

---

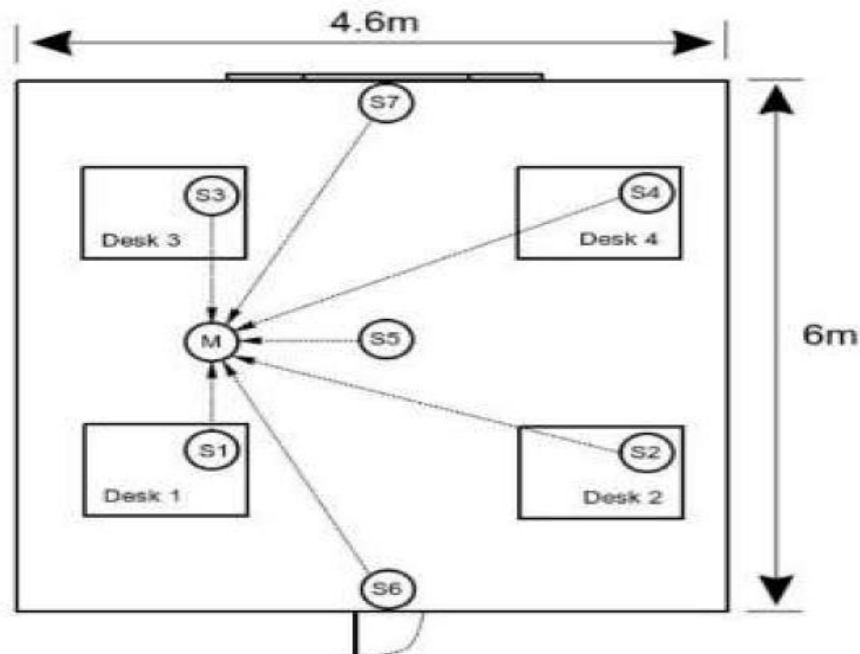
# MONITORAGGIO DI PRESENZE PER UNA RIDUZIONE DEGLI SPRECHI



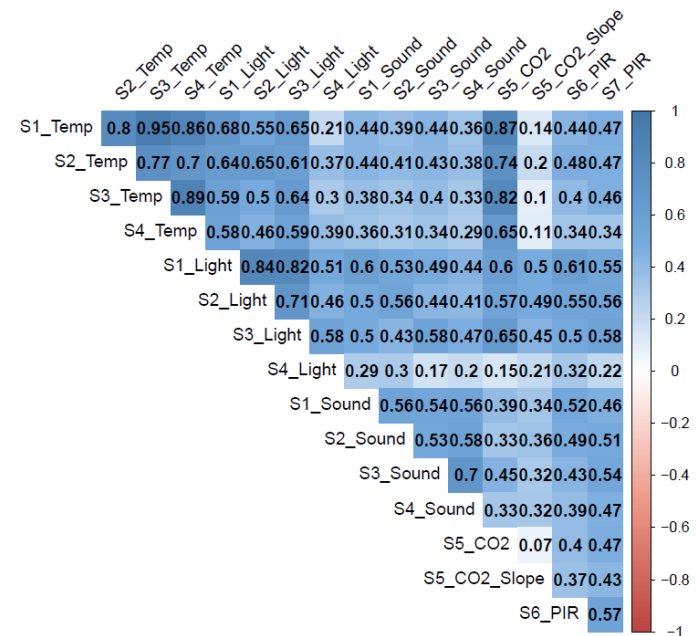
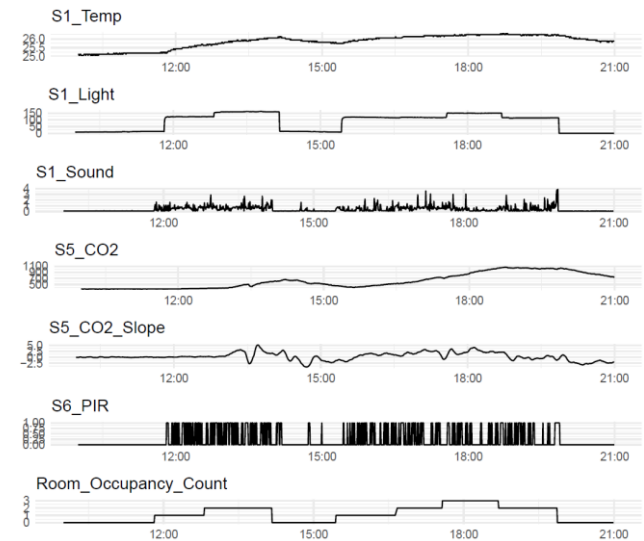
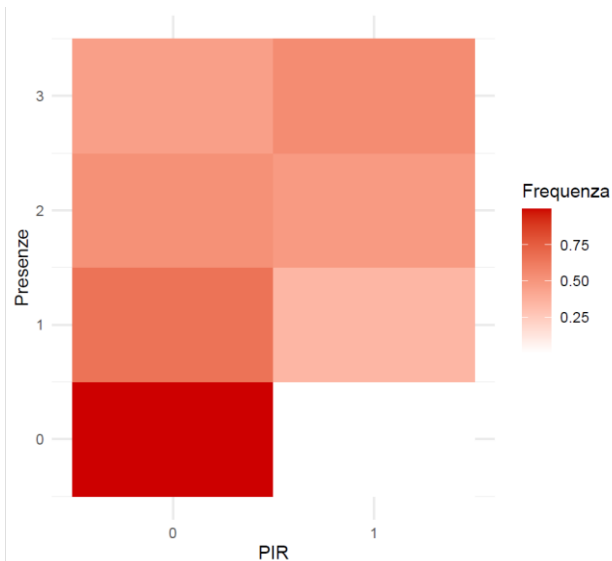
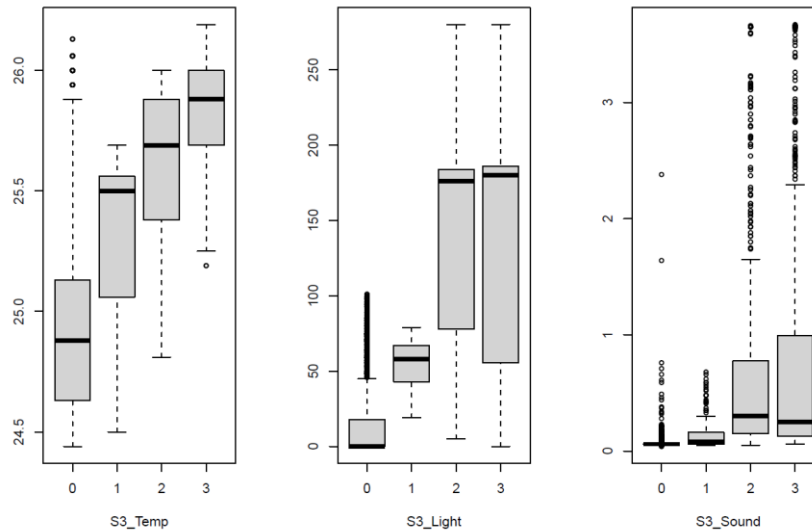
Cecilia Trevisi - mat.862637  
Giacomo Rabuzzi - mat.864452  
Riccardo Pajno - mat.864557

**Dataset** di 10122 osservazioni e 20 variabili:

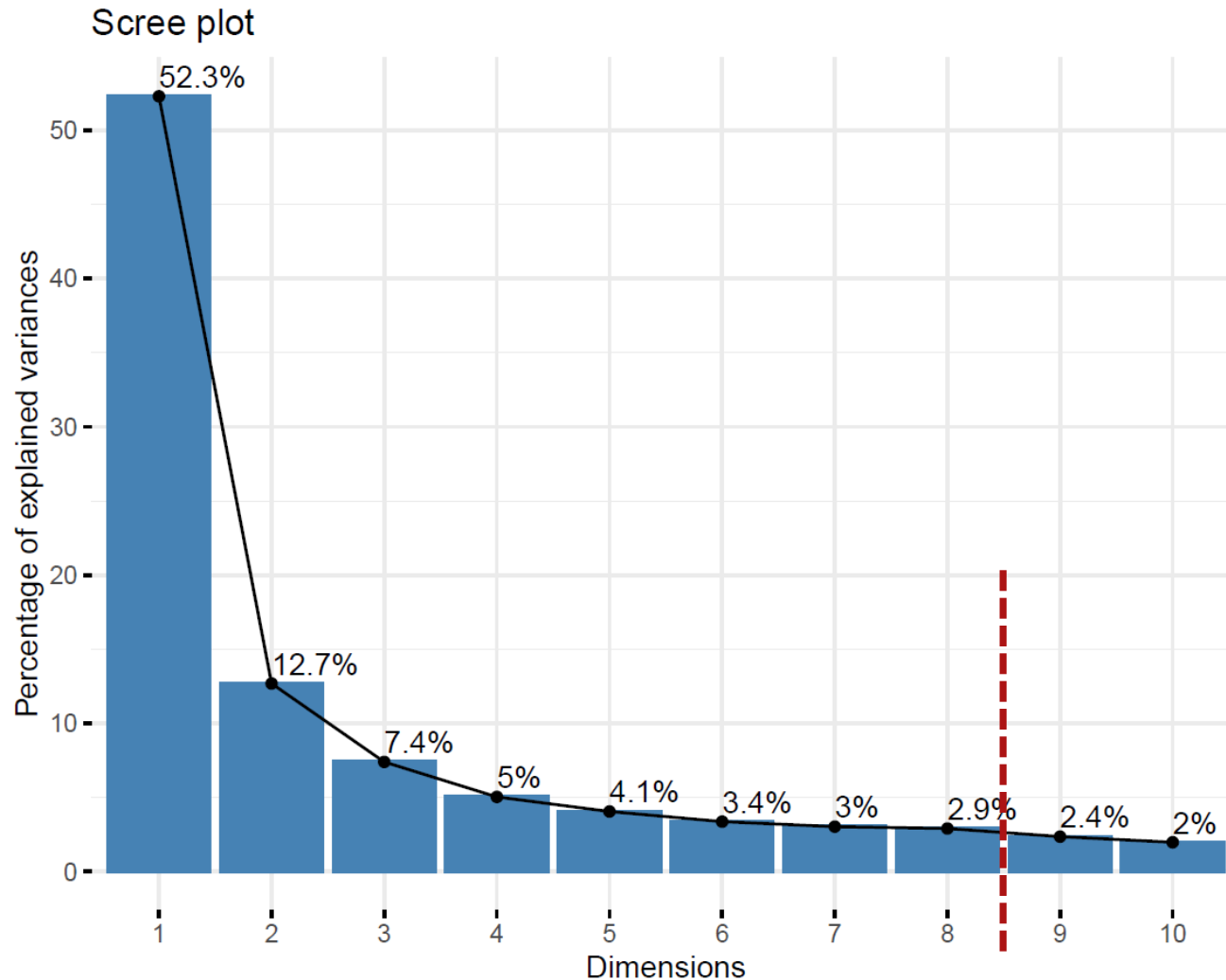
- Date: data della rilevazione
- Time: tempo della rilevazione espresso in ore/minuti/secondi
- S1-4 Temp: temperatura in gradi Celsius dei quattro sensori
- S1-4 Light: luce in 1 Lux dei quattro sensori
- S1-4 Sound: suono in 0.01V\* dei quattro sensori
- S5 CO2: CO2 in 5ppm dei quattro sensori
- S5 CO2 Slope: pendenza della CO2
- S6-7 PIR: presenza -1 o assenza-0 di movimento
- Room\_Occupancy\_Count: numero di persone all'interno della stanza



## Analisi dei *legami* tra variabili



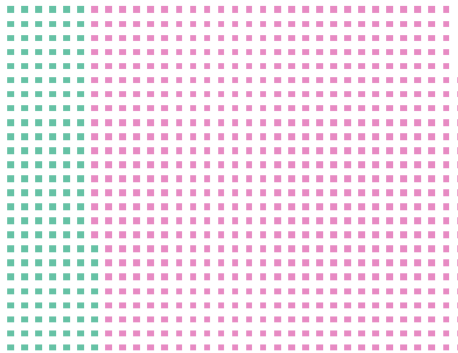
Preparazione dei dati → **PCA** sull'intero dataset



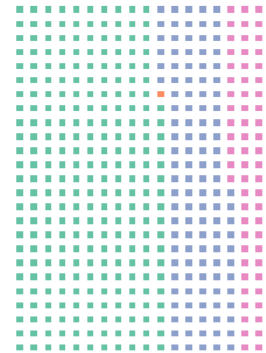
## 1. K-MEANS

### K-Medie

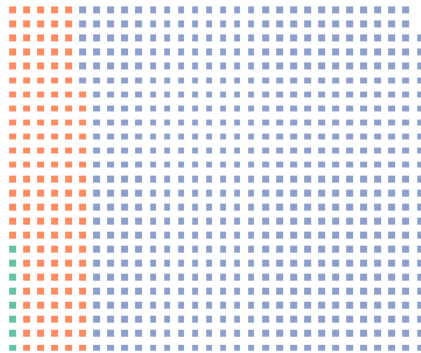
0 presenze



1 presenza



2 presenze



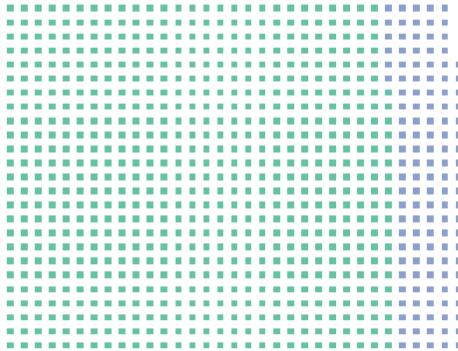
3 presenze



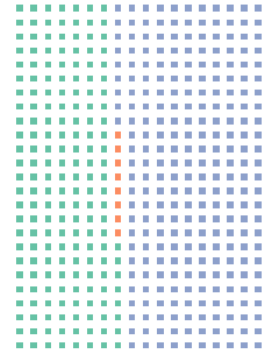
## 2. CLUSTERING GERARCHICO

### Clustering gerarchico

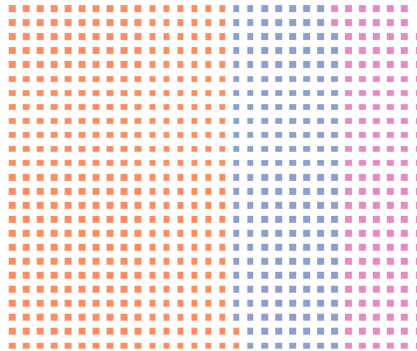
0 presenze



1 presenza



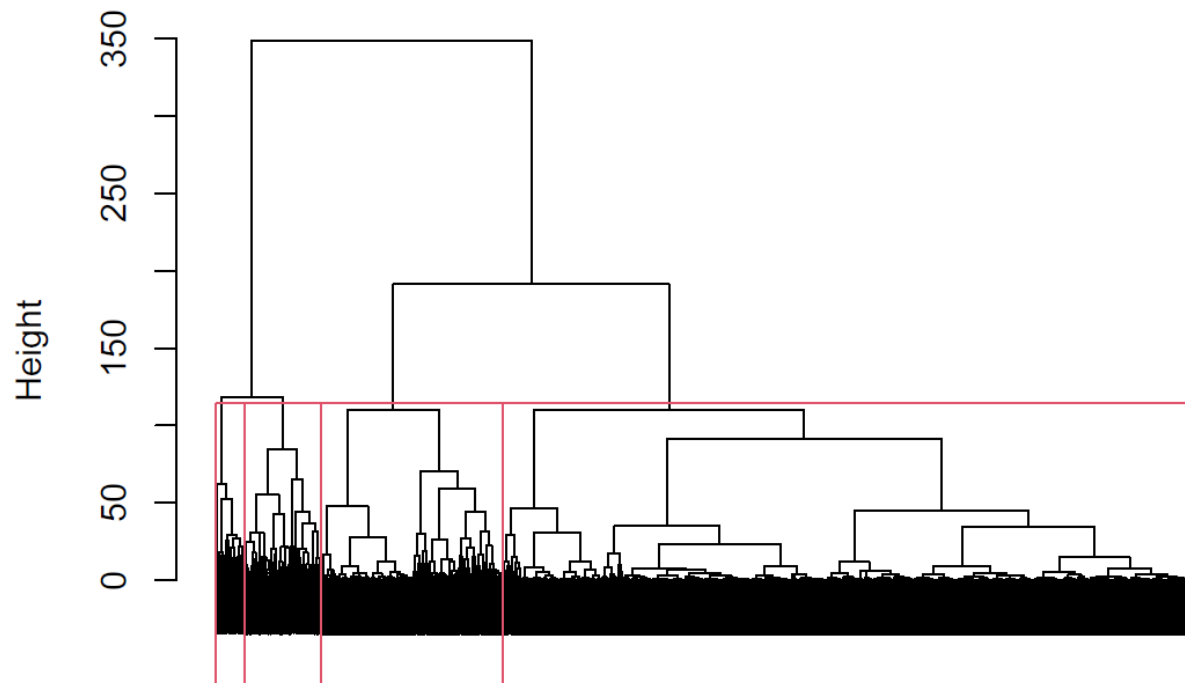
2 presenze



3 presenze



### Cluster Dendrogram

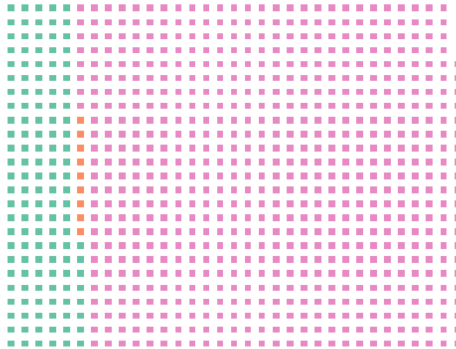


di.1  
hclust (\*, "ward.D2")

### 3. MODEL-BASED CLUSTERING

#### Model based clustering (struttura VVV)

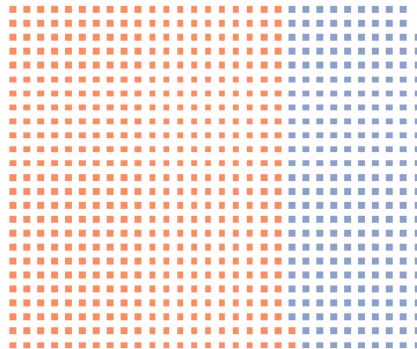
0 presenze



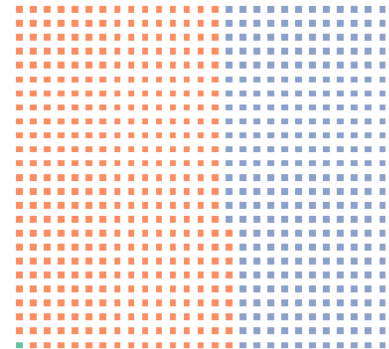
1 presenza



2 presenze



3 presenze





**Preparazione** dei dati:

- Suddivisione del dataset completo in training e test con rispettive proporzioni 80% e 20%
- Applicazione della PCA sul training e calcolo degli scores sul test
- Addestramento degli algoritmi tramite una k-fold cross validation con  $k=5$

## 1. K-NN

Modello ottimale con  $k=5$

Metrica	Classe 0	Classe 1	Classe 2	Classe 3	Media
F1	0.997	0.951	0.944	0.940	0.958
BA	0.994	0.980	0.970	0.9725	0.979

**Tabella 2.1:** Risultati di F1 e Balanced Accuracy (BA)

## 2. SUPPORT VECTOR MACHINE

Modello ottimale con kernel radiale, costo = 27.89 e un  $\gamma = 0.014$

Metrica	Classe 0	Classe 1	Classe 2	Classe 3	Media
F1	0.998	0.975	0.948	0.936	0.964
BA	0.995	0.981	0.974	0.967	0.979

**Tabella 2.2:** Risultati di F1 e Balanced Accuracy (BA)

### 3. RANDOM FOREST

Modello ottimale con 355 alberi e 2 features

Metrica	Classe 0	Classe 1	Classe 2	Classe 3	Media
F1	0.997	0.938	0.931	0.937	0.951
BA	0.994	0.962	0.963	0.976	0.974

**Tabella 2.3:** Risultati di F1 e Balanced Accuracy (BA)

## 4. RETE NEURALE

Modello ottimale con 16 neuroni nell'hidden layer e coefficiente di regolarizzazione  $\lambda = 0.07$

Metrica	Classe 0	Classe 1	Classe 2	Classe 3	Media
F1	0.999	0.981	0.955	0.953	0.972
BA	0.995	0.981	0.974	0.981	0.983

**Tabella 2.4:** Risultati di F1 e Balanced Accuracy (BA)

## 5. PROPORTIONAL ODDS

$$\text{logit}(P(G \leq \mathcal{G}_j \mid \mathbf{X} = \mathbf{x})) = \beta_{0j} + \beta^\top \mathbf{x}, \quad j = 1, \dots, k - 1,$$

Metrica	Classe 0	Classe 1	Classe 2	Classe 3	Media
F1	0.982	0.389	0.705	0.652	0.682
BA	0.925	0.664	0.829	0.805	0.806

**Tabella 2.5:** Risultati di F1 e Balanced Accuracy (BA)

EFFICACIA

Algoritmo	Purity	Silhouette Media
K-medie	0.838	0.460
K-medoidi	0.713	0.328
Kernel K-medie	0.730	-0.260
Gerarchico	0.828	0.469
Model-based	0.839	0.414

Tabella 3.1: Valori di Purity e Silhouette per vari algoritmi.

EFFICIENZA

Algoritmo	Tempo
K-medie	0.2 s
K-medoidi	35.66 s
Kernel K-medie	603.88 s
Gerarchico	5.06 s
Model-based	2.78 s

Tabella 3.4: Algoritmi di classificazione e tempo computazionale

Algoritmo	Sensitivity Classe 0	Media sensitivity altre classi
KNN	0.996	0.948
Support Vector Machine	0.998	0.951
Random Forest	0.996	0.938
Rete neurale	0.999	0.960
Proportional Odds	0.993	0.550

Tabella 3.2: Valori di Sensitivity per i vari algoritmi.

Algoritmo	Tempo
KNN	1.48 s
Support Vector Machine	0.63 s
Random Forest	1.73 s
Rete neurale	4.04 s
Proportional Odds	1.83 s

Tabella 3.5: Algoritmi di classificazione e tempo computazionale