ADVANCED ANALYTICS APPLIED TO PROBLEM SOLVING TECHNIQUE ON A PULP MILL

F. DE CARLI†‡, R.I.G. MEJIA‡ and G.A.T. DE ARAÚJO†

† *Gerencia de Excelencia Operacional, CMPC Celulosa, Av Julio Hemmelmann 320, CHILE.*
felipe.decarli@cmpcrs.com.br, guilherme.araujo@cmpcrs.com.br
‡ *Technological Institute, Univ. do Vale do Rio dos Sinos, Av. Unisinos, 950, BRASIL.*
rmejia@unisinos.br

***Abstract*— In this article, advanced analytics will be applied to the lean 8 step problem solving technique aiming to increase the data analysis reliability. The modelling techniques consider in this work are boosted and bagging decision trees, artificial neural networks, relief and stepwise regression. The proposal is validated by the evaluation of a bleaching process problem on a pulp mill. The high pressure events on bleaching washer feeding stage resulted on bleaching non planned shutdowns. The lean 8 steps problem solving technique allied to advanced analytics modeling tools reduced the high pressure events on 63%. As a result, the plant reduced its loss of production due to bleaching plant and allowed to increase the pulp production rate. On this example the relief, decision trees and stepwise regression algorithms proved to be a well-tuned packet for problem solving in a pulp mill.**

***Keywords*— Lean Manufacturing, Problem Solving, Advanced Analytics, Process Modelling, Pulp Mill.**

I. INTRODUCTION

The availability of natural resources has been reducing year by year, causing an increase of direct cost for many companies. This impact is higher in commodities company, where the cost is one of the most important issue. This scenario shows a growing demand for management systems driven by continuous improvement and problem solving. Lean manufacturing is usually accompanied by a shift towards exposure and problem solving (Forrester, 1995). Working as groups, while utilizing appropriate problem-solving techniques, it will increase efficiency in work improvement outputs (Gatchalian, 1997). A typical lean technique for trouble shooting is the 8 steps problem solving methodology. The problem solving techniques can vary slightly according to the industry segment and problem complexity.

In present case for complex problems, it is used the following steps sequence (Fig. 1).

The problem-solving efficacy among the organization is also related to the lean principles that drive the lean transformation. According to (Puvanasvaran *et al.*, 2008) these principles are meant to provide a framework to focus the direction in enhancing problem solving capability among employees by forming as people development system in lean process management.

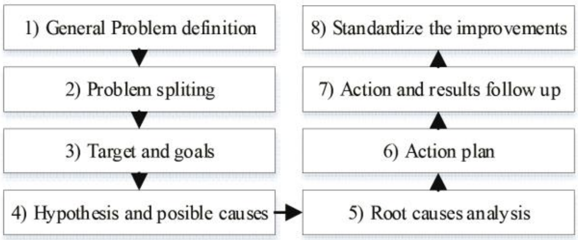

Figure 1. 8 steps problem solving technique.

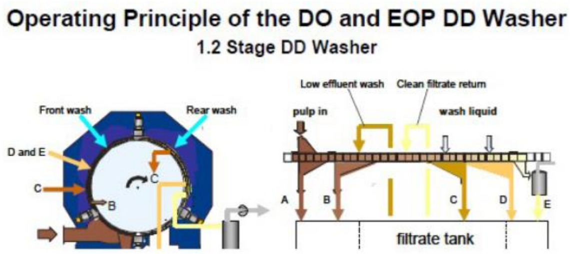

Figure 2. Displacement drum washer (DDW). Adapted from Andritz operational manual (Andritz, 2017).

Considering one of the principles, the scientific approach, this paper aims to use advanced analytics tools on step 4 (Hypothesis and possible causes). As a result, it is expected that this approach can provide a more efficient use of data on the actual problem solving technique.

For this purpose, a couple of non-parametric and parametric techniques are applied to a real problem solving session. The study case is the non-planned shutdowns of Eop bleaching stage washer of a pulp mill.

II. BACKGROUND

At the beginning, it is desirable to review the main processes principles corresponding to the study case and problem associated.

A. EOP stage on bleaching process

Eop stage is responsible to extract the residual of lignin, which has reacted in the first stage. Also, it has a second function, which is increase brightness using oxygen and hydrogen peroxide. From the Eop reactor top, pulp is discharged with help of the Eop reactor flow discharger and blown to the Eop stage DD washer. Hot water and acid filtrate (D1 stage) are used for pulp washing. From the Eop DD washer pulp drops to the Eop DD washer discharge standpipe.

The main purpose of Eop DD washer is remove all residual of lignin from pulp solution and send it to the effluent treatment (Fig. 2).

The washer operates with the following loop controls:

1. Pulp feed inlet flow: Control the feeding pulp flow from the Eop reactor by adjusting the washer inlet valves. If the flow setpoint increases the valves increases the opening.
2. Inlet filtrate pressure: It controls the filtrate pressure, which flows from the D1 stage to the DDW showers using changing the setpoint of the inlet filtrate flow loop control. If the pressure increases the flow setpoint decreases.
3. Inlet filtrate flow: Control the inlet filtrate flow using the filtrate inlet valves, it operates as a slave control for the inlet filtrate pressure. If the setpoint increases the valve increases the opening.
4. Filtrate level control: Control the vacuum vase level using the filtrate outlet valve. If the valves open the level reduces.
5. Pulp feed inlet pressure: Control the washer feeding stage pressure. If this pressure increases above the higher limit the washer trips.

The problem is associated to the pulp feed inlet pressure loop control, which many times suddenly increases reaching the DD washer interlock value. Between 2015 and 2017, the high pressure feeding events frequency was 13.7 events/month. These events cause bleaching shutdowns and production losses in the mill. For instance, the theoretical pulp production losses sum the amount of 9761 tons per year, based on 2016 OEE (Overall Equipment Effectiveness). Furthermore, this problem is a well-known bottleneck for mills pulp production.

Given this, it is proposed a literature review regarding the process modelling techniques applied to the Eop DD washer problem solving session.

B. Adaboost decision trees

The Adaboost algorithm developed by Freund and Schapire (1996) is a non-parametric modelling technique. It maintains a set of weights over the original training set S and adjusts these weights after each classifier is learned by the base learning algorithm. The adjustments increase the weight of examples that are misclassified by the base learning algorithm and decrease the weight of examples that are correctly classified.

There are two ways that Adaboost can use these weights to construct a new training set S' to give to the base learning algorithm. In boosting by sampling, examples are drawn with replacement from S with probability proportional to their weights. The second method, boosting by weighting, can be used with base learning algorithms that can accept a weighted training set directly. With such algorithms, the entire training set S (with associated weights) is given to the base learning algorithm. Both methods have been shown to be very effective (Quinlan, 1996).

C. Bagging decision trees

Bootstrap aggregating (bagging) algorithms are applied for decision trees resulting in non-parametric models. It is often compared to boosting technique. Neither of these two approaches has a clear advantage over the other. On average boosting seems to provide a better predictive power. Bagging tends to perform better in the presence of outliers and significant noise (Bauer, 1999).

In bagging, each training set is constructed by forming a bootstrap replicate of the original training set. In other words, given a training set S of m examples, a new training set S' is constructed by drawing m examples uniformly (with replacement) from S.

Bagging generates diverse classifiers only if the base learning algorithm is unstable which is, if small changes to the training set cause large changes in the learned classifier. Breiman (1994) explores the causes of instability in learning algorithms and discusses ways of reducing or eliminating it. Bagging (and to a lesser extent, boosting) can be viewed as ways of exploiting this instability to improve classification accuracy. Adaboost requires less instability than bagging, because Adaboost can make much larger changes in the training set (e.g., by placing large weights on only a few of the examples).

D. Artificial neural networks (ANN

According to Haykin (1999), a neural network is a massively parallel distributed processor that has a natural propensity for storing experimental knowledge and making it available for use, it resembles brain in two ways i) knowledge is acquired by the network through learning process and ii) Interneuron connection strengths known as synaptic weights are used to store the knowledge. Learning in NN means a dynamic process which modifies the weights of the network in some desirable way. In terms of learning, NNs can be divided basically into two classes: Non-Supervised Networks and Supervised Networks.

As presented by Pasquotto (2010), in the Supervised NNs, there is an association between the Input and output values, which can be, for example, real data of a system and that will be used as reference in the training of the network. When there is a divergence between the NN output and the desired output, the values of the synaptic weights of the network are adjusted until the error between the output and the reference reaches an acceptable value. This learning process is called learning by error correction. These networks use perceptrons multilayers with at least one hidden layer and one output layer.

According to Evaldt (2001), the best-known learning algorithm for this type of network is the Error backpropagation algorithm, which employs the learning rule by error correction. In summary, the error backpropagation algorithm consists on the synaptic weights adjust as a function of an error signal.

In this work, the author emphases on Supervised Networks, so no more details for Non-Supervised Networks are necessary.

E. Relief

According to Robnik-Sikonja and Kononenko (2003) the Relief algorithms (Relief, ReliefF and RReliefF) can separate the interaction between attributes. They are efficient, aware of the contextual information, and can correctly estimate the quality of attributes in problems with strong dependencies between attributes. These algorithms are commonly used as a preprocessing step before implement the process model because it can calculate the attributes weight, to select splits and guide the constructive induction in learning of the regression trees and other pre-modelling features (Robnik-Sikonja and Kononenko, 1997).

A key idea of the original Relief algorithm (Kira and Rendell, 1992), is to estimate the quality of attributes according to how well their values distinguish between instances that are near to each other. The original Relief can deal with nominal and numerical attributes. However, it cannot deal with incomplete data and is limited to two-class problems. Its extension, which solves these and other problems, is called ReliefF.

F. Stepwise Regression

According to Seber and Lee (2003) in systems where multiple regression analysis involves a set of variables, it may be necessary to use a method to select the main independent variables to be used on the model. This procedure reduces the computational processing for calculations, as well as the need to acquire new data to update or maintain the model. In addition, it can be shown that, in some cases, closer to reality models can be obtained by excluding input variables from the process of analysis. The personal expertise may be useful to reduce the number of independent variables.

Stepwise regression is a linear parametric modelling technique and can be helpful to predict the process behavior. Thus, it allows further analysis regarding the optimal operational ranges for main independent variables. The linear function resultant is described by

$$Y(x_1, x_2, \dots, x_k) = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \dots + \beta_k * x_k \qquad (1)$$

where: Y is output variable, also called the predicted variable. The x represents the input variables or predictors. β_0 is the vertical axis interception and β are the regression coefficients, which correlates the input variables with the output. The k is variable number.

This technique was first described by Efroymson (1960). It is based on the joint application of two selection techniques: the inclusion Selection and the Backward Elimination of variables, step by step, in the model. Thus, it requires that two levels of significance be established, one for each technique.

Alternatively, other parameters can be used to test whether a variable is important in model definition, citing the coefficient of determination (R^2) and analysis of variance.

III. METHODS

The methods focus specially on data collecting and analysis for use on the fourth step of eight step problem-solving technique. The data was collected from PI Data Archive software®, which storage all mill process data. The sampling time defined was 30s, due to mainly fast response of pressure increase events. The chosen independent variables are based on professional experience and design. The software used for all data analysis and for the predicting algorithms is Matlab®.

First, the Relieff algorithm is used to present the main independent variables ranked by weight, as a pre-modeling approach. This is a pre-modeling approach. The goal is to understand the main weighted variables that varies when pressure suddenly increases on washer feeding stage. For this purpose, it is used data from 16/06/17 to capture a short period with enough pressure variations. As the operational upper control limit is 0.32 bar, the relief algorithm was programed to classify the pressure on binary categories: “p” pressurized, it means that the measured pressure is above the upper limit of control and “n” which means that the washer is not pressurized, and it operates below the upper limit.

The decision trees and the ANN algorithms are being used to evaluate how the chosen attributes can predict the feeding stage pressure based on a learning period. Depending on the results, another variable can be applied to model the process variations. For this purpose, the data was separated into two blocks. Each block has three days of duration. The first one is used as the learning period. On one hand, the chosen period features some high-pressure events, and this is especially important for the learning algorithms techniques. On the other hand, it is necessary some variability on the attributes for decision tree and ANN algorithms capture different kinds of examples. The Fig. 3 presents the DDW feeding stage pressure.

The second period represents a normal process behavior, and it is used as the test period for all the nonparametric techniques. The Fig. 4 presents the real pressure values.

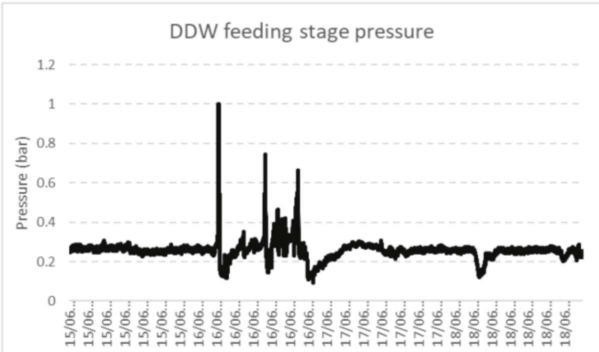

Figure 3. Learning period for the predicted variable, the DDW feeding stage pressure.

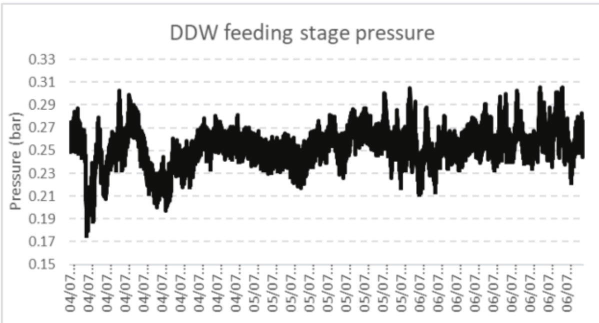

Figure 4. Test period for the predicted variable, the DDW feeding stage pressure.

A MAPE (mean absolute percentage error) and regression coefficient R^2 are the metrics chosen to evaluate the predicting power of the decision trees and ANN algorithms.

For the stepwise regression algorithm, the data used is the sum of the learning and test periods. This decision is to obtain a parametric model that presents a good but not optimal applicability for a wide operational condition. Therefore, a model with this property can be implemented on a supervisory control system and can be used as a tool for optimize the variables control range.

The problem solving efficacy is measured by the high pressure events on a monthly frequency. The criteria for counting is when the washer feeding stage pressure is above 0.45 bar.

IV. RESULTS AND DISCUSSION

The relief technique presents the weight-based rank for the dependent variables to predict the suddenly increase pressure. The Table 1 shows the prediction results.

It is suggested that for those specific high-pressure events the main variables that explain the phenomenon was the first six ones, according to the rank. These variables were responsible for 78% of relative contribution. However, this prediction is highly correlated with the variables that had behaviors bounded to feeding stage pressure. For example, the drum speed is used for the pressure control. In this case, the feedback control parameters can be evaluated for search for a better control performance. The variations on Eop production, top reactor pressure and air pressure were the main issues related to these high-pressure events. It was found that one of root causes for the top reactor variations was the lack of piping cleaning standards. For Eop production variations, it can be explained by the low gap between the actual production rate and the design washer capacity. Thus, variations on production can be critical for this equipment. The air pressure it is related to the discharge washer capacity and it was evaluated to increase its compressor capacity.

Relieff is effective as a diagnostical and pre-modelling tool, as it showed reliable results based on operational experience. This technique is simple and can be applied for similar problem and fits perfectly as a diagnostical tool for the lean 8 steps problem solving technique.

Table 1. Relieff algorithm results for variables weight prediction.

Rank	Weight	Relative contribution	Independent variables
1	0.11	22%	Eop production
2	0.07	15%	Top reactor pressure
3	0.06	13%	Air pressure
4	0.05	11%	Drum speed
5	0.04	9%	Washer vacuum
6	0.04	8%	Condensate 3 flow
7	0.03	6%	Condensate 2 flow
8	0.03	5%	Condensate 1 flow
9	0.02	5%	Condensate 4 flow
10	0.02	4%	Reactor venting aperture
11	0.01	2%	Pulp consistency
12	0.01	1%	Oxygen charge

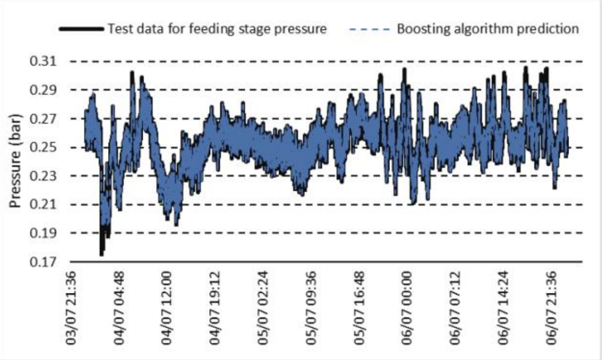

Figure 5. Test and boosting decision tree curves.

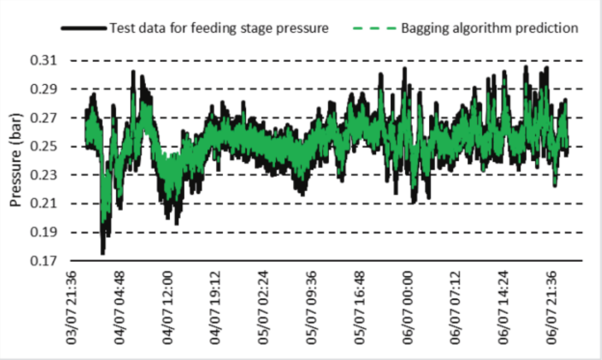

Figure 6. Test and bagging decision tree curves.

The use of boosted, bagging decision trees and ANNs are based on the evaluation of how the attributes can predict the inlet washer pressure.

The Fig. 5 shows the results for the test period using the boosting tree algorithm.

The boosting algorithm presented a precise regression result for this process with a regression coefficient of 99.8%. The number of cycles used was 10 and it is possible use more cycle to increase the prediction quality. However, it would require more computational processing and the cost-benefit relation in this case should be evaluated.

The bagging algorithm presented a similar behavior to the boosting. The error obtained was higher than the boosting tree but it is very effective as a modelling technique. Thus, the processing time was lower than boosting. This pros and cons must be considered depending which are the main modeling targets. In this case, as far as both modeling techniques presented high precision, a lower processing time can be a significant advantage. The Fig. 6 presents the bagging algorithm regression result.

Both decision trees show difficult to predict the peaks and the lower pressure values, mainly below 0.20 bar. It is suggested this issue occurred due to lack of these examples in the learning period.

A way to eliminate this effect would be choose a larger learning period with a large variety of different examples, however it would be necessary a large computational processing capacity. For this specific case, it is a drawback.

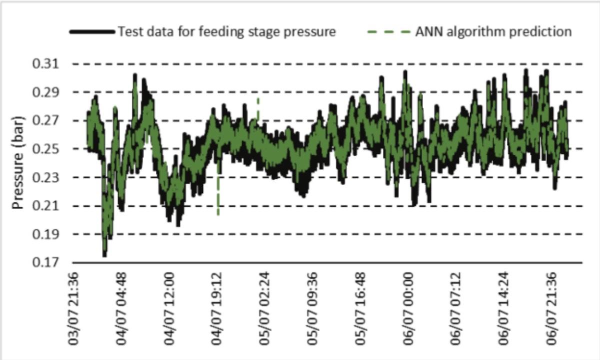

Figure 7. Test and ANN curves.

Table 2. Regression coefficients and MAPE results for non-parametric modelling techniques.

	R^2	MAPE
Boosting tree	99.8%	0.13%
Bagging tree	95.1%	1.73%
ANN	83.1%	2.18%

The Matlab® ANN toolbox uses the "dividerand" function to divide the targes into three sets, using random indices. For training, it is used the "trainlm" function that updates weight and bias values according to Levenberg-Marquardt optimization. The ANN final algorithm considered 100 perceptrons on the hidden layer and 1 perceptron on the output layer (Fig. 7).

A few simulations with less perceptrons on the hidden layer showed poor results compared to decision trees techniques. As the simulation time with 100 perceptrons was like the decision trees algorithms, the ANN configuration was kept this way. The Table 2 summarize the non-parametric models performance based on regression coefficient (R^2) and the mean absolute percentage error (MAPE).

The summary shows that the boosting tree algorithm presented the best prediction power for this process, followed by the bagging tree and the ANN. Considering that MAPE coefficients are lower than 5% for all techniques, it is suggested that all techniques was successful to model the washer inlet feeding pressure. However, the ANN provided the poorest performance according to R^2. It is suggested that this result could be improved if more perceptrons were included on the hidden and output layer requiring more time for processing the model.

The stepwise regression is applied to obtain a linear parametric equation for pressure prediction that can be used in wide operational scenarios. The first order equation with the 12 independent variables presented an R^2 of 0.56 when applied to both learning and test period.

The variables air pressure and reactor vent aperture showed the lowest confidence intervals for the regression coefficients. It is suggested that these results are related with the low variance of these parameters considering the both training and test periods.

One advantage of the technique is that it was generated a first order equation which can be applied to a wide operational condition with acceptable precision. Furthermore, these results can be used as a tool to simulate and search for optimal dependent variables range considering

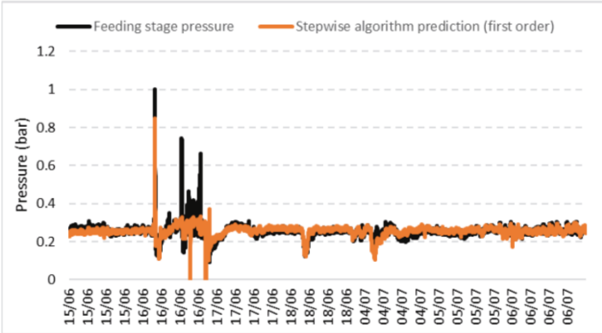

Figure 8. Learning, test period and stepwise curves.

the predicted variables. An alternative to improve its precision it would be to obtain the washer mathematical equation with the technologic supplier aiming to customize the stepwise regression algorithm. It is suggested that this approach could deliver better results for some of the peaks of day 16/06. The Fig. 8 presents the linear equation applied to the attributes values.

Finally, the 8 steps problem solving technique sessions found 10 root causes for suddenly high pressure events and four were confirmed by advanced analytics tools. After implementing nine short term actions there was 63% of reduction on pressuring frequency, it means, it was reduced from 13.7 to 5 events per month. It is suggested that the remaining high pressure events will be eliminated with the other long term actions.

The main achieved benefit is the reduction of mill losses of production due to the Eop DD washer pressurization and the elimination of the ones mill bottlenecking.

V. CONCLUSIONS

The reduction of suddenly high pressure events on DD washer feeding stage shows the 8 steps problem solving technique efficacy. This allows the mill to increase its availability and production without investment, critical matters for a pulp mill.

The data analysis on this technique can be reinforced using the advanced analytics modelling techniques. On this example the relief, decision trees and stepwise regression algorithms proved to be a well-tuned packet for problem solving in a pulp mill, especially focused on hypothesis generation and validation.

REFERENCES

Andritz, *Bleaching System Operating Manual*, CMPC celulosa S.A., Santa Fe, Chile (2017).

Bauer, E and R. Kohavi, "An empirical comparison of voting classification algorithms: Bagging, boosting, and variants," *Machine Learning*, **36**, 105-142 (1999).

Breiman, L., *Heuristics of instability and stabilization in model selection*, Technical Report, Department of Statistics, University of California, Berkeley, CA, **416** (1994).

Efroymson, M.A., *Multiple regression analysis*, Ralston HS Wilf (Ed. 1.), New York (1960).

Evaldt, M.C., *Uma metodologia para a identificação de perdas não técnicas em grandes consumidores rurais*, Master Degree Thesis, Technology Center, Universidade Federal de Santa Maria (2001).

Forrester, R., "Implications of lean manufacturing for human resource strategy," *Work Study*, **44**, 20-24 (1995).

Freund, Y., and R.E. Schapire, "Experiments with a new boosting algorithm," *Proc. 13th International Conference on Machine Learning*, 148–156 (1996).

Gatchalian, M.M., "People empowerment: the key to TQM success," *The TQM Magazine*, **9**, 429-433 (1997).

Haykin, S., *Neural Networks- a comprehensive foundation*, Prentice Hall India (1999).

Kira, K., and L.A. Rendell, "A practical approach to feature selection," *Proceedings of International Conference Machine Learning*, 249–256 (1992).

Pasquotto, J.L.D., *Previsão de Séries Temporais no Varejo Brasileiro: Uma Investigação Comparativa da Aplicação de Redes Neurais Recorrentes de Elman*, Dissertação (Graduação em Administração), Universidade de São Paulo, São Paulo (2010).

Quinlan, J.R., "Bagging, boosting, and C4.5," *Proceedings of the Thirteenth National Conference on Artificial Intelligence Cambridge*, MA: AAAI Press/MIT Press, 725–730 (1996).

Robnik-Sikonja, M., and I. Kononenko, "An adaptation of relief for attribute estimation in regression," *Proceedings of the Fourteenth International Conference Machine Learning*, 296–304 (1997).

Robnik-Sikonja, M. and I. Kononenko, "Theoretical and Empirical Analysis of ReliefF and RReliefF," *Machine Learning*, **53**, 23–69 (2003).

Seber, G.A.F. and A.J. Lee, *Linear regression analysis*, (2nd Ed.), Hoboken, NJ, Wiley (2003).

Puvanasvaran, A.P., M.H.M.A. Megat, S.H. Tang, M.R. Muhamad and A.M.S. Hamouda, "A Review of Problem Solving Capabilities in Lean Process Management," *American Journal of Applied Sciences*, **5**, 504-511 (2008).

Received: October 24, 2017.
Sent to Subject Editor: February 1, 2018.
Accepted: February 4, 2019.
Recommended by Subject Editor: José Guivant