

PROGETTO DATA SCIENCE LAB

MATTIA VENTOLA
ALESSANDRO MOTTA

Previsione della domanda di
energia nel mercato italiano

ANALISI DEL PROBLEMA

Introduzione al tema e presentazione
del problema di business al quale
si è voluto rispondere

01

EDA E FEATURES ENGINEERING

Exploratory Data Analysis e
analisi della features

02

TEST STATISTICI

Analisi di Stazionarietà,
eteroschedasticità del dataset

03

CONTENUTI

04

MODELLI

Presentazione di tutti i modelli
applicati

05

CONCLUSIONI

Qualità dei modelli e analisi di
MAPE e MAE

06

SVILUPPI FUTURI

Possibili implementazioni al
lavoro effettuato



01

ANALISI DEL PROBLEMA



DATI:

- **TEMA:** DOMANDA DI ENERGIA DEL MERCATO ITALIANO (GME)
- **PERIODO:** DAL 2016 AL 2019
- **GRANULARITÀ:** DATI A LIVELLO ORARIO

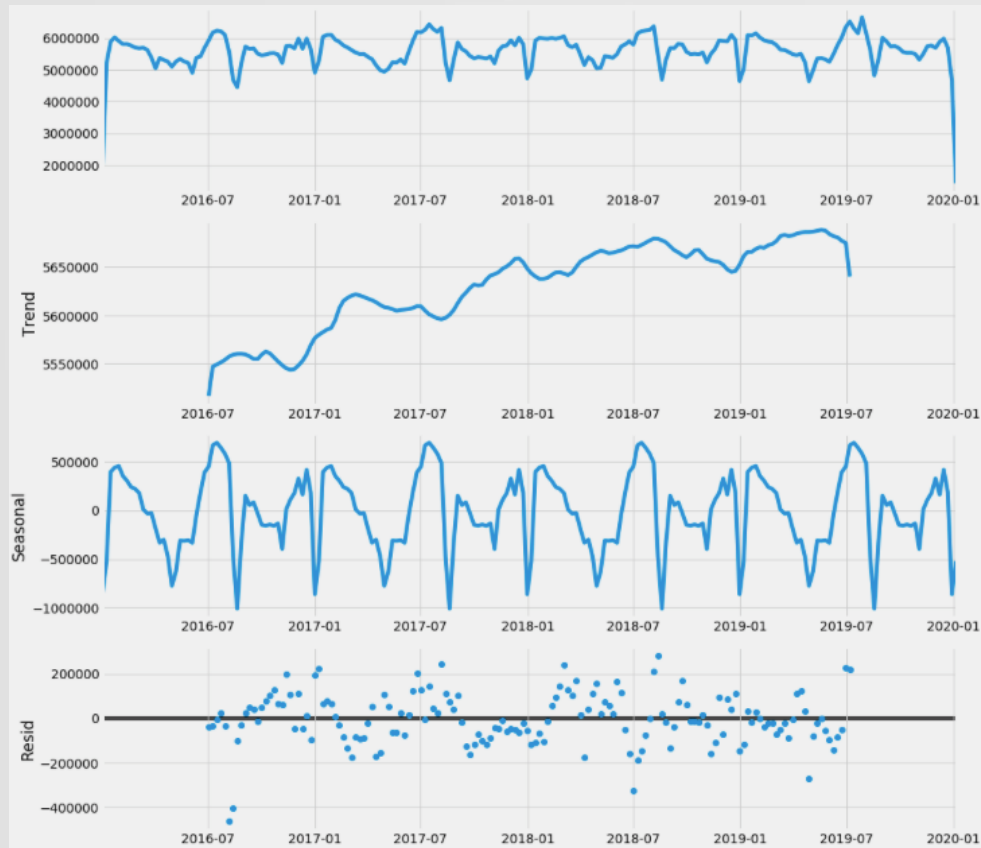
OBIETTIVO:

IMPLEMENTAZIONE DI MODELLI PREDITTIVI PER LA
DOMANDA DI ENERGIA PER ULTIMA SETTIMANA DEL 2019



02

**EDA
E FEATURES
ENGINEERING**



SEASONAL DECOMPOSITION

- Time Series
- Trend
- Stagionalità
- Residui

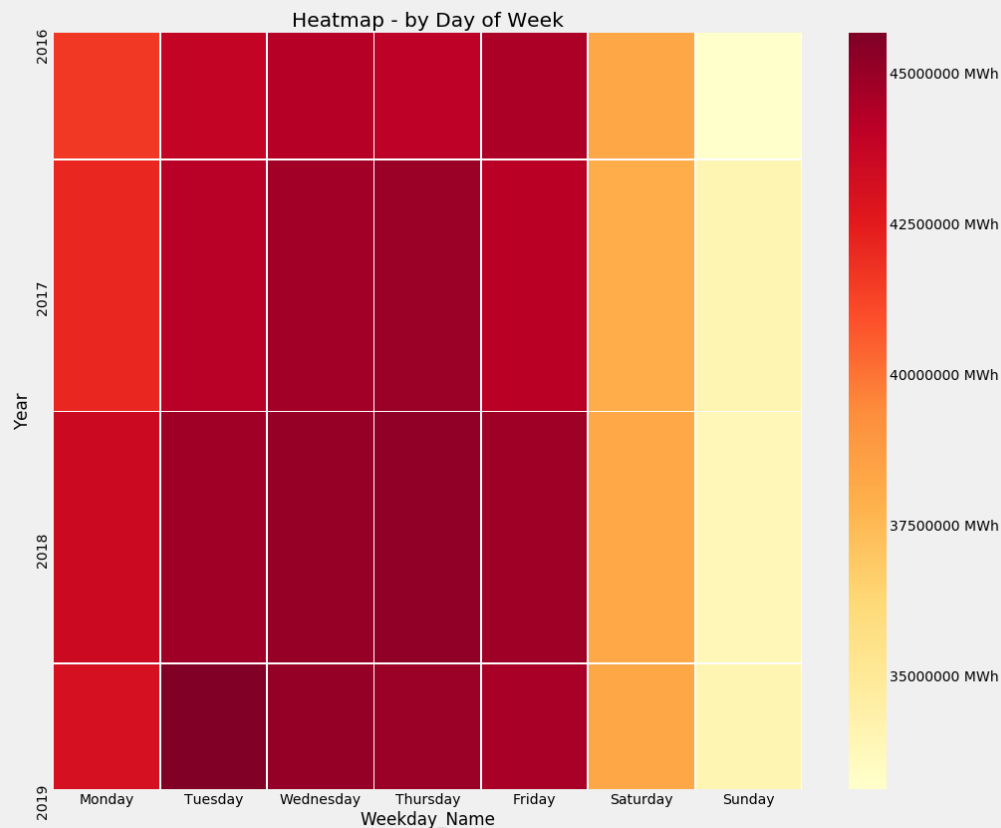
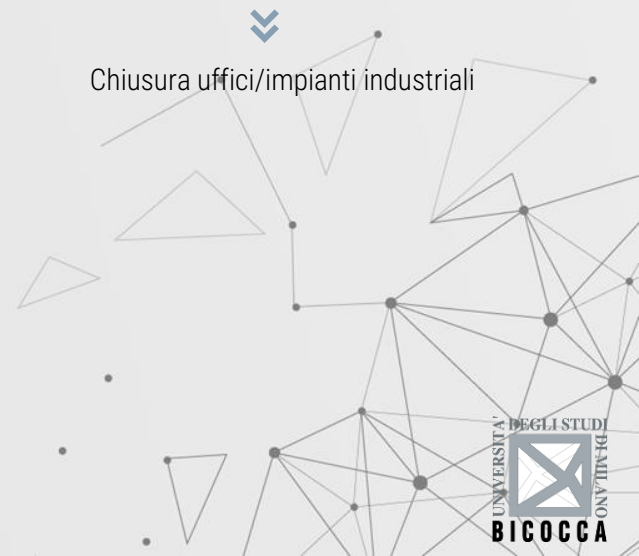
SEASONAL DECOMPOSITION (WEEKLY)



Nel weekend c'è un minor utilizzo di energia



Chiusura uffici/impianti industriali



SEASONAL DECOMPOSITION (MONTHLY)



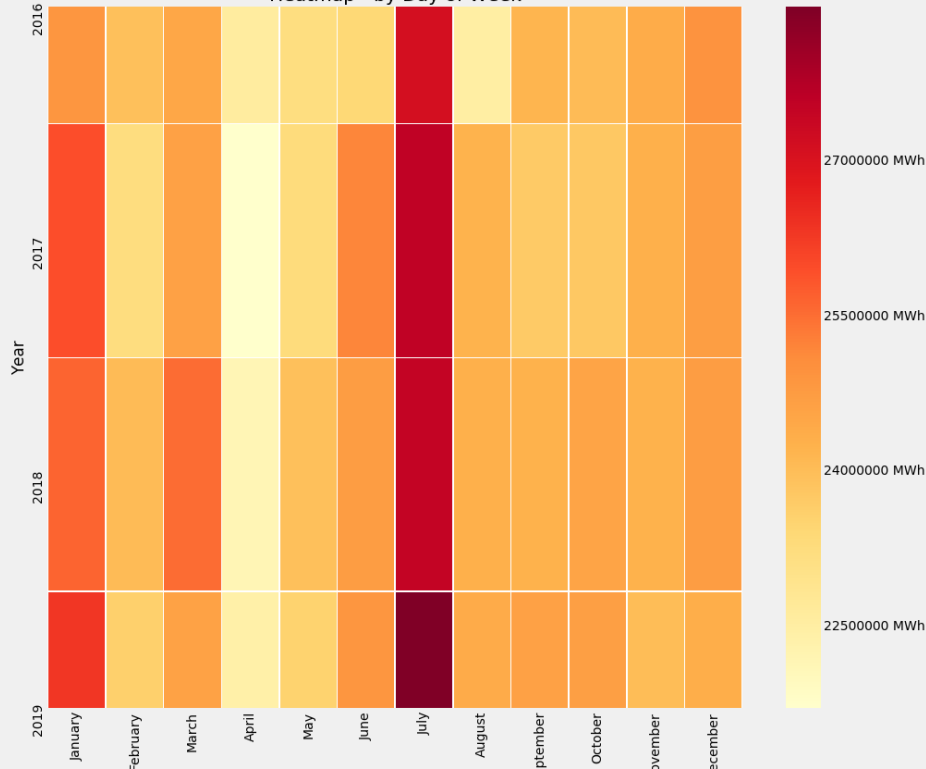
Si può notare che durante Luglio e Gennaio c'è un maggior utilizzo di energia



Uso di riscaldamento d'inverno e dell'aria condizionata d'estate



Heatmap - by Day of Week



TEST STATISTICI

TEST	IPOTESI NULLA	IPOTESI ALTERNATIVA
DICKEY FULLER	Dati non stazionari	Dati stazionari
BREUSH PAGAN	Dati non eteroschedastici	Dati eteroschedastici
P-VALUE	Maggiore di 0.05	Minore di 0.05

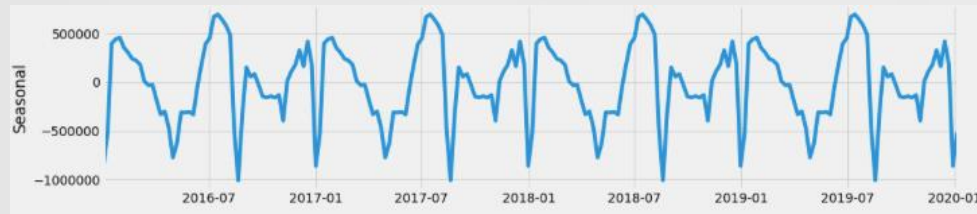
RISULTATI

I dati sono Stazionari

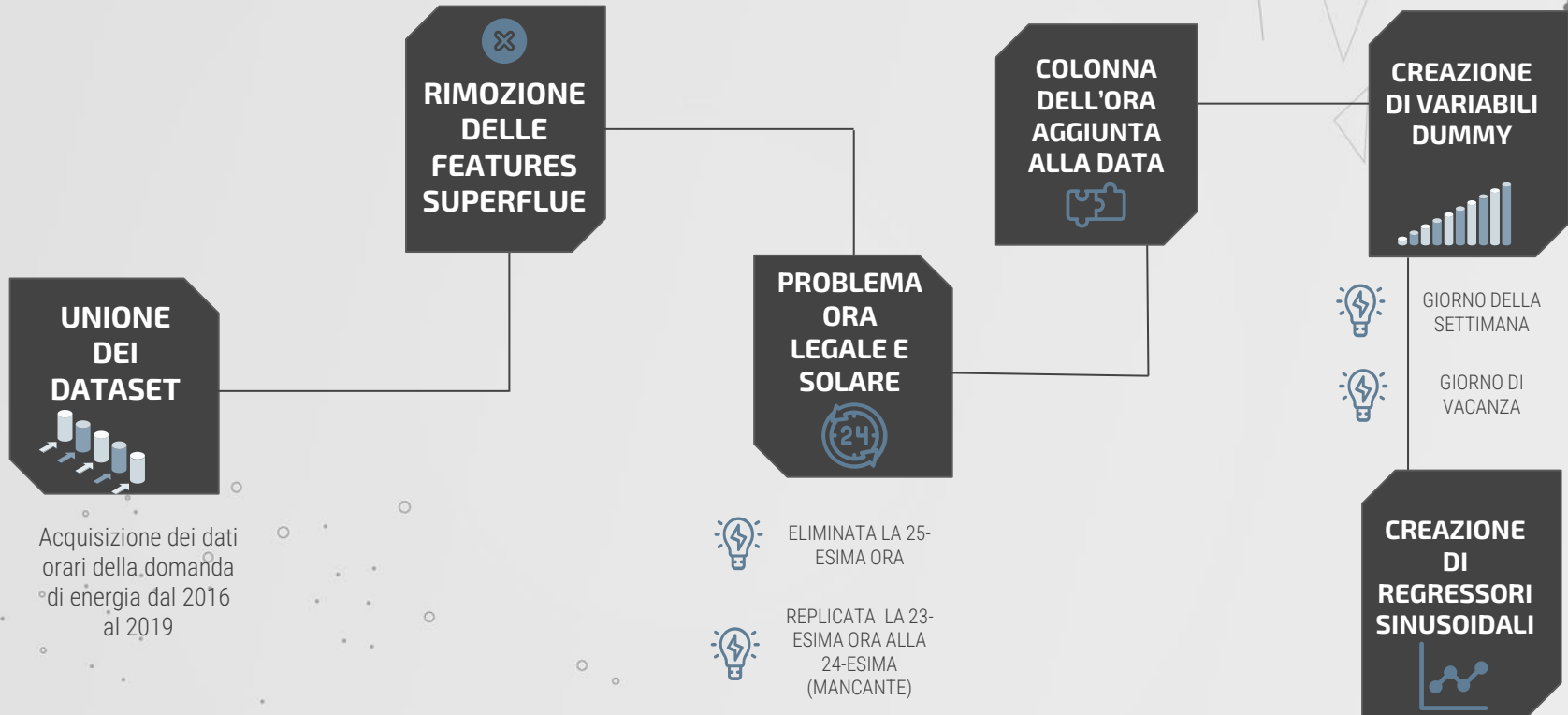
P_VALUE = 0.0000001554280


I dati sono omoschedastici

P_VALUE = 0.50432667557999



PREPROCESSING





04 MODELLI

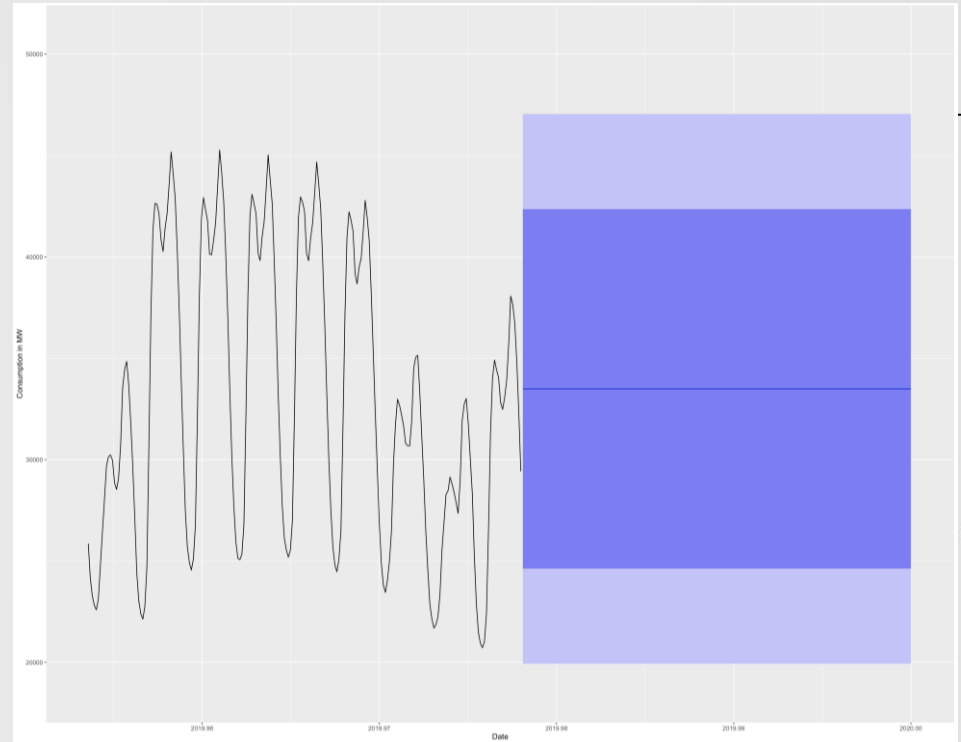
Presentazione di tutti i modelli applicati

BASELINE



PERFORMANCE

TRAIN MAE: 6087,64
TRAIN MAPE: 19,26%
TEST MAE: 7381,29
TEST MAPE: 25,35%



MSTL



PARAMETRI

IMPOSTATO CON $H = 24 \times 7$



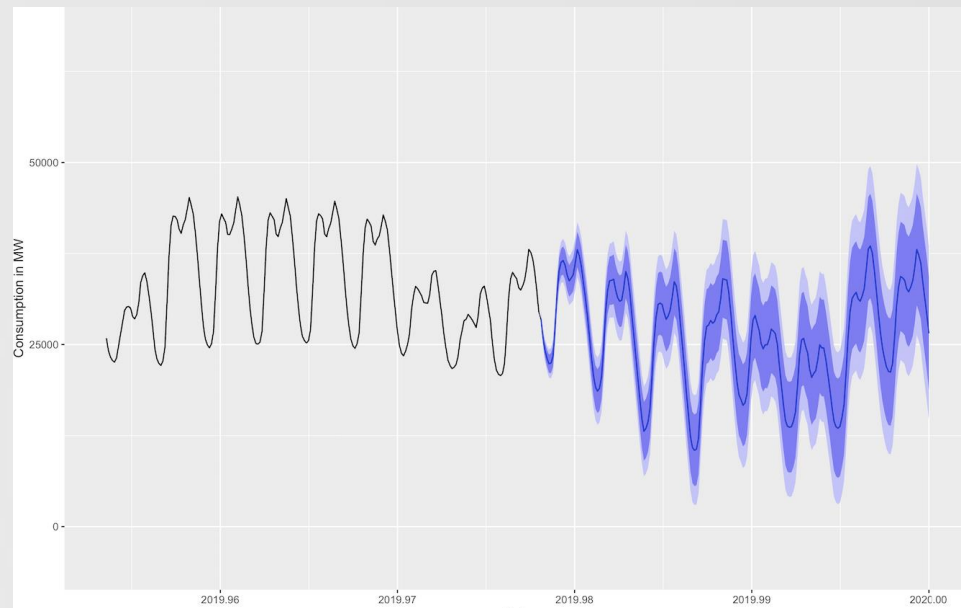
PERFORMANCE

TRAIN MAE: 3288,9

TRAIN MAPE: 5,53%

TEST MAE: 5258,04

TEST MAPE: 19.56%



TBATS



PARAMETRI

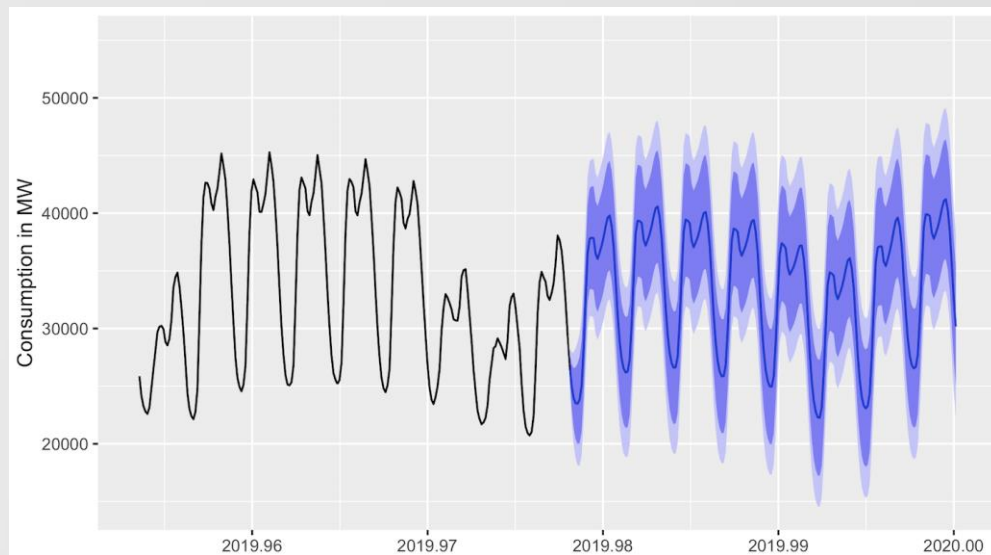
IMPOSTATO CON $H = 24 \times 7$

**TEMPO DI
APPRENDIMENTO: 35 MIN**



PERFORMANCE

TRAIN MAE: 3957,75
TRAIN MAPE: 15,26%
TEST MAE: 5742.88
TEST MAPE: 21.55%



XGBOOST



PARAMETRI

MAX DEPTH = 6

ETA = 0,3

STOPPING ROUNDS = 100

BOOSTER = 'GBTREE'

TEMPO DI

APPRENDIMENTO: 60 MIN



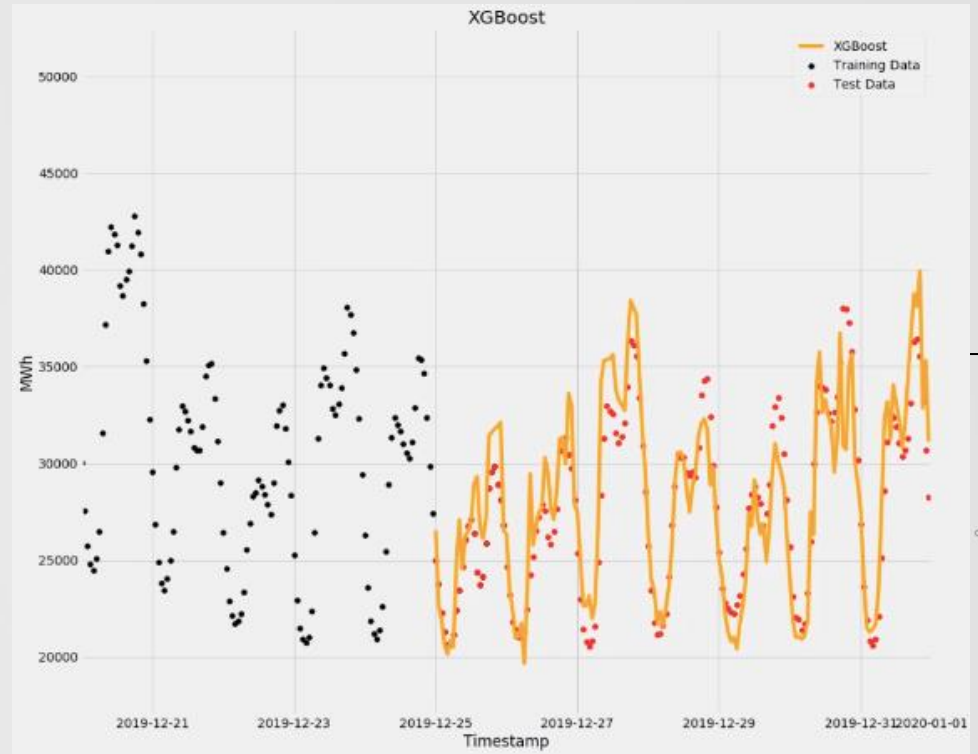
PERFORMANCE

TRAIN MAE: 1012,14

TRAIN MAPE: 4,32%

TEST MAE: 1597,77

TEST MAPE: 6,66%



PROPHET



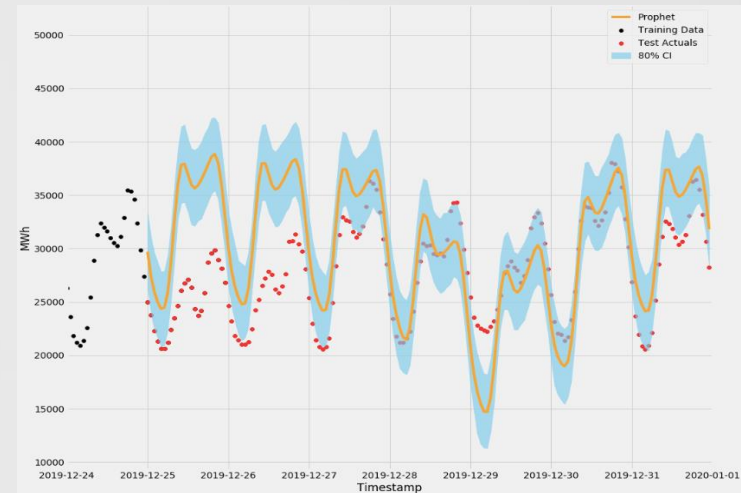
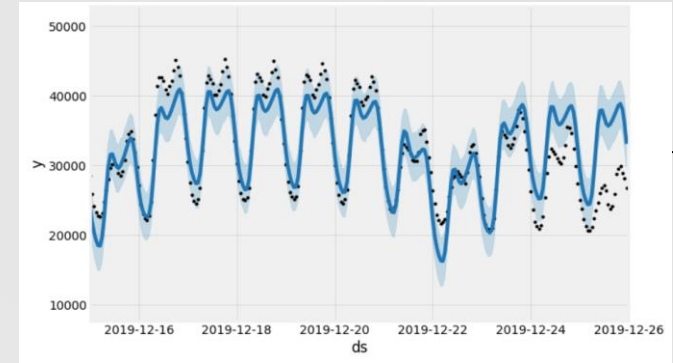
PARAMETRI

25 CHANGEPOINTS
SEASONALITY = 'ADDITIVE'
I.C. = 0,8



PERFORMANCE

TRAIN MAE: 1943,52
TRAIN MAPE: 5,25%
TEST MAE: 2148,21
TEST MAPE: 6.13%



SARIMA



PARAMETRI

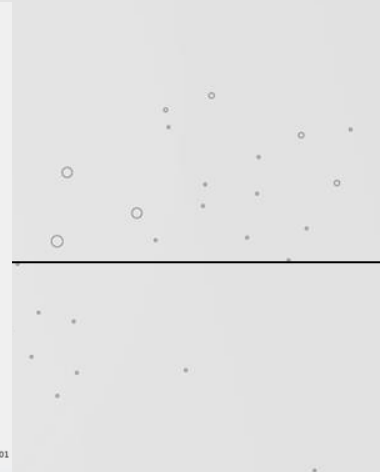
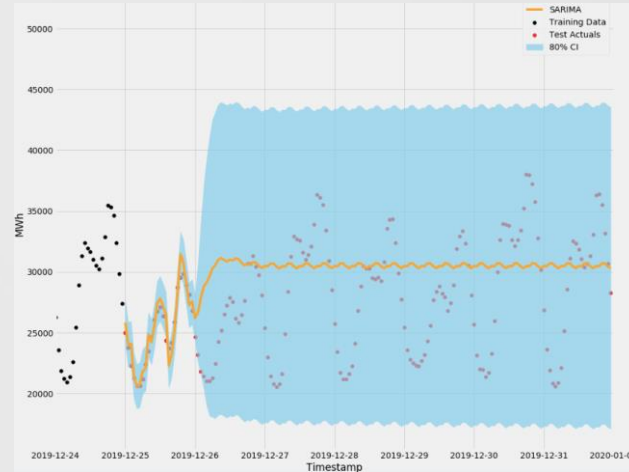
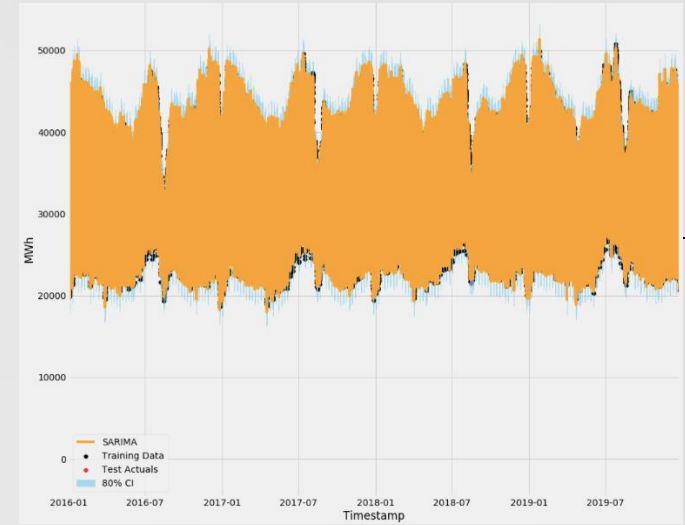
$p = 5$
 $d = 0$
 $q = 5$

$P = 0$
 $D = 0$
 $Q = 2$
 $m = 6$



PERFORMANCE

TRAIN MAE: 1123,14
TRAIN MAPE: 5,02%
TEST MAE: 3714.83
TEST MAPE: 14.83%



SUPPORT VECTOR REGRESSION



PARAMETRI

KERNEL = 'LINEARE'

C = 5

GAMMA = $1E^{-7}$

EPSILON = 1,5

TEMPO DI

APPRENDIMENTO: 40 MIN



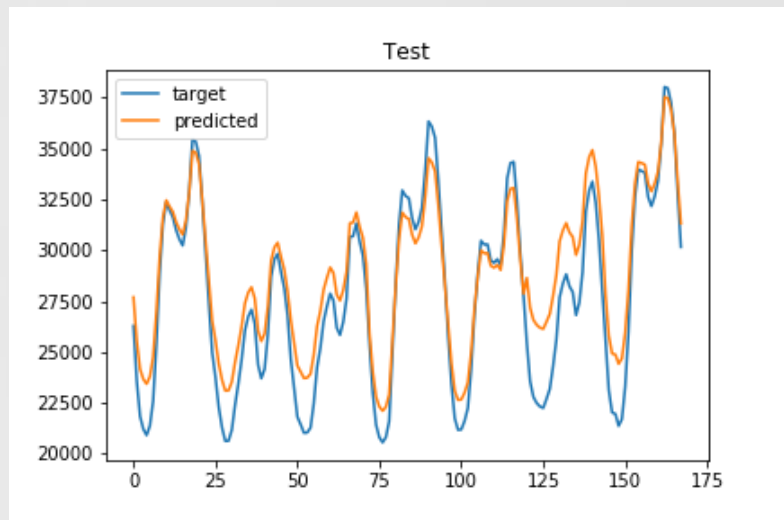
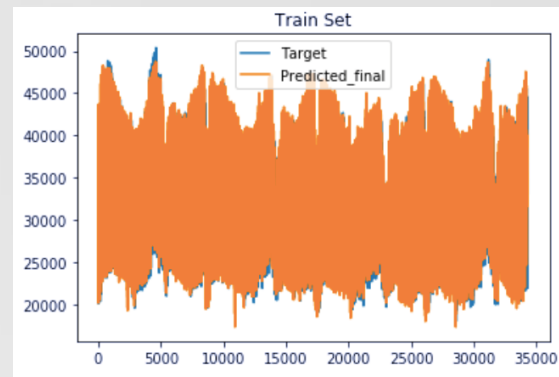
PERFORMANCE

TRAIN MAE: 1014,27

TRAIN MAPE: 4,12%

TEST MAE: 1861,64

TEST MAPE: 7,83%



ELASTIC NET



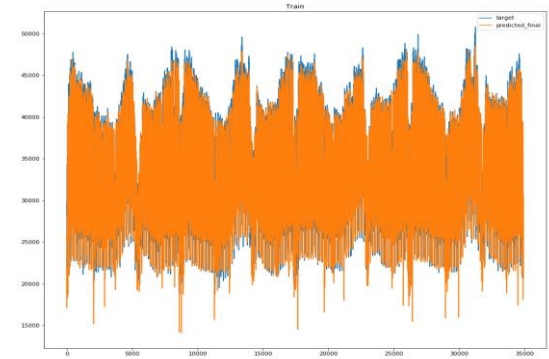
PARAMETRI

ALPHA = 0,6

L1_RATIO = 0,4

TEMPO DI

APPRENDIMENTO: 45 MIN



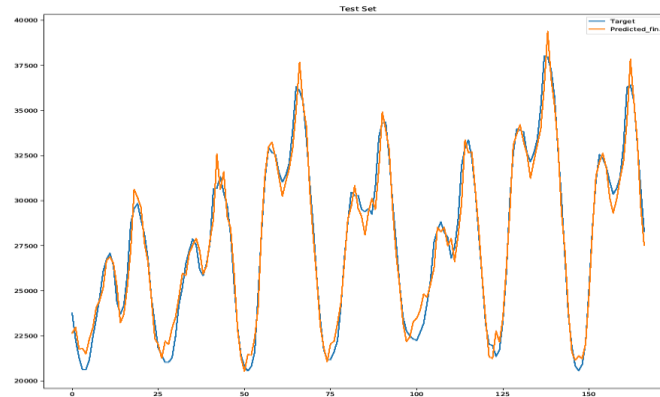
PERFORMANCE

TRAIN MAE: 1440,86

TRAIN MAPE: 4,01%

TEST MAE: 2080,74

TEST MAPE: 8,36%



05

CONCLUSIONI

Qualità dei modelli e analisi di MAE e MAPE

CONCLUSIONI

PER POTER COMPARARE I RISULTATI È STATO EFFETTUATO UN CONFRONTO TRA I RISPETTIVI MAPE:

DALL'ANALISI EMERGE CHE:

- **TUTTI I METODI RISULTANO AVERE UN MAPE INFERIORE ALLA BASELINE**
- **XGBOOST E PROPHET HANNO EFFETTUATO PREVISIONI PIÙ ACCURATE**
- **MODELLI COME SARIMA, MSTL E TBATS SI SONO DIMOSTRATI POCO PERFORMANTI**

**MAPE Prophet: 6.13 %
MAPE XGBoost: 6,66 %
MAPE SVR: 7,83 %
MAPE Elastic Net: 8,36 %
MAPE SARIMA: 14.83 %
MAPE MSTL: 19.56%
MAPE TBATS: 21.55%
MAPE Baseline: 25,35%**

06

SVILUPPI FUTURI

Possibili implementazioni al lavoro effettuato

SVILUPPI FUTURI

AVENDO A DISPOSIZIONE ANCHE I DATI ORARI SULLE TEMPERATURE E CONDIZIONI METEOROLOGICHE ITALIANE, SAREBBE INTERESSANTE POTER INSERIRE QUESTO TIPO DI VARIABILI NEI MODELLI PER RENDERE PIÙ ACCURATE LE PREVISIONI.

INFATTI DALL'ANALISI SI È NOTATO CHE I MAGGIORI CONSUMI DI ENERGIA ELETTRICA AVVENGONO IN CORRISPONDENZA DELL'INVERNO E DELL'ESTATE, E CHE QUINDI LA TEMPERATURA È STRETTAMENTE CORRELATA ALLA DOMANDA DI ENERGIA.



**GRAZIE PER
L'ATTENZIONE**