

ANÁLISE DE DADOS RECONHECENDO EMOÇÕES EM TEXTOS COM OBJETIVO DE OBTER INFORMAÇÕES ÚTEIS PARA VENDAS UTILIZANDO AS FERRAMENTAS NAIVE BAYES E NLTK COM PYTHON

Autor: Gustavo de Castro Minhaneli

Orientador: Maicon Vinícius Ribeiro

Caratinga, 2018

Agenda

- Introdução
- Referencial teórico
- Objetivo geral
- Metodologia
- Análise dos resultados
- Conclusão
- Trabalhos Futuros
- Principais referências

Introdução

- Mineração de textos
- Fonte de coleta de dados
- Com a superficialidade que domina as relações humanas, cada vez mais, o ser humano acaba por expressar opiniões nas redes sociais, tendo em vista que as emoções podem ser demonstradas e percebidas de diversas formas.

Referencial teórico

- ▶ Segundo WEISS, S. M e et al (2010) as técnicas utilizadas na Mineração de Textos, são semelhantes as utilizadas na Mineração de Dados, faz-se o uso dos mesmos métodos de aprendizagem, independente se uma técnica utiliza dados textuais com Mineração de Textos e a outra com dados numéricos utilizando Mineração de Dados.
- ▶ Em minerações de emoções não se realiza a análise do conteúdo que os textos expressa, mas sim a emoção predominante no mesmo (RODRIGUES et al., 2011).

Referencial teórico

- ▶ As emoções básicas ocorrem a partir de eventos externos ao indivíduo e são rapidamente compartilhadas entre membros de um mesmo grupo (SANTOS; GRAMINHA, 2006).
- ▶ LOPER; E; S (2002) descrevem o *NLTK(Natural Language Toolkit)* como uma suíte de aplicativos e módulos de código-fonte aberto, para prover o aprendizado da linguagem natural. Essa ferramenta é utilizada em aplicações de tratamento de texto, como por exemplo, análise de sentimento em discursos e Mineração de Opiniões do público em publicações (KOTHAPALLI; SHARIFAHMADIAN; SHIH, 2016).

Objetivo geral

- ▶ A concepção foi caracterizar frases emocionalmente utilizando a biblioteca *NLTK*(Natural Language Toolkit) com *Python*, para estimar em qual classificação emocional básica cada frase refere-se, auxiliando na tomada de decisão, coletiva ou individual.

Metodologia

- Descrição do estudo de caso
- Algoritmo de coleta de dados
- Limpeza dos dados
- Algoritmo de pré-processamento e classificação
- Algoritmo de comparação de bases textuais
- Avaliação do algoritmo

Metodologia

- Criação do algoritmo para coleta de textos no *Twitter*. (Apêndice A)
- *Consumer Key Secret*
- *Token Secret*

The screenshot displays the Twitter API application settings interface. At the top, there are four tabs: 'Details', 'Settings', 'Keys and Access Tokens', and 'Permissions'. The 'Keys and Access Tokens' tab is active.

Application Settings
Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.

Consumer Key (API Key) [Redacted]

Consumer Secret (API Secret) [Redacted]

Access Level: Read, write, and direct messages (modify app permissions)

Owner: [Redacted]

Owner ID: [Redacted]

Application Actions

Regenerate Consumer Key and Secret | Change App Permissions

Your Access Token
This access token can be used to make API requests on your own account's behalf. Do not share your access token secret with anyone.

Access Token: [Redacted]

Access Token Secret: [Redacted]

Access Level: Read and write

Owner: [Redacted]

Owner ID: [Redacted]

Token Actions

Regenerate My Access Token and Token Secret | Revoke Token Access

Fonte: Página da API do Twitter <https://apps.twitter.com/app>

Metodologia

- Dados foram coletados no período entre o dia 05 de setembro à 23 de outubro de 2018.
- 357 foram utilizadas na classificação.

Quadro 2 - Amostra dos textos coletados na rede social Twitter.

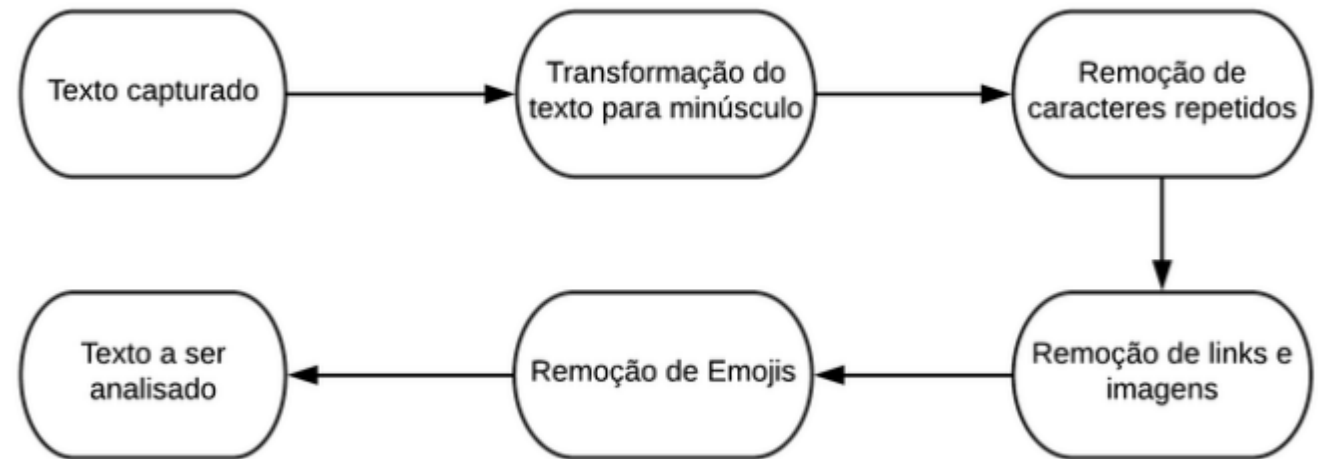
Usuários	Texto Publicado
Usuário 1	quero comprar um body novo para ir no jiad
Usuário 2	quero comprar logo esse ingresso do sítio
Usuário 3	não precisa ter tanto livro eu que gosto de comprar
Usuário 4	estou furioso com estes valores
Usuário 5	you faz por qual preço em meu docinho?
Usuário 6	vou comprar um lanche e um sorvete e ver um filme
Usuário 7	mês que vem não vou gastar um real no cartão está dito
Usuário 8	onde posso comprar um xbox?
Usuário 9	finalmente comprei um celular
Usuário 10	quero comprar tantas coisas

Fonte: O autor (2018)

Metodologia

- Fluxo de Limpeza dos Textos

Figura 11 - Fluxo de pré-processamento



Fonte: Adaptado de Marques (2014)

Metodologia

- Mapeamento de abreviações
- Limpeza dos dados

quero comprar meu cel hj <https://t.co/3o4rz59412>



quero comprar meu celular hoje



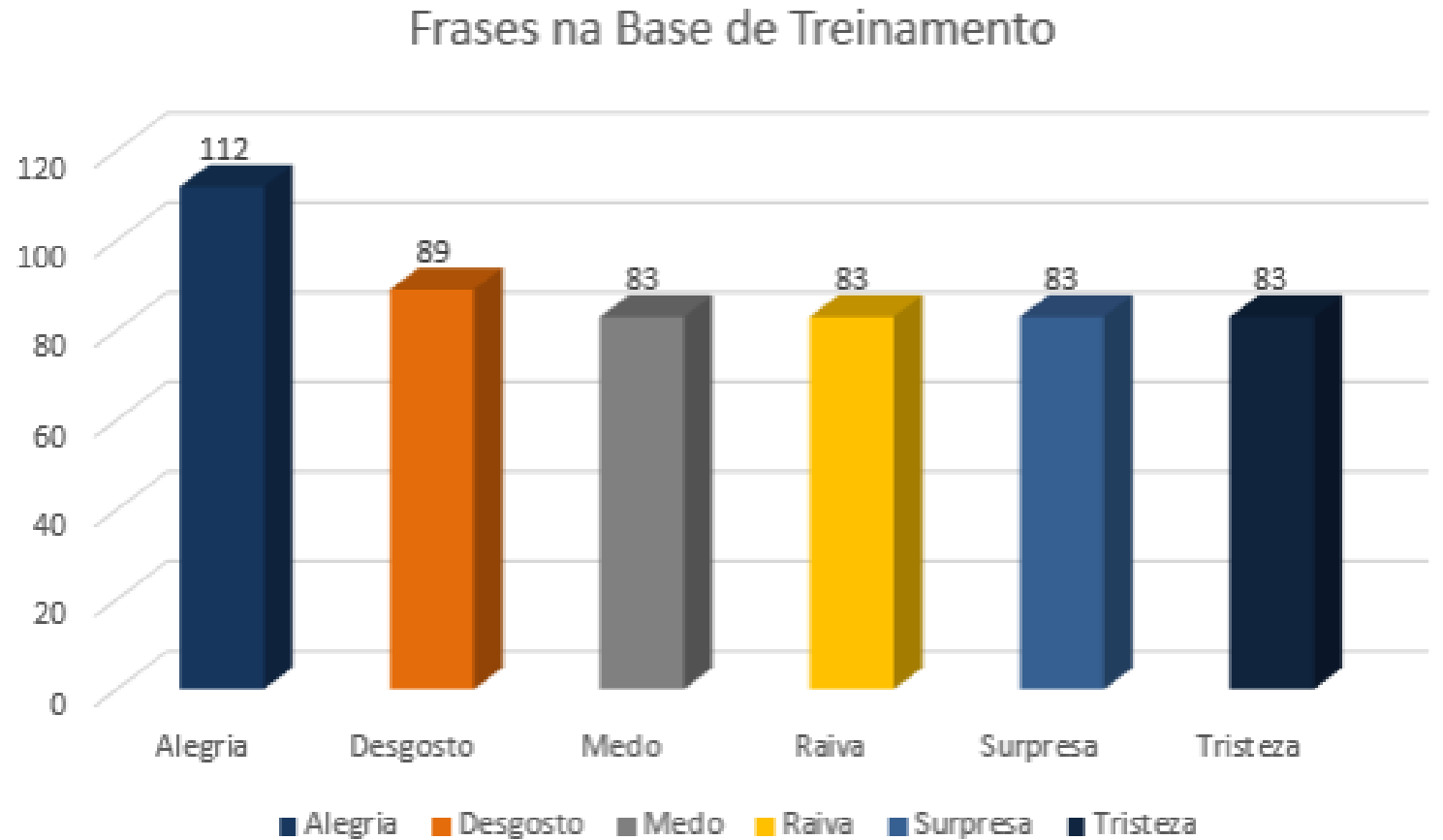
quer', 'compr', 'celul', 'hoj'

Abreviações	Significado	Abreviações	Significado
qd	quando	sqn	só que não
pq	porque	qse	quase
tb	também	c/	com
tbm	também	fds	fim de semana
qnd	quando	obs	observação
ngm	ninguém	hrs	horas
pfv	por favor	seg	segunda
algm	alguém	segs	segundos
kd	cadê	ter	terça
pqp	(palavrão)	qua	quarta
vc	você	qui	quinta
qm	quem	sex	sexta
q	que	sab	sábado
vcs	vocês	dom	domingo
p	para	tel	telefone
p/	para	n	não
cel	celular	ñ	não
dm	mensagem direta	tds	todos

Fonte: Adaptado de Marques (2014)

Metodologia

- Base previamente classificada (Anexo A)
- Total de 543 frases, utilizadas como a base de para classificar como 70% e os outros 30%, resultante em 54 frases



Metodologia

- Algoritmo de comparação das bases textuais

Base Coletada



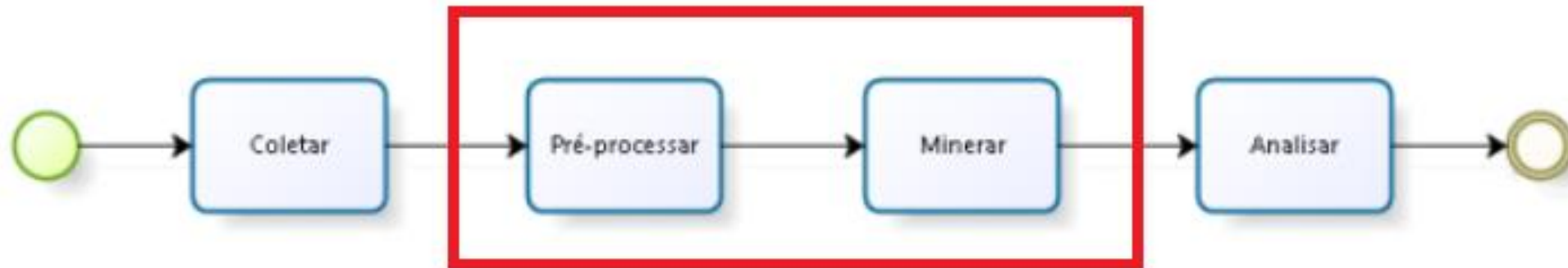
Base já Treinada



Metodologia

- Algoritmo de pré-processamento e classificação

Figura 2 - Processo de Descoberta de Conhecimento em Textos



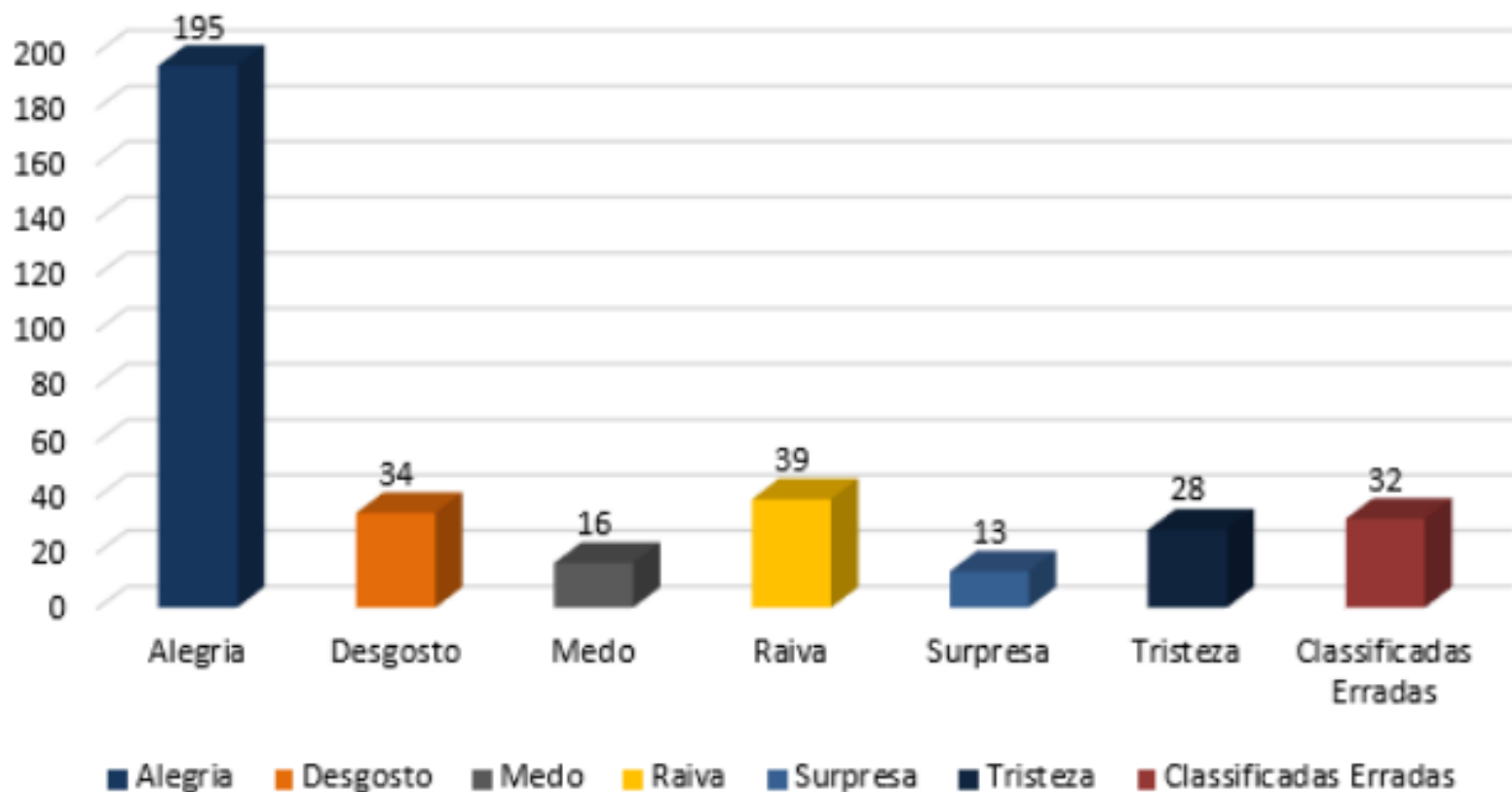
Fonte: O autor (2018) ARANHA e VELLASCO (2007)

Análise de resultados

- ▶ Classes que as frases foram classificadas
- ▶ Frases coletadas de cada emoção básica com a palavra “comprar”

Resultado da classificação das frases

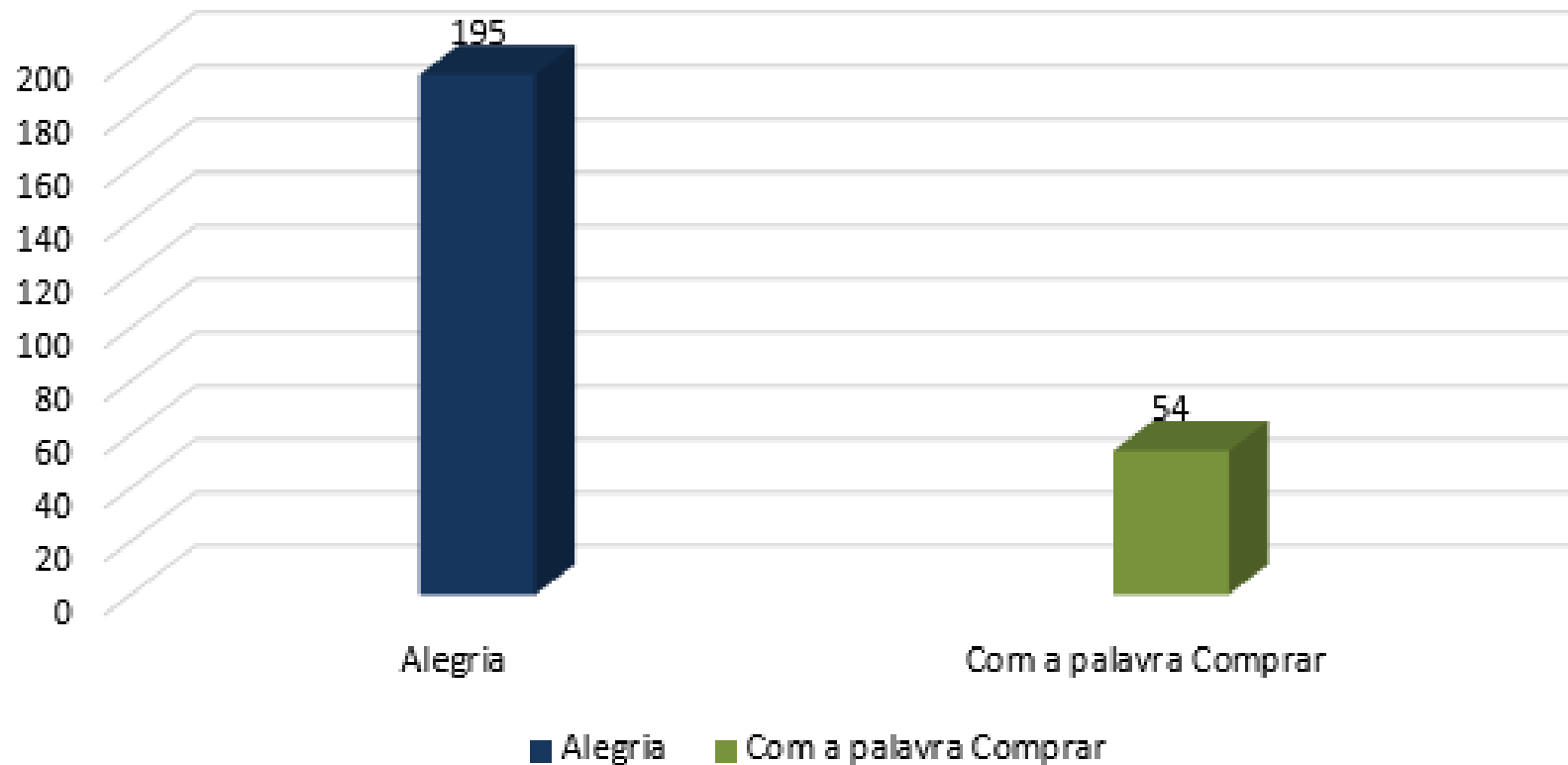
Quantidade de Frases Classificadas



Fonte: O autor (2018)

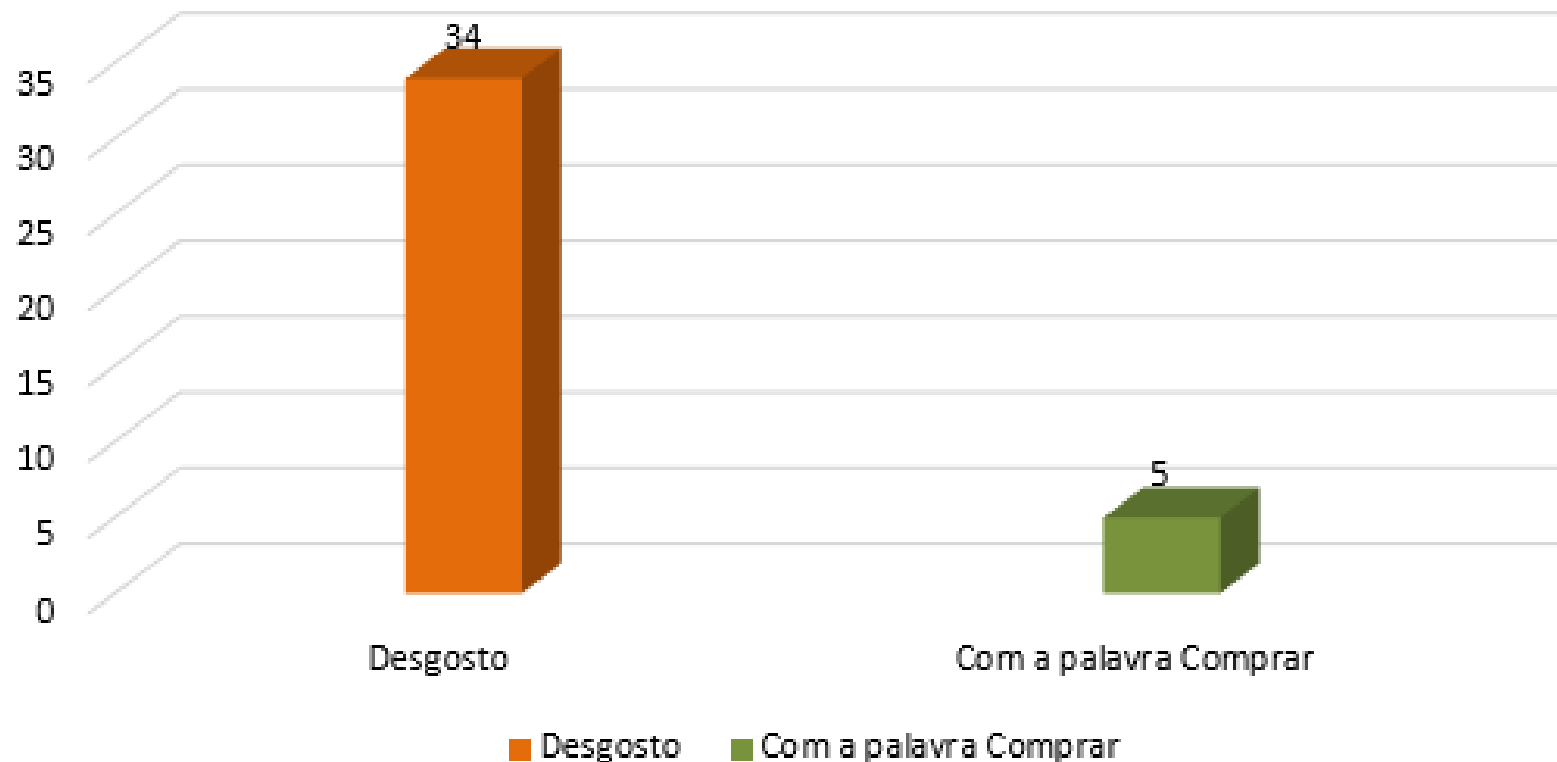
Resultado da verificação da palavra comprar nas frases de Alegria

Quantidade de Frases de Alegria que Continham a Palavra Comprar



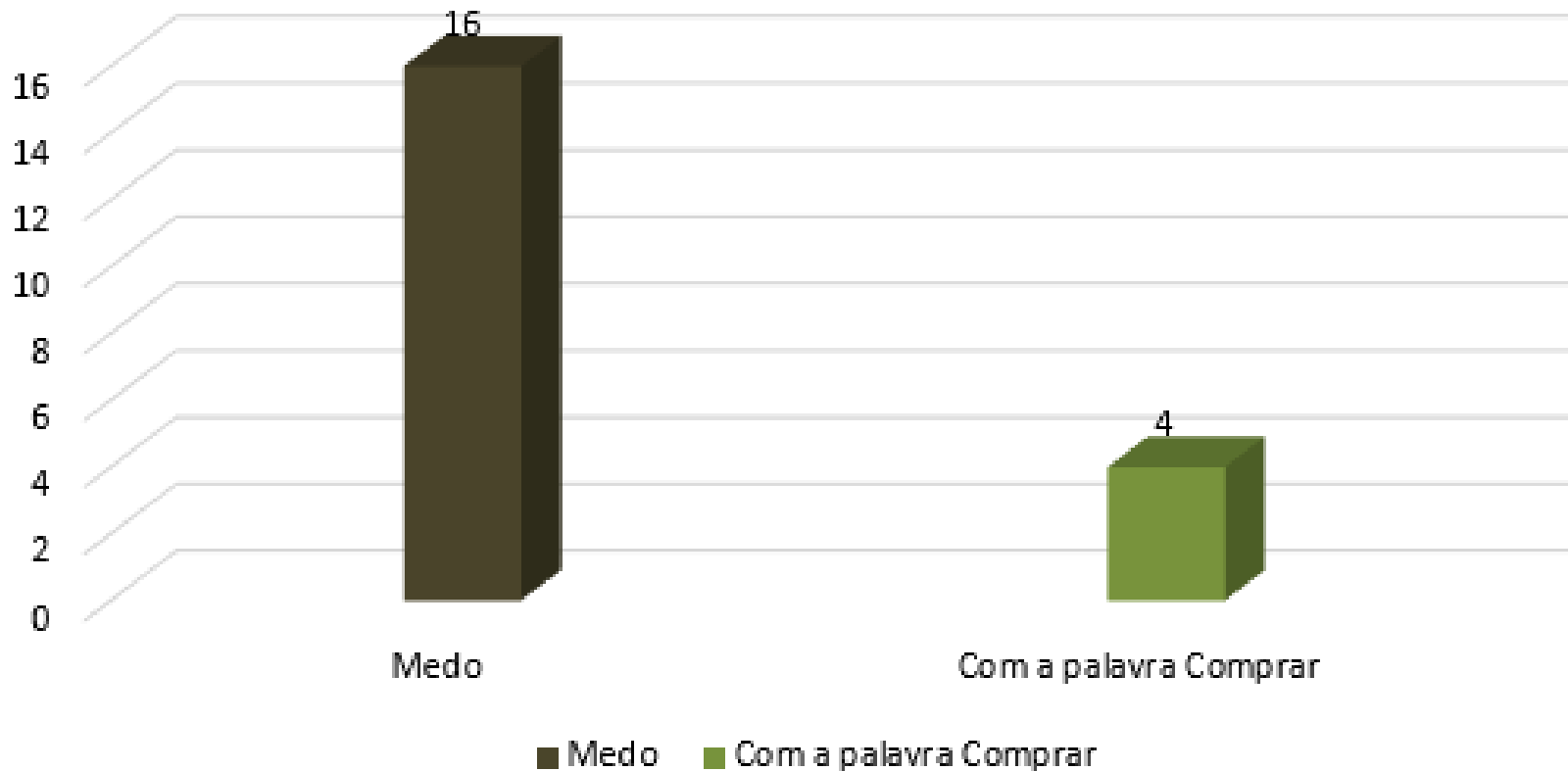
Resultado da verificação da palavra comprar nas frases de Desgosto

Quantidade de Frases de Desgosto que Continham a Palavra Comprar



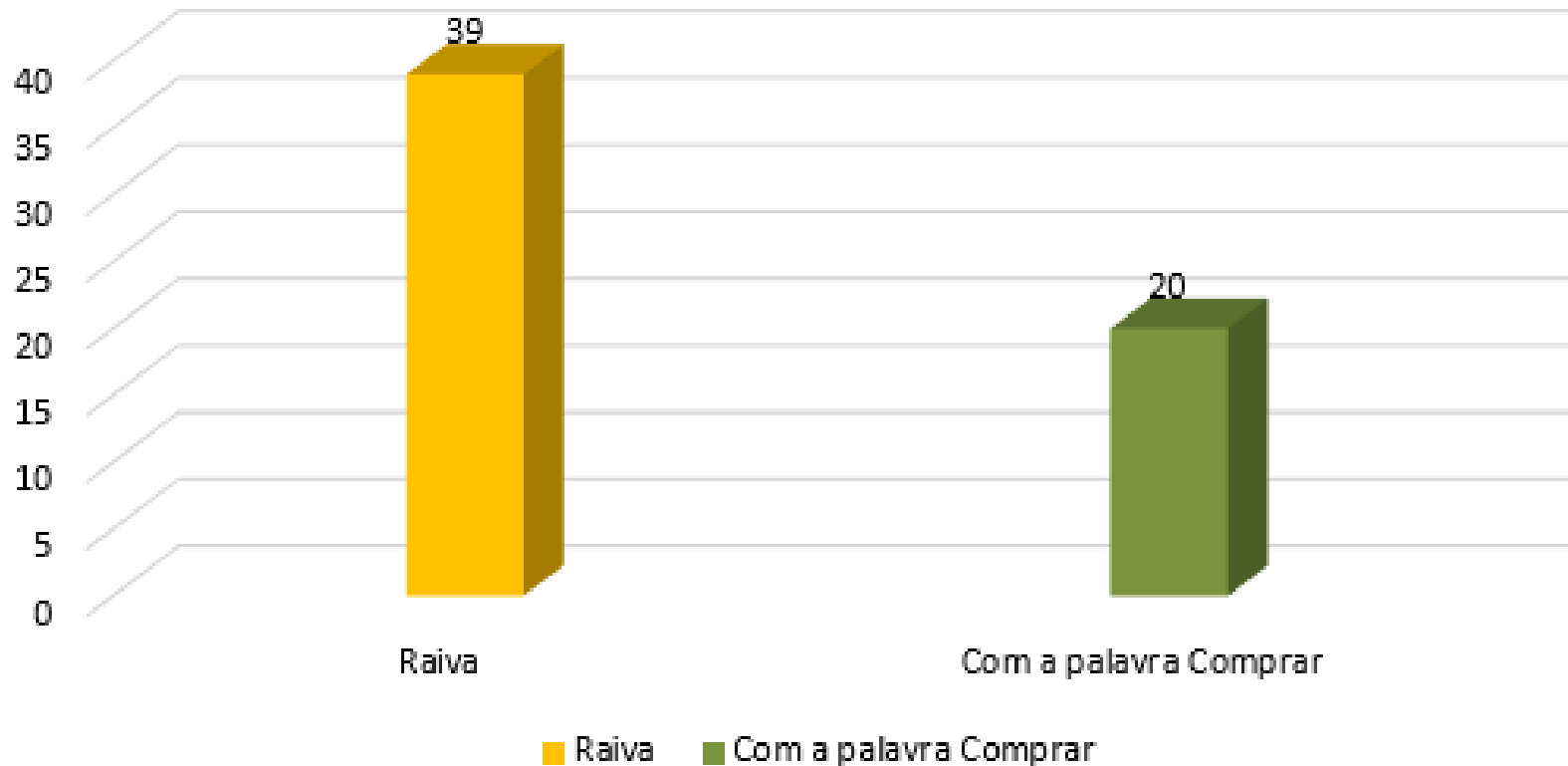
Resultado da verificação da palavra comprar nas frases de Medo

Quantidade de Frases de Medo que Continham a Palavra Comprar



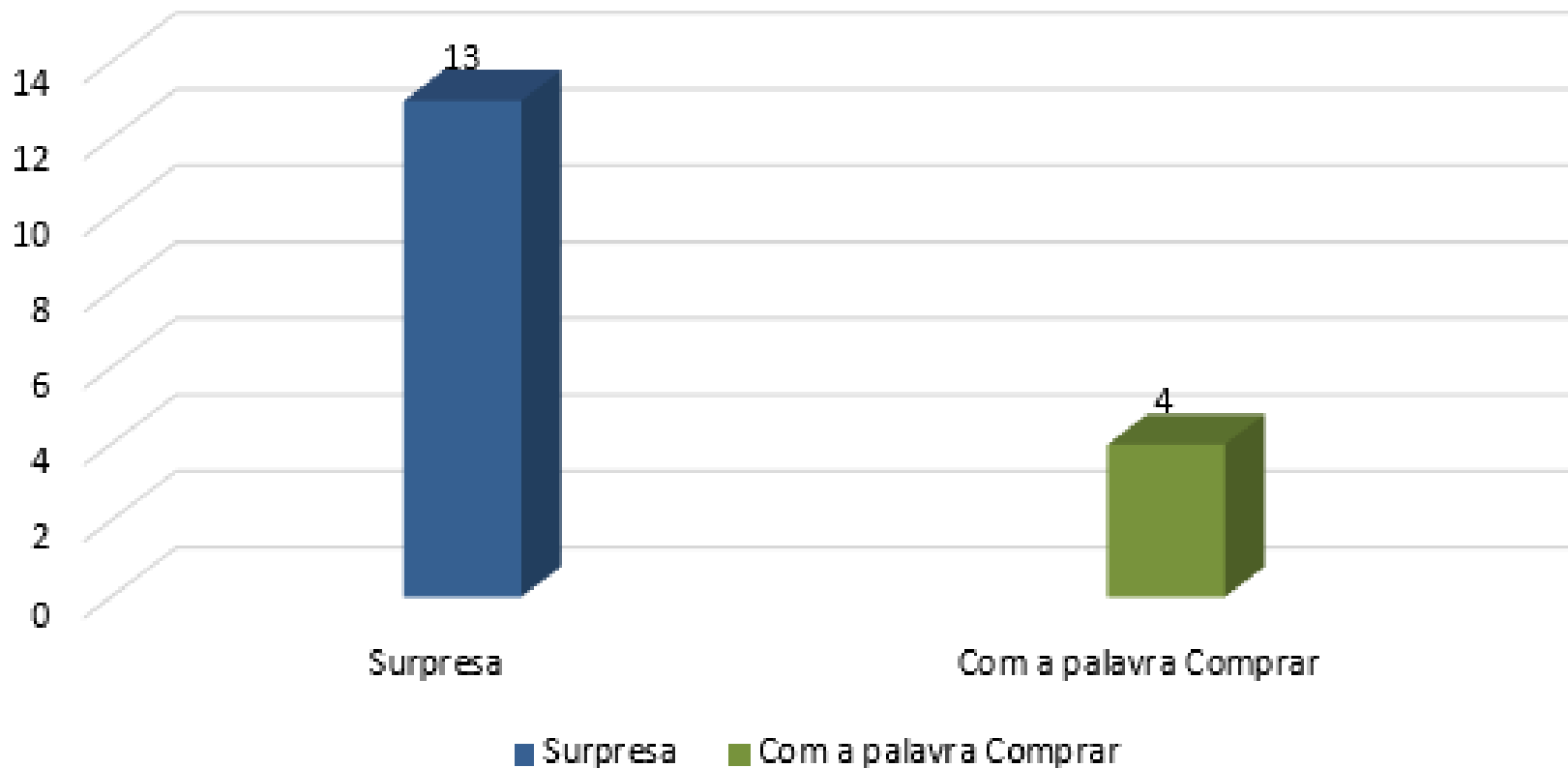
Resultado da verificação da palavra comprar nas frases de Raiva

Quantidade de Frases de Raiva que Continham a Palavra Comprar



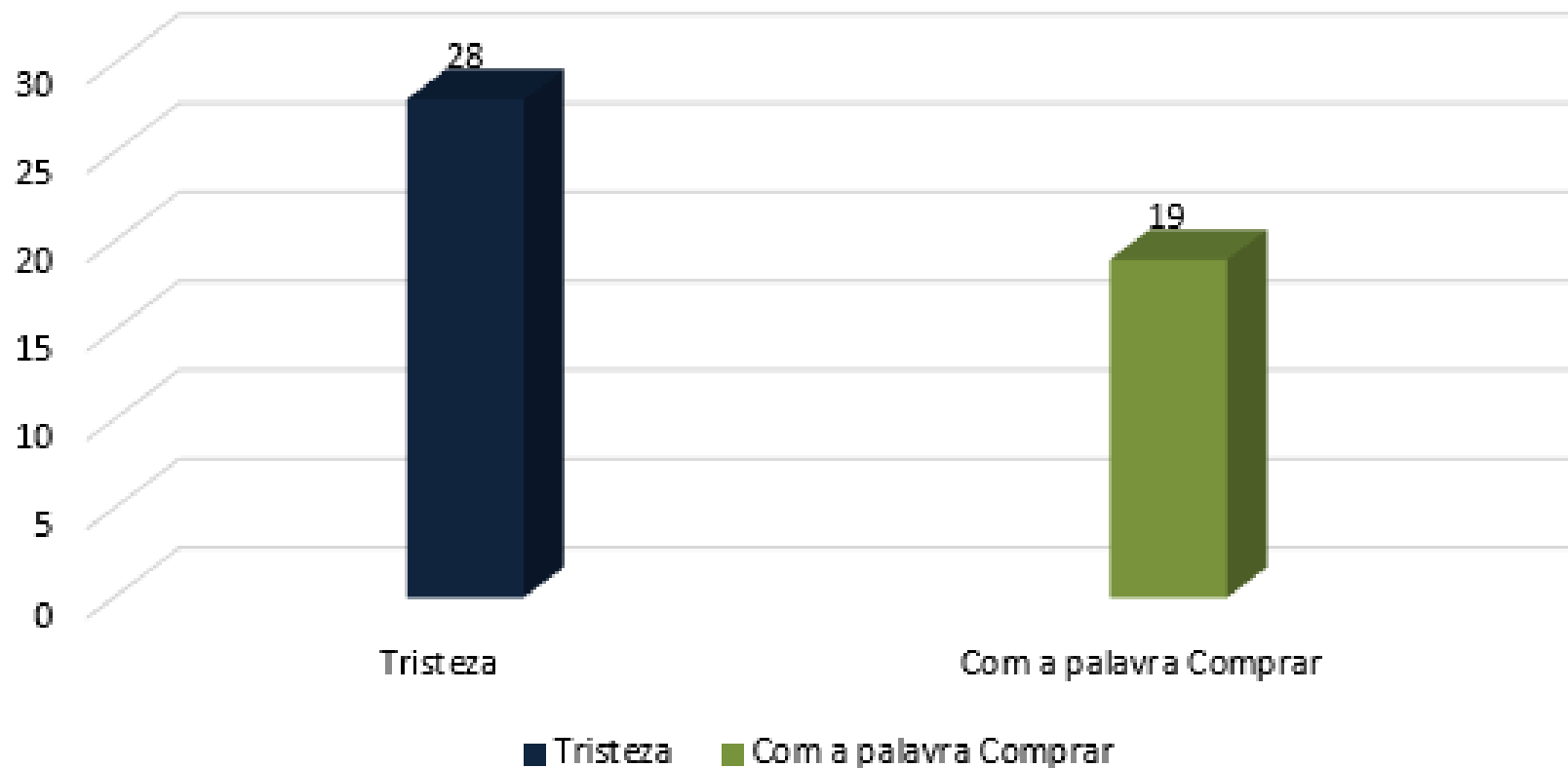
Resultado da verificação da palavra comprar nas frases de Surpresa

Quantidade de Frases de Surpresa que Continham a Palavra Comprar



Resultado da verificação da palavra comprar nas frases de Tristeza

Quantidade de Frases de Tristeza que Continham a Palavra Comprar



Conclusão

Conclusão

- ▶ Existem poucos trabalhos específicos de mineração de textos objetivando encontrar emoções em bases textuais, proporcionando um novo leque de pesquisas.
- ▶ Cenários onde se pode aplicar: sumarização de opiniões, termômetros de popularidade, ações de marketing ou relações públicas e previsão de indicadores a partir de sentimentos.
- ▶ Necessidade de melhoras na lista de *stopwords* da biblioteca NLTK e na técnica utilizada para limpeza dos dados.

Conclusão

- ▶ Mineração de Emoções podendo ser realizada com um baixo custo financeiro, caso venha a necessitar contratar um profissional da área de Neurolinguística pode ser um custo elevado. Sugere-se que tente achar uma pessoa que esteja interessado na área e queira contribuir, assim não terá o gasto com a contratação.

Trabalhos futuros

- ▶ Criação de um algoritmo para identificar os textos da base classificada que foram classificados erroneamente.
- ▶ Obtenção de uma base maior já classificada por um profissional da área de Neurolinguística.

Principais referências

ARANHA, C. N; VELLASCO, M. ***Uma abordagem de pré-processamento automático para mineração de textos em português: sob o enfoque da inteligência computacional.*** Pontifícia Universidade Católica do Rio de Janeiro, 2007.

LIU, B. ***Sentiment analysis and opinion mining, Morgan and Claypool Publishers.*** 2012.

GONÇALVES, T; et al. ***Analysing part-of-speech for portuguese text classification. In: Computational Linguistics and Intelligent Text Processing. Springer, 2006, p. 551-562.***

Obrigado !