

PREVISIONI AL BUIO

PROGETTO A CURA DI LAURA CAVENATI

CORSO DI STREAMING DATA MANAGEMENT AND TIME SERIES ANALYSIS

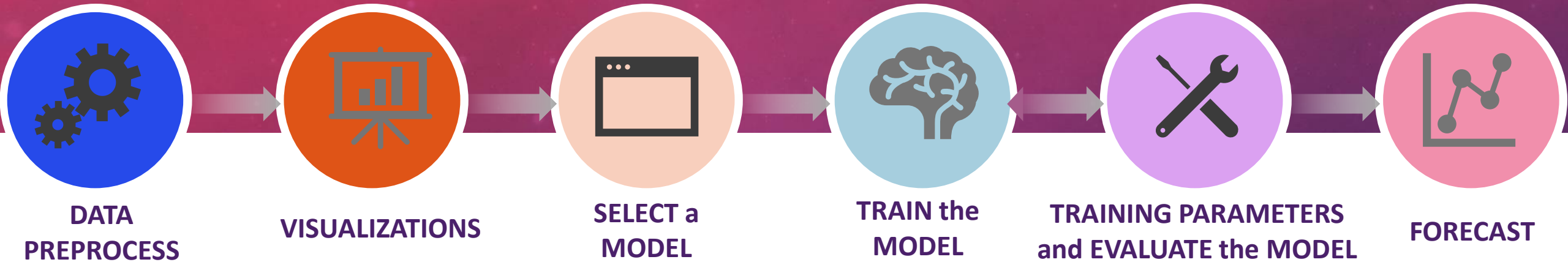
A.A: 2020/2021

LAUREA MAGISTRALE IN DATA SCIENCE

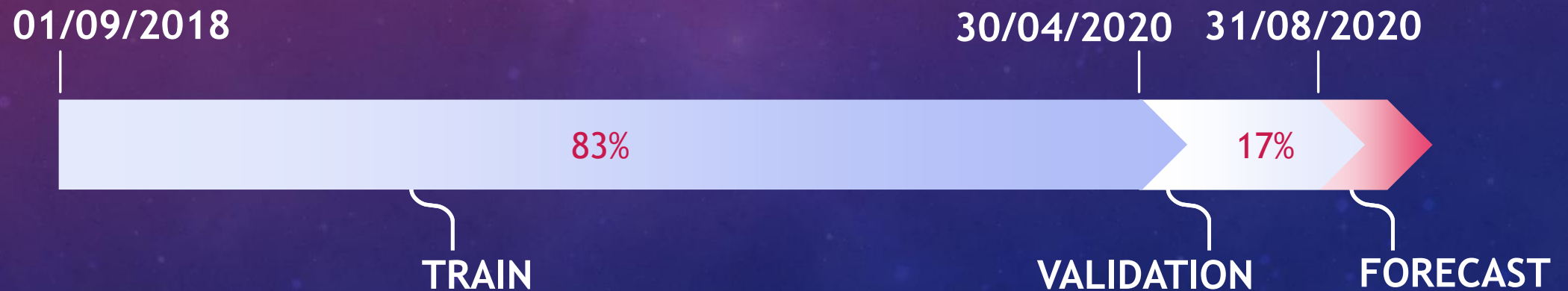
UNIVERSITÀ DEGLI STUDI DI MILANO - BICOCCA



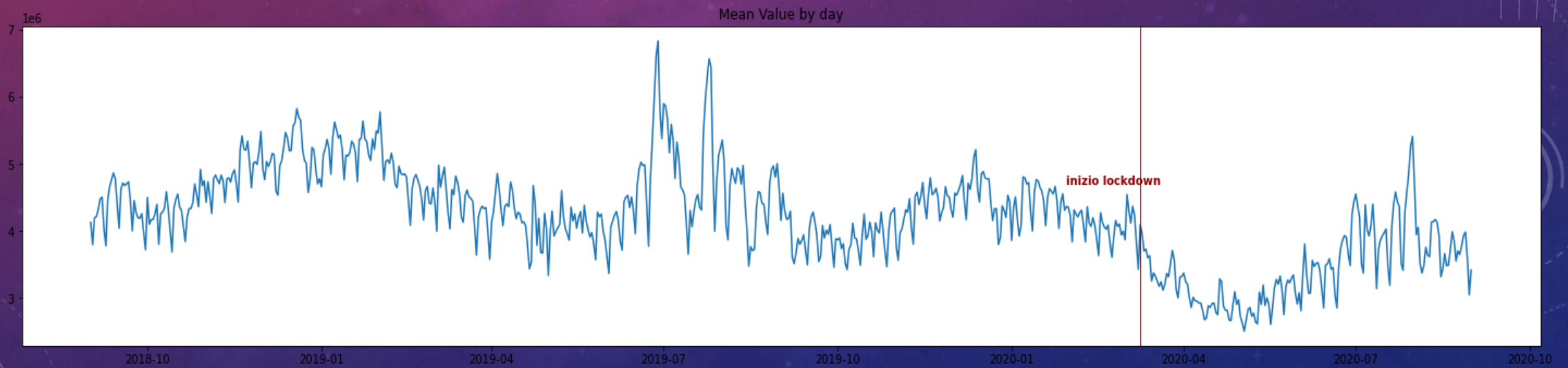
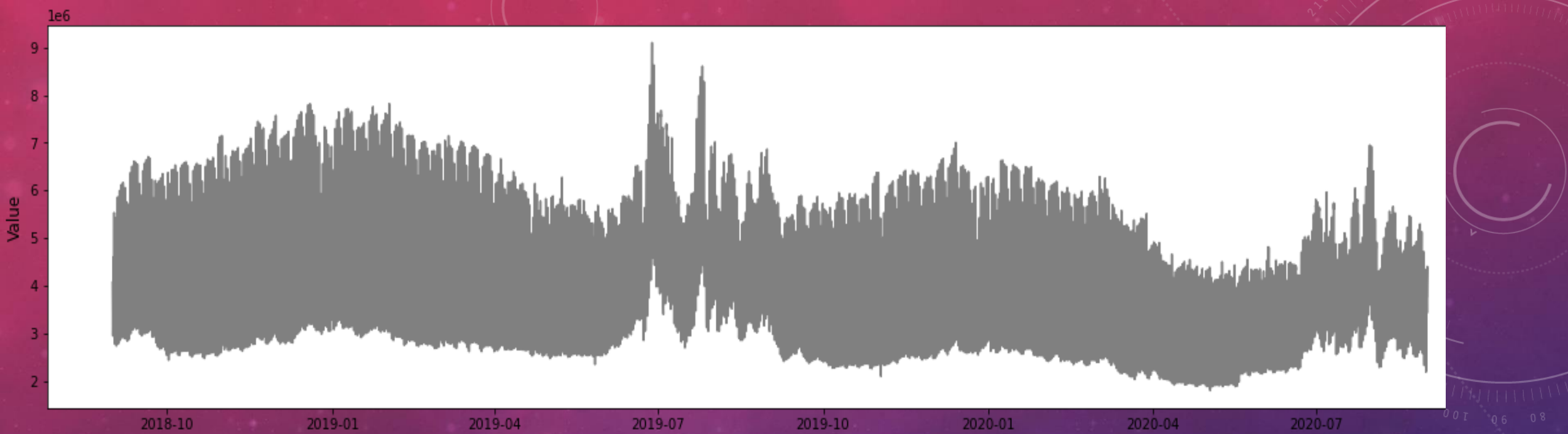
FLOWCHART



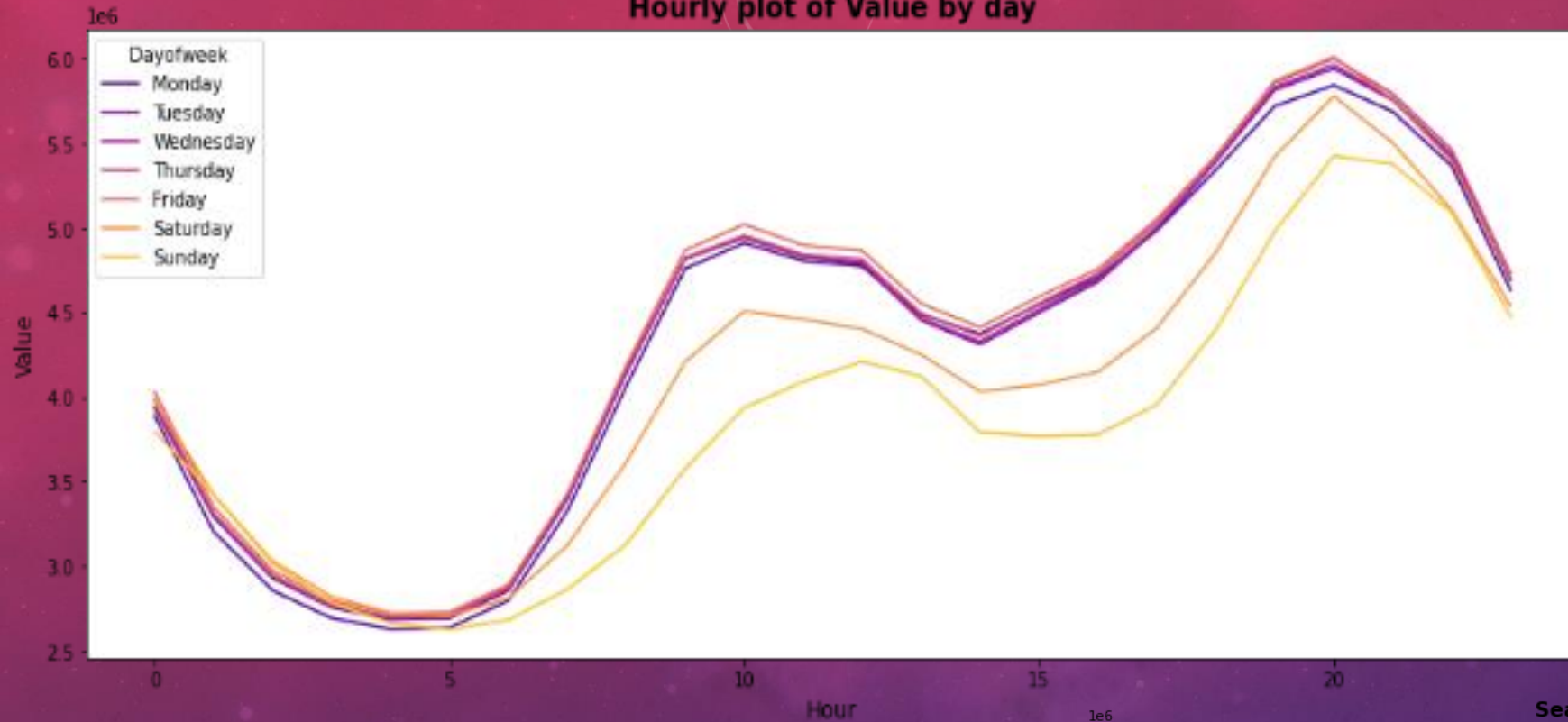
- MISSING VALUES: 2019-03-31 ore 3, 2020-03-29 ore 3 e 2020-05-31
- TRAIN/VALIDATION SPLIT:



VISUALIZZAZIONI INIZIALI



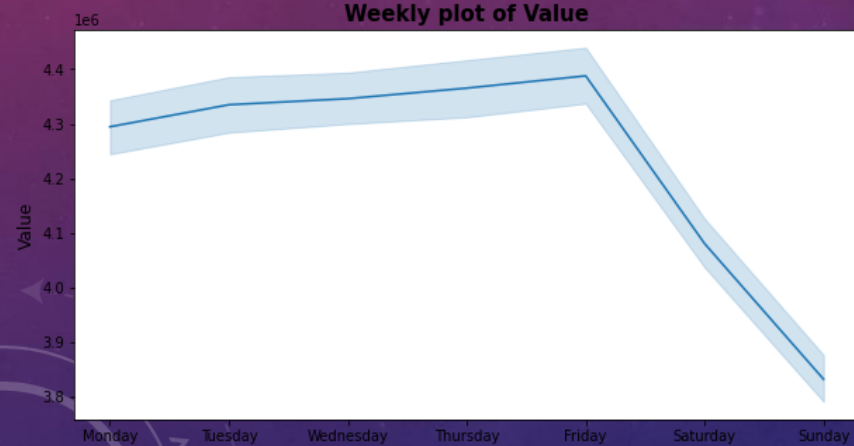
Hourly plot of Value by day



Stagionalità

- GIORNALIERA
- SETTIMANALE
- ANNUALE

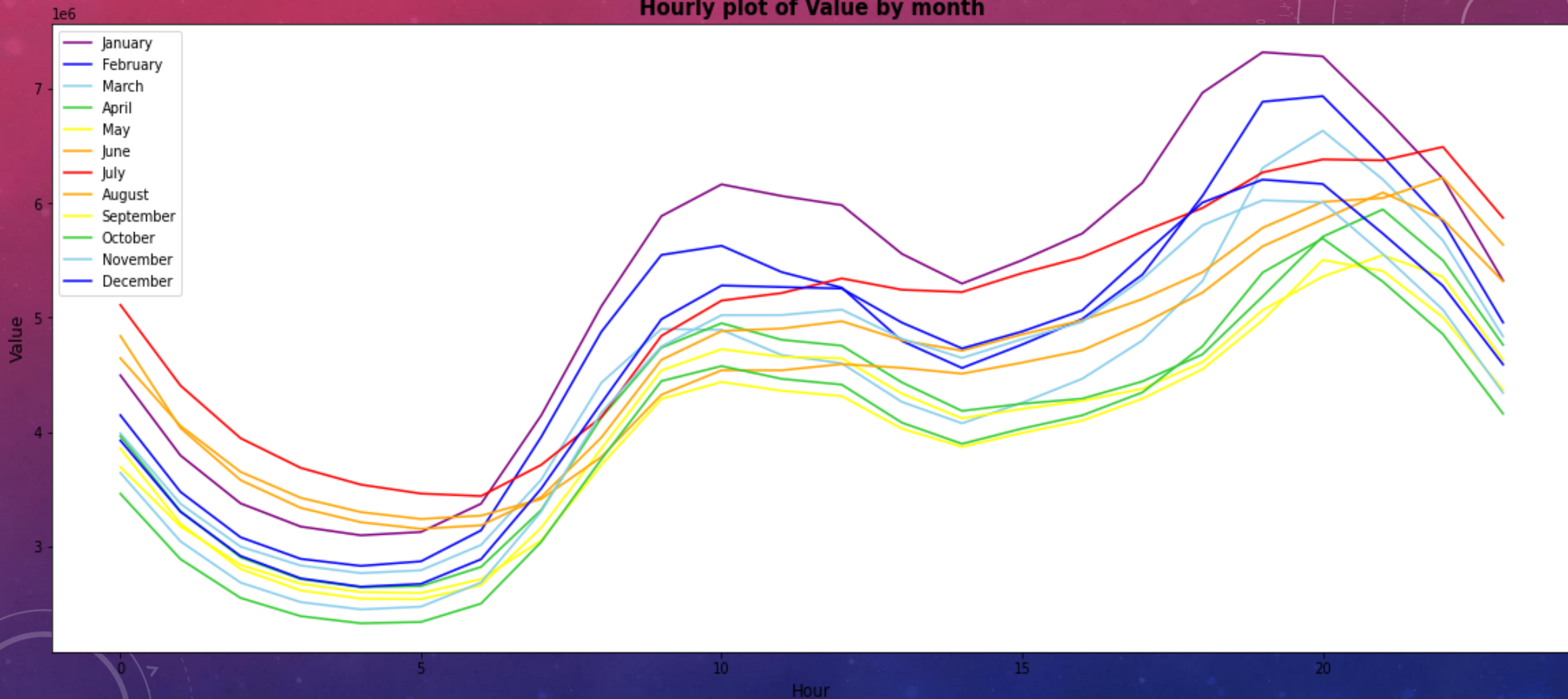
Weekly plot of Value



Seasonal plot of Value



Hourly plot of Value by month

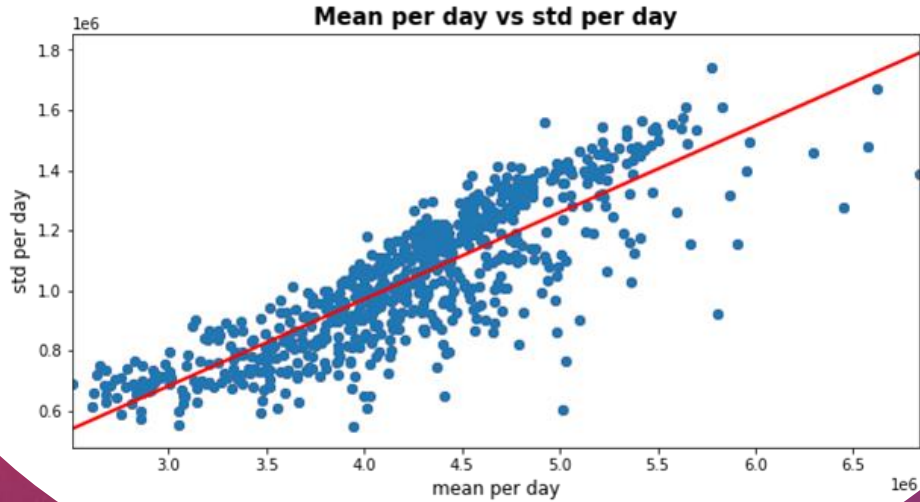


MODELLI

The background is a solid dark blue color. It is decorated with several faint, light blue circular and semi-circular patterns. These include concentric circles, arcs, and dashed lines, some of which have small arrowheads pointing in various directions. Some of the larger circles have numerical labels around their perimeters, such as 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, and 260. The overall aesthetic is technical and geometric.

ARIMA

Stazionarietà in varianza



LA SERIE è STAZIONARIA?

test

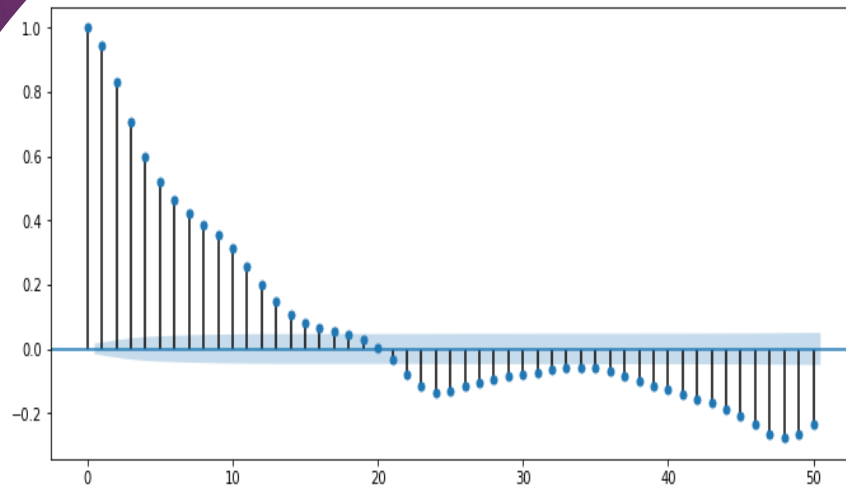
Results of Dickey-Fuller Test:

Test Statistic	-5.810651e+00
p-value	4.409576e-07
Lags Used	4.200000e+01
Number of Observations Used	1.454900e+04
Critical Value (1%)	-3.430800e+00
Critical Value (5%)	-2.861739e+00

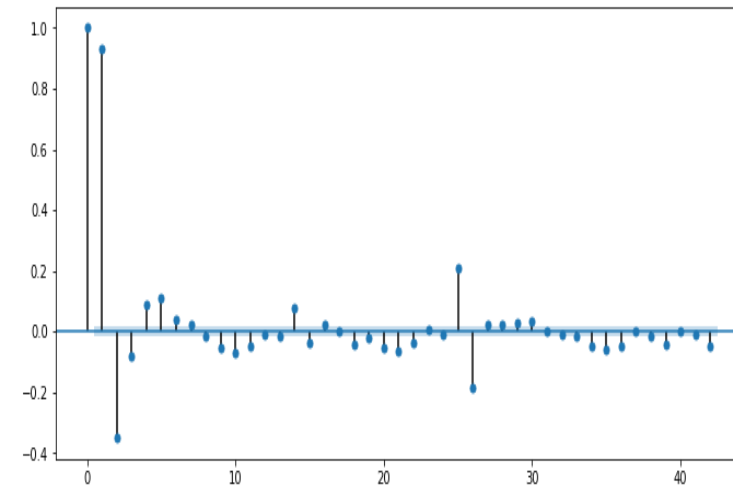
Results of KPSS Test:

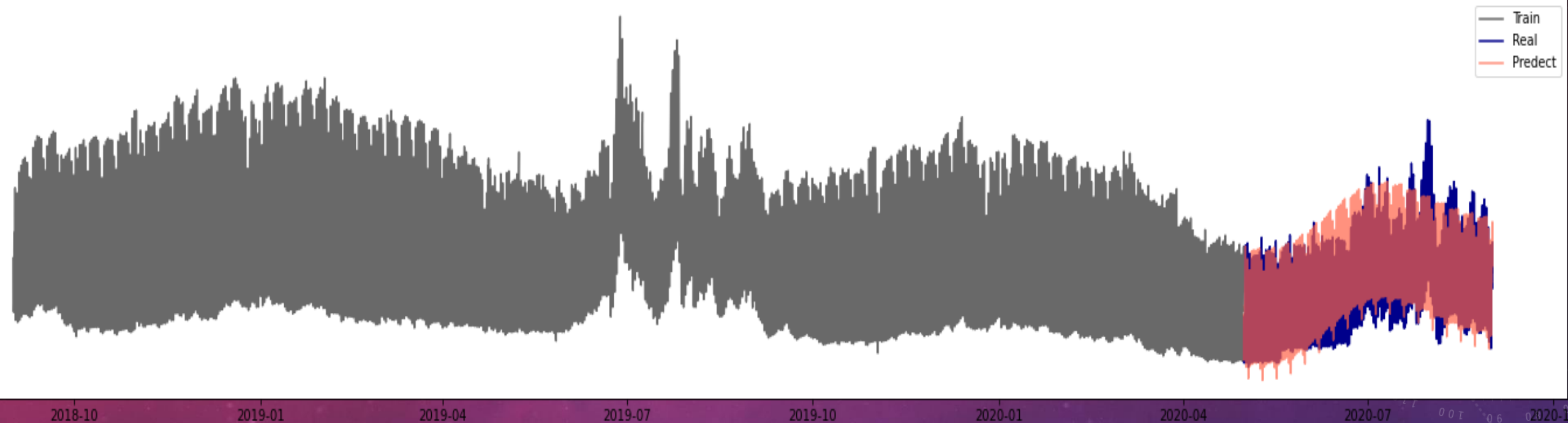
Test Statistic	9.070323
p-value	0.010000
Lags Used	42.000000
Critical Value (10%)	0.347000
Critical Value (5%)	0.463000

ACF - 24 Order Diff



PACF - 24 Order Diff



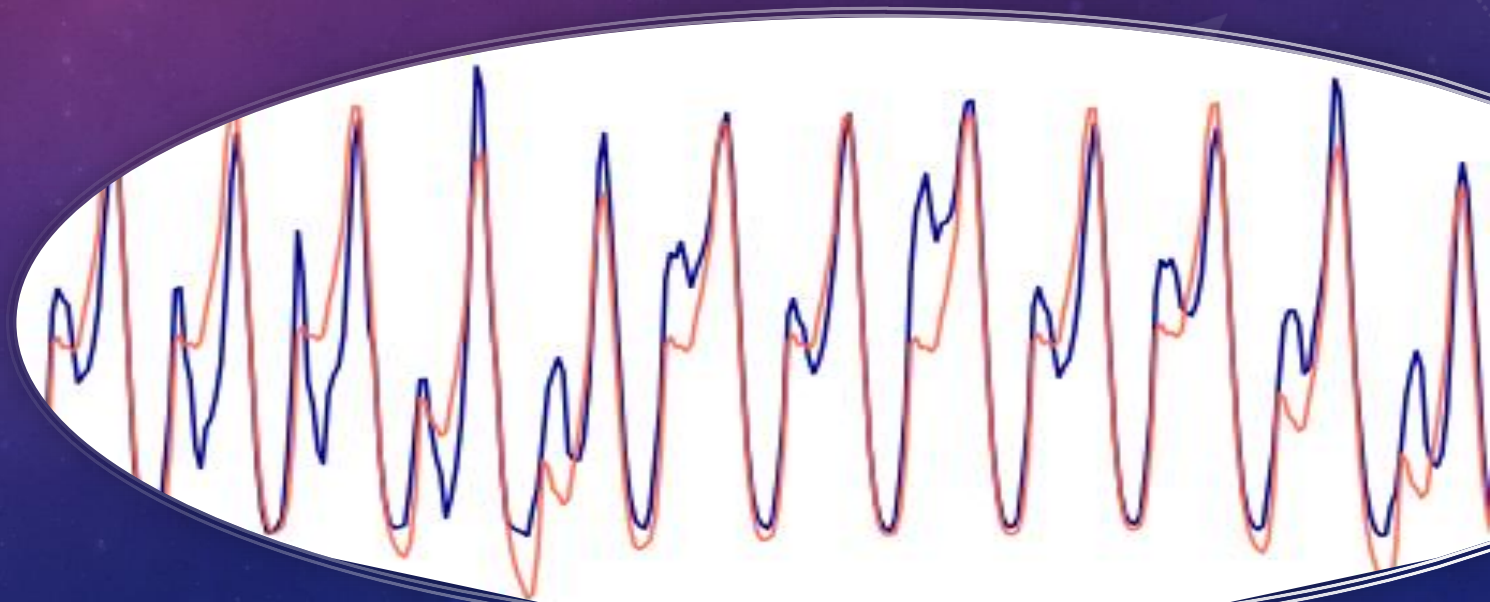


Miglior modello:

- SARIMAX (2,0,2)(1,1,1)₂₄, parametri determinati tramite approccio grid-search
- stagionalità settimanale: 10 sinusoidi
- stagionalità annuale: 5 sinusoidi
- no variabile COVID
- no log-trasformazione
- check sui residui ok

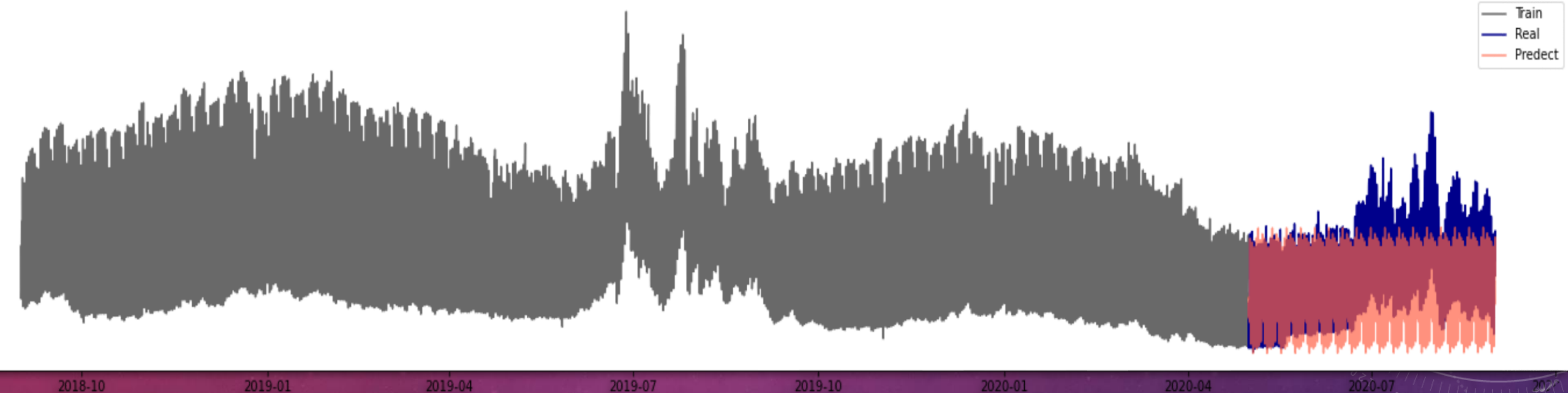
MAE train: 308301.8

MAE validation: 334117.9



The background is a solid dark blue color. It is decorated with several white geometric elements: concentric circles of varying sizes, some with dashed outer rings, and circular arcs. Many of these arcs are accompanied by degree markings (e.g., 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260) and small white arrows indicating a clockwise direction of rotation. The overall aesthetic is technical and modern.

UCM

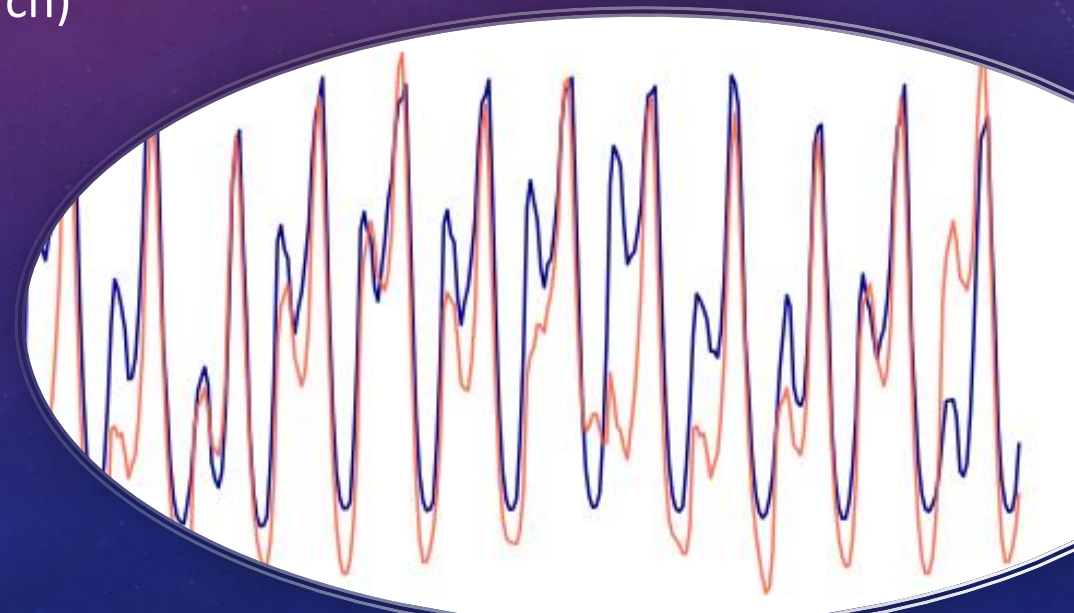


Miglior modello:

- trend: random walk (determinato tramite approccio grid-search)
- stagionalità giornaliera: variabili dummy
- stagionalità settimanale: 15 sinusoidi
- no variabile COVID

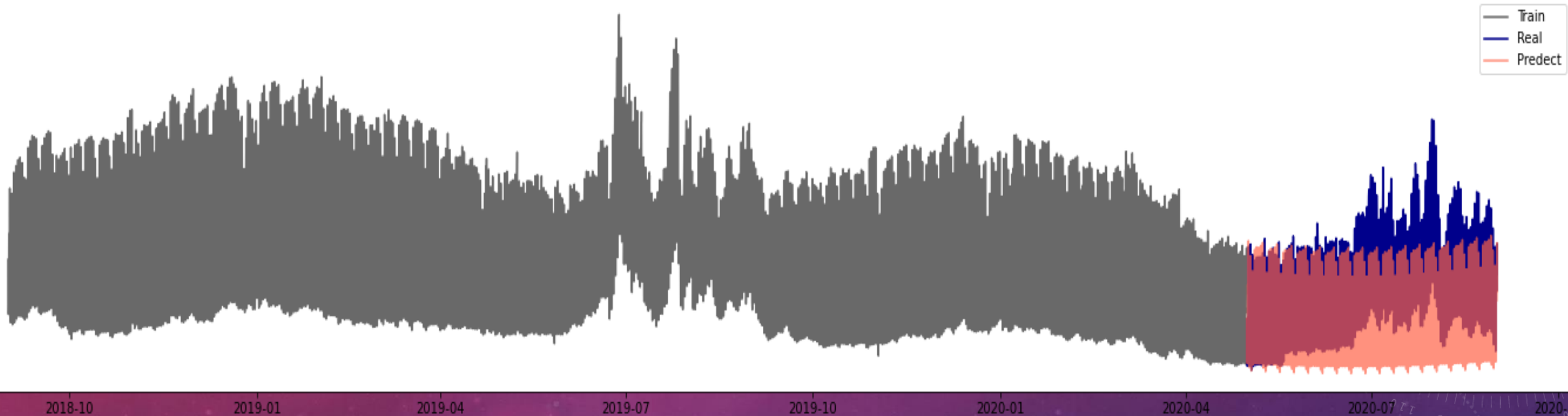
MAE train: 102623.3

MAE validation: 743458.6



The background is a solid dark blue color. It features several abstract geometric elements: concentric circles of varying sizes, some with dashed lines, and circular arcs with degree markings (e.g., 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260). Some of these arcs have small white arrows indicating a direction of movement or rotation. The overall design is technical and modern.

TBATS

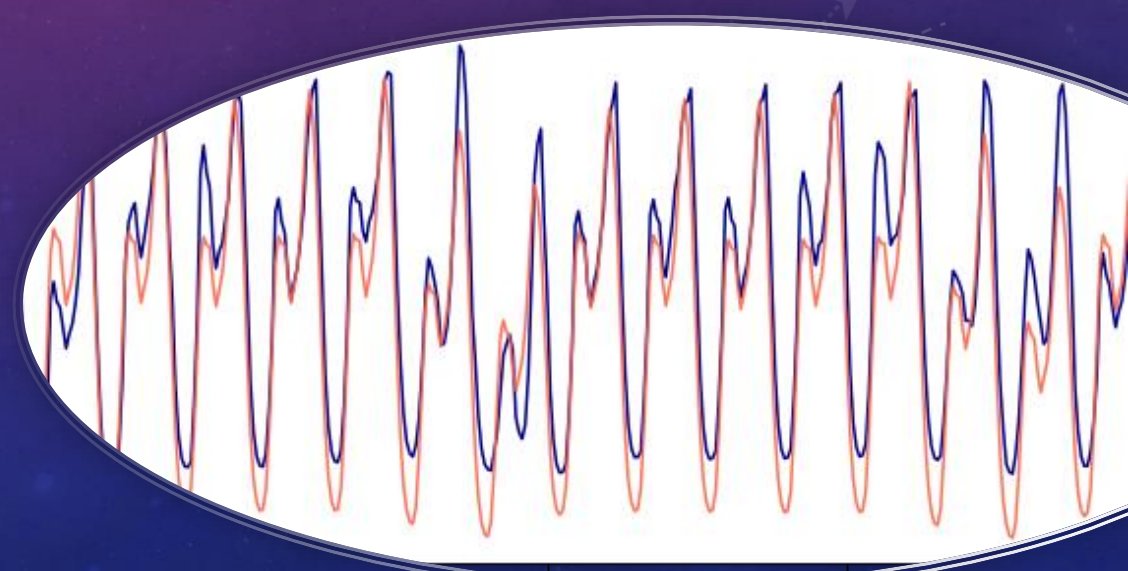


Miglior modello:

- trasformazione Box-Cox
- modellazione degli errori tramite processo ARMA
- stagionalità multipla con periodo 24, 168 e 8766 ore modellata tramite serie di Fourier.

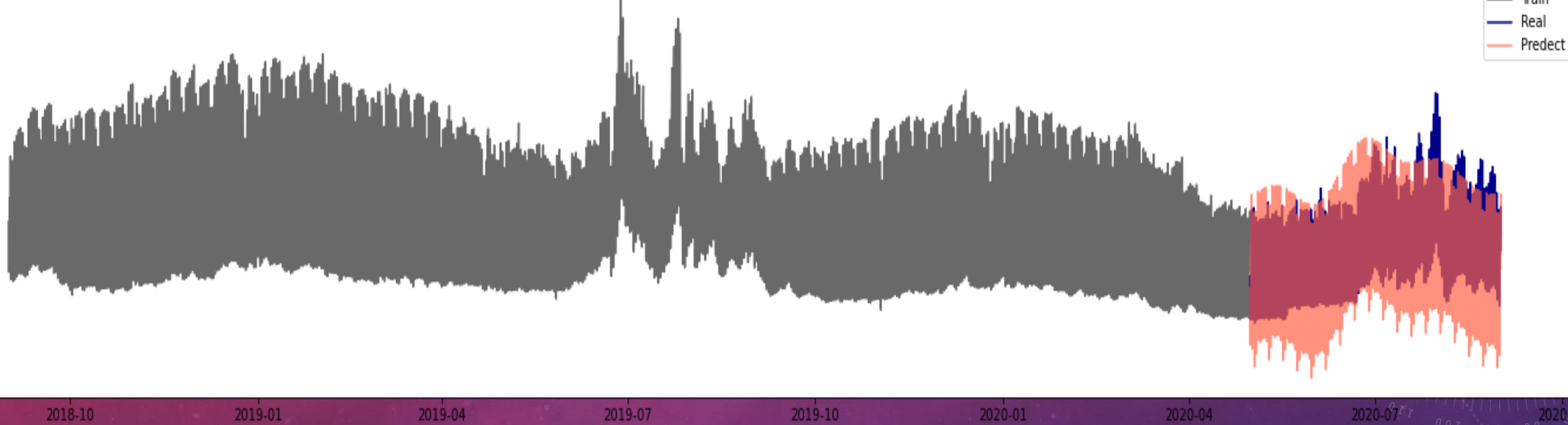
MAE train: 112372.4

MAE validation: 666606.0



The background is a solid dark blue color. It features several abstract geometric elements: concentric circles of varying sizes, some with solid lines and others with dashed lines. Small white arrows are placed at various points along these circles, indicating a clockwise direction of movement. Some of the larger circles have numerical labels (e.g., 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260) along their outer edges, suggesting a scale or measurement. The overall composition is dynamic and suggests a process of rotation or progression.

PROPHET

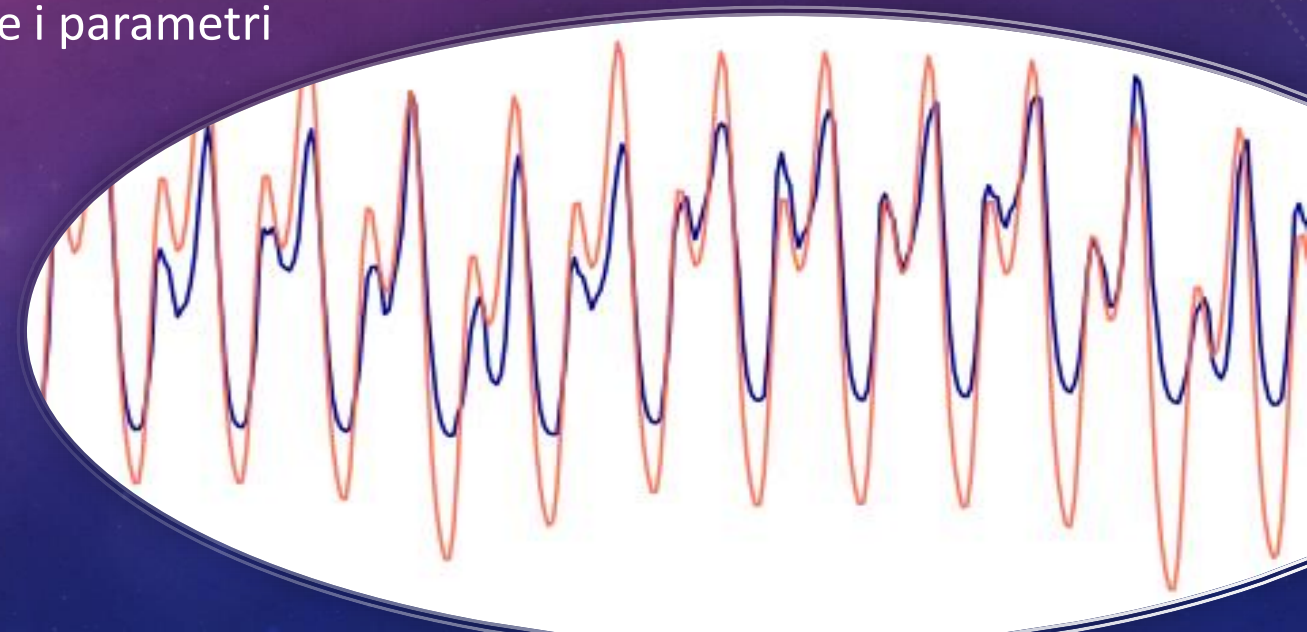


Miglior modello:

- cross-validazione sull'intero dataset per ottimizzare i parametri
- stagionalità giornaliera: 4 sinusoidi
- stagionalità settimanale: 3 sinusoidi
- stagionalità annuale: 10 sinusoidi

MAE train: 366467.4

MAE validation: 545685.0



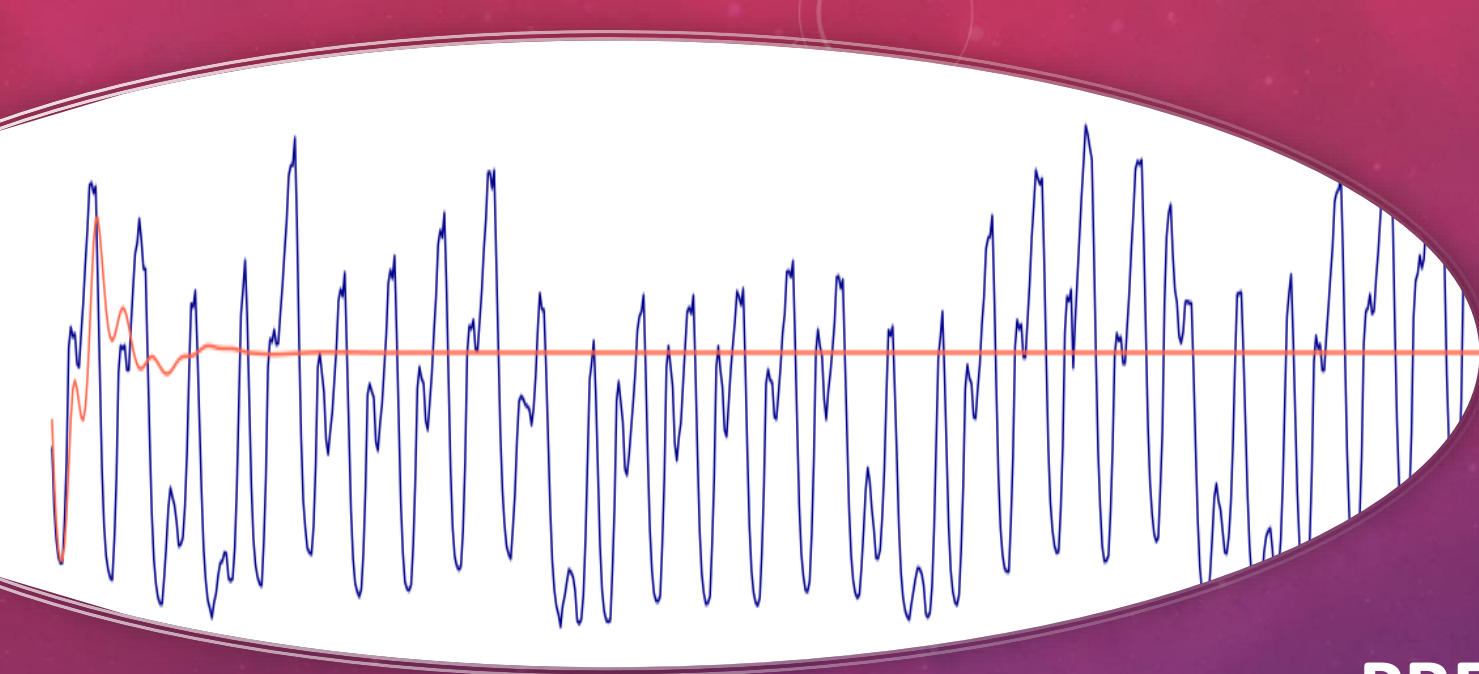
The background is a solid dark blue color. It features several abstract, light blue geometric patterns. These include concentric circles of varying sizes, some with dashed lines. Arcs and segments of circles are also present, some with small arrowheads indicating a direction. Degree markings are visible on some of the larger arcs, ranging from 140 to 260. The overall design is technical and modern, resembling a stylized compass or a diagram from a technical manual.

LSTM

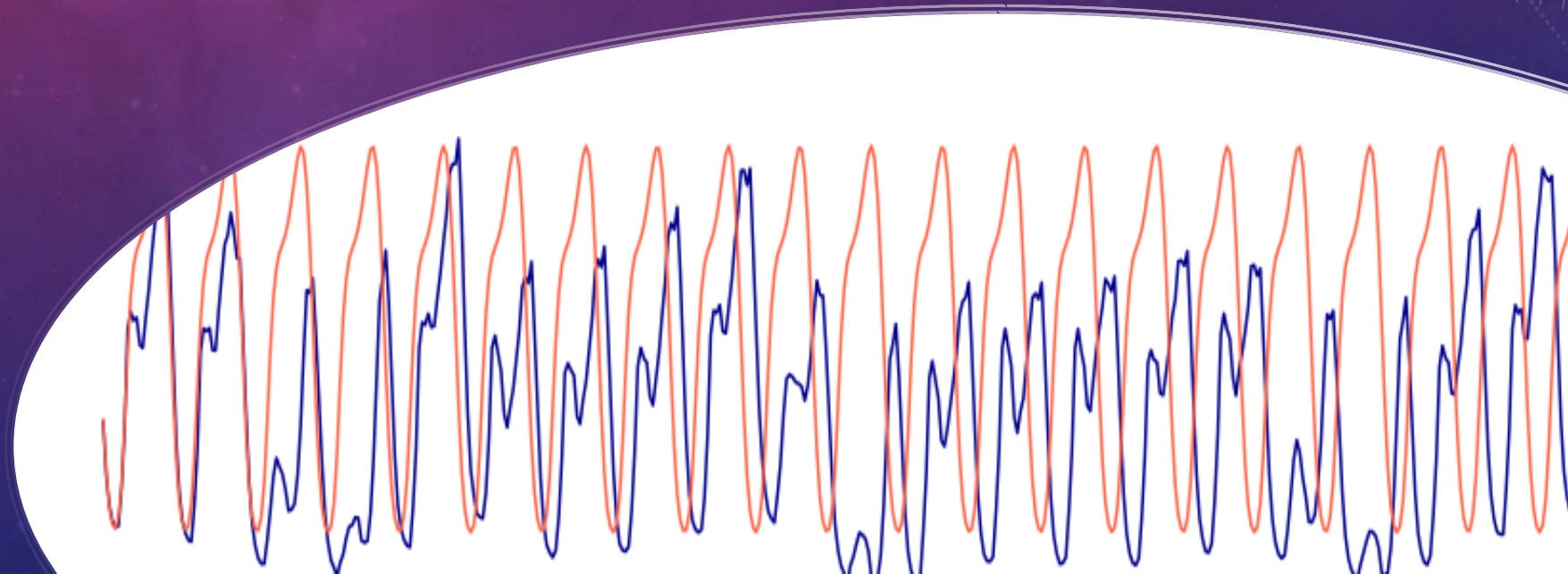
- validation set: ultimi 61 giorni, lunghezza pari alla finestra di forecast
- due diverse implementazioni: stateful e stateless
- tre diverse strategie per le previsioni multi-step:
 - previsioni one-step-ahead
 - previsioni direct output vector
 - previsioni ibride

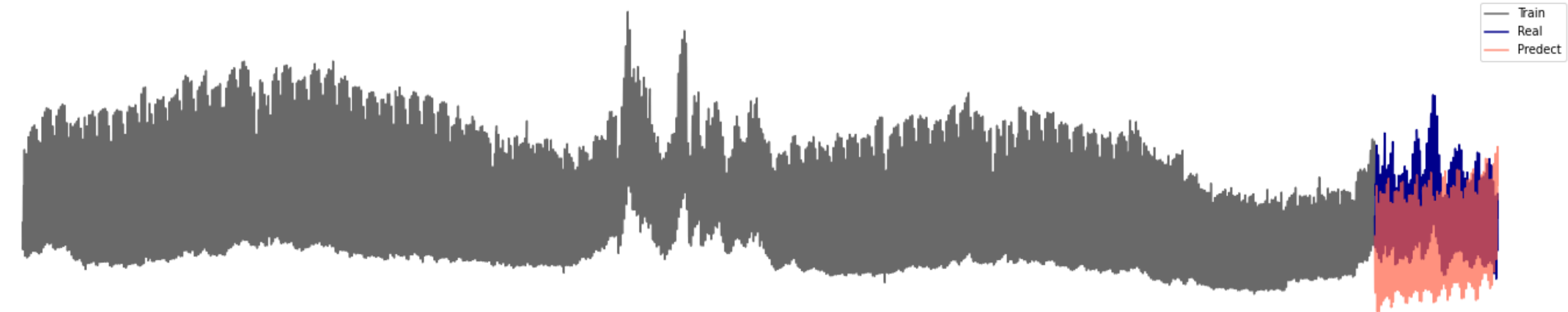
Layer (type)	Output Shape	Param #
lstm_1 (LSTM)	(None, 100)	40800
leaky_re_lu_1 (LeakyReLU)	(None, 100)	0
dropout_1 (Dropout)	(None, 100)	0
dense_1 (Dense)	(None, 744)	75144
Total params: 115,944		
Trainable params: 115,944		
Non-trainable params: 0		

I dati sono stati scalati nell'intervallo tra 0 a 1

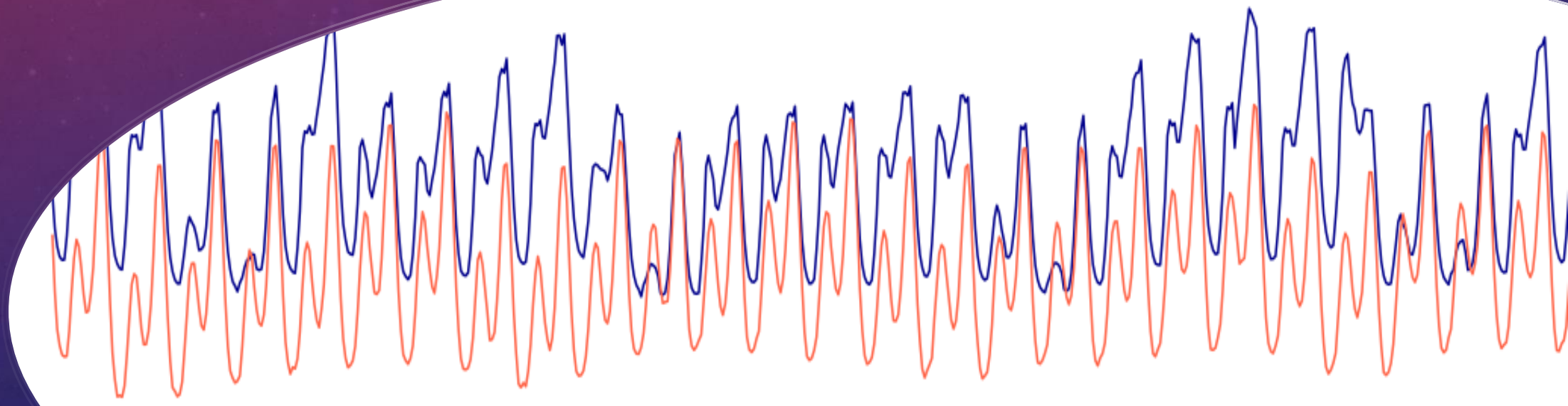


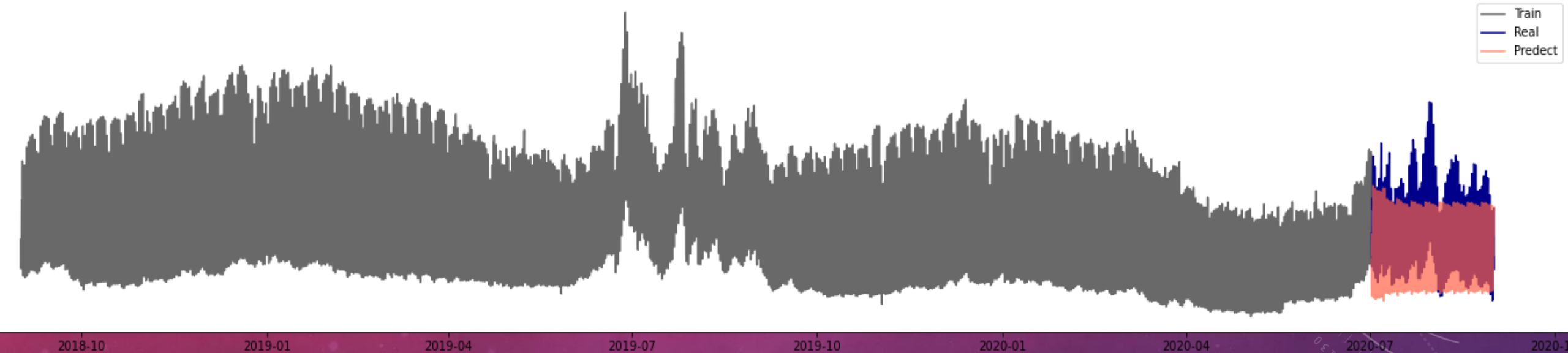
PREVISIONI ONE STEP AHEAD





PREVISIONI DIRECT OUTPUT VECTOR





Miglior modello:

- Number steps in = 24×31
- Number steps out = 24×31
- Modo stateless
- Molto instabile

MAE train: 1077271.0

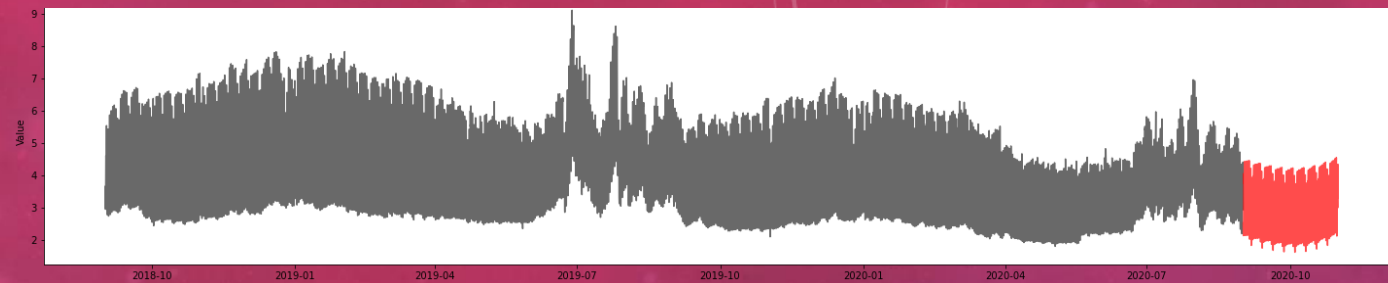
MAE validation 634172.1

PREVISIONI IBRIDE

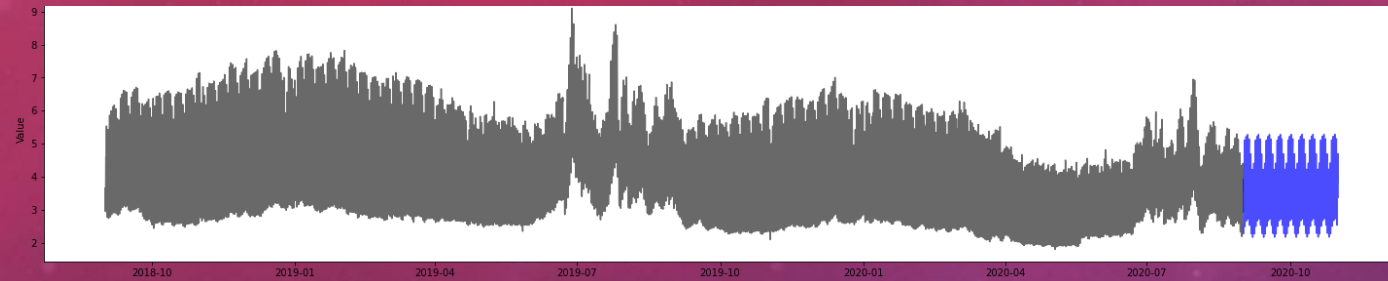
I modelli provati non sono eccessivamente performanti.

- si potrebbe provare a lavorare con la serie resa stazionaria
- si potrebbero applicare degli aggiustamenti stagionali prima della modellazione

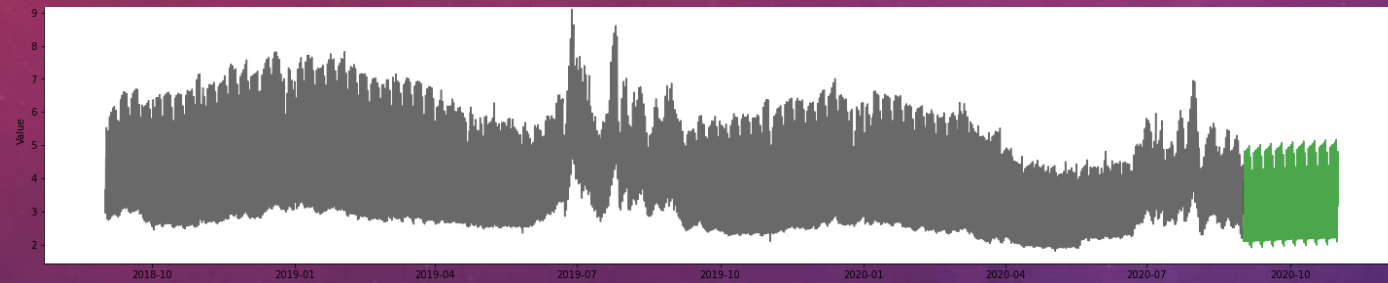
PREVISIONI



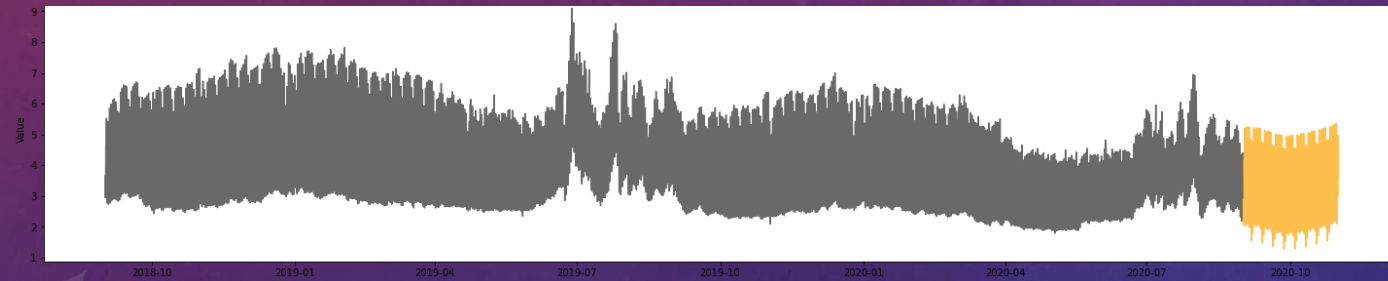
SARIMAX



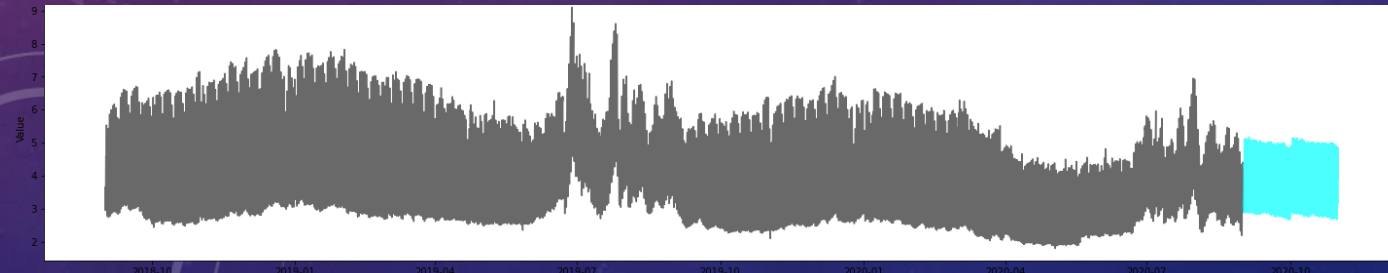
UCM



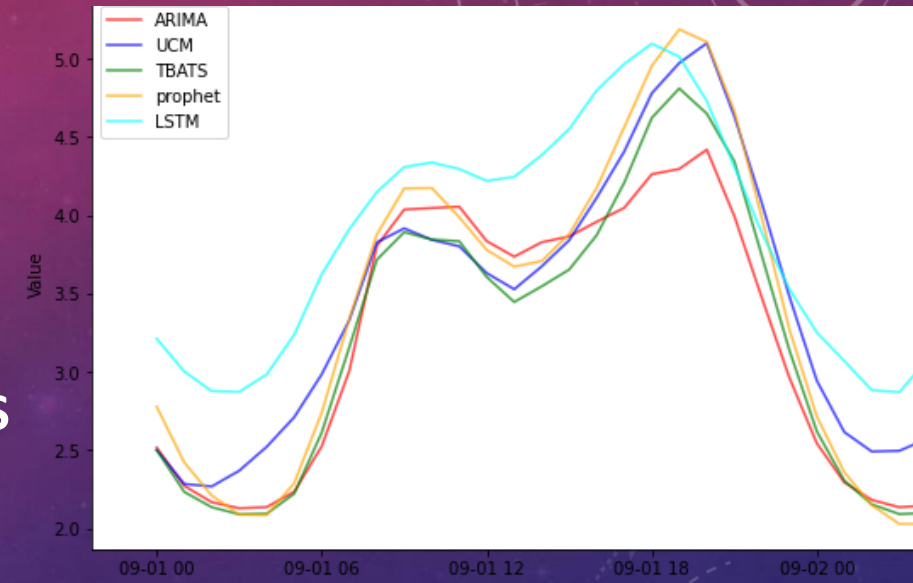
TBATS



PROPHET

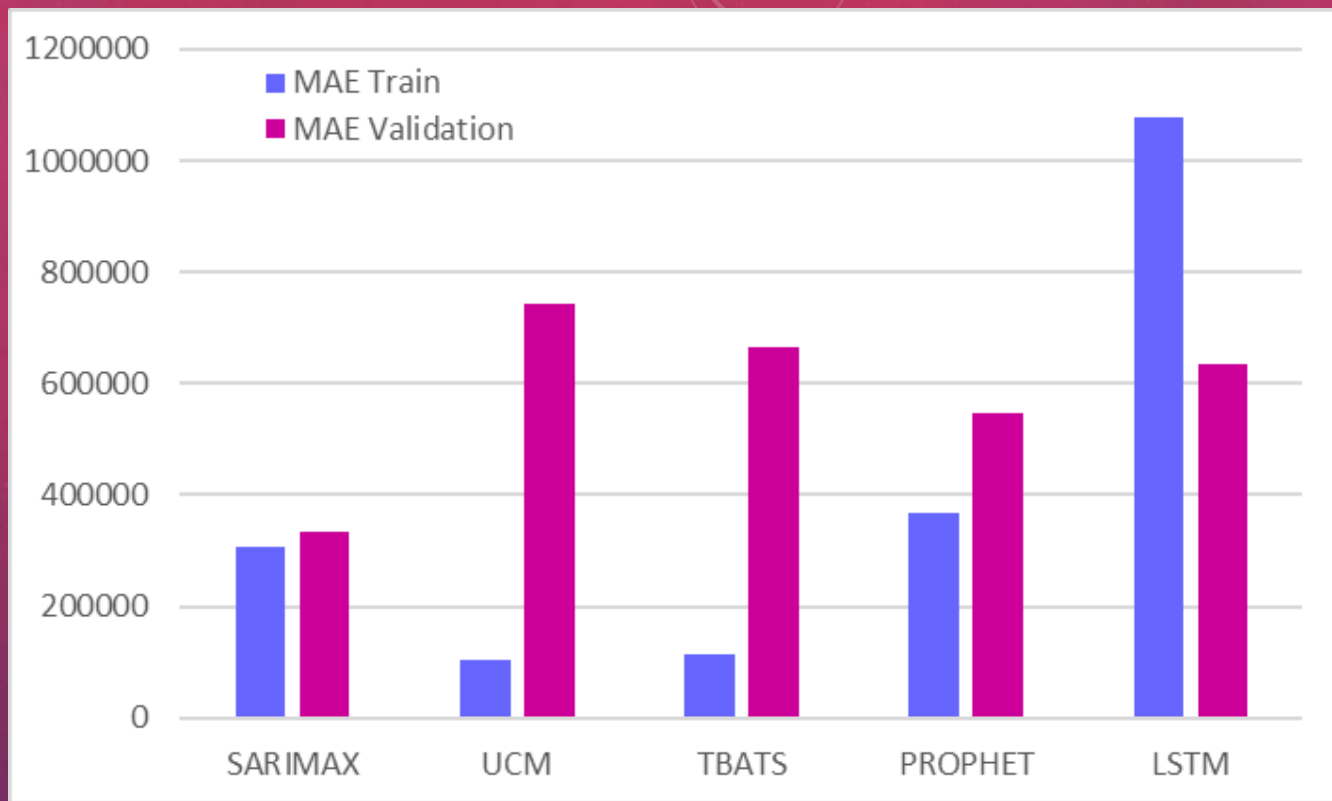


LSTM



CONCLUSIONI E SVILUPPI FUTURI





Il modello SARIMAX risulta il migliore

Per ottenere risultati migliori dovremmo sapere cosa rappresentano i dati.

Sarebbe utile anche conoscere l'andamento della serie negli anni precedenti al 2018 per modellare le anomalie dell'estate 2019 e il fattore covid in modo più performante.

Come lavoro futuro, i modelli potrebbero essere implementati in un ambiente più potente e con più RAM a disposizione. In tal modo si potrebbero implementare dei metodi per ottimizzare gli iperparametri che altrimenti risultano troppo lenti.

The background is a gradient of deep purple and blue, filled with out-of-focus light spots (bokeh). Overlaid on this are several faint, white geometric patterns: concentric circles, arcs, and a large circular scale with numerical markings from 140 to 260. The scale is positioned on the left side, curving around the top and bottom. The word 'GRAZIE!' is written in a large, white, sans-serif font on the right side.

GRAZIE!

Cavenati Laura - matricola 864000
l.cavenati1@campus.unimib.it

BIBLIOGRAFIA

- <http://proceedings.mlr.press/v37/jozefowicz15.pdf>
- <https://facebook.github.io/prophet/>
- [https://www.researchgate.net/publication/341609757 Comparative Analysis of Recurrent Neural Network Architectures for Reservoir Inflow Forecasting](https://www.researchgate.net/publication/341609757_Comparative_Analysis_of_Recurrent_Neural_Network_Architectures_for_Reservoir_Inflow_Forecasting)
- Materiale del corso di Streaming Data Management and Time Series Analysis
- Deep Learning for Time Series Forecasting - Predict the Future with MLPs, CNNs and LSTMs in Python, Jason Brownlee