

## Validação de modelos de machine learning por experimentos estatísticos de campo

### Validation of machine learning models through statistical field experiments

### Validación de modelos de aprendizaje automático mediante experimentos estadísticos de campo

#### Como citar:

Toaldo, Alessandro; Vallim Filho, Arnaldo R. A.; Oyadomari, José C. T. & Mendonça Neto, Octavio R. (2024). Validação de modelos de machine learning por experimentos estatísticos de campo. Revista Gestão & Tecnologia, vol. 24, nº 5, p. 125-153

Alessandro Toaldo

Doutor em Finanças e Controladoria, pela Universidade Presbiteriana Mackenzie

<https://orcid.org/0000-0003-3132-9020>

Arnaldo Rabello de Aguiar Vallim Filho

Professor Titular, Pesquisador e consultor, Universidade Presbiteriana Mackenzie

<https://orcid.org/0000-0003-4100-4975>

José Carlos Tiomatsu Oyadomari

Professor adjunto da Universidade Presbiteriana Mackenzie Mackenzie

<https://orcid.org/0000-0003-3059-3102>

Octavio Ribeiro de Mendonça Neto

Doutor em Contabilidade e Atuária na FEA / USP (2007) e Professor Adjunto I da Universidade Presbiteriana Mackenzie

<https://orcid.org/0000-0002-6123-6733>

Os autores declaram não haver qualquer conflito de interesse pessoal ou institucional entre a pesquisa e as organizações em estudo.

Editor Científico: José Edson Lara  
Organização Comitê Científico  
Double Blind Review pelo SEER/OJS  
Recebido em 03/02/2024  
Aprovado em 04/11/2024



This work is licensed under a Creative Commons Attribution – Non-Commercial 3.0 Brazil



## Resumo

**Objetivo** – Este artigo apresenta uma aplicação prática com o desenvolvimento de um experimento estatístico de campo em uma indústria de latas *premium* de alumínio nos Estados Unidos, visando validar estatisticamente resultados de modelos de *machine learning* (ML), previamente construídos.

**Metodologia:** Usou-se conceitos de pesquisa intervencionista, que envolve experimentos de campo onde pesquisador e organização anfitriã atuam em conjunto buscando experimentar no sistema em estudo, e por meio da observação gerar conhecimento.

**Originalidade/Relevância:** Sobre originalidade, não é frequente na literatura modelos de ML validados por experimento planejado de campo, seguido de análise estatística rigorosa. E a relevância da proposta se deve à sua contribuição para a literatura e pelas possibilidades de replicações do estudo em escala maior, na própria empresa ou em qualquer outra com desafios similares.

**Principais Resultados:** Em fase anterior do estudo modelos de ML identificaram as variáveis de maior impacto em ineficiências (geração de sucata) em um processo de produção de latas de alumínio. Essas variáveis foram validadas nesta fase do estudo, através de experimento estatístico de campo, confirmando a significância estatística dos resultados do modelo de ML.

**Contribuições Teóricas e Práticas:** A pesquisa contribui em termos práticos e científicos, pois a validação estatística de modelos de ML por experimentos planejados de campo é uma contribuição para a literatura de ciência aplicada, além de usar possibilidades práticas. Da mesma forma, apesar de amplamente utilizadas em diferentes áreas, pesquisas de cunho intervencionista ainda apresentam lacuna importante nas ciências sociais aplicadas, principalmente na gestão de processos industriais.

**Palavras-Chave:** machine learning, pesquisa intervencionista, experimento estatístico de campo, eficiência de processos industriais.

## Abstract

**Objective** – This article presents a practical application with the development of a statistical field experiment in a premium aluminum can industry in the United States, aiming to statistically validate results from machine learning (ML) models, built in a previous phase of the study.

**Methodology:** This study uses concepts of interventionist research, which involves field experiments where the researcher and host organization work together seeking to experiment in the system under study, and through observation generate knowledge.

**Originality/Relevance:** Regarding originality, it is not common to find ML models validated by planned field experiments, followed by rigorous statistical analysis. And the proposal relevance is due to its contribution to the literature and the possibilities of replicating the study on a larger scale, in the company itself or in any other company that faces similar challenges.

**Main Results:** In a previous phase of the study, ML models identified the variables with the greatest impact on inefficiencies (scrap generation) in an aluminum can production process.



These variables were validated in this phase of the study, through a statistical field experiment, confirming the statistical significance of the ML model results.

**Theoretical and Practical Contributions:** The research contributes in practical and scientific terms, as the statistical validation of ML models by planned field experiments is a contribution to the applied science literature, in addition to practical possibilities. Likewise, despite being widely used in different areas, interventionist research still presents an important gap in applied social sciences, especially in the management of industrial processes.

**Keywords:** machine learning, interventionist research, statistical field experiment, industrial process efficiency

## Resumen

**Objetivo:** Este artículo presenta un experimento de campo en una industria de latas de aluminio premium en los Estados Unidos, buscando validar estadísticamente los resultados de los modelos de aprendizaje automático (AM), creados previamente.

**Metodología:** Se utilizaron conceptos de investigación intervencionista, donde el investigador y la organización anfitriona trabajan juntos buscando experimentar en campo en el sistema en estudio y, a través de la observación, generar conocimiento.

**Originalidad/Relevancia:** En cuanto a la originalidad, no es común encontrar modelos de AM validados mediante experimentos de campo planificados, seguidos de análisis estadísticos rigurosos. Y la relevancia de la propuesta se debe a su aporte a la literatura y a las posibilidades de replicar el estudio a mayor escala, en la empresa o en otra empresa con desafíos similares.

**Principales Resultados:** En una fase previa del estudio, los modelos AM identificaron las variables con mayor impacto en las ineficiencias (generación de chatarra) en la producción de latas de aluminio. Estas variables fueron validadas en esta fase del estudio, mediante un experimento estadístico de campo, confirmando la significancia estadística de los resultados del modelo AM.

**Contribuciones Teóricas y Prácticas:** La investigación contribuye en términos prácticos y científicos, ya que la validación estadística de los modelos AM mediante experimentos de campo planificados es una contribución a la literatura científica aplicada, además de las posibilidades prácticas. Asimismo, a pesar de ser ampliamente utilizada en diferentes áreas, la investigación intervencionista aún presenta un vacío importante en las ciencias sociales aplicadas, especialmente en la gestión de procesos industriales.

**Palabras clave:** aprendizaje automático, investigación intervencionista, experimentos estadísticos de campo, eficiencia de procesos industriales.

## 1. INTRODUÇÃO

O *Machine Learning* (ML) tem sido amplamente utilizado nas mais diferentes áreas de atividade, entretanto nem sempre se encontra na literatura uma validação de campo dos resultados obtidos em um modelo preditivo de ML. Este artigo tem por foco justamente esse tipo de validação por meio de um experimento de campo, projetado com base em protocolos consolidados de planejamento de experimentos estatísticos, e de análises estatísticas clássicas voltadas para resultados de experimentos. Busca-se assim, verificar se resultados obtidos em modelos preditivos de ML, construídos previamente, se confirmam em campo.

Neste estudo, modelos preditivos foram desenvolvidos previamente, para prever uma medida da perda de eficiência em processos de produção industrial. E esses modelos, forneceram como resultado complementar, as variáveis mais importantes para a geração das previsões que foram feitas. E foi, justamente, esse conjunto de variáveis mais importantes, que se procurou validar em campo, verificando-se o efeito dessas variáveis na variável de resposta do experimento de campo.

A aplicação prática, se desenvolveu em um caso real de perda de eficiência em um processo industrial de produção de latas de latas premium de alumínio, que foi medida pelo nível de geração de sucata no processo, correspondente a perda de material. Foram explorados alguns tipos de modelos preditivos de forma a se selecionar aqueles que seriam mais adequados para modelar o processo. Os resultados desses modelos mostraram um conjunto de variáveis que se destacaram como relevantes na geração daquela sucata, e os indicadores de acurácia desses modelos mostraram valores significativos, dando robustez aos resultados. Na segunda fase dessa pesquisa, procurou-se então, validar as variáveis mais relevantes identificadas nos modelos de ML, por meio de um experimento estatístico de campo.

Para mostrar toda a abrangência do estudo, o artigo apresenta inicialmente, um relato do trabalho desenvolvido na fase 1 da pesquisa, descrevendo os modelos de ML que foram construídos e os seus respectivos resultados. Com isso, pode-se ter uma visão clara dos objetivos do experimento de campo, que representa a fase 2 da pesquisa, e que foi desenvolvido no mesmo processo industrial em que os modelos de ML foram construídos. Métodos clássicos

de análise estatística foram aplicados para dar significância estatística aos resultados do experimento, e, conseqüentemente, para validação dos resultados dos modelos de ML.

Em termos de estrutura, este artigo tem seqüência na seção 2, com uma breve revisão de conceitos associados ao tema em estudo. Em seguida, na seção 3, apresenta-se a descrição da primeira parte da pesquisa, mostrando os modelos de ML desenvolvidos e os resultados obtidos. Na seção 4, adentra-se ao foco principal deste artigo, apresentando-se a metodologia aplicada ao experimento de campo. Em seguida na seção 5, tem-se toda a descrição do experimento de campo, e da avaliação estatística de seus resultados. Finalmente, na última seção, são apresentadas as conclusões da pesquisa e recomendações para trabalhos futuros.

## 2. REVISÃO DE CONCEITOS

Nesta seção, será feita uma revisão conceitual sobre os pontos principais envolvidos no artigo. Inicialmente tem-se uma subseção sobre ML, em seguida, fala-se de pesquisa intervencionista, pesquisa-ação e pesquisa construtivista, que em termos conceituais sustentaram a proposta de metodologia levada a efeito na aplicação de campo e, por último, tem-se uma subseção que trata de planejamento estatístico de experimentos, que norteou toda a coleta e análise de dados do experimento de campo.

### 2.1 *Machine Learning*

O ML é uma das áreas da inteligência artificial (IA) em que um algoritmo “aprende” a partir de características observadas em um conjunto de dados. Russel e Norvig (2020), distinguem esse processo de aprendizado em três tipos:

- **Aprendizado Supervisionado (*Supervised Learning*):** onde algoritmos são treinados a partir de conjuntos de dados com exemplares que apresentam uma série de características (atributos) que descrevem o exemplar, e que um dos atributos represente uma categoria, que, que categorize o exemplar, e que é chamado de rótulo, do exemplar. O objetivo é que o algoritmo aprenda a estabelecer correspondências entre os atributos descritivos dos exemplares e seus respectivos rótulos, criando assim, um modelo preditivo. O aprendizado supervisionado pode ser subdividido em Modelos de Classificação, quando os possíveis valores do atributo rótulo

forem categorias (ou classes) ou Modelo de Regressão, quando os possíveis valores do rótulo forem valores numéricos contínuos.

- **Aprendizado Não-Supervisionado (*Unsupervised Learning*):** Neste caso, os algoritmos exploram conjuntos de dados sem rótulos para identificar padrões, estruturas ou agrupamentos intrínsecos.
- **Aprendizado por Reforço (*Reinforcement Learning*):** que é um tipo de aprendizado em que um agente autônomo aprende a melhorar seu comportamento interagindo com seu ambiente e recebendo um *feedback* na forma de recompensas ou penalidades. O agente deve aprender a escolher as ações que maximizam a recompensa ao longo do tempo.

A modelagem preditiva utilizada neste artigo foi definida como sendo um problema de classificação de dados, que busca justamente, prever categorias (classes) para exemplares de uma base de dados. A classificação de dados é um problema clássico de ML, em que o modelo preditivo “aprende” por meio de indução, induzindo um conjunto de regras de classificação a partir dos atributos descritivos associados a cada exemplar do conjunto de dados. O modelo preditivo se constitui assim, nesse conjunto de regras de decisão (Han et al., 2012). Neste estudo foram utilizadas quatro diferentes técnicas, com o objetivo de explorar diferentes tipos de modelos e assim garantir que os resultados obtidos são efetivamente representativos do processo estudado. São elas: *Decision Tree (DT)*, *Random Forest (RF)*, *Gradient Boosted Trees (GBT)* e *Artificial Neural Net (ANN)*, que são descritas a seguir.

### ***Decision Tree - DT***

Esta é uma técnica clássica, que é assim chamada, porque é uma estrutura hierárquica, na forma de uma árvore, que todos os exemplares de um conjunto de dados percorrem, um a um. Nessa estrutura tem-se nós e ramos, a exemplo de uma árvore. Cada nó da árvore (nó de decisão) representa um teste em um atributo do exemplar, cada ramo a partir de um nó, representa uma saída (um resultado) do teste. Ao final da árvore, tem-se uma definição de uma classe (um rótulo) para os exemplares que chegam àquele nó. Esse final da árvore é um nó folha (ou nó terminal) e representa o rótulo de uma classe.

### ***Random Forest – RF***

A RF proposta por Breiman (2001) é um tipo de método de aprendizado que combina diversas árvores de decisão para produzir uma previsão mais precisa. O algoritmo começa criando muitas árvores de decisão, cada uma das quais divide os dados em diferentes subconjuntos. O algoritmo então seleciona o melhor ponto de divisão para cada subconjunto, e cria outra árvore com base nesse subconjunto. Esse processo continua até que não haja mais pontos de dados para dividir (Han et al., 2012; Cutler et al. 2012).

### ***Gradient Boosted Tree - GBT***

Os conceitos desta técnica foram introduzidos por Schapire (1990), sendo também estudado em várias obras, como: Hastie et al. (2009); Natekin e Knoll (2013) e Krauss et al. (2017). O algoritmo opera por meio de ajustes repetidos de uma árvore de decisão aos dados de treinamento do algoritmo, com o objetivo de minimizar uma função de perda em cada iteração. A cada iteração, o modelo é ajustado conforme a acurácia da previsão. As iterações classificadas incorretamente têm seus pesos aumentados, e os exemplos classificados corretamente, têm seus pesos diminuídos. A árvore é então ajustada aos últimos exemplares de treinamento (Krauss et al., 2017).

### **Artificial Neural Net - ANN**

Uma ANN, assim como outras técnicas de IA, imita um processo da natureza. Uma ANN, especificamente, é um algoritmo computacional que corresponde a um modelo matemático inspirado nas redes neurais biológicas dos cérebros de organismos inteligentes. Permite implementar de forma simplificada, em um algoritmo computacional o funcionamento do cérebro, buscando assim solucionar um problema complexo (Haykin2008). A ANN recebe informações, processa as informações e retorna uma resposta. Uma ANN está estruturada em uma rede de nós (chamados de neurônios) conectados em camadas: uma camada de entrada de dados, camadas intermediárias (pode-se ter zero, uma ou múltiplas camadas intermediárias) e uma camada de saída. As conexões entre os neurônios são chamadas de sinapses (assim como ocorre no cérebro). O arranjo em camadas de neurônios e o padrão de conexões é chamado de arquitetura de rede. Uma arquitetura muito comum é a chamada rede perceptron na qual não existe uma camada intermediária de neurônios, apenas entrada e saída. Uma ANN perceptron com camadas intermediárias é chamada de perceptron multicamadas ou *multilayer perceptron* - MLP (Haykin2008), e é uma das ANN mais utilizadas.

### **2.2 Pesquisa Intervencionista, Pesquisa-Ação e Pesquisa Construtivista**

O estudo aqui desenvolvido propôs uma metodologia de abordagem do experimento de campo que representa uma mescla de pesquisa Intervencionista, pesquisa-Ação e pesquisa Construtivista, o que não é muito comum de se encontrar em artigos voltados para modelos de ML. Assim, é importante que se faça uma breve discussão de cada uma dessas abordagens, em particular, para aqueles que não estão familiarizados com essas metodologias.

A pesquisa Intervencionista pode ser entendida como uma forma mais ampla de estudo de caso, um experimento de campo, tratando-se de um grupo ou conjunto de métodos e técnicas de pesquisa. É um tipo de experimento onde o pesquisador busca experimentar em campo por meio da observação, atuando em conjunto com a organização anfitriã, onde observa processos e resultados e analisa os achados à luz da literatura pertinente. (Roberts, Westin e Dumay, 2010); (Jönsson, 2010). O diferencial fundamental dessa abordagem é a interação entre os pesquisadores e o objeto de estudo, com possibilidades de transformação desse objeto de



estudo. (Oyadomari et al., 2013). Diferentes autores, como Lukka e Vinnari (2017); Suomala e Lyly-Yrjänäinen (2012); e Suomala, Lyly-Yrjänäinen e Lukka (2014), colocam que a pesquisa Intervencionista é uma abordagem de estudo de caso longitudinal com uso ativo da observação do participante, visando não apenas explicar, mas também interferir na realidade estudada para modificá-la. Jönsson & Montgomery (2010) sugerem que a pesquisa intervencionista pode ser encontrada em várias áreas apresentando diferentes variações (nomenclatura, abordagens culturais, geográficas etc.), podendo se dar de formas distintas, como: Pesquisa-ação; Ciência-Ação; Ciência-Design; Pesquisa clínica e Pesquisa construtivista. Suomala, Lyly-Yrjänäinen e Lukka (2014), por outro lado, posicionam esse tipo de pesquisa entre a pesquisa-Ação e a pesquisa Construtivista.

Sobre a pesquisa-Ação, este é um tipo de abordagem que busca unir a pesquisa à ação (à prática), e com isso produzir conhecimento (Engel, 2000; Jönsson, 2010; Tripp, 2005), e é considerada a precursora da pesquisa Intervencionista em ciências sociais. Com algumas variações na metodologia, a pesquisa-Ação, em geral, segue um ciclo de melhoria e resolução de problemas (Tripp, 2005), com algumas etapas estabelecidas. Trata-se de uma investigação de campo a partir da participação conjunta do pesquisador e dos envolvidos no processo no qual a intervenção ocorre. A pesquisa Intervencionista é uma variação da pesquisa Ação (Jönsson, 2010; Lukka e Vinnari, 2017), mas também envolve elementos da pesquisa Construtivista.

A pesquisa Construtivista, por sua vez, tem uma abordagem em que a resolução de problemas se dá por meio da construção de modelos, diagramas, planos etc., buscando vincular o problema e sua solução ao conhecimento teórico. Este tipo de pesquisa foi desenvolvido por pesquisadores finlandeses (Kasanen e Lukka, 1993), como uma opção para estudos no campo da solução de problemas. A exemplo da pesquisa-Ação, segue algumas etapas para a sua aplicação (Labro e Tuomela, 2003). Na abordagem desse tipo de pesquisa, o pesquisador também trabalha em conjunto com os membros da organização anfitriã para desenvolver um construto, testar sua utilidade e fundamentar o processo com base na teoria (Roberts e Jönsson, 2010).

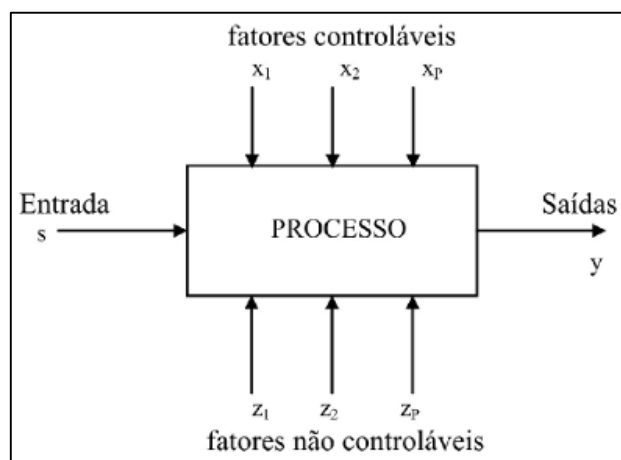
Tem-se assim, um breve quadro dos conceitos metodológicos que nortearam a proposta de metodologia utilizada no experimento de campo, e que será descrita na seção 4.

### 2.3 Planejamento de Experimentos de Campo

Como a pesquisa intervencionista é uma modalidade de experimento de campo, os dados associados ao problema devem ser obtidos por meio de observação em campo do processo sob análise (Roberts, Westin e Dumay, 2010). Entende-se um experimento como sendo um teste ou um conjunto de testes planejados nos quais são feitas alterações em variáveis de entrada, sob as quais se tem controle, com o objetivo de se observar o impacto dessas alterações em uma dada variável de resposta do sistema (Montgomery, 2017). Com isso, se avalia os efeitos desses diferentes fatores nas respostas da variável em estudo, e assim identifica-se as causas das variações naquelas respostas.

Uma forma já consolidada para se implementar esse tipo de experimento, é seguindo os procedimentos clássicos de um Delineamento de Experimentos (*Design of Experiments* - DOE) ou Projeto de Experimentos. O DOE tem sido exaustivamente utilizado em estudos de campo nas mais variadas áreas de atividade. O DOE se constitui de um conjunto de procedimentos que permite que se planeje experimentos com confiabilidade estatística, definindo quais dados, em qual quantidade e sob quais condições esses dados devem ser coletados para um dado experimento, buscando basicamente chegar a dois objetivos maiores: a acuracidade estatística nas respostas do experimento e o menor custo (Laureani & Antony, 2019; Montgomery, 2017; Lawson, 2015; Johnson & Montgomery, 2010).

A Figura 1 representa de forma esquemática o modelo de um experimento em que o DOE pode ser empregado.



**Figura 1:** Modelo geral de um experimento

Fonte: adaptado de Barra Montevechi et al., 2012

Para se avaliar o impacto de fatores controláveis (figura 1) em uma dada variável de resposta, há diferentes tipos de planejamento dos experimentos de campo que podem ser aplicados. As possibilidades de planejamento de experimentos se constituem na verdade, em um campo muito amplo, mas apenas a título de ilustração, pode-se pensar que um experimento pode ir desde algo bem simples, em que se analisa o impacto em uma variável de resposta de apenas um fator, até experimentos bem mais complexos, em que se busca analisar os impactos de múltiplos fatores na variável de resposta, analisando de forma individual cada fator, e de forma combinada, em que se analisa a interação entre fatores e seu efeito na variável de resposta. E neste caso, há muitas formas diferentes de se planejar tais experimentos.

Uma das formas de planejar um experimento que é das mais completas, é o chamado projeto fatorial completo (Montgomery, 2017). Este tipo de experimento permite que se analise todas as combinações possíveis de fatores, pois para cada repetição de um teste com um fator, todas as possíveis combinações dos níveis de todos os outros fatores são investigadas, bem como as principais interações e efeitos entre os fatores controláveis. Um projeto fatorial é importante quando interações entre fatores podem estar presentes, evitando assim, eventuais conclusões errôneas, quando não se considera a existência de tais interações. Por outro lado, quando se deseja analisar um grande número de fatores, a quantidade de combinações e interações vai aumentar significativamente, tornando-se eventualmente, inviável trabalhar com esse tipo de experimento. E nesses casos, há outras alternativas. Uma opção é o uso de projetos

fatoriais fracionados, que analisam apenas uma parte ou fração das combinações possíveis, diminuindo consideravelmente a quantidade de recursos necessários para condução dos experimentos (Montgomery, 2017).

Um dos projetos fatoriais mais importante é aquele com  $k$  fatores, onde para cada fator considera-se apenas dois níveis. Assim, uma replicação completa deste projeto requer  $2^k$  observações. Daí, este tipo de projeto ser chamado Projeto Fatorial  $2^k$ . É uma estratégia que permite identificar interações entre fatores bem como os principais efeitos individuais de cada um (Barra Montevechi et al., 2012). No caso do presente estudo em que o número de fatores não era muito grande, o experimento de campo foi conduzido utilizando-se o projeto fatorial completo.

Para a análise de resultados desses experimentos, costuma-se utilizar a análise de variância (*Analysis of Variance - ANOVA*), que é um procedimento clássico em inferência estatística, consolidado como um padrão para análise de resultados de experimentos (Devore, 2016; Montgomery, 2017). O princípio por trás de uma ANOVA é que quando se testa o efeito de diferentes fatores (tratamentos) em uma variável, sempre há uma variação entre as médias dos testes correspondentes a cada tratamento devido a um erro experimental, que sempre existe. Porém, se essa variação é “grande”, é provável que esse aumento foi devido a esse erro experimental acrescido do efeito dos tratamentos. Quando há apenas uma “pequena” diferença entre as médias dos testes, então não haveria efeito dos tratamentos. Pela ANOVA isto é testado de forma a garantir uma significância estatística para o resultado do teste. A ANOVA define quando que uma diferença entre as médias é suficientemente grande, do ponto de vista estatístico, para que possa ser considerada como efeito dos tratamentos. Permite, inclusive, que se identifique o nível de significância estatístico do impacto de um dado fator, bem como, de suas possíveis interações com outros fatores, na variável de resposta. É, portanto, um procedimento considerado seguro para análise de resultados de experimentos.

Do ponto de vista de procedimento, a ANOVA, apresenta os resultados de um “Teste de Hipótese”, que é um procedimento clássico de inferência estatística (Devore, 2016; Montgomery, 2017) em que é testada uma hipótese nula ( $H_0$ ), que seria a hipótese que se desejaria rejeitar, conforme abaixo:

*H<sub>0</sub>: Médias das Respostas obtidas sob o efeito de cada Tratamento são iguais*

H<sub>0</sub> constituída dessa forma, diz que não há efeito dos tratamentos (dos fatores). Se houvesse efeito, H<sub>0</sub> não seria verdadeira e, portanto, deveria ser rejeitada.

### **3. FASE 1: MODELAGEM E PREDIÇÃO POR *MACHINE LEARNING***

Nesta seção descreve-se a primeira parte da pesquisa, onde com o apoio de técnicas de ML e de uma ferramenta computacional, desenvolveu-se uma modelagem preditiva, capaz de prever níveis de perdas de matéria prima em um processo produtivo, e ainda identificar as variáveis mais relevantes na predição. Essa aplicação se deu em um caso real de uma fabricante de embalagens premium de alumínio, no estado de Ohio, EUA. A aplicação ocorreu em uma linha de produção da empresa, para dois tipos de produtos: latas de bebidas com dimensões de 28mm e 38mm. Modelos preditivos mostraram resultados satisfatórios para predição dos níveis de geração de sucata (*scrap*) no processo de produção, e ainda identificaram os fatores mais relevantes nessa geração de sucata. A modelagem se desenvolveu com quatro tipos diferentes de algoritmos de ML, já discutidos na subseção 2.1: *Decision Tree (DT)*, *Random Forest (RF)*, *Gradient Boosted Tree (GBT)* e *Artificial Neural Network (ANN)*.

A performance dos modelos foi avaliada pelo índice de Acurácia e pelo coeficiente de Kappa, além de outros indicadores complementares. A Acurácia mede o nível de acerto global das predições do modelo, variando entre 0,0 e 1,0 (Theissler et al., 2022). O Coeficiente Kappa descreve o grau de concordância com a realidade na classificação de dados obtida por meio de modelos, pois os resultados de um modelo podem ser provenientes simplesmente do acaso e não da estrutura do modelo propriamente dita. O Kappa procura avaliar se os resultados são efetivamente provenientes do modelo preditivo ou se são obra do acaso. Nesse sentido, uma acurácia, por exemplo, de 90% pode ser enganosa. Trata-se assim, de um indicador fundamental. Landis & Koch (1977) propuseram uma escala numérica para o Kappa, para definir a força da concordância, e essa escala vem sendo utilizada desde então. Nessa escala um coeficiente Kappa acima de 0,6 indica que os resultados obtidos têm uma concordância “substancial”.

Para as latas de 28mm, considerando-se os quatro algoritmos explorados, a Acurácia ficou entre 79% e 99% e o Kappa entre 67,5% e 98,6%. O melhor resultado foi para o algoritmo DT, e o pior foi da ANN. No caso das latas de 38mm, os resultados foram semelhantes, sendo que o melhor, desta vez foi para o GBT, ficando o DT em segundo lugar. Ambos com valores bem altos, acima de 94% para a Acurácia e acima de 88% para o Kappa.

Esses modelos de ML geraram adicionalmente um conjunto “pesos relativos” associados a cada variável considerada no modelo, que medem o nível de importância de cada variável no processo de predição da geração de sucata na produção das latas. Pela análise dos pesos relativos de cada variável, verificou-se que as variáveis *Pessoas*, *Velocidade da linha de produção* e *Tamanho e Espessura e Formato das latas* se mostraram as mais relevantes na predição do nível de geração de sucata. Por outro lado, analisando-se os pesos especificamente no que tange as técnicas DT e RF, foi possível identificar que a variável mais importante foi aquela que se refere às *Datas* de produção de cada lote de latas. Esta variável se explica porque por trás de cada data específica de produção há determinados fatores, que na verdade, são aqueles que efetivamente impactam a geração de sucata. Através de discussões conduzidas diretamente com os gestores responsáveis pelo chão de fábrica e melhoria contínua dos processos, foi identificado que a data de produção está diretamente relacionada à alta rotatividade de colaboradores, sendo este o principal fator por trás das datas. E este é um fato que já vem ocorrendo há algum tempo. A empresa vem passando por importantes desafios na retenção de seus colaboradores, principalmente aqueles que trabalham diretamente na produção das latas. Esse fator impacta de forma importante a geração de sucata, uma vez que o treinamento desses colaboradores não chega a atingir os níveis de qualidade esperados, pois permanecem por pouco tempo na empresa.

Alguns fatores que contribuem para essa alta rotatividade dizem respeito ao mercado de trabalho norte americano, que já há algum tempo mantém um percentual abaixo de 5% de desemprego, e logo, a competitividade natural na busca de profissionais qualificados é alta. Outro ponto, é a localização das instalações da empresa objeto deste estudo; uma região estratégica com grande número de indústrias, onde a procura por profissionais experientes que trabalham no chão de fábrica é alta. A combinação desses fatores contribui para uma

instabilidade e crescente *turnover* dos colaboradores desde o ano de 2019. Isso de fato dificulta a consistência do conhecimento para a fabricação de latas, pois se tratando de um processo contínuo e repetitivo que opera em 24x7 (vinte e quatro horas por sete dias na semana), o aprendizado é fortalecido pelo treinamento e prática contínua, consolidado pelos anos de experiência adquiridos por cada colaborador.

Seguindo a análise das variáveis mais importantes, verificou-se que as demais variáveis importantes são: *Line Speed – Actual*, correspondente a velocidade de cada linha produtiva, *FG\_Class*, que define cada classe e tipo de lata fabricada, *Crew 4*, definido pelo grupo de turno de pessoas que trabalham nas linhas de produção, *Time 2*, referente ao horário exato das produções das latas, *Slug 3*, referente ao tipo de matéria prima utilizada. Estas são todas variáveis controláveis, e, portanto, passíveis de ajustes para melhoria do processo produtivo.

Portanto, como síntese da fase 1 da pesquisa, identificou-se que pessoas e seus respectivos turnos de trabalho, bem como, a velocidade da linha de produção correspondem às principais variáveis na predição da geração de sucata. Na Fase 2 do estudo, com base nas variáveis identificadas, foi realizada uma intervenção em uma das linhas de produção da empresa com um experimento de campo, planejado segundo os princípios de um DOE, a para a análise dos resultados foi utilizada a ferramenta estatística ANOVA, com o apoio do software R-Studio.

#### **4. FASE 2: METODOLOGIA APLICADA AO EXPERIMENTO DE CAMPO**

Esta seção e a próxima apresentam uma aplicação em um caso real, visando validar os resultados dos modelos preditivos de ML. O presente estudo propôs uma abordagem para as ações em campo baseada em uma mescla dos tipos de pesquisa apresentados na subseção 2.2, estabelecendo uma sequência de etapas, que são descritas a seguir:

**Etapa 1 - Diagnóstico:** Nesta etapa identificou-se o problema prático, que é relevante do ponto de vista de pesquisa;

**Etapa 2 - Pré-intervenção:** Aqui obteve-se junto à organização objeto do estudo, uma compreensão abrangente do problema;

**Etapa 3 - Intervenção:** Nesta etapa teve-se a intervenção no sistema em estudo, que é subdividida em alguns passos;

**Etapa 4: Experimento de Campo:** Esta etapa tratou do experimento de campo propriamente dito, envolvendo sua abordagem técnica.

**Etapa 5: Avaliação de Resultados:** Esta foi a etapa final em que se avaliou os resultados do experimento

O detalhamento das etapas 1 a 3 é apresentado na sequência desta seção 4. E na seção 5 são apresentadas as etapas 4 e 5, que requerem um detalhamento maior.

#### 4.1 Etapa 1: Diagnóstico

Esta etapa trata da identificação de um problema prático importante para a organização e relevante para pesquisa, foi subdividida em duas fases:

*a) Chamada para colaboração:* neste caso, um dos pesquisadores se envolveu com o projeto por meio de um projeto de tese de doutorado na área e do vínculo empregatício com a organização de acolhimento . O pesquisador conversou com os seus dois principais líderes diretos; diretor presidente da subsidiária objeto do estudo e vice-presidente de finanças da região das Américas da empresa anfitriã, sugerindo o desenvolvimento de um projeto conjunto, que pudesse ajudar a organização e colaborar com os objetivos acadêmicos do pesquisador. Assim, uma das principais dificuldades da pesquisa intervencionista foi superada, uma vez que o acesso a organizações que se dispõem a participar de estudos acadêmicos é uma condição limitante para o desenvolvimento da pesquisa intervencionista. (Roberts, Westin e Dumay, 2010; Lukka e Vinnari, 2017; Suomala, Lyly-Yrjänäinen e Lukka, 2014; Suomala, Lyly-Yrjänäinen e Lukka, 2014).

*b) Diagnóstico preliminar:* aqui verificou-se que havia uma questão relevante associada à geração de sucata (*scrap*), e em um primeiro momento, avaliou-se como sendo excessiva, gerando perda de produtividade, elevando custos e causando problemas ambientais. Assim, avaliando-se preliminarmente o problema entendeu-se que havia ali uma



**oportunidade para ganhos de produtividade e que merecia uma possível intervenção nos processos.**

#### **4.2 Etapa 2: Pré-intervenção**

Nesta etapa, o que se busca é obter junto à organização objeto do estudo, uma compreensão abrangente e geral do problema a ser tratado. Nesse sentido, chegou-se a uma melhor especificação do problema da empresa, identificando-se que a questão se concentrava na quantidade de sucata gerada na produção contínua no processo da prensa de extrusão, e devia-se, portanto, desenvolver um estudo para identificar as causas da geração de sucata naquele processo. Preliminarmente, já foram identificados alguns fatores, e posteriormente, no processo de intervenção, parte deles foi confirmada e outros descartados.

De forma mais específica, buscava-se respostas para as seguintes questões:

- a) Quais seriam as variáveis passíveis de observação/mensuração no processo da prensa de extrusão que pudessem influenciar na geração de sucata naquele processo?
- b) Quais seriam as técnicas/modelos que pudessem ser utilizados em uma intervenção no processo da prensa de extrusão, e que permitissem confirmar em campo as variáveis significativas e controláveis que pudessem propiciar uma redução da sucata gerada?

#### **4.3 Etapa 3: Intervenção**

Esta etapa tem algumas subdivisões, expostas na sequência.

##### **4.3.1 Tripé da Intervenção**

O que foi possível aprender nesse processo de intervenção é que a fundação de todo o problema está relacionada à **Empresa, Pessoas e Processos**. Sendo esse último desmembrado em duas grandes áreas: linha de produção objeto desta intervenção e seus próprios processos.

##### **4.3.2 A Empresa**

Sobre a empresa objeto desta intervenção, esta pode ser caracterizada como uma organização internacional presente em mais de 60 locais em todo o mundo. É fabricante de

produtos premium em alumínio. conta com cerca de 7.500 colaboradores e tem uma receita anual de aproximadamente US\$ 2,8 bilhões.

#### 4.3.3 Subsidiária escolhida para a Intervenção

Inicialmente, foi definido que a região escolhida para realizar a intervenção deveria se localizar nas Américas. Existem aproximadamente dez subsidiárias e plantas produtivas nas Américas, incluindo Brasil, Canadá e os Estados Unidos. O pesquisador vinculado à empresa, atua nos Estados Unidos, assim, o projeto teve seu foco direcionado para uma das oito subsidiárias existentes naquele país. A subsidiária escolhida, atua em quatro grandes negócios e a representatividade de sucata gerada mensalmente é demonstrada na tabela 1 abaixo:

**Tabela 1**  
Negócios da Subsidiária Selecionada e % de *Scrap* do Processo Produtivo

Negócio	% de sucata gerada
<i>Aerosol</i>	10%
<i>Beverage</i>	16%
<i>Cheese</i>	7%
<i>Specialties</i>	15%

Fonte: Elaborado pelos autores

#### 4.3.4 Definição da área operacional e seus Processos

Em termos de área operacional foi definida uma Linha de Produção e um de seus processos. O *plant manager* da empresa apresentou as linhas produtivas, bem como os negócios que são gerados naquela fábrica. Atualmente, existem 16 linhas produtivas que atendem os seguintes negócios: *Aerosol*, *Beverage*, *Cheese* e *Specialties*. A Linha 14 escolhida para a intervenção, produz as latas para o segmento de *Beverage*, que atende basicamente dois principais clientes e segmentos (consumo de energéticos) e *Burgundy Wine* (vinho de Borgonha), conforme Figura 2 a seguir:

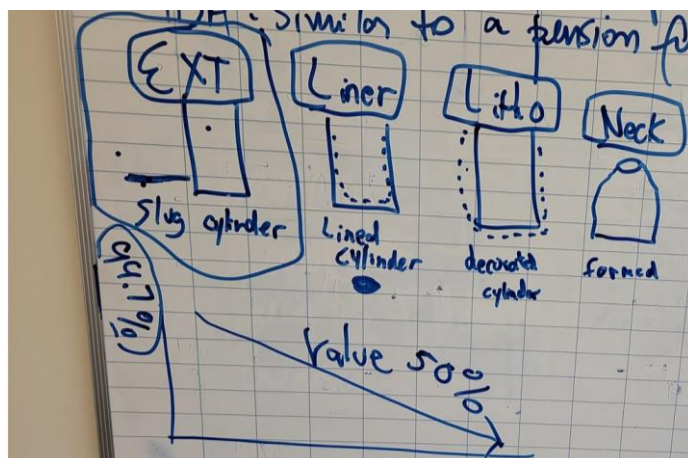


**Figura 2:** Produtos objeto desta análise de intervenção  
Fonte: Elaborado pelos autores

São basicamente quatro os processos responsáveis pela geração total de sucata, sendo eles:

- *Extrusion*
- *Liner*
- *Lithografy*
- *Necker*

O processo de impacto por extrusão representa acima de 90% de toda a sucata gerada no processo produtivo de latas de *Beverage*, o que reforça que o alumínio tem o maior valor agregado por ser uma *commodity* de natureza pura. Após o processo de extrusão, o valor agregado de *scrap* gerado é reduzido em mais de 50%. Isso se explica pela foto 2 abaixo, gerada durante reuniões com o time da qualidade da empresa, em que se desenvolveu um processo de abdução, o qual em síntese consiste na adoção provisória de hipóteses que deverão ser confirmadas pelo experimento de campo.



**Figura 3:** Processo de geração de scrap de alumínio  
Fonte: Elaborado pelos autores

Desta forma, o foco da intervenção foi todo voltado para a linha 14, no processo da prensa por extrusão.

## 5. FASE 2: EXPERIMENTO DE CAMPO E RESULTADOS

Esta seção apresenta as etapas 4 e 5 da metodologia, correspondentes ao desenvolvimento do experimento de campo e à avaliação de seus resultados. Apresenta-se aqui um maior detalhamento de todos os procedimentos dessas duas etapas.

### Etapa 4: Experimento de Campo

Esta etapa foi subdividida em quatro passos, descritos na sequência.

#### Passo 4.1: Definições de Datas para o Experimento

Durante o primeiro trimestre de 2023, foram conduzidas diversas reuniões com o time multifuncional de operações da fábrica com o objetivo de se encontrar o melhor momento para realizar a intervenção. O objetivo era de trazer o mínimo (ou nenhum) impacto à rotina do time da produção. A terceira semana do mês de abril foi definida como o melhor momento para realizar o experimento de campo, e exatamente como definido com o time multifuncional de operações, na terceira semana de abril de 2023 foi realizado o experimento de campo na linha 14.

### **Passo 4.2: Componentes do Modelo Conceitual para o Experimento**

Como base de análise foi proposto um modelo conceitual em que se mapeia o processo em que será desenvolvido o experimento identificando-se seus componentes a considerar no experimento. Assim, utilizou-se a mesma linha de produção que serviu de base para a construção dos modelos de ML (Linha 14), e verificou-se que os componentes a considerar seriam: uma máquina extrusora, uma máquina de corte, quatro equipes de extrusão, dois turnos diurnos e dois turnos noturnos. Este modelo foi validado na linha de produção, viabilizando sua utilização para este trabalho.

### **Passo 4.3: Estrutura do Experimento de Campo**

O experimento foi planejado de forma a avaliar os fatores identificados nos modelos de ML. A coleta dos dados se deu na produção contínua de latas, e foi realizada por um período de 48 horas, 2 dias seguidos, 25/04 e 26/04/2023, separadas por um período de 1 hora, subdivididos em 4 diferentes turnos de produção: A1, A2, B1 e B2. Foram coletados aproximadamente 300.000 registros, correspondentes às produções de latas, com a correspondente geração de sucata. Ao final de cada hora, o operador de produção ajustou a velocidade manualmente no painel eletrônico do Controlador Logico Programável (CLP) da máquina. Esse processo manual foi acompanhado pelo pesquisador de campo. Foram ajustadas 4 diferentes velocidades no CLP da máquina durante o experimento: 125, 127, 129 e 131. Esse processo foi replicado por 2 vezes durante os turnos A e B. A Tabela 2 demonstra a estrutura do experimento de campo.

Tabela 2  
Estrutura do Experimento de Campo

<b>Dia</b>	<b>Ação</b>	<b>Equipe</b>	<b>Turno</b>	<b>Ajuste de Velocidades no CLP da máquina</b>
25/04	Experimento e Réplica	B1	Diurno	125, 127, 129, 131
25/04	Experimento e Réplica	B2	Noturno	125, 127, 129, 131
26/04	Experimento e Réplica	A1	Manhã	125, 127, 129, 131
26/04	Experimento e Réplica	A2	Noturno	125, 127, 129, 131

### **Passo 4.4: Execução do Experimento**

O experimento foi planejado segundo o projeto Fatorial Completo com replicações, onde se avaliam todos os fatores e as interações entre eles (Montgomery, 2017; Kleinert et al., 2019). Foram analisados três fatores identificados nos modelos de ML (Velocidade da Linha, Pessoas e Turno de Trabalho), coletando-se o valor da variável de resposta (geração de sucata). Na Tabela 3 têm-se os resultados obtidos (matriz experimental), sendo apresentado o percentual de sucata (*scrap*) gerada em cada uma das condições testadas.

**Tabela 3**

Matriz Experimental e Percentual de Sucata (*scrap*) Gerada

Velocidade	Turno Matutino		Turno Noturno	
	Equipe A1	Equipe A2	Equipe B1	Equipe B2
125	2.23%	2.33%	2.23%	2.30%
127	2.21%	2.33%	2.41%	2.27%
129	2.33%	2.32%	2.23%	2.23%
131	2.24%	2.32%	2.22%	2.30%

## 5.2 Etapa 5: Análise de Resultados do Experimento

A análise de resultados foi conduzida com base na Análise de Variância (ANOVA), com o apoio da ferramenta computacional, RStudio.

A tabela 4 representa os resultados de uma primeira análise, considerando-se apenas o fator Velocidade da Linha. Esta é a chamada “one-way ANOVA”, análise a um fator (tratamento). A coluna “Pr(>F)” na tabela 4, representa a probabilidade de  $H_0$  ser verdadeira, que é determinada a partir de *F-Value*. Quanto maior *F-Value*, menor será “Pr(>F)”, e maior será a probabilidade de efeito dos Tratamentos. Os demais dados da ANOVA representam indicadores e parâmetros intermediários necessários para o cômputo de *F-Value*.

Como padrão, os testes estatísticos consideram que se rejeita  $H_0$  para  $\text{Pr}(>F) < 0,005$ . Portanto, 0,005 ou 5%, é o valor crítico de  $\text{Pr}(>F)$  para rejeição de  $H_0$ . Para valores abaixo de 0,005, entende-se que há impacto significativo do tratamento.

**Tabela 4**

Tabela ANOVA a um Fator (one-way)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Veloc	1	9.60e-06	9.648e-06	2.208	0.139
Residuals	238	1.04e-03	4.370e-06		

Fonte: Elaborado pelos autores

Na Tabela 4, tem-se  $Pr(>F) = 0.139$ . Assim, quando analisada isoladamente, a Velocidade não apresenta efeito na geração de sucata, ao nível de significância estatístico de 5%, pois  $Pr(>F) > 0,005$ .

Uma nova análise foi realizada, “two-way ANOVA”, desta vez considerando-se dois fatores como variáveis independentes: Velocidade e Equipes (Tabela 5).

**Tabela 5**

Tabela ANOVA a dois Fatores (two-way)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Veloc	1	0.0000096	9.648e-06	2.250	0.135
Equipe	3	0.0000322	1.073e-05	2.503	0.060 .
Residuals	235	0.0010079	4.289e-06		
---					
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					

Fonte: Elaborado pelos autores

Note-se que agora,  $Pr(>F)$  da Velocidade variou para 0.135, e o resultado de  $Pr(>F)$  do fator Equipe foi de 0.06, demonstrando que este fator já se aproxima bem do valor crítico de 0,05.

Outra análise a dois fatores foi desenvolvida, mas desta feita considerando-se a interação entre Velocidade e Equipes (tabela 6).

**Tabela 6**

Tabela ANOVA a dois Fatores com Interação entre Fatores

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Veloc	1	0.0000096	9.648e-06	2.299	0.1308
Equipe	3	0.0000322	1.073e-05	2.557	0.0559 .
Veloc:Equipe	3	0.0000341	1.137e-05	2.708	0.0460 *
Residuals	232	0.0009738	4.197e-06		
---					
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					

Fonte: Elaborado pelos autores

Agora, a interação entre Velocidade e Equipe se mostrou estatisticamente significativa ao nível de 5%, e o fator Equipe se aproximou mais ainda da significância crítica de 5%.

Uma quarta análise foi realizada, com três fatores, “*three-way* ANOVA”, desta vez considerando-se os fatores Turno, Velocidade e Equipes. Os resultados são apresentados na Tabela 7.

**Tabela 7**  
Tabela ANOVA a três Fatores (*three-way*)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Turno	1	0.0000229	2.294e-05	5.349	0.0216 *
Veloc	1	0.0000096	9.648e-06	2.250	0.1350
Equipe	2	0.0000093	4.630e-06	1.080	0.3414
Residuals	235	0.0010079	4.289e-06		

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Fonte: Elaborado pelos autores

Vê-se agora melhora significativa no modelo, com o valor de Pr(>F) do fator Turno, que ficou em 0.0216, mostrando que é significativa ao nível de 5%. O fator Velocidade manteve-se no nível, um pouco acima de 10% de significância, e o fator Equipe aparentemente foi substituído pelo turno de trabalho, onde equipes diferentes atuam.

Finalmente, a última análise desenvolvida considerou 3 fatores, mas desta feita considerando-se a interação entre Turno, Velocidade e Equipes (tabela 8).

**Tabela 8**  
Tabela ANOVA a três Fatores com Interação entre Fatores

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Turno	1	0.0000229	2.294e-05	5.466	0.0202 *
Veloc	1	0.0000096	9.648e-06	2.299	0.1308
Equipe	2	0.0000093	4.630e-06	1.103	0.3335
Veloc:Equipe	3	0.0000341	1.137e-05	2.708	0.0460 *
Residuals	230	0.0010079	4.383e-06		

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

E note-se que foi identificada pelo modelo a interação entre Velocidade e Equipe, e que novamente se mostrou estatisticamente significativa ao nível de 5%, e o fator Turno também se mostrou estatisticamente significativa ao nível crítico de 5%. Tem-se assim, impacto dos três fatores, de forma isolada ou por interação entre eles.

Como observação final, saliente-se que esses fatores, Velocidade, Equipe e Turno, que se mostraram relevantes, apareceram também com relevância nos modelos de ML, o que mostra



que os resultados do experimento de campo levaram à validação de achados dos modelos de ML.

## 6. CONCLUSÕES E RECOMENDAÇÕES

Conforme já salientado, na fase 1 deste estudo construiu-se modelos de ML, e foram identificadas as variáveis velocidade da máquina, pessoas e seus respectivos turnos de trabalho, como sendo as mais importantes na predição de geração de sucata no processo de produção de latas de alumínio. Na fase 2 do estudo, essas variáveis foram testadas e validadas na pesquisa intervencionista, aqui descrita, através de um experimento de campo em uma das linhas produtivas da empresa objeto do estudo. O processo produtivo de dois tipos de latas, de 28mm e 38mm, no chão de fábrica, foi acompanhado durante um período contínuo de 48 horas, na terceira semana do mês de abril/2023.

O planejamento do experimento de campo foi baseado em DOE, tendo sido utilizado um projeto fatorial completo na análise dos fatores. A ANOVA utilizada para análise dos resultados, confirmou os fatores identificados nos modelos de ML, como sendo os mais relevantes em termos de impacto na variável de resposta do experimento. Evidenciando, portanto, que o estudo inicial por modelos de ML foi capaz de identificar os fatores relevantes na geração de *scrap*, e validando, assim, aqueles modelos de ML.

Como continuidade deste trabalho, recomenda-se que em termos práticos, com base no experimento realizado, o processo de produção de latas de bebidas com 28mm e 38mm na empresa seja ajustado. E que, no caso das variáveis associadas a pessoas (representadas pelas equipes e turnos), a empresa incremente sua atuação na questão da rotatividade e retenção de seus colaboradores, conforme discutido anteriormente neste artigo. Do ponto de vista científico, por se mostraram promissores, a modelagem por ML e os tipos de experimentos desenvolvidos poderiam ser expandidos para outras linhas de produção na própria empresa, bem como para outras empresas que estejam passando por desafios similares, de forma a aprimorar tais modelos e os experimentos, bem como, eventualmente, identificar outros algoritmos que pudessem trazer resultados relevantes.

## REFERÊNCIAS

- Barra Montevechi, J. A., Carvalho Miranda, R. de, & Daniel, J. (2012). Sensitivity Analysis in Discrete Event Simulation Using Design of Experiments. In: Discrete Event Simulations - Development and Applications. Edited by Eldin Wee Chuan Lim. <https://doi.org/10.5772/50196>
- Breiman, L. (2001). Random Forests. *Machine Learning* 45, 5–32 <https://doi.org/10.1023/A:1010933404324>
- Cutler, A.; Cutler, D.R.; Stevens, J.R. (2012). Random Forests. In Ensemble Machine Learning; Zhang, C., Ma, Y.Q., Eds.; Springer: New York, NY, USA, 2012; pp. 157–175. [http://dx.doi.org/10.1007/978-1-4419-9326-7\\_5](http://dx.doi.org/10.1007/978-1-4419-9326-7_5)
- Desroches, D. e Raef Lawson, R. (2013). Evolving Role of the Controller. IMA - The Association of Accountants and Financial Professionals in Business. Institute of Management Accountants. Montvale, NJ. <https://www.admaiorasemper.com/en/page7/files/RoleOfController2014.pdf>
- Devore, J. L. (2016). Probability and Statistics for Engineering and the Sciences, 9th Edition. Cengage.
- Engel, G. I. (2000). Pesquisa-ação. *Educar Em Revista*, 16. <https://doi.org/10.1590/0104-4060.214>
- Han, J.; Kamber, M.; Pei, J. (2012). Data Mining: Concepts & Techniques. 3rd Edition. Waltham, MA. Morgan Kaufmann Series in Data management Systems, 744 p
- Hastie, T., Tibshirani, R. e Friedman, J. H. (2009). The elements of statistical learning: Data mining, inference, and prediction. Springer series in statistics. Springer, New York, 2nd edition.
- Haykin, S. (2008) Neural Networks and Learning Machines, 3rd Edition. Prentice-Hall.
- Johnson, R. T., & Montgomery, D. C. (2010). Designing experiments for nonlinear models - An introduction. *Quality and Reliability Engineering International*, 26(5). <https://doi.org/10.1002/qre.1063>
- Kaplan, R. S. (1984). The evolution of management accounting. In Readings in Accounting for Management Control. [https://doi.org/10.1007/978-1-4899-7138-8\\_27](https://doi.org/10.1007/978-1-4899-7138-8_27)
- Kaplan, R. S. (2011). Accounting scholarship that advances professional knowledge and practice. In *Accounting Review* (Vol. 86, Issue 2). <https://doi.org/10.2308/accr.00000031>
- Kasanen, E., & Lukka, K. (1993a). The constructive approach in management accounting research. *Journal of Management Accounting Research*, 5(5).
- Kasanen, E., & Lukka, K. (1993b). The constructive approach in management accounting research. *Journal of Management Accounting Research*, 5(5).
- Kleinert, T. R., Fedrigo, W., ten Caten, C. S., Núñez, W. P., & Ceratti, J. A. P. (2019). Aplicação da metodologia de projeto de experimentos na dosagem de misturas de reciclagem de pavimentos com adição de cimento. *Transportes*, 27(1). <https://doi.org/10.14295/transportes.v27i1.1574>

- Krauss, C.; Do, X. A.; Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*. Vol 259, Issue 2, 1 June 2017, p. 689-702
- Labro, E., & Tuomela, T. S. (2003). On bringing more action into management accounting research: process considerations based on two constructive case studies. *European Accounting Review*, 12(3). <https://doi.org/10.1080/0963818032000083559>
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1). <https://doi.org/10.2307/2529310>
- Laureani, A., & Antony, J. (2019). Leadership and Lean Six Sigma: a systematic literature review. In *Total Quality Management and Business Excellence* (Vol. 30, Issues 1–2). <https://doi.org/10.1080/14783363.2017.1288565>
- Lawson, J. (2015). *Design and Analysis of Experiments with R*. CRC Press. 593 p.
- Lukka, K., & Vinnari, E. (2017). Combining actor-network theory with interventionist research: present state and future potential. *Accounting, Auditing and Accountability Journal*, 30(3). <https://doi.org/10.1108/AAAJ-08-2015-2176>
- Montgomery, D. C. (2017). *Design & Analysis of Experiments*. 9<sup>th</sup> edition. John Wiley & Sons. 734 p.
- Natekin, A.; Knoll, A. (2013). Gradient boosting machines, a tutorial. *Front. Neurobot.*, Vol, 7, 2013. <https://doi.org/10.3389/fnbot.2013.00021>
- Oyadomari, J. C.T, Cardoso, R.L, Mendonça Neto, O.R, & Aguiar, A. B. (2013). Criação de conhecimento em práticas de controle gerencial: análise dos estudos internacionais. *Advances in Scientific and Applied Accounting*. <https://doi.org/10.14392/asaa/2013060101>
- Roberts, H., & Baard, V. (2010). A critical review of interventionist research. In *Qualitative Research in Accounting & Management* (Vol. 7, Issue 1). <https://doi.org/10.1108/11766091011034262>
- Jönsson, S. (2010). Interventionism – an approach for the future? *Qualitative Research in Accounting & Management*, 7(1). <https://doi.org/10.1108/11766091011034307>
- Roberts, H., Westin, O., & Dumay, J. C. (2010). A critical reflective discourse of an interventionist research project. *Qualitative Research in Accounting & Management*, 7(1). <https://doi.org/10.1108/11766091011034271>
- Russel, S.J.; Norvig, P. (2020). *Artificial Intelligence: A Modern App*. 4th Edition. Pearson Series in Artificial Intelligence, 1136p.
- Sanchez, S. M., Moeeni, F., & Sanchez, P. J. (2006). So many factors, so little time...Simulation experiments in the frequency domain. *International Journal of Production Economics*, 103(1). <https://doi.org/10.1016/j.ijpe.2005.06.007>
- Schapire, R. E. (1990). *The Strength of Weak Learnability*. Machine Learning, 5, 197-227. Kluwer Academic Publishers, Boston
- Suomala, P., & Lyly-Yrjänäinen, J. (2010). Interventionist management accounting research: lessons learned. *CIMA, Research Executive Summaries Series*, 6(1).
- Suomala, P., & Lyly-Yrjänäinen, J. (2012). *Management Accounting Research in Practice*. In *Management Accounting Research in Practice*. <https://doi.org/10.4324/9780203141205>

- Suomala, P., Lyly-Yrjänäinen, J., & Lukka, K. (2014). Battlefield around interventions: A reflective analysis of conducting interventionist research in management accounting. *Management Accounting Research*, 25(4). <https://doi.org/10.1016/j.mar.2014.05.001>
- Tang, L. C., Goh, T. N., Yam, H. S., & Yoap, T. (2006). Six Sigma: Advanced Tools for Black Belts and Master Black Belts. In *Six Sigma: Advanced Tools for Black Belts and Master Black Belts*. <https://doi.org/10.1002/0470062002>
- Theissler, A.; Thomas, M.; Burch, M.; Gerschner, F. (2022). ConfusionVis: Comparative evaluation and selection of multi-class classifiers based on confusion matrices. *Knowledge-Based Systems*, Vol. 247. <https://doi.org/10.1016/j.knosys.2022.108651>
- Tripp, D. (2005). Pesquisa-ação: uma introdução metodológica. *Educação e Pesquisa*, 31(3). <https://doi.org/10.1590/s1517-97022005000300009>