

**UNIVERSITÀ DEGLI STUDI DI
MILANO BICOCCA**

PROGETTO DI DATA SCIENCES LAB

*Previsione del prezzo dell'energia
nel mercato elettrico italiano*

BORSOTTI MATTIA

PIANCONE ANDREA

PICHINI MARCELLO

PRESENTAZIONE DEL PROBLEMA

- **DATI RELATIVI AL PREZZO DELL'ELETTRICITÀ NEL MERCATO ITALIANO.**
- **DATI DISPONIBILI: DATA, INTERVALLO ORARIO, PREZZO.**
- **PERIODO DI RIFERIMENTO: TRIENNIO 2017–2018–2019**
- **OBIETTIVO: COSTRUIRE MODELLI PREDITTIVI DEL PREZZO DELL'ELETTRICITÀ NEL MERCATO ITALIANO.**

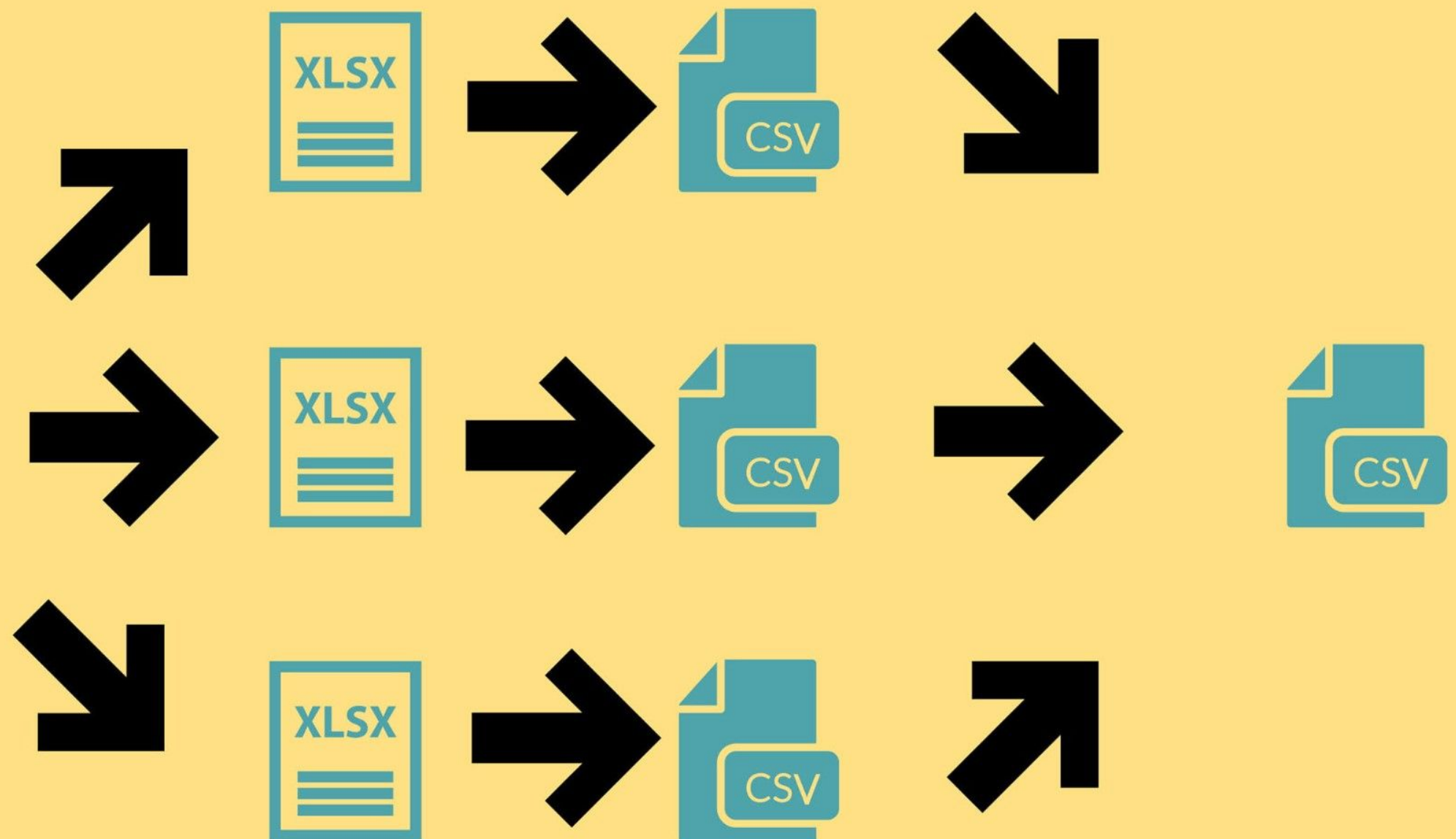


Preprocessing

PREPROCESSING

CONVERSIONE

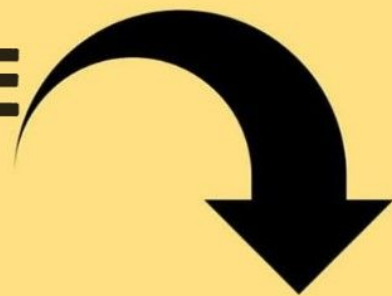
ACQUISIZIONE
DATI



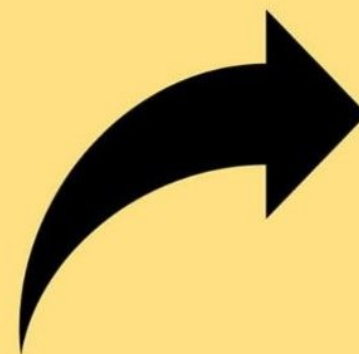
PREPROCESSING

DATA CLEANING

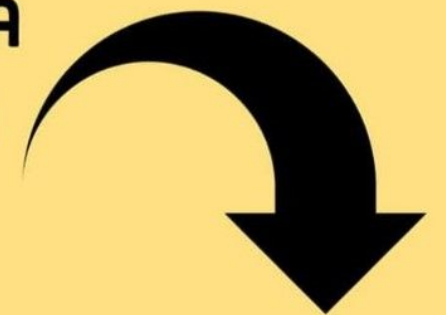
**RIMOZIONE DELLE
FEATURES
SUPERFLUE**



**A CAUSA DEL CAMBIO
ORA LEGALE-ORA
SOLARE, ALCUNI GIORNI
HANNO 23 O 25
INTERVALLI**



- **ELIMINATA LA
25-ESIMA ORA;**
- **REPLICATA LA 23-ESIMA
ORA PER OTTENERE LA
24-ESIMA.**



**LA DATA E L'INTERVALLO
ORARIO SONO
AGGREGATE
PER COSTRUIRE
UN UNICO ATTRIBUTO
DI TIPO DATETIME.**

PREPROCESSING

FEATURES ENGINEERING

PER GESTIRE LA STAGIONALITÀ
PRESENTE NELLA SERIE STORICA SI
INTRODUCONO...

**REGRESSORI
SINUSOIDALI**

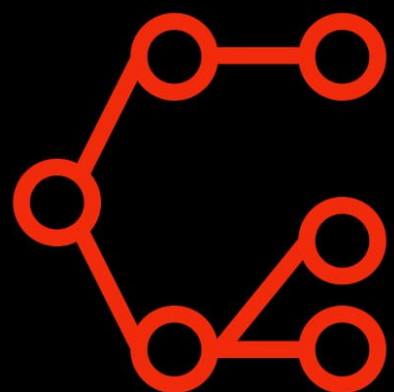
DUMMY
GIORNO DELLA
SETTIMANA

DUMMY
GIORNO
FESTIVO

PREPROCESSING

Risultato finale...

| | Data | PUN | Giorno_Vacanza | Friday | Monday | Saturday | Sunday | Thursday | Tuesday | Wednesday | ... | sinu365.23 |
|---|---------------------|-------|----------------|--------|--------|----------|--------|----------|---------|-----------|-----|------------|
| 0 | 2017-01-01 00:00:00 | 53.30 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | ... | 0.248940 |
| 1 | 2017-01-01 01:00:00 | 52.00 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | ... | -0.482206 |
| 2 | 2017-01-01 02:00:00 | 51.00 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | ... | 0.685111 |
| 3 | 2017-01-01 03:00:00 | 47.27 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | ... | -0.844881 |
| 4 | 2017-01-01 04:00:00 | 45.49 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | ... | 0.951454 |



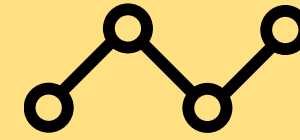
Modelli

ARIMA



PARAMETRI

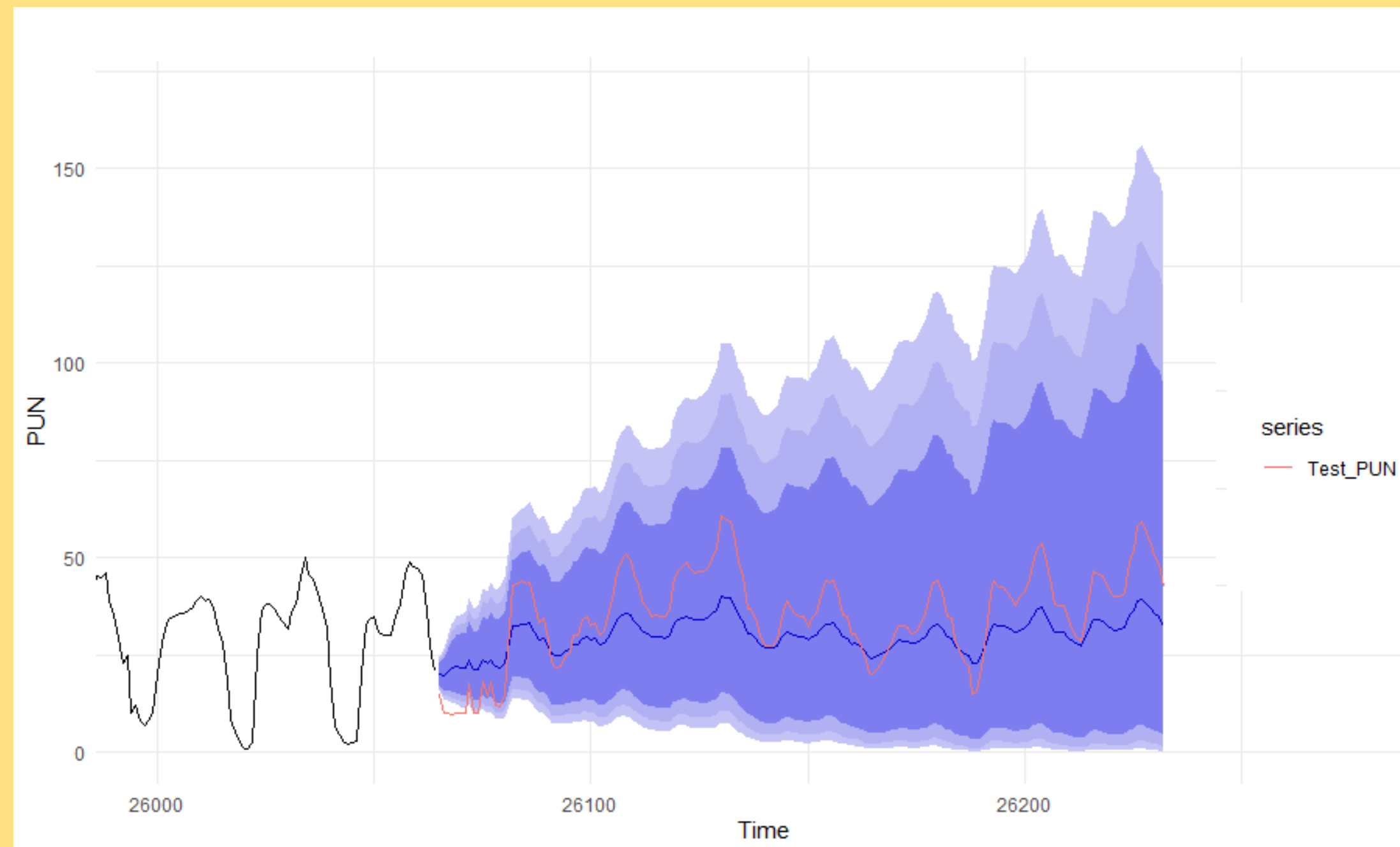
AR[3]I[1]MA[4] SAR[1,1,1]



PERFORMANCE

ALLA VARIABILE PUN È
STATA APPLICATA LA
TRASFORMAZIONE DI BOX
COX CON LAMBDA = 0.688

TRAIN MAE: 5.698
TRAIN MAPE: 10.86%
TEST MAE: 2.727
TEST MAPE: 5.737%

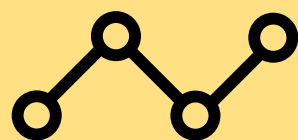


LSTM (LONG SHORT-TERM MEMORY)



PARAMETRI

4 LSTM LAYER CON DROPOUT PER LA REGOLARIZZAZIONE.
BATCH SIZE: 1024
EPOCHE: 100
TEMPO DI APPRENDIMENTO: 83 MINUTI

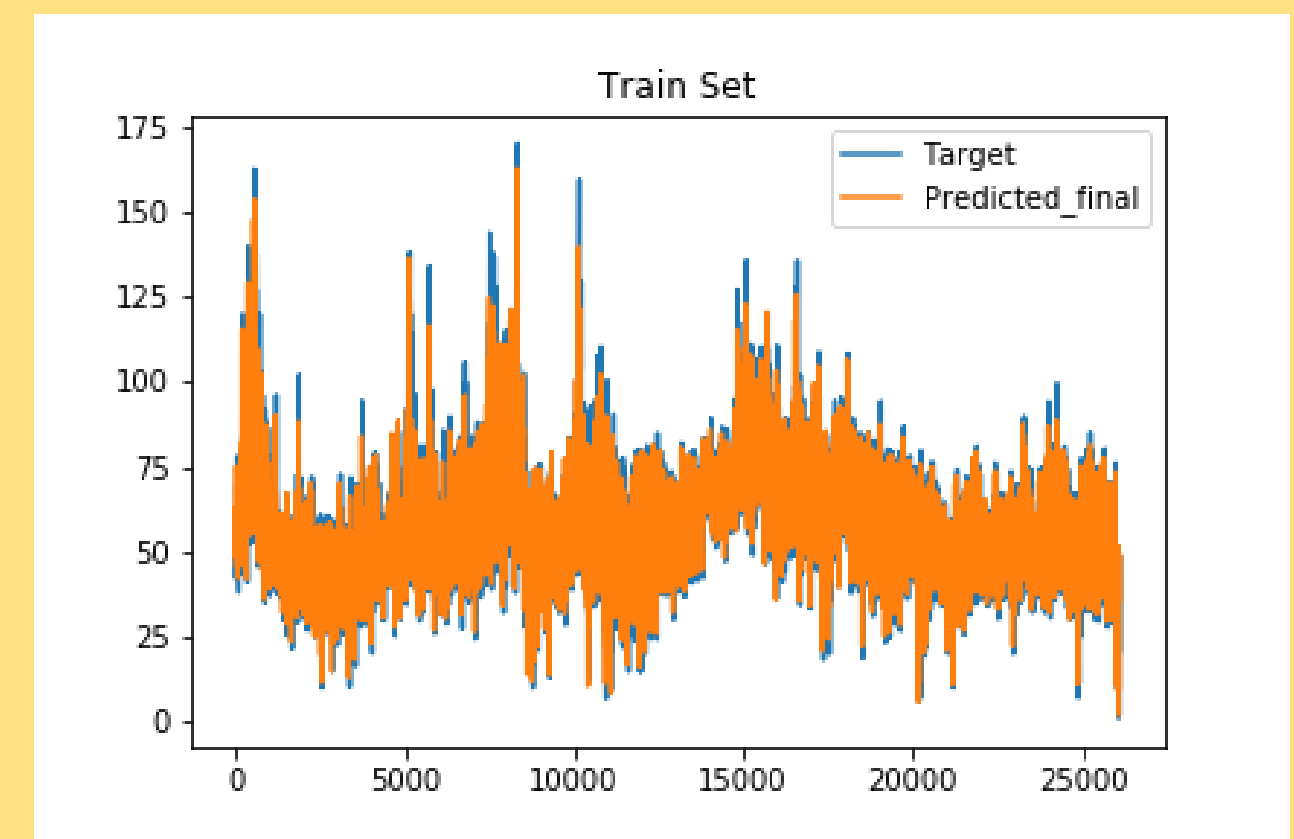
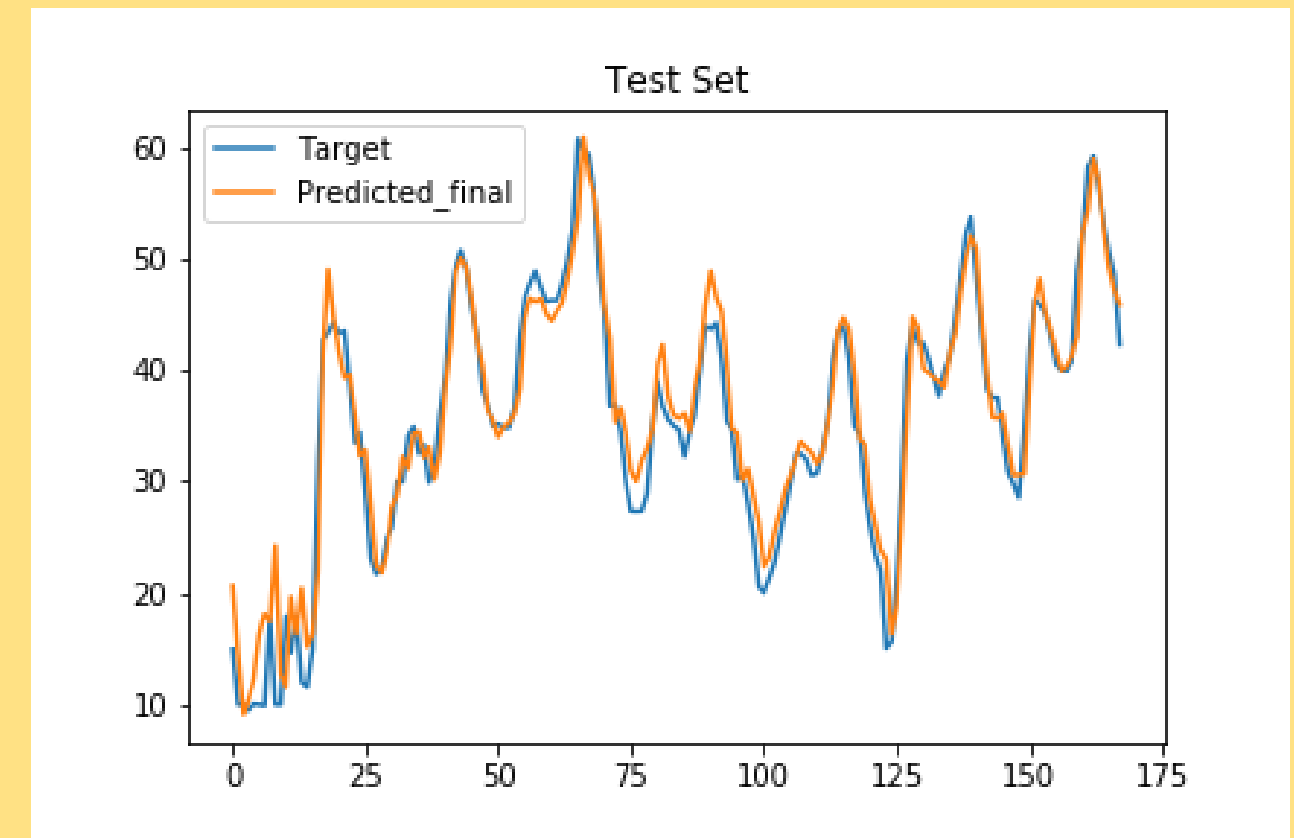


PERFORMANCE

TRAIN MAE: 6.005
TRAIN MAPE: 3.432%
TEST MAE: 2.309
TEST MAPE: 7.014%



Non c'è underfitting a causa dei diversi picchi presenti



ELASTIC NET

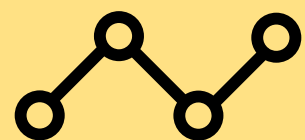


PARAMETRI

ALPHA = 0.6

L1_RATIO = 0.4

**TEMPO DI APPRENDIMENTO:
30 MINUTI**



PERFORMANCE

TRAIN MAE: 6.248

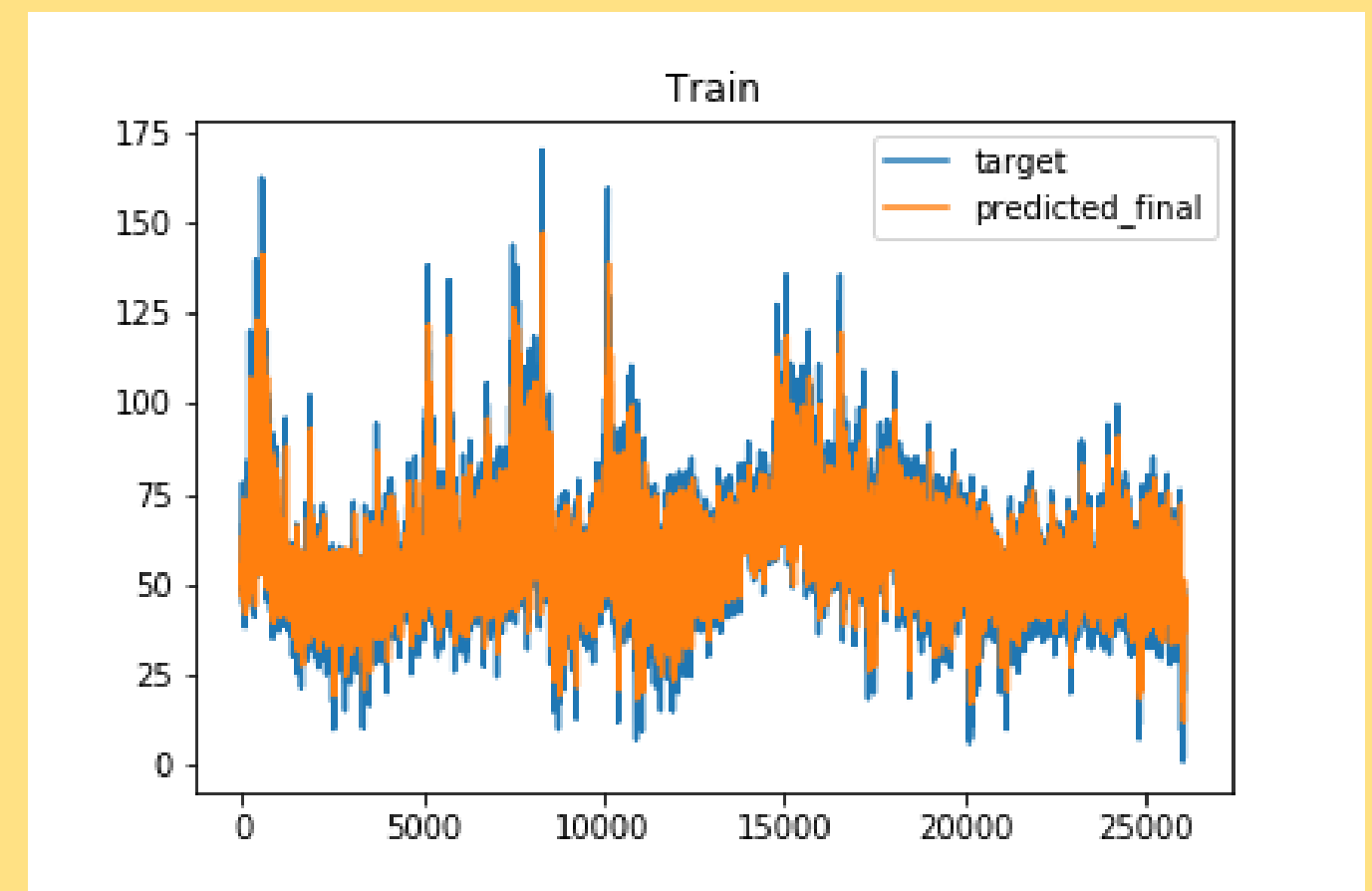
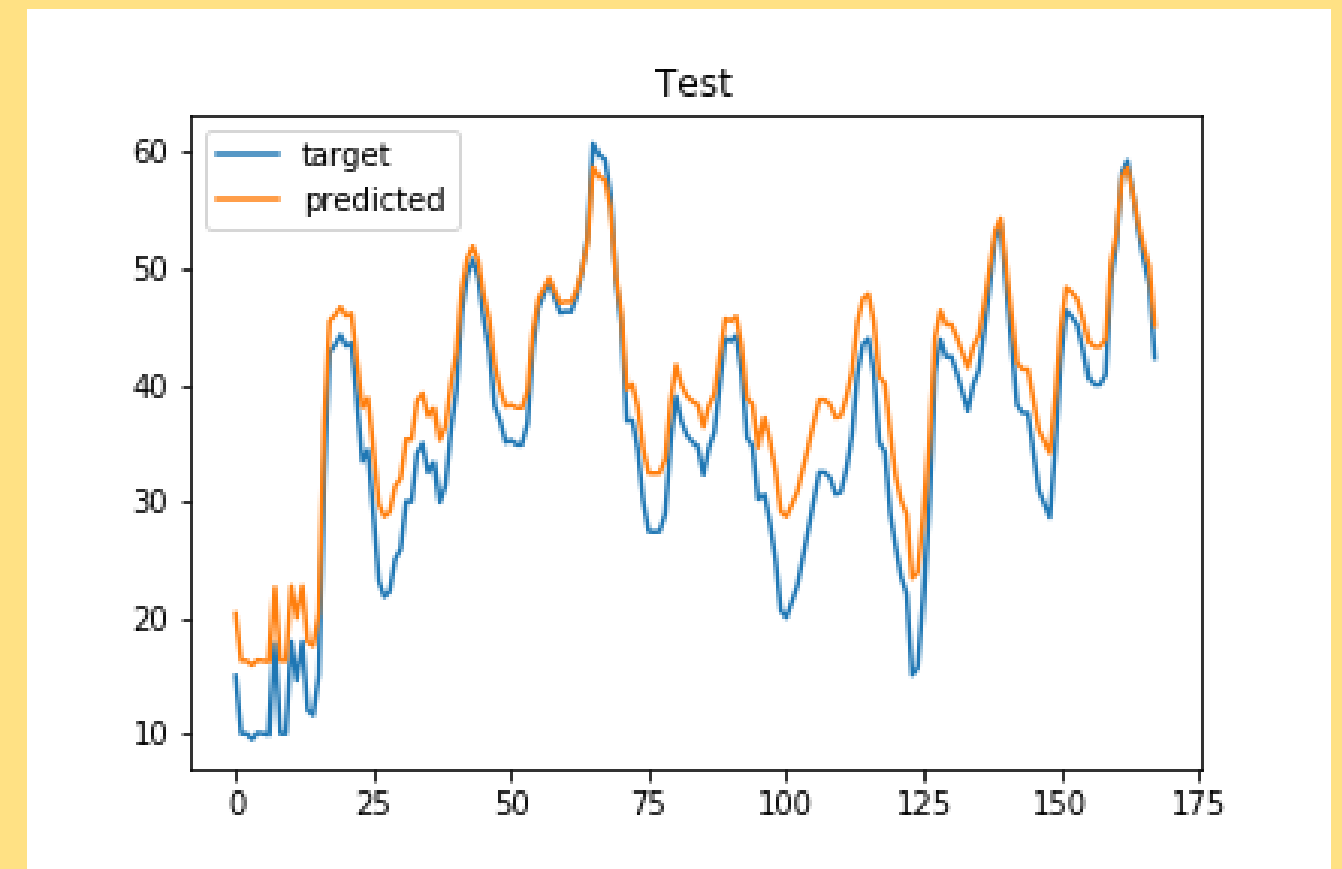
TRAIN MAPE: 12.323%

TEST MAE: 3.883

TEST MAPE: 15.563%



Non c'è underfitting a causa dei
diversi picchi presenti

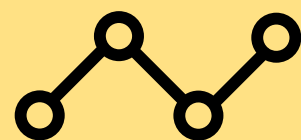


SUPPORT VECTOR REGRESSION



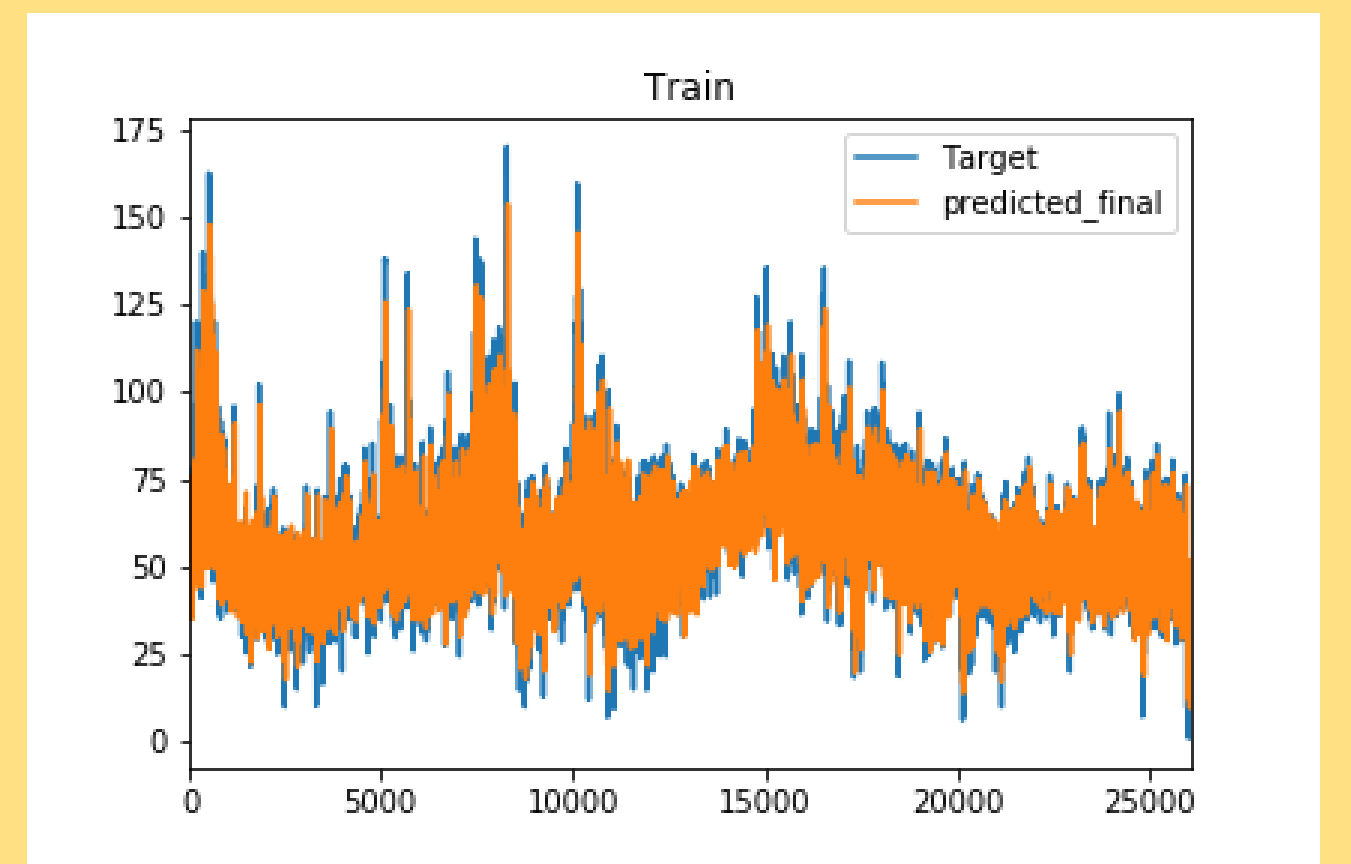
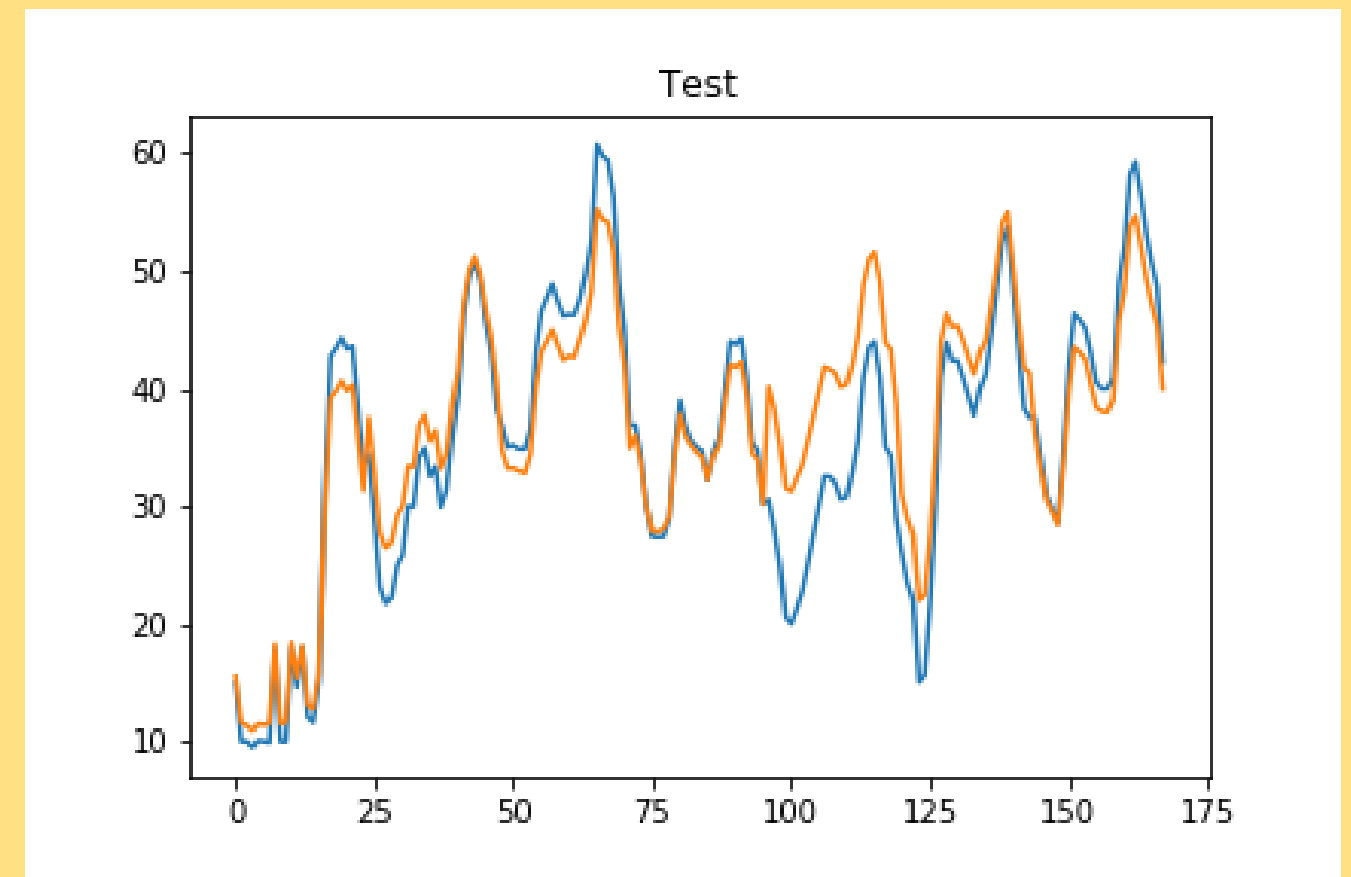
PARAMETRI

KERNEL: LINEARE
C = 5
GAMMA = 1E⁻⁷
EPSILON = 1.5



PERFORMANCE

TRAIN MAE: 5.840
TRAIN MAPE: 11.39%
TEST MAE: 3.473
TEST MAPE: 18.14%

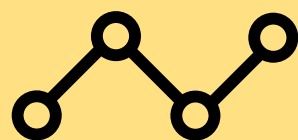


MULTILAYER PERCEPTRON



PARAMETRI

**3 LAYER CON DROPOUT PER LA
REGOLARIZZAZIONE.
BATCH SIZE: 2048
EPOCHE: 100
TEMPO DI APPRENDIMENTO: 30
MINUTI**

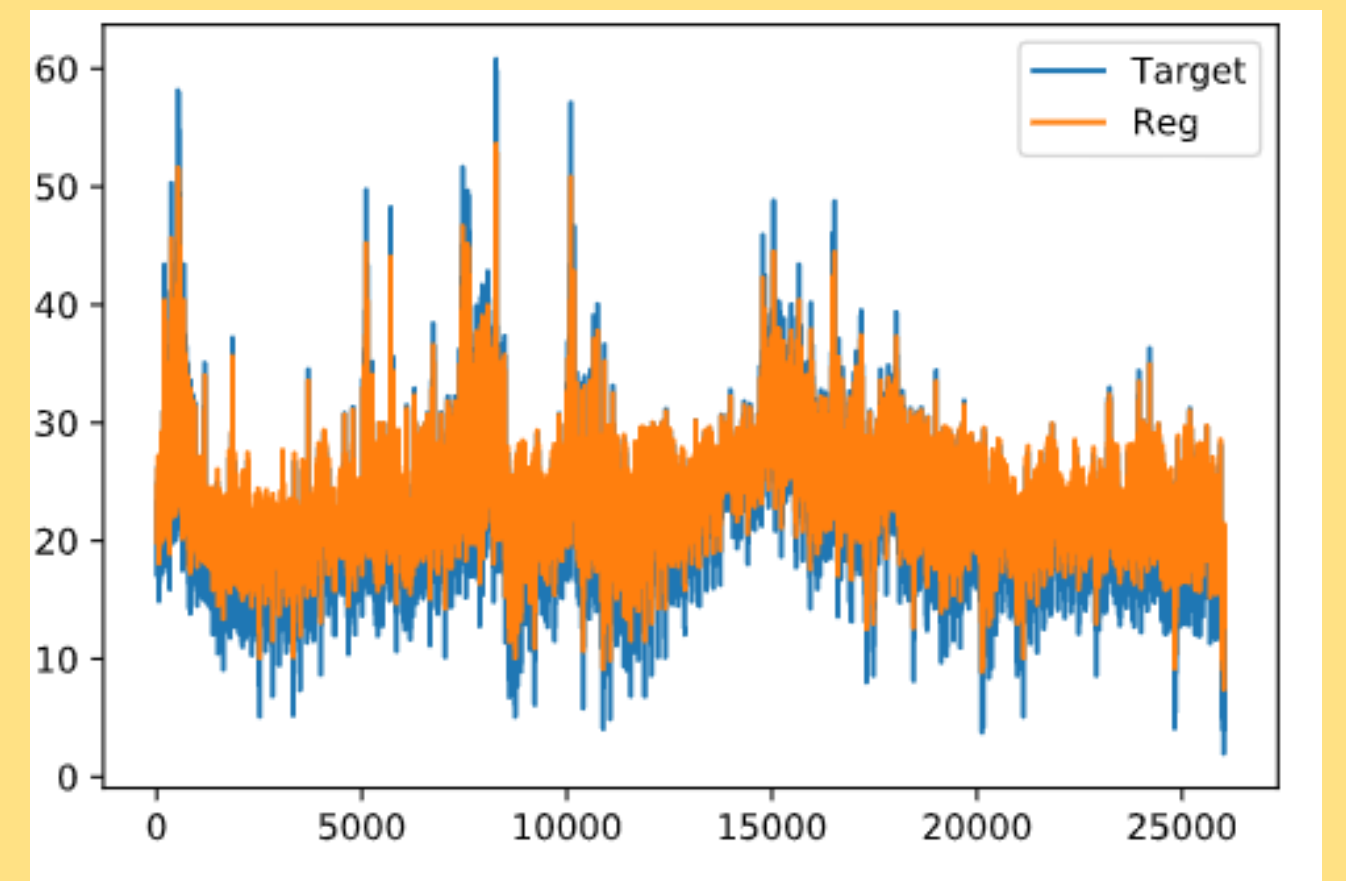
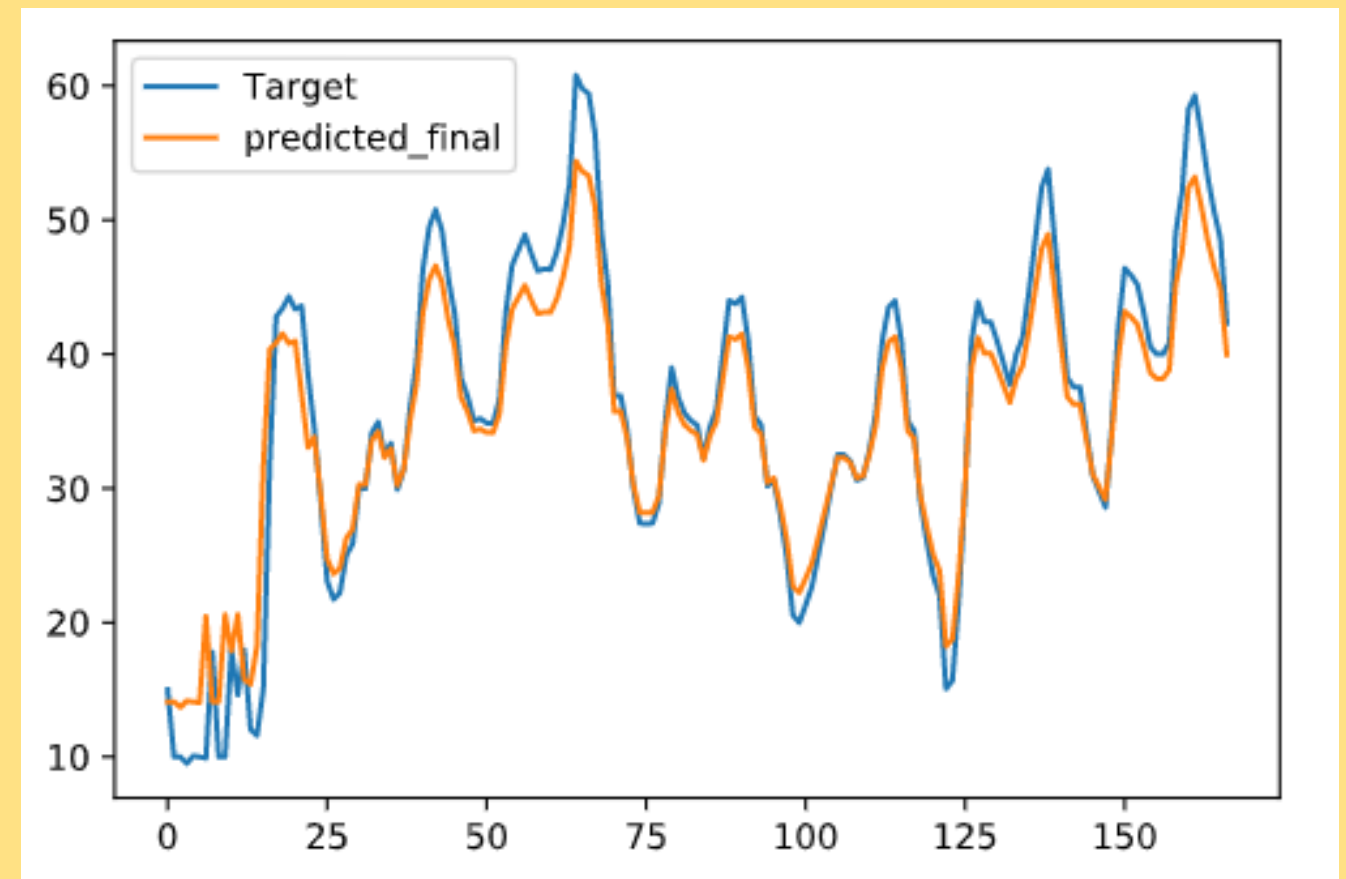


PERFORMANCE

**TRAIN MAE: 2.00
TRAIN MAPE: 9.30%
TEST MAE: 2.40
TEST MAPE: 9.30%**



Non c'è underfitting a causa dei
diversi picchi presenti

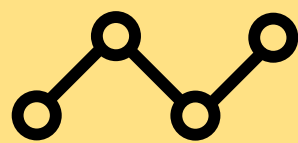


SIMPLE RECCURENT NEURAL NETWORK



PARAMETRI

**4 LAYER CON DROPOUT PER
LA REGOLARIZZAZIONE.
BATCH SIZE: 1024
EPOCHE: 100
TEMPO DI APPRENDIMENTO:
20 MINUTI**

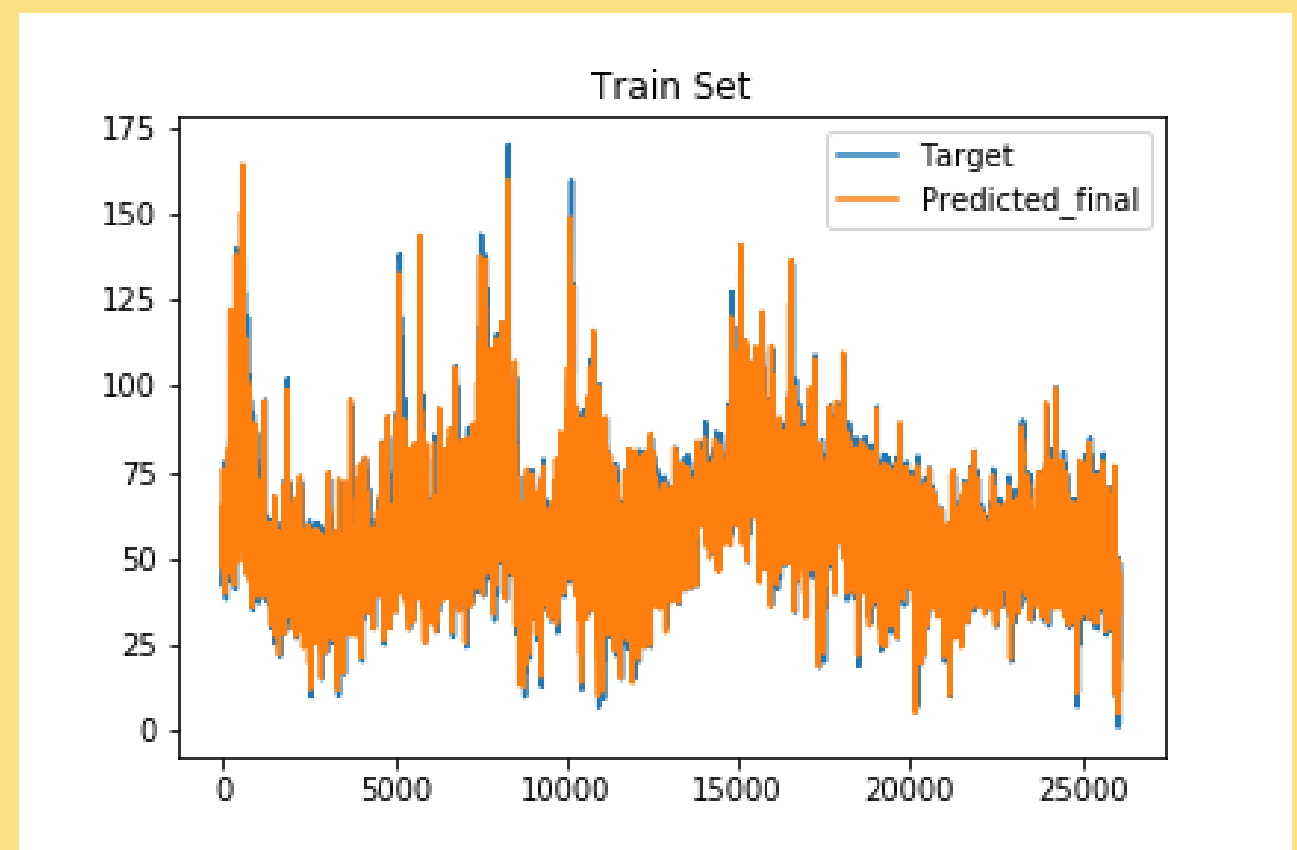


PERFORMANCE

**TRAIN MAE: 5.822
TRAIN MAPE: 3.358%
TEST MAE: 2.580
TEST MAPE: 7.025%**



Non c'è underfitting a causa dei
diversi picchi presenti

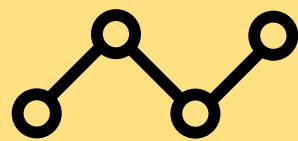


GATED RECURRENT UNIT (GRU)



PARAMETRI

**4 LAYER GRU CON DROPOUT
PER LA REGOLARIZZAZIONE.
BATCH SIZE: 2048
EPOCHE: 50
TEMPO DI APPRENDIMENTO: 40
MINUTI**

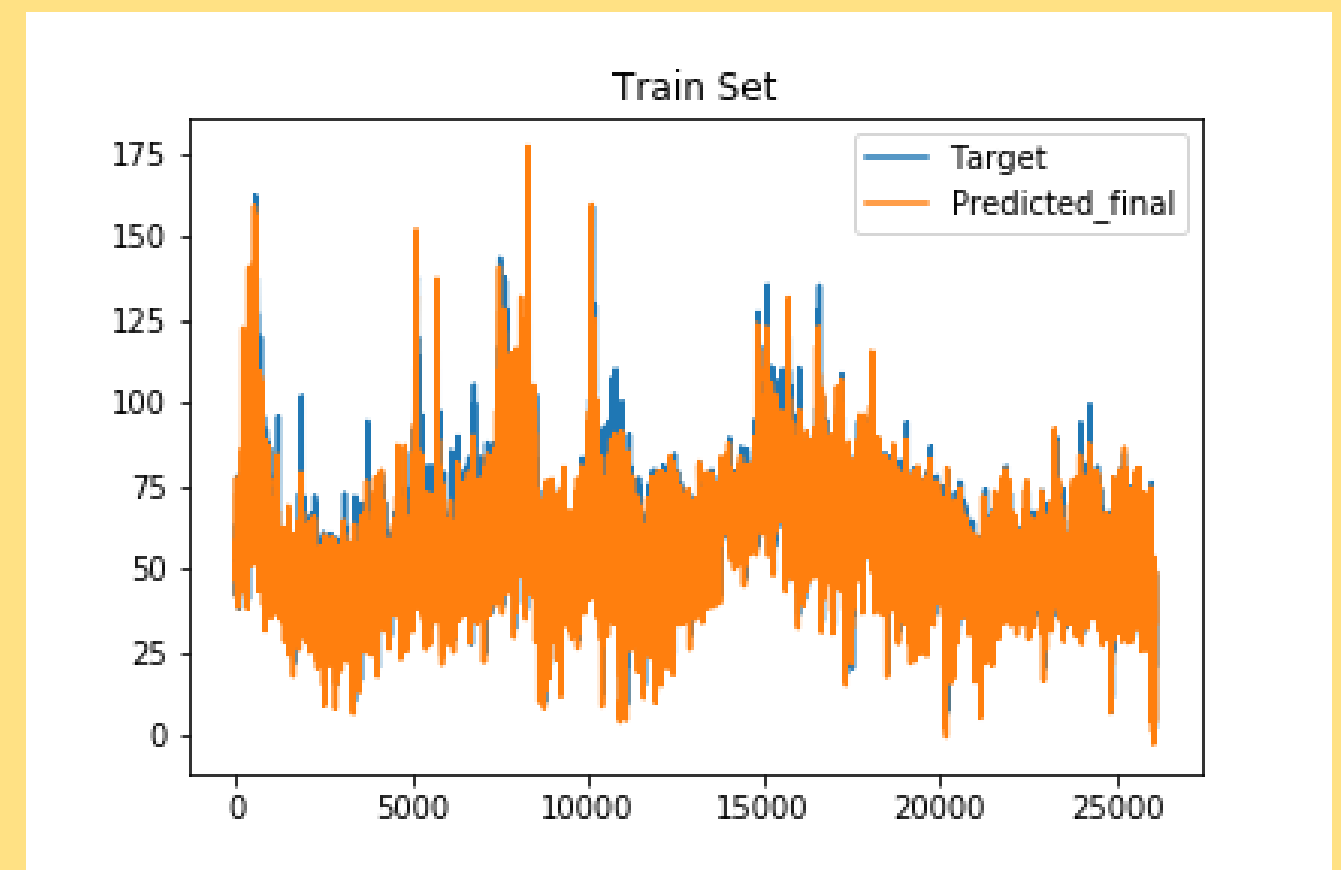
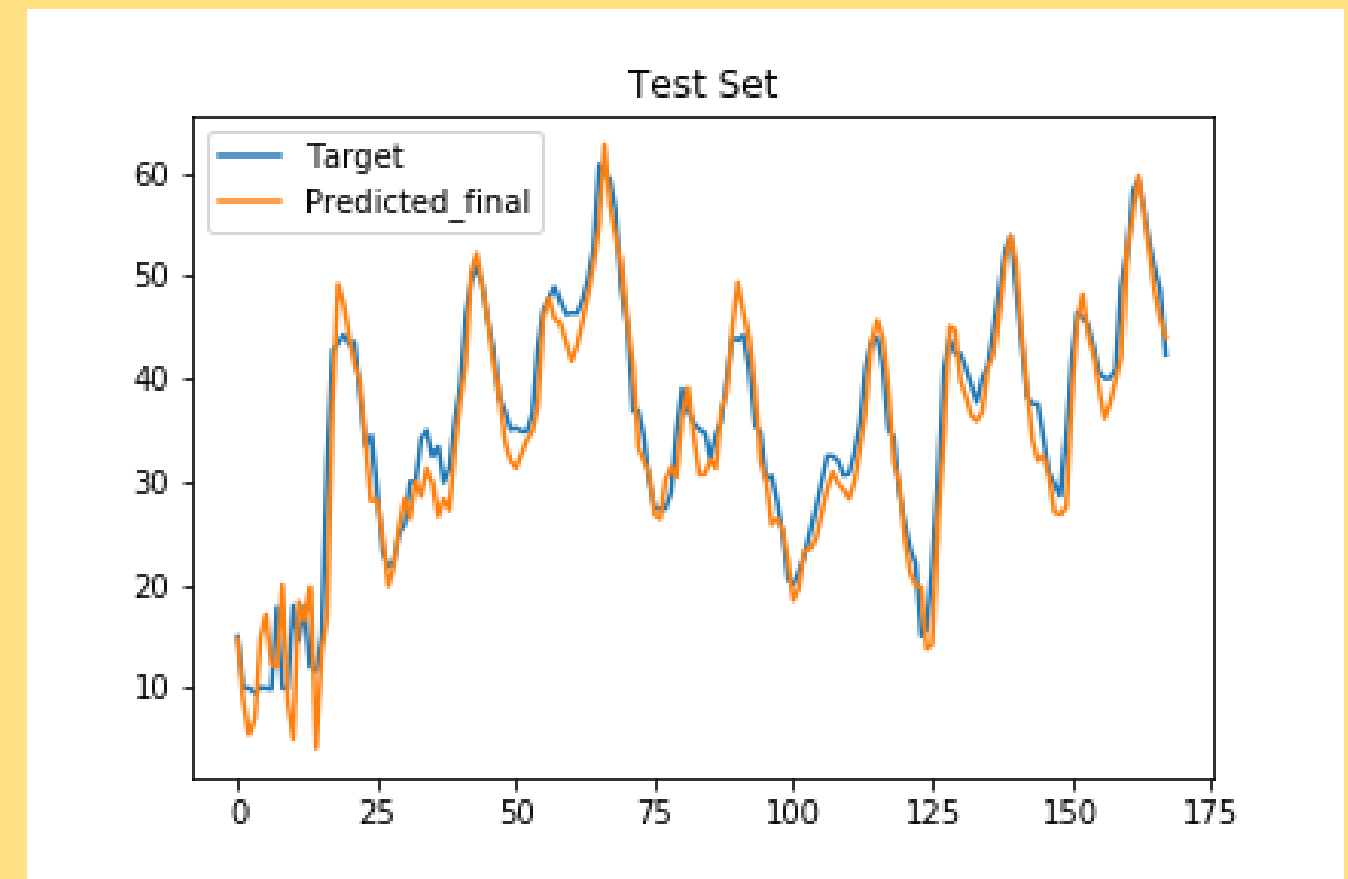


PERFORMANCE

**TRAIN MAE: 6.635
TRAIN MAPE: 3.528%
TEST MAE: 2.702
TEST MAPE: 7.064%**



Non c'è underfitting a causa dei
diversi picchi presenti





CONCLUSIONI

- SI OSSERVA COME, IN GENERALE I MODELLI CONSIGLIATI DALLA LETTERATURA PER IL TRATTAMENTO DI SERIE STORICHE (ARIMA E RETI NEURALI) OTTENGANO RISULTATI MIGLIORI RISPETTO AGLI ALTRI MODELLI.
- FRA LE RETI NEURALI RICORRENTI, SI OSSERVA CHE IN TERMINI DI PERFORMANCE IL MIGLIOR MODELLO È IL LSTM, MA È IL PIÙ LENTO AD ADDESTRARSI. IL SRNN È QUELLO CHE OTTIENE IL MIGLIOR TRADEOFF TRA EFFICACIA ED EFFICIENZA.
- IL PRINCIPALI LIMITI INCONTRATI CHE HANNO “LIMITATO” LA BONTÀ DEI MODELLI SONO I LIMITI DI TEMPO E DI POTENZA COMPUTAZIONALE A DISPOSIZIONE.



RIFERIMENTI



HYNDMAN, R.J., & ATHANASOPOULOS, G. (2018) FORECASTING: PRINCIPLES AND PRACTICE, 2ND EDITION, OTEXTS: MELBOURNE, AUSTRALIA.



CHOLLET, F., 2018. DEEP LEARNING WITH PYTHON.



[HTTPS://WWW.KAGGLE.COM/THEBROWNVIKING20/INTRO-TO-RECURRENT-NEURAL-NETWORKS-LSTM-GRU](https://www.kaggle.com/thebrownvikings20/intro-to-recurrent-neural-networks-lstm-gru)

**GRAZIE
PER
L'ATTENZIONE!**