

# **STREAMING DATA MANAGEMENT AND TIME SERIES ANALYSIS**

ANDREA MAVER, 828725

GENNAIO 2023

- 1 Introduzione
- 2 Modello baseline - Regressione lineare
- 3 ARIMA
- 4 UCM
- 5 Machine Learning
- 6 Previsioni dicembre

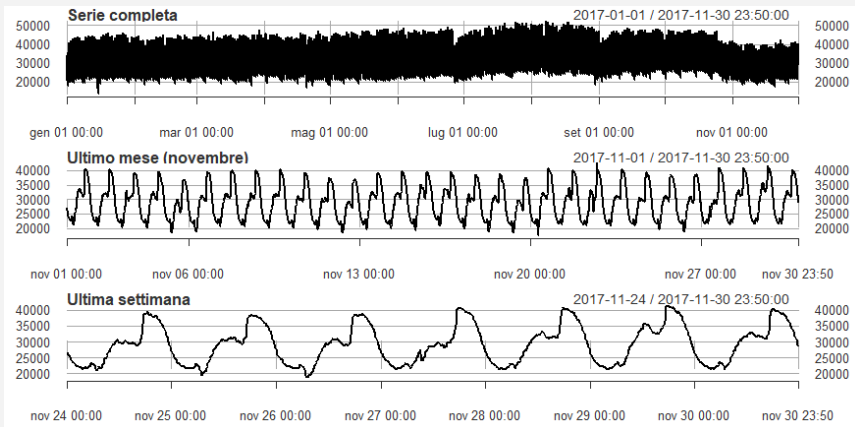
# INTRODUZIONE

Lo scopo del progetto è analizzare una serie storica di consumi elettrici che va dal 1 gennaio 2017 al 30 novembre 2017, composta da misurazioni ogni 10 minuti, per poi prevederne il mese di dicembre.

Sono state testate diverse configurazioni di modelli ARIMA, UCM e Machine Learning per trovare il modello più performante di ognuna di queste categorie.

Le valutazioni sono state effettuate tramite la metrica MAE accompagnata da controlli grafici.

# LA SERIE STORICA

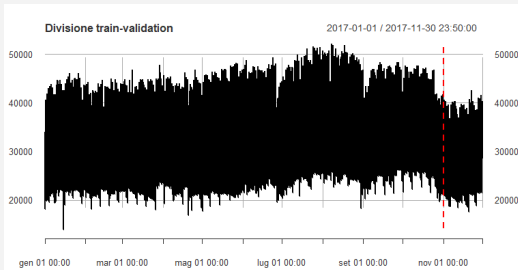


# MANIPOLAZIONE

Non viene eseguita alcuna trasformazione per non perdere interpretabilità, neanche quella logaritmica, che dopo test non porta cambiamenti sostanziali (test BoxCox = 0.22).

Divisione in train e validation:

- Train: gennaio - ottobre (91% del totale)
- Validation: novembre (9% del totale)

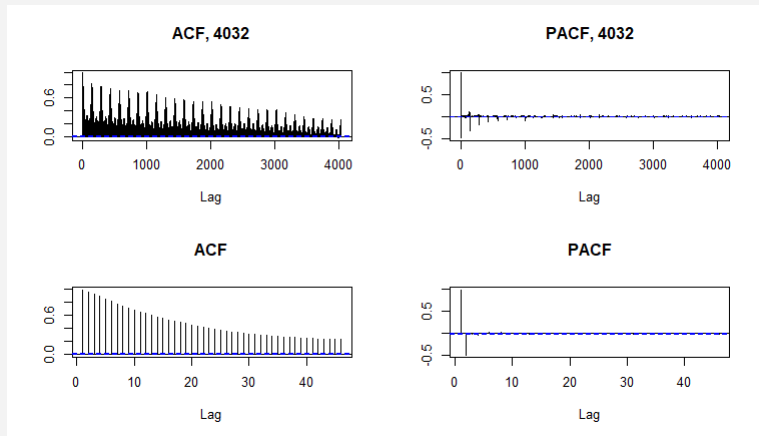


# **MODELLO BASELINE - REGRESSIONE LINEARE**

Viene eseguita una regressione lineare semplice con i seguenti regressori:

- Trend quadratico
- 6 dummy stocastiche relative ai giorni della settimana, escluso il mercoledì
- 10 seni e 10 coseni a frequenza 144 (un giorno di osservazioni)





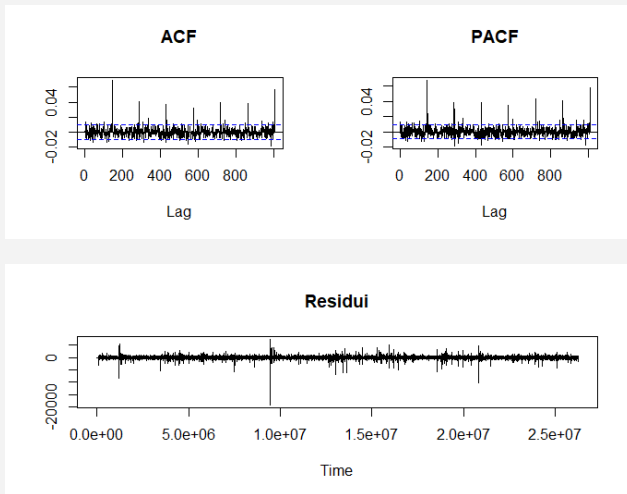
**Figure:** ACF e PACF residui della regressione

# ARIMA

Partendo da ACF e PACF della regressione sono stati sviluppati diversi modelli ARIMA, uno come sviluppo del precedente.

Di ognuno si sono osservati autocorrelogrammi, distribuzioni dei residui, bontà delle previsioni rispetto al validation e MAE. In particolare:

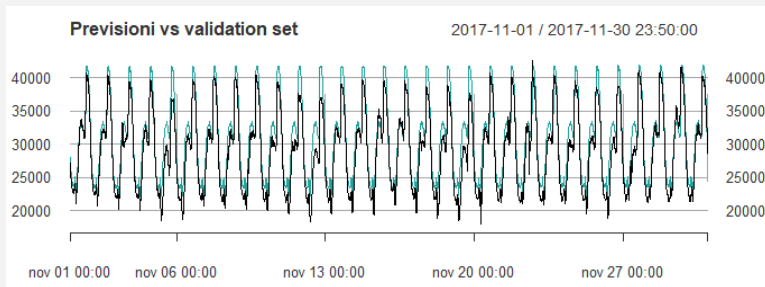
- Arimax(2,0,0) + regressori (dummy e sinusoidi frequenza 144)
- Arimax(3,0,0) + regressori (dummy e sinusoidi frequenza 144)
- Sarima(3,0,0)(0,1,0)[144]
- Sarima(3,0,0)(0,1,1)[144]



**Figure:** Correlogramma e distribuzione dei residui miglior modello

Modello	MAE	MAPE
Sarima(3,0,0)(0,1,1)[144]	1839.028	6.69136

**Table:** Risultati miglior modello ARIMA



**Figure:** Confronto tra previsioni e validation

**UCM**

Come per i modelli ARIMA vengono sviluppati diversi modelli osservando previsioni e MAE per proseguire con lo sviluppo.

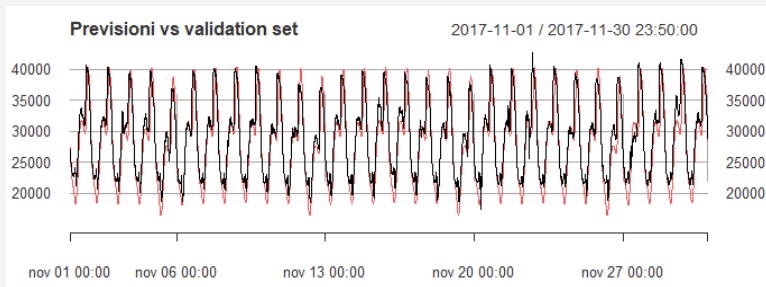
- UCM random walk + stagionalità trigonometrica 144
- UCM random walk + stagionalità trigonometrica 144 e stagionalità trigonometrica 1008
- UCM random walk + stagionalità trigonometrica 144 e 6 dummy stocastiche giornaliere

Per i modelli due e tre come parametri iniziali si sono usati quelli ottimi del modello precedente.

Sono stati effettuati anche test con local linear trend, ma hanno portato tutti a risultati peggiori.

Modello	MAE	MAPE
UCM stag trig 144 + stag trig 1008	1879.281	6.475925

**Table:** Risultati miglior modello UCM



**Figure:** Confronto tra previsioni e validation



# MACHINE LEARNING

Vengono testati diversi KNN con parametri diversi; questa tecnica è particolarmente indicata per serie molto ripetitive, come quella oggetto di studio.

Per tutti si è utilizzata la strategia MIMO e la mediana come metrica di previsione del futuro.

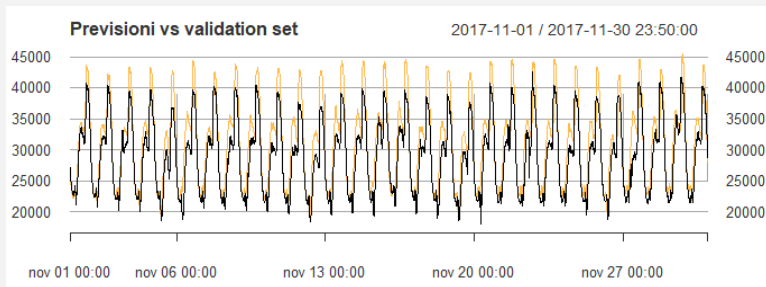
<b>Modello</b>	<b>Lags</b>	<b>K</b>	<b>MAE</b>	<b>MAPE</b>
KNN 1	1:1008	10	2346.035	8.034549
KNN 2	1:4032	10	3065.367	10.82601
KNN 3	1:1008	5	2365.022	7.99821
KNN 4	1:1008	20	2348.963	8.04137

**Table:** Risultati e configurazioni KNN

# K-NEAREST NEIGHBORS

Modello	MAE	MAPE
KNN 1	2346.035	8.034549

**Table:** Risultati miglior modello KNN



**Figure:** Confronto tra previsioni e validation

Viene creata una rete neurale ricorrente molto semplice formata dai seguenti layer:

- LSTM (25 unità)
- LSTM (20 unità)
- Dense con ottimizzatore 'adam'

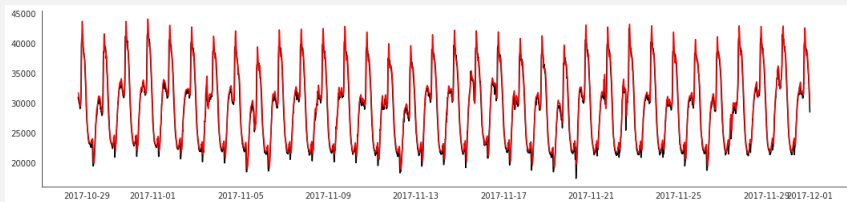
I dati vengono prima normalizzati utilizzando un MinMax Scaler e si sceglie una finestra di lookback di lunghezza 4320.

A causa dell'alto costo computazionale (dovuto al lungo lookback) la rete viene addestrata solamente per 5 epoche, sufficienti a ottenere buoni risultati.

# LSTM NETWORK

Modello	MAE	MAPE
Lstim Network	856.101	2.9414

**Table:** Risultati rete LSTM



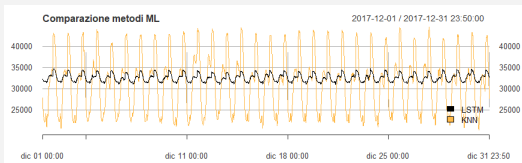
**Figure:** Confronto tra previsioni e validation

# CONFRONTO KNN E LSTM

Nonostante i buonissimi risultati della rete LSTM sul validation, nel momento della previsione dei dati di dicembre il risultato è stato poco soddisfacente.

Massimi e minimi sono risultati molto schiacciati, assumendo valori poco realistici per il tipo di dato misurato, con una varianza di due ordini di grandezza inferiore.

Per questo motivo, anche se non completamente corretto scegliere il modello migliore in base al test, per le previsioni è stato scelto KNN.



**Figure:** Confronto tra previsioni dicembre KNN e LSTM

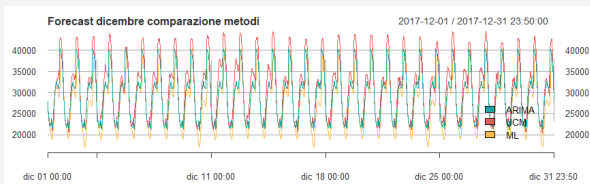
# **PREVISIONI DICEMBRE**

# PREVISIONI DICEMBRE

Si ricordano le performance dei modelli più performanti che, per la previsione del mese di dicembre, sono stati nuovamente addestrati sull'intera serie.

Modello	MAE	MAPE
ARIMA	1839.028	6.69136
UCM	1879.281	6.475925
KNN	2346.035	8.034549

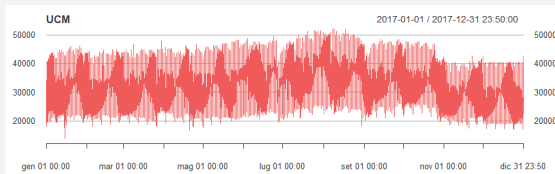
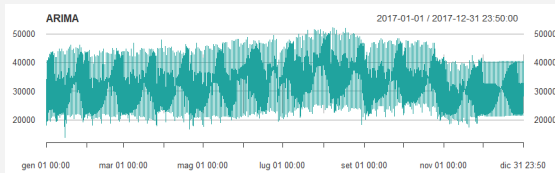
**Table:** Modelli finali e performance sul validation



**Figure:** Previsioni ARIMA, UCM e KNN mese di dicembre



# PREVISIONI DICEMBRE



GRAZIE!