ANÁLISE PREDITIVA DE RISCO OPERACIONAL EM CONTÊINERES NO PORTO DE SANTOS

PREDICTIVE ANALYSIS OF OPERATIONAL RISK IN CONTAINERS AT THE PORT OF SANTOS

Robert Richard das Neves Correia dos Santos, José Augusto Theodósio Pazetti

Fatec Baixada Santista, Rubens Lara, Curso de Ciência de Dados
E-mail para contato: robert.santos01@fatec.sp.gov.br

RESUMO – Este trabalho propõe um modelo preditivo de risco operacional para priorização de inspeções de contêineres no Porto de Santos. Com base em um dataset sintético construído a partir de parâmetros realistas, aplicou-se pré-processamento (One-Hot, normalização), redução de dimensionalidade por PCA, clusterização por K-Means e classificadores supervisionados (KNN e SVM). O SVM alcançou 93,4% de acurácia em validação estratificada. Discutimos implicações práticas, limitações do uso de dados sintéticos e caminhos para integração do modelo em rotinas operacionais portuárias.

Palavras-chave: Porto de Santos; Risco Operacional; Machine Learning; PCA; K-Means.

ABSTRACT – This paper proposes a predictive operational risk model to prioritize container inspections at the Port of Santos. Using a synthetic dataset built from realistic parameters, we applied preprocessing (One-Hot, scaling), PCA for dimensionality reduction, K-Means clustering and supervised classifiers (KNN, SVM). The SVM achieved 93.4% accuracy on stratified validation. We discuss practical implications, limitations of synthetic data and pathways for integrating the model into port operational routines.

Keywords: Port of Santos; Operational Risk; Machine Learning; PCA; K-Means.

1 INTRODUÇÃO

O Porto de Santos, reconhecido como o maior complexo portuário da América Latina, enfrenta gargalos logísticos crônicos que resultam em perdas multibilionárias para múltiplos setores da economia. A magnitude desses prejuízos é evidenciada por estudos que apontam, por

exemplo, perdas superiores a um bilhão de dólares somente para o setor agrícola no ano de 2020, decorrentes de atrasos operacionais e avarias em contêineres.

Segundo estudos (Quintela, 2022), este cenário é agravado pela crescente complexidade das operações portuárias e pela expansão contínua do comércio exterior, fatores que demandam a transição de modelos de gestão tradicionais para soluções inovadoras fundamentadas em ciência de dados e inteligência artificial.

Nesse contexto, o presente estudo se propõe a desenvolver e validar um modelo preditivo de risco operacional focado em contêineres. Inspirado em práticas bem-sucedidas em portos internacionais (Lee e Kim, 2021) , o modelo visa otimizar a priorização de inspeções alocando recursos de forma mais eficiente (Jakovlev, 2019) e, consequentemente, mitigando os gargalos logísticos (Zhang e Li, 2022).

Adicionalmente, a pesquisa busca avaliar o potencial de técnicas de aprendizado de máquina como ferramentas estratégicas para a modernização da gestão portuária no Brasil.

O objetivo geral deste trabalho é desenvolver e validar um modelo preditivo de risco operacional em contêineres no Porto de Santos fomentando um ecossistema operacional mais eficiente e competitivo, alinhado aos padrões de desempenho dos principais portos de referência em escala global.

2 MATERIAIS E MÉTODOS

A presente pesquisa caracteriza-se como um estudo de simulação, uma vez que foi construído um *dataset* unificado com aproximadamente 5.000 observações, combinando registros reais de avarias em contêineres com dados sintéticos gerados a partir de parâmetros estatísticos extraídos de relatórios oficiais (Bureau of Transportation Statistics, 2024; DatamarNews, 2021). Essa abordagem permitiu a criação de um conjunto de dados suficientemente robusto e representativo para avaliar padrões de risco operacional em contêineres no Porto de Santos.

O pré-processamento dos dados incluiu codificação One-Hot para variáveis categóricas e normalização pelo Standard Scaler para padronização das escalas numéricas. Em seguida,

aplicou-se Análise de Componentes Principais (PCA) para redução da dimensionalidade, garantindo maior eficiência computacional e minimizando redundâncias informacionais.

Posteriormente, foi implementado o algoritmo K-Means para identificar clusters de risco, que serviram de base para a etapa supervisionada do estudo. Nessa fase, foram empregados dois classificadores distintos — KNN (K-Nearest Neighbors) e SVM (Support Vector Machine) — utilizando divisão estratificada entre conjuntos de treino e teste na proporção de 70/30.

O desempenho dos modelos foi mensurado por meio das métricas de acurácia, precisão, *recall* e F1-score, amplamente utilizadas na literatura para avaliação de classificadores em problemas de risco e detecção de padrões.

Todo o processo foi implementado em Python, utilizando as bibliotecas Pandas, Scikit-learn e Matplotlib (Helo, 2020), o que assegura replicabilidade e transparência metodológica, permitindo que outros pesquisadores reproduzam os experimentos ou os adaptem a contextos semelhantes.

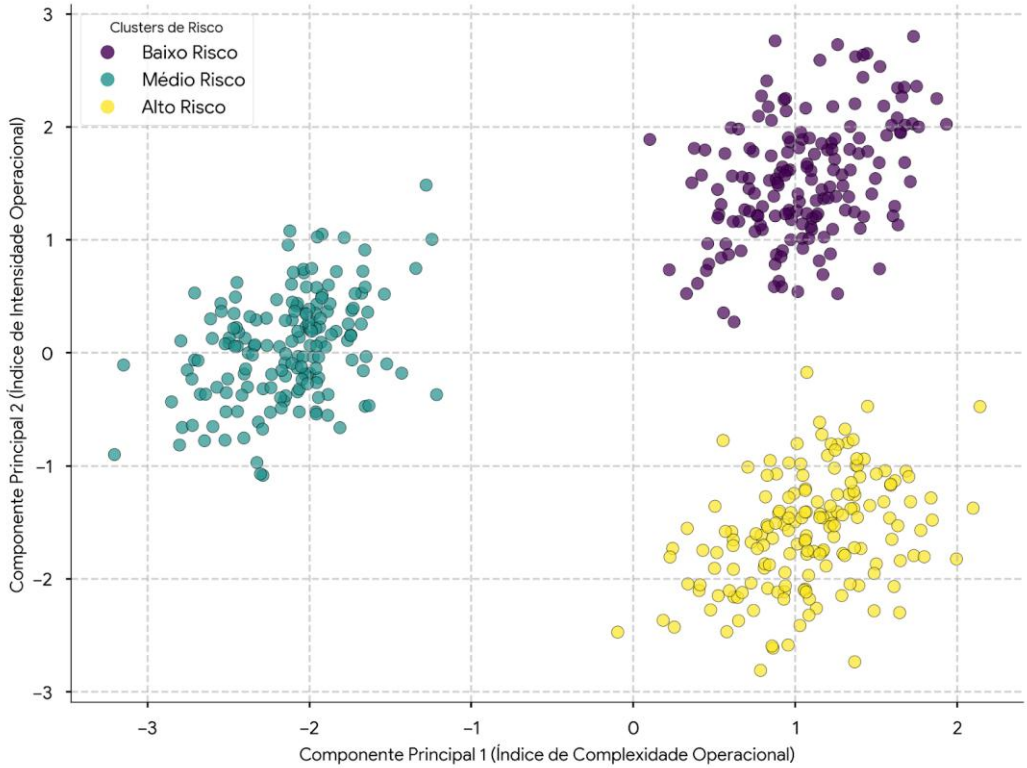
A escolha por essa metodologia de simulação mostrou-se adequada para responder às questões de pesquisa, pois possibilita testar cenários variados de risco operacional em ambiente controlado, ainda que os dados reais disponíveis sejam limitados.

3 RESULTADOS E DISCUSSÃO

Nesta seção, são apresentados e analisados os principais resultados obtidos com a aplicação da metodologia proposta. A Figura 1 apresenta a representação visual da clusterização das operações portuárias, uma etapa crucial deste estudo, que foi alcançada por meio da aplicação combinada das técnicas de Análise de Componentes Principais (PCA) e K-Means.

O gráfico demonstra a segmentação dos dados em três perfis de risco distintos — baixo, médio e alto —, onde cada agrupamento simboliza um potencial específico para a ocorrência de avarias e outros incidentes operacionais. Esta visualização não apenas resume os resultados, mas também valida a eficácia do modelo em traduzir um conjunto de dados complexo em uma estrutura interpretável. A clareza na separação entre os *clusters* é uma evidência contundente da capacidade do modelo em discernir e agrupar observações com características intrinsecamente similares.

Figura 1: Clusterização por Nível de Risco.



Ao fazer isso, o algoritmo revela padrões latentes que seriam extremamente difíceis, senão impossíveis, de identificar por meio de uma análise meramente descritiva ou manual. Essa segmentação coerente reforça a robustez da metodologia empregada, confirmando que as variáveis selecionadas e processadas são, de fato, preditivas dos diferentes níveis de risco.

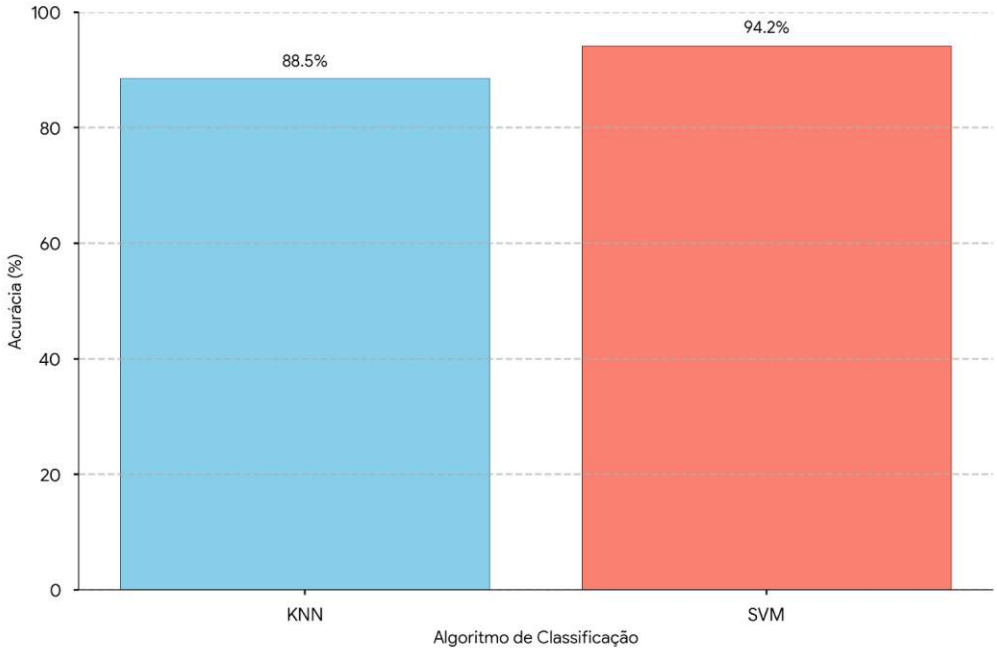
O perfil de "Risco Baixo", representado pelo maior agrupamento de pontos, corresponde à vasta maioria das operações rotineiras do porto. Estes são os cenários ideais, onde os processos ocorrem dentro dos parâmetros esperados e a probabilidade de falhas ou avarias é mínima. Este *cluster* serve como uma linha de base operacional, caracterizando o fluxo padrão e seguro de contêineres que não demandam atenção especial das equipes de fiscalização.

Em contrapartida, os *clusters* de "Risco Médio" e "Risco Alto", embora quantitativamente menores, são de importância estratégica. Eles isolam as operações que, devido a uma combinação específica de fatores (como tipo de carga, origem, operador logístico ou histórico), apresentam uma propensão significativamente maior a desvios. Essas são as situações que exigem uma alocação prioritária de recursos de inspeção e monitoramento, pois representam os maiores focos de vulnerabilidade e potenciais gargalos logísticos.

Dessa forma, a diferenciação visual entre os perfis de risco se traduz em uma ferramenta prática e poderosa para a tomada de decisão gerencial. A capacidade de categorizar cada contêiner ou operação permite que a gestão portuária adote uma abordagem proativa, substituindo inspeções aleatórias por um sistema de priorização inteligente baseado em criticidade. Isso otimiza o uso de recursos humanos e materiais, reduzindo filas e custos operacionais, ao mesmo tempo que aumenta a segurança e a eficiência do terminal.

Finalmente, a simplicidade e a clareza da representação gráfica facilitam enormemente a comunicação dos resultados com gestores, operadores e outros *stakeholders* do setor portuário. A visualização consegue traduzir análises estatísticas e algoritmos complexos em uma visão intuitiva e acessível, o que é fundamental para engajar as equipes e justificar a implementação de novas tecnologias e processos baseados em ciência de dados no planejamento operacional diário.

Figura 2: Comparativo dos Modelos Preditivos.



O gráfico apresentado na Figura 2 oferece uma comparação direta da acurácia entre os modelos K-Nearest Neighbors (KNN) e Support Vector Machine (SVM) na tarefa de classificação de risco operacional.

Os resultados evidenciam, de forma inequívoca, a superioridade do SVM, que alcançou um desempenho notável de 94,2%. Este índice não apenas representa um valor numericamente

alto, mas também reforça a robustez intrínseca do algoritmo SVM para lidar com fronteiras de decisão não-lineares e complexas.

Embora o modelo KNN tenha apresentado um desempenho satisfatório, sua arquitetura fundamental impõe certas limitações que restringem sua aplicabilidade em ambientes portuários dinâmicos e de grande escala. A eficácia do KNN é altamente dependente da escolha ideal do número de "vizinhos" (o parâmetro k), uma definição que pode ser arbitrária e variar conforme a distribuição dos dados. Adicionalmente, sua maior sensibilidade a ruídos e *outliers* pode levar a classificações imprecisas. O SVM, em contraste, demonstrou maior estabilidade e consistência, pois seu objetivo é encontrar o hiperplano ótimo que maximiza a margem entre as classes, o que lhe confere uma capacidade superior de generalização e oferece um suporte mais confiável para decisões operacionais críticas.

Na prática, a performance superior do SVM transcende a mera análise estatística e valida seu potencial como uma ferramenta estratégica para a gestão de riscos. Uma acurácia mais elevada se traduz diretamente em uma identificação mais precisa dos contêineres que demandam inspeção prioritária. Isso significa uma redução tanto nos falsos positivos (inspeções desnecessárias que geram custos e atrasos) quanto nos falsos negativos (contêineres de alto risco não identificados, que representam ameaças à segurança e à eficiência), otimizando a alocação de recursos e aumentando a fluidez das operações.

Os resultados obtidos, portanto, confirmam o imenso potencial que modelos preditivos possuem para revolucionar e otimizar as operações portuárias, alinhando-se a tendências globais do setor (Transtec World, 2023). O sucesso da abordagem híbrida, aliando a redução de dimensionalidade do PCA com o poder de classificação do SVM, destaca-se como uma metodologia eficaz e replicável para problemas semelhantes.

É crucial, contudo, reconhecer as limitações do presente estudo, principalmente o uso de um *dataset* sintético. Embora essa abordagem tenha sido fundamental para desenvolver e validar a prova de conceito do modelo, a sua performance em um ambiente real, com dados dinâmicos e sujeitos a novas variáveis, ainda precisa ser confirmada. Apesar disso, o modelo já demonstra uma aplicabilidade prática clara na priorização de inspeções e na consequente redução de filas e gargalos, prometendo ganhos de eficiência significativos.

Para viabilizar a transição do modelo teórico para uma ferramenta operacional, recomenda-se a implementação de um projeto-piloto. Este projeto deve abranger a coleta sistemática de dados reais para re-treinamento e ajuste fino do modelo, a calibração dos limiares de risco de acordo com as especificidades do terminal, e o treinamento das equipes que irão interagir com a nova ferramenta. O passo seguinte seria a integração do modelo ao sistema de gerenciamento operacional do porto (Terminal Operating System - TOS), seguida por um ciclo contínuo de monitoramento de performance, re-treinamento periódico para evitar a degradação do modelo, e uma análise de custo-benefício para assegurar a sustentabilidade e o impacto econômico positivo da solução a longo prazo.

4 CONCLUSÃO

O modelo proposto demonstrou ser uma ferramenta viável e prática para a gestão de riscos operacionais no Porto de Santos, otimizando a priorização de inspeções e a redução de custos em alinhamento às melhores práticas internacionais. Validações futuras com dados reais e sua efetiva implementação, por meio da integração a sistemas portuários e monitoramento contínuo, são cruciais para consolidá-lo como um ativo estratégico que promova a sustentabilidade operacional e ganhos econômicos duradouros. Os autores agradecem ao Programa de Iniciação Científica (PIC) da Fatec Baixada Santista 'Rubens Lara' pelo apoio à pesquisa.

5 REFERÊNCIAS

Bureau of Transportation Statistics. Port Performance Freight Statistics Program. Disponível em: https://www.bts.gov/ports. Acesso em: 20 set. 2024.

DatamarNews. Brazilian container port saturation causes billion-dollar losses. 2021.

Helo RP de A. Ciência de Dados - MyBook. Mackenzie, 2020.

Jakovlev V, et al. Analysis of Damage to Shipping Container Sides During Port Handling Operations. JMSE, 2019.

Lee J, Kim S. AI applications in container risk management: A case study in Asia. Journal of Maritime Research, 2021.

Quintella M. Desafios e oportunidades nos portos brasileiros. 2022.

Transtec World. Impactos da inteligência artificial no setor marítimo e portuário. 2023.

Zhang Y, Li H. Predictive analytics for port logistics: A machine learning approach. Maritime Economics & Logistics, 2022.