IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol-6, Special Issue-6, July 2018 E-ISSN: 2347-2693

Energy Prediction Using Data Analytics in Smart Grid

Panchami Anil[1*], Anas P V[2], Naseef Kuruvakkottil[3], Anusha K V [4], Balagopal N[5]

[1,2,3,4]Computer Science and Engineering, NSSCE, Calicut University, Palakkad, India
[5]Computer Science and Engineering, NSSCE Palakkad, India

**Corresponding Author: panchuanil89@gmail.com, Mob: 9497161658*

Available online at: www.ijcseonline.org

Abstract—A fully automated systemwhere embedded large pools of sensors in the existing electricity grid systems for monitoring and controlling it by making use of modern information technology is what is known as Smart Grid. By deriving and processing new information from these data in real time it can be made more applicable. Energy consumption prediction, which is a significant part of smart grid, may be difficult to handle with huge energy usage data in the grid. This is because the redundancy from feature selection cannot be avoided. Our aim is to predict the commercial energy consumption by a building based on its previous consumption history. First, we apply a correlation based feature selection method in order to filter out the most relevant attributes. Out of the resulting dataset so formed, for the purpose of dimensionality reduction we use a Kernel Principle Component Analysis methodology. What we obtain will be a set of principal components which will be our new dataset. To predict the energy usage, we use a Support Vector Regression method that uses kernel technique that determines a suitable point as the predicted value. Finally, we evaluate the performance of the predictor based on different evaluators to understand the efficiency of the technique.

Keywords—Smart Grid, Energy Consumption, Correlation Based Feature Selection, Kernel Principal Component Analysis, Support Vector Regression, Prediction

I. Introduction

There are many challenges faced by modern power systems in the management of power. For the purpose of handling these challenges, the smart grid concept has been introduced to make the energy systems completely automatic. The smart grid plays an important role to ensure the production of energy efficiently according to the customer demands. It ensures the production of energy efficiently and according to user requirements. It plays the role of a bridge for bringing together the power consumers, suppliers and the power producers. An excellent accuracy in short-term energy forecasting which include daily forecasting and hourly forecasting is required by energy suppliers for getting good prices for stored energy. The power suppliers had to predict the next month consumption with good precision in order to obtain good prices for traded energy.

The energy consumption patterns of different households show high variance due to the fact that their energy consumption decision arrival are usually affected by different factors. Utilities do functions in a better way such as better knowledge about behaviour of customer, savings,usage and requirement, downtime and power failures tracking etc..with the help of largeamount of data. Advanced analytics to transform data collected to information, then to knowledge and finally to actionable plans are present. For power companies, more timely, flexible and personalised marketing strategies or demand side management measures can be developed. Through the real-time interaction with the electricity company, power consumerscan adjust and optimize their electricity consumption behaviours, thus reducing their expenses.

II. Related Work

There are some data mining and research studies on fraud identification and prediction techniques have been carried out in the electricity distribution sector combination of Support Vector Machines (SVMs) is utilized [1].A cluster based classification method, which uses a set of five attributes, may also be used. The losses in a power system due to physical laws are called a technical loss. This canalso be due to unauthorized connections, non-updated database, non-measured energy consumption, fraud in measuring devices or electricity theft [2].

The process of detecting fraud contains three main steps [3]. The first step involves Cleaning and Integration.. The next step is Selection and Transformation.A method to distinguish office building energy consumption as abnormal consumption and normal consumption is adopted. It then classifies the energy consumption behaviours at normal offices and so called hierarchical classification [4]. Normal consumption can further be classified as low power consumption that can have values which are small or even close to zero.

The approach detects electricity theft at different levels which are transmission, distribution, and consumer levels [6]. If the computed value of total power transmitted exceeds the power received, then the server considers it as a theft. First, they identified four key periods which described different peak demand behaviour, concurring with common interval of the day: overnight, breakfast, daytime, and night and also found that demand in the time periods changed as a function of seasonality and days of the week, identifying two major sources of variation[7].

A Grey Correlation Analysis (GCA) based Hybrid Feature Selector is used, combining relief F algorithm and random forest (RF). For feature extraction, Kernel Principal Component Analysis (KPCA) is used further to reduce redundancy. Then a differential evolution (DE) SVM is designed to tune the super parameters of SVM [9]. The parameters considered are the mean, variance, kurtosis and skewness of the 24-hour energy consumption per day [10].

A building cooling load forecasting approach combining kernel principal component analysis (KPCA) and support vector machine (SVM) is proposed to ensure good prediction. The original inputs are firstly transformed into nonlinear principal components using KPCA. The inputs of SVR to solve the load forecasting problem use these new features [11]. A Hybrid of Supervised Correlation method and Support Vector Machine for classification of high dimensional dataset is also adopted [12].

III. METHODOLOGY

The general architecture of the system will be as follows.

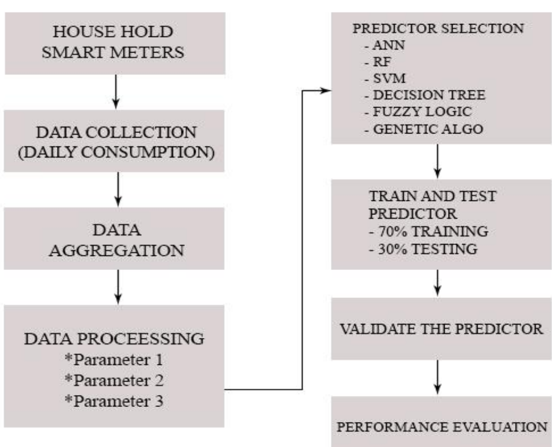

Figure. 1. System Architecture

A. Smart Meters

Smart meters are installed at the customer premises that are used by the smart grid in order to get the timely accurate measurements. An electronic device whichnotes consumption of electric energy in intervals of an hour or at least daily and conveys that information back to the utility centre for monitoring and billing is called a smart meter. More frequently than analog meters, which require a meter reader to transfer data, digital meters can send energy consumption details. Unlike home energy monitors, smart meters can gather data for remote reporting. Smart meters will help to view consumption more accurately, so more informed energy choices are made. Depending on the feature set, the meter may also acknowledge the distributor of a power flow or allow them to remotely switch electricity service on or off.

B. Data Aggregation

The daily consumption data is collected and put for aggregation. Data is searched, gathered and presented in a summarized, report-based format to achieve a particular target or processes and/or conduct human analysis in the data aggregation process, which is a kind of data and information mining process. Aggregated data can be the basis for extra calculations, combined with other datasets, used in any of the way that other data is used.

C. Data Preprocessing

It prepares raw data for further processing. Modifying the raw data into an understandable data formis included in a data mining technique called data preprocessing. Real-world data will be incomplete, inconsistent, or lacking certain behaviours. Data cleaning, integration, transformation, reduction, and discretization aresteps that it goes through.

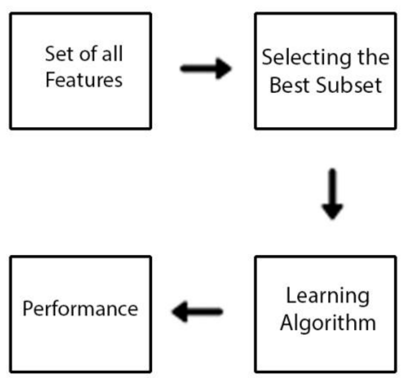

Figure. 2. Phases in the System

Correlation based feature extraction

To reduce the dimensionality attribute subset selection on the basis of relevance analysis is one method. The attributes (redundant) that do not have significant contribution in the characteristics of whole data of concern are detected and by means of correlation analysis, relevance analysis of attribute. After which the redundant attribute or attribute strongly correlated to some other attribute is disqualified to be the part of DW.

A feature Selection algorithmheuristically calculates the correlation between the attributes and feature subsets in which each feature is highly correlated with the class and uncorrelated with other subset features are rewarded.
The correlation between each attribute and the output variable is calculated. Those attributes with a low correlation (value close to zero) are removed andonly those features that have a moderate-to-high positive or negative correlation (close to -1 or 1) are chosen.

The core of the CFS algorithm is to evaluate a feature on its worth or merit. It considers the influence of features on predicting the class label together with the inter-correlation between each feature. Subset which contains features highly correlated with the class and uncorrelated with each other is the result of this algorithm.

Kernel Principal Component Analysis

The standard PCA cannot be used for non-linear principal components . To represent the data in lower dimension it always finds linear principal components. At some occasions, non-linear principal components are needed. It always finds linear principal components. It will fail to find good representative direction if we apply standard PCA for such data. Kernel PCA (KPCA) solves this disadvantage. PCA is performed in a new space with the help of Kernel PCA. Principal components in different space (Possibly High Dimensional Space) are found by using kernel trick.

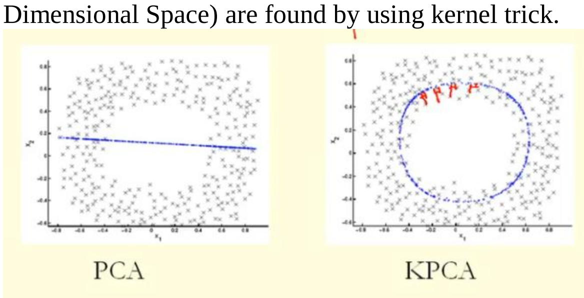

Figure. 3. Comparison of PCA and Kernel PCA

New directions based on kernel matrix are found by Kernel PCA. n (number of observations) eigenvalues can be extracted. Using few eigenvectors from total P eigenvectors.pre-image is allowed to reconstruct by PCA. In KPCA, It may not be possible. Extracting principal components by KPCA takes more time compared to Standard PCA when checking its computational complexity.

D. Predictor SelectionSupport Vector Regression

Maintaining all the main features that characterize the algorithm (maximal margin), Support Vector Machine can be utilized as a regression methodology. With only a few minor differences, the Support Vector Regression (SVR) utilizes the same ideologies as the SVM for classification,. First of all, it will be very difficult to predict the information easily,since output is a real number, which has many possibilities. The main idea is to ensure that the error is minimized, the hyperplane is individualizedsuch that the margin is maximized, noting that part of the error is tolerated.
To perform the linear separation, the kernel functions transform the data into a higher dimensional feature space,One of the most important ideas in Support Vector Classification and Regression cases,is presenting the solution by utilizing small subset of training points provides enormous computational advantages. The input X is first mapped onto a *m*-dimensional feature space using some fixed (nonlinear) mapping, and then a linear model is built in this feature space in the Support Vector Machine (SVM) regression.

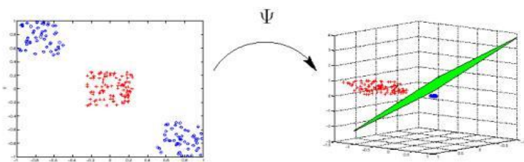

Figure. 4 Mapping to Larger Space

SVM has a technique called the kernel **trick**. These are functions that take low dimensional input space and transform it to a higher dimensional space. That is, separable problem converted by not separable problem, and these functions are called kernels. It is mostly useful in non-linear separation problem. Briefly, it does some very complex data transformations, and then determines the process to separate the data according to the labels or outputs that are defined.
The aim of a SVM is to determine where that particular hyperplane in which separates both these classes with minimum error together with also ensuring that the perpendicular distance between the two closest points from each of these two classes is the maximum, when have a linearly separable set of points of two different classes,.

The cost function includes using a kernel, which may be linear/Gaussian/polynomial based on the choice. The kernel identifies how similar different features are with respect to each other, and henceallots weights to their corresponding cost functions. So features that are closely related to each other and have the same output will get grouped together because of more weight, while the outliers will have lesser weight associated with them and get left out when we are trying to minimize our optimization objective for classification. In case of regression, the effects of these weights is somewhat similar. Contribution by outliers to the model will be very less.

E. Testing

Initially data is seperated into a training set and testing set, majority of the data is used for training, and a smaller part of the data is used for testing.Test the model by making predictions against the test set, after a model has been processed by using the training set. The "training set" is the set of data which allows the training During the training of a network the same set of data is processed many times as the connection weights are ever refined. Now, it is put for testing.

F. Performance Evaluation

A basic concept of machine learning is evaluating the performance of the data mining technique used. For understanding the quality of the model or technique,the evaluation is necessary, for choosing the most acceptable model or technique from a given set of models or techniques and for refining parameters in the iterative process of learning. For evaluating models for different tasks, there are many criteria that includes computational complexity or the comprehensibility of the model.

Explained variation

The proportion to which a mathematical model reasons for the variation (dispersion) of a given data set is measured. *X* is said to "explain" variance in *Y* even though *X* does not really cause *Y, if X* is correlated with *Y.*

A set of scores on outcome variable is present. The variance of that set of scores is then calculated. The measure of difference in the number is called as the variance which is the total variance and then fit the regression model. To calculate a predicted score, use the regression equation. The difference between the predicted scores and the actual scoresis then found. Then calculate the variance of the set of scores. It's the residual variance. The value of residual variance will be less than that of the total variance.

Explained variance = (total variance - residual variance) (1)

The proportion of explained variance is therefore: Explained variance / total variance. There are perfectly predicted the scores, and there is explained all of the variance,if the predicted scores are exactly matching the outcome scores. The residuals are all zero.

Root Mean Square Logarithmic Error

The ratio between actual and predicted is measured by RMSLE.

log(pi+1)−log(ai+1)log(pi+1)−log(ai+1) can be written as log((pi+1)/(ai+1)) (2)

When we don't want to penalize huge differences it can be used when both the values are huge numbers. When we needto penalize under estimates more than over estimates, then also it can be used.

Coefficient of determination

The proportion of the variance in the dependent variable that is predictable from the independent variable(s), is called the coefficient of determination, denoted *R*2 or *r*2 and pronounced "R squared". It is the percentage of the response variable variation that is explained by a linear model. Or:

R-squared = Explained variation / Total variation (3)

R-squared is always between 0 and 100%. 0% denoted that the model does not explains any of the variability of the response data around its average. 100% implies that the model explains every variability of the response data around its average. In brief, the higher the R-squared, the model fits the data the better.

R-squared does not denote whether a regression model is enough. One may have a lower R-squared value for a good model, or a higher R-squared value for a model that can not fit the data.

IV. Results and Discussion

A. Dataset Description

The data comprises of hourly energy consumption readings by various equipment including cooling, heating, lighting etc…, taken each day of a year. The output data that tells the total consumption details is in the Electricity column. That is the column thus that is compared with in case of performance measurement.Nine months' data will be fed as training set and the rest three months' data will be taken as the testing set. The dataset consists of twenty-four hour readings of a whole year.So the accuracy of prediction can be ensured to an extent. Here is a part of the dataset.

Table 1 Commercial Energy Consumption (Hourly)

DATE/TIME	ELECTRICITY	FANS: ELECTRICITY	COOLING: ELECTRICITY	HEATING: ELECTRICITY	INTERIOR LIGHTING
01/01 01:00	22.03598	3.586221	0	0	4.589925
01/01 02:00	14.64976	0	0	0	1.529975
01/01 03:00	14.66957	0	0	0	1.529975
01/01 04:00	14.67781	0	0	0	1.529975
01/01 05:00	14.82479	0	0	0	1.522795
01/01 06:00	22.18265	3.586221	0.013197	0	4.589925
01/01 07:00	38.13185	3.586221	0.007371	0	9.179851
01/01 08:00	45.59728	3.586221	0.00745	0	9.179851
01/01 09:00	45.60365	3.586221	0	0	9.179851
01/01 10:00	38.11376	3.586221	0	0	9.179851

This commercial electricity consumption dataset contains 8700 measurements gathered between January 1st and December 31st (1 year) in all the 24 hours.

B. Result Analysis

To develop the KPCA model, 12 variables are selected and hourly samples interval is set. Data is collected in the normal operation during the 24 hours simulation time, with total 510 samples. The linear function is also used as the kernel function of SVM. One-against-all SVM classifiers are trained for the feature vectors extracted by the above PCA and KPCA algorithms respectively. We obtain the best (C, σ) is (1e3, 0.1).The comparison between KPCA and PCA is given in figure, from which we can see the prediction accuracy of KPCA is better than that of PCA and is not influenced by the number of Principal Components.

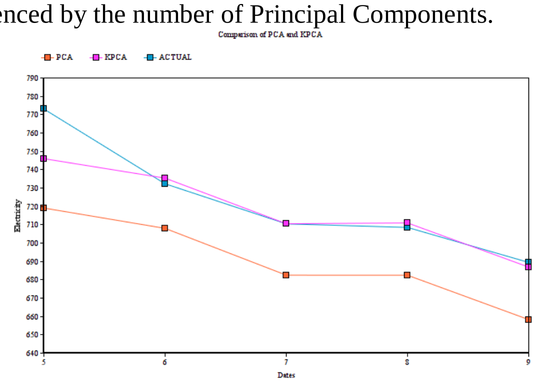

Figure 5 Prediction Results Using PCA and KPCA

To validate the effect of the model based on KPCA-SVR, the model based on PCA-SVR and SVR were also established. Root Mean Square Logarithmic Error (RMSLE), Explained Variation and Coefficient of Determination were defined as performance indices. The fault recognition performance of three models is summarized in table below.

Table 2. Performance Measurements

Method	RMSLE	Explained Variation	R^2
PCA-SVR	0.0018	0.8435	-0.1812
KPCA-SVR	0.0002	0.8435	0.8152

The experiments results show that the Prediction error of KPCA-SVM is the smallest. SVR by feature extraction using KPCA performs much better than that without feature extraction. And there is also superior performance in the KPCA than the PCA. High learning speed, good approximation and generalization ability is possessed by KPCA-SVR features compared with SVM and PCA-SVR. The KPCA-SVR based model may efficiently forecast load. While comparing with the above mentioned models, we are also in a situation that compares the model with the one using a probabilistic PCA. A dimensionality reduction technique that analyses data through a lower dimensional latent space is called Probabilistic principal components analysis (PCA). The performance measurements related to it is also carried out so as to conclude that the proposed method is the best among all.Similarly, the prediction results of the same dataset have been analysed with other prediction techniques as well. These techniques include a linear regression model and a decision tree regressor. But when compared with the actual results, it has been observed that the results are closer when we use the support vector regression technique. The plot below shows the variation.

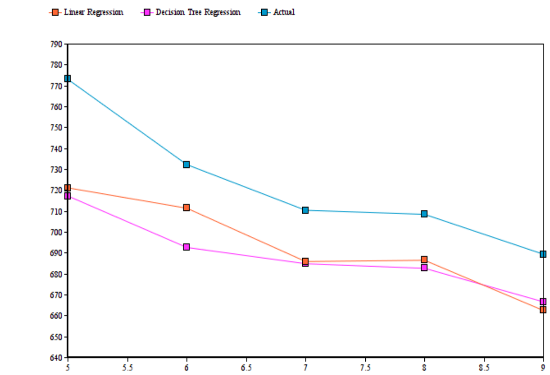

Figure 6 Prediction Results with Different Models

The evaluation measures also show better results when we use these techniques. All the three measures including RMSLE, Explained Variation, and R2 give values near to 0 in SVM. This is favourable to the required accuracy. The table given below illustrates the obtained results

Table 3 Evaluation Measures with Different Models

\Method	RMSLE	Explained Variation	R^2
SVM	0.0002	0.8435	0.8152
Linear Regression	0.0018	0.8376	-0.2071
Decision Tree	0.0025	0.8115	-0.5949

Briefly, a combination of Correlation Based Feature Selection and KPCA gives the best set of features and SVM gives accurate prediction results.

V. CONCLUSION

The main focus of the paper is to build a model for the prediction of electricity consumption of commercial apartments which will favour the system to organize energy production and consumption. It will also help to decide whether energy consumption of these buildings will be high or low (energy management). This will also be useful for the energy management system in fixing the prices of energy according to consumption of different apartments.
KPCA algorithm can solve the problem of nonlinear characteristic that the PCA algorithm can't handle with. It has a good effect on extracting minute detailed information through internal nonlinear kernel function.Support Vector Machine (SVM) has the strong ability of classification of small samples and the advantage of dealing with nonlinear and high dimension. These together form the benefits of the model over the other methodologies.
Huge quantity of datasets is collected to bring smartness to the grid. In the same time this possess challenges for utilities to manage with the nature, the distribution and the real-time constraints of the datacollected. In this paper we have

presented an overall view of the opportunities, concepts and challenges of data management in smart grids and summarized the Big Data technologies and methodologies that can be utilized to manage smart grid requirements including processing, storage and even visualization.

ACKNOWLEDGMENT

We would like to extend our sincere gratitude to our Principal Mrs. Sudha T for providing us with a good learning environment. We are deeply indebted to Associate Professor Dr.VijiRajendran V, Head of the Department of Computer Science and Engineering, N.S.S College of Engineering, Palakkad, for providing and availing all the required facilities for undertaking the project in a systematic way. We are thankful to our guide, Assistant Professor Mr. Balagopal N, for guiding us and providing good suggestions to improve the project. We are also thankful to our staff in charge Assistant Professor Mr. Anuraj Mohan, and the members of the evaluation committee Associate Professor Mrs. Sindhu.S and Assistant Professor Mrs. Maya Mohan, for providing good suggestions to improve the project. Gratitude is extended to all teaching and non-teaching staffs of Department of Computer Science and Engineering, N.S.S College of Engineering, Palakkad, for the sincere directions imparted and the cooperation in connection with the project.

We are also thankful to our parents for the support given in connection with the project. Gratitude may be extended to all well-wishers and friends who supported us.

REFERENCES

[1] Nagi, Jea, K. S. Yap, S. K. Tiong, S. K. Ahmed, and A. M. Mohammad. "*Detection of abnormalities and electricity theft using genetic support vector machines*." TENCON 2008-2008 IEEE Region 10 Conference, pp. 1-6. IEEE, 2008

[2] Angelos, Eduardo Werley S., Osvaldo R. Saavedra, Omar A. Carmona Cortés, and André Nunes de Souza. "*Detection and identification of abnormalities in customer consumptions in power distribution systems."* IEEE Transactions on Power Delivery 26, no. 4 (2011): 2436-2442.

[3] Costa, Breno C., Bruno LA Alberto, André M. Portela, W. Maduro, and Esdras O. Eler. "*Fraud detection in electric power distribution networks using an ANN-based knowledge-discovery process.*" International Journal of Artificial Intelligence & Applications 4, no. 6 (2013): 17.

[4] Bu, Li, Dongbin Zhao, Yu Liu, and Qiang Guan. "*A hierarchical classification algorithm for evaluating energy consumption behaviors.*" Neural Networks (IJCNN), 2014 International Joint Conference on, pp. 1461-1466. IEEE, 2014.

[5] Mayhew, Michael, Michael Atighetchi, Aaron Adler, and Rachel Greenstadt. "*Use of machine learning in big data analytics for insider threat detection*." Military Communications Conference, MILCOM 2015-2015 IEEE, pp. 915-922. IEEE, 2015.

[6] Jindal, Anish, Amit Dua, Kuljeet Kaur, Mukesh Singh, Neeraj Kumar, and S. Mishra. "*Decision tree and SVM-based data analytics for theft detection in smart grid."* IEEE Transactions on Industrial Informatics 12, no. 3 (2016): 1005-1016.

[7] Haben, Stephen, Colin Singleton, and Peter Grindrod. "*Analysis and clustering of residential customers energy behavioral demand using smart meter data*." IEEE transactions on smart grid 7, no. 1 (2016): 136-144.

[8] MCA, Mrs J. Sukanya. "*Applications of Big Data Analytics and Machine Learning Techniques in Health Care Sectors*." International Journal Of Engineering And Computer Science 6, no. 7 (2017).

[9] Wang, Kun, Chenhan Xu, Yan Zhang, Song Guo, and Albert Zomaya. "*Robust big data analytics for electricity price forecasting in the smart grid*." IEEE Transactions on Big Data(2017).

[10] Wahid, Fazli, RozaidaGhazali, Abdul Salam Shah, and Muhammad Fayaz. "*Prediction of energy consumption in the buildings using multi-layer perceptron and random forest.*" IJAST 101 (2017): 13-22.

[11] Xuemei, Li, Ding Lixing, LvJinhu, Xu Gang, and Li Jibin. "*A novel hybrid approach of kpca and svm for building cooling load prediction*." Knowledge Discovery and Data Mining, 2010. WKDD'10. Third International Conference on, pp. 522-526. IEEE, 2010.

[12] Dubey, Vimal Kumar, and Amit Kumar Saxena. "*Hybrid classification model of correlation-based feature selection and support vector machine*." Current Trends in Advanced Computing (ICCTAC), IEEE International Conference on, pp. 1-6. IEEE, 2016.

[13] Ince, Huseyin, and Theodore B. Trafalis. "*Kernel principal component analysis and support vector machines for stock price prediction*." IIE Transactions 39, no. 6 (2007): 629-637.

[14] Michalak, Krzysztof, and Halina Kwasnicka. "*Correlation-based feature selection strategy in neural classification.*" Intelligent Systems Design and Applications, 2006. ISDA'06. Sixth International Conference on, vol. 1, pp. 741-746. IEEE, 2006.

[15] Kallas, Maya, Gilles Mourot, Didier Maquin, and José Ragot. "*Fault estimation of nonlinear processes using kernel principal component analysis.*" Control Conference (ECC), 2015 European, pp. 3197-3202. IEEE, 2015.

.

