

# Streaming Data Management and Time Series Analysis

Corso di laurea magistrale – Data Science

Prof. Matteo Maria Pelagatti

Simone Farallo – 889719



# Introduzione

**Obiettivo dello studio:** previsione del consumo di energia elettrica per il periodo compreso tra l'1 dicembre 2017 e il 31 dicembre 2017.

**Modelli:** ARIMA, UCM e Machine Learning

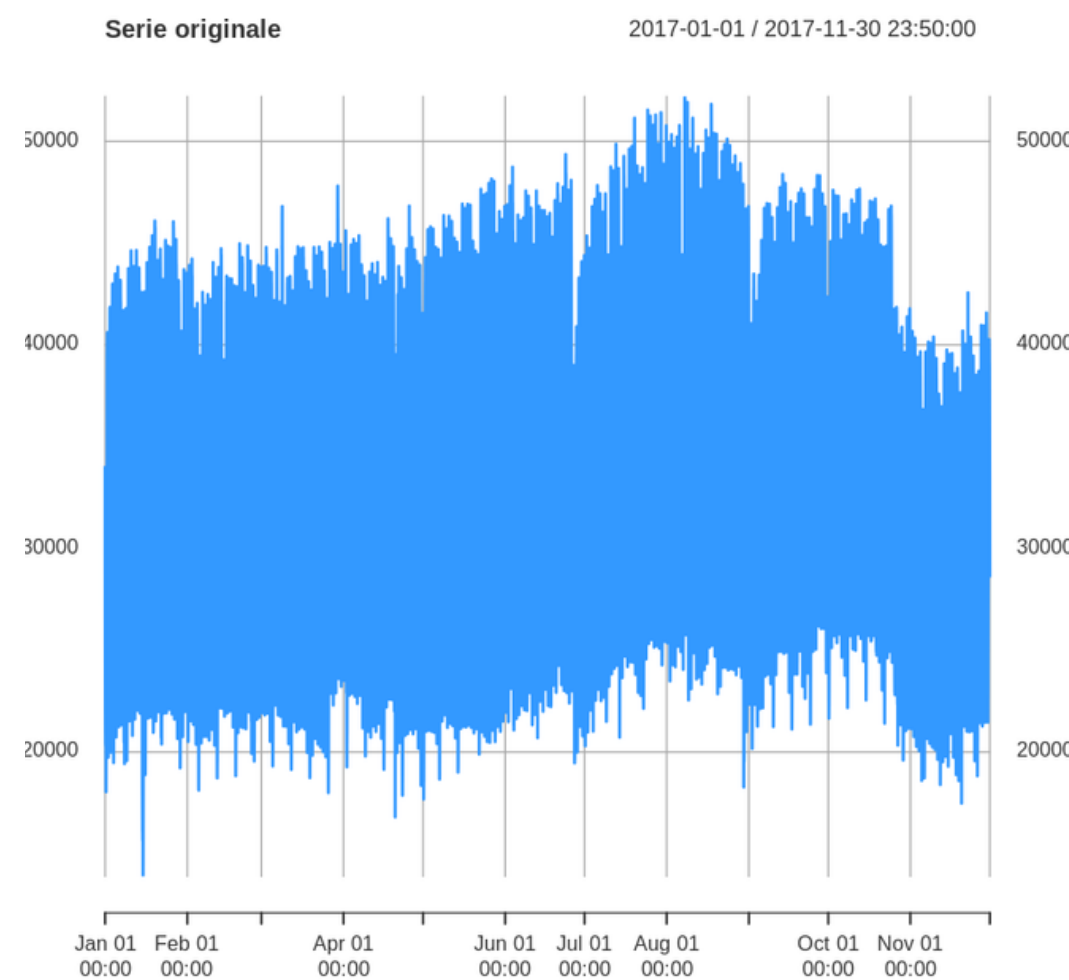
Metrica di riferimento: **Mean Absolute Error**

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

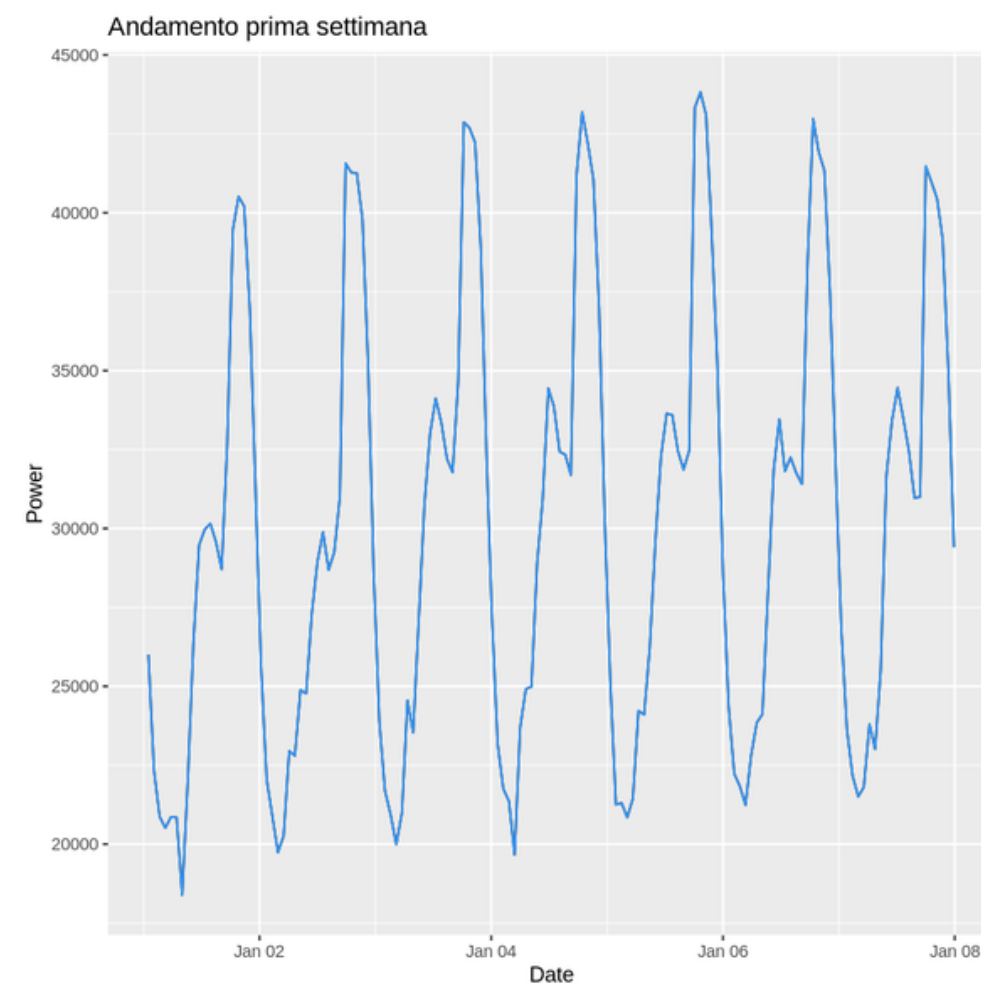
# Esplorazione dei dati

Dataset consumi energetici-Tètouan, Marocco

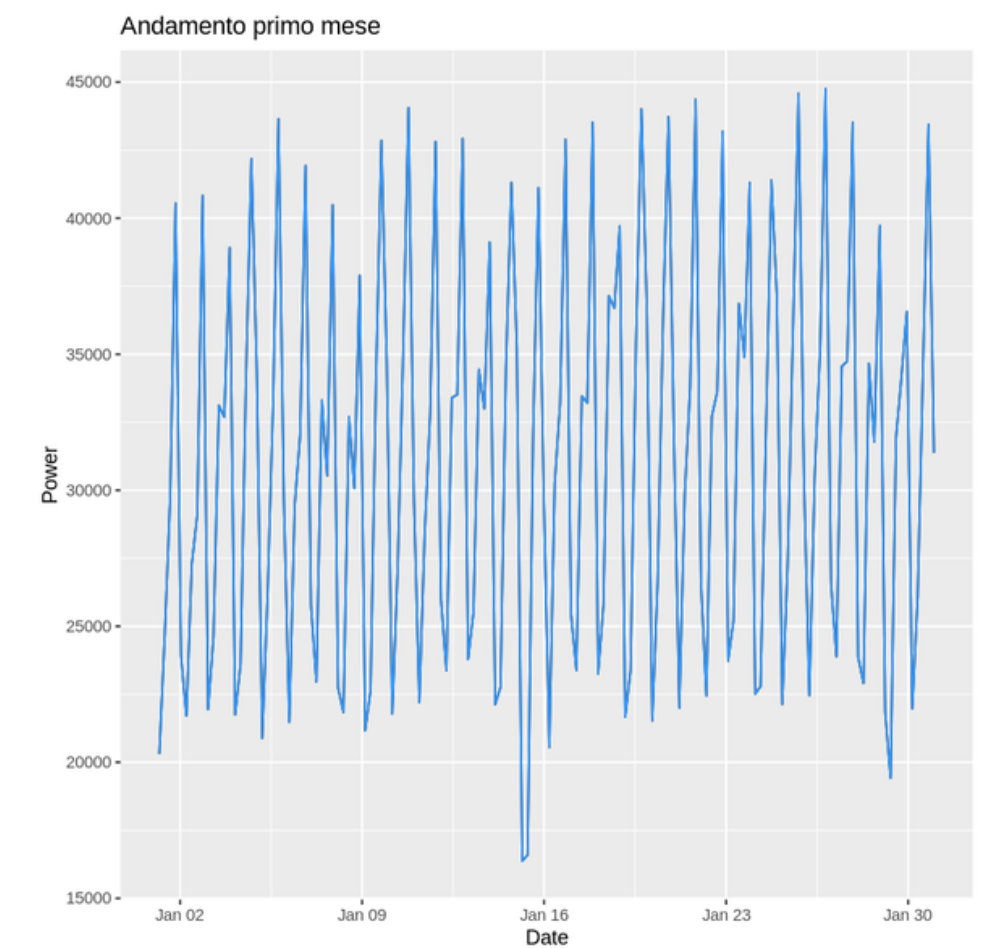
**48.096** osservazioni dal 1° gennaio 2017 alle 00:00:00 fino al 30 novembre 2017 alle 23:50:00,  
Frequenza serie: **10 minuti**



1. Serie storica originale

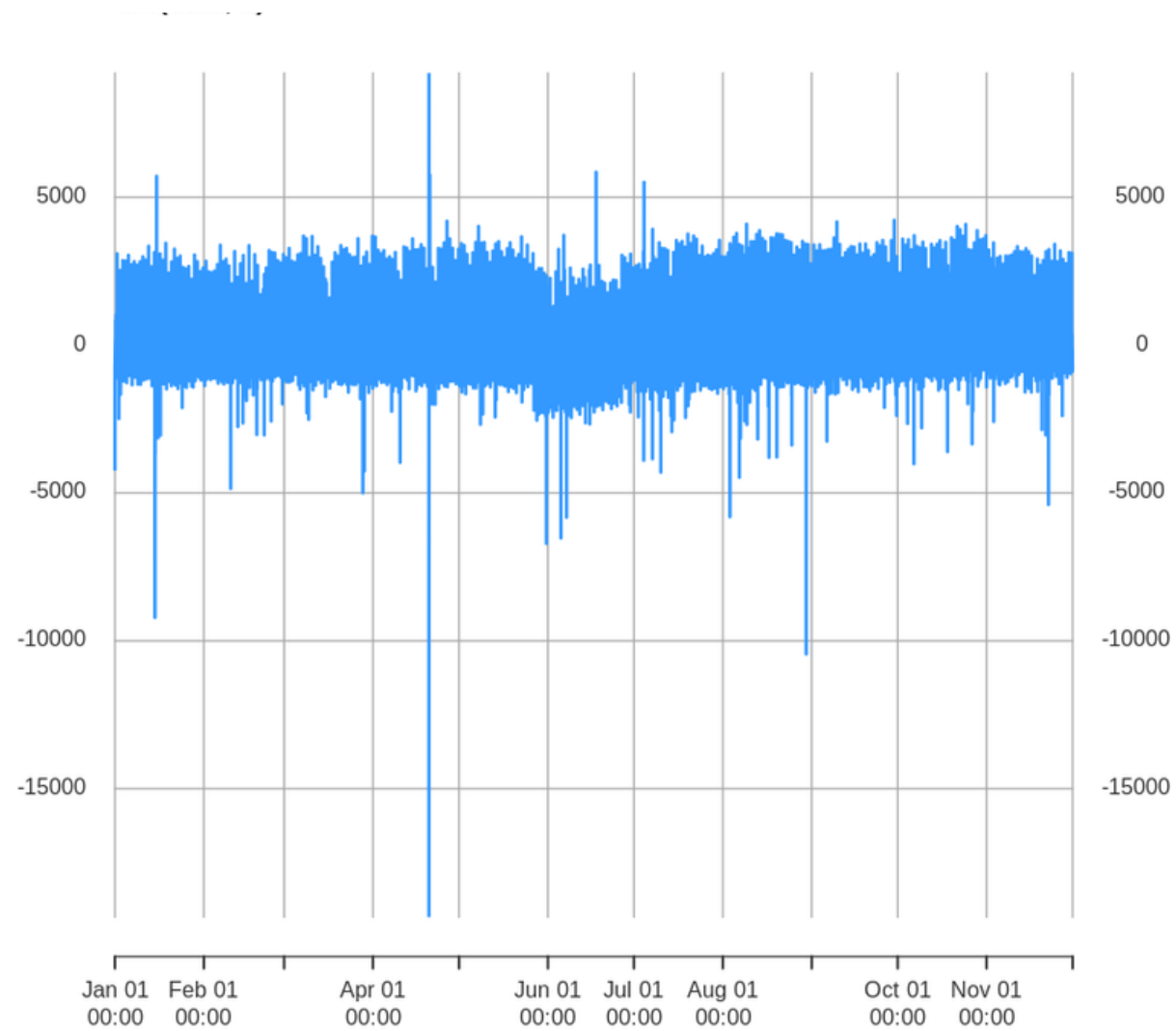


2. Prima settimana di Gennaio

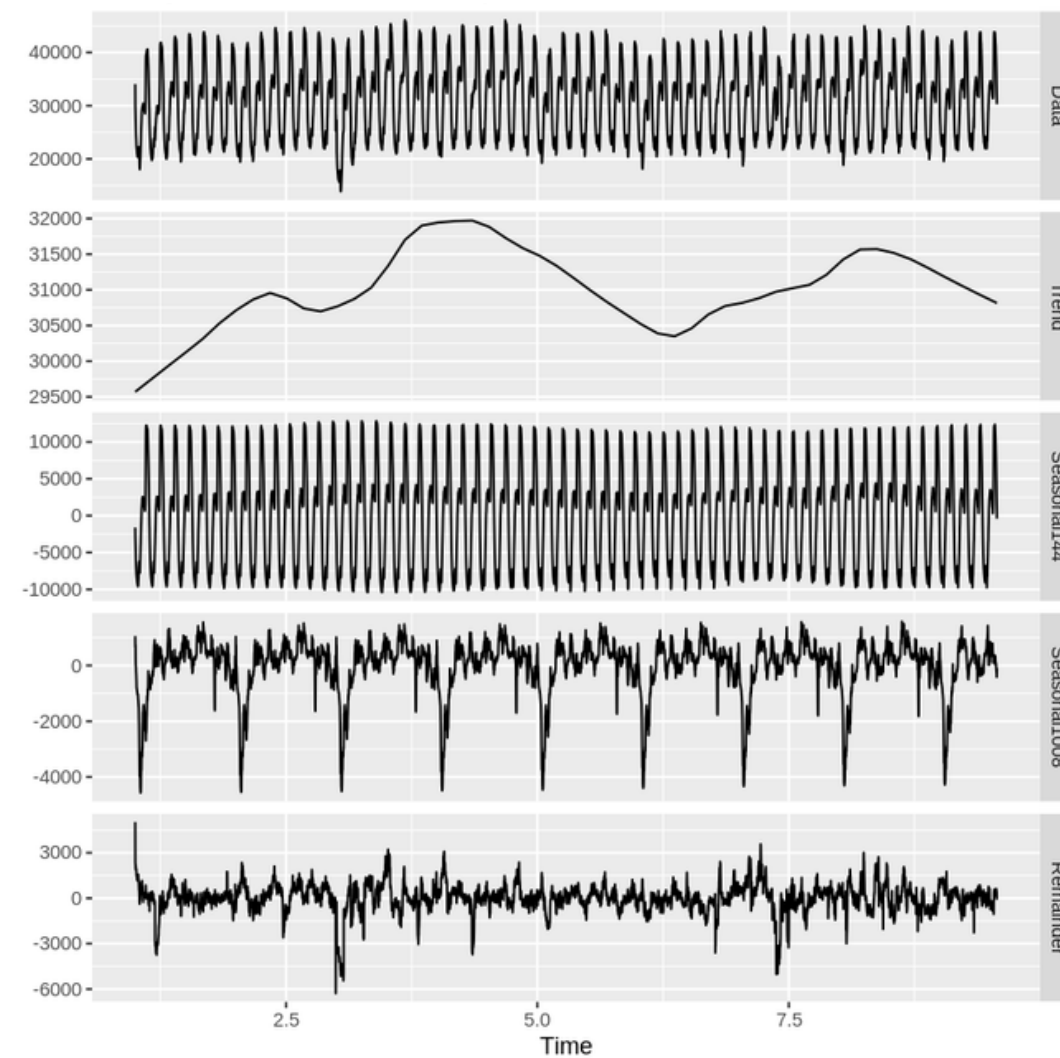


3. Mese di Gennaio

# Esplorazione dei dati



4. Outliers



5 Decomposizione serie

**Train set:** dal 1 Gennaio 2017 al 31 Ottobre 2017.

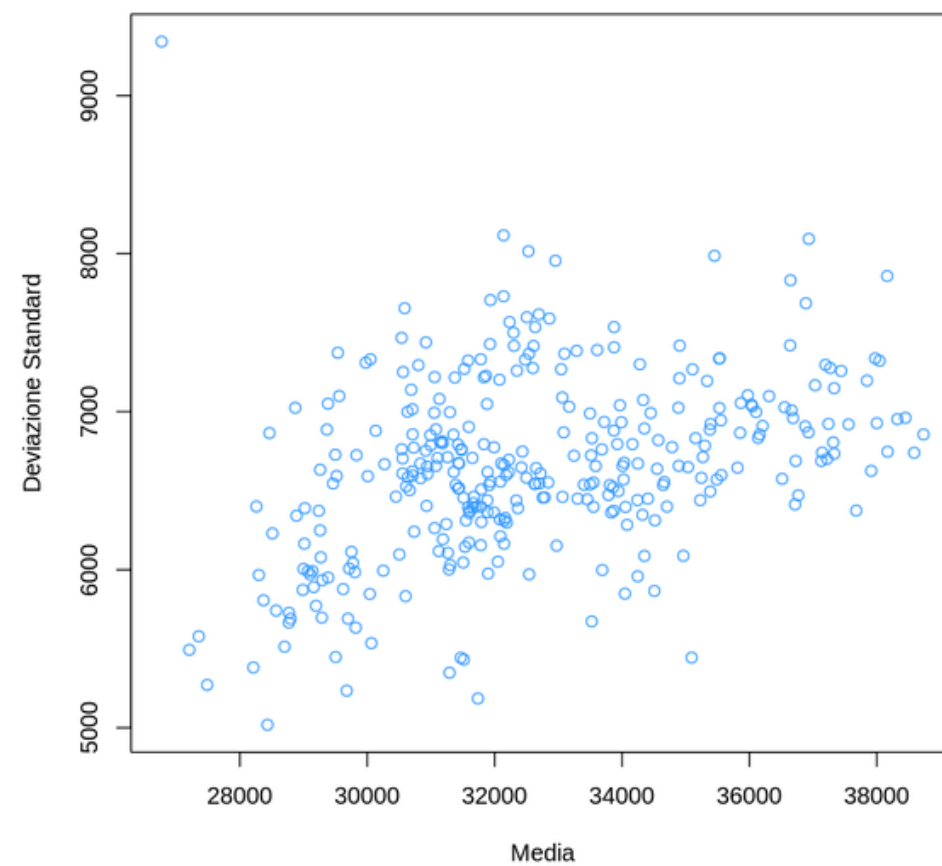
**Validation set:** dal 1 Novembre 2017 al 30 Novembre 2017.

# ARIMA

Stazionarietà del processo

Dickey-Fuller test  
KPSS test

Scatterplot prima della trasformazione di Box-Cox

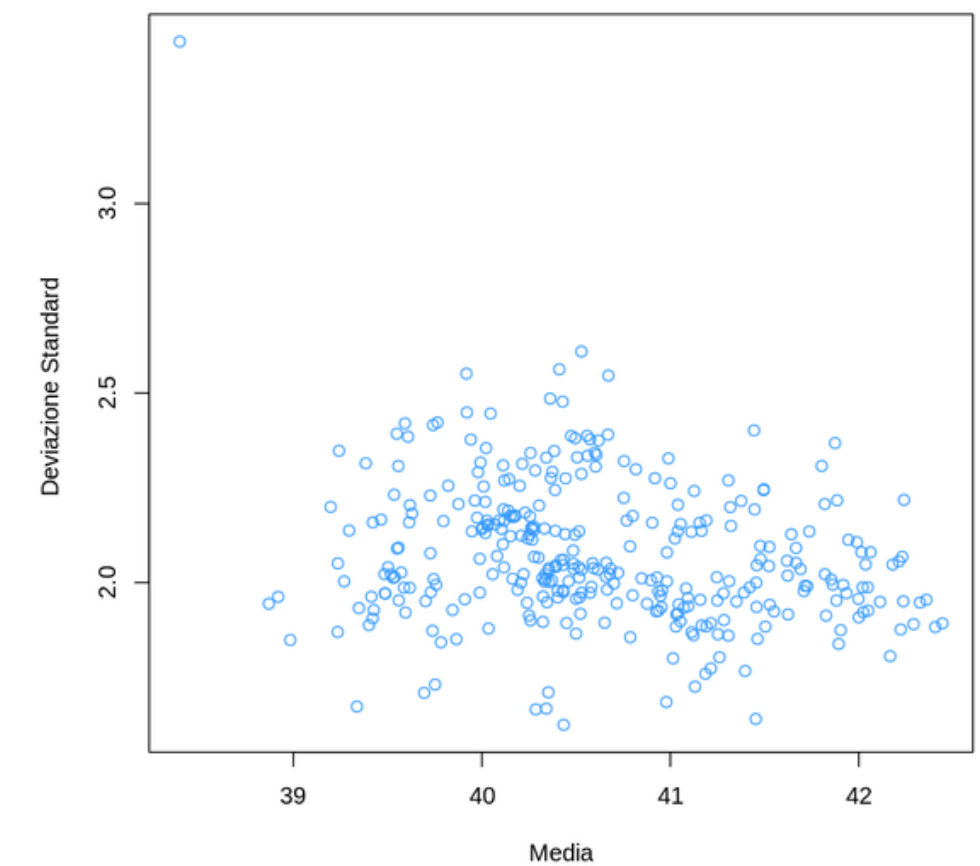


1. Serie prima della trasformazione.

Trasformazione di **Box-Cox**



Scatterplot dopo la trasformazione di Box-Cox



2. Serie dopo la trasformazione .

# ARIMA

Grafici **ACF** e **PACF**

[1] – differenza stagionale su serie originale:

$$ARIMA(\mathbf{0,0,0})(\mathbf{1,1,0})[144]$$

[2] – differenza stagionale su serie oraria

$$ARIMA(\mathbf{0,1,1})(\mathbf{0,1,1})[144]$$

**[3]** – *differenza stagionale e dummy serie originale*

$$ARIMA(\mathbf{0,0,0})(\mathbf{1,0,0})[144]$$

Modelli testati su serie originale e oraria

[1] – trend lineare + comp. stag. giornaliera (**2** armoniche) + comp. stag. settimanale(**1** armonica)

[2] – trend lineare + comp. stag. giornaliera (**10** armoniche) + comp. stag. settimanale(**1** armonica)

*[3] – trend lineare + comp stag. giornaliera(**10** armoniche) + comp. stag. settimanale (**Ciclo stoc.**)*

*[4] – trend lineare + compstag. giornaliera (**10** armoniche) + comp. stag. settimanale(**Dummy**)*

# Machine Learning

Modelli testati su serie originale e oraria

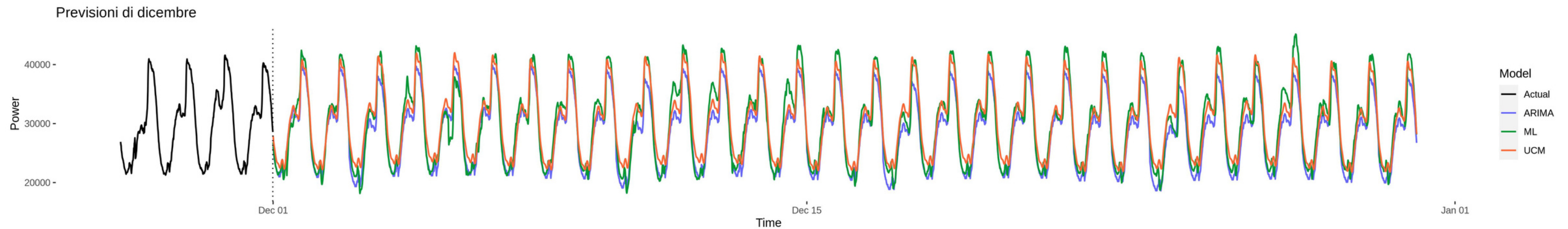
[1] – **SVM** con lag settimanale.

[2] – **Random Forest** con 350 alberi e lag settimanale.

[3] – **Xgboost** con 1000 round e lag pari a 168.



# Conclusioni



1 Previsioni sul mese di dicembre del miglior modello ARIMA,UCM e ML.

Modello	RMSE	MAPE	MAE
<b>ARIMA</b>	1342.476	3.742	1050.814
<b>UCM</b>	1703.600	4.721	1341.666
<b>ML</b>	2205.156	6.310	1748.029

2 Risultati dei modelli più performanti ottenuti sul validation set

---



**Grazie per  
l'attenzione**

