



UNIVERSIDADE
LUSÓFONA

Análise de reviews em plataformas de e-commerce

Trabalho Final de Curso

Entrega Final

Aluno: Nuno Santinhos

Orientadora: Sofia Fernandes

Trabalho Final de Curso | LCiD | Junho 2025

www.ulusofona.pt

Direitos de cópia

Análise de reviews em plataformas de e-commerce, Copyright de Nuno Santinhos, Universidade Lusófona.

A Escola de Comunicação, Arquitectura, Artes e Tecnologias da Informação (ECATI) e a Universidade Lusófona (UL) têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objectivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

Este documento foi gerado com o processador (pdf/Xe/Lua)LaTeX e o modelo ULThesis (v1.0.0) [1].

Resumo

Este trabalho explora a análise de avaliações em plataformas de e-commerce através de técnicas de aprendizagem automática não supervisionada, com o objetivo de caracterizar comportamentos de utilizadores que possam auxiliar na deteção de avaliações falsas. Com base num conjunto de dados extraído de uma plataforma de e-commerce, foram analisadas características textuais, temporais e comportamentais das avaliações. Aplicaram-se os algoritmos *K-Means* e *Isolation Forest* para segmentar perfis de utilizadores e identificar padrões de comportamento. Observaram-se diferenças relevantes entre tipos de avaliações, o que poderá contribuir para a identificação de práticas fraudulentas. A abordagem proposta visa reforçar os mecanismos automáticos de moderação, promovendo maior confiança nas plataformas digitais.

Palavras-chave: Avaliações falsas, comércio eletrónico, aprendizagem automatizada, *K-Means*, análise comportamental, características linguísticas, deteção de avaliações falsas, redes neuronais

Abstract

This work explores the analysis of reviews on e-commerce platforms using unsupervised machine learning techniques, with the aim of characterizing user behaviours that may assist in the detection of fake reviews. Based on a dataset extracted from a e-commerce platform, textual, temporal, and behavioural features of the reviews were analysed. The *K-Means* and *Isolation Forest* algorithms were applied to segment user profiles and identify behaviour patterns. Relevant differences were observed between types of reviews, which may help in identifying fraudulent practices. The proposed approach aims to strengthen automated moderation mechanisms, promoting greater trust in digital platforms.

Keywords: Fake reviews, e-commerce, machine learning, K-Means, behavioral analysis, linguistic features, fake review detection, neural networks

Conteúdo

Resumo	2
Abstract	3
Conteúdo	4
Lista de Figuras	5
Lista de Tabelas	6
Introdução	7
1.1 Contexto e Relevância	7
1.2 Objetivos	7
2 - Pertinência e Viabilidade	8
2.1 Impacto das <i>Fake Reviews</i> no Comércio Digital	8
2.2 Relevância e Benefícios da Solução	8
2.3 Inovação Proposta neste Trabalho	9
3 - Benchmarking	10
3.1 <i>Features</i>	10
3.2 Modelos de Machine Learning	11
3.3 Modelos Avançados	11
3.3 Combinação de Abordagens	12
4 - Background Teórico	13
4.1 <i>K-Means</i>	13
4.2 <i>Isolation Forest</i>	14
5 - Solução	15
5.1 Abordagem	15
5.2 Dados Utilizados	15
5.3 Pré-processamento dos Dados	16
5.4 Análise Exploratória dos Dados	17
5.5 Conclusão da Análise Exploratória	27
6 - Resultados e Discussão	28
6.1 Agrupamento dos Utilizadores	28
6.2 Detecção de Outliers	31
6.3 Análise Conjunta dos Resultados	33
6.4 Análise por Categoria de Produto	34
6.5 Limitações da Análise	35
7 - Planeamento	36
8 - Conclusão	38
Bibliografia	39

Lista de Figuras

1	Exemplo de <i>K-Means</i>	13
2	Distribuição de Métricas por Categoria	18
3	Classificação Média por Produto e Categoria	19
4	Número Médio de Avaliações por Produto	20
5	Comprimento Médio das Avaliações por Produto	21
6	Número de Avaliações por Utilizador	22
7	<i>Número Máximo de Avaliações</i>	23
8	Burstness	24
9	Percentagem de Reviews Positivas por categoria	25
10	Percentagem de Compras Verificadas por Categoria	26
11	Número de Utilizadores por Cluster	29
12	Distribuição das variáveis seleccionadas para os Clusters 2 e 6, obtidos pelo algoritmo K-Means.	30
13	Distribuição das variáveis para utilizadores classificados como comuns e como <i>outliers</i> , segundo o modelo Isolation Forest.	32
14	Distribuição dos outliers detetados pelo Isolation Forest em cada cluster	33
15	Gráfico <i>GANTT</i> do planeamento	37

Lista de Tabelas

1 Descrição do Conjunto de Dados 16

Introdução

1.1 Contexto e Relevância

As *fake reviews*, também conhecidas como avaliações falsas, são comentários criados com o objetivo de manipular a percepção de consumidores em plataformas de e-commerce. Esses comentários, que podem ser positivos ou negativos, são um fator determinante na decisão de uma compra, tornando-se um elemento crucial nos mercados digitais. As avaliações positivas podem atrair novos clientes e gerar lucros, enquanto as avaliações negativas podem afastar consumidores e prejudicar a reputação das marcas [2].

As *fake reviews* não são apenas feitas por humanos, *bots* automatizados também têm sido utilizados para criar avaliações falsas. Esses bots conseguem gerar grandes volumes de comentários imitando o estilo de escrita humana. Estudos indicam que cerca de 30% das avaliações no *TripAdvisor* são falsas, frequentemente criadas por indivíduos sem qualquer experiência real com o produto ou serviço [3]. Este tipo de manipulação prejudica tanto os consumidores quanto a integridade dos negócios.

Além disso, classificar uma avaliação como falsa ou verdadeira é uma tarefa complexa. Muitos comentários falsos apresentam uma aparência genuína, de modo a sua identificação e os métodos manuais de detecção de *fake reviews* alcançam uma precisão limitada, tornando necessário o recurso a abordagens mais sofisticadas [2].

1.2 Objetivos

Distinguir avaliações falsas de genuínas requer, antes de mais, uma compreensão profunda do que caracteriza o comportamento legítimo dos utilizadores nas plataformas de e-commerce. Este trabalho parte do princípio de que identificar padrões consistentes em avaliações autênticas pode fornecer uma base sólida para detectar desvios suspeitos.

Assim, propõe-se a análise de um *dataset* recente e real, composto maioritariamente por reviews genuínas, com o objetivo de caracterizar os comportamentos típicos de utilizadores legítimos. Através desta análise, será possível identificar perfis representativos de utilizadores honestos, que servirão de referência para a detecção de comportamentos anómalos associados a avaliações falsas.

Esta abordagem assenta na ideia de que a melhor forma de reconhecer o comportamento fraudulento é, antes de tudo, compreender em detalhe o comportamento autêntico. A estratégia adotada complementa as metodologias tradicionais, focadas apenas na fraude, promovendo uma análise mais equilibrada e robusta.

2 - Pertinência e Viabilidade

2.1 Impacto das *Fake Reviews* no Comércio Digital

As avaliações online tornaram-se um componente essencial no comércio digital, desempenhando um papel determinante nas decisões dos consumidores. Contudo, a manipulação crescente destas avaliações por meio de *fake reviews* compromete a confiança nas plataformas digitais, prejudica a transparência e distorce a concorrência [3].

Os impactos das *fake reviews* vão além dos prejuízos financeiros. Comentários falsos, sejam eles negativos ou positivos, influenciam indevidamente as escolhas dos consumidores. Um cliente pode evitar a aquisição de um produto ou serviço com base em experiências negativas inventadas, causando perdas comerciais. Por outro lado, comentários positivos enganosos podem levar à compra de produtos que não correspondem às expectativas, resultando em frustração e insatisfação [4].

Para enfrentar este problema, é necessário superar vários desafios. Um dos principais é a dificuldade em identificar com precisão quais avaliações são realmente falsas (*ground truth*). A criação de bases de dados fiáveis e bem anotadas é um processo moroso, dispendioso e suscetível a erro humano [2]. A diversidade linguística das avaliações complica ainda mais esta tarefa, uma vez que muitos métodos existentes focam-se apenas numa língua, limitando a sua eficácia em contextos multilíngues. Acrescem ainda as limitações tecnológicas e computacionais, que dificultam o tratamento de grandes volumes de dados e, por conseguinte, a exploração abrangente do problema [3].

Perante este cenário, torna-se fundamental desenvolver abordagens inovadoras, escaláveis e adaptáveis que permitam restaurar a confiança nas plataformas digitais e promover práticas comerciais justas.

2.2 Relevância

O combate às *fake reviews* não se limita à sua deteção, estando também relacionado com a promoção de práticas comerciais justas e a proteção dos consumidores contra informações enganosas. Este tema é particularmente relevante num cenário onde o impacto destas avaliações vai além do comércio digital, influenciando o turismo, os serviços e as plataformas de recomendação. Estudos mostram que práticas enganosas podem prejudicar severamente tanto consumidores como empresas, reforçando a necessidade de métodos eficazes para enfrentar este problema [3].

Além disso, a resolução deste problema contribui para o avanço de outras tecnologias, como sistemas de recomendação mais confiáveis e análises preditivas baseadas em dados limpos e fiáveis. Essas melhorias não só aumentam a confiança dos consumidores nas plataformas, mas também incentivam práticas comerciais éticas, promovendo um mercado mais sustentável e competitivo [2].

2.3 Inovação Proposta neste Trabalho

Este trabalho distingue-se dos estudos anteriores por duas razões principais. Em primeiro lugar, recorre a um *dataset* recente e ainda pouco explorado na literatura, permitindo uma análise mais atualizada dos padrões associados a *fake reviews*. Em segundo lugar, propõe-se a exploração combinada de diferentes tipos de informação — como dados textuais, temporais e comportamentais — com o objetivo de compreender em profundidade as características que definem as avaliações genuínas.

Este trabalho visa gerar *insights* relevantes sobre o comportamento dos utilizadores legítimos, com base em um conjunto diversificado de *features*. A intenção é contribuir para uma caracterização mais clara das avaliações autênticas, oferecendo uma base sólida para distinguir entre avaliações genuínas e falsas. Estes conhecimentos poderão ser futuramente utilizados para apoiar o desenvolvimento de modelos mais robustos, fiáveis e explicáveis, adaptados aos desafios reais enfrentados pelas plataformas de *e-commerce*.

3 - Estado da Arte

A construção de um sistema eficaz para a detecção de *fake reviews* requer uma análise cuidadosa das abordagens já existentes. Este capítulo apresenta um levantamento das principais técnicas utilizadas, tanto ao nível da extração de *features* como da aplicação de modelos de *machine learning*.

3.1 Features

A eficácia dos modelos depende da qualidade e relevância das *features* extraídas, ou seja, dos atributos ou métricas calculadas a partir dos dados das avaliações. Estas *features* servem como entrada para os modelos de *machine learning*, permitindo a detecção de padrões associados a comportamentos suspeitos.

Número Máximo de Avaliações (*Maximum Number of Reviews*) calcula o número máximo de avaliações publicadas por um utilizador num dado intervalo de tempo (por exemplo, num dia). Este valor é normalizado pela maior contagem encontrada entre todos os utilizadores, segundo a fórmula 1 [5, 6, 7]:

$$\frac{MaxRev(a)}{\max(MaxRev)} \quad (1)$$

Em que $MaxRev(a)$ representa o maior número de avaliações que o utilizador a publicou num único intervalo de tempo (como um dia), e $\max(MaxRev)$ é o maior valor observado entre todos os utilizadores. Esta *feature* permite identificar utilizadores que publicam mais avaliações num curto período, um comportamento típico de *spammers*.

Porcentagem de Avaliações Positivas (*Percentage of Positive Reviews*) mede a proporção de avaliações com pontuação alta (4 ou 5 estrelas) feitas por um utilizador, em relação ao total de avaliações que publicou. É calculada segundo a fórmula 2 [8]:

$$\frac{\sum_{x=1}^{|R_a|} |\star(r_x) \in \{4, 5\}|}{|R_a|} \quad (2)$$

Nesta expressão, R_a é o conjunto total de avaliações feitas pelo utilizador a , $|R_a|$ é o número total de avaliações desse utilizador, e $\star(r_x)$ representa a pontuação de uma avaliação r_x . Esta métrica destaca utilizadores com padrões de avaliação tendencialmente positivos.

Comprimento Médio das Avaliações (*Average Review Length*) classifica cada avaliação com base no seu comprimento. A ideia é que avaliações falsas tendem a ser mais curtas. A classificação é feita segundo a fórmula 3 [5, 6, 7]:

$$\begin{cases} 1, & \text{se } len(r_a) < X \\ 0, & \text{caso contrário.} \end{cases} \quad (3)$$

Em que $len(r_a)$ indica o número de palavras da avaliação r_a , e X é um limiar definido manualmente. Se a avaliação for mais curta do que X , é classificada como curta (1); caso contrário, como longa (0). Esta *feature* é útil para distinguir comentários superficiais ou automáticos.

Burstness avalia se um utilizador publicou muitas avaliações num curto espaço de tempo (por exemplo, nas últimas 24 horas), o que pode indicar atividade suspeita. A métrica é calculada com a fórmula 4 [9]:

$$\begin{cases} 1, & \text{se } \sum_{x=1}^{|R_a|} |r_x \in R_a \cap (\text{últimas 24 horas})| > X \\ 0, & \text{caso contrário.} \end{cases} \quad (4)$$

Em que R_a é o conjunto de avaliações do utilizador a , e o termo $R_a \cap (\text{últimas 24 horas})$ representa as avaliações que foram publicadas nesse intervalo temporal. A soma verifica quantas avaliações foram feitas recentemente, e se esse número ultrapassar o limiar X , classifica-se como atividade anómala (1).

Estas *features*, em conjunto com características linguísticas e do produto, fornecem aos modelos uma base sólida para identificar padrões de comportamento suspeito — desde que se disponha de dados previamente anotados (*ground truth*).

3.2 Modelos de Machine Learning

Os métodos de *machine learning* para deteção de *fake reviews* dividem-se, de forma geral, em duas categorias principais: supervisionados e não supervisionados.

3.2.1 Modelos Supervisionados

Os métodos supervisionados utilizam um conjunto de dados rotulados para treinar o modelo, onde cada *review* é previamente classificado como falsa ou genuína [10]. Exemplos de modelos supervisionados incluem *Support Vector Machines* (SVM) [11], Regressão Logística [12] e *Random Forest* [13]. Estes modelos funcionam bem quando têm acesso a um grande volume de dados rotulados, mas tendem a ser menos eficazes em situações onde os dados estão desbalanceados ou são escassos.

3.2.2 Modelos Não Supervisionados

Por outro lado, os métodos não supervisionados não necessitam de dados rotulados. Em vez disso, baseiam-se na análise de padrões ou agrupamentos nos dados. Estes métodos são úteis em situações onde não é possível obter etiquetas confiáveis, como em plataformas que geram milhares de avaliações diariamente.

Uma das técnicas utilizadas neste contexto é o *K-Means*, que organiza os dados em grupos (*clusters*) com base na similaridade das suas características. No caso da deteção de *fake reviews*, o *K-Means* pode ser usado para identificar agrupamentos de avaliações que partilhem padrões suspeitos, como textos semelhantes ou comportamentos suspeitos por parte de utilizadores [2]. Apesar de serem menos dependentes de dados anotados, os métodos não supervisionados enfrentam desafios na interpretação dos agrupamentos e podem produzir resultados menos precisos quando aplicados isoladamente.

3.2.3 Modelos Avançados

Modelos mais complexos têm sido explorados para análise do texto das *reviews*, nomeadamente redes neurais convolucionais (CNN) e recorrentes (RNN e LSTM) [14]. As CNN identificam padrões no texto, como repetições ou frases genéricas. As RNN e LSTM destacam-se na análise de relações entre palavras em textos mais longos. Embora estas redes pertençam à categoria de modelos supervisionados, algumas implementações têm combinado técnicas não supervisionadas para melhorar a robustez da deteção [3].

3.3 Combinação de Abordagens

Embora os métodos supervisionados e não supervisionados apresentem diferenças claras, eles não são mutuamente exclusivos. Muitas abordagens atuais combinam técnicas de ambos os tipos para maximizar a precisão e minimizar as limitações individuais. Esta integração tem sido essencial para lidar com os desafios do volume crescente de dados e a constante evolução das estratégias por parte dos *fake reviewers*. Além disso, a flexibilidade desta combinação permite que os modelos sejam ajustados para diferentes contextos e línguas [2].

4 - Conceitos de Modelos

4.1 K-Means

O algoritmo *K-Means* [15] é uma técnica de agrupamento não supervisionada que organiza os dados em grupos (*clusters*) com base na semelhança entre observações. O seu objetivo é particionar um conjunto de instâncias em k grupos distintos, minimizando a variância intra-*cluster*. A ideia central do algoritmo é simples: elementos com características semelhantes devem ser agrupados em torno de um centróide representativo.

Cada observação é representada como um vetor numérico, independentemente da forma como esses vetores estão armazenados (por exemplo, como linhas de uma matriz de dados). O algoritmo assume que é possível medir a proximidade entre vetores, sendo a distância euclidiana a métrica mais utilizada por defeito, embora outras métricas possam ser aplicadas.

O processo inicia-se com a seleção aleatória de k centróides iniciais. Em seguida, cada observação é atribuída ao *cluster* cujo centróide está mais próximo (fase de atribuição). Após esta atribuição, os centróides são atualizados para refletir a média dos vetores pertencentes a cada grupo (fase de atualização). Estes dois passos repetem-se de forma iterativa até à convergência, ou seja, até que as atribuições deixem de variar significativamente.

Apesar da sua simplicidade, o *K-Means* é uma ferramenta eficaz na descoberta de padrões em grandes volumes de dados e é amplamente utilizado em tarefas exploratórias, como a identificação de perfis de comportamento em avaliações online.

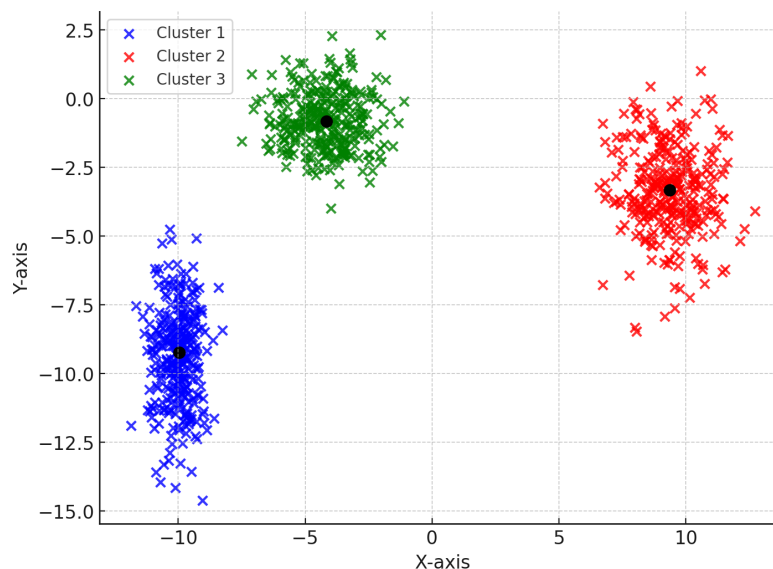


Figura 1: Exemplo de *K-Means*

Na figura 1, podemos observar o funcionamento do algoritmo *K-Means*. A imagem representa o resultado de um agrupamento realizado com o algoritmo *K-Means*, aplicado a um conjunto de dados bidimensionais. Cada símbolo “x” corresponde a uma observação individual, posicionada no plano de acordo com duas variáveis numéricas representadas nos eixos X e Y. As diferentes cores dos pontos indicam a pertença de cada observação a um dos três *clusters* identificados: azul para o *Cluster 1*, vermelho para o *Cluster*

2 e verde para o *Cluster 3*. Os pontos pretos a negrito representam os centróides de cada grupo, ou seja, os pontos médios calculados pelo algoritmo em torno dos quais se distribuem as observações pertencentes a cada *cluster*. Estes centróides são atualizados iterativamente durante o processo até estabilizarem numa posição que minimiza a distância média entre os elementos e o respetivo centro. A imagem permite observar visualmente a separação entre os grupos formados e a concentração das observações em torno dos seus centróides.

O *K-Means* é um algoritmo eficiente e rápido, especialmente útil para trabalhar com grandes volumes de dados. Apesar disso, o *K-Means* tem certas limitações. Os resultados dependem fortemente da escolha inicial dos centróides, o que pode levar a soluções inconsistentes. Uma das outras dificuldades, é a necessidade de definir o número de clusters antes de executar o algoritmo, algo que pode ser difícil sem um conhecimento prévio dos dados. No contexto das avaliações falsas, esta técnica permite para identificar padrões comuns em grupos de avaliações, permitindo destacar semelhanças nos comportamentos dos utilizadores.

4.2 Isolation Forest

O algoritmo *Isolation Forest* é uma técnica não supervisionada utilizada para a deteção de *outliers* em conjuntos de dados [16]. Ao contrário de métodos que procuram modelar o comportamento comum para, posteriormente, identificar desvios, o *Isolation Forest* baseia-se diretamente no princípio de que os *outliers* são observações mais fáceis de isolar do que as restantes.

A lógica subjacente ao algoritmo assenta no facto de que os *outliers* tendem a ser escassos e diferentes, tornando possível separá-los do resto do conjunto com menos decisões. Para o efeito, o *Isolation Forest* constrói múltiplas árvores binárias aleatórias, denominadas *isolation trees*. Cada nó destas árvores corresponde a uma divisão aleatória de uma das variáveis, particionando recursivamente os dados. A profundidade média necessária para isolar uma observação serve como estimativa da sua anomalia: observações isoladas em poucas divisões (isto é, com caminhos curtos) são mais prováveis de serem *outliers*.

Este método apresenta diversas vantagens: é eficiente em conjuntos de dados de grande dimensão, funciona bem com dados de alta dimensionalidade e não requer rótulos (*labels*) para ser aplicado. Estas características tornam-no particularmente adequado para tarefas de deteção de padrões fora do comum, como perfis de utilizadores com comportamentos excecionais.

No contexto deste trabalho, os *outliers* identificados correspondem a utilizadores cujo comportamento se desvia estatisticamente do padrão dominante. Embora o algoritmo não forneça uma justificação direta para cada caso, estas observações assumem maior relevância analítica quando ocorrem em simultâneo com variáveis frequentemente associadas a manipulação de avaliações, como frequência de publicação elevada, baixa diversidade ou extensão textual reduzida.

5 - Solução

5.1 Abordagem

Este trabalho recorre a uma abordagem não supervisionada baseada em duas técnicas principais: o algoritmo *K-Means* e o *Isolation Forest*. O *K-Means* será utilizado para identificar agrupamentos de utilizadores com base em características textuais, temporais e comportamentais extraídas das avaliações, com o objetivo de detetar perfis distintos, nomeadamente aqueles que representam comportamentos típicos de utilizadores legítimos. Por sua vez, o *Isolation Forest* será aplicado para identificar *outliers*, isto é, utilizadores cujos padrões de comportamento se desviam da maioria, podendo indiciar atividade suspeita ou artificial.

Através dos agrupamentos e dos *outliers* identificados, analisam-se os diferentes tipos de utilizadores e as suas práticas de avaliação. Estes conhecimentos poderão, numa fase posterior, informar o desenvolvimento de modelos mais robustos e adequados para a deteção automática de *fake reviews*.

5.2 Dados Utilizados

Neste estudo, é utilizado um *dataset* de avaliações provenientes de uma plataforma de *e-commerce*, contendo comentários publicados por utilizadores reais. A seleção das observações recaiu sobre reviews escritas nos anos de 2022 e 2023, de forma a garantir a atualidade dos dados analisados. Esta escolha permite explorar padrões recentes de comportamento e refletir melhor as práticas atuais dos utilizadores nestas plataformas.

Embora o *dataset* utilizado neste estudo se encontre anonimizado, identificaram-se algumas questões relacionadas com a sua sensibilidade. Por precaução e em alinhamento com boas práticas de responsabilidade no tratamento de dados, optou-se por não tornar público o código desenvolvido, garantindo assim uma abordagem mais cautelosa face a potenciais implicações. Esta decisão visa assegurar o respeito por princípios éticos e pela integridade da investigação.

Cada observação no *dataset* representa uma avaliação individual feita por um utilizador para um produto. Estas observações contêm informações textuais, categóricas e numéricas relacionadas com a avaliação e o utilizador.

No contexto deste estudo, o dataset inclui várias colunas que fornecem informações úteis para a análise das avaliações e dos utilizadores. A tabela 1 descreve as colunas mais relevantes presentes no conjunto de dados.

Tabela 1: Descrição do Conjunto de Dados

Campo	Descrição
Classificação	Avaliação numérica do produto, entre 1 e 5 estrelas.
Comentário	Texto completo da opinião deixada pelo utilizador.
ID Produto	Identificador único do produto avaliado.
ID Utilizador	Identificador único do autor da avaliação.
Data	Momento em que a avaliação foi publicada.
Compra Verificada	Indica se o produto foi efetivamente comprado pelo utilizador.

Por razões computacionais, é necessário reduzir o volume dos dados. Logo, optou-se por escolher categorias que mantivessem o equilíbrio e a qualidade dos dados.

As análises realizadas neste estudo concentram-se em categorias de produtos selecionadas, com base na proporção de *reviews* falsas e genuínas, para garantir uma abordagem equilibrada e eficaz na identificação de padrões de comportamento. Essas categorias incluem eletrodomésticos (*appliances*), telemóveis e acessórios (*cell phone and accessories*), roupas, sapatos e joias (*clothing, shoes and jewelry*), produtos de escritório (*office products*), desporto e atividades ao ar livre (*sports and outdoors*) e livros (*books*).

Entre as categorias escolhidas, algumas destacam-se pela elevada proporção de avaliações genuínas, sendo essas categorias: produtos de escritório, telemóveis e acessórios e livros. Por exemplo, a categoria de livros tem 81% de avaliações consideradas genuínas, o que faz desta categoria uma excelente referência para identificar padrões típicos de avaliações autênticas [4].

Por outro lado, algumas categorias foram selecionadas devido à maior suspeita de avaliações falsas, estas sendo: eletrodomésticos, roupas, sapatos e joias e desporto e atividades ao ar livre. Estas categorias tornam-se úteis para o estudo de padrões fraudulentos. Eletrodomésticos, roupas, sapatos, joias, e desporto e atividades ao ar livre apresentam uma quantidade significativa de *reviews* suspeitas.

5.3 Pré-processamento dos Dados

Para iniciar a manipulação dos dados, os ficheiros em formato JSON foram convertidos para CSV, facilitando a sua leitura e tratamento. De seguida, procedeu-se à seleção das avaliações escritas em inglês, uma vez que representavam a maioria no conjunto original, resultando num total de cerca de 12 milhões de reviews.

Filtragem e análise a nível do produto

Durante a análise inicial, verificou-se que 63,64% dos produtos possuíam apenas uma avaliação. Como uma única opinião não oferece contexto suficiente para análise comparativa nem permite inferir padrões fiáveis, estes casos foram removidos. Foram também calculadas várias métricas agregadas por produto, como a classificação média, o número médio de avaliações por produto e o comprimento médio das avaliações. Adicionalmente, foi analisada a percentagem de avaliações associadas a compras verificadas, por categoria de produto. Produtos com uma baixa proporção de compras verificadas podem estar mais suscetíveis à presença de avaliações artificiais ou menos fiáveis.

Filtragem e análise a nível do utilizador

Verificou-se que aproximadamente 69,58% dos utilizadores tinham submetido apenas uma única review, o que comprometeria a fiabilidade da extração de padrões comportamentais. Por esse motivo, esses registos foram também excluídos. Para os utilizadores restantes, foram calculadas diversas métricas com o objetivo de caracterizar o seu comportamento, incluindo: a classificação média atribuída, o número total de avaliações submetidas, o comprimento médio dos comentários, e a diversidade de categorias avaliadas.

Adicionalmente, foram exploradas métricas temporais e relacionais mais avançadas, como o número máximo de avaliações num único mês, a percentagem de avaliações positivas publicadas, e a métrica de *burstness*, que quantifica concentrações de atividade em curtos intervalos de tempo. Foi também considerada a relação entre o número total de avaliações realizadas e o comprimento médio dos respetivos comentários, permitindo identificar padrões de participação mais intensiva ou atípica.

Resumo do processo de filtragem

Dado que identificar utilizadores ou produtos com comportamento invulgar com base numa única avaliação seria extremamente desafiante e pouco fiável, decidiu-se pela remoção de todos os casos associados a apenas uma review. Após este processo de filtragem, o conjunto de dados final ficou reduzido a cerca de 5,5 milhões de avaliações.

5.4 Análise Exploratória dos Dados

Durante esta etapa, foram analisadas características gerais do conjunto de dados, como a distribuição de produtos, avaliações e utilizadores por categoria. Estas análises iniciais serviram como base para compreender a estrutura dos dados e identificar aspetos relevantes para a deteção de avaliações falsas.

Foi efetuada uma análise das avaliações duplicadas, considerando exclusivamente aquelas com um mínimo de 10 palavras, pois assim diminui-se o "ruído" dos dados eliminando as avaliações muito curtas. Para além disso, para cada avaliação duplicada, analisou-se o número de utilizadores que a submeteram e o número de categorias em que foi registada.

Durante a análise, constatou-se que as 10 avaliações mais repetidas no *dataset* foram todas submetidas por um único utilizador. Estas avaliações reaparecem em múltiplas categorias de produtos. A maioria destes textos é notavelmente curta e apresenta um conteúdo genérico. As categorias onde estas repetições ocorrem com maior frequência incluem Telemóveis e Acessórios, Produtos de Escritório e Roupas, Sapatos e Joalharia.

A análise exploratória teve como principal objetivo compreender as características gerais das avaliações, dos utilizadores e dos produtos presentes no conjunto de dados. Para tal, os dados foram organizados em diferentes categorias de produto, o que permitiu uma comparação mais estruturada entre diferentes domínios.

Visão Geral

A distribuição de métricas essenciais, como o número de produtos, o número de avaliações e o número de utilizadores, foi analisada para cada categoria. Esta visualização (ver Figura 2) permite observar disparidades significativas entre categorias, bem como compreender a proporção entre os elementos analisados. Estes dados fornecem um enquadramento quantitativo inicial que orienta as etapas seguintes da investigação.

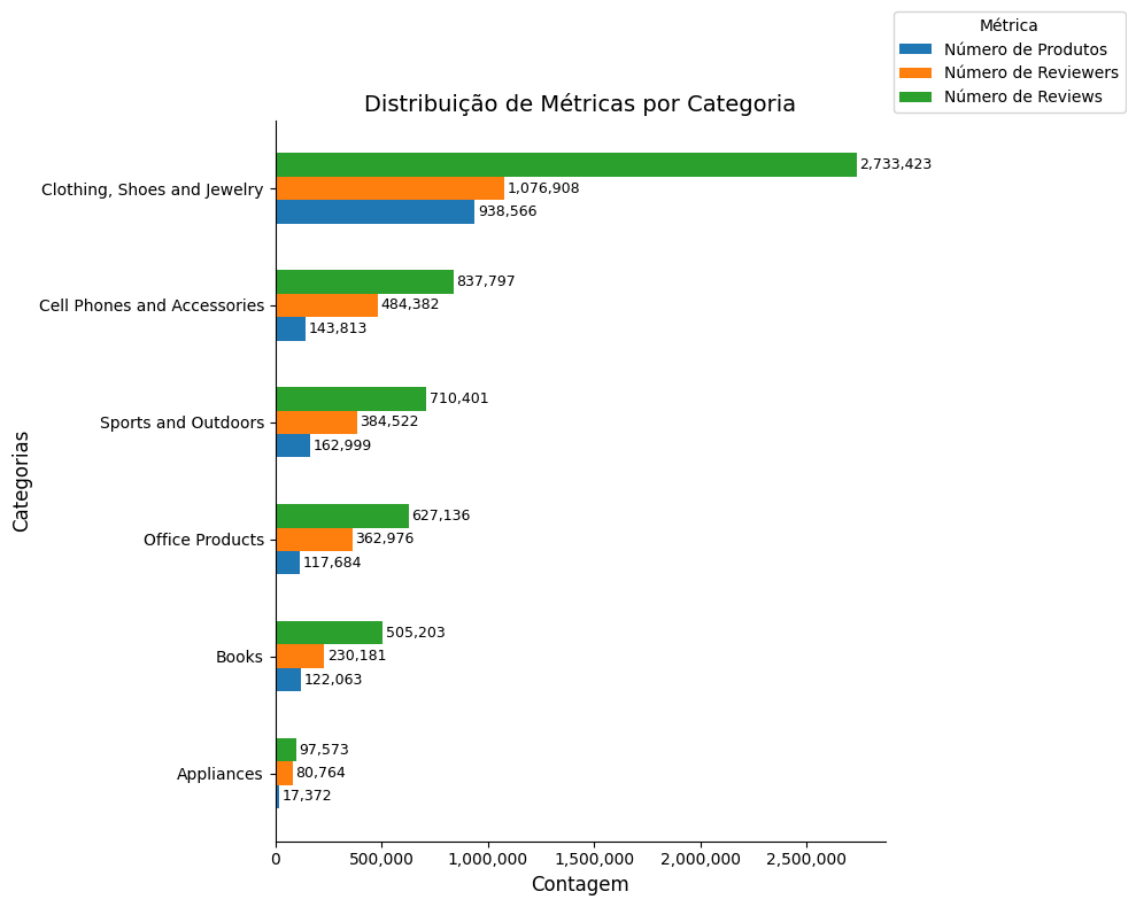


Figura 2: Distribuição de Métricas por Categoria

A análise da distribuição das métricas por categoria revela que a categoria de Roupas, Sapatos e Joalheria se destaca com uma predominância clara em todas as dimensões analisadas. Esta categoria regista aproximadamente 2,7 milhões de avaliações, realizadas por cerca de 1,1 milhões de utilizadores distintos, abrangendo cerca de 939 mil produtos. Em contraste, a categoria de Eletrodomésticos apresenta os valores mais baixos, com cerca de 97 mil avaliações, 81 mil utilizadores e 17 mil produtos, representando uma fração muito reduzida do total do *conjunto de dados*.

Observa-se ainda um padrão transversal às categorias, em que o número de avaliações é substancialmente superior ao número de utilizadores, o que sugere que muitos indivíduos publicam múltiplas reviews. Esta disparidade é especialmente evidente em categorias como Telemóveis e Acessórios e Desporto e Atividades ao Ar Livre.

Foram analisadas métricas relacionadas com os produtos, incluindo a classificação média, o número de avaliações por produto e o comprimento médio das avaliações. Esta análise permitiu compreender melhor as características associadas aos produtos e identificar possíveis padrões nos dados.

Como Varia o Rating dos Produtos por Categoria?

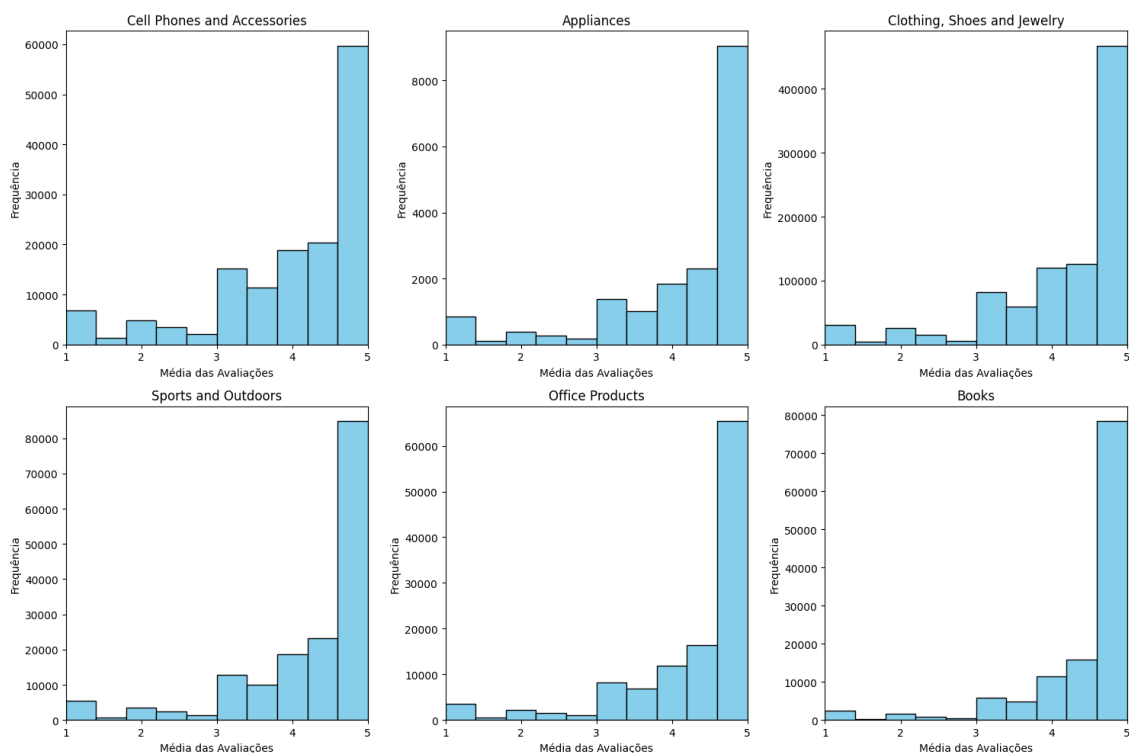


Figura 3: Classificação Média por Produto e Categoria

A Figura 3 mostra a distribuição da média das avaliações por produto em diferentes categorias. Em todas elas, observa-se uma concentração acentuada no valor máximo da escala (5 estrelas), evidenciando uma assimetria positiva. Esta tendência pode dever-se ao viés de positividade comum nas plataformas de *e-commerce*[3].

A categoria de Livros destaca-se particularmente por apresentar uma proporção muito elevada de produtos cuja média de avaliação é exatamente 5, sugerindo uma forte tendência para classificações máximas. Embora a categoria de Roupas, Sapatos e Joalheria também apresente um volume elevado de avaliações com classificações muito altas, o caso dos Livros é ainda mais expressivo em termos proporcionais. Esta uniformidade na tendência geral reforça a ideia de que a simples média das avaliações pode não ser, por si só, um indicador fiável da veracidade ou qualidade de uma review, justificando a necessidade de explorar outras métricas complementares para uma análise mais robusta.

Quantas reviews têm os produtos de cada categoria?

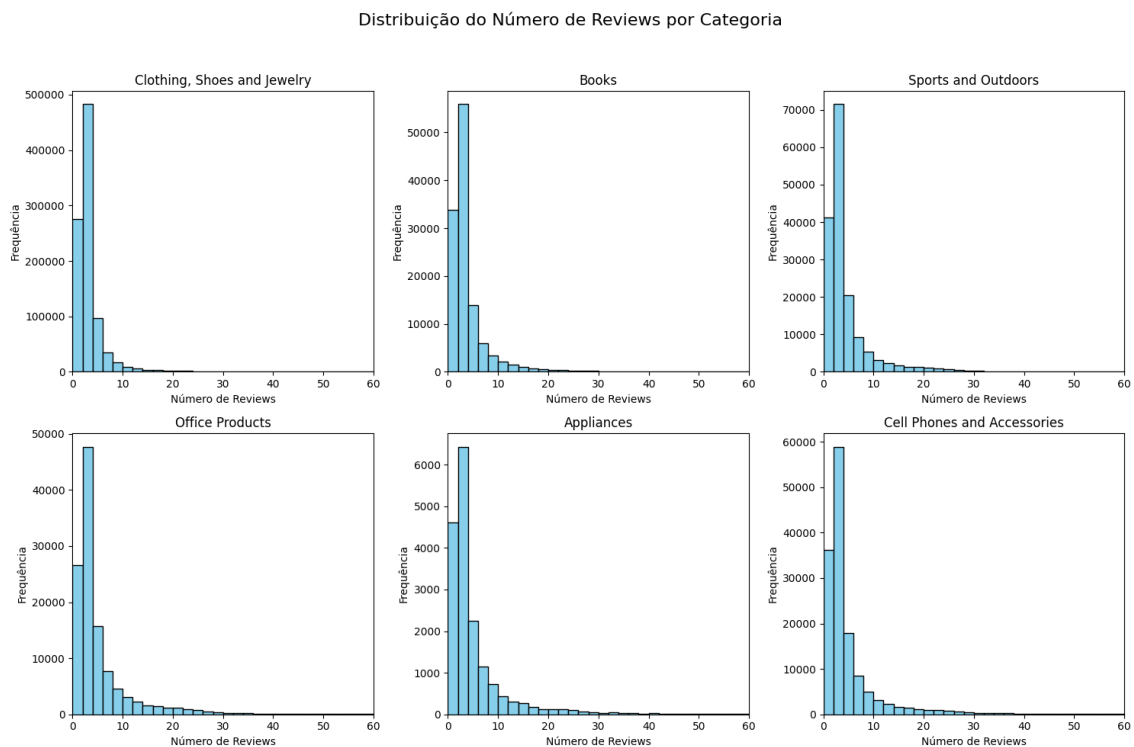


Figura 4: Número Médio de Avaliações por Produto

A Figura 4 mostra a distribuição do número de avaliações por produto em diferentes categorias. Observa-se uma tendência fortemente assimétrica em todas as categorias: a maioria dos produtos possui um número reduzido de reviews, geralmente inferior a 10, enquanto apenas uma minoria concentra um número elevado de avaliações.

Qual o tamanho das reviews?

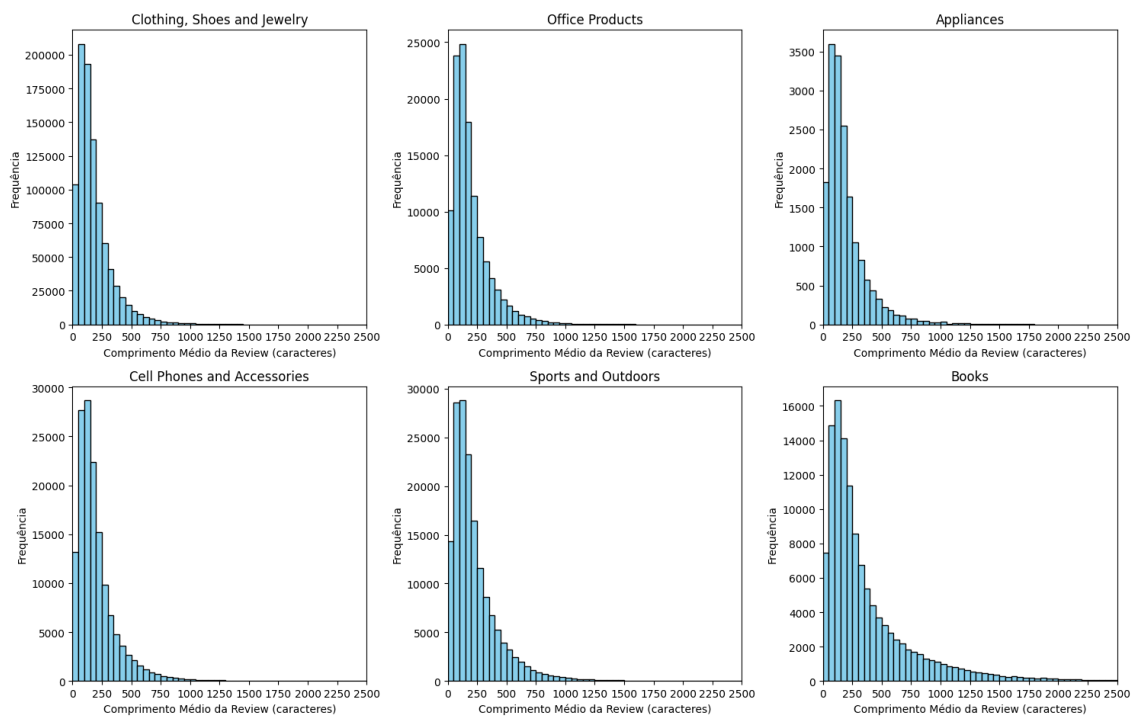


Figura 5: Comprimento Médio das Avaliações por Produto

A Figura 5 apresenta a distribuição do comprimento médio das avaliações por produto, medido em número de caracteres, para cada uma das categorias analisadas. De forma consistente, observa-se que a maioria das avaliações possui entre 100 e 200 caracteres, evidenciando uma forte concentração em textos curtos. A distribuição é assimétrica à direita, com uma cauda longa que representa um número reduzido de produtos cujas avaliações são mais extensas.

Este padrão sugere que, na generalidade, os utilizadores optam por comentários breves, com apenas uma minoria a redigir avaliações mais desenvolvidas. No entanto, a categoria de *Livros* destaca-se das restantes por apresentar uma maior proporção de produtos com avaliações mais longas. Este comportamento pode refletir uma maior motivação dos utilizadores para partilhar opiniões mais detalhadas sobre conteúdos literários, em contraste com outras categorias de produtos, onde os comentários tendem a ser mais diretos ou funcionais.

Quantas reviews faz cada utilizador?

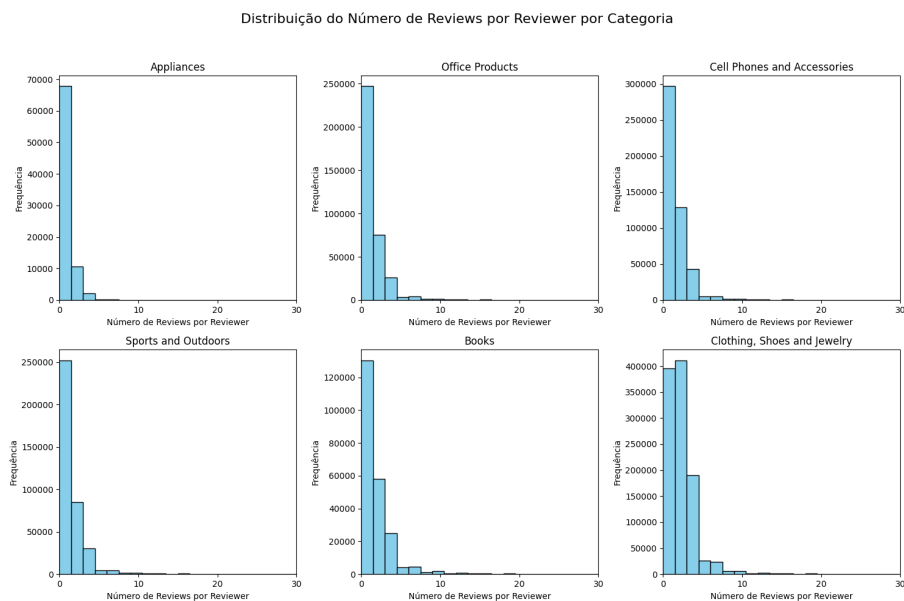


Figura 6: Número de Avaliações por Utilizador

A Figura 6 apresenta a distribuição do número de avaliações publicadas por utilizador em cada uma das categorias analisadas. Os dados revelam uma distribuição fortemente assimétrica: a esmagadora maioria dos utilizadores escreve apenas uma ou poucas reviews, enquanto apenas uma pequena fração contribui com volumes mais elevados de avaliações.

Este padrão é consistente em quase todas as categorias, refletindo uma participação geralmente limitada. No entanto, a categoria de *Livros* distingue-se por apresentar uma maior concentração de utilizadores com atividade elevada, ou seja, com um número consideravelmente superior de avaliações publicadas. Este comportamento pode indicar um maior envolvimento dos utilizadores nesta categoria.

A presença deste subconjunto de utilizadores altamente ativos pode ser relevante tanto para o enriquecimento do conteúdo da plataforma como para a análise de padrões menos comuns, que merecem atenção adicional em contextos de deteção de comportamentos invulgares.

Qual o nível de atividade dos reviewers durante um mês?

Distribuição do Máximo de Reviews por Mês por Categoria

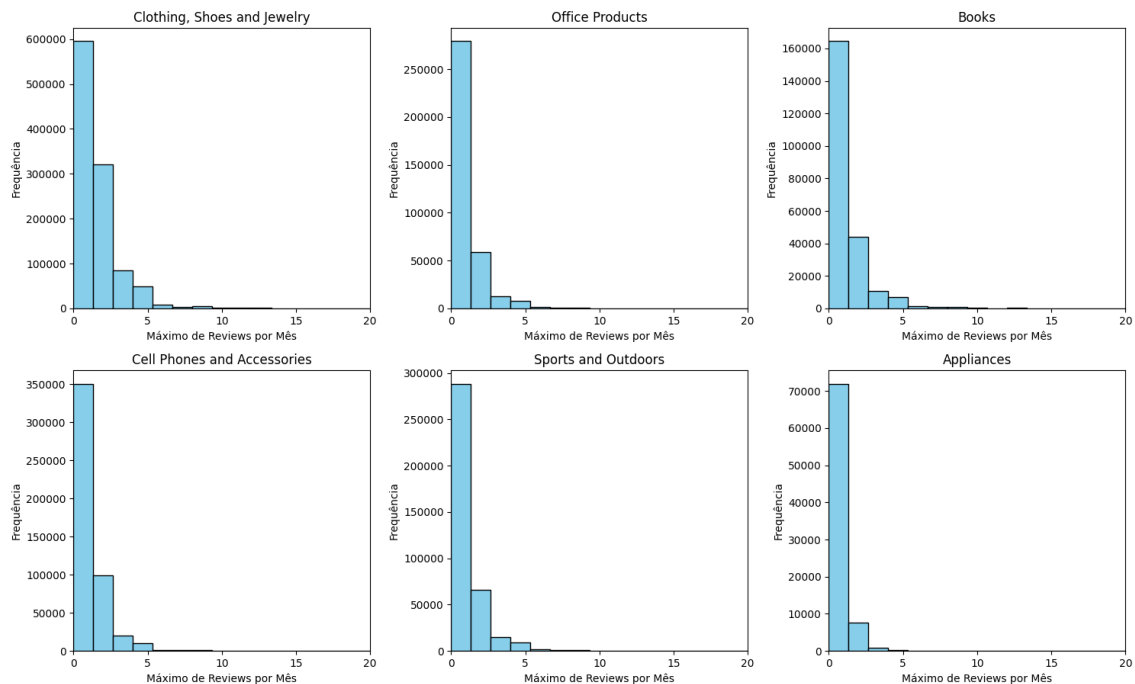


Figura 7: Número Máximo de Avaliações

A Figura 7 mostra a distribuição do número máximo de avaliações realizadas por cada utilizador num único mês, analisada por categoria de produto. Esta métrica permite avaliar a intensidade de atividade dos utilizadores ao longo do tempo, funcionando como um indicador de envolvimento.

Em todas as categorias, verifica-se que a grande maioria dos utilizadores nunca ultrapassa duas ou três reviews por mês, revelando padrões de participação moderados e compatíveis com comportamento comum. Contudo, observa-se também a presença de uma cauda longa, onde alguns utilizadores atingem níveis de atividade superiores (mais de 10 ou 15 avaliações num único mês), o que pode indiciar perfis invulgares.

Entre as categorias analisadas, destaca-se *Roupas, Sapatos e Joalharia*, onde se observa uma proporção mais elevada de utilizadores com maiores níveis de atividade. Esta particularidade torna a categoria especialmente relevante para análise mais detalhada no contexto da deteção de padrões de comportamento atípico.

Quantas avaliações são feitas por utilizador numa hora?

Distribuição do Máximo de Reviews por Hora (Burstness) por Categoria

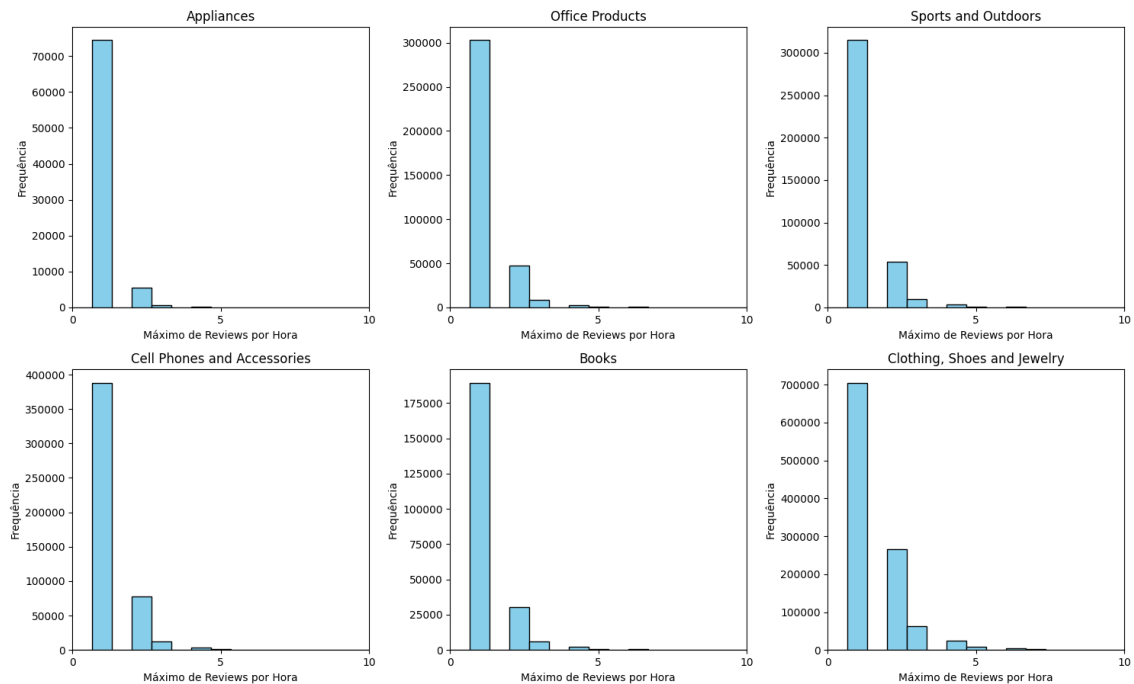


Figura 8: Burstness

A Figura 8 apresenta a distribuição do número máximo de avaliações submetidas por cada utilizador numa única hora, uma métrica conhecida como *burstness*. Este indicador permite identificar picos de atividade invulgarmente concentrada num curto espaço temporal.

Os resultados revelam que a grande maioria dos utilizadores, em todas as categorias, publica no máximo entre uma e duas avaliações por hora. Contudo, observam-se casos menos frequentes em que os utilizadores submetem cinco ou mais avaliações dentro do mesmo intervalo, sendo estes considerados potenciais *outliers*.

Qual a categoria com mais reviews positivas?

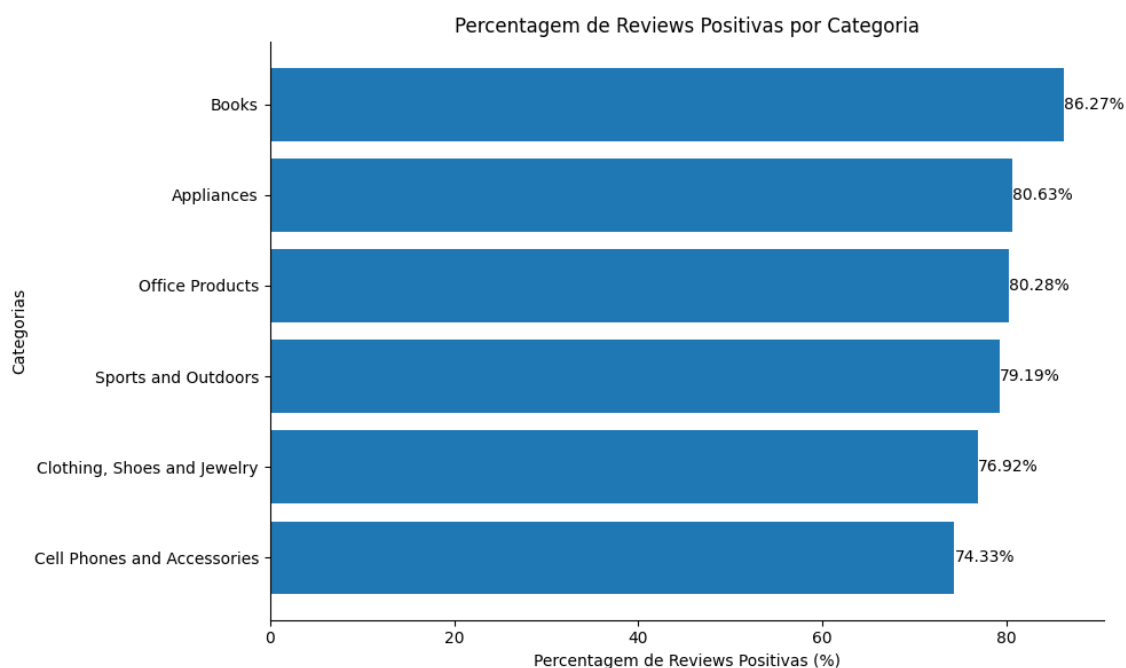


Figura 9: Percentagem de Reviews Positivas por categoria

A Figura 9 apresenta a percentagem de avaliações positivas (classificações elevadas) por categoria de produto. Observa-se uma tendência clara para avaliações tendencialmente favoráveis em todas as categorias analisadas, com valores que variam entre aproximadamente 74% e 86%.

A categoria de Livros destaca-se com a maior proporção de avaliações positivas (86,27%), seguida de Eletrodomésticos e Produtos de Escritório, ambas com cerca de 80%. Por outro lado, a categoria com menor percentagem de avaliações positivas é Telemóveis e Acessórios, com 74,33%

Qual a categoria com mais compras verificadas?

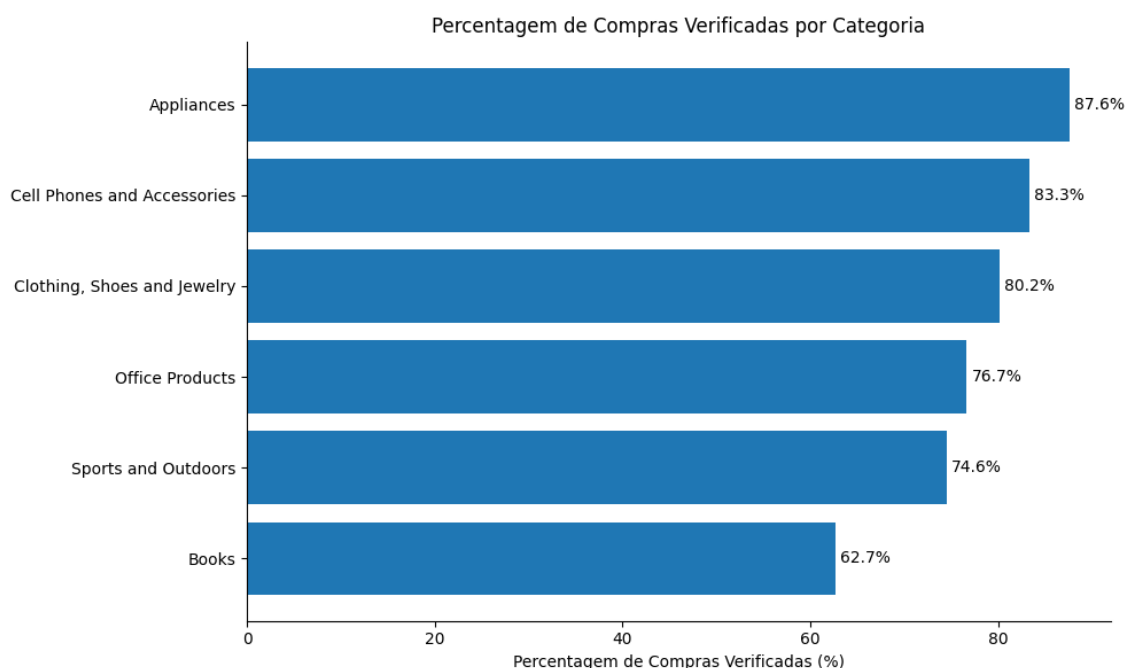


Figura 10: Percentagem de Compras Verificadas por Categoria

A Figura 10 apresenta a percentagem de avaliações associadas a compras verificadas em cada uma das categorias analisadas. Esta métrica é frequentemente utilizada como um indicador da autenticidade das avaliações, partindo do pressuposto de que reviews associadas a compras reais são, à partida, mais fiáveis.

Verifica-se que a categoria de Eletrodomésticos apresenta a maior percentagem de compras verificadas (87,6%), seguida por Telemóveis e Acessórios (83,3%) e Roupas, Sapatos e Joalharia (80,2%). Por contraste, a categoria de Livros apresenta o valor mais baixo (62,7%).

No entanto, importa sublinhar que a presença de compras verificadas, embora relevante, não deve ser interpretada isoladamente como prova de veracidade. A categoria de Livros, apesar de apresentar a percentagem mais baixa de verificações, é justamente aquela que, segundo um estudo da Mozilla, contém a maior proporção de avaliações genuínas [4].

Assim, esta métrica deve ser interpretada em conjunto com outras variáveis comportamentais e textuais, no âmbito de uma análise mais abrangente.

5.5 Conclusão da Análise Exploratória

A análise exploratória realizada permitiu identificar diferenças entre as várias categorias de produtos no que diz respeito às características das avaliações, dos produtos e dos utilizadores. Os padrões de comportamento observados são, em grande medida, semelhantes entre categorias. Embora se tenham verificado algumas variações — por exemplo, na frequência de publicação ou na percentagem de compras verificadas.

Este resultado sugere que a dimensão categorial pode influenciar certas métricas,

Entre as *features* analisadas, aquelas que apresentaram maiores diferenças entre categorias foram:

- **Comprimento médio das avaliações:** a categoria de Livros destacou-se com valores claramente superiores, tanto em média como em variabilidade, indicando comentários mais elaborados e detalhados. Em contraste, categorias como Roupas, Sapatos e Joalheria apresentaram textos substancialmente mais curtos.
- **Número médio de avaliações por produto:** Telemóveis e Acessórios e Eletrodomésticos apresentam valores mais altos.
- **Burstness:** esta métrica revelou maior concentração de atividade num curto intervalo de tempo na categoria, como Roupas, Sapatos e Joalheria.
- **Percentagem de reviews positivas e compras verificadas:** as diferenças aqui também são relevantes, com Livros a destacar-se pela elevada proporção de avaliações positivas, apesar de ter a menor percentagem de compras verificadas.

Estas variações sugerem que os padrões de comportamento dos utilizadores estão fortemente ligados ao tipo de produto, o que deve ser considerado no desenvolvimento de modelos de deteção de *reviews* falsas. A compreensão destas diferenças é essencial para ajustar os critérios de deteção de *reviews* falsas e minimizar falsos positivos ou negativos decorrentes das especificidades de cada categoria.

6 - Resultados e Discussão

O objetivo central é identificar diferentes perfis de utilizadores com base no seu comportamento ao escrever avaliações, bem como analisar características que podem estar associadas a padrões invulgares ou suspeitos.

Para tal também aplicadas técnicas de agrupamento e deteção de *outliers* para segmentar os utilizadores com base nas suas características.

Os resultados são apresentados com recurso a gráficos e explicações que contextualizam os dados face aos objetivos do estudo. Sempre que se justifica, são destacadas tendências inesperadas ou padrões consistentes entre diferentes grupos de utilizadores.

No final, discute-se o valor de cada métrica para a identificação de comportamentos fora do comum e o seu contributo potencial para o desenvolvimento de modelos de deteção de avaliações falsas.

6.1 Agrupamento dos Utilizadores

Após a análise exploratória, foi aplicado o algoritmo de agrupamento *K-Means* ao conjunto de dados. O principal objetivo desta abordagem não supervisionada foi identificar grupos com comportamentos semelhantes.

As variáveis utilizadas incluíram: número médio de estrelas atribuídas (*avg rating*), número total de avaliações realizadas, comprimento médio das avaliações em palavras, número máximo de avaliações por dia e por hora, número de categorias distintas avaliadas e percentagem de compras verificadas.

Após testes com vários valores de k , onde k representa o número de grupos a formar no algoritmo de agrupamento *K-Means*, a escolha de $k = 7$ revelou-se mais informativa, permitindo distinguir grupos com características distintas.

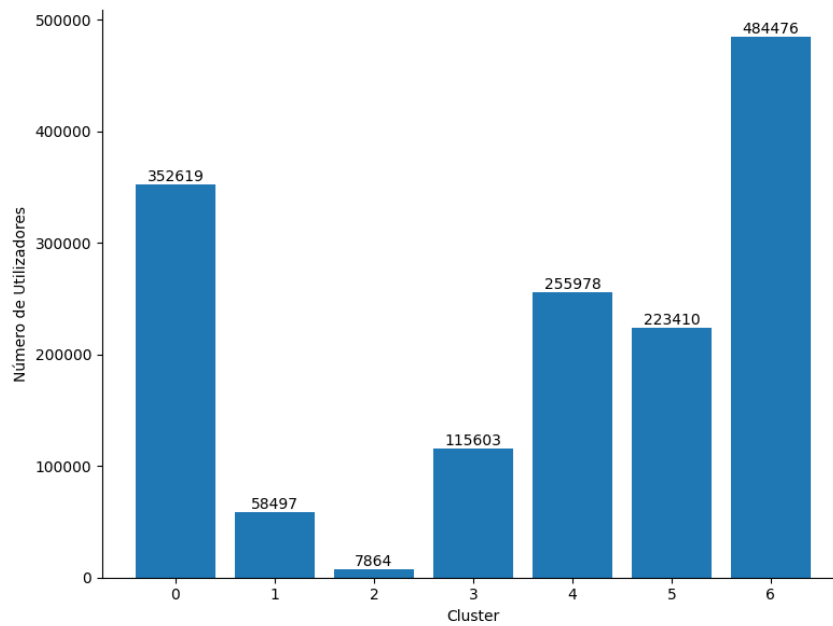


Figura 11: Número de Utilizadores por Cluster

A Figura 11 apresenta a distribuição do número de utilizadores por cluster, resultante da aplicação do algoritmo K-Means com $k = 7$. Observa-se que os Clusters 0 e 6 concentram a maioria dos utilizadores, representando, em conjunto, uma proporção significativa do total analisado. Por outro lado, o Cluster 2 inclui um número muito mais reduzido de utilizadores, o que pode refletir um grupo com características comportamentais menos comuns ou simplesmente menos representado nos dados. Esta distribuição assimétrica evidencia a diversidade de padrões de utilização presentes no conjunto de dados.

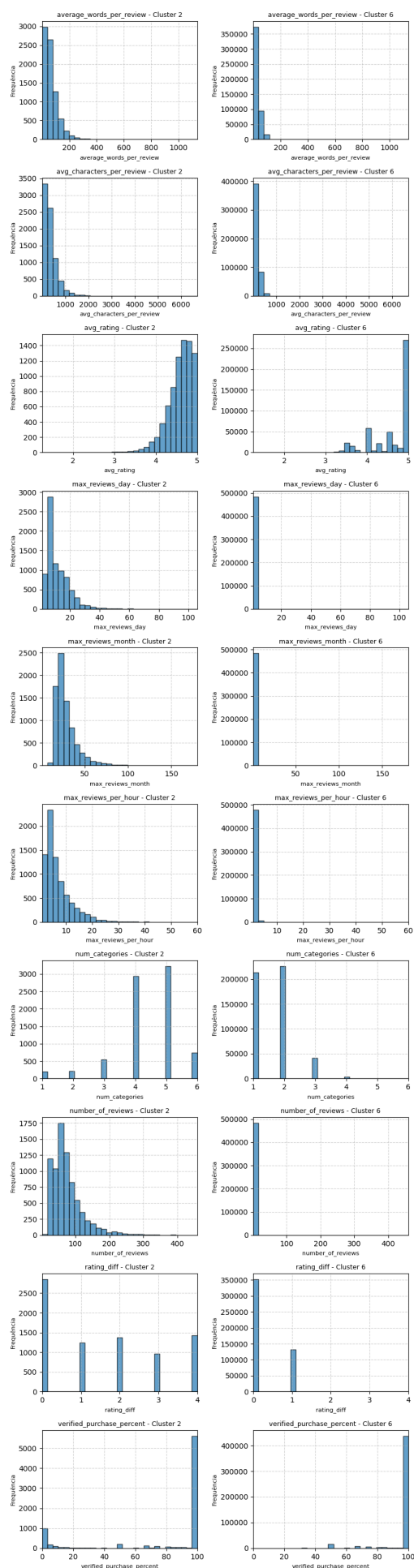


Figura 12: Distribuição das variáveis seleccionadas para os Clusters 2 e 6, obtidos pelo algoritmo K-Means.

A Figura 12 mostra a distribuição das principais variáveis comportamentais e textuais utilizadas na segmentação dos utilizadores, com destaque para os Clusters 2 e 6. Estes foram selecionados por apresentarem características contrastantes e por permitirem ilustrar com clareza a diversidade de perfis existentes.

O Cluster 6 representa uma parte significativa dos utilizadores e reflete comportamentos típicos: número total elevado de avaliações, baixa frequência horária e diária, alta percentagem de compras verificadas e avaliações geralmente homogêneas. O Cluster 0, embora numericamente semelhante ao Cluster 6, apresenta um perfil praticamente sobreponível, não oferecendo valor adicional para a análise comparativa.

O Cluster 2, por sua vez, é o menos representado, mas destaca-se por agregar utilizadores com comportamentos diferenciados: maior intensidade de atividade em curtos períodos de tempo, avaliações com maior dispersão nas classificações (ratings) e maior diversidade de produtos avaliados. Estas características fazem deste grupo um caso interessante para estudo mais aprofundado, sobretudo no contexto de deteção de comportamentos atípicos ou suspeitos.

Esta análise demonstra que o algoritmo K-Means foi eficaz na identificação de grupos de utilizadores com perfis distintos. Embora não se atribuam juízos de valor aos segmentos, estas diferenças podem contribuir para uma compreensão mais profunda da heterogeneidade nas interações em plataformas de avaliação, servindo de base para análises futuras, como a deteção de comportamentos anómalos ou fraudulentos.

6.2 Deteção de Outliers

Para além da segmentação comportamental realizada com o algoritmo *K-Means*, foi aplicada uma abordagem complementar baseada no modelo *Isolation Forest*, com o objetivo de identificar utilizadores cujo perfil de atividade se destaca significativamente em relação ao comportamento geral da amostra.

A deteção de *outliers* revela-se especialmente útil neste contexto, uma vez que permite isolar casos que se afastam dos padrões mais comuns, seja por uma elevada intensidade de participação, por características textuais invulgares ou por combinações atípicas de variáveis comportamentais. Importa reforçar que a classificação como *outlier* não implica qualquer julgamento sobre a intenção do utilizador, mas apenas assinala um desvio que poderá justificar uma análise mais atenta.

Para compreender melhor as características destes utilizadores, procedeu-se à análise da distribuição das principais variáveis, comparando os perfis classificados como *outliers* com os restantes. A Figura 13 ilustra essas diferenças. De forma geral, os *outliers* tendem a apresentar, em vários casos, maior frequência de publicações em curtos intervalos de tempo e uma maior diversidade de categorias avaliadas. Estes padrões, embora não concludentes, destacam perfis cuja atividade se afasta do padrão dominante e que, por isso, podem influenciar de forma particular a dinâmica das avaliações.

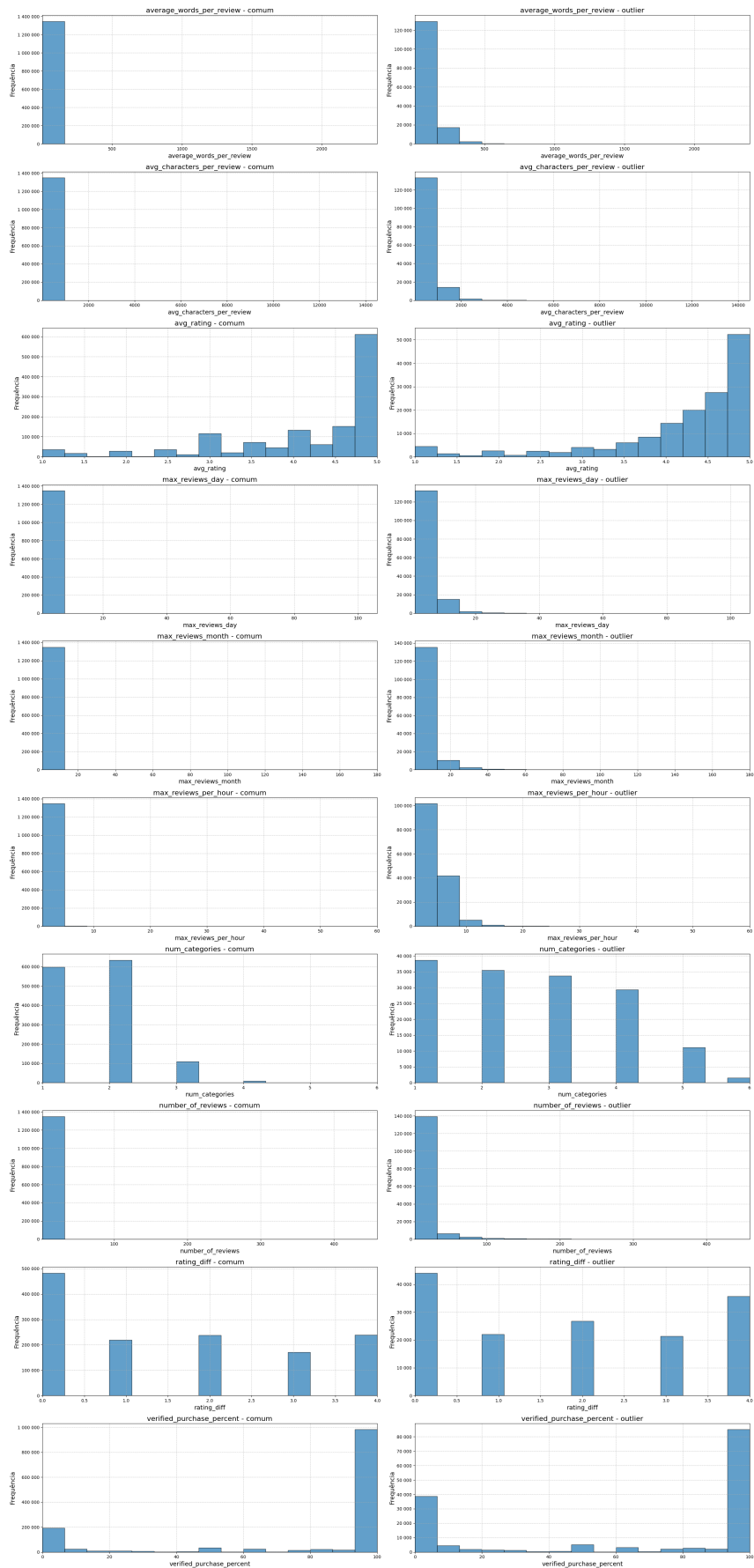


Figura 13: Distribuição das variáveis para utilizadores classificados como comuns e como *outliers*, segundo o modelo Isolation Forest.

6.3 Análise Conjunta dos Resultados

A Figura 14 apresenta a distribuição dos utilizadores classificados como *outliers* em cada um dos clusters previamente definidos através do algoritmo K-Means. De forma geral, observa-se que a maioria dos clusters é dominada por utilizadores classificados como comuns, sendo esse o caso dos Clusters 0, 4, 5 e 6, onde a proporção de *outliers* é bastante reduzida. Tal poderá indicar que estes grupos representam padrões de comportamento mais típicos no contexto do conjunto de dados analisado.

Por outro lado, os Clusters 1, 2 e 3 destacam-se por conterem proporções mais elevadas de *outliers*. Estes são também, coincidentemente, os três clusters com menor número total de utilizadores, o que sugere que agregam perfis comportamentais menos comuns. Análises anteriores já haviam mostrado que o Cluster 1 inclui utilizadores com elevada frequência de atividade e um número elevado de avaliações em curtos intervalos. O Cluster 3, por sua vez, agrupa utilizadores que escrevem comentários muito curtos, com baixa diversidade de categorias avaliadas e padrões de publicação mais concentrados.

Características como picos de publicação em curtos períodos (*burstness*), avaliações genéricas e pouco detalhadas, e baixa diversidade de categorias têm sido associadas na literatura a padrões frequentemente observados em contextos de *fake reviews* [9]. No entanto, importa sublinhar que o conjunto de dados utilizado neste trabalho foi previamente filtrado para excluir, tanto quanto possível, avaliações artificiais. Assim, a presença de perfis com estas características coloca duas hipóteses plausíveis: por um lado, podem corresponder a utilizadores legítimos com estilos de participação menos comuns; por outro, é possível que os mecanismos de filtragem aplicados não tenham sido totalmente eficazes na remoção de conteúdos falsos. Não existe, contudo, informação suficiente no conjunto de dados para confirmar ou refutar qualquer uma destas possibilidades, sendo a classificação como *outlier* um indício exploratório que justifica atenção adicional, mas não uma prova de comportamento indevido.

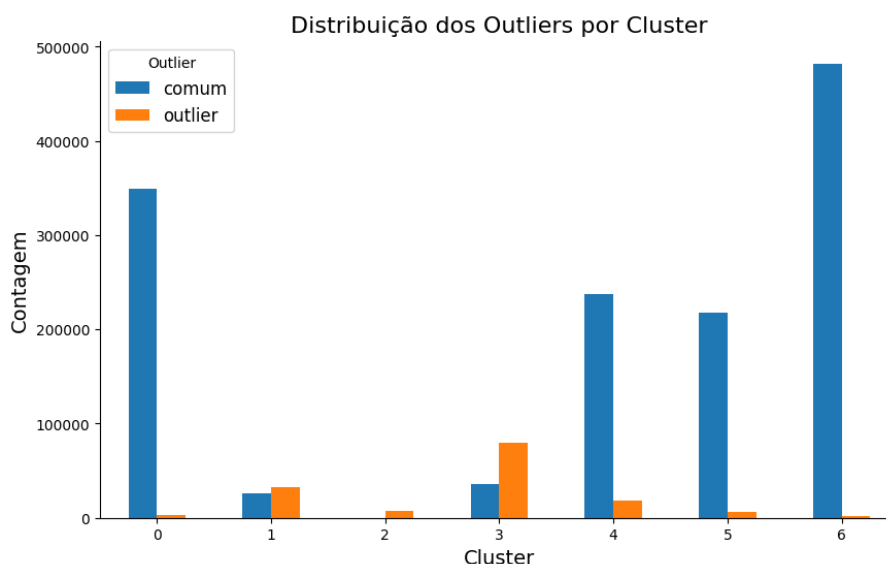


Figura 14: Distribuição dos outliers detetados pelo Isolation Forest em cada cluster

6.4 Análise por Categoria de Produto

Para aprofundar a análise comportamental dos utilizadores, foi realizada uma abordagem segmentada por categoria de produto. Esta distinção permite observar se o comportamento dos utilizadores varia consoante o tipo de produto avaliado, o que é particularmente relevante num cenário tão diverso como o das plataformas de *e-commerce*.

Segmentação com K-Means

Inicialmente, o algoritmo *K-Means* foi aplicado individualmente a cada categoria. Os resultados revelam que, embora a estrutura de clusters partilhe algumas semelhanças com a análise global, surgem diferenças relevantes ao nível da distribuição das variáveis dentro de cada grupo. Verificaram-se alterações na proporção de utilizadores por cluster, bem como nos perfis identificados como menos comuns.

Em algumas categorias, como Telemóveis e Acessórios ou Livros, determinados clusters assumem características mais vincadas, como comprimentos médios de avaliação mais curtos, atividade mais intensa ou baixa diversidade de categorias avaliadas. Estas diferenças demonstram que certos comportamentos considerados atípicos na análise global podem, em contexto categórico, assumir maior frequência ou, inversamente, revelar padrões ainda mais extremos. Este resultado evidencia que o comportamento dos utilizadores é, até certo ponto, dependente do tipo do produto.

Deteção de Outliers com Isolation Forest

De forma complementar, foi também usada a técnica *Isolation Forest* separadamente em cada categoria, com o objetivo de identificar utilizadores cujo comportamento se destaca dos restantes dentro de cada grupo de produtos. Tal como na análise global, a classificação como *outlier* não implica qualquer juízo sobre a veracidade das avaliações, servindo apenas como sinal de comportamento excecional.

As distribuições obtidas mostram que os padrões de *outliers* variam de forma considerável entre categorias. Em Roupas, Sapatos e Joalharia, por exemplo, observam-se utilizadores com reviews muito curtas, alta frequência de submissão e baixa percentagem de compras verificadas. Já em Produtos de Escritório, os *outliers* destacam-se sobretudo por concentrarem a sua atividade em poucas categorias. Por outro lado, categorias como Livros ou *Appliances* apresentam perfis menos extremos, o que poderá sugerir maior dispersão ou diversidade nos estilos de participação.

Estas diferenças reforçam a importância de considerar o tipo de produto ao desenvolver modelos de deteção de comportamentos invulgares. Um modelo treinado sobre o conjunto global poderá não captar nuances importantes presentes em contextos mais específicos.

Por fim, importa lembrar que os dados analisados foram previamente filtrados com o objetivo de remover avaliações artificiais. Assim, os perfis destacados nesta secção podem corresponder tanto a utilizadores legítimos com comportamentos fora do comum como a casos que escaparam aos mecanismos de filtragem. Sem informação adicional, não é possível validar a natureza exata destes perfis, sendo a sua identificação útil apenas como ponto de partida para análise mais aprofundada.

6.5 Limitações do Trabalho

Durante a fase de experimentação, foram considerados outros métodos de *clustering* com potencial para complementar ou melhorar os resultados obtidos com o K-Means e com o Isolation Forest. Entre os métodos avaliados encontram-se o *Spectral Clustering* [17] e o DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) [18], que identifica *clusters* com base na densidade dos pontos e permite, simultaneamente, detetar *outliers*.

No entanto, tanto o *Spectral Clustering* como o DBSCAN revelaram-se inviáveis de aplicar de forma eficaz neste estudo, devido à elevada dimensão do conjunto de dados. As limitações computacionais, nomeadamente o consumo excessivo de memória e o tempo de execução prolongado, impossibilitaram a sua utilização prática na análise. Assim, optou-se por desenvolver a análise com base no K-Means e no Isolation Forest, métodos mais adequados ao volume dos dados disponíveis.

Para além dos constrangimentos computacionais, existem também limitações associadas ao próprio conjunto de dados utilizado. Embora este tenha sido previamente filtrado para remover avaliações falsas, o processo de filtragem não é conhecido. Tal compromete uma validação rigorosa dos perfis detetados como atípicos: não é possível saber com certeza se representam comportamentos legítimos, mas invulgares, ou se são avaliações artificiais que escaparam ao processo de remoção.

Esta incerteza compromete, em parte, a interpretação dos padrões mais extremos identificados ao longo da análise. A presença de utilizadores com elevada frequência de publicações, comentários muito curtos ou baixa diversidade de categorias pode refletir tanto um estilo legítimo de participação como um indício de comportamento manipulativo. Como já referido sem informação adicional, não é possível validar essas hipóteses com rigor.

7 - Planeamento

A elaboração do planeamento levou em consideração os prazos estipulados e a complexidade de cada tarefa, permitindo que o trabalho progreda de forma eficiente e organizada. A planificação geral é apresentada na Figura 15 através de um gráfico de Gantt.

A fase inicial do projeto, entre 13 de outubro e 1 de dezembro de 2024, foi dedicada à definição do problema e à revisão da literatura. Durante este período, procedeu-se à identificação clara do tema, à formulação dos objetivos e à análise crítica de trabalhos anteriores, de modo a contextualizar e fundamentar a relevância do estudo.

Entre 1 de novembro e 31 de dezembro de 2024 decorreu o pré-processamento dos dados. Esta etapa envolveu a recolha, organização e limpeza inicial do conjunto de dados, com o objetivo de garantir a sua qualidade e consistência para etapas futuras.

No mês de janeiro de 2025, realizou-se a análise exploratória dos dados. Esta fase permitiu uma compreensão inicial dos dados, identificação de padrões e possíveis problemas, e serviu de base para a escolha das abordagens a aplicar.

O desenvolvimento da solução teve lugar entre 1 de fevereiro e 15 de abril de 2025. Durante este período, foram implementados e ajustados os algoritmos de deteção de *fake reviews*, acompanhados por testes e validações preliminares.

Por fim, entre 15 de abril e 27 de junho de 2025, decorreu a fase de análise e discussão dos resultados. Esta etapa consistiu principalmente no aperfeiçoamento da abordagem desenvolvida, com o objetivo de obter resultados mais robustos e representativos. Em vez de se focar nas implicações práticas da investigação, esta fase foi orientada para a afinação de parâmetros, revisão de técnicas e análise crítica da eficácia dos métodos utilizados face aos objetivos inicialmente definidos.

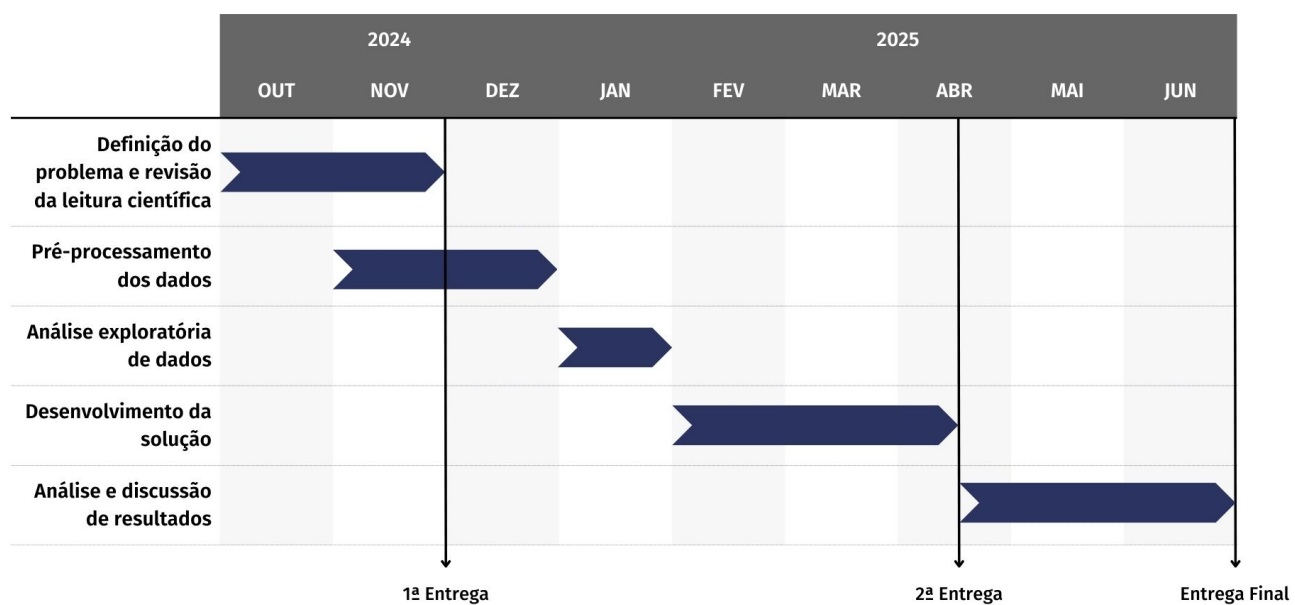


Figura 15: Gráfico GANTT do planeamento

8 - Conclusão

O principal objetivo deste trabalho foi identificar padrões de comportamento associados à publicação de avaliações em plataformas de *e-commerce*, recorrendo a métodos de análise não supervisionada. Pretendeu-se compreender como diferentes métricas comportamentais e textuais podem ajudar a distinguir entre perfis de utilizador comuns e perfis com atividade invulgar, sem depender de dados previamente rotulados.

A análise exploratória detalhada revelou diferenças entre categorias de produtos, mostrando que variáveis como o comprimento das avaliações, a frequência de publicação e a percentagem de compras verificadas são fundamentais para caracterizar comportamentos.

A aplicação do *K-Means* permitiu segmentar os utilizadores em grupos com perfis distintos, destacando padrões de participação mais comuns e outros mais invulgares. Por sua vez, o *Isolation Forest* mostrou-se eficaz na deteção de *outliers*. A combinação destas abordagens revelou-se valiosa, permitindo identificar perfis atípicos sem depender de dados rotulados.

Do ponto de vista prático, os resultados desta análise podem ser integrados em plataformas de comércio digital como um sistema auxiliar de moderação de conteúdo. Especificamente, os perfis de utilizador identificados como invulgares ou atípicos podem ser sinalizados automaticamente, permitindo que equipas de moderação analisem esses casos com maior prioridade. Esta abordagem não substitui a validação humana, mas funciona como um mecanismo de triagem que melhora a eficiência na deteção de possíveis abusos ou manipulação de avaliações.

Por exemplo, quando o sistema deteta uma *review* suspeita, a plataforma poderá optar por ocultá-la temporariamente, sinalizá-la com um alerta de possível falsidade ou submetê-la a uma verificação adicional antes da sua publicação. Esta abordagem contribuiria não só para proteger os consumidores de informações enganosas, mas também para reforçar a confiança nas plataformas e promover práticas comerciais mais justas.

Por fim, a investigação realizada fornece uma base sólida para futuros trabalhos que explorem modelos híbridos supervisionados/não supervisionados ou a integração de dados qualitativos que permitam validar as suspeitas levantadas, contribuindo para a criação de sistemas automáticos mais precisos.

Bibliografia

- [1] João P. Matos-Carvalho. *The Lusófona L^AT_EX Template User's Manual*. Lusófona University. 2024. URL: <https://github.com/jpmcarvalho/UL-Thesis>.
- [2] Rami Mohawesh et al. "Fake reviews detection: A survey". Em: *Ieee Access* 9 (2021), pp. 65771–65802.
- [3] Ramadhani Ally Duma et al. "Fake review detection techniques, issues, and future research directions: a literature review". Em: *Knowledge and Information Systems* (2024), pp. 1–42.
- [4] Mozilla. *Fakespot reveals the product categories with the most and least reliable product reviews for summer and back-to-school shopping*. <https://blog.mozilla.org/en/mozilla/news/fakespot-reveals-the-product-categories-with-the-most-and-least-reliable-product-reviews-for-summer-and-back-to-school-shopping/>. 2024.
- [5] H. Li et al. "Bimodal Distribution and Co-Bursting in Review Spam Detection". Em: *Proceedings of the 26th International Conference on World Wide Web (WWW)*. Abr. de 2017, pp. 1063–1072.
- [6] Y. Li. "Highlighting the fake reviews in review sequence with the suspicious contents and behaviours". Em: *Journal of Information and Computational Science* 12.4 (mar. de 2015), pp. 1615–1627.
- [7] A. Mukherjee et al. "Spotting opinion spammers using behavioral footprints". Em: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. Ago. de 2013, pp. 632–640.
- [8] F. H. Li et al. "Learning to Identify Review Spam". Em: *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*. 2011, p. 16.
- [9] A. Mukherjee, B. Liu e N. Glance. "Spotting Fake Reviewer Groups in Consumer Reviews". Em: *Proceedings of the 21st International Conference on World Wide Web (WWW)*. 2012, pp. 191–200.
- [10] Ahmed M Elmogy et al. "Fake reviews detection using supervised machine learning". Em: *International Journal of Advanced Computer Science and Applications* 12.1 (2021).
- [11] R Poonguzhali et al. "Fake reviews detection using support vector machine". Em: *2022 International conference on sustainable computing and data communication systems (ICSCDS)*. IEEE. 2022, pp. 1509–1512.
- [12] Kaushik Daiv et al. "An approach to detect fake reviews based on logistic regression using review-centric features". Em: *Int. Res. J. Eng. Technol.(IRJET)* 7.06 (2020), pp. 2107–2112.
- [13] Huy Le e Ben Kim. "Detection of fake reviews on social media using machine learning algorithms". Em: *Issues in Information Systems* 21.1 (2020), pp. 185–194.
- [14] Bhuvaneshwari P., Rao A. N. e Robinson Y. H. "Spam Review Detection Using Self-Attention Based CNN and Bi-Directional LSTM". Em: *Multimedia Tools and Applications* 80.12 (2021), pp. 18107–18124. DOI: 10.1007/s11042-021-10602-y. URL: <https://doi.org/10.1007/s11042-021-10602-y>.

- [15] Mohiuddin Ahmed, Raihan Seraj e Syed Mohammed Shamsul Islam. "The k-means Algorithm: A Comprehensive Survey and Performance Evaluation". Em: *Electronics* 9.8 (2020). ISSN: 2079-9292. DOI: 10.3390/electronics9081295. URL: <https://www.mdpi.com/2079-9292/9/8/1295>.
- [16] Fei Tony Liu, Kai Ming Ting e Zhi-Hua Zhou. "Isolation forest". Em: *2008 eighth ieee international conference on data mining*. IEEE. 2008, pp. 413–422.
- [17] Ulrike Von Luxburg. "A tutorial on spectral clustering". Em: *Statistics and computing* 17 (2007), pp. 395–416.
- [18] Shuo Yu et al. "Graph learning for fake review detection". Em: *Frontiers in Artificial Intelligence* 5 (2022), p. 922589.