazione 4).

$$w^T x_i + b \geq +1, \Rightarrow y_i = +1 \quad (3)$$

$$w^T x_i + b \leq -1, \Rightarrow y_i = -1 \quad (4)$$

La distanza tra i due limiti è definita come $\delta=2/w$, e i campioni positivi e negativi all'interno della fascia d'intervallo sono i vettori di supporto.

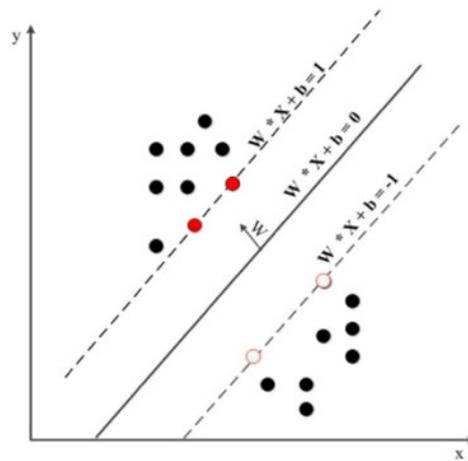


Figura 22: fascia di distinzione, intervallo di distinzione, vettori di supporto.

Il modo per ricercare l'intervallo massimo con cui dividere l'iperpiano è trovare w e b tali che soddisfino le equazioni 5 e 6:

$$\max_{w,b} \frac{2}{\|w\|} \quad (5)$$

$$s.t. y_i (w^T x_i + b) \geq 1, i = 1, 2, \dots, N. \quad (6)$$

Nel caso di problemi non linearmente separabili, vengono introdotte iper-superfici nello spazio delle caratteristiche per separare le classi positive da quelle negative.

In questo contesto, le funzioni kernel a base radiale (RBF) vengono utilizzate per descrivere lo spazio delle caratteristiche originali in uno spazio di dimensioni superiori

(Equazione 7). La funzione kernel RBF evita il calcolo diretto del prodotto interno, utilizzando il parametro σ per definire la larghezza del kernel radiale.

$$K(x_i, x_j) = \exp\left(-\frac{\|X_i - X_j\|^2}{2\sigma^2}\right) = \exp\left(-\gamma\|X_i - X_j\|^2\right) \quad (7)$$

x e y corrispondono a due vettori di input distinti nello spazio delle caratteristiche. $\gamma=1/2\sigma^2$ rappresenta un parametro specifico del kernel RBF per migliorare le prestazioni del modello. Nelle applicazioni pratiche, è spesso difficile determinare una funzione kernel appropriata per rendere i campioni di addestramento linearmente separabili nello spazio delle caratteristiche. Si introducono quindi due variabili che consentano di correggere eventuali errori. La prima è la variabile di rilassamento ξ che consente ad alcuni punti di situarsi nel lato sbagliato della superficie. La seconda è un fattore di errore C , il quale bilancia la tolleranza degli errori. L'equazione 5 può essere riscritta nella forma illustrata nell'equazione 8.

$$\min_{w,b,\xi_i} \frac{1}{2}\|w\|^2 + C \sum_{i=1}^N \xi_i \quad (8)$$

Nel caso di situazioni in cui sono presenti problemi di multi-classificazione, viene applicata la strategia "uno-contro-tutti", in cui un certo numero di classificazioni stabiliscono un certo numero di confini decisionali, ognuno dei quali giudica contro una singola classe. Ad esempio, nel caso corrente si hanno 7 livelli di sensazione termica, quindi si ha bisogno di 7 confini decisionali, ognuno per ogni classe specifica. Per migliorare ulteriormente il modello, è stata applicata una soluzione che permettesse di individuare il valore ottimale del parametro C , precedentemente mostrato. Questo metodo è chiamato "Grid Search", un algoritmo che consiste nel confrontare differenti valori di C , applicandoli al modello di training e osservando quale potesse ottenere la miglior performance. In particolare, più alto è il valore di C , minore è la tolleranza di errore concessa, ma maggiore è il rischio di overfitting, ovvero di adattamento ai dati reali del

modello. Viceversa, più basso è il valore di C , maggiore è il rischio di underfitting. Effettuata la prova, il valore ottimale trovato è stato di $C=100$.

Il modello SVC è stato applicato ad entrambi i dataset sia utilizzando la grid search, sia omettendo questa prova, utilizzando un valore C di default all'interno dell'algoritmo. Sono state effettuate diverse prove cambiando di volta in volta il tipo di variabili ambientali, utilizzando prima tutte le caratteristiche (temperatura interna, temperatura esterna, umidità relativa ed sPMV), poi rimuovendo il sPMV oppure i parametri ambientali interni.

Il secondo algoritmo di classificazione prende il nome di Random Forest Classifier (RFC), facente parte della famiglia dei modelli basati su alberi decisionali [30]. Questo gruppo di tecnologie di machine learning usa "modelli ad albero" per valutare le decisioni e le loro conseguenze. I metodi Random Forest (RF) utilizzano gli alberi decisionali come classificatori di base ed applica il principio dell'apprendimento integrato per costruire un classificatore combinatorio [22]. Un certo numero di sottoinsiemi di addestramento viene campionato in maniera casuale dal dataset originale, e per ciascun sottoinsieme viene selezionato allo stesso modo un sottoinsieme di caratteristiche per costruire un modello basato su alberi decisionali. La previsione di nuovi dati si basa sul conteggio dei risultati di classificazione forniti da tutti gli alberi decisionali. Per determinare il risultato finale della classificazione, viene applicata una strategia di voto a maggioranza. La costruzione della RF inizia con l'estrazione di alcuni campioni per i sottoinsiemi di addestramento, che saranno un numero pari a quante volte viene ripetuta l'operazione di estrazione. Viene poi costruito un modello di albero decisionale per ogni sottoinsieme, scegliendo alcune tra le variabili contenute nel dataset, e selezionando quella con il coefficiente di Gini (un indice utilizzato come criterio di separazione) più basso, per effettuare la suddivisione. Questo ciclo viene iterato fino a soddisfare la condizione di arresto. Una volta terminato, si costruisce l'albero decisionale corrispondente a ciascun sottoinsieme. Infine, tutti gli alberi decisionali vengono combinati per formare la RF.

Anche nel caso del modello RF sono state effettuate più prove scegliendo le variabili secondo il criterio illustrato per SVC, ovvero selezionando di volta in volta parametri ambientali diversi da utilizzare per confrontare i risultati.

3.3.2 Predizione TSV con regressori

Dopo aver effettuato le prove sulla previsione del TSV attraverso i classificatori, si è deciso di sperimentare anche l'utilizzo dei regressori. La differenza principale con i classificatori consiste nel fatto che la variabile target, nel caso di questo studio il TSV, che rappresenta anche l'output della previsione, è composta da valori continui anziché discreti [31].

Esistono due tipi di regressione, lineare e non lineare, di cui fanno parte molteplici algoritmi di machine learning. In questo articolo, i due modelli di regressione utilizzati sono gli stessi presi in considerazione per la classificazione, ovvero regressori basati sul vettore di supporto (SVR) e il metodo dei regressori basati su alberi decisionali (RFR) [32]. I due dataset sono stati divisi in due, in ordine temporale: il primo periodo rappresentava l'80% dei dati ed è stato utilizzato per l'addestramento del modello; il restante 20% dei dati, corrispondenti agli ultimi giorni di acquisizione dati, è stato utilizzato per la validazione del modello precedentemente addestrato.

3.4 Metriche di valutazione

L'accuratezza dei vari modelli è stata valutata attraverso due diverse categorie di metriche di valutazione, una riferita ai modelli di classificazione, l'altra adatta a valutare i modelli di regressione. Le metriche che valutano le classificazioni sono l'accuratezza (acc), il F1 score, la precisione, il richiamo e l'Area Under Curve [10].

L'accuratezza è definita come il rapporto tra i campioni classificati e il campione completo: maggiore è il fattore acc, maggiore è l'accuratezza del modello. L'equazione è la seguente:

$$acc = \frac{TP + TN}{TP + FP + TN + FN} \quad (9)$$

TP rappresenta il numero dei veri positivi, FP quello dei falsi positivi, FN sono i falsi negativi, e TN i veri negativi.

Il F1 score rappresenta la media armonica di precisione e richiamo, i quali sono due misure contrastanti tra loro: se il primo è alto, il secondo è basso e viceversa. Il valore di

F1 può valutare precisione e richiamo del modello in maniera completa. Per ottenere un modello il più accurato possibile, il valore di F1 deve essere prossimo ad 1. Vengono sotto riportate le equazioni:

$$F1 = 2 * \frac{precision * recall}{precision + recall} \quad (10)$$

$$precision = \frac{TP}{TP + FP} \quad (11)$$

$$recall = \frac{TP}{TP + FN} \quad (12)$$

Il valore di AUC rappresenta l'area sotto la curva chiamata receiver operating characteristic curve (ROC), costruita in un grafico composto dal tasso valori dei veri positivi nella previsione in ordinata e tasso dei falsi positivi in ascissa. Idealmente, il tasso dei veri positivi dovrebbe essere vicino ad 1, mentre quello dei falsi positivi vicino a 0, il che comporta ad ottenere valori di AUC prossimi a 1. L'accuratezza del modello è maggiore quanto più l'AUC è vicina ad 1. I primi quattro fattori sono stati calcolati sia per SVC che per RFC, mentre il valore di AUC è stato valutato solo per il metodo dei vettori di supporto.

Per quanto riguarda i modelli di regressione, sono quattro le metriche di valutazione dei modelli: l'errore assoluto medio (mean absolute error, MAE), l'errore quadratico medio (mean square error, MSE) e l'errore quadratico relativo (relative square error, RSE).

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (13)$$

Il MAE calcola la media dei valori assoluti degli errori, dove y rappresenta il valore reale, mentre \hat{y} segnato corrisponde alla previsione. Questo fattore non penalizza in modo eccessivo gli errori grandi, rendendolo meno sensibile agli **outlier**. È una metrica particolarmente utile quando si vuole ottenere un'indicazione chiara e interpretabile della deviazione media del modello rispetto ai valori reali.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (14)$$

Il MSE è una misura che calcola la media dei quadrati degli errori, con y valore reale e \hat{y} segnato quello predetto. Il MSE penalizza gli errori più grandi elevando al quadrato le differenze, mettendo in risalto i valori che si discostano di molto dalla realtà. È adatto alla minimizzazione degli errori di ampia entità.

$$\text{RSE} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (15)$$

Il RSE confronta l'errore quadratico totale del modello con quello di un modello di riferimento, che utilizza come previsione la media di tutti i valori osservati. Se RSE è inferiore a 1, significa che il modello di previsione migliora quello di riferimento, viceversa in caso il valore di RSE superi 1. Questa metrica consente di confrontare modelli diversi e di capire quanto uno si discosti da una strategia base di previsione.

3.5 Sviluppo ed applicazione di protocollo di misura per raccolta feedback utenti

Una volta valutati i risultati della previsione del comfort attraverso gli algoritmi di ML, è emerso il bisogno di applicare il modello ad un nuovo dataset per testare la validità della metodologia proposta. È stato quindi sviluppato di un protocollo di misura finalizzato alla raccolta dei feedback (espressi come TSV) degli utenti in ambiente d'ufficio.

3.5.1 Protocollo di misura

Durante la campagna di misura i dati raccolti sono i parametri ambientali e feedback degli utenti. I parametri ambientali misurati sono i seguenti:

- temperatura interna dell'aria (°C),
- temperatura esterna (°C),
- umidità relativa interna (%),
- concentrazione CO₂ nella stanza (ppm),
- velocità dell'aria nella stanza (m/s),
- rumorosità della stanza (dB).

I parametri sono stati misurati utilizzando una strumentazione dedicata, descritta dettagliatamente di seguito:

- Stazione microclimatica Delta Ohm HD 32.1
- Termocoppie
- Stazione Meteo Intelligente Netatmo
- Sensibo Elements

Sensore	Parametro	Range di misura	Risoluzione	Accuratezza
Stazione microclimatica	Temperatura [°C]	-30/+100 °C	0,01 °C	±0,1 °C
	Umidità relativa [%]	0% - 100%	0,1%	±1,5%
	Velocità dell'aria [m/s]	0,01 - 5 m/s	0,01 m/s	±0,02 m/s
Netatmo	Temperatura [°C]	0/50°C	0,1°C	±0,3°C
	Umidità relativa [%]	0% - 100%	0,1%	±3%
	Quantità di CO ₂ [ppm]	0/5000 ppm	1 ppm	±50 ppm
	Rumorosità [dB]	35 - 120 dB	0,1 dB	±0.5 dB
Sensibo	Temperatura [°C]	0/55°C	0,1°C	±0,5°C
	Umidità relativa [%]	5% - 95%	1%	±5%
	Quantità di CO ₂ [ppm]	0-5000 ppm	1 ppm	±50 ppm
Termocoppie	Temperatura [°C]	-200/+350 °C	0,1 [°C]	±0,5 °C

Tabella 1: specifiche tecniche dei sensori utilizzati

La **stazione microclimatica HD 32.1** è uno strumento di acquisizione composto da diverse sonde e da una centralina, la quale raccoglie le informazioni sui valori misurati dalle sonde. Le sonde installate sono: *sonda HP3217* per temperatura e umidità relativa; *sonda TP3275* munita di globo-termometro per la radiazione termica; *sonda AP3203* per la misura della velocità dell'aria. Una volta installate le sonde e sistemata la struttura su un cavalletto, queste vengono collegate alla centralina. La centralina poi viene collegata alla rete elettrica e programmata per l'acquisizione e il salvataggio dei dati.



Figura 23: installazione della stazione microclimatica

Le **due termocoppie** utilizzate sono del tipo T (rame-costantana), utilizzate per la misurazione della *temperatura esterna*. Sono state installate sul montante di una finestra all'interno della stessa stanza in cui è stata posizionata la stazione microclimatica, poi sono state cablate ad un computer per la registrazione dei dati.



Figura 24: termocoppia di tipo T

La **stazione meteo intelligente Netatmo** è uno strumento progettato per acquisire e monitorare i principali parametri ambientali interni di una stanza. Il dispositivo viene posizionato in un punto fisso della stanza, collegato alla rete elettrica e connesso alla rete Wi-Fi. Per accedere ai dati acquisiti dalla stazione Netatmo e configurare il dispositivo, è necessario utilizzare l'applicazione dedicata Netatmo Home Coach, disponibile sia per sistemi operativi Android che iOS. Tramite questa applicazione, è possibile effettuare la configurazione iniziale e visualizzare in tempo reale le misurazioni direttamente sul proprio dispositivo mobile. I parametri ambientali rilevati dalla stazione Netatmo includono *temperatura interna dell'aria, umidità relativa, concentrazione CO₂ nell'aria e rumorosità*.



Figura 25: Stazione Meteo Intelligente

Il sensore **Sensibo Elements** è un dispositivo che consente il controllo a distanza della qualità dell'aria e fornisce suggerimenti per migliorare l'ambiente in cui si trova. Una volta collegato il dispositivo ad una rete elettrica, è necessario installare l'app Sensibo, disponibile sia per Android che per iOS, all'interno della quale deve essere creato un account personale. In seguito, per associare il dispositivo all'app deve essere scannerizzato il QR code presente in esso, ed assegnare al sensore una posizione specifica all'interno della stanza. Successivamente, il dispositivo va collegato ad una rete wifi, ed una volta fatto ciò attraverso l'app sarà possibile monitorare la qualità dell'aria e ricevere suggerimenti su come migliorarla.



Figura 26: dispositivo Sensibo Elements.

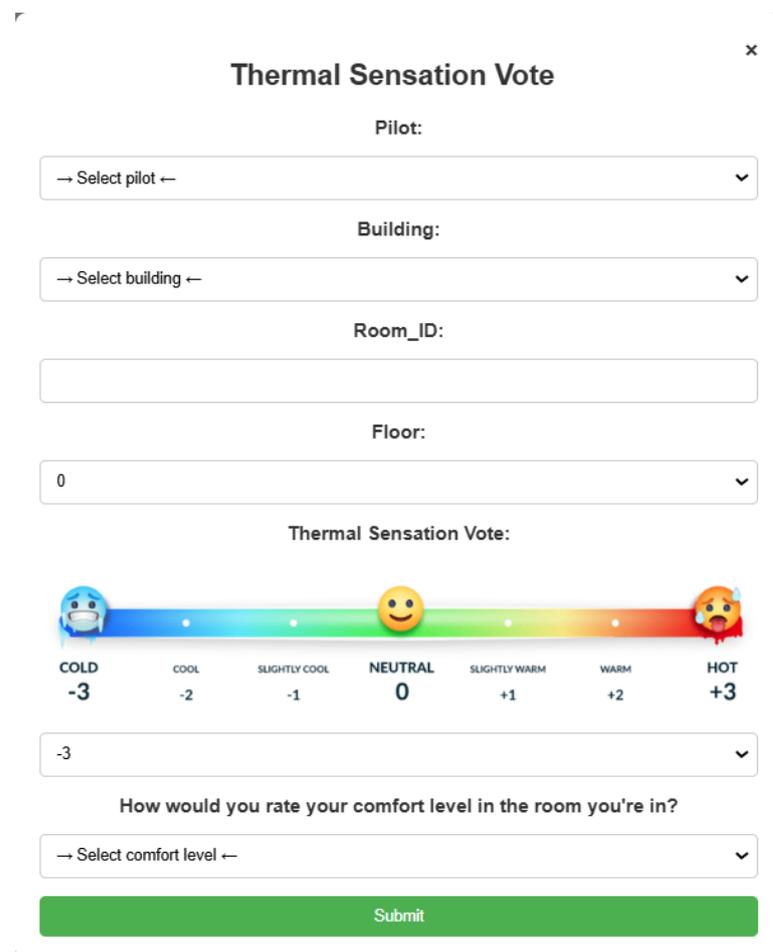
Per l'acquisizione del feedback degli utenti, il primo step ha riguardato la raccolta di informazioni personali quali genere, altezza, peso, Body Mass Index (BMI), età ed ufficio dei partecipanti (Tabella 1).

Utente	Genere	Età	Altezza	Peso	Stanza	BMI
Soggetto 1	F	28	1,78 m	55 kg	019	17,36 kg/m ²
Soggetto 2	F	27	1,68 m	52 kg	019	18,42 kg/m ²
Soggetto 3	M	24	1,80 m	83 kg	019	25,62 kg/m ²
Soggetto 4	M	27	1,73 m	73 kg	019	24,39 kg/m ²
Soggetto 5	M	31	1,80 m	70 kg	019	21,60 kg/m ²
Soggetto 6	M	27	1,73 m	71 kg	019	23,72 kg/m ²
Soggetto 7	M	22	1,71 m	69 kg	019	23,60 kg/m ²
Soggetto 8	F	32	1,75 m	59 kg	014	19,27 kg/m ²
Soggetto 9	F	51	1,60 m	49 kg	012	19,14 kg/m ²
Soggetto 10	F	35	1,66 m	51 kg	016	18,51 kg/m ²

Tabella 1: informazioni personali degli utenti

In seguito, è stato definito il metodo di raccolta delle risposte relative al comfort termico. A tale scopo, è stato creato un gruppo Telegram in cui sono stati aggiunti tutti i partecipanti. Gli utenti ricevevano periodicamente una notifica contenente un link, che apriva una pagina web con un questionario da compilare. In Figura 27, è riportata la pagina web appena citata. Le informazioni richieste sono le seguenti:

- Il pilot (in cui è stato inserito: UNIVPM),
- L'edificio (in cui è stato inserito: UNIVPM facoltà di ingegneria),
- La stanza,
- Il piano,
- Sensazione termica (TSV),
- Livello di comfort globale, in una scala da 0 a 10, che comprende la valutazione complessiva del comfort termico, acustico e luminoso



The image shows a web form titled "Thermal Sensation Vote" with a close button (x) in the top right corner. The form contains several input fields and a rating scale:

- Pilot:** A dropdown menu with the text "→ Select pilot ←".
- Building:** A dropdown menu with the text "→ Select building ←".
- Room_ID:** A text input field.
- Floor:** A dropdown menu with the value "0".
- Thermal Sensation Vote:** A horizontal scale with a color gradient from blue to red. It features seven points with corresponding labels and values: COLD (-3), COOL (-2), SLIGHTLY COOL (-1), NEUTRAL (0), SLIGHTLY WARM (+1), WARM (+2), and HOT (+3). Each point is accompanied by an emoji: a blue face with cold sweat for COLD, a yellow face with a neutral expression for NEUTRAL, and a red face with sweat drops for HOT.
- Below the scale is a dropdown menu with the value "-3".
- The question "How would you rate your comfort level in the room you're in?" is followed by a dropdown menu with the text "→ Select comfort level ←".
- A green "Submit" button is located at the bottom of the form.

Figura 27: questionario da compilare.

La campagna di misura è stata condotta tra il 20 novembre 2024 e il 27 novembre 2024 in quattro stanze situate al sesto piano del Dipartimento Di Ingegneria Industriale e Scienze Matematiche (DIISM) della Facoltà di Ingegneria dell'Università Politecnica delle Marche (UNIVPM).

Nello specifico, nella stanza denominata '019' (Figura 28) è stato installato un sensore Netatmo, mentre nella stanza '014' sono stati collocati un sensore Netatmo, un set di termocoppie e la stazione microclimatica, posizionata al centro della stanza. Nella stanza '012' è stato installato un Sensibo, mentre l'altro Sensibo è stato posizionato nella stanza '016'. Nella Tabella 2 sono riportate queste informazioni.

Sono stati coinvolti dieci partecipanti nell'esperimento, ai quali è stato richiesto di fornire un feedback personale sul comfort termico e sul comfort globale con cadenza oraria, tramite il sistema di raccolta feedback precedentemente descritto. I feedback sono stati acquisiti durante il periodo della giornata lavorativa, che andava dalle ore 9:00 alle ore 18:00.

Stanza	Stazione Microclimatica	Netatmo	Sensibo	Termocoppie
019		X		
016			X	
014	X	X		X
012			X	

Tabella 2: Disposizione dei sensori nelle stanze incluse nella campagna di misura.

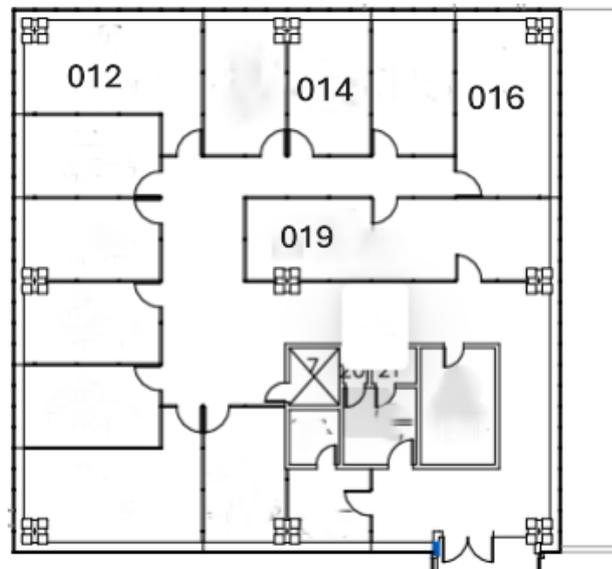


Figura 28: Pianta del luogo di raccolta dati.

3.6 Analisi database ASHRAE e ricerca correlazioni

L'ultima attività di ricerca svolta è l'analisi statistica del ASHRAE Global Thermal Comfort database II con l'obiettivo di evidenziare correlazioni tra la sensazione termica umana, espresso come TSV, e le caratteristiche personali dell'individuo (come, età, peso, altezza e genere). Dal database sono stati selezionati i dati raccolti nel file **'db_measurements_v2.1.0'**. In primo luogo, il dataset è stato filtrato per visualizzare le colonne con i parametri necessari all'analisi.

building_id	timestamp	age	gender	ht	wt	ta	tr	rh	vel	met	clo	mal_sensal	pmv
1	1995-05-18T00:00:00Z	nan	female	nan	nan	22.3	22.95	61	0.03	1.70648	0.95	2	0.5
1	1995-05-18T00:00:00Z	nan	female	nan	nan	23	24.1343	59	0.08	1.10922	1.07	2	0.4
1	1995-05-18T00:00:00Z	nan	female	nan	nan	22	22.0914	61	0.04	1.2116	0.88	0	-0.07
1	1995-05-18T00:00:00Z	nan	female	nan	nan	23.2	24.3303	58	0.03	1.77474	0.59	0	0.31
1	1995-05-18T00:00:00Z	nan	male	nan	nan	22.2	23.0771	62	0.15	1.48464	0.83	0	0.05
1	1995-05-18T00:00:00Z	nan	male	nan	nan	21.2	21.0402	65	0.15	1.50171	0.66	-1	-0.56

Figura 29: dataset filtrato contenente i parametri di interesse

Attraverso un altro dataset presente nel database, denominato **'db_metadata'**, potevano essere visualizzate le informazioni corrispondenti ai valori letti sulla colonna 'building_id'. In particolare, ogni valore corrispondeva ad una specifica raccolta dati in un certo edificio di una determinata nazione del mondo.

building_id	uilding_id_i	contributor	publication	region	country	city	lat	lon	climate
140	no	John Busch	https://doi...	asia	thaila...	bangkok	13.7563	100.502	tropical savanna
141	no	Nick Baker	https://doi...	europa	greece	athens	37.9838	23.7275	mediterranean
142	no	Nick Baker	https://doi...	europa	greece	athens	37.9838	23.7275	mediterranean
143	no	Nick Baker	https://doi...	europa	greece	athens	37.9838	23.7275	mediterranean
144	no	Nick Baker	https://doi...	europa	greece	athens	37.9838	23.7275	mediterranean
145	no	Nick Baker	https://doi...	europa	greece	athens	37.9838	23.7275	mediterranean
146	no	Nick Baker	https://doi...	europa	greece	athens	37.9838	23.7275	mediterranean

Figura 30 visualizzazione di alcune colonne del dataset 'db_metadata'

Una volta filtrato il dataset, si è proceduto con l'identificazione delle righe in cui fossero presenti tutti i dati relativi al peso, all'altezza, all'età, al genere e al feedback, poiché nella maggior parte dei casi uno o più di questi non erano registrati all'interno della riga. Procedendo in questo modo, si è quindi ottenuta una tabella di 1818 samples.

Al dataset ottenuto, è stata aggiunta una colonna corrispondente al valore di body mass index (BMI), che è l'indice di massa corporea (kg/m^2), la cui formula è riportata nell'equazione 16.

$$\text{BMI} = \frac{\text{Weight (kgs)}}{[\text{Height (m)}]^2}$$

Equazione 16: BMI (kg/m^2)

Questo indice è stato calcolato a partire dai dati circa il peso (colonna 'wt') e altezza (colonna 'ht') contenuti nel dataset.

building_id	timestamp	age	gender	ht	wt	ta	tr	rh	vel	met	clo	mal_sensal	pmv	BMI
728	2015-03-10T00:00:00Z	35	male	1.7	68	25.2	25.8	49	0.11	1.2	0.5	-2	0.03	23.5294
728	2015-03-10T00:00:00Z	24	female	1.6	52	25.2	25.8	49	0.11	1.2	0.42	-2	-0.14	20.3125
728	2015-03-10T00:00:00Z	27	female	1.65	65	25.2	25.8	49	0.11	1.2	0.5	-1	0.03	23.8751
728	2015-03-10T00:00:00Z	28	male	1.78	70	25.2	25.8	49	0.11	1.2	0.5	2	0.03	22.0932
728	2015-03-10T00:00:00Z	28	male	1.73	55	25.2	25.8	49	0.11	1.2	0.5	-1	0.03	18.3768
728	2015-03-10T00:00:00Z	29	male	1.67	68	25.2	25.8	49	0.11	1.2	0.58	0	0.19	24.3824
728	2015-03-10T00:00:00Z	33	female	1.48	47	25.2	25.8	49	0.11	1.2	0.49	0	0.01	21.4573
728	2015-03-10T00:00:00Z	40	male	1.65	60	25.2	25.8	49	0.11	1.2	0.39	0	-0.21	22.0386

Figura 31: dataset ottenuto dall'analisi con colonna BMI

Lo step successivo è stato quello di iniziare la ricerca delle correlazioni tra parametri ambientali, personali e i feedback degli utenti. Le diverse tipologie di grafici proposti per la visualizzazione delle relazioni tra parametri sono i grafici box plot e gli istogrammi.

In primo luogo, i vari parametri sono stati divisi in fasce: la colonna del BMI è stata suddivisa secondo i range di valori 0-18.5, 18.5-25, 25-30 e 30-40, che rappresentano le classi sottopeso, normopeso, sovrappeso e obesità; la colonna delle età è stata categorizzata secondo le fasce 18-30, 30-50 e 50-65; infine, per la colonna di temperatura sono state scelte 4 fasce, 18-21 °C, 21-24 °C, 24-27 °C e superiori a 27 °C. Questo meccanismo di divisione in classi ha reso possibile l'analisi di correlazione del dataset, che ha evidenziato una certa dipendenza dei valori di TSV dall'età e dal BMI (quindi da peso e altezza) dell'utente.

4. Risultati

4.1 Metriche di valutazione dei modelli ML

Come è stato già illustrato nel capitolo precedente, per valutare l'efficacia dei modelli di ML sono state definite delle metriche adatte per ogni tipo di algoritmo di previsione applicato. Vengono sotto riportati i valori ottenuti.

Le tabelle 3 e 4 contengono i risultati ottenuti dall'applicazione dei modelli di classificazione SVC e RFC nel contesto del primo caso studio (Focchi).

SVC 'FOCCHI'	simulazione 1	simulazione 2	simulazione 3	simulazione 4	simulazione 5
Kernel	RBF	RBF	RBF	RBF	RBF
Grid search	no	no	C=100	C=100	C=100
Features utilizzate	Indoor T, indoor RH, outdoor T, sPMV	Indoor T, indoor RH, outdoor T	Indoor T, indoor RH, outdoor T, sPMV	Indoor T, indoor RH, outdoor T	Outdoor T, sPMV
Accuracy	76%	79%	79%	79%	79%
Precision	79%	0.81	79%	0.80	78%
Recall	76%	79%	78%	79%	79%
F1 score	72%	78%	76%	78%	78%
ROC AUC	91%	94%	93%	94%	91%

Tabella 3: metriche del SVC per il Dataset 'Focchi'.

RFC 'FVH'	simulazione 1	simulazione 2	simulazione 3
Estimators	30	30	30
Trees depth	15	15	15
Features utilizzate	Indoor T, indoor RH, outdoor T, sPMV	Indoor T, indoor RH, outdoor T	Outdoor T, sPMV
Accuracy	97%	97%	87%
Precision	97%	97%	87%
Recall	97%	97%	87%
F1 score	97%	97%	87%

Tabella 4: RFC per il Dataset 'Focchi'.

Nelle tabelle 5 e 6 sono contenute le metriche ottenute dall'applicazione dei modelli di classificazione SVC e RFC per il secondo caso studio (FVH).

SVC 'FVH'	simulazione 1	simulazione 2	simulazione 3
Kernel	RBF	RBF	RBF
Grid search	C=100	C=100	C=100
Features utilizzate	Indoor T, indoor RH, outdoor T, sPMV	Indoor T, indoor RH, outdoor T	Outdoor T, sPMV
Accuracy	69%	72%	67%
Precision	58%	61%	56%
Recall	69%	72%	67%
F1 score	63%	66%	61%
ROC AUC	80%	83%	79%

Tabella 5: SVC per il Dataset 'FVH'.

RFC 'FVH'	simulazione 1	simulazione 2	simulazione 3
Estimators	30	30	30
Trees depth	15	15	15
Features utilizzate	Indoor T, indoor RH, outdoor T, sPMV	Indoor T, indoor RH, outdoor T	Outdoor T, sPMV
Accuracy	76%	79%	77%
Precision	76%	80%	79%
Recall	76%	79%	77%
F1 score	75%	80%	77%

Tabella 6: RFC per il Dataset 'FVH'.

Le tabelle 7 e 8 contengono le metriche di valutazione dei modelli di regressione SVR e RFR applicati al primo caso studio Focchi.

SVR 'Focchi'	Training	Validazione
MAE	0.32	0.73
MSE	0.23	0.77
RSE	0.48	0.86

Tabella 7: SVR per il Dataset 'Focchi'

RFR 'Focchi'	Training	Validazione
MAE	0.05	0.67
MSE	0.03	0.80
RSE	0.16	0.90

Tabella 8: RFR per il Dataset 'Focchi'

Per quanto riguarda i risultati ottenuti da SVR e RFR applicati al secondo caso studio FVH, le metriche sono osservabili all'interno delle tabelle 9 e 10.

SVR 'FVH'	Training	Validazione
MAE	0.40	0.49
MSE	0.31	0.37
RSE	0.56	0.60

Tabella 9: SVR per il Dataset 'FVH'

RFR 'FVH'	Training	Validazione
MAE	0.30	0.71
MSE	0.27	0.72
RSE	0.52	0.85

Tabella 10: RFR per il Dataset 'FVH'

Vengono di seguito illustrati gli andamenti delle previsioni ottenute dai modelli che hanno registrato le metriche migliori. Le figure 32, 33 e 34 si riferiscono al SVC per il dataset 'Focchi'. In figura 32 viene illustrato l'andamento dei valori predetti e reali, in figura 33 il grafico a punti che evidenzia il confronto tra i valori, in figura 34 la matrice di confusione (confusion matrix) che mette in evidenza l'accuratezza del modello. Quest'ultima mostra quanto i valori predetti si discostano da quelli reali corrispondenti: le cifre all'interno dei riquadri sulla diagonale indicano le previsioni di TSV perfette.

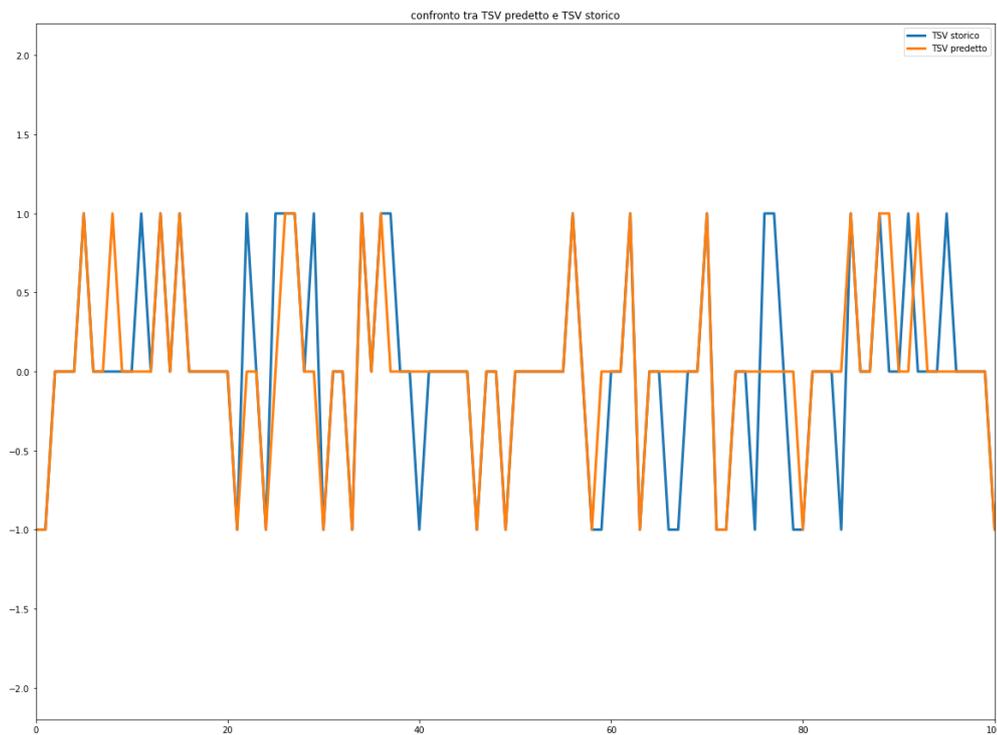


Figura 32: andamento del TSV predetto con SVC a confronto con quello reale per il Dataset 'Focchi'

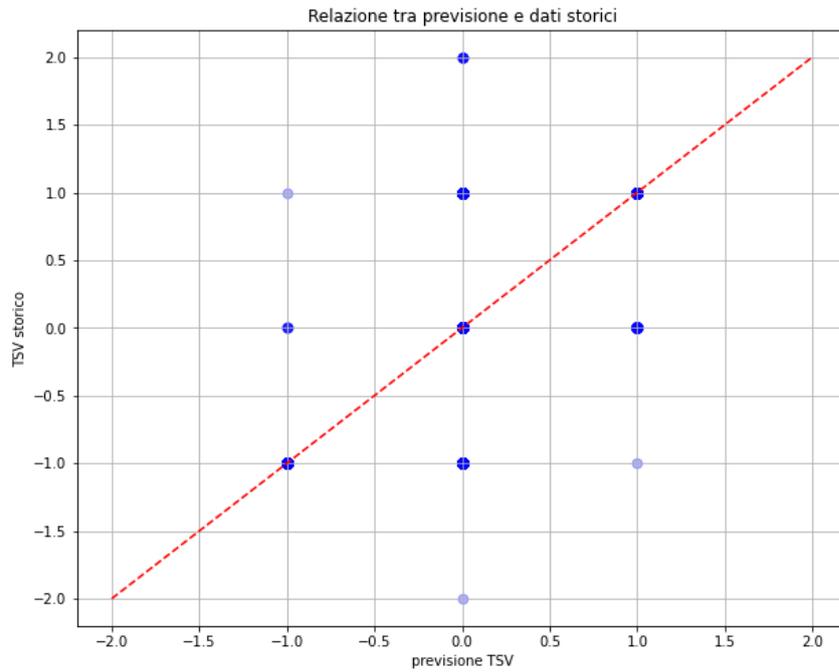


Figura 33: grafico a punti della previsione del TSV con SVC per il Dataset 'Focchi'.

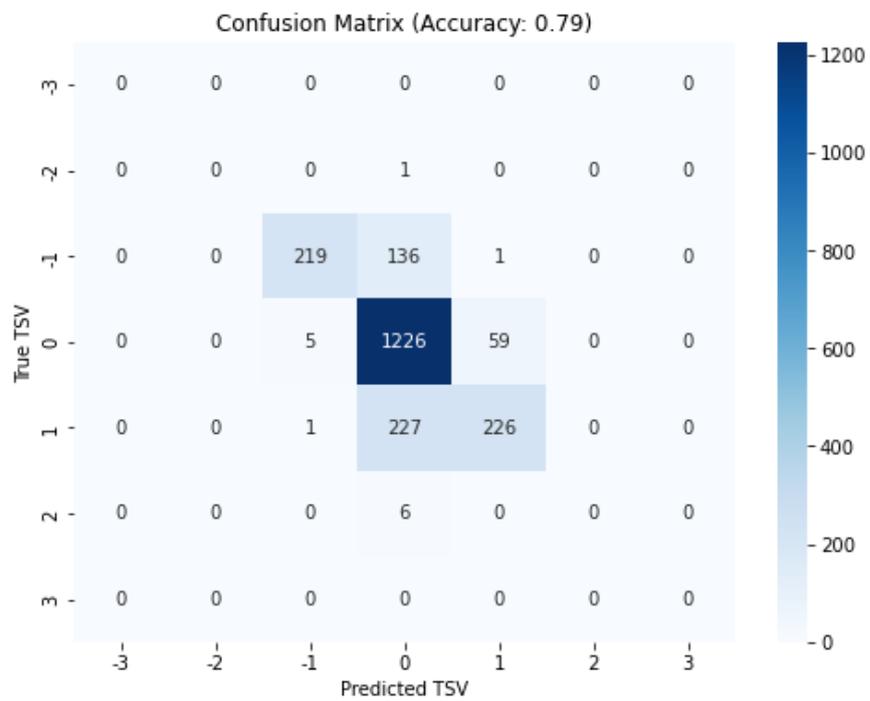


Figura 34: matrice di confusione della previsione con SVC per il Dataset 'Focchi'.

Le figure 35, 36 e 37 si riferiscono ai risultati ottenuti dal modello RFC applicato al dataset 'Focchi'.

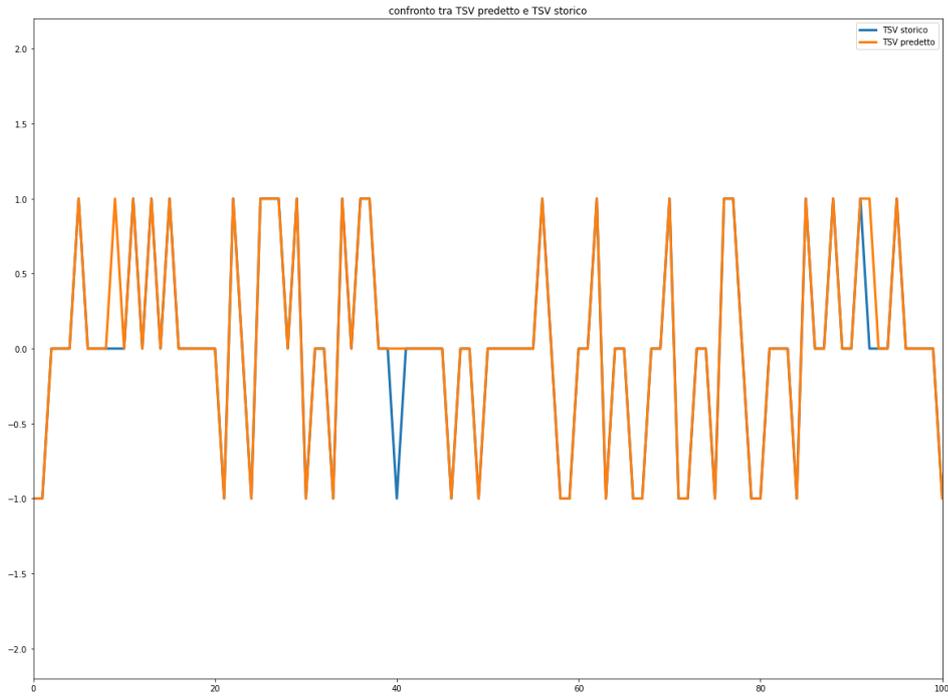


Figura 35: Andamento del TSV predetto con RFC a confronto con quello reale per il Dataset 'Focchi'.

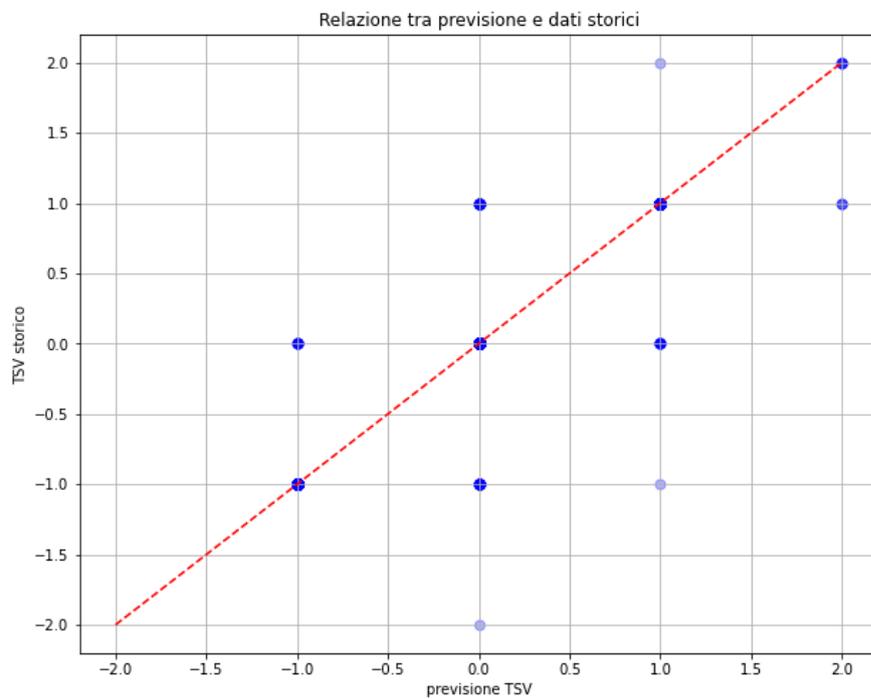


Figura 36: grafico a punti della previsione del TSV con RFC per il Dataset 'Focchi'

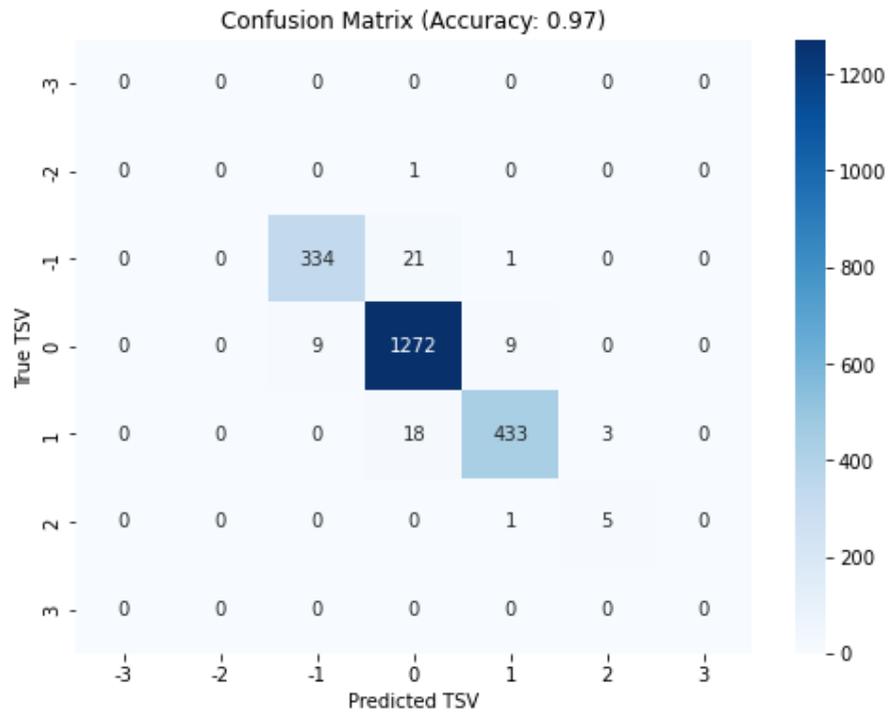


Figura 37: matrice di confusione della previsione con RFC per il Dataset 'Focchi'.

Vengono di seguito illustrati gli andamenti del TSV predetto dal metodo RFR nella fase di addestramento (figura 38) e di validazione (figura 40), insieme ai rispettivi grafici a punti per il confronto tra i valori reali e quelli ottenuti dalla previsione (figure 39 e 41).

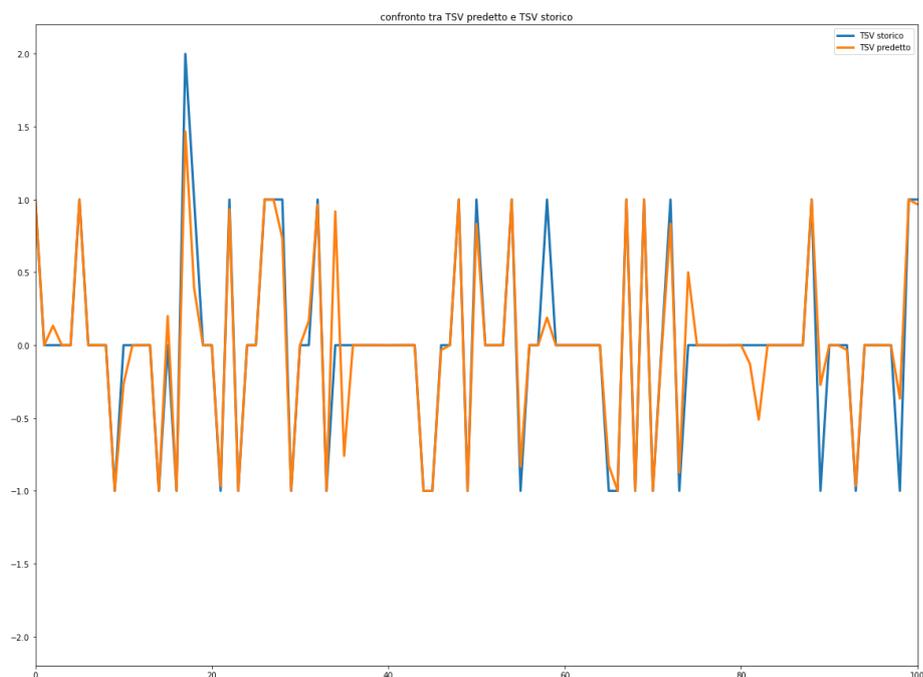


Figura 38: andamenti del TSV predetto con RFR e del TSV reale per il Dataset 'Focchi' di addestramento

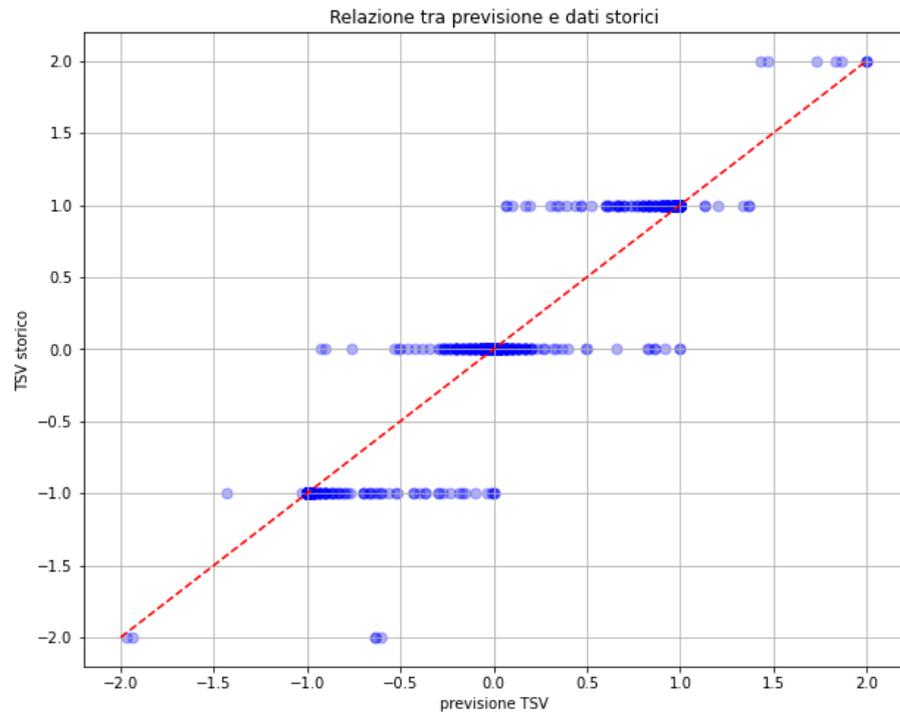


Figura 39: grafico a punti dal RFR al Dataset 'Focchi' di addestramento

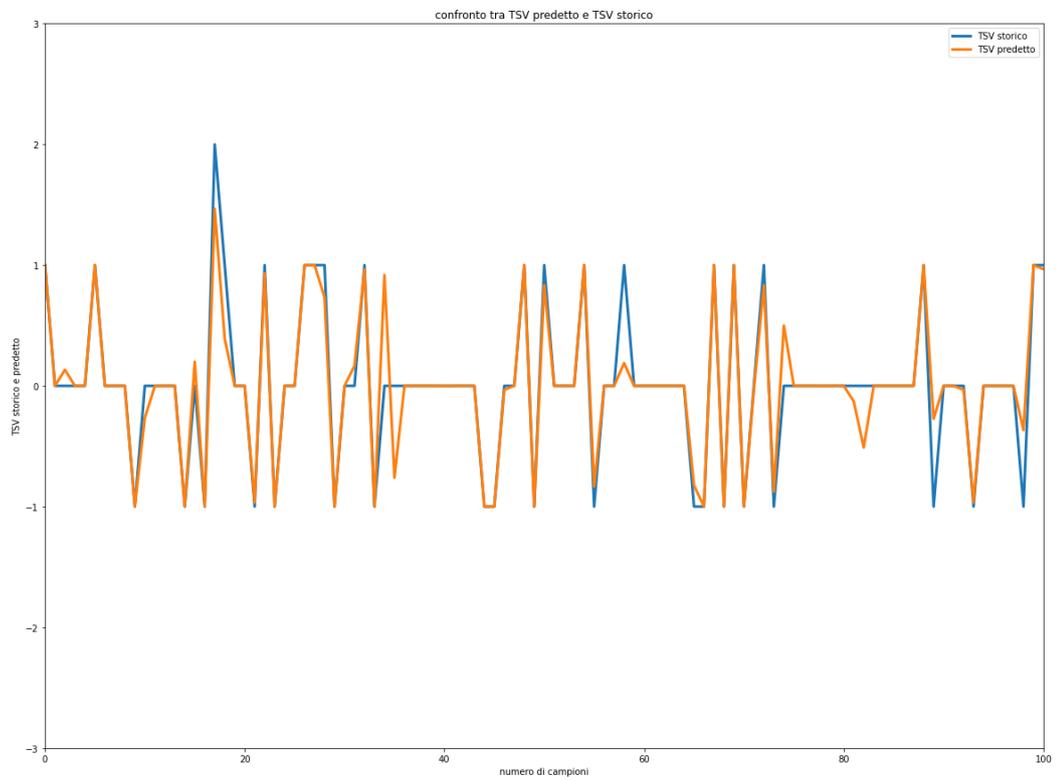


Figura 40: andamento della previsione RFR sul Dataset 'Focchi' di validazione

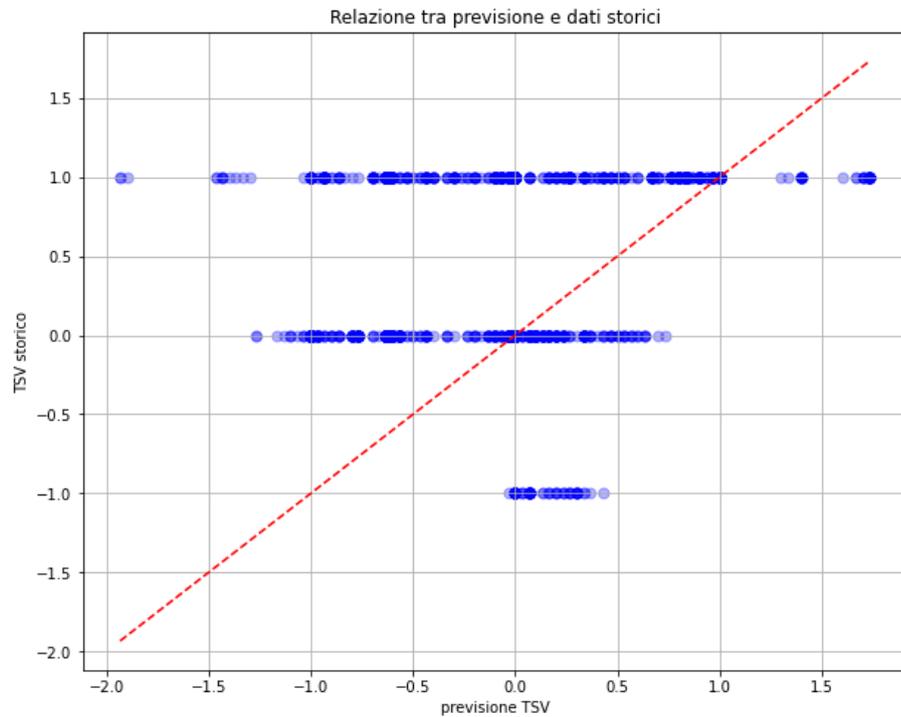


Figura 41: grafico a punti dal RFR al Dataset 'Focchi' di validazione

4.2 Analisi di correlazione del database ASHRAE

In questo paragrafo vengono mostrati i risultati ottenuti dall'analisi di correlazione dei parametri personali contenuti nel dataset analizzato presente nel database ASHRAE. Nella figura 42 viene illustrata la matrice di correlazione tra i parametri di temperatura interna, umidità relativa interna, peso, altezza, età, BMI, TSV e PMV.

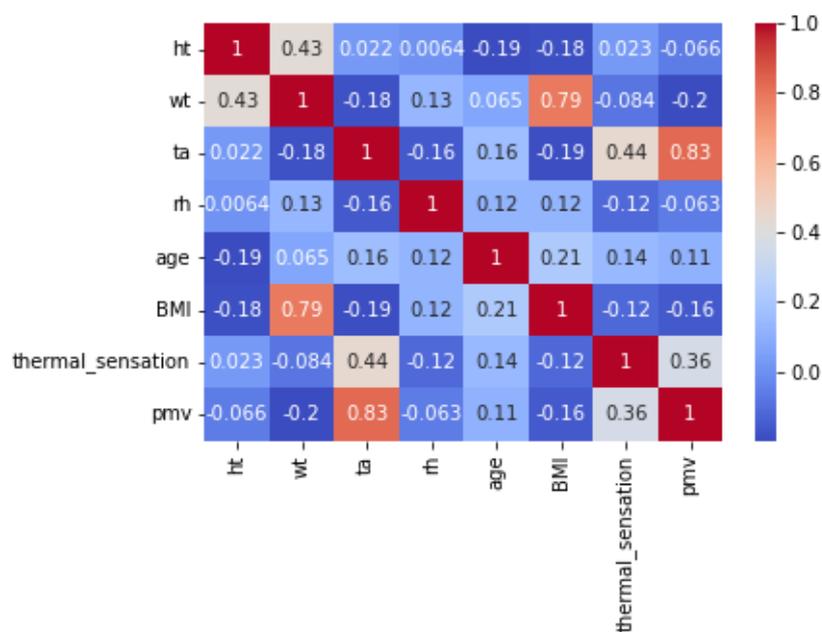


Figura 42: matrice di correlazione ottenuta dal dataset di ASHRAE database

Nella figura 43 sono contenuti i boxplot che evidenziano il trend del feedback in base alla fascia di BMI. È stato costruito un grafico per ognuna delle fasce di temperatura precedentemente scelte.

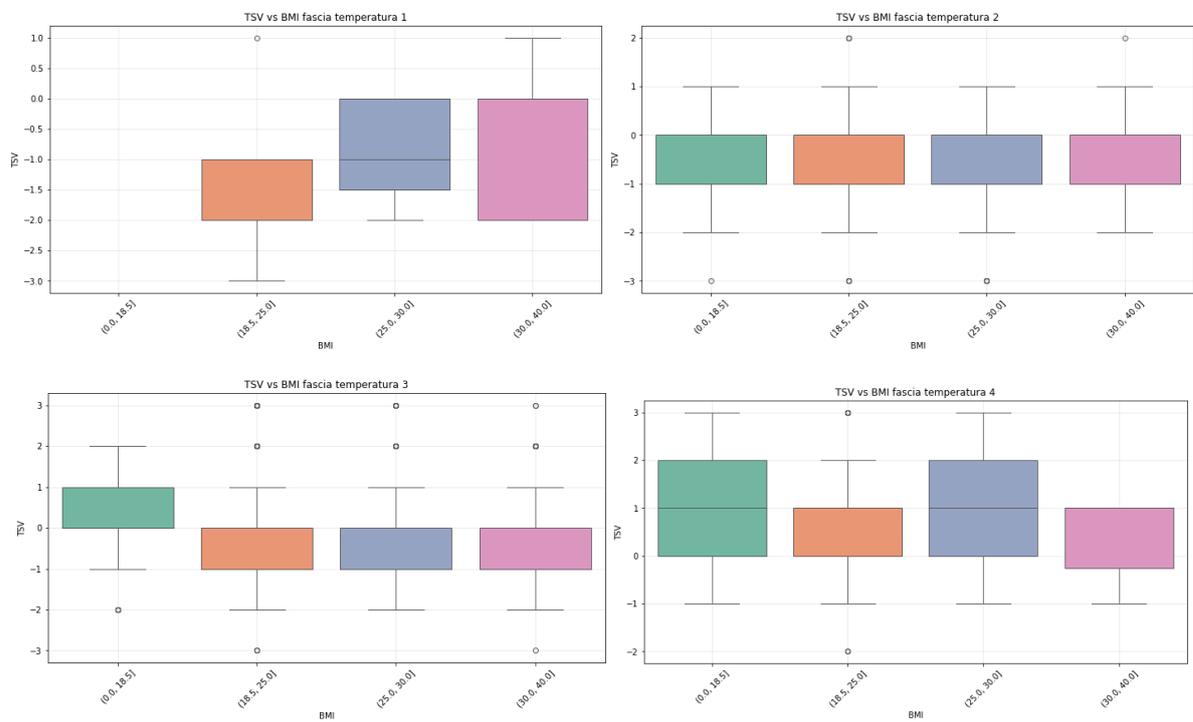


Figura 43: boxplot dei BMI in relazione al TSV secondo le 4 fasce 18-21°C (1), 21-24°C (2), 24-27°C (3), >27°C (4).

I boxplot in figura 44 mostrano il trend del TSV in base alla fascia di età, ed i grafici costruiti corrispondono alle fasce di temperatura.

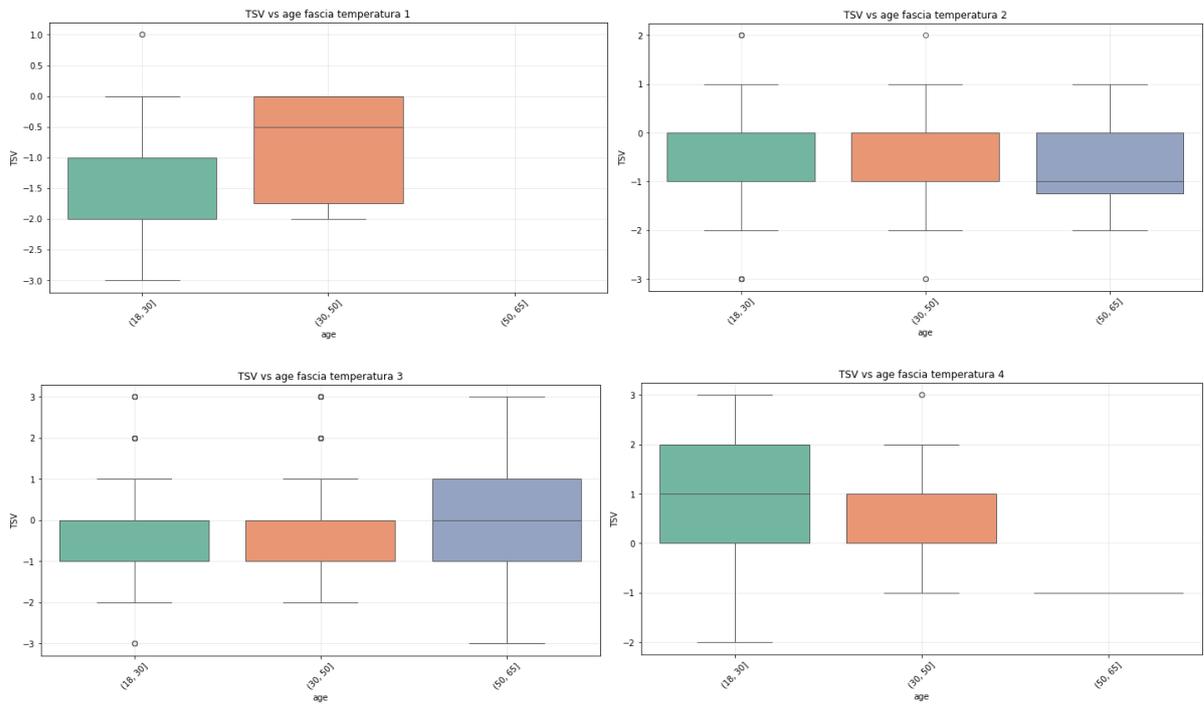


Figura 44: boxplot delle età in relazione al TSV secondo le 4 fasce 18-21°C (1), 21-24°C (2), 24-27°C (3), >27°C (4).

Nelle figure 45 e 46 sono stati riportati gli istogrammi in tre dimensioni che illustrano la frequenza del feedback (TSV) inviato dagli utenti in base alle fasce di BMI (figura 45) o alle fasce di età (figura 46) a cui appartenevano. Anche in questo caso, è presente un grafico per ogni fascia di temperatura scelta durante lo studio per l'analisi degli andamenti.

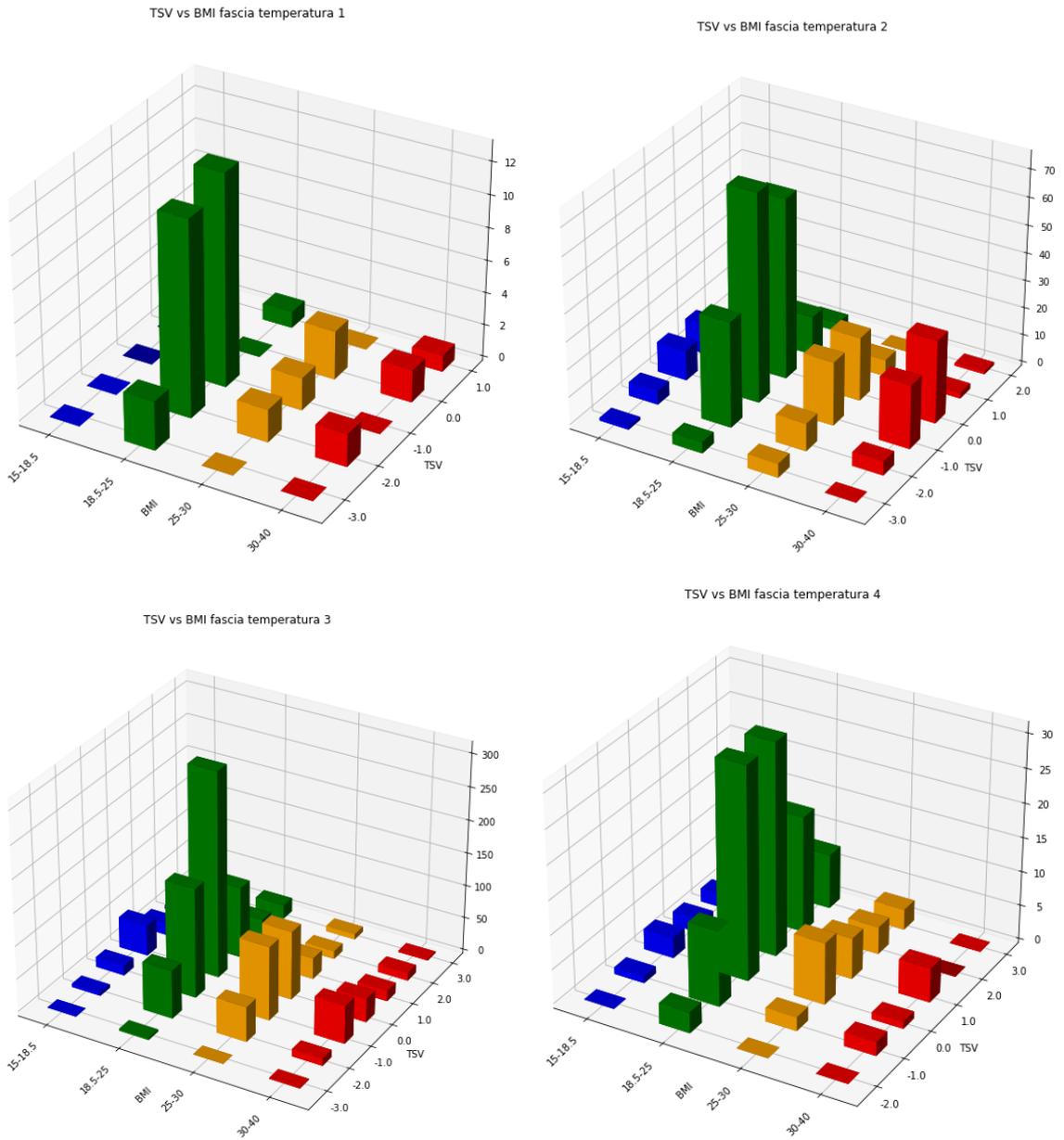


Figura 45: istogrammi dei BMI in relazione al TSV secondo le 4 fasce 18-21°C (1), 21-24°C (2), 24-27°C (3), >27°C (4).

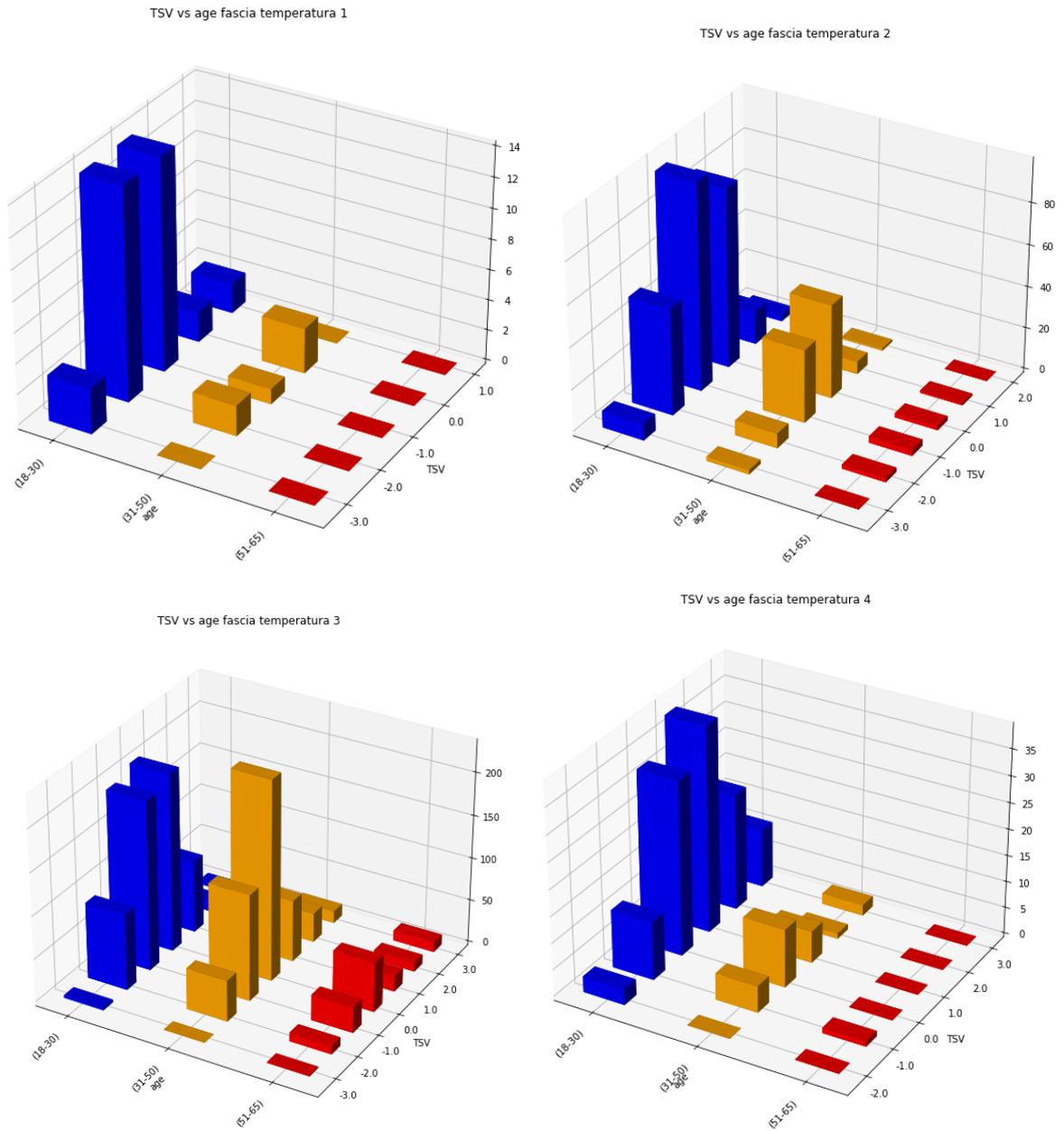


Figura 46: istogrammi delle età in relazione al TSV secondo le 4 fasce 18-21°C (1), 21-24°C (2), 24-27°C (3), >27°C (4).

5. Discussione

Dall'analisi delle metriche descritte nel capitolo precedente, emerge chiaramente come i modelli di previsione basati sulla classificazione risultino più efficaci rispetto a quelli basati sulla regressione. Tra i classificatori testati, il modello RF ha ottenuto le performance migliori, registrando un'accuratezza del 97% nel caso del dataset 'Focchi', un valore nettamente superiore rispetto al 79% raggiunto dal Support Vector Classifier (SVC).

L'efficacia del modello RFC è evidenziata dal grafico in figura 35, che riporta il confronto tra l'andamento del TSV reale e predetto, mostrando una stretta corrispondenza. Anche dalla matrice di confusione (figura 37) è ben visibile come la quasi totalità delle previsioni si trovino all'interno della diagonale, il che significa che il valore predetto corrisponde a quello reale. Tuttavia, nel caso del dataset 'FVH', le performance si sono rivelate inferiori: l'accuratezza più elevata è stata ottenuta nuovamente dal RFC, con un valore del 79%, mentre l'algoritmo SVC ha raggiunto una performance massima del 72%.

Passando all'analisi delle metriche relative ai regressori, i risultati migliori sono stati registrati per il modello Random Forest Regressor (RFR) applicato al dataset 'Focchi', con un MAE di 0,05, un MSE di 0,03 e un RSE di 0,16. Sebbene in fase di validazione del modello su dati nuovi, si sia riscontrato un aumento delle metriche di validazione sopra citate, il modello ha mantenuto la sua capacità predittiva come si può osservare in Figura 40 e 41. In particolare, dalla figura 40 si osserva che il valore di TSV predetto segue l'andamento dei valori reali. Inoltre, in alcuni tratti, la previsione e il valore reale rientrano nella stessa zona di comfort termico, fornendo una stima della sensazione termica percepita dal soggetto.

In generale, si può affermare che i classificatori producono risultati migliori dei regressori. Questo è probabilmente dovuto al fatto che, essendo il TSV un valore discreto, per la sua previsione sono maggiormente efficaci le tecniche di classificazione, le quali forniscono come output dei valori discreti, piuttosto che le tecniche di regressione, le quali risultano più adatte a situazioni di analisi di valori continui. Inoltre, attraverso il Dataset 'Focchi' le prestazioni della previsione sono migliori rispetto a quelle ottenute dal Dataset 'FVH'.

Questo risultato potrebbe essere attribuibile alle correlazioni significativamente più deboli tra i parametri riscontrate in FVH, come emerso dalla matrice di correlazione (figura 20) contenuta capitolo di analisi dati.

Per quanto riguarda l'analisi del dataset estratto dal database ASHRAE, dalla matrice di correlazione (figura 42) è possibile osservare che, in generale, il TSV (thermal sensation vote) è poco correlato con le caratteristiche personali degli utenti, mentre si registra una maggiore relazione con la temperatura interna ('ta'). Passando infatti ai boxplot (figure 43 e 44), si nota come il trend dei voti sia differente da grafico a grafico, i quali ognuno si riferisce ad una diversa fascia di temperatura.

In particolare, all'aumentare della temperatura, cresce anche il valore di TSV medio registrato dagli utenti, mentre non si riscontrano significative differenze di sensazione tra le varie fasce di BMI o di età. Dagli istogrammi presenti nelle figure 45 e 46 si legge un trend dei voti molto simile sia nel caso delle diverse fasce di BMI, sia per quanto riguarda le fasce d'età. Come già osservato dai boxplot, il valore di TSV maggiormente registrato si trova sempre all'interno delle fasce centrali di comfort, anche se aumenta leggermente all'aumentare delle fasce di temperatura. Emerge in particolare il fatto che, per la fascia di temperatura più bassa, i soggetti della fascia di BMI 18.5-25 (figura 45) tendano a soffrire maggiormente il freddo rispetto agli altri. Lo stesso si può dire per coloro che rientrano nella fascia d'età 18-30 (figura 46).

6. Conclusione

Il tema del comfort termico è ormai di grande rilevanza all'interno dell'ambito dell'ingegneria, dato che le persone passano sempre più tempo all'interno di luoghi chiusi. Studi precedenti [8][9] hanno messo in evidenza come la soggettività della sensazione termica dell'utente, la quale è influenzata da fattori fisiologici e psicologici, renda complesso stabilire i parametri ambientali da impostare per soddisfare la maggior parte dei soggetti, e hanno sviluppato tecniche sempre più sofisticate per la misura personalizzata del comfort termico, cercando anche di ottimizzare il consumo energetico da parte dei sistemi di climatizzazione. In particolare, i modelli Machine Learning si sono rivelati significativamente efficaci nell'ambito della previsione del comfort termico degli utenti.

Questo studio si è concentrato sullo sviluppo e sull'implementazione di una metodologia personalizzata basata sul ML per la previsione della sensazione termica espressa in thermal sensation vote (TSV), per misurare il comfort all'interno di un ambiente chiuso. L'obiettivo è stato quello di cercare di superare i limiti emersi dagli studi che lo hanno preceduto, quali ad esempio la necessità di analizzare parametri fisiologici, non sempre possibili da ottenere, il problema di dover analizzare dataset eterogenei, di cui spesso non si è a disposizione, e il fatto che richiedere in maniera periodica un riscontro sulla sensazione termica risulta invasivo per gli utenti. Questa ricerca ha impiegato dataset contenenti parametri ambientali e feedback degli utenti, ai quali, una volta analizzati, sono stati applicati metodi di ML per la previsione del comfort termico. Dai risultati ottenuti, è emerso come i meccanismi di classificazione si siano rivelati maggiormente efficaci rispetto a quelli di regressione, con buoni valori di accuratezza (97% per RFC, 79% per SVC), e che la correlazione tra i parametri in gioco sia fondamentale nello sviluppo di un modello di previsione del comfort. Data l'importanza della raccolta di feedback in merito alla sensazione termica, è stato poi stilato un protocollo di misura per la raccolta dei dati ambientali e del feedback degli utenti all'interno dell'ambiente di ricerca, il quale rappresenta uno sviluppo in ottica futura, in quanto potrà essere utilizzato per l'acquisizione di un dataset personale, che potrà servire da base per l'applicazione dei modelli di previsione sviluppati. Con l'obiettivo di capire l'influenza delle caratteristiche personali degli utenti sul loro feedback, è stata infine condotta un'analisi su dati provenienti dal database online "ASHRAE Global Thermal Comfort Database". I risultati ottenuti dimostrano come ci sia poca correlazione tra questi parametri ed il feedback, il cui andamento è molto simile nei vari casi presi in considerazione. Ricerche future potrebbero basarsi sui modelli di machine learning sviluppati in questo studio ed applicarli a dataset più ampi, sperimentare l'implementazione di differenti modelli presenti in letteratura ponendo particolare attenzione sull'analisi di correlazione tra le variabili scelte per la previsione. Inoltre, a partire dai dati raccolti durante la campagna presentata, si potrebbe condurre un'analisi approfondita in merito a questi. Infine, un approfondimento dello studio fatto sul database potrebbe essere rappresentato dall'applicazione di ulteriori tecniche di clustering.

Bibliografia

- [1] ASHRAE, ANSI/ASHRAE Standard 55 - «Thermal Environmental Conditions for Human Occupancy, American society of heating, refrigerating and air-conditioning engineers,» Atlanta, 2010.
- [2] N.E. Klepeis, W.C. Nelson, et al., «The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants,» J. Expo. Anal. Environ. Epidemiol. 11 (3) (2001) 231–252.
- [3] Sokratis Papadopoulos, Constantine E. Kontokosta, Alex Vlachokostas, Elie Azard, «Rethinking HVAC temperature setpoints in commercial buildings: The potential for zero-cost energy savings and comfort improvement in different climates,» 2019.
- [4] Panayiotis M. Papadopoulos, Ioanna Kyprianou, Muhammad Salman Shahid, Silvia Erba, Frédéric Wurtz, Benoit Delinchant, Peter Riederer, Mohammadreza Aghaei, Salvatore Carlucci, «Indoor thermal comfort analysis for developing energy-saving strategies in buildings,» 2023.
- [5] Ghezlane Halhoul Merabet, Mohamed Essaaidi, Mohamed Ben Haddou, Basheer Qolomany, Junaid Qadir, Muhammad Anan, Ala Al-Fuqaha, Mohamed Riduan Abid, Driss Benhaddou, «Intelligent building control systems for thermal comfort and energy-efficiency: A systematic review of artificial intelligence-assisted techniques,» 2021.
- [6] E. C. F. STANDARDIZATION, «UNI EN ISO 7730,» 2005.
- [7] Nan Ma, Dorit Aviv, Hongshan Guo, William W. Braham, «Measuring the right factors: A review of variables and models for thermal comfort and indoor air quality,» University of Pennsylvania, Philadelphia, 2021.
- [8] Noël Djongyang, René Tchinda, Donatien Njomo, «Thermal comfort: A review paper,» 2010.
- [9] Wenjie Song, John Kaiser Calautit, «Inclusive comfort: A review of techniques for monitoring thermal comfort among individuals with the inability to provide accurate subjective feedback,» UK, 2024.
- [10] Tingzhang Liu, Linyi Jin, Chujun Zhong, Fan Xue, «Study of thermal sensation prediction model based on support vector classification (SVC) algorithm with data

preprocessing,» School of Mechatronic Engineering and Automation, Shanghai University, China, 2022.

[11] Changyong Yu, Baizhan Li, Simo Kilpelaine, Hong Liu, Yuxin Wu, Baofan Chen, Risto Kosonen, «Performances of machine learning algorithms for individual thermal comfort prediction based on data from professional and practical settings,» 2022.

[12] E. C. O. STANDARDIZATION, «UNI EN ISO 7726,» 2002.

[13] G. Barone, A. Buonomano, C. Forzano, G.F. Giuzio, A. Palombo, G. Russo, «A new thermal comfort model based on physiological parameters for the smart design and control of energy-efficient HVAC systems,» Department of Industrial Engineering - University of Naples Federico II, Naples, Italy, 2023

[14] Liu W, et al., «Evaluation of calculation methods of mean skin temperature for use in thermal comfort study,» 2011.

[15] Tansey EA, Johnson CD. «Recent advances in thermoregulation,» Adv Physiol Educ, 2015.

[16] Joshi A, et al. «A three-dimensional thermoregulatory model for predicting human thermophysiological responses in various thermal environments,» Build Environ, 2022.

[17] Lim CL, Byrne C, Lee JKW. «Human thermoregulation and measurement of body temperature in exercise and clinical settings,» Ann Acad Med Singapore, 2008.

[18] Luo M, et al. «Human metabolic rate and thermal comfort in buildings: The problem and challenge,» Build Environ, 2018.

[19] Yeom DJ, Delogu F. «Local body skin temperature-driven thermal sensation predictive model for the occupant's optimum productivity,» Build Environ, 2021.

[20] Kevin D. Hall, Steven B. Heymsfield, Joseph W. Kemnitz, Samuel Klein, Dale A. Schoeller, and John R. Speakman, «Energy balance and its components: implications for body weight regulation,» From the American Society of Nutrition, 2023.

[21] Qiantao Zhao, Zhiwei Lian, Dayi Lai, «Thermal comfort models and their developments: A review,» Department of Architecture, School of Design, Shanghai Jiao Tong University, China, 2021.

- [22] P. O. Fanger, «FUNDAMENTALS OF THERMAL COMFORT,» 1988.
- [23] C. Buratti, P. Ricciardi e V. M., «HVAC systems testing and check: A simplified model to predict thermal comfort conditions in moderate environments,» 2013.
- [24] Sara Casaccia, Kristiina Jokinen, Riccardo Naccarelli, Gian Marco Revel, «Well-being and comfort of ageing people based on indoor environmental conditions: preliminary study on human-coach conversation,» 2022.
- [25] Vittoria Cipollone, Nicole Morresi, Sara Casaccia, Gian Marco Revel, «Robot-based measurement of comfort through thermal infrared imaging and wearable sensors,» 2023.
- [26] Hansaem Park , Dong Yoon Park, «Prediction of individual thermal comfort based on ensemble transfer learning method using wearable and environmental sensors,» South Korea, 2022.
- [27] Nicole Morresi, Sara Casaccia, Matteo Sorcinelli, Marco Arnesano, Amaia Uriarte, J. Ignacio Torrens-Galdiz, and Gian Marco Revel, «Sensing Physiological and Environmental Quantities to Measure Human Thermal Comfort Through Machine Learning Techniques,» 2021.
- [28] Ilaria Pigliautile, Sara Casaccia, Gian Marco Revel, Nicole Morresi, Marco Arnesano, Anna Laura Pisello, «Assessing occupants' personal attributes in relation to human perception of environmental comfort: Measurement procedure and data analysis,» 2020.
- [29] <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.interpolate.html>
- [30] Hongchao Zhang, Xu Yang, Rang Tu, Jian Huang, Yiran Li, «Thermal Comfort Modeling of Office Buildings Based on Improved Random Forest Algorithm,» China, 2022
- [31] <https://www.geeksforgeeks.org/ml-classification-vs-regression/>
- [32] Wenqiang Jing, Jiayan Liu, Zeming Qin, Teng Mu, Zhemin Ge, Yuting Dong, «Evaluating and enhancing the applicability of thermal comfort indices in diverse outdoor environments using Bayesian and random forest regression,» 2024.