

# Análise da Polaridade Presente no Conteúdo Textual

Leonardo de Jesus Silva<sup>1</sup>, Gilton José Ferreira da Silva<sup>1</sup>, Héctor Julián Tejada Herrera<sup>1</sup>

<sup>1</sup>Universidade Federal de Sergipe (UFS)  
São Cristóvão – SE – Brasil

**Abstract.** *This work refers to Sentiment Analysis which has become an important topic on the Web, especially in Social Networks. This work will utilize a Machine Learning algorithm with supervised machine learning, Naive Bayes. Making a prototype with the help of the Natural Language Toolkit (NLTK) and Natural Language Processing (NLP) gives the machine the ability to understand the natural language of the human being whether it is written or spoken. Getting a hit rate of approximately 71 % when rating feelings into two categories, joy and sadness.*

**Resumo.** *Este trabalho refere-se a Análise de Sentimento a qual tem se tornado um importante tópico na Web, especialmente em Redes Sociais. Este trabalho utilizará um algoritmo de Machine Learning com aprendizagem de máquina supervisionado, Naive Bayes. Confeccionando um protótipo com auxílio da biblioteca Natural Language Toolkit (NLTK) e Processamento de Linguagem Natural (PLN) dando a máquina a capacidade de compreender a linguagem natural do ser humano seja ela escrita ou falada. Obtendo uma taxa de acerto de aproximadamente 71% ao classificar sentimentos em duas categorias, alegria e tristeza.*

## 1. Introdução

A grande revolução tecnológica do século XXI ocorre principalmente pelo desenvolvimento das tecnologias móveis a exemplo o advento dos *smartphones*, *gadgets* de *Internet of Things* (IoT) e as tecnologias vestíveis. Outro fator é o crescente número de pessoas com acesso a Internet utilizando ferramentas, as quais permitem diminuir distância entre elas. Essas ferramentas são intituladas de Redes Sociais [Recuero 2012].

O eMarketer estimou que 2,46 bilhões de pessoas usaram as Rede Sociais pelo menos uma vez por mês em 2017, um aumento de 8,2% em relação a 2016 [eMarketer 2017]. Consequentemente esse acesso expressivo as Mídias Sociais é capaz de gerar um volume exponencial de informação em um minuto [Campos 2017].

Desta forma a criação e compartilhamento de informações na Internet supera a capacidade de um indivíduo ou organização de consumirem na mesma velocidade essas informações [Gabriel 2017]. No entanto, essas informações são de grande valia para as organizações, pois uma vez analisada possibilitaria conhecer um indivíduo ou o conjunto de indivíduos bem como suas respectivas opiniões e sentimentos em relação aos serviços ou produtos fornecidos pelas organizações por meio das informações publicados nas Redes Sociais [Gomes 2013].

Um indivíduo inserido em um grupo ou organização, pode gerar mudanças no mesmo. Seja essas mudanças nos sentimentos dos indivíduo ou no comportamento

do mesmo. A longo prazo essas podem afetar seu lado profissional e/ou pessoal [Santos et al. 2005]. A exemplo disso são as Instituições de Ensino Superior (IES) responsáveis pelo ensino superior no sistema educacional dentro do país, a qual se configura como um ambiente de estresse para os universitários [Evans et al. 2018]. Desta forma este trabalho adotará duas IES, uma IES de natureza administrativa pública e outra privada.

Pois o ambiente universitário em específico a pós-graduação é 6 vezes maior que um individuo venha a se tornar depressivo ou sofra de ansiedade em relação a qualquer outro ambiente [Evans et al. 2018]. Como também há trabalhos os quais apontam as IES como responsáveis pelo aumento de sentimentos negativos nos estudantes de graduação [Polydoro et al. 2005, Dutra 2012, Galvão et al. 2013]. Desta forma coloca as IES sendo um dos principais fatores responsáveis pelos pensamentos suicidas entre universitários [Franco et al. 2017].

Uma IES têm como objetivos as seguintes finalidades: estimular a cultura, formar diplomados em diferentes saberes, incentivar a pesquisa, promover a divulgação de conhecimento de todo gênero e forma, suscitar a melhora cultural e profissional, instigar conhecimentos dos problemas do mundo para possíveis soluções e Promover a extensão bem como abertura da instituição à participação da população [de Diretrizes 1996].

No entanto, pesquisas supracitadas caracterizam as IES como um ambiente de adversidades ao invés de um ambiente de oportunidades. Nesse sentido este trabalho buscará responder a questão: qual tipo de polaridade de sentimento, seja ele negativo ou positivo, se faz presente na população do estado de Sergipe tendo como campo de pesquisa a Mídia Social, Twitter. Considerando uma IES de ordem administrativa pública e outra privada. Desta forma foi utilizado o Censo da Educação Superior 2014 [da Educação Superior 2014], para escolha das duas IES os pré-requisitos da seguem demonstrados na Tabela 1.

**Table 1. Instituições de Ensino Superior.**

<b>Classificação</b>	<b>IES A</b>	<b>IES B</b>
<b>Natureza administrativa</b>	Pública	Privada
<b>Possui Pós-Graduação</b>	SIM	SIM
<b>Possui Semana de Prova</b>	NÃO	SIM
<b>Número de alunos</b>	28.052	22.692
<b>Número de cursos</b>	113	62
<b>Índice Geral dos Cursos</b>	3,02	2,78

## **2. Objetivos**

Uma vez a problemática vivenciada dentro das Instituições de Ensino Superior (IES), sejam elas particulares ou públicas, a comunidade, e em especial os discentes sofrem com alto índice de estresse e sentimento negativos durante seu vínculo com essas instituições. Porém, as IES não conseguem acompanhar ou averiguar o estado emocional dos alunos. É nesse intuito e no sentido de contribuir para o desenvolvimento do ambiente acadêmico mais receptivo a saúde mental. Os objetivos geral e específico deste trabalho são:

## 2.1. Geral

Criar uma Análise de Sentimento por meio dos *tweets* publicados a respeito da IES A e IES B dentro da região do espaço geográfico de Sergipe. Verificando a reverberação dos sentimentos positivos e negativos que ambas causam na comunidade sergipana.

## 2.2. Específicos

A fim de alcançar o objetivo acima citado, este trabalho tem como objetivos específicos:

- Realizar um levantamento das necessidades do cenário estudado por meio de uma Revisão Sistemática da Literatura (RSL) dos trabalhos registrados em bases acadêmicas que tratem da Análise de Sentimento em Mídias Sociais;
- Consolidar o conhecimento sobre a problemática do ambiente universitário, e o que caracteriza as Instituições de Ensino como um dos ambientes mais estressantes e adversos;
- Desenvolver um protótipo funcional a fim de adquirir maior compreensão sobre Análise de Sentimento;
- Desenvolver uma *Minimum Viable Product* (MVP), a qual terá como função analisar os tweets relacionados a IES A e IES B verificando suas respectivas polaridade de sentimento, servindo como indicadores para tomadas de decisões.

## 3. Procedimentos Metodológicos

A presente pesquisa é quanti-qualitativa, exploratória, descritiva, com enfoque nas informações publicadas nas Mídias Sociais pelos usuários. Do ponto de vista de sua natureza, classifica-se como aplicada, tomando por base o interesse em se conhecer qual a polaridade de sentimento presente no conteúdo textual postado nas Redes Sociais [Zanella 2006, Gerhardt and Silveira 2009, Gil 2008].

Quanto aos procedimentos técnicos adotados, portanto, se concentram na observação e análise de dados sobre a exploração dos textos e suas respectivas polaridades de sentimento, referentes às instituições de ensino superior definidas neste trabalho.

Em primeira instância foi realizada uma pesquisa bibliográfica e documental [Wazlawick 2017, Fonseca 2002], a fim de estabelecer a fundamentação teórica deste trabalho e a revisão de conceitos bem como contextos referentes à Análise de Sentimento, Mídias Sociais e os estudos das emoções, tal como o comportamento dos indivíduos nos meios digitais. Foram utilizados teses, livros e artigos indexados em revistas científicas bem como material complementar como filmes, vídeos entre outras fontes de informação.

No intuito de verificar, no tempo estabelecido, a relação da polaridade do sentimento das postagens de usuários relacionados as IES estipulados neste trabalho, será empregada a pesquisa *ex-post-facto*, a qual tem por objetivo investigar possíveis relações de causa e efeito entre um determinado fato identificado pelo pesquisador e um fenômeno que ocorre posteriormente [Fonseca 2002].

## 4. Trabalhos Relacionados

Existem diferentes aplicações as quais usam Análise de Sentimento para compreender um perfil de um indivíduo ou conjunto de indivíduos. Bem como analisar marcas e produtos de organizações por meio das informações compartilhadas nas Mídias Sociais.

A exemplo a cobertura de Mídia Social sobre "Game of Thrones", uma série de televisão de drama americana criada para a rede de televisão HBO <sup>1</sup> baseada na obra de George R. R. Martin. O sistema analisa dados coletados nas Redes Sociais: Twitter, Facebook, Goggle+ e Youtube. Para saber qual sentimento reverberado sobre a série [Scharl et al. 2016].

Outra aplicação é do modelo computacional para prever o surgimento de depressão e pós-traumático de stress nos usuários do Twitter [Reece et al. 2017]. Além disso tem aplicação a qual mensura a opinião pública por meio da Análise de Sentimento relacionado ao governo durante a COPA realizada no Brasil [Oliveira and de Souza Bermejo 2017].

## 5. Fundamentação Teórica

Para realizar uma Análise de Sentimento em conteúdo textual, emprega-se sub-áreas do campo de pesquisa da Inteligência Artificial (IA) como Aprendizagem de Máquina advinda do termo *Machine Learning* (ML), Processamento de Linguagem Natural (PLN) e Mineração em Texto advinda do termo *Text Mining* juntamente a fonte de informação, a qual será analisada as Mídias Sociais.

### 5.1. Aprendizagem de Máquina

Aprendizagem de Máquina da tradução livre do termo *Machine Learning*(ML) de acordo com Gabriel (2017) este termo foi criado em 1959 por Arthur Samuel, um pioneiro no campo de *games* computacionais e IA [Gabriel 2017]. Esse definiu a ML como "um campo de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados" [Gabriel 2017, p. 197].

A proposta da ML é equivalente ao processo de aprendizado do ser humano, o qual é baseado na experiência por meio dos dados. Desta forma a ML é um campo da IA trabalha com algoritmos os quais permitem a um programa "aprenda", ou seja, os programadores humanos não precisam especificar um código o qual determina as ações ou previsões que irá realizar em determinada situação [Gabriel 2017].

Essa aprendizagem pode ser configurada por Aprendizagem Supervisionada, a qual alimenta-se o algoritmo com dados sobre situações as quais já se sabem as respostas para assim esse ser capaz de responder futuras questões; Aprendizagem Não Supervisionada dado um volume imenso de dados, o qual não temos nenhum conhecimento prévio sobre as respostas das situações, o algoritmo reconheça os padrões.

Aprendizagem por Esforço este modelo de aprendizagem não é estabelecido somente na experiência vivenciada, mas também na alteração do resultado mediante uma tentativa frustrada. Ao deparar-se com uma resposta equivocada para um determinado problema ele aprende e efetua uma nova tentativa.

### 5.2. Processamento de Linguagem Natural

O Processamento de Linguagem Natural (PLN) advindo do termo em inglês *Natural Language Processing*(NLP) é um campo de pesquisa, o qual consiste em conceder capacidade de entender a linguagem natural humana falada ou escrita as máquinas.

---

<sup>1</sup>Página da Emissora HBO: <https://www.hbo.com/>

O Processamento da Linguagem Natural é um conjunto de técnicas computacionais para analisar e representar ocorrências naturais de texto em um ou mais níveis de análise linguística com o objetivo de se alcançar um processamento de linguagem similar ao humano para uma série de tarefas ou aplicações [Liddy 2001, p. 1].

Desta forma PLN possui um processo cujo grau de dificuldade é alto, paralelo à complexidade da linguagem natural. Segundo Liddy (2001) a solução para isso é granular esse nível de dificuldade ao processar a linguagem natural [Liddy 2001]. Existem diversos níveis como léxico, morfológico e semântico.

No intuito de auxiliar esse processo, utilizou-se uma etapa de pré-processamento de dados textuais. A exemplo remoção de *stopwords*, as palavras as quais aparecem com frequência porém carregam pouco significado para análise. Segmentação de palavras dentre outras técnicas para proporcionar maior compreensão a máquina da linguagem natural.

Existem diversas ferramentas *open-source* disponíveis as quais auxiliam na construção de aplicações das quais utilizem PLN a exemplo GATE<sup>2</sup>, Mallet<sup>3</sup>, OpenNLP<sup>4</sup>, UIMA<sup>5</sup>, Genisim<sup>6</sup> e NLTK<sup>7</sup>.

Este trabalho fará uso da ferramenta *Natural Language Toolkit*(NLTK). A NLTK é uma biblioteca para PLN e Aprendizagem de Máquina. A escolha justifica-se pela disseminação do uso da NLTK, bem como o fato dessa ser escrita na linguagem de programação Python a qual foi considerada a quinta linguagem de programação mais popular entre os desenvolvedores [Overflow 2017].

### 5.3. Mineração de Texto

A Mineração de Texto segundo Tan (1999) do inglês *Text Mining* refere-se a um processo de extrair padrões a partir de dados no formato de texto para apresentar informações ou conhecimentos revelantes mediante a inferências feitas em grande volume de dados do tipo texto [Tan et al. 1999].

Existe uma grande variabilidade na internet de textos para ser analisados como livros, jornais, revistas, paginas web, postagens de Redes Sociais entre outras. Sendo Redes Sociais uma das fontes de textos a qual gera mais conteúdo oriundo de perfis de usuários. Este trabalho utilizará os dados extraídos da Rede Social Twitter por intermédio da sua *Application programming interface* (API).

O processo de Mineração de Texto é composto por etapas. Essas sendo seleção de documentos, definição do tipo de abordagem dos dados - análise semântica ou estatística, preparação dos dados, indexação e normalização, cálculo da relevância dos termos, seleção dos termos e pós-processamento - análise de resultados.

---

<sup>2</sup>GATE: <https://gate.ac.uk/>

<sup>3</sup>Mallet: <http://mallet.cs.umass.edu/>

<sup>4</sup>OpenNLP: <https://opennlp.apache.org/>

<sup>5</sup>UIMA: <https://uima.apache.org/external-resources.html>

<sup>6</sup>Genisim: <https://radimrehurek.com/gensim/>

<sup>7</sup>NLTK: <http://www.nltk.org/>

## 5.4. Mídias Sociais

As Mídias Sociais estão presentes constantemente no dia-a-dia transformando a vida dos indivíduos não somente na interação social digital mas também para fins profissionais, acadêmicos e de entretenimento. Essas Redes Sociais têm um papel fundamental no conceito de Web 2.0, que emerge com características de ubiquidade e colaboratividade a partir do ano 2000 [Santaella 2014].

Segundo Recuero(2012) define que uma Rede Social é uma abstração dos padrões observáveis de conexão de um grupo social, a partir das conexões estabelecidas entre os diversos indivíduos presentes no mundo físico ou virtual [Recuero 2012]. Como Mídias Sociais a exemplo Facebook <sup>8</sup> e o Twitter<sup>9</sup>, permitem aos usuários interagirem com suas respectivas redes de contatos. Essas construídas por outros usuários e não pré-determinadas pelas ferramentas como definida por Recuero [Recuero 2012].

## 5.5. Análise de Sentimento

A identificação de sentimentos em textos é uma das áreas de pesquisas mais eminente em PLN desde o início do século XXI, quando se tornou uma área de pesquisa muito ativa [Liu 2010]. O estudo de opiniões, sentimentos e emoções expressas em texto é conhecido como Análise de Sentimento ou Mineração de Opinião dos termos em inglês respectivamente *Sentiment Analysis* ou *Opinin Mining*.

De acordo com Pang, Lee et al. (2008) o surgimento das Mídias Sociais bem como avanços tecnológicos em sub-campos da IA como ML e PLN proporcionaram o crescimento de trabalhos produzidos a respeito de Análise de Sentimento [Pang et al. 2008]. Analisar os sentimentos presentes em texto tem como propósito identificar o conteúdo subjetivo do texto e classificá-lo como positivo, negativo ou neutro [Oliveira and de Souza Bermejo 2017].

Este propósito contém uma dificuldade a qual dá origem a um problema multidisciplinar na computação envolvendo a Mineração de Dados, PLN e Aprendizagem de Máquina. A fim de extrair informações dos dados textuais de forma quantitativa e qualitativa sobre os sentimentos expressos em um texto ou conjunto de texto [Benevenuto et al. 2015].

## 6. Resultados Parciais e Trabalhos Futuros

No intuito de consolidar os conhecimentos sobre Análise de Sentimento em texto e aplicá-los, utilizou-se da biblioteca NLTK juntamente com algoritmo de aprendizagem supervisionada Naive Bayes para conceber um protótipo. A palavra protótipo deriva do grego a qual **Protós** significa primeiro e **Typos** significa tipo; em livre tradução seria : primeiro modelo, o qual está em fase de testes, estudo, ou planejamento.

O protótipo é uma ferramenta muito útil na pré-fase de desenvolvimento, pois ele pode mostrar os problemas técnicos do projeto, antes mesmo da fase de produção, evitando assim custos desnecessários com desenvolvimento.

Foi adotado uma classificação de sentimento segundo o trabalho relacionado as seis emoções - alegria, desgosto, medo, raiva, surpresa, tristeza - proposto por Ekman

---

<sup>8</sup>Página Inicial do Facebook: <https://www.facebook.com/>

<sup>9</sup>Página Inicial do Twitter: <https://www.twitter.com/>

(1992) [Ekman 1992]. Para isso foi utilizado uma base, a qual intitulamos de corpus de texto, cujo número de frases corresponde a 760 registros. Desses 160 caracterizam como frases de Alegria ao passo que as outras 5 emoções cada uma respectivamente tem 120 frases classificadas. Essa base foi fornecida pelo professor doutor Jones Granaty fundador do portal IA Expert <sup>10</sup>.

No intuito de treinar o algoritmo Naive Bayes e avaliar seu desempenho para detectar a emoção, foi utilizado 70% de cada classe para treinamento e os 30% restante foi utilizado para os testes. Segue um trecho do código do protótipo presente no GitHub <sup>11</sup>.

O Código 1 foi implementado utilizando a biblioteca NLTK com o ambiente de desenvolvimento PyCharm<sup>12</sup> fazendo-se uso da linguagem Python 3.4 . Neste código temos função de remoção das *stopwords* que são palavras de grande frequência no texto, no entanto, não tem muita relevância na análise. Bem como a extração de *Stemming*, ou seja o radical das palavras as quais tem relevância para análise. Sendo um trecho do código representando o pré-processamento da base para o treinamento e teste.

Para saber a efetividade do algoritmo e detectar as seis emoções estipulados pelo Psicólogo Ekman [Ekman 1992], foi verificado o índice de acurácia do algoritmo. Realizou-se também uma avaliação por meio de uma matriz de confusão, a qual consiste em uma tabela a qual permite a visualização do desempenho de um algoritmo de aprendizado, em específico um algoritmo de aprendizagem supervisionada.

```
stopwordsNLTK = nltk.corpus.stopwords.words('portuguese')
def removestopword(texto):
    frases = []
    for (palavras, emocao) in texto:
        semstop = [p for p in palavras.split() if
        p not in stopwordsNLTK]
        frases.append((semstop, emocao))
    return frases
```

Listing 1: Parte do Código de Pré-Processamento em Python

Cada coluna da matriz de confusão representa as instâncias de uma classe prevista, enquanto as linhas representam os casos de uma classe real. O nome é originado do fato de que a matriz torna mais factível visualmente verificar se o algoritmo está confundindo duas classes ou mais. A Figura 1.A demonstra os resultados para esta avaliação com as 6 emoções sendo classificadas.

O resultado da diagonal principal demonstra o número de acertos do algoritmo ao classificar. Pois uma vez treinado realiza a classificação nos 30% da base que foi utilizada para verificar a Acurácia do algoritmo a qual foi aproximadamente de 32% e percebeu que houve muitos erros ao tentar classificar as seis classes. Uma das hipóteses para esse resultado seja a classificação das frases na base de treinamento tenham sido feito

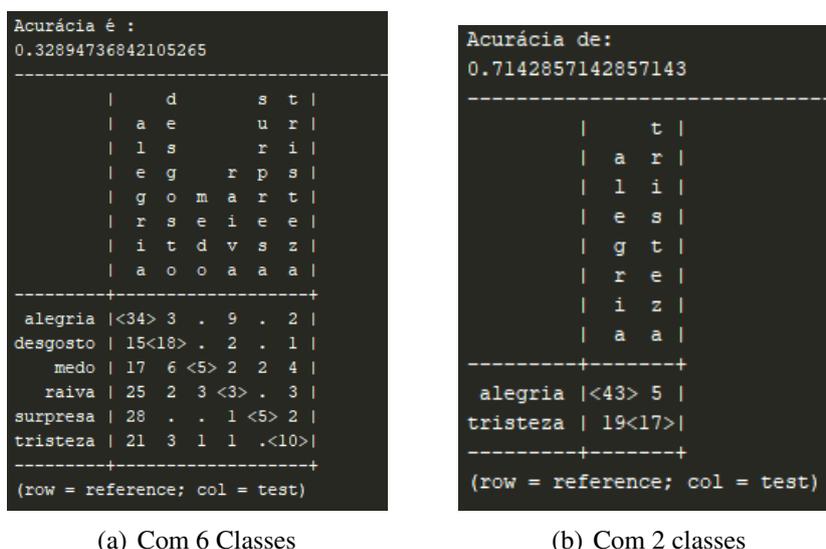
<sup>10</sup><http://iaexpert.com.br/index.php/sobre/>

<sup>11</sup>Repositório do Código do Protótipo: <https://github.com/leoproject/Prot-tipo-TCC>

<sup>12</sup>Página do PyCharm: <https://www.jetbrains.com/pycharm/>

incorretamente.

No intuito de verificar se há relação do número de classes influência nos resultados da acurácia. Reduziu-se as classes de emoções para duas a saber: alegria e tristeza. De acordo com a Figura 1.B a acurácia elevou-se para aproximadamente 71% de acerto. Todavia, mesmo com duas classes constatou-se um número expressivo de erro ao tentar classificar frases do tipo tristeza. Esse fato corroborou a hipótese, a qual a classificação das frases da base de treinamento possua ruídos, ou seja não tenha sido classificada corretamente. Devido a classificação não ter sido realizada por especialistas linguísticos ou da psicologia. Desta forma afetou o desempenho da aprendizagem do algoritmo para distinguir a qual classe a frase da base de teste pertence.



**Figure 1. Resultados do Protótipo**

Como trabalhos futuros, mediante os resultados do protótipo pretende-se utilizar a Análise de Sentimento através da polaridade de sentimento negativo e positivo. A fim de treinar o algoritmo para o desenvolvimento de uma aplicação a qual utilizará da base publica Tweets para Análise de Sentimentos em Português (TAS-PT)<sup>13</sup>, a qual possui 38119 tweets com sentimento positivo e a mesma quantia para sentimento negativo.

Bem como planejar a extração dos *tweets* referentes a IES A e IES B para verificar a polaridade de sentimento das respectivas IES em um determinado período de tempo estabelecido. Auxiliando como indicadores de o quanto a instituição propagou em determinada Mídia Social as polaridades de sentimento, seja ela positiva ou negativa.

## References

Benevenuto, F., Ribeiro, F., and Araújo, M. (2015). Métodos para análise de sentimentos em mídias sociais.

Campos, T. (2017). O que passa em 60 segundos na internet. Matéria em Jornal.

da Educação Superior, I. C. (2014). Notas estatísticas. *Instituto Nacional de Estudos e*

<sup>13</sup>Link da Base TASP-PT: <https://github.com/pauloemmilio/dataset>

- de Diretrizes, L. (1996). Bases da educação nacional.
- Dutra, E. (2012). Suicídio de universitários: o vazio existencial de jovens na contemporaneidade. *Estudos e Pesquisas em Psicologia*, 12(3):924–937.
- Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200.
- eMarketer (2017). Worldwide social network users: emarketer's estimates and forecast for 2016–2021.
- Evans, T. M., Bira, L., Gastelum, J. B., Weiss, L. T., and Vanderford, N. L. (2018). Evidence for a mental health crisis in graduate education. *Nature biotechnology*, 36(3):282.
- Fonseca, J. J. S. (2002). Metodologia da pesquisa científica.
- Franco, S. A., Gutiérrez, M. L., Sarmiento, J., Cuspoca, D., Tatis, J., Castillejo, A., Barríos, M., Ballesteros-Cabrera, M. d. P., Zamora, S., and Rodríguez, C. I. (2017). Suicide in university students in bogotá, colombia, 2004–2014. *Ciencia & saude coletiva*, 22(1):269–278.
- Gabriel, M. (2017). *Você, Eu e os Robôs – Pequeno manual do mundo digital*, volume 1. Atlas.
- Galvão, A., Câmara, J., and Jordão, M. (2013). Estratégias de aprendizagem: reflexões sobre universitários. *Revista Brasileira de Estudos Pedagógicos*, 93(235).
- Gerhardt, T. E. and Silveira, D. T. (2009). *Métodos de pesquisa*. Plageder.
- Gil, A. C. (2008). *Métodos e técnicas de pesquisa social*. 6. ed. Editora Atlas SA.
- Gomes, H. J. C. (2013). Text mining: Análise de sentimentos na classificação de notícias.
- Liddy, E. (2001). Natural language processing. encyclopedia of library and information science . new york, marcel decker. *Inc Google Scholar*.
- Liu, B. (2010). Sentiment analysis and subjectivity. *Handbook of natural language processing*, 2:627–666.
- Oliveira, D. J. S. and de Souza Bermejo, P. H. (2017). Mídias sociais e administração pública: Análise do sentimento social perante a atuação do governo federal brasileiro. *Organizações & Sociedade*, 24(82).
- Overflow, S. (2017). Developer survey results 2017.
- Pang, B., Lee, L., et al. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2):1–135.
- Polydoro, S. A. J., Santos, A., Medeiros, V., and Natario, E. (2005). Percepção de estudantes evadidos sobre sua experiência no ensino superior. *Questões do cotidiano universitário*. São Paulo: Casa do Psicólogo, pages 179–199.
- Recuero, R. (2012). A conversação em rede: comunicação mediada pelo computador e redes sociais na internet. *Porto Alegre: Sulina*, 201.
- Reece, A. G., Reagan, A. J., Lix, K. L., Dodds, P. S., Danforth, C. M., and Langer, E. J. (2017). Forecasting the onset and course of mental illness with twitter data. *Scientific reports*, 7(1):13006.

- Santaella, L. (2014). *Comunicação ubíqua: repercussões na cultura e na educação*. Pia Sociedade de São Paulo-Editora Paulus.
- Santos, A. d., Noronha, A., Amaro, C., and Villar, J. (2005). Questionário de vivência acadêmica: estudo de consistência interna do instrumento no contexto brasileiro. *Questões do cotidiano universitário*, pages 159–177.
- Scharl, A., Hubmann-Haidvogel, A., Jones, A., Fischl, D., Kamolov, R., Weichselbraun, A., and Rafelsberger, W. (2016). Analyzing the public discourse on works of fiction—detection and visualization of emotion in online coverage about hbo’s game of thrones. *Information processing & management*, 52(1):129–138.
- Tan, A.-H. et al. (1999). Text mining: The state of the art and the challenges. In *Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases*, volume 8, pages 65–70. sn.
- Wazlawick, R. (2017). *Metodologia de pesquisa para ciência da computação*, volume 2. Elsevier Brasil.
- Zanella, L. C. H. (2006). *Metodologia da pesquisa*. SEAD/UFSC.