UNIVERSIDADE FEDERAL DO CEARÁ

CAMPUS DE QUIXADÁ

CURSO DE ENGENHARIA DE SOFTWARE

LUIZ ISAIAS DE SOUZA SILVA

DETECÇÃO DE COMENTÁRIOS PEJORATIVOS EM VÍDEOS INFANTO-JUVENIS DO YOUTUBE

QUIXADÁ

2018

LUIZ ISAIAS DE SOUZA SILVA

DETECÇÃO DE COMENTÁRIOS PEJORATIVOS EM VÍDEOS INFANTO-JUVENIS DO YOUTUBE

Monografia apresentada ao Curso de Engenharia de Software do Campus Quixadá da Universidade Federal do Ceará, como requisito parcial para obtenção do Título de Bacharel em Engenharia de Software.
Área de concentração: Computação

Orientadora: Profª. Dra. Ticiana Coelho da Silva

QUIXADÁ

2018

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

S581d Silva, Luiz Isaias de Souza.
Detecção de comentários pejorativos em vídeos infanto-juvenis do Youtube / Luiz Isaias de Souza Silva. – 2018.
54 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Quixadá, Curso de Engenharia de Software, Quixadá, 2018.
Orientação: Profa. Dra. Ticiana Coelho da Silva.

1. Crianças-Adolescentes-Proteção. 2. Sistemas de Recuperação da Informação. 3. Mineração de Dados (Computação). 4. Software-Desenvolvimento. I. Título.

CDD 005.1

LUIZ ISAIAS DE SOUZA SILVA

DETECÇÃO DE COMENTÁRIOS PEJORATIVOS EM VÍDEOS INFANTO-JUVENIS DO YOUTUBE

Monografia apresentada ao Curso de Engenharia de Software do Campus Quixadá da Universidade Federal do Ceará, como requisito parcial para obtenção do Título de Bacharel em Engenharia de Software.
Área de concentração: Computação

Aprovada em: __/__/____

BANCA EXAMINADORA

Profª. Dra. Ticiana Coelho da Silva
(Orientadora)
Universidade Federal do Ceará (UFC)

Prof. Msc. Regis Pires Magalhães
Universidade Federal do Ceará (UFC)

Prof. Dra. Carla Ilane Moreira Bezerra
Universidade Federal do Ceará (UFC)

Ao meu pai Luiz Claudio.

À Minha mãe Pergentina Rodrigues.

À todos da minha família que me deram apoio e acreditaram.

Aos meus amigos de longa data em Fortaleza, aos novos amigos que fiz em Quixadá, e à todos que virão.

AGRADECIMENTOS

Primeiramente gostaria de dedicar esse trabalho à minha Mãe, por ter dado seu sangue para que eu pudesse chegar onde cheguei, me provendo uma educação básica de qualidade, cursos e preparação para o futuro. Apesar de nossas diferenças em pensamento e atitude, saiba que eu te amo do fundo do meu coração, e eu jamais estaria aqui se não fosse por você. Quero sempre ter motivos para continuar te orgulhando.

Ao meu Pai que sempre me orientou e me educou sobre a vida. Me ajudou a descobrir minha grande paixão, que é programação, me mostrou o que é ter de "se virar", e me mostrou o quanto é bom ter liberdade e caminhar com as próprias pernas. Você é meu modelo, principalmente para as atitudes boas e nobres. Eu jamais teria tido a força e a coragem para sair de casa, se não fosse por você. Agora o céu é o limite!

À todos da minha família que me apoiaram, em especial meus padrinhos e meus irmãos e minha avó D. Teresa.

Um "muito obrigado!" especial para a pessoa que me aturou, mesmo com minhas crises e inconstâncias, e me mostrou o que é amar e ser amado durante os últimos 4 anos, Letícia Aguiar, você foi e sempre será alguém especial, e apesar de termos de seguir caminhos diferentes, saiba que o que sinto é verdadeiro e apesar de não ser imortal, visto que é chama, é eterno enquanto durar.

Aos meus amigos que ficaram em Fortaleza e à todos os amigos que fiz em Quixadá, em especial Caio Melo, Gabriel Jorge, Ana Carmélia, Gustavo Carneiro, Sérgio Gadelha, Matheus Rios, Bruna Kelvyla, Adson Rodrigues, Emerson Vieira e Felipe Souza R.

Eu acredito que uma pessoa pode sempre evoluir e chegar numa versão melhor de si mesmo. Basta crer em suas atitudes e na sua capacidade, e continuar buscando.

Em memória do meu eterno melhor amigo, meu dog, Gohan.

"A gente destrói aquilo que mais ama
em campo aberto, ou numa emboscada;
alguns com a leveza do carinho
outros com a dureza da palavra;
os covardes destroem com um beijo
os valentes destroem com a espada.
(Oscar Wilde, A Balada do Cárcere de Reading)

RESUMO

A Internet, rede mundial de computadores, conecta cada vez mais pessoas ao redor do mundo. Em um dos sites mais acessados do mundo, a plataforma de vídeos Youtube, é notável a participação ativa de jovens e crianças nos comentários dos vídeos. Também é notável a quantidade de ofensas direcionadas aos usuários, nesses comentários. Quando os vídeos são direcionados às crianças e adolescentes, isso pode influenciar negativamente os jovens e ferir o Estatuto da Criança e do Adolescente brasileiro que afirma: "Deve ser respeitado a integridade física, moral e psíquica da criança e do adolescente". Com o objetivo de detectar se um vídeo no Youtube, direcionado aos jovens, possui muitos comentários negativos desenvolvemos um classificador textual Naïve Bayes e uma extensão para o navegador Google Chrome, que indica a taxa de comentários pejorativos contidos no vídeo. Para analisar comentários de vídeos do Youtube e montar o modelo de classificação textual, foram utilizadas ferramentas de mineração e classificação de dados NLTK, scikit-learn e SentiStrength. Para auxiliar e automatizar os passos de coleta e extração dos comentários de vídeos do Youtube, o autor desenvolveu vários scripts na linguagem Python. Uma extensão para Google Chrome foi desenvolvida, utilizando tecnologias web, para aproveitar os dados de classificação gerados, e facilitar a realização de novas classificações. Ao total, 87.094 comentários foram coletados, e 62.121 foram analisados, apresentando 48.773 comentários considerados neutros e 13.348 comentários considerados pejorativos. Observou-se que vídeos direcionados à crianças não possuem altas taxas de comentários negativos, e por outro lado, vídeos direcionados a jovens, possuem uma taxa de 34% de comentários pejorativos, ou seja vídeos direcionados à adolescentes possuem mais termos ofensivos em seus comentários.

Palavras-chaves: Sistema de Recuperação da Informação, Mineração de Dados, Proteção de Crianças e Adolescentes.

ABSTRACT

Internet or World Wide Web is connecting more and more people each day. In Youtube, a video streaming platform and one of Internet's most popular websites, is noticeable how teenagers and kids are actively engaged in the comments section. Also noticeable are the offenses in these comments, directed to the users. When the videos are targeted to children and teenagers audience, this may negatively influence those younger ones, and it goes against Brazil's "Estatuto da Criança e do Adolescente", a constitutional law that protects children, as it says: "It should be respected the physical, moral and psychical integrity of children and teenagers alike." With the goal to detect if a Youtube video, targeted at kids and teenagers, has lots of offensive comments, it was built a Naïve Bayes textual classifier and a Google Chrome Browser Extension, which indicates the rates of offensive comments for the video. To analyze the Youtube video comments and to build the text classification model, it was used data classification and data mining tools: NLTK, scikit-learn, and Sentistrength. To help and automate the data collection and extraction steps, many scripts in Python were developed. Finally, a Google Chrome Extension was built, using web technologies, to display the resultant data for the public and to make easier to have new classifications. 87.094 comments were collected, and 62.121 were analyzed and classified. Showing that 48.773 comments could be considered neutral and 13.348 could be considered offensive. It was noted that videos targeted at children did not have high offensive comments rate, while videos targeted at teenagers had a considerable rate, about 34% for offensive comments. This means that videos targeted at teenagers tend to have more offensive comments than videos targeted at children.

Key-words: Information Retrieval System, Data Mining, Child and Teenagers Safeguarding.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

A internet é uma rede mundial que interliga milhões de computadores em todo o mundo, servindo como um grande fator de comunicação e integração social (FALCÃO, 2015). Todos os dias, novas páginas na internet surgem, sejam de relacionamento, humor, entretenimento, notícias, ou redes sociais, como exemplo: Facebook, Twitter, Instagram, Google+, Vimeo, DailyMotion e Youtube.

É pertinente que em nossa sociedade contemporânea, as pessoas estejam mais próximas da tecnologia, principalmente as crianças. Essas crianças adquirem habilidades diferentes das crianças de antigamente: enquanto uma criança da década de 80 possuía uma maior facilidade para construir ou modelar um brinquedo, as crianças da geração atual possuem habilidades para lidar com a informática, devido ao convívio rotineiro com a mesma (FALCÃO *et al.*, 2016).

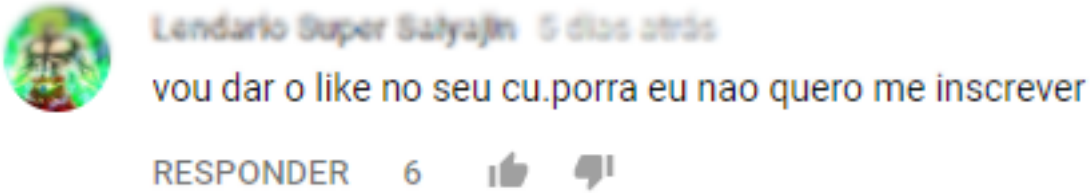
A plataforma de vídeos Youtube é o segundo site mais acessado do mundo (ALEXA, 2017). Em tal plataforma, é possível postar e compartilhar vídeos com uma enorme e abrangente gama de conteúdos, servindo também como site de buscas. É possível encontrar conteúdo musical, *vlogs*, animações, curta-metragens, conteúdo educativo, conteúdo com público alvo adulto e conteúdo voltado especificamente ao público jovem e infantil, como por exemplo os canais **Felipe Neto** [1] e **Galinha Pintadinha** [2].

Como parte de suas características de rede social, o Youtube disponibiliza, em seus vídeos, a possibilidade de usuários comentarem. Porém, tendo em vista as crianças e adolescentes estão conectadas na rede cada vez mais cedo (OLIVEIRA *et al.*, 2017), ao navegar pela plataforma em vídeos infantis, nota-se que há uma certa liberdade para realizar comentários de qualquer natureza, até mesmo ofensivos ou inapropriados para o público o qual o vídeo é destinado, um exemplo disso pode ser visto na Figura 1.

Segundo o artigo 17 do **Estatuto da Criança e do Adolescente** (ECA): "O direito ao respeito consiste na inviolabilidade da integridade física, psíquica e moral da criança e do adolescente, abrangendo a preservação da imagem, da identidade, da autonomia, dos valores, ideias e crenças, dos espaços e objetos pessoais"(ECA, 1990). Assim, é possível através da tecnologia buscar meios para proteger moralmente e psiquicamente o público mais jovem que utiliza a plataforma.

[1] https://www.youtube.com/channel/UCV306eHqgo0LvBf3Mh36AHg
[2] https://www.youtube.com/channel/UCBAb_DK4GYZqZR9MFA7y2Xg

Figura 1 – Comentário em um vídeo infantil no Youtube



Fonte: Youtube - 13 de Setembro de 2017

A Figura 1 foi retirada do vídeo **UPA CAVALINHO - Clipe Música Oficial - Galinha Pintadinha DVD 4**, um vídeo direcionado a crianças de 2 a 7 anos postado pelo canal *Galinha Pintadinha*, que é direcionado ao público infantil. Por questões de anonimato o usuário não é identificado.

O objetivo deste trabalho é identificar comentários pejorativos semelhantes ao apresentado na Figura 1 em vídeos postados no Youtube, direcionados a crianças e adolescentes. Para alcançar este objetivo, primeiramente foi criado um *script* na linguagem *Python*, para coleta de comentários dos vídeos através da API do Youtube. Outro *script Python* também foi utilizado para realizar o pré-processamento dos comentários, através da biblioteca de processamento de linguagem natural **NLTK**. Com o *dataset* de comentários coletados e pré-processados, foi criado um modelo de classificação utilizando as ferramentas **SentiStrength** e **scikit-learn**. A **scikit-learn** também foi utilizada para avaliar a eficiência do modelo criado. A utilização das ferramentas é descrita de forma detalhada no Capítulo 5.

Utilizando tecnologias web (HTML, CSS, Javascript), foi desenvolvido também uma extensão para o navegador da web **Google Chrome**. A extensão exibe os resultados da classificação previamente realizada, para o vídeo sendo assistido naquele momento. Os dados da classificação ficam salvos em um banco de dados.

Este trabalho está estruturado da seguinte forma. No Capítulo 2, são apresentados os objetivos do trabalho; no Capítulo 3, são apresentados os principais conceitos que são utilizados durante o projeto; no Capítulo 4, serão apresentados os principais trabalhos relacionados, que tratam dos temas de mineração e de proteção de crianças online; o Capítulo 5 apresenta a metodologia de trabalho; o Capítulo 6 apresenta os resultados obtidos pelo modelo de classificação textual; O Capítulo 7 apresenta uma conclusão do autor sobre o trabalho e seus resultados; E o Capítulo 8 apresenta sugestões de melhorias assim como linhas de execução para

trabalhos futuros.

2 OBJETIVOS

Para de concretizar os resultados desse trabalho, os seguintes objetivos foram estabelecidos.

2.1 Objetivo Geral

Elaborar um modelo de classificação textual, que classifica se um comentário é apropriado ou não para crianças e adolescentes.

2.2 Objetivos específicos

a) Classificar os comentários pejorativos em vídeos infantis do Youtube;

b) Analisar os comentários classificados;

c) Indicar os vídeos que possuem mais comentários pejorativos para que os responsáveis protejam suas crianças e adolescentes na plataforma Youtube;

d) Avaliar o modelo de classificação proposto;

e) Desenvolvimento de uma extensão para o navegador da web Google Chrome que permite a visualização dos resultados das classificações realizadas.

3 FUNDAMENTAÇÃO TEÓRICA

A seguir, serão detalhados os principais conceitos envolvidos e utilizados durante este trabalho.

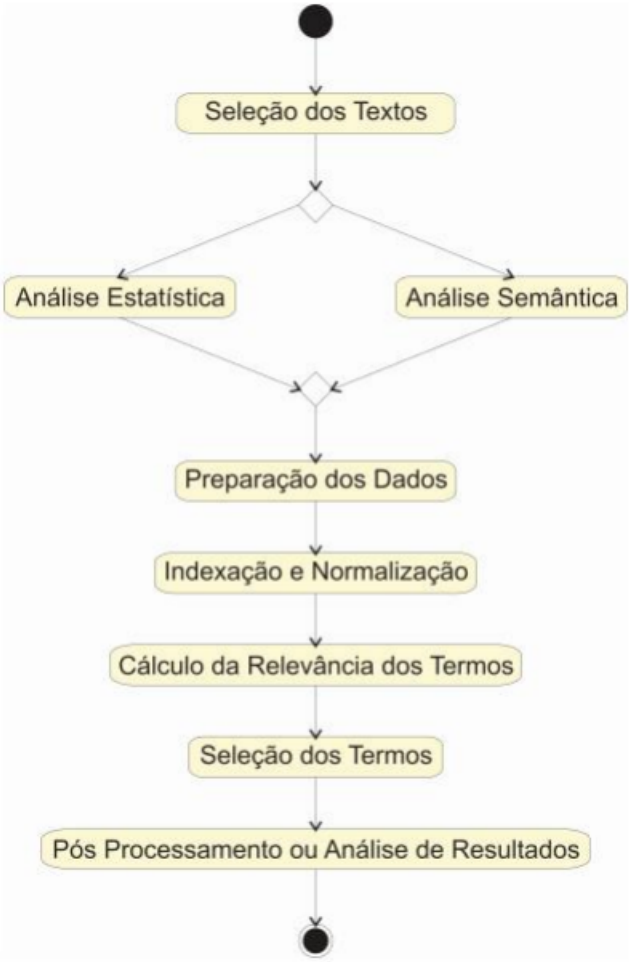
3.1 Mineração de Textos

Mineração de Texto ou *Text Mining* é definido como uma técnica de análise e extração de conhecimentos a partir de textos, frases ou palavras, com o objetivo de identificar informações úteis e implícitas contidas nos dados armazenados em formato não estruturado. Envolve a aplicação de algoritmos computacionais que processam textos e identificam as informações úteis e implícitas, que normalmente não poderiam ser recuperadas utilizando métodos tradicionais de consulta, pois a informação contida nestes textos não pode ser obtida de forma direta (MORAIS; AMBRÓSIO, 2007).

Dados em formato não estruturado representam uma grande quantidade de informações nos mais variados ambientes. Esses dados são constituídos de informações que não estão presentes em bancos de dados organizados, mas sim em e-mails, cartas, contratos, e até mesmo comentários da internet. Por serem escritos por humanos para leitores humanos, e não são acessíveis diretamente para computadores, precisam de processamento de Linguagem Natural (NLP), para que tenham sua informação extraída (DÖRRE; GERSTL; SEIFFERT, 1999). (ARAÚJO, 2015) Afirma que a prática de mineração de textos pode ser realizada em qualquer domínio que utiliza-se de textos, normalmente contidos em documentos, aplicando-se algoritmos computacionais para processar os textos e conseguir obter conhecimento contido no formato de dados não estruturados.

De forma geral as etapas da mineração de texto são: seleção de documentos, definição do tipo de abordagem dos dados (análise semântica ou estatística), preparação dos dados, indexação e normalização, cálculo da relevâncias dos termos, seleção dos termos e pós-processamento (análise dos resultados), como mostrado na Figura 2 (MORAIS; AMBRÓSIO, 2007).

Figura 2 – Etapas do Processo de Mineração de Texto



Fonte: (MORAIS; AMBRÓSIO, 2007)

3.1.1 Abordagem dos dados

(MORAIS; AMBRÓSIO, 2007) apresenta dois tipos de abordagem dos dados textuais na área de mineração de textos: Análise Semântica, baseada na funcionalidade dos termos encontrados no texto, e Análise Estatística, baseada na frequência dos termos encontrados no texto. Estas abordagens podem ser utilizadas de forma separada ou em conjunto. A abordagem utilizada neste trabalho é a Análise Semântica.

3.1.1.1 Análise Semântica

Análise que avalia a sequência dos termos no texto sendo analisado, para identificar sua função, fundamentada em técnicas de *Processamento de Linguagem Natural*. A análise

semântica se dá pelo pelo conhecimento do significado das palavras de forma individual, independente do contexto, e também pelo conhecimento da estrutura dessas palavras, por exemplo, ao determinar os limites da palavra que determinam seu radical. A utilização de Análise Semântica se dá pela melhoria na qualidade dos resultados do processo de mineração de textos (MORAIS; AMBRÓSIO, 2007).

Como as técnicas de análise semântica de textos procuram identificar a importância das palavras dentro da estrutura de orações (MORAIS; AMBRÓSIO, 2007), e por estar diretamente relacionada ao *processamento de linguagem natural*, esse conceito é utilizado neste trabalho durante o pré-processamento dos dados, detalhado no Capítulo 5.

3.1.2 Preparação dos dados

A preparação dos dados envolve a seleção de dados que constituem a base de textos de interesse e o trabalho inicial para tentar selecionar o núcleo que melhor expressa o conteúdo desses textos. Além de prover uma redução dimensional, esta etapa procura identificar similaridades em função da morfologia ou do significado dos termos nos textos. O primeiro passo do processo de prepação dos dados é a *Recuperação de Informação* (MORAIS; AMBRÓSIO, 2007). Um *Sistema de Recuperação de Informação Textual* é um sistema desenvolvido para indexar e recuperar documentos do tipo textual. Nesse tipo de sistema as consultas são descritas através de termos e os documentos relevantes são recuperados de acordo com esses termos.

Outra etapa da preparação dos dados é a Análise de Relevância, onde o usuário pode entrar com um termo e obter como resposta um resultado de um problema em particular. A Análise de Relevância é utilizada principalmente para filtrar dados não pertencentes ao conjunto obtido. Essa etapa não é realizada, tendo em vista que os comentários coletados já compõem a base de dados desejada.

O ytCommentMiner [3], software desenvolvido para coleta dos comentários, é responsável pela etapa de preparação dos dados, até o momento da coleta dos textos online. A ferramenta não implementa a etapa de Análise de Relevância sobre o conjunto obtido.

3.1.3 Indexação e Normalização

O objetivo principal da indexação e normalização dos textos é facilitar a identificação de similaridade de significado entre suas palavras, considerando variações morfológicas e

[3] https://github.com/ssisaias/ytCommentMiner

problemas de sinonímia (MORAIS; AMBRÓSIO, 2007). Este processo tem como resultado a geração de um índice, construído através de um processo de indexação. Um documento pode ser indexado por termos diferentes que são correspondentes ao vocabulário utilizado em sua área. Nesse caso, geralmente, há um conjunto de termos predefinidos e específicos para cada assunto da área em questão (MORAIS; AMBRÓSIO, 2007).

Em mineração de textos, a indexação é um processo automático. Suas principais fases são: *identificação de termos* (simples ou composto); remoção de *stopwords* (palavras irrelevantes); e *normalização morfológica* (stemming) (MORAIS; AMBRÓSIO, 2007).

Figura 3 – Etapas do Processo de Indexação Automática

Fonte: (MORAIS; AMBRÓSIO, 2007)

A identificação dos termos consiste em identificar os termos contidos nos textos, sejam palavras simples ou termos compostos por duas ou mais palavras, também pode ser realizada uma correção dos erros gramaticais atravé de um dicionário de termos (MORAIS; AMBRÓSIO, 2007).

Remoção de *stopwords* consiste em eliminar palavras que não devem ser consideradas no documento, conhecidas como *stopwords*. *Stopwords* são palavras não relevantes ao texto, por não traduzirem sua essência. Normalmente fazem parte da lista de *stopwords* preposições, pronomes, artigos, advérbios e outras classes de palavras auxiliares (MORAIS; AMBRÓSIO, 2007).

Também durante o processo de indexação, torna-se interessante remover as variações morfológicas de uma palavra, através da identificação do seu radical. Os prefixos e os sufixos são removidos, e apenas o radical resultante é adicionado ao índice. Essa técnica é chamada de lematização ou *stemming* (MORAIS; AMBRÓSIO, 2007).

A identificação dos termos, remoção dos stopwords e *steeming*, neste trabalho são realizados automaticamente através de *scripts* e descritos no Capítulo 5.

3.1.4 Cálculo de relevância dos termos

Com exceção das *stopwords*, os termos mais frequentemente utilizados no texto, costumam ter maior importância, assim como palavras constantes em títulos ou em outras estruturas, uma vez que foram colocadas ali por serem consideradas relevantes para a idéia do documento (MORAIS; AMBRÓSIO, 2007).

O cálculo de relevância de uma palavra em relação ao texto que está inserida pode se basear na frequência da mesma, na análise estrutural do documento, ou na posição sintática da palavra. Ao grau de relacionamento da palavra com o texto dá-se o nome de *peso*. Logo é o peso que indica a importância da palavra em relação ao texto (MORAIS; AMBRÓSIO, 2007).

(MORAIS; AMBRÓSIO, 2007) apresenta algumas formas para cáculo do peso, que utilizam cálculos simples de frequência: *frequência absoluta*, *frequência relativa*, *frequência inversa de documentos*. Nenhuma dessa técnicas é utilizada neste trabalho. Neste trabalho, o cálculo da relevância dos termos é realizado ao preparar o dicionário da ferramenta *Sentistrength*, onde é possível dar um nível de relevância para cada palavra individualmente.

3.1.5 Seleção dos termos

Seleção de termos corresponde à etapa de seleção das palavras retiradas do texto, após os processos de pré-processamento e cálculo da relevância. Esta técnica pode ser baseada no peso dos termos ou na sua posição sintática em relação ao texto. Entre as principais técnicas de seleção dos termos está a *Seleção por análise de linguagem natural* (MORAIS; AMBRÓSIO, 2007), utilizada neste trabalho.

3.1.5.1 Seleção por análise linguagem natural

A *seleção por análise de linguagem natural* consistem em aplicar técnicas de análise sintática ou semântica para identificar palavras em um documento. A análise semântica baseia-se no princípio de que as partes mais relevantes de um documento já estão de alguma forma demarcadas por estruturas de formatação específicas para isso (MORAIS; AMBRÓSIO, 2007).

A *Seleção por análise de linguagem natural* é utilizada pelo autor tanto no momento da extração dos textos dos conjuntos de comentários obtidos, como durante a montagem do modelo de classificação, tendo em vista que há uma classificação prévia do grupo de treino dos

comentários e uma triagem dos comentários classificados antes de serem utilizados para criação de fato do modelo.

3.1.6 Pós processamento ou análise dos resultados

Esta fase envolve a aplicação de técnicas de análise dos resultados de um Sistema de Recuperação de Informações de Texto, particularmente os resultados do processo de mineração de textos. A análise pode ser utilizada como forma de avaliação do *Sistema de Recuperações de Informação de Texto*, para saber se funcionou como deveria ou não (MORAIS; AMBRÓSIO, 2007).

Entre as principais métricas de análise de *Sistema de Recuperações de Informação de Texto* estão: *recall* e *precision*.

3.1.6.1 *Recall*

O *recall* (abrangência ou revocação) mede a habilidade do sistema em recuperar os documentos mais relevantes para seu usuário, com base no termo ou expressão utilizado na formulação de sua busca (MORAIS; AMBRÓSIO, 2007).

Sua fórmula consiste em:

$$recall = \frac{n - recuperados - relevantes}{n - possiveis - relevantes}, onde: \quad (1)$$

n-recuperados-relevantes: é o número de documentos relevantes recuperados.

n-possíveis-relevantes: é o número total de documentos relevantes do sistema. Essa informação geralmente não é conhecida e só pode ser estimada estatisticamente.

3.1.6.2 *Precision*

A *precision* (precisão) mede a habilidade do sistema em manter os documentos irrelevantes fora do resultado de uma consulta (MORAIS; AMBRÓSIO, 2007).

Sua fórmula consiste em:

$$precision = \frac{n - recuperados - relevantes}{n - total - recuperados}, onde: \quad (2)$$

n-recuperados-relevantes: é o número de documentos relevantes recuperados.

n-total-recuperados: é o número total de documentos do sistema.

3.2 Processamento de Linguagem Natural

(LIDDY, 2001) define *Processamento de Linguagem Natural (Natural Language Processing - NLP)* como um conjunto de técnicas para analisar e representar textos de origem natural, em um ou mais níveis de análise linguística com o propósito de atingir processamento linguístico similar ao humano para um conjunto de tarefas ou aplicações.

Para mineração de textos armazenados em formato não estruturados, são necessárias técnicas e ferramentas específicas da área de *Descoberta de Conhecimento em Textos (Knowledge Discovery from Text - KDT)* (MORAIS; AMBRÓSIO, 2007). Para recuperação de informação, KDT e mineração de textos possuem alto grau de dependência de Processamento de Linguagem Natural. Neste trabalho, NLP é utilizada tendo em vista que o ato de interpretar e manipular palavras como parte de uma linguagem é considerado Processamento de Linguagem Natural (MORAIS; AMBRÓSIO, 2007).

3.3 SentiStrength

SentiStrength é um classificador léxico que utiliza regras de linguística para detectar a força de um sentimento em uma frase (THELWALL; BUCKLEY; PALTOGLOU, 2012).

Para cada texto classificado, a ferramenta gera dois valores inteiros que variam de 1 a 5 numa escala positiva e 1 a 5 numa escala negativa, sendo o valor 1, um indicador de neutralidade para o sentimento. Por exemplo, uma classificação que retorna 3, 5 indica uma nota 3 para o sentimento positivo e nota 5 para o sentimento negativo, nesse caso o texto tem mais força negativa do que positiva (THELWALL; BUCKLEY; PALTOGLOU, 2012).

O classificador SentiStrength foi concebido para ser utilizado com diversos idiomas, porém seu dicionário padrão é em inglês. É possível adaptar seu dicionário para outros idiomas. O dicionário utilizado pelo autor neste trabalho, parte de um dicionário português limitado que é fornecido no site da ferramenta SentiStrength, porém adaptado com mais palavrões e algumas correções, para uma detecção mais abrangente. A lista completa dos palavrões adicionados, assim como sua relevância na classificação é indicada no Apêndice B. O dicionário utilizado, em sua completude, pode ser encontrado no Apêndice A. Além disso, foi feito um script [4] para converter a classificação gerada pela ferramenta, para a entrada esperada pelo classificador Naïve Bayes, essa etapa é descrita na seção de procedimentos.

[4] https://gist.github.com/ssisaias/08b2c8494a4553612987c9d4ae94f86c

3.4 Naive Bayes

Segundo (TAN; STEINBACH; KUMAR, 2009), uma técnica de classificação é uma maneira de construir modelos de classificação a partir de um *Dataset*. Exemplos de classificadores são: Árvore de decisão; Rede Neurais; SVM (Support Vector Machines); e classificadores Naïve Bayes. Cada uma dessas técnicas aplica um algoritmo de aprendizagem para identificar o modelo que melhor descreve um relacionamento entre o conjunto de atributos e as classes dos dados de entrada. Sendo assim o modelo gerado pelo algoritmo de aprendizagem deve ser capaz de, a partir de um conjunto de entradas, classificar corretamente registros que não eram conhecidos até então.

Naïve Bayes é um dos algoritmos mais eficientes para *machine learning* e mineração de textos. Esse algoritmo de aprendizagem supervisionado utiliza do Teorema de Bayes, porém considerando haver independência entre as *features* (característica das variáveis de entrada) do conteúdo sendo classificado (ZHANG, 2004).

De acordo com com (ZHANG, 2004), o Teorema de Bayes com *features* dependentes pode ser definido da seguinte maneira, onde por exemplo, a probabilidade P de um $comentario$ ser de classe c é:

$$P(c|comentrio) = P(c)\frac{P(comentario|c)}{P(comentario)} \tag{3}$$

Considerando existem duas classes, **positivo** e **negativo**, o evento $comentario$ pertence a classe $positivo$ se somente se:

$$f_b(comentario) = \frac{P(positivo|comentario)}{P(negativo|comentario)} \geq 1 \tag{4}$$

onde $f_b(comentario)$ é um classificador Bayesiano.

Ao assumir que as *features* são independentes dados os valores de classe, ou seja:

$$p(comentario|c) = p(comentario_1, comentario_2, ..., comentario_n|c) = \prod_{i=1}^{n} p(comentario_i|c), \tag{5}$$

O classificador resultante é:

$$f_nb(comentario) = \frac{P(positivo)}{P(negativo)} \prod_{i=1}^{n} \frac{P(comentario_i|positivo)}{P(comentario_i|negativo)} \tag{6}$$

A função $f_nb(comentario)$ é chamada de classificador Naive Bayes (Naïve Bayesian Classifier).

Neste trabalho é utilizada uma instância de Naïve Bayes chamada *Multinomial Naïve Bayes*. *Multinomial Naïve Bayes* indica que $p(comentario_i|c)$ possui uma distribuição multinomial. *Multinomial Naïve Bayes* é altamente recomendado para processamento de texto não estruturado onde há contagem de palavras ao considerar a relevância dos termos (METSIS; ANDROUTSOPOULOS; PALIOURAS, 2006a). A criação de um modelo Multinomial Naïve Bayes se dá através da ferramenta *scikit-learn*, onde o processo de criação do classificador já é previamente implementado. *scikit-learn* é uma biblioteca de aprendizagem de máquina desenvolvida na linguagem *Python* (PEDREGOSA *et al.*, 2011).

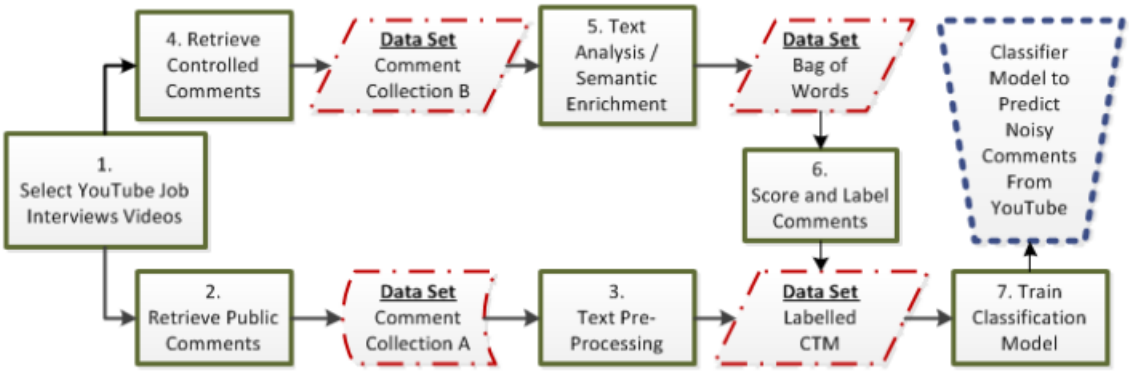
4 TRABALHOS RELACIONADOS

Explorar comentários de vídeos do Youtube é uma área de pesquisa recente e ainda em crescimento. Os trabalhos de Ammari, Dimitrova e Despotakis (2011) e (SCHULTES; DORNER; LEHNER, 2013) são pesquisas relativamente recentes que possuem coleta, mineração e classificação desses comentários.

O trabalho de (AMMARI; DIMITROVA; DESPOTAKIS, 2011) faz parte de um grupo maior de pesquisas, que visam elaborar um modelo de usuário que possa ser utilizado em simuladores de aprendizagem. Seu foco é apresentar uma técnica de filtragem de comentários com baixo valor semântico para o domínio selecionado, coletados de mídias sociais. A técnica proposta combina aprendizagem de máquina, mineração de dados, e análise semântica, a fim de obter comentários que sejam relevantes para o domínio desejado. A construção do modelo foi feita utilizando os classificadores Naïve Bayes Multinomial e Árvore de Decisão C4.5.

(AMMARI; DIMITROVA; DESPOTAKIS, 2011) coletou 1159 comentários de 17 vídeos do Youtube, dos quais 5 vídeos tiveram 193 comentários selecionados para o grupo de controle do modelo. A técnica apresentada e utilizada por (AMMARI; DIMITROVA; DESPOTAKIS, 2011) é mostrada na Figura 4:

Figura 4 – Metodologia de filtragem para comentários de baixo valor semântico do Youtube.



Fonte: (AMMARI; DIMITROVA; DESPOTAKIS, 2011)

A metodologia da Figura 4 consiste em:

1. Selecionar vídeos sobre entrevistas de emprego no Youtube.

2. Coletar os comentários públicos dos vídeos selecionados.

3. Pré-Processar os comentários obtidos.

4. Selecionar um grupo de comentários para grupo de controle, esses comentários são considerados relevantes para o domínio de entrevistas de emprego selecionado.

5. Analisar os comentários do grupo de controle, a fim de obter um Corpo de Palavras enriquecido semanticamente, relevante para o domínio.

6. Pontuar os comentários para facilitar sua classificação, que no caso do trabalho (AMMARI; DIMITROVA; DESPOTAKIS, 2011) é *relevant* (relevante) ou *noisy* (ruidoso - com baixo valor semântico).

7. Com os comentários devidamente pontuados, treinar um modelo de classificação supervisionado que irá indicar se um comentário é relevante ou não para o domínio proposto.

Os resultados apresentados por (AMMARI; DIMITROVA; DESPOTAKIS, 2011) indicam uma alta taxa de acerto dos classificadores para os comentários obtidos e pré-processados com a metodologia proposta, sendo 86,7% de corretude para o algoritmo C4.5 e 83,7% para o Naïve Bayes Multinomial.

A principal diferença entre (AMMARI; DIMITROVA; DESPOTAKIS, 2011) e este trabalho é que apesar de focar inicialmente em vídeos com público infantil e adolescente, este trabalho é capaz de aplicar o modelo em qualquer grupo de comentários de vídeos do Youtube. Dessa forma, expandindo o domínio de aplicação. Considerando também que modelo de (AMMARI; DIMITROVA; DESPOTAKIS, 2011) foi testado com 17 vídeos da plataforma porém um total de 1159 comentários, este trabalho avaliou um total de 87.094 comentários pertencentes a canais difundidos entre jovens e crianças, como **Galinha Pintadinha** e **Felipe Neto**.

(SCHULTES; DORNER; LEHNER, 2013) afirmam em seu texto que comentários em vídeos do Youtube são, de forma geral, mal vistos pelo público: em uma pesquisa com 95 participantes, 64% consideram comentários do Youtube "irrelevantes", 42% consideram agressivos e 51% os consideram "estúpidos", sendo que somente 6% dos entrevistados consideram os comentários em vídeos do Youtube "De essencial importância". Por outro lado, (SCHULTES; DORNER; LEHNER, 2013) também afirma que 34% dos entrevistados leem os comentários dos vídeos e que 53% lê os primeiros três comentários antes de começar a assistir o vídeo, além de estimar um total de 96 milhões de autores de comentários ativos na plataforma, chegando a conclusão de que a seção de comentários em vídeos do Youtube é uma funcionalidade essencial e uma das mais usadas em vídeos online. Levando esses fatores antagônicos em consideração, (SCHULTES; DORNER; LEHNER, 2013) primeiramente

procura verificar se comentários em vídeos do Youtube geram algum valor agregado e como medir esse valor. (SCHULTES; DORNER; LEHNER, 2013) também prova que através dos comentários é também possível obter uma análise semântica do vídeo em questão.

No período de 15/03/2012 à 21/03/2012, (SCHULTES; DORNER; LEHNER, 2013) coletou um total de 136.854 comentário dentre 304 vídeos do Youtube de categorias variadas. Para tentar descobrir se os comentários agregam algum valor para o leitor, foi proposto agregá-los em três tipos de comentários:

- **Discussão:** Comentários que geram debates dentre os usuários da plataforma;
- **Comentários inferiores:** Comentários com ofensas ou conteúdo irrelevante para o vídeo em questão;
- **Comentários substanciais:** Comentários não ofensivos que contém informação relevante e estão relacionados com o tema do vídeo em questão.

Após definir os tipos de comentários, ainda foram definidos dez subtipos a fim de tornar a classificação mais concisa. Durante a validação do seu *Dataset*, (SCHULTES; DORNER; LEHNER, 2013) concluiram que "não existe um tipo de comentário dominante em vídeos do Youtube" e também que "30% dos comentários são do tipo **Comentários inferiores**, o que pode explicar a má impressão dos usuários em relação aos comentários em vídeos do Youtube.

Apesar de ter utilizado um grupo maior de comentários do que este trabalho, (SCHULTES; DORNER; LEHNER, 2013), os vídeos obtidos são de diversas categorias tornando o modelo gerado menos específico.

Sobre proteção de crianças online, (FALCÃO *et al.*, 2016) aborda em sua pesquisa o problema da falta de acompanhamento por parte dos pais, quanto a utilização da internet por seus filhos.

(FALCÃO *et al.*, 2016) propos um sistema multiagente composto por agentes que coletaram e analisaram dados do Facebook. Os agentes trabalham juntos para trazer dados relevantes para o modelo de classificação, que auxilia em detectar casos de aliciamento infantil. O algoritmo utilizado para a classificação foi de árvore de decisão J48 (QUINLAN, 1986), implementado no WEKA, uma ferramenta de mineração de dados gratuita. O modelo foi validado com o perfil do Facebook de duas crianças, com o consentimento dos pais, e os resultados foram satisfatórios onde o modelo mostrou que uma das crianças apresentava-se em grupo de risco e estaria possivelmente sendo aliciada por adultos. Em comparação, o foco deste trabalho não é

detectar possíveis aliciadores, mas sim determinar se a seção de comentários de um vídeo de Youtube é seguro para a criança que está assistindo o vídeo, sendo uma outra camada de proteção para as crianças. O trabalho (FALCÃO *et al.*, 2016) também chegou a verificar a eficácia do seu modelo com outros algoritmos de classificação, entre eles o Naive Bayes que é utilizado neste trabalho.

(OLIVEIRA *et al.*, 2017) estende o trabalho de (FALCÃO *et al.*, 2016), buscando detectar automaticamente quando uma mensagem suspeita é trocada entre uma criança e um adulto. O algoritmo de classificação utilizado é Naive Bayes Multinomial (METSIS; ANDROUTSOPOULOS; PALIOURAS, 2006b).

Os dados utilizados para treinar o modelo com mensagens "perigosas" foram coletados do site: *perverted-justice* [5]. Os dados com mensagens consideradas "normais" foram retirados do site: *vircio* [6]. Foram utilizadas 1610 mensagens ao todo, com um total de 9450 palavras, para as quais o Naive Bayes Multinomial obteve uma taxa de acerto de 86,33%. Após aplicar o modelo em conversa real entre uma criança e um suspeito de aliciamento, (OLIVEIRA *et al.*, 2017) mostrou que seu modelo é mais robusto que o de (FALCÃO *et al.*, 2016), pois pode detectar frases de aliciadores fora de contexto e além disso, detecta frases como um todo e não somente palavras-chave.

A Tabela 1 apresenta uma comparação entres os trabalhos citados anteriormente e este trabalho em questão:

[5] http://www.perverted-justice.com
[6] http://vircio.net

Tabela 1 – Comparação entre trabalhos relacionados e este trabalho

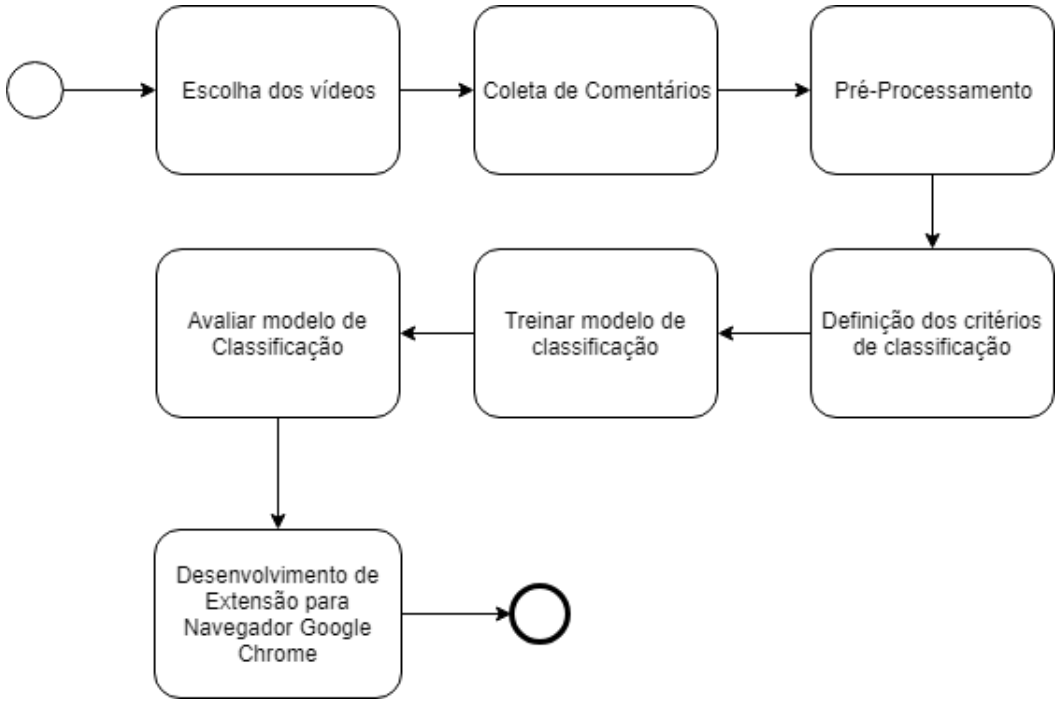
Trabalho	Tipo de Aplicação	Objetivo
(AMMARI; DIMITROVA; DESPOTAKIS, 2011)	Minerar comentários do Youtube	Filtrar comentários do Youtube com pouco valor semântico.
(SCHULTES; DORNER; LEHNER, 2013)	Minerar comentários do Youtube	Verificar a relevância de comentários do Youtube.
(FALCÃO *et al.*, 2016)	Proteção de crianças online	Medir nível de exposição de crianças no Facebook.
(OLIVEIRA *et al.*, 2017)	Proteção de crianças online	Detectar facilmente aliciadores de crianças na internet.
Este trabalho	Mineração de comentários do Youtube e proteção de crianças online	Verificar se um comentário de vídeo do Youtube é adequado ou não para crianças.

Fonte: Elaborado pelo autor

5 PROCEDIMENTOS METODOLÓGICOS

Para alcançar o objetivo de classificar os comentários pejorativos em vídeos infantis do Youtube, os passos a seguir descritos na Figura 5 foram planejados. A execução dos passos é descrita no Capítulo 6.

Figura 5 – Ilustração do procedimento metodológico



Fonte: autor

5.1 Escolha dos vídeos

O principal critério para a escolha do vídeo para este trabalho, é que seu público alvo seja infantil ou adolescente. Naturalmente, quanto mais comentários o vídeo possuir, maior a quantidade de dados para análise, logo o ideal é buscar vídeos infantis que apresentam um número expressivo de comentários, em torno de pelo menos mil. Porém, tendo em consideração que nem sempre é possível encontrar vídeos com essa quantidade ideal de comentários, podem ser utilizados vídeos com menos comentários. A popularidade dos canais que postaram os vídeos é um fator a ser considerado, visto que um canal com maior popularidade, tende a possuir vídeos com mais comentários.

5.2 Coleta de Dados

A fim de ter os dados para o ponto de partida da pesquisa e tendo em conta a grande quantidade de comentários em vídeos com público-alvo jovem, foi desenvolvida uma ferramenta em Python para coleta dos comentários em vídeos do Youtube. O objetivo da ferramenta é ser simples e objetiva na coleta desses comentários. A ferramenta é responsável pela etapa de preparação dos dados, dentro dos conceitos de mineração de texto.

5.3 Pré-Processamento

Uma vez que os dados armazenados não estão em formato adequado para extração do conhecimento, faz-se necessária a aplicação de métodos para *extração, integração, transformação*, *limpeza*, *seleção e redução* de volume desses dados, antes da etapa de mineração (MORAIS; AMBRÓSIO, 2007).

Inicialmente, observa-se uma grande quantidade de comentários sem sentido em vídeos infantis no Youtube, compostos quase que inteiramente por espaços em branco ou símbolos e letras aleatórios. Em vídeos destinados à adolescentes, há menor ocorrência de comentários sem sentido.

Faz necessário a utilização de um método ou ferramenta que fará a limpeza dos textos, bem como a aplicação da ferramenta NLTK para que seja feito o pré-processamento dos textos. Um *script* foi criado pelo autor e sua utilização é detalhada no Capítulo 6.

5.4 Definição dos Critérios de Classificação

A classificação dos comentários é definida manualmente, levando em consideração comentários de amostragem, que serão classificados como **pejorativos** ou **não pejorativos**.

Dado a enorme quantidade de comentários coletados, um dicionário de classificação também será criado para a ferramenta **SentiStrength**, a fim de facilitar a criação do conjunto de treino a ser dado de entrada para gerar o modelo de classificação.

Após a classificação manual e a classificação assistida utilizando a *SentiStrength*, um modelo deve ser construído a partir do treinamento com o conjunto de dados coletados, para que se possa classificar automaticamente os novos comentários obtidos do Youtube.

5.5 Treinamento do Modelo de Classificação

O classificador Naïve Bayes recebe um conjunto de dados de treino, e um conjunto de testes para averiguar a precisão do modelo de classificação (ZHANG; LI, 2007). Esses conjuntos de dados são escolhidos de forma aleatória dentre os comentários obtidos.

Por meio da ferramenta *Sentistrength*, pode-se obter um grupo de comentários classificados automaticamente, que deve ser analisado e melhorado pelo autor, a fim de obter um modelo de classificação mais preciso.

5.6 Definição dos critérios de classificação

A definição dos critérios de classificação corresponde à avaliação do modelo de classificação, ou seja, da sua qualidade. O modelo gerado poderá ser testado, utilizando os comentários que não foram utilizados na fase de treino, ou seja, utilizando os comentários selecionados para compor o conjunto de teste.

A avaliação do modelo é feita através da ferramenta *Scikit-learn*, que fornece métodos de *score* prontos para as medidas de *precision* e *recall*.

5.7 Desenvolvimento de Extensão para Navegador Google Chrome

Por meio do desenvolvimento de um *plugin* para o Google Chrome é possível utilizar o modelo construído e indicar vídeos com comentários pejorativos. Nessa etapa, são avaliados os resultados do processamento dos comentários, agregando agora pelos vídeos dos quais foram coletados, indicando quais os vídeos tem maiores taxas de comentários pejorativos e que não são considerados seguros para crianças e adolescentes.

6 AVALIAÇÃO EXPERIMENTAL

Para elaboração do modelo de classificação, a partir dos comentários obtidos para o vídeo **Não Faz Sentido! - Crepúsculo [+13]**, foram escolhidos de forma aletória um grupo de 50% (32.131) dos comentários para elaboração do modelo de classificação e 50% (32.130) dos comentários para elaboração do grupo de validação (ou grupo de testes). Neste Capítulo é descrita toda a execução dos passos mencionados na Figura 5 e no Capítulo 5, utilizando os grupos de comentários para elaboração e validação do modelo.

6.1 Escolha dos vídeos

Levando em consideração os critérios de escolha mencionados anteriormente, o tamanho do canal que postou o vídeo e a quantidade de comentários, foram escolhidos os vídeos a seguir:

- **PARABÉNS DA GALINHA PINTADINHA - Clipe Música Oficial - Galinha Pintadinha DVD 4** [7]
- **Galinha Pintadinha 4 - Clipe Música Oficial - Galinha Pintadinha DVD 4** [8]
- **UPA CAVALINHO - Clipe Música Oficial - Galinha Pintadinha DVD 4** [9]
- **Pintinho Amarelinho - DVD Galinha Pintadinha** [10]
- **Galinha Pintadinha 3 - Trailer - OFICIAL** [11]
- **Não Faz Sentido! - Crepúsculo [+13]** [12]

6.2 Coleta de Dados

Para coletar os comentários dos vídeos, uma aplicação na linguagem *Python* foi desenvolvida pelo autor. O Software ytCommentMiner [13] utiliza a API do Youtube (Versão 3) e permite tanto obter os comentários de topo (*top level comments* ou *Comment Threads*),

[7] https://www.youtube.com/watch?v=ei2-RjJDBHc
[8] https://www.youtube.com/watch?v=gWI4qV2H8VY
[9] https://www.youtube.com/watch?v=Fn9adh4HWUU
[10] https://www.youtube.com/watch?v=59GM_xjPhco
[11] https://www.youtube.com/watch?v=FTC-MEUmZLw
[12] https://www.youtube.com/watch?v=2Lp7XO6oWCM
[13] https://github.com/ssisaias/ytCommentMiner

como as réplicas à esses comentários, dado o ID do vídeo que pode ser encontrado na sua URL, permitindo obter todos os comentários públicos disponíveis no vídeo, até o momento da coleta.

Os comentários são armazenados na máquina onde ytCommentMiner está sendo executado no formato de dados JSON, contendo meta informações relacionadas aos comentários: ID do comentário, nome do autor, texto original da postagem, texto atual da postagem, data de publicação e total de réplicas. A Figura 6 mostra um comentário coletado pela ferramenta em formato JSON, o nome do autor do comentário foi omitido.

Figura 6 – Exemplo de comentário obtido pela ferramenta ytCommentMiner


```
}, {
    "id": "z23ddj2plpuzxd5iv04t1aokge1qbgjkhywgitl3zlnprk0h00410",
    "snippet": {
        "topLevelComment": {
            "snippet": {
                "authorDisplayName": [redacted],
                "textDisplay": "Minha filhinha ja gosta!!!!!",
                "textOriginal": "Minha filhinha ja gosta!!!!!",
                "publishedAt": "2017-09-01T01:25:28.000Z"
            }
        },
        "totalReplyCount": 0
    },
    "replies": {
        "comments": []
    }
}, {
```


Fonte: autor

6.3 Pré-Processamento

Nesta etapa, um *script* [14] escrito em Python foi executado para extrair somente os textos dos comentários obtidos através da ferramenta de extração, o script também é capaz de remover caracteres que não são considerados, dentro do contexto desta pesquisa, como *emojis*.

Em seguida, através da ferramenta NLTK, o pré-processamento foi realizado de foram automática: os acentos foram removidos e os textos foram *tokenizados*, através de um processo que remove palavras sem valor semântico para o classificador e reduz palavras com valor semântico ao seu radical, através dos métodos *word_tokenize()* e *RSLPStemmer()* presentes na

[14] https://gist.github.com/ssisaias/d6dd83361d6fa64c0341427b1f6f3f22

NTLK. Ao final dessa etapa, um arquivo contendo os comentários devidamente pré-processados é gerado.

6.4 Treinamento do Modelo de Classificação

Em posse dos comentários pré-processados, foi obtido um modelo de classificação da seguinte forma:

Primeiramente, o dicionário da *SentiStrength* deve estar previamente estabelecido, o autor obteve um dicionário pronto para a língua portuguesa já fornecido no site da ferramenta e o melhorou e adaptou para o contexto deste trabalho, removendo algumas palavras em inglês que estavam erroneamente no dicionário e adicionando os palavrões necessários para a classificação de comentários com termos pejorativos. Os palavrões adicionados podem ser encontrados no Apêndice B.

A versão do *SentiStrength* utilizada foi a 2.3, específica para Sistemas Operacionais Windows. Através de sua interface gráfica, o arquivo foi gerado pelo *script* [15] da etapa de pré-processamento. Após a classificação gerada pelo *SentiStrength*, alguns comentários foram verificados manualmente pelo autor, esse procedimento se refere à classificação assistida. Frases que foram classificadas como pejorativas erroneamente por possuírem, por exemplo, o seguinte *emoticon* **:(**, foram reclassificadas como neutras, pelo próprio autor. Ou seja, não expressão sentimento pejorativo.

Outro *script*[16] é então executado para "**traduzir**"a classificação feita pela *SentiStrength*. O funcionamento desse *script* se dá do seguinte modo: Quando a classificação de um comentário é dada como positiva pela ferramenta, é então convertida para **0** (ou classificação neutra); Quando a classificação é negativa, espera-se que haja um termo pejorativo no comentário e esse é classificado com o valor **1**. Isso faz com que o modelo de classificação apenas se preocupe com as duas classes determinadas que são neutra e pejorativa.

De posse dos comentários que compõem o grupo de treino devidamente classificados, utilizou-se a biblioteca *scikit-learn*, para criação do modelo de classificação. O procedimento está descrito a seguir.

No trecho de código abaixo, as variáveis **counts** e **targets** representam a lista com os comentários e a lista com suas classificações, respectivamente. A variável **classifier** recebe a

[15] https://gist.github.com/ssisaias/d6dd83361d6fa64c0341427b1f6f3f22
[16] https://gist.github.com/ssisaias/08b2c8494a4553612987c9d4ae94f86c

nova instância do classificador MultinomialNB, e que em seguida é "alimentada" com os dados dos comentários e sua classificação. O modelo de classificação é então criado.

```
from sklearn.naive_bayes import MultinomialNB
classifier = MultinomialNB()
classifier.fit(counts, targets)
```

6.5 Avaliação do Modelo de Classificação

Nesta etapa, os comentários separados para o grupo de teste foram submetidos ao classificador. Também foram submetidos todos os outros comentários obtidos, tendo estes passado pelo mesmo procedimento de pré-processamento, classificação na ferramenta *SentiStrength*, análise humana, e enfim, classificados pelo modelo gerado anteriormente.

A exemplo do grupo de testes mencionado na criação do modelo de classificação, uma matriz com os comentários é passada para o método **predict()** do classificador. A variável **predictions** contém a classificação gerada, que agora pode ser avaliada de acordo com os valores esperados.

```
predictions = classifier.predict(teste_counts)
```

A avaliação do modelo é feita através da ferramenta Scikit Learn, que fornece métodos de *score* prontos para as medidas de *precision* e *recall*. Um *script* [17] auxiliar foi criado com os passos de validação. Os resultados da avaliação são apresentados no próximo capítulo.

6.6 Extensão do Google Chrome - SafeYoutube

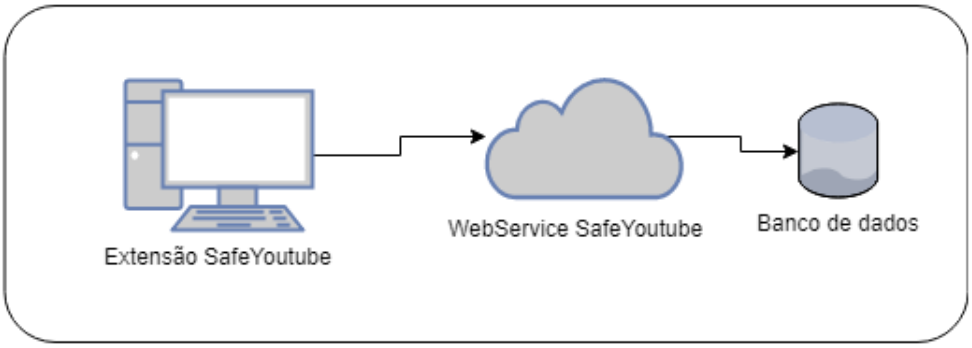
Após gerar e avaliar o modelo de classificação, e obter os resultados, foi desenvolvida uma extensão do navegador da web Google Chrome que se utiliza de um *Web Service* que retorna a margem de comentários negativos e positivos para o vídeo sendo assistido naquele momento.

A extensão **SafeYoutube** foi criada utilizando as tecnologias web HTML, CSS e Javascript. Como referência para o desenvolvimento foi utilizada a documentação oficial para desenvolvedores Chrome [18]. Sua arquitetura é descrita no diagrama da Figura 7. Os códigos-fonte

[17] https://gist.github.com/ssisaias/b9521c910d66c440c9e90d21f5360536
[18] https://developer.chrome.com/extensions

da extensão [19] e do serviço web [20], estão disponíveis publicamente e podem ser encontrados na plataforma Github.

Figura 7 – Arquitetura da extensão desenvolvida



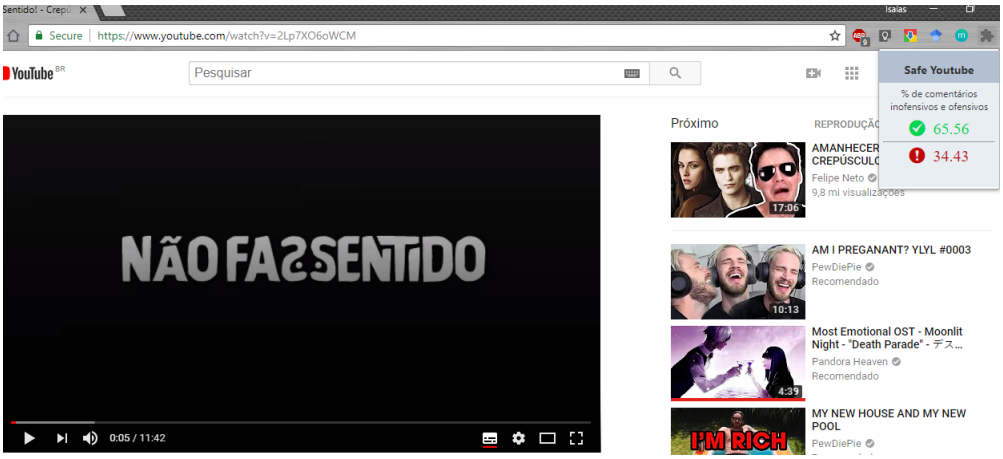
Fonte: Autor

Ao entrar um vídeo do youtube com a extensão **SafeYoutube** instalada, ela permite que o usuário a selecione, exibindo assim os resultados de uma classificação realizada previamente para aquele vídeo. Um exemplo do funcionamento é visto na Figura 8.

Caso os comentários do vídeo ainda não tenham sido classificados, o serviço irá iniciar um procedimento automático de classificação, tornando os resultados disponíveis dentro de um intervalo de tempo. O procedimento automático de classificação segue os mesmo procedimentos listados nas etapas de **pré-processamento** e **avaliação do modelo**, ou seja, os comentários são coletados, pré-processados, e classificados com o modelo de classificação gerado na etapa de **Treinamento do Modelo**.

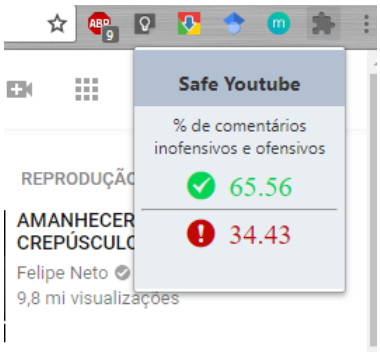
[19] https://github.com/ssisaias/safe-youtube
[20] https://github.com/ssisaias/safe-youtube-service

Figura 8 – Extensão SafeYoutube



Fonte: Autor

Figura 9 – Extensão SafeYoutube - Imagem ampliada



Fonte: Autor

Nota-se pelo *plugin* no canto superior direito da Figura 9, que o vídeo em questão possui 65.56% de comentários positivos e 34.43% de comentários negativos.

7 RESULTADOS

Foram coletados 87.094 comentários ao todo, dentre 5 vídeos infantis e 1 vídeo com público adolescente. A Tabela 2 apresenta os meta dados da base adquirida: nome dos vídeos, total de visualizações no momento da coleta, data da coleta e quantidade de comentários por vídeo.

Tabela 2 – Vídeos selecionados e comentários obtidos

Título do Vídeo	**Visualizações**	**Data da Coleta**	**Comentários obtidos**
PARABÉNS DA GALINHA PINTADINHA - Clipe Música Oficial - Galinha Pintadinha DVD 4	388.515.807	20/11/2017 12:00	11.669
Galinha Pintadinha 4 - Clipe Música Oficial - Galinha Pintadinha DVD 4	90.240.721	20/11/2017 11:30	2.320
UPA CAVALINHO - Clipe Música Oficial - Galinha Pintadinha DVD 4	112.389.836	27/09/2017 23:38	2.798
Pintinho Amarelinho - DVD Galinha Pintadinha	496.289.054	28/09/2017 00:05	12.906
Galinha Pintadinha 3 - Trailer - OFICIAL	13.849.870	27/09/2017 11:28	298
Não Faz Sentido! - Crepúsculo [+13]	15.768.111	15/05/2018 07:14	57.103

Fonte: Elaborado pelo autor

Por conta do vídeo **Não Faz Sentido! - Crepúsculo [+13]** ser o com maior quantidade de comentários e devido ao seu público alvo ser composto por adolescentes, o conjunto de treino e de testes da classificação partiu dele.

A Tabela 3 exibe a quantidade de comentários obtidos em cada classe, para cada vídeo analisado. Esse total é obtido através do método *classification_report()* presente na biblioteca *scikit-learn*.

Tabela 3 – Quantidade de comentários em cada classe por vídeo

Título do Vídeo	**Neutros**	**Pejorativos**
PARABÉNS DA GALINHA PINTADINHA - Clipe Música Oficial - Galinha Pintadinha DVD 4	11.152	517
Galinha Pintadinha 4 - Clipe Música Oficial - Galinha Pintadinha DVD 4	2.217	103
UPA CAVALINHO - Clipe Música Oficial - Galinha Pintadinha DVD 4	2.654	144
Pintinho Amarelinho - DVD Galinha Pintadinha	11.407	1.499
Galinha Pintadinha 3 - Trailer - OFICIAL	277	21
Não Faz Sentido! - Crepúsculo [+13]	21.066	11.064

Fonte: Elaborado pelo autor

Como foi descrito no Capítulo 6, o modelo gerado foi utilizado em todos os comentários obtidos, com exceção dos comentários utilizados no grupo de treino, tendo sido a etapa de geração do modelo de classificação, o único passo a ser executado somente uma vez. Além disso, apesar do grupo a partir do qual foi gerado o modelo de classificação ter partido de apenas um vídeo, os comentários negativos são similares dentre todos os vídeos, seja de conteúdo infantil ou juvenil. Seguindo a metodologia de 50% de comentários para treino e 50% para teste do classificador, foram utilizados 32.131 comentários para treino e 32.130 comentários para teste.

Em relação à analise de dados com conjuntos de dados de entrada desbalanceados, para um modelo de classificicação, (TAN; STEINBACH; KUMAR, 2009) afirma que as métricas de ***precision*** e ***recall*** se sobressaem à métrica de *taxa de acertos*, pois esta última não considera o peso das classes sendo analisadas.

Na Tabela 4 estão os valores de *precision*, descritos por vídeo, após a classificação. Os valores são obtidos com o uso da biblioteca *sklearn-metrics* (PEDREGOSA *et al.*, 2011). A coluna **precision-pos** descreve a precisão de acerto do classificador para a classe positiva (ou neutra), enquanto a coluna **precision-neg** descreve a precisão de acerto do classficador para a classe negativa, onde estão os termos pejorativos.

Na Tabela 5 estão os valores de *recall*, também descritos por vídeo e classe. A coluna **recall-pos** descreve a pontuação de *recall* para a classe de comentários positivos (ou neutra), enquanto a coluna **recall-neg** descreve a pontuação de recall para a classe de comentários pejotarivos.

Nas definições a seguir, T_p (*True positive* ou positivo verdadeiro) indica o número de comentários da classe de comentários positivos (neutros) que foram corretamente detectados como positivos; F_n (*False negative* ou falso negativo) denota a quantidade de comentários pejorativos detectados erroneamente como positivos; F_p (*False positive* ou falso positivo) indica a quantidade de comentários pejorativos detectados erroneamente como positivos; e T_n (*True negative* ou negativo verdadeiro) indica a quantidade de comentários pejorativos detectados corretamente (TAN; STEINBACH; KUMAR, 2009).

Precision pode ser definida para a classe de comentários positivos, por exemplo, como a proporção do número de acertos pelo modelo na classe positiva pelo número de instâncias que foram classificadas como positiva pelo modelo.

$$P = \frac{T_p}{T_p + F_p} \tag{7}$$

Enquanto o *recall*, corresponde a proporção do número de acertos, por exemplo, da classe positiva divido pelo que é realmente positivo.

$$R = \frac{T_p}{T_p + F_n} \tag{8}$$

Tabela 4 – Resultados de *precision*

Título do Vídeo	**Classificados**	**precision-pos**	**precision-neg**
PARABÉNS DA GALINHA PINTADINHA - Clipe Música Oficial - Galinha Pintadinha DVD 4	11.669	0.98	0.18
Galinha Pintadinha 4 - Clipe Música Oficial - Galinha Pintadinha DVD 4	2.320	0.98	0.14
UPA CAVALINHO - Clipe Música Oficial - Galinha Pintadinha DVD 4	2.798	0.98	0.16
Pintinho Amarelinho - DVD Galinha Pintadinha	12.906	0.95	0.35
Galinha Pintadinha 3 - Trailer - OFICIAL	298	0.95	0.14
Não Faz Sentido! - Crepúsculo [+13]	31.130	0.88	0.71

Fonte: Elaborado pelo autor

Tabela 5 – Resultados de *recall*

Título do Vídeo	**Classificados**	**recall-pos**	**recall-neg**
PARABÉNS DA GALINHA PINTADINHA - Clipe Música Oficial - Galinha Pintadinha DVD 4	11.669	0.86	0.71
Galinha Pintadinha 4 - Clipe Música Oficial - Galinha Pintadinha DVD 4	2.320	0.84	0.57
UPA CAVALINHO - Clipe Música Oficial - Galinha Pintadinha DVD 4	2.798	0.81	0.64
Pintinho Amarelinho - DVD Galinha Pintadinha	12.906	0.83	0.67
Galinha Pintadinha 3 - Trailer - OFICIAL	298	0.78	0.48
Não Faz Sentido! - Crepúsculo [+13]	31.130	0.83	0.79

Fonte: Elaborado pelo autor

A partir dos resultados apresentados nota-se que o classificador possui uma alta taxa de *precision* para a classe positiva, enquanto apresenta taxas menores para a classe de comentários pejorativos, onde se encontram os termos pejorativos.

Essa diferença se dá principalmente pela quantidade de comentários negativos detectados e presentes, pois pode-se notar na Tabela 3, vídeos com público alvo infantil possuem menor quantidade de comentários negativos detectados.

Por outro lado, a pontuação de *recall* para comentários negativos foi satisfatória para 4 dos 6 vídeos analisados, mostrando que dentre os comentários pejorativos de fato existentes, o classificador gerado conseguiu detectar corretamente sua maior parte. Novamente, é possível notar que as menores taxas de *recall* se dão nos vídeos com menos comentários e com menor quantidade de comentários negativos, vide Tabela 3.

8 CONCLUSÃO E TRABALHOS FUTUROS

O problema que o trabalho propôs solucionar é o de classificar comentários em vídeos destinados à crianças e adolescentes, tendo em conta a baixa moderação que ocorre nesses comentários, e a incidência de ofensas e termos de baixo calão.

Através do modelo elaborado foi possível verificar que vídeos destinados especificamente a crianças não possuem um quantidade alta de comentários ofensivos. Segundo a Tabela 3, o vídeo **PARABÉNS DA GALINHA PINTADINHA - Clipe Música Oficial - Galinha Pintadinha DVD 4** do **Canal Galinha Pintadinha**, possui aproximadamente 95,56% comentários neutros e apenas 4,44% comentários pejorativos.

Por outro lado, vídeos destinados a adolescentes possuem uma taxa bem maior de comentários negativos, tendo em vista que os comentários do vídeo **Não Faz Sentido! - Crepúsculo [+13]** do **Canal Felipe Neto** foram classificados com uma taxa de 65,56% de comentários neutro e 34,43% de comentários pejorativos.

O modelo não foi amplamente aplicado em vários vídeos, tendo em conta a quantidade de comentários já coletados e o tempo necessário para a realização do trabalho. Mas em contrapartida, a utilização da extensão diminui esse peso, viso que com ela, é possível classificar outros vídeos no Youtube.

Os dados obtidos e o modelo gerado são considerados úteis para a classificação de comentários pejorativos de vídeos no Youtube, mais precisamente vídeos com público alvo adolescente. O que torna relevante o uso da extensão criada para o navegador da web Google Chrome, principalmente quando o uso é direcionado a adolescentes.

Com o objetivo de aumentar a relevância da pesquisa, algumas melhorias podem ser realizadas, principalmente no que diz respeito à aplicação das técnicas utilizadas, assim como na extensão desenvolvida:

a) Aumentar o número de vídeos classificados;

b) Avaliar o uso da extensão criada, assim como realizar melhorias de UX e trazer para outros navegadores;

c) Permitir que o público avalie a classificação realizada, tornando classificação mais precisa;

d) Melhorar o modelo de classificação criado através do *feedback* obtido;

REFERÊNCIAS

ALEXA. **Youtube.com Traffic, Demographics and Competitors - Alexa**. 2017. [Online; Acesso em: 5 Out.2017]. Disponível em: <https://www.alexa.com/siteinfo/youtube.com>.

AMMARI, A.; DIMITROVA, V.; DESPOTAKIS, D. Semantically enriched machine learning approach to filter youtube comments for socially augmented user models. **UMAP**, p. 71–85, 2011.

ARAÚJO, J. **MINERAÇÃO DE TEXTOS DO TWITTER UTILIZANDO TÉCNICAS DE CLASSIFICAÇÃO**. 2015. 41 f. Monografia (graduação) - Campus Quixadá, Universidade Federal do Ceará, Quixadá, Brasil.

DÖRRE, J.; GERSTL, P.; SEIFFERT, R. Text mining: Finding nuggets in mountains of textual data. In: **Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: ACM, 1999. (KDD '99), p. 398–401. ISBN 1-58113-143-7. Disponível em: <http://doi.acm.org/10.1145/312129.312299>.

ECA. **Estatuto da Criança e do Adolescente**. 1990. [Online; Acesso em: 21 Mai.2018]. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/l8069.htm>.

FALCÃO, M. **Análise cognitiva para proteção da criança nas redes sociais**. 2015. 50 f. Monografia (graduação) - Campus Quixadá, Universidade Federal do Ceará, Quixadá, Brasil.

FALCÃO, M. b. ; GONÇALVES, E.; SILVA, T. Coelho da; OLIVEIRA, M. D. Behavioral analysis for child protection in social network through data mining and multiagent systems. 04 2016.

LIDDY, E. D. Natural language processing. 2001.

METSIS, V.; ANDROUTSOPOULOS, I.; PALIOURAS, G. Spam filtering with naive bayes-which naive bayes? In: MOUNTAIN VIEW, CA. **CEAS**. [S.l.], 2006. v. 17, p. 28–69.

METSIS, V.; ANDROUTSOPOULOS, I.; PALIOURAS, G. Spam filtering with naive bayes-which naive bayes? In: **CEAS**. [S.l.: s.n.], 2006. v. 17, p. 28–69.

MORAIS, E. A. M.; AMBRÓSIO, A. P. L. Mineraçao de textos. **Relatório Técnico–Instituto de Informática (UFG)**, 2007.

OLIVEIRA, M. D.; VIANA, T. Salathiel de S.; SILVA, T. Coelho da; GONçALVES, E.; JR, M. S. R. F. Textual analysis for the protection of children and teenagers in social media: Classification of inappropriate messages for children and teenagers. 04 2017.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

QUINLAN, J. R. Induction of decision trees. **Machine learning**, Springer, v. 1, n. 1, p. 81–106, 1986.

SCHULTES, P.; DORNER, V.; LEHNER, F. Leave a comment! an in-depth analysis of user comments on youtube. 2013.

TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introdução ao datamining: mineração de dados**. [S.l.]: Ciência Moderna, 2009.

THELWALL, M.; BUCKLEY, K.; PALTOGLOU, G. Sentiment strength detection for the social web. **Journal of the Association for Information Science and Technology**, Wiley Online Library, v. 63, n. 1, p. 163–173, 2012.

ZHANG, H. The optimality of naive bayes. **AA**, v. 1, n. 2, p. 3, 2004.

ZHANG, H.; LI, D. Naïve bayes text classifier. In: **2007 IEEE International Conference on Granular Computing (GRC 2007)**. [S.l.: s.n.], 2007. p. 708–708.

APÊNDICE A

DICIONÁRIO UTILIZADO NA FERRRAMENTA SENTISTRENGTH:

ADVERS, AGU, ALOL, AOK, Admir, Admoni, Agraciados, Apaixonado, Apesar, Avaric, Blur, Cuidado, Céu, DAMAG, DIN, ECSTA, EXIGÍVEL, Excel, Faixas, Feroc, Festiv, Fiel, Fuckface, Graci, Grati, Harmon, Honesto, InVigor, Joll, Lazie, Lucki, Mar, Moody, Muah, Nast, PRIVILEG, Pettie, RAID, Romântico, SAP, Shaki, Splend, Tortur, Tosspot, TrEMBL, Treacher, UGL, Virtuo, abafado, abandonar, abate, abatido, abdicar, abençoar, aberração, abertura, abestado, abismo, abjeto, abolir, abominar, abominável, aborrecer, aborrecido, aborrecimento, abrasivo, abraço, abraços, abrupto, absurdo, abusi, abuso, abutre, acalentar, acaso, accus, acidente, acossado, acostar, acrimon, acrobacia, acusação, adepto, adjudicação, adoecer, ador, adora, adoração, adoro, adorável, adulterar, adulterat, adultério, adventur, adversário, afetação, afeto, afiado, afligir, aflição, afogar, afundou, agarrar, aggravat, agitado, agitar, agitat, agitação, agonia, agoniz, agradeceu, agradável, agredir, agreeab, ai, alarme, alegar, alegação, alegrar, alegremente, alegria, aleijado, alienat, almejar, altercação, altivo, alívio, ama, amado, amante, amarfanhar, amargo, amarrotar, amavelmente, amaz, ambiguit, ambivalente, ambíguo, ameaça, amor, amoroso, amputar, amus, analfabeto, anarquia, anarquista, anexo, angr, angústia, animosidade, aniquilar, aniquilação, anomalia, anormais, ansiar, ansioso, antagoni, anti-social, antieconômico, antinatural, antipatia, antiquado, antitrust, anular, anulação, anus, anxi, anômalo, apagar, aparência, apath, aplicar, apoio, aposentar, appreciat, apprehens, apreciar, apreender, apressado, aprisionar, apuro, apóia, arbitrário, ardente, argh, argumentos, arma, armadilha, arraste, arrebatar, arrepio, arriscado, arrogan, arrogante, arruda, arsehole, arteiro, artificial, asham, asinino, assaltar, assalto, assassinato, assassino, assinalada, assombrar, assustador, assustando, asswipe, astuto, ataque, atarantado, atencioso, aterroriza, aterrorizado, aterrorizar, atordoa, atrair, atrasado, atraso, atrofia, atroz, aturdido, ausente, austero, ausência, autocrata, autocrático, avaliado, avarento, aversi, aviltar, azarado, azedo, açougue, baba, baba-ovo, babaca, babaca, babaovo, baboseira, bacura, bagos, bagunça, bagunçado, baitinga, baitola, baixa, baixar, bala, balbuciar, banal, banana, bandido, banir, baque, baranga, baranga, barato, barbari, barf, barreira, barulho, bastardo, batalha, bater, batida, beaut, bebum, bebê, bebês, beicinho, beijo, beijoca, beligerante, beligerantes, bem, berk, berrante, besta, besta, bestial, bff, bg, bicha, bicha, birdbrain, bisbilhotar, bisca, biscate, bixa, bizarro, blah, blam, blefando, blefe, bloco, blurt, boa, bobagem, bobagens, bobo, bocejo, boceta, boco, bocó, boiola, bolagato, bom, bomba, bondade, bonehead, bonito, boquete, bosseta, bosta, bostas, brandir,

brando, bravata, breve, briga, brilho, brillian, brincadeira, brincalhão, brincando, brioco, brocha, bronha, broxa, brusco, bruta, brutal, bruto, bruxa, bruxaria, buceta, buceta, bufão, bug, bulir, bullshit, bumhole, bunda, bunduda, burgl, burlista, burro, burro, busseta, bélico, bêbado, bônus, cachorra, cadela, cadela, caga, cagado, cagao, cagona, cagão, calamidade, calandra, calma, calmaria, calmarias, calúnia, canalha, canalha, cancelar, canhão, canibal, cansado, cansativo, cansaço, caos, capitular, capotar, caprichoso, captura, cara de pau, caralho, carenagem, carga, carinho, carniça, caro, carranca, carrancudo, carrossel, casseta, cassete, cataclismo, catástrofe, cauteloso, cavalo, caverna, caçador, caótico, ceder, cegos, censor, censura, cercear, cerco, cerda, chafurdar, chama, chamuscar, chantagem, charlatanismo, charlatao, charlatão, chata, chateado, chatice, chato, chato, cheio, chereca, chicote, chifruda, chifrudo, chochota, choque, chora, choradeira, choramingar, chorando, chorar, chorou, chota, chuckl, chupada, chupado, chupar, chás, cicatriz, cinzento, ciumento, clamor, claro, coaxar, cobiçar, cocaina, cocaína, cocksucker, coercitivo, colapso, colidir, colisão, combate, combatente, comed, comemorar, cometer, comiseração, como, comoção, compaixão, companheiro, compartilhada, compartilhar, compelir, competir, complicar, complicação, compulsão, comum, cona, concurso, condenar, condenação, condescen, confiante, confiança, confinar, confiscar, confisco, confissão, conflito, conforto, confrontar, confundir, confus, confusão, congestionado, congestionamento, conluio, conspirador, conspirar, conspiração, consternação, constranger, contagioso, contenda, contentamento, contente, contorcer, contorcer-se, contra, contrabandear, contradições, contrariar, controvers, contrário, contusão, convencido, cool, corna, corno, corroer, corrosivo, corrosão, corrupto, corrupção, cortar, corte, courag, covarde, coxo, craZ, crap, crappy, crasso, crepitar, cretina, cretino, crime, crise, critici, cruel, cruz, crye, crédulo, crítica, crítico, crônica, cu, cuidados, cuidar, culpa, culpado, culpável, cumplicidade, curalho, cuzao, cuzuda, cuzudo, cuzão, câncer, cãibra, cético, cínico, cú, dEspis, danado, danos, debil, debiloide, decadente, decadência, decapitar, declínio, decompos, defeito, defenc, defensiva, deficiente, deficiência, definhar, defunto, degenerar, degradantes, deitado, delectabl, delgado, delicioso, delicious, deligh, delinquente, delinquência, delírio, demissão, demitido, demitir, demitir-se, demolir, demonio, demônio, demônio, denegrir, dente, denunciar, dependentes, deplorar, deplorável, depor, depravado, depreciar, depreciativo, depreciação, deprimir, derramar, derrota, derrubada, derrubar, desabilitar, desacostumado, desacreditar, desafiador, desafiar, desafio, desagradável, desajeitado, desajustamento, desalojar, desamparado, desanima, desanimador, desanimar, desanimei, desaparecer, desapontar, desarmar, desastre, desastroso, descarado, desconcertado, desconfiança,

desconforto, desconhecido, desconsiderar, descontentamento, descontrole, descortês, descrença, descuidado, desculpe, desdém, desempenhado, desempregados, desenraizar, deserto, deserção, desespero, desfavoráveis, desfazer, desfeito, desfigurado, desfiladeiro, desgosto, desgracado, desgracados, desgraça, desgraçado, desgraçado, desgraçados, desgrenhado, desigual, desigualdade, desilusão, desinformados, desinteresse, desistente, desisti, desleal, desleixado, deslocar, deslumbrante, desmaiar, desmentir, desmoralizar, desmoronar, desnecessária, desobediente, desobediência, desolado, desolador, desolação, desonesto, desonra, desordem, desorganizado, desorientar, despejar, desperat, despercebido, despertaram, despesa, despreocupado, desprezo, desprezível, desproporcionada, desprotegido, desprovido, destemido, destruct, destruir, desumano, desvantagem, desvantajoso, desviar, desvio, desânimo, deter, determinado, determinação, detestar, deturpar, devastadores, devastar, devastação, deve, devedor, devot, diabo, diabólico, dickhead, difamar, difamação, diferem, dificuldade, difunto, difícil, digni, dilema, dilúvio, diminuir, diminuição, dinâmi, direito, disagre, disapprov, discordante, discrepante, discriminar, discriminação, discutível, discórdia, disfarce, disparar, disparate, dispendioso, dispensabilidade, dispensar, dispor, disputa, disputável, dissatisf, dissidência, dissimulado, dissipar, dissolução, dissuadir, distanciamento, distrair, ditar, ditatorial, divagar, divertido, divin, divisão, divórcio, diâmetro, doce, docemente, doces, doente, doentes, doentio, doença, doida, doido, dominar, dominação, dor, dores, dork, douchebag, downhearted, doçura, duvidoso, débil, défice, dúvida, easie, eejit, egotis, egoísta, ei, elegan, eliminar, eliminação, elogio, emaranhar, embaraçar, emboscada, emocional, emoção, empe, empinar, empt, empurrão, empurrões, encanto, encargos, encoberto, encontrão, encorajando, enemie, energ, enervar, enferrujado, enfurecer, enga, enganador, enganando, enganar, enganação, engano, enganosa, enganoso, engolfar, enjôo, enlouquecedora, enrag, enrijecer, enrolação, enrolão, ensolarado, enterrada, enterrar, enthus, entorpecido, entorse, entravar, entupir, envergonhado, envie, epidemia, epíteto, equívoco, erosão, erradicar, errado, errar, erro, errôneo, esbanjar, esboçado, escaldadura, escandaloso, escape, escaramuça, escassez, escasso, esconder-se, escoriáceo, escravidão, escravizar, escrota, escroto, escrutinar, escurecer, escuro, escárnio, escândalo, escória, esfregaço, esgotar, esgoto, esgueirar-se, esmagado, esmagador, esmagar, esnobe, esotérico, espancado, espancar, especial, espera, esperado, esperando, esperança, esperançosamente, esperançoso, espinhoso, esporrada, esporrado, esporro, espreitar, espástico, esquecimento, esquisitice, estafado, estagnado, estelionatario, estelionato, estelionatário, estigma, estola, estragar, estranged, estrangeiro, estrangeiros, estrangulamento, estrangular,

estranho, estremecimento, estresse, estridente, estrita, estupida, estupidez, estupido, estupor, estuprador, estuprando, estupro, estáticas, estéril, estúpida, estúpido, estúpido, evasão, evitar, exagerando, exasperat, excedente, excentricidade, excessiva, excesso, excluir, exclusão, excommunicat, excremento, excruciat, excêntrico, executar, execuções, expediente, expelir, explodir, explosiva, explosão, expor, exprobrar, expulsar, exterminat, extinguir, extinto, extravagante, extraviados, extraviar-se, exílio, fab, fabricat, fabuloso, faca, facada, facilidade, facilmente, fake, falecimento, falha, falhado, falido, fallout, falsear, falsidade, falsificação, falso, falta, falácia, famintos, fantástico, fanático, faroleiro, farrapo, farsa, fascinação, fascista, fastidioso, fatal, fathead, fatigu, favor, fdp, febre, febril, fedelho, feder, fedida, fedido, fedor, fedor, fedorenta, feia, feio, feiosa, feioso, feioza, feiozo, fel, fela, felacao, felação, feliz, ferida, ferido, ferimento, ferir, feroz, fervilhar, festa, feudo, fiança, fiasco, fiesta, filho da puta, fingir, fiofó, fiscal, flagrante, flertar, flexib, flexibilização, foda, foda-se, fodao, fode, foder, fodida, fodido, fodido, fodão, fofoca, fome, forgiv, forjado, formidável, fornica, forte, força, fraco, fratura, fraude, fraude, frenético, fricção, frio, frustrante, frustrar, frágeis, frígido, frívolo, fucker, fucks, fuckwit, fuctard, fud, fude, fudecao, fudendo, fudeção, fudida, fudido, fuga, fugaz, fugir, fugitivo, fugiu, fugly, fulera, fuleragem, fumaça, fuming, fundador, funn, furioso, furnica, furnicar, futilidade, fácil, fé, fúria, fútil, gaguejar, galo, ganhou, ganido, ganância, garança, garra, gaseado, gasto, gay, geek, gemer, gemido, genero, gentlest, germe, gibberi, giggl, glori, glória, gmbo, golpe, golpeado, golpista, gonorrea, gonorreia, gordura, gosta, gostei, gosto, gracioso, grande, gratef, gratuito, grave, graves, gravidade, graxos, graça, graças, grelinho, grelo, greve, griev, grinn, gritar, grosseiro, grotesco, grr, guerra, guerras, guerreado, guerrilha, gueto, guincho, h8, ha, habitar, haha, handsom, happi, hater, heartbroke, heartwarm, hediondo, hehe, herói, hesita, hijack, hilário, hipocondríaco, hipocrisia, hipócrita, histeria, histérico, hogwash, hoho, homossexual, honra, hooey, horda, horr, horrorizado, horrível, hostil, hugg, humilhante, humor, hungr, hurra, ideal, idiota, idiotas, idiotice, ignor, ignóbil, ilegal, ilegalidade, iludir, ilusão, ilógico, imaturo, imbecil, imobilidade, imoral, impasse, impatien, impedimento, impedir, imperfeito, impessoal, impetuoso, impiedoso, implacável, implicar, implorar, impopular, impor, importan, impotente, imprecisão, impressionar, imprevistas, imprudente, impróprio, impulsivo, impureza, impuro, imundo, imóveis, imóvel, inacessível, inacreditável, inadequa, incapacidade, incapaz, incentivo, incerto, incessante, incivil, inclinação, incoerências, incomodar, incompatib, incompeten, inconstante, inconvenien, incorreta, incrível, incurável, incómodo, indecente, indecis, independentemente, indesejáveis, indetermin, indeterminado, indevido, indiferente,

indiferença, indigente, indignação, indigno, indisciplinado, indisposição, indizível, ineffect, ineficiência, inesperadamente, inevitável, inexatidão, inexato, inexperiente, inexplicável, infame, infantil, infectar, infecção, infeliz, infelizmente, inferior, infernal, inferno, inferno, infestar, infiel, infiltração, inflamado, inflamar, inflação, infligir, infortúnio, infracção, infração, infrutífero, ingratidão, ingrato, ingénuo, inimigo, injunção, injustificada, injusto, innocen, inofensivo, inquietante, inquieto, insalubre, insano, insatisfatória, insecur, insegurança, inseguro, insensível, insidioso, insignificante, insincer, insinuar, insolente, insolência, inspir, instabilidade, instável, insufficien, insulto, insuportável, inteligente, intell, intempestivo, interferência, interrupção, intimidar, intimidat, intolerante, intolerável, intoxicar, intrometer-se, intrometido, intrud, intrusão, inundar, invadir, inveja, invejoso, invisível, involuntário, inválido, inábil, inépcia, inútil, ira, ironia, irracional, irrefletido, irregular, irresponsável, irrisório, irritar, irritação, irritável, irônico, isca, iscrota, iscroto, isentar, isola, jerked, jittery, jogar, jurar, juros, jurou, kenga, kindn, labuta, ladra, ladrao, ladroeira, ladrona, ladrão, ladrão, laidback, lalau, lamacento, lamentar, lamentação, lamento, lança, lapso, laço, legal, lento, leprosa, leproso, lesma, ligeira, likeab, limitação, linguado, liquidar, liquidação, litig, livel, livrar, livrará, lixo, lmao, lodo, lol, lolol, lorota, loucamente, louco, loucura, lous, loveless, lucked, lucks, lucro, ludibriacao, ludibriar, ludibriação, lunkhead, lunático, luta, lutado, lutar, luv, lágrima, lágrimas, machorra, machucar, macumba, maddest, magnific, mal, mal-estar, malandragem, malandro, malcriado, maldito, maldição, malicioso, maligno, maltratar, maluquinho, malícia, mancada, mancar, mancha, mandriar, manguaca, manguaça, manhoso, mania, manipulat, mano, manso, maníaco, maravilha, maravilhosa, maravilhoso, marginal, masmorra, masochis, massacre, masturba, matador, matar, meanie, medo, medonho, medos, medíocre, melanchol, melancolia, melhor, melhorar, melodramático, mendigar, mendigo, mentira, mentiras, mentiroso, mentiu, merda, merda, merdoso, meretriz, merr, mija, mijada, mijado, mijo, mijo, minado, minar, miserável, misses, mistak, misunderst, mocrea, mocreia, mocréa, mocréia, mogoloid, moleca, moleque, molestar, monotonia, monstro, monstruoso, monótono, morrer, mortal, mortalha, morte, mortif, mortos, mourejar, mudo, muito, muleke, muleque, multa, multicor, mundano, murchar, murmurar, musaranho, muthafuck, mácula, mártir, mérito, míope, namorada, narcótico, naufrágio, nazista, nebulosidade, nebuloso, necessitado, negado, negar, negativo, negação, negligen, negligenciar, negligente, negligência, nerd, nervoso, neurótico, neutralizar, neutralização, ninhada, nix, nojeira, nojenta, nojento, nojo, nonsens, notório, novato, ntre, nucklehead, numpty, nurtur, nutter, não, obcecar, obes, obliterar, oblíquo, obnóxio, obrigada,

obrigado, obscurecer, obsoletos, obstinado, obstruir, obstrução, obstáculo, ociosidade, oco, ocultar, odeia, odeio, odiando, odiava, odio, odioso, ofender, ofensa, offens, ok, okays, oks, omfg, omg, omissão, omitir, openminded, opor, oposição, opportun, oprimir, optimi, orgulho, orgulhoso, original, otaria, otario, otarios, otária, otário, otários, ousadia, outrag, padrão, painf, paining, painl, paixão, palatabl, palhaçada, pandemônio,
panelinha, parafuso, paralisadas, paralisados, paralisação, paralisia, paranoi, parar, parasita, paraíso, partes, partie, partilha, partição, parto, paspalha, paspalhao, paspalho, paspalhão, pateta, patetice, patife, patético, pau, pavor, paz, pecado, pecados, pecaminoso, peculiar, peido, pena, penal, pendurar, penis, pensativo, pequenino, pequeno, perambular, perda, perde, perdedor, perder, perdida, perdido, perdição, perdoou, perecer, perereca, perfeito, perigo, perigosas, perigosos, perplexidade, persecut, perseguição, perturbado, perturbar, perverso, perversão, pesadelo, pesado, pesar, pessimis, petrif, petulância, phobi, piada, pica, picada, picao, picão, pilantra, pilantragem, pillock, pinhead, pior, piranha, piroca, piroco, piru, pisar, pistoleiros, pitada, piti, pleasur, plebeu, plonker, pneu, pobres, pobreza, podre, podridão, poluentes, pomposo, pontapé, popa, populares, porra, porra, porte, positiv, pqp, praga, prais, prannock, prat, precioso, precipitado, precário, predicamento, prega, preguiça, preguiçosamente, preguiçoso, prejudic, prejudica, prejudicado, prejudicando, prejudicar, prejudiciais, prejudicial, preocupação, presidiário, pressentimento, pressur, presunçoso, presunção, pretensioso, pretensão, prettie, prisão, problema, procrastinar, procrastinação, prod, proibir, promis, prontidão, pronto, propaganda, prostituta, protelando, protesto, provação, provocar, provocação, prêmio, puk, punheta, punhetao, punhetão, punho, punir, puro, pus, puta, putaria, puto, puxa-saco, puxasaco, pária, pânico, pênis, quebrou, queda, queimar, queixar-se, queixoso, quenga, quentes, querida, querido, questionável, quitação, rabao, rabo, rabuda, rabudao, rabudo, rabudona, rabudão, rabão, racha, rachadinha, racis, radiano, radical, raiva, ralhar, ralo, rampante, rancor, rançoso, rapariga, raptar, rasgar, raso, raspar, rastejar, rato, ração, reacionário, reassur, reativa, rebelde, recalcitrante, recaída, recessão, recompensa, recorrida, recuo, recusa, recusar, redundan, refrão, refugiados, refutar, regozijar, regresso, rejeitar, relaxar, reluctan, relutante, remorso, renunciar, renúncia, repartição, repreensão, repreensível, reprimir, reprovação, repudiar, repugnante, repulsa, repulsivo, resistente, resmungar, resmungão, resolv, respeito, responsável, ressaca, ressentir, restringir, resíduos, retaliar, retardada, retardado, retardar, reter, retido, retiro, retroceder, reverter, revigor, revogar, revogação, revolta, revolução, ricos, ridicul, ridicula, ridicularizar, ridícula, ridículo, ridículo, rigor, rigorosa, rigorosas, riqueza, rir, risco, risível, rival,

rivalidade, rixa, rock, rofl, rogar, rola, romanc, romper, ronco, rosnado, rosnar, roubados, roubar, roubo, rude, rufião, ruga, ruim, rumor, ruptura, ruído, ruína, rígida, sabedoria, sabotar, sacana, sacanagem, saciar, saco, sacudiu, sadde, safada, safado, safados, salafrario, salafrário, saltitante, salvar, sangrento, sanguinário, sapatao, sapatão, sapoha, sarcas, sarcasmo, satan, satisf, saudade, saída, scariest, se, seca, secessão, secur, sede, sedento, sedentário, seduzir, segredo, segregação, seguro, selvagem, senil, sentimental, separadamente, separar, seqüestrar, servidão, servil, sexista, sexy, sigilo, silli, simples, simplista, simplório, simulado, simulação, sincer, sincero, sinistro, sintoma, siririca, sitiar, skank, slur, smart, smil, sociab, soco, sofre, sofredor, sofrem, sofrido, sofrimento, solene, solitário, solteirona, soluçando, soluço, soluços, soluçou, sombrio, sonolento, sonolência, sorri, sorriso, sorte, sorumbático, soulmate, sozinho, spam, spaz, startl, straggler, struggl, stunk, suave, suavemente, subjugat, sublevação, submisso, suborno, subserviência, substituição, subterrâneo, subtrair, subversão, subverter, sucesso, sucky, sucumbir, sufocar, suga, sugado, suja, sujeira, sujeição, suk, sunshin, sup, super, superficial, superficialidade, superior, superstit, supervisão, suplicar, suportada, suprem, supressão, suprimir, supérfluo, surdez, surdo, surpreendente, surto, suscetível, suspeito, suspender, suspensão, suspicio, susto, sweetie, sábio, sério, tabu, talento, tapa, tarada, tarado, tard, tardio, tarifa, tedioso, tehe, teimoso, temendo, temeroso, temido, temperamento, temperamentos, tempestade, tempestuoso, temporariamente, tenso, tensão, tentação, ternamente, terribl, terrifyi, terror, terrori, terrível, tesouro, testuda, tezao, tezuda, tezudo, tezão, thankf, thanx, thnx, thrash, tipo, tirania, tiro, toleran, tolices, tolo, toma no cu, tomar no cu, tombar, tonto, tormento, torpor, torrent, torto, tosser, traged, traidor, trair, traição, tranquilo, transbordamento, transgredir, transgressão, transtorno, trapaceiro, tratar, trauma, travessura, treasur, tresandar, trevas, triste, tristeza, triunfo, trivi, trocadilho, trocha, tropeço, troubl, troucha, trouchas, trouxa, trouxas, troxa, truant, trudg, truque, trágico, trêmulo, tubarão, tuberculoso, tumulto, turbulento, twat, twunt, tédio, tímido, ugh, ultimato, unauthentic, uncomfortabl, undependability, undependable, undid, uneas, unhapp, unimpress, unlov, unsavo, unwelcom, upchuck, urdidura, usado, usurpar, uva, vaca, vacilar, vadio, vaga, vagabunda, vagabundo, vagabundo, vagina, vaguear, vaidade, valente, valor, valores, valorização, valuabl, vangloriar, vangloriar-se, vazamento, vazio, veada, veadao, veado, veadão, veemente, velhaco, velho, veneno, venenoso, veracidade, verdade, verdadeiramente, verdadeiro, vergonha, verme, veto, viada, viadao, viado, viados, viadão, vice, viciado, vicioso, vigor, vil, vilão, vingança, vingar, violados, violar, violação, violento, violentos, violência, vital, vitória, vitórias, volatilidade, vom, vomitar, vulgar, vulnerab, vão, víbora, vítima, vômito, wack, wanker,

wassock, wazzock, welcom, wickedn, winn, wonderf, worr, wow, wtf, x, xana, xaninha, xavasca, xerereca, xexeca, xibiu, xochota, xota, xox, xoxota, xx, yay, yays, zomba, zombado, zombando, zombar, álibi, árduo, áspero, ânus, ódio, ódio, órfãos, útil

APÊNDICE B

Lista de palavrões adicionados ao dicionário SentiStrength e o peso atribuído manualmente a cada um:

abestado -2; anus -3; ânus -3; babaca -3; babaovo -2; baba-ovo -2; bacura -2; bagos -2; baitinga -3; baitola -4; baranga -3; baranga -3; bebum -2; besta -3; bicha -4; bisca -2; bixa -4; blefando -2; blefe -2; boceta -4; boco -2; bocó -2; boiola -3; bolagato -3; boquete -4; bosseta -4; bosta -4; bostas -4; brioco -3; brocha -3; bronha -3; broxa -3; buceta -4; bunda -3; bunduda -3; burro -2; busseta -3; buceta -4; chata -2; cachorra -2; cadela -2; caga -3; cagado -3; cagao -3; cagão -3; cagona -3; canalha -3; cara de pau -2; caralho -5; carniça -3; casseta -3; cassete -3; charlatanismo -2; charlatao -2; charlatão -2; chatice -2; chato -2; chereca -3; chifruda -2; chifrudo -2; chochota -3; chota -3; chupada -2; chupado -2; cocaina -2; cocaína -2; corna -3; corno -3; cretina -3; cretino -3; cu -5; cú -5; cu -5; curalho -4; cuzao -4; cuzão -4; cuzuda -4; cuzudo -4; debil -3; debiloide -3; defunto -2; demonio -2; demônio -2; desanima -2; desanimei -2; desgracado -2; desgraçado -2; desgracados -2; desgraçados -2; desisti -2; difunto -2; doida -2; doido -2; enganação -2; enganando -2; enganosa -2; enrolação -2; enrolão -2; escrota -4; escroto -4; esporrada -3; esporrado -3; esporro -3; estelionatario -2; estelionatário -2; estelionato -2; estupida -3; estúpida -3; estupidez -3; estupido -3; estúpido -3; fake -2; fdp -5; fedida -3; fedido -3; fedor -2; fedorenta -3; feia -3; feio -3; feiosa -3; feioso -3; feioza -3; feiozo -3; fela -3; felacao -3; felação -3; filho da puta -5; fiofó -4; foda -5; foda-se -5; fodao -4; fodão -4; fode -4; fodida -4; fodido -4; fornica -3; fraude -2; fudecao -4; fudeção -4; fudendo -4; fudida -4; fudido -4; fulera -2; fuleragem -2; furnica -2; furnicar -2; golpe -2; golpista -2; gonorrea -2; gonorreia -2; grelinho -3; grelo -3; idiota -3; idiotas -3; idiotice -3; imbecil -3; inferno -2; iscrota -3; iscroto -3; kenga -4; ladra -2; ladrao -2; ladrão -2; ladroeira -2; ladrona -2; lalau -2; leprosa -3; leproso -3; lorota -2; ludibriacao -2; ludibriação -2; ludibriar -2; machorra -2; malandragem -2; malandro -2; manguaca -2; manguaça -2; masturba -3; merda -4; mija -3; mijada -3; mijado -3; mijo -3; miserável -2; mocrea -2; mocréa -2; mocreia -2; mocréia -2; moleca -2; moleque -2; muleque -2; muleke -2; nojeira -2; nojenta -2; nojento -2; nojo -4; otaria -3; otária -3; otario -3; otário -3; otarios -3; otários -3; palhaçada -2; paspalha -2; paspalhao -2; paspalhão -2; paspalho -2; pau -2; peido -3; penis -4; pênis -4; perereca -2; pica -5; picao -4; picão -4; pilantra -3; pilantragem -3; piranha -3; piroca -4; piroco -4; piru -4; porra -5; pqp -5; prega -3; punheta -4; punhetao -4; punhetão -4; pus -2; puta -5; puto -5; puxasaco -2; puxa-saco -3; quenga -2; rabao -2; rabão -2; rabo -2; rabuda -2; rabudao -2; rabudão -2; rabudo -2; rabudona -2; racha -3; rachadinha -3;

rapariga -4; retardada -3; retardado -3; ridicula -2; ridícula -2; ridículo -2; rola -3; sacana -3; sacanagem -3; saco -2; safada -3; safado -3; safados -3; salafrario -3; salafrário -3; sapatao -4; sapatão -4; siririca -4; tarada -3; tarado -3; tédio -2; testuda -2; tezao -2; tezão -2; tezuda -2; tezudo -2; toma no cu -3; tomar no cu -3; trocha -2; troucha -2; trouchas -2; trouxa -2; trouxas -2; troxa -2; vaca -2; vagabunda -3; vagabundo -3; vagina -3; veada -3; veadao -3; veadão -3; veado -3; verme -2; viada -3; viadao -3; viadão -3; viado -3; viados -3; xana -3; xaninha -3; xavasca -3; xerereca -3; xexeca -3; xibiu -3; xochota -3; xota -3; xoxota -3.