

# Proposta de aumento na produtividade aplicando algoritmo de aprendizado de máquina em uma atividade administrativa de uma empresa de elevadores

**Autor: Álvaro Kroeff Santos**  
alvaro.kroeff@edu.pucrs.br, PUCRS, Brasil

**Orientador: Edson Zilio Silva**  
edson.silva@pucrs.br, PUCRS, Brasil

**Resumo:** O presente trabalho apresenta uma alternativa redução tempo de um processo administrativo, com sua análise apoiada em algoritmo de aprendizado de máquina em uma empresa do ramo de transportes verticais. A empresa visualizava uma possibilidade de implementar aprendizado de máquina no processo de cadastro de material, visando a redução de tempo em um processo que não agrega valor. Para avaliar e propor soluções para o problema da empresa foi estruturada uma sistemática de treino, teste e validação, com finalidade de avaliar o desempenho da sistemática apoiando o dia-a-dia dos colaboradores. Foi realizado o treino e teste da sistemática, e o cenário final do processo apresentou uma redução de 48% do tempo de cadastro de produtos da empresa.

**Palavras-chave:** Aprendizado de máquina, *machine learning*, algoritmo e cadastro de material.

**Abstract:** This text presents an alternative to reduce the time of an administrative process based on an artificial intelligence algorithm to support the analysis in a company in the vertical transport sector. The company was looking for a possibility to implement artificial intelligence in the material registration process in order to reduce the time of a process that does not add value to the service. To evaluate and propose solutions to the company's problem, this research structured a training, testing and validation system, with the purpose of evaluating performance, improving the day-to-day of employees. The research performed the training and testing of the proposed scheme. The result showed a 48% reduction in the company's product registration time.

**Keywords:** Machine learning, algorithm and material registration.

## 1. Introdução

Segundo Mueller e Massaron (2019), o nascimento da inteligência artificial como se conhece atualmente teve início com a publicação de *Computing Machinery and Intelligence*, de Alan Turing, em 1950. Nesta pesquisa, Turing buscou entender como definir se máquinas são capazes de pensar (MUELLER; MASSARON, 2019). Desde então, os avanços na tecnologia apoiam o desenvolvimento do cenário industrial e o aperfeiçoamento dos processos, a fim de gerar um menor risco de falhas (OLIVEIRA *et al.*, 2017). Por sua vez, na era da informação, o controle dos dados tornou fundamental para a orientação das empresas (KHATRI; BROWN, 2010; URBINATI *et al.*, 2019).

Com base nesses avanços, a ciência de dados redirecionou a produtividade, a inovação e a concorrência (CHEN; CHIANG; STOREY, 2012; MAZZEI; NOBLE, 2017). Prova disto é

que os investimentos em análise de dados possuem uma relação de melhora de 3% até 7% no lucro e na eficiência das organizações (MÜLLER; FAY; VOM BROCKE, 2018). Para Mazzei e Noble (2017), as corporações podem gerar valor por meio dos dados ao eliminar distúrbios na cadeia de valor com mais eficiência e eficácia. Isto porque, atualmente, os dados são produzidos de forma abundante e com muita velocidade, servindo de material de apoio para realizar a tomada de decisão em grandes corporações (THE WORLD'S, 2017).

Segundo Brandão *et al.* (2021), o *machine learning* é uma das áreas do aprendizado de máquina que engloba conhecimento de setores diversos, tais como estatística, engenharia e ciência de dados, no intuito de encontrar soluções diversas para as diferentes questões que a sociedade possui. Por exemplo, nas empresas o aprendizado de máquina vem se tornando cada vez mais uma ferramenta valiosa, uma vez que serve de apoio para a tomada de decisões. Sobre o assunto, Varian (2018) e Mueller e Massaron (2019) ensinam que o aprendizado de máquina consegue realizar análises preditivas de maneira mais rápida e eficiente do que um ser humano. Prova disto é a utilização de algoritmos capazes de analisar dados de ruído e a vibração de uma máquina a fim de identificar se há algum defeito ou mau dimensionamento do equipamento e, por conseguinte, ter a chance de realizar sua manutenção antes mesmo de o equipamento falhar e podendo gerar transtorno em toda a operação da empresa.

Vayghan *et al.* (2007) classifica os dados em três categorias, dados mestres, dados transicionais e dados históricos. Os dados mestre são definidos como as características básicas de entidades empresariais, tais como clientes, produtos, funcionários e fornecedores. Normalmente, os dados mestres são criados uma vez e são utilizados diversas vezes e não sofrem alterações com frequência (KNOLMAYER; ROTHLIN, 2006).

Além disso, sabe-se que o cadastro de dados mestres de produtos é vital para as atividades de uma empresa. Afinal de contas, se a base de dados mestre de material de uma organização está comprometida, todas as análises posteriormente realizadas com base nesses dados e nessas decisões poderão estar prejudicadas também. Ressalte-se que o fato de diversas empresas possuírem reincidência de falhas neste processo não é anormal, uma vez que, na maioria das vezes, a essencialidade do cadastro de dados mestre somente é lembrada após a ocorrência de problemas em algum processo. Neste sentido, algumas pesquisas demonstram que este problema pode ser levado até o ponto de gerar transtornos generalizados nos dados da empresa (FARIA, 2004; IMAM, 2021).

Como os dados mestres são utilizados em menus ou transações nos ERP das empresas, não é apenas a análise de dados que fica comprometida, mas também os processos das empresas. Por exemplo, se o endereço de um cliente estiver errado, isto pode resultar em

pedidos e contas enviadas para o endereço errado, se o preço de um produto estiver errado, pode acarretar que um produto seja vendido abaixo do preço pretendido, se a NCM estiver incorreta, pode gerar multa para a empresa. Portanto, mesmo que uma pequena parcela de dados mestres esteja incorreta, isso pode acarretar uma perda grande da receita de uma empresa (HAUG; ARLBJØRN, 2011).

A partir das leituras e observações anteriormente expostas, este trabalho possui a natureza metodológica de estudo de caso. Para tanto, esta pesquisa partirá do estudo sobre uma empresa do ramo de elevadores, na qual 80% dos produtos possuem alguma característica especial para o cliente. Desta forma, observa-se a necessidade de criação de um novo produto no sistema *Enterprise Resource Planning* (ERP) para que seja possível gerar o custo e a demanda do produto criado, uma vez que o cliente pode requisitar inúmeras características, de maneira a tornar seu elevador singular. Com isso, observa-se que a permissão da singularidade do objeto afeta as áreas de cadastro do produto, pois ele pode vir a ter um material diferente do usual, fator que pode dificultar a localização de informações para realizar a tarefa de cadastrar o material no ERP. Inclusive, caso o *lead time* desta tarefa seja excedido, o cliente pode vir a não receber o produto solicitado no prazo acordado no momento da venda.

Portanto, por três motivos o processo de cadastro de produto foi escolhido para este estudo pela necessidade de a empresa reduzir o custo indireto aplicado nos elevadores vendidos. Primeiro, porque não há previsão para reduzir a demanda de criação de produtos e o *lead time* deste processo. Depois, em razão de toda a etapa de cadastro de produtos desta empresa equivaler, em média, a 120 horas, perfazendo um lapso temporal elevado quando comparado a outras etapas de venda de produto, as quais necessitam de cerca de 100 horas para sua finalização. Por fim, por conta da percepção acerca da possibilidade de o aprendizado de máquina poder fazer parte do processo.

A fim de realizar um recorte de conteúdo para a condução do presente artigo, foi selecionada uma das etapas do cadastro de materiais, qual seja a inclusão da Nomenclatura Comum do Mercosul (NCM). Atualmente, este processo necessita de dois colaboradores responsáveis por realizarem apenas esta demanda. Com base nessas informações e na observação dos fatos, depreendeu-se que o tempo necessário para realizarem suas tarefas é, em média, de 16 horas, equivalente a 13,6% do tempo utilizado para cadastrar um produto no sistema.

Com a finalidade de propor soluções para o problema da companhia, foi determinada a seguinte questão de pesquisa: como diminuir o *lead time* do processo de cadastro de produtos da empresa? Para responder à questão de pesquisa, este estudo tem como objetivo geral propor

a utilização de aprendizado de máquina na tarefa de cadastro de NCM de produto para a redução do *lead time* do processo de cadastro da empresa. Por sua vez, os objetivos específicos buscam apoiar a empresa na solução do problema, tendo sido definidos como (i) propor a redução de custo indireto da empresa; e (ii) validar a acurácia da sistemática proposta.

Por fim, o trabalho possui as seguintes delimitações: (i) este trabalho abrange apenas o processo de cadastro de produtos da empresa, não extrapolando para sua implementação; (ii) o trabalho não possui como foco o processo de desenvolvimento do algoritmo aplicado no processo de cadastro de produto; (iii) este estudo também não contempla a aplicação de aprendizado de máquina de todo o processo de cadastro de produto utilizado na empresa, de tal maneira que apenas abrange a etapa de cadastro de NCM; e (iv) foi aplicado uma constante nos tempos de cadastro, por conta da política da empresa sobre fornecimento de dados

Nesta seção, foram discutidos: o contexto em que o estudo foi desenvolvido, seus objetivos e suas delimitações. Por sua vez, na segunda seção, tratar-se-á a determinação da pesquisa e o método de trabalho utilizado para a construção do estudo. Na terceira, constar-se-á a aplicação prática, a análise dos resultados obtidos e a proposta das melhorias. Na quarta seção, encontrar-se-ão as considerações finais a respeito do estudo.

## **2. Método**

A presente seção está dividida em duas subseções, quais sejam método de pesquisa e método de trabalho. A primeira subseção abordará as principais características da pesquisa e a segunda, a descrição do método de trabalho utilizado.

### **2.1. Método de pesquisa**

Este trabalho é classificado como um estudo de natureza aplicada. Com isso, o conhecimento elaborado para a aplicação prática tem o intuito de resolver um problema específico contido no processo de cadastro de produtos (GERHARDT; SILVEIRA, 2009). Mais especificamente, nesta pesquisa, o conhecimento gerado pela aplicação de um algoritmo de aprendizado de máquina possui efeito prático na solução de distúrbios da empresa investigada.

Os objetivos da pesquisa são considerados exploratórios. O estudo é considerado de cunho exploratório quando há o intuito de viabilizar um maior entendimento do autor em relação ao problema analisado, fazendo com que seja possível a criação de hipóteses para o problema estudado (SILVA; MENEZES, 2005). Neste sentido, a hipótese que orienta este trabalho é que com o auxílio de aprendizado de máquina, possibilita uma maior chance de decisão correta no trabalho.

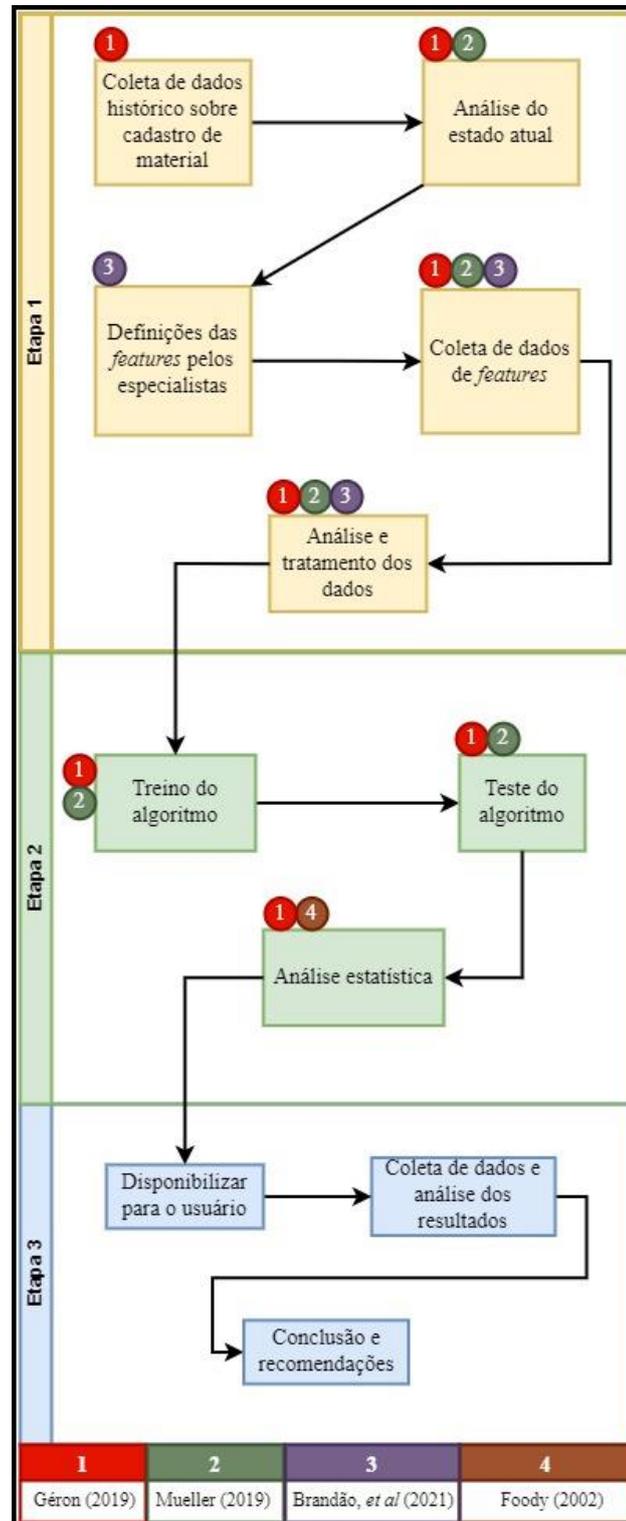
No que diz respeito à abordagem, a pesquisa possui caráter qualitativo e quantitativo. Isto porque as informações que o algoritmo utilizou para realizar as decisões, denominadas de *features*, foram definidas por especialistas da empresa, juntamente com a utilização de bases de dados históricos para a aplicação do algoritmo. Por outro lado, para avaliar os resultados, foram utilizados recursos matemáticos para gerar as evidências.

Segundo Yin (2010) e Gil (2007), considera-se estudo de caso quando um ou poucos objetos são analisados, a fim de obter um vasto detalhamento da situação. Portanto, este trabalho é definido como estudo de caso, pois possui a proposta de analisar a aplicação do aprendizado de máquina em uma empresa real do ramo de transportes verticais (GIL, 2007; YIN, 2010). Outrossim, o estudo engloba um método que possui etapas de coleta e de análise de dados. Em razão disso, o trabalho é considerando um estudo longitudinal, pois os dados utilizados na aplicação do algoritmo são de um longo período, o que permite a comparação entre antes e depois (SILVA; MENEZES, 2005).

## **2.2. Método de trabalho**

Nesta pesquisa, o método do trabalho é constituído por três macroetapas, conforme a Figura 1. Para alcançar o objetivo, foi estabelecida uma sistemática de simulação computacional que possibilitou a aplicação de um algoritmo de computador, cuja fundamentação teve base nos estudos de Géron (2019); Mueller e Masaron (2019); Brandão *et al.* (2021) e Foody (2002). O método proposto (Figura 1) é composto por três fases: (i) concepção; (ii) implementação e (iii) análise. Essas fases são detalhadas nas subseções a seguir.

Figura 1 - Método proposto



Fonte: Adaptado de Géron (2019); Mueller (2019); Brandão, *et al* (2021); Foody (2002).

### 2.2.1. Coleta de dados e análise do estado atual

Na primeira atividade desta etapa do trabalho, foram coletados os dados históricos de seis meses do processo, a fim de analisar estatisticamente o panorama atual do negócio a partir de dados quantitativos (GÉRON, 2019). A coleta de dados ocorre a partir do sistema ERP utilizado pela empresa. Ele consiste em uma tabela com quatro colunas: (i) material; (ii) departamento; (iii) tempo e data de entrada; (iv) tempo e data de saída, na qual o departamento sempre estará igual ao Departamento Tributário (DTRI), haja vista se tratar do local no qual é realizado o cadastro de NCM, que será o foco deste estudo.

Com a coleta realizada, é necessário definir o tempo de cadastro por material. Ele está estabelecido a partir da Equação 1 e do seu tempo médio amostral, definido pela Equação 2. Com a finalidade de identificar se a amostra coletada possui alta variabilidade e quão dispersa está em relação ao tempo médio, foi também calculado o desvio padrão, representado pela Equação 4, e a variância representada pela Equação 3.

Para a realização da análise do estado atual, foi necessário realizar o tratamento dos dados, pois pode haver diversos *outliers* na amostra coletada, por conta de os dados serem de apenas uma variável, que no caso em questão, é o tempo de cadastramento. Foi adotado o método de *Turkey* para determinação de *outliers*, que, segundo Bento *et al* (2018), é um método robusto para a detecção de *outliers* univariados. Para a aplicação deste método, faz-se necessário calcular a amplitude *inter-quartis* representado pela equação 5, onde Q1 é o primeiro quartil e Q3 é o terceiro quartil. E, por fim, os limites inferiores e superiores para o tratamento dos *outliers* representado na equação 6 e 7.

$$\text{Tempo de cadastro} = \text{data e tempo de entrada} - \text{data e tempo de saída} \quad (1)$$

$$\text{Tempo de cadastro médio} = \frac{\sum_{i=1}^n X_i}{n} \quad (2)$$

$$\sigma^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \quad (3)$$

$$\sigma = \sqrt{\sigma^2} \quad (4)$$

$$IQR = Q_3 - Q_1 \quad (5)$$

$$l_{inf} = Q_1 - (1,5 * IQR) \quad (6)$$

$$l_{sup} = Q_3 + (1,5 * IQR) \quad (7)$$

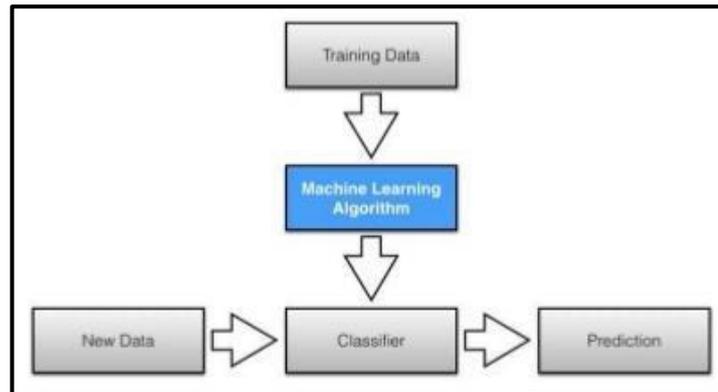
Fonte: BENTO; SANTOS (2018).

### 2.2.2. Definição das *features* pelos especialistas

O processo de aprendizado de máquina de uma sistemática simples é dividido em duas etapas: treinamento e testes Nasteski (2017). Neste processo de treinamento, uma certa

quantidade de dados de amostra são coletadas com as características necessárias para que o algoritmo seja capaz de aprender (SANDHYA; DHAGE, 2016). No processo de testes utiliza-se os dados utilizados de testes junto ao algoritmo para realizar previsões para os dados de teste. Na figura 2, é possível visualizar como ocorre este fluxo de treino e teste de algoritmo de aprendizado de máquina.

Figura 2 – Processo de aprendizado e teste de aprendizado de máquina



Fonte: Nasteski (2017).

Nesta etapa, também é realizado o levantamento das *features* ou atributos que, segundo Brandão *et al.* (2021), são atividades fundamentais para a aplicação do aprendizado de máquina, uma vez que a sistemática utilizará estes dados como balizador da tomada de decisão. Segundo Géron (2019), o sistema só será capaz de aprender se os dados de treino tiverem características relevantes. Neste caso, as seguintes informações foram discutidas em grupo: (i) como o especialista que realiza a tarefa identifica a NCM dos produtos analisados; e (ii) quais informações são necessárias para realizar a escolha correta da NCM. Ambas as questões são relevantes, pois um atributo bem definido auxilia em uma melhor performance do programa, tendo em vista que o algoritmo analisará apenas os dados que, de fato, são essenciais para a tomada de decisão. Para isto, utilizou-se o *brainstorming* para levantar ideias – neste caso possibilidade de *features*. O *brainstorming* promove a criatividade, ou seja, proporciona o compartilhamento de ideias sem que haja críticas (KING; SCHLICKSUPP, 2002).

### 2.2.3. Coleta de dados de *features*

A partir da definição dos atributos, é realizada a coleta dos dados para a aplicação de treino e de teste do algoritmo. De acordo com Géron (2019), por mais simples que a tarefa seja, é necessário o recolhimento e a análise de uma grande quantidade de dados de treino e de teste para que o algoritmo possa, de fato, aprender. Nesta perspectiva, quanto mais complexa a atividade, maior a quantidade de dados utilizados. Em outras palavras, para garantir que a

sistemática esteja aprendendo de fato, faz-se necessário tratá-lo e conduzi-lo para que apenas dados com alta qualidade sejam utilizados na sistemática. Com isso, evita-se que o algoritmo utilize dados cheios de erros e *outliers* para aprender. Exemplo disto é o aprendizado de máquina para a identificação de uma certa nota musical e a utilização de áudios repletos de ruídos como teste, de modo que a máquina terá dificuldades em identificar o som por conta destas falhas.

#### **2.2.4. Análise e tratamento dos dados para aplicar no algoritmo**

Nesta secção é realizado o tratamento dos dados de treino. Existe a possibilidade de haver materiais sem NCM, o que impossibilita saber qual seria a nomenclatura correta do Mercosul. Sendo assim, deverão ser excluídos os dados que possuem a NCM 84313110 que significa partes e peças de elevador e uma grande parte dos dados utilizam esta NCM. Além disso, é de conhecimento da empresa que ela é utilizada quando não se tem total certeza do que se trata o produto, para não afetar a entrega do produto ao cliente. A não realização deste processo de tratamento de dados pode acarretar o erro de *overfitting* ou *underfitting*. *Overfitting*, segundo Faceli *et al.* (2021), refere-se à situação em que um algoritmo apresenta baixa capacidade de realizar uma generalização de decisão. A regra que o algoritmo utiliza para realizar a decisão encontra-se super ajustada aos dados, isto é, que o algoritmo pode estar utilizando dados desbalanceados. O erro de *underfitting* é quando os dados utilizados para o treino possuem pouca representatividade, podendo acarretar uma baixa capacidade de predição.

#### **2.2.5. Treino e teste do algoritmo**

Por se tratar de um algoritmo de aprendizado de máquina, é necessário o fornecimento de um conjunto de dados de treino rotulado. Isto porque o algoritmo não terá um bom desempenho se a quantidade de dados de treino for pequena, conforme indicado na subsecção 2.2.1. Com isso, a única forma de saber se, de fato, o algoritmo terá uma boa generalização em novos casos, é por meio da realização de testes e da aplicação prática. Nestas situações, há duas formas de saber se o algoritmo está bem dimensionado. Uma delas é aplicá-lo em produção e realizar o monitoramento a respeito da qualidade do desempenho. Muito embora esta forma opere bem, se o algoritmo estiver apresentando falhas, os usuários poderão não o aprovar. A melhor opção é dividir o conjunto de dados em dois grupos de treinamento e de teste. Segundo Faceli *et al.* (2021) usualmente é utilizado 70% do conjunto de dados para treino e os 30% restantes para teste. Ao avaliar o algoritmo no conjunto de testes, é possível obter uma estimativa de erro, de tal forma que este valor é capaz de informar o desempenho do algoritmo (GÉRON, 2019).

Segundo Mueller e Massaron (2019), a divisão entre treino e teste deve ser realizada de forma aleatória. Caso contrário, o teste pode não ser confiável, haja vista que a ordenação do conjunto de dados pode causar uma superestimação (quando há significância na ordem dos dados) ou, até mesmo, uma subestimação (quando os dados distribuídos são muito diferentes). Durante a etapa de tratamento de dados de treino, a quantidade de dados que será gerado de treino e teste, precisa ser verificada. Caso finalize em um valor baixo, a forma de treino e teste deverá ser diferente. Na literatura não há um valor de referência para quantidade de dados de treino. Com isto deverá ser avaliado qual tipo de treino utilizar, validação cruzada ou *holdout*.

Segundo Géron (2019) a validação cruzada se adequa bem quando não há grandes quantidades de dados de treino, porém o seu custo operacional é maior, pois é necessário refazer os processos de treino e teste algumas vezes, segundo Rabello (2019), normalmente este processo é realizado entre 5 e 10 vezes. Na figura 3 é possível compreender como os dados de treino e teste devem ser divididos. De acordo com Faceli *et al.* (2021) este método retorna uma aproximação mais segura a respeito do desempenho do algoritmo, por conta de utilizar todos os dados para treino e teste.

Figura 3: Validação cruzada

|              | Dados totais |        |        |        |        |
|--------------|--------------|--------|--------|--------|--------|
| 1ª interação | Teste        | Treino | Treino | Treino | Treino |
| 2ª interação | Treino       | Teste  | Treino | Treino | Treino |
| 3ª interação | Treino       | Treino | Teste  | Treino | Treino |
| 4ª Interação | Treino       | Treino | Treino | Teste  | Treino |
| 5ª interação | Treino       | Treino | Treino | Treino | Teste  |

Fonte: Adaptado de Shaikh (2018)

O método *holdout*, diferente da validação cruzada, realiza apenas uma interação, ou seja, ele acaba sendo mais rápido e menos custoso computacionalmente, visto que não é necessário repetir diversas vezes a etapa de treino e teste. Conforme Faceli *et al.* (2021), na maioria das vezes esta sistemática não possui o melhor desempenho comparado aos demais, contudo, se há uma grande quantidade de dados de treino e teste, o algoritmo acaba possuindo um bom desempenho preditivo. Após realizada a segregação de dados e aplicados tanto o treino quanto o teste, é necessário identificar, a partir de avaliações estatísticas, se o algoritmo está atingindo um bom desempenho.



do algoritmo em formato de executável desenvolvido pela equipe implementadora. Deve-se solicitar que no momento em que o usuário executar o programa, deverá abrir uma janela na área de trabalho do computador do usuário, requerendo o arquivo contendo o código do item e sua respectiva descrição, após isto o algoritmo irá interpretar os dados em que o usuário forneceu em formato *MS Excel*, e o programa deverá disponibilizar um outro arquivo no mesmo formato e local contendo código do item, descrição e NCM escolhida. Ao final disto, quando os usuários poderem realizar a utilização do aplicativo criado, deve-se solicitar que utilizem em um período mínimo de um mês, para que seja possível analisar com mais dados se houve melhora no tempo de cadastro do departamento.

### **2.2.8. Coleta de dados e análise dos resultados**

Para a empresa estudada o mix de produto de um mês pode ser representada pelos produtos analisados em um único mês, visto isto deve-se solicitar que os usuários do algoritmo disponibilizado utilizem o programa desenvolvido pela equipe de tecnologia e informação da empresa em um período mínimo de 1 mês, para que seja possível realizar uma avaliação do tempo de cadastro de produtos utilizando algoritmos.

Nesta etapa é realizada uma nova coleta de dados, semelhante à executada na subseção 2.2.1. para que seja viável analisar se houve ou não diminuição no tempo após a utilização do programa disponibilizado pelos usuários. Juntamente à redução de tempo, fará com que seja possível analisar a viabilidade de realocar ou desligar um dos funcionários, visto que a ideia principal é reduzir o tempo de cadastro de produto, o que poderá resultar em um tempo excedente dos funcionários. E, por fim, identificar se houve melhoria no tempo de cadastro.

## **3. Resultados**

A empresa visualizou a necessidade de reduzir custos indiretos, visto que não era prevista a redução de demanda. Com isso, o processo de cadastro de produtos foi selecionado para a realização do estudo por apresentar sinais claros de ineficiência e com alto custo indireto apropriado. O processo atual consiste em 2 colaboradores que atuam em tempo integral na tarefa.

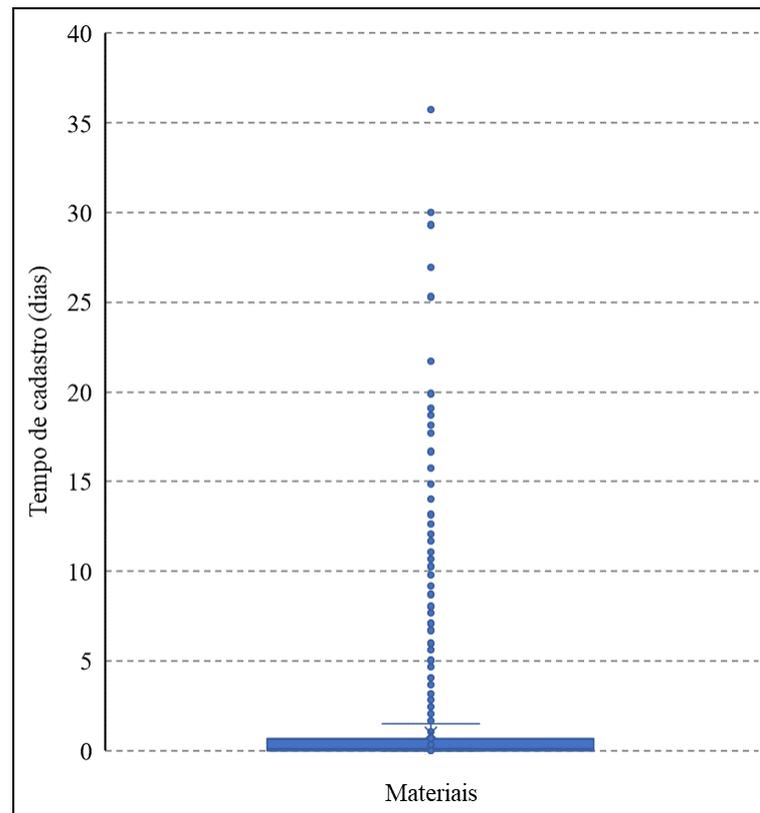
Atualmente a empresa possui diversas necessidades, que geram demandas de criação de produtos como: criação de elevadores completos, modernizar um elevador antigo que está em operação, reparo de um equipamento parado, e isto gera, em média, 2.200 novos materiais no sistema por mês. O método proposto tem como objetivo aplicar aprendizado de máquina fazendo com que melhore a eficiência, reduza o custo indireto da empresa e diminua o tempo de cadastro de produtos.

### 3.1 Coleta de dados e análise do estado atual

A primeira etapa do presente trabalho teve como foco inicial a realização de coleta de dados para possibilitar a análise do estado atual do processo de cadastro de produtos. Para coletar os tempos de cadastro do departamento tributário, foi utilizado um banco de dados do ERP da empresa no qual ficam gravadas as informações de tempo de cadastro. Os dados coletados são referentes às criações de materiais dentre os dias 1º de julho de 2021 e 1º de janeiro de 2022.

O banco de dados da empresa referente à data e tempo de cadastro de produtos foi extraído e armazenado como uma planilha em formato *MS Excel*. Apoiado nesta coleta de dados, calcula-se o tempo de cadastro para cada material conforme equação 1, o tempo médio de cadastro de produtos conforme equação 2, a variância atual do processo de acordo com a equação 3 e o desvio padrão segundo equação 4. O apêndice A representa uma amostra dos dados de tempo extraídos e calculados.

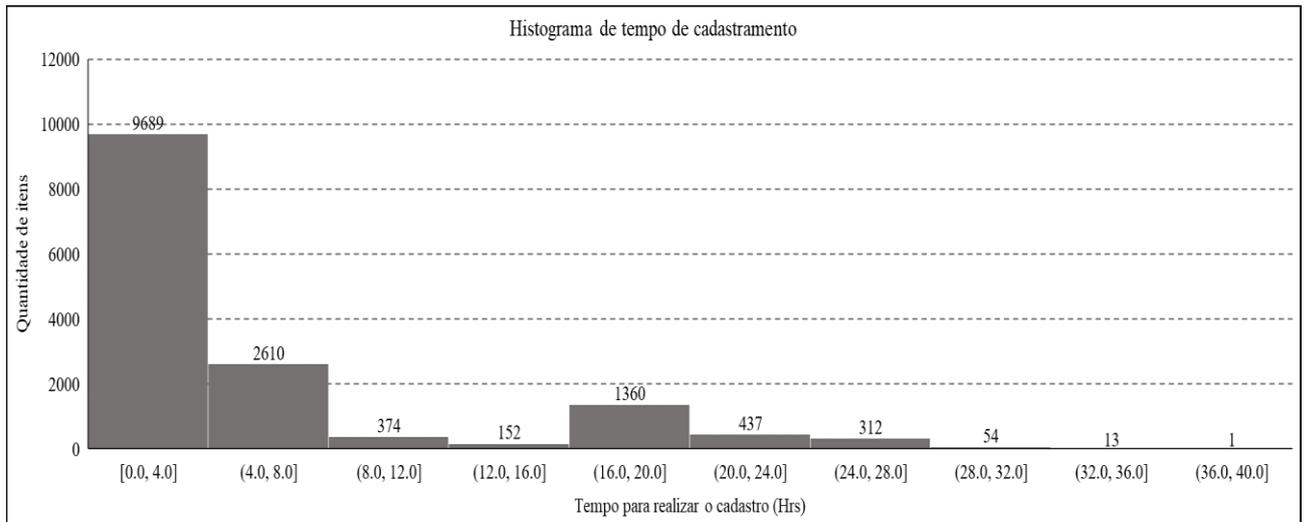
Para verificar a presença de *outliers* na amostra coletada, é utilizado o diagrama de caixa para visualizar a presença de dados além dos limites, pois segundo Rousseeuw e Hubert (2018), todos os dados que estiverem fora destes limites serão definidos como *outlier*. A partir da figura 4 é possível notar a presença de *outliers* na amostra coletada. Com base nisto foi utilizado o método de *Turkey* para identificar os limites e realizar o tratamento dos dados, de modo que todos os dados que estejam ultrapassando os limites são considerados *outliers* e não são utilizados para a análise do estado atual.

Figura 4: *Boxplot* tempo de cadastramento de material

Fonte: Elaborado pelo autor (2022)

Realizado o tratamento é necessária a identificação do processo atual. Utilizando as equações 2, 3 e 4, é possível constatar que o processo possui um tempo médio de cadastro de material de 5 horas, o desvio padrão é igual a 6,9 horas e a sua variância é de aproximadamente de  $47,77h^2$ . Ao plotar os dados já tratados, é possível identificar um maior volume de dados concentrado na média, o que dá a entender que o processo está estável. Porém, como houve a necessidade da exclusão dos *outliers*, que neste trabalho tiveram uma grande representatividade, pode-se concluir que o processo não está estável. De 17.209 dados coletados, 2207 foram considerados estar além dos limites de acordo com o método de *Tukey*, o que significa que 12,8% da amostra foi considerada *outlier*. Os limites encontrados foram calculados conforme as equações 6 e 7. Para o limite superior foi encontrado 39,5 horas e para o inferior foi de -22,7 horas, porém, como não é possível que se tenha um tempo negativo, o limite inferior não foi aplicado no tratamento.

Figura 5: Histograma de tempo de cadastro de material



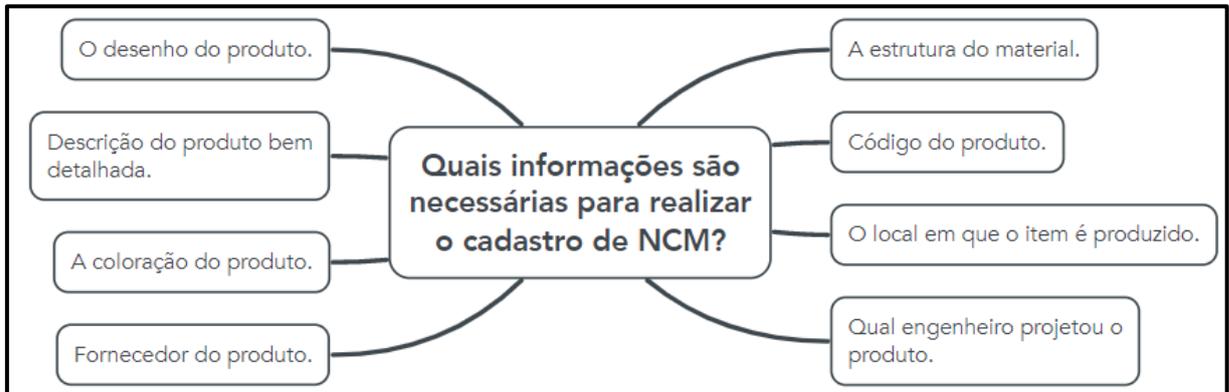
Fonte: Elaborado pelo autor (2022)

### 3.2 Definição das *features* pelos especialistas

De acordo com Faceli *et al.* (2021) a etapa de definição das *features* deve ser definida com cuidado, pois a tendência é que quanto mais dados utilizados, maior a acurácia do algoritmo desenvolvido. Porém, considerando que isto pode afetar o desempenho do algoritmo, por conta de o processamento de dados ser muito longo, é preciso definir bem seus atributos. Visto a necessidade de dar seguimento no projeto, é feito o levantamento dos atributos necessários para a aplicação do *machine learning* no processo.

Estas definições foram realizadas a partir de um grupo, onde continha o autor deste artigo, os dois funcionários que realizam a atividade, o coordenador e supervisor da área de tributos e o coordenador da equipe implementadora. Ao inicializar o levantamento das ideias através do *brainstorming*, foi feito o seguinte questionamento para possibilitar o direcionamento da reunião: quais informações são necessárias para realizar o cadastro de NCM dos novos produtos? Cada participante contribuiu durante 8 minutos suas ideias em uma folha de papel A4. No final deste tempo, as informações foram agrupadas na plataforma digital *Mindmeister* a fim de possibilitar que todos os presentes na reunião visualizassem o que havia sido levantado. O resultado do *brainstorming* foi o seguinte, conforme a figura 6.

Figura 6: *Brainstorm* realizado na reunião



Fonte: Elaborado pelo autor (2022)

Após a realização do brainstorming, equipe que está implementando a aprendizagem de máquina no processo definiu as *features* que serão utilizadas. Foram adicionadas descrições detalhadas do produto, pois esta já informa do que o produto é constituído e suas principais características. E também o código do produto, pois na empresa estudada é utilizada uma lógica para cadastro de produto que possui diversas informações sobre o produto. A figura 7 exemplifica a lógica de codificação utilizada pela empresa.

Figura 7: Lógica de codificação de produto

| Grupo   | Subgrupo                   | Família                                     | Sequencial   | Código                   |
|---|----------------------------|---|--|--------------------------|
| I   | A                          | 5300  | AF   | 1A.5300.AF               |
| Localização onde o produto é aplicado no elevador | + Principal função do item | + Família de produto que o item se enquadra | + Dois caracteres alfanuméricos, para permitir repetição das categorias passadas | = Versão final do código |

Fonte: Elaborado pelo autor (2022)

### 3.3 Coleta de dados de *features*

Após a equipe que irá implementar a melhoria realizar a definição dos atributos necessários, foi necessário realizar o levantamento dos dados de treino e teste, foi utilizada a definição realizada no capítulo anterior. Um arquivo em formato *MS Excel* com todos os materiais criados dentro o período de 1º de janeiro de 2021 até 31 de dezembro de 2021 foi retirado do sistema ERP, o que resultou em 46.609 resultados. Estes dados podem ser visualizados no apêndice B.

Estes dados já possuíam NCM cadastrada anteriormente pelos especialistas, com isso não houve a necessidade de identificar a NCM dos itens, para termos a certeza de que o algoritmo estava de fato errando ou acertando. Uma vez que os dados coletados já foram analisados pelos especialistas anteriormente, entende-se que os dados estão corretos e sem divergências, podendo ser aplicados no treino e no teste do algoritmo para que se inicie o aprendizado e fazendo com que seja possível mensurar a eficiência do algoritmo.

### **3.4. Análise e tratamento dos dados para aplicar no algoritmo**

Conforme citado na subseção 2.2.4, há necessidade de tratar os dados para a realização de treino e teste, e, com isto, foram aplicados os filtros para possibilitar a aplicação do treino e teste do algoritmo. O resultado gerado foi que, dos 46.609 itens retirados do sistema, apenas 10.137 estavam aptos a serem aplicados, pois havia 29.473 produtos com a NCM de partes e peças de elevador, o que deixaria a base de dados desbalanceada. Também havia 6.999 materiais sem NCM, que são considerados dados de treino não representativos, por conta de não ter certeza da sua NCM e não ser possível treinar o algoritmo.

### **3.5. Treino e teste do algoritmo**

Após retirados os dados de treino e teste, que resultou em um total de 10.137 dados, é necessário separar os dados para treino e teste. Conforme indicado na subseção 2.2.5, que se utilizasse 70% para a realização de treino e 30% para teste, o resultado foi de 7.096 materiais para treino e 3.041 produtos para teste. O método definido para esta aplicação foi a de validação cruzada, devido a quantidade de dados ser considerada baixa; ao separar por classes, algumas acabam ficando sem muitos dados para treino e teste, fazendo com que o algoritmo esteja sob efeito de *overfitting* e *underfitting*.

### **3.6. Análise estatística do algoritmo**

Para se identificar como o algoritmo está comportando é necessário realizar a análise quantitativa dele, que pôde ser feita a partir das equações 8, 9 10 e 11. Estas equações foram aplicadas a partir do algoritmo desenvolvido pela equipe implementadora. Parte do algoritmo pode ser visto no apêndice D e os dados utilizados para aplicar este algoritmo foram como o da amostra do apêndice C, pois ela representa bem os dados utilizados nesta sistemática.

A saída do algoritmo do apêndice D resulta uma matriz, com os cabeçalhos de coluna igual a precisão, sensibilidade e *F1-score*, e seus cabeçalhos de linha são as respectivas classes utilizadas, que no trabalho em questão são as NCM que o programa irá prever, estes dados poderão ser vistos no apêndice E. Ao final desta matriz será possível verificar a precisão, sensibilidade e *F1-score* médio deste algoritmo. Na tabela 1 estão os valores médios que o algoritmo encontrou e sua acurácia. No apêndice E, e nas figuras 8, 9 e 10 é possível verificar

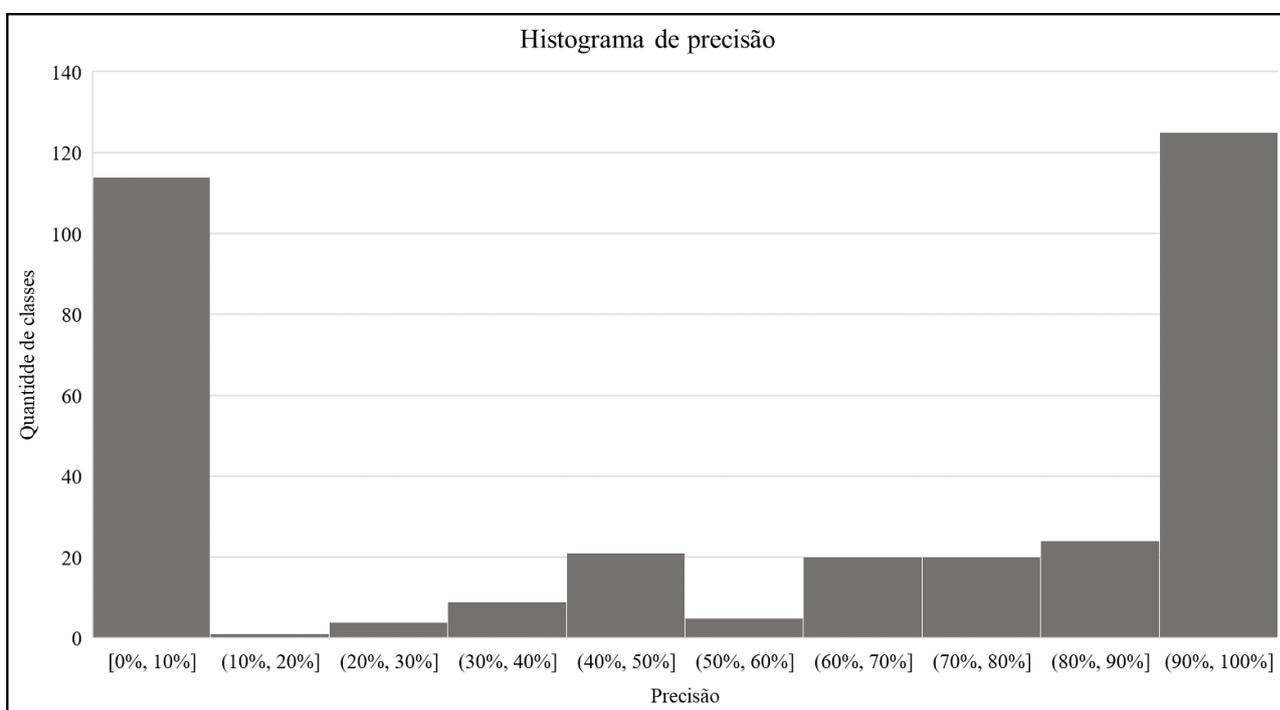
os valores de precisão, sensibilidade e *F1-score* para cada classe em forma de tabela e histograma.

Tabela 1: Resultado do algoritmo para as equações 8, 9, 10 e 11

|                  | <b>Precisão</b> | <b>Sensibilidade</b> | <b>F1-Score</b> |
|------------------|-----------------|----------------------|-----------------|
| Média aritmética | 56%             | 56%                  | 55%             |
| Média ponderada  | 82%             | 84%                  | 82%             |
| Acurácia         | 84%             |                      |                 |

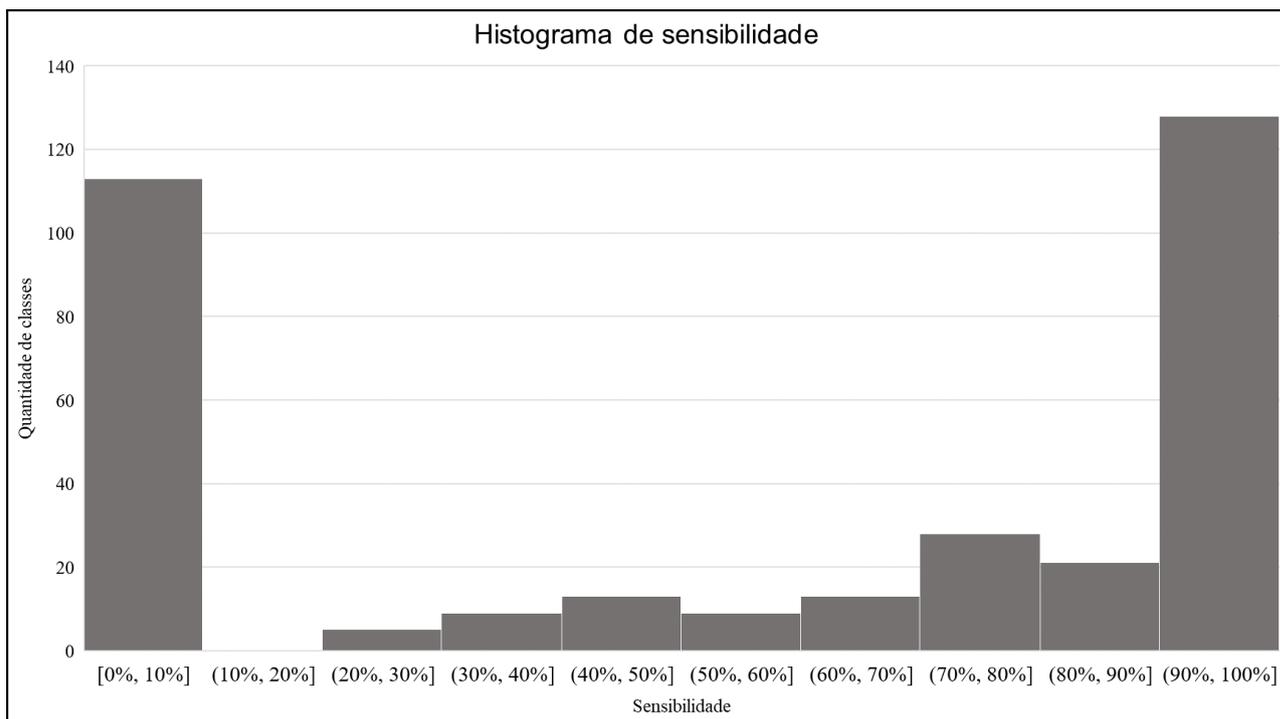
Fonte: Elaborado pelo autor (2022)

Figura 8: Histograma de precisão do teste.



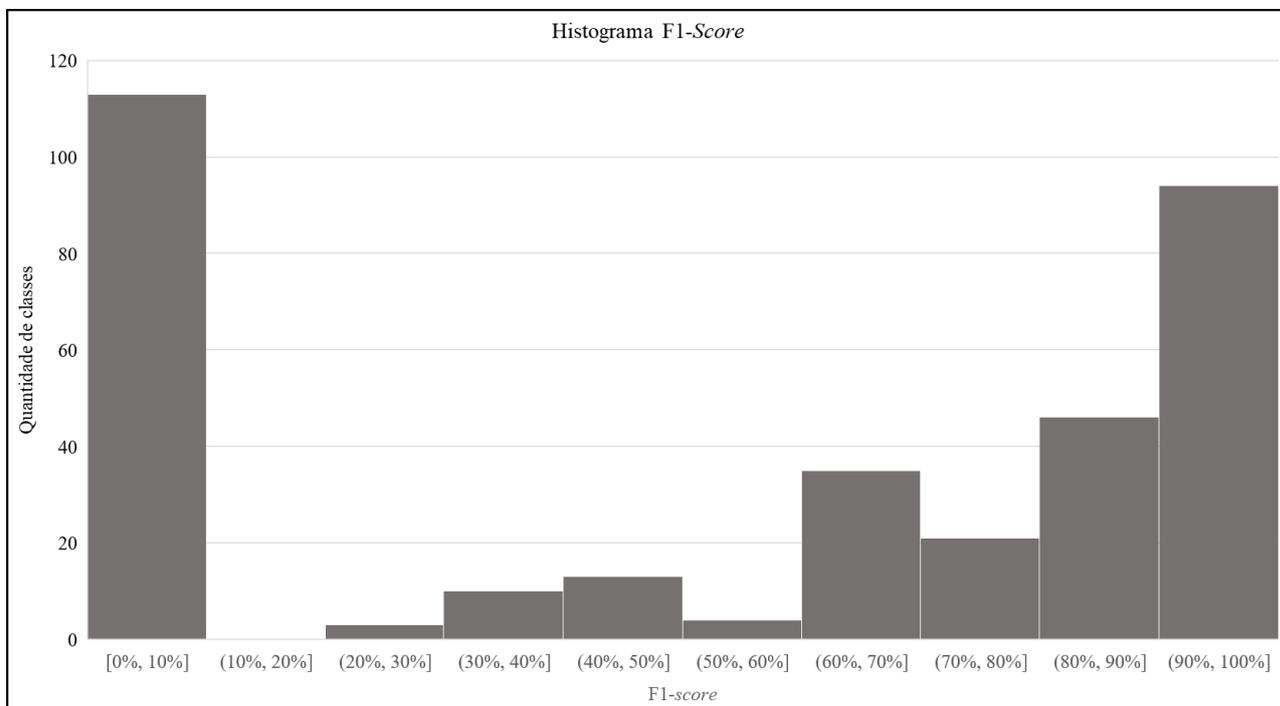
Fonte: Elaborado pelo autor (2022)

Figura 9: Histograma de sensibilidade do teste.



Fonte: Elaborado pelo autor (2022)

Figura 10: Histograma de F1-score do teste.



Fonte: Elaborado pelo autor (2022)

### 3.7. Disponibilizar para o usuário

A equipe de tecnologia e informação foi contatada para que realizar a formatação do algoritmo em forma executável e foi solicitado que a única entrada que o usuário deve fornecer para que o programa rode seja uma tabela em formato *MS Excel* contendo apenas 2 colunas, código e descrição. Estes foram os atributos definidos pelos especialistas para que o programa consiga sugerir uma NCM. Com isso o usuário terá de dispor de um arquivo no formato *MS Excel* em sua área de trabalho com os produtos cuja NCM precisa descobrir. A saída é de uma cópia do arquivo com a NCM decidida pelo algoritmo em uma coluna ao lado das de código e descrição.

Ao salvar o arquivo em sua área de trabalho contendo as informações necessárias, o usuário terá acesso ao algoritmo em formato de executável. Ao clicar neste aplicativo o programa solicita que seja selecionado o arquivo dos materiais em que se deseja descobrir a NCM. Ao realizar este direcionamento, o algoritmo, com base nos treinos e testes feitos, decidirá qual é a melhor NCM para o produto em questão, e disponibilizará um arquivo no mesmo local e formato que o primeiro – apenas com uma diferença, neste conterá uma coluna com a NCM decidida pelo algoritmo de *machine learning*.

### 3.8. Coleta de dados e análise dos resultados

Após a disponibilização dos produtos aos dois usuários que realizam a tarefa de cadastro de NCM no dia-a-dia, foi solicitado a eles que utilizassem a ferramenta criada em um período constante de no mínimo um mês, fazendo com que a coleta de dados desta seção possibilitasse verificar se houve redução de tempo no processo. Com isso foi retirado o tempo de todos os materiais cadastrados pela equipe no período de 1º de março de 2022 até 31 de março de 2022 para a realização desta análise.

Conforme comentado na subseção 2.2.8 a coleta de dados foi semelhante à realizada na subseção 3.1. O sistema *ERP* da empresa foi utilizado para a realização desta coleta, pois é neste sistema que fica gravado o tempo que os usuários levam para a realização do processo. No apêndice F, que é uma representatividade destes dados coletados, é possível identificar o tempo que cada material levou para a realização deste processo.

Foi constatado que houve uma diminuição no tempo de cadastro de produtos da empresa. Anteriormente foi encontrado um tempo médio de cadastro de NCM de 5 horas, contudo para retirar esta média foram retirados os tempos de cadastro que representavam *outliers*. Para esta nova coleta de dados foi encontrada uma média de cadastro de 2,4 horas, ou

seja, houve uma diminuição no tempo de cadastro de 48,28% e neste caso não houve tratamento para a retirada de *outliers*.

Por conta do tempo de cadastro ter reduzido consideravelmente, foi proposto para os gestores da área realizar o desligamento de um dos funcionários, com isso seria possível reduzir o custo indireto da área, contudo, não foi possível realizar o manejo dos funcionários para que não estivessem mais alocados no centro de custo do departamento tributário por conta dos resultados estatísticos encontrados na seção 3.6. Para algumas classes o algoritmo apresentou-se bem robusto, para alguns outros, no entanto, o algoritmo não teve uma boa previsão de NCM, podendo acarretar erro de cadastro e fazendo com que os dados mestres dos materiais estejam errados. No entanto, foi possível organizar as tarefas de um dos dois funcionários de modo que em uma parte do tempo contribuísse com o cadastro de materiais e em outra parte do tempo com demais atividades do setor.

#### **4. Considerações Finais**

Para avaliar e propor soluções para o problema da empresa, foi definida a seguinte questão de pesquisa: como diminuir o *lead time* do processo de cadastro de produtos da empresa? Para responder à questão de pesquisa com base na literatura, foi definido que um *software* de aprendizado de máquina conseguiria prever informações. Foi escolhida a utilização de *machine learning*, pois com ela se pode realizar treinamentos e testes do algoritmo sem que isso afete o processo atual.

O presente trabalho teve como objetivo principal propor uma sistemática de aprendizado de máquina na tarefa de cadastro de NCM de produto para a redução do *lead time* do processo de cadastro da empresa. E teve como objetivos específicos reduzir o custo indireto do processo e encontrar a acurácia da sistemática proposta. Os objetivos principais e específicos do trabalho foram parcialmente atingidos, pois foi possível analisar o processo sem a sistemática e com a sistemática e foi constatado que houve diminuição no tempo de cadastro o que resulta em um decréscimo de *lead time*. Com esta sistemática, também foi possível encontrar a acurácia do algoritmo, contudo não foi possível reduzir o custo indireto da empresa.

A utilização de aprendizado de máquina se mostrou eficiente, pois ela permitiu gerar opções de NCM para os funcionários que realizam as tomadas de decisão, podendo fazer com que realizem as tarefas em um menor tempo e com boa acuracidade. Considerando o mercado atual e a rapidez da informação, essa ferramenta se demonstra cada vez mais útil para as empresas.

Como recomendação, sugere-se que a empresa acompanhe a criação de novos dados de treino e teste do algoritmo, para verificar se a realização de mais treinos com as classes que obtiveram um pior desempenho estatisticamente possibilitaria a redução de custo na empresa, pois foi notado que houve um bom desempenho para demais classes. Para estudos futuros, o trabalho pode ser ampliado, sugerindo a comparação entre diferentes algoritmos de aprendizado de máquina e verificando se haverá alteração de desempenho.

## Referências

BENTO, G. M.; SANTOS, R. T. Avaliação de métodos de remoção de outliers e seus impactos na precisão dos métodos de interpolação. *In: 1º SIMPÓSIO MATO-GROSSENSE DE MECANIZAÇÃO AGRÍCOLA E AGRICULTURA DE PRECISÃO – SIMAP. Anais [...]*. Universidade Federal de Mato Grosso, Sinop, 2018. Disponível em: [http://www.lapmec.com.br/upload/mod\\_publicacoes/251/5b3e66947f420.pdf](http://www.lapmec.com.br/upload/mod_publicacoes/251/5b3e66947f420.pdf). Acesso em: 21 abr. 2022.

BRANDÃO, M. S.; GODINHO-FILHO, M.; AZZOLINI JUNIOR, W.; BATTISSACCO, B.C.; MARÇOLA, J.A. Melhoria da categorização de produtos a partir do uso de algoritmos de aprendizado de máquina e medidas de similaridade. **Revista Produção Online**. Florianópolis, v. 21, n.4, p. 2093-2124, 2021. Disponível em: <https://producaoonline.org.br/rpo/article/view/4483/2125>. Acesso em: 06 jun. 2022.

CHEN, H.; CHIANG, R. H. L.; STOREY, V. C. Business intelligence and analytics: from big data to big impact. **Management Information Systems Quarterly**. Minneapolis, v. 36, n. 4, p. 1165-1188, dez. 2012. Disponível em: <https://www.jstor.org/stable/41703503>. Acesso em: 21 abr. 2022.

DUARTE, G. Como detectar e tratar outliers – Parte 1. **Medium**, [S.l.] 5 mar. 2021. Disponível em: <https://medium.com/datarisk-io/como-detectar-e-tratar-outliers-parte-1-e9395c2d05de#:~:text=Outro%20m%C3%A9todo%20muito%20comum%20para,observa%C3%A7%C3%A3o%20%C3%A9%20marcada%20como%20outlier>. Acesso em: 05 jun. 2022.

FACELI, K.; LORENA, A.C.; GAMA, J.; CARVALHO, A.C.P.L.F. **Inteligência artificial: Uma abordagem de aprendizado de máquina**. 2. ed. Rio de Janeiro: Livros Técnicos e Científicos Editora Ltda, 2021.

FARIA, N. C. **Cadastro de Materiais - Um Tesouro Ignorado pelas Empresas**. Guia Log, maio 2004. Disponível em: <https://www.guialog.com.br/Y542.htm>. Acesso em: 04 abr. 2022.

FOODY, G. M. Status of land cover classification accuracy assessment. **Remote Sensing of Environment**. [S.l.], v. 80, n. 1, p. 185-201, abr. 2002. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0034425701002954>. Acesso em: 21 abr. 2022.

FOLZ, J. K. **Aprendizado de máquina para detecção de anomalias em máquinas rotativas por sinais de vibração**. 2021. Trabalho de Conclusão de Curso (Graduação em

Engenharia Mecânica) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2021. Disponível em: <https://lume.ufrgs.br/handle/10183/234141#>. Acesso em: 21 abr. 2022.

GERHARDT, T. E.; SILVEIRA, D. T. (orgs.). **Métodos de pesquisa**. Porto Alegre: Editora da UFRGS, 2009. Disponível em: <https://lume.ufrgs.br/handle/10183/52806>. Acesso em: 21 abr. 2022.

GÉRON, A. **Mãos à obra**: Aprendizado de máquina com Scikit-Learn & TensorFlow. Rio de Janeiro: Alta Books, 2019.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2007.

HAUG, Anders; ARLBJØRN, Stentoft. Barriers to master data quality. **Journal of Enterprise Information Management**, [S.l.], v. 24, n. 3, p. 288-303, 2011.

IMAM. Padrão Descritivo de Materiais – PDM. **IMAM**, 2021. Disponível em: <https://www.imam.com.br/consultoria/artigo/pdf/padrao-descrtivo-de-materiaispedm.pdf>. Acesso em: 4 abr. 2021.

KING, B.; SCHLICKSUPP, H. **Criatividade**: Uma vantagem competitiva. Rio de Janeiro: Qualitymark Ed., 2002.

KHATRI, V.; BROWN, C. V. Designing data governance. **Communications of the ACM**. [S.l.], v. 53, n. 1, p. 148-152, jan. 2010. Disponível em: <https://dl.acm.org/doi/10.1145/1629175.1629210>. Acesso em: 21 abr. 2022.

KNOLMAYER, G; ROTHLIN, M. Quality of material master data and its effect on the usefulness of distributed ERP systems. **Lecture Notes in Computer Science**, Berlim, v. 4231, pp. 362-71, 2006.

MAZZEI, M. J.; NOBLE, D. Big data dreams: A framework for corporate strategy. **Business Horizons**. Bloomington, v. 60, n. 3, p. 405-414, maio/jun. 2017. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0007681317300101>. Acesso em: 21 abr. 2022.

MUELLER, J. P.; MASSARON, L. **Aprendizado de máquina para leigos**. Rio de Janeiro: Alta Books, 2019.

MÜLLER, O.; FAY, M.; VOM BROCKE, J. The effect of big data and analytics on firm performance: An econometric analysis considering industry characteristics. **Journal of Management Information Systems**. Londres, v. 35, n. 2, p. 488-509, maio 2018.

OLIVEIRA, D. L.; SANTOS, P. P. P.; MAZZUCO, D. E.; FRAZZON, E. M. Regras de despacho aplicadas ao problema do roteamento de veículos: uma comparação via simulação. In: XXXVII Encontro Nacional de Engenharia de Produção, 2017, Joinville. **Anais do Encontro Nacional de Engenharia de Produção**. Rio de Janeiro: ABEPRO, 2017. p. 1-14. Disponível em: [https://www.abepro.org.br/biblioteca/TN\\_STO\\_238\\_376\\_33491.pdf](https://www.abepro.org.br/biblioteca/TN_STO_238_376_33491.pdf). Acesso em: 21 abr. 2022.

RABELLO, E. B. **Cross Validation: avaliando seu modelo de machine learning**. Medium. 8 jul. 2019. Disponível em: <https://medium.com/@edubrazrabello/cross-validation-avaliando-seu-modelo-de-machine-learning-1fb70df15b78>. Acesso em: 05 jun.2022.

ROUSSEEUW, P. J.; HUBERT, M. Anomaly detection by robust statistics. **WIREs Data Mining and Knowledge Discovery**. [S.l.] v. 8, n. 2, p.1-14, e1236, 2018. Disponível em: <https://wires.onlinelibrary.wiley.com/doi/full/10.1002/widm.1236>. Acesso em: 05 jun. 2022.

SILVA, E. L.; MENEZES, E. M. **Metodologia da pesquisa e elaboração de dissertação**. 4. ed. Florianópolis: Laboratório EAD da UFSC, 2005.

SHAIKH, R. **Cross validation explained: evaluating estimator performance**. Medium. 26 nov. 2018. Disponível em: <https://towardsdatascience.com/cross-validation-explained-evaluating-estimator-performance-e51e5430ff85>. Acesso em 06 jun. 2022.

THE WORLD'S most valuable resource is no longer oil, but data. **The Economist**. Londres, 06 maio 2017. Disponível em: <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>. Acesso em: 21 abr. 2022.

URBINATI, A.; BOGERS, M.; CHIESA, V.; FRATTINI, F. Creating and capturing value from big data: A multiple-case study analysis of provider companies. **Technovation**. [S.l.], v. 84-85, p. 21-36, jun./jul. 2019. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0166497218300361>. Acesso em: 21 abr. 2022.

VARIAN, H. Artificial intelligence, economics, and industrial organization. *In*: AGRAWAL, Ajay; GANS, Joshua; GOLDFARB, Avi (eds.). **The economics of artificial intelligence: an agenda**, pp. 399-419. Chicago: University of Chicago Press, 2018.

VAYGHAN, J. A.; GARFINKLE, S. M.; WALENTA, C.; HEALY, D. C.; VALENTIN, Z. The internal information transformation of IBM. **IBM Systems Journal**, [S.l.], v. 46, n. 4, p. 669-84, 2007.

YIN, R. K. **Estudo de caso: planejamento e métodos**. 4. ed. Porto Alegre: Bookman, 2010.

## Apêndices

### Apêndice A: Dados coletados de tempo de cadastro

| Cenário EPDM                  | Material       | Visão | Entrada_total    | Saida_total      | Tempo de cadastro (dias) | Tempo de cadastro (Hrs) |
|-------------------------------|----------------|-------|------------------|------------------|--------------------------|-------------------------|
| NEW                           | X26.043.884    | DTRI  | 04/11/2021 14:15 | 05/11/2021 15:56 | 1.07                     | 25.69                   |
| NEW                           | X26.155.143    | DTRI  | 07/03/2022 11:18 | 08/03/2022 12:58 | 1.07                     | 25.66                   |
| NEW                           | X26.015.167    | DTRI  | 09/12/2021 13:11 | 10/12/2021 14:43 | 1.06                     | 25.53                   |
| NEW                           | 5F.0599.AY.2   | DTRI  | 05/05/2022 07:50 | 06/05/2022 09:18 | 1.06                     | 25.46                   |
| NEW                           | 178841-ATR-004 | DTRI  | 30/11/2021 13:36 | 01/12/2021 15:00 | 1.06                     | 25.40                   |
| NEW                           | 4F.0599.BA.5   | DTRI  | 04/05/2022 17:17 | 05/05/2022 18:38 | 1.06                     | 25.34                   |
| NEW                           | 4F.0599.BA.4   | DTRI  | 04/05/2022 17:17 | 05/05/2022 18:38 | 1.06                     | 25.34                   |
| NEW                           | 4F.0599.BA.6   | DTRI  | 04/05/2022 17:17 | 05/05/2022 18:38 | 1.06                     | 25.34                   |
| NEW                           | 3Y.0084.VX.1   | DTRI  | 13/12/2021 11:42 | 14/12/2021 12:58 | 1.05                     | 25.27                   |
| NEW                           | X26.123.041    | DTRI  | 29/03/2022 11:20 | 30/03/2022 12:34 | 1.05                     | 25.24                   |
| NEW                           | X26.130.055    | DTRI  | 29/03/2022 11:20 | 30/03/2022 12:33 | 1.05                     | 25.21                   |
| NEW                           | X26.149.535    | DTRI  | 04/05/2022 09:40 | 05/05/2022 10:53 | 1.05                     | 25.21                   |
| NEW                           | X26.102.680    | DTRI  | 02/03/2022 10:59 | 03/03/2022 11:59 | 1.04                     | 25.00                   |
| NEW                           | X26.102.679    | DTRI  | 02/03/2022 10:59 | 03/03/2022 11:54 | 1.04                     | 24.91                   |
| NEW                           | 181917-IRP-020 | DTRI  | 04/05/2022 09:49 | 05/05/2022 10:43 | 1.04                     | 24.89                   |
| NEW                           | X26.102.678    | DTRI  | 02/03/2022 10:59 | 03/03/2022 11:52 | 1.04                     | 24.88                   |
| NEW                           | 181917-IRP-002 | DTRI  | 04/05/2022 09:49 | 05/05/2022 10:41 | 1.04                     | 24.87                   |
| NEW                           | X26.102.677    | DTRI  | 02/03/2022 10:59 | 03/03/2022 11:51 | 1.04                     | 24.86                   |
| NEW                           | MG0005DEFU0002 | DTRI  | 07/12/2021 16:53 | 08/12/2021 17:43 | 1.03                     | 24.82                   |
| NEW                           | MG0005DEFU0006 | DTRI  | 07/12/2021 16:54 | 08/12/2021 17:43 | 1.03                     | 24.82                   |
| NEW                           | MG0005DEFU0003 | DTRI  | 07/12/2021 16:53 | 08/12/2021 17:43 | 1.03                     | 24.82                   |
| NEW                           | MG0005DEFU0004 | DTRI  | 07/12/2021 16:54 | 08/12/2021 17:43 | 1.03                     | 24.82                   |
| NEW                           | MG0005DEFU0005 | DTRI  | 07/12/2021 16:54 | 08/12/2021 17:43 | 1.03                     | 24.82                   |
| NEW                           | 178282-BEM-001 | DTRI  | 02/05/2022 08:48 | 03/05/2022 09:34 | 1.03                     | 24.77                   |
| NEW                           | 178282-BEM-010 | DTRI  | 02/05/2022 08:48 | 03/05/2022 09:34 | 1.03                     | 24.77                   |
| NEW                           | 178282-CPC-007 | DTRI  | 02/05/2022 08:49 | 03/05/2022 09:33 | 1.03                     | 24.75                   |
| NEW                           | 178282-CPC-071 | DTRI  | 02/05/2022 08:49 | 03/05/2022 09:33 | 1.03                     | 24.75                   |
| NEW                           | 150208-PPC-110 | DTRI  | 04/05/2022 10:05 | 05/05/2022 10:48 | 1.03                     | 24.72                   |
| NEW                           | 178282-CFC-012 | DTRI  | 02/05/2022 08:48 | 03/05/2022 09:31 | 1.03                     | 24.71                   |
| Tempo médio de cadastro (Hrs) | 23.05          |       |                  |                  |                          |                         |
| Mediana (Hrs)                 | 2.36           |       |                  |                  |                          |                         |
| Desvio padrão (Hrs)           | 65.86          |       |                  |                  |                          |                         |
| Variância (Hrs <sup>2</sup> ) | 4336.94        |       |                  |                  |                          |                         |
| P25                           | 0.61           |       |                  |                  |                          |                         |
| P75                           | 16.17          |       |                  |                  |                          |                         |
| IQR                           | 15.55          |       |                  |                  |                          |                         |
| Limite superior               | 39.50          |       |                  |                  |                          |                         |
| Limite inferior               | -22.72         |       |                  |                  |                          |                         |

## Apêndice B: Dados coletados para treino e teste

| Material      | TxtBreveMaterial                         | Cód.controle |
|---------------|--|--------------|
| 4C.0430.WK.3  | PATS - WI = 1752,6, ALE                  | 84313110     |
| 4C.0297.SB.2  | CR - OP = 1066,8 , ALD                   | 84313110     |
| 4B.0403.AX.10 | HASTE ACIONAMENTO DO APARELHO DE         | 84313110     |
| 4B.0403.AX.10 | HASTE ACIONAMENTO DO APARELHO DE         | 84313110     |
| 4B.0403.AX.10 | HASTE ACIONAMENTO DO APARELHO DE         | 84313110     |
| 4B.0403.AX.10 | HASTE ACIONAMENTO DO APARELHO DE         | 84313110     |
| 4B.0403.AX.10 | HASTE ACIONAMENTO DO APARELHO DE         | 84313110     |
| 3Y.0158.AN.57 | CONJUNTO TRAVESSA SUPERIOR               | 84313110     |
| 4C.0053.F.2   | RTSR; LAT. DIR.                          | 84313110     |
| 3Z.0094.WT.4  | SUPORTE - MC                             | 84313110     |
| 3Y.0447.AW.27 | TRAVESSA DO FUNDO; WI:2150 CONJ.PLATAFOR | 84313110     |
| 3Y.0215.FV.5  | PLACA PROTEÇÃO FUNDO; WI: 1500           | 84313110     |
| 3Y.0309.AT.10 | RODAPE P/ BALAUSTRADA                    | 84313110     |
| 3Y.0309.AT.14 | RODAPE P/ BALAUSTRADA                    | 84313110     |
| 3Y.0309.AT.18 | RODAPE P/ BALAUSTRADA                    | 84313110     |
| 4C.0316.AG.9  | PLACA TETO; FRENTE                       | 84313110     |
| X26.201.120   | BOTÃO VENUS 2.0 AZUL GRAV.SOBE PRAT.CIRC | 85365090     |
| 4C.0316.XF.4  | PLACA TETO 6-8P SYNERGY SMP; WI-1100 x D | 84313110     |
| 4C.0316.XH.2  | PLACA TETO 9-10P SYNERGY SMP; WI-1200 X  | 84313110     |
| 2A.0059.JA.2  | BASE MAQUINA; 3500#,4CINTAS,NS           | 84313110     |
| 2B.0106.YE.1  | SUPORTE SUPERIOR FITA APS - "L"          | 84313110     |
| 2B.0114.V.1   | MORDENTE PASSANTE SUPORTE - 1 FURO       | 84313110     |
| 2B.0114.V.2   | MORDENTE PASSANTE SUPORTE - 2 FUROS      | 84313110     |
| 3Z.0314.KA.5  | PLACA                                    | 84313110     |
| 4C.0158.BK.1  | TRAVESSA SUPERIOR ESTRUTURA DE CABINA EV | 84313110     |

## Apêndice C: Dados coletados tratados para treino e teste

| Codigo        | Material                                 | NCM      |
|---------------|--|----------|
| X26.201.120   | BOTÃO VENUS 2.0 AZUL GRAV.SOBE PRAT.CIRC | 85365090 |
| 1Z.0465.BK.2  | ARRUELA                                  | 73182200 |
| 4B.0167.FR.1  | ARRUELA ESPAÇADORA Di 13,00 De 33,00 Esp | 73182200 |
| X08.072.098   | COMPUTADOR QOTOM MINI PC Q401S           | 84715010 |
| X08.072.098   | COMPUTADOR QOTOM MINI PC Q401S           | 84715010 |
| X08.072.098   | COMPUTADOR QOTOM MINI PC Q401S           | 84715010 |
| X26.033.762   | CONJ CLP OMRON CQM1H CPU21 -UPGRADE      | 85389010 |
| 4C.0106.ZL.1  | SUPORTE CAIXA PLUGAÇÃO; SYN 200          | 85389090 |
| 4B.6110.A.1   | CONJ. SENSOR APS -                       | 85437099 |
| 2B.0424.EG.1  | SUPORTE BUFFER - L - 254 [10,000]        | 73269090 |
| X08.012.621   | BOTÃO IMPULSO 3NA                        | 85365090 |
| 3Z.5300.BA.4  | CONJUNTO AMORTECEDOR DE BORRACHA         | 40169990 |
| X26.135.200   | CJ.ROLAMENTO TENSOR SUP/INF GUIA PERSIAN | 84835090 |
| X26.035.831   | BOTAO DE EMERGENCIA C/ CHAVE             | 85365090 |
| X26.160.814   | LONA DE FREIO MAQUINA 140VAT TAA416P4    | 68138190 |
| X26.160.814   | LONA DE FREIO MAQUINA 140VAT TAA416P4    | 68138190 |
| X26.160.814   | LONA DE FREIO MAQUINA 140VAT TAA416P4    | 68138190 |
| 6I.0905.B.1   | KIT FIXAÇÃO DA BASE DE MÁQUINA           | 73181500 |
| 3Y.6503.SG.10 | Banco Resistores 6 15R 4kW s/cabo vent   | 85332900 |
| 3Y.6503.PV.12 | Banco Resistores 4 12R 4kW c/bucha       | 85332900 |
| 3Y.6503.PV.2  | Banco Resistores 2 22R 4kW               | 85332900 |
| 3Y.6503.PV.4  | Banco Resistores 2 22R 4kW c/cabo        | 85332900 |
| 3Y.6503.SG.3  | Banco Resistores 2 9R 4kW s/cabo vent    | 85332900 |
| 3X.0382.AY.7  | POLIA 5 RAIOS 540mm                      | 84835090 |
| 4L.6152.DR.2  | SENSOR APS; ALE                          | 85437099 |
| 3Y.6503.SG.5  | Banco Resistores 4 12R 4kW s/cabo vent   | 85332900 |
| 3Y.0366.L.1   | EIXO (POLIA)                             | 84831090 |
| X26.149.425   | KIT DE FIXAÇÃO ES200/ES200 EASY          | 83026000 |
| X26.149.425   | KIT DE FIXAÇÃO ES200/ES200 EASY          | 83026000 |
| X26.149.425   | KIT DE FIXAÇÃO ES200/ES200 EASY          | 83026000 |
| X26.149.425   | KIT DE FIXAÇÃO ES200/ES200 EASY          | 83026000 |
| X26.149.425   | KIT DE FIXAÇÃO ES200/ES200 EASY          | 83026000 |
| 3Z.5802.AY.2  | CONJ. LUMINARIA LED QUADRADA AMAREL.     | 94051099 |
| 3Z.5802.AY.2  | CONJ. LUMINARIA LED QUADRADA AMAREL.     | 94051099 |
| 3Z.5802.AY.2  | CONJ. LUMINARIA LED QUADRADA AMAREL.     | 94051099 |
| 3Z.5802.AY.2  | CONJ. LUMINARIA LED QUADRADA AMAREL.     | 94051099 |

|                |   |          |
|----------------|---|----------|
| 3Z.5802.AY.2   | CONJ. LUMINARIA LED QUADRADA AMAREL.    | 94051099 |
| 4B.5293.AM.1   | PESADOR DE CARGA - SYNERGY 200          | 90318060 |
| 4C.0392.K.1    | FILTRO DE PROTEÇÃO DA CINTA EVO100 4 CI | 59111000 |
| 3W.6513.VL.948 | Consistencia M25 TW 7,5kW 380V H. Free  | 85371090 |
| 3W.6513.VL.949 | Consis. M25 TW 11/11,2kW 380V H. Free   | 85371090 |

**Apêndice D: Função que realiza as equações 8, 9, 10 e 11.**

```
#EXIBIR RESULTADOS - precesion, recall, f1-score, support, accuracy, average
print(classification_report(teste_NCM,clf_predictions))
```

## Apêndice E: Retorno do algoritmo por classe

| Classe   | Precisão | Sensibilidade | f1-score |
|----------|----------|---------------|----------|
| 27101932 | 100%     | 100%          | 100%     |
| 35061090 | 100%     | 100%          | 100%     |
| 39033020 | 100%     | 100%          | 100%     |
| 39169090 | 0%       | 0%            | 0%       |
| 39173100 | 0%       | 0%            | 0%       |
| 39174090 | 0%       | 0%            | 0%       |
| 39191000 | 50%      | 100%          | 67%      |
| 39191010 | 0%       | 0%            | 0%       |
| 39191020 | 0%       | 0%            | 0%       |
| 39199000 | 0%       | 0%            | 0%       |
| 39199010 | 100%     | 100%          | 100%     |
| 39199020 | 68%      | 73%           | 70%      |
| 39199090 | 67%      | 71%           | 69%      |
| 39201099 | 100%     | 100%          | 100%     |
| 39202090 | 0%       | 0%            | 0%       |
| 39203000 | 0%       | 0%            | 0%       |
| 39205100 | 69%      | 60%           | 64%      |
| 39205900 | 33%      | 33%           | 33%      |
| 39206100 | 86%      | 91%           | 88%      |
| 39206219 | 0%       | 0%            | 0%       |
| 39232110 | 100%     | 100%          | 100%     |
| 39269021 | 100%     | 100%          | 100%     |
| 39269069 | 88%      | 75%           | 81%      |
| 39269090 | 75%      | 60%           | 67%      |
| 40059190 | 100%     | 100%          | 100%     |
| 40081100 | 0%       | 0%            | 0%       |
| 40081900 | 0%       | 0%            | 0%       |
| 40092190 | 0%       | 0%            | 0%       |
| 40092210 | 0%       | 0%            | 0%       |
| 40093290 | 50%      | 100%          | 67%      |
| 40103100 | 0%       | 0%            | 0%       |
| 40103200 | 0%       | 0%            | 0%       |
| 40103300 | 100%     | 100%          | 100%     |
| 40103900 | 67%      | 100%          | 80%      |
| 40169300 | 91%      | 78%           | 84%      |
| 40169990 | 95%      | 97%           | 96%      |
| 44071100 | 100%     | 100%          | 100%     |
| 44151000 | 100%     | 100%          | 100%     |
| 44152000 | 96%      | 100%          | 98%      |
| 48191000 | 100%     | 95%           | 97%      |
| 48211000 | 0%       | 0%            | 0%       |
| 48219000 | 0%       | 0%            | 0%       |
| 49089000 | 40%      | 67%           | 50%      |
| 49111010 | 100%     | 80%           | 89%      |
| 49111090 | 33%      | 25%           | 29%      |
| 49119900 | 0%       | 0%            | 0%       |
| 56075090 | 0%       | 0%            | 0%       |
| 58110000 | 0%       | 0%            | 0%       |
| 59111000 | 0%       | 0%            | 0%       |
| 59119000 | 33%      | 50%           | 40%      |
| 68053090 | 100%     | 98%           | 99%      |
| 68080000 | 0%       | 0%            | 0%       |
| 68109900 | 76%      | 100%          | 87%      |
| 68138190 | 100%     | 100%          | 100%     |
| 70031900 | 33%      | 100%          | 50%      |
| 70032000 | 50%      | 50%           | 50%      |
| 70071900 | 67%      | 67%           | 67%      |
| 70072900 | 75%      | 33%           | 46%      |
| 70099100 | 0%       | 0%            | 0%       |
| 70169000 | 0%       | 0%            | 0%       |
| 70191900 | 0%       | 0%            | 0%       |
| 70199090 | 0%       | 0%            | 0%       |
| 72085100 | 0%       | 0%            | 0%       |
| 72085200 | 0%       | 0%            | 0%       |
| 72085400 | 67%      | 67%           | 67%      |
| 72089000 | 0%       | 0%            | 0%       |
| 72104910 | 71%      | 100%          | 83%      |
| 72122090 | 0%       | 0%            | 0%       |
| 72149100 | 100%     | 100%          | 100%     |
| 72149910 | 25%      | 100%          | 40%      |
| 72149990 | 0%       | 0%            | 0%       |
| 72155000 | 80%      | 100%          | 89%      |
| 72162100 | 20%      | 100%          | 33%      |
| 72163100 | 90%      | 90%           | 90%      |

|          |      |      |      |
|----------|------|------|------|
| 72164010 | 0%   | 0%   | 0%   |
| 72164090 | 0%   | 0%   | 0%   |
| 72166110 | 100% | 100% | 100% |
| 72166190 | 0%   | 0%   | 0%   |
| 72169100 | 100% | 100% | 100% |
| 72192400 | 100% | 100% | 100% |
| 72193100 | 0%   | 0%   | 0%   |
| 72193300 | 89%  | 96%  | 92%  |

## Apêndice F: Coleta de novos dados de tempo de cadastro

| Cenário EPDM | Material      | Visão | Entrada total    | Saída total      | Tempo de cadastro (dias) | Tempo de cadastro (Hrs) |
|--------------|---------------|-------|------------------|------------------|--------------------------|-------------------------|
| NEW          | 1A.0053.CC.1  | DTRI  | 15/03/2022 13:51 | 15/03/2022 14:23 | 0.0227                   | 0.5450                  |
| NEW          | 1A.0053.CD.1  | DTRI  | 15/03/2022 13:51 | 15/03/2022 14:23 | 0.0227                   | 0.5450                  |
| NEW          | 1A.0059.JM.3  | DTRI  | 15/03/2022 13:51 | 15/03/2022 14:23 | 0.0227                   | 0.5450                  |
| NEW          | 1A.0059.JV.1  | DTRI  | 15/03/2022 13:51 | 15/03/2022 14:23 | 0.0227                   | 0.5450                  |
| NEW          | 1A.0059.JW.1  | DTRI  | 15/03/2022 13:51 | 15/03/2022 14:23 | 0.0227                   | 0.5450                  |
| NEW          | 1A.0059.JX.1  | DTRI  | 15/03/2022 13:51 | 15/03/2022 14:23 | 0.0227                   | 0.5450                  |
| NEW          | 1A.0059.JX.2  | DTRI  | 15/03/2022 13:51 | 15/03/2022 14:23 | 0.0227                   | 0.5450                  |
| NEW          | 1A.0059.JX.3  | DTRI  | 15/03/2022 13:51 | 15/03/2022 14:23 | 0.0227                   | 0.5450                  |
| NEW          | 1A.0106.PA.2  | DTRI  | 10/03/2022 14:47 | 10/03/2022 22:25 | 0.3181                   | 7.6356                  |
| NEW          | 1A.0107.GA.1  | DTRI  | 15/03/2022 13:51 | 15/03/2022 14:23 | 0.0227                   | 0.5450                  |
| NEW          | 1A.0107.GA.2  | DTRI  | 15/03/2022 13:51 | 15/03/2022 14:23 | 0.0227                   | 0.5450                  |
| NEW          | 1A.0107.GL.1  | DTRI  | 15/03/2022 13:51 | 15/03/2022 14:23 | 0.0227                   | 0.5447                  |
| NEW          | 1A.0222.EJ.1  | DTRI  | 15/03/2022 13:51 | 15/03/2022 14:23 | 0.0227                   | 0.5447                  |
| NEW          | 1A.0318.EH.1  | DTRI  | 15/03/2022 13:51 | 15/03/2022 14:23 | 0.0227                   | 0.5447                  |
| NEW          | 1A.0430.XT.1  | DTRI  | 15/03/2022 13:51 | 15/03/2022 14:23 | 0.0227                   | 0.5447                  |
| NEW          | 1A.0430.XT.2  | DTRI  | 15/03/2022 13:51 | 15/03/2022 14:24 | 0.0227                   | 0.5447                  |
| NEW          | 1B.0215.YA.1  | DTRI  | 15/03/2022 13:51 | 15/03/2022 14:24 | 0.0227                   | 0.5447                  |
| NEW          | 1B.0215.YA.2  | DTRI  | 15/03/2022 13:51 | 15/03/2022 14:24 | 0.0227                   | 0.5447                  |
| NEW          | 1B.0215.YC.1  | DTRI  | 15/03/2022 13:51 | 15/03/2022 14:24 | 0.0227                   | 0.5450                  |
| NEW          | 1B.0215.YC.2  | DTRI  | 15/03/2022 13:51 | 15/03/2022 14:24 | 0.0227                   | 0.5450                  |
| NEW          | 1B.0215.YD.1  | DTRI  | 15/03/2022 13:51 | 15/03/2022 14:24 | 0.0227                   | 0.5450                  |
| NEW          | 1B.0215.YD.2  | DTRI  | 15/03/2022 13:51 | 15/03/2022 14:24 | 0.0227                   | 0.5450                  |
| NEW          | 1E.0316.GS.1  | DTRI  | 10/03/2022 14:47 | 10/03/2022 22:25 | 0.3181                   | 7.6356                  |
| NEW          | 1E.0901.JD.1  | DTRI  | 29/03/2022 11:15 | 29/03/2022 11:53 | 0.0264                   | 0.6339                  |
| NEW          | 1E.0901.JD.10 | DTRI  | 29/03/2022 11:15 | 29/03/2022 11:53 | 0.0264                   | 0.6339                  |
| NEW          | 1E.0901.JD.11 | DTRI  | 29/03/2022 11:15 | 29/03/2022 11:53 | 0.0264                   | 0.6339                  |
| NEW          | 1E.0901.JD.12 | DTRI  | 29/03/2022 11:15 | 29/03/2022 11:53 | 0.0264                   | 0.6339                  |
| NEW          | 1E.0901.JD.13 | DTRI  | 29/03/2022 11:15 | 29/03/2022 11:53 | 0.0264                   | 0.6339                  |
| NEW          | 1E.0901.JD.14 | DTRI  | 29/03/2022 11:15 | 29/03/2022 11:53 | 0.0264                   | 0.6339                  |
| NEW          | 1E.0901.JD.2  | DTRI  | 29/03/2022 11:15 | 29/03/2022 11:53 | 0.0264                   | 0.6342                  |
| NEW          | 1E.0901.JD.3  | DTRI  | 29/03/2022 11:15 | 29/03/2022 11:53 | 0.0264                   | 0.6342                  |
| NEW          | 1E.0901.JD.4  | DTRI  | 29/03/2022 11:15 | 29/03/2022 11:53 | 0.0264                   | 0.6342                  |
| NEW          | 1E.0901.JD.5  | DTRI  | 29/03/2022 11:15 | 29/03/2022 11:53 | 0.0264                   | 0.6342                  |
| NEW          | 1E.0901.JD.6  | DTRI  | 29/03/2022 11:15 | 29/03/2022 11:53 | 0.0264                   | 0.6342                  |
| NEW          | 1E.0901.JD.7  | DTRI  | 29/03/2022 11:15 | 29/03/2022 11:53 | 0.0264                   | 0.6342                  |
| NEW          | 1E.0901.JD.8  | DTRI  | 29/03/2022 11:15 | 29/03/2022 11:53 | 0.0264                   | 0.6342                  |
| NEW          | 1E.0901.JD.9  | DTRI  | 29/03/2022 11:15 | 29/03/2022 11:53 | 0.0264                   | 0.6342                  |
| NEW          | 1E.6505.EK.1  | DTRI  | 02/03/2022 11:10 | 02/03/2022 11:26 | 0.0112                   | 0.2694                  |
| NEW          | 1E.6505.EK.2  | DTRI  | 02/03/2022 11:10 | 02/03/2022 11:26 | 0.0112                   | 0.2694                  |
| NEW          | 1E.6505.EM.1  | DTRI  | 17/03/2022 08:35 | 17/03/2022 08:39 | 0.0032                   | 0.0769                  |
| NEW          | 2A.0345.CZ.1  | DTRI  | 29/03/2022 08:54 | 29/03/2022 09:23 | 0.0201                   | 0.4819                  |
| NEW          | 2B.0107.GD.1  | DTRI  | 02/03/2022 14:41 | 09/03/2022 11:33 | 6.8698                   | 164.8750                |
| NEW          | 2B.0107.GE.1  | DTRI  | 02/03/2022 14:41 | 09/03/2022 11:33 | 6.8698                   | 164.8750                |

|                               |              |      |                  |                  |        |          |
|-------------------------------|--------------|------|------------------|------------------|--------|----------|
| NEW                           | 2B.0107.HM.1 | DTRI | 10/03/2022 17:02 | 11/03/2022 10:21 | 0.7215 | 17.3156  |
| NEW                           | 2B.5330.BB.1 | DTRI | 02/03/2022 14:41 | 09/03/2022 11:33 | 6.8698 | 164.8750 |
| NEW                           | 2B.5330.BB.2 | DTRI | 02/03/2022 14:41 | 09/03/2022 11:33 | 6.8698 | 164.8750 |
| Tempo médio de cadastro (Hrs) | 11.1335      |      |                  |                  |        |          |
| Mediana (Hrs)                 | 1.8525       |      |                  |                  |        |          |
| Desvio padrão (Hrs)           | 23.5084      |      |                  |                  |        |          |
| Variância (Hrs <sup>2</sup> ) | 552.6436     |      |                  |                  |        |          |