

# Exploração de métodos simbólicos para detecção de emoções para o português

**Stephanie Briere Americo**  
Núcleo Interinstitucional de  
Linguística Computacional (NILC)  
Instituto de Ciências Matemáticas  
e de Computação (ICMC)  
Universidade de São Paulo  
stephanieb.americo@usp.br

**Thiago Alexandre Salgueiro Pardo**  
Núcleo Interinstitucional de  
Linguística Computacional (NILC)  
Instituto de Ciências Matemáticas  
e de Computação (ICMC)  
Universidade de São Paulo  
taspardo@icmc.usp.br

## Resumo

Este trabalho investiga métodos simbólicos para a detecção de emoções em textos em português, considerando múltiplos corpú, domínios e diferentes configurações de pré-processamento. Os resultados mostram grande variação no desempenho absoluto entre domínios, mas estabilidade no desempenho relativo entre os métodos, evidenciando a influência das propriedades do corpú e o gradiente entre complexidade e interpretabilidade. A inclusão da classe neutra tende a degradar o desempenho ao aumentar a ambiguidade e, frequentemente, o desbalanceamento entre classes, enquanto um pré-processamento mais extensivo beneficia especialmente abordagens simbólicas. A análise qualitativa indica que parte dos erros decorre de ambiguidades linguísticas, do grande espaço para subjetividade no processo de anotação e das próprias nuances emocionais, reforçando a importância de avaliações comparativas multi-domínio.

## 1 Introdução

A detecção automática de emoções em textos é uma tarefa relevante do Processamento de Linguagem Natural, com aplicações em mídias sociais, sistemas de recomendação, educação, atendimento ao cliente e monitoramento de saúde mental (Mohammad, 2016). Diferentemente da análise de sentimentos baseada apenas em polaridade, essa tarefa busca identificar estados afetivos mais complexos, como alegria, tristeza, raiva e medo.

A ambiguidade da linguagem natural e a subjetividade das expressões emocionais dificultam tanto a interpretação humana quanto a análise automática de emoções. As amostras da Tabela 1 ilustram esse fenômeno: na amostra 1, um *review* de filme combina uma opinião positiva com uma ressalva negativa, evidenciando a subjetividade da avaliação; na amostra 2, um *review* de jogo de horror utiliza emoções tipicamente negativas, como

o medo, para expressar uma experiência positiva. Esses exemplos também mostram que a polaridade textual não está necessariamente alinhada às emoções expressas, pois ambas são fortemente dependentes do contexto e, embora relacionadas, não são mutuamente determinantes.

	<i>Texto</i>
1	O filme é muito bom, mas não recomendaria para minha família.
2	Assustador e viciante, não consegui parar de jogar.

Tabela 1: Exemplos de textos nos quais a detecção de emoções não é trivial.

Modelos neurais de larga escala têm alcançado desempenho superior ao de métodos clássicos de aprendizado de máquina (Devlin et al., 2019; Maruf et al., 2024). No entanto, por serem treinados sobre grandes volumes de dados potencialmente enviesados, esses modelos podem herdar preconceitos humanos e produzir decisões injustas. Além disso, sua natureza opaca dificulta a interpretação, auditoria e responsabilização dos sistemas automáticos. Em contraste, métodos simbólicos oferecem transparência, controle explícito do conhecimento e previsibilidade do comportamento do modelo, características historicamente centrais para a Inteligência Artificial e ainda cruciais em aplicações que exigem resultados interpretáveis por humanos.

É neste contexto que abordagens simbólicas permanecem relevantes, especialmente quando interpretabilidade, baixo custo computacional e facilidade de adaptação a novos domínios são requisitos centrais. Em áreas sensíveis, como saúde mental, educação e sistemas de tomada de decisão automatizada, compreender por que uma emoção foi atribuída é uma exigência ética tão importante quanto o desempenho preditivo (Guidotti et al., 2018; Ribeiro et al., 2016).

Este estudo integra um esforço mais amplo que busca avaliar a detecção de emoções em português de forma abrangente, considerando todos os corpú

publicamente disponíveis e diferentes esquemas de anotação, sem restringir o conjunto de emoções ou o paradigma dos modelos de classificação avaliados. O presente artigo constitui um recorte inicial dessa investigação, com foco em métodos simbólicos.

Consideramos desafios centrais da tarefa, como subjetividade e ambiguidade linguística, cenários *multilabel*, desbalanceamento de classes, variação entre esquemas de anotação, adaptação a domínios e escassez de recursos (Maruf et al., 2024). As contribuições são: (i) uma comparação empírica entre métodos com diferentes níveis de interpretabilidade em múltiplos corpú; (ii) uma análise qualitativa da interpretabilidade, com foco nas regras e estruturas aprendidas; e (iii) uma discussão de limitações e oportunidades de melhoria.

## 2 Trabalhos relacionados

Em nossa revisão da literatura, não identificamos trabalhos em português que utilizem abordagens simbólicas para a detecção de emoções. Esta seção apresenta estudos relevantes que exploram outras abordagens para o português, incluindo trabalhos baseados em corpú também utilizados neste estudo e descritos em mais detalhes na Seção 3, bem como a menção a outros trabalhos com abordagens semelhantes.

O estudo de da Silva (2020) avaliou classificadores SVM com *kernel* de árvore para detecção de emoções em textos em um cenário de transferência de domínio, utilizando um corpú com publicações de domínio livre para treinamento e um corpú reduzido com publicações sobre o mercado de ações (DANTEStocks) para teste. Foram treinados classificadores binários para pares de emoções opostas segundo a roda de Plutchik, alcançando Medida- $F$  média ponderada de 0,56. Os resultados evidenciaram o impacto do tamanho e do desbalanceamento das classes, bem como a degradação do desempenho com a inclusão da classe neutra, indicando limitações de generalização entre domínios e dificuldades na classificação de textos ambivalentes.

Um modelo pré-treinado para o português (BERTimbau (Souza et al., 2020)) foi ajustado ao corpú GoEmotions-BR para detectar as seis emoções básicas de Ekman e a classe neutra (Oliveira and Sichman, 2024), seguindo procedimento análogo ao adotado no GoEmotions original em inglês. Embora os resultados não sejam diretamente comparáveis devido ao uso de corpú distintos, os valores

de Medida- $F$  apresentam ordem de magnitude e consistência semelhantes entre os modelos, com média geral de 0,57. Observou-se desempenho significativamente superior na detecção da classe “alegria” em relação a “nojo”, atribuível principalmente à maior representatividade da primeira no treinamento e à sua expressão mais explícita em textos, enquanto “nojo” é menos frequente e mais dependente de nuances contextuais.

Outros trabalhos em português também investigam a detecção de emoções em textos, adotando abordagens semelhantes às descritas anteriormente. Hammes and de Freitas (2021) utilizou o modelo BERTimbau no corpú GoEmotions traduzido com o Google Tradutor, resultando em um recurso distinto do anteriormente mencionado (GoEmotions-BR). Diferentemente de trabalhos que mapeiam as emoções para o conjunto reduzido das seis emoções básicas de Ekman acrescidas da classe neutra, esse estudo considerou as 28 classes originais do GoEmotions. Os resultados obtidos foram compatíveis com os reportados para o modelo BERT original em inglês, alcançando Medida- $F$  média de 0,48.

Dosciatti et al. (2013) aplicou classificadores Naive Bayes, kNN e SVM ao corpú Notícias Curtas, reportando que o SVM obteve o melhor desempenho, com acurácia de 0,61. Outros estudos relevantes incluem Cortiz et al. (2021), que também empregou modelos baseados em BERT, e Santos (2019), que comparou algumas das mesmas abordagens clássicas de aprendizado de máquina.

## 3 Conjuntos de dados

Os corpú utilizados foram selecionados após uma revisão abrangente dos recursos disponíveis para a língua portuguesa, obtidos por disponibilização pública ou por contato direto com seus autores. Alguns corpú descritos na literatura não foram considerados neste trabalho devido a restrições de acesso ou licenciamento, ou por redundância em relação a recursos já adotados. Estes corpú são descritos em Hammes and de Freitas (2021); Cortiz et al. (2021); Santos (2019); Duarte (2019). Essa escolha reflete limitações práticas de disponibilidade, preservando ainda diversidade de domínios, esquemas de anotação e modelos de emoção adotados.

O corpú DANTEStocks<sup>1</sup> contém 4.277 publi-

<sup>1</sup>Disponível em: <https://www.kaggle.com/datasets/fernandojvdasilva/stock-tweets-ptbr-emotions>.

cações em português da rede social X, coletadas em 2014 e relacionadas a ações do índice Ibovespa (da Silva et al., 2020). Os textos foram anotados de forma supervisionada com a classe neutra e as oito emoções primárias da roda de Plutchik (Plutchik, 2001), organizadas em quatro pares de opostos: confiança e nojo; alegria e tristeza; antecipação e surpresa; irritação e medo. O córpus apresenta forte desbalanceamento entre classes.

da Silva et al. (2020) disponibilizou também outro córpus com um subconjunto reduzido do primeiro, contendo apenas 334 textos que obtiveram concordância total entre anotadores, sendo majoritariamente composto por exemplos neutros. Para facilitar a apresentação dos resultados, denominamos esta variação do córpus **DANTEStocks Concordância**.

O córpus **Domínio Livre** é composto por 230.857 publicações da rede social X, coletadas sem restrição temática e anotadas automaticamente por supervisão distante, a partir de *hashtags* associadas às emoções de Plutchik. É um conjunto desbalanceado e não possui publicações “neutras” (da Silva, 2020).

**Notícias Curtas**<sup>2</sup> é um córpus composto por 1.750 notícias do jornal virtual *O Globo*, anotadas manualmente com a classe neutra e as seis emoções básicas de Ekman (Ekman and Keltner, 1997): alegria, tristeza, medo, surpresa, nojo e raiva. O conjunto é balanceado, com 250 textos por classe, e a anotação foi realizada por especialistas com resolução de divergências por consenso (Dosciatti et al., 2013).

O **GoEmotions-BR** (Oliveira and Sichman, 2024) deriva da tradução automática do córpus GoEmotions (Demszky et al., 2020), originalmente composto por um conjunto desbalanceado de 58 mil sentenças em inglês anotadas em 27 emoções (“alegria”, “diversão”, “entusiasmo”, “luto”, “tristeza”, “raiva”, “aborrecimento”, entre outras) e neutro. A tradução foi realizada com o modelo *chatGPT3.5-turbo*.

Nos experimentos de Demszyk et al. (2020), além da versão original do GoEmotions com 27 emoções granulares, foi proposta uma variante reorganizada em que as emoções são mapeadas para as 6 categorias básicas de Ekman, acrescidas da classe neutra (Oliveira and Sichman, 2024). Adotamos também essa variação (**GoEmotions-BR-**

<sup>2</sup>Disponível para solicitação em: <https://www.pggia.pucpr.br/~paraíso/mineracaodeemocoes/recursos.php>.

**Ekman**), pois ela permite comparar diretamente, em um mesmo domínio, o impacto de diferentes modelos de emoção sobre os resultados.

Com exceção do córpus Notícias Curtas, os córpus descritos nesta seção possuem anotação *multilabel*. Nesse cenário, cada amostra pode receber um ou mais rótulos de classes emocionais, como ilustrado na Tabela 2. Essa característica reflete de forma mais fiel a complexidade emocional humana, mas também aumenta a complexidade e a dificuldade da tarefa de detecção automática.

	<i>Texto</i>	<i>Rótulo</i>
1	VALE5 nao passa de 29,90	Antecipação
2	Algo me diz que vou ver a #PETR4 na casa dos 12 hoje ainda.. vamos aguardar.. :)	Confiança, Alegria, Antecipação

Tabela 2: Exemplo de amostra que recebeu apenas um rótulo de emoção e de amostra que foi anotada com múltiplos rótulos (*multilabel*). As amostras são do córpus DANTEStocks (da Silva, 2020).

É importante mencionar que não há consenso científico sobre a natureza e a delimitação das emoções humanas, e os modelos adotados na literatura baseiam-se em diferentes teorias e métodos (como expressões faciais em Ekman e *crowdsourcing* no GoEmotions). A classe “neutra” é ainda mais controversa, pois pressupõe a ausência de emoção, uma condição para a qual não há evidência empírica clara. Embora cada córpus apresente sua própria definição, essa classe tende a agrupar casos heterogêneos, como baixa intensidade emocional, ambiguidade ou ausência de pistas suficientes para a anotação. Além disso, frequentemente constitui a classe predominante em córpus desbalanceados, possivelmente por concentrar casos de difícil caracterização. Consequentemente, trata-se de uma categoria pouco definida, cuja utilização pode introduzir ruído e dificultar a interpretação e a modelagem das emoções.

A Tabela 3 resume os córpus utilizados neste trabalho, que diferem quanto ao domínio, esquema de anotação, modelo de emoções e balanceamento de classes.

## 4 Metodologia

Esta seção descreve os métodos avaliados e a configuração experimental adotada, fornecendo as informações necessárias para garantir a reprodutibilidade dos resultados. Detalhes adicionais, bem como esclarecimentos sobre decisões de implementação, podem ser obtidos mediante contato com os

<i>córpus</i>	<i>Modelo de emoções</i>	<i>Tamanho</i>	<i>Anotação</i>
DANTE	Plutchik	4.277	Superv.
DANTEConcordancia	Plutchik	334	Superv.
DomínioLivre	Plutchik	230.857	Automat.
NoticiasCurtas	Ekman	1.750	Superv.
GoEmotionsBR	27 e neutro	~58.000	Superv.
GoEmotionsBREkman	Ekman	~58.000	Superv.

Tabela 3: Resumo dos corpúis analisados

autores.

#### 4.1 Métodos avaliados

Avaliamos cinco métodos de detecção de emoções em texto, em sua maioria classificadores simbólicos, com diferentes níveis de interpretabilidade. A seleção prioriza modelos consolidados, de fácil implementação e interpretação, permitindo uma comparação consistente entre abordagens simbólicas, supervisionadas e parcialmente interpretáveis.

O **ZeroR** é adotado como *baseline* por sempre prever a classe majoritária do conjunto de treinamento, estabelecendo um limite inferior de desempenho (Alnuaimi and Albaldawi, 2024).

Propomos um método simbólico baseado em regras e léxico (**LexicoR**), inspirado em Seal et al. (2020). O método utiliza o léxico emocional *Emocionário* (Ramos, 2021), no qual termos emocionais são manualmente identificados e anotados por especialistas com base em obras de referência da língua portuguesa. A partir desse recurso, o LexicoR atribui pontuações às emoções conforme a presença e a frequência de termos emocionais, produzindo decisões totalmente transparentes. O objetivo é avaliar o desempenho alcançável por uma abordagem puramente simbólica, baseada exclusivamente em conhecimento linguístico e regras manuais, e verificar se ela supera de forma consistente um *baseline* ingênuo (ZeroR) em múltiplos domínios.

Para cada *token* do texto que esteja presente no *Emocionário*, considera-se uma janela de contexto simétrica de quatro palavras anteriores e posteriores, na qual são identificados modificadores semânticos de negação, intensificação e redução. A pontuação base é 1; intensificadores multiplicam esse valor por 3, redutores o dividem por 3, e a negação inverte o sinal.

Diferentemente de Seal et al. (2020), não assumimos relações explícitas de oposição entre emoções. As pontuações são acumuladas por emoção ao longo do texto, e apenas as emoções dominantes — aquelas com pontuação máxima — são previstas. Em caso de empate, todas as emoções cor-

respondentes são atribuídas, evitando a definição arbitrária de limiares para classificação *multilabel*. O Apêndice A detalha o funcionamento da solução.

A **árvore de decisão** é um modelo supervisionado que aprende regras hierárquicas por meio de testes sobre atributos (Murthy, 1998). Trata-se de um método simbólico, pois pode ser expresso como um conjunto de regras, sendo sua interpretabilidade maximizada na representação em forma de árvore: cada caminho da raiz até uma folha corresponde a uma regra explícita. Como o modelo não oferece suporte nativo à classificação *multilabel*, adotamos a estratégia um-contratodos, induzindo uma árvore independente para cada emoção.

O **RIPPER** induz regras proposicionais diretamente de dados rotulados, expressas como condições lógicas simples e facilmente auditáveis (Cohen, 1995). Embora não seja nativamente *multilabel*, empregamos novamente a estratégia um-contratodos.

O **Random Forest** combina múltiplas árvores de decisão treinadas sobre subconjuntos aleatórios de instâncias e atributos (Breiman, 2001). Embora cada árvore individual possua uma estrutura simbólica e interpretável, a agregação de dezenas ou centenas de árvores torna a lógica global do modelo difícil de inspecionar, o que o afasta de uma abordagem puramente simbólica e o aproxima de um modelo de “caixa-preta”. Neste trabalho, o Random Forest é incluído como um método de transição entre paradigmas simbólicos e estatísticos, alinhado ao objetivo mais amplo de comparar abordagens com diferentes níveis de interpretabilidade. Além disso, o método oferece suporte natural à classificação multiclasse, com extensões que permitem seu uso em cenários *multilabel*.

#### 4.2 Configuração experimental

O pré-processamento é frequentemente apontado como uma limitação de abordagens clássicas em contraste com modelos profundos que operam sobre texto bruto. Para avaliar seu impacto, cada corpúis é analisado em duas configurações: com e sem pré-processamento extensivo. Considerando o forte desbalanceamento observado na maioria dos corpúis, sobretudo pela predominância da classe neutra, avaliamos adicionalmente um cenário em que todas as instâncias rotuladas como neutras foram excluídas do conjunto de dados. Sendo assim, os experimentos foram conduzidos sobre os corpúis originais e também sobre subconjuntos contendo apenas amostras associadas a classes emocionais,

sem qualquer ocorrência da classe neutra.

Os corpúscos são divididos em 80% para treino e 20% para teste, com amostragem estratificada. Os experimentos são conduzidos de forma independente por corpúscos e por configuração. Quando aplicado, o pré-processamento extensivo inclui: (i) remoção de *hashtags* emocionais, evitando a trivialização da tarefa; (ii) remoção de *stop words*, acentos e caracteres especiais; (iii) normalização de *stocks*<sup>3</sup>, URLs, valores monetários e porcentagens (da Silva et al., 2020); e (iv) lematização com *Stanza* (Qi et al., 2020).

Nos métodos supervisionados, os textos são representados por vetores TF-IDF (Manning et al., 2008), utilizando *scikit-learn* (Pedregosa et al., 2011). O RIPPER é aplicado diretamente sobre essa representação por meio da biblioteca *Wittgenstein* (Moscovitz, 2020). Árvores de decisão e *Random Forest* são treinados com *scikit-learn*, com seleção de hiperparâmetros via *Grid Search*. O espaço de busca inclui número de árvores (50, 100, 200, 500), profundidade máxima (nenhuma, 10, 20) e parâmetros mínimos de divisão (2, 5) e de folhas (1, 2).

## 5 Avaliação

Esta seção apresenta uma visão geral dos resultados experimentais, destacando tendências de desempenho entre os métodos avaliados e as diferenças observadas entre corpúscos e domínios. Também comparamos as abordagens sob a perspectiva do compromisso entre desempenho e interpretabilidade.

### 5.1 Resultados quantitativos

A Tabela 4 resume o desempenho dos métodos em cada corpúscos, utilizando *Medida-F macro* como métrica principal, por atribuir peso igual às classes e ser adequada a cenários desbalanceados (Sokolova and Lapalme, 2009). No corpúscos Domínio Livre sem pré-processamento, a anotação baseada em *hashtags* emocionais torna a tarefa trivial; sua remoção explica a queda acentuada de desempenho após o pré-processamento. De modo geral, observa-se um padrão estável de desempenho relativo: árvore de decisão e *Random Forest* obtêm os melhores resultados médios, seguidos por RIPPER, *LexicoR* e *ZeroR*. Essa ordenação se mantém na maioria dos corpúscos e configurações, com variações atribuídas ao domínio e ao esquema de anotação.

Embora o *Random Forest* generalize a árvore de decisão, esta última apresenta desempenho médio superior. Esse resultado é consistente com o desenho experimental: os modelos operam sobre representações esparsas (TF-IDF), com conjuntos de treinamento relativamente pequenos. Nessas condições, a combinação de múltiplas árvores com amostragem aleatória tende a diluir sinais lexicais fortes, enquanto uma única árvore determinística (por emoção) explora mais diretamente os atributos emocionalmente informativos. Esse resultado reforça que maior complexidade e custo computacional não implicam necessariamente melhor desempenho, especialmente quando o domínio não favorece a estrutura de modelos mais complexos específicos.

A Figura 1 apresenta o desempenho médio por corpúscos. A inclusão da classe neutra impacta negativamente os resultados ao intensificar o desbalanceamento e a ambiguidade entre classes. Apesar de resultados mistos nas árvores de decisão, o pré-processamento beneficia os métodos de forma geral, com ganhos mais consistentes nas abordagens simbólicas. Domínios em que as emoções são expressas por marcadores lexicais explícitos (Domínio Livre) ou vocabulário previsível (Notícias Curtas) apresentam melhor desempenho, enquanto domínios mais específicos e dependentes de contexto (DANTEStocks), exibem quedas generalizadas. Esses resultados indicam grande variação no desempenho absoluto entre domínios, mas estabilidade no desempenho relativo entre os métodos, reforçando a importância de avaliações comparativas multi-domínio.

### 5.2 Análise qualitativa

Do ponto de vista da interpretabilidade, observa-se um gradiente claro entre os métodos avaliados. O *LexicoR* e o RIPPER produzem resultados diretamente interpretáveis por meio de regras explícitas que associam termos ou combinações lexicais a emoções. Enquanto o desempenho do *LexicoR* depende fortemente da qualidade e cobertura do léxico, o que é um desafio diante da escassez desses recursos, o RIPPER surge como uma alternativa particularmente atraente.

As regras induzidas pelo RIPPER são simples e interpretáveis (Tabela 5): cada rótulo é definido por um conjunto de regras disjuntivas, bastando a satisfação de uma delas. Essas regras capturam termos emocionais (“incrível”, “ótimo”, “animado” e “empolgado”), modificadores de sentido (como

<sup>3</sup>Códigos de ações do mercado financeiro.

córpus	ZeroR	LexicoR	Arv. Dec.	RIPPER	Rand. Forest
DANTE_ComNeutro	0,0476	0,0921	<b>0,3120</b>	0,0950	0,1699
DANTE_ComNeutro_Preprocessado	0,0476	0,1269	<b>0,3152</b>	0,0715	0,1673
DANTE_SemNeutro	0,0689	0,0708	<b>0,3418</b>	0,0764	0,1887
DANTE_SemNeutro_Preprocessado	0,0689	0,0891	<b>0,3149</b>	0,0785	0,2110
DANTEConcordancia_ComNeutro	0,0966	0,1705	<b>0,2010</b>	0,1289	0,0961
DANTEConcordancia_ComNeutro_Preprocessado	0,0966	0,1172	<b>0,1711</b>	0,1191	0,0966
DANTEConcordancia_SemNeutro	0,0962	0,0000	0,1462	0,1197	<b>0,1855</b>
DANTEConcordancia_SemNeutro_Preprocessado	0,0962	0,0812	<b>0,2318</b>	0,1215	0,2150
DominioLivre	0,0807	0,0853	<b>0,9858</b>	0,9850	0,9777
DominioLivre_Preprocessado	0,0807	0,1299	<b>0,5022</b>	0,3613	0,4547
GoEmotionsBR_ComNeutro	0,0177	0,0422	<b>0,3141</b>	0,2506	0,1867
GoEmotionsBR_ComNeutro_Preprocessado	0,0177	0,0540	<b>0,2842</b>	0,2237	0,2073
GoEmotionsBR_SemNeutro	0,0088	0,0309	<b>0,3436</b>	0,2886	0,2087
GoEmotionsBR_SemNeutro_Preprocessado	0,0088	0,0442	<b>0,3236</b>	0,2827	0,2373
GoEmotionsBREkman_ComNeutro	0,0822	0,1809	<b>0,3959</b>	0,2007	0,2856
GoEmotionsBREkman_ComNeutro_Preprocessado	0,0822	0,2727	<b>0,3886</b>	0,2195	0,3427
GoEmotionsBREkman_SemNeutro	0,1213	0,1474	<b>0,4755</b>	0,2567	0,3830
GoEmotionsBREkman_SemNeutro_Preprocessado	0,1213	0,2631	<b>0,4777</b>	0,2788	0,4361
NoticiasCurtas_ComNeutro	0,0357	0,1041	<b>0,5083</b>	0,1881	0,4042
NoticiasCurtas_ComNeutro_Preprocessado	0,0357	0,1741	<b>0,5273</b>	0,2642	0,4698
NoticiasCurtas_SemNeutro	0,0476	0,1163	<b>0,4766</b>	0,2456	0,4441
NoticiasCurtas_SemNeutro_Preprocessado	0,0476	0,1501	<b>0,5289</b>	0,3352	0,5027
<b>Media Geral</b>	0,0639	0,1155	<b>0,3893</b>	0,2359	0,3123

Tabela 4: Resumo de desempenho (Medida- $F$  macro) de todos os métodos de classificação por córpus.

negação) e expressões compostas frequentes (“bom trabalho”). Contudo, mudanças substanciais de domínio exigem nova indução. Em contraste, o LexicoR é estático e dispensa retreino, mas requer ampla cobertura e atualização contínua. Por não demandar recursos extras, como um léxico, o RIPPER é especialmente adequado a cenários com recursos limitados e alta exigência de interpretabilidade.

Emoção	Regra	
admiracao	[incrivel=V]	OU
	[otimo=V]	OU
	[bom=V E trabalho=V]	
empolgacao	[nao=F E parecer=F E estar=V E animado=V]	OU
	[nao=F E empolgado=V]	

Tabela 5: Regras fornecidas pelo RIPPER para o córpus GoEmotionsBR pré-processado (=V indica a presença e =F a ausência de cada termo, respectivamente).

Entre os métodos de melhor desempenho, a árvore de decisão apresenta interpretabilidade intermediária: embora seus caminhos de decisão sejam compreensíveis, o conhecimento encontra-se distribuído em múltiplos ramos, o que dificulta uma visão global do comportamento do modelo. A Figura 2, que representa parte da subárvore dos quatro primeiros níveis para a classe “medo” no

córpus Domínio Livre pré-processado, mostra que o modelo identifica termos emocionais (“medo”, “assustador” e “terror”) e modificadores de sentido relevantes, como negação. Destaca-se a seleção do termo “estranho”, cuja relevância é coerente com o domínio do córpus, composto por publicações sobre o mercado de ações, em que variações inesperadas podem ser associadas ao medo dos acionistas. Essas evidências também ilustram o rápido crescimento do modelo, o que aumenta a complexidade interpretativa. O Random Forest, por sua vez, é substancialmente mais opaco, dificultando a inspeção direta de decisões individuais.

A análise qualitativa dos erros (Tabela 6) revela dois padrões principais nas falhas do método de melhor desempenho (árvores de decisão): (i) erros por omissão, nos quais um ou mais rótulos em exemplos *multilabel* não são previstos; e (ii) confusões em casos com maior nuance ou ambiguidade emocional. Em ambos, os erros estão associados a sinais emocionais fracos (linhas 4 e 5), textos curtos com pouca informação contextual ou ambíguos (linhas 2 e 4) e à sobreposição semântica entre emoções (linhas 1, 3, 4 e 5), evidenciando limites intrínsecos da tarefa.

### Desempenho Médio Geral por Corpus (F1\_macro)

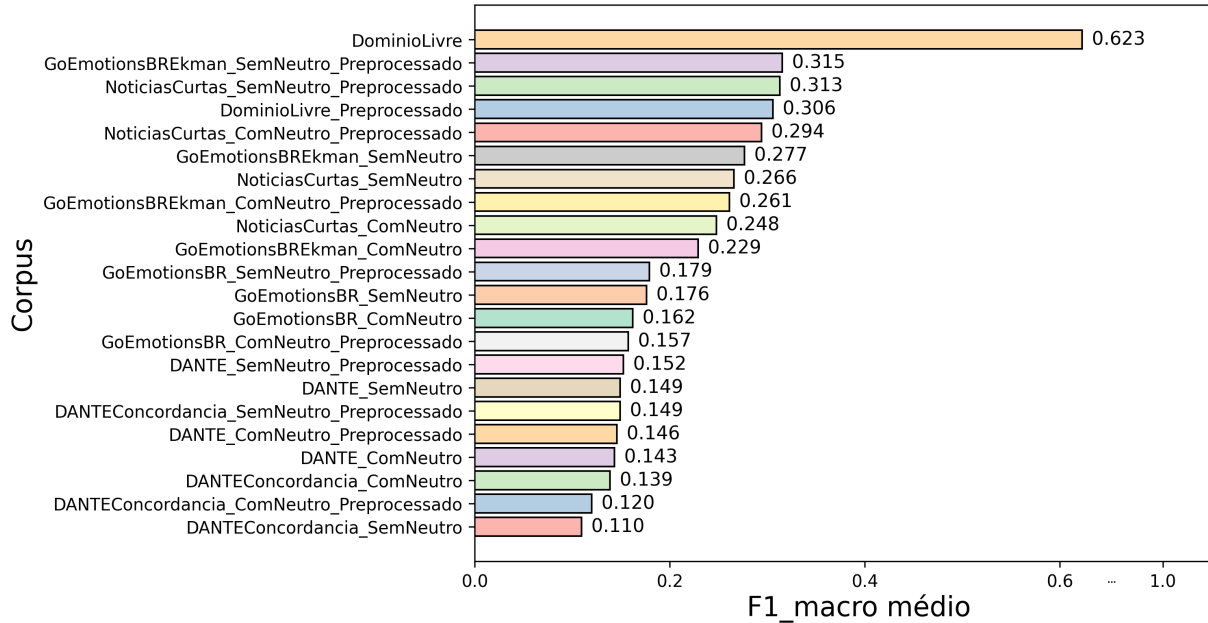


Figura 1: Desempenho (Medida- $F$  macro) médio de cada corpúsculo avaliado.

Texto	Esperado	Previsto
ela desrespeitou [NOME], não gosto dela	nojo, raiva	raiva
Parece deliciosamente entediante.	raiva	alegria
Sem contar que o vídeo completo faz parte da campanha, segundo o idiota do [NOME].	tristeza	raiva
Tenho certeza de que isso subiu para pelo menos 10 mil inscritos. Pode ser até mais alto agora!	alegria	surpresa
Filhotes de caracal nascem em cativeiro: Espécie está ameaçada de extinção em Israel.	surpresa	alegria

Tabela 6: Exemplos de erros da árvore de decisão. O marcador [NOME] foi introduzido apenas neste artigo para fins de reprodução, com o objetivo de evitar a identificação e associação com as figuras públicas citadas.

## 6 Impressões e discussões teóricas

Os corpúsculos analisados abrangem desde textos formais, como manchetes jornalísticas, até conteúdos informais de redes sociais, permitindo avaliar o efeito do estilo textual na detecção de emoções. Textos mais estruturados e emocionalmente ricos, como obras literárias, seriam relevantes, mas não encontramos recursos públicos em português com essas características.

Os resultados mostram que métodos simbólicos são mais eficazes em domínios com emoções explícitas, vocabulário previsível e menor ambiguidade. Nesses cenários, maior complexidade não implica

melhor desempenho, como no caso das árvores de decisão frente ao Random Forest. Observa-se, assim, um compromisso entre desempenho e interpretabilidade: modelos mais complexos tendem a obter melhores médias, enquanto métodos baseados em regras oferecem resultados mais transparentes, fundamentais em aplicações sensíveis.

A análise qualitativa dos erros evidencia limites intrínsecos da tarefa, sobretudo em cenários *multi-label* e com sinais emocionais sutis ou ambíguos. Os corpúsculos avaliados adotam modelos de emoção com pressupostos distintos, que vão de abordagens baseadas em evidências fisiológicas inatas (Ekman) a esquemas derivados de *crowdsourcing* (GoEmotions) e modelos que assumem exclusividade mútua entre estados emocionais (Roda de Plutchik), evidenciando a dificuldade de definir emoções de forma consensual. O domínio e o esquema de anotação influenciam fortemente os resultados, e corpúsculos desbalanceados ou com expressão emocional sutil permanecem desafiadores.

Observa-se também forte desbalanceamento de polaridade: o GoEmotions-BR é o modelo mais granular, com múltiplas emoções positivas, enquanto os demais corpúsculos assumem, em geral, apenas a classe “alegria”. Modelos mais granulares ampliam a ambiguidade e a subjetividade, dificultando distinguir erros de anotação de limitações do classificador. A classe neutra é particularmente indefinida, frequentemente associada a expressões

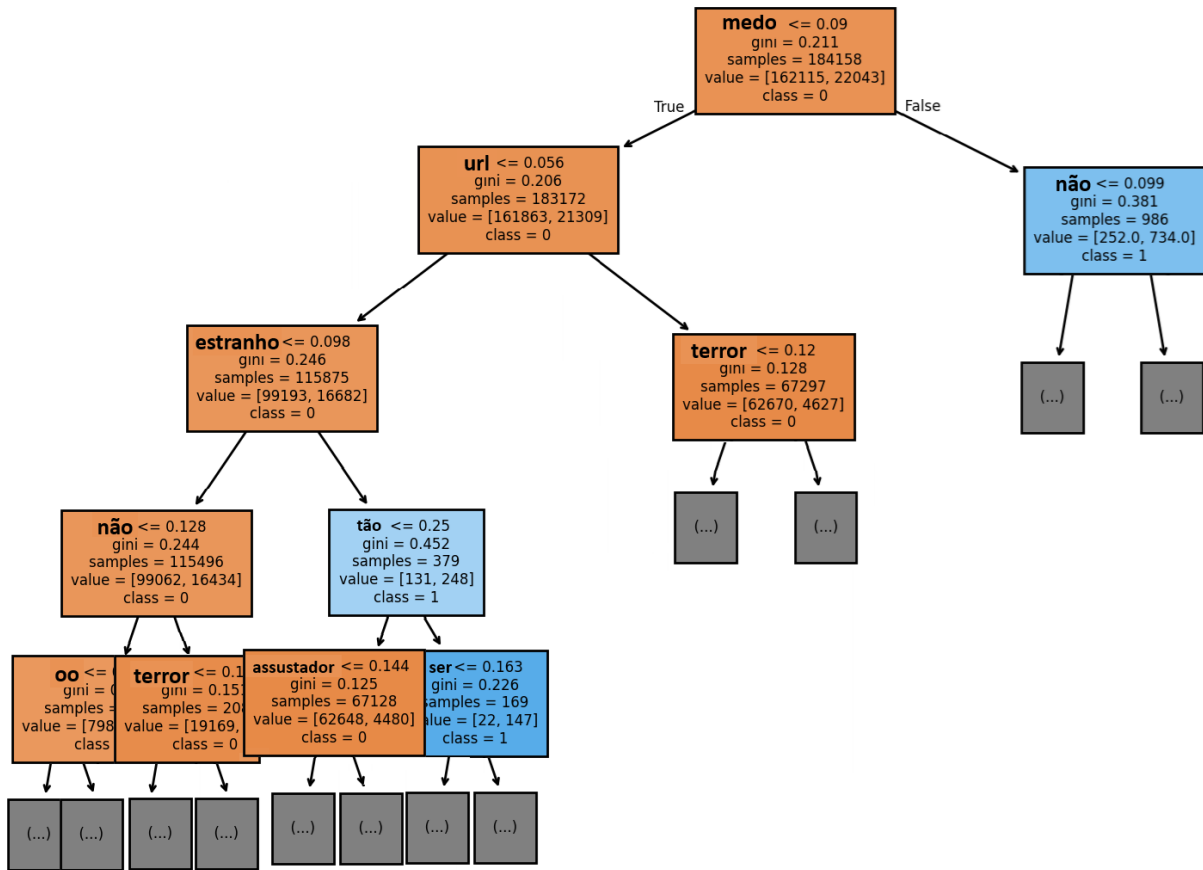


Figura 2: Árvore de decisão da classe “medo” para o corpus Domínio Livre pré-processado. A cor de cada nó indica a classe majoritária prevista pela subárvore enraizada nesse nó. Neste caso, as classes correspondem à presença (azul) ou à ausência (laranja) do rótulo “medo”. A intensidade da cor é proporcional à pureza do nó, definida pela proporção de amostras da classe majoritária associadas a esse nó.

fracas ou ambíguas, o que intensifica a confusão entre classes. Em síntese, não há um método universalmente superior; a escolha depende do contexto, do modelo de emoção e dos requisitos de interpretabilidade.

Como trabalhos futuros, destacam-se abordagens híbridas simbólico-neurais e a indução automática de regras a partir de representações linguísticas mais ricas que o TF-IDF, como *word embeddings* aprendidos por modelos neurais, visando conciliar desempenho, robustez e interpretabilidade. Pretendemos também avançar no estudo mais amplo do qual este trabalho faz parte, ampliando a comparação entre paradigmas de classificação, modelos de emoção e domínios, para analisar sistematicamente suas vantagens e limitações no contexto do português.

**Limitações**

As principais limitações incluem sensibilidade ao domínio, dependência de vocabulário emocional

explícito e menor transparência nos modelos de melhor desempenho. No método LéxicoR, o desempenho depende diretamente da qualidade e cobertura dos recursos, ainda escassos para a língua portuguesa.

**Considerações éticas**

Os classificadores de emoções aqui apresentados baseiam-se em modelos teóricos com limitações inerentes, dados os desafios de obter consenso científico sobre a definição e a delimitação das emoções humanas. Assim, suas previsões não devem ser tratadas como evidências objetivas nem utilizadas em contextos sensíveis ou para tomada de decisão automatizada, devendo restringir-se a uso auxiliar e informativo, sob responsabilidade e supervisão humana.

Ademais, essas ferramentas suscitam riscos éticos relevantes, incluindo a possibilidade da informação ser utilizada para manipulação ou indução de estados emocionais para fins comerciais e/ou

políticos. Questões adicionais envolvem vieses nos dados e nos modelos, falta de transparência e riscos à privacidade e ao consentimento, reforçando a necessidade de cautela quanto à interpretação e ao uso desses sistemas.

## Agradecimentos

Agradecemos aos autores dos recursos utilizados por disponibilizarem os dados para pesquisa e contribuírem no avanço da área para o português.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Este projeto também foi apoiado pelo Ministério da Ciência, Tecnologia e Inovações, com recursos da Lei N. 8.248, de 23 de outubro de 1991, no âmbito do PPI-Softex, coordenado pela Softex e publicado como Residência em TIC 13, DOU 01245.010222/2022-44.

## Referências

- Amer F. A. H. Alnuaimi and Tasnim H. K. Albaldawi. 2024. [An overview of machine learning classification techniques](#). *BIO Web of Conferences*, 97:00133.
- Leo Breiman. 2001. [Random Forests](#). *Machine Learning*, 45(1):5–32.
- William W. Cohen. 1995. [Fast Effective Rule Induction](#). In Armand Frieditis and Stuart Russell, editors, *Machine Learning Proceedings 1995*, pages 115–123. Morgan Kaufmann, San Francisco (CA).
- Diogo Cortiz, Jefferson Silva, Newton Calegari, Ana Freitas, Ana Soares, Carolina Botelho, Gabriel Rêgo, Waldir Sampaio, and Paulo Boggio. 2021. [A Weakly Supervised Dataset of Fine-Grained Emotions in Portuguese](#). In *Anais do XIII Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*, pages 73–81, Porto Alegre, RS, Brasil. SBC.
- Fernando J. Vieira da Silva. 2020. [Cross-domain emotion detection in tweets](#). Tese de doutorado, Universidade Estadual de Campinas, Instituto de Computação, Brasil.
- Fernando J. Vieira da Silva, Norton T. Roman, and Ariadne M. B. R. Carvalho. 2020. [Stock market tweets annotated with emotions](#). *Corpora*, 15(3):343–354.
- Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. [GoEmotions: A Dataset of Fine-Grained Emotions](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4040–4054, Online. Association for Computational Linguistics.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Mariza Miola Dosciatti, Lohann Paterno Coutinho Ferreira, and Emerson Cabrera Paraiso. 2013. [Identificando emoções em textos em português do brasil usando máquina de vetores de suporte em solução multiclasse](#). *X ENIAC: Encontro Nacional de Inteligência Artificial e Computacional. Fortaleza, Brasil*.
- Luís Carlos Fernandes Duarte. 2019. Reconhecimento automático de emoções em texto com recurso a emojis. Dissertação de mestrado, Universidade de Coimbra, Portugal.
- Paul Ekman and Dacher Keltner. 1997. [Universal facial expressions of emotion: An old controversy and new findings](#). In Ullica C. Segerstråle and Peter Molnár, editors, *Nonverbal Communication: Where Nature Meets Culture*, pages 27–46. Lawrence Erlbaum Associates, Mahwah, NJ.
- Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. 2018. [A Survey of Methods for Explaining Black Box Models](#). *ACM Comput. Surv.*, 51(5).
- Luiz Otávio Alves Hammes and Larissa Astrogildo de Freitas. 2021. [Utilizando BERTimbau para a Classificação de Emoções em Português](#). In *Anais do XIII Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana (STIL)*, pages 56–63, Porto Alegre, RS, Brasil. SBC.
- C. D. Manning, P. Raghavan, and H. Schütze. 2008. [Introduction to Information Retrieval](#). An Introduction to Information Retrieval. Cambridge University Press.
- Abdullah Al Maruf, Fahima Khanam, Md. Mahmudul Haque, Zakaria Masud Jiyad, M. F. Mridha, and Zeyar Aung. 2024. [Challenges and Opportunities of Text-Based Emotion Detection: A Survey](#). *IEEE Access*, 12:18416–18450.
- Saif M. Mohammad. 2016. [9 - Sentiment Analysis: Detecting Valence, Emotions, and Other Affectual States from Text](#). In Herbert L. Meiselman, editor, *Emotion Measurement*, pages 201–237. Woodhead Publishing.
- Ilan Moscovitz. 2020. [wittgenstein: Ruleset covering algorithms for transparent machine learning \(RIPPER\)](#). <https://github.com/imoscovitz/wittgenstein>. Acesso em: 2026-01-29.
- Sreerama K. Murthy. 1998. [Automatic construction of decision trees from data: A multi-disciplinary survey](#). *Data Mining and Knowledge Discovery*, 2(4):345–389.

Francisco Bráulio Oliveira and Jaime Simão Sichman. 2024. [Portuguese Emotion Detection Model Using BERTimbau Applied to COVID-19 News and Replies](#). In *Anais da XXXIV Brazilian Conference on Intelligent Systems (BRACIS)*, pages 265–280, Porto Alegre, RS, Brasil. SBC.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. [Scikit-learn: Machine Learning in Python](#). *Journal of Machine Learning Research*, 12(85):2825–2830.

Robert Plutchik. 2001. [The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice](#). *American Scientist*, 89(4):344–350.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. [Stanza: A Python Natural Language Processing Toolkit for Many Human Languages](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 101–108. Association for Computational Linguistics.

Barbara Ramos. 2021. [Descrição de uma metodologia desenvolvida para revisão de um léxico de palavras de emoção](#). In *Anais do XIII Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*, pages 389–397, Porto Alegre, RS, Brasil. SBC.

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. [“Why Should I Trust You?”: Explaining the Predictions of Any Classifier](#). In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’16*, page 1135–1144, New York, NY, USA. Association for Computing Machinery.

Allisfrank dos Santos. 2019. [Análise de sentimento multiclasse: uma abordagem com o uso de aprendizado de máquina](#). Dissertação de mestrado, Universidade Federal de São Carlos, Brasil.

Dibyendu Seal, Uttam K. Roy, and Rohini Basak. 2020. [Sentence-level emotion detection from text based on semantic rules](#). In *Information and Communication Technology for Sustainable Development: Proceedings of ICT4SD 2018*, pages 423–430. Springer Singapore.

Marina Sokolova and Guy Lapalme. 2009. [A systematic analysis of performance measures for classification tasks](#). *Information Processing Management*, 45(4):427–437.

Fábio Souza, Rodrigo Nogueira, and Roberto Lotufo. 2020. [BERTimbau: Pretrained BERT Models for Brazilian Portuguese](#). In *Intelligent Systems: 9th Brazilian Conference (BRACIS 2020)*, pages 403–417. Springer International Publishing.

## A Implementação do LexicoR

Com o objetivo de facilitar a compreensão e a reprodutibilidade do trabalho, este apêndice detalha o funcionamento do algoritmo LexicoR, proposto na Seção 4.1. O pseudo-código para a classificação de uma amostra é apresentado no Algoritmo 1.

O léxico de emoções (Emocionário), que associa palavras a emoções, é lido e a associação é mantida em uma estrutura de dicionário. São definidas as listas de modificadores semânticos de negação, intensificação e redução, apresentadas na Tabela 7.

Tipo de Modificador	Lista de Palavras
Negação	"jamais", "nada", "nem", "nenhum", "ninguém", "ninguem", "nunca", "não", "nao", "tampouco", "sem", "sequer"
Intensificação	"mais", "muito", "demais", "completamente", "absolutamente", "totalmente", "definitivamente", "extremamente", "frequentemente", "bastante", "super", "altamente", "incrivelmente", "intensamente", "profundamente", "terrivelmente", "tão", "tao", "enormemente"
Redução	"pouco", "quase", "menos", "apenas", "levemente", "ligeiramente", "parcialmente", "raramente", "ocasionalmente", "moderadamente"

Tabela 7: Listas de palavras associadas a cada tipo de modificação.

Para classificar uma amostra, cada *token* da amostra é analisado individualmente. Se o *token* também estiver presente no Emocionário, então essa palavra representa alguma das emoções mapeadas pelo léxico emocional. Nesse caso, o algoritmo incrementa a pontuação da classe (emoção) em questão.

Para definir o valor do incremento, uma janela com as quatro palavras que apareceram antes do *token* atual e com as quatro palavras que aparecem depois do *token* é considerada. Se não houver modificadores de negação, intensificação ou redução na janela considerada, a pontuação da classe é incrementada em 1. Caso haja algum modificador de negação, o incremento da pontuação é -3. Os modificadores de intensificação e redução alteram o incremento para 3 e 1/3, respectivamente. Se um modificador de negação aparecer juntamente de algum dos outros modificadores, o incremento é o inverso: 3 para 1/3 e vice-versa.

Depois de repetir este processo para cada *token* da amostra, o algoritmo escolhe e prevê as classes com a maior pontuação para a amostra. Caso a amostra não possua nenhum marcador de emoção

do Emocionário, a pontuação de todas as classes será zero e a classe neutra será atribuída à amostra.

---

**Algoritmo 1** Pseudo-código do LexicoR

---

**Entrada:** Amostra a ser classificada  $a$

**Entrada:** Léxico emocional  $emoc$

**Saída:** Classes de emoção preditas  $c$

```
1:  $neg \leftarrow$  conjunto de modificadores de negação
2:  $inten \leftarrow$  conjunto de modificadores de intensificação
3:  $reduc \leftarrow$  conjunto de modificadores de redução
4:
5:  $scores \leftarrow$  vetor com o valor 0 para todas as classes de
   emoções em  $emoc$ 
6: para cada token  $p$  de  $a$  do
7:   se  $p \in emoc$  então
8:      $label \leftarrow$  emoção associada a  $p$  no léxico  $emoc$ 
9:      $ctx \leftarrow$  4 palavras anteriores e 4 palavras posteriores a  $p$  em  $a$ 
10:    se  $ctx$  contém alguma palavra de  $inten$  então
11:      se  $ctx$  contém alguma palavra de  $neg$  então
12:         $peso \leftarrow 1/3$ 
13:      senão
14:         $peso \leftarrow 3$ 
15:      fim se
16:    senão se  $ctx$  contém alguma palavra de  $reduc$ 
então
17:      se  $ctx$  contém alguma palavra de  $neg$  então
18:         $peso \leftarrow 3$ 
19:      senão
20:         $peso \leftarrow 1/3$ 
21:      fim se
22:    senão se  $ctx$  contém alguma palavra de  $neg$  então
23:       $peso \leftarrow -3$ 
24:    senão
25:       $peso \leftarrow 1$ 
26:    fim se
27:     $scores[label] \leftarrow scores[label] + peso$ 
28:  fim se
29: fim para
30:
31: se todos os valores de  $score$  forem 0 então
32:   retorne "neutro"
33: fim se
34:
35:  $maior \leftarrow \max(\{score\})$ 
36: retorne todas as classes  $label$  tal que  $score[label] = maior$ 
```

---