

Modelos Transformers para a Análise Automática de Satisfação na Plataforma Consumidor.gov.br

Daniel G. Silva¹, William B. A. M. Betker¹, Daniel P. Gonçalves¹, Ugo S. Dias¹

¹Universidade de Brasília – Faculdade de Tecnologia – Departamento de Engenharia Elétrica
Brasília – DF – Brasil

{danielgs, udias}@unb.br

Abstract. *In Brazil, Consumidor.gov.br platform is a relevant tool for Consumer's Rights: consumers communicate with suppliers to solve conflicts between the parts; in the end, the consumer writes his/her opinion about the process. This work proposes a sentiment analysis model through Transfer Learning, fine-tuning a BERT model via a training set which aggregates well-known datasets of customer reviews in Portuguese, which is subsequently evaluated in the task of sentiment analysis of Consumidor.gov.br complaints, through a test set exclusively labeled for this purpose. The model is deployed in SENACON data analysis environment and is able to perform sentiment analysis of the finalized disputes on the platform, on a daily basis, with F1-score of 0,72.*

Resumo. *A plataforma Consumidor.gov.br é um recurso relevante para os Direitos do Consumidor, no Brasil: consumidores contatam os fornecedores para resolver conflitos entre as partes; ao final, o consumidor escreve sua opinião sobre o processo. Este trabalho propõe um modelo de análise de sentimentos via Aprendizado por Transferência, realizando o fine-tuning de um modelo BERT em um conjunto de treino que agrega reconhecidas bases de dados de avaliações de consumidores em português. A seguir, avalia-se o modelo na classificação das reclamações do Consumidor.gov.br, por meio de um conjunto de testes exclusivamente rotulado para este fim. O modelo é implantado no ambiente de análise de dados da SENACON, onde consegue realizar análise de sentimentos das reclamações finalizadas, diariamente, com F1-score de 0,72.*

1. 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 on-line 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. Além disso, diversas entidades de defesa do consumidor trabalham com plataformas digitais para não só coletar a experiência, mas também para intermediar a resolução de conflitos entre consumidores e fornecedores.

Neste contexto, uma das ferramentas mais relevantes no Brasil é a plataforma *Consumidor.gov.br* [SENACON 2023b], um serviço da Secretaria Nacional do Consumidor (SENACON) do Ministério da Justiça e Segurança Pública (MJSP). Esta Secretaria, criada em 2012, tem suas atribuições estabelecidas no Código de Defesa do Consumidor

e no Decreto nº 2.181/97, as quais envolvem o planejamento, elaboração, coordenação e execução da Política Nacional das Relações de Consumo [SENACON 2021].

A plataforma *Consumidor.gov.br* permite a comunicação entre consumidores e fornecedores de produtos e serviços com o objetivo de solucionar problemas na relação de consumo. Entre as diversas funcionalidades disponíveis, 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, de quando criou a reclamação. Ademais, a plataforma é monitorada em conjunto pelos Procons, Defensorias Públicas, Ministérios Públicos, Agências Reguladoras e por toda a sociedade em potencial, uma vez que as informações de cada processo compõem uma base de dados pública, à disposição de qualquer interessado, independentemente de solicitação e em formato aberto, em conformidade com as diretrizes de acesso à informação e transparência ativa, tratadas no Decreto 8.777 [Brasil 2021a] e na Lei 12.527 [Brasil 2021b].

Em termos de alcance, a plataforma possui um universo bastante representativo de reclamações. Lançada em junho de 2014, ela já registrou até o final de 2022 mais de 6,3 milhões de reclamações, e conta com uma base de 4,2 milhões de usuários cadastrados e mais de 1275 empresas credenciadas. Desde 2020, mais de 1 milhão de reclamações são processadas anualmente [SENACON 2023a]. Este sucesso traz consigo desafios tecnológicos. O crescimento de sua base de dados e o amadurecimento de seus processos culminaram em uma grande necessidade de automatização de análises, além da otimização dos fluxos de trabalho. Naturalmente, diversas aplicações na área 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.

Uma das informações passíveis de automatização é a opinião dos consumidores sobre os atendimentos realizados pelas empresas. A plataforma *Consumidor.gov.br* permite que o consumidor registre sua opinião sobre o atendimento recebido após a finalização da reclamação. Surge daí 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 (i) enorme volume de reclamações, (ii) a impossibilidade de um rotulamento ostensivo destas, e (iii) a necessidade de que o modelo realize em tempo real a classificação de sentimentos para novas reclamações finalizadas, é necessária a associação de técnicas modernas de Aprendizado de Máquina. Portanto, este trabalho traz a proposta de construir um modelo de análise de sentimentos por meio do Aprendizado por Transferência [Souza and Filho 2022], realizando o processo de *fine-tuning* do modelo BERTimbau, a variante pré-treinada em língua portuguesa do modelo *Bidirectional Encoder Representations from Transformers* (BERT), através de uma base de reclamações de clientes correlata, para a seguir testar o seu emprego na inferência dos sentimentos das reclamações da plataforma *Consumidor.gov*, por meio de um conjunto de testes exclusivamente rotulado para este fim. Finalmente, o modelo validado passa por um processo de implantação no ambiente de análise de dados da SENACON, tornando-se apto a realizar análise de sentimentos das reclamações finalizadas na plataforma em

tempo real.

O resto deste trabalho se organiza da seguinte forma: a Seção 2 apresenta os trabalhos relacionados; a Seção 3 descreve todo o processo de preparação dos dados, seguido do treinamento, validação e implantação do modelo em ambiente de produção; a Seção 4 apresenta os resultados; e a Seção 5 traz as conclusões e trabalhos futuros.

2. Trabalhos Relacionados

A literatura associada às técnicas de análise de sentimentos é bastante extensa, uma vez que é um problema estudado há bastante tempo, dentro da área de Processamento de Linguagem Natural (PLN). Tal conceito também vale para a análise de sentimentos de consumidores e clientes quanto a produtos e serviços adquiridos. No entanto, a análise de sentimentos dos consumidores quanto à resolução (ou não) de um conflito de consumo é um problema ainda pouco explorado. Com este contexto, nesta seção busca-se analisar trabalhos que envolvam a análise de sentimentos em *big data*, voltados à resolução de conflitos de consumo ou à utilização de modelos baseados em Transformers.

[Hajiali 2020] realiza um trabalho de revisão de artigos sobre análise de sentimentos em *big data*. Nele, emprega o método de revisão sistemática da literatura e, com base em diferentes filtros, chega-se a 23 artigos de relevância para analisar. As desvantagens e vantagens das técnicas investigadas são estudadas e suas questões-chave são enfatizadas. Para o treinamento dos classificadores de sentimentos, os trabalhos exploram estratégias clássicas de Aprendizado de Máquina como as Redes Perceptron Multicamadas ou Máquinas de Vetores-Suporte. Entre as conclusões obtidas, o autor aponta que uma melhor análise de *big data* textual em termos de sentimento aumenta a eficiência, flexibilidade e inteligência para um negócio obter ideias comerciais.

[Zhou et al. 2018] analisam a influência das revisões de produtos por consumidores nas estratégias de desenvolvimento de novos produtos. Para isso, utilizam estratégias de *big data* em um conjunto de dados de 3 milhões de avaliações *online* obtidas de um aplicativo para dispositivos móveis. O estudo descobre que o volume de avaliações tem uma relação curvilínea com a agilidade na resposta ao cliente. Além disso, esta agilidade tem uma relação curvilínea com o desempenho do produto. O trabalho contribui ao demonstrar a influência da capacidade das empresas de utilizar avaliações de clientes *online* e seu impacto no desempenho do produto.

[Liu and Wan 2023] utilizam dados de comentários nas redes sociais para estudar a opinião pública em relação à resolução de conflitos de compras na maior plataforma de comércio de bens usados da China, por meio de técnicas de análise de frequência de palavras, análise temática e análise de sentimentos. O estudo descobriu que as disputas se concentram principalmente nas questões de devolução e reembolso, além de indicar que a plataforma necessita aperfeiçoar a concepção do sistema e aumentar os canais de recurso do serviço.

[Souza and Filho 2022] estudam o modelo BERT para a análise de sentimentos na língua portuguesa. O trabalho compara diferentes formas de agregação dos vetores de atributos produzidos na saída do modelo, as suas variantes para o português brasileiro e para múltiplas línguas, e a combinação de diferentes conjuntos de dados para treinamento e teste. Os resultados indicam o modelo BERTimbau como a melhor variante do BERT

para a referida tarefa, e também se verifica que os modelos treinados com um conjunto de dados distinto do conjunto de teste ainda assim apresentam resultados satisfatórios.

Nota-se, pelos trabalhos supramencionados, que a aplicação de modelos baseados em Transformers para a análise de sentimentos é uma perspectiva já iniciada, ainda que recentemente. O mesmo pode-se concluir do uso de métodos de *big data* na referida tarefa. No entanto, a literatura a respeito de análise de sentimentos na resolução de conflitos de consumo é bastante escassa e, no conhecimento dos autores, inexistente quando se trata da língua portuguesa e de plataformas/serviços gerais de intermediação de conflitos, como a plataforma Consumidor.gov.br. Além disso, também não se encontrou na literatura trabalhos que utilizem modelos de Transformers para este contexto específico. Tais lacunas indicam a direção de pesquisa deste trabalho, conforme será apresentada em mais detalhes nas seções subsequentes.

3. Processo

Como já mencionado na seção anterior, a plataforma Consumidor.gov.br é uma ferramenta de intermediação entre consumidores e empresas para a resolução de conflitos. O consumidor, previamente cadastrado, registra a sua reclamação na ferramenta e, a partir daí, inicia-se a contagem do prazo de até 10 dias para manifestação da empresa. Após a resposta da empresa, é garantida ao consumidor a chance de comentar a resposta recebida, classificar a demanda como Resolvida ou Não Resolvida e, ainda, indicar o grau de satisfação com o atendimento recebido [SENACON 2023a].

Na proposta deste trabalho, interessa-nos analisar o sentimento associado ao comentário final deixado pelo consumidor. Uma vez que a quantidade de reclamações é muito grande, não é possível realizar o rotulamento manual de todos os comentários. Assim, a estratégia adotada foi a de rotular uma amostra de comentários para ser utilizada como conjunto de testes e treinar o modelo BERT através de bases públicas de avaliação de produtos e serviços por consumidores. Após feita a validação do modelo e avaliação do desempenho no conjunto rotulado, foram realizadas as adaptações e integrações necessárias para que o modelo pudesse operar em ambiente de produção. A seguir, descrevemos cada uma destas etapas.

3.1. Base de Dados da plataforma Consumidor.gov.br

Para fins de desenvolvimento da solução em conjunto com os autores deste trabalho, a equipe de dados do MJSP forneceu acesso a uma cópia somente leitura do banco de dados da plataforma. Entre as informações disponíveis, é possível citar:

- número de protocolo;
- situação da reclamação;
- relato do consumidor;
- resposta do fornecedor;
- indicador de solução;
- texto da avaliação final do consumidor.

No processo de extração, transformação e carregamento (ETL - *Extract, Transform and Load*) os dados das reclamações são extraídos do banco de dados por intermédio de uma solução de integração de dados: o Azure DataFactory. Esta ferramenta permite o

tratamento do grande volume de dados envolvidos, os quais, após extraídos, são transportados para um Data Lake [K.V.N et al. 2016] dividido em diversas zonas. A zona inicial de armazenamento é chamada de RawZone, onde estão dispostos os dados brutos da maneira original, sem transformações.

Em seguida, os dados são processados através da ferramenta DataBricks, sendo realizadas as limpezas necessárias para sua utilização: a remoção de dados pessoais – em conformidade com a LGPD [Brasil 2018] –, a deduplicação dos registros e a aplicação de algoritmos para garantir a integridade dos dados. Após essa etapa, os dados são armazenados na TrustedZone, uma zona em que já foram realizados tratamentos para assegurar a qualidade dos dados, de forma que possam ser considerados exatos e confiáveis.

Por fim, são realizados processamentos voltados para prover o dado em seu formato final, estruturado da melhor maneira para a aplicação que vai utilizá-lo. Após ser feito um enriquecimento que envolve etapas como a junção de tabelas relacionadas e o pré-processamento dos textos das reclamações, chega-se à RefinedZone, uma zona que contém dados refinados e prontos para o consumo. Essa estruturação permite tanto extrair, na íntegra, os dados necessários para a criação do conjunto de testes que validará o treinamento do modelo, quanto aplicar, posteriormente, o modelo já treinado aos novos dados no ambiente de BI do do MJSP.

Vale frisar que, para atender ao objetivo deste trabalho, somente o texto da avaliação final do consumidor foi utilizado. Não foram consideradas as demais informações disponíveis no banco de dados, em especial o indicador de solução (“Resolvido” / “Não Resolvido”) ou o grau (nota) de satisfação, porque há, segundo relatos de experiência da equipe do MJSP, uma confiabilidade reduzida destes indicadores, no sentido de que é comum observar, por exemplo, textos de avaliação final indicando insatisfação com o tratamento dado pela empresa, mas que o consumidor marca no sistema como “Resolvido”.

3.2. Rotulação dos dados para o conjunto de teste

O processo de rotulação foi feito por 3 voluntários, que antes de visualizarem os textos a serem rotulados definiram os critérios para atribuir a cada registro uma categoria. Nesse mesmo momento, algumas reclamações foram avaliadas em conjunto para esclarecer possíveis dúvidas. A rotulação consistia na leitura dos textos escritos pelos consumidores, quando finalizado o processo de reclamação, e identificação do sentimento que mais se adequava ao texto: satisfeito, insatisfeito ou neutro. O sentimento neutro foi utilizado quando não era possível identificar no texto do consumidor alguma mínima expressividade de sentimento. A Tabela 1 apresenta exemplos de textos rotulados para cada sentimento.

Após a primeira rodada de rotulação, as reclamações que não foram classificadas com o mesmo sentimento por todos os rotuladores foram separadas. Iniciou-se então uma nova rodada somente com estes registros, em que os rotuladores tinham acesso aos rótulos que foram dados previamente e tinham que, juntos, definir qual mais se adequava. Caso não fosse possível chegar a um consenso, o texto era descartado, isto é, não entraria no conjunto final de exemplos rotulados.

Ao final deste processo, foram obtidas 496 reclamações rotuladas, as quais formaram o conjunto de testes. A Tabela 2 apresenta a sua distribuição dos rótulos, em que se

Tabela 1. Exemplos de textos do conjunto de testes.

Sentimento	Exemplo de comentário
Satisfeito	“Problema resolvido com rapidez e eficiência. Obrigado.”
Insatisfeito	“O objeto foi entregue com muito tempo de atraso e só consegui ser respondida após o mesmo ser entregue.”
Neutro	“Não consigo ler a resposta do fornecedor mas consegui as passagens.”

pode notar a predominância da classe de consumidores insatisfeitos.

Tabela 2. Distribuição de classes no conjunto de testes.

Classe	Quantidade	Percentual
Satisfeito	145	29,2%
Insatisfeito	244	49,1%
Neutro	107	21,5%

3.3. Bases de Dados para os conjuntos de treino e validação

Como mencionado anteriormente, devido à impossibilidade de uma rotulação extensiva dos dados originais da plataforma Consumidor.gov.br, recorreu-se ao emprego de bases de dados correlatas à tarefa de análise de sentimentos em textos de consumidores para se compor os conjuntos de treino e validação. Neste sentido, construiu-se uma base a partir da união dos seguintes conjuntos de dados públicos: Olist [Olist and Sionek 2018], Buscapé [Hartmann et al. 2014] e B2W [B2W 2018].

A base *Brazilian E-Commerce Public Dataset by Olist* [Olist and Sionek 2018] é composta por 100 mil pedidos de 2016 a 2018, feitos em diversos *marketplaces* no Brasil. Cada pedido possui informações de localização, status, preço, entre outros. Neste trabalho, estamos considerando somente os textos de avaliação final dos pedidos, emitidos pelos consumidores, o que compreende ao redor de 30 mil exemplos.

A base Buscapé [Hartmann et al. 2014] é composta por mais de 80 mil textos de avaliações de produtos feitas por consumidores, capturadas do site de busca e comparação de preços Buscapé¹ no mês de Setembro de 2013.

A base *B2W-Reviews01* [B2W 2018] é composta por mais de 130 mil avaliações de produtos feitas por consumidores, coletadas do site de comércio eletrônico Americanas² entre os meses de janeiro e maio de 2018.

Todas estas bases são públicas e contém, além do comentário sobre o produto, uma nota dada pelo consumidor no intervalo de 1 a 5, sendo 5 a melhor avaliação possível, enquanto 1 é a pior avaliação possível. Uma vez que a tarefa de análise de sentimentos neste trabalho envolve a classificação em três níveis (Satisfeito, Neutro e Insatisfeito), há uma conversão das notas nas classes de acordo com a seguinte regra: notas 1 e 2 são

¹<https://www.buscape.com.br>

²<https://www.americanas.com.br>

consideradas como “Insatisfeito”, nota 3 é considerada como “Neutro” e, por fim, notas 4 e 5 são consideradas como “Satisfeito”. Com a agregação das três fontes de dados, formase uma base única com 259.107 exemplos, dos quais 181.374 são selecionados para o conjunto de treino (70%), enquanto 77.733 exemplos são selecionados para o conjunto de validação (30%).

A Tabela 3 apresenta a distribuição de classes nos conjuntos de treino e de validação. Nota-se que a distribuição de classes é desbalanceada, com uma predominância de exemplos da classe “Satisfeito”.

Tabela 3. Distribuição de exemplos entre as classes nos conjuntos de treino e de validação.

Classe	Treino	Validação
Satisfeito	121.680 (67,1%)	52.107 (67,0%)
Insatisfeito	37.831 (20,9%)	16.145 (20,8%)
Neutro	21.863 (12,0%)	9.481 (12,2%)

3.4. Modelo

Baseado na arquitetura Transformer [Vaswani et al. 2017], o BERT é um modelo grande de linguagem proposto em 2019 e desenhado para inúmeras aplicações de processamento de linguagem natural, entre elas a análise de sentimentos [Devlin et al. 2019]. Trata-se de um *multi-layer bidirectional Transformer encoder*: um modelo capaz de receber o texto de entrada de uma só vez, ao invés da tradicional forma sequencial das arquiteturas recorrentes, ao mesmo tempo que explora com competência os mecanismos de atenção e representação de palavras em espaço latente já advindos de estratégias anteriores. O modelo base é composto por uma pilha de 12 *encoders* Transformer. Cada bloco Transformer possui 6 camadas idênticas, em que cada uma destas possui duas subcamadas, com a primeira sendo um *multi-head self-attention mechanism* e a segunda sendo uma rede *feedforward* simples totalmente conectada. É então empregada uma conexão residual em torno de cada uma das duas subcamadas, seguida por uma técnica de *Layer Normalization*.

O uso de um modelo de linguagem como o BERT consiste de dois passos, sendo eles o pré-treino e o ajuste-fino (*fine-tuning*). No pré-treino, o modelo passa por um treinamento não-supervisionado em duas tarefas genéricas. Já no ajuste-fino, o modelo é iniciado com os parâmetros pré-treinados, os quais continuam a ser ajustados usando dados rotulados de tarefas-fim (*downstream task*), como a análise de sentimentos [Souza and Filho 2022]. O BERT está disponível em dois tamanhos: o BERT_{BASE}, com 110 milhões de parâmetros, e o BERT_{LARGE}, com 340 milhões.

O pré-treino de modelos BERT para a língua portuguesa foi realizado recentemente, tais variantes foram denominadas BERTimbau [Souza et al. 2020]. Para construí-las, utilizou-se do corpus brWaC [Wagner Filho et al. 2018], até então o maior corpus aberto de textos em português. O BERTimbau também é disponibilizado em dois tamanhos, respectivamente pré-treinados nas versões BASE e LARGE do BERT, por meio das mesmas tarefas. No treinamento feito para este trabalho, considerou-se o modelo BERTimbau_{BASE}.

3.4.1. Treinamento

Dada a grande complexidade do modelo BERTimbau, o treinamento (*fine-tuning*) para a tarefa de classificação foi feito em uma máquina dedicada com uma GPU Nvidia GTX 1080Ti com 11 Gigabytes de memória GDDR5X. Por conta do próprio desenho da arquitetura da rede, os documentos (sequências de *tokens*) precisam ser truncados em um tamanho máximo fixo. De forma a selecionar um valor adequado de tamanho máximo, é computado para cada entrada do conjunto de treino o comprimento da sequência de *tokens* e um histograma é montado, conforme a Figura 1. Assim, definiu-se como critério escolher o comprimento de sequência que engloba o percentil 95 do conjunto: neste caso, o valor aproximado que atende é 128. A partir daí, todos os documentos são truncados para até 128 *tokens*.

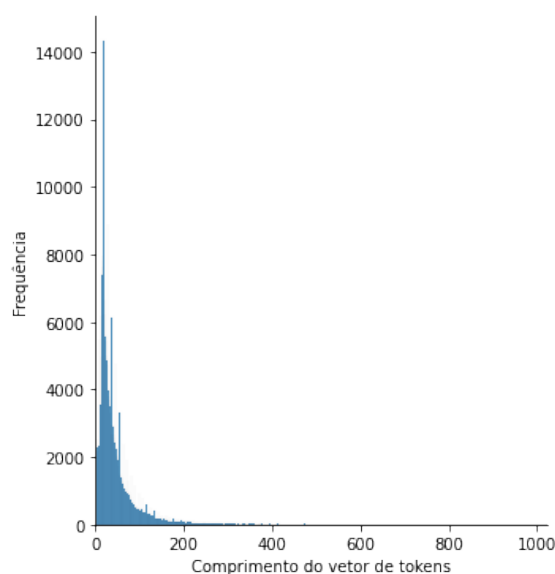


Figura 1. Histograma do tamanho, em número de *tokens*, dos documentos.

O tamanho de *batch* utilizado, para adequar a convergência do treinamento simultaneamente com o maior uso possível de memória da GPU, foi de 83 instâncias. O treinamento foi feito por 10 épocas, com o modelo final sendo o da época em que se obteve o melhor valor de F1-Score no conjunto de validação. Utilizou-se a função de entropia cruzada como função custo, com a camada de saída softmax.

Finalizado o treinamento, é realizada a predição de novos exemplos do conjunto de teste, o qual foi construído conforme descrito na Seção 3.2. A avaliação consiste no cálculo das métricas precisão, revocação (*recall*) e F1-Score.

4. Resultados

O treino teve duração total de 355 minutos. O melhor resultado para F1-Score ocorreu na época 3, tendo valor igual a 0,7221. Após isso, o valor F1-Score no conjunto de validação entra em declínio, em contraste ao mesmo para o conjunto de treino, indicando uma potencial situação de sobreajuste (*overfitting*), conforme se observa na Figura 2a. Um resultado semelhante ocorre na métrica de acurácia, vide Figura 2b.

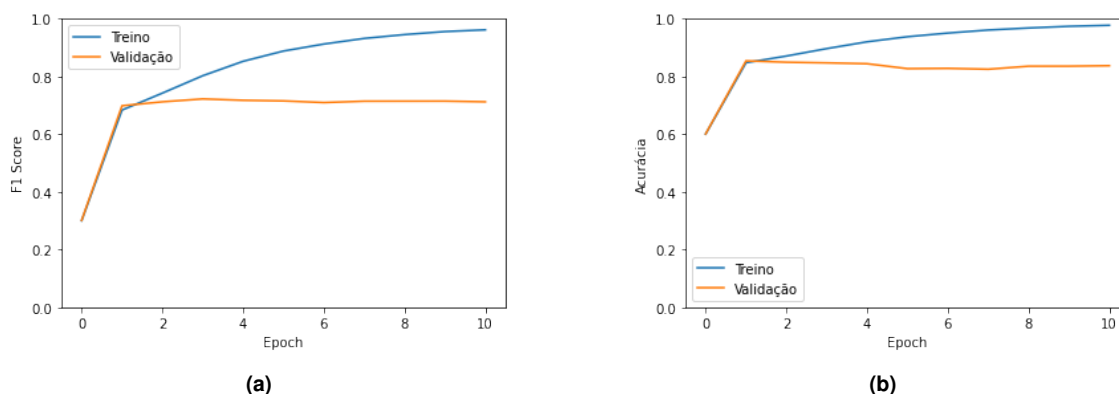


Figura 2. Desempenho do modelo, em termos de F1-Score (a) e acurácia (b), ao longo do treinamento.

Treinado o modelo, faz-se as predições sobre o conjunto de teste. Os resultados estão disponíveis na Tabela 4. Nota-se que, apesar de um F1-Score médio bom, o mo-

Tabela 4. Resultados da Inferência sobre o conjunto de testes.

Classe	Precisão	Recall	F1-Score
Insatisfeito	87%	77%	82%
Neutro	40%	45%	42%
Satisfeito	75%	83%	79%
Média*	73%	72%	72%

*Ponderada conforme a proporção da classe no conjunto de teste.

delo apresenta um desempenho inferior para a classe neutro, o que pode ser explicado pela baixa quantidade de dados de treino pertencentes a esta classe. No entanto, podemos considerar como positivos os resultados atingidos especificamente para as classes “Satisfeito” e “Insatisfeito”, especialmente ao se lembrar que o conjunto de treino não foi construído a partir da plataforma Consumidor.gov.br, mas sim com dados de consumidores em circunstâncias similares, porém distintas da ferramenta de destino, na qual eles estão tecendo comentários a respeito da resolução (ou não) de um conflito de consumo com uma certa empresa.

Também é importante de se registrar o desempenho levemente superior do modelo na classe “Insatisfeito”. É justamente esta classe que pode ser primeiramente analisada pela SENACON, uma vez que priorizar a análise de reclamações de consumidores insatisfeitos com o processo na plataforma pode indicar melhorias nas políticas da secretaria.

Por fim, a Figura 3 apresenta a matriz de confusão, para uma melhor visualização da capacidade de classificação do modelo.

4.1. Implantação do modelo em ambiente real

Encerrado o treinamento e avaliação do modelo, passa-se para a sua integração ao painel interno de *Business Intelligence* (BI) da SENACON, para que possa subsidiar o cálculo de métricas associadas a satisfação dos consumidores com as empresas.

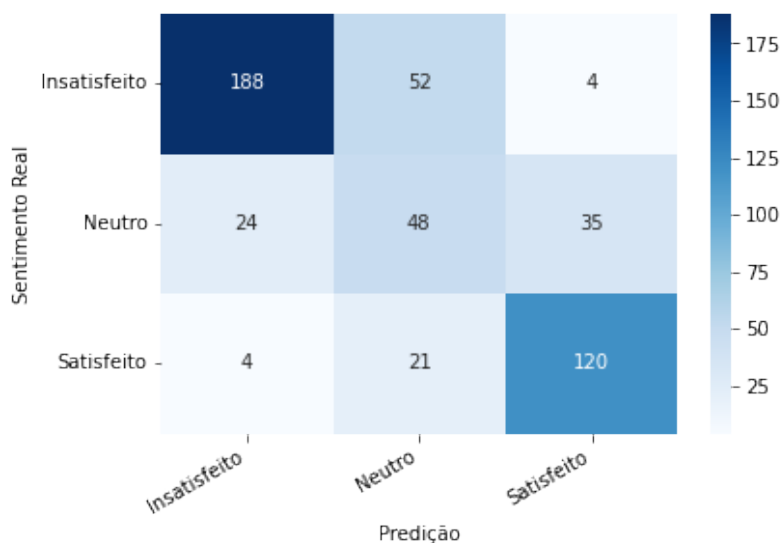


Figura 3. Matriz de confusão do modelo junto ao conjunto de testes.

Os arquivos contendo o modelo treinado e sua estrutura foram convertidos para o formato ONNX Runtime³, de forma que o processo de inferência fosse acelerado para um tempo viável de execução, considerando o tamanho da base da plataforma. Feito isso, os mesmos foram importados pelos scripts de processamento dos dados do painel de BI dentro da plataforma Azure.

Cada texto de avaliação da reclamação de consumidores nos últimos 5 anos dentro da plataforma foi classificado em um processo que levou cerca de 35 horas para ser finalizado, devido à grande quantidade de registros da base histórica. Em seguida, foi implementado um processo de atualização incremental, executado diariamente de forma automática, classificando apenas os novos textos de avaliação de reclamações recém-finalizadas. Este processo, por sua vez, é executado em poucos minutos e garante que os dados que são utilizados na tomada de decisões reflitam períodos recentes.

Para a visualização dos resultados obtidos, foram utilizadas as probabilidades produzidas para cada classe, na saída do modelo. Uma vez que a análise individual dos textos não teria interesse prático pelo corpo técnico da SENACON, as reclamações são agrupadas e a estatística média de cada uma das classes é apresentada em uma aba do painel de BI, construído com a ferramenta PowerBI. A Figura 4 apresenta um exemplo de visualização dos resultados obtidos, com os nomes dos fornecedores anonimizados. Observe que há possibilidade de analisar textos agrupados por fornecedor, grupos econômicos ou segmentos de mercado, o que eventualmente auxiliará a identificar fornecedores que estão com maior grau de insatisfação na resolução das reclamações, por exemplo.

5. Conclusão

Este trabalho apresentou uma proposta de solução para a análise de sentimentos das avaliações finais referentes aos processos de reclamação contra empresas, cadastrados por consumidores na plataforma Consumidor.gov.br. A solução utiliza o reconhecido modelo

³<https://onnxruntime.ai/index.html>



Figura 4. Exemplo de visualização de métricas calculadas com base no modelo de análise de sentimentos.

de linguagem baseado na arquitetura Transformer, o BERT, que passou pelo processo de refinamento (*finetuning*) para a tarefa de classificação de textos em três classes: Satisfeito, Neutro e Insatisfeito. Utilizou-se especificamente a variante do modelo BERTimbau, pré-treinado para a língua portuguesa.

Uma vez que a plataforma dispõe de uma base de reclamações que cresce diariamente na ordem de milhares e que não possui rotulamento prévio algum, desenhou-se uma estratégia de refinamento do modelo através de um conjunto de dados formado por bases de dados públicas de avaliações de produtos e serviços por consumidores, que em seguida foi validado por meio de um conjunto de testes rotulado manualmente, desta vez com os textos de avaliação de consumidores da plataforma Consumidor.gov.br.

Os resultados indicaram um desempenho satisfatório do modelo, com F1-Score médio de 72%, Precisão média de 73% e Revocação média de 72% no conjunto de testes. Por fim, o modelo validado pôde ser incorporado ao ambiente interno de análise de dados da SENACON, permitindo a visualização de métricas de satisfação dos consumidores com as empresas, com atualização diária e em tempo hábil para a tomada de decisões.

Como trabalhos futuros, pretende-se aprimorar o modelo de análise de sentimentos, com a inclusão de mais dados de treino ou a incorporação de técnicas de balanceamento de classes, de forma a melhorar o desempenho geral, especialmente na classe “Neutro”. Também é desejável a realização de um estudo de viabilidade para a utilização de modelos de linguagem mais recentes, com desempenho inclusive superior ao BERT em diversas tarefas de processamento de linguagem natural.

Referências

B2W (2018). B2w-reviews01. <https://github.com/americanas-tech/b2w-reviews01>.

Brasil (2018). Lei 13.709. *Diário Oficial da República Federativa do Brasil*. Lei Geral de Proteção de Dados Pessoais (LGPD).

Brasil (2021a). Decreto 8.777. *Diário Oficial da República Federativa do Brasil*. Institui a Política de Dados Abertos do Poder Executivo federal.

- Brasil (2021b). Lei 12.527. *Diário Oficial da República Federativa do Brasil*. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North*, volume 1, pages 4171–4186. Association for Computational Linguistics.
- Hajiali, M. (2020). Big data and sentiment analysis: A comprehensive and systematic literature review. *Concurrency and Computation: Practice and Experience*, 32(14):e5671.
- Hartmann, N., Avanço, L., Balage, P., Duran, M., das Graças Volpe Nunes, M., Pardo, T., and Aluísio, S. (2014). A large corpus of product reviews in portuguese: Tackling out-of-vocabulary words. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 3865–3871. European Language Resources Association (ELRA).
- K.V.N, R., HOD, Department of Information Technology, Vignan's Institute of Information Technology, Visakhapatnam, India., K.V.N., R., and Project Manager, Tech Mahindra, Visakhapatnam, India. (2016). An introduction to data lake. *i-manager's Journal on Information Technology*, 5(2):1.
- Liu, Y. and Wan, Y. (2023). Consumer satisfaction with the online dispute resolution on a second-hand goods-trading platform. *Sustainability*, 15(4).
- Olist and Sionek, A. (2018). Brazilian e-commerce public dataset by Olist. <https://www.kaggle.com/dsv/195341>.
- SENACON (2021). Página da SENACON - Secretaria Nacional do Consumidor. <https://www.gov.br/mj/pt-br/assuntos/seus-direitos/consumidor>.
- SENACON (2023a). Boletim Consumidor.gov.br. Technical report, Ministério da Justiça e Segurança Pública. <https://consumidor.gov.br/pages/publicacao/externo/>.
- SENACON (2023b). Plataforma - Consumidor.gov.br. <https://www.consumidor.gov.br/>.
- Souza, F., Nogueira, R., and Lotufo, R. (2020). BERTimbau: Pretrained BERT models for brazilian portuguese. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12319 LNAI:403–417.
- Souza, F. D. and Filho, J. B. d. O. e. S. (2022). BERT for sentiment analysis: Pre-trained and fine-tuned alternatives. In Pinheiro, V., Gamallo, P., Amaro, R., Scarton, C., Batista, F., Silva, D., Magro, C., and Pinto, H., editors, *Computational Processing of the Portuguese Language*, Lecture Notes in Computer Science, pages 209–218. Springer International Publishing.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors,

Advances in Neural Information Processing Systems, volume 30, pages 5998–6008. Curran Associates, Inc.

Wagner Filho, J. A., Wilkens, R., Idiart, M., and Villavicencio, A. (2018). The brWaC corpus: A new open resource for Brazilian Portuguese. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Zhou, S., Qiao, Z., Du, Q., Wang, G. A., Fan, W., and Yan, X. (2018). Measuring customer agility from online reviews using big data text analytics. *Journal of management information systems*, 35(2):510–539.