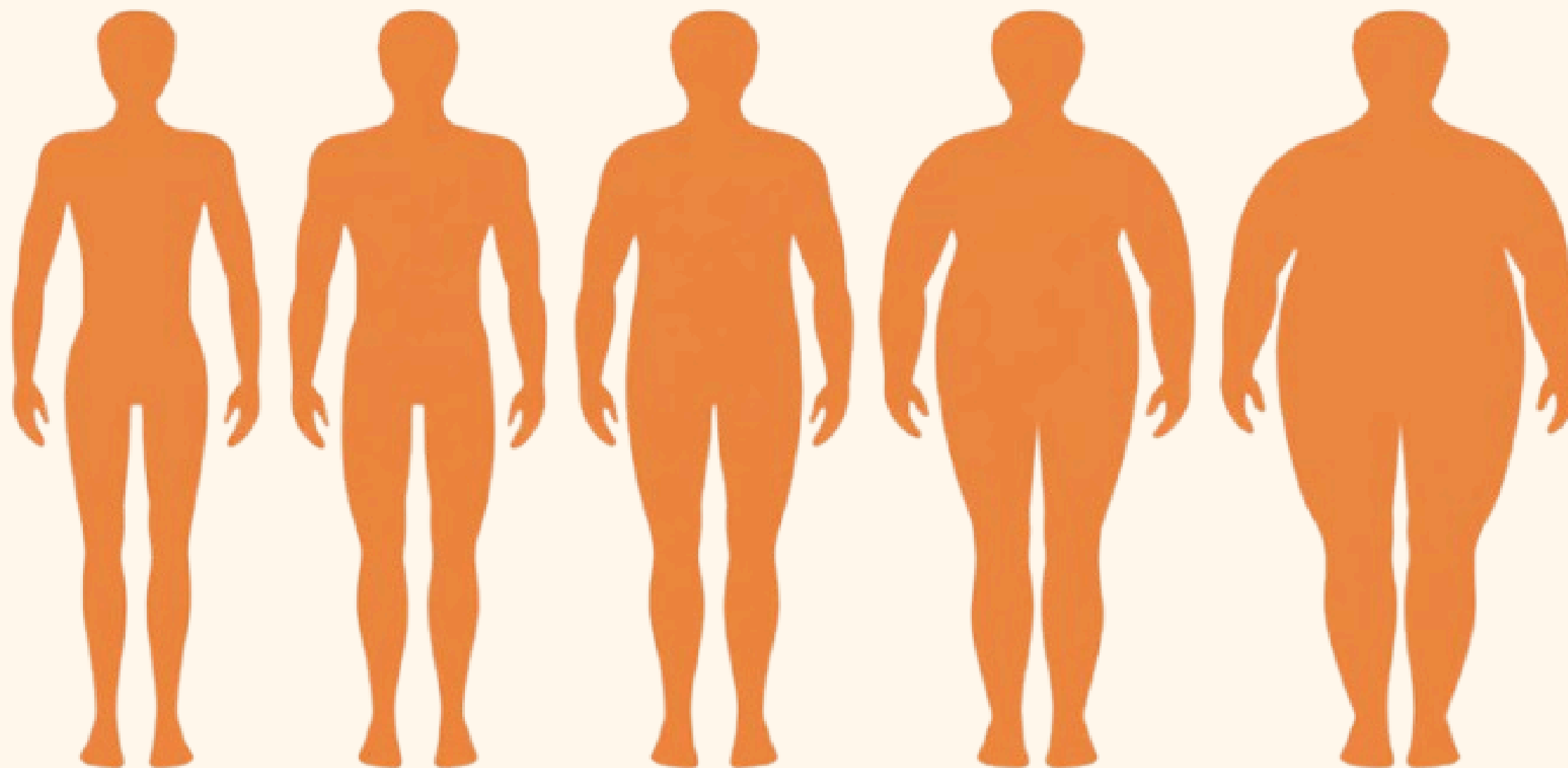
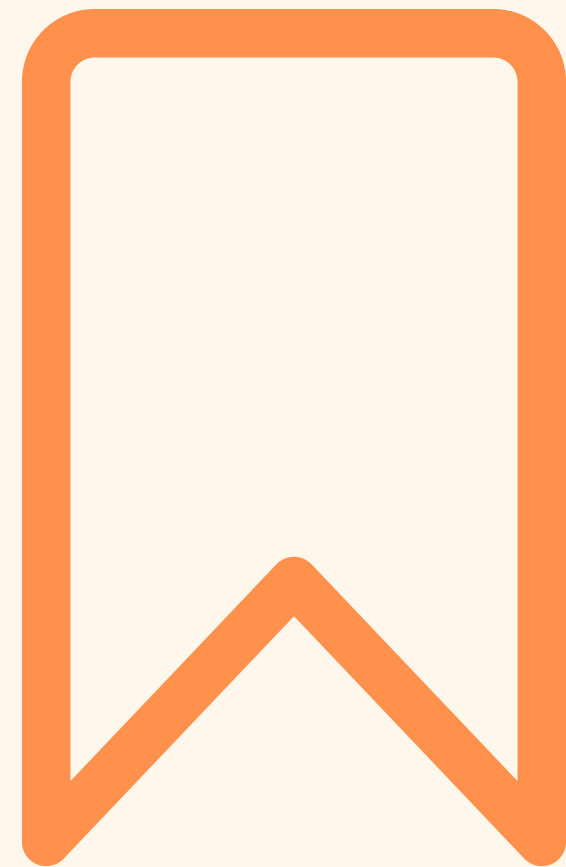


# ANALISI E CLASSIFICAZIONE DELL'OBESITÀ TRAMITE DATI COMPORTAMENTALI E FISICI



# INDICE



**1**

**OBIETTIVO**

**2**

**DATASET**

**Spiegazione  
e visualizzazione**

**3**

**CLASSIFICAZIONE**

**Algoritmo e risultati**

**4**

**VALUTAZIONE**

**del modello**

# OBIETTIVO

Effettuare un'analisi dei dati relativi allo stile di vita, alle abitudini dietetiche e ai parametri fisici di un gruppo di persone. Lo scopo è determinare quali specifici tratti e comportamenti sono predittivi di una diagnosi di obesità. In seguito, sviluppare un modello di machine learning capace di classificare gli individui esaminati in due gruppi distinti: normopeso e obeso



# COSA RAPPRESENTA IL DATASET?



Il dataset in questione contiene dati su oltre 2 mila individui descrivendo caratteristiche fisiche, abitudini alimentari e comportamentali legate allo stile di vita. Le feature includono indicatori come frequenza dell'attività fisica, consumo di cibi calorici, idratazione e utilizzo della tecnologia. Il target categorizza gli individui in due categorie:

- **Sottopeso e normopeso con 0**
- **Sovrappeso e obeso con 1**

# DATASET

2111 ISTANZE 16 FEATURES



## Feature categoriche

<b>FCVC</b>	frequenza consumo verdure
<b>NCP</b>	numero pasti giornalieri
<b>CH2O</b>	frequenza consumo di acqua
<b>CAEC</b>	consumo di cibo tra i pasti
<b>FAF</b>	frequenza attività fisica
<b>TUE</b>	frequenza uso dispositivi tecnologici
<b>MTRANS</b>	mezzo di trasporto utilizzato maggiormente
<b>CALC</b>	frequenza consumo di alcolici

## Feature continue

**ETÀ**

**ALTEZZA**

**PESO**

## Feature binarie

**Storia familiare  
con sovrappeso**

se un membro della famiglia ha  
sofferto o soffre di obesità o  
sovrappeso

**SCC**

controllo delle calorie giornaliero

**SMOKE**

fa utilizzo di sigarette

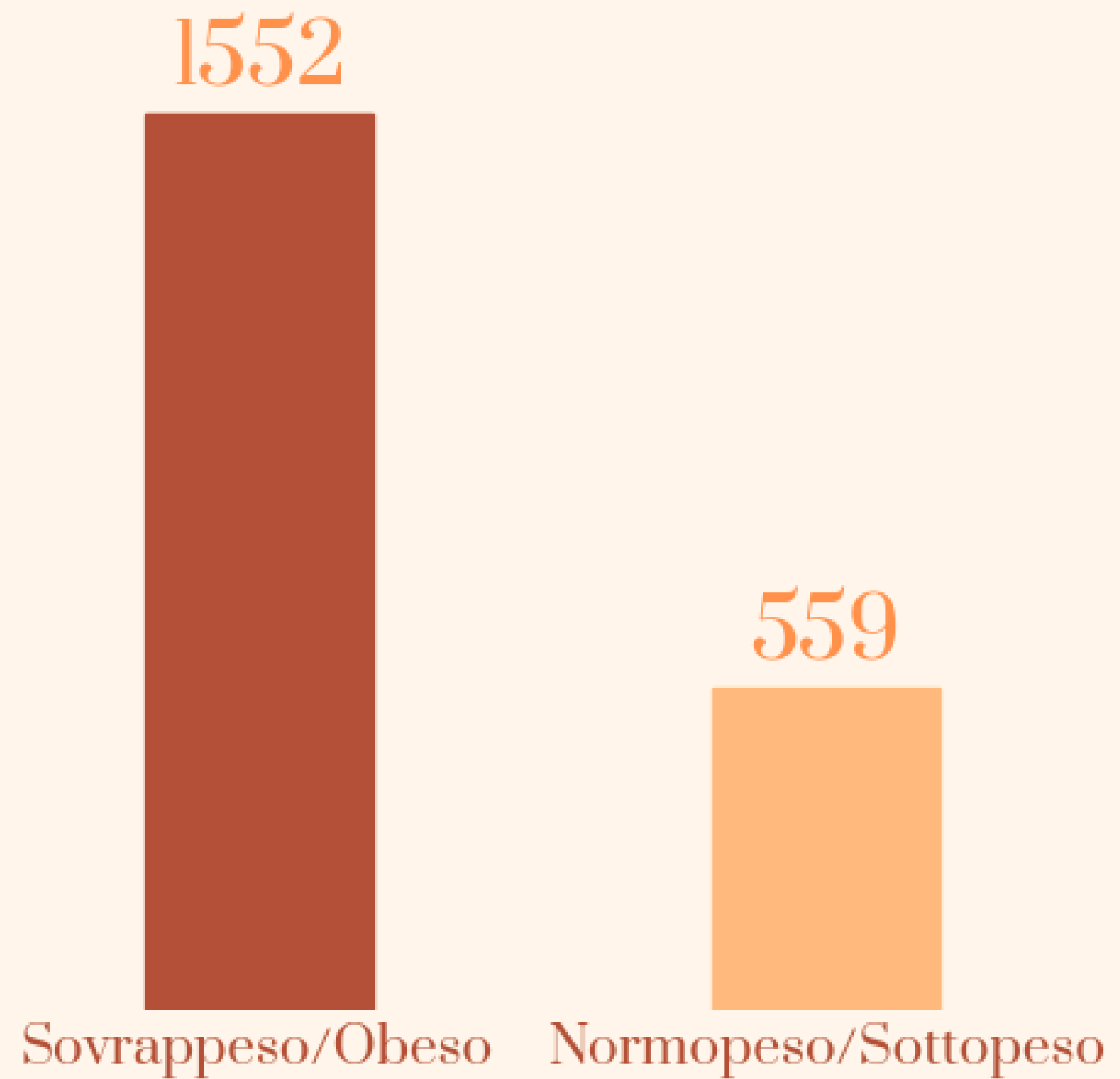
**FAVC**

consumo di cibi altamente calorici

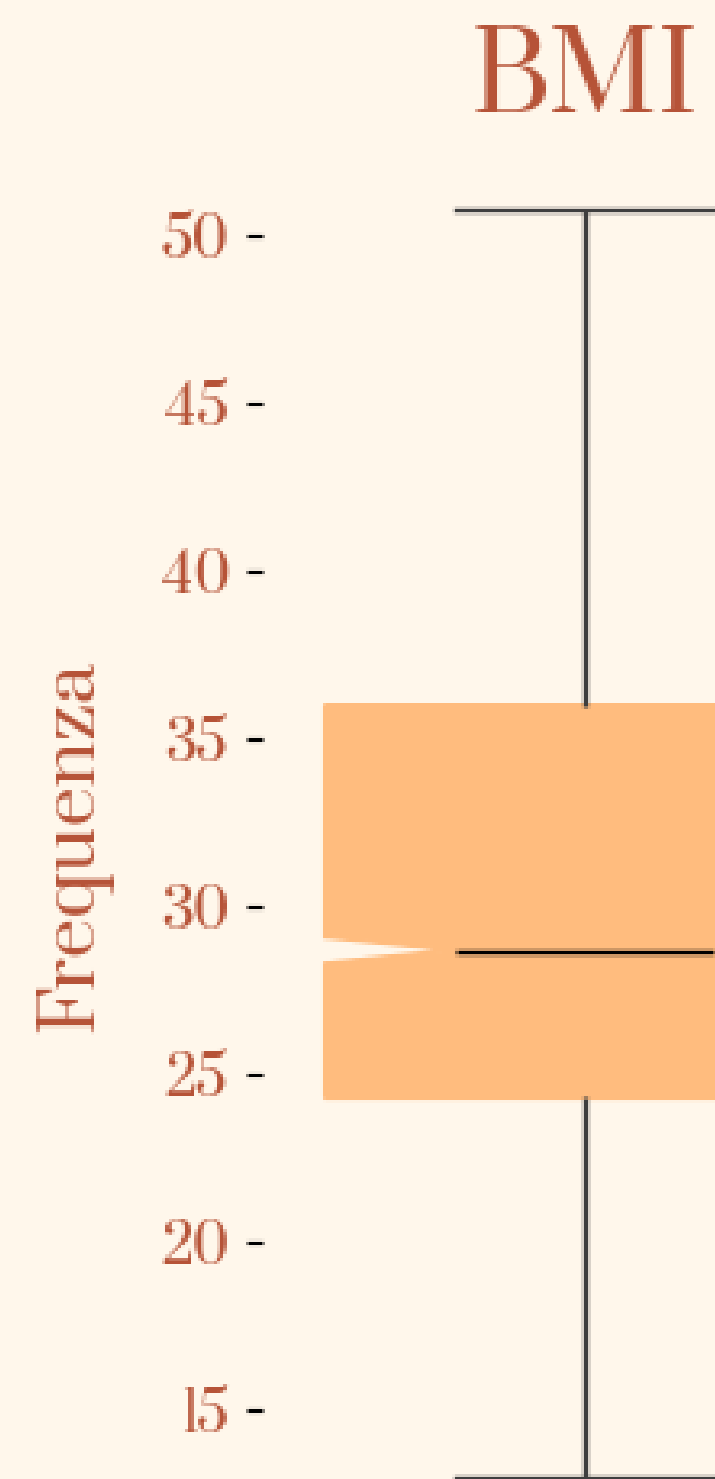
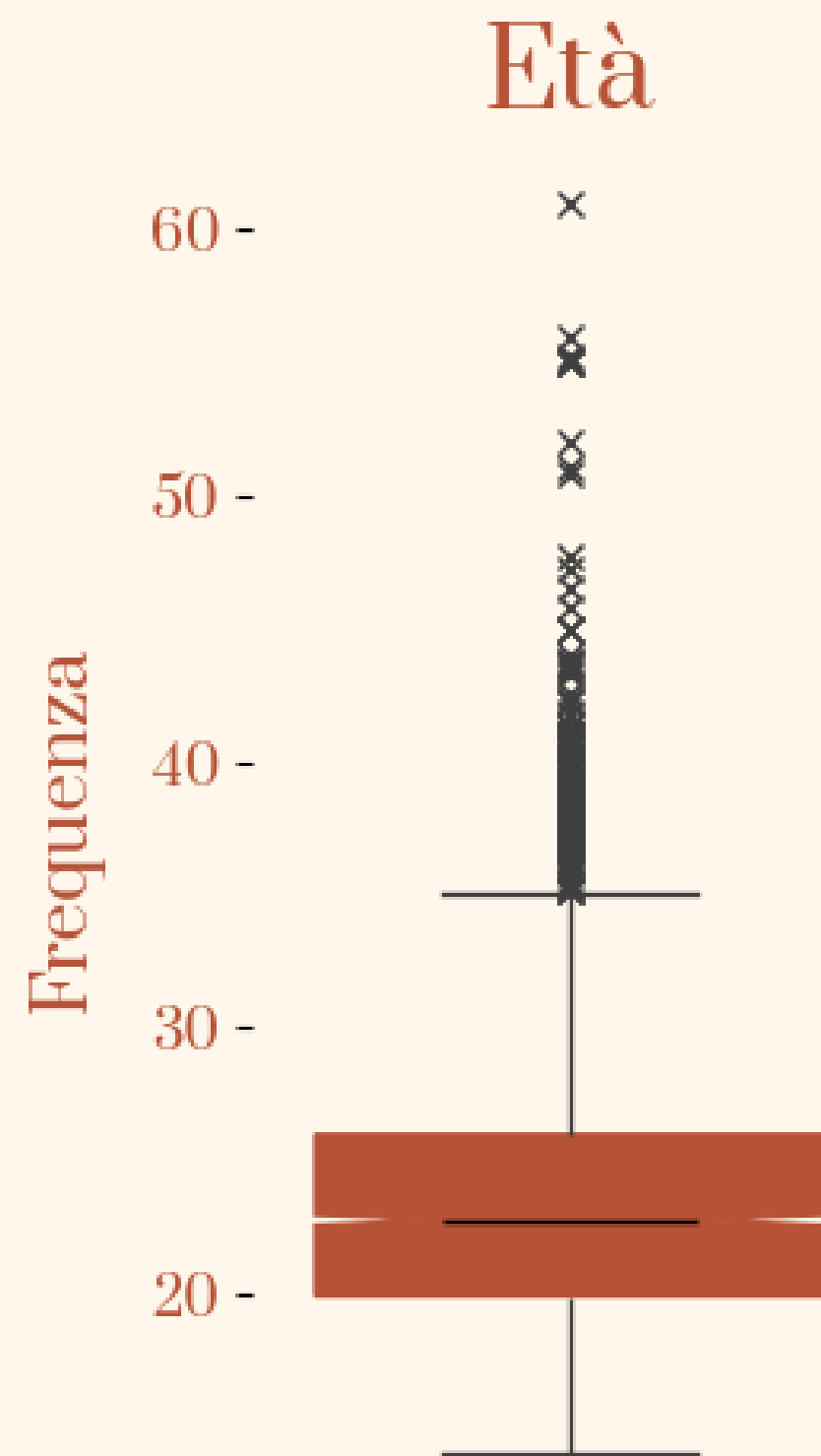
**GENERE**

maschio / femmina

# TARGET

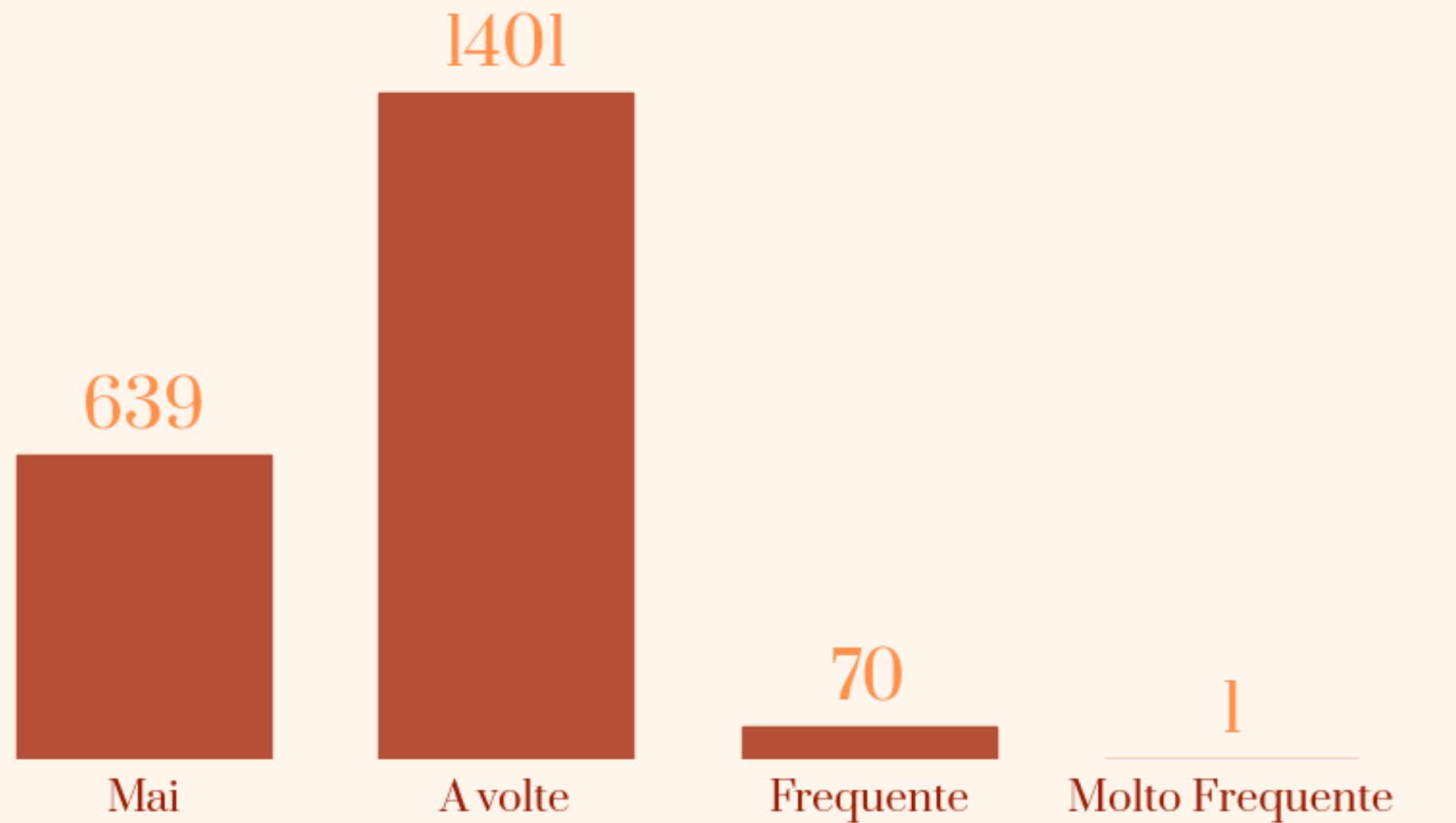


# FEATURE CONTINUE

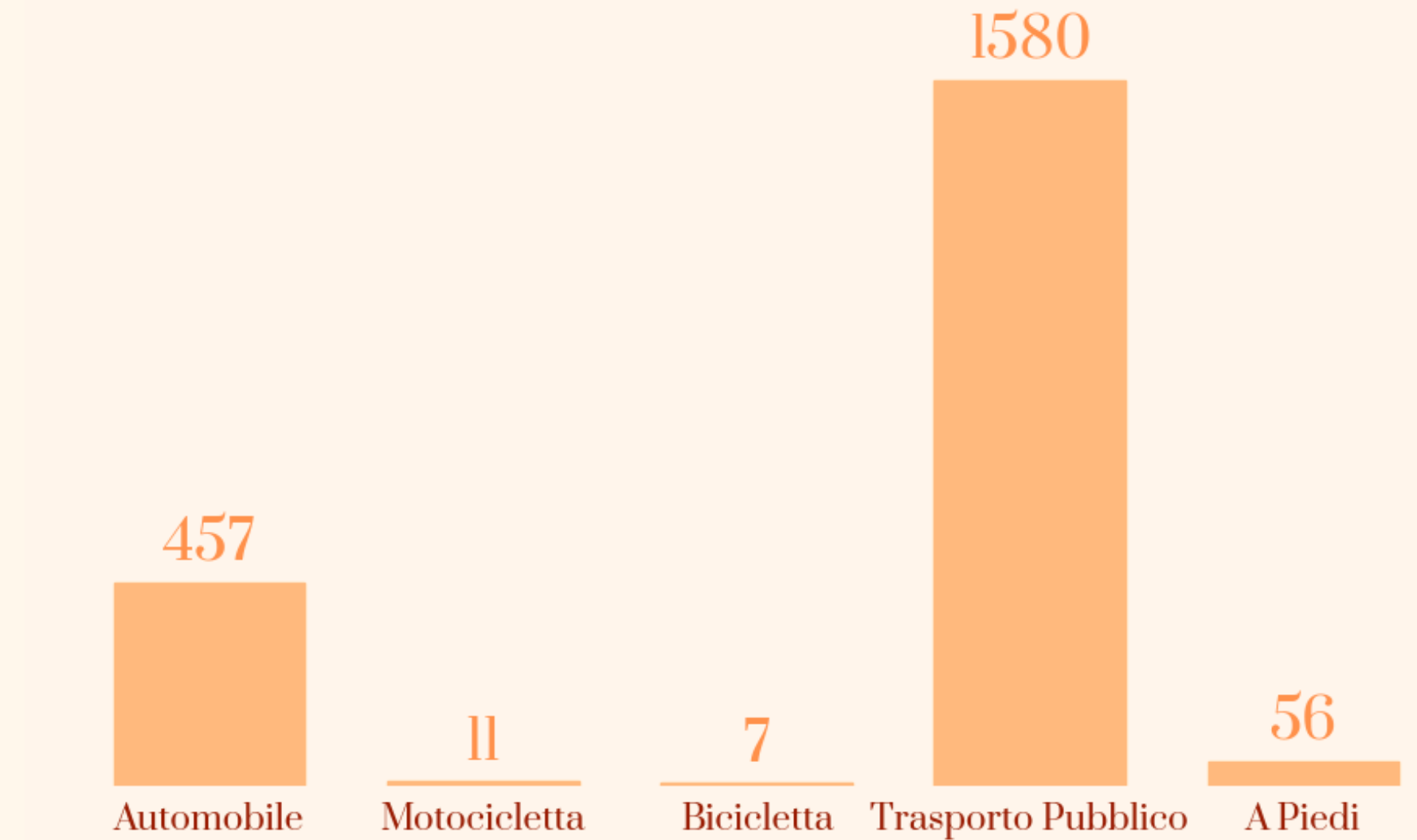


# FEATURE CATEGORICHE

Consumo di Alcol

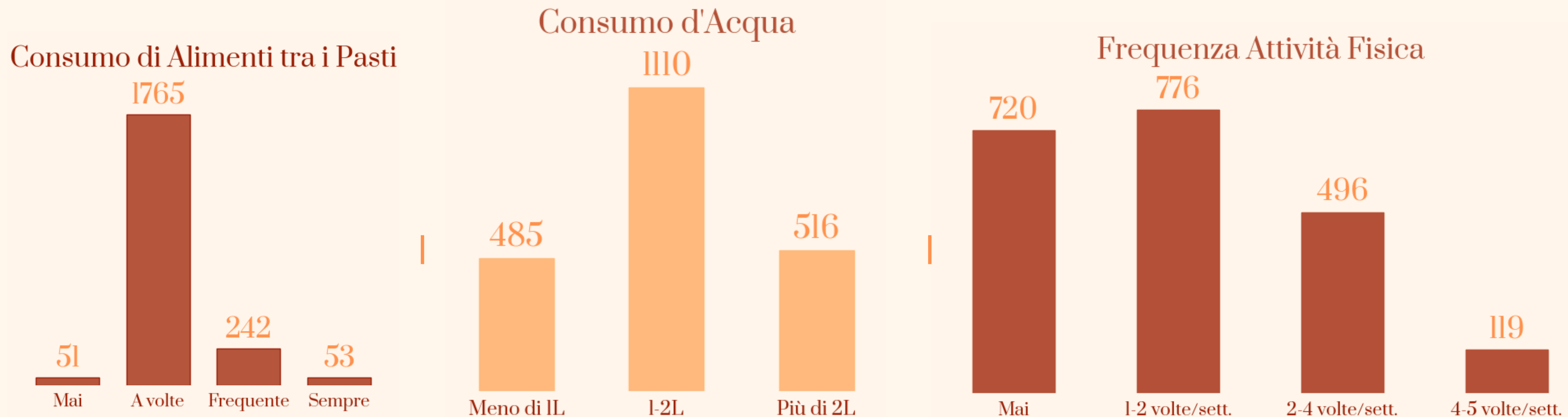


Mezzo di Trasporto Principale



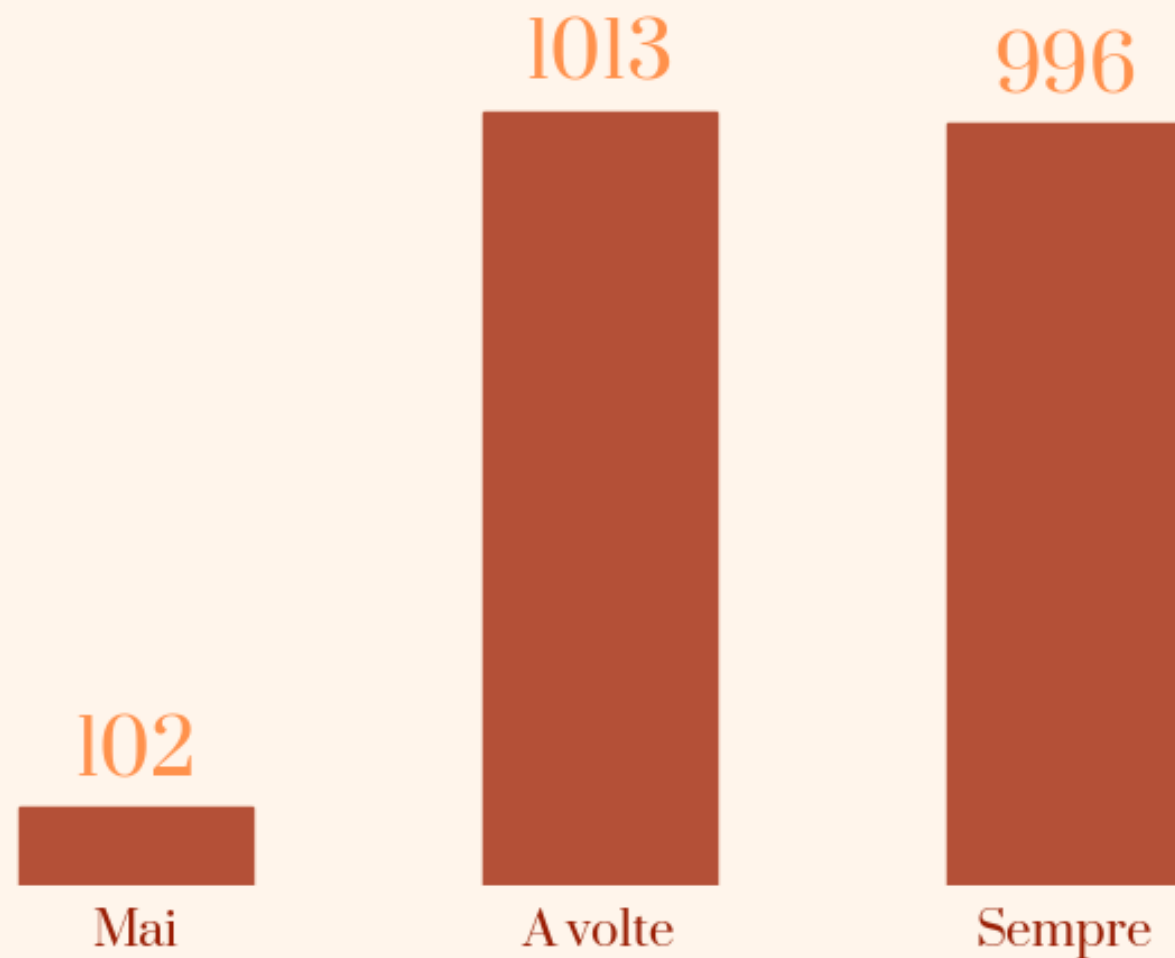


# FEATURE CATEGORICHE

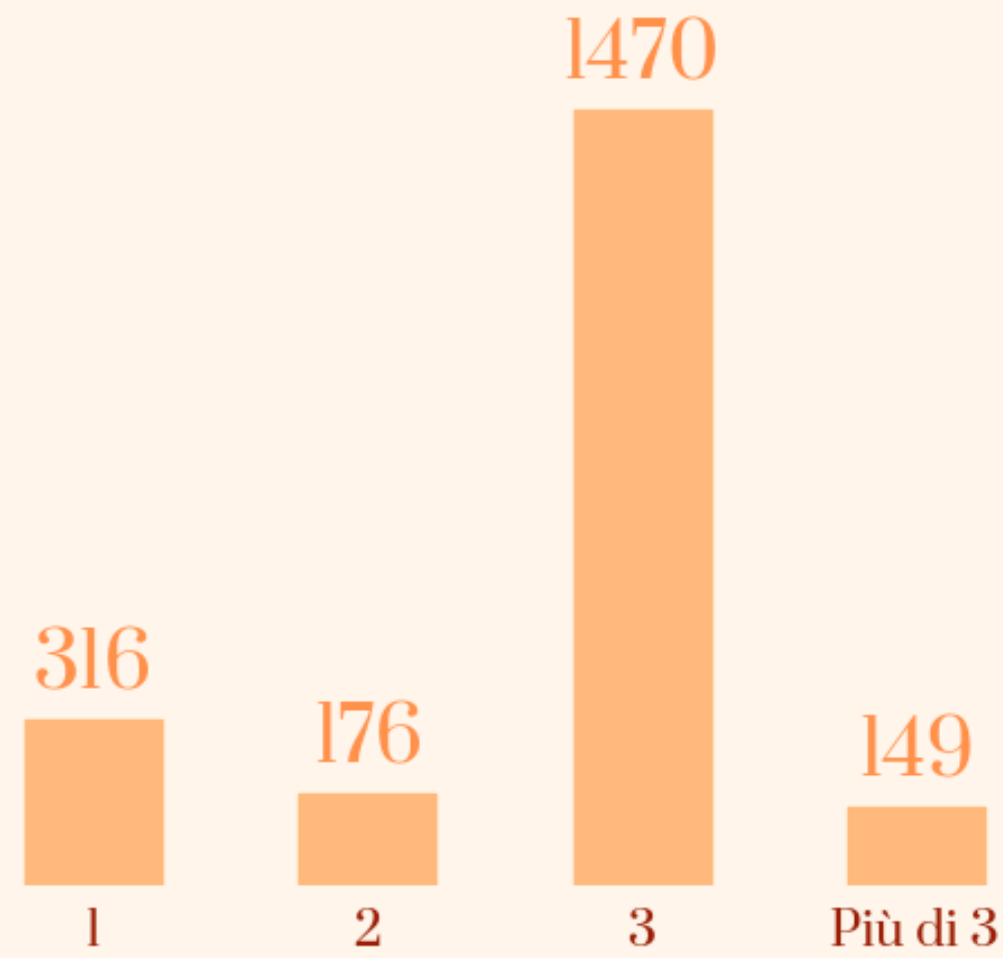


# FEATURE CATEGORICHE

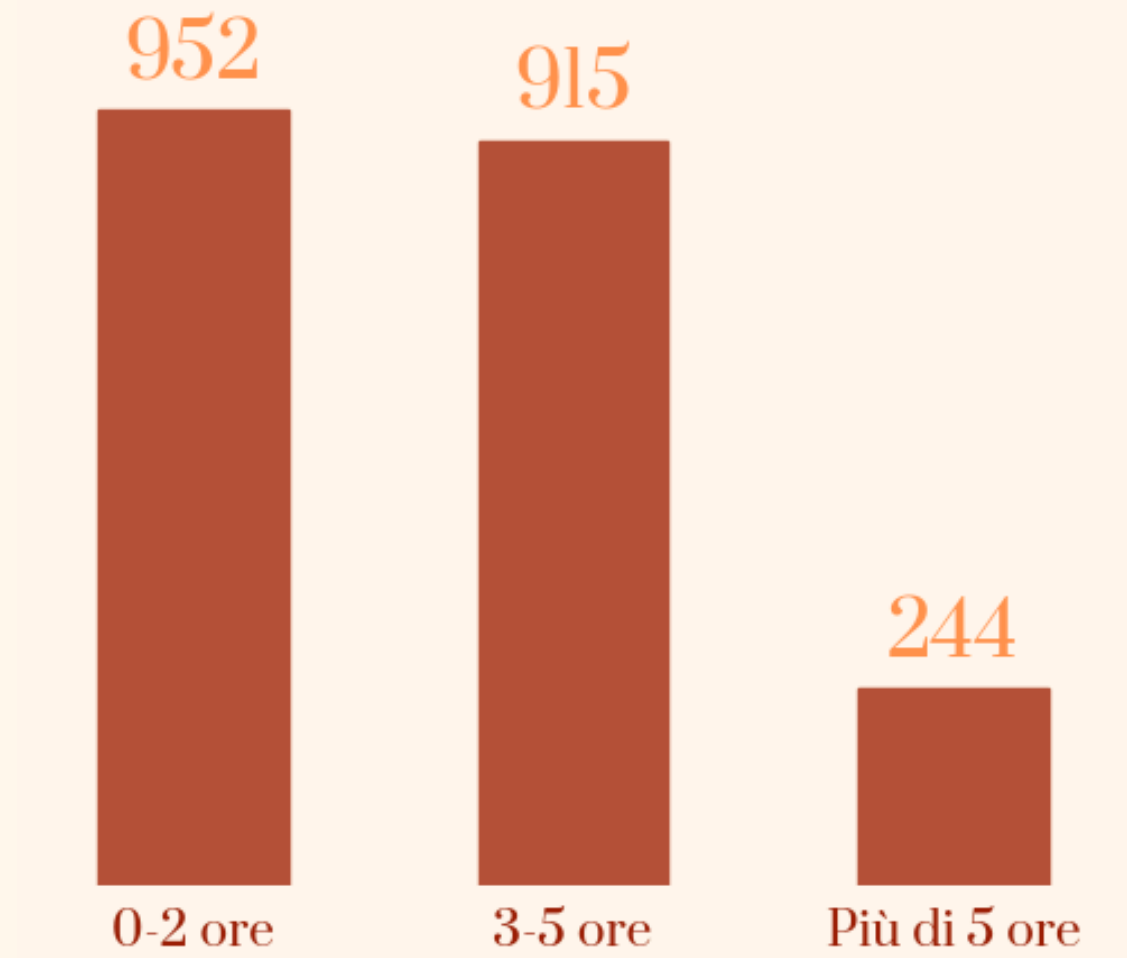
Frequenza Consumo di Vegetali



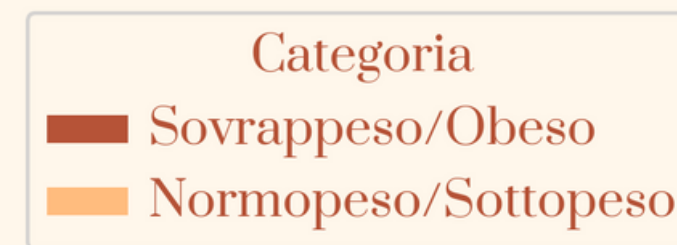
Numero di Pasti Principali



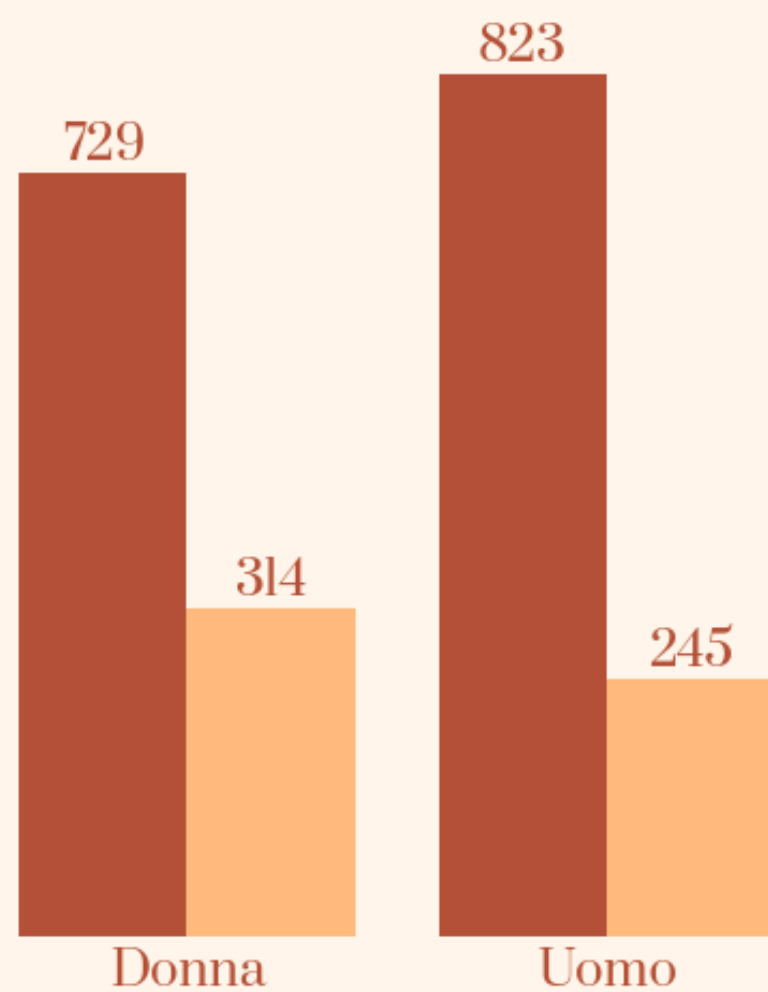
Uso di Dispositivi Tecnologici



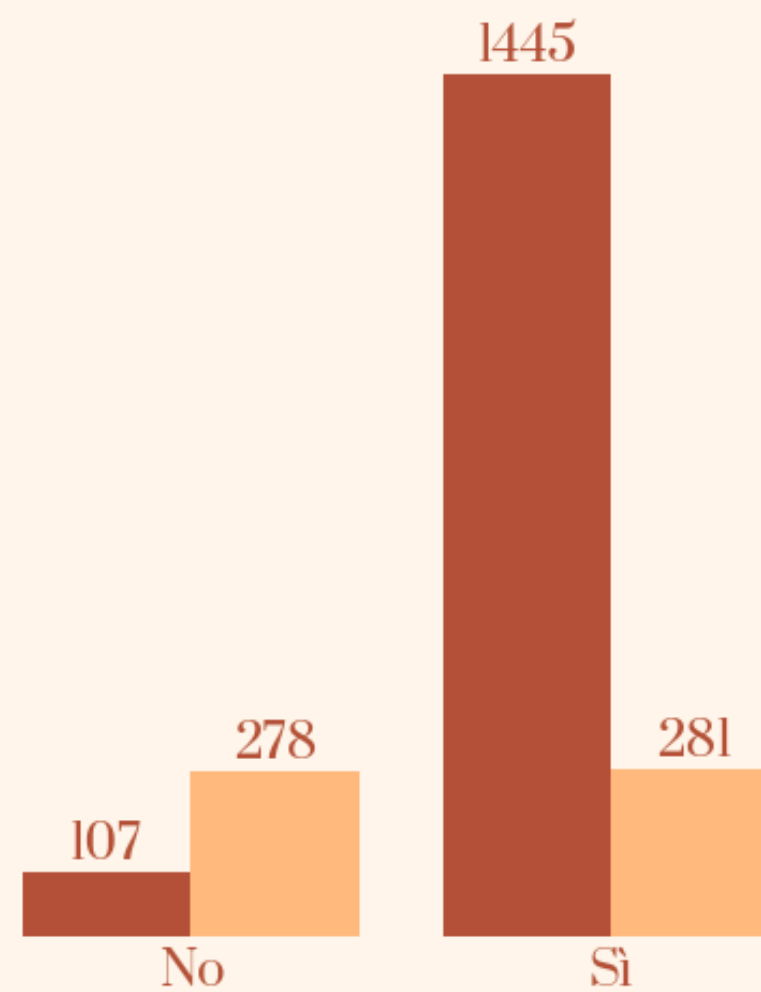
# FEATURE BINARIE



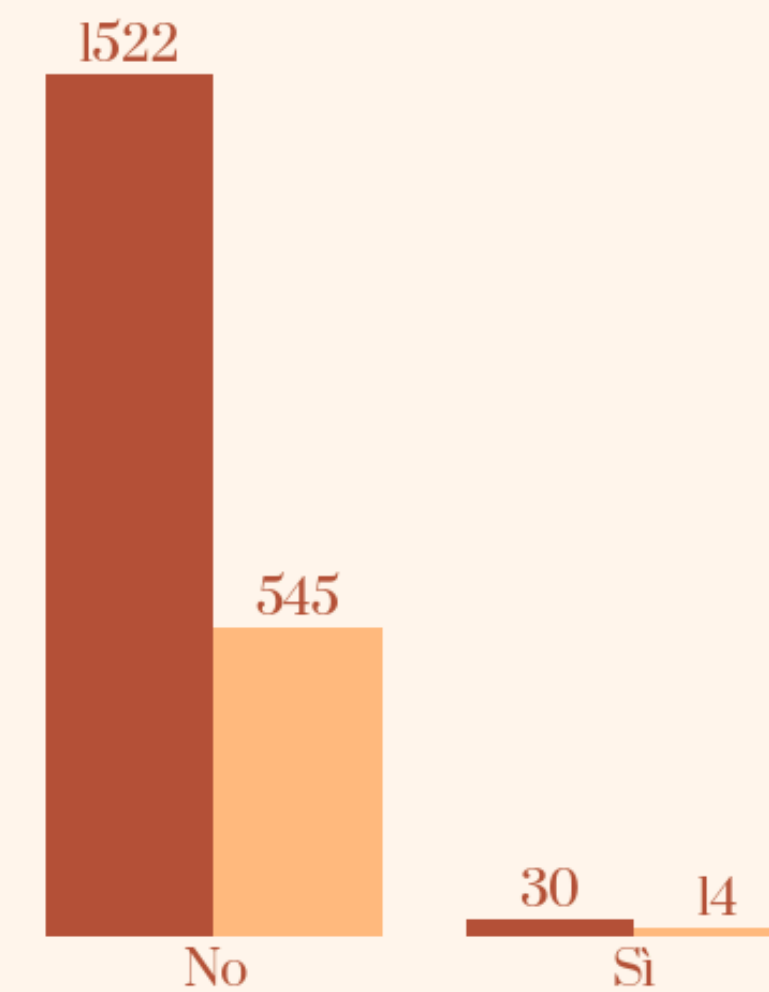
Genere



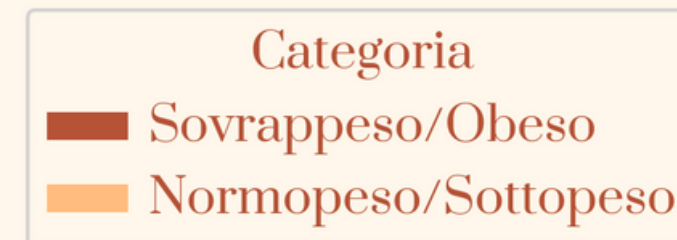
Famiglia con storia di obesità



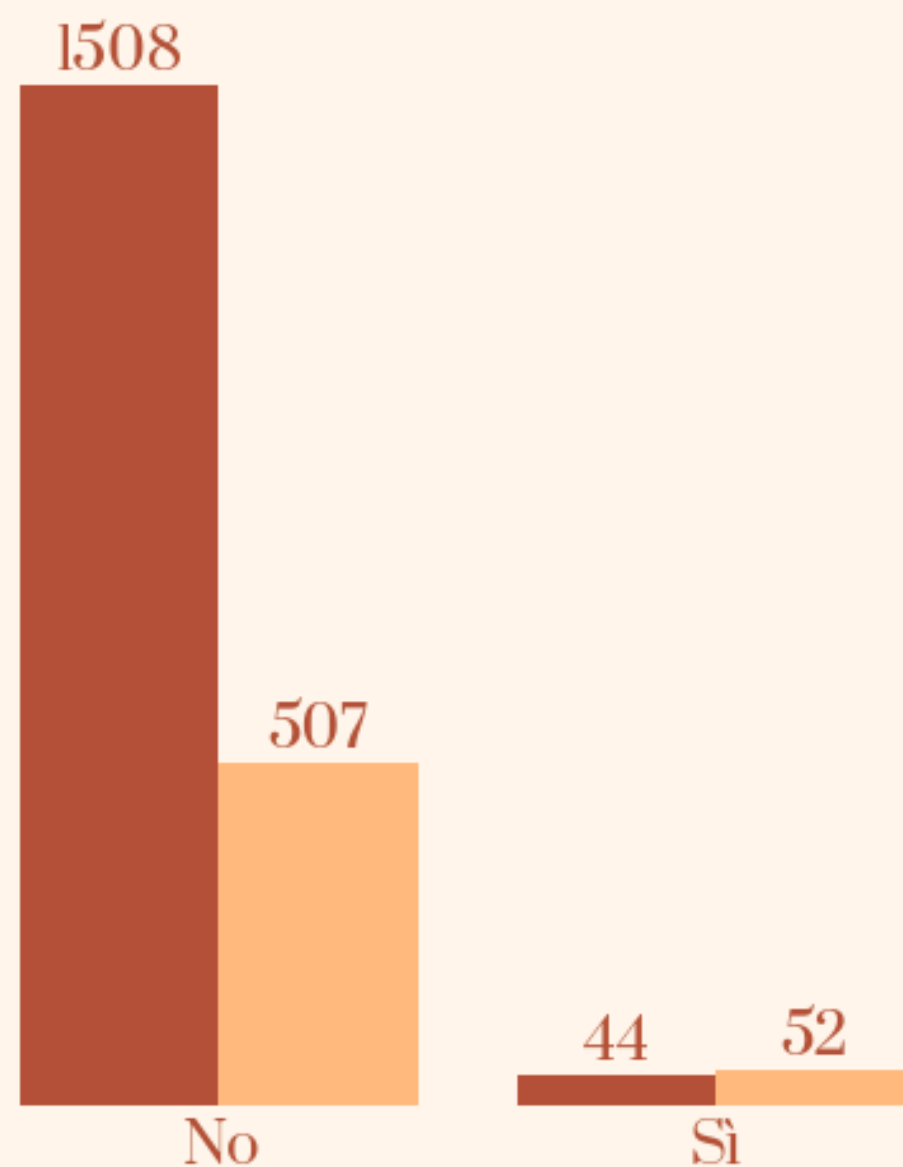
Fumatore



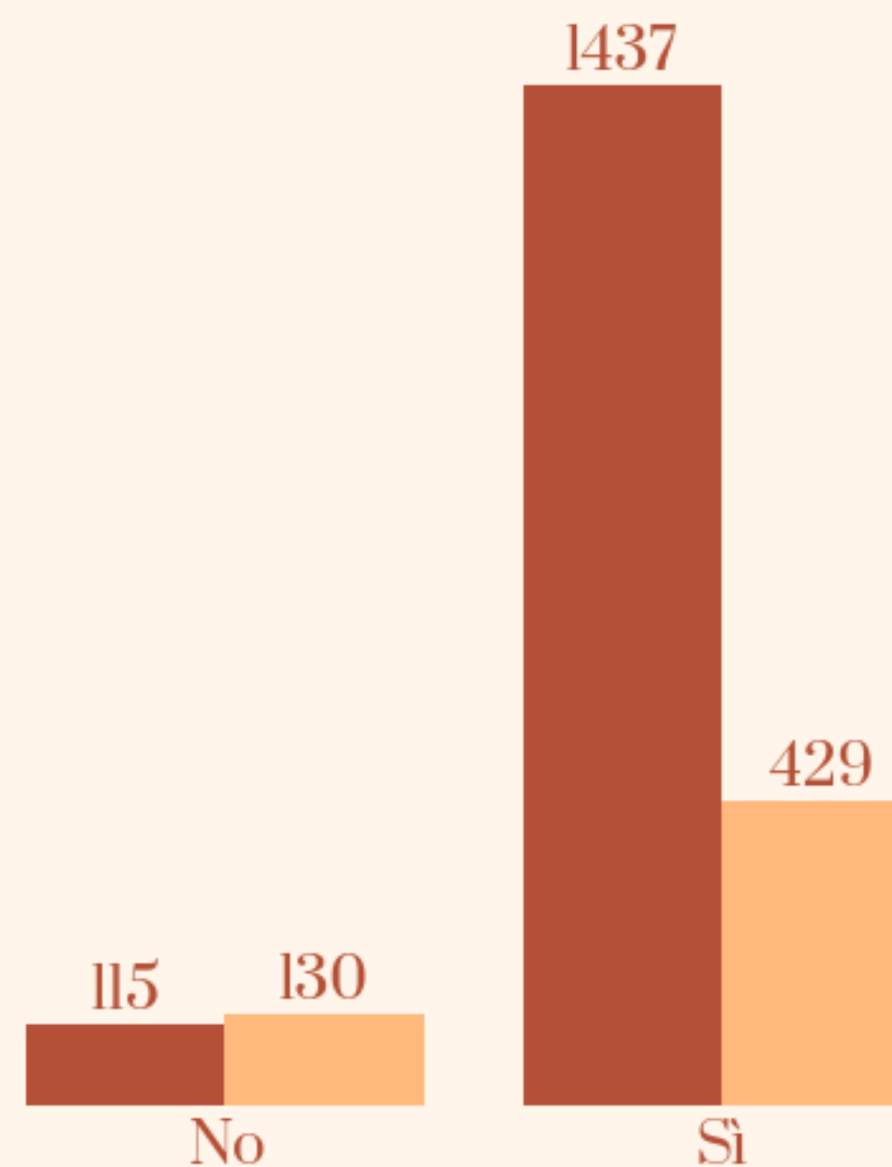
# FEATURE BINARIE



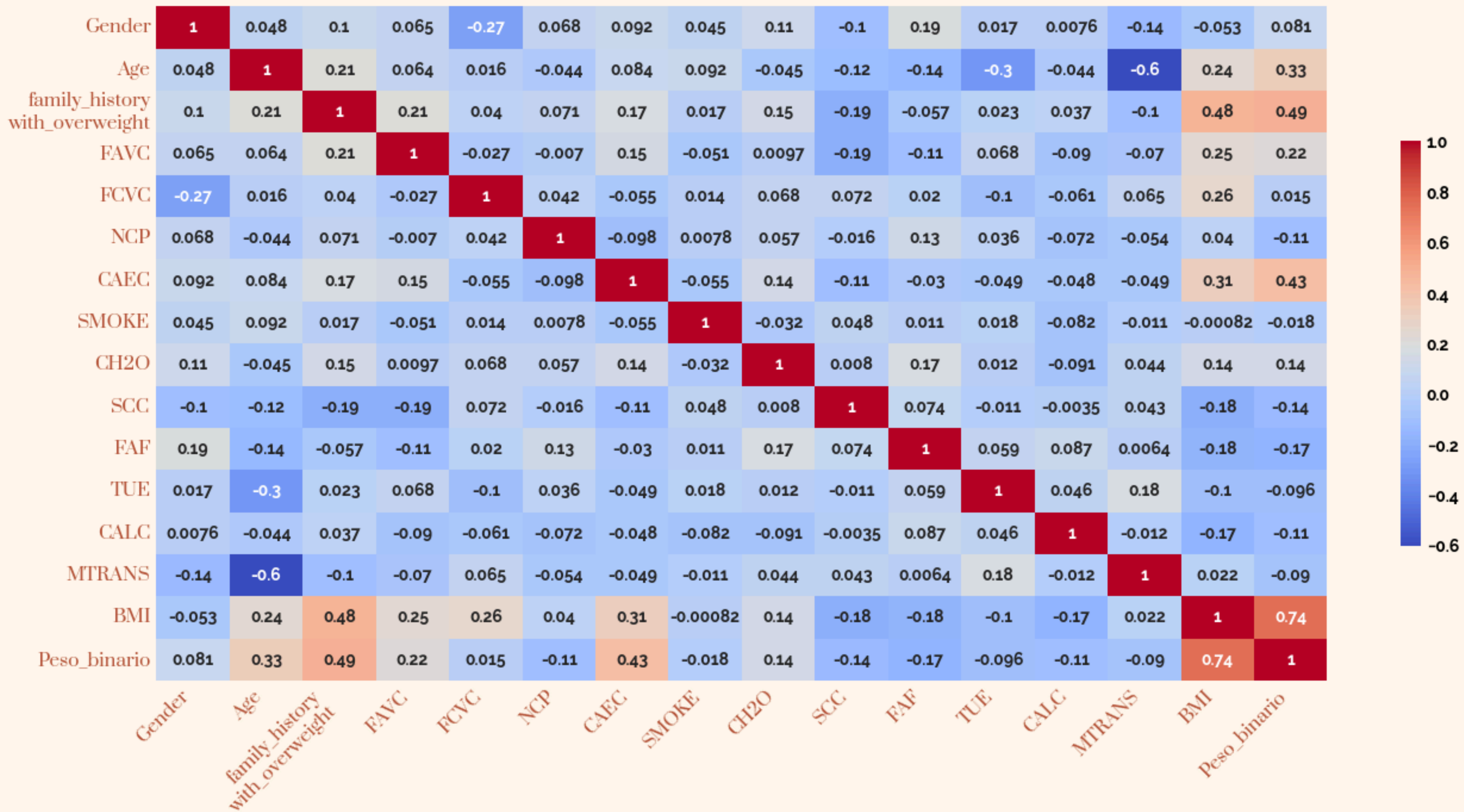
Controllo delle calorie giornaliero



Consumazione di cibo altamente calorici



# HEATMAP



# CLASSIFICAZIONE

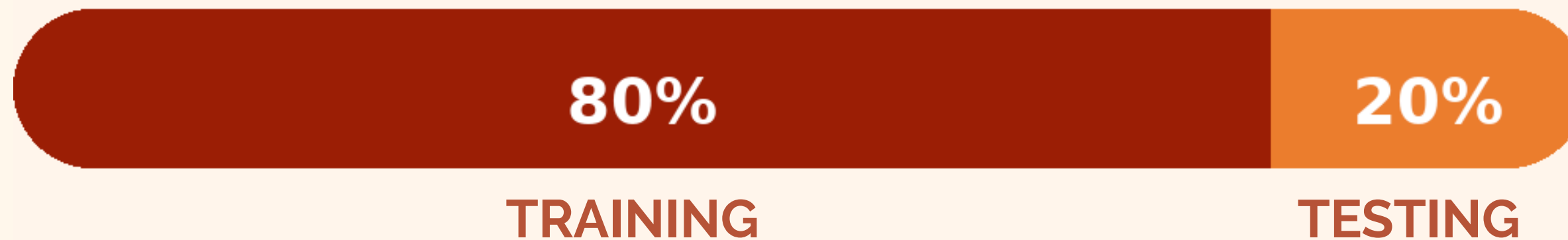
---

TRAMITE RANDOM FOREST



# SUDDIVISIONE

La prima cosa effettuata è stato suddividere il dataset in due parti, training e testing. Le percentuali sono le seguenti, elencate come sono state scritte:



# IPERPARAMETRI

IPERPARAMETRO	VALORE	SIGNIFICATO
n_estimators	100	Indica il numero di alberi da generare
max_depth	None	Controlla la massima profondità che l'albero può raggiungere.
min_samples_split	5	Numero minimo di campioni richiesti affinché un nodo interno (non una foglia) possa essere ulteriormente diviso
min_samples_leaf	1	Numero minimo di campioni richiesti per trovarsi in un nodo foglia
random_state	3	Serve per ottenere gli stessi alberi ogni volta che vengono eseguiti

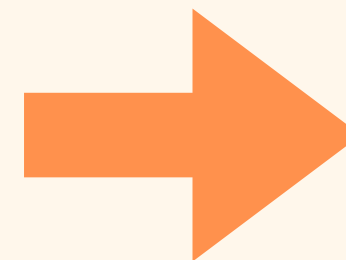


# HYPERPARAMETER TUNING

L'hyperparameter Tuning è un processo che mira a trovare la combinazione migliore di iperparametri per migliorare le prestazioni del modello. Gli iperparametri non vengono appresi dal modello ma impostati prima dell'addestramento.

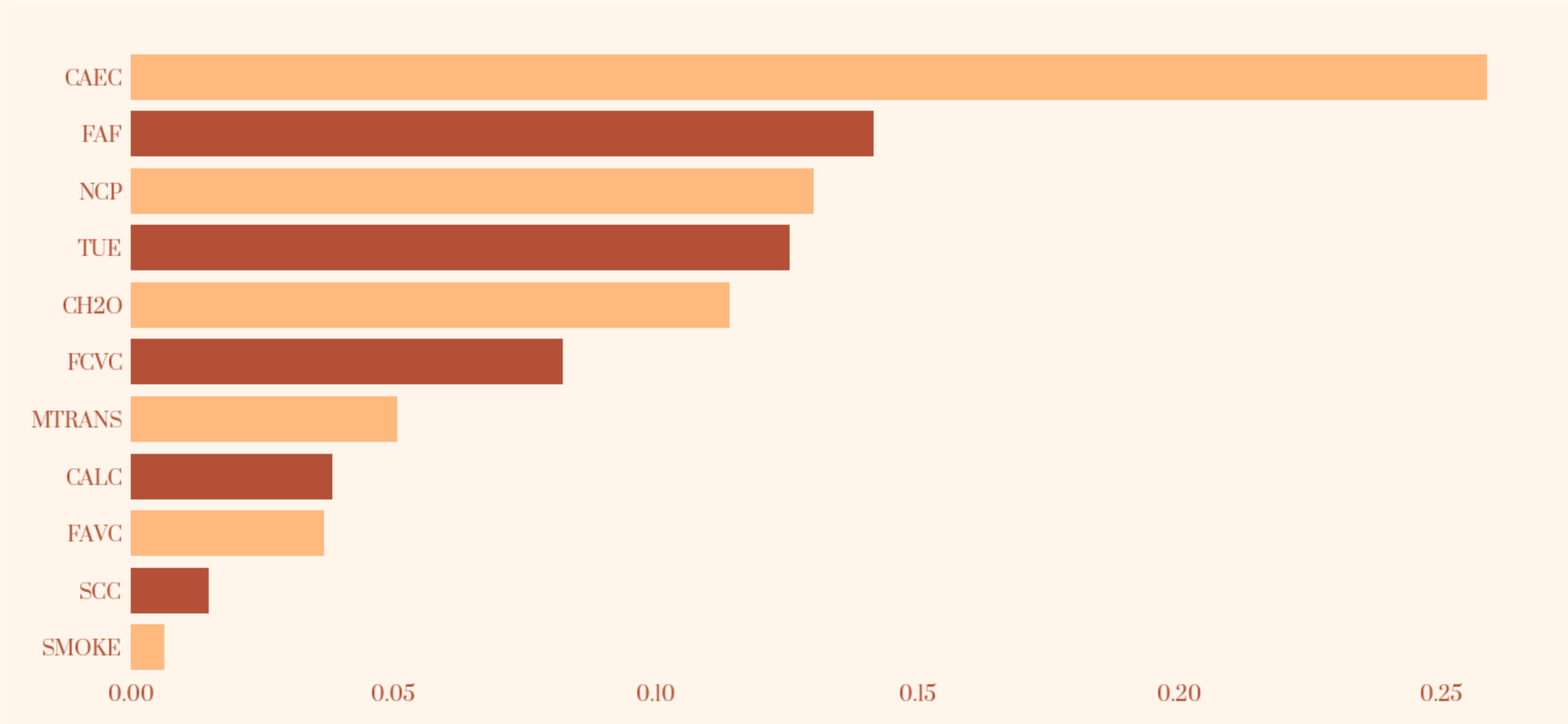
Nel nostro caso abbiamo utilizzato la tecnica del **Grid Search**. Essa consiste nel testare sistematicamente tutte le combinazioni possibili di un insieme predefinito di iperparametri

```
param_grid = {  
    'n_estimators': [50, 100, 200],  
    'max_depth': [5, 10, 15, None],  
    'min_samples_split': [2, 5, 10],  
    'min_samples_leaf': [1, 2, 4]  
}
```



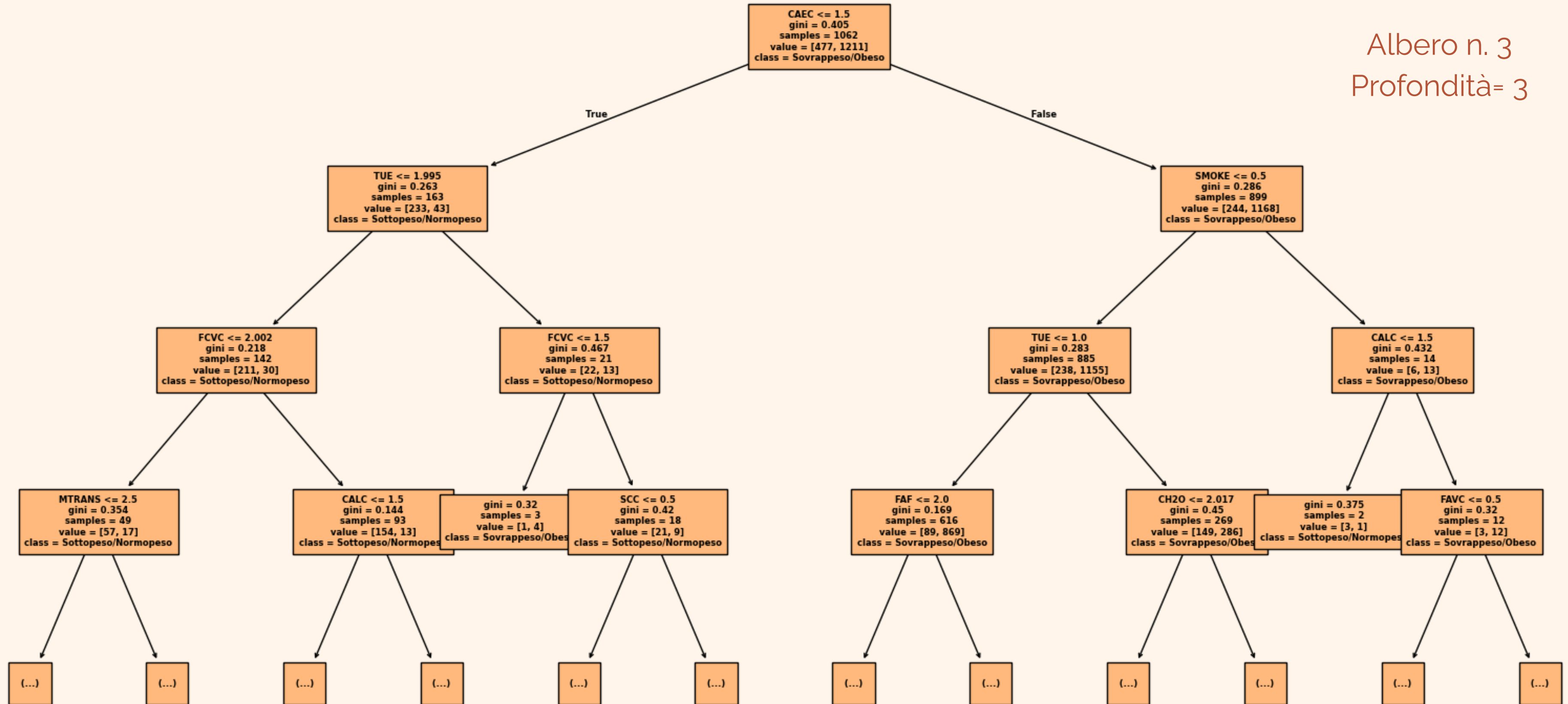
**108**  
**possibili modelli**

# FEATURE IMPORTANCE



# DECISION TREE

Albero n. 3  
Profondità= 3



# VALUTAZIONE

---

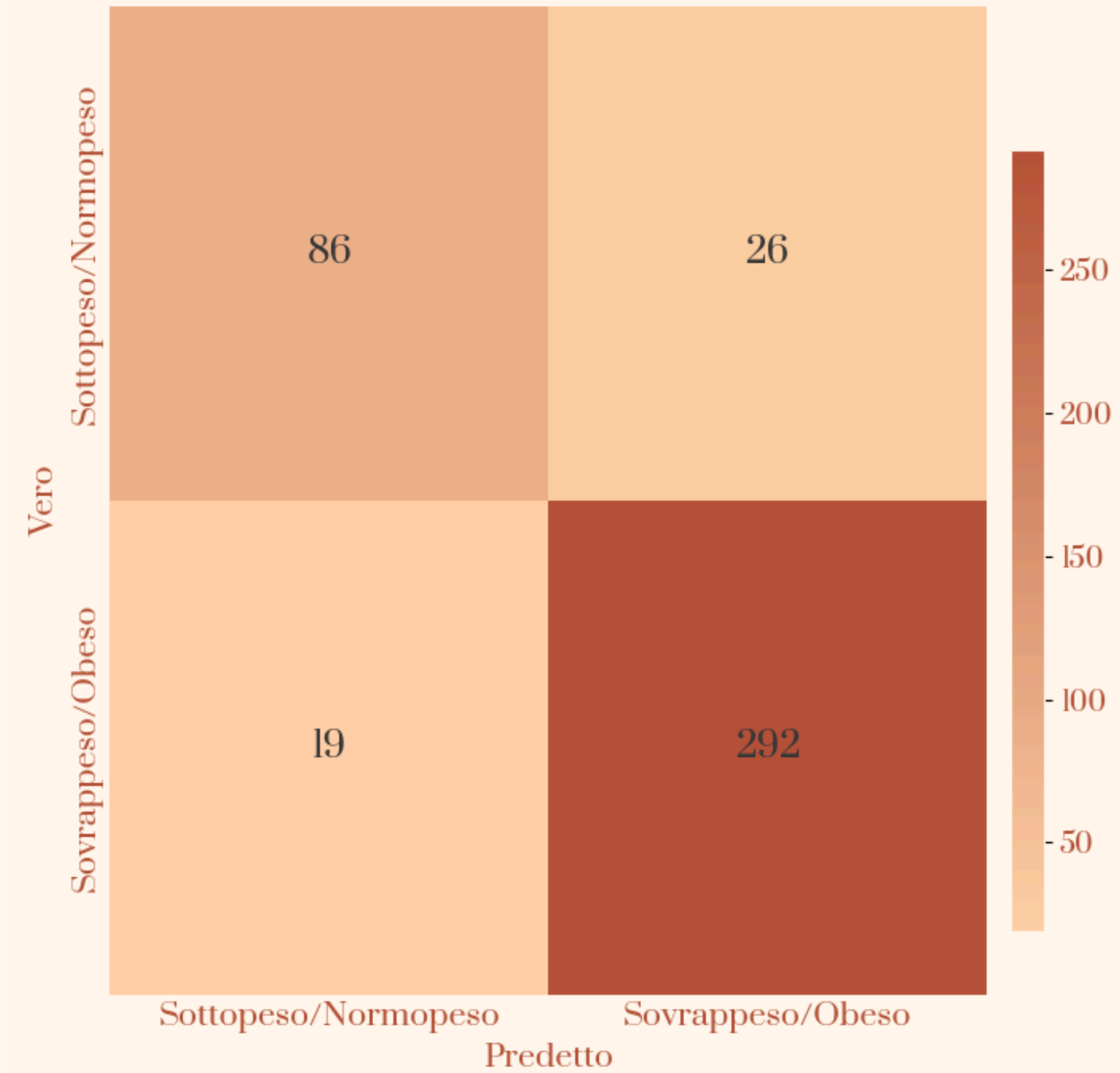


# METRICHE

<b>Precision</b>	Percentuale di predizioni positive corrette sul totale delle predizioni positive	$\frac{Tp}{Tp + Fp}$
<b>Recall</b>	Percentuale di casi positivi correttamente individuati dal modello sul totale dei positivi reali	$\frac{Tp}{Tp + Fn}$
<b>F1 Score</b>	Quantifica l'efficacia del modello nel riconoscere la classe positiva minimizzando falsi positivi e falsi negativi. Metrica utile quando si hanno classi sbilanciate	$\frac{2Tp}{2Tp + Fp + Fn}$
<b>ROC</b>	Rappresenta la relazione tra TPR e FPR al variare della soglia di decisione	$TPR = \frac{Tp}{Tp + Fn} \quad FPR = \frac{Fp}{Fp + Tn}$
<b>ROC-AUC</b>	Area sotto la curva ROC e valuta il modello nel complesso	
<b>Accuracy</b>	Percentuale di predizioni corrette sul totale delle predizioni	$\frac{Tp + Tn}{Tp + Tn + Fp + Fn}$

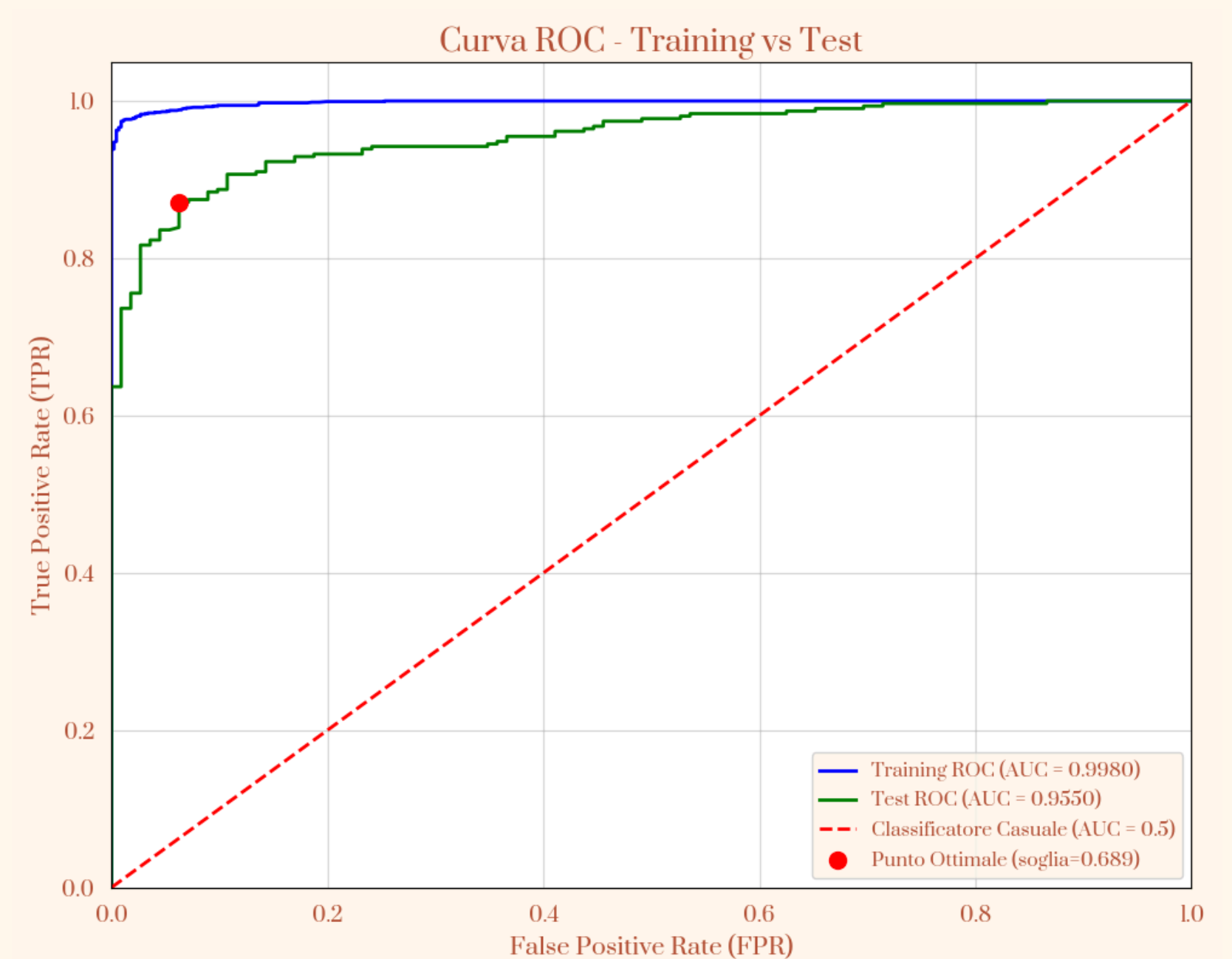
# VALUTAZIONE MODELLO

	Precision	Recall	F1 Score	Support
Sovrappeso/Obeso	0.82	0.77	0.79	112
Sottopeso/Normopeso	0.92	0.94	0.93	311
Macro Avg	0.87	0.85	0.86	423
Weighted Avg	0.89	0.89	0.89	423
Accuracy: 0.8936				

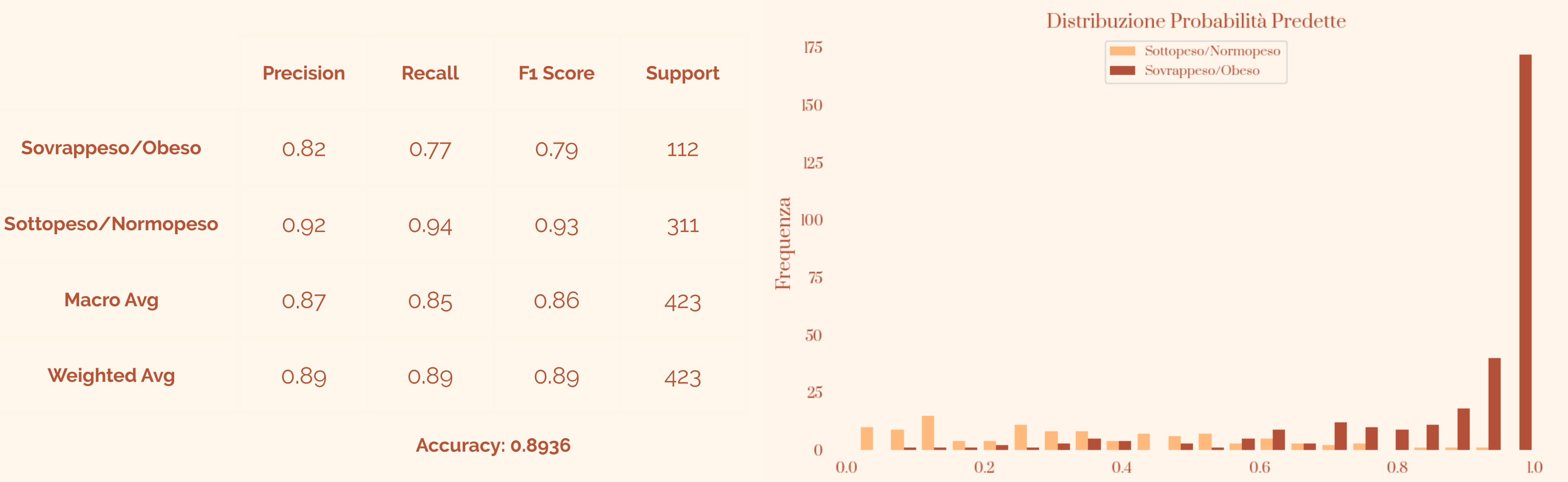


# VALUTAZIONE MODELLO

	Precision	Recall	F1 Score	Support
Sovrappeso/Obeso	0.82	0.77	0.79	112
Sottopeso/Normopeso	0.92	0.94	0.93	311
Macro Avg	0.87	0.85	0.86	423
Weighted Avg	0.89	0.89	0.89	423
Accuracy: 0.8936				



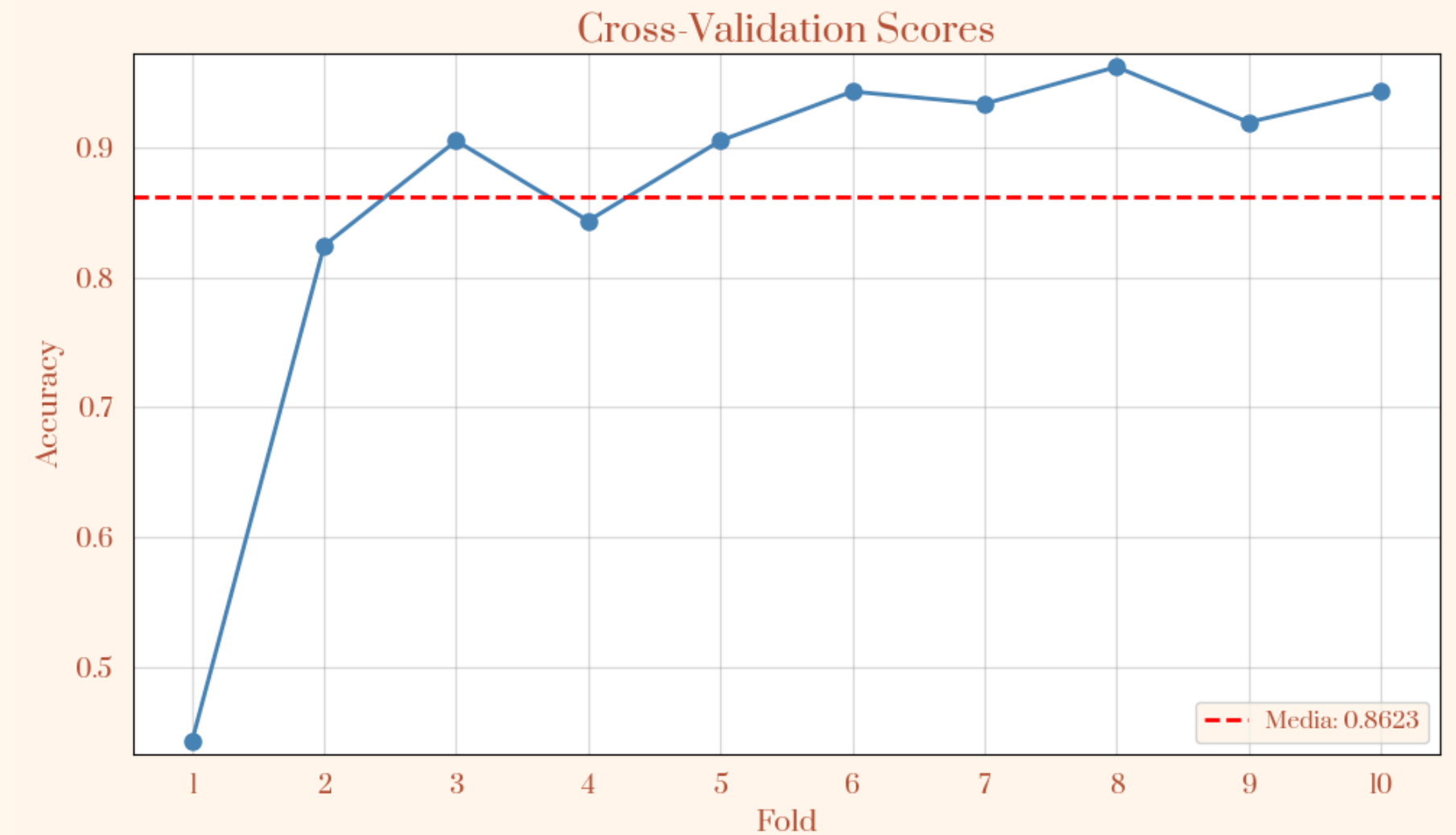
# VALUTAZIONE MODELLO





# VALUTAZIONE MODELLO

	Precision	Recall	F1 Score	Support
Sovrappeso/Obeso	0.82	0.77	0.79	112
Sottopeso/Normopeso	0.92	0.94	0.93	311
Macro Avg	0.87	0.85	0.86	423
Weighted Avg	0.89	0.89	0.89	423
Accuracy: 0.8936				



Fine