General Forecasting Techniques and Machine Learning in Industrial Applications

Jordi Nadal on behalf of Grupo AIA

January 22th, 2018
Chennai

1. General introduction to load forecasting
2. Conventional forecasting techniques
3. Initiatives using machine learning techniques for forecasting
4. *‘State of the art’* methods for prediction
5. Grupo AIA experience
6. Conclusions
7. Bibliography

1- General introduction to load forecasting

Importance of load forecasting

- **Load forecasting** (LF) has become one of the most **important** areas of research in the power industry, especially in **power system design, planning, operation and development** and it is the base of economic studies of energy distribution and power market.
- **Accurate** LF plays a key role in **reducing generation costs** and deals with the **reliability** of the power system.
- In addition, **renewable energy** LF (wind, solar, etc.) become even **more difficult** since new descriptive variables with different phenomenology come into play.

Forecasting classification

- **By periods**: The period of the load forecasting can go from years (long-term) to weeks/months (medium) or day/hours (short).
- **By methodology used**:
  - Statistical approach: Mathematical models that predict the relationship between load and several input factors.
  - Expert systems: Previous knowledge (if-then rules).

Forecasting model procedure

Independently of the **methodology to be applied**, tasks are divided into 4 parts:

- Identification of input variables (feature engineering).
- Pre-processing of data (outlier/peak detection).
- Selection of training set.
- Testing the model (define a metric).

2- Conventional forecasting techniques

Summary of the three most frequently used short-term forecasting methods:

ARIMA

- **Input variables:** Past values of load, and past forecast errors.
- **Advantages:** Dynamic, can correct for correlation of errors across observations.
- **Disadvantages:** Assumption of linearity, cannot capture impact of weather variables.

Linear models

- **Input variables**: Weather and calendar variables.
- **Advantages:** Can capture impact of weather on load, easier model estimation, easier interpretation of model parameters.
- **Disadvantages**: Restrictive assumption of linear relationship between load and other variables.

Artificial Neural Networks

- **Input variables**: Weather and calendar variables, nonlinear transformations of those.
- **Advantages:** Flexible, can capture nonlinearity, allows interaction among load affecting factors.
- **Disadvantages**: Requires large data samples, estimation takes longer, network architecture may be usually utility dependent. Problematic with over-fitting.

3- Initiatives using ML: GEFCom2012/2014

- Recently, efforts have been focused on the use of machine learning algorithms to solve forecasting-like problems.
- Those initiatives carry out a significant increase of the performance (lower values of MAPEs), but also a less computational time, more flexibility and interpretability.
- The methods used in the different events:

Winning Methods

GEFCom2012
Wind Forecasting

Techniques

1. Gradient boosting, regression
2. K-Nearest Neighborhood, regression
3. Gradient boosting, regression
4. Neural networks, Gaussian process

The magic of
"Gradient Boosting + REGRESSION"

GEFCom2014

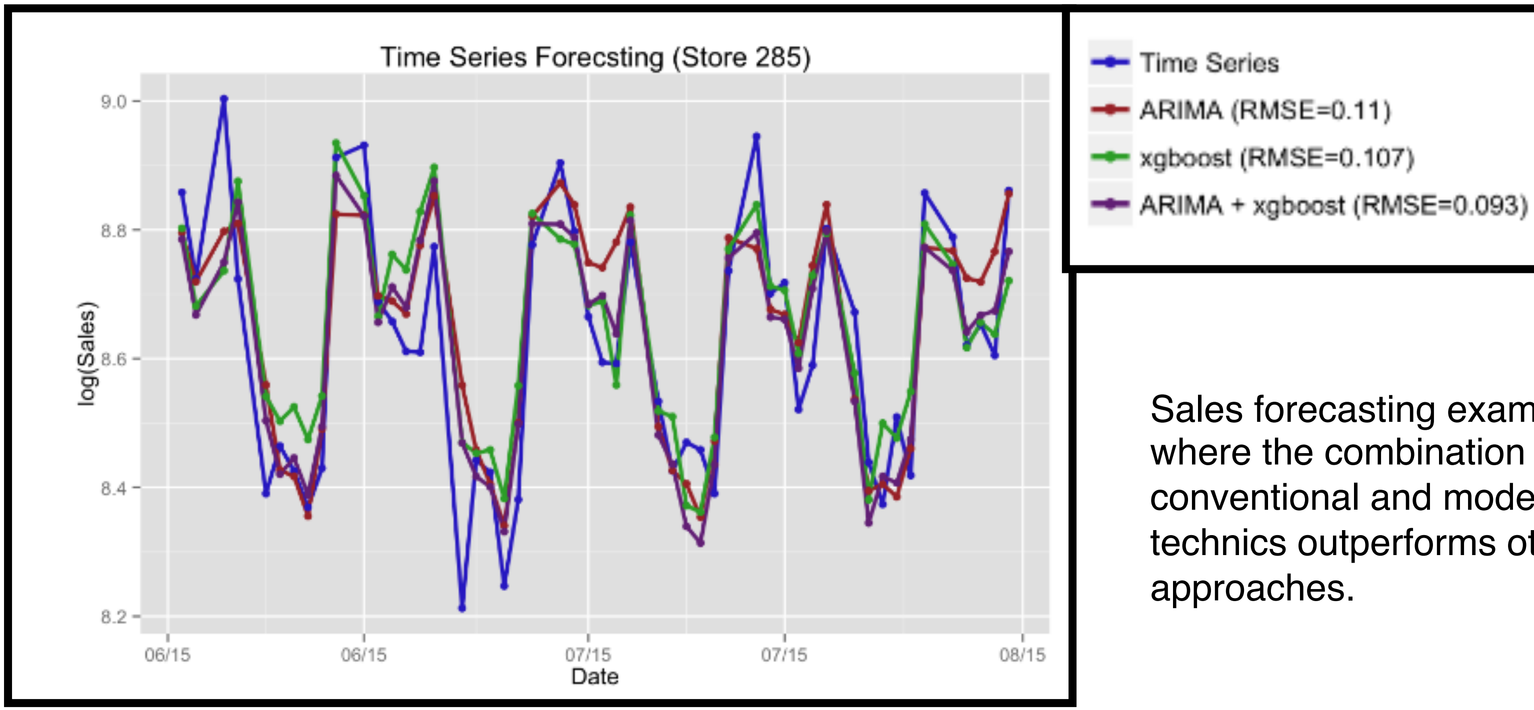
teams in the wind track of GEFCom2014.

ic	Forecasting models and techniques	Generalization ability (preventing overfitting)	Input variables and features (most important)
	Gradient Boosting Machines (GBM), organized in two layers	Cross validation	Forecasts of wind speed and direction at 10 and 100 m, at target and neighboring lead times, time of day
	Quantile Regression Forest (QRF) and Gradient Boosting Decision Trees (GBDT)	Cross validation	Forecasts of wind speed and direction at 10 and 100 m, and features derived from these
	k-Nearest Neighbor (*k*-NN)	Training data selection	Wind components, wind speed and direction at 10 and 100 m (filtered), time of day

- **GEFCom** is a competition that brings together the state of the art techniques and bridges the gap between academic research and industrial practice.

3- Initiatives using ML: Kaggle

- Recently, efforts have been focused on the use os machine learning algorithms to solve forecasting-like problems.
- Those initiatives carry out a significant increase of the performance (lower values of MAPEs), but also less computational time, more flexibility and interpretability.
- The methods used in one forecasting competition:

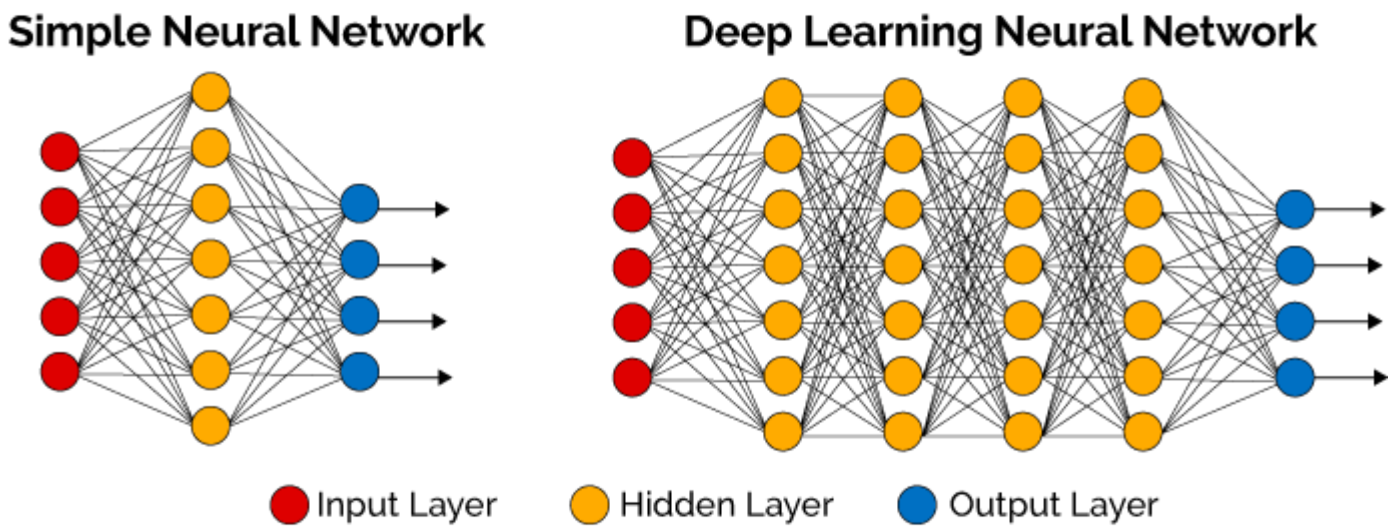


Sales forecasting example where the combination of conventional and modern technics outperforms other approaches.

- **Kaggle** is a platform for data science competitions with a community of more than 600000 data scientist from all over the world, solving difficult real problems.

4- ‘*State of the art’* methods: Definitions

Deep Learning

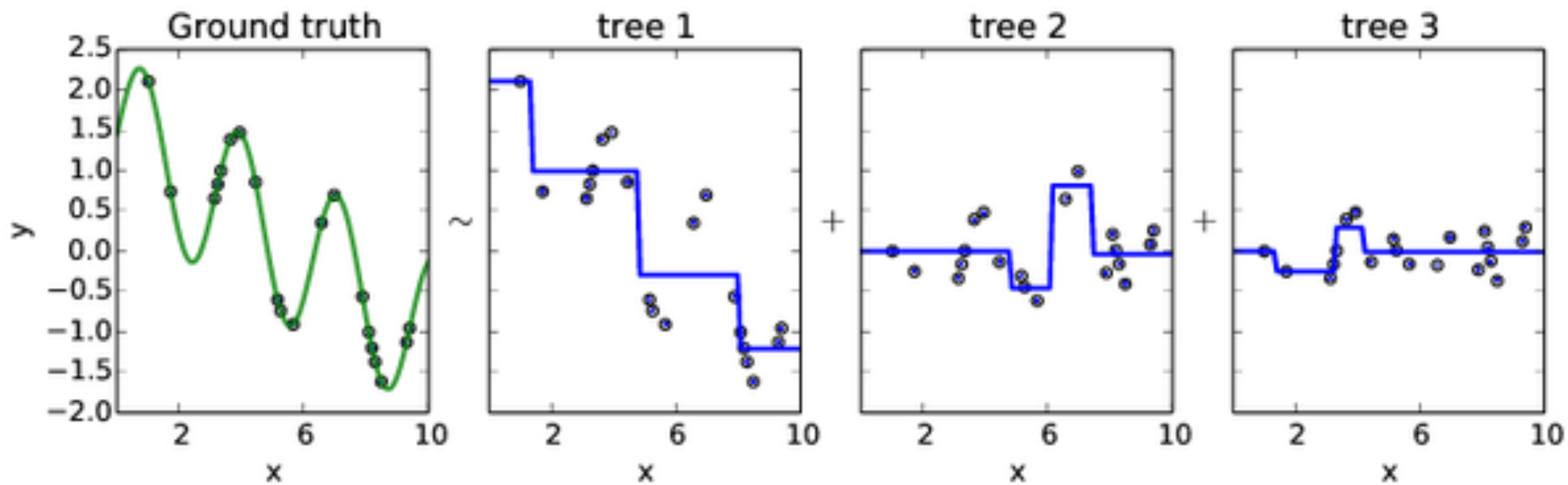


Deep learning:

- is a specific approach used for building and training **neural networks**, which are considered highly promising decision-making nodes.
- An algorithm is considered to be deep if the input data is passed through a series of nonlinearities or nonlinear transformations (NN layers) before it becomes output.
- Deep learning removes the manual identification of features in data and, instead, relies on whatever training process it has in order to discover the useful patterns in the input examples.
- This makes training the neural network easier and faster, and it can yield a better result that advances the field of artificial intelligence.

4- ‘*State of the art’* methods: Definitions

Gradient Boosting Machine



Gradient boosting machine :

- Machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees.
- It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function.
- It is flexible and interpretable framework and allows the possibility of parallelism.

4- '*State of the art'* methods: Example 1

Article

Wind Speed Prediction Using a Univariate ARIMA Model and a Multivariate NARX Model

Erasmo Cadenas [1], Wilfrido Rivera [2,†], Rafael Campos-Amezcua [2,*,†] and Christopher Heard [3,†]

Published: February'2016

Table 5. Statistical errors generated by the ARIMA and NARX models.

Model	La Mata		Metepec	
	MAE	MSE	MAE	MSE
ARIMA	0.91	1.51	0.44	0.39
NARX	0.86	1.35	0.43	0.34
Improved Percentages	5.5%	10.6%	2.3%	12.8%

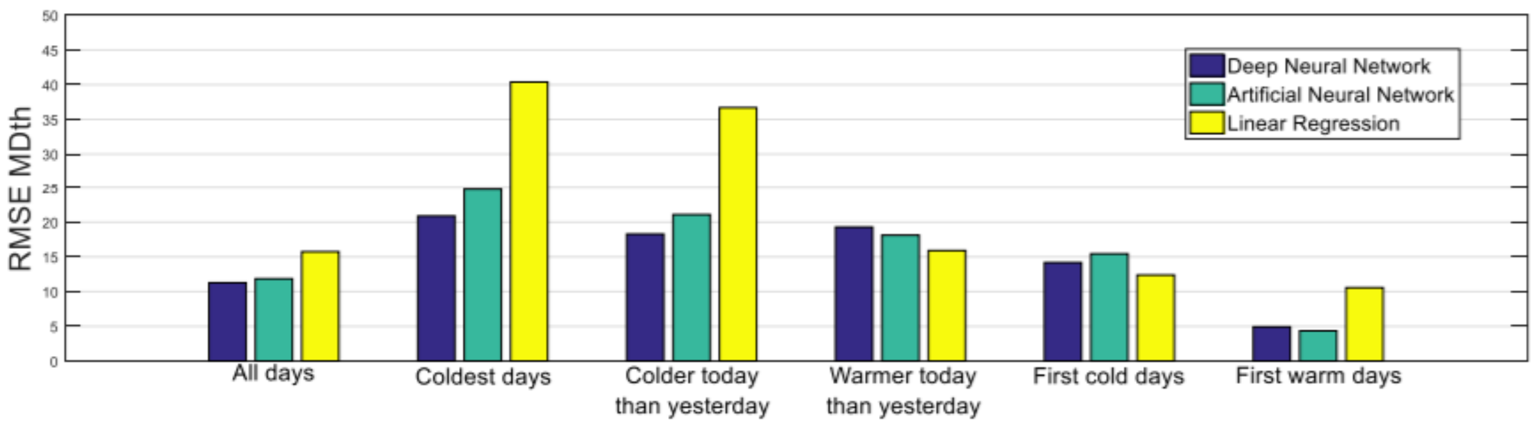
- The results obtained show that using a multivariate NARX model, more accurate results were obtained.
- The inclusion of additional meteorological variables is thus recommended in wind speed forecasting models if they are available.
- As well as being a multivariate model, the NARX neural network is a class of discrete-time non-linear techniques that can represent a variety of non-linear dynamic systems, as in the case of wind speed time series.

4- ‘*State of the art’* methods: Example 2

Deep Neural Network Regression for Short-Term Load Forecasting of Natural Gas

Gregory D. Merkel Richard J. Povinelli Ronald H. Brown

Published: July’2017



- Shows the performance of the median operational area when comparing DNNs to ANNs on several hard to forecast days.
- DNN is performing better than the others except for the situation where the lack of data helps linear regressions.

4- ‘*State of the art’* methods: Example 3

Deep Learning for Solar Power Forecasting – An Approach Using Autoencoder and LSTM Neural Networks

André Gensler, Janosch Henze, Bernhard Sick Nils Raabe

Published: February’2017

Data	P-PVFM		MLP		LSTM		DBN		Auto-LSTM	
	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train
Avg. RMSE	0.1086	0.1148	0.0724	0.0682	0.0724	0.0675	0.0714	0.0663	**0.0713**	**0.0642**
Avg. MAE	0.0560	0.0585	0.0372	0.0344	0.0368	0.0337	0.0367	0.0334	**0.0366**	**0.0323**
Avg. Abs. Dev.	0.4368	0.5123	0.2809	0.2891	0.2786	0.2834	0.2772	0.2813	**0.2765**	**0.2714**
Avg. BIAS	0.0399	0.0463	**-0.0011**	**-0.0031**	-0.0073	-0.0042	-0.0024	-0.0043	-0.0021	-0.0041
Avg. Corr.	0.9294	0.9160	0.9344	0.9361	0.9352	0.9375	**0.9363**	0.9399	0.9362	**0.9431**

Conventional approach

Modern approach

- P-PVFM refers to the physical reference model. Expert system model (no details).
- The Auto-LSTM network is the best performing DNN. It combines the feature extraction ability of the AutoEncoder with the forecasting ability of the LSTM.
- There is a significant gain on the performance of the error comparing the conventional approach (PVFM) and the ML approach using DNN.

4- '*State of the art'* methods: Example 4

Review

Recent Trends in Load Forecasting Technology for the Operation Optimization of Distributed Energy System

Published: August'2017

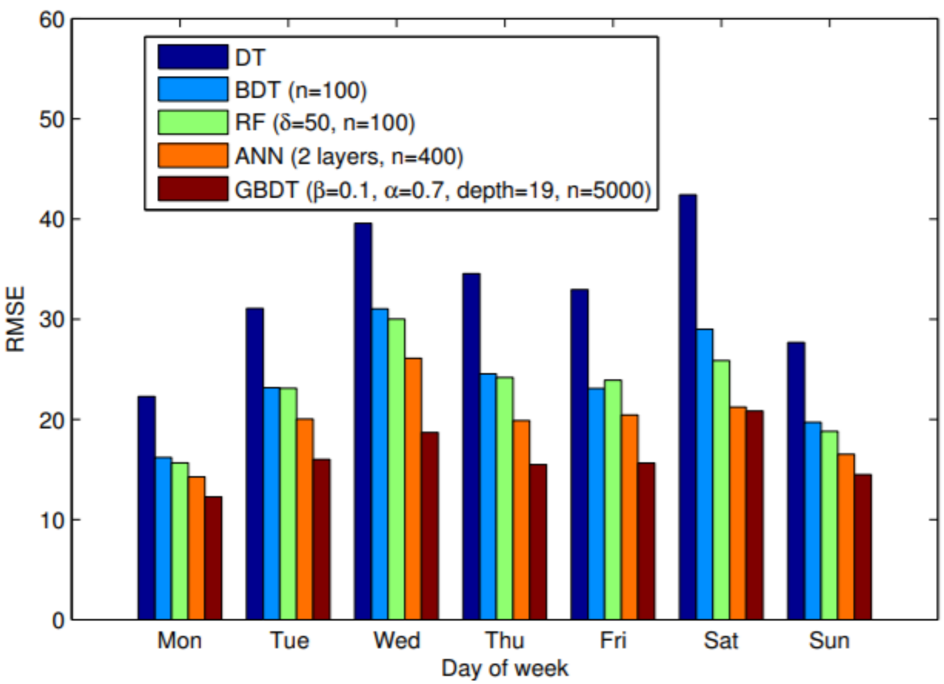
Year	Author	Prediction Model	Load Type	Time Scale
2017	Bianchi et al. [17]	Echo State Networks and PCA Decomposition	electricity	10 min
2016	Gupta et al. [18]	wavelet neural network	electricity	hour
2016	Idowu et al. [19]	support vector machine	heating	hour
2016	Deb et al. [20]	artificial neural network	cooling	day
2015	Chitsaz et al. [21]	Self-Recurrent Wavelet Neural Network	electricity	hour
2015	Protic et al. [22]	SVM-WAVELET	heating	hour
2015	Abdoos et al. [23]	support vector machine	electricity	hour
2015	Bianchi et al. [24]	Recurrent Neural Networks	electricity	hour
2014	Chou et al. [25]	SVM-ANN	cooling and heating	hour
2014	Rodrigues et al. [26]	artificial neural network	electricity	hour
2013	Bacher et al. [27]	time-series	heating	hour
2012	Ilić et al. [28]	artificial neural network	electricity	hour

4- '*State of the art*' methods: Example outside sector

An investigation into machine learning approaches for forecasting spatio-temporal demand in ride-hailing service *

Ismaïl Saadi † Melvin Wong ‡ Bilal Farooq § Jacques Teller ¶ Mario Cools ‖

March 8, 2017



- SVM, after performing some tests, the results revealed that the algorithm was not adapted to handle the current ride-hailing problem and it is, SVM for regression, computationally extremely expensive.
- GBDT clearly provides the best performances in terms of general predictive capabilities.

5- Grupo AIA experience: Energy sector

- **Symbolic forecasting** use functions as input variables for, among others, capturing tendencies, take into account the calendar/laborability, the context, etc. It is based on the historical data and it is trained continuously (stream data). And a key point is that it weights the historical data using exponential smoothing.
  - Advantages: Once the functions are calibrated, predictions have good performance.
  - Disadvantages: Hard to calibrate functions. Need special treatment for every service point. Need some initial period of data to estipulate what are the best functions to use.

- **Conventional methods**: Since the early 1990s, Grupo AIA is using symbolic forecasting technique with great success for different applications. Electric and natural gas forecasting demand applying this methodology in a daily basis to predict a very large amount of service points.

- **Modern methods**: Nowadays, AIA is focusing on applying modern techniques like, GBM for natural gas and LSTM for electric load forecasting, that in combination with the previous approach led to better performance of the predictions.
  - As an example, GBM can handle multiple service points without the need to individualize the calibration. Also, provides the important matrix variables that help in the interpretability of the model showing the most important variable for each SP.

6- Conclusions

- Conventional techniques provide a reasonable load forecasting prediction, but, in general, **machine learning techniques** being one of the latest approaches and **provide greater accuracy** of the forecasts as compared to conventional techniques.

- Furthermore, **modern techniques** have some **advantages** like **interpretability**, **parallelism** which increase efficiency on large datasets and **flexibility**.

- So far, it seems that, **the best approach combines both worlds**, conventional and modern techniques, methodologies since one complements the other.
    - Small data set with ARIMA and linear models and large dataset with non-linear correlations of deep learning and gradient boosting.

HQ Barcelona
Av. de la Torre Blanca, 57
08172 Sant Cugat del Vallès
Barcelona
Tel. +34 93 504 49 00

San Francisco
48 Terra Vista Ave. # D
San Francisco, CA 94115
Tel. 1 415 978 98 00
Fax. 1 415 978 98 10

Washington
1801 Swann Street NW, Apt. 302
Washington. DC 20009

Bibliography

1. EPRI. 2001. “*Day-Ahead/Hour-Ahead Forecasting for Demand Trading: A Guidebook*”. Technical report.
2. A.K. Srivastava et al. “*Short-Term Load Forecasting Methods: A Review*”. ICETEESES. 2016.
3. V. Singh et al. “*An Introduction to Load Forecasting: Conventional and Modern Technologies*”. IRACST-2014.
4. F.M. Bianchi et al. “*An overview and comparative analysis of Recurrent Neural Networks for Short Term Load Forecasting*”. May-2017.
5. I. Saadi et al. “*An investigation into machine learning approaches for forecasting spatio-temporal demand in ride-hailing service*”. Mar-2017.
6. T.Hong et al. “*Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond*”. 2016.
7. Jingrui Xie and Tao Hong. “*Wind Speed for Load Forecasting Models*”. *May 2017.*
8. T. Hong .”*Global Energy Forecasting Competition Past, Present and Future*”. June-2014.
9. E. Cadenas et al. ”*Wind Speed Prediction Using a Univariate ARIMA Model and a Multivariate NARX Model*”. February-2016.
10. W. He. “*Deep neural network based load forecast*”. March-2014.
11. D. Marino et al. “*Building Energy Load Forecasting using Deep Neural Networks*”.2016.
12. G. Merkel et al. “*Deep Neural Network Regression for Short-Term Load Forecasting of Natural Gas*”. July-2017 and reference therein.
13. A. Gensler et al. “*Deep Learning for Solar Power Forecasting – An Approach Using Autoencoder and LSTM Neural Networks*”. 2016.
14. P. Su et al. “*Recent Trends in Load Forecasting Technology for the Operation Optimization of Distributed Energy System*”. August 2017.