Electricity Demand Forecasting based on smart meter feeds

GitHub /GR3AT

Project Objective

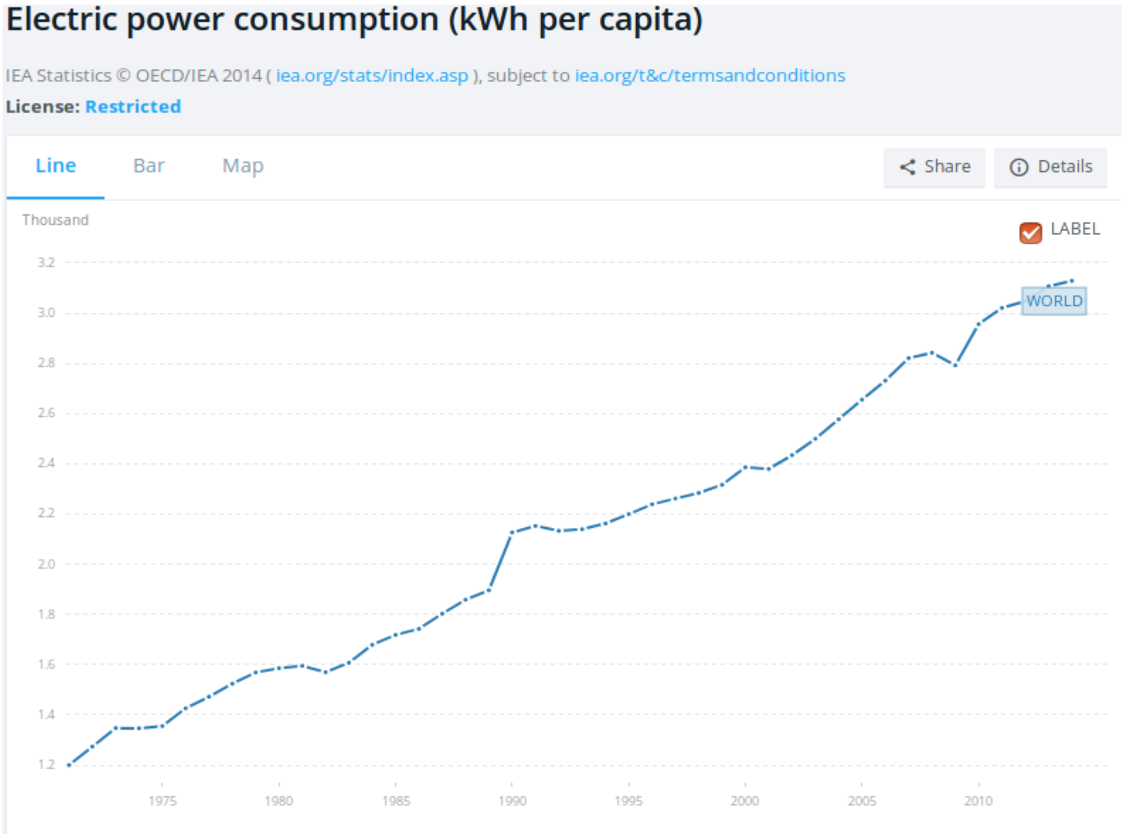
Compare several machine learning algorithms based on their performance on electricity demand (time series) forecasting

Why is this important - Motivation

Every moment :
Electricity Demand = Power Supply

Reasons which make this match difficult :

- Renewable energy penetration
- Unknown customer behaviour
- Constantly changing needs
- Unknown weather conditions



Mismatch Problems

Electrical Power Losses

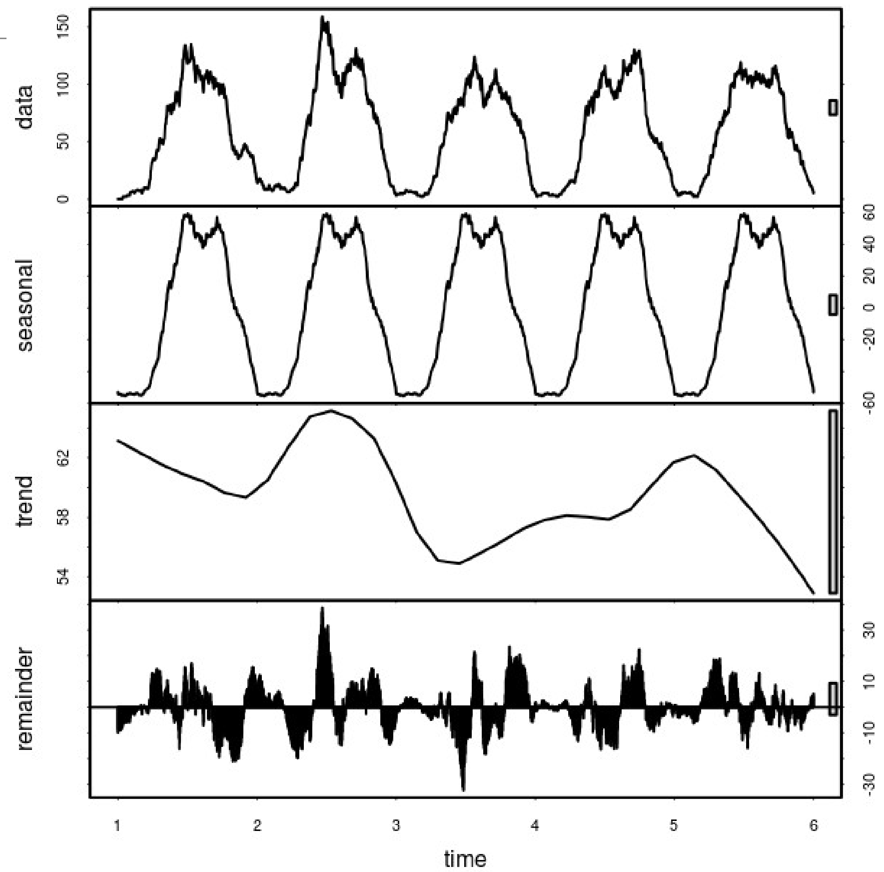
1% Short-term load forecasting improvement for a Power Plant of 1-GigaWatt $300,000 annually [2]

Power Outages (Blackout)

Solution : System Monitoring and Development of better Forecasting Models

Forecasting Model - ARIMA

About time series:
- Seasonality
- Trend
- remainder



Autoregressive (AR) Integrated (I) Moving Average (MA) model parameters :

1. p, the autoregressive term (number of time lags to consider)
2. d, the integrated term (differencing degree to achieve stationarity)
3. q, the moving average term (model error as combination of previous lags error).

Case Study

Compare models based on machine learning algorithms, using two evaluation metrics, for the study of short term forecasting.

Regression algorithms	Evaluation metrics	Short term forecasting frequency
Random Forest	Root Mean Square Error (RMSE)	Hourly forecasting
Decision Tree		
Linear Regression	R-Squared Score (R2)	
Support Vector Machines		
Multi-Layer Perceptron	Execution Time	

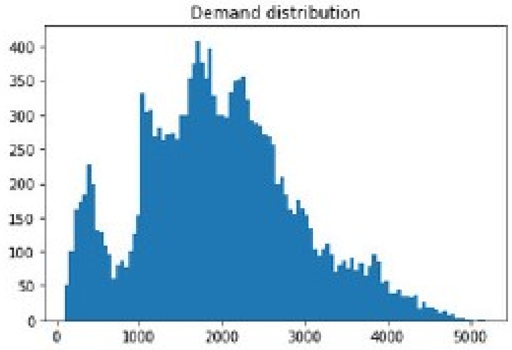
Datasets

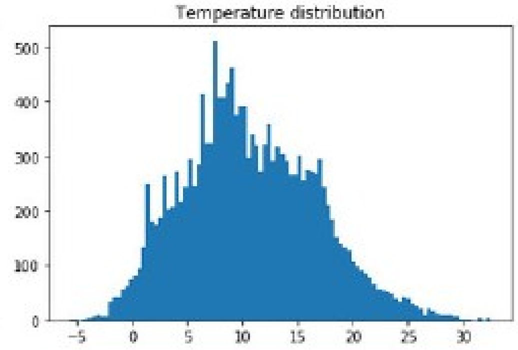
- **Energy Dataset** : Half-hour Energy Demand data for more than 5 thousand households in London, collected between 2011 and 2014 using in house smart meters.
  - 167 million records (~10 GB)
  - Sparse data

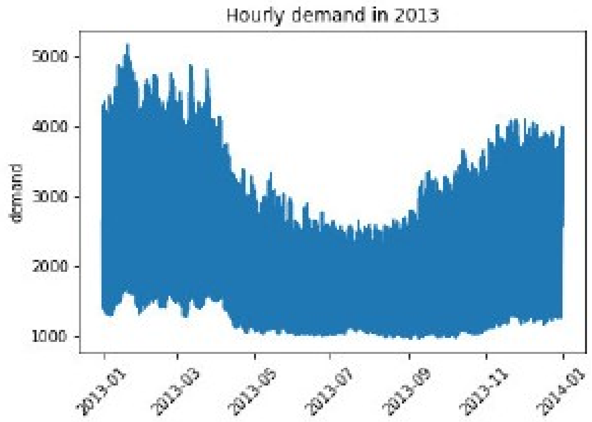
- **Weather Dataset** : London Weather Dataset downloaded from Kaggle.
  - Contains Half-hour weather related data for the period between 2011 and 2014
  - Clean dataset, ready to use

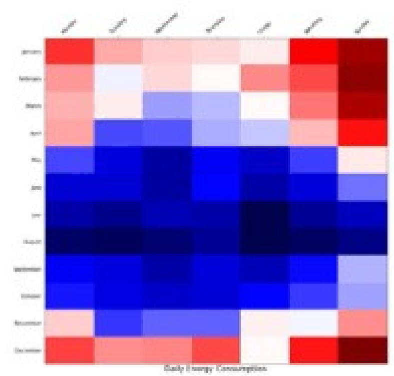
Preprocessing

- **Energy dataset Preprocessing**
  - Resample 30m-->60min
  - Group by civilian gorups (ACORNs)
  - Create new features
- Result : 167 million to 735 thousand records

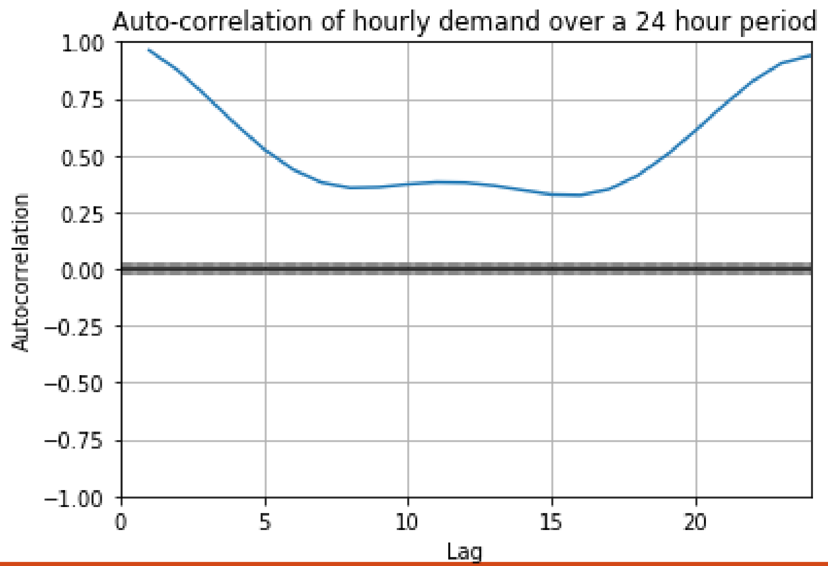








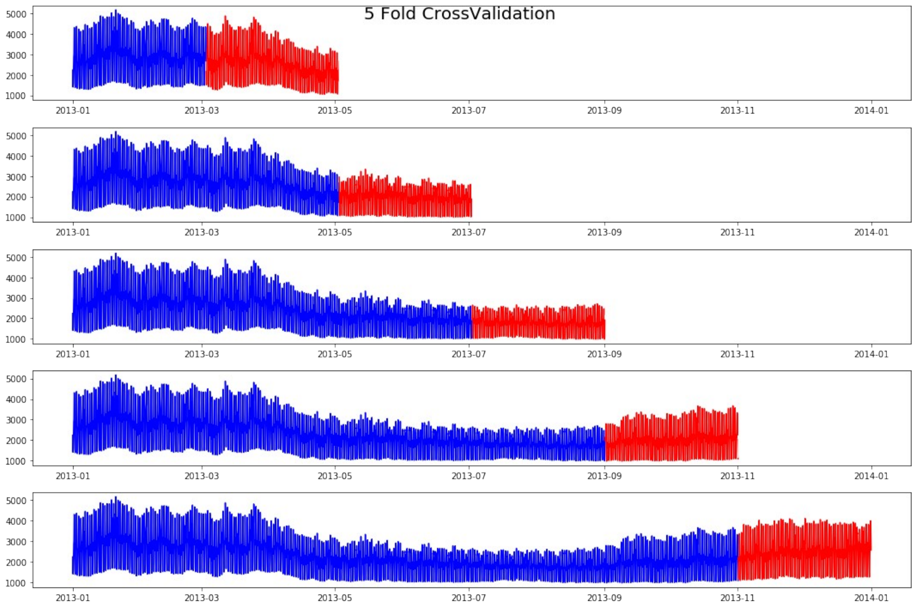
- **Weather dataset Preprocessing**
  - Resample 30m-->60min
  - Keep only features of interest (temperature, windSpeed, Humidity)
  - Filter odd values (humidity > 100%)
- Combine datasets
- Normalize values
- Remove Nan values



Algorithms – Best Model Selection

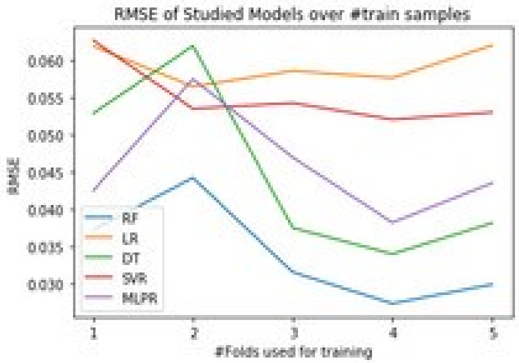
Grid Search Approach using a few tunning parameters and 5 fold Cross-Validation

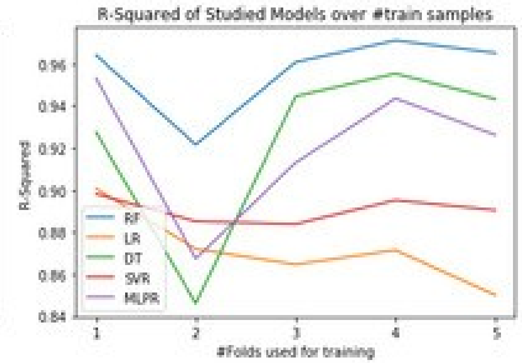
Algorithm	Tunning Parameters
Random Forest	#trees, #features, maximum tree depth
Decision Tree	Split criterion, #features, maximum tree depth
Support Vector Machines	Kernel, kernel coeficint (gamma), penalty (C)
Multi-Layer Perceptron	#hidden layer, #perceptrons per layer, activation function and solver
ARIMA	Order and Seasonal order (p,d,q)

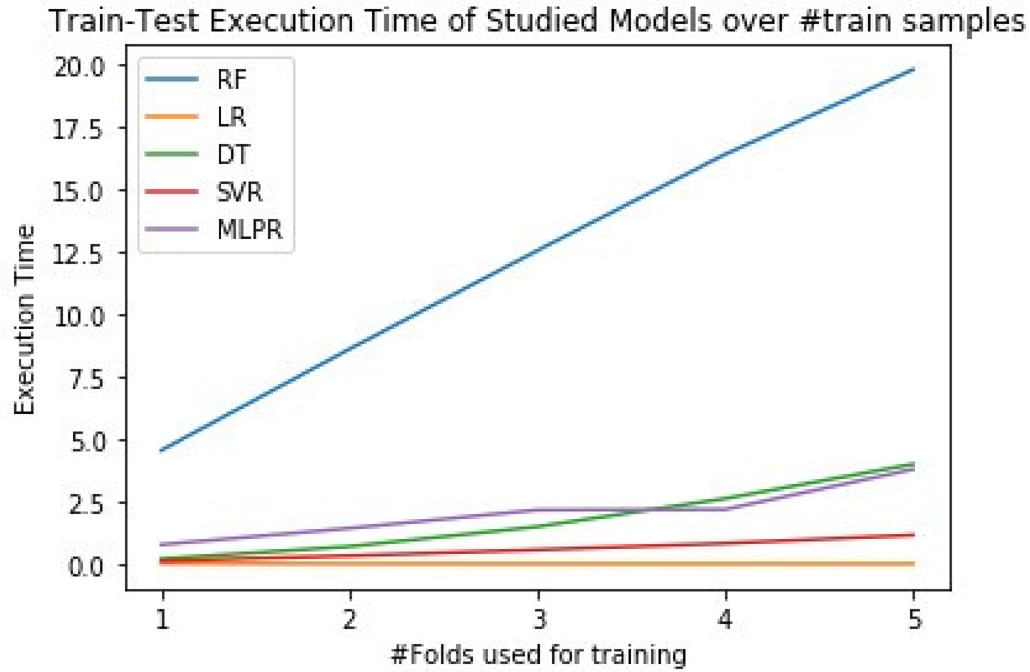


Results

Model	RMSE	R-Squared	Elapsed Time
SARIMA	0.088	-0.076	8.324
RF	0.024	0.977	65.098
LR	0.036	0.946	0.010
DT	0.036	0.947	8.920
SVM	0.062	0.840	0.254
MLP	0.033	0.954	2.672

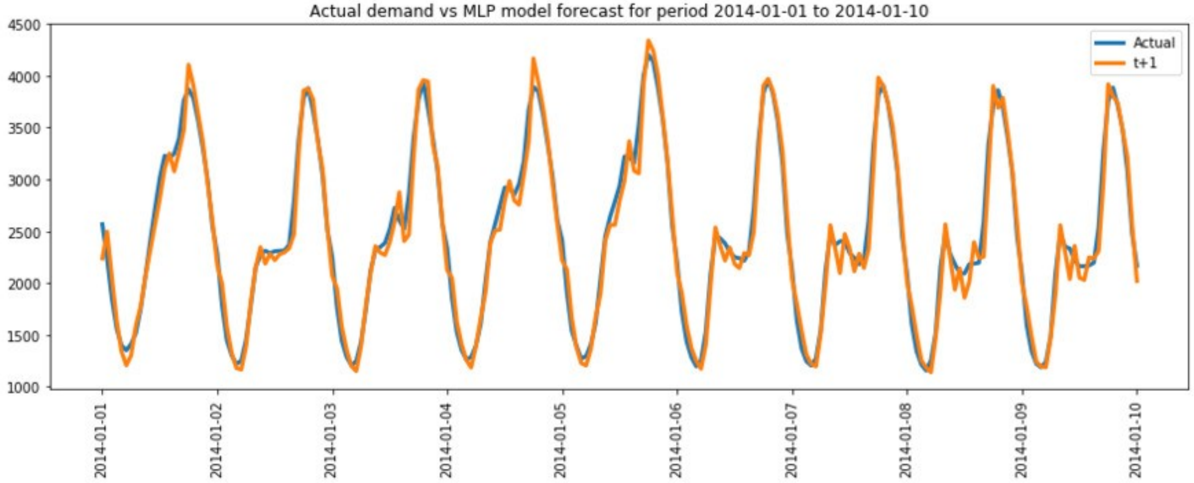


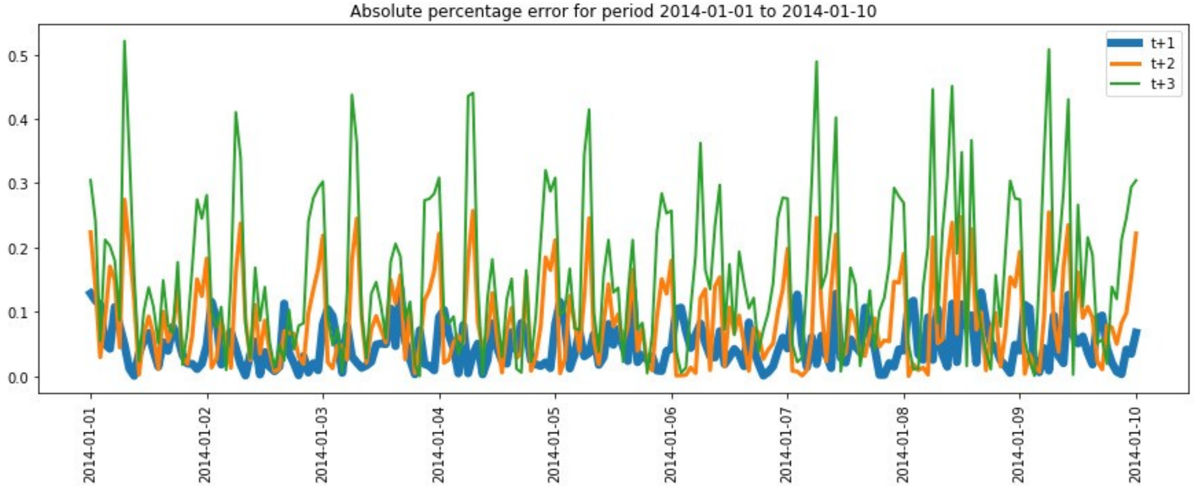




Conclusion - Next Hour Forecasting

- We have shown that appart from ARIMA statistical methods there are more ML Algorithms which can perform equally or better in Energy Demand Forecasting.
- There is a lot of space for improvement in the field of electricity demand forecasting, which can be filled with the help of machine learning algorithms.





Thank You

References

[1] WorldBank Graph : https://data.worldbank.org/indicator/EG.USE.ELEC.KH.PC
[2] Hong, Tao (2015). "Crystal Ball Lessons in Predictive Analytics". EnergyBiz Magazine. Spring: 35–37.
[3] Power Outage pictures : https://goo.gl/images/Kx5AGN, https://goo.gl/images/CXqzCS