Database for storing electrical power signatures of household devices

C.Cernazanu-Glavan*, M. Marcu**
* University "Politehnica" of Timisoara/Computer Science, Timisoara, Romania, cosmin.cernazanu@cs.upt.ro
** University "Politehnica" of Timisoara/Computer Science, Timisoara, Romania, marius.marcu@cs.upt.ro

***Abstract*— This paper describes the storage of electrical signals from a software framework that provides an automated diagnosis of household devices, based on power signature analysis. Storing power signatures of household devices into a database helps us in two ways: it meets the criteria for operation of the automatic diagnosis algorithm and ensures the possibility of performing future benchmarks tests to establish the performance of the algorithm. The initial electrical signal is stored into the database both in normal and in normalized format (according to amplitude). The results for each stage of signal processing and signal analysis are stored in different tables, together with the best results obtained and the methods that led to them. The article describes the creation and maintenance of this database and details the methods of data accessing. A series of stored procedures allows the user quick access to desired data whether he wants to access the full signal or only a part of it.**

I. Introduction

The use of electrical network for collecting, monitoring and analyzing data led to what we call today a smart power grid.[1] The introduction of this term in 2005 by Amin and Wollenberg shaped the main objective of the field: the establishment of a strong connection between the providers of electric energy and the consumers by means of several de dedicated applications.

The aim of these applications consists in supplying specific services for both parts, such as: monitoring the power consumption, finding a most favorable price list and analyzing the electrical signals for finding new information.[3]

In the same time, we witness a development of the household devices which become more and more capable of using several services supplied by the smart power grid.[2] Consequently, they are capable of selecting a price plan that ensures optimum values of the power consumption and supplies information to the user about the current state or the possible changes of programs that may occur in the near future.

Starting from a scenario in which a device has an active role in providing useful information, the authors have suggested and developed a framework which would help to automatically diagnose the problems (the functioning errors) that may occur at the electrical devices.

An automatic diagnosis for the power signal supplied by the current household device can be achieved through an analysis process. It is an attempt of finding irregularities of the power signal that may indicate possible problems or anomalies of functioning.

Power Signature Analysis Database (PSAD) represents a means of storing all the data that will be used by the framework in the process of detecting the anomalies of functioning. Our aim is to store, a long time, all the information supplied by the household device which will be the basis of workflow detecting anomalies. In the future, the information will be used for measuring the performances of several methods and for comparing their results.

This article aims at pointing out both the structures and the tables that will be used by the PSAD and the methods of administrating it. We will present in detail the possibilities of accessing the data and the way in which the storing procedures provide a fast access to the entire signal or only to a part of it.

A. Data collecting

The data that will be stored in PSAD are power signatures of different devices that have been recorded/supplied at various times. The data were obtained by using some smart meters capable of analyzing and measuring different characteristic parameters for the electric energy (strength, amperage, voltage) and having the possibility of sending the gathered information to the utility company.

The most common used monitoring systems supply services for measuring the power consumption with the purpose of distributing the total cost between the electric appliances existing in a house. Kill A Watt or Watts Up Pro devices [7] can be used, because they are easily integrated in the electric network and the measured data are sent wireless.

A solution of this type is described in [5] and consists in obtaining power consumption per device by using a method of decomposition of the final energy into constituent parts. Although the authors have the same goal with ours (obtaining specific data for each electric appliance), the difference resides in the way we will use these data.

In our case, the measuring devices are used distinctly for each household appliance. This solution has already been implemented by the authors and it successfully manages to supply power signatures in real time [8]. The hardware infrastructure necessary for this solution is outlined in Fig. 1. It can be seen that each household appliance is monitored by a measuring device interposed between the appliance and the AC power line.

These measuring devices form together a network of sensors capable of monitoring the total energy consumed in a private residence

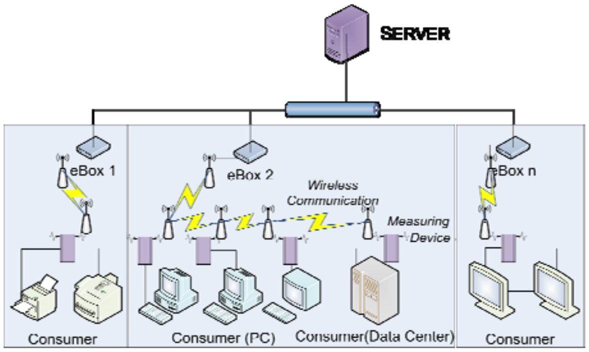

Fig. 1 Overall smart monitoring infrastructure

The data are sent through wireless technology to the server. For a unitary approach, the authors define the notion of node as an ensemble made up of a controller specialized in measuring AC line parameters, a wireless communication device and a microcontroller which controls the whole activity of the node.

Once the information is on the server, it is used by a framework for helping to detect some events that do not normally appear in the workflow of the respective electric device. The workflow of the power signature and failure detection software application is shown in Fig. 2 [4].

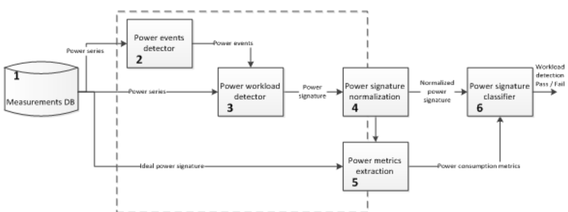

Fig. 2 Power signature and failure detection workflow

It can be seen that PSAD has an important role in the workflow being responsible both for storing the measurements performed by the smart meters in real time and for storing some ideal power signatures (standard signals) supplied by the producer and used in the workload detection. For a better understanding of the data storing method, it is necessary to make a presentation of the power signature.

In the next chapter we will analyze the parameters that are characteristic for the power signatures and the methods of differentiating them. Based on these parameters, we will suggest a data structuring and for each structure we will present its own method of access.

II. Power Signature

The use of power signatures for obtaining some information does not represent a new method of analysis, because power signatures have been widely used for nondestructive testing of integrated circuits and printed circuits boards [6]. The power signature for a device represents a variation in time of its energy consumes. Since in the great majority of time, the household devices are unplugged or in stand-by, the value of power consumption is constant and has a zero value (or close to 0).

The time intervals, we are interested in, are those in which the device performs a certain action (or a pre-established program). Base of these time intervals we can notice some disorders of the electric signal which consist in power events and power workloads. (Fig. 3)

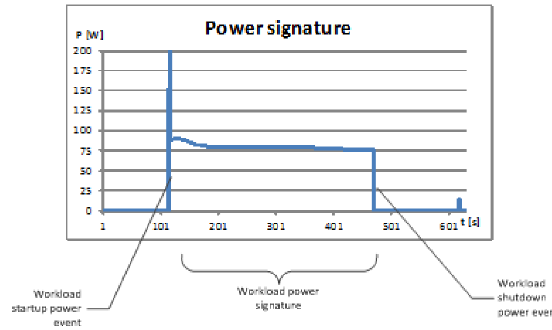

Fig. 3 Power events and power workloads

If we consider the case of a washing machine that executes a program (a washing program consisting in two phases for heating the water, one washing phase, one rinsing phase and one spinning phase), we will see from Fig. 4 that its power signature presents a complex, irregular form.

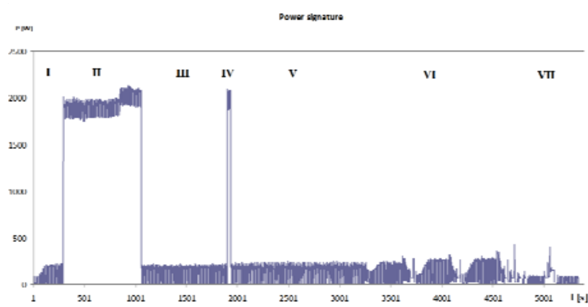

Fig. 4 Power signature for a washing machine

The electric signal is characterized by several parameters. In table I, we can see a selection of its most important parameters that are used for defining and differentiating more power signatures.

TABLE I.
Power Signature Parameters

Number of samples	4591
Duration (sec.)	4591
Amplitude (maximum)	2215 watt
Amplitude (minimum)	0 watt
No. of program phases	7
Device	Washing machine
Type of program	Washing program type 3
PSD (db/Hz)	0.143

We can notice that the majority of the parameters are calculated only after the signal was recorded and stored in the database.

The great majority of the parameters are calculated relatively easy, but there are whatsoever some parameters (No. of program phases and PSD frequency) that require a major computing analysis of the entire signal.

The method through which the signal is stored in the database and the place where its parameters can be found will be presented in the next chapter.

III. STRUCTURE OF DATABASE

A. Why a database and not a file?

A great number of users that work in the field of pattern recognition or use other methods in the same field are accustomed to the existence of databases that allow the performance of some comparisons between several methods and the supply of some relevant results. [9] The databases are text files containing all the data necessary for running experiments.

The work with these files is hard and time consuming and requires separate directories for each dataset and as well, some standard files that must contain data characterizing the signal. This working method is extremely difficult to implement especially when we need more datasets because it implies many search actions in all the existing files and directories.

Another problem refers to the method of access and partition of data. The file storage requires the creation and administration of a group of directories to which only certain users may have access. Another problem arises when a user wants to save the data. The user must have specific access rights and throughout the entire process the data must be invisible for the rest of the users.

Moreover, the methods of processing and analysis of power signature electric signals use a large amount of data. Each signal is characterized by several parameters and it is represented by more forms of signal (normalized signal, smoothing signal, FFT signal). In addition, parts of signal will be extracted for highlightening more methods of splitting the electric signal into components.

In Fig. 5, we can notice the main methods of data processing. This way, there are emphasized the methods of access and the type of data that must be supplied by the database.

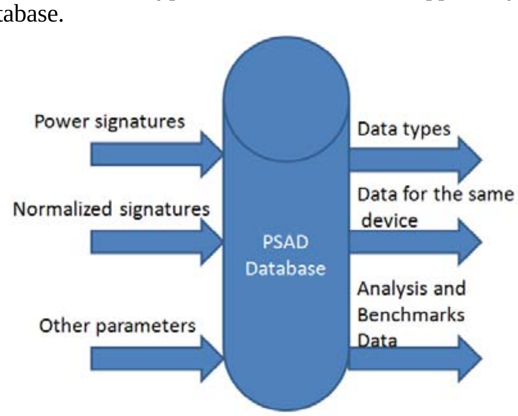

Fig. 5 Methods of accessing the power signature data

The great majority of these methods can be performed only if we use a database for storing the data. The use of files and directories is not adequate in this case.

B. Main tables of Database

PSAD is structured on several layers. In this database there were defined four layers:

- The "Device" layer contains all the data referring to devices. We have information about the manufacturers of the devices and the types of devices used.
- The "Signals_<Device_Id>" layers include all the information about signals. This type of layer is specific for each device, that is why, all the tables included have a suffix containing the id of the device. For each device, there is a number of layers equal to the number of signals recorded for that device. The tables on this layer include information about the nature, samples, normalized samples and the histogram of the signal.
- The "Components_<Device_id>_<Signal_id" layers contain all the information about a certain component of the signal. All the tables on these layers have two suffixes: the id of the device that provided the signal and the id of the signal that contains the components. The tables on this layer contain detailed information about the analyzed component and its samples.
- The "Performance" layer. On this layer there are recorded all the results of all the analyzes and recognition methods that have been applied on signals. On this layer we have information about the nature of the methods applied, the fields that supplied these methods and the date when they were applied. Information about the benchmarks applied is also stored in these tables.

An overview of this database can be observed in Fig. 6 and includes all the tables existing in this database and the connections between them. We can notice a massive interdependency between these tables (we have 14 tables) due especially to the nature of the data stored in them.

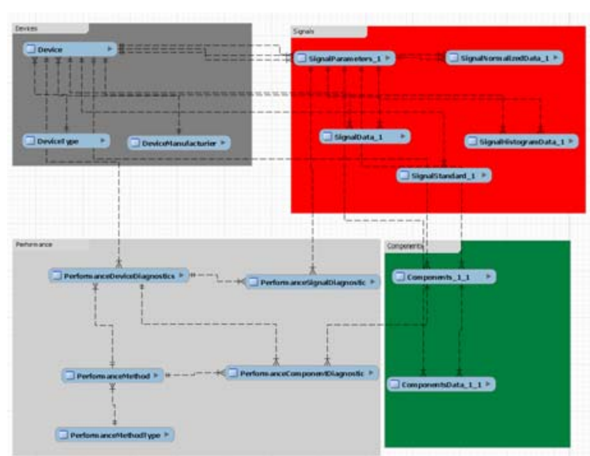

Fig. 6 Overview of the database

1) Device Layer

As a consequence of the fact that the data samples from the signal occupy lots of space, it is necessary to build a set of tables specific for each device. The format of this layer and the tables contained can be seen in Fig. 7.

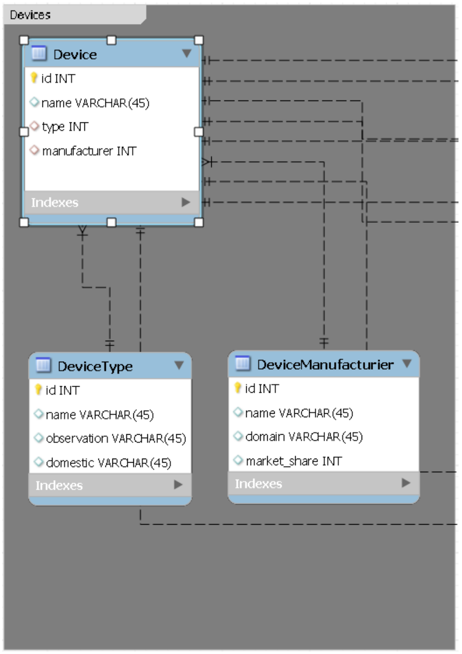

Fig. 7 The "Device" layer structure

The table containing data about the device is "Device". This table contains information about all the devices that supplied the electric signals saved in PSAD. The field "id" from this table will be used as suffix for all the tables existing on the layers "Signals" and "Components". The "DeviceType" table is used for the future interrogations performed in the entire database for finding some similar devices (washing machines, refrigerators, ovens, etc.).

The "DeviceManufacturier" table is used for classifying the devices according to the manufacturer that built them.

This organization helps at finding similar data and allows a better view of the results obtained.

2) Signals Layer

We must underline that a layer of this type is created for each device. From this reason, we can see that all the tables existing on this layer have as suffix the id of the device (Fig. 8 presents the layer for the device having as *id* the value 1)

This is necessary because in this database we want to store some signals belonging to several devices and their number is relatively big. The tables from this layer include information about the signals supplied by a single device.

The main table is "SignalParameters_<ID>" which contains all the data concerning a certain signal. We meet again here a part of the values presented in Table 1 because they characterize a certain signal and can be used for certain selection techniques for signals of the same type.

"SignalData_<ID>" and "SignalNormalizedData_<ID>" are tables containing samples of the signal measured in real time from the current device and the values of the normalized signal.

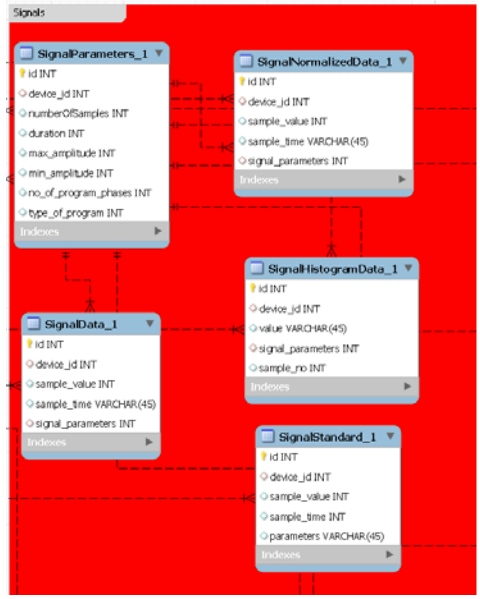

Fig. 8 The structure of the "Signals" layer for the device with *id*=1

The <ID> is the same with the "id" field from the "Device" table and it clearly identifies the device that supplies the signals. The normalization of the signal is performed by taking into consideration a number of 10.000 samples and it is done on the basis of the real signal. The normalization is performed with the help of an external program, at a time ulterior to the process of recording the initial signal.

Another table is "SignalHistogramData_<ID>" which contains the histogram of the original signal, extremely useful for certain applications.

As these signals will be subject to a diagnosis process, we will also need a standard signal (supplied by a producer) which should contain the values a signal must have if the device functioned perfectly. This is the reason why the "SignalStandard_<ID" table is responsible with storing all the information that characterizes this signal and all its samples.

3) Components Layer

As we mentioned in the introduction, a power signature is made up of several components. Each component must be clearly defined for rendering the way in which it was identified by the recognition techniques. We showed that a component is part of a signal, which, in its turn, is specific to a device, consequently, all the tables on this layer have two suffixes: the first suffix is the id of the device for which the signal was measured and the second suffix belongs to the signal out of which the component was extracted. We can see in Fig. 9 a type of layer for components that were extracted from the signal with id = 1 and which, in its turn was recorded as belonging to the device with *id* = 1.

There are two tables on this layer (Components_<Device_id>_<Signal_id> and

ComponentsData_<Device_id>_<Signal_id>) that contain the characteristics of a component and the samples that frame it. The idea of component of a signal is not recognized by all the techniques applied to the power signatures and from this reason, we must specify that this layer may not be consulted by certain applications. In the framework created by the authors, the components of the signal play an important role, that is why, this layer was created.

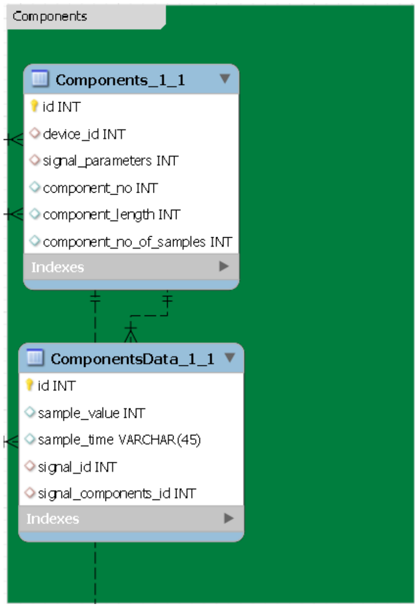

Fig. 9 The structure of the "Components" layer for the device with *id* =1 and the signal with *id* = 1

The main table is Components_<Device_id>_<Signal_id> and it contains the main characteristics of the component: its number, length, importance in the program of the device and the number of samples that contain it.

The samples of the component are saved in the ComponentsData_<Device_id>_<Signal_id> and they can belong to the real or to the normalized signal.

4) *Performance Layer*

When we created this database, our aim was to compare the results of several diagnose methods applied to the signal. Consequently, it was necessary to develop a structure where to store all this information. This layer contains tables in which the following information is stored: the diagnosis set at a certain time for a certain device, the method that was applied for finding this diagnosis, the type of method and the results obtained through this method. (Fig. 9)

The diagnose methods are divided according to the dataset they are working with. Hence, if the method is applied to the entire signal, its performances are saved in the "PerformanceSignalDiagnostic" table and, if the method works only with certain components of the signal, the results are saved in the "PerformanceComponentDiagnostic" table. Each of these tables contains data regarding the performance obtained, the type of diagnosis used and possible observations. The problems that may occur or possible necessary processings are saved, too, because this information is vital for comparing the methods.

If the user wants to perform a benchmark between multiple methods, all that he must do is to select only those methods that he considers necessary by viewing not only the aspects that are common but the different characteristics of each method, too.

The existence of this layer gives vital importance to PSAD, by monitoring the performances on a long time interval and offering the possibility of selecting the best diagnosis method for a pre-established scenario.

Fig. 10 The structure of the Performance layer

C. *Views and stored procedures*

As PSAD contains a large amount of information, their grouping is done through numerous views especially created for helping the user. Thus, there will be selected several signals or components according to the user's preferences.

There are 10 existing views and they allow the selection of data depending on the specific device, its type, its manufacturer, the method applied and the existing components. In addition, these views allow making fast reports and they are useful especially for the experienced users that want to extract some common information out of the entire database.

The stored procedures are used when the user wants to add in the database some information that must be previously processed. The stored procedures perform a data processing and supply new information, especially those that summarize more data.

Their importance resides in finding statistic information that can generally characterize data found in several tables.

D. *Triggers*

As the database is created on a tree-structure (a number of tables containing the signals of the device are dynamically created for each device, and for each signal, other tables are created for saving its components), it is necessary to define some triggers that must perform the "administrative" work.

The triggers are divided in two categories: triggers used for generating the tables and triggers used for deleting these tables. The triggers used for generating the tables build the entire set of tables when a new device is introduced in the database or when a new signal is recorded. Thus, in the first case, we must generate a series of tables that will appear on the "Signals" layer and, in the second case, we must generate a series of tables that will appear on the "Components" layer. In addition, when some information is deleted, other triggers must take care that the remaining tables are consisting to the entire remaining dataset.

The triggers, views and stored procedures ensure an automation level absolutely necessary for the user. Thus, the user is not responsible for the integrity of the database and avoids exhausting work. In the same time, they ensure a fast access to the necessary data and there are eliminated the synchronization problems that appear when several applications run in the same time with the database.

E. Database maintenance

PSAD database runs on a MySQL platform due to its performance and to the fact that one must not pay for a license. The access to the database is based on a username and password. The levels of access and the rights for each user are also specified. For checking its consistency, scripts are running periodically for randomly verify some data.

Unfortunately now the access is limited only for some applications but, we want a promotion of it with the aim of being accessed by other applications too, especially by the ones used for research. In this respect, there are some security problems. As a solution, we want to introduce a way of access only for certain views that supply specific data without the right of accessing the main tables.

IV. CONCLUSIONS

The use of power signatures supplied by the household devices for monitoring and diagnosing certain problems is in continuous growing. For this reason, it is necessary to create a database with more sets of power signatures in order to monitor the performances acquired in this field. We wish PSAD to be a reference mark database for all the applications and methods that use these signals.

Power signatures are electric signals that are monitored for a long period of time and contain a large number of data. As an example, only one device that is monitored for 2h supplies approximately 104 samples that must be stored. In addition, the amount of data associated to a signal increases 5 times during the signal processing. Taking into consideration the number of existing devices and their functioning time, it is absolutely necessary to create a database that supplies specific data for each type of device.

The internal structure of PSAD is based on a number of layers that group the information according to several criteria. On each layer, there is a series of tables dynamically created for storing only one type of information. The access is very simple and it is done with the help of the views, which allow the user to select only the information he needs.

The development of the database is done automatically with the help of triggers and stored procedures and the administrative work of the user is highly reduced. Also, the consistency of the database and its structures is ensured by a periodical checking with the help of some scripts.

The PSAD database is successfully used in many applications, but our aim is to promote it in more research projects. Due to the existence of a layer that monitors the performances of the methods from the diagnosis field and the detection of the failures, this database will be helpful for all the researchers in the field.

A future possible development aims at integrating it to the providers of energy services and to the users so that they could see reports concerning the degree of functionality for the household devices and some future projections about the deterioration in time of devices.

ACKNOWLEDGMENT

This work was partially supported by the strategic grant POSDRU/89/1.5/S/57649, Project ID 57649 (PERFORM-ERA), co-financed by the European Social Fund – Investing in People, within the Sectoral Operational Programme Human Resources Development 2007-2013.

REFERENCES

[1] S. Massoud Amin and Bruce F. Wollenberg, "Toward a Smart Grid", IEEE Power and Energy Magazine, Vol. 3, Number 5, Sep.-Oct. 2005

[2] A. Ipakchi and F. Albuyeh, “Grid of the Future”, IEEE Power and Energy Magazine, Mar.-Apr. 2009

[3] R. Kazman, L. Bass, J. Ivers, and G. A. Moreno, “Architecture Evaluation without an Architecture: Experience with the Smart Grid”, 33rd International Conference of Software Engineering, May 2011.

[4] Marcu M., Cernazanu C., "Dynamic Analysis of Electronic Devices' Power Signatures", International Instrumentation and Measurement Technology Conference (I2MTC), Graz, May 13-16, 2012, ISBN: 978-1-4577-1773-4, pag. 117-122, DOI: 10.1109/I2MTC.2012.6229562

[5] J. Froehlich, E. Larson, S. Gupta, G. Cohn, M. S. Reynolds, and S. N. Patel, “Disaggregated End-Use Energy Sensing for the Smart Grid”, IEEE Pervasive Computing, Special Issue on Smart Energy Systems, Jan–Mar 2011.

[6] I. C. Miller, “IDDQ testing in deep submicron integrated circuits”, Proceedings of International Test Conference, ITC 1999, Atlantic City, USA, Sep. 1999.

[7] K. Fehrenbacher, “10 Monitoring Tools Bringing Smart Energy Home”, Business Week, Apr. 2009, http://www.businessweek.com/technology/content/apr2009/tc20090414446611.htm.

[8] M. Marcu, C. Stangaciu, A. Topirceanu, D. Volcinschi, and V. Stangaciu, “Wireless Sensors Solution for Energy Monitoring, Analyzing, Controlling and Predicting”, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Volume 57, 2011.

[9] A. Frank and A. Asuncion, "UCI Machine Learning Repository", Irvine, CA, University of California, School of Information and Computer Science, 2010, url:http://archive.ics.uci.edu/ml