

Carolina Nascimento Silva
Nicolý de Oliveira Avelino

Projeto: Análise de Sentimento em Feedbacks de Clientes a partir de Áudio

Objetivo

A detecção de sentimentos em feedbacks de clientes permite analisar opiniões espontâneas que frequentemente refletem de forma mais autêntica a percepção dos clientes. Isso possibilita identificar pontos fortes e fracos de produtos ou serviços que poderiam passar despercebidos em análises textuais ou formulários estruturados. Com este projeto, iremos analisar áudios de feedback, transcrevê-los e classificá-los quanto ao sentimento (positivo, negativo ou neutro), fornecendo uma visão abrangente sobre o tom dessas avaliações e contribuindo para uma compreensão mais precisa e estratégica das opiniões de clientes.

Pesquisa e Seleção da Biblioteca de Transcrição

1. Transkriptor

Transkriptor é uma biblioteca focada em fornecer transcrições rápidas e precisas, com suporte para mais de 100 idiomas. É ideal para empresas que precisam de resultados imediatos e lidam com uma base de clientes diversificada. Apesar de sua precisão, que pode chegar a 99%, e do bom custo-benefício para usos pontuais, a ferramenta possui capacidades limitadas de personalização, o que pode ser uma restrição para casos mais específicos.

2. Deepgram

Conhecida por sua alta precisão e velocidade, Deepgram é uma opção robusta para transcrever áudios em grande escala. Ela oferece suporte a múltiplos idiomas e possui excelente custo-benefício para grandes volumes de dados. Além disso, sua capacidade de personalização permite ajustes para contextos específicos, tornando-a uma escolha confiável para empresas que buscam escalabilidade e flexibilidade.

3. Microsoft Azure Speech-to-Text

Microsoft Azure Speech-to-Text combina precisão e velocidade em transcrições, sendo altamente recomendada para projetos de larga escala. Com suporte para mais de 100 idiomas e excelente capacidade de personalização, permite criar modelos ajustados às necessidades de negócios. É uma escolha sólida para empresas que já utilizam o ecossistema Azure e precisam de soluções integradas.

4. Google Cloud Speech-to-Text

Com suporte para mais de 125 idiomas, o Google Cloud Speech-to-Text é uma das opções mais abrangentes para transcrições globais. Ele combina alta precisão e boa velocidade, sendo ideal para aplicações em tempo real. Sua integração com outros serviços do Google Cloud oferece vantagens para usuários desse ecossistema, embora a personalização seja limitada em comparação a outras ferramentas.

5. Amazon Transcribe

Amazon Transcribe se destaca pela alta precisão e suporte a fluxos de áudio em tempo real. Com vários idiomas e dialetos disponíveis, é uma solução viável para contextos multilíngues. Seu custo-benefício é competitivo para empresas que já utilizam a AWS, e sua capacidade de personalização a torna adequada para cenários específicos de transcrição.

6. Speechmatics

Speechmatics oferece transcrições precisas e em tempo real, sendo especialmente útil para fluxos de trabalho instantâneos. Embora suporte mais de 50 idiomas, esse número é limitado em relação a outras opções. A biblioteca tem boa capacidade de personalização e é recomendada para empresas que valorizam precisão em tempo real com uma gama moderada de idiomas.

7. IBM Watson Speech to Text

IBM Watson Speech to Text é uma solução versátil que combina precisão e suporte a vários idiomas. Sua alta velocidade de processamento e boa capacidade de personalização a tornam adequada para cenários multinacionais e usos frequentes. É uma escolha confiável para empresas que precisam de soluções contínuas e ajustáveis para transcrição.

8. Rev.AI

Rev.AI é reconhecida por sua precisão em transcrição de áudio e vídeo, além de oferecer uma boa velocidade de processamento. No entanto, com suporte para apenas 36 idiomas, ela pode não ser a melhor escolha para contextos multilíngues. Sua personalização é limitada, mas a relação custo-benefício é atraente para empresas com necessidades específicas em áudio e vídeo.

9. OpenAI Whisper

OpenAI Whisper se destaca por sua precisão e flexibilidade, especialmente em projetos de baixo orçamento, já que é de código aberto e gratuito para uso. Apesar de sua velocidade ser moderada em configurações comuns, sua alta capacidade de personalização permite ajustes profundos, tornando-o ideal para quem busca flexibilidade e controle total sobre o processo de transcrição.

Conclusão

Cada biblioteca tem suas próprias vantagens e desvantagens, mas OpenAI Whisper se destaca como a mais adequada para transcrição de feedbacks de clientes, devido à sua alta precisão, flexibilidade de personalização, e custo zero de licenciamento. Essa combinação permite maior controle sobre a adaptação do modelo para diferentes contextos, sendo ideal para empresas que buscam soluções eficientes e econômicas.

Ferramenta	Precisão	Velocidade	Suporte a Idiomas	Custo-Benefício	Capacidade de Personalização
Transkriptor	Alta (até 99%)	Rápida	+100 idiomas	Bom para transcrições rápidas	Limitada em personalização
Deepgram	Alta	Muito rápida	Suporte global, diversas línguas	Custo-benefício elevado para uso em escala	Altamente personalizável
Microsoft Azure	Alta	Rápida	+100 idiomas	Boa relação custo-benefício	Permite criação de modelos personalizados
Google Cloud	Alta	Muito rápida	+125 idiomas	Escalável, mas pode ter custo alto em grande uso	Permite customização de modelos
Amazon Transcribe	Alta	Rápida	+70 idiomas e dialetos	Flexível, com opções de pagamento por uso	Personalização limitada, mas com alguns ajustes possíveis
Speechmatics	Alta	Rápida	+50 idiomas	Boa relação custo-benefício	Modelos customizáveis para necessidades específicas

IBM Watson	Boa	Rápida	Suporte global	Competitiva em preços para várias demandas	Alta personalização para casos específicos
Rev.AI	Muito alta	Rápida	36 idiomas	Boa relação custo-benefício	Limitada para personalização, foco na precisão
OpenAI Whisper	Alta	Moderada	+80 idiomas	Gratuito e de código aberto	Altamente flexível, ideal para personalização

Coleta e Processamento de Dados

Os áudios utilizados neste projeto foram criados especificamente para fins de análise e compreendem um conjunto de dez gravações fictícias. Para garantir variedade e realismo, utilizamos duas fontes distintas: cinco áudios foram gerados com o auxílio de ferramentas de inteligência artificial para síntese de voz, enquanto os outros cinco foram gravados com as vozes de pessoas de nosso convívio familiar. Os feedbacks foram intencionalmente equilibrados quanto ao sentimento: quatro áudios expressavam sentimentos positivos, quatro negativos e dois neutros. Essa abordagem permitiu simular uma diversidade de cenários reais, mantendo controle sobre o conteúdo e a classificação dos feedbacks, assegurando a robustez do processo de análise de sentimentos.

Projeto e codificação

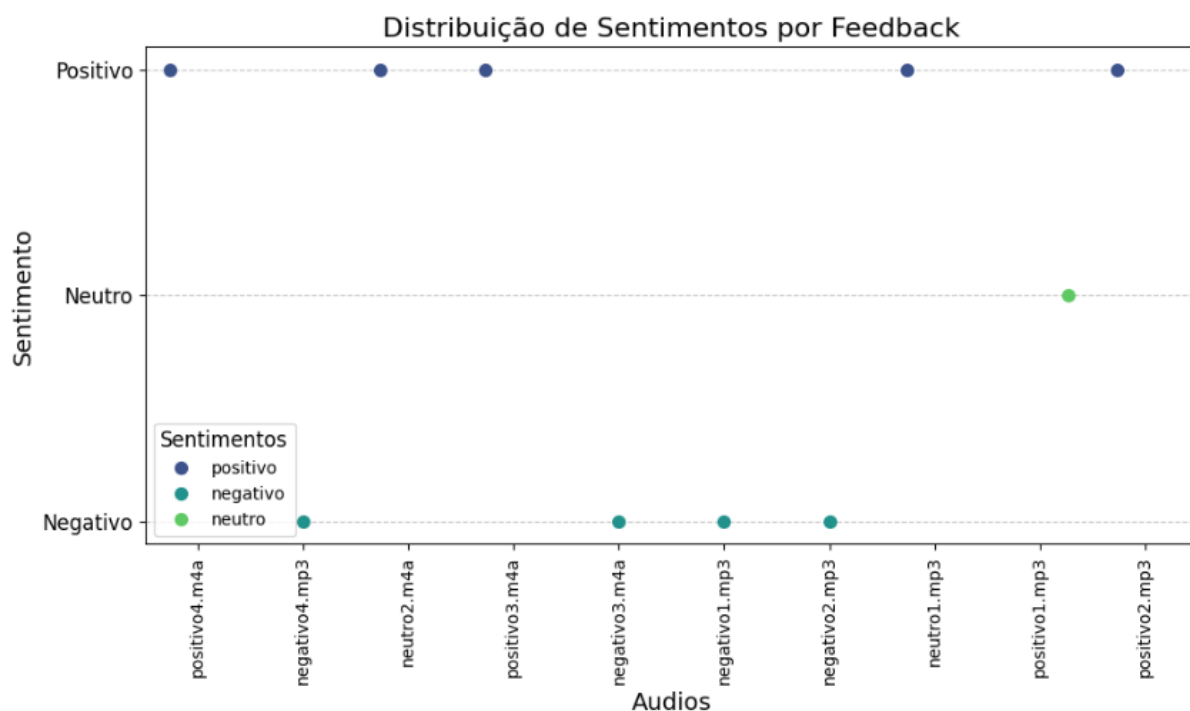
Todo o desenvolvimento do código e a execução do projeto foram realizados utilizando a plataforma Google Colab, que oferece um ambiente de programação baseado na nuvem, acessível para projetos de machine learning e processamento de linguagem natural. Essa escolha permitiu o uso de recursos computacionais sem necessidade de configuração local.

O ambiente pode ser acessado através do link: [Colab](#)

Análise dos Resultados

Durante a análise dos feedbacks, foi necessário traduzir os textos originais do português para o inglês devido às limitações do desempenho do TextBlob em português. Sem a tradução, a biblioteca frequentemente apresentava uma tendência a classificar muitos feedbacks como neutros, independentemente de seu real tom emocional. Essa limitação ocorre porque o TextBlob possui um modelo pré-treinado mais robusto e otimizado para a língua inglesa, com um vocabulário e estrutura semântica mais abrangentes.

Após a tradução, realizamos uma avaliação comparando os sentimentos classificados pelo TextBlob com os sentimentos esperados. Os resultados indicaram uma precisão geral de 70%, o que demonstra um desempenho razoável, mas com algumas limitações. Abaixo estão os detalhes da precisão por categoria:



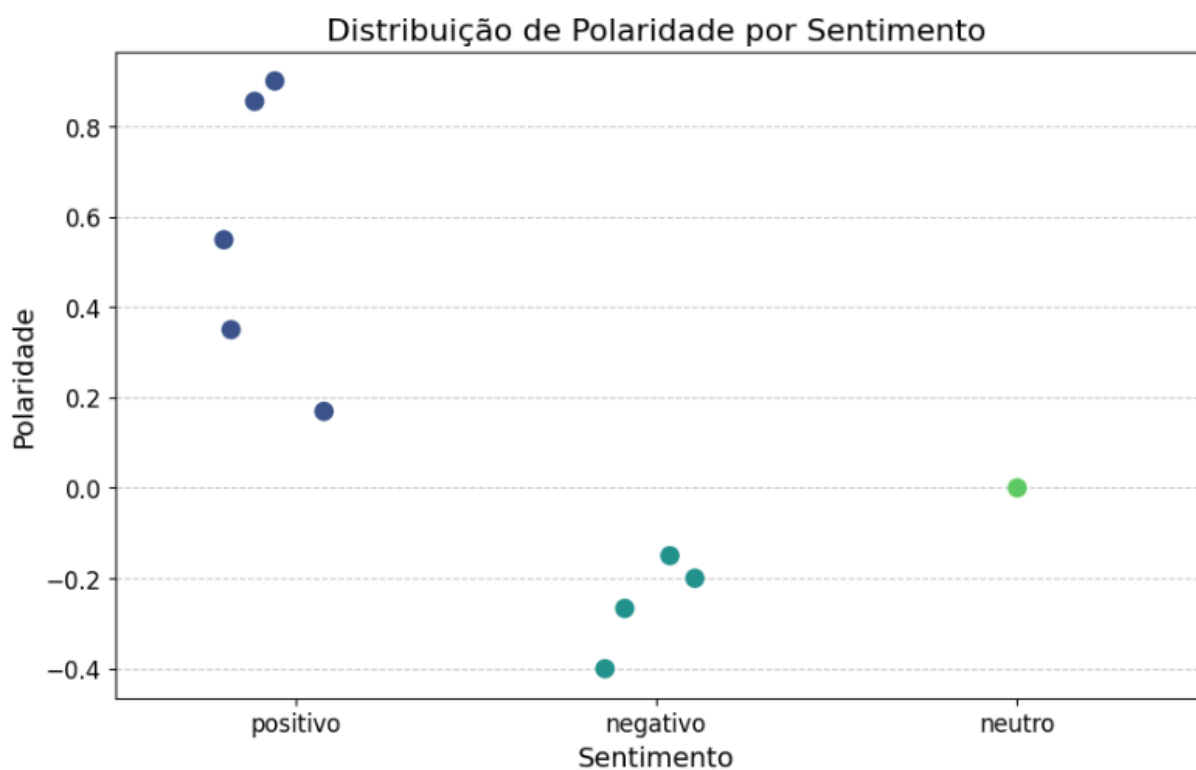
Negativo: 100%

Neutro: 0%

Positivo: 75%

Esses resultados mostram que o TextBlob é eficaz para identificar sentimentos negativos e, em menor grau, positivos. No entanto, a classificação de sentimentos neutros foi um ponto crítico, com erros que indicam a dificuldade do modelo em distinguir nuances mais equilibradas.

Os scores para sentimentos positivos e negativos estão bem representados na análise, com valores extremos geralmente bem categorizados. As frases neutras, no entanto, apresentam scores próximos de zero, mas frequentemente são confundidas com positivas. Isso pode ocorrer porque frases neutras, como "atendeu às expectativas", podem ter uma leve carga positiva dependendo do contexto.



Conclusão

A análise de sentimentos realizada demonstrou resultados satisfatórios em termos de precisão geral, alcançando 70% de acerto na classificação dos feedbacks. Contudo, observou-se uma disparidade no desempenho entre as diferentes categorias: enquanto a classificação de sentimentos negativos foi precisa (100% de acerto), houve dificuldades em distinguir sentimentos neutros (0% de acerto), com uma tendência de classificá-los como positivos. Os scores de polaridade atribuídos pelo TextBlob forneceram insights úteis sobre a intensidade emocional dos feedbacks, mas destacaram limitações na identificação de nuances linguísticas, especialmente em frases traduzidas do português para o inglês. Essa etapa de tradução, embora tenha melhorado o desempenho, pode ter causado perda de informações contextuais, comprometendo a classificação em algumas instâncias.

Para aprimorar futuras análises, é possível fazer a adoção de modelos avançados de Processamento de Linguagem Natural (PLN) melhores para o português, como o BERT, que possui maior capacidade de interpretar nuances e contextos complexos da língua. Esse modelo, baseado em aprendizado profundo, têm o potencial de melhorar significativamente o reconhecimento de sentimentos neutros e proporcionar uma análise mais ajustada às particularidades da língua portuguesa. Essas melhorias poderiam aumentar a precisão geral da análise, contribuindo para uma compreensão mais confiável dos feedbacks, auxiliando na tomada de decisões estratégicas com base nos dados coletados.

Referências

"Whisper: Robust Speech Recognition via Large-Scale Weak Supervision".

Disponível em: <https://github.com/openai/whisper>

"googletrans: Python Library to Use Google Translate API".

Disponível em: <https://pypi.org/project/googletrans/>

"TextBlob: Simplified Text Processing". Disponível em: <https://textblob.readthedocs.io>

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In Proceedings of NAACL-HLT 2019.

Disponível em: <https://arxiv.org/abs/1810.04805>