

SUPPORT VECTOR MACHINES IN PROJECTS RISK CLASSIFICATION

***Domingos M. R. Napolitano, Marcio Romero and Renato José Sassi**

Postgraduate Program in Computer Science and Knowledge Management, Universidade Nove de Julho (PPGIC-UNINOVE)

ARTICLE INFO

Article History:

Received 19th August, 2018
Received in revised form
08th September, 2018
Accepted 18th October, 2018
Published online 28th November, 2018

Key Words:

Support Vector Machines,
Project Management, Risk Management,
Risk Matrix, Probability Impact Matrix.

ABSTRACT

Projects are fundamental to organizations, but they are subject to the occurrence of risks, which can affect their success. Therefore, project managers seek to manage risks by demanding decisions about their treatment. A widely used tool in this process is the Probability and Impact Matrix (PIM) that is popular, but deficient. This paper aimed to answer the question "How to classify risks in projects using Support Vector Machines (SVM)?" To answer this question, SVMs were applied to process risks, previously classified using an PIM. The results show that the SVM application returns more accurate results than the PIMs.

Copyright © 2018, Domingos M. R. Napolitano et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Citation: Domingos M. R. Napolitano, Marcio Romero and Renato José Sassi. 2018. "Support vector machines in projects risk classification", *International Journal of Development Research*, 8, (11), 23915-23920.

INTRODUCTION

A project is a temporary effort undertaken to create a product, service or exclusive result, are essential enterprises for the implementation of organizational strategies and fundamental for the growth of companies (PMI, 2017). Risks are events whose occurrence is uncertain, which can affect the success of a project and in its evaluation must consider two dimensions: the degree of uncertainty and its impact on the objectives (Hillson, 2009). Therefore, it is important to make decisions about which risks should be addressed by applying risk management techniques (Kerzner, 2011). Risk management is characterized by a sequence of activities that begins with the planning, followed by its identification, qualitative and quantitative analyzes and its monitoring and control (PMI, 2017). Project risk management is often related to project success, but, paradoxically, little adopted (Rabechini Junior and Carvalho, 2013). Probabilities are used to describe uncertainty and the term impact is most often used to describe the effect on objectives and the product of both is employed for the quantification of risks (Carvalho and Rabechini Júnior, 2011). In order to decide which risks can be addressed, monitored or tolerated, tools called Probability and Impact

Matrix (PIMs), also called risk matrices (Cox, 2008). The studies of Cox (2008); Markowski and Mannan (2008), Ni, Chen and Chen (2010) and Duijim (2015) showed that PIMs have deficiencies that can lead to a wrong decision and seek to use different techniques to minimize such deficiencies, but are quite popular and easy to apply. This paper aims to evaluate the use of Support Vector Machines (SVM) in the project's risks classification. In order to reach this aim, the application of an SVM data obtained through an PIM proposed by Cox (2008).

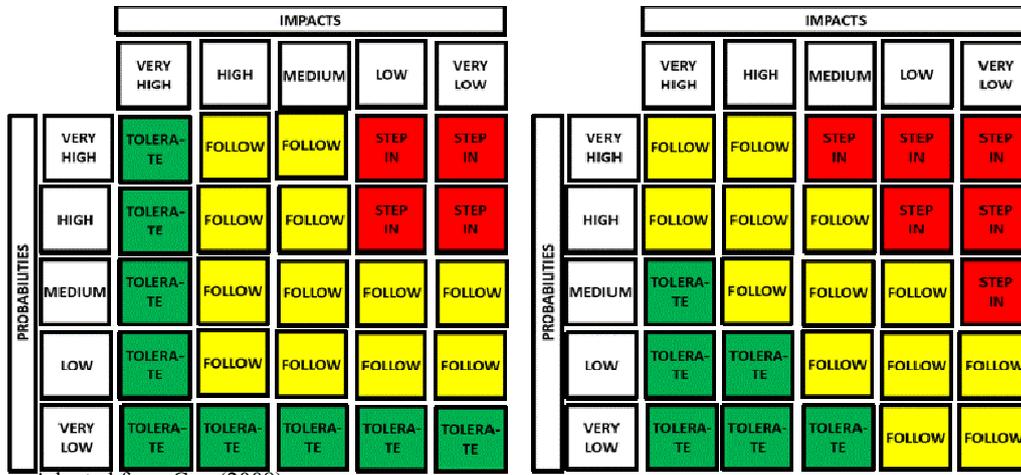
Theoretical Reference

According to PMI (2017), the first stage of risk management involves planning the processes that will be used in its management, followed by the identification, mapping and characterization. The identified risks are then analyzed qualitatively and quantitatively, following a scale of priorities. After the analysis, a plan of responses to the risks should be planned and implemented, which should establish the allocation of resources for the proper treatment of risks (PMI, 2017). The risk management process involves several strategies, such as acceptance, monitoring, administration or resolution (Carvalho and Rabechini Júnior, 2011). One of the tools most used in the decision on the treatments of risks are the Probability of Impact Matrix (PIM) (Hillson, 2009 and Cox, 2008). The application of PIM was analyzed by authors

***Corresponding author: Domingos M. R. Napolitano**
Postgraduate Program in Computer Science and Knowledge Management,
Universidade Nove de Julho (PPGIC-UNINOVE)

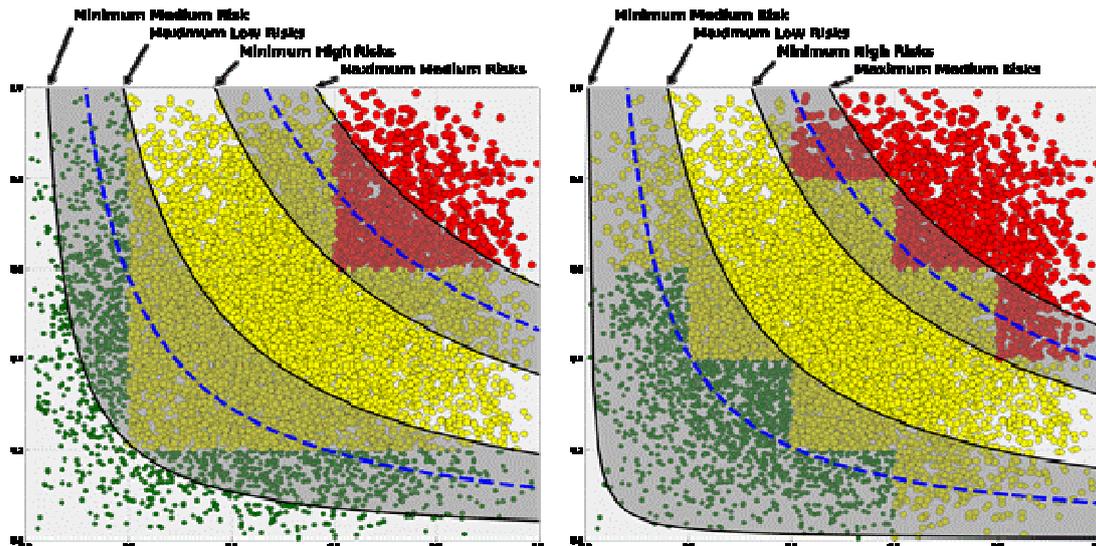
such as Cox (2008), Baybutt (2015), Ale, Burnap and Slater, (2015). In order to improve the performance of PIMs, Cox (2008) proposes three axioms: weak consistency, interrelationship and consistent coloring, which can be summarized in a theorem. In Figure 1 two PIMs are presented, the first one following the Cox theorem and the second violating it.

PIMs have become very popular as they are perceived as a convenient and understandable way of presenting the risks(Ale, Burnap and Slater, 2015). However, this convenience, derived from the simplicity of a PIM, implies loss of information, which can be observed in Figure 2 below, each point represents a risk positioned in the probability and impact plan with the respective classifications.



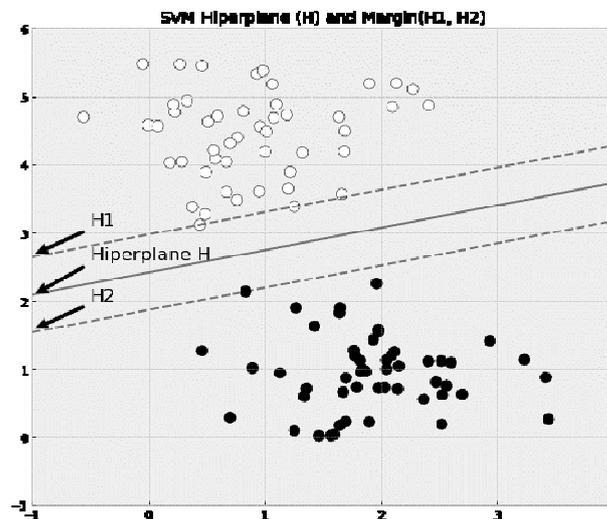
Source: Adapted from Cox (2008)

Figure 1. In the left image PIM following the Cox's theorem and violates the above theorem at the right



Source: Authors

Figure 2. PIMs and risk curves

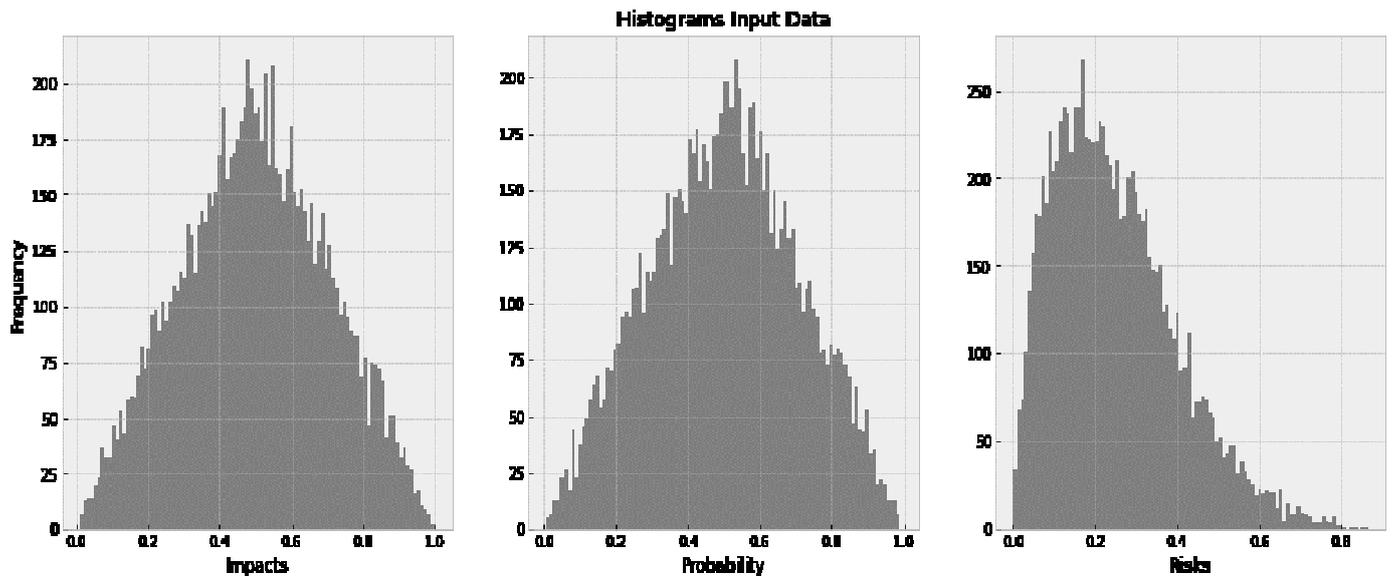


Source: Authors

Figure 3. Hyperplane and margin of an SVM

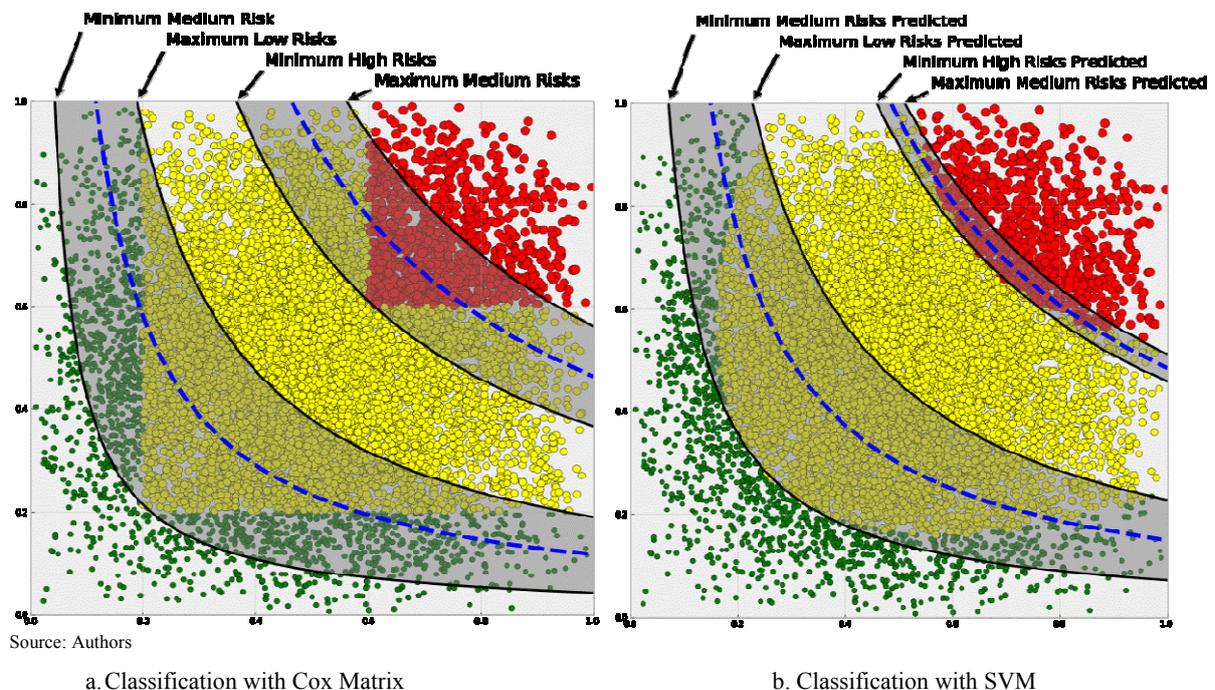
In Figure 2, the chart to the left shows the classification by an PIM following the Cox theorem and to the right are shown the risks classified by means of a conventional PIM. They are also plotted the risk curves representing the points with the same result of the product, likelihood and impact and quantitative definition of risk. In principle, the points with the same risk value should have the same classification. However, Figure 2 shows that values under shaded areas, even having the same risk value, can be classified differently, depending on the shape of the PIM. This classification can be understood as imprecise, since events with the same risk values.

It can be said that in the shaded regions there is a problem of ambiguity in the hierarchy of risks, that is to say, it is not possible by means of the classification to distinguish between two risks of the same value or even to establish a prioritization among the risks contained in this region. For the purpose of this paper the classifications in this situation will be denominated ambiguous classifications. The idea of using computational intelligence techniques to solve the problem of risk classification with PIMs was initially suggested by Cox (2008) and implemented by authors such as Markowski and Mannan (2008), who proposed the application of fuzzy logic in



Source: Authors

Figure 4. Histogram of the input data



Source: Authors

a. Classification with Cox Matrix

b. Classification with SVM

Figure 5. Results with PIM of Cox and SVM

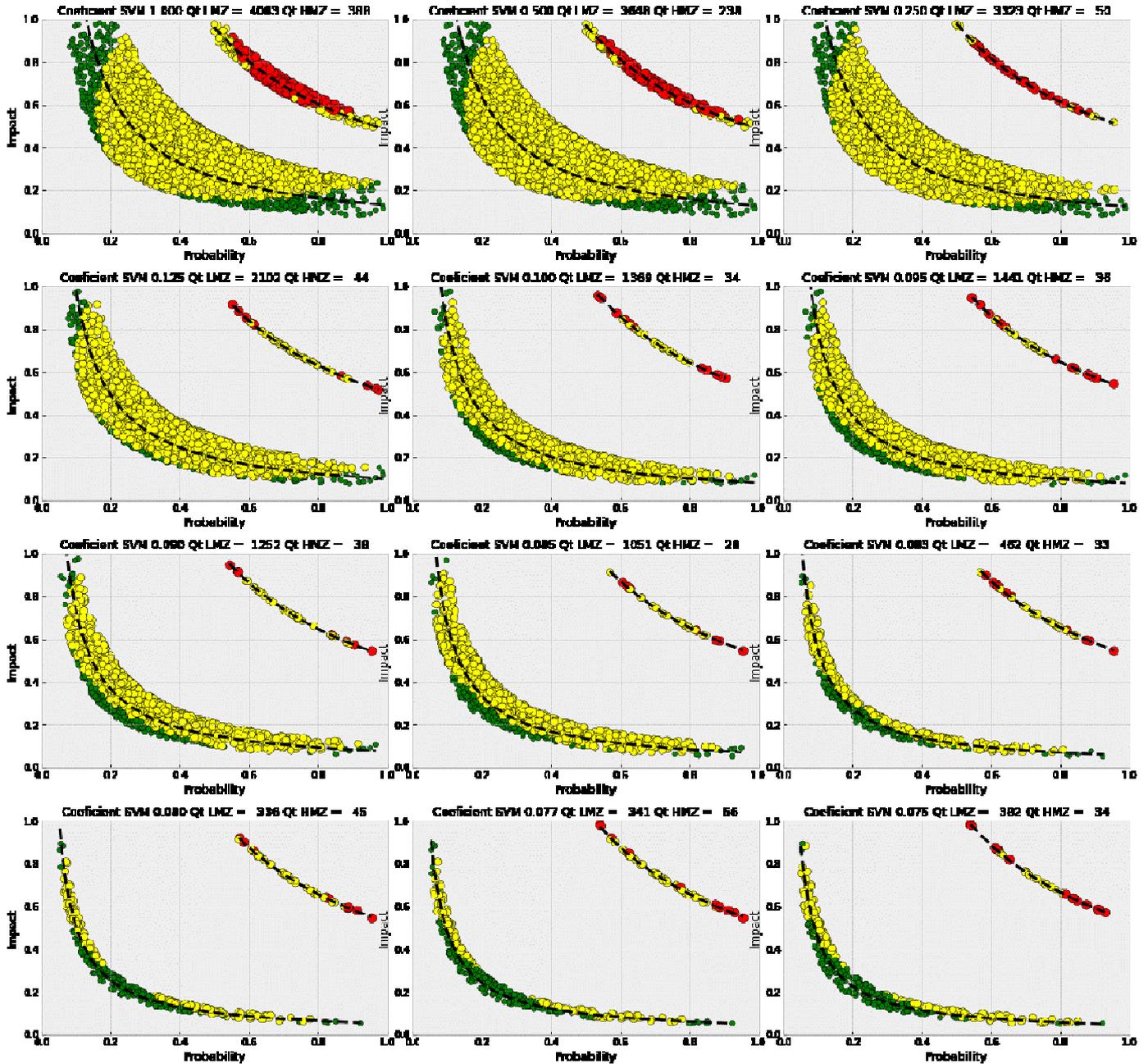
The effect presented in Figure 2 is pointed out as a great fragility of PIMs by Cox (2008) and Baybutt (2016) because it results in the possibility of reverse classification of a risk, leading to prioritizing minor risks to the detriment of larger ones.

PIMs and Jiang and Chen (2014) who employed Support Vector Machines in supply chain risks. Support Vector Machines or simply SVMs are the basis of a machine learning theory, developed by Cortes and Vapnik (1995) and adopts the

principle of Structure Risk Minimization (Jiang and Chen, 2014). The basic idea of SVMs is to map nonlinearly linearly inseparable sets of data are mapped into a space of high-dimensional characteristics through the function, making them linearly separable; (Li et al., 2009; Elish and Elish, 2008) expressed in Figure 3. In Figure 3, ○ and ● represent, respectively, different classes. H is the so-called optimal separation hyperactive plane.

$$\xi_i \geq 0, i = 1, 2, \dots, n$$

Where, w = O coefficient of the optimum hyperplane, b = threshold value and ξ = non-negative clearance variable. According to Jiang and Chen (2014) the input vector X can be mapped to a characteristic space of high dimension by a given non-linear mapping.



Source: Authors

Figure 7. Results of classification with SVM varying the penalty factor

H1 and H2, respectively, represent a straight line through the nearest point of H and they equate to H (the distances of H1 and H2 to H are the same). According to Jiang and Chen (2014) given the set of data $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, $y_i \in \{+1, -1\}$, for both the situations, that is, linearly separable or linearly inseparable, these data can be separated by a hyperplane. The hyperplane satisfies:

$$\text{minimize } \frac{1}{2}(w)^T w \dots \dots \dots (1)$$

$$\text{subjected to } y_i[(w \cdot x_i) + b] - 1 + \xi_i \geq 0,$$

And then the optimal hyperplane of separation can be found in this space, the problem to seek optimal classification is transformed into:

$$\text{minimize } \alpha \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n x_i y_j \alpha_i \alpha_j K(x_i \cdot x_j) - \sum_{j=1}^n \alpha_{ij} \quad (2)$$

$$\text{subjected to } \sum_{i=1}^n y_i \alpha_i = 0,$$

$$0 \leq \alpha_i \leq C, i = 1, 2, \dots, n$$

Where, $K(u, v)$ = The kernel function and C = Othe penalty coefficient. Since $C > 0$ and a higher value of C means a higher penalty for classification error. According to Eq. (2), the optimal solution can be obtained $\alpha^* = (\alpha_1^*, \dots, \alpha_n^*)$ e $b^* = y_i \sum_{i=1}^n \alpha_i^* K(x_i, x_j)$, and then the decision-making is $f(x) = sigmoid(\sum_{i=1}^n \alpha_i^* K(x_i, x_j) + b^*)$. Is fundamental to the function of the pre-determined kernel or kernel and to define the appropriate parameters to discover the optimal hyperplane. In this study, the radial base core function (RBF).

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

MATERIALS AND METHODS

To perform the experiments, an SVM was implemented for classification using the machine learning library scikit learn of the Python 3.5 language, for the generation and processing of data the Scipy, Numpy and Pandas libraries were used to visualize the data used to Matplotlib library. The experiment was conducted in accordance with the following steps:

function setting the optimize. curve_fit function of the SciPy package was used.

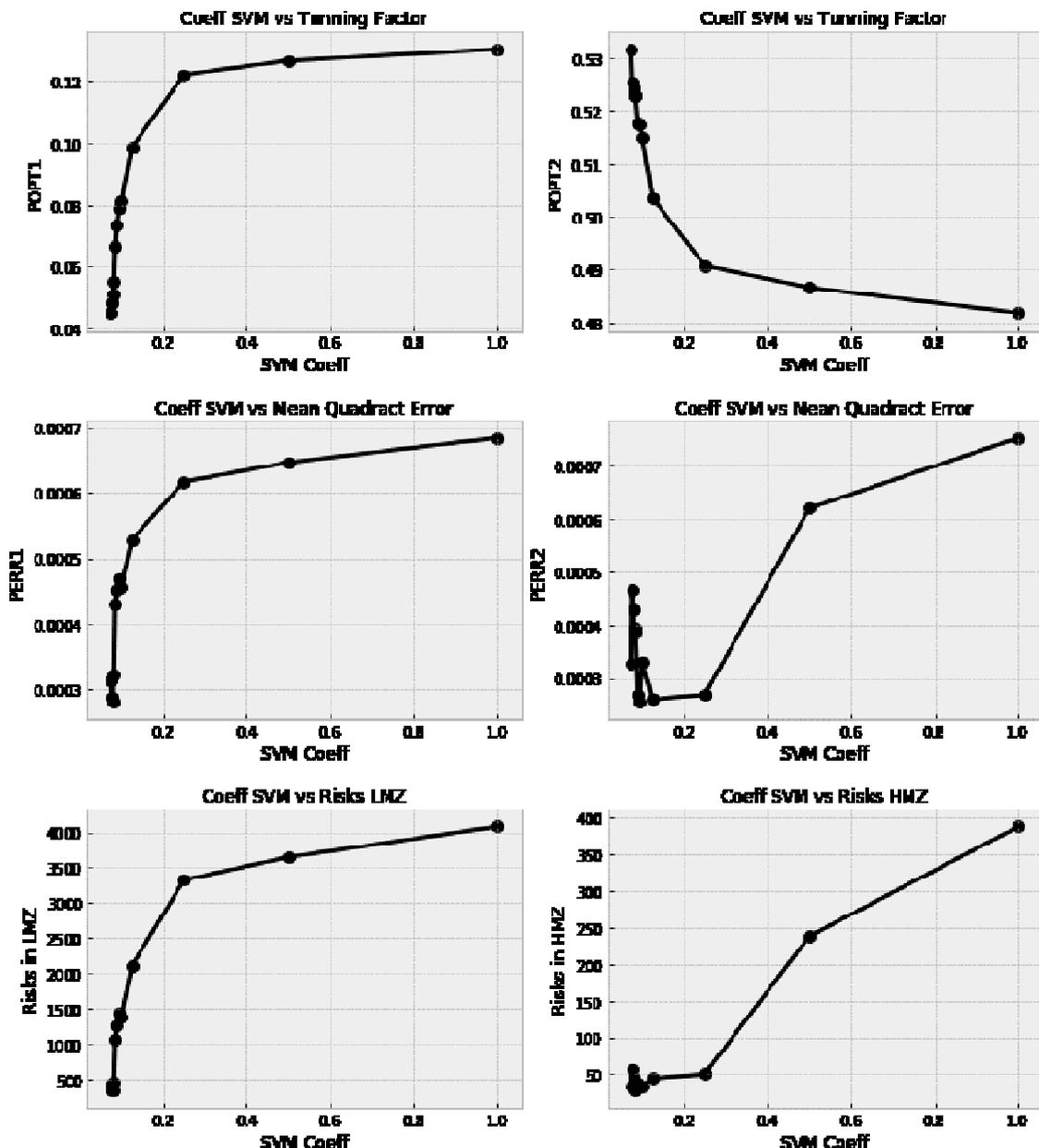
Step 4. Comparison of results.

Step 5. Reclassification of the data set with SVM by varying the penalty factor C for the values in the set 1.0, 0.5, 0.25, 0.125, 0.1, 0.095, 0.09, 0.085, 0.0825, 0.08, 0.0775 and 0.075.

Step 6. Analysis of the results found, evaluating the sensitivity to C of the adjustment factor A and its mean square error and the number of instances in the zones of ambiguity.

RESULTS AND DISCUSSION

Figure 4 shows the input data for the classification processes. The data were generated randomly following a triangular distribution with parameters minimum 0.0, mean 0.5 and maximum 1,0. The results of step 2 were consolidated with the results of step 3, in the graph shown in Figure 5 below. Observing the graph of Figure 5a, it can be seen that employing an PIM of Cox instances with similar risk values, present



Source: Authors

Figure 8. Classification relationships with SVM varying the penalty factor

different classifications, which can be observed in the shaded areas and delimited by risk lines with minimum average risk limits (represented per ●) and maximum low risk (represented by ●) and maximum average risk and minimum high risk (represented by ●). Figure 5b presents the results for classification using an SVM, although classification problems are observed, it was possible to find decision functions for each of the zones where the classification is ambiguous and it can also be noticed that the shaded regions are smaller which indicates a lower probability of misclassifications. In step 4, the classification with the SVM was repeated, varying the coefficient of penalty C and keeping the classification obtained using the PIM of Cox as a target. The results are expressed in Figure 6. To facilitate the visualization only the values were included where the classification is within the margins of the SVM. The results show that the variation of the penalty factor C influences the adjustment factor A of the relation $IMPACT = A / PROBABILITY$, which delimits the decision classes. This variation changes the regions where the classification is ambiguous and the mean squared error of the adjustment of A also undergoes variations as well as the number of instances in the zone of ambiguity. The results presented in Figure 8 show that the sensitivity to factor C variation (SVM Coef) of the adjusted factor A (POPT1 and POPT2), the mean square error (PERR1 and PERR2) of this adjustment and the number of instances in the zones that the classifications are ambiguous (LMZ and HMZ). The sensitivity analysis showed that to reduce the classifications within the margin it is necessary to decrease the adjustment factor, which leads to more conservative classifications than with a conventional PIM. In the PIM of Cox, used in the experiment, the risk that delimits the Low Middle zones is 0.2, but as a function of the discretization of the data this value can reach 0.04 which is the minimum of an average risk or 0.19 that is the maximum of a low risk. With less intensity the problem also occurs in the border between high and medium. The main advantage in the use of SVM is that it is possible to minimize the problems with the PIMs, being possible to define an exact threshold of the classification that is the adjusted factor A.

Conclusions

The results of the experiments allow us to affirm that it is possible to perform the risk classification using Support Vector Machines with advantages over traditional methods, since effects such as reverse classification and ambiguity in the risk hierarchy are minimized. It was also possible to verify that the SVMs allow to define decision thresholds with narrower limits of undefinition.

As future research is proposed the joint use of other techniques such as fuzzy logic, to reduce the ambiguity in the results, and evolutionary algorithms for the optimization of parameters of the SVM. The authors would like to thank Uninove and CAPES for the scholarships granted, which allowed the realization of this research.

REFERENCES

- Ale, B.; Burnap, P. and Slater, D. 2015. "On the origin of PCDS-(Probability consequence diagrams)". *Safety science*, v. 72, p. 229-239.
- Baybutt, P. 2016. Designing risk matrices to avoid risk ranking reversal errors. *Process Safety Progress*, v. 35, n. 1, p. 41-46.
- Carvalho, M. M; Rabechini Júnior, R. 2011. "Fundamentosemgestão de projetos". 3ª. ed. São Paulo, SP: Atlas, 2011.
- Cortes, C.; Vapnik, V. 1995. "Support vector networks". *Machine learning*, v. 20, p. 273-297.
- Cox, L. 2008. "What's wrong with risk matrices?". *Risk analysis*, v. 28, n. 2, p. 497-512.
- Duijin, N. 2015. "Recommendations on the use and design of risk matrices". *Safety science*, v. 76, p. 21-31.
- Elish K. O.; Elish, M. O. 2008. "Predicting defect-prone software modules using support vector machines". *Journal of Systems and Software*, v. 81, n. 5, p. 649-660.
- Hillson, D. 2009. "Managing risk in projects". Farhan, Surrey (UK), Gower Publishing, Ltd.
- Jiang, F; Chen, J. 2014. "A hybrid support vector machines and two-dimensional risk matrix model for supply chain risk assessment". *Research Journal of Applied Sciences, Engineering and Technology*, v. 7, n. 11, p. 2193-2199.
- Kerzner, H. 2011. "Gerenciamento de projetos - Uma abordagem sistêmica para planejamento, programação e controle". 10. ed. São Paulo: Edgard Blucher Ltda., 2011.
- Li, H.; Liang, Y.; XU, Q. 2009. "Support vector machines and its applications in chemistry". *Chemometrics and Intelligent Laboratory Systems*, v. 95, n. 2, p. 188-198.
- Markowski, A.; Mannan, S. 2008 "Fuzzy risk matrix". *Journal of hazardous materials*, v. 159, n. 1, p. 152-157.
- Ni, H; Chen, A.; Chen, N. 2010. "Some extensions on risk matrix approach". *Safety Science*, v. 48, n. 10, p. 1269-1278.
- Project Management Institute (PMI). (2017). "Um Guia do Conhecimento no Gerenciamento de Projetos (PMBOK)" Sexta Edição em Português. Newton Square, PA: Project Management Institute, Inc.
- Rabechini Junior, R.; Carvalho, M. M. 2013. "Relacionamento entre gerenciamento de risco e sucesso de projetos". *Prod.*, São Paulo, v. 23, n. 3, p. 570-581.
