

Avaliação Automática da Qualidade das Respostas para Reclamações no Setor de Telefonia no Brasil

Gustavo N. Ubeda¹, João Monteiro¹, Solange O. Rezende¹, Fabio Lobato²

¹Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo
Caixa Postal 668 – 13560-970 – São Carlos – SP – Brasil

²Instituto de Engenharia e Geociências – Universidade Federal do Oeste do Pará
Santarém – PA – Brasil

{gustavo.ubeda, joaovitormonteiro}@usp.br

solange@icmc.usp.br, fabio.lobato@ufopa.edu.br

Abstract. *Complaint platforms in Brazil influence how consumers look at reviews to make their purchasing decisions, especially in the telecommunications sector, where companies' responses can attract or alienate customers. Many of these companies face difficulties in providing effective responses, losing competitive advantages. This paper proposes an AI model to classify customer satisfaction based on responses to complaints. Textual representation methods, such as TF-IDF, BoW, and BERT, and machine learning with hyperparameter adjustments were used to optimize the performance of the trained models. The results showed that removing the neutral class from the reviews and adjusting BERT increased accuracy, highlighting its importance in analyzing reviews and improving customer service, with the implementation of an API for commercial use.*

Resumo. *As plataformas de reclamações no Brasil influenciam como consumidores observam opiniões para tomarem suas decisões de compra, especialmente no setor de telecomunicações, onde as respostas das empresas podem atrair ou afastar clientes. Muitas dessas enfrentam dificuldades em oferecer respostas eficazes, perdendo vantagens competitivas. Este artigo propõe um modelo de IA para classificar a satisfação dos clientes com base nas respostas às reclamações. Foram utilizados métodos de representação textual, como TF-IDF, BoW e BERT, e aprendizado de máquina com ajustes de hiperparâmetros para otimizar o desempenho dos modelos treinados. Os resultados mostraram que remover a classe neutra das avaliações e ajustar o BERT aumentaram a precisão, destacando sua importância na análise de opiniões e no aprimoramento do atendimento ao cliente, com a implementação de uma API para uso comercial.*

1. Contextualização

Os Sistemas de Informação (SI) desempenham um papel estratégico na gestão corporativa, especialmente por meio de sistemas de Gestão de Relacionamento com o Consumidor (CRM), que integram tecnologia e redes sociais para lidar com o crescente volume de conteúdo gerado por usuários. Nesse cenário, destaca-se o *Electronic Word-of-Mouth*

(eWoM), que engloba relatos de experiências de consumidores sobre produtos e serviços. A integração de CRMs com plataformas sociais permite às empresas monitorar o comportamento do consumidor, identificar padrões de sentimento e responder de forma estratégica às expectativas do público [Bahtar and Muda 2016, Schmäh et al. 2017].

No Brasil, plataformas como Reclame Aqui e Consumidor.gov destacam-se como plataformas voltadas para o eWoM, acumulando milhões de reclamações sobre empresas de diversos setores, especialmente telecomunicações. Estas não apenas conectam consumidores em busca de soluções, mas também exigem respostas rápidas e eficazes das empresas. A análise detalhada desse grande volume de dados é crucial para ajustar práticas de atendimento, fortalecer a reputação corporativa e melhorar a gestão das interações com os consumidores [rec 2024, con 2024].

Diversos estudos têm contribuído para o avanço da mineração de opiniões e análise de sentimentos em dados textuais, abordando diferentes contextos e técnicas. Em [Jain et al. 2021], foi proposta uma abordagem sistemática baseada em Aprendizado de Máquina (AM) para classificar sentimentos e detectar reclamações falsas. No contexto de redes sociais, a revisão de [Cortis and Davis 2021] destacou metodologias híbridas e técnicas amplamente utilizadas, como *Naive Bayes* e *Support Vector Machine*. Métodos lexicais e modelos de AM foram integrados em [Qi and Shabrina 2023], enquanto [Saha et al. 2020] demonstrou a superioridade de BERT associado a LDA na identificação de entidades em textos clínicos. Aplicações de BERT também foram exploradas em [Jin and Aletras 2020], com resultados expressivos na classificação de reclamações no Twitter, e em [Albineli and de Oliveira 2023], para avaliar sentimentos em avaliações turísticas no Brasil. Em relação ao atendimento a reclamações, [Peixoto 2021] empregou AM para otimizar o processo, evidenciando desafios na análise de dados textuais não estruturados. No setor industrial, [Abas et al. 2020] propôs o modelo FGAOM, destacando a adaptação de domínio em análises baseadas em aspectos, enquanto [Bayazed et al. 2024] revisou técnicas de mineração de opiniões comparativas, enfatizando a eficácia de BERT e GPT-3 em dados de língua árabe.

Neste contexto, o estudo propõe um modelo preditivo baseado em AM para avaliar como os consumidores percebem as respostas das empresas às reclamações. O objetivo central é comparar métodos de *embedding* textual clássicas e *transformers*, além de realizar ajustes de hiperparâmetros, com o fim de selecionar o melhor modelo de AM. As contribuições incluem recomendações práticas para empresas, identificação de áreas de melhoria no atendimento e uma metodologia replicável para a análise de sentimentos em plataformas de eWoM, promovendo avanços na interação entre consumidores e organizações.

2. Processo adotado

O presente estudo adotou uma metodologia estruturada em etapas, do pré-processamento dos dados até a avaliação de modelos preditivos aplicados a reclamações de consumidores. Os dados foram coletados por meio de um *scraper* desenvolvido especificamente para extrair informações da base Reclamações do *Consumidor.gov*. Este processo resultou em um conjunto robusto de mais de 900 mil reclamações envolvendo as principais empresas de telefonia do Brasil: Vivo, Tim, Claro e Oi. E as classes extraídas para a qualidade das respostas das empresas foram: Boa, Neutra e Ruim.

Na etapa de pré-processamento, os dados foram inicialmente limpos para remoção de valores ausentes e *stop words*, assegurando consistência analítica. Para a representação textual, foram empregadas abordagens baseadas em vetorização lexical, *Bag of Words* e *TF-IDF*, complementadas pelo uso do algoritmo Porter para *stemming*. Adicionalmente, foi utilizado o modelo pré-treinado *BERT*, permitindo a geração de vetores que capturam relações semânticas e sintáticas mais complexas. Para lidar com o desequilíbrio das classes nos dados, a classe Neutra foi removida e um *undersampling* foi aplicado no conjunto de treinamento. O processo de vetorização das reclamações e suas respectivas respostas pode ser observado na Figura 1, sendo que as classes foram definidas como: Boa (Notas 5 e 4), Neutra (3) e Ruim (1 e 2).

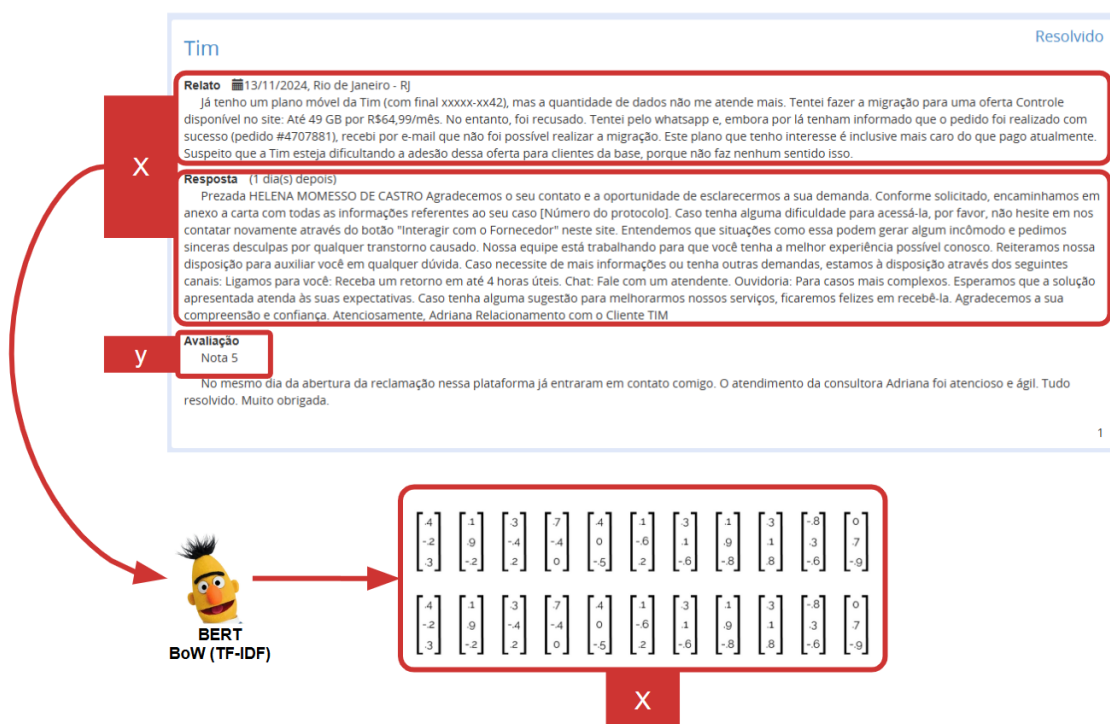


Figura 1. Exemplo de vetorização das reclamações

O treinamento dos modelos preditivos incluiu Redes Neurais *MLP*, *Random Forest* e Máquinas de Vetores de Suporte. Os hiperparâmetros de cada modelo foram otimizados via *GridSearchCV*, utilizando validação cruzada e a métrica F1 como critério principal de avaliação. Os dados foram divididos em três conjuntos: treino (80%), teste (10%) e validação (10%).

Por fim, a robustez das soluções foi avaliada em experimentos específicos para diferentes empresas e em toda a base. Ferramentas como *scikit-learn*, *imblearn* e *transformers* foram utilizadas, juntamente com bibliotecas de visualização, como *Seaborn* e *Matplotlib*.

3. Solução

Os experimentos realizados avaliaram três modelos de AM utilizando diferentes representações textuais e estratégias de ajuste de hiperparâmetros. No primeiro experimento, combinações com *Bag-of-Words* (*BoW*) e *TF-IDF* apresentaram limitações na

captura de nuances semânticas, especialmente para a classe neutra, resultando em um F1 médio de 0,42. Embora o *MLP* com *BoW* tenha sido o mais equilibrado (F1 de 0,4290 e acurácia de 47,62%), os modelos em geral enfrentaram dificuldades em generalizar adequadamente para essa classe.

No segundo experimento, a representação baseada em *BERT* foi introduzida, oferecendo avanços modestos na captura semântica. Apesar de melhorias pontuais, como o *Random Forest* atingindo um F1 de 0,4220, os resultados ainda foram limitados, particularmente na distinção da classe neutra. Essa dificuldade revelou a necessidade de se remover a classe neutra do conjunto, que também produz poucos *insights* para as empresas, além de explorar abordagens mais avançadas, como o *fine-tuning* de modelos pré-treinados, para superar as barreiras impostas pela complexidade dos dados.

No terceiro experimento, a exclusão da classe neutra marcou um ponto de virada, simplificando a tarefa de classificação. O *Random Forest*, ajustado e combinado com *BERT*, alcançou um F1 médio de 0,7667, refletindo melhorias significativas em precisão e revocação para as classes positiva e negativa. Empresas como *ClaroResidencial* e *ClaroFixo* apresentaram F1 acima de 0,85.

Os resultados evidenciam que a remoção da classe neutra e o refinamento das representações textuais foram essenciais para melhorar o desempenho dos modelos. O *Random Forest* destacou-se como a solução mais robusta, alcançando um equilíbrio entre as métricas. Além disso, a análise por domínio (empresa) sugere que abordagens personalizadas para diferentes contextos de dados podem ser decisivas para maximizar a eficiência dos modelos. Por fim, foi desenvolvida uma *API*, com auxílio da biblioteca *FastAPI*, que utiliza os modelos gerados para classificar a probabilidade de uma resposta da empresa ser avaliada de forma positiva ou negativa pelos clientes, permitindo uma avaliação automatizada das interações com os consumidores em tempo real. Uma prévia da tela de documentação da *API* pode ser observada na Figura 2.



Figura 2. Documentação da API desenvolvida

Referências

- (2024). Página do consumidor.gov. Acessado em: 29 de outubro de 2024.
- (2024). Página do reclame aqui. Acessado em: 29 de outubro de 2024.
- Abas, A. R., El-Henawy, I., Mohamed, H., and Abdellatif, A. (2020). Deep learning model for fine-grained aspect-based opinion mining. *IEEE Access*, 8:128845–128855.
- Albineli, I. C. and de Oliveira, R. (2023). Análise de sentimentos de publicações em plataformas on-line sobre turismo no brasil.
- Bahtar, A. Z. and Muda, M. (2016). The impact of user-generated content (ugc) on product reviews towards online purchasing—a conceptual framework. *Procedia Economics and Finance*, 37:337–342.
- Bayazed, A., Almagrabi, H., Alahmadi, D., and Alghamdi, H. (2024). Acom: Arabic comparative opinion mining in social media utilizing word embedding, deep learning model & llm-gpt. *IEEE Access*.
- Cortis, K. and Davis, B. (2021). Over a decade of social opinion mining: a systematic review. *Artificial Intelligence Review*, 54(7):4873–4965.
- Jain, P. K., Pamula, R., and Srivastava, G. (2021). A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer Science Review*, 41:100413.
- Jin, M. and Aletras, N. (2020). Complaint identification in social media with transformer networks. *arXiv preprint arXiv:2010.10910*.
- Peixoto, L. H. R. (2021). *Aprendizado de Máquina Aplicado no Atendimento de Reclamações de Clientes*. PhD thesis, Universidade de São Paulo.
- Qi, Y. and Shabrina, Z. (2023). Sentiment analysis using twitter data: a comparative application of lexicon- and machine-learning-based approach. *Social Network Analysis and Mining*, 13(1):31.
- Saha, B., Lisboa, S., and Ghosh, S. (2020). Understanding patient complaint characteristics using contextual clinical bert embeddings. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 5847–5850. IEEE.
- Schmäh, M., Wilke, T., and Rossmann, A. (2017). *Electronic word-of-mouth: a systematic literature analysis*. Gesellschaft für Informatik, Bonn.