



BUSINESS ANALYTICS & DATA MINING CHAMPIONSHIP 2019

BADM_1002

- Varan Singh Rohila
- Aryan Khandal
- Palak Sood



Contents

- Objective
- Dataset Provided
- Data Preprocessing
- Data Analysis
- Approach and Building of the Model
- Conclusion



Objective

A renowned energy corporation approached our firm with the following requirements:

- Since it is crucial to match supply with demand, what pattern or schedule they should adopt for their energy/power purchases?
- What are the estimated energy consumption rates of their consumers for the upcoming years (long-term forecast)?

Hence, our case objective defined is **“Energy Demand Forecasting for Power Industry”**



Dataset Provided

DATE	TOTAL Load	AVGTEMP	AVGHUM	HOL_IND	SUNRISE	SUNSET	daylight
2004-01-01 00:30:00	253080.0	68	55	1	1900-01-02 07:01:00	1900-01-01 18:22:00	12:39:00
2004-01-01 01:00:00	250560.0	68	55	1	1900-01-02 07:01:00	1900-01-01 18:22:00	12:39:00
2004-01-01 01:30:00	248400.0	68	55	1	1900-01-02 07:01:00	1900-01-01 18:22:00	12:39:00
2004-01-01 02:00:00	245880.0	68	55	1	1900-01-02 07:01:00	1900-01-01 18:22:00	12:39:00
2004-01-01 02:30:00	244720.0	68	55	1	1900-01-02 07:01:00	1900-01-01 18:22:00	12:39:00



Data Preprocessing

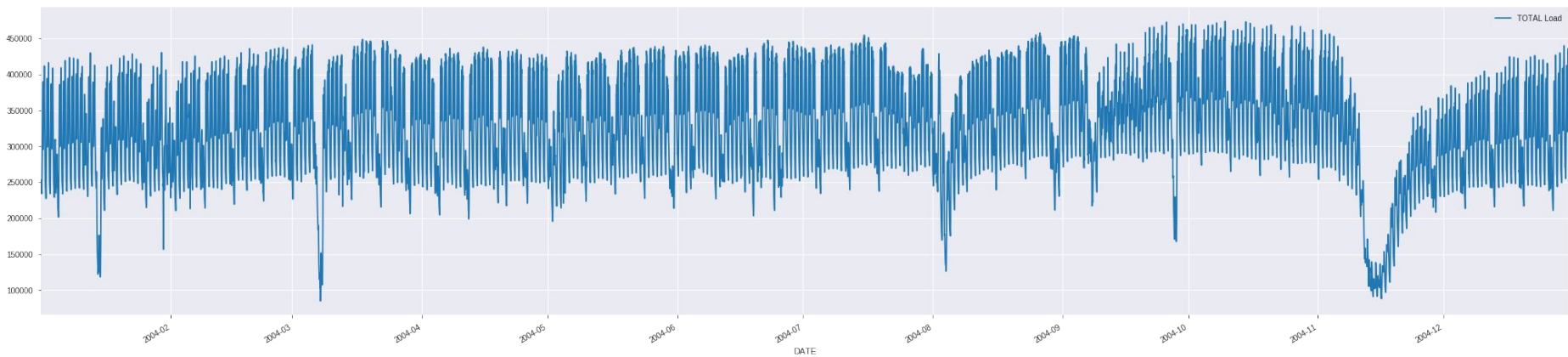
	TOTAL Load	AVGTEMP	AVGHUM	HOL_IND	daylight
DATE					
2004-01-01 00:30:00	0.507698	68	55	1	759.0
2004-01-01 01:00:00	0.502234	68	55	1	759.0
2004-01-01 01:30:00	0.497550	68	55	1	759.0
2004-01-01 02:00:00	0.492085	68	55	1	759.0
2004-01-01 02:30:00	0.489569	68	55	1	759.0



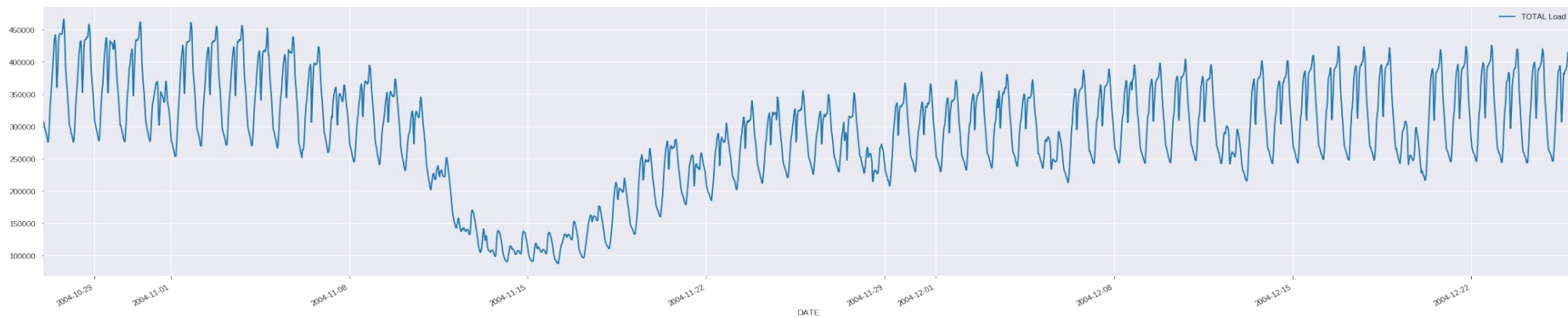
Data Analysis

After the required preprocessing, EDA was used to generalize the data. Various trends in the provided data were observed and some conclusions were drawn as follow:

1. Drops in power consumption are observed usually when there is a holiday.
2. There is a decreasing and increasing trend at the end of every year which probably indicates a season change.



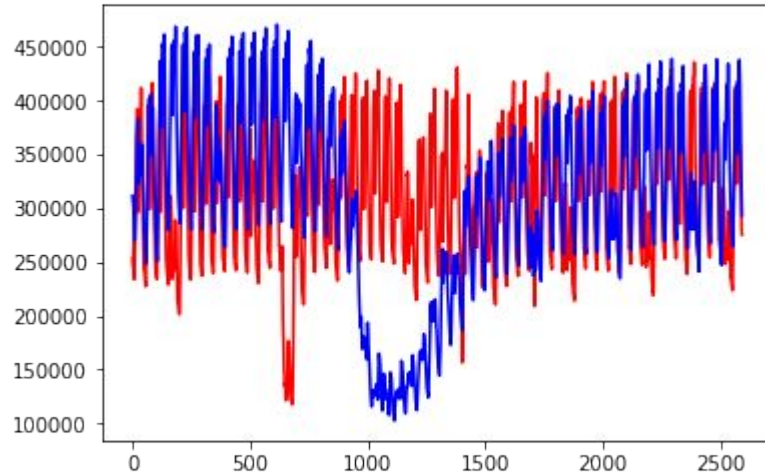
1 year's Power Consumption



Expanded view of the decreasing curve

Approach and Building Model

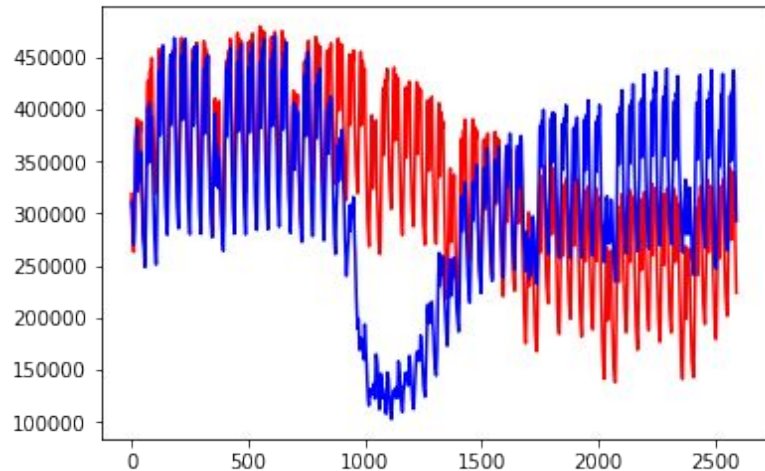
SARIMA adds three new hyperparameters to specify the autoregression (AR), differencing (I) and moving average (MA) for the seasonal component of the series, as well as an additional parameter for the period of the seasonality. The **MAPE** obtained for the **validation dataset** is 28.



Approach and Building Model

Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data.

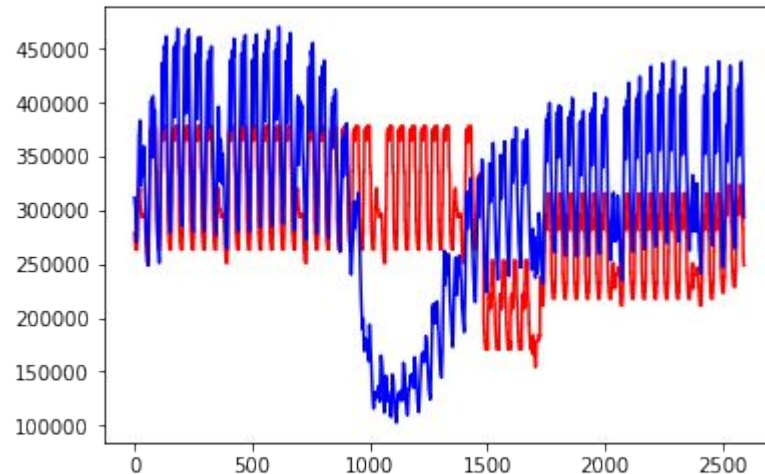
The **training MAPE** for the dataset is **10** while the **validation MAPE** is **32**.



Approach and Building Model

XGBoost, a Gradient Boosted tree algorithm can be used for the forecasting problems.


The **testing MAPE** observed is **32**.





Conclusion

- Since the energy consumption is low on holidays, the energy corporation should generate less power for those days.
- The dip in the graph is not getting detected by the model as the dataset is for only 3 years, so the drop in the energy consumption occurs only three times in the respective two years which is very less compared to the granularity of data i.e. for 30 minute intervals.
- The dip in the graph probably represents a season change since there is a sharp decrease in the power consumption. The corporation can limit the generation of power during that period.
- The training MAPE for the dataset is 10 while the validation MAPE is 32 for prophet, 28 for SARIMA model.

- 
- Day of the year was most commonly used to split trees, while hour and year came in next thus stating their respective importance.
 - The forecasted energy demand values can be used to optimize the power generation process thus decreasing wastage and increasing the profits of the corporation.
 - The SARIMA model has given the closest forecasted values and will also prove to be effective in the production environment.



THANK YOU