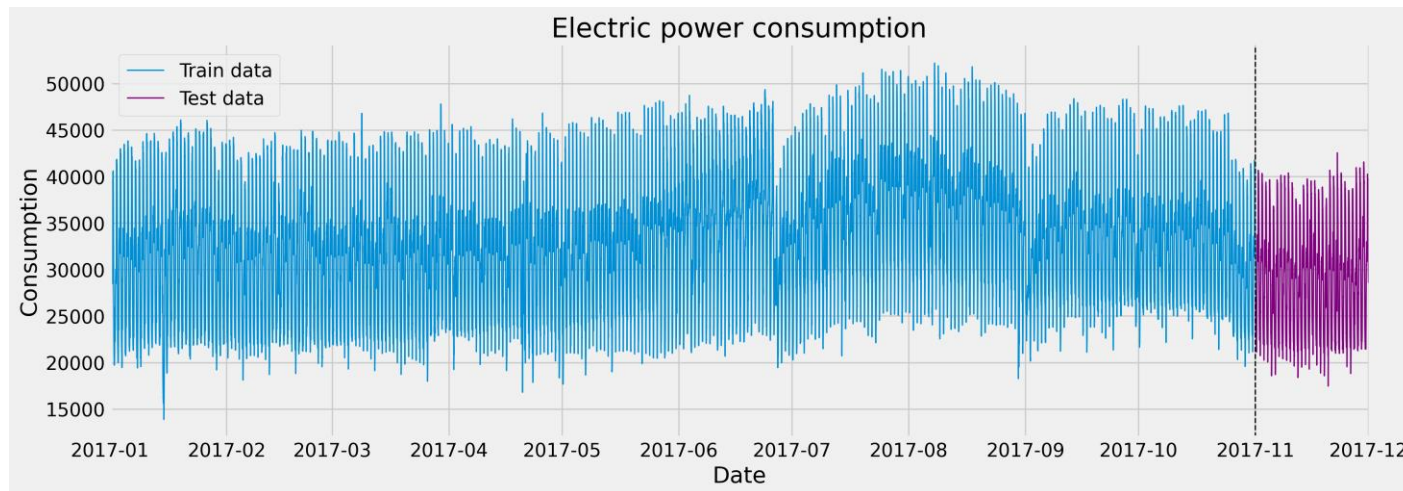


Streaming data management and Time Series Analysis

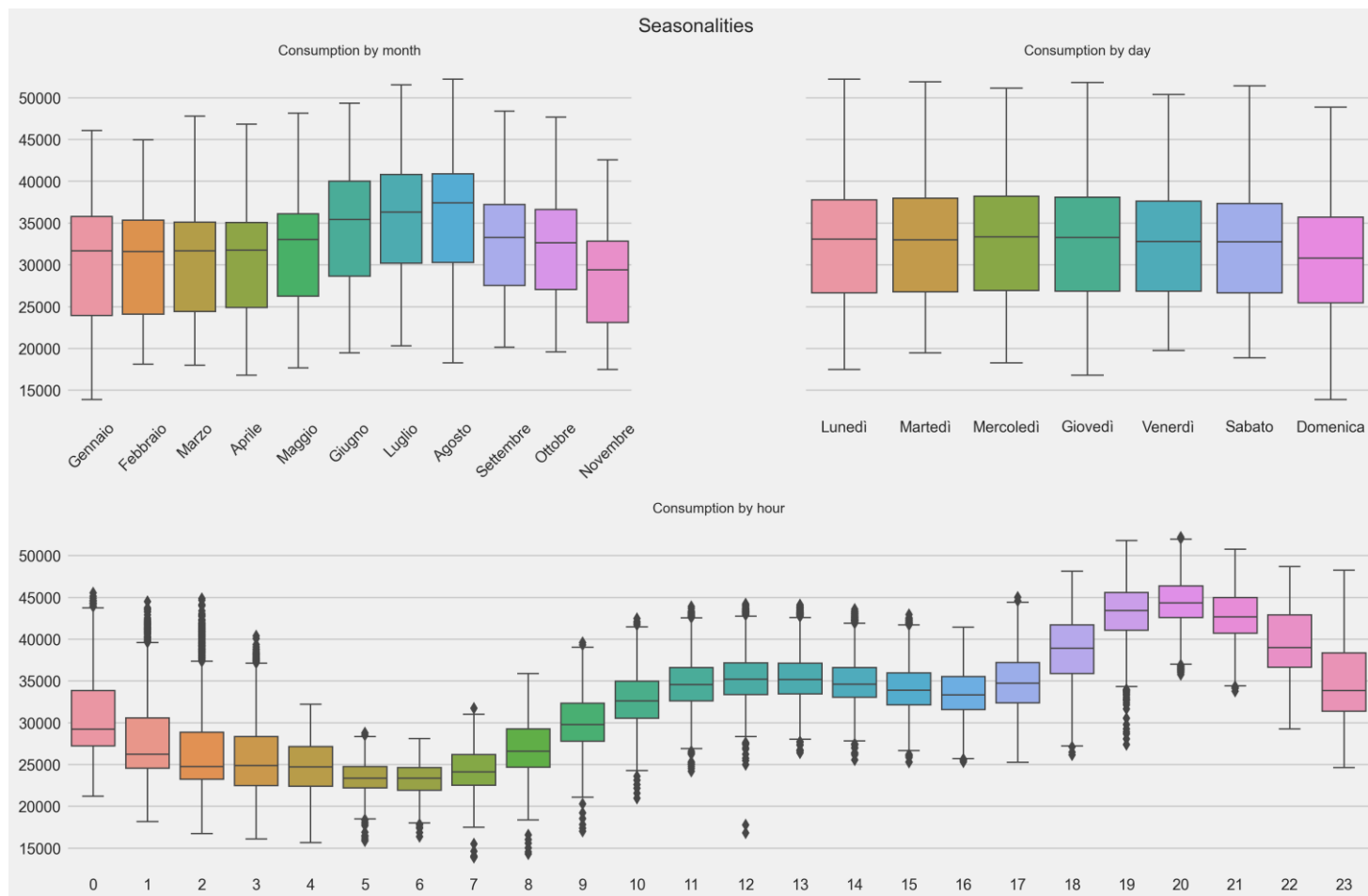
Consumo di energia elettrica in Marocco

Introduzione



L'obiettivo di questa analisi è stato quello di sviluppare 3 modelli, appartenenti alle tre famiglie tipicamente utilizzate nell'analisi di serie storiche, per studiare e prevedere correttamente il consumo di energia elettrica in Marocco in un anno.

Panoramica dei dati



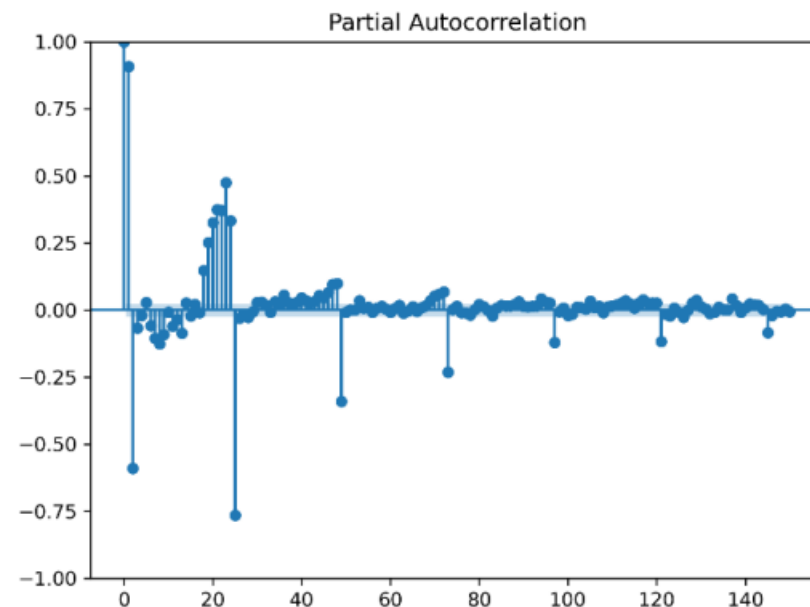
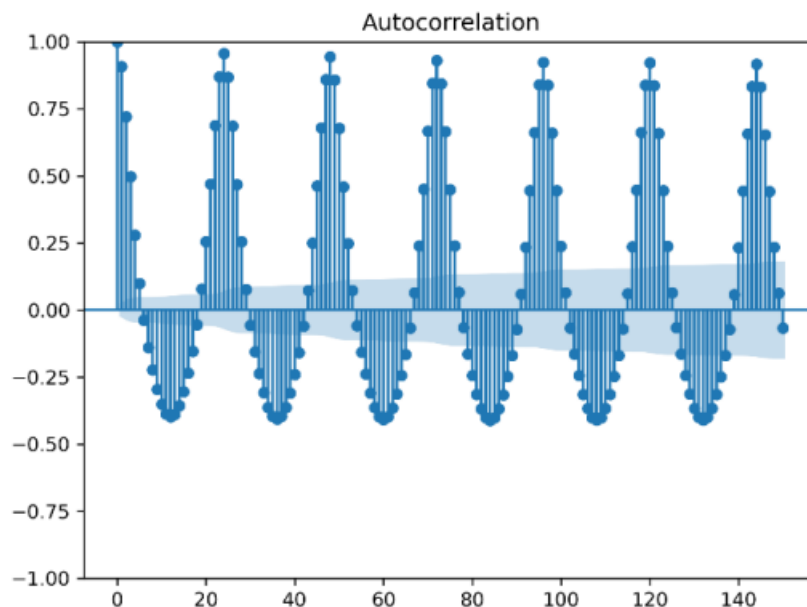
- Dati disponibili per il periodo dal 1/1/2017 al 30/11/2017
- Frequenza di un'osservazione ogni 10 minuti
- Non ci sono missing data
- Andamento stagionale, giornaliero e settimanale.

Preprocessing

- Divisione dei dati in training (dal 1/1 al 31/10), e test set (dal 1/11 al 30/11)
- Campionamento dei dati, prendendo la prima osservazione di ogni ora (solo per ARIMA e UCM).

Analisi dell'autocorrelazione

I grafici dell'ACF e della PACF hanno evidenziato la già rilevata stagionalità giornaliera.



Feature Aggiunte

Giorno della settimana	Categorica → Dummies	Una colonna per giorno della settimana.
Stagione	Categorica → Dummies	Una colonna per stagione.
Festività	Dummy	Se il giorno è festivo.
Ramadan	Dummy	Periodo di Ramadan (dal 2017-5-26 al 2017-6-24).
Ore di luce	Numerica (float)	Ore medie di luce, un valore per ogni mese.
Temperatura media diurna	Numerica (int)	Temperatura media di giorno, un valore per ogni mese.
Temperatura media notturna	Numerica (int)	Temperatura media di notte, un valore per ogni mese.

Modello ARIMA

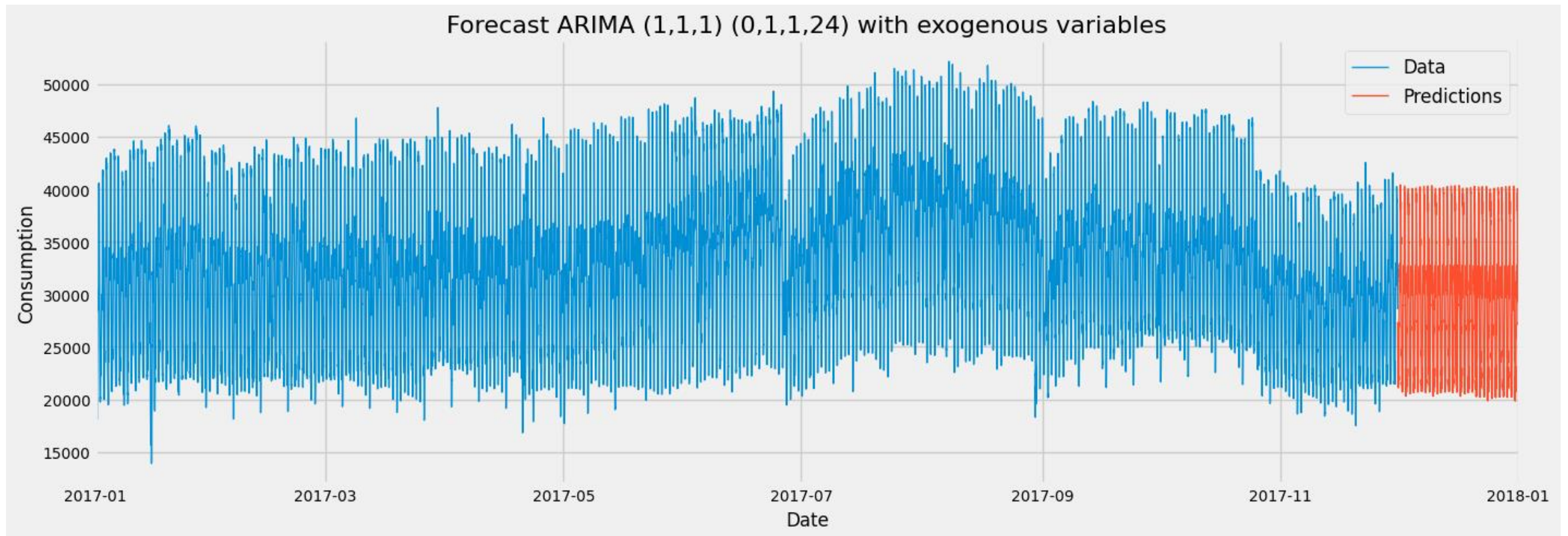
Tramite differenze successive, è stato trovato che il modello ARIMA che meglio si adatta alla serie storica utilizzata è il modello (1, 1, 1), con una parte stagionale (0, 1, 1, 24). Il modello è stato testato sia con che senza le feature esogene appena descritte. Nella seguente tabella sono riassunti i principali risultati:

Metriche	Senza variabili esogene	Con variabili Esogene
AIC	119436.71	119450.702
MAE sul training set	587.41	586.95
MAE sul test set	1395.88	1232.88

I due modelli all'incirca si equivalgono per quanto riguarda i risultati ottenuti sul training set, come dimostrano l'AIC e il MAE sui dati di training. La differenza tra i due modelli sta nella capacità di generalizzazione: il modello con variabili esogene tende a overfittare in misura minore rispetto al modello senza.

Modello ARIMA - Previsioni

Per effettuare le previsioni sono stati addestrati 6 modelli, tutti con le stesse caratteristiche, uno per ogni minuto di osservazione (:00, :10, :20, :30, :40, :50). Le previsioni sono poi state concatenate in un'unico vettore.



Modello UCM

Per stimare il modello UCM sono state utilizzate le seguenti componenti:

- Local Linear Trend
- Stagionalità stocastica di ordine 24
- Componente autoregressiva di ordine 1
- Variabili Esogene

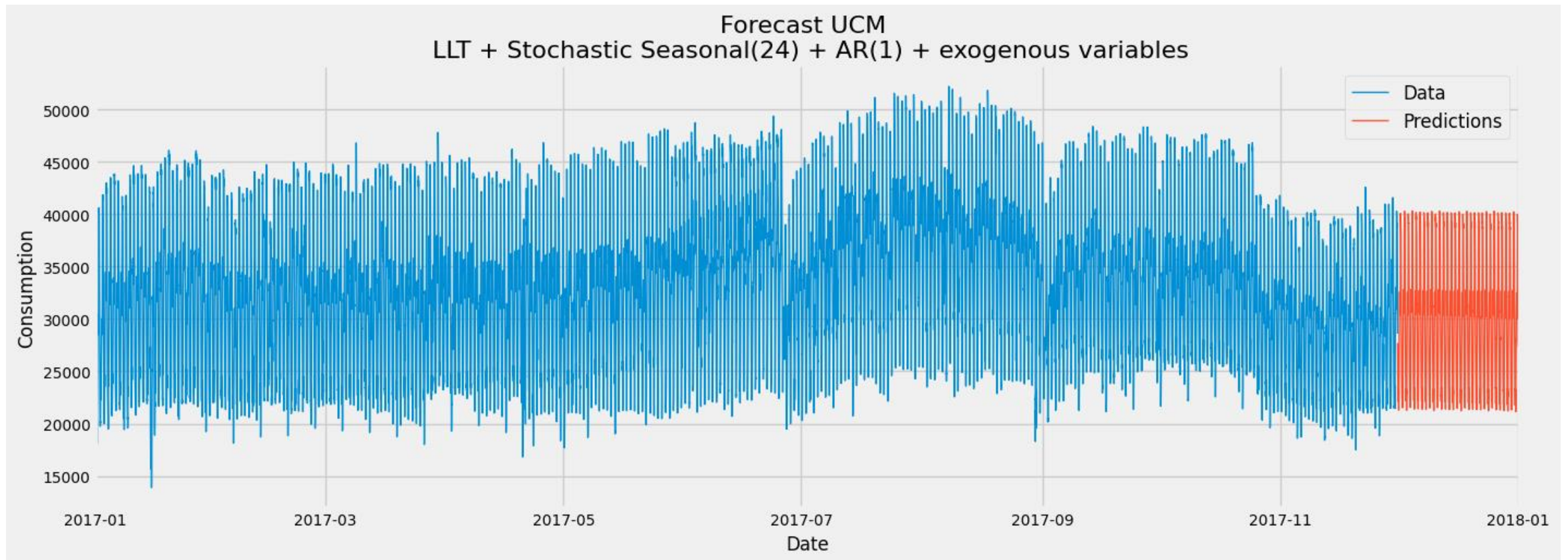
Di seguito vengono riassunti i risultati del modello:

UCM	
AIC	122389.07
MAE sul training set	797.19
MAE sul test set	1200.65

Anche qui siamo in presenza di overfitting del modello, imputabile al drastico cambio di andamento che si è verificato nel mese di Novembre, che corrisponde proprio ai dati di test.

Modello UCM - Previsioni

Per le previsioni è stata svolta la stessa procedura usata per il modello ARIMA, ossia la stima di 6 modelli tutti con le stesse componenti.



Modello Machine Learning - XGBoost

Per la famiglia di modelli di Machine Learning è stato utilizzato l'algoritmo XGBoost, che impara dalle feature a prevedere il valore desiderato. In particolare, alle feature già create precedentemente, sono state aggiunte le feature nella tabella. Il modello è stato quindi allenato a tenere in considerazione i 6 lag precedenti, per prevedere il valore successivo.

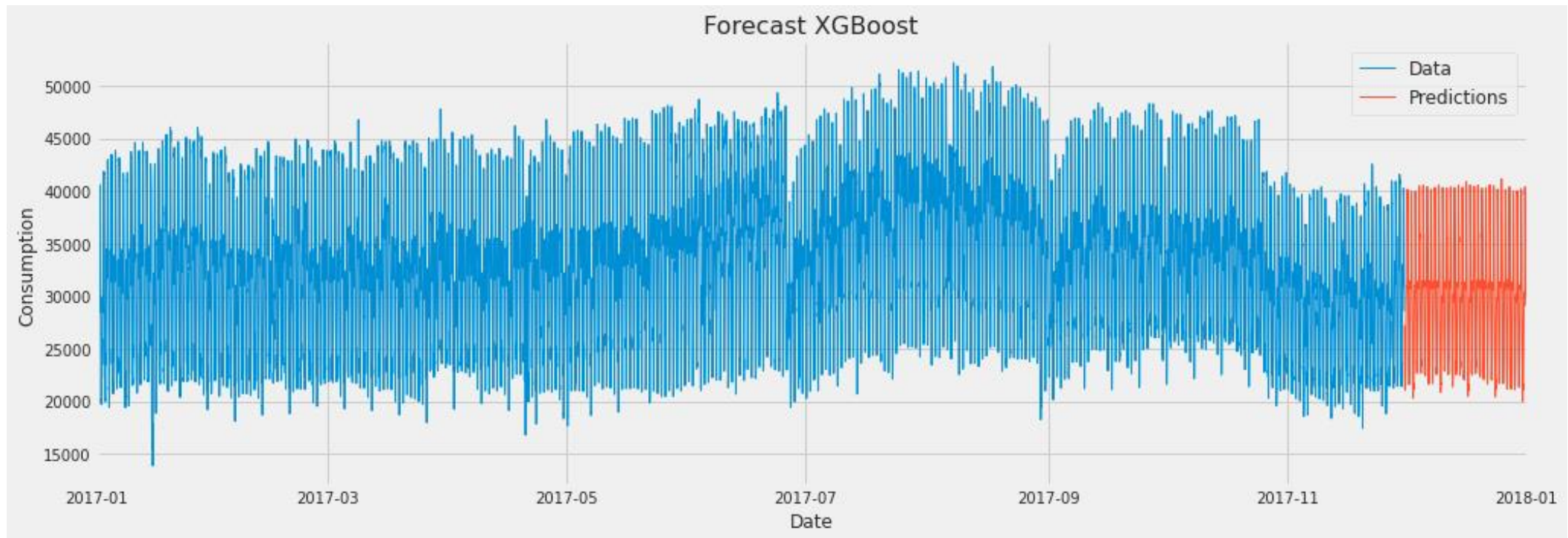
FEATURE	DESCRIZIONE	FEATURE	DESCRIZIONE
Lags	Una colonna per ciascuna delle 6 osservazioni precedenti	Settimana dell'anno	In quale settimana dell'anno è stata fatta l'osservazione
Giorno del mese	Indice del giorno del mese	Ora	Ora del giorno dell'osservazione
Trimestre	In quale trimestre è stata fatta l'osservazione	Mese	Mese dell'osservazione

Modello Machine Learning - XGBoost

I parametri utilizzati sono stati i seguenti:

- 130 stimatori
- Learning rate di 0.05
- Profondità degli alberi pari a 50

I risultati ottenuti sul training set sono stati molto scarsi (**MAE 2962.12**), tuttavia le previsioni su Dicembre sono sembrate verosimili:



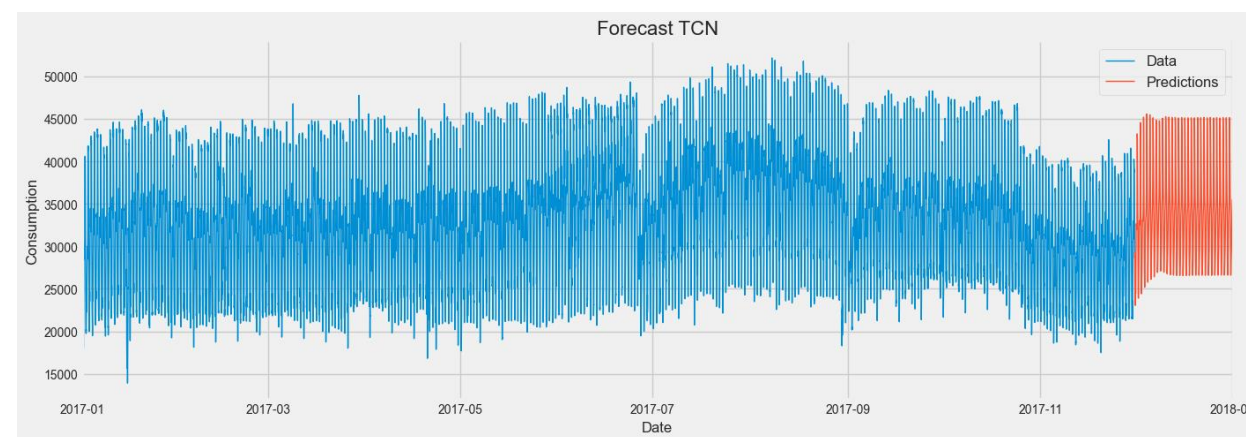
Modello Deep Learning – Temporal Convolutional Network

Per svolgere questo task, è stato anche sperimentato un approccio di Deep Learning: la Temporal Convolutional Network.

La Temporal Convolutional Network (TCN) è una tipologia di rete neurale artificiale che si concentra sull'elaborazione di dati temporali. La logica alla base della TCN consiste nell'utilizzo di convoluzioni per catturare relazioni a lungo termine tra i dati, permettendo alla rete di imparare a prevedere eventi futuri.

La rete utilizzata è stata impostata con i seguenti parametri:

Parametri	
Lags	4
Dilations	1,2,3,23,24,25,47,48,49, 142,143,144
Dropout rate	0.2
Activation	ReLU
Optimizer	Adam con learning rate = 0.001
Loss	Mean squared error
Epochs	56



I risultati ottenuti tuttavia sono inverosimili e pertanto è stato preferito l'algoritmo XGBoost.

Fine della presentazione

Grazie dell'attenzione