

## Avaliação Intermediária

Marco Moliterno Pena Piacentini (Insper, São Paulo, Brasil)

### 1 Introdução

Descobrir os fatores que mais aborrecem os clientes em suas compras online é um desafio para as empresas varejistas, pois essas informações podem ser difíceis de identificar a partir de dados não estruturados, como comentários de clientes. No entanto, entender o que os clientes não gostam ou o que pode ser melhorado em suas experiências de compra é fundamental para melhorar a satisfação do cliente e aumentar a fidelidade do cliente. Nesse sentido, as empresas podem recorrer a técnicas de análise de sentimentos e outras técnicas de processamento de linguagem natural para identificar os fatores que mais afetam a experiência do cliente em compras online.

### 2 Metodologia usada

Foi usada a base de dados disponibilizada pela Olist em:

[www.kaggle.com/datasets/olistbr/brazilian-ecommerce](https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce)

Primeiramente, foi feito um join entre "olist\_orders\_dataset.csv" e "olist\_order\_reviews\_dataset.csv", para associar as reviews aos dados de seus pedidos.

Então o dataset foi separado de acordo com o review score. Assim permitindo a análise individual de cada nível de satisfação do cliente.

Para identificar as maiores causas de remorso nos clientes, foi calculada uma vetorização TFIDF das palavras usadas nas review de pedidos com nível de satisfação 1. Então foi calculada a soma do TFIDF de cada palavra para todas as reviews.

Com isso, foi possível identificar as palavras que apareceram com maior frequência nas reviews baixas. Dentre elas, foram filtradas palavras conectivas e genéricas (de, do, etc), finalmente possibilitando a análise das 30 palavras mais relevantes.

### 3 Resultados

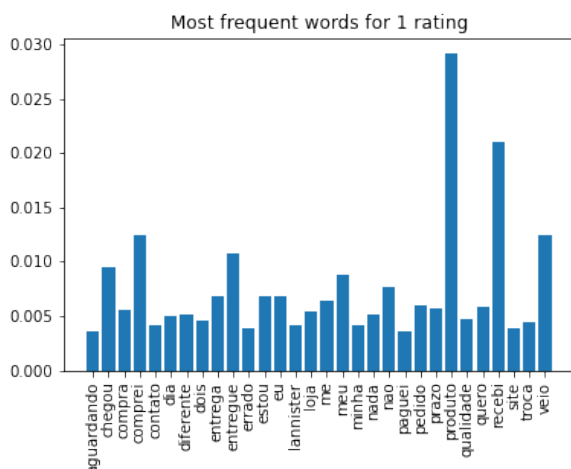


Figure 1: Soma de TFIDF das 30 palavras mais relevantes em reviews de valor 1

Analisando o gráfico, seis palavras se destacam: "produto", "recebi", "veio", "comprei", "entregue" e "chegou".

Dessas seis, três ("veio", "recebi", "entregue" e "chegou"), indicam que as reviews negativas tinham alguma relação com a entrega, além disso a presença de palavras como "aguardando", "prazo" e "dia" apontam para uma grande fonte de aborrecimento dos clientes ser o tempo de entrega.

### 4 Validação

Para validar essa hipótese, foi calculado o tempo entre a compra e a entrega de cada pedido, ou seja, o tempo que o cliente teve que esperar.

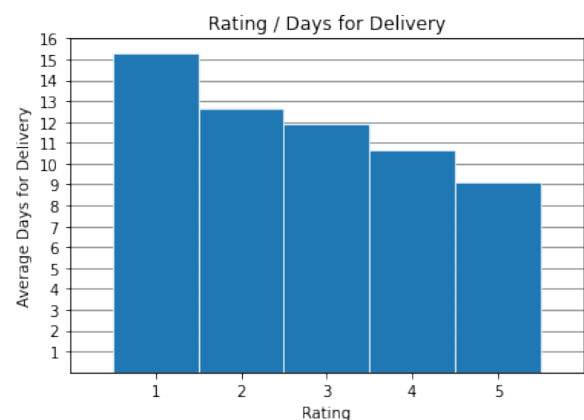


Figure 2: Média de dias para entrega para cada rating

Como pode ser visto no gráfico acima, há uma clara correlação entre satisfação do cliente e tempo de entrega, já que em média clientes que tiveram que esperar mais deram notas piores em suas reviews. Isso é consistente com a hipótese tirada da análise do TFIDF.

Portanto, pode-se concluir que: **A maior fonte de aborrecimento dos clientes é o tempo de espera para entrega.**

### 5 Outras considerações

A hipótese proposta não é a única causa de aborrecimento dos clientes.

Outras possibilidades incluem algo de errado no produto entregue (indicado pela presença de palavras como "errado", "diferente", "troca" e "qualidade"), o preço pago (indicado pela palavra "paguei"), ou algum problema nos recursos da loja online (indicado pelas palavras "loja", "site" e "contato").

### 6 Repositório do código

[https://github.com/Marcompnp/nlp\\_avalia-o](https://github.com/Marcompnp/nlp_avalia-o)