Anais do Congresso Brasileiro de Iniciação Científica
(https://periodicos.unisanta.br/COB)

ISSN 2317-1316

OTIMIZAÇÃO INTELIGENTE DE CONTÊINERES NO PORTO DE SANTOS: INTEGRANDO IoT E *MACHINE LEARNING* PARA EFICIÊNCIA OPERACIONAL

INTELLIGENT CONTAINER OPTIMIZATION AT THE PORT OF SANTOS: INTEGRATING IoT AND MACHINE LEARNING FOR OPERATIONAL EFFICIENCY

Robert Richard das Neves Correia dos Santos; Priscila Bayer de Oliveira Simões; José Augusto Theodosio Pazetti

Fatec Baixada Santista, Rubens Lara, Curso de Ciência de Dados
Universidade de São Paulo - USP
E-mail para contato: robert.santos@aluno.cps.sp.gov.br

RESUMO – Este trabalho descreve os resultados do projeto "Otimização Inteligente de Contêineres no Porto de Santos". Inicialmente focada em sensores de IoT, a pesquisa evoluiu para a análise preditiva de risco operacional e, em seguida, para a Governança Preditiva, ao constatar que os maiores desafios à eficiência estão na fragmentação informacional, e não na infraestrutura física. Compilou-se um dataset de 17.364 registros e aplicaram-se técnicas de Machine Learning (PCA, K-Means e XGBoost) para analisar o Lead Time. Os resultados evidenciaram que os tempos burocráticos exercem impacto dominante no atraso logístico. Conclui-se que modelos preditivos e a explicabilidade analítica (SHAP Values) fundamentam a Governança Preditiva e a tomada de decisão compartilhada.

Palavras-chave: Governança Preditiva; Porto de Santos; *Machine Learning*; Ciência de Dados; Eficiência Operacional.

ABSTRACT – This paper describes the results of the project "Intelligent Container Optimization at the Port of Santos". Initially focused on IoT sensors, the research evolved into predictive operational risk analysis and then into Predictive Governance, after finding that the main barriers to efficiency lie in information fragmentation rather than in physical infrastructure. A dataset of 17,364 records was compiled and Machine Learning techniques (PCA, K-Means and XGBoost) were applied to analyze Lead Time. The results showed that bureaucratic times have a dominant impact on logistics delays. The study concludes that predictive models and analytical explainability (SHAP Values) underpin Predictive Governance and shared decision-making.

Keywords: Predictive Governance; Port of Santos; Machine Learning; Data Science; Operational Efficiency.

ISSN 2317-1316

1. INTRODUÇÃO

O Porto de Santos, maior terminal da América Latina, convive com processos de inspeção manual que elevam custos e gargalos logísticos, gerando perdas estimadas em US$ 555 milhões (DATAMARNEWS, 2025).

A proposta inicial visava o uso de sensores de IoT e algoritmos preditivos para reduzir custos operacionais e o tempo de inspeção de contêineres físicos. Com o avanço da pesquisa, constatou-se que os maiores desafios transcendem a infraestrutura física, o que levou à reorientação do foco para a Análise Preditiva de Risco Operacional e, na fase final, para a Governança Preditiva, diante da evidência de que os gargalos críticos residem na fragmentação informacional entre órgãos anuentes, entes públicos e privados.

Nesse contexto, é essencial o uso da Ciência de Dados para analisar a eficiência dos processos burocráticos, entendendo como a transparência analítica pode fundamentar a tomada de decisão e reduzir os gargalos no complexo portuário.

2. METODOLOGIA

A primeira etapa da pesquisa, vinculada ao Programa de Iniciação Científica (PIC) e posteriormente contemplada pela bolsa MIDTI, adotou abordagem aplicada com dados simulados (Google Colaboratory, Python), cobrindo seis meses de operação (out. 2024 a mar. 2025). O risco operacional foi compreendido como a consolidação de fatores de integridade estrutural, segurança física, integridade da carga, falha de monitoramento e risco logístico.

Diante da necessidade de validação prática, a pesquisa avançou para um *dataset* real de aproximadamente 8 mil registros, compilado a partir de fontes públicas internacionais (BTS, DatamarNews/DataLiner, Port of Hamburg Authority, PSA Singapore e o CPPI do Banco Mundial). Após o pré-processamento (*One-Hot Encoding* e *Standard Scaler*), aplicou-se PCA para redução a duas dimensões e K-Means para agrupar as operações em três *clusters* de risco, usados para treinar e comparar os classificadores KNN e SVM (divisão 70/30).

Na fase final, a metodologia baseou-se em pesquisa aplicada e quantitativa, estruturada sob o ciclo de vida da Ciência de Dados, em Python: (1) construção de um *dataset* inédito

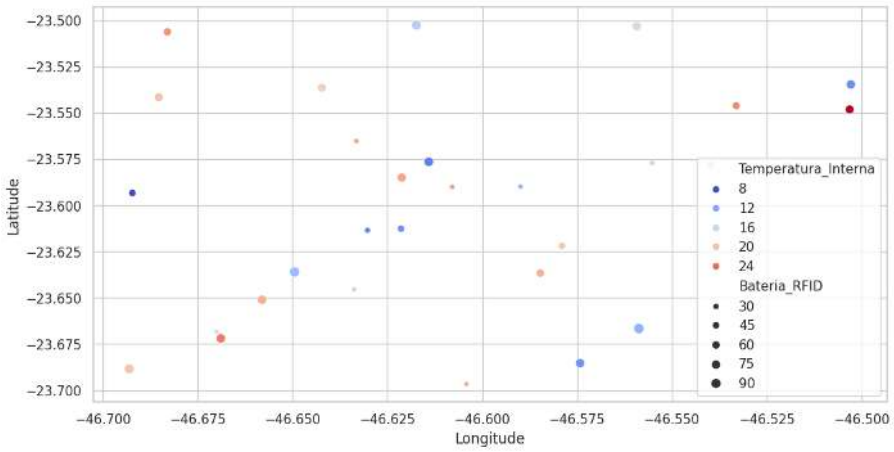
ISSN 2317-1316

(2023-2025) com 17.364 registros, integrando ANTAQ, Estatísticas de Anuências e Porto Sem Papel; (2) pré-processamento com *KNN Imputer*, *One-Hot Encoding* e *Standard Scaler* (PEDREGOSA et al., 2011); (3) redução de dimensionalidade via PCA e clusterização via K-Means, traçando perfis dos *stakeholders* (HASTIE; TIBSHIRANI; FRIEDMAN, 2009; AMAZON WEB SERVICES, [s.d.]); (4) predição do *Lead Time* com XGBoost, com otimização via Optuna (PINHEIRO, 2023); e (5) explicabilidade via *SHAP Values*, para quantificar os impactos e fundamentar a tomada de decisão.

3. RESULTADOS E DISCUSSÃO

Os resultados obtidos ao longo da pesquisa são apresentados a seguir.

Figura 1: Mapa de Calor (Temperatura e Bateria).

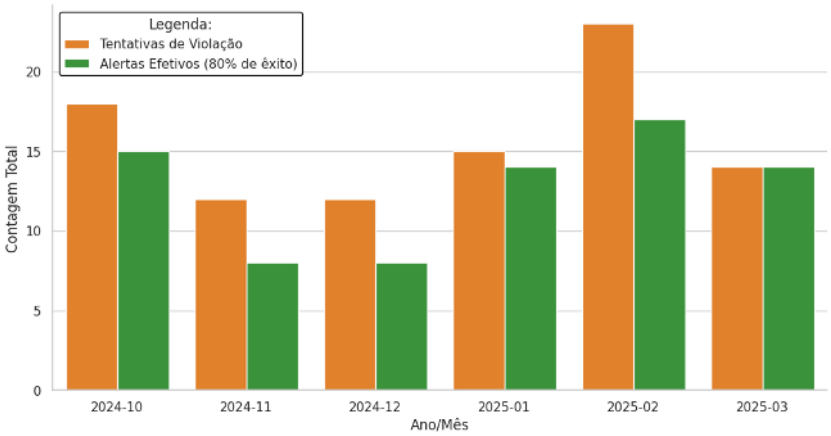


Fonte: Elaborado pelos autores, 2025.

Na Figura 1, o mapa de calor georreferenciado, que cruza a temperatura interna e a carga da bateria dos contêineres, mostrou-se um recurso eficaz para a identificação de *clusters* críticos e para o suporte a decisões de manutenção preventiva.

ISSN 2317-1316

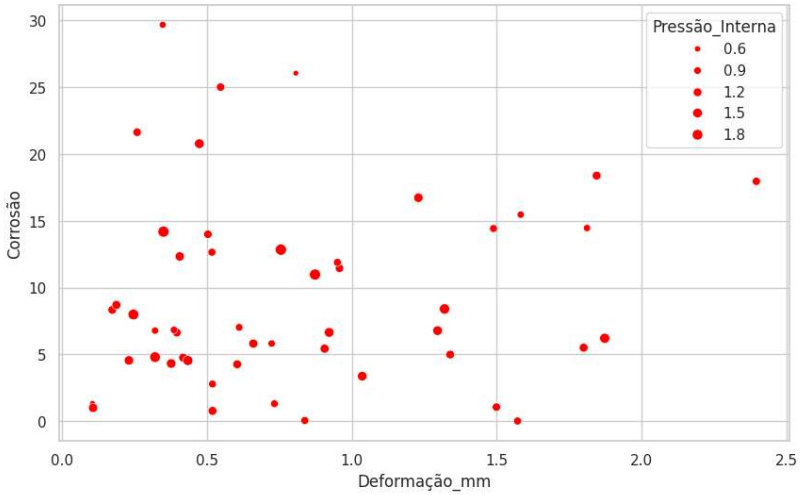
Figura 2: Gráfico de Barras (Violações e Alertas).



Fonte: Elaborado pelos autores, 2025.

O gráfico de desempenho de alertas da Figura 2 indicou uma taxa de sucesso de 80% na detecção de tentativas de violação, com picos de incidentes que evidenciaram fragilidades temporárias no monitoramento.

Figura 3: Risco de Acidentes.



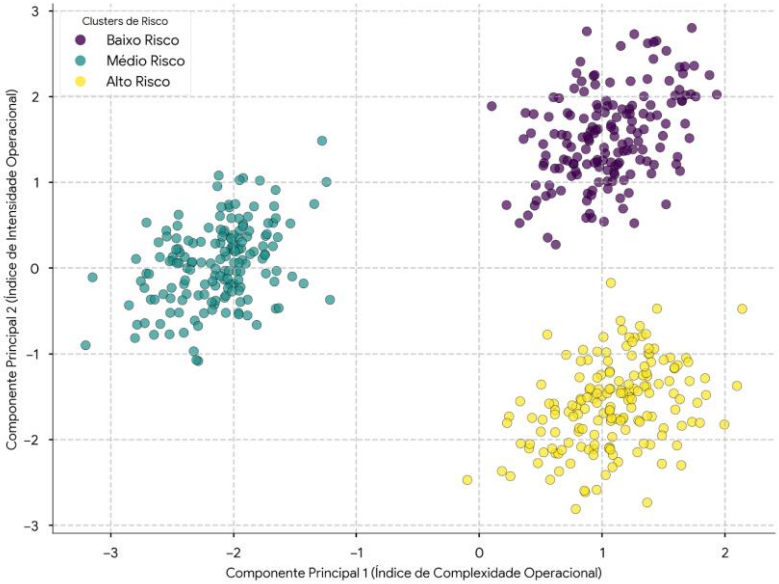
Fonte: Elaborado pelos autores, 2025.

A análise de correlação entre deformação, corrosão e pressão interna, vista na Figura 3, revelou que a combinação de alta pressão e alta corrosão configura risco estrutural crítico,

ISSN 2317-1316

indicando que a IoT contribui para a eficiência e a segurança portuária, ainda que os dados simulados exigissem validação com dados reais.

Figura 4: Clusterização por Nível de Risco.

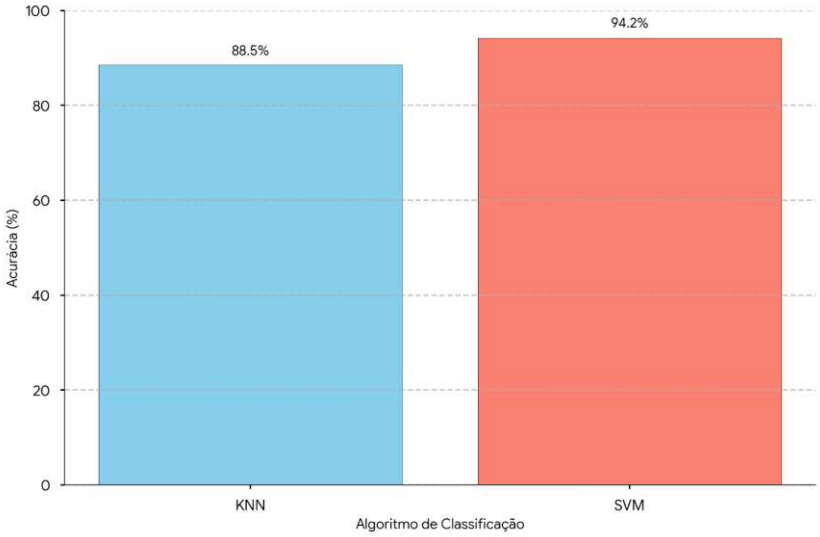


Fonte: Elaborado pelos autores, 2025.

Na Figura 4, referente à etapa seguinte, conduzida sobre o *dataset* real de aproximadamente 8 mil registros, o PCA e o K-Means evidenciaram padrões distintos de risco operacional.

ISSN 2317-1316

Figura 5: Comparativo dos Modelos Preditivos.



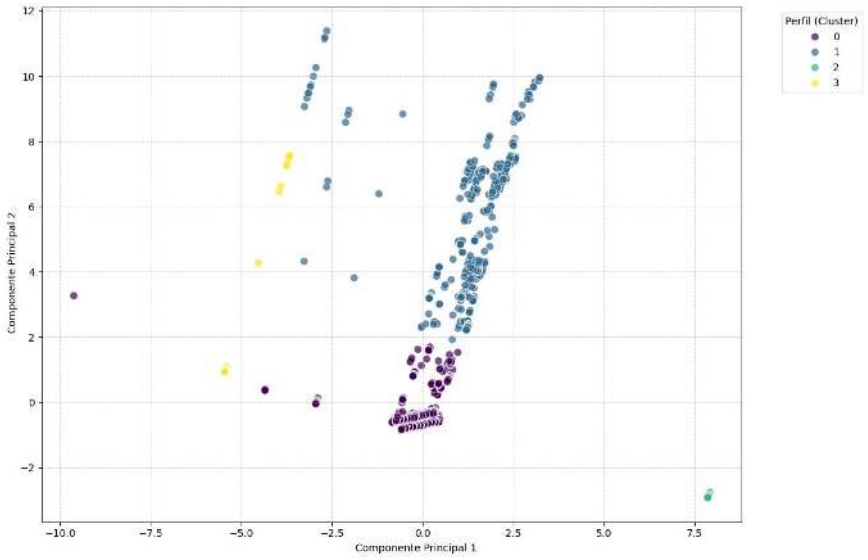
Fonte: Elaborado pelos autores, 2025.

Na comparação apresentada na Figura 5, o SVM atingiu acurácia de 94,2%, superando o KNN (88,5%), o que reforça o potencial de modelos preditivos para apoiar decisões e reduzir perdas operacionais, resultado coerente com o contexto de perdas associadas aos gargalos logísticos (DATAMARNEWS, 2025).

Os resultados preditivos e estatísticos finais, obtidos sobre o *dataset* de 17.364 registros, evidenciaram que a fragmentação informacional constitui um gargalo organizacional grave, mas que pode ser revertido por meio da visibilidade analítica.

ISSN 2317-1316

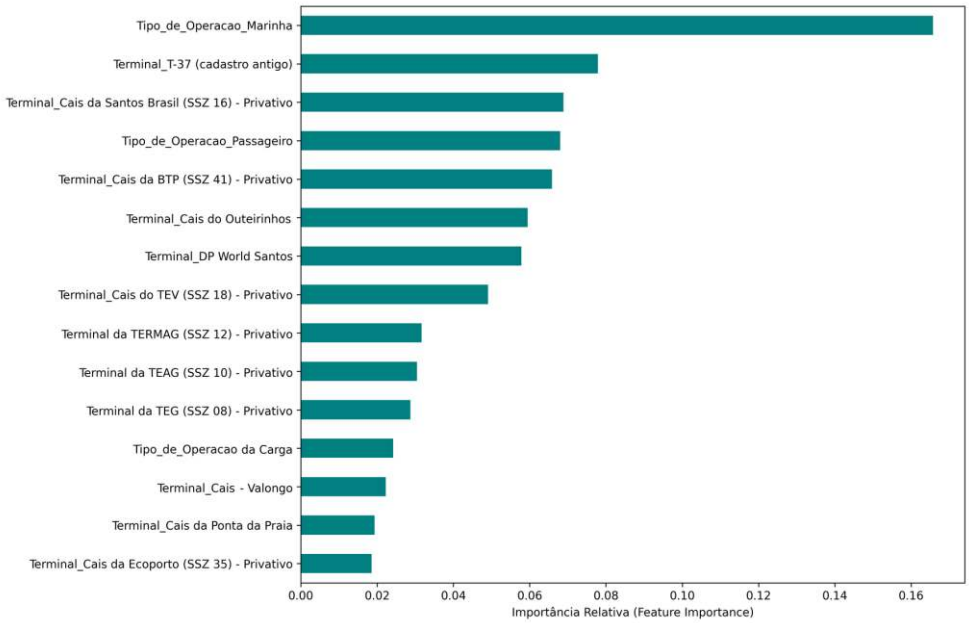
Figura 6: Agrupamento das Operações Portuárias.



Fonte: Elaborado pelos autores, 2026.

Na Figura 6, a clusterização via K-Means dividiu o ecossistema em quatro grupos, demonstrando que a capacidade técnica de integração de dados influencia diretamente a eficiência na movimentação das cargas.

Figura 7: Ranking de Importância das Variáveis.

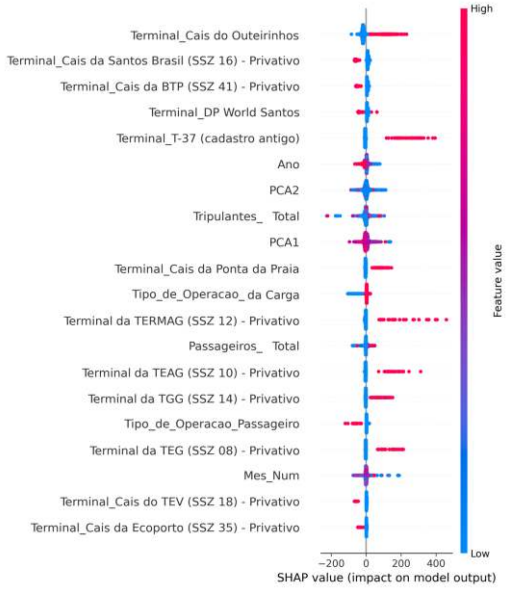


Fonte: Elaborado pelos autores, 2026.

ISSN 2317-1316

O modelo XGBoost otimizado, referente à Figura 7, expôs a hierarquia de importância das variáveis de operação, revelando que os tempos de liberação burocrática exercem dominância estatística sobre o atraso final.

Figura 8: Explicabilidade do Impacto das Variáveis.



Fonte: Elaborado pelos autores, 2026.

O diagnóstico, visto na Figura 8, evidenciou como a falta de integração dos anuentes prolonga a operação, deslocando-a para zonas críticas de ineficiência, embasamento que pode ser usado para criar políticas de compartilhamento de dados.

4. CONCLUSÃO

O projeto evoluiu de um foco em componentes físicos para o diagnóstico de que o verdadeiro gargalo é organizacional: a lacuna de governança das informações portuárias. A Ciência de Dados mostrou-se base robusta para a Governança Preditiva, mediando discussões entre *stakeholders* com dados unificados. Como projeção futura, planeja-se integrar esse motor preditivo a *frameworks* de monitoramento em tempo real, permitindo agir contra o *Lead Time* excessivo antes de sua ocorrência.

ISSN 2317-1316

5. DECLARAÇÃO DE USO DE IAG (INTELIGÊNCIA ARTIFICIAL GENERATIVA)

"Foi utilizada a ferramenta de Inteligência Artificial Generativa Claude Sonnet 5.5 (Anthropic) nas etapas de formatação do artigo conforme as normas do edital do 18º COBRIC, tradução do título e do resumo para o inglês e revisão textual, com a finalidade de adequação às normas, correção ortográfica e de pontuação e melhoria da clareza do texto. A ferramenta não foi listada como autora, e a responsabilidade intelectual, ética e legal pelo conteúdo é exclusiva dos autores".

6. REFERÊNCIAS

AMAZON WEB SERVICES. Como funciona o clustering do k-means. [S. l.]: Amazon Web Services, [s. d.]. Disponível em: https://docs.aws.amazon.com/pt_br/sagemaker/latest/dg/algo-kmeans-tech-notes.html. Acesso em: 13 nov. 2025.

DATAMARNEWS. Brazilian container port saturation causes billion-dollar losses. [S. l.]: DatamarNews, 2025.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. The elements of statistical learning: data mining, inference, and prediction. 2. ed. New York: Springer, 2009.

PEDREGOSA, Fabian et al. Scikit-learn: machine learning in Python. Journal of Machine Learning Research, v. 12, p. 2825–2830, 2011.

PINHEIRO, João Manoel Herrera. Um estudo sobre algoritmos de boosting e a otimização de hiperparâmetros utilizando Optuna. 2023. 147 f. Trabalho de Conclusão de Curso (Graduação em Engenharia Mecatrônica) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2023.