

PROGETTO SDMTSA

Introduzione

Il progetto ha come obiettivo la definizione, lo sviluppo e l'implementazione di diversi algoritmi con lo scopo di effettuare una previsione dei prezzi del mercato energetico. In particolare, vengono sviluppati diversi modelli appartenenti a 3 categorie:

- Modelli ARIMA
- Modelli UCM
- Modelli non-lineari (ML)

Per realizzare la previsione viene utilizzato un dataset che include i prezzi dell'energia elettrica aggregati a livello giornaliero ed orario. I dati si riferiscono ad un periodo che va dal 01/09/2018 al 01/09/2020. Il file messo a disposizione è costituito semplicemente da tre colonne:

- Data (dd/mm/yyyy)
- Ora (1-...-24)
- Valore

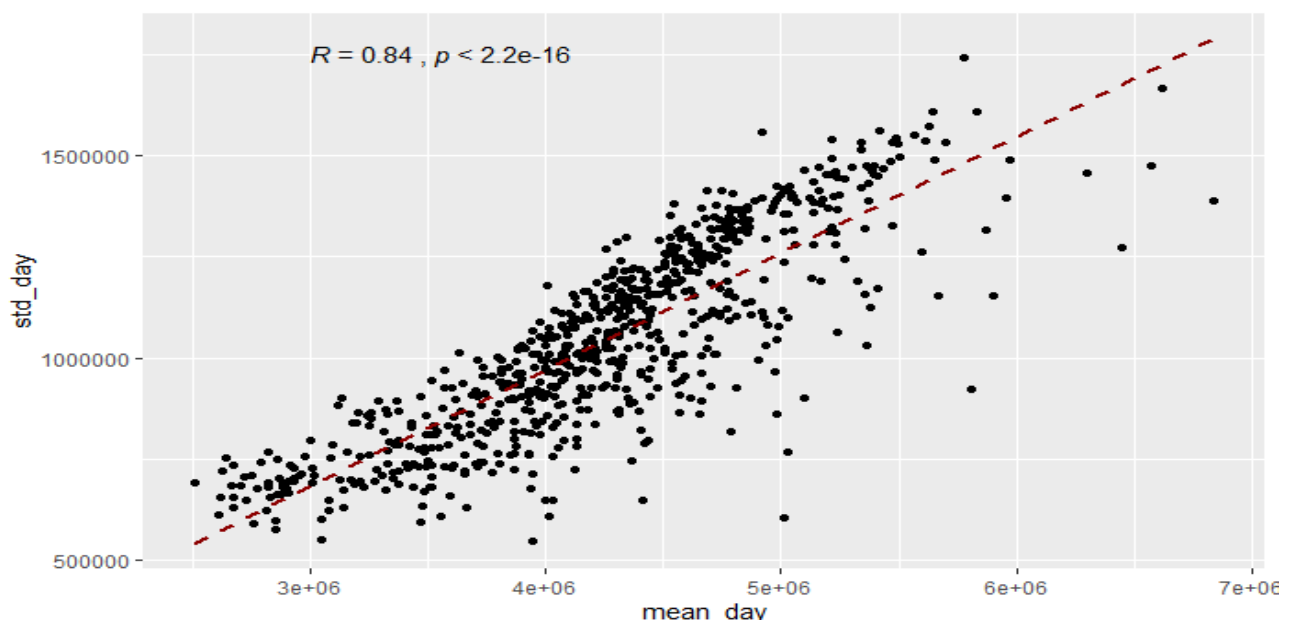
L'obiettivo finale dell'analisi consiste nel prevedere l'andamento della serie storica fino al 01/02/2021 cioè due mesi in avanti.

Analisi e Preprocessing

Esplorando la serie storica è emerso che il giorno 31/05/2020 è completamente mancante. Per risolvere tale problema si è deciso di sostituire le ore mancanti con la media della stessa ora dei giorni 30 Maggio e 1 Giugno del 2020.

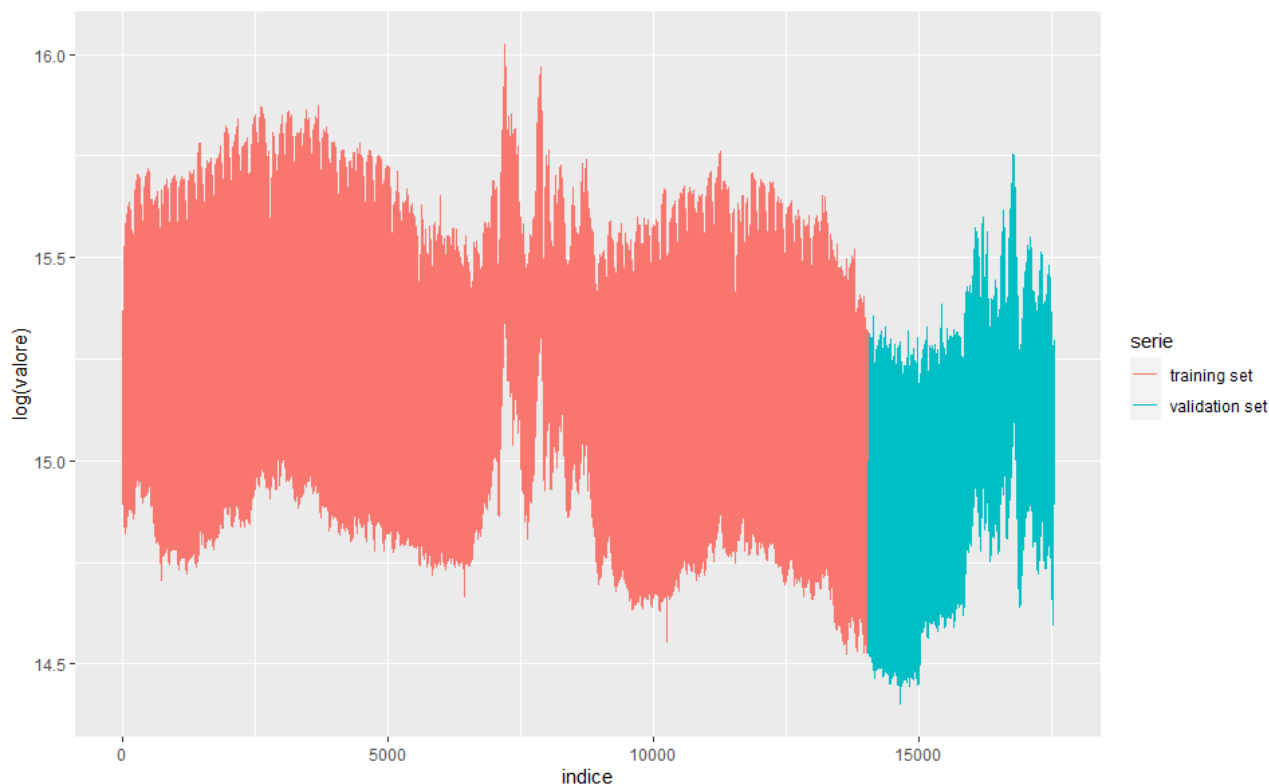
Inoltre, in corrispondenza dei giorni 31/03/2019 e 29/03/2020 si osserva la mancanza di un'ora (3 AM), ciò dovuto al cambio ora legale-ora solare. Si è deciso allora di sostituire questo valore mancante replicando l'ora precedente (2 AM).

Successivamente, si è plottato uno scatterplot dove si confronta la media giornaliera con la deviazione standard giornaliera al fine di verificare se sia necessario applicare una trasformazione ai dati.



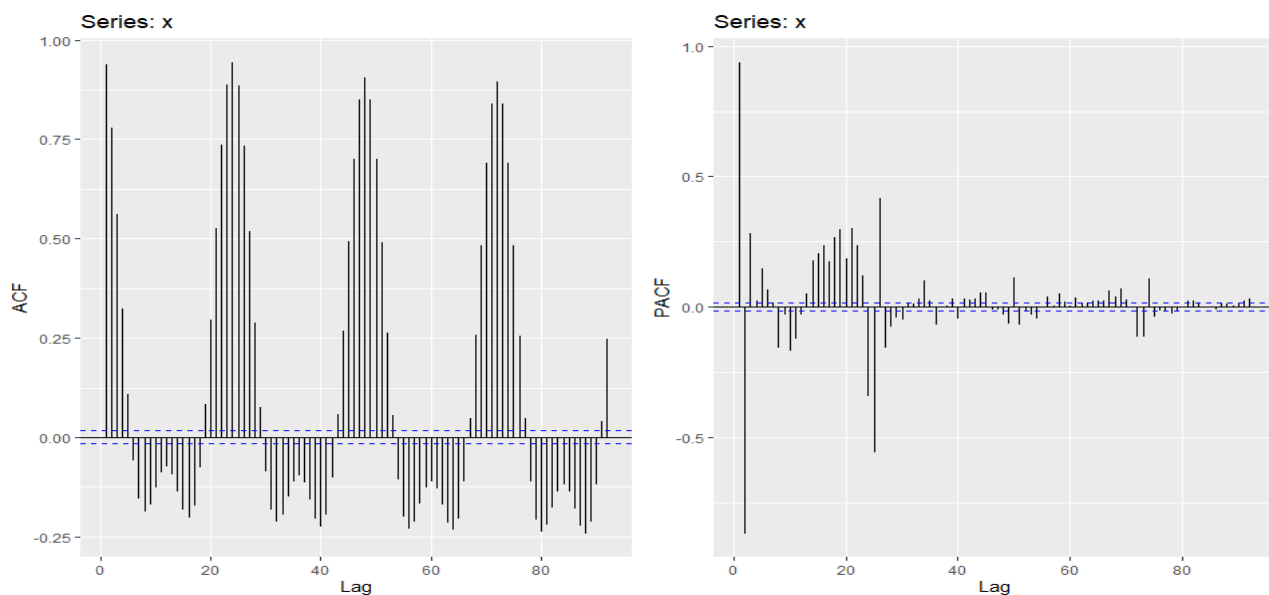
Come si osserva dalla figura, la relazione lineare tra la media giornaliera e la deviazione standard giornaliera è molto forte, come segnala il coefficiente di correlazione lineare pari a 0.84. Per tale ragione, si è ritenuto necessario applicare la trasformazione logaritmica ai dati.

Infine, si è suddiviso il dataset in train e validation set con il primo contenente l'80% delle osservazioni e il secondo il restante 20%.

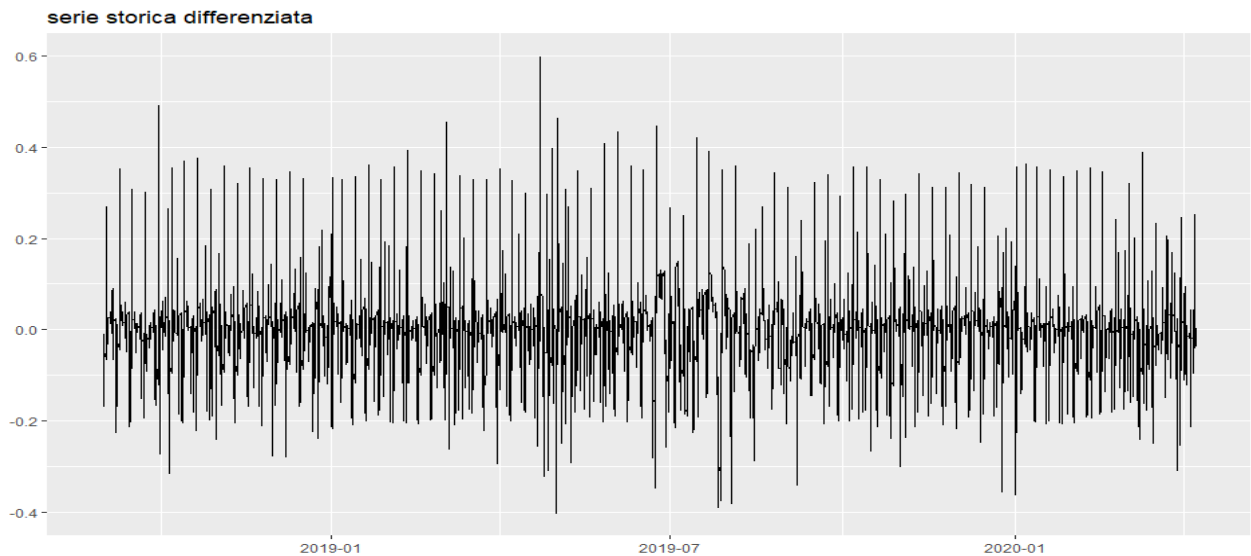


ARIMA

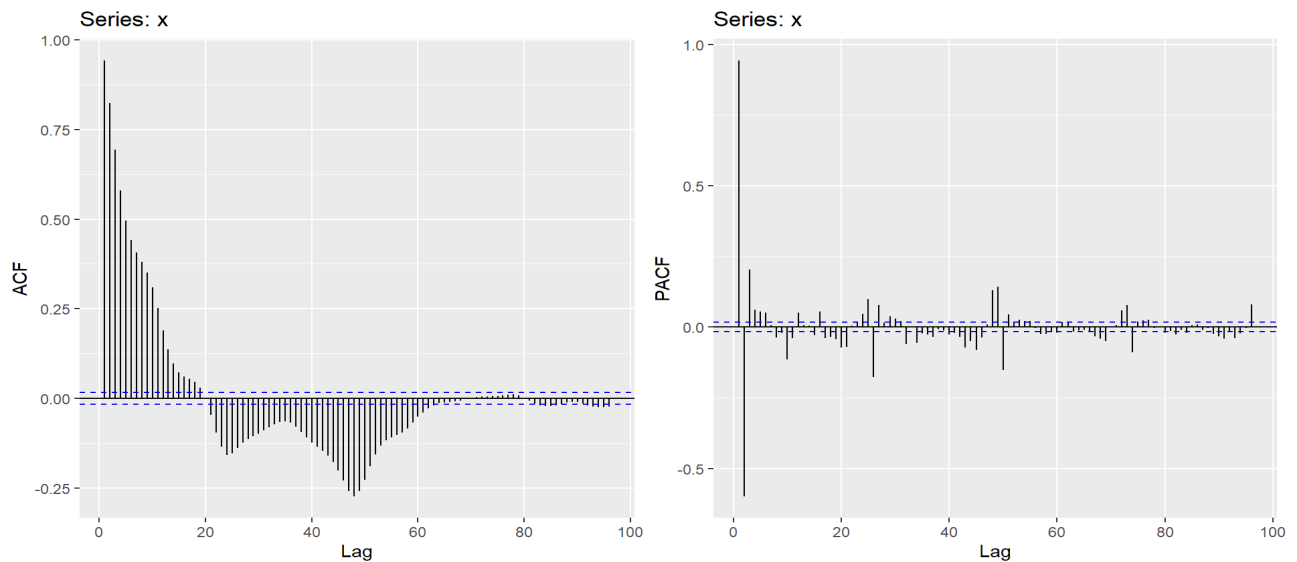
Per stimare i coefficienti del modello Arima è stata eseguita la procedura di Box-Jenkins che prevede l'iniziale analisi dei correlogrammi.



Dai grafici ACF e PACF si osservano dei picchi ogni 24 periodi (ore). Pertanto, si procede ad effettuare una differenza stagionale di ordine 24.



Da tale grafico la serie integrata sembra essere stazionaria. Tale impressione visiva è stata confermata dal test di Dickey-Fuller.



Dai correlogrammi della serie integrata sembra emerge la presenza di un possibile MA(2) stagionale in quanto nel grafico della ACF il 24-esimo e 48-esimo ritardo spiccano mentre nella PACF i ritardi multipli di 24 rientrano gradualmente verso lo zero.

Inoltre, nel PACF si nota che i primi 3 ritardi sono molto alti. Pertanto si procede a stimare un modello $ARIMA(3,0,0)(0,1,2)[24]$.

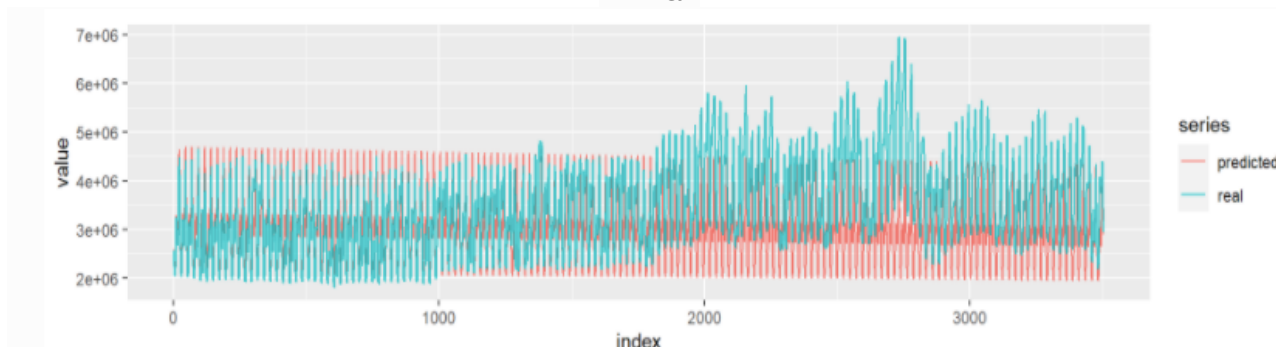
Successivamente, per il secondo modello, si è deciso di aggiungere a questo modello dei regressori sinusoidali per poter cogliere le stagionalità infra-annuali non catturata dalla semplice integrazione stagionale.

Infine, il terzo modello oltre ad utilizzare alle sinusoidi, utilizza anche delle variabili dummy associate ad:

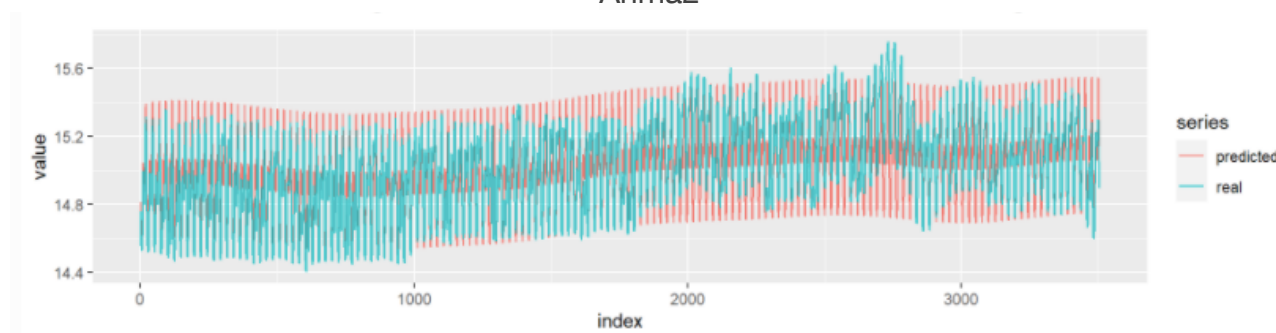
- Weekend
- Primo Gennaio
- Pasqua e Pasquetta
- Ferragosto
- Natale e Santo Stefano

Di seguito vengono inseriti i grafici delle previsioni (sul validation set) per i tre modelli con i relativi punteggi di MAE (mean absolute error).

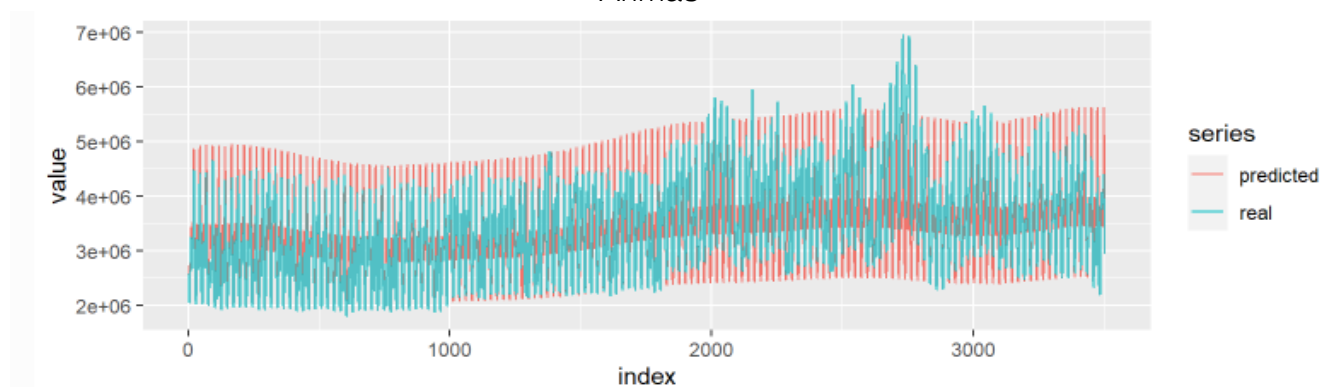
Arima1



Arima2



Arima3



Modello	MAE
Arima 1	65367
Arima 2 (sinusoidi)	419742
Arima 3 (sinusoidi+dummy)	419985

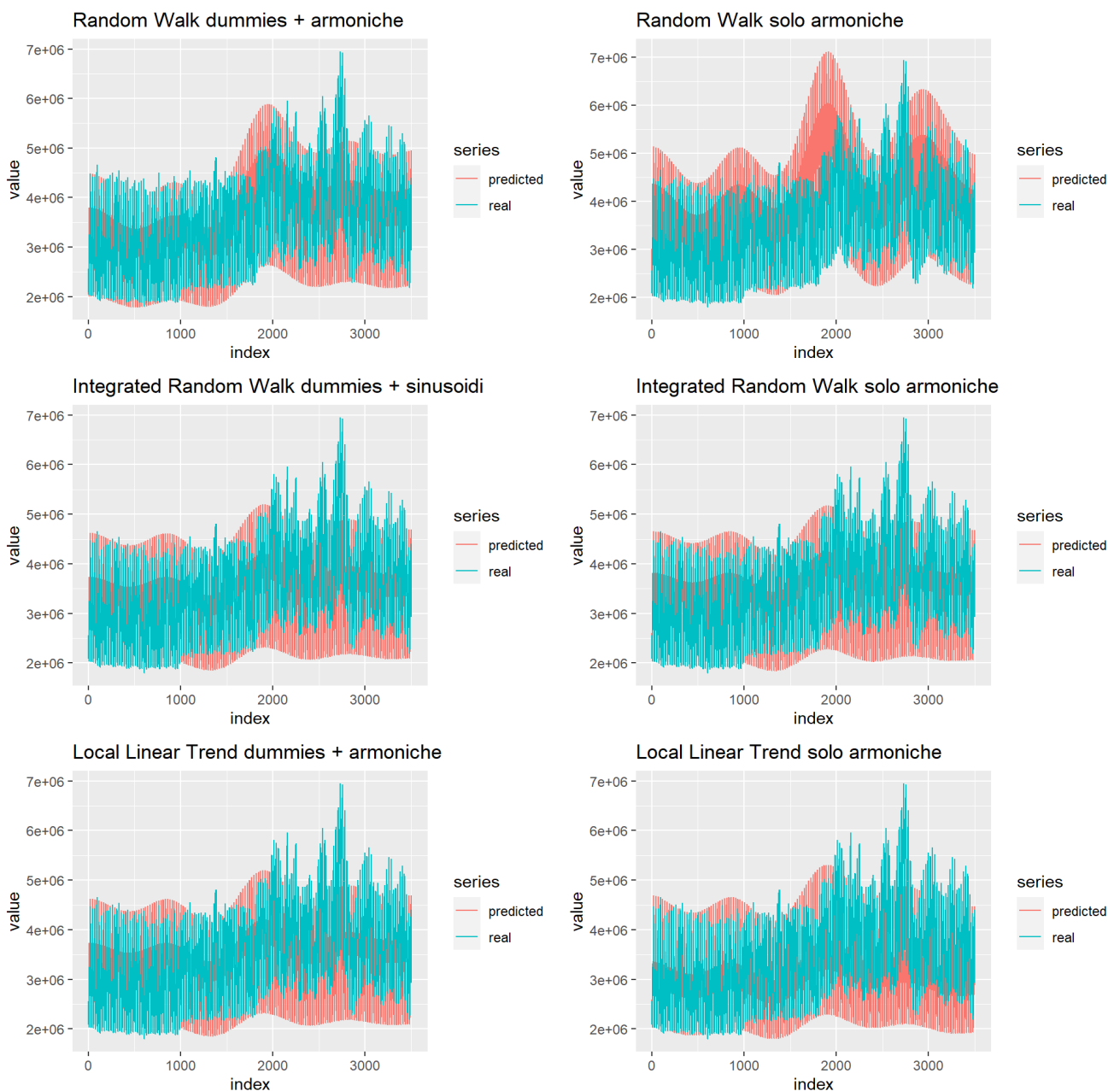
Modelli UCM

I modelli UCM vengono utilizzati per combinare diversi componenti quali ciclo, trend e stagionalità. In particolare, in questo progetto sono stati utilizzati:

- Local Linear Trend
- Random Walk
- Integrated Random Walk

Per ciascuno di essi si utilizzano due diversi approcci per gestire la stagionalità giornaliera ed annua. Il primo approccio consiste nell'usare le dummies stocastiche per la stagionalità giornaliera e 10 sinusoidi stocastiche per la stagionalità annua. Il secondo approccio consiste nell'utilizzo di 5 armoniche per la stagionalità giornaliera e ancora 10 armoniche per la stagionalità annua.

Come per i modelli ARIMA, sono riportati di seguito le previsioni (sul validation test) e la tabella dei punteggi (MAE).



Modello	Armoniche	Armoniche+Dummy
Random Walk	422541	612673
Integrated Random Walk	485634	494310
Local Linear Trend	484865	564000

Modelli Machine Learning

Per i modelli non lineari sono state utilizzate due diverse tecniche per effettuare la previsione: RNN (con due diverse architetture) e KNN.

In particolare, per quanto riguarda la prima tipologia sono stati scelti le architetture LSTM (Long Short-Term Memory) e GRU (Gated Recurrent Unit) poiché permettono di conservare le informazioni sul passato analizzando i dati in maniera sequenziale.

Sia per LSTM che per GRU la struttura è identica: batch size pari a 24, architettura a due hidden layers con il primo strato contenente 128 neuroni mentre il secondo 64 con dropout rate del 20% per evitare overfitting. I modelli sono stati allenati per 15 epoche con ottimizzatore Adam e funzione di perdita il MAE.

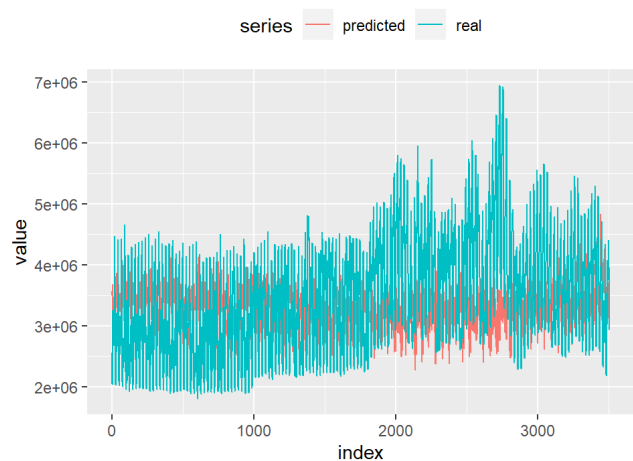
Per quanto riguarda il KNN viene settato il parametro k uguale a 2 e scelti come storia per la previsione 15 giorni e per la previsione si sceglie la strategia multioutput.

Come per i modelli precedenti si verificano le previsioni sul validation set.

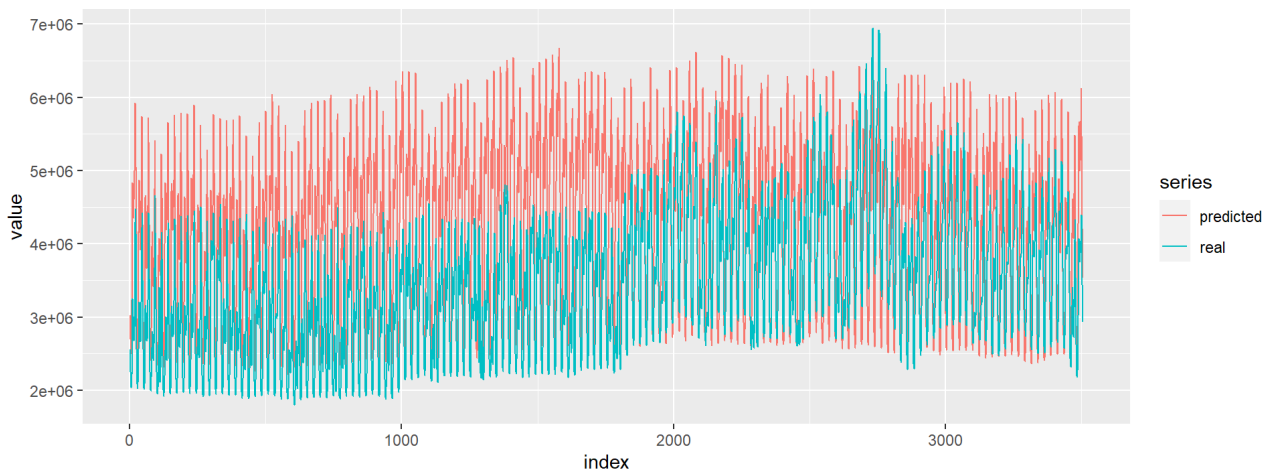
LSTM



GRU



KNN



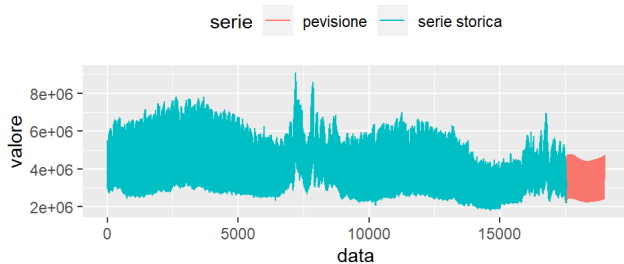
Modello	MAE
LSTM	974566
GRU	918021
KNN	908322

Confronto Finale

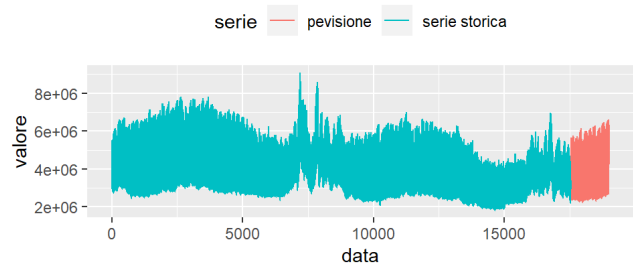
L'obiettivo è stato fare la previsione della serie dal 01/09/20 al 01/11/20 utilizzando il miglior modello (sui risultati nel validation set) per ciascuna categoria.

Il migliore dei non lineari è il KNN mentre per gli UCM è il Random Walk con dummy stocastiche per la stagionalità giornaliera e sinusoidi stocastiche per la stagionalità annua. Infine, il miglior modello ARIMA (che risulta migliore rispetto ai ML e UCM) è nella forma $(3,0,0)(0,1,2)[24]$ con 10 regressori sinusoidali. Di seguito i relativi grafici.

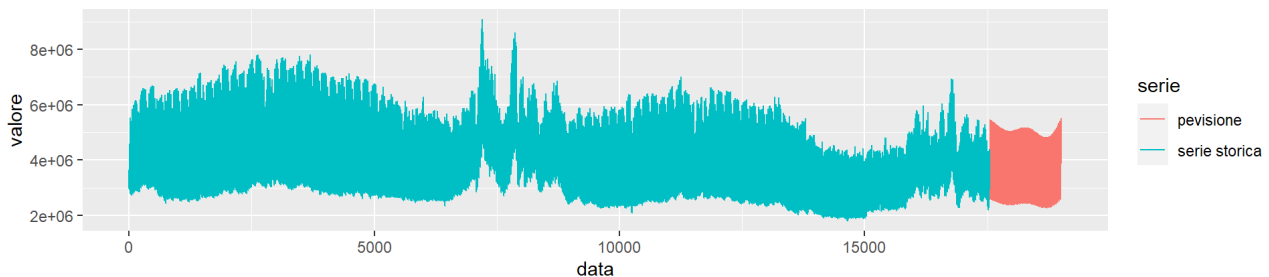
previsioni ARIMA



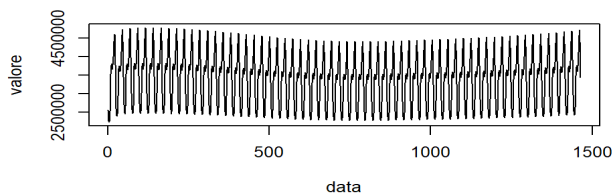
previsioni KNN



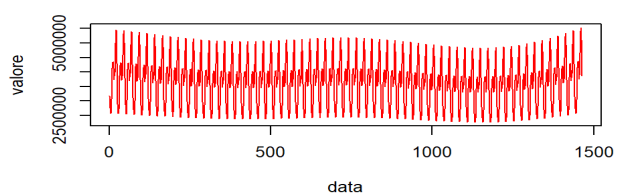
previsioni UCM



previsioni ARIMA



previsioni UCM



previsioni KNN

