Cognizant

Forecast Power Consumption for a Household

Amardeep Bhagat

Background

Problem summary:

Predict global power consumption for a household based on the past history of usage from 2007 till 2010

Available data set:

- Data source: UC Irvine individual power consumption data from a location near Paris, France
- Minute-averaged active and reactive power for a
- Voltage and intensity
- Distribution of power by set of kitchen appliances, laundry appliances and temperature affected units (e.g. A/C, water heater etc.)

Missing Data Set:

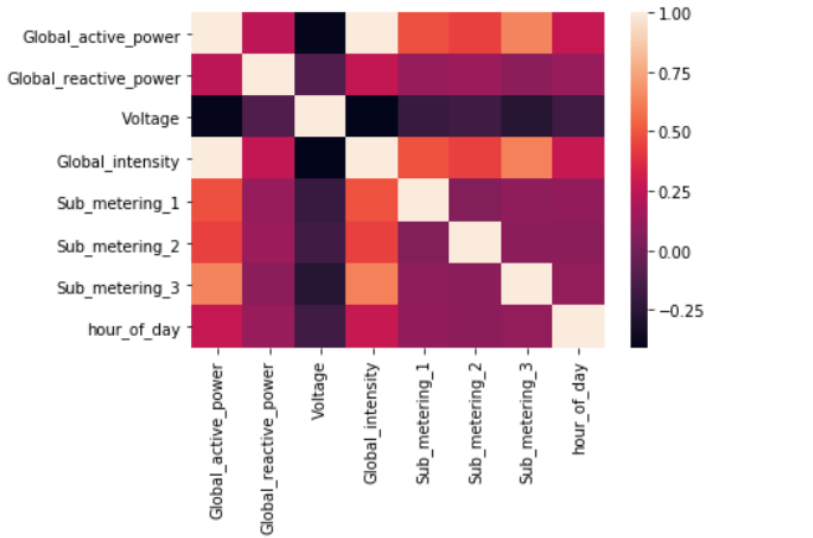
- Hourly temperature of the targeted location

Key Observations in the data in EDA

Observations

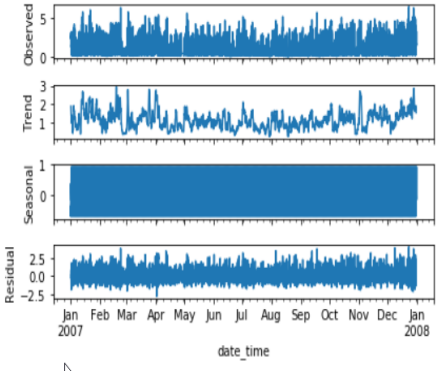
- Non-numeric values: converted to null
- Missing values: used forward fill logic to fill the gap
- Absence of features impacting power consumption (except intensity which is proportional to power consumption by laws of physics)
- Seasonality of hourly consumption is the biggest factor that can be used for auto-correlation

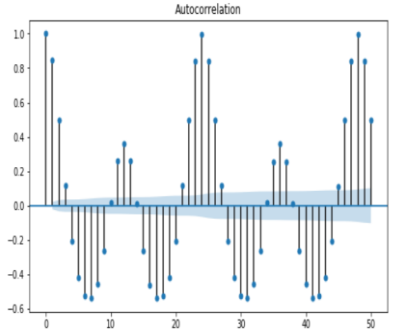
Correlation of different features

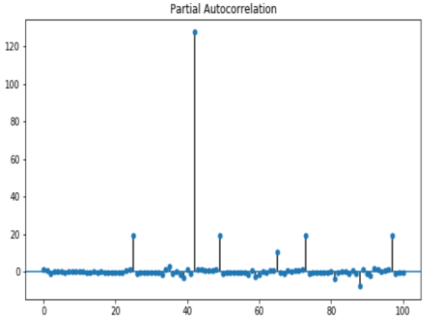


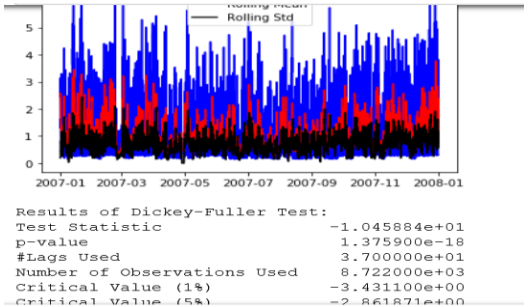
Cognizant

Additional Observations for Model selection









- ✓ Decomposition of power usage data shows trend, seasonality and residual values
- ✓ ACF value of seasonal shows a cyclic behavior of 4 hours lag
- ✓ PACF value of seasonal shows a lag of 24 hours
- ✓ Results of Dickey Fuller's test show a very high test statistic value compared to critical values and hence the stationarity in data

Cognizant

Model Preparation

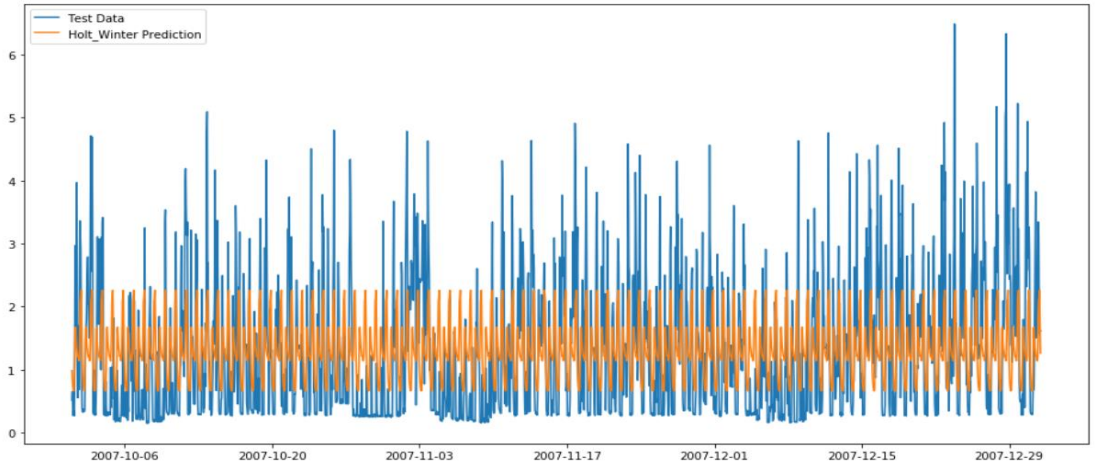
Test and Training Data Set

- Q1-Q3 2007 data was used as training data set while Q4 2007 data was used as test data set

Model Selection

- Models that didn't work
  - Naïve, simple average, moving average, simple linear smoothening and Holts Linear models were discarded based on linear way of prediction
  - SARIMA model didn’t provide any conclusive prediction based on high variability in order and seasonal parameters. Also the RMS error was way too high
- Model that forecasted the closest
  - Holt Winters Model

Observations from Model



Observations

- Holt Winters model was able to provide the foundational prediction and mostly was affected by hourly usage change behavior
- Any spikes in the test data which was mainly based on temperature factor was unaccounted for in the prediction based on the missing temperature data
- Correlation of data was more hourly and no impact of long term seasonality (based on season changes over the year)
- RMS error was 0.93 which was in medium acceptance limit

Cognizant

Potential Next Steps

Data Set Enrichment:

- Choose a data set where both usage and temperature data available for the applicable locality

Additional opportunities in forecasting:

- Multi-variate analysis based on temperature and seasonality
- Predict individual appliance consumption (that are dependent and independent of temperature) and aggregate to derive overall forecast of power consumption

Additional Model considerations:

- Enhance SARIMA model output through auto selection of order and seasonal factors
- Apply deep learning algorithms
- Explore possibilities of ensemble methods based on linear regression (temperature affected forecast) and other time series based models

Cognizant

Thank you

Amardeep Bhagat
amardeep.bhagat@cognizant.com