

**ANTÔNIO VIEIRA CAETANO DE MATTOS  
CAIO JARDIM SCARPA COSTA**

Modelo de previsão de sucesso de produtos lançamento:  
um estudo de caso da CienaLab

PROJETO DE GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO  
APRESENTADO AO DEPARTAMENTO DE ENGENHARIA INDUSTRIAL  
DA PUC-RIO, COMO PARTE DOS REQUISITOS PARA OBTENÇÃO  
DO TÍTULO DE ENGENHEIRO DE PRODUÇÃO

Orientador: Rodrigo Goyannes Gusmão Caiado  
Coorientador: Renan Silva Santos

Departamento de Engenharia Industrial  
Rio de Janeiro, 24 de novembro de 2025.

## **AGRADECIMENTOS**

À minha mãe, Jeane, que deu tudo de si para que eu pudesse ser quem sou hoje, sem ela eu nada seria. Me formar é uma forma de retribuir em forma de orgulho o amor, carinho e dedicação disponibilizados por ela durante todos esses anos.

Ao meu padrasto, Cleber, por ser um espelho de profissional que eu busco ser, referência para todos e que durante boa parte da minha vida me teve como filho.

À minha namorada, Adrielle, que foi minha companheira, conselheira e ouvinte, que me deu abrigo sempre que precisei e que me apoiou nos momentos cruciais dessa jornada.

Aos meus amigos, em especial Caio Sanches, que sempre esteve comigo em todos os momentos da vida, neste não seria diferente. Ao Caio Scarpa, fiel escudeiro dessa jornada na PUC-Rio, e também João Pedro Araújo e Matheus Sousa, que me ajudaram muito. E à todos meus colegas de trabalho, que sempre me deram apoio e forças, em especial meu maior incentivador dentre eles, Salvio Formisano.

**Antônio Mattos**

Aos meus pais, cujos pés foram impedidos de correr e deram a vida aos filhos com asas para voar. Que compreenderam a minha ausência, acolheram as minhas escolhas e não mediram esforços para que este momento se tornasse possível.

À minha irmã, que sempre foi a minha referência, o ponto mais alto da régua, o norte seguro, a prova viva de onde eu poderia chegar.

Aos meus amigos da faculdade, que tornaram o caminho possível, sobretudo nos períodos mais desafiadores da graduação para um bolsista. Em especial, agradeço a José Chaves, Eduarda Santis, Breno Elehep, Júlia Monnerat, Amanda Hilbert, Antonio Mattos, Lucas Barros e Pedro Saloio, por nunca me deixarem caminhar sozinho.

À minha avó, que, mesmo enfrentando dificuldades que eu talvez nunca compreenda por completo, sempre me apontou a educação como o único caminho capaz de transformar a realidade. Formar-me engenheiro sempre foi um sonho nosso; gostaria que estivesse aqui hoje para ver esta conquista, que também é dela.

**Caio Scarpa**

## RESUMO

A consolidação do *streetwear* como um dos segmentos mais relevantes da moda contemporânea, especialmente entre consumidores pertencentes a Geração Z, intensificou os desafios na sua previsão de demanda, em um contexto marcado por lançamentos em *drops*, com ciclos de vida curtos e com uma forte influência das redes sociais. Nesse contexto, marcas que operam com coleções enxutas e estoque limitado enfrentam uma elevada incerteza quanto ao desempenho de suas peças, impactando diretamente o planejamento da produção e o resultado financeiro das coleções lançadas. Este trabalho tem como objetivo desenvolver e avaliar um modelo preditivo de sucesso de produtos lançamento no contexto da marca carioca de *streetwear* CienaLab, com foco apenas na categoria de produtos de camisetas. A pesquisa adota uma abordagem aplicada, quantitativa e preditiva, estruturada como estudo de caso. Foi utilizado um banco de dados histórico de coleções anteriores da marca, no qual cada camiseta é descrita por atributos visuais e fabris, como a sua cor, seus materiais e seus tipos e tamanhos de estampa aplicados e por indicadores de desempenho de vendas em duas janelas, sendo a primeira após os primeiros 15 dias de lançamento da coleção e a segunda no total do ciclo de vendas. A partir desses indicadores é construída uma variável categórica de desempenho, que classifica as camisetas em categorias como *best performer*, *hyped* e *normal*. Sobre essa base analisada, o problema é formulado como uma tarefa de classificação supervisionada multiclasse. Com diferentes algoritmos testados e comparados, por fim selecionando o modelo *Random Forest*, ajustado por validação cruzada estratificada e busca em grade de hiper parâmetros. A interpretação do modelo escolhido é realizada com o uso de valores SHAP, permitindo analisar a contribuição dos atributos visuais para as previsões de demanda. Adicionalmente, é implementado um procedimento de mapeamento entre coordenadas de cor em CIELAB ( $L^*a^*b^*$ ) e as categorias de cor utilizadas pela marca, viabilizando o uso do modelo no processo de criação de peças para o uso comum. Por fim, o classificador é integrado a uma interface desenvolvida em Python e validado em uma coleção recente da CienaLab, discutindo-se seu potencial como ferramenta de apoio ao planejamento de produtos em ambientes de alta volatilidade de demanda.

**Palavras-chave:** *streetwear*; previsão de demanda; aprendizado de máquina; *Random Forest*; moda.

# Product launch success prediction model: a case study of CienaLab

## ABSTRACT

The consolidation of *streetwear* as one of the most relevant segments in contemporary fashion, especially among Generation Z consumers, has intensified classic demand forecasting challenges in a context marked by short product life cycles, limited-stock *drops* and strong influence from digital media. In this scenario, brands that operate with small collections and planned scarcity face high uncertainty regarding the performance of individual items, which directly affects production planning and the financial results of each collection. This study aims to develop and evaluate a predictive model for the success of new product launches in the context of the Brazilian *streetwear* brand CienaLab, focusing on the T-shirt category. The research adopts an applied, quantitative and predictive approach, structured as a single case study. A historical database of past collections is used, in which each T-shirt is described by visual and production-related attributes (color, materials, types and sizes of prints) and by sales performance indicators in two time windows: the item's share of collection sales in the first 15 days and in the total sales of the collection. Based on these indicators, a categorical performance variable is constructed, classifying items into groups such as best performer, hyped and normal. The problem is formulated as a multiclass supervised classification task. Different algorithms are tested and compared, and a Random Forest model is selected and tuned using stratified cross-validation and grid search for hyperparameters. Model interpretation is conducted with SHAP values, allowing the assessment of the relative contribution of visual attributes to the predictions. In addition, a mapping procedure between CIELAB (L\*a\*b\*) color coordinates and the brand's discrete color categories is implemented, enabling use of the model during the product design process. Finally, the classifier is integrated into a Python-based graphical interface and validated on a recent CienaLab collection, discussing its potential as a decision-support tool for product planning in highly volatile *streetwear* environments.

**Keywords:** *streetwear*; demand forecasting; machine learning; Random Forest; fashion.

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>8</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO.....</b>	<b>11</b>
<b>2.1</b>	<b>Moda <i>streetwear</i>, <i>drops</i> e consumo da Geração Z .....</b>	<b>11</b>
2.1.1	<i>Drops</i> , escassez e ciclos curtos .....	12
<b>2.2</b>	<b>Previsão de demanda na indústria da moda.....</b>	<b>133</b>
<b>2.3</b>	<b>Atributos visuais e identidade de produto na decisão de compra .....</b>	<b>13</b>
<b>2.4</b>	<b>Aprendizado supervisionado e modelos de classificação na moda .....</b>	<b>14</b>
<b>2.5</b>	<b>Modelos baseados em árvore e <i>Random Forest</i> .....</b>	<b>16</b>
<b>2.6</b>	<b>Representação de variáveis categóricas em aprendizado de máquina.....</b>	<b>17</b>
<b>2.7</b>	<b>Seleção de modelos, validação e ajuste de hiper parâmetros .....</b>	<b>19</b>
<b>2.8</b>	<b>Explicabilidade em modelos preditivos: XAI e SHAP na moda .....</b>	<b>20</b>
<b>2.9</b>	<b>Representação métrica da cor em modelos preditivos: o uso do espaço CIELAB .....</b>	<b>21</b>
<b>3</b>	<b>METODOLOGIA .....</b>	<b>23</b>
<b>3.1</b>	<b>Tipo de pesquisa, dados e variáveis .....</b>	<b>23</b>
<b>3.2</b>	<b>Estudo de caso: CienaLab .....</b>	<b>24</b>
3.2.1	CienaLab .....	24
3.2.2	Modelo de negócios .....	25
3.2.3	Estrutura operacional e produtiva .....	26
3.2.4	Vantagens competitivas e riscos .....	26
3.2.5	Relevância do caso para o estudo.....	27
<b>3.3</b>	<b>Preparação dos dados e divisão treino-teste .....</b>	<b>28</b>
<b>3.4</b>	<b>Construção e seleção do modelo preditivo .....</b>	<b>29</b>
<b>3.5</b>	<b>Interpretação do modelo (XAI) e variáveis de cor .....</b>	<b>30</b>
<b>3.6</b>	<b>Implementação da ferramenta e validação em nova coleção .....</b>	<b>31</b>
<b>4</b>	<b>RESULTADOS E DISCUSSÃO .....</b>	<b>34</b>
<b>4.1</b>	<b>Caracterização da base e da variável de desempenho .....</b>	<b>34</b>
<b>4.2</b>	<b>Resultado dos modelos avaliados .....</b>	<b>35</b>
<b>4.3</b>	<b>Modelo final e seu desempenho.....</b>	<b>36</b>
4.3.1	Análise das métricas de classificação .....	37
4.3.2	Importância das variáveis.....	38

4.3.3	Análise de explicabilidade com SHAP .....	39
<b>4.4</b>	<b>Teste do modelo em uma coleção lançada posteriormente .....</b>	<b>40</b>
<b>4.5</b>	<b>Implementação prática: interface CustomTkinter .....</b>	<b>41</b>
4.5.1	Tratamento de cor pelo espaço LAB.....	41
4.5.2	Travamentos de seleção .....	42
4.5.3	Previsão individualizada e cadastro de coleção .....	43
<b>4.6</b>	<b>Implicações gerenciais e operacionais .....</b>	<b>45</b>
<b>5</b>	<b>CONCLUSÕES .....</b>	<b>46</b>
<b>5.1</b>	<b>Atendimento dos objetivos .....</b>	<b>47</b>
<b>5.2</b>	<b>Contribuições do estudo .....</b>	<b>48</b>
<b>5.3</b>	<b>Limitações do estudo.....</b>	<b>49</b>
	<b>REFERÊNCIAS .....</b>	<b>51</b>

## LISTA DE ILUSTRAÇÕES

Figura 1 – Princípio do aprendizado supervisionado .....	15
Figura 2 – Processos de <i>machine learning</i> na indústria da moda.....	17
Figura 3 – Comparativo visual entre os modelos avaliados no Grid Search CV .....	36
Figura 4 – Matriz de confusão do modelo final.....	37
Figura 5 – Importância das variáveis visuais no modelo.....	38
Figura 6 – Importância das variáveis na saída do modelo (via SHAP).....	39
Figura 7 – Interface inicial da aplicação.....	43
Figura 8 – Resultado de previsão individual .....	44
Figura 9 – Resumo da coleção atual .....	44
Quadro 1 – Variáveis utilizadas no modelo preditivo .....	24
Quadro 2 – Mapeamento de cores do modelo (escolhido pelo time criativo da CienaLab)..	42

## LISTA DE TABELAS

Tabela 1 – Critério de classificação das peças e distribuição das classes .....	34
Tabela 2 – Top 5 Modelos segundo LazyPredict (via <i>Balanced Accuracy</i> ) .....	35
Tabela 3 – Resultados do Grid Search CV dos modelos finalistas .....	35
Tabela 4 – Resultados do Grid Search CV do modelo final.....	36
Tabela 5 – Desempenho do modelo nas coleções posteriores.....	40
Tabela 6 – Categorização de cores no espaço CIELAB .....	41



# 1 INTRODUÇÃO

O comportamento de consumo na indústria da moda passa por transformações relevantes nas últimas décadas, impulsionada principalmente pela ascensão das redes sociais, com a digitalização dos canais de venda e pela influência consolidada das novas gerações. A Geração Z, em especial destaque, representa um público que valoriza autenticidade, identidade e sua expressão pessoal e procura esse reflexo nos produtos carregados, buscando uma forma de comunicar pertencimento e estilo (Grigoreva; Garifova; Polovkina, 2021). Esse perfil de consumo redefiniu a lógica clássica do mercado da moda e impulsionou a expansão do *streetwear*, que deixou de ser uma manifestação cultural de nicho para se tornar um movimento cultural e econômico global (Patel; Mehta, 2024).

O segmento de moda de nicho conhecido como *streetwear* tem sua origem na fusão entre a cultura urbana periférica, música e esportes de rua que emergiram entre as décadas de 1970 e 1980. Inicialmente o movimento era associado às comunidades de *surf*, *skate* e *hip-hop* e representava uma forma de resistência e identidade coletiva voltada à autenticidade e à valorização das expressões pessoais locais. Com o passar dos anos, marcas independentes e sem investimentos iniciais transformaram esse estilo em um modelo de negócios baseado na escassez de produtos e na exclusividade, com o valor simbólico das peças superando o seu custo de produção. O *streetwear* evoluiu de um fenômeno contracultural e nichado para um dos segmentos mais lucrativos da moda atual.

Inserida nesse cenário global, a CienaLab é uma marca brasileira e carioca que se consolida no segmento *streetwear*, com um modelo de operação baseado em lançamentos periódicos, com curto espaçamento de tempo e de curta duração, conhecidos como “*drops*”. Essa estratégia reflete as mesmas dinâmicas de exclusividade, pertencimento e escassez simbólica que são encontradas na literatura internacional, no qual o valor de um produto está fortemente associado à sua limitação e à narrativa que o acompanha (Patel; Mehta, 2024).

A marca atua com foco em camisetas como produto central, buscando traduzir também visualmente os códigos culturais da Geração Z. Sua comunicação é mediada por influenciadores e redes sociais, o que constitui um dos principais vetores de influência sobre o comportamento de compra dessa geração (Erwin; Saununu; Rukmana, 2023). Esse posicionamento reforça a importância da identidade e da imagem de marca, elementos que determinam o desempenho competitivo das empresas de moda e o nível de conexão emocional com o consumidor, relação conhecida como Amor a Marca. (Ferreira; Rodrigues; Rodrigues, 2019).

Ao mesmo tempo, o modelo produtivo da CienaLab, sustentado por coleções com estoque limitado e pela alta rotatividade de produtos, evidencia o desafio descrito quanto à previsão de demanda em setores sujeitos a forte volatilidade e curta vida útil dos itens (Thomassey, 2010). Com uma forte sazonalidade, ausência de históricos relevantes e uma forte influência de variáveis externas, como o clima e as tendências sociais, o cenário complexo limita a aplicabilidade de modelos estatísticos tradicionais. Dessa forma, o uso de modelos de previsão que consideram múltiplas dimensões da moda, integrando dados qualitativos, como atributos visuais dos produtos, e quantitativos, como volume de vendas, é extremamente importante para criar previsões mais precisas e representativas da realidade de consumo (Giri; Chen, 2022). Esses desafios tornam a marca um objeto de estudo relevante para investigar como atributos qualitativos e quantitativos se relacionam com o desempenho comercial no contexto da moda *streetwear*.

Nesse contexto, este trabalho propõe o desenvolvimento de um classificador supervisionado baseado em Random Forest, escolhido por sua robustez em bases com variáveis categóricas e por sua capacidade de oferecer interpretabilidade sobre os fatores de desempenho do produto. O modelo atua, assim, como uma ferramenta de suporte à decisão, elevando o processo decisório da CienaLab da intuição criativa à evidência quantitativa.

Para garantir o rigor técnico e a aplicabilidade gerencial da solução, ferramentas especializadas foram empregadas. A codificação da cor foi realizada utilizando o espaço CIELAB ( $L^*a^*b^*$ ), essencial para converter a percepção estética em coordenadas métricas e capturar a distância perceptual entre as cores, superando as limitações do simples código nominal. Adicionalmente, para transformar o modelo em uma ferramenta transparente, foi utilizada a técnica de explicabilidade SHAP (SHapley Additive exPlanations). O SHAP permite quantificar a contribuição de cada atributo visual para a previsão de sucesso, gerando insights de negócio valiosos que foram integrados a uma interface gráfica em Python, consolidando a solução como um Sistema de Apoio à Decisão (SAD) para o planejamento de coleções futuras.

A pesquisa parte de um banco de dados de coleções anteriores da marca, no qual cada camiseta é descrita por suas características qualitativas, como cor, tecido, tipo e posicionamento da estampa, aplicação em frente ou costas e proporção visual, associadas aos seus respectivos resultados de venda. A análise considera, de forma conjunta, duas janelas de desempenho, sendo a primeira as vendas realizadas nos primeiros 15 dias após o lançamento da coleção e a segunda as vendas acumuladas ao longo de todo o ciclo da coleção. A primeira janela reflete principalmente a resposta inicial às características estéticas e visuais das peças, enquanto a

segunda captura o desempenho consolidado ao longo do tempo, incluindo influências sociais, como por exemplo o uso por influenciadores digitais e artistas. A combinação desses dois indicadores permite construir categorias de desempenho que diferenciam peças que tenham um forte impacto inicial, com peças que ganham tração ao longo do tempo.

O modelo será então testado e validado a partir do lançamento mais recente da CienaLab, realizado no dia 23 de outubro de 2025, comparando o comportamento real de vendas com os padrões observados nas coleções anteriores. A intenção é avaliar se as estimativas obtidas são coerentes com a realidade de mercado e, assim, construir uma ferramenta que auxilie no processo de planejamento de coleções futuras, permitindo estimar, já na fase de criação, o potencial comercial de cada peça.

Como objetivos específicos, pretende-se:

- a) organizar e caracterizar o banco de dados das camisetas da CienaLab, classificando atributos visuais e materiais;
- b) analisar o comportamento de vendas das peças nos períodos de 15 dias e acumulado até o presente;
- c) identificar correlações entre elementos estéticos e resultados quantitativos; e
- d) validar o modelo construído a partir de coleções anteriores, aplicando-o ao lançamento mais recente da marca.

A motivação deste trabalho reside na possibilidade de transformar os dados históricos da marca em conhecimento aplicável geral, aprimorando o processo decisório na criação e no planejamento de produtos de moda, principalmente do segmento *streetwear*. Ao propor uma integração entre análise estética, comportamento de consumo e gestão da produção, o estudo contribui para reduzir incertezas, minimizar riscos e aumentar a eficiência operacional em um ambiente instável e caracterizado pela rapidez e volatilidade.

O trabalho está estruturado da seguinte forma: o Capítulo 1 apresenta a introdução, abordando o contexto, o problema, os objetivos e a metodologia geral. O Capítulo 2 desenvolve a fundamentação teórica, tratando de comportamento do consumidor, identidade de marca e previsão de demanda no setor de moda. O Capítulo 3 descreve detalhadamente os procedimentos metodológicos, incluindo o tratamento dos dados qualitativos e quantitativos. O Capítulo 4 apresenta os resultados obtidos e a análise do modelo aplicado. Por fim, o Capítulo 5 reúne as conclusões, limitações do estudo e recomendações para pesquisas futuras.

## 2 REFERENCIAL TEÓRICO

### 2.1 Moda *streetwear*, *drops* e consumo da Geração Z

O *streetwear* vai muito além de ser apenas um estilo de roupa e de como se vestir, têm se transformado em um fenômeno cultural e de comunicação que reflete a identidade das pessoas, autenticidade e senso de pertencimento. Dentro desse contexto, surge como uma expressão ligada a movimentos urbanos, música e esportes, consolidando-se nas últimas décadas como algo global, capaz de dialogar com as transformações digitais, sociais e tecnológicas. No Brasil, esse movimento é especialmente forte entre a Geração Z, composta por jovens que cresceram imersos na conectividade, na cultura do consumo imediato e na necessidade de pertencimento.

Os consumidores dessa geração valorizam intensamente a autoexpressão, utilizando a moda como meio de comunicação social (Patel; Mehta, 2024). A escolha de uma peça de roupa ultrapassa a função prática básica de se vestir, passa a carregar outros significados, como a sua personalidade, seu estilo de vida e a identidade do grupo ao qual estão inseridos (Subramaniam; Rathnapriya, 2024). Além disso, a aprovação e o reconhecimento em comunidades digitais têm um papel crucial nesse processo, influenciando diretamente nos padrões de consumo (Shi; Li; Chumnumpan, 2020).

Os jovens dessa geração cresceram em um contexto de hiper conectividade e exposição contínua à informação, tendo sido a primeira completamente moldada num ambiente digital, em que o acesso instantâneo às tecnologias transformam suas relações de consumo e percepção de valor (Wood, 2013). Grigoreva, Garifova e Polovkina (2021) complementam essa visão ao afirmarem que o comportamento de consumo da Geração Z é guiada por uma lógica baseada na comparação constante e busca pela autenticidade, em que esses consumidores tendem a preferir empresas que adotam transparência, coerência entre discurso e práticas de marca e comunicação direta.

Essa busca por autenticidade e validação imediata se materializa nas redes sociais, em especial mais recentemente o Instagram e o TikTok, que funcionam como vitrines em tempo real, que acabam ditando tendências e influenciando diretamente a percepção de relevância de um produto, fazendo com que nesse ambiente a vida útil de um item se torne extremamente curta, já que em questão de poucas horas podem determinar se este item será desejado ou esquecido, reforçando o conceito de “modernidade líquida” de Bauman (2001), em que as

relações e os valores se tornam frágeis, temporários e muitas das vezes descartáveis. No contexto do *streetwear*, um produto pode simbolizar status e reconhecimento em um dia, e no outro se tornar irrelevante.

Pesquisas recentes demonstram que um número expressivo de 85% dos consumidores reconhece que os influenciadores de mídias sociais exercem um impacto considerável nas suas decisões de compra (Erwin; Saununu; Rukmana, 2023), evidenciando o papel que esses influenciadores desempenham na formação do comportamento do consumidor. Ainda no mesmo estudo, é ressaltada que os conteúdos de estilo de vida e avaliações de produtos surgem como as principais preferências, com 65% daqueles que responderam demonstrando uma forte afinidade por essas categorias, reforçando a ideia de que a autenticidade e a identificação na apresentação dos produtos são fatores cruciais para a eficácia de produtos divulgados através de influenciadores.

Essa conexão entre a Geração Z, a cultura digital e o consumo acelerado convergem com a lógica de *drops*. Essa estratégia transforma lançamentos limitados em eventos culturais, usando a escassez e a urgência para criar no consumidor a necessidade de adquirir um produto, baseado no seu desejo por autenticidade e pertencimento, sendo assim a materialização do consumo contemporâneo: líquido, imediato e social (Kim; Huang, 2020).

### 2.1.1 *Drops*, escassez e ciclos curtos

No contexto de marcas independentes, os lançamentos pontuais, também conhecidos como *drops*, tornaram-se prática estratégica já consolidada em todo o mercado, consistindo em coleções de tiragens limitadas, em que são anunciadas e vendidas principalmente em canais digitais (Kim; Huang, 2020). A escassez programada cria uma sensação de exclusividade e incentiva o engajamento, fortalecendo a comunidade em torno da marca.

Entretanto, o modelo de *drops* apresenta frequentes desafios operacionais, devido a sua característica de tiragens reduzidas e margens de lucro apertadas, em que qualquer erro de previsão pode gerar excesso de estoque ou escassez de produtos, causando frustração entre os consumidores e perda de oportunidades de vendas. O curto ciclo dos produtos torna essa tomada de decisão ainda mais crucial, exigindo que as marcas ajustem rapidamente seus estoques e estratégias de *marketing* (Cachon; Swinney, 2011). Syntetos *et al.* (2016) destacam que, em produtos de ciclo curto, como os de moda, o tempo de mercado é insuficiente para consolidar

padrões históricos de vendas, inviabilizando o uso de métodos tradicionais baseados em séries longas de dados.

Diante desse cenário, existe uma maior necessidade de previsibilidade de demanda, em que as ferramentas analíticas e modelos preditivos permitem que as marcas independentes conciliem criatividade, engajamento e eficiência operacional, minimizando riscos e aproveitando melhor as oportunidades de mercado.

## **2.2 Previsão de demanda na indústria da moda**

Prever a demanda significa estimar o volume futuro de vendas de produtos ou serviços com base nas informações e dados disponíveis (Nenni; Giustiniano; Pirolo, 2013). No setor da moda, isso é essencial na hora da tomada de decisão relacionadas à produção, aquisição de insumos, logística e claro, planejamento financeiro.

No entanto, a previsão da demanda na moda apresenta desafios próprios e uma grande dificuldade em utilizar os métodos tradicionais, já que a alta volatilidade das preferências dos consumidores junto com o apelo simbólico dos produtos e curto ciclo de vida impactam na obtenção de dados históricos. Junto disso, o lançamento de novas coleções significa não existir dados confiáveis para embasar estimativas (Nenni; Giustiniano; Pirolo, 2013).

Portanto, prever a demanda nesse contexto exige a integração de abordagens quantitativas, como histórico de vendas e sazonalidade, com variáveis qualitativas, como percepção de marca e atributos estéticos e referenciais dos produtos. Essa combinação é essencial para capturar a complexidade do consumo contemporâneo de moda *streetwear*, especialmente considerando a Geração Z, cuja tomada de decisão é fortemente influenciada por fatores sociais e simbolismos (Chen; Lu, 2021).

## **2.3 Atributos visuais e identidade de produto na decisão de compra**

No setor de moda, e principalmente no segmento *streetwear*, os produtos são avaliados e tem a sua demanda inicial pela sua aparência visual. A literatura mostra que elementos como cor, estampa, material, textura e proporção constituem estímulos fundamentais no julgamento estético do consumidor, fatores que impactam diretamente a decisão de compra. Yu *et al.* (2021) demonstram que os usuários analisam características visuais como cor, forma, padrão gráfico e acabamentos para inferir a partir disso o estilo, qualidade e adequação pessoal, tornando esses

elementos determinantes do interesse inicial por um produto. Segundo os autores, os atributos estéticos são fatores decisivos no processo de escolha, principalmente em categorias guiadas por tendências e identidade cultural, como a moda.

Essa relação entre aparência e intenção de compra também é observada em estudos experimentais sobre design de produto. Kuo, Chang e Lai (2022) em seus estudos mostram que as decisões relacionadas com a cor, brilho, volumetria e acabamentos geram respostas emocionais diferentes, capazes de alterar a atratividade percebida pelo cliente. Certas combinações cromáticas ou formas mais marcadas tendem a comunicar modernidade, tecnologia ou delicadeza, influenciando a forma como o consumidor interpreta o produto antes mesmo de considerar os seus outros aspectos. Os autores reforçam que elementos de forma e cor compõem um vocabulário visual que orienta a interpretação emocional do objeto e a sua demanda.

Dessa forma, a literatura converge ao apontar que atributos visuais são variáveis explicativas legítimas e com impacto mensurável sobre o desempenho de produtos de moda. A cor, o material utilizado, a presença e o tamanho de estampas e o estilo de seus elementos gráficos constituem fatores capazes de modular tanto o interesse inicial quanto o comportamento de compra ao longo do ciclo de vida do produto. Esses achados justificam a inclusão dessas variáveis no modelo desenvolvido neste estudo, uma vez que elas influenciam não apenas a percepção subjetiva de estilo, mas também o desempenho quantitativo de vendas em coleções de moda.

## **2.4 Aprendizado supervisionado e modelos de classificação na moda**

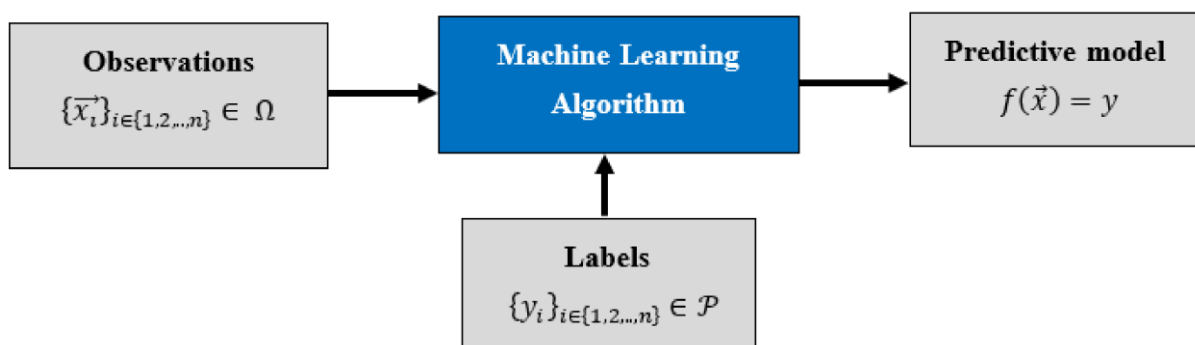
Diante dos desafios apresentados no setor da moda, torna-se essencial compreender como a literatura tem tratado a previsão de sucesso de novos produtos e a análise de fatores que influenciam seu desempenho. Essa discussão envolve tanto entender os modelos de classificação quanto as variáveis e abordagens utilizadas para mensurar o potencial de sucesso dos produtos.

O diagnóstico de desempenho dos produtos de moda não se restringe à análise de vendas totais, exigindo a compreensão de como os atributos específicos inerentes aos produtos se relacionam com o comportamento dos consumidores. Para lidar com essa complexidade, a literatura tem avançado em direção a abordagens quantitativas baseadas em aprendizado de máquina, que permitem identificar padrões e gerar previsões em contextos de alta incerteza.

A previsão de sucesso de produtos baseia-se, em grande parte, em técnicas de aprendizado supervisionado, nas quais um modelo é treinado com exemplos históricos contendo as entradas (características dos produtos) e saídas (resultados de sucesso ou fracasso) conhecidas. A partir desse aprendizado, o modelo é capaz de classificar novos casos, prevendo seu provável desempenho futuro (Hastie; Tibshirani; Friedman, 2008).

Conforme ilustrado na Figura 1, o aprendizado supervisionado parte de um conjunto de dados rotulados, utilizados para treinar um modelo capaz de prever novas observações com base nos padrões aprendidos (Toche Tchio *et al.*, 2024).

Figura 1 – Princípio do aprendizado supervisionado



Fonte: Toche Tchio *et al.* (2024).

Entre os principais modelos aplicados nesses contextos de classificação estão as árvores de decisão e suas variações. Esses algoritmos utilizam regras de hierarquia para segmentar os dados com base nas variáveis mais relevantes, permitindo identificar de forma clara os fatores que mais contribuem para o sucesso de determinado produto.

O estudo de Breiman (2001) foi um marco nesse campo ao propor o algoritmo *Random Forest*, que combina múltiplas árvores de decisão independentes para aumentar a precisão das previsões. Esse método apresenta excelente desempenho em cenários com grande número de variáveis categóricas e amostras limitadas, que são características comuns no mercado da moda.

Segundo Thomassey (2010), a aplicação de algoritmos supervisionados em moda mostra-se eficiente pois permite incorporar variáveis qualitativas, como categoria do produto, cor, estilo, estampa, em conjunto com dados quantitativos de vendas, possibilitando compreender padrões de comportamento de consumo e construir previsões mesmo para lançamentos inéditos. Esse modelo permite distinguir produtos com alta e baixa aceitação de mercado e traduzir esse conhecimento em decisões práticas sobre estoque e lançamento



## 2.5 Modelos baseados em árvore e *Random Forest*

Os modelos baseados em árvore são amplamente valorizados pela sua capacidade de representar processos de decisão de maneira intuitiva, fazendo com que a relação entre as variáveis preditoras e resultados estejam mapeados através de uma estrutura hierárquica de regras. Cada “nó” representa uma pergunta sobre uma variável (por exemplo: “A cor é preta?”), e cada ramificação indica o caminho tomado conforme as respostas, resultando em uma classificação final. Essa lógica reflete o próprio raciocínio humano, o que torna as árvores úteis em áreas que exigem interpretação dos resultados (Breiman, 2001).

A principal evolução dessa técnica foi a criação do *Random Forest*, que combina múltiplas árvores de decisão independentes treinadas com subconjuntos aleatórios dos dados. Ainda de acordo com Breiman (2001), essa abordagem é menos suscetível a ruídos e variações aleatórias, o que aumenta sua capacidade de generalização, fator imprescindível em mercados de alta incerteza como o da moda. Em lançamentos do tipo *drops*,

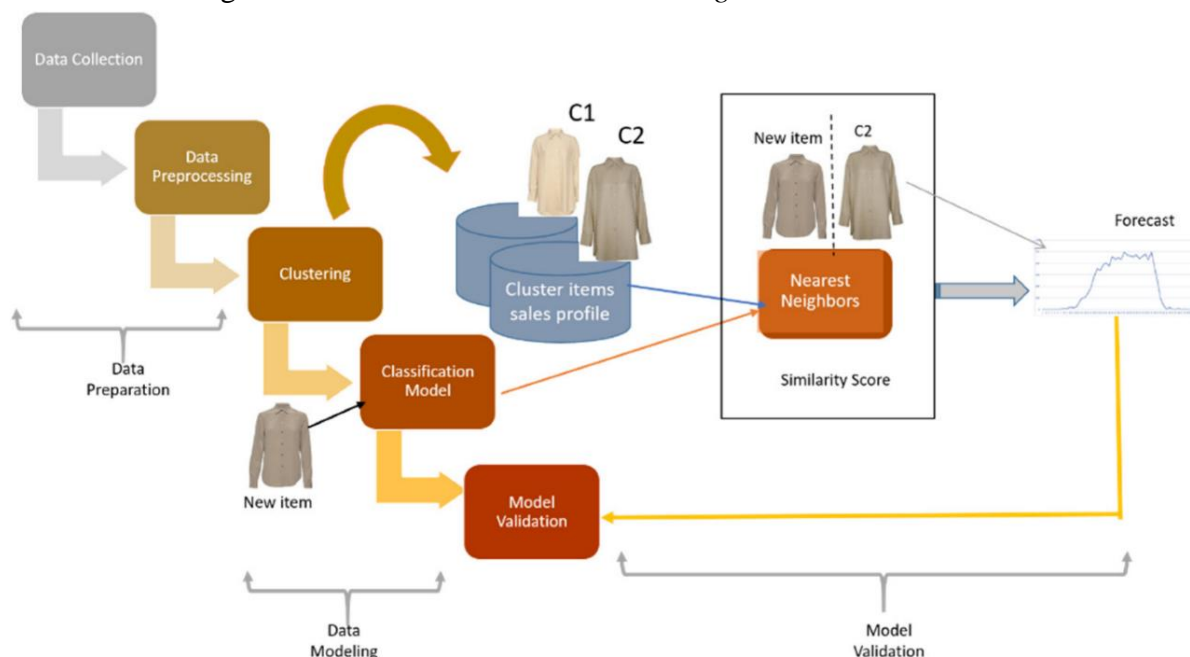
Na área da moda, Thomassey (2010) e Anitha e Neelakandan (2024) reforçam que modelos de árvore são especialmente eficazes para analisar produtos com atributos predominantemente categóricos como cor, estilo e tipo de lançamento. Essas variáveis, que refletem aspectos culturais e simbólicos, muitas vezes têm peso decisivo na percepção de valor por parte dos consumidores. A capacidade do *Random Forest* de lidar naturalmente com dados categóricos sem necessidade de transformações complexas o torna ideal para o mundo da moda, onde as características como “cor: preto”, “estampa: centralizada”, e “modelagem: *oversized*” são fundamentais.”

Enquanto abordagens avançadas como as de Avogaro *et al.* (2024) e Kunz *et al.* (2023) mostram resultados impressionantes em larga escala, sua complexidade e requisitos computacionais os tornam menos acessíveis para marcas independentes de *streetwear*. Esta lacuna justifica a exploração de métodos mais simples, porém eficazes, como *Random Forest*, que melhor se adequam ao contexto de dados limitados e recursos restritos. porém eficazes, como *Random Forest*, que melhor se adequam ao contexto de dados limitados e recursos restritos.

De forma geral, os modelos de *machine learning* aplicados à moda seguem um fluxo estruturado que parte da coleta de dados, preparação, geração e avaliação das previsões. Um

exemplo que contextualiza bem pode ser representado na Figura 2, de Giri e Chen (2022), que mostra as etapas principais de um processo preditivo.

Figura 2 – Processos de *machine learning* na indústria da moda



Fonte: Giri e Chen (2022).

## 2.6 Representação de variáveis categóricas em aprendizado de máquina

As bases de dados de *marketing* e varejo de moda costumam ser compostas por uma combinação de variáveis contínuas, como por exemplo preço, vendas e tempo e variáveis discretas como cor, código do produto ou tipo de material. Guo e Berkahn (2016) destacam que as variáveis discretas como a idade, cor ou material podem ser representadas por inteiros, mas, em muitos casos, esses inteiros são apenas rótulos nominais, sem qualquer significado ordinal, como por exemplo, atribuir 1, 2 e 3 para vermelho, azul e amarelo não implica que a cor azul é maior que a cor vermelha, ou que a média entre vermelho e amarelo resulta no azul. Essa distinção entre números nominais e ordinais é fundamental para o pré-processamento, pois impede que o modelo interprete essas relações numéricas que não existem de fato entre categorias.

No contexto do aprendizado de máquina, essa representação dessas variáveis categóricas precisa ser também adequada ao tipo de algoritmo e às restrições de cardinalidade. Para atributos de baixa cardinalidade, uma abordagem clássica é a codificação 1-para-N (*one-hot encoding*), em que cada categoria é mapeada para um vetor binário com um único elemento igual a 1 e os demais iguais a 0. Micci-Barreca (2001), em artigo publicado no *SIGKDD*

*Explorations*, aponta que esse esquema é mais utilizado quando o número de categorias é pequeno, mas se observa pouco viável quando a cardinalidade cresce, já que aumenta excessivamente o número de entradas do modelo e, conseqüentemente, a quantidade de parâmetros a serem estimados. Esse problema é particularmente relevante em bases de moda, nos quais os atributos da base de dados como a cor, a estampa ou o seu material podem assumir dezenas de valores distintos ao longo do tempo.

Dessa mesma forma, para atributos de alta cardinalidade, Micci-Barreca (2001) propõe um esquema de pré-processamento baseado em estatísticas do alvo (*target statistics encoding*). Em vez de representar cada categoria por um vetor, o autor mapeia cada valor categórico para uma estimativa da probabilidade ou do valor esperado da variável dependente, o que transforma o atributo original em um escalar quase que contínuo. Essa transformação é formulada combinando a estatística específica de cada categoria com uma probabilidade ou uma média global por meio de um fator de “*shrinkage*”, o que permite lidar melhor com categorias raras e reduzir o risco de *overfitting*, que é quando um modelo de *machine learning* se ajusta muito bem aos dados de treinamento, a ponto de quase memorizar os seus ruídos e detalhes específicos, impedindo de generalizar e fazer previsões precisas em dados novos e desconhecidos. O resultado é uma representação numérica efetiva que preserva boa parte do poder preditivo do atributo original e é adequada a modelos de regressão e classificação sensíveis à escala dos dados.

Além desses esquemas, Guo e Berkhahn (2016) trazem na literatura as limitações de abordagens ingênuas, como alimentar diretamente redes neurais com códigos inteiros de categorias, e descrevem o uso de *one-hot encoding* e *entity embeddings* como alternativas mais adequadas. Segundo esses autores, o *one-hot encoding* tende a exigir muitos recursos computacionais quando há diversas variáveis de alta cardinalidade e, além disso, trata cada categoria como completamente independente, ignorando relações potencialmente informativas entre elas. Para contornar esse problema, existe o método de *entity embedding* que mapeia cada estado de uma variável categórica para um vetor em um espaço multidimensional, de modo que valores com efeitos semelhantes na função a ser aproximada fiquem próximos nesse espaço, revelando uma continuidade latente nas categorias.

Guo e Berkhahn (2016) mostram, em experimentos com a base Rossmann, que diferentes algoritmos lidam de formas distintas com as suas variáveis categóricas. O k-NN recebe as entradas *one-hot* codificadas, enquanto *random forests* e *gradient boosted trees* operam diretamente com as variáveis categóricas codificadas como inteiros. Embora esses

inteiros continuem sendo números nominais, os modelos de árvore particionam o espaço de estados por meio de divisões em valores discretos, os chamados splits, o que permite explorar a informação categórica sem assumir uma relação ordinal contínua. Essa evidência empírica mostrou-se relevante para o presente estudo, na medida em que sustenta o nosso uso de codificações inteiras, os *Label Encoding*, em conjunto com algoritmos de árvore, como o *Random Forest*, que foi adotado no modelo da CienaLab, desde que se reconheça explicitamente o caráter nominal dessas variáveis. Dessa forma, a literatura sobre representação de variáveis categóricas estabelece um bom plano de fundo teórico para as decisões de pré-processamento tomadas neste trabalho e no modelo proposto.

## **2.7 Seleção de modelos, validação e ajuste de hiper parâmetros**

Em problemas de previsão de demanda com múltiplas variáveis categóricas, a etapa de seleção de modelos e ajuste de hiper parâmetros é de extrema importância para garantir que o desempenho observado não seja apenas resultado de sobre ajuste ao conjunto de treino, mas reflita corretamente a capacidade real de generalização do algoritmo utilizado.

Em estudos relevantes sobre vendas no varejo, Guo e Berkahn (2016) comparam diferentes modelos supervisionados, o *k-nearest neighbors* (KNN), o *random forests* e o *gradient boosted trees*, utilizando a biblioteca *scikit-learn* para implementação dos algoritmos e padronização dos experimentos. Nesse trabalho, os autores deixam evidente que os hiper parâmetros de cada modelo, como por exemplo a profundidade máxima das árvores, número de estimadores e parâmetros de amostragem, foram definidos empiricamente a partir da otimização do desempenho em um conjunto de validação, e não apenas no conjunto de treino, justamente para reduzir o risco de ajustes oportunistas aos dados disponíveis.

Além da escolha adequada dos algoritmos, a literatura também enfatiza a importância de métricas corretas para comparar os modelos sob as mesmas condições, de modo justo. No estudo de Guo e Berkahn (2016), por exemplo, a avaliação é feita por meio do *Mean Absolute Percentage Error* (MAPE), adotado em substituição ao *Root Mean Square Percentage Error* (RMSPE) por apresentar um comportamento mais estável na presença dos outliers em séries de vendas. Esse cuidado na hora da metodologia de definir um critério de desempenho claro e utilizá-lo para comparar diferentes configurações é muito coerente com as recomendações mais gerais da literatura de previsão em cadeias de suprimentos, que nos apontam as necessidades

de ferramentas e processos para testar modelos de forma padronizada e robusta antes de sua efetiva implementação em ambientes reais.

Neste trabalho, a lógica adotada para a seleção de modelos segue a mesma direção. Após a construção da base de dados com as variáveis visuais e os indicadores de vendas, são identificados, com o auxílio da biblioteca *scikit-learn*, os modelos com melhor desempenho preditivo e, em seguida, realiza-se o ajuste dos hiper parâmetros, como a profundidade máxima das árvores (*max\_depth*) e o número de estimadores (*n\_estimators*), por meio de busca em grade, o chamado *Grid Search*, combinada a validação em múltiplas partições dos dados. Esse procedimento permite comparar de forma justa algoritmos como *Random Forest*, métodos de *boosting* e classificadores de base simples, sob o mesmo conjunto de variáveis e métricas, para então selecionar o modelo que apresenta maior estabilidade e acurácia sobre os dados históricos da Cienalab.

Dessa forma, a etapa de escolha do modelo final não é arbitrária, mas principalmente é fundamentada em práticas corretas da literatura de aprendizado de máquina aplicado a dados de vendas.

## **2.8 Explicabilidade em modelos preditivos: XAI e SHAP na moda**

A literatura argumenta que sistemas baseados em *machine learning* não podem ser tratados como caixas-pretas quando suas decisões afetam os processos como a previsão de demanda ou a gestão de estoques. Essa preocupação é explicitada no trabalho de Ribeiro, Singh e Guestrin (2016), que demonstram que a ausência de explicabilidade reduz a confiança de usuários e dificulta a validação de decisões corretas, especialmente em ambientes sensíveis a variações comportamentais, como o estudado consumo de moda.

Nesse contexto surge o campo de *Explainable Artificial Intelligence* (XAI), cujo objetivo é permitir que os modelos complexos sejam compreendidos em termos de como utilizam as variáveis de entrada para gerar cada predição. A interpretabilidade é especialmente relevante em domínios onde atributos subjetivos, como a estética das peças, o seu design e a cor do produto influenciam diretamente o comportamento do consumidor. Para o setor de moda, onde pequenas diferenças visuais podem alterar significativamente o desempenho de um produto, modelos explicáveis se tornam instrumentos cada vez mais essenciais para apoiar decisões criativas e operacionais nas marcas.

Entre as abordagens que emergem dentro do XAI, o método SHAP (*SHapley Additive exPlanations*) é um destaque por oferecer uma decomposição da previsão em contribuições individuais das variáveis. Lundberg e Lee (2017) apresentam o SHAP como uma estrutura baseada na teoria dos valores de *Shapley*, permitindo atribuir, a cada *feature*, um valor que expressa sua influência positiva ou negativa em uma determinada predição. Essa abordagem é particularmente adequada para problemas de classificação na moda, nos quais não apenas o resultado, como por exemplo a categoria de desempenho de um produto é relevante, mas também a compreensão dos fatores que levaram o modelo chegar até essa classificação.

Assim, o uso do SHAP no presente estudo permite interpretar de forma mais eficaz como as variáveis dos produtos lançados, como a cor, o material utilizado e as características estéticas das estampas influenciam a probabilidade de uma peça ser classificada como *best performer*, *hyped*, *normal* ou *late performer*. Essa capacidade de decompor a decisão do modelo é coerente com as recomendações da literatura de XAI (Ribeiro; Singh; Guestrin, 2016) e com a abordagem teórica de explicabilidade proposta por Lundberg e Lee (2017). Além disso, estudos sobre estética visual e comportamento do consumidor, como o de He *et al.* (2018), reforçam que atributos visuais são determinantes e, portanto, precisam ser analisados individualmente, justamente o que a aplicação de SHAP possibilita neste trabalho.

## **2.9 Representação métrica da cor em modelos preditivos: o uso do espaço CIELAB**

A cor é uma das variáveis estéticas mais influentes no desempenho de vendas principalmente de camisetas, especialmente no segmento de moda *streetwear*. Para que atributos cromáticos possam ser utilizados com variáveis quantitativas em modelos de aprendizado de máquina, é necessário convertê-los para um sistema numérico que o represente de forma clara as diferenças perceptuais entre as suas cores. Nesse sentido, o espaço CIELAB ( $L^*$ ,  $a^*$ ,  $b^*$ ) constitui um dos modelos mais amplamente empregados para análise computacional da cor.

Wang *et al.* (2006) demonstram que o CIELAB é capaz de representar informações cromáticas de forma significativa, permitindo comparar cores com base em métricas universais de luminosidade ( $L^*$ ), eixo vermelho-verde ( $a^*$ ) e eixo amarelo-azul ( $b^*$ ). No estudo, os autores desenvolveram um algoritmo para categorizar automaticamente cores em 11 nomes universais, como vermelho, azul, amarelo e preto, utilizando justamente valores derivados do espaço CIELAB. Os experimentos realizados pelos autores confirmam que a transformação de cores

para coordenadas  $L^*a^*b^*$  permitem estabelecer fronteiras cromáticas e aplicáveis a sistemas de algoritmos, possibilitando análises quantitativas mesmo em domínios com o seu alto grau de subjetividade perceptual de acordo com as crenças de cada cliente.

Em seu trabalho, Wang *et al.* (2006) ressaltam que o CIELAB é utilizado em aplicações de análise de imagem, reprodução de cor e classificação visual devido à sua capacidade de aproximar as diferenças perceptuais reais e observadas por humanos. Essa característica torna o espaço claro para tarefas de classificação automatizada que exigem precisão e comparabilidade entre amostras cromáticas, exatamente o caso de modelos preditivos orientados para avaliar o desempenho de produtos a partir de atributos visuais. Os autores demonstram ainda que, apesar de alternativas mais recentes como o modelo CIECAM02 apresentarem um desempenho as vezes superior e apenas em certos cenários, o CIELAB continua sendo uma ferramenta consolidada e eficiente para representar cor de forma não ordinal em suas bases de dados.

No presente estudo, a conversão da variável “cor” para códigos numéricos, que foram operadas por meio do *Label Encoding*, está sustentada pela premissa estabelecida por Wang *et al.* (2006). A cor pode ser mensurada e reduzida a atributos matemáticos e numéricos sem a perda de significado perceptual. Assim, ao transformar cores nomeadas, como “branco”, em representações numéricas consistentes, torna-se possível utilizá-las como variáveis independentes em modelos, permitindo que algoritmos como o *Random Forest* capturem padrões entre esses atributos cromáticos e o seu desempenho comercial das peças quando lançadas para o público geral.

Dessa forma, além de viabilizar a modelagem, essa conversão possibilita a interpretação posterior via métodos explicáveis, como o mencionado SHAP, que podem indicar como que determinadas dimensões de cor contribuem positivamente ou negativamente para o desempenho das camisetas analisadas dentro das coleções lançadas, conectando, assim, métricas reais a percepções estéticas observadas no comportamento de compra dos clientes.

## 3 METODOLOGIA

### 3.1 Tipo de pesquisa, dados e variáveis

Este trabalho caracteriza-se como uma pesquisa aplicada, de abordagem quantitativa e natureza preditiva, já que tem como objetivo desenvolver e avaliar um modelo de aprendizado de máquina capaz de apoiar nas decisões no desenvolvimento de produtos do segmento de moda *streetwear*. A estratégia adotada é um estudo de caso, realizado na CienaLab, buscando analisar o comportamento de vendas de suas camisetas a partir dos seus atributos visuais e de materiais. Esse enquadramento é coerente com a literatura de previsão de demanda na moda, que destaca a relevância de modelos quantitativos baseados em dados históricos para lidar com a incerteza e a volatilidade do setor (Giri; Chen, 2022; Koren; Shnaiderman, 2023; Nenni; Giustiniano; Pirolo, 2013; Thomassey, 2010).

A base de dados utilizada foi construída a partir dados internos de vendas da marca, do período de janeiro de 2022 até setembro de 2025, organizados em produtos e coleções. Cada linha da base corresponde a uma peça específica com a sua coleção correspondente, incluindo informações de identificação, atributos de design e composição, além de indicadores de seu desempenho comercial. Entre esses indicadores, é importante destacar a participação percentual da peça nas vendas da coleção nos primeiros 15 dias e no total do ciclo de vendas da coleção, essas métricas permitem observar tanto a resposta inicial do mercado quanto o desempenho consolidado do produto ao longo do tempo. A integração de dados operacionais e históricos de vendas em sistemas analíticos é apontada na literatura como fator de extrema importância para o aumento da acurácia de previsões em cadeias de moda e varejo (Koren; Shnaiderman, 2023; Nenni; Giustiniano; Pirolo, 2013).

No presente estudo, as variáveis independentes são restritas a atributos que podem ser definidos antes do lançamento da coleção, ou seja, características sob responsabilidade das áreas do time de produto e design da marca. Foram consideradas no total a cor da peça, o material principal, o material secundário, o tipo de estampa frontal, o tamanho da estampa frontal, o tipo de estampa nas costas e o tamanho da estampa nas costas (Quadro 1). A escolha por focar tanto nos atributos visuais quanto nos materiais está alinhada à literatura que relaciona elementos de design, identidade de marca e percepção de valor do consumidor, indicando que essas características podem também influenciar diretamente o desempenho comercial de produtos de moda (Clotey *et al.*, 2023; Ferreira; Rodrigues; Rodrigues, 2019).



Quadro 1 – Variáveis utilizadas no modelo preditivo

Variável	Tipo	Descrição resumida
cor	Categórica	Cor principal da peça
material1	Categórica	Material principal do corpo da peça
material2	Categórica	Material secundário ou de complemento
estampaF	Categórica	Tipo de estampa na parte frontal
tamanhoEstF	Categórica	Tamanho da estampa frontal
estampaC	Categórica	Tipo de estampa na parte das costas
tamanhoEstC	Categórica	Tamanho da estampa nas costas
%vdacolecao15	Numérica	Porcentagem das vendas nos primeiros 15 dias
%vdacolecaotot	Numérica	Participação da peça no total de vendas da coleção
performance_category	Categórica	Classe de desempenho atribuída à peça.

Fonte: Autoria própria (2025).

A variável-alvo, dependente, do modelo foi construída a partir da junção dos indicadores de desempenho nas janelas vendas de 15 dias e no total da coleção. Inicialmente, foram calculada as médias da participação percentual das peças nas vendas dos primeiros 15 dias e nas vendas totais da coleção. Em seguida, cada produto foi classificado por sua faixa de desempenho, resultando na variável categórica *performance\_category*, que foi a responsável por agrupar as peças nas categorias *best performer*, *hyped* e *normal*, conforme o posicionamento relativo em relação às médias dos indicadores. Essa categorização permite transformar um problema originalmente contínuo, de previsão de volumes, em um problema de classificação, facilitando a interpretação dos resultados, permitindo uma melhor tomada de decisão por parte de todos os integrantes do time da marca, e aproximando a análise das propostas da literatura que empregam modelos de classificação para apoiar decisões de sortimento e planejamento em varejo de moda (Giri; Chen, 2022; Koren; Shnaiderman, 2023).

### 3.2 Estudo de caso: CienaLab

#### 3.2.1 CienaLab

A CienaLab foi fundada em 2021, no Rio de Janeiro, em Bangu, com a proposta de operar como um laboratório criativo. A peça de roupa é apenas um suporte para a construção de um estilo de vida. Desde a origem, a marca assumiu como princípio a noção de comunidade, sendo essa sintetizada no lema *Family Business*, e a prática de lançar coleções em *drops* de curta duração, explorando sempre uma estética urbana por meio de design, além de ilustrações e posicionamentos de estampa que são característicos da marca. Essa orientação criativa se

materializa em coleções sempre temáticas, como por exemplo, referências locais como o futebol de Bangu ou universos ficcionais como “St. Ciena”, nas quais a narrativa guia as decisões de design e de comunicação da marca.

A trajetória da marca também dialoga em maioria com o perfil da Geração Z, grupo que valoriza principalmente a autenticidade, identidade e o sentimento de participação em comunidades simbólicas, articulando suas decisões de consumo em ambientes digitais (Grigoreva; Garifova; Polovkina, 2021). Nesse ecossistema criado, os influenciadores, artistas e criadores desempenham papel central na legitimação dessas tendências e na aceleração dos ciclos de desejo para o consumidor final, afetando diretamente a recepção de novos produtos e a dinâmica das vendas ao longo do tempo (Erwin; Saununu; Rukmana, 2023).

### 3.2.2 Modelo de negócios

O modelo de negócios da CienaLab é pensado a partir da lógica de lançamentos sazonais e de curta duração, conhecidos como *drops*, formato característico do mercado *streetwear*. Cada uma é pensada e planejada com base em um conceito narrativo central, que orienta e baseia tanto o desenvolvimento estético quanto as decisões de comunicação da coleção. Essa estrutura permite equilibrar a criatividade e previsibilidade operacional, uma vez que cada coleção possui um número limitado de peças, sendo em média entre 30 e 40, e um cronograma de lançamento de aproximadamente três meses entre coleções. O modelo de *drops* é sustentado pela escassez planejada, os produtos não são esgotados sem motivo, o que gera desejo e exclusividade, sendo amplamente adotado por marcas direcionadas à Geração Z (Patel; Mehta, 2024).

As coleções são organizadas segundo uma estrutura piramidal de produtos, composta por três níveis básicos: base, meio e topo. A base representa aproximadamente 60% da coleção e inclui peças de maior volume de produção e alta rotatividade, como camisetas e shorts, que sustentam financeiramente o restante da linha. O meio da pirâmide, cerca de 30%, é formado por produtos que auxiliam na construção narrativa das coleções, com melhores designs e materiais, ela apresenta jaquetas, calças e moletons que expressam o tema central da campanha, mesmo com menor rentabilidade. Por fim, o topo, que corresponde a cerca de 10% das peças, engloba itens de alto valor agregado e produção reduzida e exclusiva, destinados para gerar desejo e posicionar a marca no imaginário aspiracional do consumidor. Essa organização reflete o equilíbrio entre objetivos comerciais e simbólicos, característica comum às marcas de moda baseadas em histórias contadas. (Ferreira; Rodrigues; Rodrigues, 2019).

As camisetas ocupam posição central na estratégia da CienaLab e são responsáveis pela maior parcela de receita das coleções. Por serem peças de alta demanda e acessíveis, elas funcionam como ponto de entrada para novos consumidores e, ao mesmo tempo, como sustentação financeira para o restante do portfólio da coleção. O foco nesse produto também se justifica pelo potencial de diferenciação com a sua estética, a marca pode explorar combinações específicas de cores, materiais, tipografias e posicionamentos de estampas, elementos que, são determinantes na construção de uma identidade visual consistente e reconhecível (Clottey *et al.*, 2023).

### 3.2.3 Estrutura operacional e produtiva

A operação da Ciena tem base em seus lançamentos com ciclos médio de três meses, envolvendo etapas de pesquisa, design, produção e lançamento. O processo é coordenado por ferramentas digitais, como o ClickUp, utilizado para o gerenciamento de tarefas, e o Shopify, como plataforma de vendas e responsável pela coleta de dados e comportamento de consumo. Essa integração entre controle operacional e análise digital é coerente com o movimento de transformação da indústria da moda, no qual a digitalização dos fluxos produtivos tem ampliado a capacidade de resposta e reduzido custos de coordenação (Shi; Li; Chumnumpan, 2020).

A produção é realizada em lotes reduzidos e planejados conforme a expectativa inicial e previsão de desempenho comercial, priorizando flexibilidade e minimização de estoques. Essa abordagem está também alinhada à filosofia da produção enxuta, que vai tentar reduzir desperdícios e otimizar recursos de acordo com a demanda prevista (Thomassey, 2010). No caso da marca, o planejamento de produção é diretamente condicionado à incerteza de aceitação das peças, uma vez que o comportamento de compra da sua base, sobretudo da Geração Z é altamente volátil e sensível à influência social. Essa volatilidade é um dos principais desafios da cadeia de suprimentos da moda, em especial quando há grande variedade de produtos e ciclos de vida curtos (Nenni; Giustiniano; Pirolo, 2013).

### 3.2.4 Vantagens competitivas e riscos

A principal vantagem competitiva da Ciena reside na coerência entre identidade de marca, a sua estética e a conexão com a sua comunidade. A empresa consolidou o seu posicionamento dentro do *streetwear* brasileiro, utilizando narrativas culturais e linguagem

própria para gerar identificação com o público jovem. Essa combinação entre identidade forte e consistência simbólica é reconhecida na literatura como um dos principais fatores de diferenciação e desempenho em empresas de moda (Clottey *et al.*, 2023). Outro ponto de vantagem é o uso das mídias sociais e de influenciadores de moda, sendo explorado principalmente nas campanhas de lançamento. Essa estratégia é coerente com o comportamento da Geração Z, cuja decisão de compra é fortemente influenciada por personalidades digitais e redes de pertencimento (Erwin; Saununu; Rukmana, 2023). A comunicação próxima com o público e a autenticidade percebida fortalecem a fidelização, além de promover o efeito que diferencia a Ciena das marcas convencionais do mercado.

Contudo, o mesmo conjunto de fatores que leva vantagens à marca também a expõe a riscos operacionais. A dependência da influência digital e a alta velocidade dos ciclos de tendência tornam a previsão de demanda particularmente complexa. Peças que não recebem destaque imediato podem ter vendas aquém do esperado, enquanto outras podem esgotar rapidamente após a exposição de um influenciador, distorcendo o planejamento inicial. Esse comportamento pode ser um dos principais gargalos da cadeia de suprimentos da moda, sendo a discrepância entre a previsão e o consumo real, agravada por fatores externos (Nenni; Giustiniano; Pirolo, 2013). Em coleções de curta duração essa volatilidade implica riscos de estoque encalhado e a perda de oportunidade de venda.

Além disso, a estrutura de *drops* com estoque limitado reduz a margem de correção após o lançamento. Qualquer erro de estimativa no volume ou na aceitação de um produto impacta diretamente o faturamento da coleção. A imprevisibilidade da demanda é uma das principais fontes de ineficiência em sistemas produtivos de moda e sugerem o uso de modelos preditivos baseados em dados históricos para mitigar esses riscos (Koren; Shnaiderman, 2023). Em síntese, a marca apresenta um conjunto de vantagens competitivas associadas à força de marca e a sua coerência estética, mas, entretanto, enfrenta riscos estruturais decorrentes da volatilidade do mercado e da sua imprevisibilidade da demanda.

### 3.2.5 Relevância do caso para o estudo

A escolha da CienaLab como objeto para esse estudo fundamenta-se na possibilidade de aplicar métodos quantitativos de previsão em um ambiente real caracterizado por alta incerteza e múltiplas variáveis qualitativas. A marca representa um caso da indústria da moda na qual as suas decisões de produção e de lançamento são diretamente afetadas por fatores

culturais, estéticos e em um novo ambiente digital. Esse contexto oferece condições ideais para investigar a aplicação de modelos preditivos, capazes de identificar padrões complexos entre atributos visuais e desempenho comercial (Giri; Chen, 2022; Koren; Shnaiderman, 2023).

### 3.3 Preparação dos dados e divisão treino-teste

Para a etapa de modelagem, foi utilizada a base de dados contendo as informações por produto e coleção. Essa base foi importada para o ambiente Python e deu origem a um conjunto de dados no qual foram preservadas apenas as variáveis relevantes descritas na Seção 3.1, bem como os indicadores de desempenho necessários para a construção da variável-alvo. A utilização das bases históricas em nível de produto é coerente com os estudos de previsão de demanda em moda e varejo, que enfatizam a importância de dados consistentes e organizados para a aplicação de técnicas quantitativas (Giri; Chen, 2022; Koren; Shnaiderman, 2023; Nenni; Giustiniano; Pirollo, 2013; Thomassey, 2010).

As variáveis categóricas relacionadas ao design das peças, como a cor, os materiais, os tipos de estampa e os tamanhos de estampa, foram convertidas para códigos numéricos inteiros. Esse procedimento permitiu que os dados fossem utilizados diretamente pelos algoritmos empregados ao estudo, preservando a informação qualitativa dos atributos visuais na forma de variáveis numéricas. Para garantir a reprodutibilidade e o uso posterior do modelo, o dicionário de categorias associado a cada variável foi armazenado juntamente com o classificador treinado, possibilitando que novas peças sejam codificadas segundo o mesmo padrão.

A partir desse conjunto de dados preparado, definiu-se como variável-alvo a *performance\_category* e como preditoras os códigos numéricos das variáveis listadas no Quadro 1. Após isso, o conjunto de dados foi dividido em dois subconjuntos, um para treinamento e ajuste dos modelos propostos e um outro para teste e avaliação do desempenho preditivo. A divisão foi realizada por meio de amostragem aleatória estratificada, para preservar a proporção das classes de desempenho em ambos os subconjuntos, com aproximadamente 60% das observações destinadas ao treino e 40% ao teste. Essa estratégia de separação entre dados de treinamento e dados de teste, associada à estratificação por classe, é consistente com as boas práticas de estudos de previsão de demanda e análise de desempenho de produtos em contextos de moda e varejo (Koren e Shnaiderman, 2023; Giri e Chen, 2022).

### 3.4 Construção e seleção do modelo preditivo

A partir da base de dados tratada e da definição da variável-alvo *performance\_category*, o problema foi formulado como uma tarefa de classificação supervisionada multiclasse. O objetivo desse modelo é estimar, para cada camiseta lançada nas coleções, a probabilidade de pertencer a uma das classes de desempenho definidas como *best performer*, *hyped* ou *normal*. Para isso, obtendo como entrada apenas atributos visuais e de fabricação das camisetas. Essa abordagem está alinhada com estudos que utilizam modelos de aprendizado de máquina para apoiar decisões de previsão de demanda, sortimento e planejamento em contextos de varejo de moda sujeitos a alta incerteza e ciclos de vida curtos (Giri; Chen, 2022; Koren; Shnaiderman, 2023; Nenni; Giustiniano; Pirolo, 2013; Thomassey, 2010).

No primeiro momento, foi realizado uma triagem exploratória dos algoritmos de classificação, com o uso de uma automação que testa diferentes modelos supervisionados em condições corretas de treino e validação. Essa etapa incluiu, entre outros classificadores baseados no *Random Forest*, métodos lineares e modelos probabilísticos. Para cada algoritmo testado, foram calculadas métricas de desempenho, destacando o *balanced accuracy*, que considera de forma mais adequada cenários em que há um desbalanceamento entre as classes da variável-alvo escolhida. Essa comparação inicial permitiu identificar quais famílias de modelos apresentavam um desempenho mais promissor para uma futura assertividade, destacando em particular os métodos baseados em árvores de decisão, que são frequentemente empregados em aplicações de previsão de demanda e classificação de produtos em varejo (Giri; Chen, 2022; Koren; Shnaiderman, 2023).

Na sequência, os modelos candidatos selecionados por essa triagem inicial foram submetidos a um processo de ajuste de seus hiperparâmetros com a ajuda do Grid Search, combinada com uma validação cruzada de dados. Para isso, o conjunto de treinamento foi particionado em cinco subconjuntos, *k-folds*, com a proporção de classes preservadas em cada partição. Em cada combinação de hiperparâmetros, o modelo era treinado e avaliado sucessivamente nesses *folds*, e a *balanced accuracy* média obtida servia como critério de comparação entre essas configurações mencionadas. Foram considerados, nesse processo, modelos como *Random Forest*, *Extra Trees*, *Gaussian Naive Bayes*, *Nearest Centroid* e *Ridge Classifier*, com grades de parâmetros específicas para cada família.

Ao final do processo de ajuste, o algoritmo *Random Forest* foi selecionado como modelo preditivo principal, por apresentar o melhor equilíbrio entre o desempenho médio na

validação cruzada e estabilidade dos resultados entre os *folds*. Com os hiperparâmetros definidos na busca, o classificador foi então treinado em seu conjunto completo, utilizando como preditoras os códigos numéricos das variáveis descritas na Tabela 1 e como saída a variável *performance\_category*, classificando assim as camisetas da coleção entre *best performer*, *hyped* ou *normal*.

### 3.5 Interpretação do modelo (XAI) e variáveis de cor

O modelo *Random Forest* é composto por um conjunto de árvores de decisão e, sendo assim, apresenta um comportamento interno de difícil interpretação direta apenas a partir dos parâmetros do algoritmo. Dado que o objetivo do trabalho não é apenas obter boa acurácia preditiva, mas também compreender de que forma esses atributos visuais e fabris das camisetas se relacionam com a quantidade de vendas de cada uma, optou-se pelo uso de técnicas de explicabilidade de modelos, XAI. A adoção de abordagens de XAI permite aproximar a modelagem preditiva das necessidades gerenciais, o que nos fornece também evidências concretas sobre a importância relativa das variáveis e os sentidos das contribuições de cada atributo para as previsões do modelo.

Para interpretar o modelo de *Random Forest* treinado, foram calculados valores de SHAP, *Shapley Additive exPlanations*, coerente com a abordagem teórica de explicabilidade proposta por Lundberg e Lee (2017), utilizando um explicador para modelos de árvores. Os valores de SHAP são baseados na teoria dos valores de *Shapley*, originalmente formulada na teoria dos jogos cooperativos, e nos permitem quantificar, para cada observação feita, uma contribuição marginal de cada variável preditora para o desvio da previsão em relação a um valor de referência. Para esse estudo, esses valores foram calculados para o conjunto de teste e sintetizados em gráficos de resumo, que mostram, em termos agregados, quais atributos das peças exercem maior influência sobre a sua classificação nas diferentes categorias de desempenho. Esse procedimento complementa a análise de importância de variáveis do *Random Forest* e nos oferece uma visão mais nítida sobre o papel de cada atributo dentro das coleções.

Além da interpretação total do modelo, foi necessário também tratar as variáveis de cor de cada camiseta, já que ela é um atributo visual de extrema importância no design das camisetas e relevante para a percepção de valor do consumidor. A base histórica utilizada na modelagem trabalha com categorias discretas de cor, como “preto”, “*off-white*”, “azul” e outras,

enquanto, no processo de desenvolvimento de novos produtos, a cor é frequentemente especificada em termos de coordenadas em espaços de cor uniformes. Para aproximar essas duas representações, foi implementado o padrão de cor CIELAB ( $L^*a^*b^*$ ), convertendo essas categorias discretas em um procedimento de mapeamento entre coordenadas  $L^*a^*b^*$ , no qual a luminosidade ( $L^*$ ), eixo vermelho-verde ( $a^*$ ) e eixo amarelo-azul ( $b^*$ ) representam uma cor específica lida em um código de coordenada.

Na ferramenta desenvolvida, o usuário informa os valores de  $L^*$ ,  $a^*$  e  $b^*$  correspondentes à cor desejada da peça. A partir desses valores, são calculados o croma e o ângulo de matiz, que permitem classificar a cor básica seguindo regras derivadas do modelo de categorização adotado. Em seguida, cada cor básica é associada a uma ou mais categorias de cor efetivamente presentes na base histórica da marca, selecionando-se a opção mais coerente com a paleta utilizada pela Cienalab. Esse mapeamento garante consistência entre a forma como as cores são especificadas no processo de design e a forma como são representadas na entrada do modelo de classificação, permitindo que a previsão de desempenho seja sensível a variações de cor de maneira compatível com os dados que fundamentaram o treinamento do modelo utilizado.

### 3.6 Implementação da ferramenta e validação em nova coleção

Após a definição do modelo final, o classificador *Random Forest* foi exportado para uso operacional por meio da biblioteca *joblib*. O arquivo gerado armazena, de forma conjunta, o modelo treinado, o dicionário de categorias correspondente às variáveis preditoras e os valores médios utilizados na construção da variável *performance\_category*. A disponibilização do modelo desta forma permite que ele seja integrado a outras aplicações sem a necessidade de repetir todo o processo de treinamento, viabilizando seu uso cotidiano como ferramenta de apoio à decisão no desenvolvimento das futuras coleções da marca. A literatura de previsão de demanda em moda e varejo destaca a importância de soluções que não se restrinjam ao ambiente de pesquisa, mas que sejam incorporadas nas rotinas empresariais, contribuindo para reduzir incertezas e apoiar decisões em contextos de alta variabilidade de demanda (Giri; Chen, 2022; Koren; Shnaiderman, 2023; Nenni; Giustiniano; Pirolo, 2013; Thomassey, 2010).

Com base nesse modelo, foi desenvolvida uma interface gráfica em Python, utilizando a biblioteca CustomTkinter, com o objetivo de tornar a ferramenta acessível aos responsáveis pelo processo de criação de produtos. Essa interface permite o registro de novas peças de



camiseta com o preenchimento de campos com os atributos visuais e fabris, incluindo as coordenadas de cor em CIELAB ( $L^*$ ,  $a^*$ ,  $b^*$ ), seleção da cor correspondente utilizada pela marca, materiais principal e secundário, tipos de estampa na frente e nas costas e respectivos tamanhos. Foram implementadas regras de validação para garantir a consistência das combinações selecionadas, como por exemplo a vinculação entre a opção “sem estampa” e o tamanho igual a zero. A partir dos valores informados, a ferramenta codifica as categorias segundo o mesmo dicionário utilizado no treinamento do modelo e realiza uma previsão, apresentando ao usuário a classe de desempenho estimada e as probabilidades associadas a cada categoria. Adicionalmente, a interface permite agrupar várias peças em uma coleção “falsa” para gerar um resumo com a distribuição de produtos entre as classes de desempenho, o que facilita o uso do modelo como apoio à composição da pirâmide de produtos dentro de uma coleção de moda. Essa implementação aproxima o modelo preditivo de um sistema de suporte à decisão, em linha com propostas que defendem a integração entre modelos quantitativos e práticas de planejamento em varejo de moda (Koren; Shnaiderman, 2023; Nenni; Giustiniano; Pirolo, 2013; Thomassey, 2010).

Para avaliar a capacidade de generalização do modelo também em um cenário distinto daquele utilizado no treinamento, foi conduzida uma validação na nova coleção da Cienalab. Para isso, foi utilizada uma segunda base de dados, contendo o desempenho de venda dessas peças lançadas. Inicialmente, foram recalculados, nessa nova coleção, os indicadores de participação percentual das peças nas vendas nos primeiros 15 dias e no total da coleção. Em seguida, aplicaram-se as mesmas regras e os valores armazenados no modelo para reconstruir a variável *performance\_category* real da nova coleção. As variáveis categóricas de cor, materiais e estampas foram então mapeadas para códigos numéricos utilizando o dicionário de categorias salvo no arquivo do modelo. Categorias eventualmente não presentes na base original receberam um código específico, com o intuito de registrar a sua diferença em relação ao conjunto de treinamento.

Sendo assim, com a base de dados preparada com o conjunto da nova coleção, o modelo *Random Forest* foi utilizado para prever a categoria de desempenho de cada peça, e essas previsões foram comparadas com as classes reais construídas a partir dos indicadores de vendas. A avaliação foi realizada utilizando como base o cálculo da acurácia e a análise da matriz de confusão, observando assim a capacidade do modelo em separar corretamente as categorias de desempenho em um contexto temporalmente diferente daquele utilizado no treinamento. Essa forma de validação externa, em base de dados de outra coleção, complementa a validação

cruzada conduzida na fase de construção do modelo e segue recomendações da literatura para avaliação de modelos de previsão de demanda em ambientes de moda e varejo, nos quais a mudança de contexto entre coleções pode afetar o desempenho preditivo (Giri; Chen, 2022; Koren; Shnaiderman, 2023). Dessa maneira, a combinação entre a implementação da ferramenta e sua validação em nova coleção reforça o potencial de uso prático do modelo proposto como instrumento de apoio ao planejamento de produtos lançamento em marcas do segmento de moda *streetwear*.

## 4 RESULTADOS E DISCUSSÃO

### 4.1 Caracterização da base e da variável de desempenho

Conforme descrito na Seção 3.1, foi utilizada uma base de dados que organiza as camisetas da CienaLab em nível de produto, onde cada peça tem um conjunto de atributos visuais, de fabricação e comerciais. Foi construída a variável categórica *performance\_category* que a partir dos indicadores de participação percentual nos primeiros quinze dias de venda e no total do ciclo de vendas, classifica as camisetas em faixas de desempenho.

A Tabela 1 mostra os critérios utilizados e a distribuição das peças entre cada classe, seguindo os critérios apresentados na metodologia.

Tabela 1 – Critério de classificação das peças e distribuição das classes

Classe	Critério de classificação	Descrição	Contagem	Proporção
<i>best performer</i>	%vdacolecao15 > m <sub>1</sub> e %vdacolecaotot > m <sub>2</sub>	Alto desempenho no início e no total da coleção	16	32%
<i>hyped</i>	%vdacolecao15 > m <sub>1</sub> e %vdacolecaotot < m <sub>2</sub>	Forte início, mas queda no desempenho acumulado	5	10%
<i>normal</i>	%vdacolecao15 < m <sub>1</sub> e %vdacolecaotot < m <sub>2</sub>	Desempenho consistentemente abaixo da média	29	58%
<i>late performer</i>	%vdacolecao15 < m <sub>1</sub> e %vdacolecaotot > m <sub>2</sub>	Início fraco e crescimento posterior	0	0%
<b>Total</b>			<b>50</b>	<b>100%</b>

Fonte: Autoria própria (2025).

Legenda: m<sub>1</sub>: média de %vdacolecao15 da coleção; m<sub>2</sub>: média de %vdacolecaotot da coleção; %vdacolecao15: participação percentual da peça nas vendas da coleção nos primeiros 15 dias; %vdacolecaotot: participação percentual da peça no total de vendas da coleção.

Observa-se que grande parte das camisetas se concentram na classe *normal* (58%), representando as peças que servem de base de sustentação de cada coleção. Em contrapartida, a classe *best performer*, com 32% das observações, indicando um grupo relevante de produtos de alto desempenho. A classe *hyped*, embora menor que o restante, é muito característica do mercado, o que a torna interessante do ponto de vista gerencial, já que compreender a motivação do que pode ocasionar sua queda ao longo do ciclo é importantíssimo.

Essa distribuição confirma o comportamento esperado em coleções do segmento que é gerida pela prática dos *drops*, que consiste em poucas peças concentrarem a maior parte da atenção e demanda, enquanto a maior parcela da coleção cumpre uma função de suporte,

alicerce para as demais. A ausência da amostra late performer, corrobora com o perfil da empresa e seus consumidores, afinal, em ciclos de vida curtos há pouco tempo para que produtos inicialmente discretos se tornem relevantes no futuro.

## 4.2 Resultado dos modelos avaliados

Inicialmente foi realizada uma etapa de seleção baseado no pacote *LazyPredict*, que compara de forma rápida dezenas de algoritmos para modelos de classificação. Devido ao desbalanceamento de classes explicito em 4.1, seguiu-se com a escolha de *Balanced Accuracy* como variável de medida. Dessa forma, foi obtida a Tabela 2, filtrada pelos 5 melhores resultados preliminares obtidos através do *LazyPredict*.

Tabela 2 – Top 5 Modelos segundo LazyPredict (via *Balanced Accuracy*)

Modelo	<i>Balanced Accuracy</i>
GaussianNB	0,78
Random Forest	0,75
Extra Trees	0,72
Nearest Centroid	0,61
Ridge Classifier	0,61

Fonte: Autoria própria (2025).

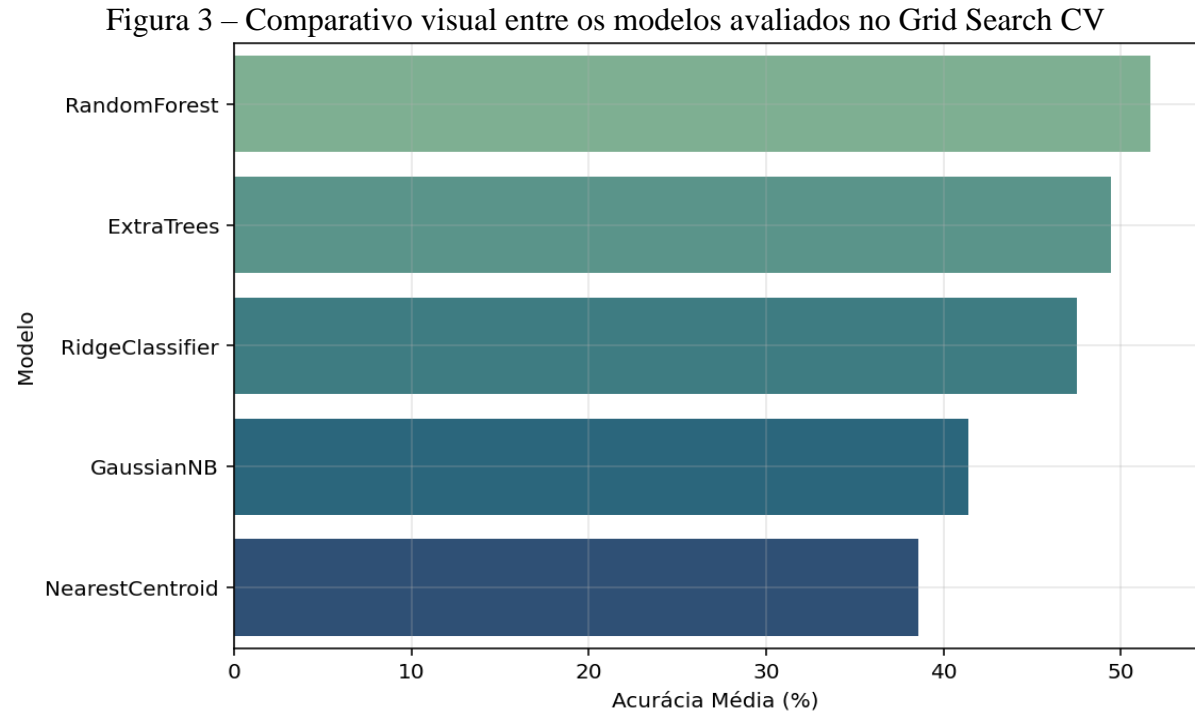
Em seguida, foi realizada o ajuste fino com o Grid Search CV (com 5 *folds*) para os modelos finalistas, buscando os melhores parâmetros possíveis para a base de dados, dessa forma, o Grid Search CV nos retornou a melhor combinação para os modelos e a acurácia alcançada, que pode ser vista na Tabela 3.

Tabela 3 – Resultados do Grid Search CV dos modelos finalistas

Modelo	Acurácia média (CV)	Melhores hiperparâmetros encontrados
Random Forest	0,5167	class_weight='balanced', max_depth=5, min_samples_leaf=4, min_samples_split=2, n_estimators=300
Extra Trees	0,4944	class_weight=None, max_depth=5, min_samples_split=5, n_estimators=100
Ridge Classifier	0,475	alpha=1.0, class_weight=None
Gaussian Naive Bayes	0,4139	(sem hiperparâmetros ajustáveis)
Nearest Centroid	0,3861	metric='manhattan', shrink_threshold=0.1

Fonte: Autoria própria (2025).

Como pode ser visto, embora o *Gaussian Naive Bayes* tenha obtido melhor desempenho inicial, o *Random Forest* apresentou melhor performance e estabilidade quando treinado com a ponderação de classes, sendo selecionado como o modelo final, o que justifica o uso do Grid Search CV como decisor da escolha do modelo final, a Figura 3 pode trazer esse resultado de forma visual comparando os modelos entre si.



Fonte: Autoria própria (2025).

4.3 Modelo final e seu desempenho

Após a etapa de otimização apontar o *Random Forest* como melhor modelo, ele foi treinado com os melhores hiperparâmetros e avaliado sobre a base de teste, que resultou na Tabela 4, que traz as métricas de classificação do modelo.

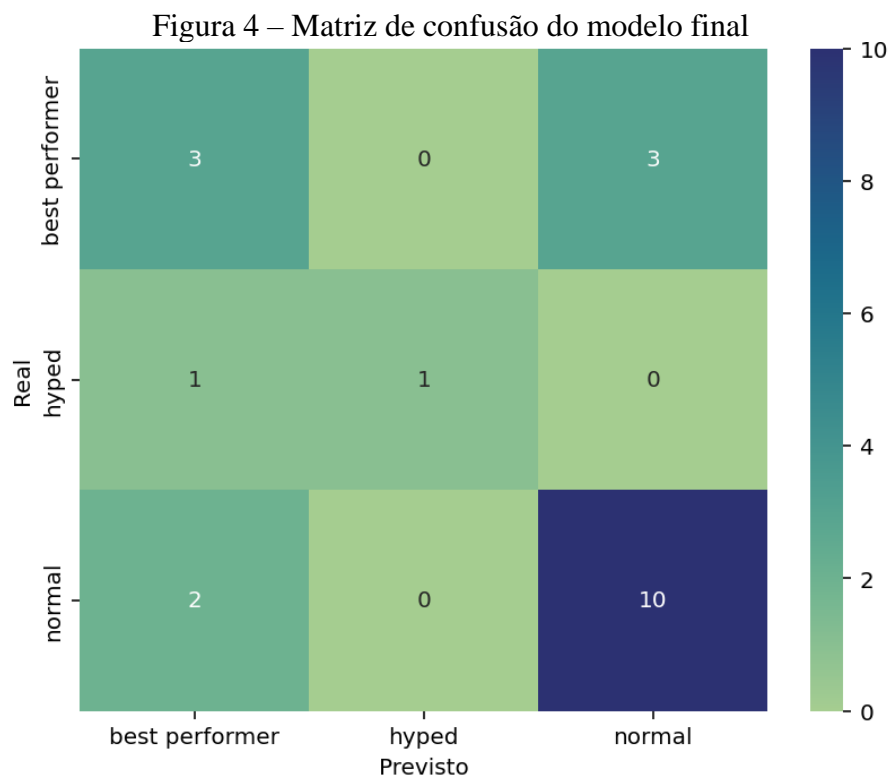
Tabela 4 – Resultados do Grid Search CV do modelo final				
Classe	Precision	Recall	F1-score	Suporte
best performer	0.50	0.50	0.50	6
hyped	1.00	0.50	0.67	2
normal	0.77	0.83	0.80	12
Acurácia global	—	—	0.70	20
Macro avg	0.76	0.61	0.66	20
Weiglited avg	0.71	0.70	0.70	20

Fonte: Autoria própria (2025).

#### 4.3.1 Análise das métricas de classificação

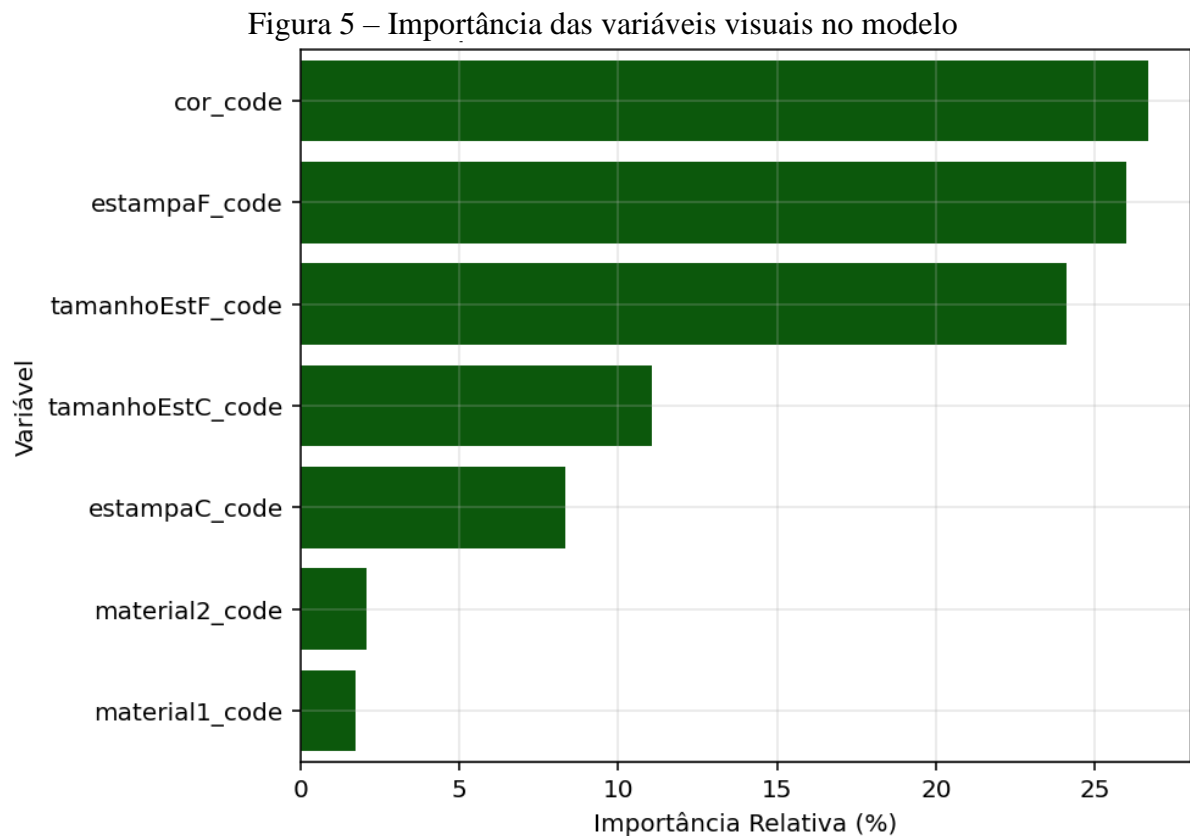
As métricas apresentadas na Tabela 5, revelam que o modelo *Random Forest* alcançou uma Acurácia Global de 70% sobre a base de teste de 20 amostras, mas devemos detalhar analisando classe a classe. O modelo demonstrou o melhor desempenho na classe “*normal*”, que é a majoritária, tendo um *F1-score* de 0,8. O desempenho mais fraco ocorreu na classe “*best performer*”, com precisão e recall de 0,5. Este resultado pode ser explicado como consequência do pequeno suporte de 6 amostras para essa classe.

Esses resultados também podem ser expressos visualmente pela matriz de confusão final do modelo, exemplificados através da Figura 4, que fornece insights sobre a natureza dos erros de classificação. Os principais desafios se encontram nas classes minoritárias, “*hyped*” e “*best performer*”, sendo uma confirmação do padrão de confusão típico em problemas com desbalanceamento de classes.



#### 4.3.2 Importância das variáveis

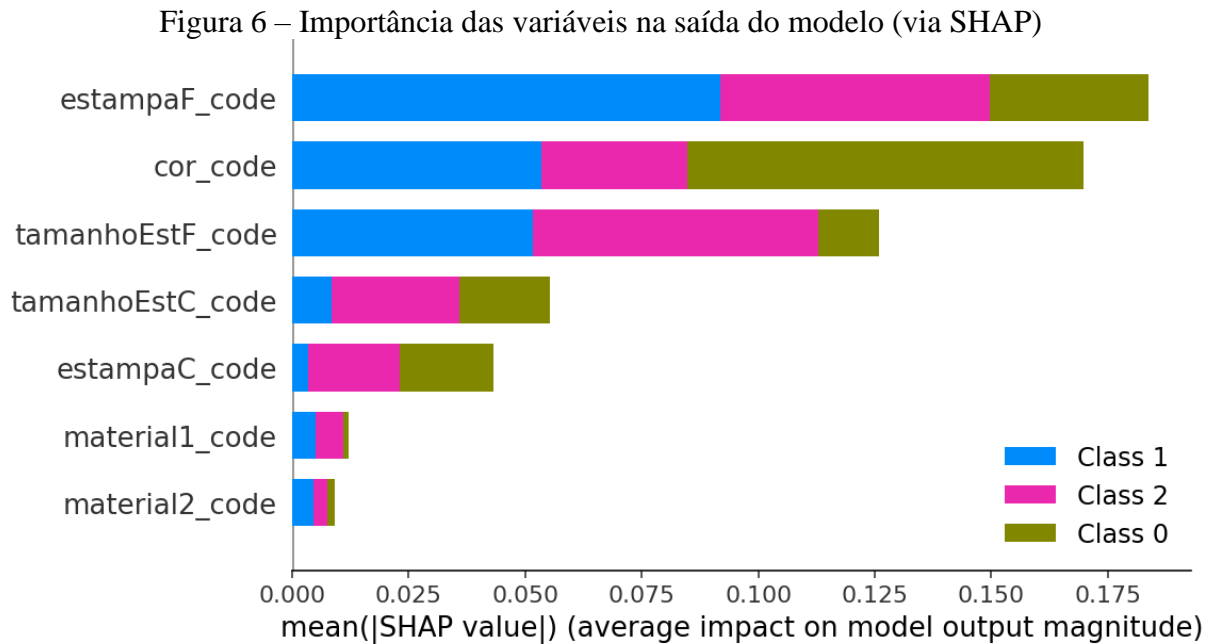
Uma das vantagens que o *Random Forest* traz é a possibilidade de quantificar a importância das variáveis, que mede o quanto cada *feature* contribuiu para o poder preditivo do modelo. A Figura 5 apresenta esse *ranking* de importância.



A análise revela que a cor, o posicionamento da estampa frontal e o tamanho da estampa frontal, contribuem com mais de 75% da importância total. Esse resultado confirma o que a literatura sobre a moda no contexto do *streetwear* e consumo da geração Z já aponta. A variável cor ser considerada a mais importante está alinhado com Yu *et al.* (2021) e Kuo, Chang e Lai (2022), que a destacam como o estímulo fundamental no julgamento estético. O tamanho e posicionamento da estampa também muito presentes na importância relativa do modelo reforçam seu papel como meio de comunicação social (Patel; Mehta, 2024).

### 4.3.3 Análise de explicabilidade com SHAP

Para aprofundar a análise de relevância, foi utilizado o método SHAP, uma técnica de XAI que fornecesse o valor de impacto de cada *feature* na previsão final para cada categoria (Figura 6).



Fonte: Autoria própria (2025).

É notável o contraste que existe entre o *ranking* da Importância Relativa, que classifica a Cor como o fator primário, e o *ranking* SHAP. Essa divergência é esperada e demonstra a robustez do SHAP: enquanto a Importância Relativa mede a utilidade da variável na redução de impureza (colocando a Cor em primeiro por ser fundamental nas divisões iniciais do *dataset*), o SHAP mede o impacto ponderado na probabilidade de classificação final. O fato de o código da estampa frontal ter subido para a primeira posição no SHAP indica que essa *feature* possui o maior poder preditivo líquido, sendo o elemento mais decisivo para modular o desempenho do produto nas diferentes categorias (“*normal*”, “*hyped*”, “*best performer*”). Este achado confirma que a Estampa é o principal driver de valor estético e simbólico para o consumidor de *streetwear*, conforme a literatura de moda e identidade.



#### 4.4 Teste do modelo em uma coleção lançada posteriormente

Para avaliar a capacidade de generalização do modelo em um contexto operacional, foi obtido junto aos diretores da CienaLab os dados das coleções mais recentes, Local Order pt. 1 e Local Order pt. 2, onde esta base de dados foi utilizada no modelo *Random Forest* aplicado ao conjunto de dados de treinamento. Essas duas coleções eram compostas por 12 novos produtos. O modelo manteve seus limites de vendas definidos pelas médias iniciais e histórias (15 dias e total) e classificou essas peças, onde posteriormente, essa classificação predita do modelo foi comparada com a classificação real do produto. O desempenho do modelo nessa nova amostra é detalhado na Tabela 5.

Tabela 5 – Desempenho do modelo nas coleções posteriores

Classe	Real	Previsto	Diferença
best performer	4	2	-2
hyped	2	1	-1
normal	6	9	3

Fonte: Autoria própria (2025).

A acurácia do modelo foi de 67%, o que demonstra que o modelo manteve um desempenho razoável, ligeiramente menor que o 70% observado na base de teste original. A análise por classe nos revela que o modelo exibiu um viés conservador, onde classificou normalmente todas as pessoas da classe “normal” corretamente e tendeu a incluir os “*best performer*” e “*hyped*” como “normal” também, como uma forma de minimizar o risco de falsos positivos. De mesma forma, o modelo se mostra confiável em prever as classes mais raras, tendo precisão de 100% para as classes “*best performer*” e “*hyped*”, significando que o modelo estava correto todas as vezes que previu essas classes.

Esse resultado se mostra coerente com o SHAP da seção 4.3.3, onde o sucesso da classe alvo, a “*best performer*”, depende de uma diferenciação estética entre estampas, cor e seu posicionamento. Embora o modelo tenha sucesso em prever os padrões gerais, a dificuldade em detectar os casos de “*best performer*” reflete a extrema sutileza e volatilidade do consumo da Geração Z. Isso sugere que a performance máxima depende de detalhes estéticos subjetivos e do contexto temático da estampa, que são fatores de difícil estruturação em variáveis quantitativas discretas.

## 4.5 Implementação prática: interface CustomTkinter

Com o objetivo de gerar um instrumento operacional acessível para o time criativo e comercial da CienaLab, foi desenvolvida uma aplicação na biblioteca CustomTkinter, que acessa as variáveis do modelo e prevê a categoria de desempenho de peças novas antes de seu lançamento. Essa interface serve para dar respostas rápidas sem precisar com que os dados estejam em base de dados, uma vez selecionados os campos na ferramenta, o modelo calcula com base no preenchimento das informações.

### 4.5.1 Tratamento de cor pelo espaço LAB

A interface permite que o usuário informe os valores de:

- L\*: Luminosidade;
- a\*: Verde-Vermelho;
- b\*: Azul-Amarelo.

Esses valores são convertidos via modelo perceptual proposto por Wang *et al.* (2006) em uma cor básica, reproduzido na Tabela 6.

Tabela 6 – Categorização de cores no espaço CIELAB

Stages	Basic colours	L*		C <sub>ab</sub> *		h <sub>ab</sub>	
		<	≥	≤	>	≤	>
1	White	100	92	5	0	-	-
	Grey	92	35	5	0	-	-
	Black	35	0	5	0	-	-
2	Brown	60	10	45	5	20	80
	Pink	100	50	85	5	320	5
	Yellow	100	83	-	-	85	105
3	Red	-	-	-	-	5	45
	Orange	-	-	-	-	45	80
	Green	-	-	-	-	80	210
	Blue	-	-	-	-	210	285
	Purple	-	-	-	-	285	5

Fonte: Wang *et al.* (2006).

No entanto, para classificar conforme Wang *et al.* (2006), as coordenadas cartesianas (a\*, b\*) necessitam de uma conversão para as coordenadas polares, onde obtém-se o Croma (C<sub>ab</sub>\*) que representa a saturação da cor, e o Ângulo de Matriz (H<sub>ab</sub>), que define a tonalidade. O sistema já realiza essa conversão automaticamente através das equações 1 (Cálculo do Croma) e 2 (Cálculo do Ângulo de Matiz), a seguir:

$$C_{ab}^* = \sqrt{(a^*)^2 + (b^*)^2} \quad (1)$$

$$h_{ab} = \arctan\left(\frac{b^*}{a^*}\right) \quad (2)$$

Como a base histórica da marca não contém todas as cores universais (por exemplo: Rosa, Laranja ou Amarelo), foi implementada uma lógica de mapeamento de cores próximas. Após o cálculo do LAB determinar para qual cor básica aquela camiseta está inserida, o sistema consulta uma lista de correspondências pré-definida pela equipe da marca, que traduz a cor universal para as categorias aceitas pelo modelo (Quadro 2). Sendo assim, o sistema não permite uma entrada arbitrária, mas restringe as opções para garantir que o valor inserido seja compatível com o do modelo, mantendo a coerência com o histórico de produtos da empresa.

Quadro 2 – Mapeamento de cores do modelo (escolhido pelo time criativo da CienaLab)

Cor básica identificada (Wang)	Opções de mapeamento no modelo
<i>White</i>	<i>Off-white</i> , Branco
<i>Grey</i>	Cinza, Preto, Branco
<i>Black</i>	Preto
<i>Brown</i>	Marrom, Cinza
<i>Red</i>	Vermelho
<i>Green</i>	Verde
<i>Blue</i>	Azul
<i>Purple</i>	Roxo, Azul, Vermelho
<i>Pink</i>	Vermelho, <i>Off-white</i>
<i>Yellow</i>	<i>Off-white</i> , Branco, Marrom
<i>Orange</i>	Vermelho, Marrom

Fonte: Autoria própria em parceria com o time criativo da CienaLab (2025).

#### 4.5.2 Travamentos de seleção

A interface incorpora a lógica de validação de dados e travamentos inteligentes para evitar o *garbage in, garbage out*. Esse tipo de lógica assegura que os atributos inseridos na interface respeitem as restrições lógicas do produto analisado:

- Se o atributo *estampaF\_code* for definido como “Sem Estampa”, o campo de *tamanhoEstF\_code* é automaticamente convertido para 0. O mesmo vale para as costas.
- Se o produto possui estampa, a opção de tamanho “0” é bloqueada de ser prevista.

- Os campos de material1 e material2 devem sempre estar presentes, garantindo que todas as *features* sejam preenchidas.
- A cor do modelo é sempre derivada do cálculo LAB, onde é selecionada a partir da lista disponibilizada para preenchimento.

#### 4.5.3 Previsão individualizada e cadastro de coleção

A interface foi projetada para centralizar todas as operações em uma única tela intuitiva. Conforme ilustrado na Figura 7, o painel principal apresenta campos de entrada para as coordenadas LAB, onde exibe automaticamente a cor básica associada ao modelo Wang e a lista suspensa de escolha da cor, além disso, aparecem também os menus de seleção para materiais e estampas.

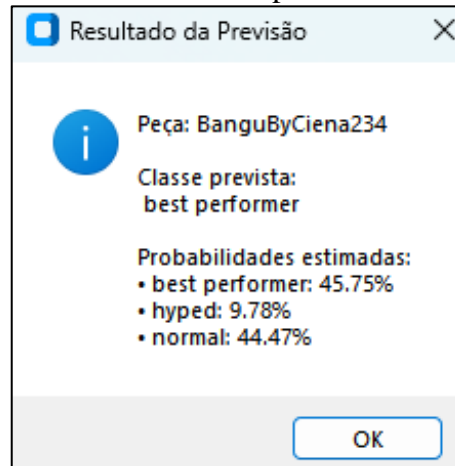
Figura 7 – Interface inicial da aplicação

Fonte: Autoria própria (2025).

A aplicação permite o uso de duas formas distintas. A primeira é a previsão individual, na qual o usuário preenche os dados de uma peça e solicita uma análise imediata (Figura 8). O

sistema processa essas informações e retorna um *pop-up* contendo a classe predita (ex.: “normal”).

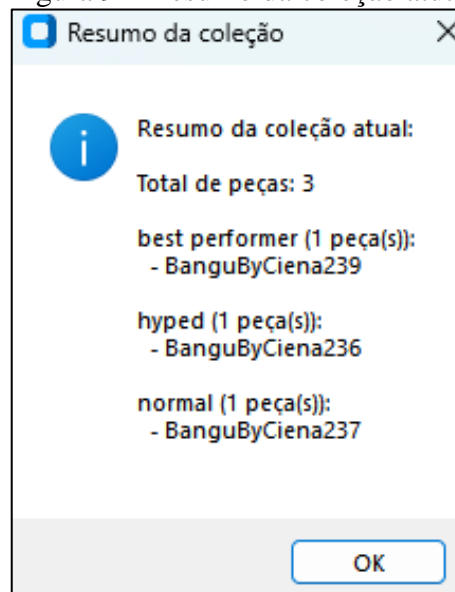
Figura 8 – Resultado de previsão individual



Fonte: Autoria própria (2025).

A segunda modalidade é a do resumo da coleção, onde o usuário pode ir agrupando as peças, adicionando essas peças a uma lista temporária. Ao comando do usuário, essa ferramenta gera um relatório consolidado que exhibe a contagem de peças em cada categoria e os SKUs inseridos em cada grupo, a Figura 9 apresenta esse resumo, que serve de instrumento de avaliação geral.

Figura 9 – Resumo da coleção atual



Fonte: Autoria própria (2025).

## 4.6 Implicações gerenciais e operacionais

A implementação do modelo e da interface introduz uma mudança qualitativa no processo de tomada de decisão da CienaLab. Até então baseada na intuição criativa e na análise empírica de vendas passadas, a definição passa a contar com um suporte analítico mais estruturado.

A capacidade de classificar antecipadamente os produtos nas variáveis de performance oferece diretrizes claras para a gestão de estoques e estratégias de *marketing*:

- 1) Otimização de Estoque: Para produtos identificados como “Best Performer”, a marca pode adotar uma estratégia de estoque mais agressiva, principalmente pelo fato do modelo ser mais conservador.
- 2) Alocação de Verba de *Marketing*: A identificação prévia dos produtos “*Hyped*” que possuem alto potencial de engajamento e desejo imediato, permite direcionar o orçamento de mídia paga especificamente para esses itens, potencializando o potencial de hype, onde a peça atrai tráfego para o site, impulsionando a vendas dos produtos de base.

Dessa forma, a ferramenta não substitui a visão criativa, que é a essência do *streetwear*, mas atua como uma bússola de risco, permitindo que a CienaLab concilie a inovação estética com a eficiência operacional necessária para que o negócio caminhe bem.

## 5 CONCLUSÕES

Este trabalho teve como objetivo principal desenvolver e avaliar um modelo preditivo do sucesso de novos produtos no contexto da marca carioca de *streetwear* CienaLab, com foco exclusivo na categoria de camisetas. Partiu-se do desafio de previsão de demanda em um ambiente marcado por coleções em *drops*, com seus ciclos de vida curtos, uma elevada volatilidade e com forte influência das redes sociais, cenário em que métodos tradicionais, baseados em séries longas e estáveis, tendem a enfrentar dificuldades. A proposta central foi investigar de que maneira atributos qualitativos relacionados ao design, como a cor, os materiais e as estampas, poderiam ser combinados aos indicadores quantitativos de vendas para apoiar nas decisões de planejamento criativo e produtivo de coleções futuras lançadas.

A pesquisa adotou uma abordagem aplicada, quantitativa e preditiva, estruturada como estudo de caso único na CienaLab. A partir de um banco de dados históricos de vendas em coleções anteriores, organizou-se um conjunto de informações dos produtos lançados, no qual cada camiseta é descrita por seus atributos visuais e fabris e por indicadores de desempenho de vendas em duas janelas, a primeira sendo a sua participação nas vendas da coleção nos primeiros quinze dias e a segunda no seu total do ciclo da coleção. Com base nesses indicadores apontados, foi construída uma variável categórica de desempenho, *performance\_category*, que permite classificar as peças em faixas como *best performer*, *hyped* e *normal*, aproximando o problema da previsão de demanda de um problema de classificação supervisionada multiclasse.

Sobre essa base, diferentes algoritmos de aprendizado de máquina foram testados e comparados entre si, e posteriormente com a seleção do modelo *Random Forest* ajustado por validação cruzada estratificada e busca em grade de hiper parâmetros. Em seguida, foram aplicadas técnicas de explicabilidade, XAI, em especial valores SHAP, para interpretar o modelo e analisar a contribuição dos atributos visuais para as previsões de demanda. Também foi desenvolvido um procedimento de mapeamento entre coordenadas de cor em CIELAB ( $L^*a^*b^*$ ) e as categorias discretas de cor utilizadas na base histórica, permitindo que o modelo seja utilizado de forma mais fiel ao processo de criação cotidiano da marca. Por fim, o classificador foi integrado a uma interface gráfica em Python e validado na coleção mais recente da CienaLab, chamada *Local Order*, avaliando seu potencial como ferramenta de apoio ao planejamento de produtos.

## 5.1 Atendimento dos objetivos

Como objetivo geral, este trabalho propôs desenvolver um modelo preditivo de sucesso de produtos lançamento com base em dados históricos reais da marca carioca de *streetwear* CienaLab, focado apenas em camisetas. O objetivo foi atendido por meio da construção do modelo *Random Forest*, treinado sobre a base organizada, e de sua posterior implementação em uma ferramenta interativa voltada ao uso gerencial cotidiano e para o uso comum dos funcionários da empresa.

Quanto aos objetivos específicos, é possível sintetizar o atendimento da seguinte forma:

- a) **Organizar e caracterizar o banco de dados das camisetas da CienaLab, classificando atributos visuais e fabris:** Esse objetivo foi contemplado na etapa de preparação dos dados, no qual o banco foi estruturado em nível de produto, com a sua identificação de variáveis, com a padronização de atributos visuais como cor, materiais, tipos e tamanhos de estampa e a definição dos indicadores de desempenho de vendas. A síntese das variáveis utilizadas foi apresentada na metodologia, juntamente com a descrição dos seus papéis no modelo proposto.
- b) **Analisar o comportamento de vendas das peças nos períodos de quinze dias e no acumulado total:** As duas janelas de desempenho foram utilizadas em conjunto para caracterizar o comportamento das camisetas ao longo do tempo. A partir delas, foi possível diferenciar peças com o seu impacto inicial elevado, peças que ganham tração de forma mais gradual e itens de desempenho mais discreto, o que fundamentou a construção da variável de desempenho utilizada como alvo do modelo.
- c) **Identificar correlações entre elementos estéticos (design) e resultados quantitativos (vendas):** Esse objetivo foi atendido de forma indireta por meio da modelagem preditiva. A utilização do *Random Forest*, combinada à análise de importância de variáveis e aos valores SHAP, permitiu verificar como que os atributos visuais e materiais utilizados nas peças contribuem para a classificação das peças em suas diferentes categorias de desempenho, indicando associações entre as decisões de design e os resultados encontrados de vendas.
- d) **Validar o modelo construído aplicando ao lançamento mais recente da marca:** Por fim, o modelo treinado sobre a base histórica foi aplicado a coleção mais recente da CienaLab, chamada de “*Local Order*”. Nessa etapa, a variável de desempenho foi



reconstruída para a nova coleção com base nos mesmos critérios utilizados no treinamento anterior e as previsões do modelo foram comparadas às categorias reais atribuídas às peças. Esse processo permitiu avaliar a capacidade de generalização do classificador proposto em um contexto temporalmente distinto, em linha com a proposta de utilizar o modelo como instrumento de apoio ao planejamento de futuras coleções para o time criativo e de produção.

## 5.2 Contribuições do estudo

Este trabalho apresenta contribuições em três dimensões principais: prática, metodológica e acadêmica. Do ponto de vista prático, o estudo oferece à CienaLab uma interface estruturada para apoiar decisões no desenvolvimento de camisetas em coleções futuras, de forma prática e de uso comum. A ferramenta construída no Capítulo 4 permite que a equipe de produto simule novas peças a partir de atributos de design e possa receber uma estimativa de desempenho antes do lançamento, auxiliando na composição da pirâmide de produtos e na priorização de cada um dos itens dentro da coleção. Ainda que não substitua o julgamento criativo e o acompanhamento de tendências, o modelo acrescenta uma camada de evidência quantitativa ao processo decisório.

No plano metodológico, o trabalho propõe uma integração entre diferentes componentes:

- a) Estruturação de uma base histórica em nível de produto com foco em atributos visuais;
- b) Formulação do problema de desempenho em termos de janelas de vendas iniciais e totais;
- c) Aplicação de um modelo de árvore de decisão em ensemble, o *Random Forest*, com uma seleção por validação cruzada;
- d) Uso de técnicas de explicabilidade, SHAP, para interpretar a contribuição das variáveis;
- e) Implementação de uma interface gráfica que operacionaliza o modelo no contexto da marca.

O procedimento realizado no mapeamento entre as cores definidas em CIELAB e categorias discretas da base histórica também se destaca como uma contribuição importante

para aproximar a linguagem mais técnica de design da realidade dos modelos preditivos encontrados.

Em termos acadêmicos, o estudo se insere na literatura de previsão de demanda e análise de desempenho em moda, com foco em um caso de *streetwear* brasileiro, no qual possui uma baixa literatura. Ao articular conceitos de comportamento do consumidor, identidade de marca, atributos de produto e aprendizado de máquina, o trabalho ilustra como técnicas quantitativas podem ser aplicadas mesmo em setores mais intensivos em seus simbolismos e estéticas com influência social. Além disso, ele reforça a importância de considerar atributos qualitativos, em destaque os elementos de design, na modelagem de desempenho de produtos, contribuindo para o debate sobre o uso de dados ricos em contexto principalmente para ambientes com uma alta volatilidade, como é o caso do segmento de moda *streetwear*.

### **5.3 Limitações do estudo**

Como todo estudo aplicado, este trabalho apresenta limitações que devem ser consideradas na interpretação dos resultados encontrados e em eventuais aplicações práticas do modelo em outras marcas.

A primeira limitação está relacionada ao escopo da base de dados inicial. O modelo foi treinado e validado exclusivamente com dados da CienaLab, em um recorte específico de tempo e com foco apenas na categoria de camisetas. Dessa forma, seus resultados não podem ser generalizados automaticamente para outras marcas, segmentos ou categorias de produto, nem para outros períodos muito distantes daqueles representados na base histórica.

Em segundo lugar, apesar de incorporar múltiplos atributos de design, o modelo trabalha apenas com um subconjunto das variáveis potencialmente interessantes para o desempenho das peças. Outros elementos como o preço, ações promocionais, clima, calendário de lançamentos, intensidade de divulgação paga e orgânica e outros fatores externos não foram explicitamente modelados. Parte desses efeitos podem ser capturados implicitamente pelos padrões de vendas presentes na base histórica, mas a ausência dessas variáveis limita a capacidade do modelo de explicar ou antecipar mudanças muito bruscas de comportamento ligadas a outros eventos externos.

Outra limitação é com relação ao tamanho e ao desbalanceamento das classes da variável de desempenho. Em bases de moda o número de observações por classe tende a ser um pouco mais reduzido, o que impacta a robustez estatística de alguns resultados e exige um maior

cuidado na interpretação dos seus desempenhos por classe. Embora tenham sido adotadas estratégias como o uso de *balanced accuracy* e a validação cruzada estratificada, trata-se de um problema inerente a esse tipo de contexto.

Por fim, a própria natureza do segmento de moda *streetwear*, altamente sensível a tendências e à dinâmica das redes sociais, impõe um horizonte de validade de tempo limitado aos modelos. A relevância dos padrões extraídos de coleções passadas tende a reduzir-se à medida que novas referências estéticas e de comportamentos surgem, o que implica na necessidade de atualização periódica da base de dados e um novo treinamento do modelo proposto para manter sua aderência à realidade.

Entretanto, apesar dessas limitações citadas, os resultados encontrados indicam que a integração entre os atributos de design e os dados históricos de vendas podem fornecer ferramentas de extrema importância para o planejamento de produtos em marcas do segmento *streetwear*. O modelo proposto e a ferramenta desenvolvida não substituem o olhar criativo e nem o acompanhamento cauteloso de tendências, mas se somam a essas dimensões como um novo apoio quantitativo, que pode ser capaz de reduzir incertezas em um ambiente tão volátil. Nesse sentido, o estudo reforça o potencial e a necessidade do uso de técnicas de aprendizado de máquina dentro da moda como um caminho interessante para aproximar a análise de dados das decisões gerenciais na prática criativa.

## REFERÊNCIAS

- ANITHA, S.; NEELAKANDAN, R. A demand forecasting model leveraging *machine learning* to decode customer preferences for new fashion products. **Complexity**, [Frisco, TX], v. 2024, n. 1, S425058, p. 1-10, July 2024. DOI: <https://doi.org/10.1155/2024/8425058>. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1155/2024/8425058>.
- AVOGARO, A. *et al.* Dif4FF: leveraging multimodal diffusion models and graph neural networks for accurate new fashion product performance forecasting. **Arxiv**, [Ithaca, NY], arXiv:2412.05566v1, 7 Dec. 2024. DOI: <https://doi.org/10.48550/arXiv.2412.05566>. Disponível em: <https://arxiv.org/abs/2412.05566>.
- BAUMAN, Z. **Modernidade líquida**. Rio de Janeiro: Zahar, 2001.
- BREIMAN, L. Random forests. **Machine Learning**, [Berlin], v. 45, n. 1, p. 5-32, Oct. 2001. DOI: <https://doi.org/10.1023/A:1010933404324>. Disponível em: <https://link.springer.com/article/10.1023/A:1010933404324>.
- CACHON, G. P.; SWINNEY, R. The value of fast fashion: quick response, enhanced design, and strategic consumer behavior. **Management Science**, [Rio de Janeiro], v. 57, n. 4, p. 778-795, Apr. 2011. DOI: <https://doi.org/10.1287/mnsc.1100.1303>. Disponível em: <https://pubsonline.informs.org/doi/10.1287/mnsc.1100.1303>.
- CHEN, I. F.; LU, C.-J. Demand forecasting for multichannel fashion retailers by integrating clustering and machine learning algorithms. **Processes**, [Basel], v. 9, n. 9, art. 1578, p. 1-14, Sept. 2021. DOI: <https://doi.org/10.3390/pr9091578>. Disponível em: <https://www.mdpi.com/2227-9717/9/9/1578>.
- CLOTTEY, S. N. T. *et al.* Impact of brand identity and brand image on fashion entrepreneurs: a case study of greater accra Metropolis. **International Journal of Vocational and Technical Education Research**, [London], v. 9, n. 2, p. 50-77, Fev. 2023. DOI: <https://doi.org/10.37745/ijvter.15/vol9n25077>. Disponível em: <https://eajournals.org/ijvter/vol-9-issue-2-2023/impact-of-brand-identity-and-brand-image-on-fashion-entrepreneurs-a-case-study-of-greater-accra-metropolis/>.
- ERWIN, E.; SAUNUNU, S. J.; RUKMANA, A. Y. The Influence of social media influencers on generation Z consumer behavior in Indonesia. **West Science Interdisciplinary Studies**, [Jakarta], v. 1, n. 10, p. 1040-1050, Oct. 2023. DOI: <https://doi.org/10.58812/wsis.v1i10.317>. Disponível em: <https://wsj.westsciences.com/index.php/wsis/article/view/317>.
- FERREIRA, P.; RODRIGUES, P.; RODRIGUES, P. Brand love as mediator of the brand experience-satisfaction-loyalty relationship in a retail fashion brand. **Management and Marketing**, [São Paulo], v. 14, n. 3, p. 278-291, Oct. 2019. DOI: <https://doi.org/10.2478/mmcks-2019-0020>. Disponível em: <https://reference-global.com/article/10.2478/mmcks-2019-0020?medium=article>.

GIRI, C.; CHEN, Y. Deep learning for demand forecasting in the fashion and apparel retail industry. **Forecasting**, [Basel], v. 4, n. 2, p. 565-581, June 2022. DOI: <https://doi.org/10.3390/forecast4020031>. Disponível em: <https://www.mdpi.com/2571-9394/4/2/31>.

GRIGOREVA, E. A.; GARIFOVA, L. F.; POLOVKINA, E. A. Consumer behavior in the information economy: generation Z. **International Journal of Financial Research**, [Ontario], v. 12, n. 2, p. 164-171, Jan. 2021. DOI: <https://doi.org/10.5430/ijfr.v12n2p164>. Disponível em: <https://ideas.repec.org/a/jfr/ijfr11/v12y2021i2p164-171.html>.

GUO, C.; BERKHAHN, F. Entity embeddings of categorical variables. **Arxiv**, [Ithaca, NY], arXiv:1604.06737v1, 22 Apr. 2016. DOI: <https://arxiv.org/abs/1604.06737>. Disponível em: <https://arxiv.org/abs/1604.06737>.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning: data mining, inference, and prediction**. 2. ed. Berlin: Springer, 2009.

HE, W. *et al.* Mask R-CNN. **Arxiv**, [Ithaca, NY], arXiv:1703.06870v3, 24 Jan. 2018. DOI: <https://arxiv.org/abs/1703.06870v3>. Disponível em: <https://arxiv.org/abs/1703.06870>.

KIM, J.; HUANG, O. Transformative potentials of drop marketing and drop retail strategies in the fashion industry. **International Textile and Apparel Association Annual Conference Proceedings**, [Cary, NC], v. 77, n. 1, p. 1-3, Dec. 2020. DOI: <https://doi.org/10.31274/itaa.11791>. Disponível em: <https://www.iastatedigitalpress.com/itaa/article/id/11791/>.

KOREN, M.; SHNAIDERMAN, M. Forecasting in the fashion industry: a model for minimising supply-chain costs. **International Journal of Fashion Design, Technology and Education**, [Abingdon, UK], v. 16, n. 3, p. 308-318, Apr. 2023. DOI: <https://doi.org/10.1080/17543266.2023.2201508>. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/17543266.2023.2201508>.

KUNZ, M. *et al.* Deep learning based forecasting: a case study from the online fashion industry. **Arxiv**, [Ithaca, NY], arXiv:2305.14406v1, 23 May 2023. DOI: <https://doi.org/10.48550/arXiv.2305.14406>. Disponível em: <https://arxiv.org/abs/2305.14406>.

KUO, L.; CHANG, T.; LAI, C.-C. Affective psychology and color display of interactive website design. **Displays**, [New York], v. 71, art. 102134, Jan. 2022. DOI: <https://doi.org/10.1016/j.displa.2021.102134>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0141938221001347>.

LUNDBERG, S.; LEE, S.-I. A unified approach to interpreting model predictions. **Arxiv**, [Ithaca, NY], arXiv:1705.07874v2, 25 Nov. 2017. DOI: <https://doi.org/10.48550/arXiv.1705.07874>. Disponível em: <https://arxiv.org/abs/1705.07874>.

MICCI-BARRECA, D. A preprocessing scheme for high-cardinality categorical attributes in classification and prediction problems. **ACM SIGKDD Explorations Newsletter**, [New York], v. 3, n. 1, p. 27-32, July 2001. DOI: <https://doi.org/10.1145/507533.507538>. Disponível em: <https://dl.acm.org/doi/10.1145/507533.507538>.

NENNI, M. E.; GIUSTINIANO, L.; PIROLO, L. Demand forecasting in the fashion industry: a review. **International Journal of Engineering Business Management**, [New York], v. 5, p. 1-6, Jan. 2013. DOI: <https://doi.org/10.5772/56840>. Disponível em: <https://journals.sagepub.com/doi/10.5772/56840>.

PATEL, A.; MEHTA, A. Gen Z reshaping the streetwear sneaker market. **GLS Kalp: Journal of Multidisciplinary Studies**, [Ahmedabad], v. 1, n. 2, p. 52-56, Apr./June 2024. DOI: <https://doi.org/10.69974/glskalp.01.02.55>. Disponível em: <https://www.glskalp.in/index.php/GLSKALP/article/view/55>.

RIBEIRO, M. T.; SINGH, S.; GUESTIN, C. “Why should i trust you?” explaining the predictions of any classifier. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 22., 2016, San Francisco, CA. **Proceedings** [...]. New York: Association for Computing Machinery, 2016. p. 1135-1144. Disponível em: <https://dl.acm.org/doi/10.1145/2939672.2939778>.

SHI, X.; LI, F.; CHUMNUMPAN, P. The use of product scarcity in marketing. **European Journal of Marketing**, [Leeds], v. 54, n. 2, p. 380-418, Jan. 2020. DOI: <https://doi.org/10.1108/EJM-04-2018-0285>. Disponível em: <https://www.emerald.com/ejm/article-abstract/54/2/380/72663/The-use-of-product-scarcity-in-marketing?redirectedFrom=fulltext>.

SUBRAMANIAM, A.; RATHNAPRIYA, B. Customer preference towards nostalgic products. **Journal of Lifestyle and SDGs Review**, [São Paulo], v. 4, n. 4, e03622, p. 1-12, nov. 2024. DOI: <https://doi.org/10.47172/2965-730X.SDGsReview.v4.n04.pe03622>. Disponível em: <https://sdgsreview.org/LifestyleJournal/article/view/3622>.

SYNTETOS, A. A. *et al.* Supply chain forecasting: theory, practice, their gap and the future. **European Journal of Operational Research**, [Amsterdam], v. 252, n. 1, p. 1-26, July 2016. DOI: <https://doi.org/10.1016/j.ejor.2015.11.010>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0377221715010231>.

THOMASSEY, S. Sales forecasts in clothing industry: the key success factor of the supply chain management. **International Journal of Production Economics**, [Amsterdam], v. 128, n. 2, p. 470-483, Dec. 2010. DOI: <https://doi.org/10.1016/j.ijpe.2010.07.018>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0925527310002598>.

TOCHE TCHIO, G. M. *et al.* A comprehensive review of supervised learning algorithms for the diagnosis of photovoltaic systems, proposing a new approach using an ensemble learning algorithm. **Applied Sciences**, [Basel], v. 14, n. 5, art. 2072, p. 1-29, Mar. 2024. DOI: <https://doi.org/10.3390/app14052072>. Disponível em: <https://www.mdpi.com/2076-3417/14/5/2072>.

WANG, Z. *et al.* An algorithm for categorising colours into universal colour names. In: CONFERENCE ON COLOUR IN GRAPHICS IMAGING AND VISION, 3., 2006, Washington, DC. **Proceedings** [...]. Washington, DC: Society for Imaging Science and Technology, 2006. p. 426-430. Disponível em: <https://doi.org/10.2352/CGIV.2006.3.1.art00087>.

WOOD, S. **Generation Z as consumers: trends and innovation**. Raleigh, CA: Institute for Emerging Issues, 2013.

YU, X. *et al.* A new form of brand experience in online social networks: an empirical analysis. **Journal of Business Research**, [New York], v. 130, p. 426-435, June 2021. DOI: <https://doi.org/10.1016/j.jbusres.2020.02.011>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0148296320301053>.