

Internet of Reviews: um dataset de características e avaliações de objetos inteligentes no mercado brasileiro

João G. A. Bradachi^{1,2}, Philipe de Freitas Melo¹, Thais R. M. B. Silva^{1,2}

¹Instituto de Ciências Exatas e Tecnológicas
Universidade Federal de Viçosa (UFV) – Florestal – MG

²Manna Team

{joao.bradachi, philipe.freitas, thais.braga}@ufv.br

Abstract. *With the expansion of the Internet of Things (IoT), analyzing the consumer journey becomes strategic. This paper presents the construction and analysis of a robust dataset in Portuguese, integrating reviews (post-sale) and questions (pre-sale) from the smart object e-retail market. By applying NLP techniques, we identify an asymmetry within this ecosystem: while the overall reception is positive, frustrations are concentrated on software and connectivity challenges, which outweigh hardware failures. The analysis of questions further reveals deep pre-purchase insecurity regarding interoperability, indicating that the lack of standardization and clarity in product specifications acts as the primary barrier to IoT technology adoption.*

Resumo. *Com a expansão da Internet das Coisas (IoT), a análise da jornada do consumidor torna-se estratégica. Este trabalho apresenta a construção e análise de um dataset robusto em português, que integra tanto o pós-venda (avaliações) e pré-venda (perguntas) do varejo eletrônico de objetos inteligentes. Aplicando técnicas de PLN, identificamos uma assimetria deste ecossistema: enquanto a recepção geral é positiva, as frustrações concentram-se em desafios de software e conectividade, que superam falhas de hardware. A análise das perguntas revela ainda uma profunda insegurança prévia sobre interoperabilidade, indicando que a falta de padronização e clareza nas especificações dos produtos atua como a principal barreira para a adoção da tecnologia IoT.*

1. Introdução

A Internet das Coisas (IoT) tem mudado cada vez mais nossa rotina ao interligar objetos comuns, equipados com sensores, software e conectividade que coletam, trocam e agem sobre dados pela internet, chamados frequentemente de objetos inteligentes (Kortuem et al. 2009). O mercado global de IoT teve uma rápida expansão, com US\$ 139,51 bilhões em 2024 e em plena expansão (Mordor Intelligence 2025). A IoT tem sido considerada a base da transformação digital global, incentivada por tecnologias adjacentes, desde a integração do 5G, soluções de água inteligente até inovações na área da saúde, como cirurgiões robôs e hospitais inteligentes (da Silva and Baranauskas 2023).

Impulsionada por seu crescimento exponencial (Fabri et al. 2025), observa-se uma proliferação de dispositivos de IoT e “objetos inteligentes” nas plataformas de comércio eletrônico. Neste cenário, conteúdo gerado por usuários consolidou-se

como um indicador fundamental da percepção e do sentimento dos consumidores (Xiao et al. 2021), exercendo um papel importante nos processos de decisão de compra. Compreender as dificuldades e indignações expressas por avaliações permite que as empresas aprimorem seus serviços e identifiquem desafios e oportunidades deste ecossistema complexo, como por exemplo a previsão de *churn* (Beckhauser and Fileto 2024). Isso inclui desde a clareza nas especificações anunciadas até a eficiência no pós-venda, a fidelização e a redução de custos operacionais com devoluções e suporte (Sebrae 2022). Entretanto, pesquisadores e desenvolvedores carecem de dados acerca de como tais produtos são recebidos e utilizados pelo público geral, dificultando o avanço da área de IoT.

Com base nisto, o objetivo deste artigo é construir um conjunto de dados sólidos sobre produtos considerados como objetos inteligentes e, a partir dele, discutir o panorama da satisfação e as principais barreiras de adoção enfrentadas pelos consumidores de objetos inteligentes no varejo no contexto latino-americano. O estudo é norteado pelas seguintes questões de pesquisa: **Q1:** É possível construir uma base de dados robusta e em larga escala de avaliações reais de objetos inteligentes na América Latina? **Q2:** Como e quais os produtos caracterizados como inteligentes estão sendo consumidos pelo público? **Q3:** Quais os principais desafios e oportunidades enfrentados pelos produtos IoT?

Para responder a essas questões, este trabalho propõe uma abordagem automatizada que combina uma coleta de dados sólida com técnicas de Processamento de Linguagem Natural (PLN) para construção de um dataset inovador na área de IoT. Nossa coleta de dados foi estruturada em duas etapas, com o uso de uma API para caracterização dos produtos e o desenvolvimento de um *Web crawler* via Selenium para a extração de perguntas e avaliações. Em seguida os dados foram enriquecidos por meio de técnicas de análise de sentimento e modelagem de tópicos (BERTopic).

Nossos dados e resultados, além de evidenciar padrões de consumo de objetos inteligentes no contexto latino-americano, revelam que, embora a recepção deste tipo de produto seja majoritariamente positiva (76,33%), a experiência do usuário é frequentemente comprometida por desafios de configuração de software e falhas de pareamento na conectividade, que superam em volume as queixas sobre defeitos físicos de hardware. A etapa de pré-compra é marcada por uma grande insegurança técnica quanto à interoperabilidade dos produtos, evidenciando uma oportunidade de aprofundamento dos dados para melhorar a adoção plena da IoT pelos consumidores latino-americanos.

2. Trabalhos Relacionados

As avaliações de produtos em plataformas de comércio eletrônico constituem uma rica fonte de dados para o aprimoramento de dispositivos e estratégias comerciais (Rohman et al. 2020). Nesse cenário, técnicas de PLN e mineração de texto tornaram-se essenciais para capturar as sutilezas semânticas do usuário (Barbosa 2018). Diversos estudos demonstram que a polaridade de sentimento predominante em grandes volumes de *reviews* é um indicador confiável de aceitação de mercado (Brownfield and Zhou 2020; Ojeda et al. 2024; Wang et al. 2025). Contudo, a análise de sentimentos isolada, ao nível de documento, pode falhar em detectar insatisfações pontuais (Usmani et al. 2021). Um consumidor pode expressar satisfação geral, mas reportar problemas críticos em componentes específicos, tornando a extração de tópicos uma camada essencial, frequentemente abordada pela análise de sentimentos ao nível de aspecto (Tripathy and Sharaff 2026).

No contexto brasileiro, o trabalho de (Santos et al. 2024) propôs uma metodologia de análise de pós-venda em *marketplaces*. A abordagem proposta combina técnicas de *embeddings* textuais, como o uso do BERTopic; análise de sentimentos com o modelo BERTimbau e extração de entidades nomeadas via modelo RoBERTa para identificar tendências e entidades nomeadas em produtos e serviços. Um dos principais desafios consiste na busca por conjuntos de dados confiáveis e de alta qualidade em língua portuguesa para validação da metodologia. Trabalhos como os de (de Melo 2024) e (Siqueira et al. 2024) apresentam *datasets* abrangentes, adotando técnicas de PLN na análise exploratória. Contudo, esses estudos abordam domínios distintos, focando, respectivamente, em restaurantes brasileiros estrelados pelo Guia Michelin e em comentários de softwares de dispositivos móveis. O estudo de (Santana et al. 2025), por sua vez, reuniu informações de objetos inteligentes extraídas manualmente de *marketplaces* e da literatura, disponibilizando-os em uma ontologia específica.

Embora técnicas de PLN sejam eficazes para produtos gerais de consumo massivo, objetos inteligentes IoT apresentam desafios específicos de conectividade, segurança e interoperabilidade que exigem abordagens dedicadas. Como observado em seu trabalho de *Smart Homes* por (Yu et al. 2024), em avaliações de robôs aspiradores da Amazon, a satisfação tende a ser impulsionada por funcionalidades (“inteligência”), enquanto a insatisfação é causada por fatores operacionais, como falhas de conectividade e suporte ineficiente. Já o estudo de (Chanda et al. 2024), por exemplo, mapeou a “jornada de frustração” na configuração de dispositivos IoT, identificando que problemas de usabilidade na fase de *setup* e pareamento Wi-Fi são barreiras críticas de adoção. Similarmente, (Truelove et al. 2019) analisaram 90 mil avaliações de dispositivos IoT e concluíram que a instabilidade de software e rede são problemas frequentes. Além disso, a heterogeneidade de protocolos introduz riscos de “*Shadow IoT*” e falhas de segurança que permeiam as discussões de usuários, como destacado por (Protick et al. 2026) ao analisarem preocupações de privacidade em *reviews* de câmeras e assistentes de voz.

Apesar dos avanços citados, observa-se uma lacuna significativa na literatura: a ausência de bases de dados robustas, em larga escala e em língua portuguesa, que capturem a jornada completa do consumidor de IoT de forma geral, desde as dúvidas pré-compra até a experiência de uso (pós-venda). Enquanto a maioria dos trabalhos foca em produtos genéricos fora do escopo ou muito específicos, como celulares e caixas de som, este estudo diferencia-se ao disponibilizar um grande dataset integrado que abrange milhares de objetos inteligentes diferentes. Além disso, esse trabalho faz uma investigação aprofundada do panorama do consumo de produtos IoT no varejo latino-americano.

3. Materiais e Métodos

O desenvolvimento deste estudo se fundamenta na extração, consolidação, tratamento e análise de dados em larga escala. Conforme ilustrado na Figura 1, o processo se divide em quatro etapas principais: (i) Coleta de dados via API e automação; (ii) Caracterização exploratória; (iii) Filtragem e seleção de objetos inteligentes; (iv) Processamento de Linguagem Natural (PLN).

3.1. Coleta de Dados

Para viabilizar a construção do dataset, foi realizada uma estratégia de coleta em duas etapas complementares: a primeira utilizou interfaces oficiais de programação (API) para

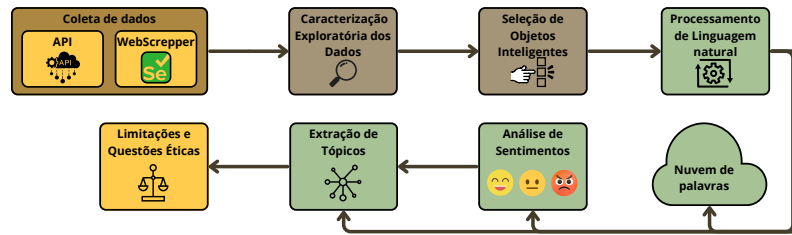


Figura 1. Metodologia do trabalho.

o levantamento dos metadados técnicos dos objetos inteligentes presentes na plataforma, enquanto a segunda fase empregou automação de navegação para a extração de interações (perguntas e avaliações).

Fase 1: Mapeamento via API e Palavras-Chave: Inicialmente, para identificar produtos IoT, foi realizado um levantamento de palavras-chave específicas do domínio, obtidas a partir do estudo de (Santana et al. 2025), que mapeou diversas instâncias de objetos inteligentes e as modelou na rede semântica OD4CoT. Adicionalmente, também foram acrescentados manualmente alguns termos semelhantes para enriquecer as buscas¹. Posteriormente, foi utilizado o *endpoint* da API do *e-commerce* de varejo para buscar os produtos propriamente ditos, baseado nos termos definidos. Os dados retornados pelas API são estruturados no formato JSON com campos que descrevem o produto, como seu nome e ID na plataforma; data de criação e o status do anúncio; o id de domínio; e atributos. Estes atributos, por sua vez, são objetos JSON que compreendem os Contextos das Coisas, como descrito por (Santana et al. 2025).

Fase 2: Extração de dados via WebScrapper: A partir de cada um dos IDs obtidos na etapa anterior, foi feito o mapeamento das avaliações de cada um dos produtos coletados. A vantagem da plataforma usada é a presença de um campo de perguntas e respostas (Q&A) dos produtos, também coletado por este trabalho. Este dado também permite, portanto, analisar tanto a pós-vendas (avaliações) dos produtos, como também os dados de pré-venda. Para isso, foi necessário implementar um Web Crawler usando o *framework* Selenium para automatizar a navegação no site e coletar os dados faltantes, uma vez que o *endpoint* da API não retornava tais informações. Uma vez acessada a página de cada produto, a automação de coleta de dados era executada. O processo consiste em acessar o *modal* das avaliações e extrair trechos HTML que contêm os dados requisitados. O mesmo procedimento foi replicado para a extração das perguntas e respostas.

3.2. Filtragem dos Objetos Inteligentes

Dada a natureza abrangente do termo “inteligente” no comércio online, que frequentemente é empregado para descrever atributos de design ergonômico, versatilidade ou simplesmente como marketing para chamar atenção de um produto, em detrimento da conectividade digital, foi necessária uma etapa de validação dos dados ao paradigma da IoT. Dessa forma, foi implementado um filtro de validação baseado na presença explícita de especificações técnicas de rede para evitar a inclusão de itens como “guarda-roupas inteligentes”. Para que um objeto fosse mantido no dataset final, o algoritmo de filtragem exigiu a confirmação de ao menos um protocolo de conectividade, (i.e. Wi-Fi, Bluetooth ou interfaces de conexão cabeada). Este procedimento de limpeza foi fundamental para garantir a integridade semântica do corpus e a precisão das análises subsequentes.

¹ A lista final das palavras-chave está disponível no repositório do projeto.

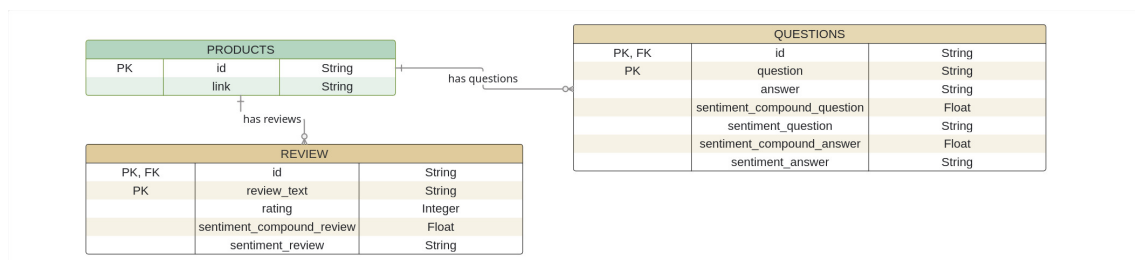


Figura 2. Modelo relacional do conjunto de dados

3.3. Descrição do Dataset

A coleta, após a aplicação da filtragem, resultou em 5.094 objetos distintos. Os dados coletados foram estruturados pela *URL* dos produtos, com ele é possível obter os nomes dos produtos, identificadores na plataforma, códigos de domínio e palavras-chave. Ademais, é possível extrair o conjunto de características dos produtos, o qual abrange, ao todo, 108.637 ocorrências distribuídas em 633 categorias distintas. Um diferencial deste conjunto de dados foi a coleta das Avaliações somada às Perguntas/Respostas dos usuários, que resultou em 3.093 avaliações e 8.877 perguntas distintas. Os registros das avaliações são compostas pelo comentário textual e uma avaliação quantitativa variando de 0 à 5. Já as perguntas são compostas por um conteúdo textual e a resposta associada. Enquanto as avaliações nos permite investigar as opiniões pós-venda, o acréscimo de perguntas nos oferece mais uma camada de análise, permitindo um aprofundamento na pre-venda dos produtos.

O conjunto de dados foi estruturado conforme o modelo apresentado na Figura 2. Nele, é possível identificar as três tabelas que compõem o *dataset*, bem como os nomes das colunas e seus respectivos tipos de dados. Além disso, mapeamos os relacionamentos entre as tabelas com base no id do produto e na principal informação armazenada. O *dataset* está disponível em <https://zenodo.org/records/20964552>.

3.4. Processamento de Linguagem Natural

A fim de aprofundar a investigação dos dados e endereçar as questões de pesquisa, este estudo propõe ainda três etapas de enriquecimento dos dados para analisar os itens com técnicas de Processamento de Linguagem Natural (PLN), incluindo uma análise exploratória léxica com nuvens de palavras; a classificação de sentimentos nos dados textuais; e extração de tópicos utilizando *embeddings* semânticos.

Análise de sentimentos: Visando determinar a aceitação dos produtos no mercado (Q1), os dados foram submetidos à análise de sentimentos para inferência de emoções e atitudes. Para a execução desta etapa, foi utilizada a ferramenta LeIA (*Léxico para Inferência Adaptada*) (Almeida 2018), uma adaptação para o português do léxico e analisador VADER (*Valence Aware Dictionary and sEntiment Reasoner*) (Gilbert 2014). O modelo foi selecionado por sua eficácia em textos curtos e informais de redes sociais e *e-commerce*, permitindo classificar as interações em polaridades positiva, neutra e negativa.

Modelagem de tópicos com *embeddings*: Por fim, para compreender os principais temas discutidos pelos usuários, realizamos a extração de tópicos a partir do algoritmo BERTopic (Grootendorst 2022), que utiliza arquiteturas baseadas em Transformers para criar representações densas de texto. Para garantir a precisão semântica em português, aplicou-se o modelo de *embedding paraphrase-multilingual-MiniLM-L12-v2* (Reimers and Gurevych 2019), seguido pelas etapas de redução de dimensionalidade

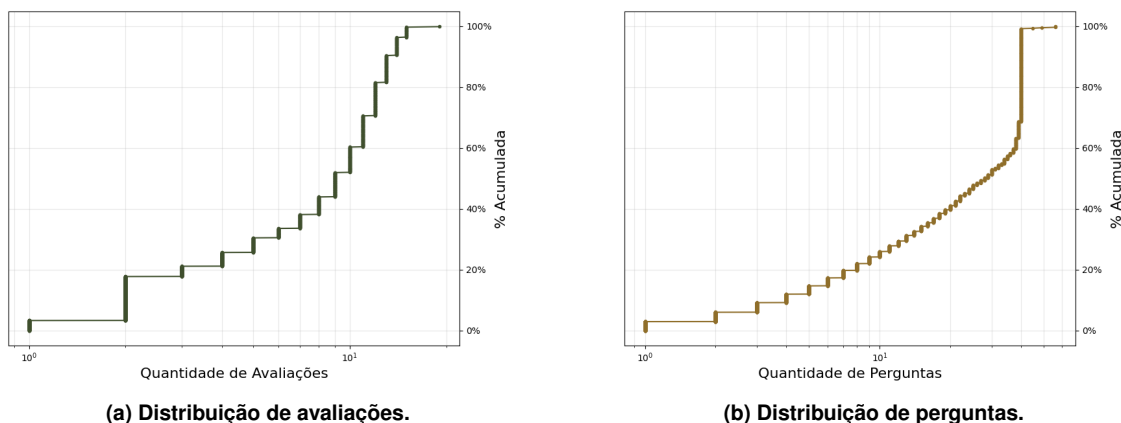


Figura 3. CDF da densidade de interações (avaliações e perguntas) por objeto.

com UMAP e de clusterização com o HDBSCAN. Esta extração semântica foi aplicada individualmente sobre os subconjuntos de sentimentos positivos e negativos. O modelo foi configurado para gerar até 40 tópicos distintos, dos quais os cinco mais relevantes de cada categoria foram analisados mais profundamente.

4. Caracterização Exploratória dos Dados e Discussão

Na primeira fase da análise dos resultados, focamos na viabilidade de construção do *dataset* (Q1) e na identificação dos produtos consumidos (Q2). A robustez da base, composta de mais de 5 mil objetos, é evidenciada também pelas quantidades de avaliações e perguntas totais exibidas pelas Figuras 3a e 3b, que mostram uma distribuição acumulada típica de grandes conjunto de dados de *marketplaces*. Observa-se que apenas 20% dos produtos possuem duas avaliações ou menos. A maioria dos objetos possui até 10 avaliações (60%), mostrando um bom balanceamento dos dados entre os itens coletados e 40% possuem mais do que isso. Já a Figura 3b mostra um volume ainda maior de perguntas, com mais da metade dos produtos tendo mais de 30 perguntas cada.

Em seguida, a Figura 4 detalha as características mais presentes nos produtos totais coletados, revelando que, além de identificadores básicos (marca e modelo) e atributos físicos (cor, tamanho, peso), a conectividade (*Wi-Fi* e *Bluetooth*) emerge como o principal diferencial técnico declarado, justificando a filtragem deste tipo de produto para refinamento dos dados IoT. Esses dados evidenciam os principais tipos de metadados cadastrados para objetos inteligentes, embora seja possível notar que nem todos os objetos possuem suas características completas cadastradas na plataforma.

Ao mapear a relação de objetos por categorias (i.e. domínios do *e-commerce* de varejo) na Figura 5, observa-se que o mercado de IoT é liderado por objetos do tipo *Smart Speakers*, Vestíveis (*Smartwatches*), Iluminação (*Light Bulbs*) e Segurança (*Home Alarms*). Este panorama responde à Q2, indicando uma concentração de consumo em dispositivos de entrada mais básicos e de uma automação residencial mais simplificada. De forma complementar, vimos também uma extensa lista de objetos, desde ar-condicionado até *Air Fryer* e chuveiros inteligentes, sugerindo um maior nível de automatização residencial.

A análise temporal (Figura 6) sugere um interesse por IoT fortemente sazonal. O pico de interações em novembro (salto de 207 para 1.840 avaliações e 1.376 para 3.626 para perguntas) durante o período da coleta coincide com a *Black Friday*, o que sugere que a barreira de preço ainda é um fator determinante para a adoção desses objetos. Ao obser-

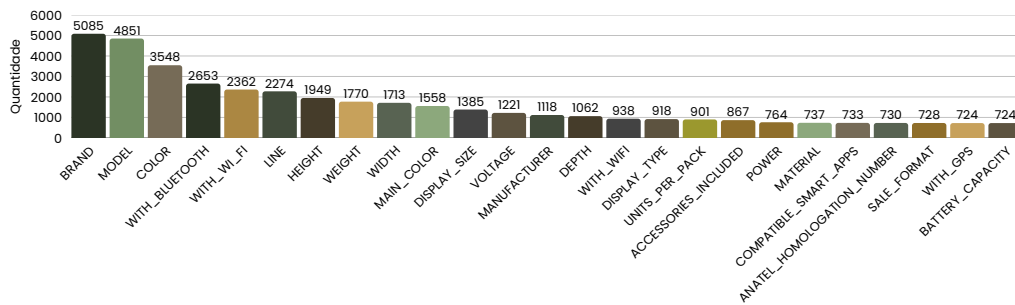


Figura 4. Top 25 atributos técnicos mais presentes nos objetos coletados.

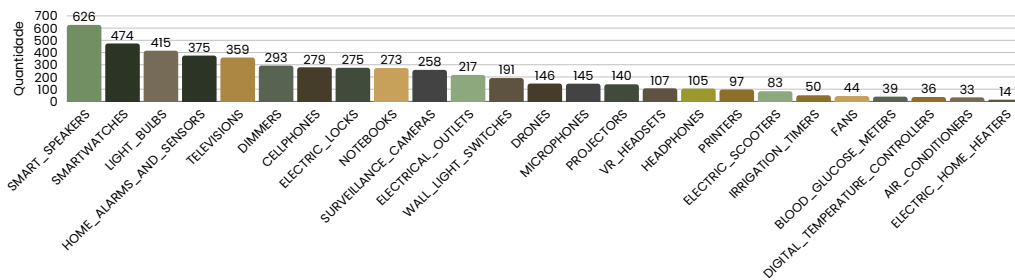


Figura 5. Top 25 categorias de objetos IoT com mais produtos encontrados.

var as notas das avaliações dadas pelos usuários, na Figura 7, nota-se que 89,10% delas correspondem à maior nota (5), indicando uma alta satisfação em relação aos produtos.

A satisfação em relação ao sentimento é apresentada pela Figura 8. Quanto às avaliações, a Figura 8a mostra que 75,43% das interações são positivas, o que aponta uma alta aceitação dos objetos inteligentes no mercado. Em relação às perguntas, observamos um comportamento esperado na Figura 8b com alta quantidade de classificações neutras, visto que perguntas geralmente não possuem uma polaridade de sentimento, uma vez que são questões mais técnicas ou dúvidas quanto ao produto anunciado. Entretanto, notamos que 19,07% das classificações são negativas, sugerindo que o campo de perguntas é usado para realizar comentários de caráter negativo. Por fim, na Figura 8c é notável o caráter positivo das respostas, demonstrando um interesse dos vendedores de manter um tom mais favorável aos seus produtos.

Ao estratificar a análise de sentimentos por domínios de aplicação (Figura 9), identificam-se diferenças significativas na recepção tecnológica dos dispositivos. Vemos uma incidência maior de insatisfação em categorias como controladores de temperatura digital, *headsets VR* e drones, que registraram taxas de sentimento negativo superiores a 25%, sugerindo uma correlação entre o sentimento negativo e fatores como o custo mais elevado, produtos menos comuns e de maior complexidade operacional. Em contrapartida,

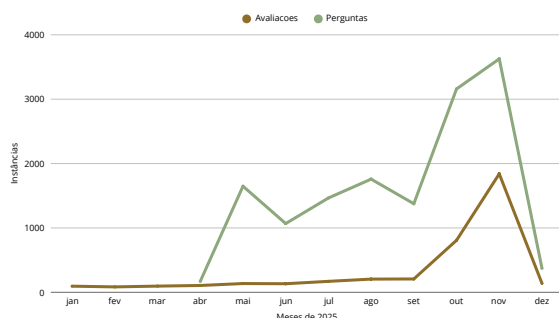


Figura 6. Evolução temporal.

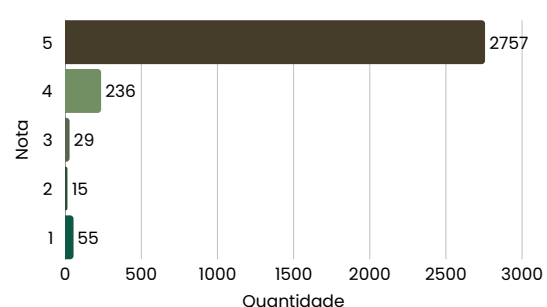


Figura 7. Notas das Avaliações.

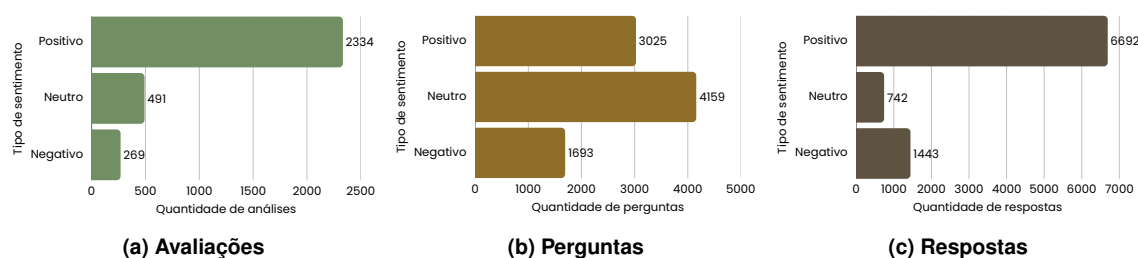


Figura 8. Analise de Sentimento.

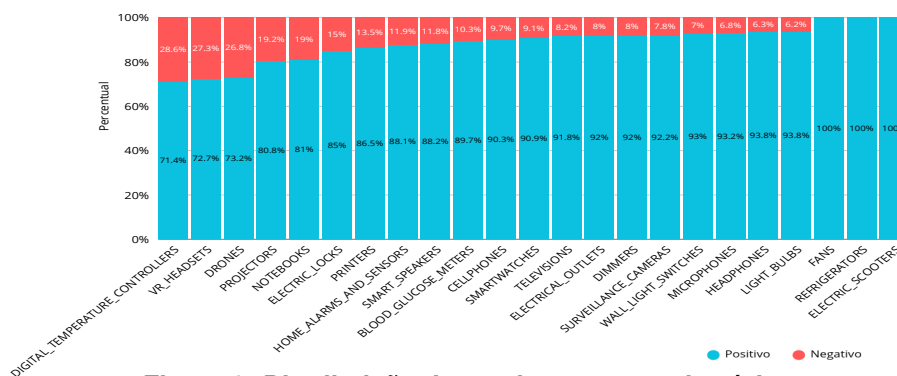


Figura 9. Distribuição de sentimentos por domínio.

tida, domínios como refrigeradores, ventiladores e *scooters* elétricas alcançaram 100% de avaliações positivas. Tais resultados indicam que uma maior simplicidade de uso são determinantes centrais para a aceitação de objetos IoT.

4.1. Barreiras de Informação e Interoperabilidade

Para compreender de forma mais detalhada as opiniões e questões dos consumidores em relação aos produtos IoT, realizamos uma análise léxica e semântica dos dados. As avaliações e dúvidas mais recorrentes são sintetizadas pelas nuvens de palavras (Figura 10) e nos tópicos de perguntas. A presença de termos como “*compatível*”, “*funciona com*”, “*consegue*” e “*preciso de*” (Figura 10b) evidencia que a principal insegurança do consumidor é a interoperabilidade, outros como “*internet*”, “*wifi*”, “*alexa*” “*tomada*” sugerem a forma como usuários querem interagir com os produtos. Em contrapartida, as respostas (Figura 10c) tendem a ser genéricas e burocráticas (foco em nota fiscal e garantia), muitas vezes não sendo satisfatórias e não provendo o suporte técnico necessário para converter o interesse em compra.

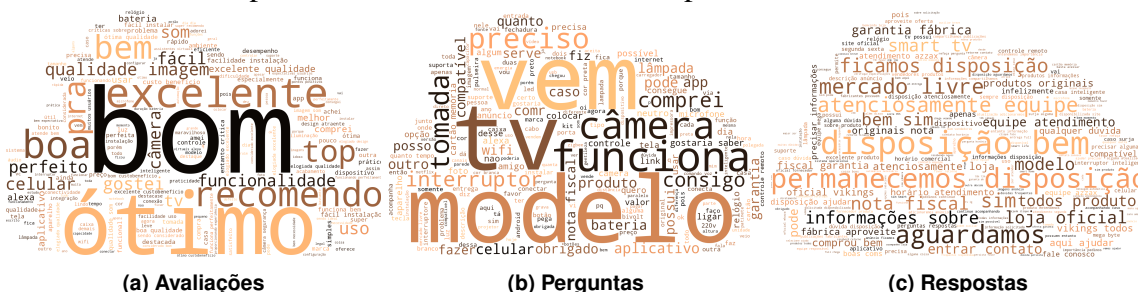


Figura 10. Análise léxica das interações no marketplace.

4.2. Desafios e Oportunidades em IoT

Para responder à **Q3**, a modelagem de tópicos com BERTopic (Tabelas 1 e 2) evidencia um interessante contraste entre as satisfações e frustrações dos usuários, revelando uma assimetria de percepção nos dados entre hardware e software. Enquanto os elogios

Tabela 1. Tópicos Extraídos das Avaliações de Produtos

Sentimento	#	Quantidade	Principais Palavras-Chave	Rótulo Atribuído
Positivo	1	179	fácil, instalar, instalação, muito	Facilidade de Instalação
	2	57	alexa, assistentes, como, além	Conectividade com Assistentes de Voz
	4	55	microfone, som, echo, não	Qualidade de Áudio
	3	40	perfeitamente, funcionou, funciona, funcionando	Funcionamento Geral / Confiabilidade
	5	35	custo, benefício, ótimo, benefício	Custo-Benefício
Negativo	6	42	não, que, produto, com	Mau Funcionamento
	7	32	ainda, não, sem, produto	Avaliações Feitas Sem Teste
	8	30	de, para, problemas, entanto	Design Atrativo e Limitações Técnicas
	9	26	não, que, no, celular	Desafios de Configuração
	10	13	bateria, pra, não, muito	Defeitos de Bateria/Carregamento

Tabela 2. Tópicos Extraídos com BERTopic das Perguntas

Categoria	#	Quantidade	Principais Palavras-Chave	Rótulo Atribuído
Positivo	11	164	câmera, câmeras, cartão, camera	Compatibilidade de Câmeras
	12	149	tv, essa, netflix, boa	Condição do Produto, Garantia e Procedência
	13	126	lâmpada, ventilador, lâmpadas, capacitor	Instalação Elétrica e Requisitos Técnicos
	14	112	teclas, botões, paralelo, interruptor	Configuração de Hardware
	15	73	app, android, aplicativo, smart	Compatibilidade de Aplicativos
Negativo	16	114	compra, não, entrega, vendedor	Problemas de Entrega / Pedido Errado
	17	92	app, android, aplicativo, baixar	Dificuldade na Configuração e Compatibilidade de Hardware
	18	73	interruptor, interruptores, tomada, conj	Compatibilidade com Smartphones
	19	48	recebi, paguei, cadê, produto	Não Entrega / Suspeita de Fraude
	20	45	não, que, responder, nao	Falha de Comunicação e Ineficiência no Atendimento

focam na “Facilidade de Instalação” (Tópico 1), as reclamações (*tópicos* 6, 9 e 10) concentram-se em “Mau Funcionamento”, “Desafios de Configuração” e “Defeitos de Bateria/Carregamento”. Enquanto os tópicos positivos são impulsionados pela tangibilidade do dispositivo, como a “Qualidade de Áudio” (Tópico 4) e a “Facilidade de Instalação Física” (Tópico 1), as barreiras de adoção são predominantemente lógicas e de conectividade. Diferente de eletrônicos tradicionais, onde a queixa principal costuma ser o defeito físico (o produto “não liga”), em IoT o gargalo é a falha de comunicação com o ecossistema digital. Alguns tópicos expõem a insegurança com desafios de configuração e compatibilidade geral e comunicação (T9 e T11). Os usuários frequentemente relatam produtos que funcionam fisicamente, mas falham em entregar a inteligência prometida devido a instabilidades de pareamento ou limitações de software não documentadas. Essa desconexão gera um cenário onde o *hardware* é aprovado, mas o produto é devolvido por incapacidade de integração.

A análise das perguntas (Tabela 2) também sugere uma fragmentação dos protocolos de IoT. Os tópicos também evidenciam que o consumidor é obrigado a possuir um conhecimento técnico avançado antes da compra. Essa questão é representada nos Tópicos 11, 13 e 14, onde usuários questionam especificidades de câmeras, instalação e *hardware*. Por fim, a análise também revela questões gerais do varejo eletrônico, como destacado pelos tópicos 16 (“Problemas de Entrega”), 19 (“Não entrega/Fraude”) e 20 (“Ineficiência no Atendimento”). Um exemplo do Tópico 20 (“*mandei mensagem... mas não respondem*”) sugere que, dada a complexidade inerente à configuração de objetos inteligentes, a ausência de um suporte pós-venda é muitas vezes percebida como uma falha do próprio produto.

Portanto, a oportunidade observada deste domínio reside menos na inovação de *hardware* e mais na simplificação do *setup* inicial e na transparência das especificações de conectividade. Sob a ótica da reutilização deste *dataset*, a base disponibilizada abre um amplo leque de aplicações práticas para desenvolvedores, pesquisadores e engenheiros de IoT. Profissionais da área podem explorar os dados para investigar o desempenho de produtos ou nichos específicos (como dispositivos de segurança ou climatização), reali-

zando análises para entender como diferentes estratégias de conectividade são percebidas pelo público. É possível, por exemplo, contrastar a satisfação de usuários limitados a certos ecossistemas em relação àqueles que utilizam plataformas genéricas. Ademais, os dados de pré-venda (perguntas) oferecem um insumo valioso para times de Engenharia de *Software*, que podem minerar as dúvidas mais recorrentes para antecipar gargalos de usabilidade, otimizar fluxos de configuração (*onboarding*) em aplicativos de controle e refinar a documentação técnica antes mesmo do lançamento de novas linhas de produtos.

4.3. Limitações e Questões Éticas

Mesmo com o rigor metodológico e da escala da coleta realizada, este estudo apresenta limitações que devem ser ponderadas. Primeiramente, o escopo da pesquisa concentrou-se em uma única plataforma de comércio eletrônico. Embora o *e-commerce* de varejo possua uma posição dominante na América Latina, a análise pode refletir particularidades e vieses específicos de seu ecossistema. Além disso, apesar do uso de uma lista exaustiva de palavras-chave baseada em ontologias e de um processo híbrido de extração e filtragem, não é possível garantir a representatividade total dos dados. Fatores tais como a plataforma indexa produtos e as restrições na disponibilização pública de dados podem ter ocultado objetos e dados que se enquadrariam no escopo da pesquisa. No que se refere aos aspectos éticos, a pesquisa seguiu rigorosamente os preceitos da Lei Geral de Proteção de Dados (LGPD), garantindo que o armazenamento e a análise dos textos não infrinjam a privacidade dos usuários originais. Ademais o uso e disponibilização dos dados estão dentro dos parâmetros estabelecidos dos termos e condições do uso da API².

5. Conclusão e Trabalhos Futuros

Este trabalho preenche a lacuna de dados estruturados sobre a adoção de IoT no Brasil ao construir e disponibilizar um *dataset* amplo e de alta qualidade com mais de 5.000 objetos únicos. O principal diferencial da base é a integração de interações de pós-venda (avaliações) com dados de pré-venda (perguntas), o que permitiu mapear uma visão panorâmica da jornada do consumidor, indo além da satisfação de uso para expor as inseguranças técnicas que antecedem a compra.

A análise validou as questões propostas (**Q1** e **Q2**), revelando um mercado de consumo sazonal focado em dispositivos de automação residencial de entrada. Constatou-se (**Q3**) que as maiores barreiras à adoção da IoT não residem em defeitos de *hardware*, mas sim em desafios de *software* e conectividade (e.g., falhas de pareamento e configuração). Além disso, as dúvidas revelaram forte insegurança dos usuários sobre a falta de padronização e clareza nos anúncios, que obriga uma investigação ativa da interoperabilidade dos sistemas, apontando uma oportunidade para aprimoramento da usabilidade e da documentação pelos fabricantes.

Como trabalhos futuros, sugere-se a expansão temporal e multiplataforma da coleta para mitigar vieses, além da aplicação de Modelos de Linguagem de Grande Escala (LLMs) em português (e.g., BERTimbau) visando refinar a captura de nuances semânticas locais. Por fim, ricas propriedades técnicas obtidas pelo *dataset* abrem portas para a sua imediata reutilização na representação fidedigna de objetos inteligentes em ambientes de simulação.

²https://developers.mercadolivre.com.br/pt_br/termos-e-condicoes

Agradecimentos

Agradecimentos ao Manna Team, a Fundação Araucária de Apoio ao Desenvolvimento Científico e Tecnológico do Estado do Paraná (FA), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - Brasil (Processos nº 421548/2022-3 e 404922/2023-6) e ao Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Uso de Inteligência Artificial

Neste trabalho utilizou-se a Inteligência Artificial (IA) Generativa Gemini³, com a finalidade de auxiliar na correção gramatical e ortográfica. O artigo foi elaborado e revisado integralmente pelos autores, que assumem total responsabilidade pelo seu conteúdo.

Referências

- [Almeida 2018] Almeida, R. J. A. (2018). LeIA - Léxico para Inferência Adaptada. <https://github.com/rafjaa/LeIA>.
- [Barbosa 2018] Barbosa, J. C. (2018). *Mineração de texto: uso de técnicas de processamento de linguagem natural para suporte à geração de projeções baseadas em opiniões do consumidor*. PhD thesis, Mestrado em Sistemas de Informação e Gestão do Conhecimento.
- [Beckhauser and Fileto 2024] Beckhauser, W. and Fileto, R. (2024). Can a simple customer review outperform a feature set for predicting churn? In *Anais do XXXIX Simpósio Brasileiro de Bancos de Dados*, pages 117–128.
- [Brownfield and Zhou 2020] Brownfield, S. and Zhou, J. (2020). Sentiment analysis of amazon product reviews. In *Proceedings of the Computational Methods in Systems and Software*, pages 739–750. Springer.
- [Chanda et al. 2024] Chanda, V., Hao, L., and Schulzrinne, H. (2024). From frustration to function: A study on usability challenges in smart home IoT devices. In *2024 IEEE 21st Consumer Communications & Networking Conference (CCNC)*, pages 1–6. IEEE.
- [da Silva and Baranauskas 2023] da Silva, J. and Baranauskas, M. (2023). Internet das coisas em hospitais: Uma revisão sistemática da literatura em ambientes ubíquos e pervasivos. *Relatório técnico, Universidade Estadual de Campinas*.
- [de Melo 2024] de Melo, T. (2024). Starrest: Um conjunto de dados de restaurantes estrelados e respectivas avaliações. In *Anais do VI Dataset Showcase Workshop*, pages 114–124, Porto Alegre, RS, Brasil. SBC.
- [Fabri et al. 2025] Fabri, E. V., Lima, M. P. F., Rodrigues, M. d. S., Silva, S. M. d., and Rangel, M. L. (2025). A revolução da internet das coisas (iot): Conectando o mundo físico e digital. *Revista Missioneira*, 27:173–183.
- [Gilbert 2014] Gilbert, C. H. E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*.
- [Grootendorst 2022] Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure.
- [Kortuem et al. 2009] Kortuem, G., Kawsar, F., Sundramoorthy, V., and Fitton, D. (2009). Smart objects as building blocks for the internet of things. *IEEE internet computing*, 14(1):44–51.
- [Mordor Intelligence 2025] Mordor Intelligence (2025). Mercado de IoT de Consumo - Tamanho e Análise de Participação - Previsões de Crescimento. Acessado em: 30 de junho de 2026.

³<https://gemini.google.com/app>

- [Ojeda et al. 2024] Ojeda, D., Zalewski, W., and Maletzke, A. (2024). Utilizando a quantificação na análise de sentimentos em reviews de produtos. In *Anais da XIX Escola Regional de Banco de Dados*, pages 71–80. SBC.
- [Protick et al. 2026] Protick, T. I., Peddinti, S. T., Taft, N., and Das, A. (2026). LLM-powered analysis of IoT user reviews: Tracking and ranking security and privacy concerns. In *Proceedings of the 20th International AAAI Conference on Web and Social Media (ICWSM)*.
- [Reimers and Gurevych 2019] Reimers, N. and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- [Rohman et al. 2020] Rohman, A. N., Luviana Musyarofah, R., Utami, E., and Raharjo, S. (2020). Natural language processing on marketplace product review sentiment analysis. In *2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS)*.
- [Santana et al. 2025] Santana, H., Bradachi, J., Andrade, M., Silva, A., Silva, T., Silva, F., and Aylon, L. (2025). Od4cot: um dataset baseado em ontologia para o contexto das coisas. In *Anais do XVII Simpósio Brasileiro de Computação Ubíqua e Pervasiva*, pages 141–150. SBC.
- [Santos et al. 2024] Santos, I., Villas, M., and Lifschitz, S. (2024). Análise de satisfação de serviços e produtos brasileiros em plataformas eletrônicas: Uma abordagem de mineração de texto. In *Anais do XXXIX Simpósio Brasileiro de Bancos de Dados*, pages 37–43. SBC.
- [Sebrae 2022] Sebrae (2022). A importância do pós-venda na fidelização e satisfação de clientes. Acessado em: 30 de junho de 2026.
- [Siqueira et al. 2024] Siqueira, V., Costa, R. H., Soares, T., Lunardi, G., and Silva, W. (2024). Dataset anotado de sentimentos a partir de comentários de aplicativos móveis. In *Anais do VI Dataset Showcase Workshop*, pages 65–76.
- [Tripathy and Sharaff 2026] Tripathy, G. and Sharaff, A. (2026). Traversing the landscape of aspect-based sentiment analysis: Delving deeper into techniques, trends, and future directions. *Computer Science Review*, 60:100885.
- [Truelove et al. 2019] Truelove, A., Chowdhury, F. N., Gnawali, O., and Alipour, M. A. (2019). Topics of concern: identifying user issues in reviews of IoT apps and devices. In *Proceedings of the IEEE/ACM 1st International Workshop on Software Engineering Research Practices for the Internet of Things (SERP4IoT 2019)*.
- [Usmani et al. 2021] Usmani, U. A., Haron, N. S., and Jaafar, J. (2021). A natural language processing approach to mine online reviews using topic modelling. In Chaubey, N., Parikh, S., and Amin, K., editors, *Computing Science, Communication and Security*, pages 82–98, Cham. Springer International Publishing.
- [Wang et al. 2025] Wang, H., Wang, S., Lu, Y., et al. (2025). Enhanced audience sentiment analysis in IoT-integrated metaverse media communication. *PLoS ONE*, 20(10):e0332106.
- [Xiao et al. 2021] Xiao, Y., Qi, C., and Hongyong, L. (2021). Sentiment analysis of amazon product reviews based on nlp.
- [Yu et al. 2024] Yu, Y., Fu, Q., Zhang, D., and Gu, Q. (2024). Understanding user experience with smart home products. *Journal of Computer Information Systems*.