

UMA ABORDAGEM COMPUTACIONAL PARA A CATEGORIZAÇÃO AUTOMÁTICA DE RECLAMAÇÕES EM COMÉRCIOS ELETRÔNICOS UTILIZANDO PERCEPTRON MULTICAMADAS

Matheus Konrad Xavier¹

Renato Balancieri²

RESUMO

O comércio eletrônico vem tendo um crescimento impressionante ao longo dos últimos anos e com o aumento da concorrência nesse mercado, os clientes começaram a querer ser ouvidos cada vez mais. Com a grande escala dos comércios eletrônicos e o número de reclamações, surge a necessidade de automatizar o processo de classificação para que elas cheguem ao setor responsável e sejam resolvidas o mais rápido possível. A proposta deste trabalho foi desenvolver uma abordagem computacional que automatizasse esse processo. Para isso, foi utilizada a metodologia CRISP-DM (*Cross Industry Standard Process for Data Mining*), que categorizou 600 reclamações reais de três comércios eletrônicos atuantes no Brasil por meio do site Reclame Aqui. Porém, como uma reclamação pode ser a respeito de inúmeros problemas diferentes, este trabalho se propôs a classificar apenas problemas relacionados à entrega, ao produto e ao pagamento. O algoritmo escolhido para realizar a tarefa de categorização foi o Perceptron de Multicamadas, o qual foi alimentado gradualmente com 25 reclamações dos assuntos relacionados à entrega, produto e pagamento. Conforme o modelo foi sendo alimentado, ele foi evoluindo na capacidade de classificar os problemas, chegando a obter 85% de acurácia e média superior a 80% entre todas as métricas coletadas, como precisão, *recall* e *F1-Score*.

PALAVRAS-CHAVE: Classificação; Aprendizado de máquina; Mineração de texto.

INTRODUÇÃO

O comércio eletrônico já vinha tendo um crescimento impressionante ao longo dos últimos anos, mas com o surgimento da pandemia de coronavírus no ano de 2020, esse crescimento se tornou ainda maior, pois, de acordo com Silva (2022), o modo como as pessoas realizavam suas transações financeiras e econômicas acabou mudando.

Segundo dados da Abcomm (2023) (Associação Brasileira de Comércio Eletrônico), o faturamento do comércio eletrônico em 2023 chegou aos 185,7 bilhões de reais, um

1 - Universidade estadual de Maringá, matheuskonradxavier@hotmail.com

2 - Universidade estadual do Paraná, renato.balancieri@unespar.edu.br

crescimento de 9,5% em relação ao ano anterior. Com esse alto faturamento, criou-se um ambiente cada vez mais competitivo (SILVA, 2022).

Entretanto, com a grande escala que os comércios eletrônicos tomaram hoje em dia, nem sempre é possível manter um atendimento de alto nível, resultando em problemas ocasionalmente. Nesse momento, com o objetivo de não perder a lealdade de um cliente, é muito importante que esse problema seja encaminhado ao setor responsável e resolvido o mais rápido possível.

A categorização automática de reclamações em comércios eletrônicos surge como o tema deste trabalho, com o objetivo de auxiliar no desafio crescente que é gerenciar e lidar com um número de reclamações cada vez maior de uma forma eficiente. Este trabalho envolve o uso de técnicas de processamento de linguagem natural (PLN), aprendizado de máquina e mineração de texto para categorizar de maneira automática as reclamações em categorias específicas.

O objetivo deste trabalho é desenvolver uma abordagem computacional capaz de classificar automaticamente reclamações realizadas por clientes em comércios eletrônicos em categorias previamente definidas. E a partir do objetivo geral, desdobram-se os seguintes objetivos específicos:

- Definir os tópicos relevantes a serem categorizados;
- Estabelecer uma base de dados com reclamações reais relacionadas aos tópicos relevantes;
- Determinar e aplicar um algoritmo de aprendizado supervisionado para aprender com reclamações reais;
- Avaliar os resultados e as métricas extraídas do modelo, evidenciando se ele obteve sucesso ou não na classificação de reclamações.

METODOLOGIA

O modelo usado de referência para esta pesquisa foi o *Cross Industry Standard Process for Data Mining* (CRISP-DM).

Na fase de compreender o contexto, foi visualizado que uma reclamação pode ser classificada de inúmeras formas, pois elas podem ser a respeito dos altos preços, atendimento, insatisfação com a compra e entre outras coisas. Então, este trabalho tem como foco classificar apenas as categorias com o maior índice de ocorrência na plataforma Reclame Aqui, que são as relacionadas a problemas com o produto, o pagamento e com a entrega.

Como não existe um padrão claro e absoluto para classificar esses tópicos, torna-se necessário o uso de um modelo de aprendizado supervisionado. Esse modelo deve ser capaz de aprender com reclamações anteriores, a fim de extrair padrões e palavras mais utilizadas, para conseguir classificar corretamente e de forma padronizada reclamações posteriores.

Na fase de compreender os dados, com o objetivo do modelo se tornar mais eficiente, torna-se necessário encontrar uma base de dados de reclamações reais dentro dos tópicos propostos. Então, para satisfazer essa necessidade, este trabalho optou por extrair reclamações reais de três comércios eletrônicos que atuam em solo brasileiro, que são: Amazon, Magazine Luiza e Submarino. Após a escolha dos comércios eletrônicos, alguns desafios foram encontrados, como a falta de padrão entre os sites, nenhum tipo de classificação prévia, a falta da concentração de reclamações em um só lugar, entre outras coisas.

A fim de solucionar todos os problemas anteriormente descritos, a fonte de reclamações reais foi alterada para a plataforma Reclame Aqui, que conta com uma base padronizada de reclamações a respeito de diversas empresas, incluindo as selecionadas anteriormente. Então, em seguida, foi utilizada uma técnica de raspagem de dados para realizar a coleta das reclamações as quais foram validadas e classificadas manualmente pelo autor.

Na fase de preparação de dados, foi realizada uma correção ortográfica em todas as reclamações, a fim de padronizar as palavras em todas as reclamações e, consequentemente, obter um melhor resultado quanto à frequência e importância de cada palavra em uma reclamação. Após todas as reclamações terem sido corrigidas adequadamente, algumas técnicas de processamento de linguagem natural foram aplicadas. Começando pela geração de *tokens*, seguida pela remoção das *stopwords* e por último o *stemming*.

Na fase de modelagem, o algoritmo escolhido foi o Perceptron de Múltiplas Camadas, escolhido pois consegue resolver problemas não-lineares e de classificação multiclasse (FALCÃO et al., 2013). Para o seu treinamento, foram usadas 300 reclamações, sendo 100 de cada tópico, classificadas manualmente pelo autor. Elas foram apresentadas ao modelo de maneira gradual, aumentando 25 reclamações de treino em cada tópico a cada rodada de teste, para medir seu desempenho e verificar quantas reclamações são necessárias para se obter um resultado satisfatório. As reclamações foram apresentadas ao modelo pelo valor resultante do algoritmo TF-IDF, para que ele aprendesse as palavras mais relevantes a respeito de cada tópico e pudesse classificar as reclamações posteriormente.

Na fase de avaliação e interpretação, foram realizados testes com 300 reclamações já rotuladas extraídas do Reclame Aqui e classificadas manualmente pelo autor, realizando a

classificação e comparando com os resultados esperados, tendo como objetivo extrair métricas estatísticas desses resultados, como a acurácia, *recall* e *F1-Score*.

A implantação é a última etapa, onde o conhecimento obtido é organizado e demonstrado para os envolvidos. Neste trabalho, os conhecimentos serão demonstrados nos resultados e discussões, por meio de métricas estatísticas extraídas do modelo.

RESULTADOS E DISCUSSÕES

Como base de dados neste experimento, foram utilizadas 600 reclamações reais extraídas da plataforma Reclame Aqui, sendo 200 de cada tópico proposto na fase de compreensão do contexto, todas essas reclamações foram classificadas manualmente pelo autor. Para o treinamento e avaliação do modelo, esse grupo de 600 reclamações foi dividido em 2 grupos de 300, ambos os grupos contavam com 100 reclamações de cada tópico.

Com o objetivo de avaliar o modelo, as reclamações de treino foram apresentadas ao modelo de forma gradual, em pequenos grupos de 25 de cada tópico. Em seguida, o modelo era colocado para classificar as 300 reclamações do outro grupo, para possibilitar a extração de métricas da evolução do aprendizado do mesmo.

Tabela 1. Resultados obtidos

	Primeira Rodada			Quarta Rodada		
	Precisão	Recall	F1-Score	Precisão	Recall	F1-Score
Problema com a entrega	69%	79%	74%	83% (+14%)	89% (+10%)	86% (+12%)
Problema com o produto	74%	91%	82%	88% (+14%)	89% (-2%)	89% (+7%)
Problema com o pagamento	75%	47%	58%	84% (+9%)	77% (+30%)	80% (+22%)
Acurácia		72%			85% (+13%)	

Na primeira rodada, os resultados foram regulares para os problemas relacionados à entrega e produto, porém obtiveram uma pontuação ainda pior para problemas relacionados ao pagamento, chegando a obter uma pontuação de 47% no *recall* e 58% no *F1-Score*.

Esses valores tão baixos nessa primeira rodada já eram esperados e provavelmente se devem à representação do texto das reclamações utilizando o algoritmo TF-IDF. Isso causa uma confusão inicial no modelo, pois com o pequeno número de amostras na base da primeira rodada, as palavras mais relevantes ainda não foram definidas com precisão. Esse problema afeta principalmente a categoria de problemas com pagamento, pois é nessa categoria onde é necessário uma maior variedade de palavras para descrever corretamente o problema, havendo então uma dificuldade para definir as palavras mais relevantes para a sua categoria com um número baixo de amostras.

Entretanto, as outras duas categorias não foram tão afetadas, isso se deve ao fato de que elas possuem algumas palavras-chave que se repetem com frequência e logo se tornam as mais relevantes para aquele tipo de documento no cálculo do algoritmo TF-IDF. No entanto, isso também acaba se tornando um problema, pois faz com que o modelo classifique um número maior de falsos positivos quando se depara com essas palavras em outras categorias. Essa situação pode ser visualizada de forma mais clara nas categorias relacionadas a problema com a entrega e com o produto, em que, apesar de obterem uma pontuação mais alta de *recall* na primeira fase em comparação à categoria de problemas com o pagamento, obtiveram uma pontuação pior de precisão, como consequência do alto número de falsos positivos.

Então, ao final da quarta rodada, obteve-se uma média superior a 80% em todas as métricas, tendo destaque na melhora de 30% obtida na métrica de *recall* da categoria de problemas relacionados ao pagamento. Isso significa que se fosse apresentado um grupo de 1000 reclamações relacionadas a problemas com pagamento na primeira rodada de testes, o modelo conseguiria identificar apenas 470, mas ao final conseguiria identificar 770, o que é uma melhora considerável.

CONCLUSÃO

Com base no trabalho realizado, pode-se concluir que o modelo obtido alcançou os objetivos propostos com sucesso. Ao longo das diferentes fases de treinamento e avaliação, observou-se uma evolução significativa na capacidade do modelo em categorizar reclamações relacionadas a problemas de entrega, produto e pagamento. Inicialmente, o modelo apresentou resultados razoáveis, especialmente em problemas relacionados ao pagamento, onde houve uma taxa de *recall* mais baixa. Isso era esperado devido à representação do texto das reclamações, onde foi utilizado o algoritmo TF-IDF, que pode gerar confusão quando há um número pequeno de amostras disponíveis para treinamento. Porém, com o aumento do número de amostras, a definição das palavras mais relevantes para cada categoria ficou mais precisa, e como consequência, o modelo conseguiu categorizar com mais eficiência os diferentes tipos de reclamações. Isso resultou em melhorias significativas nas métricas de avaliação, incluindo aumento no *recall*, precisão e F1-Score.

Em suma, os resultados obtidos demonstram que o modelo desenvolvido é eficaz na classificação automatizada de reclamações em comércios eletrônicos, proporcionando uma abordagem computacional que pode ser aplicada para otimizar o processo de atendimento ao cliente e resolver as questões mais rapidamente.

REFERÊNCIAS

ABCOMM. Principais indicadores do E-Commerce. 2023. Disponível em:
<<https://dados.abcomm.org/>>. Acesso em: 01 jan. 2024.

FALCÃO, H. S. et al. Classificação de vagas de estacionamento com utilização de rede perceptron multicamadas. Revista de Sistemas de Informação da FSMA, Visconde de Araújo, n. 12, p. 41–48, 2013. Disponível em:<https://www.fsma.edu.br/si/edicao12/FSMA_SI_2013_2_Estudantil_1.pdf>. Acesso em: 10 ago. 2023.

SILVA, A. P. C. E-COMMERCE: IMPACTOS NO CONSUMO DO SEGMENTO DE BELEZA E SAÚDE DURANTE A PANDEMIA COVID-19.30f. Monografia (Graduação em Logística) — Universidade Federal do Tocantins, Araguaína, 2022.