Dimensionamento de estoque de sobressalentes de demanda rara usando Machine Learning

A Machine Learning approach to spare parts inventory sizing under rare demand conditions

Marcello Dantas Gomes Júnior – marcello.dantas@petrobras.com.br
Petróleo Brasileiro S.A. – (PETROBRAS), Vitória, ES

Article History:

Submitted: 2026 - 01 - 07 Revised: 2026 - 01 - 14 Accepted: 2026 - 01 - 20

Resumo: A indústria tem, dentre seus diversos desafios, a necessidade de dimensionar de forma adequada o estoque de sobressalentes estratégicos. Esses itens possuem histórico de baixo consumo, porém sua falta pode resultar em atrasos nos reparos e serviços de manutenção, levando à indisponibilidade operacional. Esses efeitos podem ocorrer em diferentes escalas. Por um lado, ter um estoque volumoso de itens estratégicos pode garantir uma maior disponibilidade operacional. No entanto, isso também acarreta custos adicionais com armazenamento, preservação e capital fixo. Portanto, é necessário encontrar uma solução de compromisso. A utilização de técnicas tradicionais ou mais simples para determinar o nível ideal de estoque para cada sobressalente muitas vezes esbarra na falta de dados históricos, especialmente em instalações no início de seu ciclo de operação e manutenção. Além disso, a diversidade de aplicações para alguns materiais também representa um desafio. Nesse contexto, o presente trabalho propõe uma metodologia baseada em machine learning, utilizando dois algoritmos conhecidos como Zero Inflated Poisson (ZIP) e Zero Inflated Negative Binomial (ZINB). Essa abordagem é aplicada a aproximadamente 27 mil sobressalentes, proporcionando uma definição mais rápida dos estoques estratégicos e gerando valor para a empresa.

Palavras-chave: Sobressalentes Estratégicos, Machine Learning, Poisson Inflacionado de Zeros, Binomial Negativo Inflacionado de Zeros, Manutenção.

Abstract: The industry faces, among its many challenges, the need to appropriately size the inventory of strategic spare parts. These items typically exhibit a history of low consumption; however, their unavailability may result in delays in repair and maintenance activities, leading to operational unavailability. Such effects may occur at different scales. On the one hand, maintaining a large inventory of strategic items can ensure higher operational availability. On the other hand, it entails additional costs related to storage, preservation, and tied-up capital. Therefore, a trade-off solution must be achieved. The use of traditional or simpler techniques to determine the optimal inventory level for each spare part often faces limitations due to the lack of historical data, especially in facilities at the early stages of their operation and maintenance lifecycle. Moreover, the diversity of applications associated with certain materials further increases the complexity of this problem. In this context, the present work proposes a machine learning–based methodology using two well-established algorithms: Zero-

Inflated Poisson (ZIP) and Zero-Inflated Negative Binomial (ZINB). This approach is applied to approximately 27,000 spare parts, enabling a faster definition of strategic inventory levels and generating value for the company.

Keywords: Strategic Spare Parts; Machine Learning; Zero-Inflated Poisson; Zero-Inflated Negative Binomial; Maintenance.

1. Introdução

Um dos desafios da indústria, especialmente para plantas e ativos físicos de uso intensivo, é o dimensionamento adequado do estoque de sobressalentes estratégicos da manutenção, ou seja, itens que possuem histórico de baixo consumo, mas cuja falta pode causar atrasos nos serviços de reparo e manutenção, levando à indisponibilidade operacional, (Mor et al., 2021; Poppe et al., 2017; Sathish et al., 2019; van der Auweraer et al., 2019). Os efeitos podem se manifestar em escalas de pequenas a maiores. Se por um lado ter um estoque volumoso de itens estratégicos pode trazer maior garantia de disponibilidade operacional, por outro traz custos adicionais de armazenamento e preservação, além de custo com capital imobilizado. É necessária uma solução de compromisso (Costantino et al., 2018; Tapia-Ubeda et al., 2019; Turrini & Meissner, 2019). A utilização de técnicas tradicionais ou mais simples para inferir o nível ideal de estoque de cada peça de reposição esbarra muitas vezes na falta de dados históricos, principalmente em instalações em fase inicial do ciclo de operação e manutenção, (Farhat et al., 2018; Hu et al., 2018; Kazemi Zanjani & Nourelfath, 2014; Rezaei et al., 2018). Outro problema é a diversidade de aplicações de alguns materiais. O presente trabalho propõe uma metodologia baseada em machine learning utilizando dois algoritmos conhecidos como Zero Inflated Poisson (ZIP) e Zero Inflated Negative Binomial (ZINB).

As instalações industriais offshore apresentam um desafio adicional para a formação do estoque estratégico de seus ativos físicos: o fato de operarem em alto mar (Gomes Junior, 2020; Gomes Junior & Rocha Filho, 2021). Este ambiente operacional impõe consequências mais graves para estoques estratégicos mal dimensionados. Os almoxarifados das unidades estacionárias de produção offshore possuem espaço pequeno, o que traz limitações na quantidade de peças de reposição. No entanto, a ausência de um sobressalente estratégico a bordo pode causar enormes perturbações operacionais e de segurança. Não se pode ser demasiado conservador ou otimista no dimensionamento dos stocks (van Horenbeek et al., 2013).

Outro grande problema, mas não exclusivo da indústria offshore, é a escassez de dados ou informações (Cavalieri et al., 2008; Mirzahosseinian & Piplani, 2011; Romeijnders et al., 2012). Essa escassez pode ser decorrente do fato da planta ser nova ou ter acabado de entrar em fase de operação e manutenção, bem como pode ser devida à cultura de não registrar com precisão as informações relacionadas à manutenção, problema que se agrava em indústrias com ativos de uso intensivo. As indústrias de ativos intensivos operam 24 horas por dia, sete dias por semana, com exceção do tempo de inatividade da produção. Assim, a necessidade de enfrentar os problemas de forma corretiva sobrecarrega as equipes de todos os níveis hierárquicos, o que ofusca o olhar para o bom registro das informações, dificultando a estruturação de processos que tragam melhorias operacionais e maior segurança. Um processo que pode se beneficiar de uma cultura de bons registros é a gestão de peças sobressalentes estratégicas (Hardwick & Lafraia, 2021). Porém, mudar uma cultura em vários níveis hierárquicos é um processo demorado e, embora fundamental, leva muito tempo para funcionar bem. Desta forma, é necessário desenvolver métodos para trabalhar neste ambiente de escassez de dados, gerando resultados que não sejam excessivamente conservadores ou otimistas. Assim o presente trabalho propõe uma abordagem com uso de modelos supervisionados de machine learning utilizando informações do histórico de consumo de peças de reposição estratégicas, com o objetivo de dimensionar o estoque dessas peças de reposição para uma janela de tempo futura.

2. Revisão Bibliográfica

A gestão de sobressalentes tem sido amplamente estudada no âmbito da Pesquisa Operacional (Operational Research – OR), dada sua relevância para ambientes industriais e de serviços que buscam simultaneamente reduzir custos, aumentar a disponibilidade de equipamentos e lidar com padrões de demanda altamente irregulares. Nesse contexto, Hu et al. (2018) apresentam uma revisão sistemática e crítica da literatura sobre OR aplicada à gestão de peças sobressalentes, destacando que, embora essa área seja uma extensão natural da gestão de estoques, ela envolve desafios específicos, como o elevado número e diversidade de itens, demandas intermitentes, risco de obsolescência e forte interdependência com políticas de manutenção (Exton & Labib, 2002; Labib & Exton, 2001; Gallagher et al., 2005).

Os autores estruturam a literatura a partir de um marco conceitual baseado no ciclo de vida do produto (Klepper, 1996; Rink & Swan, 1979), relacionando objetivos, métodos e disciplinas da OR às diferentes fases do ciclo de vida dos ativos, desde a aquisição inicial até

o fim de vida. Nesse framework, destacam-se quatro grandes áreas de aplicação: classificação multicritério de sobressalentes, previsão de demanda, otimização de estoques e simulação. Observa-se que as prioridades e técnicas empregadas variam conforme o estágio do ciclo de vida, sendo essa diferenciação particularmente relevante em ambientes industriais complexos (Cavalieri et al., 2008; Driessen et al., 2015).

No que se refere à classificação de sobressalentes, a literatura evoluiu dos tradicionais métodos ABC, baseados exclusivamente no consumo anual, para abordagens multicritério conhecidas como Multiple Criteria Inventory Classification (MCIC). Essas abordagens incorporam critérios como criticidade, lead time, custo e impacto operacional, utilizando técnicas como algoritmos genéticos (Guvenir & Erel, 1998), matrizes bi-criteriais (Flores & Whybark, 1986, 1987), AHP (Partovi & Burton, 1993; Partovi & Anandarajan, 2002), lógica fuzzy (Cakir & Canbolat, 2008) e redes neurais artificiais (Yu, 2011). Apesar da sofisticação metodológica, Hu et al. (2018) ressaltam que grande parte desses estudos apresenta validação empírica limitada, frequentemente baseada em dados simulados ou exemplos observacionais, além de haver pouca integração entre os resultados da classificação e a definição efetiva de políticas de estoque (Mohammaditabar et al., 2012; Teunter et al., 2010; Zhang et al., 2001).

A previsão de demanda de sobressalentes é apontada como um dos maiores desafios da área, em função do comportamento intermitente, da baixa frequência de consumo e da grande variabilidade dos padrões de demanda (Boone et al., 2008). A literatura revisada pode ser organizada em três grandes grupos: métodos baseados em séries temporais, métodos baseados em confiabilidade e manutenção, e previsões ajustadas por julgamento. No primeiro grupo, destacam-se métodos clássicos como Suavização Exponencial, Croston e suas variantes (Syntetos & Boylan, 2001, 2005), além de redes neurais (Gutierrez et al., 2008; Kourentzes, 2013) e técnicas de bootstrapping (Viswanathan & Zhou, 2008), sendo reportada vantagem empírica das variantes de Croston em diversos contextos (Van Wingerden et al., 2014). O segundo grupo incorpora informações de confiabilidade, como MTBF e ciclos de vida, bem como planos de manutenção, para estimar a demanda futura de sobressalentes (Ghobbar & Friend, 2003; Wang & Syntetos, 2011; Ghodrati et al., 2012), ganhando relevância com o avanço da manutenção preditiva e o uso de dados operacionais e ambientais (Tracht et al., 2013; Frazzon et al., 2014). Já o ajuste por julgamento, embora amplamente utilizado na prática, ainda carece de evidências empíricas robustas quanto ao seu impacto nos resultados de previsão (Fildes et al., 2009; Syntetos et al., 2009b).

No campo da otimização de estoques, os estudos se distribuem ao longo de diferentes horizontes do ciclo de vida dos ativos. Trabalhos clássicos abordam a provisão inicial de sobressalentes sob incerteza, restrições orçamentárias e critérios de criticidade (Schweitzer, 1967; Jaquette & Osaki, 1972; Guide & Srivastava, 1997; Öner et al., 2010). Outros estudos focam na otimização de políticas de reposição, considerando sistemas multi-echelon, demandas variáveis no tempo e integração entre previsão e reposição (Bakker et al., 2012; Basten & Van Houtum, 2014). Na fase final do ciclo de vida, ganham destaque problemas relacionados à compra final, remanufatura, refabricação e canibalização de equipamentos, frequentemente integrados a modelos de logística reversa (Teunter & Fortuin, 1998, 1999; Inderfurth & Mukherjee, 2008; Krikke & Van Der Laan, 2011; Pourakbar et al., 2012, 2014). Estudos mais recentes também incorporam objetivos ambientais, integrando custos econômicos e emissões de carbono, refletindo a crescente preocupação com sustentabilidade (Dekker et al., 2012; Digiesi et al., 2015; Kader et al., 2015).

A simulação, especialmente por meio de dinâmica de sistemas e simulação de eventos discretos, é amplamente utilizada para avaliar políticas de estoque em ambientes complexos, validar modelos analíticos e capturar interações entre manutenção, falhas e reposição de sobressalentes (Rossetti & Thomas, 2006; Persson & Saccani, 2007, 2009; Tiacci & Saetta, 2011). Essas abordagens mostram-se particularmente eficazes em sistemas multi-echelon, com transferência lateral de estoques, múltiplos modos de falha e integração entre estoque e manutenção (Kabir & Farrash, 1996; Marseguerra et al., 2005; Lee et al., 2008).

Por fim, Hu et al. (2018) destacam um desalinhamento persistente entre os avanços teóricos da literatura e sua adoção em softwares corporativos e sistemas ERP, que continuam a empregar majoritariamente métodos clássicos como EOQ, ROP e MRP. Esse hiato é atribuído à falta de pesquisa aplicada, dificuldades de implementação e barreiras de conhecimento (Boylan & Syntetos, 2016). Apesar do aumento da disponibilidade de dados provenientes de sistemas de manutenção informatizados e de monitoramento preditivo, sua integração efetiva com modelos avançados de gestão de estoques ainda é limitada, configurando uma importante oportunidade de pesquisa.

De forma geral, a literatura revisada evidencia avanços significativos em classificação, previsão, otimização e simulação aplicadas à gestão de sobressalentes. No entanto, não foi identificado, até o presente momento, o uso de modelos de contagem inflacionados de zeros, como o Zero-Inflated Poisson (ZIP) ou o Zero-Inflated Negative Binomial (ZINB),

especificamente aplicados ao problema de dimensionamento de estoques de sobressalentes estratégicos com demanda rara, indicando uma lacuna relevante a ser explorada pelo presente trabalho.

3. Método Proposto

Os modelos de machine learning supervisionados Zero Inflated Poisson (ZIP) e Zero Inflated Negative Binomial (ZINB) são usados para lidar com dados que possuem excesso de zeros e distribuição discreta.

O modelo Zero Inflated Poisson (ZIP) é adequado quando há excesso de zeros nos dados e a distribuição dos valores não nulos se assemelha a uma distribuição de Poisson. Esse modelo assume que os zeros podem ser gerados por dois processos distintos: um processo de excesso de zeros e um processo de contagem de ocorrências. O ZIP é útil em situações em que existe uma proporção significativa de zeros e a contagem de ocorrências é baixa.

Já o modelo Zero Inflated Negative Binomial (ZINB) é usado quando há excesso de zeros e a distribuição dos valores não nulos se assemelha a uma distribuição de binomial negativa. Assim como o ZIP, o ZINB também considera dois processos: um processo de excesso de zeros e um processo de contagem de ocorrências. A diferença é que o ZINB permite maior flexibilidade na modelagem de dados, pois a distribuição de binomial negativa é capaz de capturar uma maior variabilidade nos dados.

No contexto do dimensionamento de estoque de sobressalentes de demanda rara, esses modelos podem ser aplicados para prever a quantidade de peças de reposição necessárias para atender a uma demanda que ocorre raramente. Os excessos de zeros nos dados podem ser devido à falta de demanda por determinadas peças em determinados períodos. Ao considerar esses excessos de zeros, os modelos ZIP e ZINB podem ajudar a estimar a probabilidade de demanda zero para determinadas peças e, ao mesmo tempo, prever a quantidade necessária em caso de demanda não nula.

No entanto, é importante ressaltar que a aplicação desses modelos depende da disponibilidade de dados históricos relevantes sobre a demanda e do entendimento adequado do contexto específico do dimensionamento de estoque de sobressalentes. Além disso, é necessário considerar outros fatores, como a incerteza na previsão da demanda e a consideração de políticas de estoque, para tomar decisões efetivas de dimensionamento de estoque.

A definição sobre a existência ou não de uma quantidade excessiva de zeros na variável dependente Y é elaborada por meio de um teste específico, conhecido por teste de Vuong (1989), que representará um importante output a ser analisado na estimação de modelos de regressão para dados de contagem, quando houver a suspeita de existência de inflação de zeros.

Para decidir entre a modelagem via ZIP e ZINB, é utilizado o teste de Cameron e Trivedi para detecção de superdispersão. O teste de Cameron e Trivedi (Cameron e Trivedi, 1991) é uma abordagem estatística utilizada para a seleção entre os modelos Zero Inflated Poisson (ZIP) e Zero Inflated Negative Binomial (ZINB). Esses modelos são comumente aplicados em situações em que os dados apresentam excesso de zeros e distribuição discreta. A escolha entre ZIP e ZINB é importante porque eles diferem em suas suposições sobre a variabilidade dos dados. O modelo ZIP assume que a variabilidade é igual à média, enquanto o modelo ZINB permite que a variabilidade seja maior do que a média. O teste de Cameron e Trivedi é baseado em uma estatística de razão de verossimilhança, que compara o ajuste dos dois modelos aos dados observados. A ideia é verificar qual modelo se ajusta melhor aos dados, considerando a presença de excesso de zeros. O teste consiste em calcular a diferença entre as log-verossimilhanças dos dois modelos, juntamente com seus respectivos graus de liberdade. Essa diferença é comparada com uma distribuição qui-quadrado para determinar se há evidências estatísticas suficientes para preferir um modelo em relação ao outro.

A Tabela 1, adaptada de Fávero e Belfiori (2014), ilustra a aplicabilidade de cada modelo.

Tabela 1 – Aplicabilidade dos modelos supervisionados para dados de contagem

Verificação	Modelos de Regressão para Dados de Contagem			
	Poisson	Binomial Negativo	Poisson Inflacionado com Zeros (ZIP)	Binomial Negativo Inflacionado com Zeros (ZINB)
Superdispersão nos Dados da Variável Dependente	Não	Sim	Não	Sim
Quantidade Excessiva de Zeros na Variável Dependente	Não	Não	Sim	Sim

Fonte: adaptado de Fávero e Belfiori (2014)

A Figura 1, adaptada de Minami et al (2007), ilustra o perfil dos dados de contagem a serem modelados pelas distribuições Zero Inflated Poisson (ZIP) e Zero Inflated Negative Binomial (ZINB).

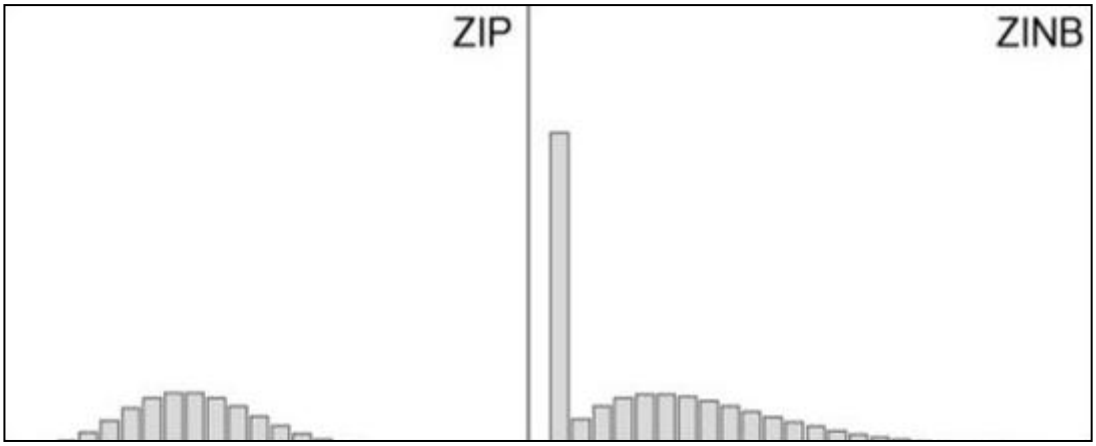


Figura 1 – Perfil do histograma das distribuições inflacionadas com zeros Poisson e Binomial Negativo
Fonte: adaptado de Minami et al (2007)

Sobressalentes estratégicos para manutenção possuem perfil de demanda rara, porém a falta deles causa indisponibilidade nas funções dos ativos físicos.

Em relação especificamente aos modelos de regressão Poisson inflacionados de zeros, podemos definir que, enquanto a probabilidade p de ocorrência de nenhuma contagem para dada observação i (i = 1, 2, ..., n, em que n é o tamanho da amostra), ou seja, p(Yi =0), é calculada levando-se em consideração a soma de um componente dicotômico com um componente de contagem e, portanto, deve-se definir a probabilidade plogit de não ocorrer nenhuma contagem devido exclusivamente ao componente dicotômico, a probabilidade p de ocorrência de determinada contagem m (m = 1, 2, ...), ou seja, p(Yi=m), segue a própria expressão da probabilidade da distribuição Poisson, multiplicada por (1-plogit), conforme Equações 1, 2 e 3.

$$\begin{cases} p(Y_i = 0) = p_{logit_i} + \left(1 - p_{logit_i}\right) \cdot e^{-\lambda_i} \\ p(Y_i = m) = \left(1 - p_{logit_i}\right) \cdot \frac{e^{-\lambda_i} \cdot \lambda_i^m}{m!}, \; m = 1, 2, \ldots \end{cases} \tag{1}$$

$$p_{logit_i} = \frac{1}{1 + e^{-(\gamma + \delta_1 W_{1i} + \delta_2 W_{2i} + \cdots + \delta_q W_{qi})}} \tag{2}$$

$$\lambda_{poisson_i} = e^{(\alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki})} \tag{3}$$

Os modelos de regressão Poisson inflacionados de zeros apresentam dois processos geradores de zeros, sendo um devido à distribuição binária (neste caso, são gerados os

chamados zeros estruturais) e outro devido à distribuição Poisson (nesta situação, são gerados dados de contagem, entre os quais os chamados zeros amostrais).

Já em relação aos modelos de regressão do tipo binomial negativo inflacionados de zeros, podemos definir que, enquanto a probabilidade p de ocorrência de nenhuma contagem para dada observação i, ou seja, p(Yi = 0), é também calculada levando-se em consideração a soma de um componente dicotômico com um componente de contagem, a probabilidade p de ocorrência de determinada contagem m (m = 1, 2, ...), ou seja, p(Yi=m), segue agora a expressão da probabilidade da distribuição Poisson-Gama., conforme Equações 4, 5 e 6.

$$\begin{cases} p(Y_i = 0) = p_{logit_i} + \left(1 - p_{logit_i}\right) \cdot \left(\dfrac{1}{1 + \theta^{-1}\lambda_{bneg_i}}\right)^{\theta} \\ p(Y_i = m) = \left(1 - p_{logit_i}\right) \cdot \left[\dfrac{\delta^{\theta} \cdot m_i^{\theta-1} \cdot e^{-m_i\delta}}{(\theta - 1)!}\right], \qquad m = 1, 2, \ldots \end{cases} \tag{4}$$

$$p_{logit_i} = \frac{1}{1 + e^{-(\gamma + \delta_1 W_{1i} + \delta_2 W_{2i} + \cdots + \delta_q W_{qi})}} \tag{5}$$

$$\lambda_{bneg_i} = e^{(\alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki})} \tag{6}$$

4. Estudo de caso e resultados

O processo de definição do mapa de sobressalentes estratégicos para manutenção é descrito por meio de um fluxograma simplificado, implementado em linguagem R e desenvolvido em ambiente RStudio. O fluxo de processamento foi concebido para operar de forma totalmente automatizada, permitindo sua aplicação em bases de dados extensas e heterogêneas, típicas de ambientes industriais complexos.

A metodologia é executada de maneira independente para cada grupo de mercadoria, entendido como a classe do material (por exemplo, bombas, válvulas, componentes de motores, entre outros). Para cada grupo, constrói-se um modelo estatístico de dados de contagem considerado mais apropriado ao comportamento de consumo observado. As unidades de análise são os sobressalentes pertencentes a cada grupo, enquanto as variáveis explicativas contemplam informações relacionadas ao histórico de consumo, ao parque instalado e às características do fabricante.

O histórico de consumo é representado a partir da quantidade consumida ao longo do tempo e do período de existência do material no sistema de gestão, considerando como referência a data mais recente entre a criação do item no sistema corporativo e o início da

operação da unidade mais antiga analisada. O parque instalado expressa a recorrência do sobressalente nas listas técnicas dos equipamentos, refletindo sua base potencial de consumo. O fabricante é tratado como uma variável categórica, incorporada aos modelos por meio de técnicas de dummização, amplamente utilizadas em modelagem estatística e em algoritmos de machine learning para dados de contagem.

Uma exceção metodológica ocorre nos casos em que todos os materiais de um determinado grupo de mercadoria apresentam histórico de consumo nulo. Nessas situações, a inexistência de variabilidade nos dados inviabiliza o ajuste de modelos probabilísticos de contagem. Para os materiais desse grupo que atendem aos critérios de elegibilidade para estoque estratégico, adota-se, de forma conservadora, a definição de uma unidade em estoque.

A etapa de preparação e tratamento dos dados (data wrangling) é fundamental para o desempenho do método. A automação desse processo permite elevado ganho de eficiência no manuseio de grandes bases de dados, envolvendo procedimentos de limpeza, padronização, transformação de variáveis, tratamento de valores ausentes e identificação de outliers. Essa etapa garante a consistência e a qualidade dos dados utilizados na modelagem, sendo crucial para a robustez dos resultados obtidos.

A seleção do modelo estatístico mais adequado para cada grupo de mercadoria é realizada a partir de testes específicos aplicados aos dados de contagem. Inicialmente, avalia-se a presença de superdispersão, o que orienta a escolha entre modelos de Poisson ou Binomial Negativo. Em seguida, aplica-se um teste adicional para verificar a necessidade de considerar inflação de zeros, determinando o uso, quando apropriado, de modelos inflacionados de zeros. Dessa forma, assegura-se que o modelo adotado represente de maneira mais fiel o padrão de consumo observado em cada grupo.

Uma vez definido o modelo mais apropriado, procede-se à previsão do consumo futuro dos sobressalentes, a qual subsidia diretamente o dimensionamento do nível de estoque. O horizonte de previsão é definido com base no lead time de cada material, entendido como o intervalo entre a solicitação e a efetiva disponibilidade do sobressalente para uso. O estoque é dimensionado para cobrir esse período com uma probabilidade associada à criticidade dos equipamentos atendidos. Sobressalentes vinculados a equipamentos de maior impacto em segurança ou continuidade operacional requerem maiores probabilidades de cobertura do lead time, resultando em níveis de estoque mais elevados, enquanto materiais associados a equipamentos de menor criticidade ou com lead times

reduzidos podem ser atendidos por estratégias alternativas de suprimento, sem formação de estoque estratégico.

A metodologia foi aplicada em duas unidades industriais distintas, envolvendo, respectivamente, mais de 30 mil sobressalentes em uma unidade e aproximadamente 70 mil sobressalentes em outra. Em ambos os casos, o processamento completo — desde a preparação dos dados até a estimação dos níveis de estoque estratégicos para todos os itens — foi realizado em tempos computacionais reduzidos. Na primeira unidade, o método levou menos de 90 segundos para estimar o nível de estoque de todos os sobressalentes elegíveis, enquanto na segunda unidade, com base de dados significativamente maior, o tempo total de processamento foi de aproximadamente três minutos. Esses resultados evidenciam a escalabilidade e a aplicabilidade prática da abordagem proposta em ambientes industriais de grande porte.

Os resultados gerados pelo algoritmo são posteriormente analisados sob a ótica técnica e gerencial antes da implementação definitiva da estratégia de sobressalentes estratégicos. Alguns materiais específicos, como itens perecíveis ou de rápida obsolescência, podem ser excluídos da formação de estoque e tratados por estratégias alternativas de suprimento sob demanda. Da mesma forma, sobressalentes de elevado custo unitário são avaliados de forma mais criteriosa, reforçando o caráter de apoio à decisão da metodologia proposta, e não sua aplicação automática e indiscriminada.

A base de dados considerada no escopo deste trabalho compreende um parque expressivo de equipamentos industriais, acompanhado por um conjunto amplo e diversificado de listas técnicas de materiais. Esses sobressalentes estão organizados em um grande número de grupos de mercadoria, refletindo a complexidade e a heterogeneidade dos ativos analisados. Nem todos os itens avaliados são elegíveis para a formação de estoque estratégico, conforme critérios previamente estabelecidos relacionados à criticidade, ao lead time e à estratégia de manutenção. Ainda assim, o algoritmo proposto estima o nível de estoque potencial para todos os sobressalentes, de modo a permitir reavaliações futuras e ajustes dinâmicos nas políticas de suprimento e manutenção.

Para cada grupo de mercadoria, é construído um modelo supervisionado para dados de contagem, de forma que o fluxo metodológico seja executado repetidamente para todas as classes de materiais. Essa abordagem modular permite capturar características específicas de consumo em cada grupo, ao mesmo tempo em que preserva a padronização do processo. O tempo total de processamento para a estimativa dos níveis de estoque de todos os

sobressalentes mostrou-se bastante reduzido, evidenciando o baixo custo computacional da metodologia e sua viabilidade para aplicações recorrentes e em larga escala. Ademais, o elevado grau de automação alcançado contribui significativamente para a redução de esforços operacionais repetitivos por parte de técnicos e engenheiros, liberando recursos humanos para atividades de maior valor agregado.

No que se refere aos estoques estratégicos já existentes em armazéns terrestres, os resultados indicam um potencial relevante de otimização do capital imobilizado em materiais ao longo de um horizonte plurianual. Tal redução ocorre sem comprometer os níveis de serviço e desconsiderando, inclusive, benefícios adicionais associados à diminuição de custos indiretos, como armazenagem, preservação, movimentação e condicionamento dos sobressalentes.

Por outro lado, a aplicação da metodologia também aponta a necessidade de inclusão de um conjunto significativo de sobressalentes que anteriormente não compunham o estoque estratégico. Esses itens estão majoritariamente associados a equipamentos de elevada criticidade, tanto do ponto de vista da segurança operacional quanto da continuidade da operação. A incorporação desses sobressalentes tende a proporcionar respostas mais rápidas às intervenções de manutenção, melhorias na mantenabilidade dos ativos e aumento da disponibilidade operacional. Parte relevante dessas recomendações já foi implementada, resultando em uma ampliação substancial da cobertura de equipamentos críticos, especialmente aqueles com maior impacto em segurança operacional, representando um avanço significativo na gestão da manutenção e dos ativos industriais.

5. Considerações finais

Embora ainda exista certo nível de ceticismo na indústria em relação à adoção de técnicas de Machine Learning, evidências crescentes demonstram que, quando aplicadas de forma criteriosa, objetiva e alinhada às necessidades operacionais, essas abordagens podem gerar resultados consistentes nas áreas de manutenção, confiabilidade e gestão de ativos. Em particular, métodos orientados por dados têm se mostrado adequados para tratar problemas caracterizados por elevada incerteza, demanda rara e limitações de histórico, cenários típicos da gestão de sobressalentes industriais.

O presente trabalho demonstra a efetividade do uso de modelos supervisionados para dados de contagem, incluindo tanto modelos tradicionais quanto aqueles com inflação de zeros, no dimensionamento do estoque estratégico de sobressalentes para manutenção. A

metodologia proposta permitiu a revisão e a adequação dos níveis de estoque existentes, promovendo uma utilização mais eficiente dos recursos destinados à manutenção. Como resultado, foi possível alcançar uma redução relevante dos custos associados à formação e à manutenção de estoques estratégicos, sem que houvesse degradação dos níveis de disponibilidade operacional, da segurança dos sistemas ou do tempo de resposta às falhas.

Os resultados indicam que o método foi capaz de atender de forma equilibrada ao trade-off entre custo de estoque e desempenho operacional, assegurando a presença de sobressalentes críticos nos horizontes necessários para intervenção, ao mesmo tempo em que evitou a manutenção de estoques excessivos ou desnecessários. Dessa forma, a abordagem contribui simultaneamente para a confiabilidade dos ativos, a continuidade operacional e a eficiência econômica do sistema de manutenção, reforçando o potencial do Machine Learning como ferramenta de apoio à decisão em contextos industriais complexos.

REFERÊNCIAS

Agresti, A. (2002). *Categorical data analysis*. Wiley.

Bacchetti, A., & Saccani, N. (2012). Spare parts classification and demand forecasting for stock control: Investigating the gap between research and practice. *Omega, 40*(6), 722–737.

Basten, R. J. I., & van Houtum, G. J. (2014). System-oriented inventory models for spare parts. *Surveys in Operations Research and Management Science, 19*(1), 34–55.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Analysis of count data*. Chapman and Hall.

Cameron, A. C., & Trivedi, P. K. (1991). Regression-based tests for overdispersion in the Poisson model. *Journal of Econometrics, 46*(3), 347–364. https://doi.org/10.1016/0304-4076(90)90014-K

Cavalieri, S., Gaiardelli, P., Ierace, S., & Maccarrone, P. (2008). A decision-making framework for managing maintenance spare parts. *Production Planning & Control, 19*(4), 379–396. https://doi.org/10.1080/09537280802034471

Costantino, F., Di Gravio, G., Shaban, A., & Tronci, M. (2018). Spare parts management for irregular demand items. *Omega, 81*, 57–66. https://doi.org/10.1016/j.omega.2017.09.009

Dekker, R., Bloemhof, J., & Mallidis, I. (2012). Operations research for green logistics: An overview of aspects, issues, contributions and challenges. *European Journal of Operational Research, 219*(3), 671–679.

Digiesi, S., Mossa, G., & Rubino, S. (2015). A sustainable EOQ model for repairable spare parts under uncertain demand. *IMA Journal of Management Mathematics, 26*(2), 185–203.

Exton, T., & Labib, A. (2002). Spare parts decision analysis – The missing link in CMMSs (Part II). *Journal of Maintenance & Asset Management, 17*(1), 14–21.

Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: An empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting, 25*(1), 3–23.

Flores, B. E., & Whybark, D. C. (1986). Multiple criteria ABC analysis. *International Journal of Operations & Production Management, 6*(3), 38–46.

Flores, B. E., & Whybark, D. C. (1987). Multiple criteria ABC analysis. *International Journal of Operations & Production Management, 7*(1), 79–85.

Ghobbar, A. A., & Friend, C. H. (2003). Evaluation of forecasting methods for intermittent parts demand in the field of aviation: A predictive model. *Computers & Operations Research, 30*(14), 2097–2114.

Gomes Junior, M. D. (2020). Deactivating a motor generator set in a stationary unit to produce offshore oil in decommissioning using reliability, availability and maintainability analysis and life cycle cost. *Revista Offshore*. https://doi.org/10.48072/2525-7579.rog.2020.248

Gomes Junior, M. D., & Rocha Filho, L. B. (2021). Estudo de confiabilidade da proposta de reorganização das votações de detectores de gás dos FPSOs da cessão onerosa. In *Anais do Congresso Brasileiro de Manutenção e Gestão de Ativos*. ABRAMAN.

Guenverir, H. A., & Erel, E. (1998). Multicriteria inventory classification using a genetic algorithm. *European Journal of Operational Research, 105*(1), 29–37.

Hu, Q., Boylan, J. E., Chen, H., & Labib, A. (2017). Spare parts classification in industrial manufacturing using the dominance-based rough set approach. *European Journal of Operational Research, 262*(3), 1136–1163.

Hu, Q., Boylan, J. E., Chen, H., & Labib, A. (2018). OR in spare parts management: A review. *European Journal of Operational Research, 266*(2). https://doi.org/10.1016/j.ejor.2017.07.058

Hosmer, D. W., Jr., & Lemeshow, S. (2000). *Applied logistic regression*. Wiley.

Kazemi Zanjani, M., & Nourelfath, M. (2014). Integrated spare parts logistics and operations planning for maintenance service provider. *International Journal of Production Economics, 158*, 1–12. https://doi.org/10.1016/j.ijpe.2014.07.012

Kollars, R. J. (2014). *Categorical data analysis for the behavioral and social sciences*. Routledge.

Lee, L. H., Kang, H. Y., & Huang, C. C. (2011). The effect of lead time on inventory levels with predetermined order quantity. *International Journal of Production Research, 49*(15), 4563–4576. https://doi.org/10.1080/00207543.2010.527112

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Sage Publications.

McCullagh, P., & Nelder, J. (1989). *Generalized linear models*. Chapman and Hall.

Minami, M., Lennert-Cody, C. E., Gao, W., & Roman-Verdesoto, M. (2007). Modeling shark bycatch: The zero-inflated negative binomial regression model with smoothing. *Fisheries Research, 84*(2), 210–221. https://doi.org/10.1016/j.fishres.2006.10.019

Mirzahosseinian, H., & Piplani, R. (2011). A study of repairable parts inventory system operating under performance-based contract. *European Journal of Operational Research, 214*(2), 256–261. https://doi.org/10.1016/j.ejor.2011.04.035

Mor, R. S., Singh, S., Bhardwaj, A., & Singh, L. P. (2021). Spare parts inventory management in the warehouse: A lean approach. *International Journal of Industrial Engineering & Production Research, 32*(2), 1–11. https://doi.org/10.22068/ijiepr.32.2.1

Nahmias, S. (2015). *Production and operations analysis* (6th ed.). Waveland Press.

Peng, R. D. (2016). *R programming for data science*. Chapman and Hall/CRC.

Poppe, J., Boute, R., & Lambrecht, M. (2017). Numerical study of inventory management under various maintenance policies. *Reliability Engineering & System Safety, 168*, 262–273. https://doi.org/10.1016/j.ress.2017.06.012

Rezaei, H., Ghodsi, R., & Jafari-Eskandari, M. (2018). A new methodology to optimize target stock level for unpredictable demand of spare parts: A case study. *IFAC-PapersOnLine, 51*(11), 538–543. https://doi.org/10.1016/j.ifacol.2018.08.374

Romeijnders, W., Teunter, R., & van Jaarsveld, W. (2012). A two-step method for forecasting spare parts demand using information on component repairs. *European Journal of Operational Research, 220*(2), 386–393. https://doi.org/10.1016/j.ejor.2012.01.019

Sathish, N., Kumar, V. S., & Prasad, K. R. (2019). Procurement methodologies to optimize the inventory levels of spare parts. *International Journal of Recent Technology and Engineering, 8*(3), 1662–1669. https://doi.org/10.35940/ijrte.C4437.098319

Silver, E. A., Pyke, D. F., & Peterson, R. (1998). *Inventory management and production planning and scheduling* (3rd ed.). Wiley.

Stokes, M. E., Davis, C. S., & Koch, G. G. (2012). *Categorical data analysis using the SAS system*. SAS Institute.

Syntetos, A. A., & Boylan, J. E. (2005). The accuracy of intermittent demand estimates. *International Journal of Forecasting, 21*(2), 303–314.

Syntetos, A. A., Boylan, J. E., & Disney, S. M. (2009). Forecasting for inventory planning: A 50-year review. *Journal of the Operational Research Society, 60*, 149–160.

Tapia-Ubeda, F. J., Miranda, P. A., & González-Prida, V. (2019). Supplier selection for spare parts supply chain networks. *IFAC-PapersOnLine, 52*(13), 2237–2242. https://doi.org/10.1016/j.ifacol.2019.11.538

Turrini, L., & Meissner, J. (2019). Spare parts inventory management: New evidence from distribution fitting. *European Journal of Operational Research, 273*(1), 118–130. https://doi.org/10.1016/j.ejor.2017.09.039

van der Auweraer, S., Boute, R., & Lambrecht, M. (2019). Forecasting spare part demand with installed base information: A review. *International Journal of Forecasting, 35*(1), 181–196. https://doi.org/10.1016/j.ijforecast.2018.09.002

van Horenbeek, A., Pintelon, L., & Muchiri, P. (2013). Joint maintenance and inventory optimization systems: A review. *International Journal of Production Economics, 143*(2), 499–508. https://doi.org/10.1016/j.ijpe.2012.04.001

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica, 57*(2), 307–333.

Wang, W., & Syntetos, A. A. (2011). Spare parts demand: Linking forecasting to equipment maintenance. *Transportation Research Part E, 47*(6), 1194–1209.

Yu, M.-C. (2011). Multi-criteria ABC analysis using artificial-intelligence-based classification techniques. *Expert Systems with Applications, 38*(4), 3416–3421.

Zipkin, P. H. (2000). *Foundations of inventory management*. McGraw-Hill.