



UNIVERSIDADE FEDERAL DO PARÁ  
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS  
FACULDADE DE COMPUTAÇÃO  
BACHARELADO EM SISTEMAS DE INFORMAÇÃO

LEONARDO DIAS MARTINS

**MINERAÇÃO DE TEXTO PARA ANÁLISE AFETIVA DA INTERAÇÃO DOS  
USUÁRIOS COM JOGOS EMPÁTICOS**

BELÉM

2019

LEONARDO DIAS MARTINS

**MINERAÇÃO DE TEXTO PARA ANÁLISE AFETIVA DA INTERAÇÃO DOS  
USUÁRIOS COM JOGOS EMPÁTICOS**

Trabalho de Conclusão de Curso apresentado à Faculdade de Computação da Universidade Federal do Pará, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Orientador (a): Prof.<sup>a</sup> Dr.<sup>a</sup> Fabíola Pantoja Oliveira Araújo.

BELÉM

2019

LEONARDO DIAS MARTINS

**MINERAÇÃO DE TEXTO PARA ANÁLISE AFETIVA DA INTERAÇÃO DOS  
USUÁRIOS COM JOGOS EMPÁTICOS**

Trabalho de Conclusão de Curso apresentado à Faculdade de Computação da Universidade Federal do Pará, como requisito parcial para obtenção de título de Bacharel em Sistemas de Informação.

Orientador (a): Prof.<sup>a</sup> Dr.<sup>a</sup> Fabíola Pantoja Oliveira Araújo.

Aprovado em: \_\_\_\_ / \_\_\_\_ / \_\_\_\_.

**Banca Examinadora:**

---

Prof.<sup>a</sup> Dr.<sup>a</sup> FABÍOLA PANTOJA OLIVEIRA ARAÚJO  
Orientadora - UFPA

---

Prof.<sup>a</sup> Dr.<sup>a</sup> LIVIANE PONTE RÊGO  
Examinadora - UFPA

---

Prof.<sup>a</sup> Dr.<sup>a</sup> MARCELLE PEREIRA MOTA  
Examinadora - UFPA

A minha família, razão da minha vida.

A Deus.

## **AGRADECIMENTOS**

Primeiramente gostaria de agradecer a Deus, o responsável pela minha existência, sempre me dando forças para nunca desistir e mostrando o caminho certo para seguir.

Gostaria de agradecer também à minha família, que sempre me incentivou e me guiou, com sabedoria e ensinamentos valiosos, para que eu pudesse fazer as escolhas certas, sempre respeitando o próximo.

Aos professores, que se dedicaram a repassar conhecimentos preciosos a nós alunos, contribuindo para nossa formação acadêmica e profissional. Em especial agradeço a minha orientadora Fabíola Araújo, pela paciência e ajuda durante todo o processo de desenvolvimento desse trabalho, sendo sempre atenciosa e disposta a esclarecer todas as minhas dúvidas.

Por último, mas não menos importante, agradeço aos meus amigos, pela amizade, companheirismo e ajuda nos momentos que precisei.

“A mente que se abre a uma nova ideia  
jamais voltará ao seu tamanho original.”

Albert Einstein

## RESUMO

Diariamente uma grande quantidade de dados circula na internet, pois, é enorme a quantidades de computadores e equipamentos eletrônicos que possuem acesso à internet, produzindo informações em forma de imagens, vídeos e textos, com isso, é necessária uma forma de analisar e extrair essas informações. Portanto, esse trabalho apresenta conceitos e técnicas para encontrar padrões de sentimentos e emoções, por meio do uso de técnicas e ferramentas, principalmente da área de mineração de textos. A finalidade dos procedimentos realizados nesse trabalho foi desenvolver um método para poder extrair os perfis emocionais e sentimentais dos comentários dos usuários do jogo *Last Day of June*, onde foram apresentados os resultados e informações extraídas da análise de sentimentos realizada. Foram utilizados também, três algoritmos de classificação, *Naive Bayes*, Máquina de vetor de suporte (*SVM*) e K-Vizinhos mais próximos (*KNN*), para prever a classe dos elementos de acordo com os sentimentos identificados na etapa de análise dos comentários, entre eles, o *SVM* com *kernel Radial* foi o que apresentou a melhor precisão, com 79%, seguido do *KNN* com 3 vizinhos mais próximos, com 75% e por último o *Naive Bayes*, com 62%.

**Palavras-Chaves:** Mineração de Texto. Mineração de Dados. Análise de Sentimentos. Emoções. Afetividade.

## **ABSTRACT**

Every day a large amount of data circulates on the Internet, because it is enormous to the quantities of computers and electronic equipment that have access to the internet, producing information in the form of images, videos and texts, with that, a way of analyzing and extracting these information. Therefore, this work presents concepts and techniques to find patterns of feelings and emotions, through the use of techniques and tools, especially in the area of text mining. The purpose of the procedures performed in this work was to develop a method to extract the emotional and sentimental profiles of the comments of the users of Last Day of June, where the results and information extracted from the analysis of feelings were presented. Three classification algorithms, Naive Bayes, Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) were also used to predict the class of elements according to the feelings identified in the comment analysis step, between the SVM with the Radial kernel was the one that presented the best precision, with 79%, followed by the KNN with 3 nearest neighbors, with 75% and finally the Naive Bayes, with 62%.

**Keywords:** Text Mining. Data Mining. Sentiment Analysis. Emotions. Affectivity.

## LISTA DE SIGLAS

KDD: *Knowledge Discovery in Databases*

KNN: K-Vizinhos mais próximos

NRC: *National Research Council*

SVM: Máquina de Vetor de Suporte

TXT: Arquivo de texto simples

RBF: *Radial Basis Function*

## LISTA DE FIGURAS

<b>Figura 1</b> - Função <i>get_nrc_sentiment</i> . .....	35
<b>Figura 2</b> - Trecho do <i>dataframe</i> DF.Geral.....	36
<b>Figura 3</b> – Gráfico de barras das frequências das palavras.....	40
<b>Figura 4</b> - Nuvem de palavras.....	41
<b>Figura 5</b> - Gráfico Pizza - Porcentagem das emoções.....	42
<b>Figura 6</b> - Porcentagem dos Sentimentos.....	43
<b>Figura 7</b> - Gráfico de barras da frequência dos sentimentos. ....	44
<b>Figura 8</b> - Elementos de um <i>Boxplot</i> . .....	45
<b>Figura 9</b> - <i>Boxplot</i> das 8 emoções. ....	45
<b>Figura 10</b> - Mapa de calor da correlação entre emoções.....	47
<b>Figura 11</b> - Resultados <i>Naive Bayes</i> . .....	49
<b>Figura 12</b> - Resultados <i>Naive Bayes</i> com <i>Kernel</i> . .....	50
<b>Figura 13</b> - <i>SVM Kernels</i> . .....	51
<b>Figura 14</b> - <i>SVM</i> com <i>kernel Linear</i> . ....	51
<b>Figura 15</b> - <i>SVM</i> com <i>kernel Radial</i> . ....	52
<b>Figura 16</b> - <i>SVM</i> com <i>kernel Polynomial</i> . .....	52
<b>Figura 17</b> - <i>SVM</i> com <i>kernel Sigmoid</i> . .....	53
<b>Figura 18</b> - <i>KNN</i> com 3 vizinhos mais próximos. ....	53
<b>Figura 19</b> - <i>KNN</i> com 5 vizinhos mais próximos. ....	54

## SUMÁRIO

<b>SEÇÃO 1.....</b>	<b>12</b>
1.1 PROBLEMÁTICA .....	12
1.2 JUSTIFICATIVA E MOTIVAÇÃO.....	14
1.3 OBJETIVOS .....	16
1.3.1 OBJETIVO GERAL .....	16
1.3.2 OBJETIVO ESPECÍFICO.....	16
1.4 ESTRUTURA DO TRABALHO .....	16
<b>SEÇÃO 2.....</b>	<b>18</b>
2.1 TECNOLOGIA E SOCIEDADE.....	18
2.2 COMPUTAÇÃO AFETIVA.....	21
2.3 MINERAÇÃO DE DADOS .....	23
2.4 MINERAÇÃO DE TEXTO .....	26
2.5 ANÁLISE DE SENTIMENTOS.....	27
<b>SEÇÃO 3.....</b>	<b>30</b>
3.1 PROCEDIMENTOS PREPARATÓRIOS.....	30
3.1.1 ESCOLHA DO JOGO.....	30
3.1.2 COLETA DOS COMENTÁRIOS.....	31
3.1.3 PREPARAÇÃO DO AMBIENTE .....	31
3.1.4 PRÉ-PROCESSAMENTO DO TEXTO.....	32
3.1.4.1 LEITURA DOS COMENTÁRIOS.....	32
3.1.4.2 TRATAMENTO DO TEXTO.....	32
3.1.4.3 TRANSFORMAÇÃO DOS DADOS.....	33

3.2	IDENTIFICAÇÃO DAS EMOÇÕES .....	33
3.2.1	MÉTODO UTILIZADO .....	34
3.2.2	PROCESSO DE IDENTIFICAÇÃO DAS EMOÇÕES.....	35
3.3	ALGORITMOS DE CLASSIFICAÇÃO .....	37
3.3.1	ESCOLHA DOS ALGORITMOS DE CLASSIFICAÇÃO .....	37
<b>SEÇÃO 4.....</b>		<b>39</b>
4.1	RESULTADOS OBTIDOS.....	39
4.1.1	FREQUÊNCIA DOS TERMOS.....	39
4.1.2	NUVEM DE PALAVRAS .....	40
4.1.3	PORCENTAGEM DAS EMOÇÕES.....	41
4.1.4	PORCENTAGEM DOS SENTIMENTOS.....	42
4.1.5	FREQUÊNCIA DAS PALAVRAS POR EMOÇÕES .....	43
4.1.6	DIAGRAMA DE CAIXAS OU <i>BOXPLOT</i> DAS EMOÇÕES .....	44
4.1.7	CORRELAÇÃO DAS EMOÇÕES.....	46
4.1.8	RESULTADOS CLASSIFICADORES.....	48
<b>SEÇÃO 5.....</b>		<b>55</b>
5.1	CONSIDERAÇÕES FINAIS.....	55
5.2	CONTRIBUIÇÕES DO TRABALHO .....	56
5.3	TRABALHOS FUTUROS.....	56
<b>REFERÊNCIAS.....</b>		<b>57</b>

## SEÇÃO 1

### INTRODUÇÃO

Nesta seção, são apresentados os capítulos introdutórios do trabalho, como a problemática, a justificativa e motivação, as contribuições, os objetivos, geral e específico, e por fim, a estrutura, que exibirá como o trabalho está organizado.

---

#### 1.1 PROBLEMÁTICA

Os seres humanos vivem em constantes mudanças sentimentais, causadas por diversos fatores, como as responsabilidades profissionais, relações com seus familiares, amigos, animais de estimação e etc. O bem-estar pessoal está relacionado diretamente com os nossos sentimentos, algumas vezes influenciado por atitudes nossas, ou também, por algo feito por outras pessoas. A realidade é que o ser humano está conectado emocionalmente com tudo que o conhece, e até mesmo o que não conhece, graças a sua imaginação e capacidade de raciocínio.

A empatia tem uma relação direta com as emoções, pois ela permite que as pessoas empatas tenham uma percepção mais exata dos sentimentos alheios, podendo assim compreender melhor o outro e entender suas necessidades e anseios. Nem sempre as pessoas são capazes de perceber a intensidade dos problemas alheios, pois não analisam a situação se imaginando no lugar da pessoa que passa por dificuldades e como se sentiria se estivesse passando pelo mesmo (KRZYNARIC, 2015).

Conforme a tecnologia evolui, mais pessoas vão tendo acesso ao que ela pode proporcionar, aumentando sua presença em variados setores da sociedade, como no ambiente residencial, comercial, pedagógico e etc. Dessa forma, aumenta gradativamente a quantidade de dados e informações geradas por meio do uso dos recursos tecnológicos, onde podem ser extraídos conhecimentos valiosos a partir da

análise desses dados, porém, conforme aumenta o seu volume, mais difícil fica de analisá-lo de forma manual, sem a utilização de técnicas computacionais automatizadas.

A humanidade gradativamente fica mais integrada e conectada, devido principalmente, aos equipamentos multitarefas com acesso à internet, como *smartphones* e computadores, que contribuem em diversos fatores, como possibilitar a comunicação de pessoas de diversos lugares, realizar compras pela internet, utilizar serviços digitais entre outros. Atualmente fica até difícil de imaginar algo em que a tecnologia não esteja presente ou tenha algum tipo de relação (VANZIN, 2018).

Um dos setores que a tecnologia é empregada com o objetivo de entreter as pessoas é o de jogos digitais, tem diversas opções disponíveis, divididas em várias categorias, em que é possível encontrar jogos de ação, aventura, simulação, tiro, corrida entre outros. Dentre essas categorias existem algumas que merecem destaque, pois além do objetivo principal, que é entreter o usuário, também estimulam o desenvolvimento pessoal. Alguns exemplos são os jogos educativos, que ajudam o aprendizado e desenvolvimento educacional dos usuários e os jogos empáticos, que estimulam os jogadores a compreenderem os problemas alheios, como se o próprio jogador estivesse-os enfrentando.

Dessa forma, durante a interação dos usuários com os jogos digitais várias sensações e emoções são sentidas pelos jogadores, devido a capacidade de alguns jogos simularem situações e características da vida humana por meio de recursos gráficos e sonoros. Alguns jogadores utilizam a internet para expressarem suas opiniões sobre os seus jogos, como em redes sociais, sites dos jogos, canais de vídeos entre outros, na maioria das vezes utilizam comentários de texto para fazer isso, pois é uma maneira simples e rápida. Entretanto, apesar dos jogadores relatarem suas experiências por meio de comentários deixados em sites da internet, há uma dificuldade de identificar quais emoções foram sentidas por eles enquanto jogavam algum jogo, pois geralmente as emoções não são citadas nos comentários de maneira explícita, sendo necessário utilizar técnicas computacionais para analisar e identificar, de forma automática, padrões emocionais presentes no texto.

## 1.2 JUSTIFICATIVA E MOTIVAÇÃO

Sabe-se que, atualmente, existem várias alternativas tecnológicas para ajudar e entreter a população, entre uma dessas, estão os jogos eletrônicos, desenvolvidos para atenderem um vasto e variado mercado consumidor, em que não são mais considerados como apenas um simples meio de diversão, estão mais complexos e realistas, proporcionam uma imersão maior do jogador com o jogo, por meio de recursos gráficos e efeitos sonoros avançados.

Existem acessórios produzidos especificamente para esse mercado do entretenimento, como os óculos de realidade virtual, controles com reconhecimento de movimentos, fones de ouvidos preparados para melhorar a experiência sonora nos jogos e etc. Esse conjunto de fatores e tecnologias agregados aos jogos digitais além de garantirem uma maior realidade e qualidade de imagens e sons, contribuem também para os jogos direcionados ao desenvolvimento pessoal serem mais eficazes, como os jogos terapêuticos e pedagógicos.

A experiência proporcionada ao usuário é muito importante, por isso os desenvolvedores de serviços digitais e jogos sempre criam um canal em que os clientes possam deixar suas avaliações, críticas e sugestões, sendo possível identificar o que está agradando ou desagradando os clientes, relatar problemas que não foram percebidos pela equipe de desenvolvimento e outros diversos relatos de usuários, também de experiências sentidas durante o tempo em que o produto ou serviço foi utilizado.

Apesar de existirem esses meios de feedback para os usuários expressarem suas opiniões por meio de mensagens, nem sempre é possível identificar explicitamente as emoções contidas nelas, em alguns casos o autor do comentário usa palavras com relação direta ao seu sentimento, como por exemplo, quando descreve que está feliz por ter utilizado um produto ou jogado um jogo, deixando claro que o sentimento é de felicidade, porém, em outras situações não é tão simples identificar o perfil emocional da mensagem, por não possuir algum termo que remeta claramente a algum sentimento, também há situações que não é possível fazer essa identificação emocional do texto.

Nos jogos empáticos, que tem por objetivo simular aos jogadores experiências e sensações vividas por outras pessoas e que colaborem para que, por meio de uma experiência virtual, o jogador possa imaginar-se em situações de dificuldades e como lidar com elas, por meio de decisões que devem ser tomadas no decorrer do jogo, é interessante realizar a análise emocional nas mensagens enviadas pelos jogadores, pois é possível identificar quais sentimentos foram mais estimulados durante a interação com o jogo, com isso, podendo verificar-se as emoções predominantes e outras estatísticas relacionadas, e também, examinar o impacto que a narrativa causou na forma de pensar do jogador.

Portanto, é preciso ser feita uma reflexão acerca de um assunto importante, os possíveis efeitos que os avanços tecnológicos causam no comportamento humano, a grande dependência dos meios tecnológicos pela grande maioria da população (PAREDES, 2019), a falta de empatia cada vez menor das pessoas, sendo apontado por algumas pessoas, pelo crescimento do uso das redes sociais (MOURA, 2017). Os jogos digitais também são alvos de estudos e debates, em que são questionados os efeitos que eles causam em relação a forma de pensar e comportamento humano, principalmente em crianças e adolescentes (SARMET e PILATI, 2016). Por isso algumas empresas criam alternativas para estimular os jogadores a agirem e pensarem de forma correta e consciente, por meio de jogos educacionais e que estimulem a prática do bem. Com isso, fica evidente a necessidade de realizar estudos por meio de coletas e análises de dados referentes à influência exercida pelos jogos digitais nos seus usuários.

Existem vários métodos que podem ser utilizados na coleta e tratamento de dados textuais durante a análise afetiva dos comentários de usuários de jogos digitais, uma delas é a mineração de texto, que pode ser feita de forma manual ou automatizada. Em alguns casos, em que a quantidade de dados coletados e analisados é pequena, torna-se possível o tratamento textual ser feito de forma manual, porém na maioria das situações a quantidade de dados a ser estudada é grande, de forma que fique inviável ser feita dessa maneira, demandando bastante tempo para isso. As técnicas automatizadas são as mais utilizadas e recomendadas, por utilizarem ferramentas computacionais que realizam o processamento e tratamento dos dados textuais de maneira ágil e automática, sendo necessário apenas

que o usuário configure a ferramenta para realizar o trabalho, de acordo com o seu objetivo desejado.

### **1.3 OBJETIVOS**

Nesta seção são apresentados os objetivos, geral e específicos, que o trabalho busca atingir.

#### **1.3.1 OBJETIVO GERAL**

O objetivo geral é realizar a análise afetiva do texto através da mineração de dados, com a finalidade de identificar as emoções contidas em comentários de usuários de jogos empáticos em páginas da internet, possibilitando conhecer e apresentar informações referentes aos padrões emocionais identificados no texto.

#### **1.3.2 OBJETIVO ESPECÍFICO**

Os objetivos específicos são:

1. Analisar os padrões emocionais presentes em comentários de usuários do jogo empático previamente escolhido, coletados a partir de páginas da internet;
2. Avaliar a técnica mais adequada para realizar a classificação das mensagens de acordo com as emoções básicas predefinidas no trabalho;
3. Apresentar informações referentes as frequências das emoções identificadas nos comentários dos usuários;
4. Avaliar o uso de algoritmos de classificação para a identificação das emoções nos comentários;
5. Elaborar e apresentar, por meio de gráficos, comparações entre as emoções identificadas nos comentários.

### **1.4 ESTRUTURA DO TRABALHO**

Este trabalho apresenta sua estrutura dividida em cinco seções, que além dessa seção inicial a qual essa subseção faz parte, estão ordenadas da seguinte maneira:

**Seção 2:** Nesta seção são abordados conceitos e informações acerca da relação da tecnologia com sociedade e também sobre alguns assuntos importantes referentes ao propósito do trabalho, como computação afetiva, mineração de dados, mineração de texto e análise de sentimentos.

**Seção 3:** Nesta seção são apresentados os procedimentos realizados nos dados coletados para a análise de sentimentos, desde a coleta dos dados até a identificação das emoções nos comentários, também, são explicados os conceitos referentes aos algoritmos de classificação.

**Seção 4:** Esta seção apresenta os resultados obtidos da análise dos sentimentos nos comentários dos usuários.

**Seção 5:** Nesta seção são apresentadas as considerações finais acerca de tudo que foi abordado no trabalho, assim como, propostas de melhorias nos métodos apresentados e utilização dos conceitos e técnicas em outras áreas.

## SEÇÃO 2

### FUNDAMENTOS ACERCA DO PROCESSO DE ANÁLISE DE SENTIMENTOS

Nesta seção, são apresentados os conceitos ligados às atividades e procedimentos utilizados na análise de sentimentos, assim como, sobre a relação que a tecnologia exerce na sociedade atual.

---

#### 2.1 TECNOLOGIA E SOCIEDADE

A tecnologia é a responsável por colaborar em diversos fatores na vida do ser humano, como no combate e prevenção às doenças, em que equipamentos e máquinas avançadas são capazes de detectar alterações no organismo humano. Algumas formas de prevenção às doenças são feitas por meio de análises sanguíneas, exames por imagens, como tomografia computadorizada e raio-X entre outras.

A tecnologia geralmente é utilizada para criar soluções para melhorar a vida humana, entretanto, nem sempre é possível avaliar os efeitos negativos que podem ser causados em algumas situações, por isso geralmente são feitos estudos e testes para garantir que o ser humano não seja prejudicado.

Com a facilidade de acesso aos meios tecnológicos como computadores, celulares, internet entre outros, pessoas com más intenções utilizam esses meios com outras finalidades, para obterem vantagens de forma ilegal, como roubo e venda de dados pessoais pela internet, falsificação de documentos, venda de produtos proibidos e diversos outros crimes cometidos com a ajuda da internet e equipamentos eletrônicos. É importante sempre analisar como a tecnologia está sendo utilizada, principalmente no meio familiar, de maneira que os impactos sociais resultantes do uso de equipamentos e serviços eletrônicos possam ser medidos, com o objetivo

de equilibrar a relação do meio real com o meio digital, com isso, evitar excessos que possam provocar problemas pessoais (SOUZA, 2016).

De acordo com Boechat (2017), junto com o avanço do desenvolvimento tecnológico, aumenta também as relações das famílias com os meios digitais, criando a necessidade de inserção e adaptação a esses meios, onde pessoas que antes não tinham um contato maior com os adventos tecnológicos passam a sentir a necessidade de se adaptarem a esse novo estilo de vida.

Nessa nova forma de viver, a presença da tecnologia é grande, com isso, muda-se a forma como os membros da família se relacionam, modificando comportamentos e atitudes interpessoais. A autora também deixa claro que, com o avanço tecnológico e conseqüentemente o desenvolvimento dos meios de comunicação e informação, desde o momento em que a informação se tornou global, alguns conceitos e paradigmas existentes sobre a família e sociedade por exemplo, foram e continuam sendo modificados, passando a terem novas características e atribuições, com a ajuda da internet, mídias sociais e outros meios de comunicação digitais.

É importante ressaltar que a sociedade atual vive na era da tecnologia e da informação e fugir dessa realidade é quase impossível, mesmo que algumas pessoas tentem evitar o uso ou contato de aparatos e serviços digitais, ainda assim, em certas ocasiões é inevitável utilizá-los, pois estão presentes em quase todas as áreas, como na de serviços, lazer, educação, econômica e em diversos outros meios essenciais para a vida humana. Ainda há alguns povos nativos de algumas regiões remotas do planeta que podem ser considerados excluídos desse mundo digital, que não possuem qualquer contato com os recursos disponíveis na sociedade globalizada, certamente, diferenciando-os radicalmente na forma de viver e pensar das pessoas que vivem nas cidades.

Engana-se quem acha que a tecnologia surgiu recentemente, que está ligada somente aos aparelhos e equipamentos modernos, desde o tempo primitivo o ser humano já desenvolvia formas e equipamentos para sobreviver, como as armas e ferramentas feitas de ferro ou pedras para uso nas caças e na manipulação dos alimentos. Somente nos últimos dois séculos que a tecnologia foi avançar em um ritmo

acelerado e de forma global, impulsionada pela criação do computador que, posteriormente, com a ajuda dos meios de comunicação como internet e celular, alavancaram mais ainda o desenvolvimento tecnológico, causando uma transformação na sociedade e conectando pessoas de todo o mundo (OLIVEIRA, CASAGRANDE e GALERANI, 2016).

Pode-se dizer que o ser humano está em constante evolução e, graças a sua capacidade de raciocínio e inteligência ele vai criando e aperfeiçoando meios para contribuir de diversas formas com o desenvolvimento da sociedade. Neste contexto, fica claro que a tecnologia tem um papel fundamental nesse cenário, pois ela ajuda em diversos fatores, como a utilização de equipamentos para a comunicação remota em tempo real, máquinas que realizam trabalhos com mais eficiência que o trabalho manual, em alguns casos sendo até impossível ser feito somente com a força humana.

O mais preocupante, contudo, é constatar que com a evolução tecnológica e as facilidades que ela proporciona, problemas surgem como efeito desse avanço rápido, em que devemos analisar e estudar formas alternativas que possam amenizar ou solucionar esses impactos que afetam a sociedade em geral, seja de forma direta ou indireta. Não é exagero afirmar que a sociedade contemporânea está cada vez mais integrada ao mundo digital e provavelmente isso é irreversível, é importante que o ser humano tenha consciência social, ética e ambiental e a capacidade de sempre avaliar os riscos antes de ser realizada qualquer ação. Porém, preocupa o fato de que algumas pessoas e empresas não se importam com os impactos resultantes de suas ações, isso porque na maioria das vezes só visam o benefício próprio, desconsiderando o bem da sociedade e do meio ambiente.

É importante ser debatido e estudado como a sociedade está se transformando devido a vários fatores, sendo um deles a presença massiva da tecnologia em quase todos os meios. É um assunto bem relevante, pois essa transformação tem impactos diretos não somente na vida humana, mas também na fauna e flora do planeta Terra, visto que o ser humano depende delas para sobreviver. Há muitos temores acerca do futuro da humanidade, devido ao grande alcance das criações tecnológicas nos meios sociais, principalmente na parte mais jovem da população, que é atraída por equipamentos eletrônicos, que podem realizar diversas tarefas e proporcionar opções

de entretenimento, como acessar internet, reproduzir músicas, acessar redes sociais digitais, entre outros (REIS, 2018).

A tecnologia moderna é uma das mais maravilhosas conquistas do homem e mostra em milhares de aspectos, o que podemos e poderemos fazer com o uso da inteligência para o bem da humanidade.

Mas o uso da tecnologia não pode deixar de ser permanentemente vigiado pelo nosso bom senso, para não gerar distorções. Os resíduos produzidos pelos inúmeros processos de transformação das nossas indústrias foram ignorados desde o início dessas atividades humanas e só foram considerados depois que os danos à natureza já eram enormes (SOUZA, 2016, p. 37).

O autor deixa claro na citação acima a importância da tecnologia para o desenvolvimento da sociedade e, conseqüentemente, contribuir com seu bem-estar. Esse é o motivo pelo qual é relevante frisar esse ponto, uma vez que, evidencia o objetivo principal do avanço tecnológico, que é a contribuição com a evolução humana de maneira positiva. Conforme citado acima não se deve desconsiderar os efeitos prejudiciais que a má utilização da tecnologia pode gerar, analisando sempre se está ocorrendo a presença de impactos negativos ou se ocorrerá no futuro.

Fica evidente, diante do exposto, que a sociedade vive em uma época que a tecnologia e seus meios tornaram-se essenciais para o ciclo do desenvolvimento humano, estando presente em variados setores da sociedade. Dessa forma, surgem os questionamentos acerca das transformações que estão ocorrendo na humanidade e os efeitos que elas podem gerar na vida humana, positivamente ou negativamente. Devido a grande capacidade de raciocínio humano, são criadas constantemente soluções para melhorar a vida em nosso planeta de uma forma geral, e também, graças a inteligência do ser humano é possível avaliar o que está errado para que possa ser corrigido ou alterado para que o mínimo de impacto seja gerado no planeta Terra e, conseqüentemente, nos seres que nele habitam.

## **2.2 COMPUTAÇÃO AFETIVA**

Antes da abordagem sobre computação afetiva, é necessário esclarecer alguns pontos referentes a afetividade, pois são essenciais para melhor compreender o que será abordado neste capítulo. De acordo com Almeida e Mahoney (2004), a afetividade é o conjunto de emoções e sentimentos causados pela influência que o ser humano e o meio em que vive exercem mutuamente, porém, a estrutura orgânica humana também tem relação direta com a percepção das emoções, até mesmo nos

primeiros anos de vida, em que a criança tem necessidades oriundas da fisiologia humana, como a percepção de dor, fome e várias outras sensações. Na maioria das pessoas, conforme elas vão desenvolvendo-se e, conseqüentemente, aprimorando suas percepções sensoriais e a capacidade de raciocínio, as emoções e os sentimentos ficam mais evidentes e naturais, devido a convivência e aprendizagem com o meio em que ela faz parte.

Os estudos e implementações feitos para que características afetivas humanas sejam processadas e interpretadas computacionalmente fazem parte da área denominada de computação afetiva, em que são utilizadas algumas técnicas, como o aprendizado de máquina, para reconhecer padrões afetivos em dados a partir da comparação com modelos de emoções previamente definidos. Também é possível simular aspectos emocionais humanos em ambientes computacionais, a exemplo dos *emoticons*, que são ícones com formas que remetem ao rosto humano, e de alguns jogos virtuais, que buscam retratar essas características em personagens virtuais (BEHAR, 2009).

Como bem nos assegura Picard (2000), o reconhecimento, compreensão e manifestação de emoções pelos computadores é essencial para que se tornem máquinas mais inteligentes, aperfeiçoando e tornando mais natural a comunicação entre o ser humano e o computador. Ainda segundo a autora, os computadores afetivos, como ela define, utilizam equipamentos de captura de imagens, áudios e sinais corporais que junto com o uso de ferramentas que processam e interpretam esses dados, auxiliam no reconhecimento de padrões emocionais, entretanto, existem muitos empecilhos que impedem um reconhecimento mais amplo das emoções humanas, como variações das intensidades emotivas, distinção das emoções, problema no reconhecimento das ações que estimularam as emoções entre outros, dificultando o mapeamento mais amplo e eficiente das emoções.

Conforme explicado anteriormente, mesmo com as dificuldades ainda encontradas para o reconhecimento mais fiel das emoções humanas pelos computadores, a computação afetiva já é utilizada em diversas áreas como na educação infantil, com a utilização de personagens animados que simulam características e comportamentos humanos, com a finalidade de atrair a atenção da criança no ambiente de aprendizagem. A computação afetiva, assim como a

inteligência artificial, são áreas bastante promissoras, pois é crescente a utilização de equipamentos eletrônicos pela população, principalmente por crianças e adolescentes, que passam a maior parte do tempo em contato com *smartphones*, relógios inteligentes e computadores, criando relações sociais e afetivas com outras pessoas e também pelos equipamentos e aplicativos que possuem.

De acordo com Behar (2009, p. 206):

Os estudos apresentados por Bercht (2001) e por Jaques e Vicari (2005) demonstram que o reconhecimento dos fenômenos afetivos e sua aplicação em sistemas computacionais resultam em sistemas adaptativos e funcionalmente mais eficientes nas interações humanas mediadas por computador. Modelos computacionais para tratamento dos fenômenos afetivos, mais precisamente das emoções do usuário, vêm sendo desenvolvidos e empregados principalmente em ambientes inteligentes de aprendizagem direcionados ao domínio de conhecimento [...].

### **2.3 MINERAÇÃO DE DADOS**

A mineração de dados consiste na busca por conhecimento em volumes de dados estruturados ou não estruturados, onde os estruturados são organizados e rotulados de acordo com as particularidades de cada dado, como tipo e quantidade, por exemplo, em uma planilha de informações de funcionários de uma empresa, em que as colunas são referentes as categorias como nome, cargo, endereço, entre outras, e as linhas são os dados de cada funcionário de acordo com as categorias (colunas). Os dados não estruturados são aqueles que não possuem esse tipo de organização dos dados estruturados, da maneira que, na maioria das vezes, possuem diversas informações dispersas em um só lugar, como encontramos nos textos, imagens e vídeos publicados em uma rede social ou outro site da internet. Essa busca é feita por meio de vários processos de manipulação dos dados, com intuito de prepará-los para serem analisados e gerarem informações relevantes à pesquisa (AMARAL, 2016a).

A mineração de dados faz parte de uma área mais abrangente que procura encontrar conhecimentos, ainda não revelados, em grandes quantidades de dados, essa área é denominada de Descoberta de conhecimento em base de dados ou KDD (*Knowledge Discovery in Databases*), em que suas etapas consistem em separar ou armazenar os dados a serem analisados, selecionar somente as partes úteis deles, ou seja, o que realmente interessa para ser explorado. Posteriormente os dados são

preparados para serem analisados, são organizados e padronizados para a próxima etapa, e também, os dados sem importância para a mineração são descartados, restando somente o que pode ser utilizado para gerar os conhecimentos desejados (SILVA, BOSCARIOLI e PERES, 2017).

Após a fase de preparação dos dados, são executadas as atividades de mineração de dados, que tem por objetivo extrair os conhecimentos a partir dos dados já pré-processados.

Existem várias maneiras de realizar o tratamento dos dados coletados para a atividade de mineração, como executar a limpeza ou descarte de elementos irrelevantes ou fora do contexto do estudo, transformações nos documentos para melhorar o desempenho computacional durante o processamento, classificação dos dados, como por exemplo, em categoria ou frequência, para facilitar no momento de analisá-los e diversas outras ações que podem ser realizadas com objetivos diversos. A mineração de dados pode ser utilizada em diversos contextos e com finalidades variadas, tal como na classificação de *e-mails* por categorias, detecção de *spams*, recomendações de produtos baseadas em compras anteriores ou buscas realizadas em sites etc. A tarefa de minerar dados também é utilizada para conhecer melhor acerca de algo, seja por meio de recursos textuais ou gráficos, disponíveis em ferramentas de mineração de dados, um exemplo disso é quando os dados são agrupados e exibidos estatisticamente em um gráfico, dividido de acordo com a porcentagem de cada grupo (CASTRO e FERRARI, 2016).

Na mineração de dados também é possível utilizar algoritmos de classificação para atribuir uma classe a um elemento ainda não classificado, usando técnicas de aprendizado de máquina, que analisam bases de dados já classificadas para relacionarem as características em comum entre elementos classificados e não classificados, e assim, prever a classe dos dados sem classificação (CASTRO e FERRARI, 2016).

Conforme explicado acima, a mineração de dados é fundamental quando o intuito é encontrar conhecimentos a partir de um determinado volume de dados, principalmente em grandes quantidades, pois as diversas operações e soluções disponíveis nas ferramentas utilizadas para minerar dados permitem que sejam

obtidos resultados que colaborem com a finalidade desejada, de uma forma eficiente e rápida se comparada a outros métodos, tal como o manual, em que o usuário teria que analisar os documentos individualmente ou em pequenos grupos, e dependendo da dimensão de dados utilizada, seria bastante custosa ou até impossível ser realizada dessa forma. Muitas empresas e profissionais já utilizam a mineração de dados como ajuda na tomada de decisão nos seus negócios, pois permite que possam ter um entendimento melhor a respeito dos seus clientes, produtos, funcionários entre outros, de acordo com as suas necessidades.

A mineração de dados tem sua aplicação cada vez mais difundida em áreas que antes sequer poderíamos imaginar uma aplicação prática, pois eram modelos de negócios em que os dados não se encontravam armazenados digitalmente. Um exemplo é a análise de sentimento utilizado por empresas para avaliar a reputação da empresa em redes sociais. Além das tradicionais aplicações em marketing, hoje a mineração de dados e o aprendizado de máquina entram aplicações na medicina, educação, processamento de linguagem natural, bioinformática, detecção de fraude, reconhecimento de fala, finanças, robótica, sistemas de recomendação, mineração de texto entre muitos outros (AMARAL, 2016a, p. 3-4).

Conforme verificado, na atividade de mineração de dados busca-se encontrar conhecimentos que não estão presentes de forma explícita em uma ou mais bases de dados analisadas, que somente torna-se possível por meio da utilização de ferramentas computacionais que realizam processos de tratamento e análise desses dados, de maneira que se consiga compreender esses dados mediante a obtenção de novos conhecimentos.

Minerar dados já se tornou comum em diversas empresas de variados setores, incorporando a atividade no processo de trabalho. A mineração de dados possui várias finalidades, podendo ser utilizada em qualquer área que se possa imaginar, desde que existam dados necessários à serem analisados e compreendidos, sendo bastante útil principalmente para o setor comercial, mas também é importante para a população de maneira geral, pois é a partir de informações obtidas na mineração de dados que é possível conhecer melhor as necessidades e preferências da sociedade. Assim, as empresas podem oferecer produtos e serviços personalizados de acordo com cada público-alvo, também é possível filtrar e bloquear conteúdos inoportunos por parte de algumas empresas e pessoas, assim como impedir golpes, fraudes ou outro tipo de cibercrime contra instituições e usuários.

## 2.4 MINERAÇÃO DE TEXTO

A mineração de texto é a coleta e união de dados textuais de uma ou mais fontes em um documento que geralmente é chamado de *corpus*, em que são feitas diversas operações e tratamentos nesses dados, com o intuito de revelar estatísticas e padrões presentes nos textos de acordo com uma finalidade desejada, como por exemplo, descobrir quais são os produtos mais buscados na internet de acordo com uma categoria específica ou analisar qual o assunto ou termo está sendo mais comentado em determinada rede social em um período específico. A mineração de texto possui outros inúmeros propósitos, que vão depender do objetivo desejado, sempre realizando ações sobre um *corpus* buscando atingir as metas e resultados esperados. As principais ações feitas no *corpus* são com a finalidade de remover termos que não tenham importância para a finalidade proposta, outra ação bastante comum na mineração de texto é a realização da contagem da frequência em que uma ou mais palavras aparecem no *corpus*, esses e outros processos fazem parte do pré-processamento do texto (SILVA, BOSCAROLI e PERES, 2017).

A etapa do pré-processamento do texto do *corpus* é importante, pois é o momento que serão definidas as operações que serão executadas para realizar a limpeza dos elementos desnecessários para o estudo realizado, como a remoção de *stopwords*, que são palavras irrelevantes para a mineração de texto, geralmente cada idioma possui uma lista com suas próprias *stopwords*, porém o usuário pode criar sua lista personalizada e utilizá-la em vez de usar as já pré-existentes. Portanto, se em alguma etapa realizada forem removidos termos essenciais para o processo de análise do documento, certamente a qualidade dos resultados será comprometida. Um exemplo disso é remover a pontuação em um contexto que ela tenha relevância para os resultados, de modo que sua remoção gere resultados diferentes do esperado (SILVA, BOSCAROLI e PERES, 2017).

De certa forma, quase tudo o que for feito durante o pré-processamento do texto implicará nos resultados da análise, portanto é importante planejar e verificar antecipadamente quais ações serão realizadas de acordo com o objetivo desejado, para alcançar resultados mais precisos e reduzir a ocorrência de erros. Conforme

explicado acima, uma ação realizada de forma errada durante o tratamento do texto pode comprometer parcialmente ou totalmente a qualidade do estudo.

De acordo com Rezende (2005, p. 338):

Mineração de Textos é um conjunto de técnicas e processos que descobrem conhecimento inovador nos textos. Ela está sendo empregada atualmente em projetos de diversas áreas, por exemplo, para descobrir fatos na genética e na pesquisa de proteínas. [...] Aplicações não tão pretensiosas já se encontram implantadas, como a categorização automática de mensagens de correio eletrônico em bancos de investimento e a extração automática de resumos a partir de documentos pesquisados, realizada por alguns mecanismos de busca na web.

Fica evidente, diante do exposto, que a mineração de texto pode ser utilizada para diversas finalidades, colaborando para revelar informações e padrões presentes em textos de um ou mais locais, que antes, sem a ajuda da mineração, não era possível identificá-los. Em alguns casos, torna-se extremamente necessária a utilização da mineração de textos, como quando a quantidade textual a ser analisada tem um tamanho muito grande, sendo inviável ser feita uma análise de forma manual e sem a ajuda de processos computacionais automatizados, que realizam esse serviço de maneira rápida e eficiente.

## **2.5 ANÁLISE DE SENTIMENTOS**

A análise de sentimentos, também chamada por alguns autores de mineração de opiniões, é um ramo da mineração de dados que atua na identificação de sentimentos e emoções extraídas de textos, com objetivo de compreender as características sentimentais dos usuários ao expressarem suas opiniões, isso é feito por meio de métodos que analisam os elementos do texto e comparam com padrões emocionais predefinidos (COOPER e SCHINDLER, 2016). Compreender o que as pessoas comentam na internet é importante, por isso empresas e pesquisadores trabalham para desenvolverem ferramentas computacionais que consigam identificar a subjetividade humana em textos produzidos em diversos lugares, como redes sociais, sites de avaliações de produtos e serviços, fóruns de discussões entre outros, por essa razão, muitas empresas já utilizam a análise de sentimentos para identificar os aspectos emocionais e subjetivos de cada cliente, assim, conseguindo oferecer soluções mais próximas do que eles desejam (CASTRO e FERRARI, 2016).

Todo e qualquer texto publicado na internet tem sua relevância, pois cada um possui características e opiniões distintas e indica o posicionamento do usuário acerca de um determinado assunto, na maioria das vezes são utilizadas palavras que expressam determinados sentimentos do autor.

Algumas empresas utilizam a análise de sentimentos para descobrir padrões presentes nos textos presentes na internet, comparando os textos publicados pelos usuários com um dicionário de termos, em que possui várias palavras, sendo cada uma ligada a uma ou mais emoções, com isso, as palavras do textos que estiverem presentes no dicionário de termos poderão ser relacionadas a emoções, podendo identificar as emoções presentes no texto e a frequência de cada uma (SOLOMON, 2016).

Na análise de sentimentos são utilizadas técnicas comuns da mineração de textos e da inteligência artificial, como respectivamente, os procedimentos de coleta e tratamento do texto e uso de aprendizagem de máquina para encontrar perfis emocionais a partir de modelos já treinados e rotulados (GUIMARÃES, 2017). A análise de sentimentos pode ser usada para diversas finalidades, como na avaliação de emoções em redes sociais (AGUIAR, FAIÇAL, *et al.*, 2018), classificação emotiva de letras de músicas (SOUZA e CAFÉ, 2018) entre outros. Conforme explicado acima, as empresas, principalmente, cada vez mais adotam a análise de sentimentos às suas práticas, pois é de extrema importância não só conhecer o que seus clientes escrevem sobre seus produtos, mas também entender os sentimentos envolvidos nos textos que são escritos.

Conforme explicam Cooper e Schindler (2016, p. 107):

O campo emergente da *análise de sentimentos e mineração de opiniões* fornece-nos meios de aplicar tratamento computacional a opiniões, sentimentos e subjetividade em forma textual. Sua meta é decodificar a atitude de quem fala ou escreve sobre algum tópico. [...] A dificuldade da maioria dos algoritmos de sentimentos é a dependência de palavras-chave simples para expressar sentimentos complexos sobre um produto, filme ou serviço. São os fatores culturais, a obscuridade linguística e os contextos que tornam difícil a transformação do texto escrito em sentimento digitalizado.

Fica evidente que a atividade de análise de sentimentos fica cada vez mais frequente, principalmente no ramo empresarial, pois a sociedade vive em uma época que grande parte da população utiliza bastante aparelhos conectados à internet para

se comunicar, assim, gerando uma grande quantidade de dados e informações, na maior parte em forma de texto, portanto, torna-se necessário o uso de técnicas que possam extrair informações que não estão de maneira explícitas nos textos, como a presença de sentimentos dos usuários. O avanço de outras áreas como a inteligência artificial e a mineração de dados, que estão diretamente ligadas à análise de sentimentos, contribui e continuará contribuindo para que os textos da internet gradativamente tornem-se fontes de conhecimentos valiosos e úteis para a sociedade, resultando em uma comunicação mais eficiente.

## SEÇÃO 3

### APRESENTAÇÃO DOS PROCEDIMENTOS REALIZADOS DURANTE A ANÁLISE DE SENTIMENTOS

Esta seção descreve os procedimentos realizados para extrair as emoções dos comentários, desde a coleta e preparação dos dados até a identificação das emoções. No último capítulo da seção são apresentados os algoritmos de classificação escolhidos no trabalho.

---

#### 3.1 PROCEDIMENTOS PREPARATÓRIOS

Nessa seção são apresentados os procedimentos realizados nos dados antes de serem utilizados na análise de sentimentos, pois, é necessário prepará-los para descartar aqueles que não tenham utilidade para o objetivo da análise afetiva.

##### 3.1.1 ESCOLHA DO JOGO

A escolha do jogo foi feita após algumas pesquisas, com o objetivo de encontrar um jogo empático que, por meio de sua história, tivesse um grande apelo emocional e potencial capacidade de estimular vários tipos de sentimentos em seus jogadores. O escolhido foi *Last Day of June*, em que apresenta um enredo em que um casal sofre um acidente de carro, em que o marido, chamado *Carl*, perde seus movimentos e fica em uma cadeira de rodas e sua esposa *June* não consegue sobreviver. No decorrer do jogo são apresentados alguns desafios a serem realizados pelo jogador, com o intuito de mudar o desfecho do dia em que aconteceu o acidente e evitar a morte de *June* (CHICKEN, 2017).

### 3.1.2 COLETA DOS COMENTÁRIOS

Para realizar a identificação das emoções foram coletados 110 comentários de usuários do jogo digital *Last Day of June*<sup>1</sup>, essa quantidade é referente ao total de comentários que foram encontrados no idioma português do Brasil durante a pesquisa. A maior parte do texto coletado foi a partir da página do jogo na plataforma de jogos *Steam*, na seção de análises, que é destinada aos usuários para expressarem suas opiniões referentes aos jogos que possuem na plataforma, mas também a coleta foi feita no *Twitter*<sup>2</sup> e *Youtube*<sup>3</sup>.

### 3.1.3 PREPARAÇÃO DO AMBIENTE

Para a análise afetiva dos comentários coletados, foi preciso preparar o ambiente, para que fosse possível realizar o desenvolvimento de um *script* na linguagem de programação R, e assim, utilizar as técnicas necessárias para realizar a análise dos sentimentos. Foram utilizadas as ferramentas *RStudio*<sup>4</sup> (versão 1.1.456), junto com o *R*<sup>5</sup> (versão 3.6), ambas são para trabalhar com a linguagem de programação R, com diversas funcionalidades disponíveis, por meio de funções e bibliotecas que podem ser utilizadas para realizar análises de dados.

As linguagens de programação *R* e *Python* estão entre as mais populares para se trabalhar com análise de dados, a primeira tendo como foco principal essa área, já a segunda, é mais abrangente e não tem um foco específico, tendo sua utilização em várias áreas, sendo uma delas a análise de dados (CANTO, 2017).

---

<sup>1</sup>< [https://store.steampowered.com/app/635320/Last\\_Day\\_of\\_June](https://store.steampowered.com/app/635320/Last_Day_of_June)>

<sup>2</sup><<https://twitter.com/>>

<sup>3</sup><<https://www.youtube.com/>>

<sup>4</sup><<https://www.rstudio.com/>>

<sup>5</sup><<https://www.r-project.org/>>

*Python* e *R* se destacam no ramo da mineração de dados pelo fato de possuírem diversas funcionalidades e ferramentas disponíveis, que podem ser usadas por meio de pacotes, sendo baixados ou carregados ao ambiente computacional (MARQUESONE, 2016).

### 3.1.4 PRÉ-PROCESSAMENTO DO TEXTO

A seguir serão apresentados alguns procedimentos realizados no *RStudio*, com o objetivo de preparar os textos para serem usados de forma mais eficiente no decorrer do desenvolvimento da solução. O principal pacote do R utilizado nas próximas etapas é o *tm*<sup>6</sup>, que é um dos pacotes pré-carregados do R e tem inúmeras funções direcionadas a atividade de mineração de texto.

#### 3.1.4.1 LEITURA DOS COMENTÁRIOS

Foram coletados 110 comentários de usuários do jogo *Last Day of June*, organizados e numerados de 1 a 110 em um arquivo de texto simples (TXT), em que cada linha corresponde ao comentário de um usuário. Foi utilizada a função *read\_lines*, para realizar a leitura do arquivo com os comentários e deixá-los prontos para serem manipulados dentro da aplicação, essa função faz parte do pacote chamado *readr*, que é utilizado para leitura de diversos tipos de dados.

#### 3.1.4.2 TRATAMENTO DO TEXTO

Uma das etapas principais do pré-processamento de texto é a remoção de termos que não tenham relevância para o objetivo proposto, como as *stopwords*, que são palavras que não tenham valor semântico, geralmente são palavras conectivas (com, para, e, a). Outros elementos do texto que geralmente são eliminados são os pontos, números, caracteres especiais entre outros, porém, essas etapas variam de acordo com a finalidade do estudo (AMARAL, 2016b).

---

<sup>6</sup> <<http://tm.r-forge.r-project.org/>>

A lista de *stopwords* removidas nesse trabalho foi uma personalizada encontrada no site *GitHub*<sup>7</sup>, possuindo 560 termos do idioma português que não tem valor semântico para o texto, como conectivos, além dos termos da lista, mais algumas palavras sem valor para a análise foram removidas manualmente. Também foram removidos os números, pontuação e espaços desnecessários no texto, como espaços duplicados, além da transformação das palavras, para ficarem totalmente em letras minúsculas.

### 3.1.4.3 TRANSFORMAÇÃO DOS DADOS

Nesta etapa, após serem feitas as modificações, utiliza-se a função *VCorpus* para salvar os termos restantes em um novo documento, denominado de *corpus*, que é um documento que reúne textos de uma ou mais fontes, com o intuito de prepará-lo para ser analisado. A primeira tarefa a ser realizada é a conversão do documento em uma matriz de frequência de termos, por meio da função *TermDocumentMatrix*. Depois é calculada a quantidade total que cada termo aparece no documento, assim, criando uma lista com os termos e os valores de suas respectivas frequências, organizada em ordem decrescente, de acordo com os valores das frequências. Ao final, os dados da frequência dos termos foram colocados em um documento do tipo *dataframe*, que é um formato de documento do R em que os dados são organizados em linhas e colunas, assim como em uma matriz, a vantagem desse tipo de documento é a possibilidade da utilização de dados de diferentes tipos de atributos, como números e caracteres no mesmo *dataframe*, também é mais fácil de realizar uma consulta dos termos presentes nele.

## 3.2 IDENTIFICAÇÃO DAS EMOÇÕES

Nessa etapa, depois de já ter sido realizada a fase de pré-processamento do texto, são executadas as ações necessárias para a obtenção dos padrões

---

<sup>7</sup> <<https://github.com/stopwords-iso/stopwords-pt>>

emocionais presentes nos comentários dos usuários e organização dos dados, para posteriormente, serem feitas as análises finais e apresentação dos resultados obtidos.

### 3.2.1 MÉTODO UTILIZADO

Para realizar a análise afetiva dos comentários, foi utilizado o pacote do *R* chamado *syuzhet*<sup>8</sup>, que disponibiliza algumas funções úteis para a identificação das emoções presentes em textos, entre elas, tem a função chamada *get\_nrc\_sentiment* que usa um dicionário de termos, denominado de *NRC Emotion Lexicon*<sup>9</sup>, que associa palavras à emoções e sentimentos, para realizar a comparação das palavras e identificar as emoções e sentimentos presentes no texto analisado. O dicionário possui aproximadamente 14000 palavras em 105 idiomas diferentes traduzidos a partir do inglês (idioma padrão do dicionário), entre eles o português do Brasil, a tradução foi feita pelos autores do dicionário com a ajuda da ferramenta *Google Translate*<sup>10</sup>. O dicionário classifica cada palavra de acordo com as 8 emoções básicas de Plutchik (1984), sendo alegria, tristeza, raiva, medo, confiança, desgosto, surpresa e expectativa (*joy, sadness, anger, fear, trust, disgust, surprise, e anticipation*).

Além das 8 emoções citadas, o dicionário também possui 2 sentimentos, positivo e negativo, onde um termo pode estar associado a uma ou mais emoções, por exemplo, uma palavra pode estar ligada a medo e tristeza, essa relação é dada de acordo com os valores presentes nos campos das 10 colunas, sendo 8 das emoções e 2 dos sentimentos, as 8 primeiras representam uma emoção cada e as 2 últimas os sentimentos, os valores são 0 ou 1 e representam, respectivamente, ausência ou presença da emoção e do sentimento.

Entre os métodos disponíveis no pacote *syuzhet*, o *get\_nrc\_sentiment* é o único que possibilita a identificação de emoções no idioma português do Brasil, sendo o mais adequado ao modelo proposto nesse trabalho, portanto justificando a sua escolha.

---

<sup>8</sup> < <https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html>>

<sup>9</sup> < <http://sentiment.nrc.ca/lexicons-for-research/>>

<sup>10</sup> < <https://translate.google.com.br/>>

### 3.2.2 PROCESSO DE IDENTIFICAÇÃO DAS EMOÇÕES

O primeiro passo é utilizar a função `get_nrc_sentiment` para comparar o corpus com o dicionário de emoções do *NRC*, a função retorna uma matriz com 11 colunas, onde a primeira é a que representa a posição das linhas do texto analisado, nesse contexto representam os comentários. As 8 colunas seguintes, a partir da segunda, representam as emoções e as 2 últimas são dos sentimentos. Cada linha da matriz apresenta valores referentes a quantidade de termos identificados por emoção (coluna) em cada comentário, logo, cada linha possui 11 campos, além do primeiro referente a linha do texto, os 8 seguintes são associados às emoções e os 2 últimos aos sentimentos, assim como podemos ver abaixo na Figura 1.

Figura 1 - Função `get_nrc_sentiment`.

	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
1	0	0	0	0	1	0	0	1	2	1
2	1	0	0	1	0	2	0	0	1	1
3	0	1	0	0	2	0	0	1	0	4
4	1	0	0	0	0	0	0	0	1	0
5	3	6	2	3	3	2	4	1	9	7
6	0	0	0	0	0	0	0	0	0	0
7	1	0	1	1	1	0	1	1	2	1
8	0	2	0	0	2	0	2	0	0	2
9	0	1	0	0	1	1	1	0	1	1
10	3	3	2	4	2	6	2	3	7	6
11	0	0	0	0	0	0	0	0	0	1
12	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	2	0	0	3	1
14	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0

Fonte: Captura de tela da aplicação *RStudio*.

Outros procedimentos foram realizados com o objetivo de melhorar a visualização dos dados, como renomear os títulos das emoções para o português, substituir os valores da quantidade de termos nos campos das emoções pelo percentual de cada uma entre as 8 disponíveis, de acordo com cada comentário. Também foi adicionada uma coluna com o sentimento predominante por comentário, com 3 classificações, “POSITIVO”, quando os números de palavras encontradas positivas são maiores que negativas, “NEGATIVO”, quando tem mais negativas que positivas e “NEUTRO”, quando os valores de positivo e negativo são iguais, ou então, quando a soma dos dois seja 0 (zero), ou seja, sem presença de sentimento.

Algumas ações foram feitas com a finalidade de reunir as informações em um único documento, como a união dos *dataframes*, juntando os comentários, as porcentagens das emoções, a emoção e o sentimento predominante em cada comentário e também a remoção de comentários sem relevância para o trabalho, como por exemplo, os que não tiveram nenhuma emoção detectada ou predominante. Abaixo, na Figura 2, é possível visualizar como as informações ficaram organizadas no *dataframe*, nomeado de “DF.Geral”, após as operações feitas.

**Figura 2 - Trecho do *dataframe* DF.Geral.**

COMENTÁRIOS	RAIVA	EXPECTATIVA	DESGOSTO	MEDO	ALEGRIA	TRISTEZA	SURPRESA	CONFIANÇA	EMOÇÃO PREDOMINANTE.	POLARIDADE
2: Acabei de concluir o game, simplesmente incrível, uma...	25.00	0.00	0.00	25.00	0.00	50.00	0.00	0.00	TRISTEZA	NEUTRO
3: Jogo muito bom , com uma trilha sonora calma e relax...	0.00	25.00	0.00	0.00	50.00	0.00	0.00	25.00	ALEGRIA	POSITIVO
4: Isso não é um jogo, mas um filme chato, parado e sem...	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	RAIVA	NEGATIVO
10: Um dos jogos mais bonitos que tive a honra de joga...	12.00	12.00	8.00	16.00	8.00	24.00	8.00	12.00	TRISTEZA	NEGATIVO
13: Compre o jogo. Por quê? Porque é barato. Porque é ...	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	TRISTEZA	NEGATIVO
18: Que jogo! Com certeza vale a pena.	25.00	0.00	0.00	25.00	0.00	50.00	0.00	0.00	TRISTEZA	NEGATIVO
20: Um dos jogos mais tocantes que eu joguei nos últim...	12.50	12.50	0.00	12.50	12.50	37.50	0.00	12.50	TRISTEZA	NEGATIVO
21: Um jogo simplesmente lindo e emocionante!! Não te...	4.55	9.09	0.00	4.55	22.73	13.64	27.27	18.18	SURPRESA	POSITIVO
25: Jogabilidade é bem legal e a historia é muito boa, g...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	CONFIANÇA	POSITIVO
29: Jogo bem curto que fala de amor, família e amigos. S...	0.00	21.05	0.00	0.00	36.84	5.26	10.53	26.32	ALEGRIA	POSITIVO
30: Um adventure que trata sobre causas e consequenci...	9.09	18.18	0.00	9.09	9.09	18.18	9.09	27.27	CONFIANÇA	POSITIVO
31: Prós: Visual magnifico; História muito MUITO legal e ...	0.00	0.00	0.00	22.22	0.00	55.56	0.00	22.22	TRISTEZA	NEGATIVO
33: História emotiva e muito bem construída, contada at...	15.79	0.00	15.79	21.05	0.00	26.32	15.79	5.26	TRISTEZA	NEGATIVO

**Fonte:** Captura de tela da aplicação *RStudio*.

Entre as 12 colunas do *dataframe* da Figura 2, a primeira e segunda são referentes, respectivamente, a posição e conteúdo dos comentários, a terceira até a décima representa os valores dos percentuais de cada uma das 8 emoções, a décima primeira coluna apresenta qual é a emoção predominante por comentário e a última apresenta a polaridade do sentimento do comentário. De forma a exemplificar, a linha 2 representa o segundo comentário, em que foram identificadas 3 emoções, sendo, a raiva representado 25% das emoções detectadas, medo com 25% e a tristeza com o maior percentual, 50%. Na coluna denominada “EMOÇÃO PREDOMINANTE” consta a tristeza como a emoção que mais foi identificada no comentário 2, assim como “NEUTRO” na coluna “POLARIDADE”, demonstrando que não houve sentimento predominante.

### 3.3 ALGORITMOS DE CLASSIFICAÇÃO

Os algoritmos de classificação servem para prever a classe de dados que não possua uma classe identificada, mas possua atributos em comum ou similares com outros dados já classificados, assim, classificando-as baseadas nas similaridades dos atributos dos dados. Existem vários algoritmos classificadores, que utilizam técnicas distintas para realizar a classificação dos dados, com isso, a eficiência de cada um vai variar de acordo com as características e dimensões dos dados. A tarefa de classificação é do tipo supervisionada, que significa que o atributo que será previsto pelo algoritmo existe no conjunto de dados analisados, diferente do que acontece nas tarefas não-supervisionadas (AMARAL, 2016a).

#### 3.3.1 ESCOLHA DOS ALGORITMOS DE CLASSIFICAÇÃO

Foram selecionados 3 algoritmos para serem utilizados nesse trabalho, o *Naive Bayes*, Máquina de Vetor de Suporte (*SVM - Support Vector Machine*) e K-vizinhos mais próximos (*KNN - K-Nearest Neighbor*). Essa seleção baseou-se na pesquisa em trabalhos anteriores, como em (AGUIAR, FAIÇAL, *et al.*, 2018), (SANTOS, 2017) e (SILVA, 2018) que utilizam os classificadores na análise de sentimentos em redes sociais e também em (NETO, PINHEIRO, *et al.*, 2017) que aborda a análise de opinião para identificar preconceito em redações. Sendo assim, os classificadores escolhidos foram:

- **Naive Bayes:** algoritmo probabilístico, que calcula a probabilidade de um evento ocorrer considerando a análise de ocorrência feita em eventos anteriores. É um algoritmo simples, entanto, apresenta bons resultados (BARI, CHAOUCHI e JUNG, 2019).
- **Máquina de Vetor de Suporte (SVM):** nesse modelo de classificação utiliza-se a técnica de separar os elementos de variadas classes em hiperplanos, com isso, a classificação de um novo elemento é feita de acordo com sua posição no hiperplano (YATES e NETO, 2013).
- **K-vizinhos mais próximos (KNN):** classifica os elementos levando em consideração os seus vizinhos mais próximos, o algoritmo calcula a posições de cada elemento em um espaço previamente definido e depois analisa a

distância dos vizinhos do elemento a ser classificado, a classe predominante dos vizinhos é a escolhida para ser atribuída ao elemento analisado, o número de vizinhos a serem analisados é determinado anteriormente, pelo usuário (YATES e NETO, 2013).

## SEÇÃO 4

### APRESENTAÇÃO DOS RESULTADOS OBTIDOS A PARTIR DA ANÁLISE DE SENTIMENTOS DOS COMENTÁRIOS.

Esta seção apresenta os resultados finais acerca da análise dos sentimentos realizada nos comentários dos usuários, por meio de informações gráficas e escritas.

---

#### 4.1 RESULTADOS OBTIDOS

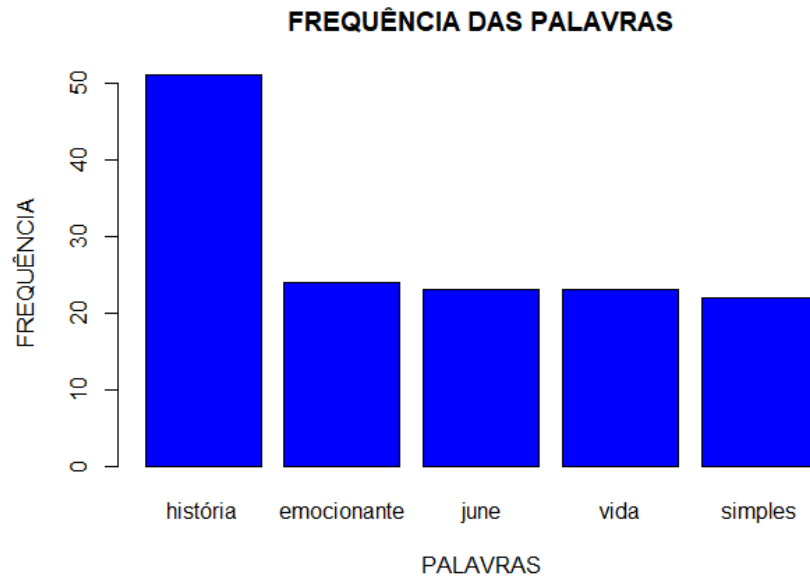
A partir das análises feitas nos capítulos anteriores, foi possível obter informações relevantes acerca dos comentários dos usuários, possibilitando esclarecer algumas questões referentes as opiniões, emoções e sentimentos dos jogadores. Essas informações foram organizadas de formas diferentes, como em gráficos e tabelas, de maneira que pudessem ser melhor visualizadas e compreendidas.

##### 4.1.1 FREQUÊNCIA DOS TERMOS

Analisar a quantidade de vezes que um ou mais termos estão presentes em um documento pode ser bastante útil, pois, com a tabela de frequência desses termos podem ser realizadas outras operações a partir desses dados, com o objetivo de encontrar novas informações acerca do conteúdo do documento (AMARAL, 2016b).

Nesse trabalho, a frequência das palavras foi utilizada para esclarecer quais termos foram mais comentados pelos usuários, assim como, para criar um gráfico de barras e uma nuvem de palavras com tais dados, logo abaixo, na Figura 3, é possível visualizar o gráfico de barras das 5 palavras mais comentadas pelos usuários do jogo *Last Day of June*.

**Figura 3** – Gráfico de barras das frequências das palavras.



Fonte: Própria (2019).

Percebe-se que a palavra mais utilizada pelos usuários foi “história”, localizada na escala de 50 ocorrências, fazendo referência a narrativa contada no jogo. Logo em seguida, com pouco mais de 20 ocorrências, está o termo “emocionante”, que faz alusão às emoções causadas pela interação do usuário com o jogo, depois vem os termos “*june*”, “vida” e “simples”, ligeiramente acima de 20 ocorrências também.

#### 4.1.2 NUVEM DE PALAVRAS

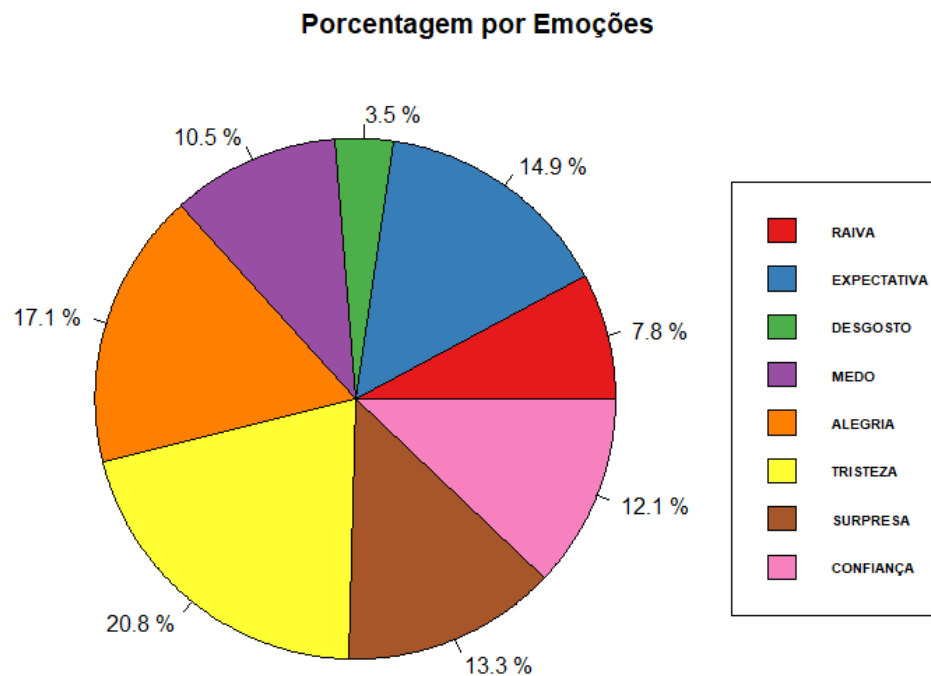
Como explicado anteriormente, os dados presentes em uma tabela de frequência de termos podem ser utilizados de várias maneiras, uma delas é para gerar uma nuvem de palavras. Conforme explica (AMARAL, 2016b), a nuvem de palavras é uma forma de representar a frequência delas por meio de um gráfico com as próprias palavras, em tamanhos que variam de acordo com suas frequências, termos mais frequentes possuem tamanhos maiores e o inverso para termos menos frequentes.



identificados em cada uma emoção. Por fim, foi utilizada a função do R chamada *pie* para exibir os percentuais de cada emoção organizados em um gráfico de pizza.

Para a apresentação das informações foi escolhido o gráfico de pizza, pois segundo (BECKER, 2015), é o modelo mais adequado quando o objetivo é exibir as informações de maneira que todas as categorias sejam abordadas, podendo compará-las mutuamente e verificar a ocorrência de cada uma no cenário geral. A Figura 5, abaixo, mostra todas as emoções e seus respectivos percentuais, exibindo a representatividade delas no contexto geral dos comentários.

**Figura 5** - Gráfico Pizza - Porcentagem das emoções.



**Fonte:** Própria (2019).

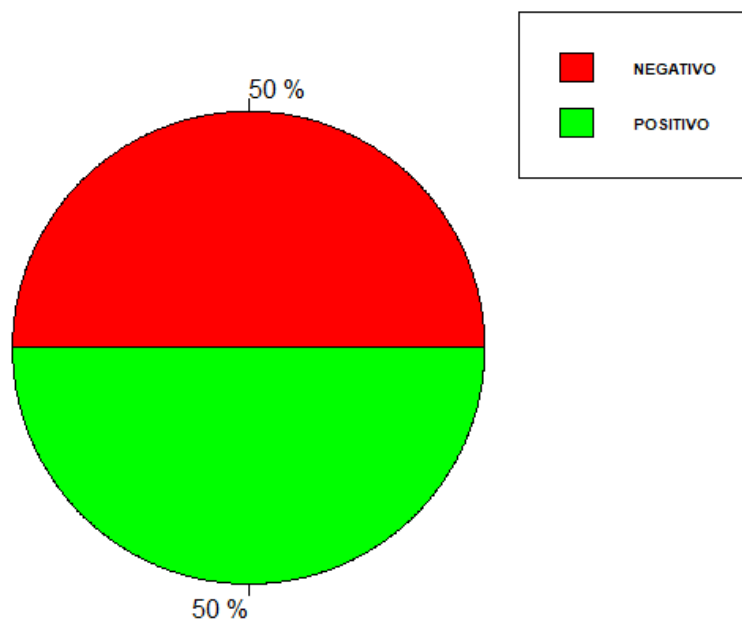
Analisando a Figura 5, nota-se que as 8 emoções são representadas no gráfico, sendo a tristeza com maior percentual entre elas, com 20.8%, em seguida aparece a alegria com 17.1%, o desgosto foi a emoção com menor percentual entre todas, apresentando apenas 3.5% da totalidade.

#### 4.1.4 PORCENTAGEM DOS SENTIMENTOS

O cálculo da porcentagem dos sentimentos foi realizado da mesma maneira como foi feito com as emoções, porém, considerando somente os 2 sentimentos, positivo e negativo, é importante explicar que a identificação dos sentimentos não

exerce influência nos resultados da análise das emoções, e nem o contrário acontece, são resultados distintos e sem relação entre eles. A Figura 6 apresenta, graficamente, a frequência de cada sentimento em relação ao total das palavras analisadas.

**Figura 6 - Porcentagem dos Sentimentos.**  
**PORCENTAGEM - SENTIMENTOS**



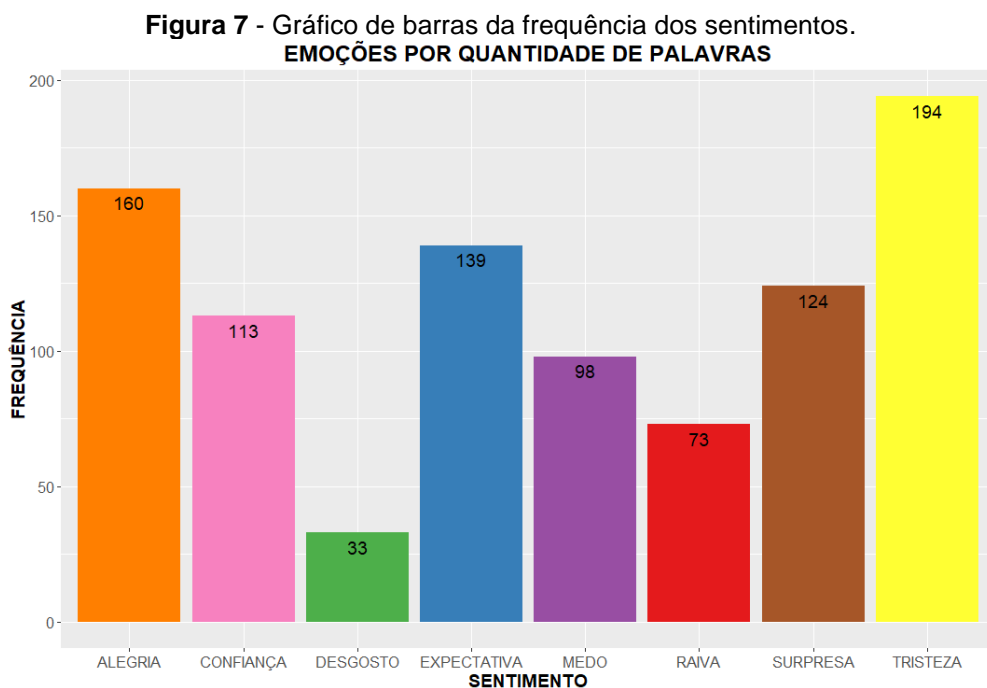
**Fonte:** Própria (2019).

Percebe-se na Figura 6, o equilíbrio dos sentimentos identificados nos comentários, pois não houve nenhum predominante, com cada um representando 50% do total.

#### **4.1.5 FREQUÊNCIA DAS PALAVRAS POR EMOÇÕES**

Uma outra forma de exibir a frequência das 8 emoções identificadas nos comentários, é exibindo-as em um gráfico de barras, que conforme (BECKER, 2015) já esclareceu anteriormente, assim como no gráfico de pizza, é possível visualizar todas as categorias e suas relações com o contexto geral.

Na Figura 7 visualiza-se as 8 emoções dispostas em barras, com tamanhos de acordo com a frequência em relação ao número de palavras atribuídas a elas, as emoções com barras maiores foram relacionadas às maiores quantidades de palavras e o mesmo raciocínio vale para as emoções menos frequentes. O valor da quantidade de palavras atribuídas a cada emoção é exibido na barra da emoção correspondente, assim, deixando as informações mais evidentes.



Fonte: Própria (2019).

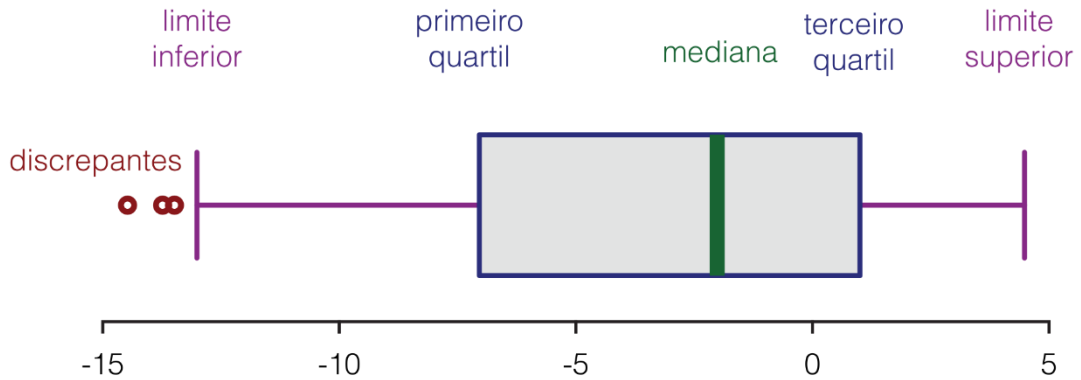
Analisando o gráfico, é nítido notar a quantidade de palavras identificadas em cada uma das 8 emoções, sendo a tristeza a de maior ocorrência, com 194 termos, logo em seguida aparece a alegria com 160 termos, a menos frequente foi o desgosto com 33 palavras identificadas.

#### 4.1.6 DIAGRAMA DE CAIXAS OU *BOXPLOT* DAS EMOÇÕES

No *Boxplot* ou diagrama de caixas, são apresentadas 5 regiões no gráfico, a mediana, dois quartis, que concentram 50% dos valores totais, um com valores maiores que a mediana e o outro com valores menores, e também, as linhas superiores e inferiores localizadas fora dos quartis, ou nomeadas também como bigodes, que representam, respectivamente, os limites máximo e mínimo. Os valores acima ou abaixo dos bigodes são considerados discrepantes, denominados também

como *outliers* (AGRESTI e FINLAY, 2012; BARROS, MATTOS, *et al.*, 2018). Na Figura 8, é mostrado um exemplo do diagrama de caixa e o detalhamento de seus elementos.

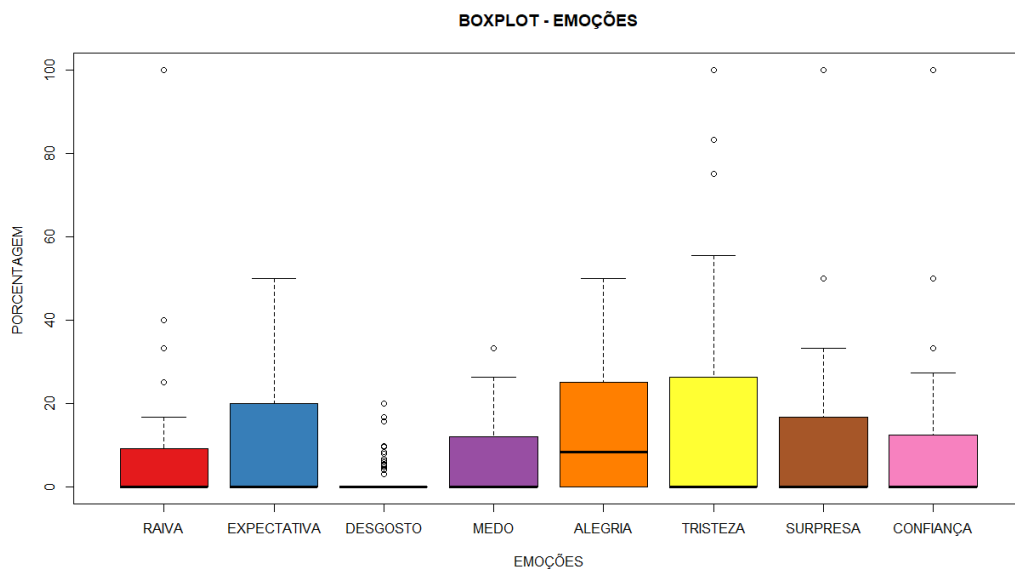
**Figura 8** - Elementos de um *Boxplot*.



Fonte: (WIKIPÉDIA, 2017).

Depois da abordagem dos conceitos do diagrama de caixa ou *Boxplot* e como interpretá-los, será apresentado na Figura 9, o *Boxplot* correspondente às oito emoções identificadas nos comentários.

**Figura 9** - *Boxplot* das 8 emoções.



Fonte: Própria (2019).

Analisando a figura acima, é possível notar que a caixa que representa a emoção tristeza possui o tamanho maior entre as 8, pois é a que possui mais termos, em relação a linha da mediana, somente na caixa da alegria tiveram elementos com valores abaixo da mediana, já referente aos limites inferiores, nenhuma emoção

apresentou valores nessa região, pois na análise das emoções e sentimentos do *NRC* não tem atribuição de valores negativos, por outro lado, nos limites superiores, apenas desgosto não teve ocorrência nessa região.

Os elementos discrepantes são aqueles que possuem valores distantes dos presentes na caixa e nos limites inferiores e superiores. Em relação a esses elementos, nenhuma emoção teve valores discrepantes abaixo das caixas, isso ocorreu devido as medianas das emoções ficarem bem próximas de 0, assim, como a função *get\_nrc\_sentiment* não utiliza valores negativos, não houve a ocorrência de valores discrepantes inferiores.

Somente as emoções expectativa e alegria não tiveram nenhum *outlier*, pois não apresentaram elementos com valores maiores que o permitido no limite superior e inferior. Desgosto foi a emoção que mais apresentou *outliers* superiores, pois como ela foi identificada em poucos comentários, a maioria dos valores associados à essa emoção foram zero, contribuindo para que sua mediana, seus quartis e bigodes, ficassem localizados na escala do valor zero. Por isso, os poucos valores referentes a identificação do desgosto nos comentários ficaram acima do limite superior, sendo posicionados na região dos valores discrepantes.

Por fim, as emoções raiva, tristeza, surpresa e confiança foram as que apresentaram elementos com maior diferença em relação aos das regiões centrais, ou os mais distantes das caixas e seus limites, isso aconteceu devido a presença de comentários em que essas emoções tiveram um alto percentual, como por exemplo, os *outliers* dessas emoções presentes na escala máxima do *boxplot*, representando os comentários com a identificação apenas uma emoção, resultando em 100% ou valor 100, com isso, ficando acima do limite superior da emoção.

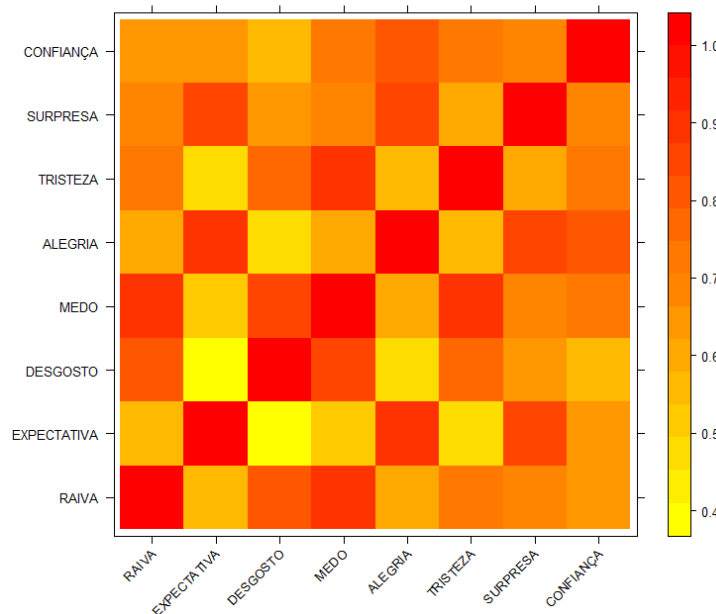
#### **4.1.7 CORRELAÇÃO DAS EMOÇÕES**

Para analisar a correlação existente entre as emoções, ou seja, a similaridade que elas possuem mutuamente, foi utilizada a função *cor* para calcular o valor da correlação entre as emoções, baseada nos dados do *dataframe* "DF.Geral". Posteriormente, foram usados os valores dessa função para gerar o gráfico do tipo

mapa de calor das correlações das emoções, em que os valores são representados por cores e suas intensidades, significando que o valor inicial é representado por uma cor e conforme ele aumenta, a cor fica mais intensa.

No contexto das correlações das emoções, foi utilizada uma escala de valor que varia do 0.4 ao 1.0, onde 0.4 representa o menor nível de similaridade entre duas emoções, correspondente a cor amarela mais suave e 1.0 equivale ao nível máximo de similaridade, representado pela cor laranja mais intensa. No intervalo de 0.4 ao 1.0 os valores são acrescidos em 0.1, criando 7 escalas, conforme aumenta o valor da escala, a cor fica mais intensa. Na Figura 10 é apresentado o mapa de calor, no formato de uma matriz quadrada, com 8 linhas e 8 colunas, em que cada linha representa uma emoção, sendo assim também para as colunas, a posição que uma linha e coluna se cruzam é representada por um quadrado de cor correspondente a intensidade da correlação que as 2 emoções (referente a linha e a coluna) possuem igualmente.

**Figura 10 - Mapa de calor da correlação entre emoções.**  
**CORRELAÇÃO DAS EMOÇÕES**



Fonte: Própria (2019).

Analisando a Figura 10, representada pelo mapa de calor da correlação entre as emoções, fica evidente a similaridade entre algumas emoções, como nos quadrantes dos pares das emoções “surpresa-expectativa”, “surpresa-alegria”, “tristeza-medo”, “alegria-expectativa”, “medo-raiva”, “medo-desgosto”, “desgosto-

raiva” e “desgosto-tristeza”, em todos esses pares de emoções o índice de correlação foi próximo de 1.0, podendo ser percebido na intensidade da cor dos quadrantes. Verificando as emoções com menos similaridades entre si, encontram-se os quadrantes dos pares de emoções “tristeza-expectativa”, “alegria-desgosto”, “medo-expectativa” e “expectativa-desgosto” que possuem a cor amarela mais clara, que representa valores mais próximos a menor escala, representando baixo índice de correlação entre elas.

O baixo valor de termos identificados em alguma emoção pode influenciar no cálculo da correlação entre essa emoção e as outras, como acontece com o desgosto, que não possui um valor alto na escala das correlações, pois foi a emoção que teve menos identificação nos comentários. Quanto maior a quantidade de dados referentes as emoções, mais preciso é o cálculo das correlações entre as emoções.

A diagonal principal da matriz, iniciada no par “raiva-raiva” e terminada em “confiança-confiança” não deve ser considerada, pois é onde os quadrantes representam a correlação da emoção com ela mesma, tendo sempre o índice máximo.

#### 4.1.8 RESULTADOS CLASSIFICADORES

Foram utilizados três algoritmos de classificação, já citados em capítulos anteriores, o *Naive Bayes*, *SVM* e o *KNN* para classificar os sentimentos a partir dos dados obtidos na análise de sentimento. Os dados foram divididos em duas partes, 60% foram destinados para formar uma base de treino do algoritmo, o 40% restantes focaram para os testes. Abaixo serão apresentados os resultados referentes a cada um algoritmo.

- **Naive Bayes:** Esse algoritmo conseguiu uma eficiência de 0.455 ou 45% na classificação dos sentimentos, tanto nos dados de treino, quanto nos de teste, abaixo, na Figura 11, é possível observar sua precisão, assim como a matriz de confusão para cada uma das bases de dados, que apresenta quantos termos foram classificados de forma errada.

**Figura 11** - Resultados *Naive Bayes*.

**TABELA DE CONTINGÊNCIA PARA DADOS DE TREINAMENTO**

		TreinoPred		
		NEGATIVO	NEUTRO	POSITIVO
NEGATIVO		5	16	2
NEUTRO		2	17	1
POSITIVO		1	14	8

**TABELA DE CONTINGÊNCIA PARA DADOS DE TESTE**

		TestePred		
		NEGATIVO	NEUTRO	POSITIVO
NEGATIVO		5	8	2
NEUTRO		0	13	0
POSITIVO		2	12	2

**PRECISÃO**

	PrecisaoTreino	PrecisaoTeste
[1,]	0.455	0.455

Fonte: Própria (2019).

Analisando a primeira matriz de confusão da Figura 11, o primeiro valor, 5, significa que 5 elementos da classe “NEGATIVO” foram avaliados corretamente, o segundo valor, 16, corresponde aos elementos da classe “NEGATIVO” classificados como “NEUTRO” e por último, o valor 2 representa os elementos de “NEGATIVO” que foram classificados como “POSITIVO”, usando esse mesmo raciocínio, podem ser avaliados quantos elementos foram classificados de forma errada de acordo com cada classe.

Uma outra alternativa é utilizar a opção de estimativa de densidade do *kernel*, em que os cálculos das densidades das probabilidades condicionais das classes são baseados em parâmetros não-numéricos (MAJKA, 2019), que é o caso da classe dos sentimentos.

De acordo com Pierson (2019, p. 87):

*A estimativa de densidade do kernel (KDE) é um método de suavização; coloca um kernel – uma função de ponderação que serve para quantificar a densidade – em cada ponto de dados no conjunto e soma os kernels para gerar uma estimativa de densidade do kernel para a região em geral.*

Com a estimativa de densidade do *kernel* (KDE) ativada, a precisão da classificação nos dados de treino aumentou quase 20 pontos percentuais em comparação ao *kernel* padrão, que é recomendado para classes do tipo

numéricas, atingindo 0.621 ou aproximadamente 62%. Abaixo, na Figura 12 é possível analisar esse crescimento.

**Figura 12** - Resultados *Naive Bayes* com *Kernel*.

TABELA DE CONTIGÊNCIA PARA DADOS DE TREINAMENTO			
	TreinoPred		
	NEGATIVO	NEUTRO	POSITIVO
NEGATIVO	7	15	1
NEUTRO	1	19	0
POSITIVO	0	8	15

TABELA DE CONTIGÊNCIA PARA DADOS DE TESTE			
	TestePred		
	NEGATIVO	NEUTRO	POSITIVO
NEGATIVO	5	8	2
NEUTRO	0	13	0
POSITIVO	2	12	2

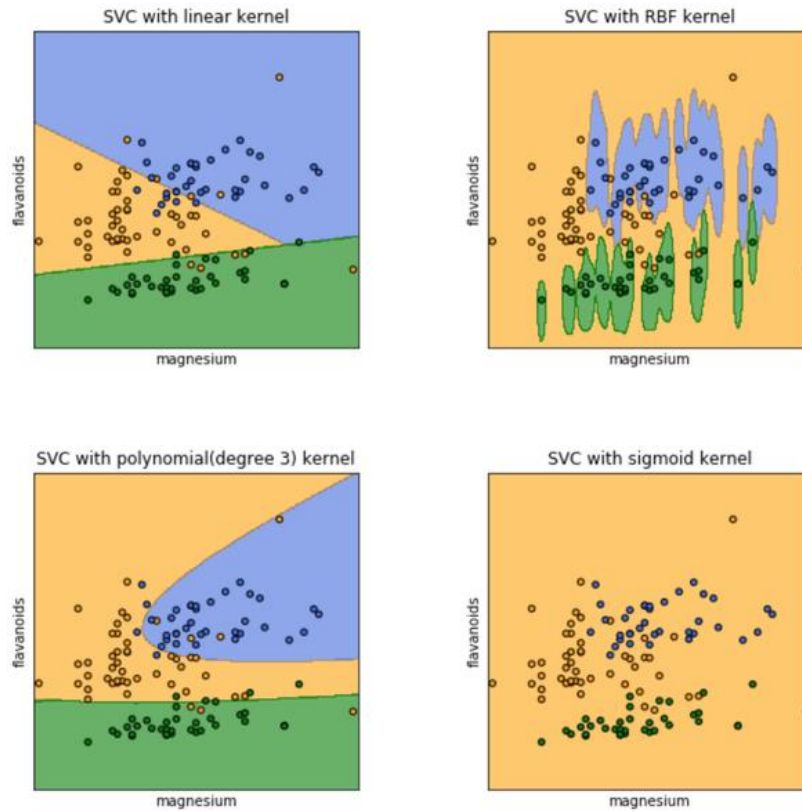
  

PRECISÃO		
	PrecisaoTreino	PrecisaoTeste
[1,]	0.621	0.455

Fonte: Própria (2019).

- **SVM:** Esse algoritmo apresenta algumas variações na maneira de classificar os elementos, assim como o *Naive Bayes*, pode-se utilizar *kernel* para auxiliar na classificação, porém, ele disponibiliza 4 opções de *kernel*, *Linear* (Padrão), *Radial* (*RBF - Radial Basis Functions*), *Polynomial* e *Sigmoid*. No *SVM* o *kernel* atua na maneira de como o espaço dos dados e os hiperplanos são dispostos, dispondo as características dos dados de forma não-linear em um espaço multidimensional, exceto na opção linear (Padrão), que dispõe os dados em planos lineares. Na Figura 13, é possível visualizar exemplos de espaços de acordo com cada tipo de *kernel*.

**Figura 13 - SVM Kernels.**



**Fonte:** Extraído de (CHOUDHARY, 2018).

Na Figura 14 são apresentados os resultados para o *kernel Linear*.

**Figura 14 - SVM com *kernel Linear*.**

#### Confusion Matrix and Statistics

	Reference		
Prediction	NEGATIVO	NEUTRO	POSITIVO
NEGATIVO	8	6	1
NEUTRO	2	10	1
POSITIVO	2	2	12

#### Overall Statistics

Accuracy : 0.6818

**Fonte:** Própria (2019).

O SVM com *kernel Linear* obteve uma precisão de 68% na classificação dos elementos.

Na Figura 15, é exibido o resultado da utilização do SVM com o *kernel Radial*, onde obteve uma precisão de 0.7955 ou 79%.

**Figura 15 - SVM com kernel Radial.****Confusion Matrix and Statistics**

		Reference		
Prediction		NEGATIVO	NEUTRO	POSITIVO
NEGATIVO		12	2	1
NEUTRO		1	11	1
POSITIVO		1	3	12

**Overall Statistics**

Accuracy : 0.7955

Fonte: Própria (2019).

Na Figura 16, os resultados apresentados são referentes ao SVM com *kernel Polynomial*, onde obteve-se uma precisão baixa, de apenas 0.3409 ou 34%.

**Figura 16 - SVM com kernel Polynomial.****Confusion Matrix and Statistics**

		Reference		
Prediction		NEGATIVO	NEUTRO	POSITIVO
NEGATIVO		14	1	0
NEUTRO		13	0	0
POSITIVO		15	0	1

**Overall Statistics**

Accuracy : 0.3409

Fonte: Própria (2019).

Por último, foi utilizado o SVM com o *kernel Sigmoid*, onde obteve-se uma precisão de 0.5455 ou 54%, como é possível visualizar logo abaixo, na Figura 17.

**Figura 17** - SVM com *kernel Sigmoid*.

**Confusion Matrix and Statistics**

	Reference		
Prediction	NEGATIVO	NEUTRO	POSITIVO
NEGATIVO	6	7	2
NEUTRO	1	11	1
POSITIVO	4	5	7

**Overall Statistics**

**Accuracy : 0.5455**

**Fonte:** Própria (2019).

Como é possível notar, entre as opções de *kernel* utilizadas, o *Radial* foi o que obteve maior precisão, com 79%. As diferenças de precisão de cada *kernel* são dadas devido a forma como o espaço e os hiperplanos são organizados, dependendo dos atributos dos dados. Assim, de acordo com as características dos dados dos sentimentos, o *kernel Radial* foi o que realizou a divisão do espaço da melhor maneira durante a tarefa de classificação, facilitando na tomada de decisão no momento de definir as classes dos elementos, sendo possível classifica-los de forma mais eficiente.

- **KNN:** Esse algoritmo também pode ser usado com configurações personalizadas, porém, não é na utilização de *kernel*, mas sim escolhendo a quantidade de elementos vizinhos a serem analisados do dado não classificado. Na Figura 17, é possível visualizar a precisão do algoritmo quando foram analisados os 3 vizinhos mais próximos.

**Figura 18** - KNN com 3 vizinhos mais próximos.

**Confusion Matrix and Statistics**

	Reference		
Prediction	NEGATIVO	NEUTRO	POSITIVO
NEGATIVO	10	4	1
NEUTRO	2	10	1
POSITIVO	1	2	13

**Overall Statistics**

**Accuracy : 0.75**

**Fonte:** Própria (2019).

Essa classificação com *KNN* com 3 vizinhos mais próximos obteve uma precisão de 0.75 ou 75%.

Na Figura 18, foram utilizados 5 vizinhos mais próximos para serem analisados durante a classificação, obtendo 0.7273 ou 72% de precisão

**Figura 19** - *KNN* com 5 vizinhos mais próximos.

**Confusion Matrix and Statistics**

	Reference		
Prediction	NEGATIVO	NEUTRO	POSITIVO
NEGATIVO	9	4	2
NEUTRO	2	10	1
POSITIVO	1	2	13

**Overall Statistics**

Accuracy : 0.7273

**Fonte:** Própria (2019).

Concluindo as análises feitas sobre o uso dos classificadores, é possível notar que alguns obtiveram bons resultados, como o *SVM* com a utilização do *kernel Radial*, que alcançou 79% de precisão, tornando-se o método mais eficiente entre todos os analisados, em contrapartida, a pior precisão também foi do algoritmo *SVM*, mas com a utilização do *kernel Polynomial*.

## SEÇÃO 5

### CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS.

Esta seção conclui o trabalho, apresentando as considerações finais e ensinamentos adquiridos ao longo do seu desenvolvimento, assim como os contratemplos que surgiram. Também, são apresentadas propostas de utilização do conhecimento e técnicas abordadas nesse trabalho.

---

#### 5.1 CONSIDERAÇÕES FINAIS

O desenvolvimento do presente estudo possibilitou uma verificação de como a análise de sentimentos pode ser utilizada para obtenção de conhecimentos em diversos contextos. Uma reflexão acerca dos benefícios dos recursos computacionais aliados as técnicas de mineração de textos na identificação de padrões emocionais em mensagens de usuários da Internet, e também, as dificuldades encontradas na tarefa de fazer um computador reconhecer padrões por meio de textos. Além disso, também permitiu mostrar a importância do conhecimento que se pode extrair dos textos, além do simples reconhecimento das palavras.

De um modo geral, foi possível perceber que é possível fazer uma máquina reconhecer emoções e sentimentos em um conjunto de textos, talvez ainda não seja tão eficiente quanto ao que se deseja, mas já é imaginável a evolução dos métodos utilizados atualmente, pois a análise de sentimentos já é bastante utilizada e atrai a atenção de diversas pessoas e empresas pelo mundo.

Dado as tarefas realizadas e conceitos abordados durante o desenvolvimento desse trabalho, foi possível encontrar padrões emocionais nas mensagens analisadas dos usuários do jogo *Last Day of June*, apresentando estatísticas sobre as emoções mais frequentes, assim como, os termos mais utilizados nos comentários e por último avaliar o desempenho de algoritmos de classificação na tarefa de classificar os

sentimentos a partir dos resultados gerados durante a mineração de texto e análise dos sentimentos.

Dessa maneira, foi possível identificar os 8 tipos de emoções abordados no trabalho, sendo que, dos dados coletados dos 110 comentários de usuários analisados, a tristeza foi a emoção predominante, com 20.8% do total. No contexto da análise dos sentimentos, houve uma igualdade entre o “POSITIVO” e o “NEGATIVO”, com 50% para cada um.

## 5.2 CONTRIBUIÇÕES DO TRABALHO

Este trabalho contribui com a análise e identificação de emoções e sentimentos em comentários feitos em páginas de jogos na internet, por meio da mineração de texto e ferramentas, e também, uso de técnicas de aprendizado de máquina para classificar emoções em palavras do texto a partir de um dicionário com termos pré-classificados.

## 5.3 TRABALHOS FUTUROS

Considerando que a quantidade de comentários coletada do jogo *Last Day of June* no idioma português não foi muito relevante, e por isso, não foi possível atingir resultados mais eficientes, seria mais eficaz caso tivesse uma quantidade maior de mensagens. Outra maneira de aumentar a eficiência da identificação as emoções, seria desenvolver ou utilizar um dicionário de emoções mais abrangente com termos em português, ou até mesmo aperfeiçoar o *NRC Emotion Lexicon*, adicionando mais termos da língua portuguesa nele, portanto, notou-se uma escassez de dicionários com essas características, fazendo com que palavras não fossem relacionadas às emoções.

Por fim, seria interessante também utilizar formas automatizadas para coletar os textos na internet, pois nesse trabalho, a coleta foi feita de forma manual, devido a dificuldade de encontrar uma ferramenta que pudesse coletar os comentários nas páginas consultadas e filtrá-los de acordo com o objetivo da pesquisa.

## REFERÊNCIAS

- 
- AGRESTI, A.; FINLAY, B. **Métodos estatísticos para as Ciências Sociais**. 4ª. ed. Porto Alegre: Penso, 2012.
- AGUIAR, E. J. D. **Análise de sentimento em redes sociais**. Dissertação (Dissertação em Ciência da Computação) - UENP. Bandeirantes, p. 74. 2017.
- AGUIAR, E. J. D. et al. Análise de Sentimento em Redes Sociais para a Língua Portuguesa Utilizando Algoritmos de Classificação. **SIMPÓSIO BRASILEIRO DE REDES DE COMPUTADORES E SISTEMAS DISTRIBUÍDOS (SBRC)**, Porto Alegre: Sociedade Brasileira de Computação, Maio 2018. Disponível em: <<https://portaldeconteudo.sbc.org.br/index.php/sbrc/article/view/2430>>. Acesso em: 14 jun 2019. Anais do XXXVI Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos.
- ALMEIDA, L. R. D.; MAHONEY, A. A. **Constituição da pessoa na proposta de Henri Wallon**. São Paulo: Edições Loyola, 2004.
- AMARAL, F. **Aprenda Mineração de Dados: Teoria e prática**. 1ª. ed. Rio de Janeiro: Alta Books, 2016a.
- AMARAL, F. **Introdução à Ciência de Dados: Mineração de Dados e Big Data**. 1ª. ed. Rio de Janeiro: Alta Books, 2016b.
- BARI, A.; CHAOUCHI, M.; JUNG, T. **Análise Preditiva Para Leigos**. 2ª. ed. Rio de Janeiro: Alta Books, 2019.
- BARROS, A. C. et al. **Análise de séries temporais em R: curso introdutório**. 1ª. ed. Rio de Janeiro: Elsevier, 2018.
- BECKER, J. L. **Estatística básica: transformando dados em informação**. Porto Alegre: Bookman, 2015.

- BEHAR, P. A. **Modelos pedagógicos em educação a distância**. Porto Alegre: Artmed, 2009.
- BERCHT, M. UFRGS - LUME - Repositório Digital. **Em direção a Agentes Pedagógicos com dimensões afetivas**, 2001. Disponível em: <<https://www.lume.ufrgs.br/handle/10183/1329>>. Acesso em: 8 Jun 2019.
- BOECHAT, I. T. **As famílias e as tecnologias digitais: a comunicação pela articulação de vieses não antes explorados**. Curitiba: Appris, 2017.
- CANTO, C. E. As linguagens mais populares em Ciência de Dados. **Propus data science**, 2017. Disponível em: <<https://propus.science/as-linguagens-mais-populares-em-ciencia-de-dados/>>. Acesso em: 17 jun 2019.
- CASTRO, L. N. D.; FERRARI, D. G. **Introdução à mineração de dados: conceitos básicos, algoritmos e aplicações**. São Paulo: Saraiva, 2016.
- CHICKEN, R. “Last Day of June”: o amor não termina com a morte. **Observador**, 2017. Disponível em: <<https://observador.pt/2017/09/08/jogo-last-day-of-june-critica/>>. Acesso em: 7 abr. 2019.
- CHOUDHARY, P. Jupyter - nbviewer. **Comparing different SVM kernels on a multi-class classification problem**, 2018. Disponível em: <[https://nbviewer.jupyter.org/github/pramitchoudhary/Experiments/blob/master/notebook\\_gallery/jupyter/SVM\\_kernel\\_comparison.ipynb](https://nbviewer.jupyter.org/github/pramitchoudhary/Experiments/blob/master/notebook_gallery/jupyter/SVM_kernel_comparison.ipynb)>. Acesso em: 6 jul 2019.
- COOPER, D. R.; SCHINDLER, P. S. **Métodos de Pesquisa em Administração**. Tradução de Scientific Linguagem Ltda. 12<sup>a</sup>. ed. Porto Alegre: AMGH, 2016.
- GITHUB. **Portuguese stopwords collection**. Disponível em: <<https://github.com/stopwords-iso/stopwords-pt>>. Acesso em: 9 abr. 2019.
- GOOGLE Tradutor. Disponível em: <<https://translate.google.com.br/>>. Acesso em: 22 jun. 2019.

- GUIMARÃES, R. G. Repositório Institucional - UFLA. **Aprimoramento da análise de sentimentos em redes sociais utilizando análise léxica e perfil de usuário**, 2017. Disponível em: <<http://repositorio.ufla.br/jspui/handle/1/15376>>. Acesso em: 14 jun 2019.
- JQUES, P. A.; VICARI, R. M. UFRGS - LUME - Repositório Digital. **Pat: um agente pedagógico animado para interagir afetivamente com o aluno**, 2005. Disponível em: <<http://hdl.handle.net/10183/559>>. Acesso em: 8 jun 2019.
- JOCKERS, M. R-Project. **Introduction to the Syuzhet Package**, 2017. Disponível em: <<https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html>>. Acesso em: 18 abr. 2019.
- KRZNARIC, R. **Empatia: Sobre a arte de viver**. Rio de Janeiro: Zahar, 2015.
- LAST Day of June. **Steam**, 2017. Disponível em: <[https://store.steampowered.com/app/635320/Last\\_Day\\_of\\_June/](https://store.steampowered.com/app/635320/Last_Day_of_June/)>. Acesso em: 08 mar. 2019.
- MAJKA, M. naivebayes: High Performance Implementation of the Naive. **R-Project**, 2019. Disponível em: <<https://CRAN.R-project.org/package=naivebayes>>. Acesso em: 6 jul 2019.
- MARQUESONE, R. **Big Data: Técnicas e tecnologias para extração de valor dos dados**. São Paulo: Casa do Código, 2016.
- MOHAMMAD, S. National Research Council Canada. **The Sentiment and Emotion Lexicons**, 2016. Disponível em: <<http://sentiment.nrc.ca/lexicons-for-research/>>. Acesso em: 20 jun. 2019.
- MOURA, R. 'Redes sociais reduzem noção de vergonha, diálogo e empatia', diz psicoterapeuta americano. **BBC**, 2017. Disponível em: <<https://www.bbc.com/portuguese/geral-42197265>>. Acesso em: 22 mai 2019.

- NETO, S. R. D. S. et al. Uma abordagem computacional de análise de opinião para identificação de preconceito em redações. **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)**, 2017. Disponível em: <<http://www.br-ie.org/pub/index.php/sbie/article/view/7647>>. Acesso em: 12 junho 2019.
- OLIVEIRA, J. D.; CASAGRANDE, N. M.; GALERANI, L. D. D. J. A EVOLUÇÃO TECNOLÓGICA E SUA INFLUÊNCIA NA EDUCAÇÃO. **Revista Interface Tecnológica**, 2016. Disponível em: <<https://revista.fatectq.edu.br/index.php/interfacetecnologica/article/view/123>>. Acesso em: 13 maio 2019.
- PAREDES, A. Avanços tecnológicos: vantagens e desvantagens. **IEBS**, 2019. Disponível em: <<https://www.iebschool.com/pt-br/blog/software-de-gestao/tecnologia/avancos-tecnologicos-vantagens-e-desvantagens/>>. Acesso em: 02 jun 2019.
- PICARD, R. W. **Affective Computing**. Cambridge: The MIT Press, 2000.
- PIERSON, L. **Data science para leigos**. Tradução de Eveline Machado. 2ª. ed. Rio de Janeiro: Alta Books, 2019.
- PLUTCHIK, R. A General Psychoevolutionary Theory. In: SCHERER, K. R.; EKMAN, P. **Approaches to Emotion**. Hillsdale: Lawrence Erlbaum, 1984.
- R Project. **The R Project for Statistical Computing**. Disponível em: <<https://www.r-project.org/>>. Acesso em: 17 mar. 2019.
- R Project. **tm - Text Mining Package**. Disponível em: <<http://tm.r-forge.r-project.org/>>. Acesso em: 17 mar. 2019.
- REIS, A. **Sociedade.com: Como as tecnologias digitais afetam quem somos e como vivemos**. Arquipelago Editorial: Porto Alegre, 2018.
- REZENDE, S. O. **Sistemas inteligentes: fundamentos e aplicações**. Barueri: Manole, 2005.

- R-PROJECT. **Wordcloud2 introduction**, 2018. Disponível em: <<https://cran.r-project.org/web/packages/wordcloud2/vignettes/wordcloud.html>>. Acesso em: 22 jun. 2019.
- RSTUDIO. **RStudio - Open source and enterprise-ready professional software for R**, 2018. Disponível em: <<https://www.rstudio.com>>. Acesso em: 06 mar. 2019.
- SANTOS, T. G. D. Análise de Opiniões Utilizando Técnicas de Mineração de Dados em Redes Sociais. Estudo de Caso: Twitter. **Repositório Institucional - AEE**, Anápolis, p. 48, 2017. Disponível em: <<http://repositorio.aee.edu.br/jspui/handle/aee/53>>. Acesso em: 15 jun 2019.
- SARMET, M. M.; PILATI, R. Efeito dos jogos digitais no comportamento: análise do General Learning Model. **Temas em Psicologia**, Ribeirão Preto, 24, n. 1, mar 2016. 17-31. Disponível em: <[http://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1413-389X2016000100002&lng=pt&nrm=iso](http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1413-389X2016000100002&lng=pt&nrm=iso)>. Acesso em: 22 mai 2019.
- SILVA, L. A. D.; BOSCARIOLI, C.; PERES, S. M. **Introdução à Mineração de Dados: Com Aplicações em R**. Rio de Janeiro: Elsevier, 2017.
- SILVA, R. M. J. D. O. Análise de Sentimento em tweets. **Repositório da Produção Científica e Intelectual da Unicamp**, Campinas, p. 89, 2018. Disponível em: <<http://repositorio.unicamp.br/jspui/handle/REPOSIP/332102>>. Acesso em: 6 jun 2019.
- SOLOMON, M. R. **O comportamento do consumidor: comprando, possuindo e sendo**. Tradução de Beth Honorato. 11ª. ed. Porto Alegre: Bookman, 2016.
- SOUZA, A. L. P. D. **Redescobrimo a Sabedoria Milenar**. 2ª. ed. São Paulo: Scortecci Editora, 2016.
- SOUZA, R. R.; CAFÉ, L. M. A. Informação & Sociedade: Estudos. **Análise de sentimento aplicada ao estudo de letras de música**, 2018. Disponível

em: <<http://periodicos.ufpb.br/index.php/ies/issue/view/2055/showToc>>.

Acesso em: 14 jun 2019.

TWITTER. **Veja o que está acontecendo no mundo agora**. Disponível em:

<<https://twitter.com/>>. Acesso em: 11 abr. 2019.

VANZIN, T. **Cibersociedade e novas tecnologias**. Erechim: Deviant, 2018.

WIKIPÉDIA. Diagrama de caixa. **Wikipédia: A enciclopédia livre**, 2017. Disponível

em: <[https://pt.wikipedia.org/wiki/Diagrama\\_de\\_caixa](https://pt.wikipedia.org/wiki/Diagrama_de_caixa)>. Acesso em: 15 jun. 2019. il. color.

YATES, R. B.; NETO, B. R. **Recuperação de Informação: Conceitos e Tecnologia das Máquinas de Busca**. 2<sup>a</sup>. ed. Porto Alegre: Bookman, 2013.

YOUTUBE , 2019. Disponível em: <<https://www.youtube.com/>>. Acesso em: 13 abr. 2019.