

CLASSIFICAÇÃO DE FALHAS DE UM CENTRO DE USINAGEM: UM ESTUDO DE CASO UTILIZANDO ÁRVORE DE DECISÃO

Higor Medina Velloso

Instituto Federal Fluminense
BR 356, km 158 - Martins Lage - Campos dos Goytacazes – RJ
hmveloso@gmail.com

Henrique Rego Monteiro da Hora

Instituto Federal Fluminense
BR 356, km 158 - Martins Lage - Campos dos Goytacazes – RJ
dahora@gmail.com

RESUMO

As empresas estão cada vez mais buscando atingir uma performance de alto nível para se manter em um ambiente caracterizado por intensa competitividade global. Uma das formas de atingir o nível qualidade e excelência desejado é dando foco nas funções de verificação das atividades de manutenção. Na área de manufatura industrial, os centros de usinagem representam o que há de mais avançado no quesito tecnologia, e paradas indesejadas podem gerar prejuízos para as companhias. Para um melhor controle de detecção e reconhecimento de falhas, a fim de atingir um maior período de disponibilidade deste maquinário, diversos autores utilizam ferramentas de inteligência artificial, como árvores de decisão e SVM, entre outros. Este trabalho tem como objetivo aplicar a técnica de árvore de decisão em um *dataset* composto pelo histórico de falhas de um centro de usinagem no período de março de 2016 a outubro de 2018, sendo que cada evento possui 4 atributos: Falha, Componente em falha, Tempo até a falha e Tempo da última manutenção. No ambiente WEKA, foi utilizado o algoritmo J48 com validação do tipo *cross-validation* com 10 subconjuntos e um número mínimo de 7 objetos por folha para gerar um modelo que possui como desfecho o atributo Falha e outro para Componente em falha. Ao fim do trabalho, obteve-se que o sistema hidráulico é o componente de maior criticidade deste tipo de maquinário, uma vez que possui o maior número de eventos tanto para o atributo Falha, quanto para Componente em falha. Apesar dos modelos terem obtido taxas razoáveis de acurácia, percebe-se que eles possuem algumas falhas.

Palavra-chave: Previsão de falha; Mineração de dados; Árvore de Decisão; Centro de Usinagem.

ABSTRACT

Companies aim to reach a high standard performance to maintain themselves in a global intense competitive environment. One way to reach the quality and excellence goal is focusing in maintenance activities verification functions. In industrial manufacturing, machining centers represent the greatest technology advance, and unexpected downtimes may result in financial loss. For a better failure detection and identification control, in order

to get a higher machine availability, many authors use artificial intelligence tools, such as decision trees and SVM, among other techniques. This paper aims to apply the decision tree technique in a dataset containing a machining center failure history from March 2016 to October 2018 and each event has 4 attributes: Failure, Failed component, Time to failure and Time since last maintenance. In WEKA environment, the J48 algorithm was used with a 10 folders cross-validation and the minimum objects per leaf set to 7 to create a model with Failure and other with Failed Component as closure. At the end of this paper, the hydraulic system was shown as the most critical component in this kind of machine, once it has the more events to both attributes Failure and Failed component. Besides models present a reasonable accuracy, some failures could be noticed.

Keywords: Failure prediction; Data mining; Decision Tree; Machining Center.

Como Citar:

VELLOSO, Higor Medina; HORA, Henrique Rego Monteiro da. Classificação de falhas de um centro de usinagem: um estudo de caso utilizando árvore de decisão. *In: SIMPÓSIO DE PESQUISA OPERACIONAL E LOGÍSTICA DA MARINHA*, 19., 2019, Rio de Janeiro, RJ. *Anais [...]*. Rio de Janeiro: Centro de Análises de Sistemas Navais, 2019.

1. INTRODUÇÃO

As empresas encaram um ambiente caracterizado por uma intensa competitividade a nível global [1] e, por causa da exigência do mercado, fazem um esforço para atingir uma performance de alto nível. Parte desse esforço tem sido voltado para a manutenção, incluindo as funções de verificação dessas atividades [2]. Neste contexto, a manutenção possui um papel essencial na garantia tanto da qualidade quanto da produtividade empresarial [3].

Segundo [4], a manutenção surgiu na época da Primeira Guerra Mundial da necessidade de uma equipe para reparos em máquinas que apresentavam falha no menor tempo possível. Ainda segundo [4], as atividades de manutenção evoluíram durante o período da Segunda Guerra Mundial, onde, além de corrigir as falhas, tornou-se importante evitar que elas ocorressem através da prevenção. Essas manutenções seriam classificadas mais tarde como manutenção corretiva e preventiva, respectivamente [5].

Em [5] é definido ainda um terceiro tipo de manutenção: a manutenção preditiva. Porém, segundo [6], as atividades de manutenção podem ser divididas em quatro tipos: manutenção de quebra (a mais ineficiente, obrigando alto estoque de sobressalentes), manutenção preventiva (uma boa abordagem para equipamentos de baixa rotatividade, mas pode acabar sendo realizada cedo ou tarde demais), manutenção preditiva (realizada apenas quando é necessário, podendo ainda aumentar a capacidade produtiva da máquina) e manutenção proativa (ataca a causa raiz das falhas, porém necessita de pessoal altamente capacitado).

Em [7] as manutenções preventiva e preditiva são definidas como os dois principais tipos e caracterizadas como manutenção baseada no tempo e manutenção baseada na condição, respectivamente. Apesar da manutenção preventiva ser realizada em intervalos fixos levando em consideração o custo médio do componente e sua distribuição de falhas durante sua vida útil [8], ela não considera as condições do sistema ao longo de seu funcionamento [7]. Assim, conforme comparação em [9], a manutenção preditiva apresenta um melhor desempenho para evitar falhas funcionais.

Na área de manufatura industrial, onde os centros de usinagem estão se tornando o principal equipamento para a tecnologia avançada [10], o objetivo é produzir produtos com lucro através de uma manutenção efetiva [11], pois paradas não agendadas destas máquinas podem atrasar *deadlines* e causar grande perda financeira [12].

Visando melhorar a detecção e reconhecimento de padrões de falha, [13] propõem a utilização de ferramentas de inteligência artificial. Esta área de monitoramento do estado das

máquinas é foco de um interesse crescente, tanto da comunidade científica quanto das indústrias, assim como a utilização de IA como ferramenta de previsão de falhas [14,15]

Em [16], é realizado um estudo estatístico com dados de falha de doze centros de usinagem entre 2005 e 2010. Os resultados encontrados evidenciam que o sistema que apresentou mais falha foi o sistema hidráulico dentre 15 sistemas diferentes, com mais de 18% do total de falhas.

Em [17], é apresentada a utilização de árvore de decisão e *Support Vector Machine* (SVM) para a detecção de furtos de energia utilizando os dados de várias casas dos EUA.

Já em [14], é realizado um experimento para avaliar a condição de rolamentos utilizando as técnicas de árvore de decisão utilizando o algoritmo J48, *Proximal Support Vector Machine* (PSVM) e *Support Vector Machine* (SVM). Os resultados encontrados por eles demonstram que a utilização de árvores de decisão em aplicações de monitoramento de condição pode ser bastante efetiva, evidenciando as características mais discriminantes.

Uma vez estabelecido o estado da arte, verifica-se muitas proposições de modelos, mas nem tantas aplicações reais. Assim, este trabalho vem dar sua contribuição, ocupando-se de experimentar a aplicação de árvores de decisão, técnica mais frequente entre as proposições de modelos preditivos, em centros de usinagem.

2. METODOLOGIA

A metodologia deste trabalho se baseou em duas etapas: coleta de dados e análise exploratória.

2.1. COLETA DE DADOS

Para trabalhos de análise de falha, a base da metodologia é a etapa de coleta de dados. Ela é fundamental para criar um modelo que melhor traduz a realidade [16]. Neste trabalho, foi coletado o histórico de falhas no período de março de 2016 a outubro de 2018 de um centro de usinagem modelo Wotan Cutmax 3-TT de uma empresa do ramo de manufatura industrial.



Figura 1 - Wotan Cutmax 3-TT. Fonte: PORTAL MÁQUINAS [18].

Este *dataset* é composto por 67 instâncias com atributos que contém atributos categóricos e numéricos. Para cada atributo foi dado um código de identificação, conforme Quadro 1. A partir deste ponto do trabalho, cada atributo será referenciado apenas utilizando seu respectivo código.

Quadro 1 - Lista de atributos do histórico de falhas

Nome do atributo	Código	Tipo de informação
Falha	(1)	Categórica

<i>Componente em falha</i>	<i>(II)</i>	<i>Categórica</i>
<i>Tempo até a falha</i>	<i>(III)</i>	<i>Numérica</i>
<i>Tempo da última manutenção (h)</i>	<i>(IV)</i>	<i>Numérica</i>

Fonte: Própria.

Dos atributos categóricos, (I) contém uma breve descrição sobre a falha e (II) possui a informação de qual o componente apresentou a falha. Já nos atributos numéricos, (III) mostra a quantidade de turnos de 16h que a máquina operou e (IV) o tempo que durou a última manutenção.

2.2. ANÁLISE EXPLORATÓRIA DOS DADOS

Para realizar a análise exploratória de dados será realizada uma análise em cada atributo, gerando histogramas para os atributos categóricos e quadros com estatísticas para os atributos numéricos. Caso a distribuição dos eventos não seja balanceada, os eventos com os três menores valores de frequência serão agrupados em no evento “Outros”. Esse agrupamento diminui o problema de classificação causado pelo desbalanceamento da distribuição, uma vez que eventos raros são mais sujeitos a serem classificados de forma errada. Por fim, será utilizado o algoritmo J48 para confecção da árvore de decisão.

O J48 é uma implementação do algoritmo de classificação C4.5 para o Waikato Environment for Knowledge Analysis (WEKA) [19]. Apesar de simples, o algoritmo é efetivo e possui melhor desempenho se comparado com o classificador Naive Bayes, por exemplo [20].

O algoritmo foi aplicado utilizando o método de validação *cross-validation* com 10 subconjuntos definidos e um número mínimo de 7 objetos por folha. Duas árvores foram geradas: uma onde as folhas são o atributo (I) e outra onde as folhas são o atributo (II).

O atributo (I) foi removido do *dataset* ao gerar a árvore para (II) e vice-versa, uma vez que o objetivo é gerar um modelo para prever qual o componente irá falhar ou qual falha irá acontecer.

3. RESULTADOS

O atributo (I) está distribuído em doze categorias, sendo elas: “Barramento não lubrificado”, “Componentes sujos”, “Concentração de óleo”, “Entupimento das caixas de óleo”, “Falta de lubrificação”, “Filtro de óleo saturado”, “Folga de componentes”, “Nível baixo de óleo”, “Óleo contaminado”, “Proteção do barramento solta”, “Quebra de componente” e “Vazamento”. A Figura 2 mostra o histograma deste atributo.

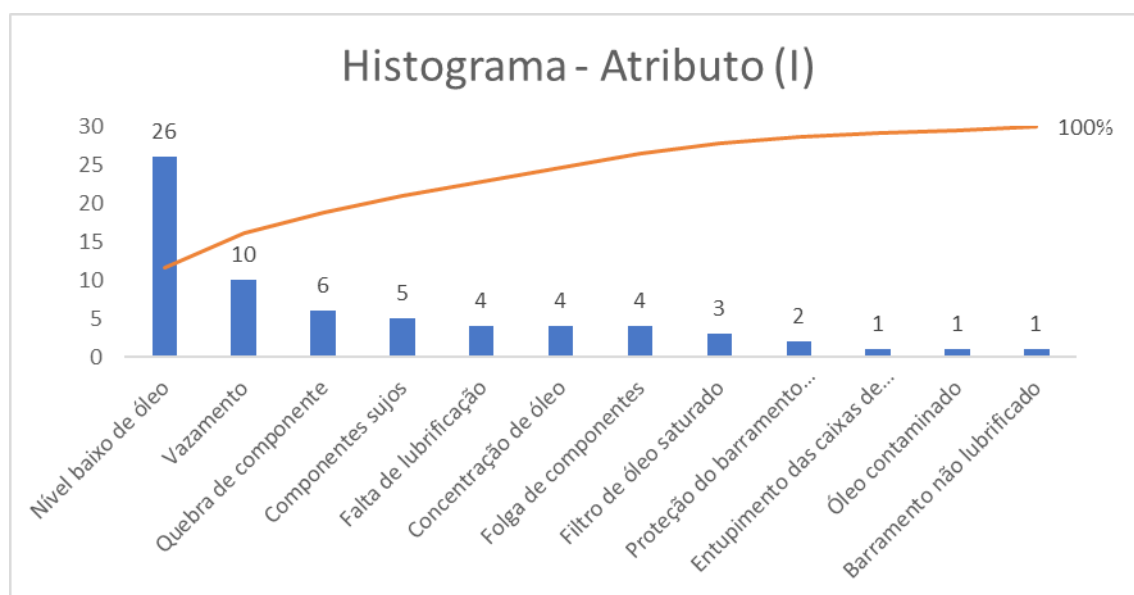


Figura 2 - Histograma do atributo (I). Fonte: Própria.

Dada a distribuição desbalanceada, todos os eventos do atributo (I) com frequência igual ou menor do que 3 foram agrupados no evento “Outros”. O histograma da nova distribuição pode ser vista na Figura 3.

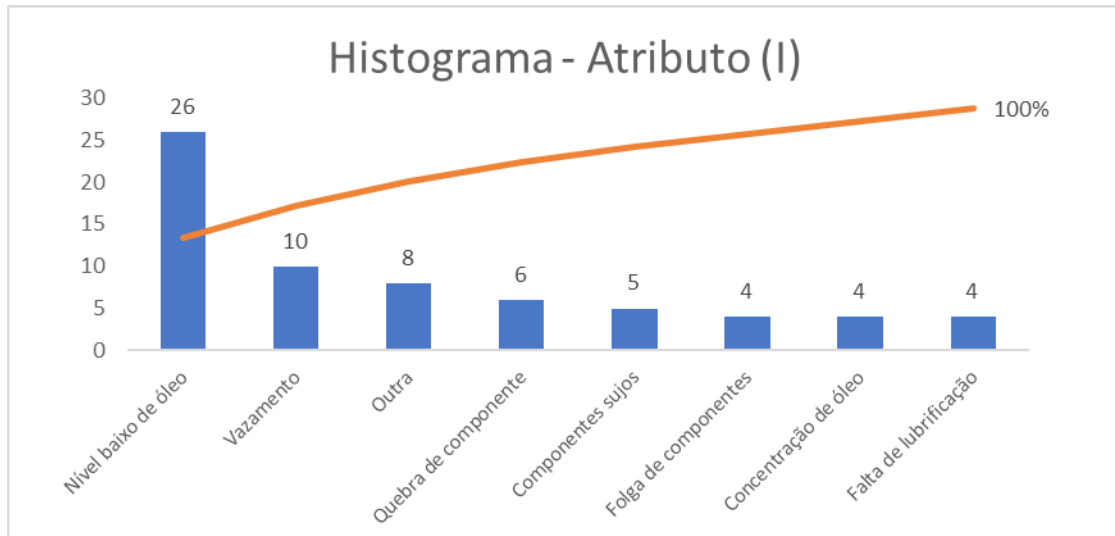


Figura 3 - Histograma do atributo (I) após tratamento de eventos raros. Fonte: Própria.

Já o atributo (II), durante o período avaliado, mostrou que treze componentes apresentaram falha. Foram eles: “Cabeçote”, “Eixos”, “Esteira de cavaco”, “Filtro”, “Lubrífil”, “Monitor do Comando Heidenhain”, “Montagem geral”, “Proteção do barramento”, “Proteção telescópica”, “Rolamento central”, “Sistema hidráulico”, “Torpedo” e “Válvula de retenção”. A Figura 4 mostra o histograma deste atributo.

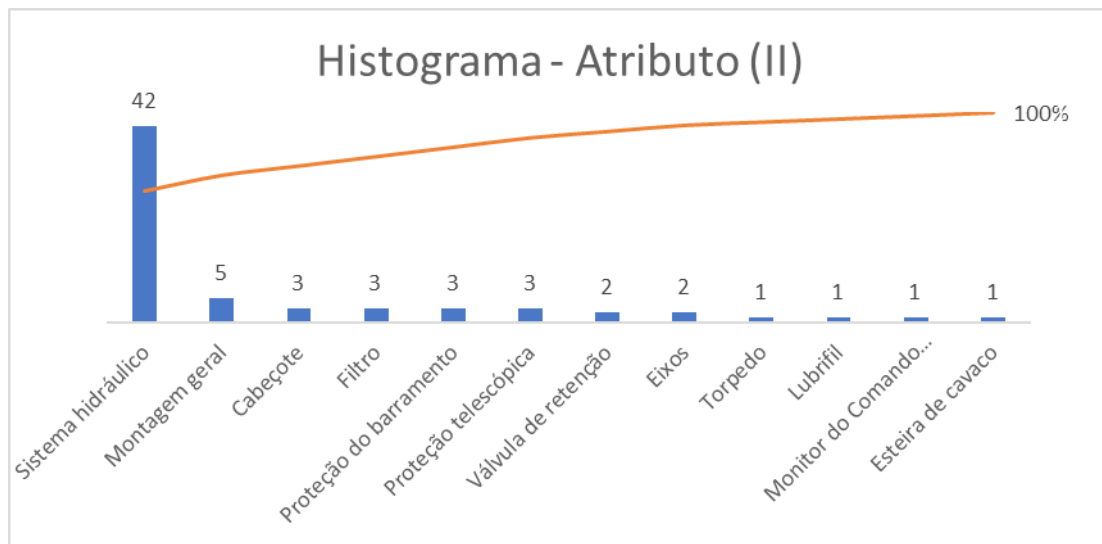


Figura 4 - Histograma do atributo (II). Fonte: Própria.

Assim como feito anteriormente para o atributo (I), os eventos que possuem frequência igual ou menor do que três foram agrupados no evento “Outro” a fim de melhorar o balanceamento da distribuição, como pode ser visto no histograma da Figura 5.

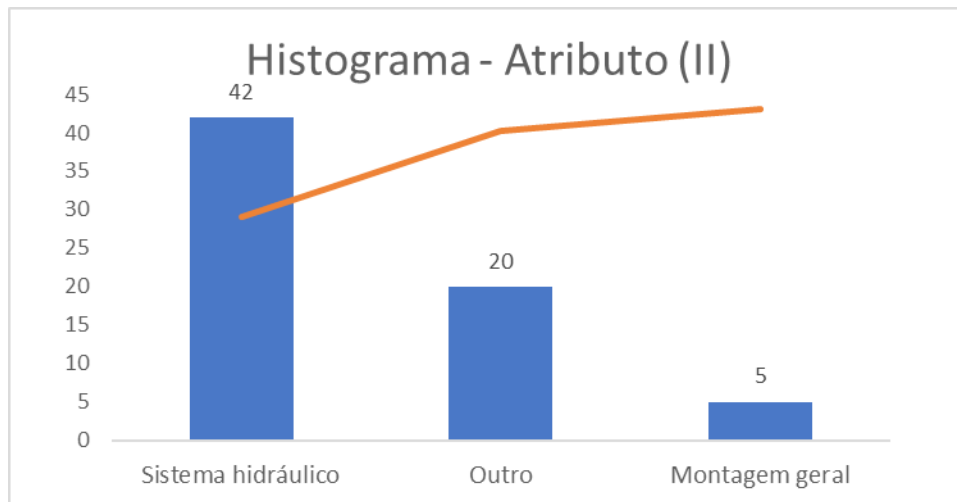


Figura 5 - Histograma do atributo (II) após tratamento de eventos raros. Fonte: Própria.

Os atributos (III) e (IV) possuem 67 valores reais e não nulos. O Quadro 2 apresenta algumas de suas estatísticas.

Quadro 2 - Informações estatísticas dos atributos numéricos

Atributo	Média	Min	Max	Desvio padrão
(III)	13,90	0,04	86,94	16,89
(IV)	2,74	0,08	26,00	4,64

Fonte: Própria.

Na Figura 6, temos a árvore de decisão que explora o atributo (I) como desfecho, ou seja, deseja-se saber como os outros atributos se comportam quando fixado o valor deste. Nela, pode-se notar que o atributo (III) é o único atributo relevante para a classificação. Nos valores entre parênteses, pode-se identificar quantos eventos alcançaram determinado desfecho (valor a esquerda) e quantos desses foram classificados de forma errada.

Esta árvore possui três folhas e tamanho cinco e teve uma acurácia de 40,3%, acertando a classificação de 27 das 67 instâncias. Apesar deste baixo desempenho, este modelo obteve uma taxa de acerto superior a 88% para falhas do tipo “Nível baixo de óleo”, conforme Figura 7. Sua matriz de confusão pode ser vista na Figura 8.

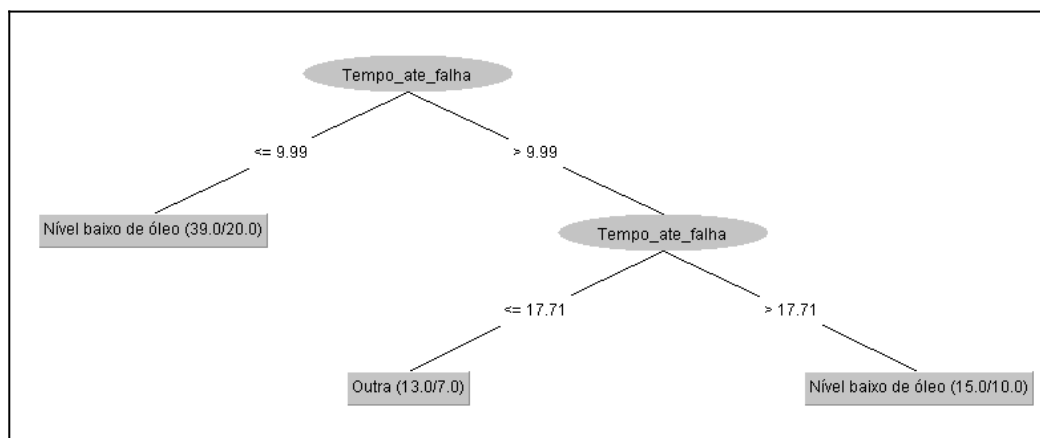


Figura 6 - Árvore de decisão com atributo (I) como folha. Fonte: Própria.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,000	0,000	?	0,000	?	?	0,413	0,085	Quebra de componente
0,500	0,085	0,444	0,500	0,471	0,395	0,729	0,327	Outra
0,885	0,829	0,404	0,885	0,554	0,076	0,546	0,407	Nível baixo de óleo
0,000	0,000	?	0,000	?	?	0,359	0,070	Falta de lubrificação
0,000	0,000	?	0,000	?	?	0,254	0,047	Concentração de óleo
0,000	0,016	0,000	0,000	0,000	-0,035	0,331	0,063	Componentes sujos
0,000	0,000	?	0,000	?	?	0,489	0,147	Vazamento
0,000	0,000	?	0,000	?	?	0,198	0,043	Folga de componentes
0,403	0,333	?	0,403	?	?	0,482	0,241	

Figura 7 - Acurácia detalhada por classe do atributo (I). Fonte: Própria.

```

a b c d e f g h <-- classified as
0 1 5 0 0 0 0 0 | a = Quebra de componente
0 4 4 0 0 0 0 0 | b = Outra
0 2 23 0 0 1 0 0 | c = Nível baixo de óleo
0 1 3 0 0 0 0 0 | d = Falta de lubrificação
0 1 3 0 0 0 0 0 | e = Concentração de óleo
0 0 5 0 0 0 0 0 | f = Componentes sujos
0 0 10 0 0 0 0 0 | g = Vazamento
0 0 4 0 0 0 0 0 | h = Folga de componentes
    
```

Figura 8 - Matriz de confusão para árvore de decisão com atributo (I) como folha. Fonte: Própria.

Já a árvore onde as folhas são o atributo (II), ficou conforme Figura 9. Nela, o modelo classificou mais de 97% de todos os eventos como falha no componente “Sistema hidráulico”. Este modelo obteve uma acurácia de 59,7% com 40 acertos em 67 eventos, como pode ser visto na Figura 10 e na Figura 11.

Esta árvore de decisão possui apenas uma regra, criando uma única regra de classificação. Assim, por mais que o componente “Sistema hidráulico” represente mais de 62% dos eventos registrados, este modelo não atende uma necessidade de predição de falhas por não conseguir prever anomalias de outros componentes.

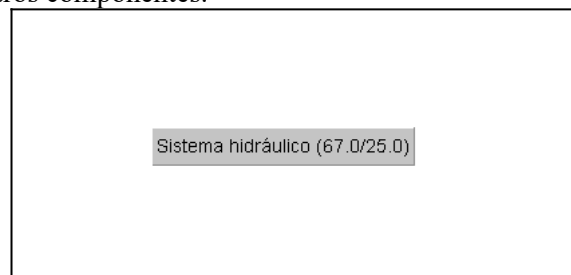


Figura 9 - Árvore de decisão com atributo (II) como folha. Fonte: Própria.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,000	0,043	0,000	0,000	0,000	-0,114	0,444	0,281	Outro
0,952	1,000	0,615	0,952	0,748	-0,135	0,388	0,571	Sistema hidráulico
0,000	0,000	?	0,000	?	?	0,374	0,102	Montagem geral
0,597	0,640	?	0,597	?	?	0,403	0,449	

Figura 10 - Acurácia detalhada por classe do atributo (II). Fonte: Própria.

a	b	c	<-- classified as
0	20	0	a = Outro
2	40	0	b = Sistema hidráulico
0	5	0	c = Montagem geral

Figura 11 - Matriz de confusão para árvore de decisão com atributo (II) como folha. Fonte: Própria.

4. DISCUSSÃO

Com os resultados apresentados na seção anterior, pode-se notar que o sistema hidráulicos de um centro de usinagem é o componente de maior criticidade, uma vez que o maior número de eventos ocorridos no período analisado foi “Nível baixo de óleo”, sendo essa uma falha desse tipo de sistema, e que ambos os modelos de predição obtiveram maior sucesso nestes mesmos tipos de falha. Este resultado corrobora com [16], que indicou este componente como o sistema que mais apresentou falhas dentre 15 sistemas diferentes.

Apesar da precisão com valor próximo a 40%, o modelo que possui o atributo (I) como desfecho demonstra a relevância das árvores de decisão e do algoritmo J48 em aplicações de monitoramento de condição e análise de falha, assim como em [14].

5. CONCLUSÃO

A aplicação da técnica de árvore de decisão cumpriu com o objetivo de avaliar a viabilidade de seu uso para análise de falhas em centros de usinagem. Este resultado se torna relevante tanto para a indústria quanto para a academia, uma vez que a utilização de técnicas de inteligência artificial para predição de falhas e agendamento de manutenções é um assunto que apresenta grande potencial.

Os resultados encontrados nas análises exploratórias estão coerentes com as pesquisas já difundidas na área. Apesar da análise realizada para o atributo (I) ter gerado um modelo de baixa precisão (ligeiramente maior do que 40%), a árvore de decisão gerada obteve uma taxa de acerto maior para o componente tido como o mais crítico. Na contramão do resultado anterior, a análise realizada para o atributo (II) obteve uma precisão superior (acima de 62%), mas sua árvore de decisão não atende o requisito de predição de falhas, uma vez que ela não consegue predizer outro componente em falha senão “Sistema hidráulico”.

Como sugestão de trabalhos futuros, propõe-se a utilização de outras técnicas de classificação, ou ainda aplicar técnicas de recomendação para apresentar qual o melhor período para executar a manutenção. A fim de aprimorar os resultados obtidos, pode-se ainda aplicar as técnicas aqui apresentadas em centros de usinagem instrumentados, a fim de realizar análises com atributos que representam a condição do equipamento antes da falha.

6. REFERÊNCIAS

- [1] A. Ali, The Impact of Innovativeness and Development Time on New Product Performance for Small Firms, *Mark. Lett.* 11 (2000) 151–163. doi:10.1023/A:1008142823872.
- [2] L. Swanson, Linking maintenance strategies to performance, *Int. J. Prod. Econ.* 70 (2001) 237–244. doi:10.1016/S0925-5273(00)00067-0.
- [3] W.R. Marcorin, C.R.C. Lima, Análise dos Custos de Manutenção e de Não-manutenção de Equipamentos Produtivos, 11 (2003) 8.
- [4] L. TAVARES, Administração Moderna da Manutenção, NOVO POLO PUBLICAÇÕES, 2011.
- [5] H.G. Xenos, Gerenciando a Manutenção Produtiva, Edição: 1, Editora GD, 1998.
- [6] C. Scheffer, P. Girdhar, Practical Machinery Vibration Analysis and Predictive Maintenance, Elsevier, 2004.
- [7] K.-A. Nguyen, P. Do, A. Grall, Multi-level predictive maintenance for multi-component systems, *Reliab. Eng. Syst. Saf.* 144 (2015) 83–94. doi:10.1016/j.res.2015.07.017.

- [8] P. Do Van, A. Barros, C. Bérenguer, K. Bouvard, F. Brissaud, Dynamic grouping maintenance with time limited opportunities, *Reliab. Eng. Syst. Saf.* 120 (2013) 51–59. doi:10.1016/j.ress.2013.03.016.
- [9] M.C. MUNION, Predictive vs. preventive: The debate—and the future, *Plant Eng.* (2017). <https://www.plantengineering.com/articles/predictive-vs-preventive-the-debate-and-the-future/> (accessed December 27, 2018).
- [10] C. Yang, J. Liu, Y. Zeng, G. Xie, Real-time condition monitoring and fault detection of components based on machine-learning reconstruction model, *Renew. Energy*. 133 (2019) 433–441. doi:10.1016/j.renene.2018.10.062.
- [11] S. Nallusamy, Enhancement of Productivity and Efficiency of CNC Machines in a Small Scale Industry Using Total Productive Maintenance, *Int. J. Eng. Res. Afr.* 25 (2016) 119–126. doi:10.4028/www.scientific.net/JERA.25.119.
- [12] S. Nandi, H.A. Toliyat, Condition monitoring and fault diagnosis of electrical machines - a review, in: 1999: pp. 197–204.
- [13] R. Liu, B. Yang, E. Zio, X. Chen, Artificial intelligence for fault diagnosis of rotating machinery: A review, *Mech. Syst. Signal Process.* 108 (2018) 33–47. doi:10.1016/j.ymssp.2018.02.016.
- [14] V. Sugumaran, V. Muralidharan, K.I. Ramachandran, Feature selection using Decision Tree and classification through Proximal Support Vector Machine for fault diagnostics of roller bearing, *Mech. Syst. Signal Process.* 21 (2007) 930–942. doi:10.1016/j.ymssp.2006.05.004.
- [15] H.M. Velloso, H.R.M. Da Hora, R.A. de Carvalho, Inteligência Artificial Para Detecção De Falhas Em Equipamentos: Um Estudo Bibliométrico No Setor De Óleo E Gás, in: Penedo/RJ, 2018.
- [16] Z. Yang, C. Chen, F. Chen, Q. Hao, B. Xu, Reliability analysis of machining center based on the field data, *Sci. Technol.* (2013) 9.
- [17] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, S. Mishra, Decision Tree and SVM-Based Data Analytics for Theft Detection in Smart Grid, *IEEE Trans. Ind. Inform.* 12 (2016) 1005–1016. doi:10.1109/TII.2016.2543145.
- [18] PORTAL MÁQUINAS, Portal Máquinas, (2018). <http://portalmaquinas.com.br/limeira/index.php/2-uncategorised/193-mandrilhadora-cnc-wotan-cut-max-3> (accessed December 26, 2018).
- [19] I.H. Witten, E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Second Edition, Elsevier, 2005.
- [20] T.R. Patil, M.S.S. Sherekar, Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification, *Open Access*. 6 (2013) 6.