

**INSTITUTO ENSINAR BRASIL
FACULDADE DOCTUM CARATINGA**

GUSTAVO DE CASTRO MINHANELI

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

CARATINGA

2018

GUSTAVO DE CASTRO MINHANELI

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

Monografia apresentada ao Curso de Ciência da Computação das Faculdades Integradas de Caratinga, como requisito parcial à obtenção do título de Bacharel em Ciência da Computação.

Área de Concentração: Mineração de dados textuais.

Orientador (a): Maicon Vinícius Ribeiro.

CARATINGA

2018

TERMO DE APROVAÇÃO

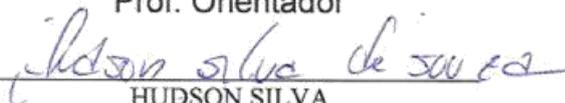
O Trabalho de Conclusão de Curso intitulado: ANÁLISE DE DADOS RECONHECENDO EMOÇÕES EM TEXTOS COM OBJETIVO DE OBTER INFORMAÇÕES ÚTEIS PARA VENDAS UTILIZANDO AS FERRAMENTAS NAIVE BAYES E NLTK COM PYTHON, elaborado pelo(s) aluno(s) GUSTAVO CASTRO MINHANELI foi aprovado por todos os membros da Banca Examinadora e aceito pelo curso de CIÊNCIA DA COMPUTAÇÃO das FACULDADES DOCTUM DE CARATINGA, como requisito parcial da obtenção do título de

BACHAREL EM CIÊNCIA DA COMPUTAÇÃO.

Caratinga 18/12/2018



MAICON RIBEIRO
Prof. Orientador



HUDSON SILVA
Prof. Avaliador 1



ELIAS DE SOUZA GONÇALVES
Prof. Examinador 2

AGRADECIMENTOS

Embora não seja uma tarefa fácil, é de extrema importância que alguns agradecimentos sejam feitos. Durante o desenvolvimento deste trabalho, muitas pessoas passaram pela minha vida e contribuíram, de maneira garantida, para que o mesmo tivesse bons resultados e se tornasse algo digno de ser mencionado e citado para outras pesquisas. Tentarei, então, agradecer a todos que me auxiliaram, procurando não me esquecer de ninguém.

Agradeço, primeiramente, a Deus por ter me dado saúde e motivação para que eu pudesse chegar até aqui. Agradeço também a toda a minha família, especialmente meus pais, por acreditarem em mim, pela dedicação e apoio de sempre.

Ao meu orientador, Maicon Vinícius Ribeiro: obrigado por acreditar em mim como seu aluno. Com sua orientação, paciência e empenho consegui analisar este trabalho com êxito.

Por fim, agradeço a todos que estão comigo desde o início da faculdade, por todos os momentos que passamos e pelos conhecimentos compartilhados durante a graduação.

“Se atribuirmos algum significado especial a um dado, este se transforma em uma informação (fato). Se os especialistas elaboraram uma norma (regra), a interpretação do confronto entre o fato e a regra constitui um conhecimento.”

Alberto Sulaiman Sade

RESUMO

Mineração de textos, conhecida como mineração de dados textuais ou descoberta de conhecimento de bases de textuais, em geral, se refere ao processo de extração de informações de interesse e padrões não triviais ou descoberta de conhecimento em documentos de texto não estruturados. Tendo em vista que os meios de difusão de informação e opinião contém vários dados não estruturados alimentados por pessoas que as utilizam, percebeu-se que, são expressas emoções nas publicações. As emoções fazem parte da vida do ser humano, são demonstradas e percebidas, através de gestos e expressões, utilizando a fala ou escrita. As redes sociais digitais são cada vez mais populares, por isso muita informação útil sobre produtos e serviços podem ser extraídas a partir de opiniões manifestadas nestes ambientes virtuais. A análise de emoções em redes sociais vem sendo explorada em diferentes pesquisas, tendo como principal intuito extrair opiniões dos usuários sobre os mais diversos assuntos, possibilitando que informações significativas sejam obtidas. Fundamentado a respeito do tema, fez-se coleta de publicações textuais feitas por usuários na rede social Twitter, estimulado a reconhecer emoções de usuários, de modo a identificar o quão provável o mesmo tem de interesse em relação a um determinado produto ou serviço. No processo de descoberta de conhecimento, foi utilizado o algoritmo de *Naive Bayes*, contido como função na biblioteca NLTK que foi utilizada integralmente com a linguagem de programação Python, buscando identificar emoções nos textos com relação interesse de compra.

Palavras-chave: Mineração de Textos.hj PNL. *Naive Bayes*. Descoberta de Conhecimento.

ABSTRACT

Text mining, known as textual data mining or knowledge discovery of textual bases, generally refers to the process of extracting information of interest and non-trivial patterns or knowledge discovery in unstructured text documents. Given that the means of diffusion of information and opinion contains several unstructured data fed by people who use them, it has been noticed that emotions are expressed in the publications. Emotions are part of the life of the human being, are expressed and perceived, through gestures and expressions, using speech or writing. Digital social networks are becoming more popular, so much useful information about products and services can be drawn from the opinions expressed in these virtual environments. The analysis of emotions in social networks has been explored in different researches, with the main purpose of extracting users' opinions on the most diverse subjects, allowing significant information to be obtained. Based on the theme, it was collected textual publications made by users on the social network Twitter, stimulated to recognize emotions of users, in order to identify how likely the same is of interest in relation to a particular product or service. In the process of knowledge discovery, the Naive Bayes algorithm was used as a function in the NLTK library, which was used in full with the Python programming language, trying to identify emotions in the texts with relation to purchase interest.

Keywords: *Text Mining. NLP. Naive Bayes. Knowledge Discovery.*

LISTA DE ABREVIATURAS E SIGLAS

AS - Análise de Sentimentos

IHC - Interação Humano Computador

KDD - *Knowledge Discovery in Databases*

KDT - *Knowledge Discovery from Text*

PLN - Processamento de Linguagem Natural

PNL - *Natural Language Processing*

LISTA DE ILUSTRAÇÕES

Figura 1 - O ciclo do processo de KDD	15
Figura 2 - Processo de Descoberta de Conhecimento em Textos	16
Figura 3 - Etapas do Processo de Indexação Automática.....	20
Figura 4 - Expressões faciais típicas de seis emoções básicas: a. alegria, b. medo, c. surpresa, d. tristeza, e. desgosto, f. raiva.....	22
Figura 5 - Página inicial do site oficial da linguagem python.....	26
Figura 6 - Lista de stopwords em português	28
Figura 7 - Eliminação de stopwords	29
Figura 8 - Teorema de Bayes.....	30
Figura 9 - Calculo para classificar se um tweet é positivo	31
Figura 10 - Chaves e tokens do Twitter para desenvolvedores.....	34
Código 1 - Chaves e tokens do Twitter para desenvolvedores	35
Código 2 - Parâmetro para coleta de textos do Twitter	35
Código 3 - Coleta e Armazenamento de posts no arquivo lista_de_dados_coletados.txt	35
Figura 11 - Fluxo de pré-processamento	37
Quadro 1 - Mapeamento abreviações	38
Gráfico 1 - Base textual classificada por profissional de Neurolinguística.....	39
Código 4 - Inserção da base de treinamento com os dados do Anexo A.....	40
Código 5 - Aplicação de Remoção de Radical e Palavras Irrelevantes.....	40
Figura 12 - Processamento de palavras das frases do arquivo de textos com postagens dos usuários	41
Quadro 2 - Amostra dos textos coletados na rede social Twitter.	42
Código 6 - Algoritmo de comparação de frases coletadas com frases já classificadas pelo profissional de Neurilinguística.	43
Código 7 - Algoritmo de classificação das frases	44
Quadro 3 - Resultado da Coleta de Dados Brutos	45
Gráfico 2 - Quantidade de Frases Classificadas	47
Gráfico 3 - Quantidade de Frases de Alegria que Continham a Palavra Comprar ...	48
Gráfico 4 - Quantidade de Frases de Desgosto que Continham a Palavra Comprar	49
Gráfico 5 - Quantidade de Frases de Medo que Continham a Palavra Comprar	50
Gráfico 6 - Quantidade de Frases de Raiva que Continham a Palavra Comprar	50
Gráfico 7 - Quantidade de Frases de Surpresa que Continham a Palavra Comprar	51
Gráfico 8 - Quantidade de Frases de Tristeza que Continham a Palavra Comprar .	52
Gráfico 9 - Classificação Geral de Frases	53

SUMÁRIO

1 INTRODUÇÃO	11
2 REFERENCIAL TEÓRICO	13
2.1 Tarefas E Técnicas Da Mineração	13
2.1.1 Mineração de Dados	13
2.1.1.1 Definição	13
2.1.1.2 Algoritmos de Mineração de Dados	14
2.1.1.3 Descoberta de Conhecimento em Base de Dados	14
2.1.2 Etapas da Mineração de Textos	15
2.1.2.1 Coleta	16
2.1.2.2 Pré-Processamentos de Textos	17
2.1.2.2.1 Processamento de Linguagem Natural	17
2.1.2.3 Indexação	19
2.1.2.4 Mineração de Textos	20
2.1.2.5 Análise	21
2.2 Emoções Básicas	21
2.2.1 Alegria	23
2.2.2 Desgosto	23
2.2.3 Medo	23
2.2.4 Raiva	23
2.2.5 Surpresa	24
2.2.6 Tristeza	24
2.3 Análise De Sentimentos	24
2.4 Ferramentas	25
2.4.1 Linguagem Python	25
2.4.1.1 História	26
2.4.1.2 Definição	27
2.4.2 Natural Language Toolkit	27
2.4.2.1 Stopwords	28
2.4.2.2 Stemming	29
2.4.2.3 Naive Bayes	30
2.5 TWITTER	31
3 METODOLOGIA	33
3.1 Criação Do Algoritmo De Coleta De Dados	33
3.1.1 Coleta dos Dados	35
3.1.1.1 Dificuldades na Estrutura dos dados Textuais	36
3.1.1.2 Preparação dos textos	36
3.1.2 Base Textual já Classificada	39
3.1.3 Algoritmo De Pré-Processamento E Classificação	39
3.1.4 Base de dados textuais coletada	42
3.2 Seleção De Algoritmos	42

3.2.1 Algoritmo de Comparação de Base Textuais	43
3.2.2 Algoritmo de Classificação de Frases Coletadas	43
4 RESULTADOS	45
4.1 Coleta De Dados	45
4.2 Pré-Processamento E Classificação	46
4.3 Análise Dos Dados	52
5 CONCLUSÃO	54
6 TRABALHOS FUTUROS	56
REFERÊNCIAS	57
APÊNDICE A — Algoritmo de Coleta de Dados Textuais	66
APÊNDICE B — Algoritmo de Pré-Processamento	67
APÊNDICE C — Algoritmo de Avaliação da Precisão	69
APÊNDICE D — Algoritmo de Comparação de Frases em Arquivos	71
ANEXO A — Base de Treinamento do Algoritmo	72

1 INTRODUÇÃO

A Mineração de textos é caracterizada como Mineração de Dados, que é absorção de informações desconhecidas através dos textos escritos em linguagem natural coletada de uma rede social, ou qualquer plataforma que contém dados, os quais possam ser coletados e ser realizado uma análise de modo a obter informações úteis particularmente ou para empresa em específico, que procura uma determinada informação para tomada de decisões.

As informações normalmente são armazenadas em formato de texto, mineração tem valor comercial alto e pode ser aplicada em diversas áreas, neste trabalho empregou-se a Mineração de Textos para obter informações relevantes para vendas de produtos ou serviços.

O número de trabalhos relacionados a completude do tema é ainda bem limitado, principalmente devido à complexidade de se trabalhar com texto, especialmente provindos de redes sociais, que é desestruturado e o processo para fazer o tratamento do mesmo exige um custo de tempo alto (SILVA, 2016). Ainda assim, com o uso intensivo das redes sociais, a análise de dados para esse meio vem ganhando cada vez mais importância (FRANÇA et al., 2014).

Há diversas origens de grandes volumes de dados presentemente. LEONEL JUNIOR et al. (2014) destaca algumas fontes, que juntas, produzem dados que alcançam a escala de *petabytes* diários. São elas: redes sociais, sistemas corporativos, serviços e sistemas web, transações financeiras, *e-commerce*, entre outros. Dentre as inúmeras fontes de dados, esse trabalho possui ênfase em dados provenientes da rede social *Twitter*.

Com este grande volume de dados gerados diariamente, percebeu-se que ao analisar essas informações, as organizações poderiam ter a vantagem de conhecer as opiniões dos usuários sobre seus serviços ou produtos a partir de dados das redes sociais (GOMES, 2013).

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. Através das mesmas, estão os gestos e expressões, demonstrados através de comunicação falada verbalmente ou descrita com palavras. Dados textuais de mídias sociais

costumam conter opiniões e sentimentos que são possíveis de ser identificadas e detectadas utilizando técnicas de mineração (GONÇALVES et al., 2015).

Análise de Sentimentos (AS) é o campo que analisa opiniões, atitudes, sentimentos, avaliações e emoções das pessoas em diversas mídias, dentre elas os textos, de forma computacional. Nesse contexto surgiu esta área, também conhecida como Mineração de Opinião, que estuda as opiniões e emoções expressas em textos através das técnicas da inteligência artificial, e na identificação automática da polaridade de opiniões (LIU, 2012). A AS é uma área que envolve Psicologia e Mineração de Textos, percorrendo através do reconhecimento de padrões chegando até a Interação Humano Computador.

O termo Interação Humano Computador refere-se a usabilidade e comunicabilidade, facilita a vida dos usuários, expondo informações de maneira clara utilizando técnicas de interação para auxílio na utilização dos usuários (NETTO, 2004).

A concepção foi caracterizar frases emocionalmente, para estimar em qual classificação emocional básica cada frase refere-se, auxiliando na tomada de decisão, seja coletivo ou individual.

Este trabalho está estruturado da seguinte maneira: Na seção 2, são apresentados conceitos relacionados sobre processo de Mineração de Textos, ferramentas utilizadas, abordagens para explicar sobre emoções básicas. Na seção 3, é mostrado a metodologia do trabalho em questão, apresentando as etapas e o modelo de aplicação da técnica proposta. Já na seção 4, resultados experimentais são apresentados e discutidos. Conclusões são descritas na Seção 5.

2 REFERENCIAL TEÓRICO

Os tópicos a seguir são base para a realização desse trabalho e contextualizam os principais conceitos que estão envolvidos. A fundamentação teórica, que explana as etapas de Mineração de Dados textuais, emoções e ferramentas utilizadas, foi necessária para a subsequência dos temas, que consistem na explicação do conceito e dos tipos de recuperação de informação.

2.1 TAREFAS E TÉCNICAS DA MINERAÇÃO

Os objetivos a serem alcançados são o fator responsável pela definição da escolha das tarefas a serem utilizadas na mineração de dados (GARCIA, 2012). Não há uma definição genérica de mineração que seja melhor em qualquer situação, cada situação é diferente de todas as outras.

Após a escolha do objetivo, define-se a técnica ou tarefa de mineração. Tarefa se diferencia de técnica de mineração pelo fato de que, a tarefa utilizada define qual a informação ou padrão deseja-se encontrar nos dados, e a técnica específica os métodos que foram aplicadas para alcançar os objetivos desejados (WITTEN; FRANK; HALL, 2011).

2.1.1 Mineração de Dados

Esta seção objetiva fornecer a visão de como surgiu Mineração de Dados, estruturas de dados e procedimentos utilizados para obter conhecimento em textos.

2.1.1.1 Definição

PINHEIRO (2008, p.19) define que “*Data Mining* ou Mineração de Dados, é o processo de descoberta de padrões e tendências existentes em repositórios”. Já TAN, STEINBACH e KUMAR (2009) em outras palavras, dizem que a Mineração de Dados é o processo de descoberta automática de informações úteis em grandes depósitos de dados.

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.

O processo de Mineração de Dados não é apenas um processo computacional ou automático, mas sim um esforço cooperativo entre pessoas e máquinas. As pessoas conseguem descrever e desenhar os bancos de dados, os problemas que devem ser solucionados para os objetivos serem alcançados (PINHEIRO, 2008).

Os padrões descobertos devem ser significantes na medida em que levam alguma vantagem, normalmente vantagem econômica (WITTEN; FRANK, 2005). Esses padrões podem ser expressos de maneira que sejam úteis e nos permitam realizar previsões não triviais gerando novos dados (NASCIO, 2004).

Utilizando mineração, dados são transformados em informações, mas precisa-se escolher qual algoritmo e técnica utilizada para chegar ao objetivo desejado. CASTANHEIRA (2008) enfatiza que a Mineração de Dados se caracteriza pela existência de um algoritmo, que diante da tarefa proposta foi eficiente em extrair o conhecimento implícito e útil de um banco de dados.

2.1.1.2 Algoritmos de Mineração de Dados

Existem diversos algoritmos para minerar textos, no entanto, não existe uma única solução para todos os problemas de inteligência artificial como também nem sempre uma única técnica será a resposta para todos os problemas de aprendizado, classificação e representação do conhecimento (WITTEN; FRANK; HALL, 2011). Na vasta variedade de conjuntos de dados existentes encontram-se diversos tipos de estruturas, e uma ferramenta de Mineração de Dados pode num ambiente demonstrar bons resultados, mas em outro ser totalmente inadequada (LAHIRI, 2006). Sendo assim, convém iniciar da parte mais básica, utilizando aplicações os métodos fundamentais e deixar as opções mais trabalhadas como opções para último caso.

2.1.1.3 Descoberta de Conhecimento em Base de Dados

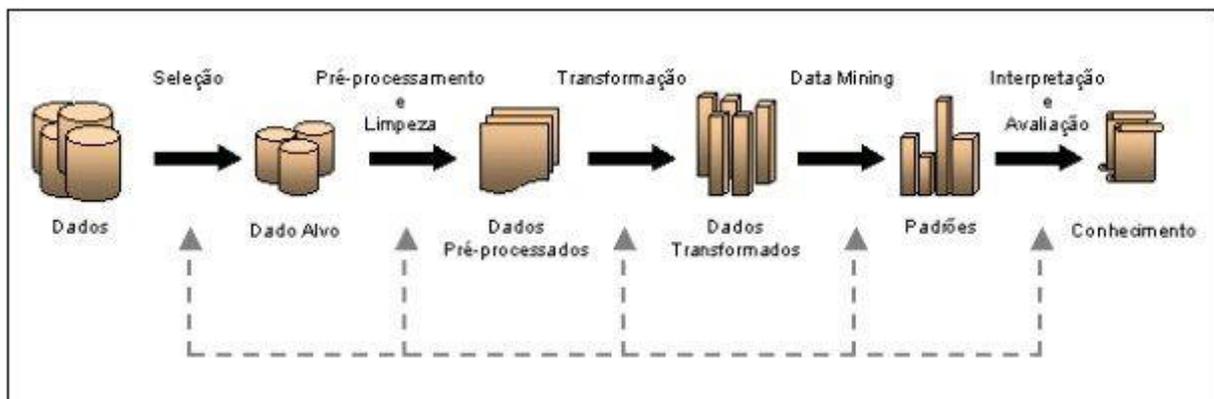
Neste capítulo são discutidos os principais conceitos envolvidos no processo de descoberta de conhecimento em bases de dados.

Segundo ELMASRI e NAVATHE (2011), o *KDD* é composto por cinco fases: seleção de dados, limpeza de dados, transformação ou codificação de dados, Mineração de Dados e o relatório e exibição da informação descoberta.

KDD (*Knowledge Discovery in Databases*) é um processo de descoberta de conteúdo em bases de dados que tem como objetivo principal extrair informações. Para isto envolve diversas áreas de aquisição do conhecimento, tais como: estatística, matemática, bancos de dados, inteligência artificial, visualização de dados e reconhecimento de padrões. São utilizadas também técnicas, em seus diversos algoritmos, oriundos dessas áreas, segundo (CASTANHEIRA, 2008).

Para se iniciar uma descoberta de conhecimento na base de dados, é necessário ter o domínio da aplicação e dos objetivos que se deseja alcançar. Esse processo é composto por cinco etapas, conforme pode ser visto na Figura 1.

Figura 1 - O ciclo do processo de KDD



Fonte: (Fayyad, Piatetsky-Shapiro e Smyth, 1996)

A Figura 1 demonstra uma fonte de dados e o processo de Mineração de textos, começando da esquerda para a direita até a obtenção de informação através dos dados coletados. Passa-se por diversas etapas, como pré-processamento dos dados para obter somente os úteis para a análise e reconhecer padrões de comportamentos, até que pudesse manipular os mesmos para realizar uma classificação utilizando um algoritmo a fim de gerar conhecimento.

2.1.2 Etapas da Mineração de Textos

Este tópico demonstra a metodologia proposta para Mineração de Textos por (ARANHA; VELLASCO, 2007). Os autores descrevem como sendo um modelo completo para adquirir conhecimentos a partir do corpo do texto. Esse tópico, visa

detalhar etapas e técnicas dessa metodologia, onde este processo é o que melhor se enquadra no presente trabalho. A Figura 2 descreve o método.

Figura 2 - Processo de Descoberta de Conhecimento em Textos



Fonte: Adaptado de ARANHA e VELLASCO (2007)

A metodologia aplica-se em cinco etapas distintas, a primeira extração dos dados, a segunda no pré-processamento com o objetivo de remoção de palavras irrelevantes contidas nos dados coletados, a terceira etapa confere os textos após o pré-processamento dos dados, a quarta na obtenção de conhecimento e, por fim, uma quinta etapa para interpretação dos resultados obtidos (ARANHA; VELLASCO, 2007).

O processo de extração e organização de conhecimento através da Mineração de Textos pode ser dividido em três fases principais: pré-processamento dos documentos, extração de padrões com agrupamentos de textos e avaliação do conhecimento (REZENDE; MARCACINI; MOURA, 2011).

2.1.2.1 Coleta

A etapa de coleta é importante, pois é responsável pela extração dos textos que servirão de base para a análise (SANTOS, 2016). Neste período são selecionados os textos conforme o domínio do problema. Considerando os objetivos que se deseja alcançar com o processo, o primeiro passo é eleger o conjunto de textos utilizados (CECI, 2010).

A coleta dos dados é um processo que pode ser custoso e burocrático, podendo ser realizada de diferentes formas, dependendo da natureza dos dados (MELO, 2010).

2.1.2.2 *Pré-Processamentos de Textos*

O principal objetivo de pré-processar um texto, consiste na filtragem e limpeza dos textos, eliminando redundâncias e informações desnecessárias para o conhecimento que se deseja extrair (GONÇALVES; ET AL, 2006).

Contudo, realizar a limpeza das palavras que possuem pouca importância para a classificação do texto e facilitou a obtenção das palavras de característica do texto (BRAGA, 2017).

O pré-processamento é útil para identificar grupos similares em um conjunto de texto. Ainda, devem-se construir representações desse conjunto, para tornar o processamento da mineração mais assertivo (ALMEIDA, 2007).

Pré-processar textos, consiste em uma das fases mais críticas do processo de Mineração de Textos. Nesse sentido, a principal contribuição deste trabalho foi a proposta de uma metodologia de pré-processamento textual, utilizando do mesmo para remoção de palavras contidas nos textos coletados que não são úteis para análise (FELDMAN; SANGER, 2006).

2.1.2.2.1 *Processamento de Linguagem Natural*

O PNL é uma área de pesquisa que surgiu em meados de 1950 e é uma das áreas de ciência da computação, inteligência artificial e linguística cuja principal preocupação é estabelecer a ligação entre computadores e a linguagem natural falada pelas pessoas. Proveniente de um ramo da inteligência artificial, atualmente, o Processamento de Linguagem Natural assume-se como uma área de pesquisa independente. Apesar disto, em problemas mais práticos continua a ser complicado aplicar o processamento de linguagem natural para a sua resolução. Existem ainda assim algumas técnicas, algoritmos e recursos linguísticos que podem ser bastantes úteis para resolver problemas em específico (D E LISBOA, 2013).

O PLN é uma área de Ciência da Computação que estuda como os computadores podem analisar e/ou gerar textos em linguagem natural (PERNA; DELGADO; FINATTO, 2010).

TURBAN e et al (2010) descreve que o processamento da linguagem natural pode ser visto como a forma de comunicação entre o homem e a máquina, sendo essa comunicação em qualquer linguagem que se fale. Os autores ainda dizem que:

Para entender uma consulta em linguagem natural, o computador precisa ter conhecimento para analisar e interpretar a entrada de informação. Isso pode significar conhecimento linguístico de palavras, conhecimento sobre áreas específicas, conhecimentos gerais e até mesmo conhecimento sobre os usuários e seus objetivos. No momento em que o computador entende a informação, ele pode agir da forma desejada (TURBAN; ET AL, 2010, p.168).

LOPES (2002) diz que o PLN não é uma tarefa trivial devida a natureza ambígua da linguagem natural. Essa diversidade faz com que o PLN se diferencie do processamento das linguagens de programação de computador, as quais são fortemente definidas para evitar a ambiguidade.

Análise de sentimento é um problema de processamento de linguagem natural (PLN). Está em todos os aspectos da PLN, resolução de correferência, lidar com negação e desambiguação de palavras, o que aumenta a dificuldade da análise pois esses problemas ainda não foram solucionados (LIU, 2012).

O Processamento da Linguagem Natural é um conjunto de técnicas computacionais para analisar e representar ocorrências naturais de texto em um ou mais níveis de análise linguística com o objetivo de se alcançar um processamento de linguagem similar ao humano para uma série de tarefas ou aplicações (LIDDY, 2001, p.57).

Liddy (2001) demonstra que para processar a linguagem natural, pode-se processar os textos em diversos níveis, tais como léxicas, morfológica, semântico, dentre outras.

A Análise Léxica é a etapa para converter uma sequência de caracteres em um seguimento de palavras que foram as palavras candidatas a serem termos do índice. O analisador léxico separa o alfabeto de entrada em caracteres de palavras (a-z) e separadores de palavras (BARION; LAGO, 2008).

Na etapa da análise léxica, ocorre a conversão de uma cadeia de caracteres (o texto da consulta) em uma cadeia de palavras. Assim, o principal objetivo desta etapa é a identificação das palavras que constituem a consulta. Na fase seguinte, artigos, preposições e conjunções são candidatos naturais à lista de *stopwords*, ou seja, a serem eliminados. Então, pode ser executada a normalização lexical através de *stemming*, pois frequentemente o usuário especifica uma palavra na consulta, mas somente variações desta palavra

estão presentes em um documento relevante (GONZALEZ; LIMA, 2003, p. 5).

A utilização de várias representações de conhecimento, tais como, um léxico de palavras, seus significados e propriedades, e um conjunto de regras gramaticais, além de muitas vezes outros recursos, como uma ontologia de entidades e ações, ou até mesmo, uma enciclopédia de sinônimos ou abreviações (KAO; POTEET, 2007).

A análise morfológica é responsável por definir artigos, substantivos, verbos e adjetivos, armazenados em um tipo de dicionário. Depois de construído o dicionário, a análise sintática faz uso dele procurando mostrar relacionamento entre as palavras e, num segundo momento, verifica sujeito, predicado, complementos nominais e verbais, adjuntos e apostos. Na análise semântica, ocorre o encontro de termos ambíguos, de sufixos e afixos, entretanto, em questões de significado associados aos morfemas componentes de uma palavra, o sentido real da frase ou palavra (SANTOS et al., 2014).

A análise semântica procura identificar a função que determinados termos exercem no texto (CORDEIRO, 2005). Segundo LOPES (2002), stop-words são palavras que não agregam valor semântico dentro do contexto da base de dados textual.

2.1.2.3 *Indexação*

Nesta etapa, são identificadas as similaridades entre os termos coletados. A indexação tem como objetivo identificar a similaridade de significado entre termos, considerando variações morfológicas e a sinonímia, tendo como resultado um índice para os dados coletados (MORAIS; AMBRÓSIO, 2007).

Com tal crescimento, surgiu a necessidade de ferramentas automáticas ou semiautomáticas de indexação para não só garantir resultados mais eficazes, mas também para facilitar as tarefas envolvidas (NARUKAWA; LEIVA; FUJITA, 2009).

Na Figura 6 demonstra-se o processo de indexação, com suas principais fases: a identificação dos termos, a remoção de stopwords e a normalização morfológica.

Figura 3 - Etapas do Processo de Indexação Automática



Fonte: WIVES (2002)

A qualidade da estratégia de busca em um texto indexado e o vocabulário são fatores importantes para a atividade (LANCASTER, 1998).

As *stopwords* devem ser removidas da coleção textual, pois são palavras comuns e sem significado para o resultado final. Após a remoção das *stopwords*, era possível ter uma base de dados melhor estruturada, o que torna as tarefas de mineração de dados mais ágil, segundo (IBRAHIM, 2006).

2.1.2.4 Mineração de Textos

A Mineração de Textos (*Text Mining*) é um conjunto de técnicas e métodos utilizados para extrair conhecimento de dados não estruturados (TICOM, 2007). As técnicas utilizadas para minerar conseguem manipular documentos em formato textual, objetivando a aquisição do conhecimento implícito presente nos textos (ARANHA; PASSOS, 2006).

Mineração de Textos pode ser definido como um processo intensivo de conhecimento em que o usuário interage com uma coleção de documentos usando ferramentas de análise com o objetivo de extrair conhecimento pela identificação e exploração de padrões, tendências e conexões (FELDMAN; SANGER, 2006).

Mineração de texto tem sido comumente usados em sistemas para recuperação de informações baseadas no contexto de pergunta e resposta, que ajudam o usuário a explorar conteúdos relacionados por tópicos (GUPTA; LEHAL, 2009).

Na coleta de textos, são reunidos documentos que formam o conjunto de textos em meio eletrônico de uma determinada área, chamado de corpus textual (BUBENHOFER; PTASHNYK; KILIAN, 2012).

2.1.2.5 *Análise*

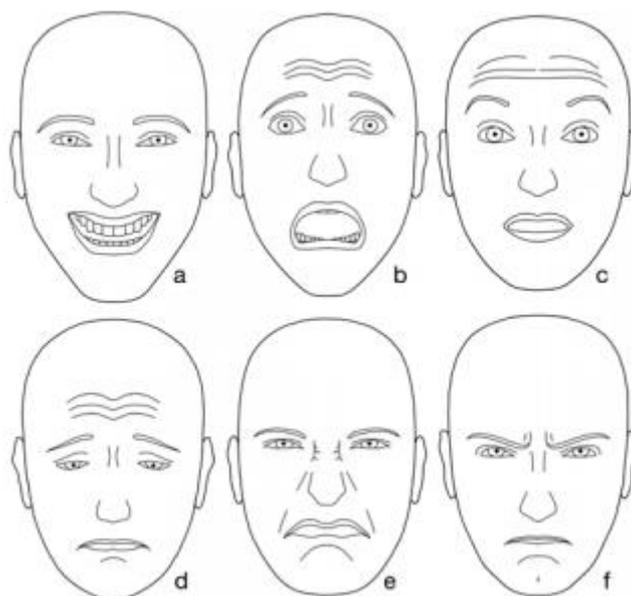
A etapa da análise realiza a validação da eficiência do processo como um todo, analisando os dados obtidos após aplicação dos algoritmos na etapa anterior (BRITO, 2016). Também, é uma situação que trata o caso de sinonímia de palavras, onde apesar de constituírem morfologias diferentes, as palavras expressam o mesmo conceito (LOPES, 2004).

Em resumo, trata-se de um processo de análise de informação que procura reduzir o espaço de busca recuperando apenas as informações que são úteis para a resolução de problemas específicos (TICOM, 2007).

2.2 Emoções Básicas

Faz-se comum encontrar na literatura a nomenclatura emoções básicas. Contudo, não há conformidade referente ao modelo que explica o funcionamento emocional, também não existe uma definição de quantas e quais são as emoções básicas. Entretanto, a maioria dos autores costuma citar as seguintes ou alguma variação delas: alegria, desgosto, medo, raiva, surpresa, tristeza. No entanto, alguns trabalhos relatam que o uso de emoções melhora significativamente os resultados da Mineração de Opiniões (BOLLEN; MAO; ZENG, 2011). Miguel (2015) demonstra na Figura 4 as expressões faciais de cada emoção básica.

Figura 4 - Expressões faciais típicas de seis emoções básicas: a. alegria, b. medo, c. surpresa, d. tristeza, e. desgosto, f. raiva



Fonte: Miguel (2015)

Tendo em vista que as emoções básicas como alegria, desgosto, medo, raiva, surpresa, tristeza, estão intrínsecas neste processo aprendido. Não é analisado o que o conteúdo dos textos expressa, mas sim a emoção predominante no mesmo, se é positiva ou negativa (RODRIGUES et al., 2011).

As emoções têm uma orientação centrífuga e uma acentuada expressividade (por exemplo: alegria, raiva, medo); têm uma forte influência sobre o sistema neurovegetativo. Emoções violentas de raiva ou medo, repetidas com frequência, induzem transtornos psicossomáticos; emoções que não se expressam acumulam-se nos órgãos (TORO, 2002).

As emoções básicas ocorrem a partir de eventos externos ao indivíduo e são rapidamente compartilhadas entre membros de um mesmo grupo (SANTOS; GRAMINHA, 2006).

As emoções de fundo são aquelas em que o sujeito tem a capacidade de decodificá-las rapidamente em diferentes contextos, sendo elas agradáveis ou desagradáveis (DAMÁSIO, 2004).

2.2.1 Alegria

A alegria é uma emoção positiva que provoca bem-estar e satisfação. Esta emoção básica conduz a sentimentos positivos e potência a inibição de pensamentos negativos (FREITAS, 2011a).

Galati, Manzano e Sotgiu (2006) afirmam que, o significado de alegria atualmente tornou-se muito próximo de satisfação com a vida e bem-estar subjetivo, sendo que esses termos têm sido frequentemente utilizados para descrever o mesmo tipo de fenômeno.

2.2.2 Desgosto

A emoção desgosto ocorre em ocasiões desagradáveis e aversivas e é, claramente, uma emoção negativa (FREITAS, 2011b).

Goleman (2012) descreve que na expressão facial de desgosto, o lábio superior é repuxado para os lados e dá-se também um ligeiro franzimento do nariz, com objetivo de tapar as narinas contra um odor ofensivo ou cuspir um alimento venenoso.

2.2.3 Medo

É a emoção do perigo. Esta emoção causa um forte impacto fisiológico, o coração começa a bater mais forte, a respiração acelera, os músculos contraem, as mãos tremem. Todas estas sensações derivam do sistema nervoso simpático, adrenalina e noradrenalina que agem sobre o nosso corpo quando temos medo (CASANOVA; SEQUEIRA; SILVA, 2009).

Esta emoção é despertada por um aumento de intensidade bastante rápido na estimulação neuronal consequente a um perigo real ou imaginário (EKMAN, 2003).

2.2.4 Raiva

A raiva é a única emoção que prepara para a defesa, proporcionando energias necessárias para o efeito (STRONGMAN, 1998). Esta é uma emoção à qual e associam, também, a revolta, a hostilidade, a irritabilidade, o ressentimento, a indignação, o ódio e a violência (FREITAS, 2011b).

Raiva faz-se acompanhar por reações físicas como o aumento da adrenalina e da noradrenalina no fluxo sanguíneo, aumento da pressão arterial e ritmo cardíaco, contração muscular e aceleração da respiração, entre outras (FREITAS; CASTRO, 2007). Ekman (2003) defende que esta emoção é a mais perigosa, uma vez que conduz à tendência para prejudicar o alvo da que a levou a ter esta emoção.

2.2.5 Surpresa

A surpresa é uma das emoções mais breves, durando apenas alguns segundos. Uma vez que se compreende o que está acontecendo, surpresa tende a se combinar com outra emoção, positiva ou negativa, dependendo do evento eliciador, ou não ser seguida por nenhuma outra, caso não seja avaliado como importante (MIGUEL, 2015).

Goleman (2012) diz que ao encurvar as sobrancelhas aumenta o campo visual e conseqüentemente há entrada de mais luz para a retina, deste modo, fornece ao indivíduo maior informação da ocasião inesperada, tornando então mais fácil a percepção do que houver para uma melhor preparação da tomada de decisão.

2.2.6 Tristeza

A tristeza é um sentimento humano que expressa desânimo ou frustração em relação a alguém ou algo. A tristeza pode causar reações físicas como depressão, choro e insônia (MATA, 2007).

Rangé (2003) pontua que a emoção tristeza provém do diagnóstico de uma doença grave e é equivalente a uma grande perda, como um objeto fonte de gratificação ou valorização.

Uma pessoa desesperançosa apresenta distorções cognitivas que a levam a acreditar que seus problemas atuais são irreversíveis e percebem sua vida como ausente de qualquer satisfação, predominando sentimento de culpa e tristeza (LOPES; SANTOS; LOPES, 2008).

2.3 Análise De Sentimentos

Informação textual pode ser dividida em dois grandes grupos: fatos e opiniões. Fatos são expressões objetivas sobre eventos e propriedades. Opiniões são, normalmente, expressões subjetivas que descrevem o sentimento das

peças, apreciações e sentimentos sobre entidades, eventos e seu pertencentes (LIU, 2010). No entanto, existe uma leve diferença de definição entre ambos, onde Mineração de Opiniões extrai e analisa as opiniões de pessoas sobre entidades enquanto a Análise de Sentimentos identifica sentimentos expressos em um texto (TSYTSARAU; PALPANAS, 2011).

A análise de sentimento vem sendo tema de diversos estudos, tanto no ambiente acadêmico como diz (DING; LIU; ZHANG, 2009), bem como na esfera da indústria, como afirma (ROSA, 2015).

Mukherjee (2012) afirma que a análise de sentimentos consiste em uma atividade que envolve Processamento de Linguagem Natural (PLN) e Extração de Informações e que tem como objetivo adquirir o sentimento das pessoas, os sentimentos que são expressos em comentários positivos ou negativos, perguntas e pedidos em documentos escritos através da análise de uma grande quantidade destes elementos. Morency, Rada e Payal (2011) têm uma visão diferente, os autores descrevem que a análise de sentimentos é uma área de estudos que foca na identificação automática de estados privados, como opiniões, emoções, sentimentos, avaliações, crenças e especulações na linguagem natural.

2.4 Ferramentas

Nesta seção foram apresentadas as ferramentas que tiveram mais importância diagnóstica de uma doença grave para o trabalho que são: Python, NLTK e Naive Bayes.

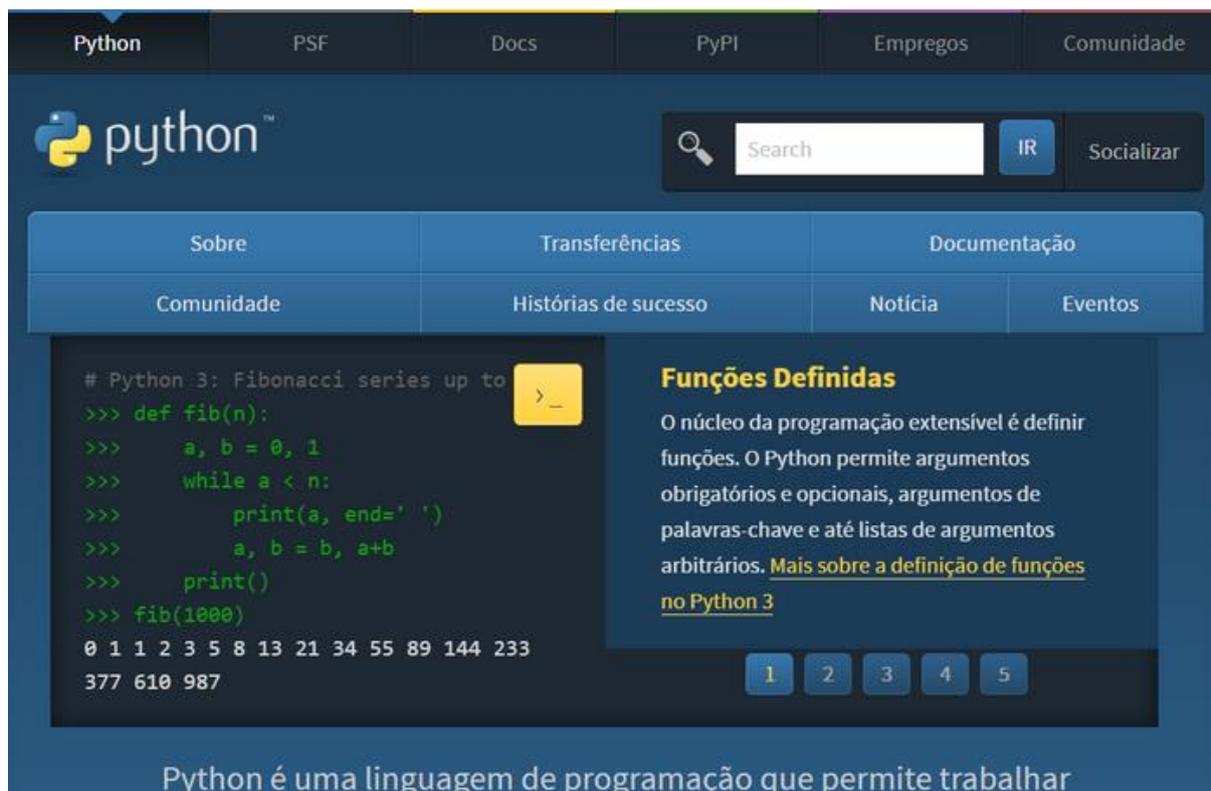
2.4.1 Linguagem *Python*

Nesta sessão é apresentado um pouco de conhecimento sobre a linguagem de programação utilizada para realizar algumas etapas deste trabalho.

O núcleo da programação extensível é definir funções. O *Python* permite recursos obrigatórios e opcionais, palavras-chave e até lista de argumentos arbitrários. O interpretador Python¹ é livre para download, cuja página inicial é apresentada na Figura 5, e existem versões para Linux, OS X e Windows.

¹Página Inicial do site da linguagem de programação *Python* <<https://www.python.org>>

Figura 5 - Página inicial do site oficial da linguagem python



Fonte: <https://www.python.org/>

No site oficial da Linguagem *Python* tem diversos conteúdos explicativos que demonstram como funciona a linguagem de programação na parte de Documentação, além de ter uma Comunidade onde outras pessoas compartilham ideias e ajudam quem expõe sua dúvida.

2.4.1.1 História

Com a característica marcante de um ambiente de desenvolvimento mais legível e limpo, pela falta de regras e sintaxes comparando-a com a linguagem C por exemplo, é muito utilizada como meio de inserção de pessoas no mundo da programação, onde a pequena quantidade de imposições de sintaxe ajuda o estudante a adquirir uma noção geral de como é programar, e estender esta noção para outras linguagens (BOGDANCHIKOV; ZHAPAROV; SULIYEV, 2013).

A linguagem de programação *Python*, criada em 1991, vêm conquistando espaço entre as outras ferramentas de programações atuais pela sua interface "amigável" de fácil aprendizado e ao mesmo tempo pela sua grande aplicabilidade e realização de tarefas tão complexas quanto o seu antecessor Java. O *Python*

apresenta-se como uma linguagem com um grande histórico de sucesso, quando o assunto é qual linguagem de programação introduzir a alunos que nunca tiveram contato com códigos-fonte e criação de softwares (MARQUES; COSTA, 2011).

O nome da linguagem origina-se do nome da série humorística britânica *Monty Python's Flying Circus* (SANTANA; GALES, 2010) e foi lançada em 1991 pelo holandês Guido Van Rossum, sendo hoje mantida pela organização sem fins lucrativos *Python Software Foundation* (PSF) sob licença livre.

2.4.1.2 Definição

Python é uma linguagem de altíssimo nível (em inglês, *Very High Level Language*) orientada a objeto, de tipagem dinâmica e forte, interpretada e interativa (BORGES, 2010). BORGES (2014) diz que, possui uma sintaxe clara e concisa, que favorece a legibilidade do código fonte, tornando a linguagem mais produtiva.

Python é uma linguagem ideal para programar. Como uma linguagem multiparadigma, ela proporciona muita flexibilidade na forma de estruturar programas. Ela pode ser usada como linguagem de script para simplesmente executar código-fonte, como linguagem de procedimentos para organizar programas como uma coleção de funções que se chamam ou como linguagem orientada a objeto para usar classes, herança e módulos para montar uma hierarquia. Essa flexibilidade permite definir o melhor estilo de programação para desenvolver um projeto particular (CARVALHO, BRANCO, 2016, p.23).

Com a característica marcante de um ambiente de desenvolvimento mais legível e limpo, pela falta de regras e sintaxes comparando-a com a linguagem C por exemplo, é muito utilizada como meio de inserção de pessoas no mundo da programação, onde a pequena quantidade de imposições de sintaxe ajuda ao estudante a adquirir uma noção geral de como é programar, e estender esta noção para outras linguagens (BOGDANCHIKOV; ZHAPAROV; SULIYEV, 2013).

2.4.2 Natural Language Toolkit

O pacote NLTK realiza o processamento de linguagem natural em *Python* "linguagem de programação de excelente funcionalidade para processamento de dados linguísticos" (BIRD; KLEIN; LOPER, 2009). NLTK foi concebido em 2001 como parte de um curso de Linguística Computacional no Departamento de Ciência da Computação e Informação da Universidade da Pensilvânia (BIRD; KLEIN; LOPER,

2009). Devido a seu caráter de software aberto e gratuito, tem sido desenvolvido e ampliado com a ajuda de dezenas de colaboradores, pelo seu uso e concepção de módulos de análise linguística (BIRD; KLEIN; LOPER, 2009).

(LOPER; BIRD, 2002) descrevem o NLTK 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).

2.4.2.1 Stopwords

O NLTK possui uma lista de stopwords para o Português, que são palavras que podem ser consideradas irrelevantes para o entendimento do sentido de um texto, ou seja, palavras semanticamente irrelevantes. Exemplos: as, e, os, de, para, com, sem, foi. Essas palavras são geralmente removidas de um texto durante a fase de pré-processamento.

A seguir, a lista de palavras que contém na stopwords.

Figura 6 - Lista de stopwords em português

De	A	O	Que	E	Do	da
um	Para	Com	não	Os	No	Se
ao	Das	À	as	Dos	Como	mas
já	Nos	Também	Só	Pelo	Pela	Até
Eu	Ela	Entre	Sem	Mesmos	Aos	Seus
depois	Nas	Me	esse	Você	Essa	Num
eles	Suas	Meu	Às	minha	Elas	Qual
numa	Lhe	Deles	Essas	Esses	Tivesse	Estivéssemos
pelos	Dele	Tu	Te	Vocês	Vos	Lhes
pelas	Minhas	Teus	Tuas	Nosso	Nossa	Nossos
este	Dela	Teu	tua	Delas	Esta	Estes
Aquele	Aquela	Aqueles	Aquelas	Isto	Aquilo	Estou
Estamos	Estão	Estive	Esteve	Estivemos	Estiveram	Estava
Estavam	Estivera	Estivéramos	Esteja	Estejamos	Estejam	Estava
Estivessem	Estiver	Estivemos	Estiverem	Haja	Hajamos	Hajam
Houvessem	Os	Houver	Sou	Houverem	Houverei	Houvera
Seja	Sejamos	Sejam	Fosse	Fóssemos	Fossem	For
Forem	Serei	Será	Seremos	Serão	Seria	Seríamos
Tenho	Tem	Temos	Tinha	Tenhamos	Tinham	Tive
Tivemos	Tiveram	Tivera	Tivéramos	Tiverem	terei	Terão
Teríamos	Teriam	Em	Na	Ele	Muito	Isso
Nem	Nós	Meus	Nossas	Estas	Está	Estávamos
Houve	Houvesse	Quem	Fôramos	Formos	Seriam	Teve
Teria						

Fonte: Adaptado de (ARANHA e VELLASCO, 2007)

Utilizando a lista da Figura 6, foram eliminados preposições, artigos, conjunções e alguns caracteres sem efeito, todos denotados como stopwords e retirados por sua alta frequência no texto, como no exemplo na Figura 7.

Figura 7 - Eliminação de stopwords

Na maioria das vezes os documentos retomados pelas ferramentas de recuperação de informações envolvem um contexto mais amplo fazendo com que o usuário tenha que garimpar ou seja especificar ou filtrar estes documentos e que demanda tempo e conhecimento a fim de obter a informação que ele realmente necessita...

Fonte: Adaptado de (ARANHA e VELLASCO, 2007)

6Usualmente no processo de remoção de *stopwords* é utilizada uma lista contendo todas as palavras que não devem ser indexadas. Tal lista é definida como *stoplist*, e as palavras contidas nelas são conhecidas como *stopwords* (CARRILHO JUNIOR, 2007).

SOARES (2008) diz que, assim que obtida a coleção de textos no qual realizou-se o processo de mineração, mesmo após a remoção de *stopwords* consideradas padrão, é interessante que seja analisada a frequência de ocorrência dos termos que estão presentes na coleção, pois termos com elevada frequência em muitos documentos constituem pouco conteúdo semântico, sendo, então, candidatos a *stopwords*.

A definição automática de *stopwords* através de um algoritmo que identifique as palavras que aparecem em grande quantidade nos textos (SANTOS, 2010).

2.4.2.2 Stemming

Stemming é uma técnica de redução de termos a um radical comum, a partir da análise das características gramaticais dos elementos, como grau, número, gênero e desinência (SILVA, 2008).

A utilização da função que realiza o *stemming* tem como objetivo reduzir as palavras agrupando-as por meio de seu sufixo. As palavras: acompanha, acompanhado, acompanhante seriam reduzidas ao radical da palavra "acompan", por exemplo. Esse processo é extremamente eficiente para a língua inglesa, por conta da simplicidade das regras gramaticais da língua, já para línguas como

espanhol ou português, o processo pode não ser muito preciso. Um exemplo pode ser observado com as palavras "barata" e "barato", ambas possuem dois valores semânticos diferentes, mas poderiam ser reduzidas ao radical "barat" (SILVEIRA, 2013). Esta redução permite a diminuição do custo computacional das etapas do processo (CORRÊA et al., 2012).

2.4.2.3 Naive Bayes

O Naive Bayes é um classificador que utiliza a probabilidade baseado na regra de Bayes. Ele é frequentemente utilizado como base na classificação de textos por ser rápido e fácil de implementar (RENNIE et al., 2003). GOMES (2013) diz que a ferramenta Naive Bayes é considerada uma das mais eficazes em situações de processamento e precisão na classificação de novas amostras. O Naive Bayes apresenta ótimos resultados para a categorização de textos (ARANHA; VELLASCO, 2007).

A técnica utilizada na ferramenta Naive Bayes, que significa Bayes Inocente, segue um modelo probabilístico que em vez de considerar o valor único de um atributo como base da decisão, considera o valor de todos os atributos de interesse e os trata como se fossem independentes entre si (MITCHELL T. M, 2006).

Como citado acima, o teorema utiliza da regra de Bayes, foi realizada uma adaptação, utilizando a abordagem de (TOIT, 2015). Portanto, para analisar se um texto extraído do *Twitter* e considerá-lo positivo, por exemplo, a fórmula ficaria conforme a apresentada na Figura 8.

Figura 8 - Teorema de Bayes

$$P(\text{positive}|\text{tweet}) = \frac{P(\text{tweet}|\text{positive}) P(\text{positive})}{P(\text{tweet})}$$

Fonte: (TOIT, 2015)

Se fosse classificar o texto extraído do Twitter como negativo, bastaria adaptar a equação demonstrada na Figura 8, alterando os campos positive para *negative*.

No exemplo da equação foi utilizado $P(\text{positive}|\text{tweet})$, sendo assim é necessário um conjunto de *tweets* coletados que já estejam classificados como

(positivo ou negativo). Assim, pode-se calcular a probabilidade de um *tweet* ser de uma classe específica.

As chances de encontrar um *tweet* específico no conjunto são relativamente baixas. Dessa forma, o *tweet* deve ser segmentado em palavras, que são os tokens, e o cálculo deve ser feito a partir de cada palavra no conjunto. A fórmula para calcular se um *tweet* é positivo é demonstrada na Figura 9:

Figura 9 - Cálculo para classificar se um *tweet* é positivo

$$P(\text{positive}|\text{tweet}) = P(T1|\text{positivo}) * P(T2|\text{positivo}) * \dots * P(Tn|\text{positivo})$$

Fonte: TOIT (2015)

Na fórmula da Figura 9, o campo T1 até Tn são todas as palavras no *tweet*. Para determinar a probabilidade de uma palavra específica ser de uma categoria precisa-se dividir o número de vezes que T ocorre em *tweets* que estão marcados como positivos no conjunto pelo número total de palavras dos *tweets*.

2.5 Twitter

As redes sociais vêm se tornando uma importante fonte para análise de dados, uma vez que nestas estão dados advindos de diferentes perfis de usuários sobre os mais diversos assuntos. Especificamente para análise de sentimento, o Twitter é uma base rica, pois concentra publicações de usuários com tamanho padronizado, e na grande maioria estão na forma textual. A partir deste cenário, surgiram abordagens baseadas em análise de sentimentos, que objetivam extrair informações úteis sobre as publicações, como a proposta do autor (GONÇALVES; DORES; BENEVENUTO, 2012).

Rodrigues Barbosa et al. (2012) pontuam que o modelo de interação do *Twitter* induz os usuários a compartilhar e expressar continuamente suas opiniões e sentimentos, que são propagados para seus respectivos seguidores.

Para Zago (2008), algumas empresas fazem o uso do Twitter com pouco incentivo, com as contas-feed e esquece de explorar outras possibilidades. Já Russel

(2013) cita a curiosidade humana e a necessidade de compartilhar ideias e experiências, fazer perguntas, interagir de maneira rápida.

O Twitter também fornece aos desenvolvedores acesso a um grande volume de informações em tempo real através de uma *Streaming* API. A *Twitter Streaming*² API fornece para desenvolvedores acessos de baixa latência para a *stream* global do *Twitter* de dados de *tweets*.

²<https://developer.twitter.com/docs/api/streaming>

3 METODOLOGIA

A pesquisa bibliográfica foi a origem para a elaboração deste trabalho. Tentou-se, assim, desvendar, recolher e analisar informações e conhecimentos prévios sobre a extração de conhecimento em documentos textuais, procurando uma resposta ou contribuição para a resolução do problema (VILAÇA, 2010).

Este capítulo expõe os métodos utilizados no processo de desenvolvimento, que tem como objetivo descobrir se há alguma relação entre as publicações e as emoções dos usuários do *Twitter* em relação a um determinado produto ou serviço.

A condução das atividades de classificação de textos, que este trabalho trata, foi desenvolvida de acordo com as etapas de coleta de dados, preparação e pré-processamento, seleção de algoritmos e classificação (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2007).

3.1 Criação Do Algoritmo De Coleta De Dados

Com o objetivo de construir uma melhor visão das possibilidades sobre a mineração de textos sociais a partir de rótulos utilizados na própria frase postada pelo usuário na rede social, em especial no *Twitter*, fez-se uma análise das emoções contidas nas mesmas.

A princípio foi necessário a criação de um algoritmo para coleta de textos na base de dados do *Twitter*. O algoritmo de coleta de dados se encontra como anexo por completo no Apêndice A. O desenvolvimento do código-fonte para obtenção de *tweets* foi a primeira etapa.

Para ter acesso aos textos postados pelos usuários é preciso solicitar as chaves e tokens na área de desenvolvedor do *Twitter*³.

³Site do Twitter para desenvolvedores <https://developer.twitter.com/>

Figura 10 - Chaves e tokens do Twitter para desenvolvedores

Details
Settings
Keys and Access Tokens
Permissions

Application Settings

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

Consumer Key (API Key)	<input style="width: 90%;" type="text"/>
Consumer Secret (API Secret)	<input style="width: 90%;" type="text"/>
Access Level	Read, write, and direct messages (modify app permissions)
Owner	<input style="width: 50%;" type="text"/>
Owner ID	<input style="width: 50%;" type="text"/>

Application Actions

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	<input style="width: 90%;" type="text"/>
Access Token Secret	<input style="width: 90%;" type="text"/>
Access Level	Read and write
Owner	<input style="width: 50%;" type="text"/>
Owner ID	<input style="width: 50%;" type="text"/>

Token Actions

kj

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

O *consumer key secret* e o *access token secret* devem permanecer secretos, pois, são responsáveis pela autenticação de segurança da aplicação, de acordo com a documentação do Twitter⁴. As chaves e tokens foram inseridas no código-fonte de coleta de textos no trecho da linha 4 à 9.

⁴<https://dev.twitter.com/rest/reference/get/search/tweets>

Código 1 - Chaves e tokens do Twitter para desenvolvedores

```

4  ts = TwitterSearch(
5      consumer_key='...',
6      consumer_secret='...',
7      access_token='...',
8      access_token_secret='...'
9  )

```

Fonte: O autor (2018)

Contudo, após ter inserido as chaves e tokens, foi adicionado como parâmetros, palavras chaves para a coleta dos textos. Esses foram inseridos um por vez manualmente, como, por exemplo: “quero comprar”, “tenho interesse”, “gostei muito do produto”, “quanto custa”, “qual valor“. Para visualizar a linha onde foram passados estes parâmetros, encontra-se no Apêndice A um exemplo na linha 12 do algoritmo.

Código 2 - Parâmetro para coleta de textos do Twitter

```

12  tso.set_keywords(['quero comprar'])

```

Fonte: O autor (2018)

Com a execução do algoritmo de coleta de dados, obtém-se no pacote um arquivo nomeado como lista_de_dados_coletados.txt conforme criado na linha 15 do Apêndice A todos os textos que foram coletados.

3.1.1 Coleta dos Dados

A partir da execução do tópico anterior, fez-se a criação do trecho de código-fonte, onde se realizou a coleta dos textos e foram armazenados no arquivo utilizando o trecho da linha 18 a 22 do Apêndice A.

Código 3 - Coleta e Armazenamento de posts no arquivo

lista_de_dados_coletados.txt

```

18  for tweet in ts.search_tweets_iterable(tso):
19      texto.append('%s\n' % tweet['text'])
20      texto.append('\n')
21      arq.writelines(texto)
22  arq.close()

```

Fonte: O autor (2018)

Ao executar o algoritmo, caso o arquivo não tenha sido gerado, foi originado com a opção de inserção no fim do documento, o que garante que os dados já coletados anteriormente não se apaguem, realizando a escrita no final.

3.1.1.1 *Dificuldades na Estrutura dos dados Textuais*

Apesar da grande potencialidade oferecida pela Mineração de Textos, este recurso vem aos poucos ganhando espaço nas organizações. Porém, fez-se uma análise para que a mineração tenha o sucesso esperado, que são:

- As relações entre os atributos precisam ser muito bem definidas, caso contrário, os resultados podem ser mal interpretados;
- Uma interpretação errada pode disfarçar as falhas nos resultados;
- Conhecer os usuários sobre os métodos e técnicas de mineração para o correto funcionamento do processo;
- Privacidade e a legislação sobre a obtenção e divulgação dos dados;
- O problema da baixa qualidade dos dados;

Houve dificuldades cruciais encontradas no processo, surgiram da própria natureza dos dados analisados, como sua dimensionalidade, e na coleta dos mesmos, sendo necessário realizar métodos para tratamento de formatos específicos.

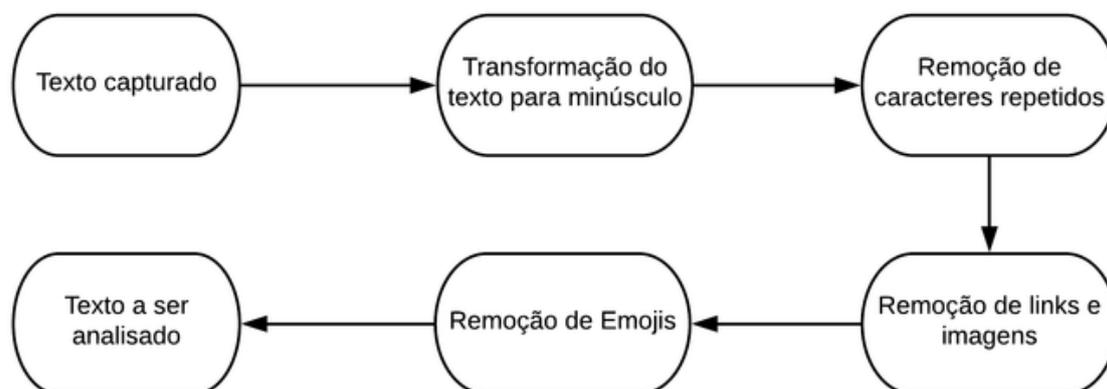
Outro aspecto que dificultou a realização de uma análise mais apurada, foi não ser possível pré-classificar alguns textos que tinham radicais das palavras que não eram semelhantes aos que havia na base previamente treinada. Por exemplo, coletou-se a frase "eu amo você", e na base pré-classificada havia frases utilizando o radical da palavra amor, que reduzido ficou amo, mas a biblioteca do NLTK não reconhece a palavra da primeira frase como o radical de amor, dificultando a classificação em casos específicos.

3.1.1.2 *Preparação dos textos*

Ao capturar os dados textuais do *Twitter*, foi imprescindível realizar reparos no vocabulário e limpeza do texto, para obter uma melhor análise. Demonstra-se na Figura 11 o fluxo realizado nesta limpeza. Após coletado os textos, converteu-os para

minúsculo. Na sequência realizou-se a remoção de caracteres repetidos mais de 3 vezes, bem como os links e imagens do texto. Posteriormente, foi feita a conversão dos *emojis*. Assim, o texto tornou-se preparado para que a análise de emoções básicas.

Figura 11 - Fluxo de pré-processamento



Fonte: Adaptado de Marques (2014)

Os *emojis* são convertidos para seu valor hexadecimal quando são extraídos da base do *Twitter*. Sendo assim, fez-se necessário a remoção dos mesmos.

A limpeza dos dados envolve uma verificação da consistência das informações e o preenchimento ou a eliminação de valores nulos e redundantes. Nessa fase são identificados e removidos os dados duplicados e/ou corrompidos. Uma boa limpeza de dados é essencial, podendo inclusive diminuir o tempo de processamento, eliminando consultas desnecessárias (CASTANHEIRA, 2008, p.17).

Na coleta de dados, normalmente não estão em um formato adequado para extração de conhecimento, por isso, faz-se necessária a aplicação de métodos para extração e integração, transformação, limpeza, seleção e redução de volume destes dados, antes da etapa de mineração (MORAIS; AMBRÓSIO, 2007).

O Quadro 1 expõe as palavras ou junção delas para efetuar a correção de suas abreviações para o seu significado.

Quadro 1 - Mapeamento abreviações

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
nd	nada		pts	pontos
vdd	verdade		dps	depois
sdd	saudade		abs	abraços
sdds	saudades		td	tudo
qro	quero		vsf	(palavrão)
br	brasil		vtnc	(palavrão)
d	de		obg	obrigado
oq	oquê		dnv	de novo
eh	é		mt	muito
neh	né		min	minuto
pgto	pagamento		hr	hora
vcto	vencimento		tô	estou
/	,		sla	sei lá
bike	bicileta			

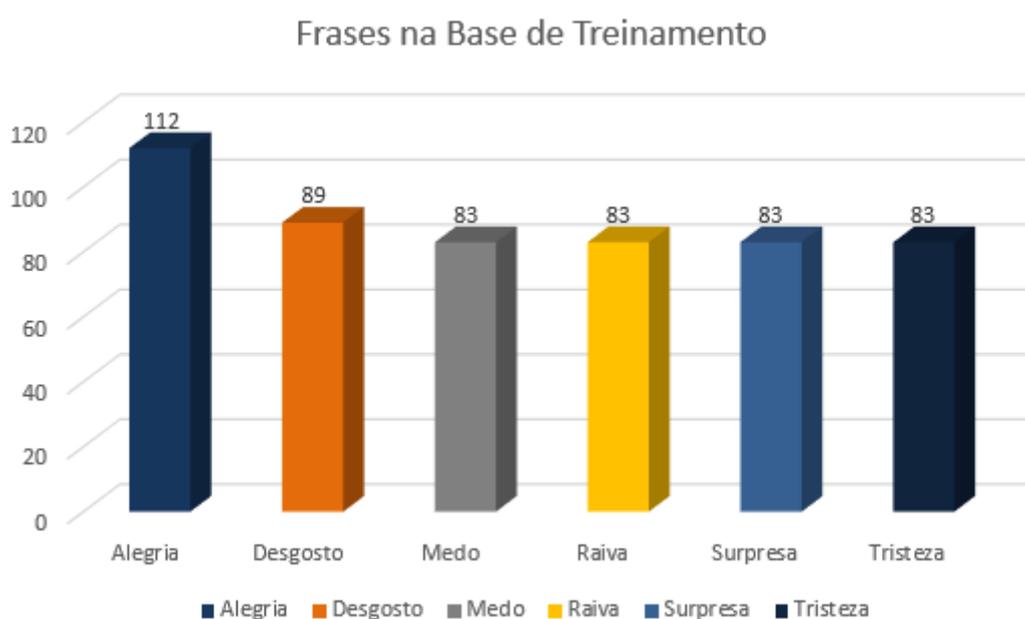
Fonte: Adaptado de Marques (2014)

As frases que se fizeram necessárias a adaptação foram realizadas as trocas das abreviações para o significado correto.

3.1.2 Base Textual já Classificada

Este trabalho utilizou uma base de dados textuais previamente classificados por profissional da área de Neurolinguística, contendo 538 frases. Essa, obtida no curso de Mineração de Textos na plataforma *Udemy*, disponibilizado pelo professor Jones Granatyr, que se encontra, no Anexo A.

Gráfico 1 - Base textual classificada por profissional de Neurolinguística.



Fonte: O autor (2018)

Das frases previamente classificadas, 112 das mesmas são classificadas como alegria, 89 de desgosto e 83 frases de cada uma das demais emoções básicas, totalizando 2935 palavras.

3.1.3 Algoritmo De Pré-Processamento E Classificação

Para elaboração deste processo, fora utilizado a biblioteca *NLTK* juntamente com a linguagem de programação *Python*, por ser um software livre, de código-fonte aberto, adaptado a língua portuguesa e que atendeu a maioria das necessidades. Essa biblioteca, contém diversas funções inseridas, assim, utilizou-se a

`nltk.corpus.stopwords.words('portuguese')`, que utiliza o conceito da *list of words*, que é uma lista que contém palavras e essas são irrelevantes para mineração. Foram deslocadas das frases, realizando o pré-processamento e minerando textos após esse processo.

Para pré-processar os textos, criou-se um algoritmo especificamente para realizar a demanda. O algoritmo por completo se encontra no Apêndice B, no fim deste trabalho.

Foi inserido a base de dados textuais na variável `basetreinamento`, essa, citada no tópico anterior, objetivando utilizá-la para classificar as frases que foram coletadas.

Código 4 - Inserção da base de treinamento com os dados do Anexo A

```
3 basetreinamento = [] # Encontra-se no Anexo A
```

Fonte: O autor (2018)

Ao criar o algoritmo, fez-se necessário a criação de uma função que removesse as palavras irrelevantes e o seu radical, que permaneceria para análise. A função "aplicastemmer", contida na linha 10 do Anexo A, realizou esse processamento.

Código 5 - Aplicação de Remoção de Radical e Palavras Irrelevantes

```
5 stopwordsnltk = nltk.corpus.stopwords.words('portuguese')
6 stopwordsnltk.append('vou') #foi inserido mais esta stopword pois não continha na biblioteca
7 stopwordsnltk.append('tão') #foi inserido mais esta stopword pois não continha na biblioteca
8
9 # função para remoção de stopwords já aplicando o stemmer na frase
10 def aplicastemmer(texto):
11     stemmer = nltk.stem.RSLPStemmer()
12     frasesstemming = []
13     for (palavras, emocao) in texto:
14         comstemming = [str(stemmer.stem(p)) for p in palavras.split() if p not in stopwordsnltk]
15         frasesstemming.append((comstemming, emocao))
16     return frasesstemming
17
18 frasescomstemmingtreinamento = aplicastemmer(basetreinamento)
```

Fonte: O autor (2018)

Removido cada palavra irrelevante e reduzindo ao seu radical, fizeram-se as funções, "buscapalavras", "buscafrequencia" e "buscapalavrasunicas". A função `buscapalavras` tem como principal função retornar as palavras já pré-processadas pelo código-fonte de "aplicastemmer", já em seu radical. A "buscafrequencia" tem como objetivo retornar a frequência das palavras da base de treinamento.

Subsequindo, criou-se a função "buscapalavrasunicas" que retorna cada palavra única contida na variável "basetreinamento", criada na linha 3 do Anexo A demonstrada no Código 5.

Figura 12 - Processamento de palavras das frases do arquivo de textos com postagens dos usuários

```

20 # função que lista todas as palavras da base
21 def buscapalavras(frases):
22     todaspalavras = []
23     for (palavras, emocao) in frases:
24         todaspalavras.extend(palavras)
25     return todaspalavras
26
27 palavrastraining = buscapalavras(frasescomstemmingtreinamento)
28
29 # extração das frequências que as palavras aparecem
30 def buscafrequencia(palavras):
31     # a função nltk.FreqDist checa a frequência de cada palavra
32     palavras = nltk.FreqDist(palavras)
33     return palavras
34
35 frequenciatreinamento = buscafrequencia(palavrastraining)
36
37 # função que retorna a lista com palavras únicas sem repetição
38 def buscapalavrasunicas(frequencia):
39     freq = frequencia.keys()
40     return freq
41
42 palavrasunicastreinamento = buscapalavrasunicas(frequenciatreinamento)
43
44 # função para extrair palavras de cada frase
45 # percorrer cada uma das palavras únicas na base de dados de
46 # treinamento e verificar quais delas existe na frase passada por parâmetro
47 def extratorpalavras(documento):
48     doc = set(documento)
49     caracteristicas = {}
50     for palavras in palavrasunicastreinamento:
51         caracteristicas['%s' % palavras] = (palavras in doc)
52     return caracteristicas
53
54 basecompletatreinamento = nltk.classify.apply_features(extratorpalavras, frasescomstemmingtreinamento)

```

Fonte: O autor (2018)

Juntas, as funções, têm como objetivo extrair as palavras para chegar a ter apenas uma de cada na base de treinamento para realizar a classificação baseando-se apenas nelas.

Foram ao total 543 frases contendo 6 emoções básicas na base de treinamento, sendo assim, utilizou-se a base de treinamento como 70% e os outros 30%, resultante em 54 frases. Deu-se entrada para classificá-las nesta proporção por vez de processamento das mesmas.

3.1.4 Base de dados textuais coletada

A base de dados textual deste trabalho é composta por textos publicados pelos usuários no *Twitter*, cujo foco foi relacionado às compras de produtos ou serviços. Foram aquelas que os usuários disponibilizaram como pública. Para extrair os textos desses grupos foi utilizada a API *Twitter Streaming*⁵. O Quadro 1 apresenta uma amostra dos textos que compõem a base de dados.

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	você 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)

Baseando-se nos dados coletados, obteve-se uma base textual composta por 2342 palavras, presentes em 357 publicações realizadas pelos usuários que postaram.

3.2 Seleção De Algoritmos

A seleção dos algoritmos é um procedimento, no qual, são considerados o que será necessário para avaliar a classificação. Nessa etapa, foram escolhidos dois algoritmos, julgados como necessários para a aplicação da técnica, que foram demonstrados nos tópicos a seguir.

⁵<https://developer.twitter.com/docs/api/streaming>

3.2.1 Algoritmo de Comparação de Base Textuais

Após a preparação dos dados textuais da coleta nas etapas anteriores, foi realizado uma comparação entre a base previamente classificada e a coletada, já limpa, objetivando checar se haviam frases coletadas que poderiam ser iguais as que já continha na base que utilizaria como pilar para classificar as frases coletadas. Esse processo foi feito com o algoritmo que se encontra no Apêndice D.

Código 6 - Algoritmo de comparação de frases coletadas com frases já classificadas pelo profissional de Neurolinguística.

```

1  def procura_palavra():
2      f1 = open("lista_de_dados_coletados.txt.txt", "r", encoding="utf-8")
3      f2 = open("base_treinamento.txt", "r+", encoding="utf-8")
4      for line1 in f1:
5          f2.seek(0)
6          for line2 in f2:
7              if line1 == line2:
8                  print("-----IGUAIS-----")
9                  print("linha1 ->", line1, "linha2 ->", line2)
10                 print("iguais\n")
11
12         f1.close()
13         f2.close()
14
15  if __name__ == '__main__':
16      procura_palavra()

```

Fonte: O autor (2018)

Com esse problema, o objetivo de o algoritmo foi comparar cada frase coletada com todas as frases já treinadas e se contivesse alguma igual ele retornava qual era idêntica, para realizar a remoção da mesma, pois, poderia atrapalhar nos resultados finais.

3.2.2 Algoritmo de Classificação de Frases Coletadas

O Algoritmo utilizado para realizar a avaliação segue em anexo no Apêndice C. Com ele, realizou-se o pré-processamento da base de treinamento e da base de coleta e demonstrou os resultados obtidos sobre a classificação.

Código 7 - Algoritmo de classificação das frases

```
61 basecompletatreinamento = nltk.classify.apply_features(extratorpalavras, frasescomstemmingtreinamento)
62 basecompletacoletada = nltk.classify.apply_features(extratorpalavras, frasescomstemmingcoletada)
63
64 # classificador de treinamento da base coletada de textos, baseando na base de dados de treinamento
65 classificador = nltk.NaiveBayesClassifier.train(basecompletatreinamento)
66
67 # informa a quantidade de precisão o algoritmo retorna
68 print("Precisão do algoritmo: ", nltk.classify.accuracy(classificador, basecompletacoletada))
69
70 # constroi a tabela de matriz de confusão
71 from nltk.metrics import ConfusionMatrix
72 esperado = []
73 previsto = []
74 for (frase, classe) in basecompletacoletada:
75     resultado = classificador.classify(frase)
76     previsto.append(resultado)
77     esperado.append(classe)
78 matriz = ConfusionMatrix(esperado, previsto)
79 print(matriz)
```

Fonte: O autor (2018)

O objetivo foi comparar o documento de texto treinado com os textos coletados nos perfis públicos, de modo a identificar, baseado na quantidade de palavras relacionadas as emoções básicas e a frequência delas, no documento, bem como retornar a porcentagem de acerto da classificação das mesmas.

4 RESULTADOS

Este capítulo objetiva demonstrar detalhadamente o resultado adquirido acerca da realização dos procedimentos citados anteriormente.

No processo de desenvolvimento algumas funcionalidades foram criadas buscando atingir o objetivo final, dentre elas, criou-se um algoritmo para comparar os dados coletados e os textos já classificados, outro algoritmo para classificação dos mesmos textos adquiridos e, por fim, um algoritmo de avaliação dos resultados.

4.1 Coleta De Dados

O primeiro passo do processo de Mineração de Textos em redes sociais é a coleta de dados. A implementação do *script* de coleta de *tweets*, criou-se para realizar a aquisição de dados da plataforma do *Twitter*. Essa implementação foi criada, pois, executa o processo de coleta de modo eficaz, fundamentado para criação através da leitura do autor (FILHO, 2014).

Para a identificação dos *tweets* relacionados com o assunto tratado, o *script* recebeu de entrada textos, como palavras-chave, relacionados ao assunto.

Como resultado dessa coleta, foi obtida uma grande quantidade de *tweets* relacionados ao assunto de interesse deste trabalho, no caso, emoções básicas. No Quadro 3, são apresentados os dados finais da coleta, contabilizando frases únicas que foram separadas para classificá-las.

Quadro 3 - Resultado da Coleta de Dados Brutos

Dias de Coleta	16
<i>Tweets</i> coletados	7348 <i>tweets</i>
Tamanho total em disco (arquivo txt)	102 Mb

Fonte: O autor (2018)

O estudo de caso foi aplicado usando apenas os dados do dia 05 de setembro à 23 de outubro de 2018, pois, foi esse o período de maior concentração de *tweets* coletados, contudo, houve coleta de dados anteriores no mês de junho e agosto, porém, foram desconsiderados para geração de resultados deste trabalho, e sim para aprendizado do próprio autor em relação ao tema. Mesmo com a obtenção de uma grande quantidade de dados, 85% dos mesmos não se encaixavam ao cenário utilizado para análise, fizeram-se necessários a remoção de diversos *tweets*.

4.2 Pré-Processamento E Classificação

Nesta fase, foi aplicado o pré-processamento de dados apenas nos *tweets* que foram selecionados para fazer parte do estudo de caso. Foram removidos links, nomes de usuários, caracteres especiais e as *stopwords*.

Foram aplicadas técnicas de processamento de linguagem natural, com a finalidade de deixar o texto mais limpo para as próximas etapas. Para isso, foi empregada a biblioteca *Natural Language Toolkit (NLTK)*⁶ do *Python*, na versão 3.6.

No projeto, foi criado um *script*. Esse algoritmo realiza todos os requisitos do pré-processamento utilizando Processamento de Linguagem Natural, bem como, classifica as frases coletadas a partir da base de treinamento contida no Anexo A. Até então, a implementação só aceita, como entrada, frases advindas de um arquivo do tipo TXT.

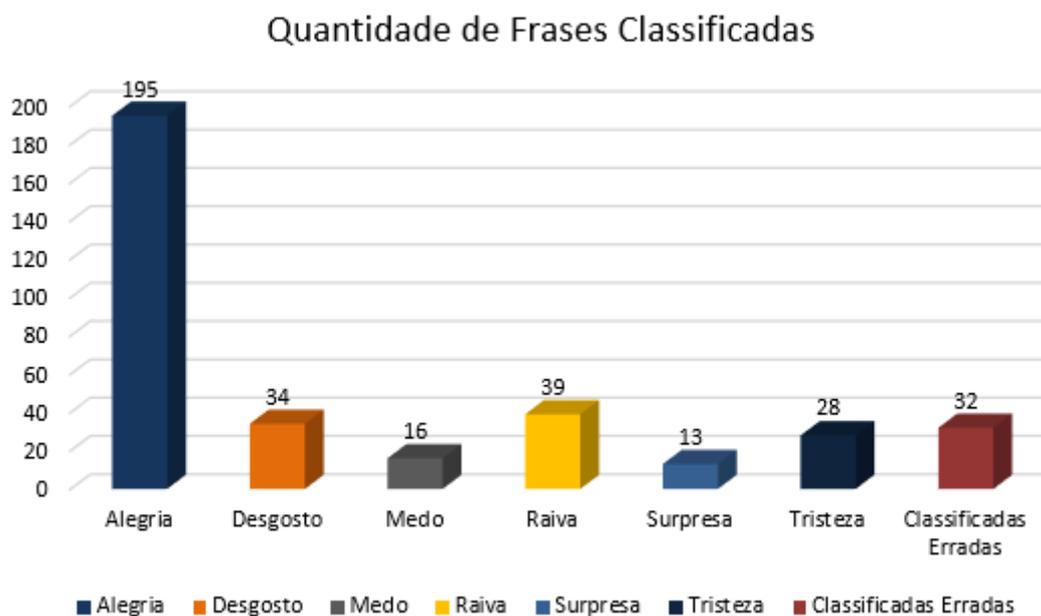
Essa técnica transforma a coleção de textos em uma tabela, que indica as palavras e o número de ocorrência dos termos dentro do corpo textual, assim o valor é transformado em numérico para ser utilizado pelos algoritmos de classificação (SILVA, 2016). E ainda, houve um árduo trabalho linguístico no pré-processamento ao tratar-se com as características da língua Portuguesa, baseado no léxico.

No pré-processamento é feita a decomposição das frases. Frases que contém um conjunto de letras que formam as palavras, números e símbolos que são agrupados em frases. Ao pré-processar esses elementos significam percorrer as frases coletadas, a cada espaço contido, uma unidade léxica, tendo como limite a quebra de linha (BARBOSA, 2013).

Em seguida aplicou-se a tokenização do texto, utilizando o espaço como separador dos tokens. As emoções foram classificadas pelas classes são: (i) Alegria, (ii) Desgosto, (iii) Medo, (iv) Raiva, (v) Surpresa, (vi) Tristeza. Como primeiro passo, aplicou-se a limpeza do texto, que consistiu em aplicar a radicalização, remoção das *stopwords*, caracteres indesejados e URLs. A partir do pré-processamento, criou-se o modelo de classificação *tweets*, classificou todos os *tweets* coletados, analisados, separados e pré-processados.

⁶<http://www.nltk.org/>

Gráfico 2 - Quantidade de Frases Classificadas

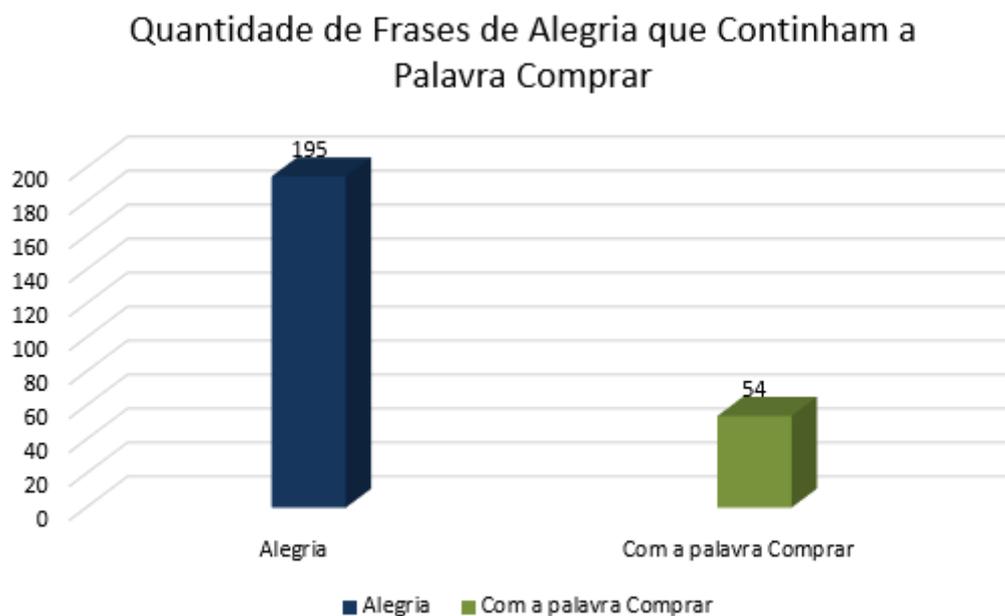


Fonte: O autor (2018)

Nos gráficos a seguir foram demonstradas a quantidade de frases que continha a palavra comprar de cada emoção básica, bem como, foi demonstrando separadamente a quantidade de frases classificadas de cada emoção e algumas das palavras chaves que as fizeram com que fosse classificadas na categoria.

O Gráfico 3 demonstra que 195 das frases classificadas foram para a classe Alegria, contendo as palavras com radical "am", "hoj", "sint", "ador", "grand", "maravilh", dentre outros. Esses radicais, foram fatores determinantes para a classificação. Do total de frases classificadas como Alegria, 54 delas continham a palavra comprar.

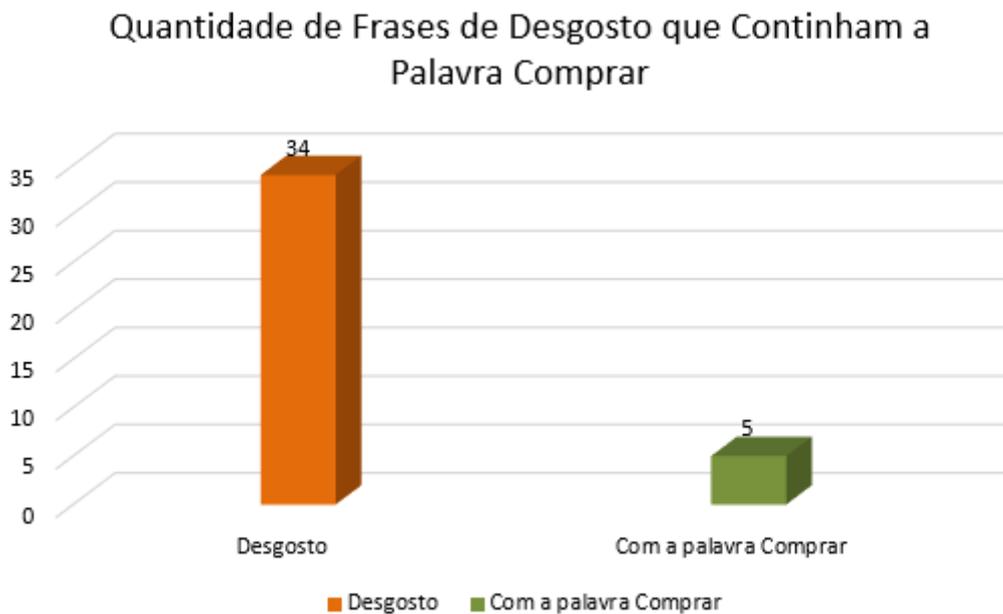
Gráfico 3 - Quantidade de Frases de Alegria que Continham a Palavra Comprar



Fonte: O autor (2018)

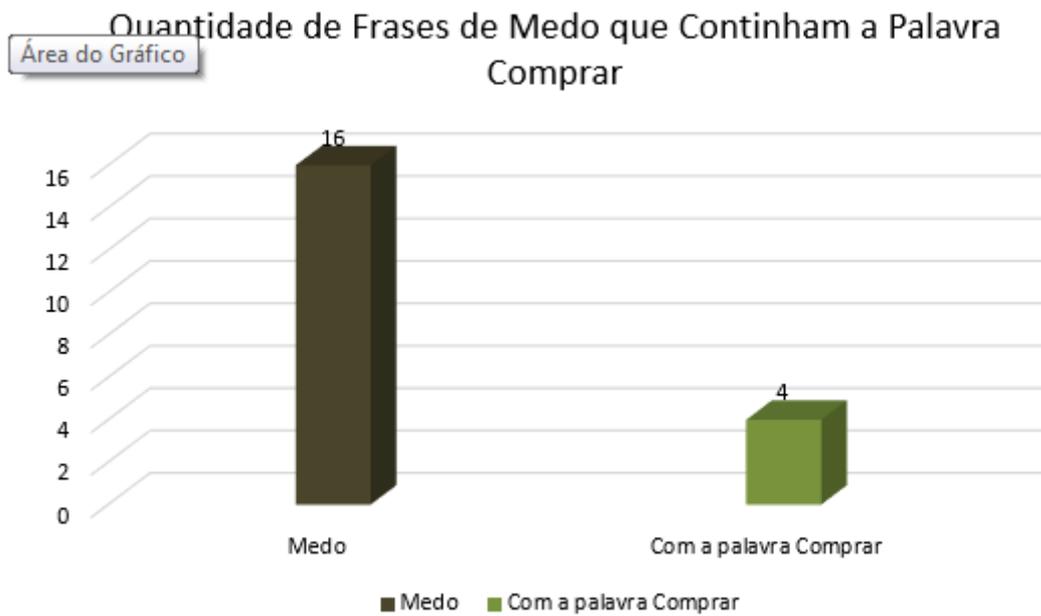
O Gráfico 4 demonstra que 34 das frases classificadas foram para a classe Desgosto, contendo as palavras com radical "abomin", "tod", "desagr", "man", "pesso", "ser", dentre outros. Esses radicais, foram fatores determinantes para a classificação. Do total de frases classificadas como Desgosto, 5 delas continham a palavra comprar.

Gráfico 4 - Quantidade de Frases de Desgosto que Continham a Palavra Comprar



O Gráfico 5 demonstra que 16 das frases classificadas foram para a classe Medo, contendo as palavras com radical "perig", "socorro", "amedront", "pod", "lug", "cuid", dentre outros. Esses radicais, foram fatores determinantes para a classificação. Do total de frases classificadas como Medo, 4 delas continham a palavra comprar.

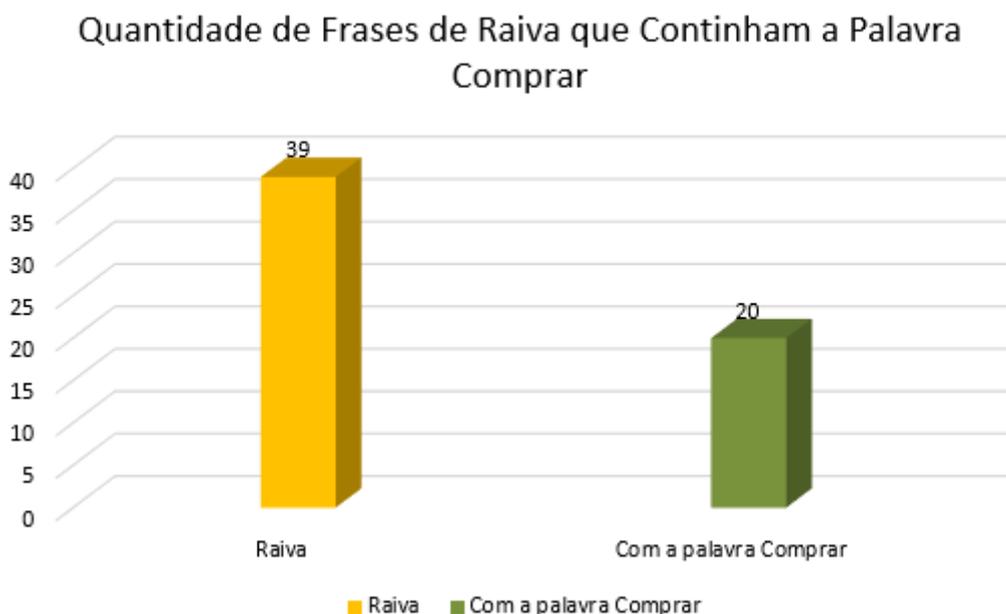
Gráfico 5 - Quantidade de Frases de Medo que Continham a Palavra Comprar



Fonte: O autor (2018)

O Gráfico 6 demonstra que 39 das frases classificadas foram para a classe Raiva, contendo as palavras com radical "frustr", "ruim", "deix", "pod", "odi", "dinh", dentre outros. Esses radicais, foram fatores determinantes para a classificação. Do total de frases classificadas como Raiva, 20 delas continham a palavra comprar.

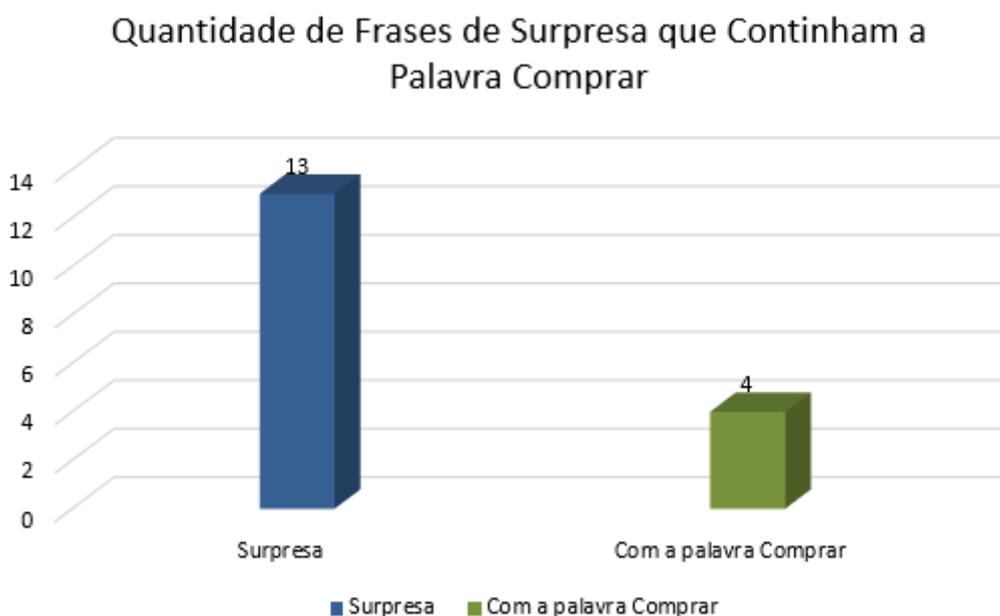
Gráfico 6 - Quantidade de Frases de Raiva que Continham a Palavra Comprar



Fonte: O autor (2018)

O Gráfico 7 demonstra que 13 das frases classificadas foram para a classe Surpresa, contendo as palavras com radical "surpreend", "acredit", "real", "expect", "sumiu", "boquiabert", dentre outros. Esses radicais, foram fatores determinantes para a classificação. Do total de frases classificadas como Surpresa, 4 delas continham a palavra comprar.

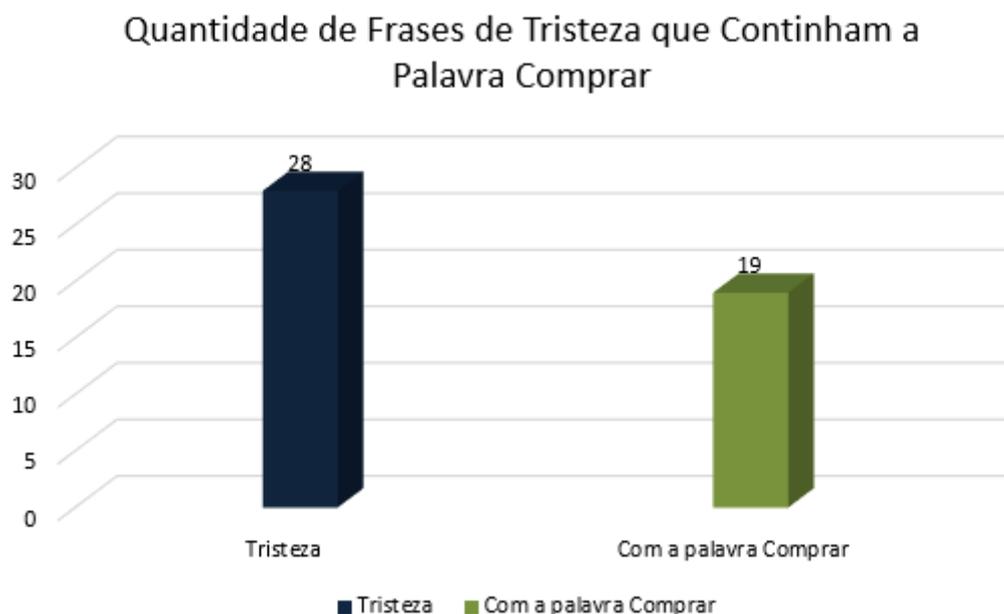
Gráfico 7 - Quantidade de Frases de Surpresa que Continham a Palavra Comprar



Fonte: O autor (2018)

O Gráfico 8 demonstra que 28 das frases classificadas foram para a classe Tristeza, contendo as palavras com radical "trist", "depress", "nad", "abat", "aborrec", "amargur", dentre outros. Esses radicais, foram fatores determinantes para a classificação. Do total de frases classificadas como Tristeza, 19 delas continham a palavra comprar.

Gráfico 8 - Quantidade de Frases de Tristeza que Continham a Palavra Comprar



Fonte: O autor (2018)

Das frases classificadas como Alegria, 54 frases continham a palavra compra. As classificadas como Desgosto, 5 frases continham a palavra comprar. Das frases de Desgosto e Surpresa, em cada emoção haviam 4 frases com a palavra comprar. Das frases de Raiva, 20 tinham a palavra comprar e as frases de Tristeza, 28 frases continham a palavra comprar.

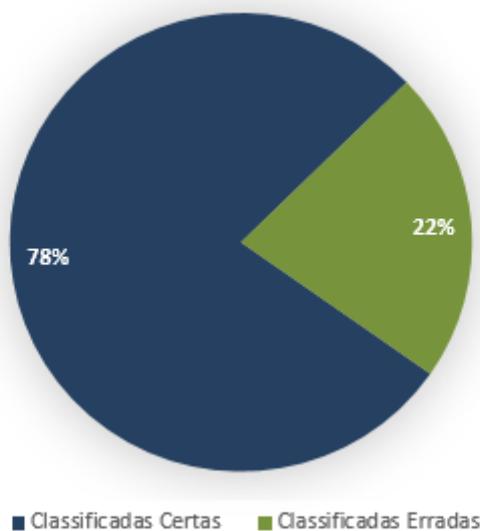
4.3 Análise Dos Dados

Após o pré-processamento e classificação dos *tweets* realizada no passo anterior, fez-se necessária a realização de uma análise das classificações geradas. Através de um algoritmo de avaliação com base em frases classificadas por um profissional da área de Neurolinguística, a classificação foi justificada.

A frequência do aparecimento de algumas palavras nos textos coletados foi primordial para demonstração os resultados.

Ao avaliar as 357 frases, observou-se que foram classificadas, 195 como alegria, 34 como desgosto, 16 como medo, 39 como raiva, 13 como surpresa e 28 como tristeza, obtendo uma precisão de 78,27868852459017%, totalizando 194 classificadas corretamente e 32 erroneamente. O Gráfico 9 demonstra a classificação Geral das frases.

Gráfico 9 - Classificação Geral de Frases

Classificação Geral de Frases

Fonte: O autor (2018)

Contudo, obteve-se uma classificação considerável das frases como demonstrado no Gráfico 9, onde, cerca de 78% das frases foram classificadas corretamente e 22% foram classificadas como erradas em relação as emoções propostas.

5 CONCLUSÃO

Este trabalho realizou o processo de Mineração de Textos no *Twitter*. Bem como, identificou emoções nas frases coletadas semelhantes às que haviam na base de treinamento do algoritmo.

A etapa que se categoriza como mais importante deste processo, é a Mineração de Dados, caracterizada por um algoritmo que diante da tarefa será eficiente em atrair conhecimento implícito e útil de uma boa base de coleta. Julga-se que mineração de dados é a fase de transformação de dados para conhecimento.

Para realização do estudo, diversas pesquisas a respeito de emoções e Mineração de Textos em redes sociais foram feitas, obtendo uma fundamentação teórica dos processos realizados. Com as ferramentas utilizadas, determinou-se a implementação de algoritmos em *Python* para mineração.

Através das pesquisas percebeu-se que, além de ter muitos trabalhos acerca do assunto de Mineração de Textos e Mineração de Dados, existem poucos trabalhos específicos de mineração de textos objetivando encontrar emoções nas bases textuais.

Unir a área de Análise de Sentimentos, identificando emoções em textos proporciona um novo leque de pesquisas. Aplicado a AS, propicia um caminho de investigação na área de conhecimento.

Com a Análise de Sentimentos em várias aplicações têm sido propostas para a Análise de Sentimentos, há alguns cenários onde se pode aplicar, como por exemplo, sumarização de opiniões prevalecente em revisões de produtos, termômetros de popularidade para marcas, pessoas ou organizações, usadas em ações de marketing ou relações públicas e previsão de indicadores a partir de sentimentos.

Mesmo não podendo basear-se inteiramente acerca de publicações para avaliar um nível emocional do usuário, são muitos os usuários que expõem as vidas pessoais e pensamentos íntimos nas redes sociais.

Alguns desafios foram enfrentados no decorrer do desenvolvimento da proposta, iniciando pela preparação dos dados textuais, dado no que lhe concerne mal estruturado.

Com a pesquisa feita e os testes realizados, pode-se perceber que, se houver uma base de treinamento maior, classificada por profissionais da área de

Neurolinguística, que está atrelada aos estudos sobre os hábitos humanos, consegue-se treinar uma base bem estruturada, tendo potencial para avaliar emoções contidas com muito mais precisão nas frases que foram adquiridas nas redes sociais.

Finalmente, considerando-se técnicas de Mineração de Texto apresentadas, mostram-se aplicáveis às atividades de processamentos de postagens coletadas *online*, o que se precisa para minerar dados das redes sociais é uma motivação de encontrar informações que possam vir a ser úteis.

Portanto, a demanda de classificar emoções objetivando de coletar informações sobre interesse de compra, com ferramentas informatizadas, necessita de melhoras em alguns aspectos, como, por exemplo, a lista de *stopwords* na língua portuguesa contida na biblioteca NLTK, utilizada no desenvolvimento, não contém todas as palavras necessárias para a realização do pré-processamento adequado, teve-se que adaptar as que não continha. Outro aspecto que necessita de melhoras é a técnica de limpeza de dados, que nessa aplicação foram realizadas manualmente.

Mesmo a 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.

6 TRABALHOS FUTUROS

Fica a sugestão de desenvolvimento posterior, a criação de um algoritmo, que poderia principalmente abordar erros de classificação de dados, bem como a criação de uma base maior já classificada por um profissional da área de Neurolinguística, para que possa realizar processamentos de bases ainda maiores e obter uma precisão mais aceitável acerca de Mineração de Emoções.

Realização de um novo estudo de caso com foco na obtenção de dados de outras plataformas de coleta e a criação de algoritmos utilizando a inteligência artificial para treinar novos dados coletados.

REFERÊNCIAS

- ALMEIDA, L. G. P. de. **Análise de algoritmos de agrupamento para base de dados textuais**. Laboratório Nacional de Computação Científica, Petrópolis, 2007.
- 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.
- ARANHA, C; PASSOS, E. **A Tecnologia de Mineração de Textos**. Laboratório ICA Elétrica PUC-Rio, 2006.
- ARRUDA, M.J.F.C. **O ABC das emoções básicas-Implementação e avaliação de duas sessões de um programa para a promoção de competências emocionais. Um enfoque comunitário**. Tese (mestrado). Universidade de Açores: Ponta Delgada, 2014.
- BARBOSA, Alexandre N. **Descoberta de Conhecimento Aplicado à Base de Dados Textual de Saúde**. São Leopoldo, 2012.
- BARBOSA, F. **Metodologia para Estruturação de Informações de Laudos Radiológicos**. Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2013.
- BARION, Eliana Cristina Nogueira; LAGO, Decio. **MINERAÇÃO DE TEXTOS**. Revista de Ciências Exatas e Tecnologia, 2008. Disponível em: <<http://pgsskroton.com.br/seer/index.php/rcext/article/viewFile/2372/2276>>. Acesso em: 29 out. 2018.
- BIRD, Stven; KLEIN, Ewan; LOPER, Edward. **Natural Language Processing with Python - Analyzing Text with the Natural Language Toolkit**. 1ª. ed. 2009. 479 p.
- BOGDANCHIKOV, A; ZHAPAROV, M; SULIYEV, R. Python to learn programming. **IOPScience ScieTech**. 2013. Disponível em: <<http://iopscience.iop.org/>>. Acesso em: 20 set. 2018.
- BOLLEN, J; MAO, H; ZENG, X. **Twitter mood predicts the stock market**. Journal of Computational Science, 2011.
- BORGES, L. E. **Python para desenvolvedores**. Rio de Janeiro: Novatec, 2014, p. 27 - 29.
- BORGES, Luis E. **Python para desenvolvedores**. 2a. ed. Rio de Janeiro: Edição do Autor, 2010.

BRAGA, L. O. P. **Mídias são parciais! Será? Uma ferramenta de análise automática dos canais de notícia em relação às figuras políticas do Brasil.** Universidade Federal Rural de Pernambuco, 2017.

BRITO, Edeleon Marcelo Nunes. **Mineração de Textos: Detecção automática de sentimentos em comentários nas mídias sociais.** Belo Horizonte, Minas Gerais, 2016.

BUBENHOFER, N; PTASHNYK, S; KILIAN, C. K. **Corpora, bancos de dados e a web: estado da arte da pesquisa em fraseologia e lexicografia assistida por computador.** Entrelinhas, v. 6, 2012, p. 83-96.

CARRILHO JUNIOR, João Ribeiro. **Desenvolvimento de uma Metodologia para Mineração de Textos.** Rio de Janeiro, 2007.

CARVALHO, BRANCO, Saullo, CASTELOK. **KExploit - Script em Python para automação de Exploração de Kernel de Linux,** 2016.

CASANOVA , Nuno ; SEQUEIRA , Sara ; SILVA, Vítor Matos. **Trabalho desenvolvido no âmbito da disciplina de ‘Psicologia Geral’** . 2009.

CASTANHEIRA, Luciana Gomes. **Aplicação de Técnicas de Mineração de Dados em Problemas de Classificação de Padrões.** Belo Horizonte, 2008.

CECI, Flávio. **Um Modelo Semiautomático para a Construção e Manutenção de Ontologias a partir de Bases e Documentos Não Estruturados.** 2010.

CORDEIRO, A. D. **Gerador Inteligente de Sistemas com Auto-aprendizagem para Gestão de Informações e Conhecimento.** UFSC: Departamento de Engenharia de Produção, 2005.

CORRÊA, Geraldo Nunes et al. **Uso da mineração de textos na análise exploratória de artigos científico.** 2012. Disponível em: <http://www.icmc.usp.br/CMS/Arquivos/arquivos_enviados/BIBLIOTECA_113_RT_383.pdf>. Acesso em: 31 out. 2018.

D E LISBOA, RSLingo. **A Formal Requirements Specification Approach based on Linguistic Patterns TeseProvis** . 2013.

DAMÁSIO, A. **Em busca de Espinosa: Prazer e dor na ciência dos sentimentos.** São Paulo, São Paulo: Companhia das Letras, 2004.

DING, X; LIU, B; ZHANG, L. **Entity discovery and assignment for opinion mining applications.** New York, USA, 2009. 1125–1134 p.

EKMAN, P. **Emotions Revealed: recognizing faces and feelings to improve communication and emotional life**. New York: Times Books, 2003.

ELMASRI, Ramez; NAVATHE, Shamkant B. **Sistemas de banco de dados**. 6. ed. São Paulo: Pearson, 2011. 788 p.

FAYYAD; PIATETSKY-SHAPIRO; SMYTH. **From Data Mining to Knowledge Discovery in Databases**. *AAAI Magazine*. 1996.

FELDMAN, Ronen; SANGER, James. **The text mining handbook: advanced approaches in analyzing unstructured data**. New York, Cambridge University Press: Cambridge University, 2006.

FILHO, J. A. C. **Mineração de textos: análise de sentimento utilizando tweets referentes à copa do mundo 2014**. Universidade Federal do Ceará, 2014.

FLACH & LACHICHE, Flach & Lachiche. **Naive Bayesian Classification of Structured Data**. Kluwer Academic Publishers, Holanda e Universidade de Bristol, 2003. Disponível em: <<http://www.cs.bris.ac.uk/~flach/papers/mlj04-1BC-final2.pdf>>. Acesso em: 20 ago. 2018.

FRANÇA, T. C et al. **Big social data: Princípios sobre coleta, tratamento e análise de dados sociais**. XXIX Simposio Brasileiro de Banco de Dados–SBBDD, 2014.

FREITAS, A. M. **A Psicologia das emoções: O fascínio do rosto humano**. Porto: Edições Universidade Fernando Pessoa, 2011a.

_____. **O código de Ekman: O cérebro, a face e a emoção**. Porto: Edições Universidade Fernando Pessoa, 2011b.

FREITAS, A. M; CASTRO, E. **Expressão facial: O reconhecimento das emoções básicas em dependentes de cocaína. Estudo empírico com portugueses**. Revista da Faculdade de Ciências da Saúde: Edições Universidade Fernando Pessoa, 2007, p. 28-37.

GALATI, D; MANZANO, M; SOTGIU, I. **The subjective components of happiness and their attainment: A cross-cultural comparison between Italy and Cuba**. *Social Science Information*, 2006.

GARCIA; ALAN CÁSSIO. **Mineração de dados aplicada a sistemas de recomendação**. Santa Cruz do Sul. Trabalho de Conclusão de Curso () - UNIVERSIDADE DE SANTA CRUZ DO SUL, 2012.

GOLEMAN, D. **Inteligência Emocional**. 17ª. ed. Lisboa: Temas e Debates, 2012.

GOMES, Helder Joaquim Carvalheira. **Text Mining: análise de sentimentos na classificação de notícias**. Lisboa, 2013.

GONZALEZ, M; LIMA, V. **Recuperação de Informação e Processamento da Linguagem Natural**. PUCRS. 2003. Disponível em: <<https://www.inf.pucrs.br/~gonzalez/docs/minicurso-jaia2003.pdf>>. Acesso em: 7 set. 2018.

GONÇALVES, J. C et al. **Discovering Intentions and Desires within Knowledge Intensive Processes**. Management Workshops, Springer International Publishing Switzerland, Austria, 2015, p. 273-284.

GONÇALVES, P; DORES, W; BENEVENUTO, F. **Uma escala psicometrica para medição de sentimentos no twitter**. 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.

GUPTA, Vishal; LEHAL, Gurpreet S. **A Survey of Text Mining Techniques and Applications**. 2009.

IBRAHIM, Abu El-Khair. **Effects of stop words elimination for Arabic information retrieval: a comparative study**. International Journal of Computing & Information Sciences, 2006.

KAO, Anne; POTEET, Steve R. **Natural language processing and text mining**. Springer Science & Business Media, 2007.

KOTHAPALLI, M; SHARIFAHMADIAN, E; SHIH, L. **Data mining of social media for analysis of product review**. 2016. 25 p.

KOTSIANTIS, Sotiris B; ZAHARAKIS, I; PINTELAS, Panayiotis. **Supervised machine learning: A review of classification techniques**. 2007.

LAHIRI, Lahiri R. COMPARISON OF DATA MINING AND STATISTICAL TECHNIQUES FOR CLASSIFICATION MODEL. **Thesis of Master of Science at the Louisiana State University, US**. 2006. Disponível em: <http://etd.lsu.edu/docs/available/etd-11012006-192748/unrestricted/Lahiri_thesis.pdf>. Acesso em: 11 ago. 2018.

LANCASTER, F. W. **Indexing and abstracting in theory and practice**. London: Library Association, 1998.

LEONEL JUNIOR, R. d. A et al. **Mineração em Dados Abertos In: IV Jornada Científica de Sistemas de Informação**. Parnaíba, PI, 2014.

LIDDY, E. **Natural Language Processing. Encyclopedia of Library and Information Science**. New York: Marcel Decker, Inc, 2001.

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

_____. **“Sentiment analysis and subjectivity”**, *Handbook of natural language processing*. 2. ed. 2010.

LOH, S. **Abordagem Baseada em Conceitos para Descoberta de Conhecimento em Textos**. Universidade Federal do Rio Grande do Sul, Instituto de Informática, 2001.

LOPER; EDWARD; BIRD. **The natural language toolkit: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS**, v. 1. 2002, p. 63-70.

LOPES, Maria Célia Santo. **Mineração de Dados Textuais Utilizando Técnicas de Clustering para o Idioma Português**. Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2004.

LOPES, I. L. **Uso das linguagens controlada e natural em bases de dados: revisão da literatura**. SciELO Brasil, 2002, p. 41-52.

LOPES, R. F. F; SANTOS, M. R; LOPES, E. J. **Efeitos do relaxamento sobre a ansiedade e desesperança em mulheres com câncer**. São Paulo: Rev. bras. ter. comport. Cogn., v. 10, 2008, p. 39-49.

MARQUES, D. L; COSTA, L. F. S. **Atraindo alunos do ensino médio para a computação: Uma experiência prática de introdução a programação utilizando jogos e python**. Departamento de Ciências Exatas-Universidade Federal da Paraíba-UFPB, 2011.

MARQUES, JADER FABIANO BATISTA. **Sistema De Monitoramento E Análise De Comentários No Twitter**. SÃO JOSE, 2014.

MATA , Andréia Silva. **Inteligência Emocional Em Crianças**. UNIVERSIDADE SÃO FRANCISCO , 2007.

MELO, MARCELO DAMASCENO. **Um Processo De Mineração De Dados Para Predição De Níveis Criminais De Áreas Geográficas Urbanas**. Fortaleza, Ceará, 2010.

MIGUEL, Fabiano Koich . **Psicologia das emoções: uma proposta integrativa para compreender a expressão emocional**. Universidade Federal de Londrina, Londrina, Brasil, 2015.

MITCHELL T. M, Mitchell. **The Discipline of Machine Learning: Association for the Advancement of Artificial Intelligence (AAAI)**. 2006. Disponível em: <<http://www.cs.cmu.edu/~tom/pubs/MachineLearning.pdf>>. Acesso em: 5 set. 2018.

MORAIS, Edilson Andrade Martins; AMBRÓSIO, Ana Paula L. **Mineração de Textos**. Goiânia: UFG, 2007.

MORENCY, Louis-Philippe; RADA, Mihalcea; PAYAL, Doshi. **Towards multimodal sentiment analysis: harvesting opinions from the web**. 2011, p. 169-176.

MUKHERJEE, Subhabrata. **Sentiment analysis - a literature survey**. Indian Institute of Technology, 2012.

NARUKAWA, C. M; LEIVA, I. G; FUJITA, M. S. L. **Indexação automatizada de artigos de periódicos científicos: análise da aplicação do software SISA com uso da terminologia DeCS na área de Odontologia**. Informação & Sociedade: Estudos, 2009.

NASCIO. **Think Before You Dig: Privacy Implications of Data Mining & Aggregation.[S.I]**: NASCIO. Representing Chief Information Officers of the States, US Government, 2004.

NETTO, A. A. **IHC – Modelagem e Gerência de Interfaces com o Usuário**. Visual Books, Rio de Janeiro, 2004.

PERNA, C. L; DELGADO, H. K; FINATTO, M. J. . **Linguagens Especializadas em Corpora: modos de dizer e interfaces de pesquisa**. EDIPUCRS, 2010.

PINHEIRO, C. A. R. **Inteligência Analítica - Mineração de Dados e Descoberta de Conhecimento**. Rio de Janeiro: Ciência Moderna Ltda, 2008.

RANGÉ, B. **Influência das cognições na vulnerabilidade ao stress**. São Paulo: In Lipp, N. M. E. Mecanismos psicofisiológicos do stress: teoria e aplicações clínicas, 2003.

RENNIE, J. D et al. **Tackling the poor assumptions of naive bayes text classifiers**. In: **ICML**. 2003, p. 616-623.

REZENDE, S. O; MARCACINI, R. M; MOURA, M. F. **Revista de Sistemas de Informação da FSMA**. 2011, p. 7-21.

RODRIGUES BARBOSA , Glívia Angélica et al. **Characterizing the effectiveness of twitter hashtags to detect and track online population sentiment. In: PROCEEDING OF THE 2012 ACM ANNUAL CONFERENCE EXTENDED ABSTRACTS ON HUMAN FACTORS IN COMPUTING SYSTEMS EXTENDED ABSTRACTS.** . ed. Austria, 2012.

RODRIGUES, Carlos Augusto S. et al. **Mineração de Opinião / Análise de Sentimentos.** UFSC, 2011.

RONALDO, G; EMMANUEL, P. **Data Mining - Um Guia Prático.** v1. ed. Rio de Janeiro: Elsevier, v .1, 2005.

ROSA, R. L. **Análise de sentimentos e afetividade de textos extraídos das redes sociais.** Universidade de Sao Paulo, 2015.

RUSSEL, Mathew A.. **Mining the social web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub and More.** 2ª. ed. Sebastopol: O'reilly Media, Inc, 2013.

SANTANA, Osvaldo; GALES, Thiago. **Python e Django: desenvolvimento ágil de aplicações web.** São Paulo: Novatec, 2010.

SANTOS , Ronnie E. S et al. **Técnicas De Processamento De Linguagem Natural Aplicadas Ao Processo De Mineração De Textos: Resultados Preliminares De Um Mapeamento Sistemático.** Revista de Sistemas e Computação. Universidade Federal de Pernambuco, 2014. Disponível em: <<https://revistas.unifacs.br/index.php/rsc/article/download/3030/2498>>. Acesso em: 22 out. 2018.

SANTOS, L.M. **Protótipo para mineração d opinião em Redes Sociais: Estudo de Casos Selecionados Usando o Twitter.** Universidade Federal de Lavras, 2010.

SANTOS, P. L; GRAMINHA, S. S. V. **Problemas emocionais e comportamentais associados ao baixo rendimento acadêmico.** Estudos de Psicologia, Natal, 2006.

SANTOS, Wilian Pereira da Silva. **Análise dos Tweets sobre a Black Friday através da Mineração de Texto e Análise de Sentimentos.** Rio de Janeiro, Brasil, 2016.

SILVA, Luiz Claudio Marini. **Uma Análise Dos Procedimentos De Mineração De Textos No Sgbd Oracle 10g.** Rio de Janeiro, Rio de Janeiro, 2008.

SILVA, N. F. F. d. **Análise de sentimentos em textos curtos provenientes de redes sociais.** Universidade de São Paulo, 2016.

SILVEIRA, GUILHERME PATRICIO . **Mineração De Textos Aplicada A Bases De Dados Jurídicas**. UNIVERSIDADE FEDERAL DO PARANÁ , 2013. Disponível em: <<https://acervodigital.ufpr.br/bitstream/handle/1884/48005/2012-2%20TCC%20Guilherme%20Patricio%20Silveira.pdf?sequence=1&isAllowed=y>>. Acesso em: 1 set. 2018.

SIMOS, G. C. **How Much Data Is Generated Every Minute On Social Media?**. 2015. Disponível em: <<http://wersm.com/how-much-data-is-generated-every-minute-on-social-media/>>. Acesso em: 13 nov. 2018.

SIQUEIRA, A. **Recirculação Jornalística No Twitter: Filtro E Comentário De Notícias Por Interagentes Como Uma Forma De Potencialização Da Circulação**. <http://cencib.org/simposioabciber/PDFs/CC/Gabriela%20da%20Silva%20Zago>. UFVVSF, 2011.

SOARES, Fábio de Azevedo. **Mineração de Textos na Coleta Inteligente de Dados na Web**. Rio de Janeiro, 2008.

STRONGMAN, K. T. **A psicologia da emoção**. 4ª. ed. Lisboa: Climepsi Editore, 1998.

TAN, P; STEINBACH, M; KUMAR, V. **Introdução ao Data Mining: Mineração de Dados..** Rio de Janeiro: Ciência Moderna Ltda, 2009.

TICOM, Antonio Alexandre Mello. **Aplicação Das Técnicas De Mineração De Textos E Sistemas Especialistas Na Liquidação De Processos Trabalhistas .** Rio de Janeiro, Rio de Janeiro, 2007.

TOIT, Jurgens du. **The Bayes Classifier: building a tweet sentiment analysis tool**. 2015. Disponível em: <<http://cloudacademy.com/blog/naive-bayes-classifier>>. Acesso em: 30 nov. 2018.

TORO, R. **Biodanza**. São Paulo: Editora Olavobrás, 2002.

TSYTSARAU, Mikalai; PALPANAS, Themis. **Survey on Mining Subjective Data on the Web**. 2011. 1–6 p.

TURBAN, E; et al. **Tecnologia da Informação para Gestão: Transformando os Negócios na Economia Digital**. 6. ed. Bookman, 2010.

VILAÇA, M.L.C. **Pesquisa e Ensino: Considerações e Reflexões**. Revista E-scrita, v. 1, 2010.

WEISS, S. M; et al. **Text mining: predictive methods for analyzing unstructured information..** Springer Science & Business Media, 2010.

WITTEN, I. H; FRANK, E. **Data Mining: practical machine learning tools and techniques**. Elsevier, 2005.

WITTEN; FRANK; HALL. **DATA MINING. Practical Machine Learning Tools and Techniques**. 3. ed. Morgan Kaufmann, 2011.

WIVES, L. **Tecnologias de descoberta de conhecimento em textos aplicadas à inteligência competitiva**. Exame de Qualificação EQ-069, PPGC-UFRGS, 2002.

ZAGO, G. **II Simpósio ABCiber**. São Paulo, 2008. Disponível em:
<<http://cencib.org/simposioabciber/PDFs/CC/Gabriela%20da%20Silva%20Zago>>.
Acesso em: 29 set. 2018.

APÊNDICE A — Algoritmo de Coleta de Dados Textuais

Algoritmo utilizado para implementar coletar os textos do *Twitter*.

```
1 from TwitterSearch import *
2
3 try:
4     ts = TwitterSearch(
5         consumer_key='[REDACTED]',
6         consumer_secret='[REDACTED]',
7         access_token='[REDACTED]',
8         access_token_secret='[REDACTED]'
9     )
10
11     tso = TwitterSearchOrder()
12     tso.set_keywords(['quero comprar'])
13     tso.set_language('pt')
14
15     arq = open('lista_de_dados_coletados.txt', 'a+', encoding="utf-8")
16     texto = []
17
18     for tweet in ts.search_tweets_iterable(tso):
19         texto.append('%s\n' % tweet['text'])
20         texto.append('\n')
21         arq.writelines(texto)
22     arq.close()
23
24 except TwitterSearchException as e:
25     print(e)
```

APÊNDICE B — Algoritmo de Pré-Processamento

Algoritmo para pré-processar textos armazenados no arquivo de coleta

```

1  import nltk
2
3  basetreinamento = [] # Encontra-se no Anexo A
4
5  stopwordsnltk = nltk.corpus.stopwords.words('portuguese')
6  stopwordsnltk.append('vou') #foi inserido mais esta stopword pois não continha na biblioteca
7  stopwordsnltk.append('tão') #foi inserido mais esta stopword pois não continha na biblioteca
8
9  # função para remoção de stopwords já aplicando o stemmer na frase
10 def aplicastemmer(texto):
11     stemmer = nltk.stem.RSLPStemmer()
12     frasesstemming = []
13     for (palavras, emocao) in texto:
14         comstemming = [str(stemmer.stem(p)) for p in palavras.split() if p not in stopwordsnltk]
15         frasesstemming.append((comstemming, emocao))
16     return frasesstemming
17
18 frasescomstemmingtreinamento = aplicastemmer(basetreinamento)
19
20 # função que lista todas as palavras da base
21 def buscapalavras(frases):
22     todaspalavras = []
23     for (palavras, emocao) in frases:
24         todaspalavras.extend(palavras)
25     return todaspalavras
26
27 palavrastrreinamento = buscapalavras(frasescomstemmingtreinamento)
28
29 # extração das frequências que as palavras aparecem
30 def buscafrequencia(palavras):
31     # a função nltk.FreqDist checa a frequência de cada palavra
32     palavras = nltk.FreqDist(palavras)
33     return palavras
34
35 frequenciatreinamento = buscafrequencia(palavrastrreinamento)
36
37 # função que retorna a lista com palavras únicas sem repetição
38 def buscapalavrasunicas(frequencia):
39     freq = frequencia.keys()
40     return freq
41
42 palavrasunicastreinamento = buscapalavrasunicas(frequenciatreinamento)
43
44 # função para extrair palavras de cada frase
45 # percorrer cada uma das palavras únicas na base de dados de
46 # treinamento e verificar quais delas existe na frase passada por parâmetro
47 def extratorpalavras(documento):
48     doc = set(documento)
49     caracteristicas = {}
50     for palavras in palavrasunicastreinamento:
51         caracteristicas['%s' % palavras] = (palavras in doc)
52     return caracteristicas
53
54 basecompletatreinamento = nltk.classify.apply_features(extratorpalavras, frasescomstemmingtreinamento)
55
56 # classificador de treinamento da entrada de textos baseando na base de dados de treinamento
57 classificador = nltk.NaiveBayesClassifier.train(basecompletatreinamento)

```

```
59 f = open("lista_de_dados_coletados.txt", 'r', encoding="utf-8")
60 texto = f.readlines()
61 x = 0
62 alegria = 0
63 desgosto = 0
64 medo = 0
65 raiva = 0
66 surpresa = 0
67 tristeza = 0
68 testestemming = []
69 stemmer = nltk.stem.RSLPStemmer()
70
71 while x < len(texto):
72     if texto[x] != "\n":
73         texto[x] = texto[x].split('\n')
74         frase = "".join(texto[x])
75         x += 1
76         for (palavrastreinamento) in frase.split():
77             comstem = [p for p in palavrastreinamento.split()]
78             testestemming.append(str(stemmer.stem(comstem[0])))
79             novo = extratorpalavras(testestemming)
80             print("(", frase, ")", classificador.classify(novo), ", ")
81             if classificador.classify(novo) == 'alegria':
82                 alegria +=1
83             if classificador.classify(novo) == 'desgosto':
84                 desgosto +=1
85             if classificador.classify(novo) == 'medo':
86                 medo +=1
87             if classificador.classify(novo) == 'raiva':
88                 raiva +=1
89             if classificador.classify(novo) == 'surpresa':
90                 surpresa +=1
91             if classificador.classify(novo) == 'tristeza':
92                 tristeza +=1
93
94 print("\n--PROCESSADAS", alegria+desgosto+medo+raiva+surpresa+tristeza, "--\n")
95 print("Alegria: ", alegria)
96 print("Desgosto: ", desgosto)
97 print("Medo: ", medo)
98 print("Raiva: ", raiva)
99 print("Surpresa: ", surpresa)
100 print("Tristeza: ", tristeza)
```

APÊNDICE C — Algoritmo de Avaliação da Precisão

Algoritmo para avaliar a precisão e mostrar matriz de confusão.

```

1  import nltk
2
3  basetreinamento = [] # Encontra-se no Anexo A
4  basecoletada = [] # base de dados coletada pré-processada pelo Apêndice A
5
6  stopwordsnltk = nltk.corpus.stopwords.words('portuguese')
7  stopwordsnltk.append('vou')
8  stopwordsnltk.append('tão')
9
10 # função para remoção de stopwords já aplicando o stemmer na frase
11 def aplicastemmer(texto):
12     stemmer = nltk.stem.RSLPStemmer()
13     frasesstemming = []
14     for (palavras, emocao) in texto:
15         comstemming = [str(stemmer.stem(p)) for p in palavras.split() if p not in stopwordsnltk]
16         frasesstemming.append((comstemming, emocao))
17     return frasesstemming
18
19 frasescomstemmingtreinamento = aplicastemmer(basetreinamento)
20 frasescomstemmingcoletada = aplicastemmer(basecoletada)
21
22 # função que lista todas as palavras da base
23 def buscapalavras(frases):
24     todaspalavras = []
25     for (palavras, emocao) in frases:
26         todaspalavras.extend(palavras)
27     return todaspalavras
28
29 palavrastreinamento = buscapalavras(frasescomstemmingtreinamento)
30 palavras_base_coletada = buscapalavras(frasescomstemmingcoletada)
31
32 # extração das frequências que as palavras aparecem
33 def buscafrequencia(palavras):
34     # a função nltk.FreqDist checa a frequência de cada palavra
35     palavras = nltk.FreqDist(palavras)
36     return palavras
37
38 frequenciatreinamento = buscafrequencia(palavrastreinamento)
39 frequencia_frases_coletadas = buscafrequencia(palavras_base_coletada)

```

```

41 # função que retorna a lista com palavras únicas sem repetição
42 def buscapalavrasunicas(frequencia):
43     freq = frequencia.keys()
44     return freq
45
46 palavrasunicastreinoamento = buscapalavrasunicas(frequencia_treinoamento)
47 palavrasunicas_base_coletada = buscapalavrasunicas(frequencia_frases_coletadas)
48
49 # função para extrair palavras de cada frase
50 # percorrer cada uma das palavras únicas na base de dados de
51 # treinoamento e verificar quais delas existe na frase passada por parâmetro
52 def extratorpalavras(documento):
53     doc = set(documento)
54     caracteristicas = {}
55     for palavras in palavrasunicastreinoamento:
56         caracteristicas['%s' % palavras] = (palavras in doc)
57     return caracteristicas
58
59 basecompletatreinoamento = nltk.classify.apply_features(extratorpalavras, frasescomstemmingtreinoamento)
60 basecompletacoletada = nltk.classify.apply_features(extratorpalavras, frasescomstemmingcoletada)
61
62 # classificador de treinoamento da base coletada de textos, baseado na base de dados de treinoamento
63 classificador = nltk.NaiveBayesClassifier.train(basecompletatreinoamento)
64
65 # informa a quantidade de precisão o algoritmo retorna
66 print("Precisão do algoritmo: ", nltk.classify.accuracy(classificador, basecompletacoletada))
67
68 # constroi a tabela de matriz de confusão
69 from nltk.metrics import ConfusionMatrix
70 esperado = []
71 previsto = []
72 for (frase, classe) in basecompletacoletada:
73     resultado = classificador.classify(frase)
74     previsto.append(resultado)
75     esperado.append(classe)
76 matriz = ConfusionMatrix(esperado, previsto)
77 print(matriz)

```

APÊNDICE D — Algoritmo de Comparação de Frases em Arquivos

Algoritmo para comparar base de dados coletada com base de treinamento.

```
1 def procura_palavra():
2     f1 = open("lista_de_dados_coletados.txt.txt", "r", encoding="utf-8")
3     f2 = open("base_treinamento.txt", "r+", encoding="utf-8")
4     for line1 in f1:
5         f2.seek(0)
6         for line2 in f2:
7             if line1 == line2:
8                 print("-----IGUAIS-----")
9                 print("linha1 ->", line1, "linha2 ->", line2)
10                print("iguais\n")
11
12    f1.close()
13    f2.close()
14
15 if __name__ == '__main__':
16    procura_palavra()
```

ANEXO A — Base de Treinamento do Algoritmo

Base de treinamento do algoritmo obtida em um curso de Mineração de Textos do professor Jones Granatyr:

```
basetreinamento = [  
( ' este trabalho e agradável ',' alegria '),  
( ' gosto de ficar no seu aconchego ',' alegria '),  
( ' fiz a adesão ao curso hoje ',' alegria '),  
( ' eu sou admirada por muitos ',' alegria '),  
( ' adoro como você e ',' alegria '),  
( ' adoro seu cabelo macio ',' alegria '),  
( ' adoro a cor dos seus olhos ',' alegria '),  
( ' somo tão amáveis um com o outro ',' alegria '),  
( ' sinto uma grande afeição por ele ',' alegria '),  
( ' quero agradecer meus filhos ',' alegria '),  
( ' me sinto completamente amado ',' alegria '),  
( ' eu amo você ',' alegria '),  
( ' que grande alivio ',' alegria '),  
( ' a dor esta amenizando finalmente ',' alegria '),  
( ' acho que me apaixonei ',' alegria '),  
( ' amar e maravilhoso ',' alegria '),  
( ' estou me sentindo muito animada ',' alegria '),  
( ' me sinto muito bem hoje ',' alegria '),  
( ' como o luar e belo ',' alegria '),  
( ' o dia esta muito bonito ',' alegria '),  
( ' nossa como sou afortunado ',' alegria '),  
( ' as maravilhas do mundo ',' alegria '),  
( ' recebi muito carinho hoje do meus colegas ',' alegria '),  
( ' estou me sentindo reconfortada hoje ',' alegria '),  
( ' e muito bom estar com os amigos ',' alegria '),  
( ' estou muito contente com o resultado dos testes ',' alegria '),  
( ' essa pintura esta bem brilhante ',' alegria '),  
( ' temos água em abundancia ',' alegria '),
```

(' que roupa delicada ',' alegria '),
(' você e um grande comediante ',' alegria '),
(' que bondade a sua em vir aqui ',' alegria '),
(' o amor e lindo ',' alegria '),
(' nossa amizade vai durar para sempre ',' alegria '),
(' estou eufórica com a notícia ',' alegria '),
(' ele e realmente fiel a mim ',' alegria '),
(' vou dar uma grande festa para comemorar meu aniversário ',' alegria '),
(' graças a deus que eu enxerguei o certo ',' alegria '),
(' essa e a melhor escolhas de todas ',' alegria '),
(' o mais incrível e você minha bela ',' alegria '),
(' e tão engraçado tentar explicar ',' alegria '),
(' e emocionante estar neste lugar ',' alegria '),
(' estou cativada pelo seu olhar ',' alegria '),
(' estou loucamente apaixonada ',' alegria '),
(' eu nunca tive duvidas ',' alegria '),
(' estou rodeada pelo seu abraço ',' alegria '),
(' eu vejo estrelas pelo caminho ',' alegria '),
(' eu sinto o sol sempre que você esta por perto ',' alegria '),
(' eu estou sorrindo de orelha a orelha ',' alegria '),
(' isso vale a pena ',' alegria '),
(' finalmente você colocou meu amor em primeiro lugar ',' alegria '),
(' nós dançamos noite adentro ',' alegria '),
(' seu amor e brilhante ',' alegria '),
(' toquei muitos corações durante o meu caminho ',' alegria '),
(' eu serei sua amiga e companheira ',' alegria '),
(' você me traz de volta a vida ',' alegria '),
(' você e como um sonho doce ',' alegria '),
(' adoro este doce de frutas ',' alegria '),
(' meu suco favorito ',' alegria '),
(' estou agradecida pela ajuda ',' alegria '),
(' e um enorme prazer ter você em nossa equipe ',' alegria '),
(' trabalhar em equipe e o melhor ',' alegria '),

(' me sinto flutuando no ar ',' alegria '),
(' a brisa esta agradável hoje ',' alegria '),
(' ótimo e compatível ',' alegria '),
(' somos compatíveis um com o outro ',' alegria '),
(' o órgão e compatível com o paciente ',' alegria '),
(' estou contente fui aceita na faculdade ',' alegria '),
(' fui aprovada no meu exame ',' alegria '),
(' fui beneficiada pela minha empresa ',' alegria '),
(' eu sou muito cativante ',' alegria '),
(' estou contente com o apoio ',' alegria '),
(' como este lugar e confortável ',' alegria '),
(' e bom estar quente neste frio ',' alegria '),
(' um elogio nunca e demais ',' alegria '),
(' vou te chamar para comemorar ',' alegria '),
(' e desejável a sua presença em nossa apresentação ',' alegria '),
(' sou muito grata a você ',' alegria '),
(' me dedico muito naquilo que faço ',' alegria '),
(' estou completamente apaixonada ',' alegria '),
(' vamos agitar essa noite ',' alegria '),
(' você significa muito para mim ',' alegria '),
(' vamos agir sem preconceitos e julgamentos ',' alegria '),
(' finalmente completei a minha coleção, maravilhoso ',' alegria '),
(' eu sou sua rainha ',' alegria '),
(' satisfatoriamente eu anuncio o vencedor dos jogos ',' alegria '),
(' você me atrai facilmente ',' alegria '),
(' aquele rapaz e extremamente atraente ',' alegria '),
(' sinto-me viva ',' alegria '),
(' sinto-me em paz ',' alegria '),
(' estamos tendo muito lucro ',' alegria '),
(' muito bem esta tudo em ordem agora ',' alegria '),
(' podemos arrumar um emprego juntos ',' alegria '),
(' a arrumação esta terminada, que alívio ',' alegria '),
(' o câncer e benigno ',' alegria '),

(' o amor e abundante ',' alegria '),
(' vamos ser caridosos este natal ',' alegria '),
(' com todo esse charme você irá atrair a todos ',' alegria '),
(' nossa como você e charmoso querido ',' alegria '),
(' sou querida pelos meu amigos ',' alegria '),
(' seja cuidadoso com os meus sentimentos ',' alegria '),
(' estou comovido com tamanha caridade ',' alegria '),
(' um chá quente e reconfortante ',' alegria '),
(' que alegria ter vocês aqui ',' alegria '),
(' vamos aplaudir o vencedor ',' alegria '),
(' palmas para a aniversariante ',' alegria '),
(' desejo a você tudo de bom ',' alegria '),
(' hora de apreciar um bom vinho ',' alegria '),
(' aprecio sua presença em minha escola ',' alegria '),
(' anseio por seus próximos trabalhos ',' alegria '),
(' maravilhoso jogo amistoso ',' alegria '),
(' e ótimo que os ânimos tenham se apaziguado ',' alegria '),
(' concretizei finalmente meu sonho ',' alegria '),

(' você e abominável ',' desgosto '),
(' abomino a maneira como você age ',' desgosto '),
(' estou adoentado ',' desgosto '),
(' meu pai esta adoentado ',' desgosto '),
(' estamos todos doentes ',' desgosto '),
(' essa situação e muito amarga ',' desgosto '),
(' disse adeus amargamente ',' desgosto '),
(' tenho antipatia por aquela pessoa ',' desgosto '),
(' como pode ser tão antipática! ',' desgosto '),
(' que horrível seu asqueroso ',' desgosto '),
(' tenho aversão agente como você ',' desgosto '),
(' isso tudo e só chateação ',' desgosto '),
(' estou muito chateada com suas mentiras ',' desgosto '),
(' tão desagradável ',' desgosto '),

(' isso me desagrada completamente ',' desgosto '),
(' te desagrada isso ',' desgosto '),
(' estou com enjôos terríveis ',' desgosto '),
(' todos estão enfermos ',' desgosto '),
(' foi uma enfermidade terrível ',' desgosto '),
(' isso e muito grave ',' desgosto '),
(' não seja tão grosseiro ',' desgosto '),
(' você fez uma manobra ilegal ',' desgosto '),
(' sua indecente, não tem vergonha? ',' desgosto '),
(' você e malvado com as crianças ',' desgosto '),
(' que comentário maldoso ',' desgosto '),
(' sem escrúpulos você manipula a tudo ',' desgosto '),
(' sinto repulsa por você ',' desgosto '),
(' e repulsivo a maneira como olha para as pessoas ',' desgosto '),
(' estou indisposta ',' desgosto '),
(' a indisposição me atacou hoje ',' desgosto '),
(' acho que vou vomitar ',' desgosto '),
(' tem muito vomito lá ',' desgosto '),
(' que incomodo essa dor ',' desgosto '),
(' não me incomode nunca mais ',' desgosto '),
(' suas bobagens estão nos incomodando ',' desgosto '),
(' que nojo olha toda essa sujeira ',' desgosto '),
(' como isso está sujo ',' desgosto '),
(' tenho náuseas só de lembrar ',' desgosto '),
(' me sinto nauseada com o cheiro desta comida ',' desgosto '),
(' você esta obstruindo a passagem de ar ',' desgosto '),
(' você esta terrivelmente doente ',' desgosto '),
(' olhe que feia esta roupa ',' desgosto '),
(' que atitude deplorável ',' desgosto '),
(' nossa como você e feio ',' desgosto '),
(' muito mau tudo isso ',' desgosto '),
(' estou desgostoso com você ',' desgosto '),
(' você cortou o meu assunto ',' desgosto '),

(' para que tanta chateação? ', ' desgosto '),
(' esse perfume e enjoativo ', ' desgosto '),
(' ser perigoso não nada bom ', ' desgosto '),
(' você e perigoso demais para minha filhas ', ' desgosto '),
(' que fedido este esgoto ', ' desgosto '),
(' que fedido você esta ', ' desgosto '),
(' que cachorro malcheiroso ', ' desgosto '),
(' hora que ultraje ', ' desgosto '),
(' e ultrajante da sua parte ', ' desgosto '),
(' situação desagradável essa ', ' desgosto '),
(' você só me da desgosto ', ' desgosto '),
(' tenho aversão a pessoas assim ', ' desgosto '),
(' antipatia e um mal da sociedade ', ' desgosto '),
(' que criatura abominável ', ' desgosto '),
(' e depressiva a maneira como você vê o mundo ', ' desgosto '),
(' me desagrada sua presença na festa ', ' desgosto '),
(' sinto asco dessa coisa ', ' desgosto '),
(' que hediondo! ', ' desgosto '),
(' vou golfar o cafe fora ', ' desgosto '),
(' hora que garota detestável! ', ' desgosto '),
(' estou nauseada ', ' desgosto '),
(' isso que você disse foi muito grave ', ' desgosto '),
(' não seja obsceno na frente das crianças ', ' desgosto '),
(' não seja rude com as visitas ', ' desgosto '),
(' esse assunto me da repulsa ', ' desgosto '),
(' que criança terrivelmente travessa ', ' desgosto '),
(' que criança mal educada ', ' desgosto '),
(' estou indisposta te dar o divorcio ', ' desgosto '),
(' tão patetico, não tem nada mais rude para dizer? ', ' desgosto '),
(' por motivo torpe, com emprego de meio cruel e com impossibilidade de defesa para a vítima ', ' desgosto '),
(' a inveja e tão vil e vergonhosa que ninguém se atreve a confessá-la ', ' desgosto '),

(' o miserável receio de ser sentimental e o mais vil de todos os receios modernos ',' desgosto '),

(' travesso gato quando fica com saudades do dono mijando no sapato ',' desgosto '),

(' isso é um ato detestável e covarde ',' desgosto '),

(' revelam apenas o que é destrutivo e detestável para o povo ',' desgosto '),

(' não sei como é a vida de um patife, mais a de um homem honesto e abominável ',' desgosto '),

(' há coisas que temos que suportar para não acharmos a vida insuportável ',' desgosto '),

(' as injúrias do tempo e as injustiças do homem ',' desgosto '),

(' odioso e desumano ',' desgosto '),

(' você não publicará conteúdo odioso, pornográfico ou ameaçador ',' desgosto '),

(' rancoroso e reprimido ',' desgosto '),

(' não há animal mais degradante, estúpido, covarde, lamentável, egoísta, rancoroso e invejoso do que o homem ',' desgosto '),

(' o virulento debate entre políticos ',' desgosto '),

(' eu imploro, não me matem! ',' medo '),

(' tem certeza que não é perigoso? ',' medo '),

(' não tenho certeza se é seguro ',' medo '),

(' tenho que correr pra não me pegarem ',' medo '),

(' socorro! ele queria roubar os meus doces! ',' medo '),

(' esse cara está me perseguindo ',' medo '),

(' não entro lá, é um lugar muito perigoso ',' medo '),

(' este lugar continua assustador ',' medo '),

(' na selva tem muitos animais perigosos ',' medo '),

(' avancem com cautela ',' medo '),

(' este lugar está silencioso de mais, cuidado! ',' medo '),

(' por favor, deixe-me viver! ',' medo '),

(' vou ficar sem mesada se tirar nota baixa ',' medo '),

(' parece que tem olhos nos vigiando ',' medo '),

(' eu temo que a sentença do juiz possa ser negativa ',' medo '),
(' mas essa missão e arriscada ',' medo '),
(' salvem-se quem puder! ',' medo '),
(' meu plano pode ser descoberto ',' medo '),
(' não tive culpa, juro não fui eu ',' medo '),
(' tenho que tomar cuidado com o lobisomem ',' medo '),
(' se eu não achar, ele vai descobrir a verdade ',' medo '),
(' meu deus, ele desapareceu! ',' medo '),
(' tomara que eles não me vejam daqui! ',' medo '),
(' mantenha isso em segredo, se descobrirem estaremos ferrados ',' medo '),
(' por favor, me soltem, eu sou inocente ',' medo '),
(' estou ouvindo passos atrás de mim ',' medo '),
(' eu vou pedir socorro! ',' medo '),
(' cuidado com as curvas na estrada ',' medo '),
(' não sei não, parece perigoso ',' medo '),
(' estou tremendo de medo! ',' medo '),
(' socorro, eu vou cair! ',' medo '),
(' eu não vou ate a floresta negra, e muito perigoso ',' medo '),
(' ouço passos na minha direção ',' medo '),
(' acho que está arriscado de mais ',' medo '),
(' vamos voltar, e muito perigoso ',' medo '),
(' fuja, se não acabaremos mortos ',' medo '),
(' receio por não me livrar desta situação ',' medo '),
(' socorro! ele está armado! ',' medo '),
(' ei cuidado, você vai bater no poste! ',' medo '),
(' socorro, nós estamos afundando ',' medo '),
(' e serio, cuidado com essa arma! ',' medo '),
(' os tubarões estão atacando! ',' medo '),
(' sinto arrepios quando fico sozinho no escuro ',' medo '),
(' calma, eu não estou com o dinheiro ',' medo '),
(' eu acho que estou sendo enganado ',' medo '),
(' ligeiro, temos que fugir depressa ',' medo '),
(' tem um crocodilo selvagem vindo para cá ',' medo '),

(' se ficarmos quietos eles não vão nos achar ',' medo '),
(' fuja! o tigre parece faminto ',' medo '),
(' estou sem saída, preciso de um milagre ',' medo '),
(' tire isso de mim! socorro! ',' medo '),
(' não sei nadar, vou me afogar! ',' medo '),
(' não tenho certeza se é seguro ',' medo '),
(' vou apanhar se meus pais verem meu boletim ',' medo '),
(' não consigo sair daqui! ',' medo '),
(' se sair tão tarde, poderei ser assaltada ',' medo '),
(' não me deixe por favor! ',' medo '),
(' espere, não pode me largar aqui sozinho ',' medo '),
(' temo pela sua segurança ',' medo '),
(' eu te entrego o dinheiro, por favor não me mate! ',' medo '),
(' ele vai levar todo o meu dinheiro ',' medo '),
(' não dirija tão rápido assim ',' medo '),
(' me descobriram, irão me prender! ',' medo '),
(' só espero que não me façam nenhum mal ',' medo '),
(' vou me afogar, me ajudem a sair da água ',' medo '),
(' não estaremos a salvo aqui ',' medo '),
(' não quero nem pensar no que pode acontecer ',' medo '),
(' nessa cidade é uma desgraça atrás da outra ',' medo '),
(' alguém está me ligando, estou assustado ',' medo '),
(' isso não é remédio, não me matem ',' medo '),
(' eu não confio nele, tenho que ter cautela ',' medo '),
(' muita cautela ',' medo '),
(' vou ser descoberto, meu deus ',' medo '),
(' receio que terei de ir ',' medo '),
(' a noite é muito perigosa ',' medo '),
(' estou estremecendo com essa casa ',' medo '),
(' olha aquela criatura se movendo monstruosamente ',' medo '),
(' não agüento este suspense ',' medo '),
(' afugente os cães ',' medo '),
(' estou chocado e amedrontado com este assassinato brutal ',' medo '),

(' e preciso afugenta com ímpeto este medo do inferno ',' medo '),
(' seu políticos usam suas forças para afugentar e amedrontar o povo ',' medo
) ,
(' o objetivo disso e apenas me amedrontar mais ',' medo '),
(' isso me apavora ',' medo '),

(' ele a feriu profundamente ',' raiva '),
(' vou despejar minha cólera em você ',' raiva '),
(' me sinto atormentado ',' raiva '),
(' não me contrarie ',' raiva '),
(' vou destruir tudo que foi construído ',' raiva '),
(' não consigo terminar este trabalho, e muito frustrante ',' raiva '),
(' me frustra a sua presença aqui ',' raiva '),
(' esta comida me parece muito ruim ',' raiva '),
(' você me destrói ',' raiva '),
(' estamos separados ',' raiva '),
(' estou odiando este vestido ',' raiva '),
(' não pude comprar meu celular hoje ',' raiva '),
(' ela e uma garota ruim ',' raiva '),
(' estivemos em um show horroroso ',' raiva '),
(' o ingresso estava muito caro ',' raiva '),
(' se eu estragar tudo vai por água a baixo ',' raiva '),
(' não possuo dinheiro algum ',' raiva '),
(' sou muito pobre ',' raiva '),
(' vai prejudicar a todos esta nova medida ',' raiva '),
(' ficou ridículo ',' raiva '),
(' este sapato esta muito apertado ',' raiva '),
(' a musica e uma ofensa aos meus ouvidos ',' raiva '),
(' não consigo terminar uma tarefa muito difícil ',' raiva '),
(' reprovei em minha graduação ',' raiva '),
(' estou muito chateado com tudo ',' raiva '),
(' eu odeio em você ',' raiva '),
(' e um desprazer conhecê-lo ',' raiva '),

(' estou desperdiçando minhas férias ',' raiva '),
(' e muito ruim este jogo ',' raiva '),
(' vamos ter muito rancor pela frente ',' raiva '),
(' não achei que seria tão terrível ',' raiva '),
(' vou vetar o orçamento ao cliente ',' raiva '),
(' meus pais não consentiram nosso casamento ',' raiva '),
(' eu odiei este perfume ',' raiva '),
(' seu descaso é frustrante ',' raiva '),
(' me sinto completamente amarga ',' raiva '),
(' desprezo muito o seu trabalho ',' raiva '),
(' estamos descontentes por nossa família ',' raiva '),
(' vou infernizar a sua empresa ',' raiva '),
(' estou furioso com estes valores ',' raiva '),
(' obrigaram o rapaz a sair ',' raiva '),
(' como ele pode deixar de lado? ',' raiva '),
(' são apenas injúrias sobre mim ',' raiva '),
(' estou enfurecido com a situação dessa empresa ',' raiva '),
(' estou com o diabo no corpo ',' raiva '),
(' isso foi diabólico ',' raiva '),
(' tenho aversão à gente chata ',' raiva '),
(' não vou perdoar sua traição ',' raiva '),
(' esse dinheiro sujo e corrupto ',' raiva '),
(' eles me crucificam o tempo todo ',' raiva '),
(' eu vou enlouquecer com todo este barulho ',' raiva '),
(' não agüento todo esse assédio ',' raiva '),
(' cólera do dragão ',' raiva '),
(' isso é ridículo! ',' raiva '),
(' da próxima vez, vou inventar tudo sozinho ',' raiva '),
(' seus tolos! deixaram ele escapar! ',' raiva '),
(' jamais te perdorei ',' raiva '),
(' o que é isso? outra multa ',' raiva '),
(' você passou dos limites! ',' raiva '),
(' sente-se e cale a boca ',' raiva '),

(' ingratosvermesvocês me pagam! ',' raiva '),
(' saiam da dai, se não arranco vocês dai! ',' raiva '),
(' você já me causou problemas suficientes ',' raiva '),
(' isso foi a gota d'agua ',' raiva '),
(' o que você tem com isso? ',' raiva '),
(' não vejo a hora de me livrar de você ',' raiva '),
(' já entendi a jogada seus safados! ',' raiva '),
(' você não merece piedade ',' raiva '),
(' saia de perto de mim ',' raiva '),
(' suma daqui, ou arranco seu couro! ',' raiva '),
(' estou revoltado com essa situação ',' raiva '),
(' seu idiota! ',' raiva '),
(' não, eu não vou te emprestar dinheiro! ',' raiva '),
(' você não passa de um cafajeste! vai embora ',' raiva '),
(' pare de frescura e vá trabalhar ',' raiva '),
(' eles merecem uma lição ',' raiva '),
(' ainda estou muito bravo com você ',' raiva '),
(' eu preciso surrar aquela chantagista ',' raiva '),
(' olha o que você fez! derramou! ',' raiva '),
(' você está pedindo pra apanhar! ',' raiva '),
(' me deixa em paz! ',' raiva '),
(' morra maldito, morra! ',' raiva '),
(' você e mais irritante de perto ',' raiva '),
(' e bom fechar o bico ',' raiva '),

(' magicamente você me surpreendeu ',' surpresa '),
(' e imenso esse globo ',' surpresa '),
(' isso e tremendamente interessante ',' surpresa '),
(' meu bilhete for sorteado, inacreditável! ',' surpresa '),
(' um assalto a mão armada! ',' surpresa '),
(' incrível, cabe em qualquer lugar! ',' surpresa '),
(' você por aqui? ',' surpresa '),
(' não dá pra acreditar no que ela me contou ',' surpresa '),

(' os convidados já estão chegando! ',' surpresa '),
(' puxa vida! nunca nos livramos de alguém tão depressa ',' surpresa '),
(' micha carteira sumiu, eu estava com ela na mão ',' surpresa '),
(' oh! um disco voador ',' surpresa '),
(' amigos, que bela surpresa! ',' surpresa '),
(' nunca pensei que veria isso e perto ',' surpresa '),
(' nem acredito que comi tanto ',' surpresa '),
(' não acredito que veio me ver ',' surpresa '),
(' não acredito que e tão descarado ',' surpresa '),
(' me surpreende sua falta de tato ',' surpresa '),
(' o prédio onde eles moravam desabou! ',' surpresa '),
(' inacreditável um bolo tão grande ',' surpresa '),
(' e serio mesmo? não dá pra acreditar ',' surpresa '),
(' como assim não vai ao nosso encontro? ',' surpresa '),
(' como assim não tem ninguém em casa? ',' surpresa '),
(' ue, mas para onde ele foi?! ',' surpresa '),
(' por essa eu não esperava ',' surpresa '),
(' nossa, olha só que mergulho ',' surpresa '),
(' minha esposa está grávida! ',' surpresa '),
(' meu dinheiro sumiu! ',' surpresa '),
(' e verdade que os dois terminaram?!? ',' surpresa '),
(' caramba, nem vi você chegar ',' surpresa '),
(' nossa, como pode alguém cozinhar tão mal? ',' surpresa '),
(' nossa que incrível ',' surpresa '),
(' a fórmula sumiu! ',' surpresa '),
(' eu nem acredito que já estou terminando o curso ',' surpresa '),
(' não acredito que esta aqui comigo novamente ',' surpresa '),
(' está escondendo algo de nós! ',' surpresa '),
(' como assim, ainda não terminou a tarefa ',' surpresa '),
(' pensei que já estivesse pronta! ',' surpresa '),
(' opa! quem apagou a luz? ',' surpresa '),
(' caramba! aonde vai tão rápido? ',' surpresa '),
(' estamos seguindo o caminho errado! ',' surpresa '),

(' quatro reais o litro da gasolina! ',' surpresa '),
(' me assustei ao vê-lo desse jeito! ',' surpresa '),
(' minha mãe está grávida, acredita nisso? ',' surpresa '),
(' parece mentira você ter crescido tanto ',' surpresa '),
(' me surpreende sua imaginação ',' surpresa '),
(' suas roupas são realmente lindas ',' surpresa '),
(' com consegue ser tão bela? ',' surpresa '),
(' essa e realmente uma casa deslumbrante ',' surpresa '),
(' superou minhas expectativas ',' surpresa '),
(' e admirável a maneira como se comporta ',' surpresa '),
(' isso e realmente chocante ',' surpresa '),
(' algumas noticias me surpreenderam no noticiário ',' surpresa '),
(' surpreendente sua festa ',' surpresa '),
(' estou tremendo de alegria ',' surpresa '),
(' chocou grande parte do mundo ',' surpresa '),
(' eu ficaria muito espantado com a sua vinda ',' surpresa '),
(' ele e admirável ',' surpresa '),
(' sua beleza me surpreendeu ',' surpresa '),
(' seus olhos são surpreendentemente verdes ',' surpresa '),
(' os políticos se surpreendem quando alguém acredita neles ',' surpresa '),
(' estou perplexa com essas denúncias ',' surpresa '),
(' fiquei perplexo com suas palavras ',' surpresa '),
(' estou abismado com sua prosa ',' surpresa '),
(' eu ficaria realmente abismado se me dissessem isso ',' surpresa '),
(' o grupo foi surpreendido enquanto lavava o carro ',' surpresa '),
(' estou boquiaberto com as imagens ',' surpresa '),
(' estou boquiaberto com essas suas palavras ',' surpresa '),
(' esse quadro e maravilhoso ',' surpresa '),
(' este carro me deixou maravilhado ',' surpresa '),
(' estou maravilhada ',' surpresa '),
(' essa expectativa esta me matando ',' surpresa '),
(' vou caminhar sempre na expectativa de encontrá-lo ',' surpresa '),
(' você emudece minhas palavras ',' surpresa '),

(' minhas palavras vão emudecer se não parar de me surpreender ',' surpresa '),

(' a mulher e um efeito deslumbrante da natureza ',' surpresa '),
 (' estou deslumbrada com essas jóias ',' surpresa '),
 (' isso é romântico e deslumbrante ',' surpresa '),
 (' isso pode ser surpreendentemente deslumbrante ',' surpresa '),
 (' trabalho deslumbrante ',' surpresa '),
 (' essas pessoas são esplêndidas ',' surpresa '),
 (' e esplêndido como o céu se encontra no momento ',' surpresa '),
 (' e um carro fantástico ',' surpresa '),
 (' um edifício realmente fantástico ',' surpresa '),

(' por favor não me abandone ',' tristeza '),
 (' não quero ficar sozinha ',' tristeza '),
 (' não me deixe sozinha ',' tristeza '),
 (' estou abatida ',' tristeza '),
 (' ele está todo abatido ',' tristeza '),
 (' tão triste suas palavras ',' tristeza '),
 (' seu amor não é mais meu ',' tristeza '),
 (' estou aborrecida ',' tristeza '),
 (' isso vai me aborrecer ',' tristeza '),
 (' estou com muita aflição ',' tristeza '),
 (' me aflige o modo como fala ',' tristeza '),
 (' estou em agonia com meu íntimo ',' tristeza '),
 (' não quero fazer nada ',' tristeza '),
 (' me sinto ansiosa e tensa ',' tristeza '),
 (' não consigo parar de chorar ',' tristeza '),
 (' não consigo segurar as lágrimas ',' tristeza '),
 (' e muita dor perder um ente querido ',' tristeza '),
 (' estou realmente arrependida ',' tristeza '),
 (' acho que o carma volta, pois agora sou eu quem sofro ',' tristeza '),
 (' você não cumpriu suas promessas ',' tristeza '),
 (' me sinto amargurada ',' tristeza '),

(' coitado esta tão triste ',' tristeza '),
(' já e tarde de mais ',' tristeza '),
(' nosso amor acabou ',' tristeza '),
(' essa noite machuca só para mim ',' tristeza '),
(' eu não estou mais no seu coração ',' tristeza '),
(' você mudou comigo ',' tristeza '),
(' quando eu penso em você realmente dói ',' tristeza '),
(' como se fosse nada você vê minhas lágrimas ',' tristeza '),
(' você disse cruelmente que não se arrependeu ',' tristeza '),
(' eu nunca mais vou te ver ',' tristeza '),
(' ela esta com depressão ',' tristeza '),
(' a depressão aflige as pessoas ',' tristeza '),
(' estar depressivo e muito ruim ',' tristeza '),
(' estou derrotada e deprimida depois deste dia ',' tristeza '),
(' e comovente te ver dessa maneira ',' tristeza '),
(' e comovente ver o que os filhos do brasil passam ',' tristeza '),
(' como me sinto culpada ',' tristeza '),
(' estou abatida ',' tristeza '),
(' a ansiedade tomou conta de mim ',' tristeza '),
(' as pessoas não gostam do meu jeito ',' tristeza '),
(' adeus passamos bons momentos juntos ',' tristeza '),
(' sinto sua falta ',' tristeza '),
(' ele não gostou da minha comida ',' tristeza '),
(' estou sem dinheiro para a comida ',' tristeza '),
(' queria que fosse o ultimo dia da minha vida ',' tristeza '),
(' você está com vergonha de mim ',' tristeza '),
(' ela não aceitou a minha proposta ',' tristeza '),
(' era o meu ultimo centavo ',' tristeza '),
(' reprovei de ano na faculdade ',' tristeza '),
(' afinal você só sabe me desfazer ',' tristeza '),
(' eu falhei em tudo nessa vida ',' tristeza '),
(' eu fui muito humilhado ',' tristeza '),
(' e uma história muito triste ',' tristeza '),

('ninguém acredita em mim ','tristeza '),
(' eu não sirvo para nada mesmo ','tristeza '),
(' droga, não faço nada direito ','tristeza '),
(' sofrimento em dobro na minha vida ','tristeza '),
(' fui demitida essa semana ','tristeza '),
(' as crianças sofrem ainda mais que os adultos ','tristeza '),
(' pra mim um dia é ruim, o outro é pior ','tristeza '),
(' de repente perdi o apetite ','tristeza '),
(' oh que dia infeliz ','tristeza '),
(' estamos afundados em contas ','tristeza '),
(' nem um milagre pode nos salvar ','tristeza '),
(' só me resta a esperança ','tristeza '),
(' pior que isso não pode ficar ','tristeza '),
(' meu salário é baixo ','tristeza '),
(' não passei no vestibular ','tristeza '),
(' ninguém se importa comigo ','tristeza '),
(' ninguém lembrou do meu aniversário ','tristeza '),
(' tenho tanto azar ','tristeza '),
(' o gosto da vingança é amargo ','tristeza '),
(' sou uma mulher amargurada depois de que você me deixou ','tristeza '),
(' estou desanimada com a vida ','tristeza '),
(' e um desânimo só coitadinha ','tristeza '),
(' a derrota é depressiva ','tristeza '),
(' discriminar é desumano ','tristeza '),
(' que desânimo ','tristeza '),
(' e uma desonra para o país ','tristeza '),
(' a preocupação deveria nos levar à ação não à depressão ','tristeza '),
(' passamos ao desalento e à loucura ','tristeza '),
(' aquele que nunca viu a tristeza nunca reconhecerá a alegria ','tristeza '),

(' cuidado com a tristeza ela é um vício ','tristeza')]