

Análise Automática de Satisfação do Consumidor na Plataforma Consumidor.gov

Anna B. S. Perotto, Eduardo C. R. Costa, William B. A. M. Betker, Daniel G. Silva e Ugo S. Dias

Resumo— Este trabalho aplica a Análise de Sentimentos, por meio de um modelo BERT pré-treinado em português do Brasil, nos comentários de finalização dos processos da plataforma *Consumidor.gov*. Empregando Transferência de Aprendizado e o rotulamento de um conjunto de dados para fins de teste da proposta, chegou-se a um desempenho de classificação com medidas F1-Score para as classes “Satisfeito” e “Insatisfeito” de 74% e 81%, respectivamente.

Palavras-Chave— Análise de Sentimentos, BERT, PLN.

Abstract— This paper studies the effectiveness of Sentiment Analysis, through a pre-trained BERT model in Brazilian Portuguese, for a consumer complaints dataset from the *Consumidor.gov* platform. Transfer Learning is employed for the fine-tuning phase and a manually-labelled test set is designed to evaluate the model. Classification results show a F1-Score of 74% for the “Satisfied” class and 81% for the “Unsatisfied” class.

Keywords— Sentiment Analysis, BERT, NLP.

I. INTRODUÇÃO

No contexto da Transformação Digital, empresas acompanham os avanços tecnológicos e a transição dessas práticas, seja no comércio físico, seja na consolidação do Mercado Digital, pela compra e venda de produtos e serviços online através de redes sociais ou pelo *e-commerce*. Após uma compra, os usuários são capazes de exteriorizar opiniões pessoais sobre sua experiência em forma de texto e mídia, em publicações compartilhadas em diferentes plataformas.

Uma dessas bases agregadoras de opiniões de consumidores é a plataforma *Consumidor.gov* [1], um serviço da Secretaria Nacional do Consumidor (SENACON) do Ministério da Justiça e Segurança Pública (MJSP), o qual permite a comunicação entre consumidores e fornecedores de produtos e serviços com o objetivo de solucionar problemas na relação de consumo. Entre diversas funcionalidades disponíveis na plataforma, o consumidor, ao fim do processo de intermediação com a empresa, pode registrar sua opinião final em forma de texto sobre o atendimento dado e o cumprimento das expectativas iniciais, quando criou a reclamação.

Esta plataforma possui um universo bastante representativo de reclamações. Por exemplo, somente no mês de Maio de 2022, ao redor de 100 mil reclamações foram registradas no sistema [2]. Naturalmente, diversas aplicações na seara da extração automática de conhecimentos podem ser desenhadas. Devido à grande quantidade de dados, faz-se necessária a construção de aplicações automatizadas capazes de extrair informações relevantes a partir dos dados. Surge então a

oportunidade de se utilizar do Processamento de Linguagem Natural (PLN) para, com base nestes relatos registrados no momento da finalização da reclamação, inferir a polaridade de sentimentos e, conseqüentemente, medir ainda que indiretamente a opinião geral sobre os atendimentos encaminhados dentro da plataforma, de uma certa empresa ou setor. No entanto, dado o enorme volume de reclamações e a impossibilidade de um rotulamento ostensivo destas, este trabalho traz a proposta de construir um modelo de análise de sentimentos por meio da Transferência de Aprendizado [3], realizando o processo de *fine-tuning* do modelo *Bidirectional Encoder Representations from Transformers* (BERT), através de uma base de reclamações de clientes correlata, para enfim testar o seu emprego na inferência dos sentimentos das reclamações na plataforma *Consumidor.gov*.

II. METODOLOGIA

O desenvolvimento do trabalho compreende as seguintes etapas: obtenção, pré-processamento e carregamento dos dados; treinamento (*fine-tuning*) do modelo, análise dos resultados de treino e inferência utilizando a base de dados obtida do *Consumidor.gov*. A implementação é feita no ambiente Google Colaboratory [4], na linguagem Python. Utiliza-se o BERTimbau, um modelo BERT pré-treinado no Português do Brasil, disponibilizado no repositório online NeuralMind [5].

O processo de *fine-tuning* do modelo utilizou da base de dados *Olist* [6] para os conjuntos de treino (67%), e de validação (33%). A *Olist - Brazilian E-Commerce Public Dataset* dispõe de 43.482 registros de consumidores fornecidos em uma pesquisa de avaliação, no período entre 2016 e 2018. Estes apresentam uma nota e um comentário. A nota, em uma escala de 1 a 5, é dada pelo consumidor à empresa. Neste trabalho a nota representa a satisfação do cliente.

Embora o *dataset* não contenha originalmente classificação de sentimentos, é possível converter a nota fornecida na pesquisa de avaliação em um de três níveis de satisfação pré-definidos. Assim, atribui-se o rótulo “Insatisfeito” aos relatos com notas entre 1 e 2, “Neutro” aos relatos com nota 3 e “Satisfeito” aos relatos com pontuação entre 4 e 5. A distribuição de classes neste conjunto de treino, conforme a Tabela I, mostra que a classe majoritária é dada por “Satisfeito”, enquanto “Neutro” é a classe minoritária.

Ainda como etapa de preparação para o *fine-tuning*, é necessária a conversão dos textos em vetores de *encodings* utilizando um *BertTokenizer*, fornecido pelo próprio modelo BERT. Dada uma *string* de entrada, a saída do *tokenizer* são 2 vetores. O primeiro é um vetor de *encodings* onde cada elemento é um número inteiro sem sinal, o segundo vetor

TABELA I: Distribuição de Sentimentos no conjunto de treino.

Sentimento	Relatos	Porcentagem
Insatisfeito	11.512	26,48%
Neutro	3.752	8,63%
Satisfeito	28.218	64,90%
Total	43.482	100%

contém a *attention mask*. O comprimento da *string* de entrada impacta diretamente na memória de hardware necessária ao treino. Por este motivo, todas as sequências utilizadas são truncadas em 50 elementos.

O treinamento do modelo é realizado por 10 épocas, com o monitoramento da acurácia e F1-Score sobre o conjunto de validação. Os resultados podem ser visualizados na Fig. 1 e na Fig. 2. Nota-se o *overfitting* do modelo a partir da época 2, caracterizado por um queda da acurácia para o conjunto de validação e um aumento da mesma no conjunto de treino.

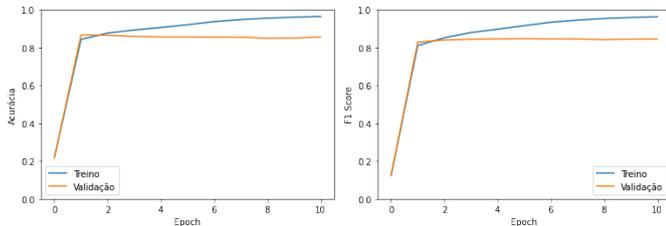


Fig. 1: Acurácia vs. épocas. Fig. 2: F1-Score vs. épocas.

Para avaliação do modelo, um conjunto de teste foi construído pelos autores através da rotulação de reclamações extraídas diretamente da base de dados da plataforma *Consumidor.gov*. Este conjunto abrange 496 reclamações de consumidores que foram previamente rotuladas pelos autores deste trabalho com os rótulos referentes aos sentimentos de Satisfeito, Neutro e Insatisfeito. Duplas de rotuladores liam cada reclamação e atribuíam um rótulo e, em caso de divergência, um terceiro rotulador emitia o julgamento de desempate.

III. RESULTADOS

Aplica-se o classificador treinado para a inferência nos textos do conjunto de teste, formado pelos relatos de finalização das reclamações na plataforma *Consumidor.gov*. As informações de Precisão, Recall e F1-Score podem ser vistas na Tabela II.

TABELA II: Resultados da Inferência sobre o conjunto de testes.

Sentimento	Precisão	Recall	F1-Score
Insatisfeito	74%	90%	81%
Neutro	0	0	0
Satisfeito	63%	88%	74%

A acurácia geral foi de 70%, razoável para uma aplicação que não foi treinada com os dados originais da plataforma *Consumidor.gov*. Entretanto, pode ser observado que o modelo foi incapaz de classificar o sentimento neutro, como fica mais

claro na matriz de confusão, apresentada na Fig. 3. Uma possível justificativa se encontra na pequena quantidade de dados desta classe no conjunto de treino, como mostra a Tabela I. Por outro lado, a acurácia para os sentimentos insatisfeito e satisfeito foram adequadas. O desempenho do modelo para dados da classe satisfeito foi inferior ao da classe insatisfeito.

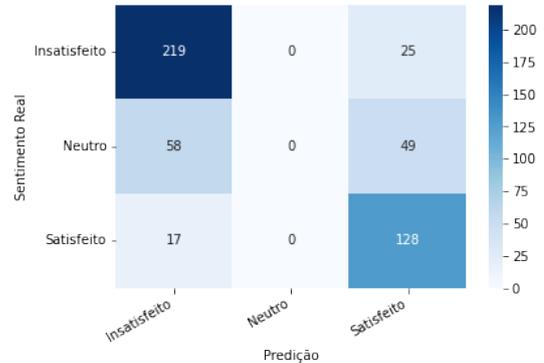


Fig. 3: Matriz de confusão

IV. CONCLUSÕES

Falta, à plataforma *Consumidor.gov*, como ferramenta pública que fornece a interlocução entre empresas e consumidores, um estudo das experiências dos consumidores no processo de mediação do conflito com a empresa. Dessa forma, fornecedores e os gestores da ferramenta poderiam inferir o sentimento do cliente referente ao atendimento dado e, assim, otimizar o processo de resolução de problemas.

Nesse contexto, este trabalho empregou Análise de Sentimentos para classificar o atendimento de reclamações publicadas no portal. Estudou-se a implementação do modelo BERTimbau que, depois de treinado, levou a um desempenho de classificação com medidas *F1-Score* para as classes “Satisfeito” e “Insatisfeito” de 74% e 81%, respectivamente. Futuramente, pretende-se testar outras transformações de notas em sentimentos para o conjunto de treino, no sentido de melhorar a identificação do sentimento “Neutro”, além de empregar o modelo para criar estatísticas automatizadas de satisfação com o atendimento das reclamações, fornecendo mecanismos de transparência aos cidadãos e aos fornecedores, e também instrumentos de monitoramento destas empresas pelos gestores na plataforma *Consumidor.gov*.

REFERÊNCIAS

- [1] SENACON. (2022) Plataforma - Consumidor.gov. [Online]. Available: <https://www.consumidor.gov.br/>
- [2] —. (2022) Página de dados abertos - Consumidor.gov. [Online]. Available: <https://www.consumidor.gov.br/pages/dadosabertos/externo/>
- [3] F. Souza and J. Filho, “BERT for sentiment analysis: Pre-trained and fine-tuned alternatives,” *CoRR*, vol. abs/2201.03382, 2022. [Online]. Available: <https://arxiv.org/abs/2201.03382>
- [4] Google. (2021) Colaboratory. [Online]. Available: <https://colab.research.google.com/>
- [5] F. Souza, R. Nogueira, and R. Lotufo, “Bertimbau: Pretrained bert models for brazilian portuguese,” in *Intelligent Systems*, R. Cerri and R. C. Prati, Eds. Cham: Springer International Publishing, 2020, pp. 403–417.
- [6] B. E.-C. P. Dataset. (06/2021). [Online]. Available: <https://www.kaggle.com/olistbr/brazilian-ecommerce/code>