

Análise de Algoritmos de Classificação para Detecção de Emoções em *Tweets* em Português Brasileiro

Title: Analysis of Classification Algorithms for Emotion Detection in Brazilian Portuguese Tweets

Daniel Kansaon¹, Michele A. Brandão², Saulo A. de Paula Pinto³

¹Departamento de Ciência da Computação
Universidade Federal de Minas Gerais (UFMG)
Belo Horizonte, MG – Brasil

²Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais (IFMG)
Ribeirão das Neves, MG – Brasil

³Instituto de Ciências Exatas e Informática
Pontifícia Universidade Católica de Minas Gerais (PUC-MG)
Belo Horizonte, MG – Brasil

daniel.kansaon@dcc.ufmg.br, michele.brandao@ifmg.edu.br,

saulo@sga.pucminas.br

Abstract. *With increasing access to the Web, large amounts of content are produced daily. The study of such contents allows the discovery of new knowledge. In this sense, this work presents an analysis of algorithms that allow the detection of emotions in tweets in the Brazilian Portuguese language. Thus, ten algorithms are considered, from decision trees to classifiers based on Bayes model, addressing altogether, seven classes of emotions: sad, upset, love, happy, anger, envy and irony. The results of the experimental evaluation are better when classifying relationships of distinct emotions, reaching 85% accuracy with a Naive Bayes algorithm. On the other hand, relations between close feelings present results inferior to 70% of correctness in some cases. Moreover, Naive Bayes-based classification algorithms present efficient results in a variety of contexts, in addition to having consistent language-independent behavior.*

Resumo. *Com o crescente acesso à Web, grande quantidade de conteúdos são produzidos diariamente. O estudo de tais conteúdos permite a descoberta de novos conhecimentos. Nesse sentido, este trabalho apresenta uma análise de algoritmos que permitem a detecção de emoções em tweets no idioma português brasileiro. Assim, são considerados dez algoritmos, desde árvores de decisão e classificadores baseados no modelo de Bayes, abordando ao todo, sete classes de emoções: triste, chateado, amor, feliz, raiva, inveja e ironia. Os resultados da avaliação experimental são melhores ao classificar relações de emoções distintas, chegando a acurácia de 85% com um algoritmo Naive Bayes. Por outro lado, relações entre sentimentos próximos apresentam resultados inferiores a 70% de acerto em alguns casos. Ademais, algoritmos de classificação baseados no Naive Bayes apresentam resultados eficientes em variados contextos, além de terem comportamento consistente independente do idioma.*

1. Introdução

Com a evolução da Web de um contexto apenas de comércio eletrônico, e-mail, buscas e acesso a conteúdo para um cenário de entendimento do pensamento humano, diferentes estudos têm sido realizados nessa direção. Por exemplo, Sharma et al. [2018] propõem uma aplicação baseada na Web cujo objetivo é não apenas analisar sentimentos, mas também mapear a intensidade dos termos geograficamente. Além disso, Moraes e Sampaio [2017] propõem uma taxonomia que utiliza análise de dados de redes sociais para auxiliar na gestão e atualização do conhecimento a partir de produções bibliográficas. Ambos os estudos permitem uma maior compreensão do comportamento e/ou pensamento humano e do conteúdo gerado por eles a partir de dados disponíveis na Web.

Os diferentes sistemas Web, como as redes e mídias sociais, são utilizados como canais de comunicação por diferentes pessoas de classes sociais, opiniões, atitudes e escolaridades distintas. Isso permite que tais sistemas sejam importantes fontes de dados e informações, conseqüentemente, de novos conhecimentos. Estima-se que a quantidade de informações mundiais, obtidas a partir de dados na Web, se duplica a cada vinte meses [Dwivedi et al. 2016], e esse prazo de duplicação tende a reduzir. Devido ao grande volume de dados disponíveis, pesquisadores têm buscado cada vez mais a exploração de tais dados para, por exemplo, encontrar informações valiosas sobre o mercado, produtos, clientes e concorrentes, o que pode permitir o aumento dos lucros e aumentar a competitividade das organizações.

Atualmente existem diferentes estratégias que permitem a análise e processamento de dados para auxiliar na extração de informações e construção de conhecimento. Por exemplo, a classificação de produtos em seus códigos para ajudar atividades contábeis e fiscais [Batista et al. 2018] e a identificação da tendência de sequências de crimes acontecerem em determinados períodos [LaFree 1999]. Outra estratégia é a análise de sentimentos e/ou emoções¹ que permite a extração e descoberta do sentimento e/ou emoção expresso em um texto. Esse tipo de análise é aplicado no contexto de redes e mídias sociais, além de estar associada à mineração de dados. A análise de sentimentos/emoções pode ser empregada em diversas situações, principalmente para entender o senso comum sobre um determinado acontecimento, como um evento em economia e política, por exemplo.

Nesse contexto, o objetivo principal deste trabalho é apresentar uma metodologia para a detecção de emoções expressas em *tweets* (postagens realizadas no Twitter²), bem como um comparativo do desempenho de vários algoritmos. Tais emoções podem ser identificadas com o auxílio de processos de mineração, cujo objetivo é extrair informações relevantes por meio de métodos de classificação para determinar as emoções expressas em *tweets*. Dessa forma, este trabalho foca em responder à pergunta de pesquisa: quais algoritmos de classificação podem melhor descobrir emoções em *tweets* em português brasileiro? Em outras palavras, é realizada uma pesquisa exploratória com o intuito de analisar o desempenho de diferentes algoritmos de classificação aplicados na análise de emoções semelhantes e distintas em português brasileiro. Ademais, também é verificado o desempenho dos algoritmos em uma base de dados com *tweets* em inglês para verificar

¹Sentimentos e emoções são conceitos distintos. Sentimentos são respostas às emoções, ou seja, refere-se a como uma pessoa se sente diante de uma emoção. Por outro lado, emoções surgem a partir de estímulos externos e são repostas químicas e neurais baseadas nas memórias emocionais [Rout et al. 2018].

²Twitter: <https://twitter.com/>

o quanto o idioma influencia no desempenho dos mesmos. Essas contribuições são uma extensão das apresentadas por Kansaon et al. [2018]. Em geral, este trabalho ampliou as análises realizadas no artigo anterior, validando as hipóteses apresentadas e os algoritmos em outros idiomas, trazendo uma discussão mais ampla sobre o desempenho dos algoritmos, bem como suas variações em diferentes contextos.

O restante deste artigo está organizado da seguinte forma. A Seção 2 apresenta os trabalhos relacionados. A Seção 3 descreve a metodologia utilizada e a Seção 4 descreve os resultados experimentais obtidos, detalhando a aplicação de cada método. Finalmente, a Seção 5 apresenta conclusões e comenta sobre trabalhos futuros.

2. Trabalhos Relacionados

Diversas áreas do conhecimento podem ser beneficiadas com a mineração de dados e a análise de sentimentos/emoções, por exemplo, política, educação e saúde, por possibilitarem a busca de informações significativas em grandes volumes de dados. Assim, existem diversos trabalhos recentes que enfatizam o uso da mineração de dados, especificamente, a aplicação de técnicas de classificação para a análise de sentimentos e/ou emoções. Note que muitos dos algoritmos aplicados na análise de sentimentos, também podem ser utilizados na análise de emoções. Por isso, também são discutidos aqui diversos trabalhos sobre essas análises.

Existem várias técnicas que podem ser empregadas nas análises de sentimento, Pandeye e Rajpoot [2016] por exemplo, utilizaram técnicas de classificação como J48, Random Forest, Decision Tree, Random Tree, Naive Bayes, Simple Naive Bayes e DecisionStump na análise de um conjunto de dados sobre o consumo de álcool por estudantes, obtendo uma acurácia de 95,44% com o Decision Stump. Além disso, existem outras abordagens, como feito por Caetano et al. [2017], que utilizaram a ferramenta *SentiStrength* para analisar sentimentos no Twitter sobre candidatos das eleições americanas. Tal estudo foi realizado com o objetivo de analisar a homofilia (a tendência de pessoas se relacionarem com semelhantes) entre usuários. Algumas comparações foram realizadas medindo o desempenho de diferentes métodos de análise de sentimentos, Araújo et al. [2015] apresentam um estudo comparando oito diferentes métodos para detecção de sentimentos em *tweets* em inglês. Por outro lado, Araújo et al. [2016] propõem uma análise de sentimento multilíngue, onde é feita tradução automática de textos para o inglês, e posteriormente, utiliza-se métodos existentes na língua inglesa para realizar a análise. Já Ribeiro et al. [2016] apresentam um Benchmark comparando os principais métodos de análise de sentimentos em diferentes cenários. Em linhas gerais não existe um método superior, visto que o desempenho varia de acordo com a base de dados e até mesmo o contexto.

Ademais, Yadollahi et al. [2017] realizam uma categorização de tarefas na análise de sentimentos e retratam a importância de uma taxonomia clara e lógica dentro dessa análise. Adicionalmente, métodos e melhorias do estado da arte para a análise de sentimentos em textos foram discutidos, além de aspectos teóricos ligados ao surgimento das emoções e sentimentos. Note que os sentimentos estão relacionados às emoções visto que são respostas a elas. Em complemento, Garimella et al. [2018] observaram que nas redes sociais, indivíduos tendem a ampliar os argumentos um dos outros principalmente em temas controversos.

Tabela 1. Modelos diferentes de emoções básicas propostas por teóricos. Fonte: adaptada de [Yadollahi et al. 2017].

| Teórico | Ano | Emoções Básicas |
|----------|------|--|
| Ekman | 1972 | raiva, desgosto, medo, alegria, tristeza, surpresa |
| Plutchik | 1986 | raiva, antecipação, desgosto, medo, alegria, tristeza, surpresa, confiança |
| Shaver | 1987 | raiva, medo, alegria, amor, tristeza, surpresa |
| Lovheim | 2011 | raiva, desgosto, angústia, medo, alegria, interesse, vergonha, surpresa |

Nesse contexto, a Tabela 1 apresenta algumas emoções básicas que podem ser encontradas em todos os seres humanos, bem como a respectiva classificação de acordo com as teorias da emoção [Yadollahi et al. 2017]. Nesse cenário, Rout et al. [2018] utilizam algoritmos supervisionados e não-supervisionados para analisar sentimentos (positivo, neutro, negativo) e emoções (raiva, medo, alegria, amor, triste, surpreso e grato) no Twitter. Em tal estudo, abordagens não-supervisionadas são utilizadas apenas na análise de sentimentos, enquanto que as supervisionadas são utilizadas na identificação de emoções e sentimentos. Ademais, Bouazizi e Ohtsuki [2016] propõem uma abordagem para quantificação de emoções no Twitter. Tal trabalho destaca a importância da detecção de emoções em *tweets* e apresenta uma maneira de extraí-las utilizando a técnica de classificação Random Forest. Foram coletados cerca de 40.000 *tweets* e atribuídas uma emoção para cada um, sendo elas: amor, felicidade, diversão, entusiasmo, alívio, ódio, raiva, tristeza, tédio preocupação, surpresa e neutralidade. A análise de tais estudos auxilia na escolha de sete emoções para identificar em *tweets* coletados neste trabalho: Triste, Chateado, Amor, Feliz, Raiva, Inveja, Ironia.

Em relação à comparação de várias técnicas de classificação, Garg e Khurana [2014] propõem um estudo em um contexto diferente da análise de sentimentos e/ou emoções. O objetivo é utilizar algoritmos de classificação para projetar um modelo eficaz de detecção de intrusão, impedindo que as redes de computadores sejam invadidas. Os métodos foram avaliados usando 41 atributos e cerca de 94.000 instâncias para o conjunto de treinamento, além de 48.000 instâncias para o conjunto de teste. Por fim, foi apresentada uma lista de 45 métodos de classificação ordenados pelos seus respectivos resultados, sendo que os cinco algoritmos com os melhores resultados foram Rotation Forest, Random Tree, Random Committee, Random Forest e IBK.

De maneira geral, grande parte dos trabalhos apresentam metodologias para a análise de sentimentos, ou seja, focam na classificação de sentimentos como positivos ou negativos. Já este trabalho, além da classificação tradicional, foca na detecção de emoções expressas em textos, apresentando um comparativo de algoritmos de classificação. A ideia é que a partir de um texto, seja possível detectar as emoções expressas (Amor, Feliz, Triste, Raiva, etc). Além do inglês, o trabalho analisa, principalmente, textos no idioma português brasileiro. Considerar tal idioma não é muito comum por não existirem muitas bases de dados disponíveis [Vilares et al. 2017], além de existirem poucos métodos e ontologias disponíveis no português para análise dos textos. Por fim, a metodologia proposta, que inclui coleta e processamento dos dados, apresenta um comparativo dos



Figura 1. Principais etapas da metodologia.

resultados dos principais algoritmos de classificação no contexto da análise de emoções.

3. Metodologia

Esta Seção apresenta os processos de coleta, processamento e preparação dos dados realizados, bem como a descrição dos algoritmos de classificação e o processo de treinamento dos classificadores. Especificamente, o trabalho busca identificar emoções expressas em *tweets* por meio de métodos de classificação. Para atingir o objetivo, é necessário realizar uma série de etapas para adaptar os dados para a execução dos diferentes algoritmos. A Figura 1 ilustra as principais etapas da metodologia, onde cada etapa pode ser executada mais de uma vez.

3.1. Seleção das Emoções

A definição das classes de emoções analisadas durante a pesquisa tem como base as diversas emoções que podem ser expressas em *tweets* e podem ser identificadas de várias formas, por exemplo, *hashtag*, *emojis* e *emoticons*. A *hashtag* é composta por uma ou algumas palavras-chave relacionadas ao assunto abordado na postagem. Elas são representadas pelo símbolo de cerquilha (#), seguida da palavra-chave. Posteriormente, as *hashtags* são indexadas para utilização dos mecanismos de busca. Os *emojis* e *emoticons* são amplamente usados no Twitter e estão diretamente associados a emoções. Os *emojis* são exibidos como imagens claras e podem conter animações. Por outro lado, os *emoticons*, são formas mais simples de expressar emoção e, geralmente, são representados por uma sequência de caracteres, por exemplo: :(, :D, :/ [Matsuda 2017].

As emoções analisadas neste trabalho foram selecionadas através do estudo feito por Yadollahi et al. [2017], que oferece uma contribuição importante, pois apresenta as emoções básicas que podem ser encontradas nos seres humanos (Tabela 1). Dessa forma, no trabalho, busca-se *tweets* que contenham as emoções: Triste, Chateado, Amor, Feliz, Raiva, Inveja, Ironia. Para isso, são utilizadas três abordagens distintas para a coleta dos dados, conforme descrito na Seção 3.2.

3.2. Coleta dos Dados

Após a seleção das emoções, é necessário realizar a extração e o armazenamento dos dados. A API oficial do Twitter fornece vários recursos para a coleta de *tweets*. Por meio dela, usuários que se cadastram como desenvolvedores conseguem coletar *tweets* a partir de parâmetros. A API possui algumas restrições para a coleta dos dados, por exemplo, a limitação de cem *tweets* como resposta para cada consulta realizada. Apesar disso, ela se apresenta suficiente para realização do estudo.

Assim, são criadas três bases de dados para comparação dos resultados. A primeira base de dados contém *tweets* coletados no idioma português brasileiro através do uso da *hashtag* (#). Já a segunda possui *tweets* em português brasileiro sem o uso da *hashtag*. Por fim, a terceira base de dados tem *tweets* no idioma inglês. O principal objetivo

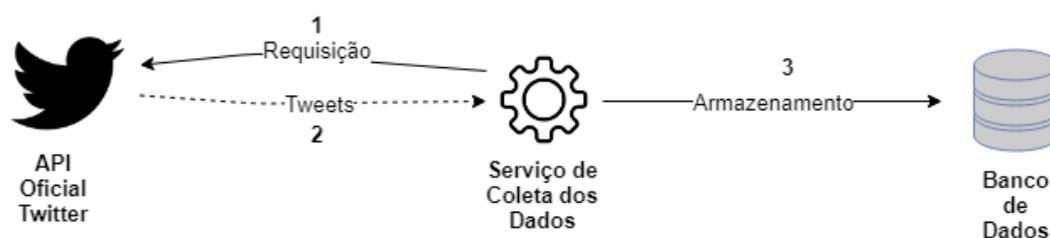


Figura 2. Fluxo da coleta dos dados.

de construir três bases de dados é validar os resultados, aplicando a mesma metodologia em diferentes quantidades de dados com características distintas. Além disso, nas três bases de dados, são armazenadas todas as informações relevantes dos *tweets*, sendo elas: texto, critério da busca, autor, data de criação, idioma, quantidade de URLs, quantidade de menções feitas, texto de URLs e tipo. Ao todo, foram coletados 381.756 mil *tweets* somando os dados disponíveis nas três bases de dados.

Coleta de *tweets* com *hashtag*. A Figura 2 apresenta o funcionamento do fluxo de coleta dos *tweets*. Toda a comunicação com a API é feita através do protocolo HTTP usando métodos de GET, onde as respostas retornadas são armazenadas na base de dados. A coleta foi feita entre os meses de agosto e outubro de 2017, contando com 12.814 *tweets* armazenados na base de dados. Para coletar *tweets* que expressam as emoções selecionadas, é definido que eles precisam conter a *hashtag* com o nome da emoção. Assim, são considerados os *tweets* que possuem as *hashtags*: #Triste, #Chateado, #Feliz, #Amor, #Raiva, #Inveja, #Ironia.

Coleta de *tweets* sem *hashtag*. Para busca, são utilizadas palavras individuais, como: Triste, Chateado, Feliz, Amor, Raiva, Inveja, Ironia. Assim, são coletados todos os *tweets* que possuem alguma dessas palavras no texto e que estejam no idioma português brasileiro. A ideia de não utilizar a *hashtag* para a busca é uma forma de obter *tweets* em contextos que não necessariamente expressam a emoção pela qual o *tweet* está classificado. A *hashtag* é comumente utilizada para agrupar conteúdos através de uma palavra-chave, onde geralmente a *hashtag* está relacionada ao assunto abordado no *tweet*. Assim, ao buscar pela *hashtag* se torna mais provável encontrar *tweets* que estão relacionados com o significado da palavra-chave. Dessa forma, a opção de utilizar somente a palavra como critério, visa ampliar a busca a fim de encontrar *tweets* em contextos variados. Para exemplificar, o texto [“RT @UOLNoticias: Bill Clinton pede a brasileiros que não votem orientados pela raiva e o ódio <https://t.co/Pj8WdDBKcD>”], que foi extraído da base de dados, contém a palavra raiva, mas o contexto mostra que se trata de uma notícia publicada no Twitter. Finalmente, no mês de setembro de 2018, foram coletados 175.116 mil *tweets*, considerando todas as emoções selecionadas.

Coleta de *tweets* em inglês. A principal diferença desta etapa em relação às anteriores está no idioma dos textos. O foco aqui é coletar *tweets* apenas em inglês, a fim de comparar os resultados com os obtidos para o português brasileiro. Para a coleta, as emoções selecionadas são traduzidas para o inglês, sendo elas: Sad, Upset, Happy, Love, Angry, Envy, Sarcasm. Dessa forma, são considerados todos os *tweets* que contém alguma dessas palavras no texto. Vale destacar, que nesta base utiliza-se a palavra Sarcasm em vez de Ironia. A escolha dessa abordagem se justifica pelo fato da palavra Irony em inglês ser

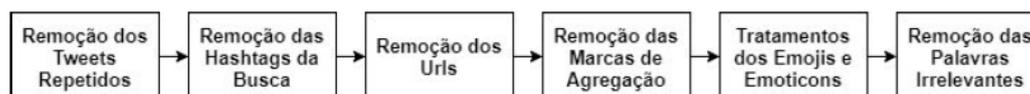


Figura 3. Etapas do processamento de cada tweets.

pouco utilizada nos tweets, logo, optamos por utilizar um sinônimo em inglês (Sarcasm). A coleta foi realizada no mês de setembro de 2018, ao todo foram obtidos 193.826 *tweets* considerando as sete emoções. A coleta é realizada utilizando as palavras sem *hashtag*, construindo uma base semelhante a base de dados em português sem *hashtag*, observando assim, os resultados dos classificadores em outro idioma.

Com as diferentes bases de dados construídas é possível comparar a acurácia dos algoritmos em diferentes cenários, observando se aspectos relacionados a quantidade de dados para treino, idioma, metodologia utilizada na coleta, afetam no desempenho dos classificadores.

3.3. Processamento dos Dados

As três bases de dados são utilizadas em todas as etapas de processamento descritas. O objetivo do processamento é tratar os dados coletados por meio da remoção das informações que não contribuem para classificação. Cada *tweet* contém uma grande quantidade de informações que são irrelevantes para a análise. Por isso, nesta etapa da metodologia é necessário a execução de vários passos conforme mostra a Figura 3, a fim de eliminar ruídos presentes no texto.

Assim, o primeiro passo foi a remoção de todos os *tweets* repetidos da base de dados, para que não influencie no resultado final. Isso se deu, pois observou-se que a API do Twitter retorna informações duplicadas. Em seguida, as *hashtags* usadas para a coleta dos *tweets* são removidas dos textos, os quais possuem links compartilhados e até mesmo notações utilizadas para mencionar outros usuários. Assim, as marcas de agregação identificadas através da sigla RT ou @ nome do usuário são removidas juntamente com os links encontrados. Ademais, o ambiente utilizado para a execução dos algoritmos não possui suporte a caracteres Unicode. Logo, por se tratarem de caracteres com formato específico, os *emojis* são transformados em palavras-chave, formadas pela junção da letra “E” maiúscula, completada posteriormente com uma palavra que representa o respectivo símbolo do *emoji*, por exemplo: ECoracao, EBravo, ESorridente. Por fim, as palavras conhecidas como *stop words* (a, as, de, para, etc) também são removidas. Para exemplificar, temos os seguinte *tweet*: [“meu deus minha empresa fez a primeira venda #feliz”], após todas as etapas de processamento, o texto fica: [“deus empresa fez primeira venda”].

3.3.1. Nivelamento dos dados

Para a primeira base de dados, que possui 12.814, uma emoção só é analisada se possuir um mínimo de mil *tweets*. Com isso, as emoções Ironia, Inveja e Raiva são excluídas. Aparentemente as *hashtags* com tais palavras não são comentários utilizadas na rede, assim não foi possível encontrar exemplos suficientes para treino dos classificadores, necessitando que fossem removidas. Após tal exclusão, a quantidade total de *tweets* é 9.631,

Tabela 2. Relações de emoções para a tarefa de classificação.

| BD com <i>hashtag</i> [12.814] | BD sem <i>hashtag</i> [175.116] | BD em Inglês [193.826] |
|--------------------------------|---------------------------------|------------------------|
| Feliz x Chateado | — | Happy x Upset |
| Triste x Chateado | — | Sad x Upset |
| Amor x Feliz | Amor x Feliz | Love x Happy |
| Feliz x Triste | Feliz x Triste | Happy x Sad |
| Amor x Triste | Amor x Triste | Love x Sad |
| — | Amor x Inveja | Love x Envy |
| — | Feliz x Raiva | Happy x Angry |
| — | Raiva x Inveja | Angry x Envy |
| — | Raiva x Triste | Angry x Sad |
| — | — | Envy x Sarcasm |
| — | — | Happy x Sarcasm |

BD: Base de Dados

assim, representando uma média de dois mil *tweets* para cada emoção.

Para as outras duas bases de dados, uma emoção só é analisada se possuir mais de dez mil *tweets*. A metodologia permitiu a obtenção de uma maior quantidade *tweets* para essas bases, sendo possível ampliar o volume de dados de cada emoção, assim analisando se a ampliação dos dados para treino reflete em uma melhora na taxa de acerto da classificação. Na base de dados sem *hashtag*, as emoções Chateado (6.482) e Ironia (5.644 mil) possuem cerca de seis mil *tweets*, com isso foram então removidas. Tudo indica que essas palavras não são muito utilizadas no Twitter, fazendo com que encontrar muitos exemplos seja uma tarefa difícil, inviabilizando uma comparação justa considerando as emoções Ironia e Chateado com as demais, que possuem cerca de vinte mil *tweets* em média. Já na base de dados de *tweets* em inglês, todas as emoções possuem quantidades representativas de dados. Assim, para cada emoção foram utilizados vinte mil *tweets* para a tarefa de classificação.

3.4. Preparação dos Dados

Após os tratamentos iniciais, os dados são divididos em algumas classes. No trabalho, definimos “Relação de emoções” como um agrupamento de duas classes de emoções, por exemplo: “Amor” é definida como uma classe, já “Amor x Feliz” é considerado uma relação de emoções, na qual os algoritmos pré-selecionados classificam/distinguem as classes ali presentes.

Com o nivelamento dos dados realizados, algumas emoções precisaram ser removidas conforme os critérios estabelecidos. Assim, para que a remoção não impactasse nas análises, as relações de emoções foram criadas visando manter uma estrutura em que seja possível comparar os resultados em diferentes bases de dados. Além disso, a seleção das relações visou contrastar as emoções separando-as em: positivas x negativas, positivas x positivas e negativas x negativas. Devido a existência de sete emoções se torna complexo realizar todas as combinações possíveis, com isso, em cada base de dados foi criado ao menos uma relação que representa as emoções contrastando-as entre semelhantes, opostas e neutras. No trabalho, as emoções Amor e Feliz são consideradas positivas, já as emoções Triste, Chateado e Raiva, representam emoções negativas. Por fim, as emoções

Inveja e Ironia representam emoções neutras, que podem ser positivas ou negativas.

A Tabela 2 apresenta as relações construídas para a classificação das emoções, onde cada coluna contém as relações construídas a partir das premissas definidas na Seção 3.3.1. Em algumas bases de dados não foi possível construir algumas relações de emoções por não existirem *tweets* suficientes para atender as premissas. Nesse caso, a notação “—” representa a não existência da relação.

3.5. Seleção dos Algoritmos de Classificação e Treinamento

Ao observar os objetivos do trabalho, percebe-se que estão diretamente associados às tarefas tradicionais de classificação. Ou seja, dado um texto, deseja detectar/classificar qual emoção está ali presente. Sendo assim, o trabalho seleciona vários algoritmos de classificação, priorizando agregar métodos com diferentes abordagens que apresentaram bons resultados em outros trabalhos.

Os algoritmos baseados no Naive Bayes são os principais utilizados no trabalho. Em geral, esses algoritmos utilizam a probabilidade condicional para criar o modelo de dados a ser trabalhado. Além do modelo tradicional, utilizou-se o Naive Bayes Multinomial, no qual a classe de um documento (*tweets*) é determinada não apenas pelas palavras existentes, mas também pela frequência que ocorrem [Witten et al. 2016]. Essa abordagem pode ser utilizada em textos em português, como França e Oliveira [2014], que através do algoritmo Naive Bayes atingiram resultados de até 90% de acurácia ao classificar polaridades expressas nos *tweets* relacionados aos protestos ocorridos no Brasil em 2013.

Além dos classificadores baseados no Naive Bayes, também foram selecionados algoritmos que apresentaram bons resultados em outros trabalho. Isso permite verificar se os resultados obtidos nesses trabalhos se repetem na classificação de *tweets* no idioma português brasileiro. Em Garg e Khurana [2014] foram avaliados o desempenho de vários métodos de classificação, conforme descrito na Seção 2. Assim, foram escolhidos os algoritmos que apresentaram bons resultados e que são compatíveis com o formato dos dados aqui trabalhados, são eles: IBK, Random Committee, Random Forest. O primeiro é encontrado na categoria Classificação Lazy. Esses classificadores são simples e eficazes, são usados principalmente para classificação de fluxos de dados. Em sequência o algoritmo Random Committee, que está na categoria Meta Classificador. Esses tipos de classificadores encontram o melhor conjunto de atributos para treinar, combinando as decisões de vários tipos diferentes métodos de classificação. Por fim, o Random Forest que está presente na categoria de Árvore de Decisão, que é uma técnica de classificação popular na qual uma árvore é montada como resultado, onde cada nó denota um valor de um atributo e os ramos apresentam o resultado. Em resumo, os algoritmos selecionados para a classificação dos *tweets* foram: Naive Bayes, Naive Bayes Multinomial, Naive Bayes Multinomial Updateable, Complement Naive Bayes, Bayesian Logistic Regression, Sparge Generative Model, DMNB Text, IBK, Random Forest e Random Committee. Inicialmente todos esses algoritmos foram aplicados na base de dados com *hashtag*. Após isso, foram selecionados os três principais algoritmos (Naive Bayes Multinomial Updateable, Naive Bayes Multinomial e Complement Naive Bayes) que apresentaram os melhores resultados, sendo aplicados nas demais bases de dados (sem *hashtag* e inglês), a fim de observar se as precisões sofrem variações com a mudança das bases e idioma.

Tabela 3. Base de dados com *hashtag*: Resultado da aplicação dos algoritmos na relação Amor x Triste.

| Algoritmo | Amor | | Triste | | Acerto (%) | Erro (%) |
|------------------------------------|-----------|----------|-----------|----------|------------|----------|
| | F-Measure | ROC Area | F-Measure | ROC Area | | |
| Naive Bayes | 0,59 | 0,87 | 0,77 | 0,87 | 70,72 | 29,27 |
| Naive Bayes Multinomial | 0,84 | 0,92 | 0,86 | 0,92 | 85,41 | 4,58 |
| Naive Bayes Multinomial Updateable | 0,84 | 0,92 | 0,86 | 0,92 | 85,54 | 4,45 |
| Sparge Generative Model | 0,82 | 0,92 | 0,84 | 0,92 | 83,84 | 6,15 |
| DMNB Text | 0,82 | 0,92 | 0,86 | 0,92 | 84,50 | 5,49 |
| Complement Naive Bayes | 0,84 | 0,85 | 0,86 | 0,85 | 85,64 | 4,35 |
| Bayesian Logistic Regression | 0,79 | 0,80 | 0,80 | 0,80 | 80,38 | 9,61 |
| IBK | 0,76 | 0,83 | 0,69 | 0,83 | 73,59 | 6,40 |
| Random Forest | 0,82 | 0,91 | 0,82 | 0,91 | 83,11 | 6,88 |
| Random Committee | 0,80 | 0,88 | 0,80 | 0,88 | 81,03 | 18,96 |

Já para o treino dos classificadores, é comum usar um terço dos dados para testes e dois terços para treino, mas o problema dessa abordagem é que, a parte usada para treino pode não ser representativa do problema e nem do conjunto de testes. A base de dados deve ser representada na proporção certa, pois se o problema não estiver bem representado, dificilmente qualquer algoritmo apresentará um bom resultado. Uma maneira de mitigar qualquer parcialidade causada pela amostragem, é utilizar o tipo de treinamento conhecido como validação cruzada (de dez partições ou vias). Nesse caso, os dados são divididos em dez partições de tamanhos aproximadamente iguais e cada uma delas é usada para testes e o restante é usado para treinamento. Esse processo é repetido dez vezes para que no final, cada instância seja usada uma vez para testar. Para obter a melhor estimativa de erro, é indicado usar dez partições [Witten et al. 2016].

4. Resultados

Esta seção apresenta a avaliação dos algoritmos quanto à sua capacidade de diferenciar as emoções expressas nos *tweets*. Especificamente, foram analisados os resultados de cada

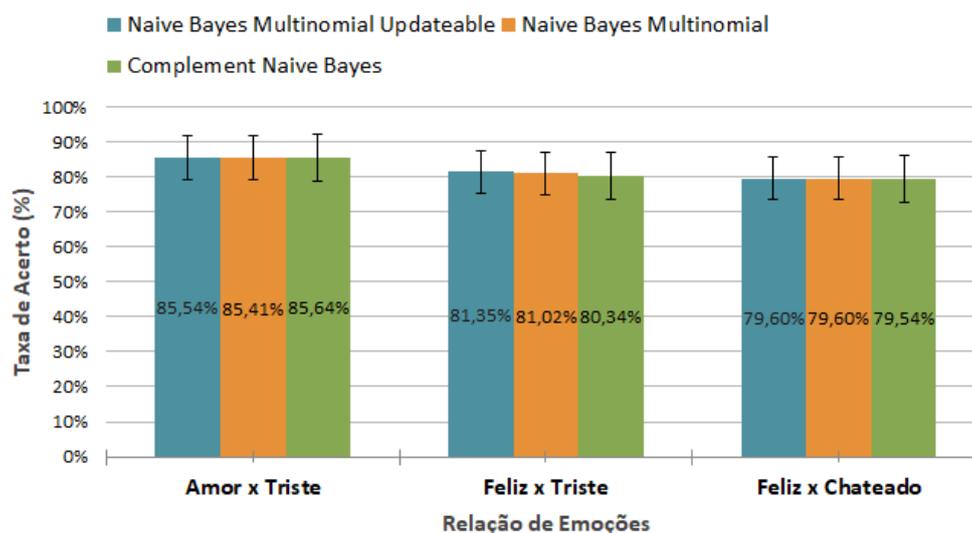


Figura 4. Base de dados com *hashtag*: Acurácia dos melhores algoritmos para detectar emoções positivas x negativas. A partir dos resultados em cada relação, os algoritmos apresentam uma média de taxa de acerto: Naive Bayes Multinomial Updateable 82,16% (+-3), Naive Bayes Multinomial 82,01% (+-3) e Complement Naive Bayes 81,84% (+-3), com 95% de confiança.

algoritmo nas bases de dados, considerando as palavras relevantes em cada emoção, bem como as características individuais de cada base de dados e as variações dos algoritmos.

Para facilitar a apresentação dos resultados, utilizaremos o termo “base de dados com *hashtag*” ao se referir aos dados coletados através da utilização de *hashtags*. Já o termo “base de dados sem *hashtag*”, se refere a base de dados construída sem a utilização das *hashtags* durante a coleta. Por fim, o termo “base de dados em inglês” está relacionado a base de dados considerando os *tweets* em inglês.

4.1. Resultados da Base de Dados com *hashtag*

Com base na aplicação dos algoritmos e divisão das diversas classes de emoções, alguns algoritmos se destacam na classificação, já outros não apresentam bons resultados. Em geral, os algoritmos baseados no modelo de Bayes apresentam os melhores resultados. Considerando o maior valor de acerto para combinação de classes que representam emoções semelhantes (positiva x positiva e negativa x negativa) e distintas (positiva x negativa), as taxas de acerto de classificação de emoções são 85,54% para o Naive Bayes Multinomial Updateable, 85,41% para o Naive Bayes Multinomial e 85,64% para Complement Naive Bayes.

As relações compostas por emoções distintas (positivas x negativas) apresentam melhores resultados. Os três algoritmos responsáveis por tal desempenho são apresentados na Figura 4. Em primeiro lugar, a relação Amor x Triste teve 85,64% de acerto na classificação com o método Complement Naive Bayes (a Tabela 3 apresenta resultados para todos os algoritmos de classificação aplicados para essa relação, o que permite observar as métricas F-Measure e ROC2 com valores maiores também para Complement Naive Bayes em relação a outros algoritmos). Uma das razões para esse resultado é a quantidade de *emojis* encontrados na relação. Os *emojis* Ecoracao e Esorridente foram

Tabela 4. Base de dados com *hashtag*: Palavras consideradas relevantes para a análise de cada emoção. As cores representam palavras semelhantes encontradas em classes diferentes.

| # | AMOR | FELIZ | TRISTE | CHATEADO |
|----|-------------|-------------|------------------|------------------|
| 1 | ECoracao | ESorridente | ERostoFranzido | ERostoDesanimado |
| 2 | ESorridente | ECoracao | EChorando | ERostoFranzido |
| 3 | Amor | Boas | ERostoDesanimado | Tristeza |
| 4 | Love | Alegria | Triste | ESemExpressao |
| 5 | Amamos | Agradecer | Acabando | Poxa |
| 6 | Paixao | Especial | Coitado | Triste |
| 7 | Amado | Feliz | ESemExpressao | Raiva |
| 8 | Desejo | Sorriso | Infelizmente | Ruim |
| 9 | Flores | Sucesso | Acabaram | Problemas |
| 10 | Fofura | Excelente | Aff | Sozinho |

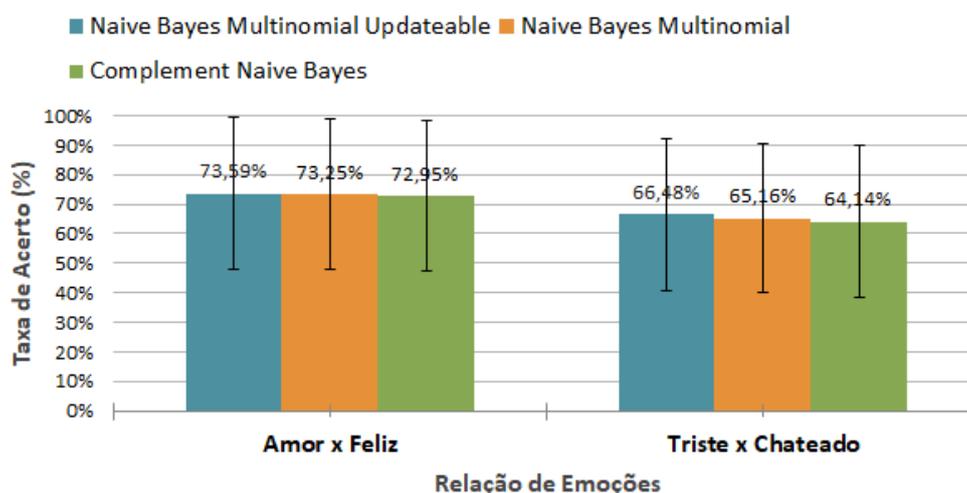


Figura 5. Base de dados com *hashtag*: Acurácia dos algoritmos para emoções negativas x negativas e positivas x positivas. A partir dos resultados em cada relação, os algoritmos apresentam uma média de taxa de acerto: Naive Bayes Multinomial Updateable 70,03% (+-31), Naive Bayes Multinomial 69,20% (+-36) e Complement Naive Bayes 68,54% (+-39), com 95% de confiança.

encontrados diversas vezes na emoção amor. Já para triste, não foi detectado nenhuma ocorrência desses *emojis*, conforme mostra a Tabela 4. Em seguida, a relação Feliz x Triste obteve 81,35% de acerto com o método Naive Bayes Multinomial Updateable. Os *emojis* ECoração e ESorridente encontrados na classe feliz, também não são localizados na classe triste. Por fim, a relação Feliz x Chateado ficou em terceiro lugar, com 79,60% de acerto na classificação, com os algoritmos Naive Bayes Multinomial e Naive Bayes Multinomial Updateable. Vale ressaltar que as palavras relevantes para a classificação das emoções (positivas x negativas) não se repetem. Consequentemente, essas relações são as que apresentam os melhores resultados comparado com as demais.

Em contrapartida, as relações compostas por emoções próximas (positivas x positivas ou negativas x negativas) apresentam os piores resultados. Os algoritmos usados

para a classificação não foram capazes de distinguir ou encontrar palavras que diferenciem essas emoções. Como consequência, os resultados das relações representadas por emoções semelhantes são inferiores às relações compostas por emoções distintas. Ademais, nas emoções triste e chateado, quatro das dez palavras consideradas relevantes se repetem, sendo elas: ERostoFranzido, ESemExpressao, Triste e ERostoDesanimado. Isso indica a semelhança entre as emoções, refletindo nos resultados obtidos.

A Figura 5 exibe a taxa de acerto dos três algoritmos que apresentam os melhores resultados nas relações de emoções semelhantes. O método Naive Bayes Multinomial Updateable obteve melhor resultado dos algoritmos aplicados na relação Chateado x Triste, com 66,48% de classificação correta. Já nas emoções feliz e amor, os *emojis*: ECoracao e ESorridente aparecem em ambas emoções, indicando a proximidade entre elas. Apesar da semelhança, a relação Amor x Feliz obteve um resultado de 73,59% de classificação correta com o método Naive Bayes Multinomial Updateable, sendo superior ao resultado obtido pela relação Triste x Chateado. Os algoritmos IBK, Random Forest, Random Committee não apresentam resultados eficientes quando comparados com os resultados apresentados no trabalho feito por Garg e Khurana [2014]. Isso se dá pelo fato do formato dos dados serem diferentes e estarem em outro idioma. Outro ponto importante que distingue os dois trabalhos é a quantidade de dados usados para as etapas de processamento e teste. Em Garg e Khurana [2014] são considerados cerca de 142.000 para o conjunto de treino e teste.

Além das etapas de treino e de processamento dos dados, que são importantes para a classificação das emoções, os *emojis* também se mostram essenciais para a classificação. Tais ideogramas representam as principais palavras para caracterizar uma emoção, diferenciando uma positiva de outra negativa.

4.2. Resultados da Base de Dados sem *hashtag*

Nesta etapa apenas os algoritmos que apresentaram os melhores resultados são aplicados, sendo eles: Naive Bayes Multinomial Updateable, Naive Bayes Multinomial e Complement Naive Bayes³. Tal abordagem é adotada, pois o objetivo é validar se com uma nova base de dados os resultados se mantêm consistentes. Por isso, para responder à pergunta, são utilizados apenas os algoritmos que foram eficientes para a classificação.

Assim como visto anteriormente, as relações compostas por emoções distintas, ou seja (positivas x negativas), apresentam os melhores resultados em relação às emoções próximas. Em contrapartida, se compararmos os resultados com os valores obtidos a partir da nova base com na base de dados com *hashtag*, verificamos que ocorre uma diminuição na taxa de acerto nas relações distintas, já nas próximas o resultado se mantém similar. Na nova base, as relações Amor x Triste (80,35%) e Feliz x Raiva (80,28%), foram as que atingiram as maiores taxas de acerto na classificação, conforme a Figura 6.

Quando observa-se os resultados em relação a base de dados com *hashtag*, observa-se uma diminuição na taxa de acerto de alguns algoritmos nas classes de emoções. Na Seção 4.1, o algoritmo Naive Bayes Multinomial Updateable alcançou uma precisão de 85,54% na classe Amor x Triste, já com a nova base de dados, na mesma classe, o algoritmo indica uma diminuição de aproximadamente 5% na taxa de acerto

³Note que esses algoritmos não possuem parâmetros específicos. Apenas utilizam distribuições de probabilidades distintas.

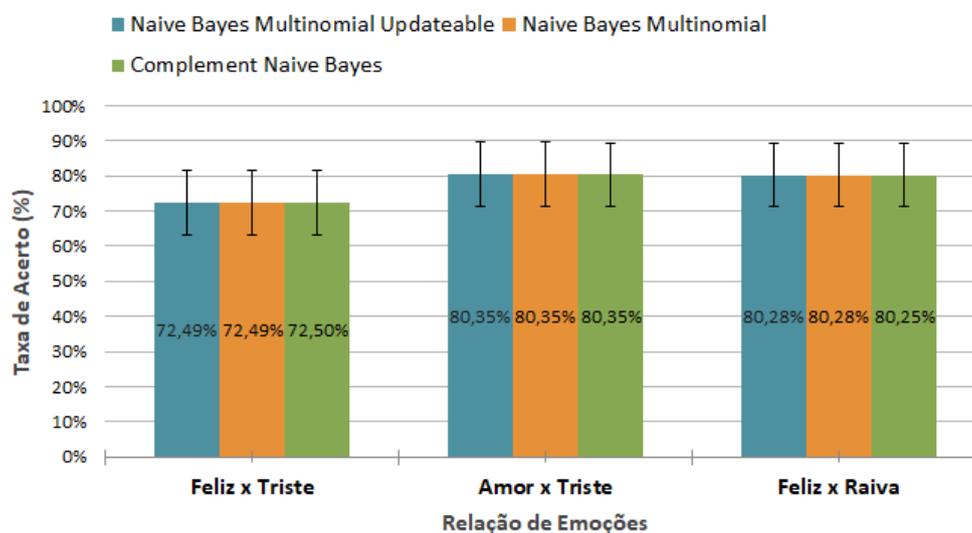


Figura 6. Base de dados sem *hashtag*. Acurácia dos melhores algoritmos para detectar emoções (positivas x negativas). A partir dos resultados em cada relação, os algoritmos apresentam uma média de taxa de acerto: Naive Bayes Multinomial Updateable 77,70% (+-9), Naive Bayes Multinomial 77,70% (+-9) e Complement Naive Bayes 77,71% (+-9), com 95% de confiança.

(80,35%), Figura 6. A classe Feliz x Triste, apresenta variações semelhantes, enquanto o Naive Bayes Multinomial Updateable apresentou 81,35% de acerto na classe (Figura 4), temos uma precisão de 72,49% nos novos dados, ou seja, uma redução de quase 10%.

Para entender a variação desses resultados, é importante observar as características individuais de cada base. O volume de dados é a principal diferença entre elas, enquanto a base anterior contém 12.814 *tweets*, a nova possui 175.116 mil. Com isso, a nova base possui mais dados para treino de cada algoritmo, logo, são mais exemplos para o aprendizado de cada emoção, que pode gerar um modelo mais bem treinado e por consequência mais preciso. Mas ao observar os resultados, vemos o oposto disso. Não houve melhora na taxa de acerto, e sim uma pequena diminuição de precisão. Os algoritmos treinados na nova base de dados com mais exemplos, tiveram uma precisão inferior aos treinados na base de dados com *hashtags*. Com isso, para entender os resultados e suas variações, é necessário analisar os dados de cada base bem como a metodologia utilizada para a coleta.

Na nova base, os dados foram coletados sem a utilização de *hashtags*, assim, encontra-se uma maior probabilidade de encontrar *tweets* que possuem as palavras definidas, mas que não expressam a emoção da busca. Como exemplo, temos o seguinte *tweet*: “Se a raiva passa e a saudade aperta, não duvide, é amor”, que foi obtido através da coleta usando a palavra raiva. Vemos que o texto contém a palavra raiva, mas não está expressando raiva ou emoção semelhante. Assim como o exemplo, na nova base, existem inúmeros outros *tweets* semelhantes, que não expressam a emoção da palavra da busca ou simplesmente são neutros. Em contrapartida, na base com *hashtag*, esses tipos de sentenças praticamente não existem, pois a busca por *hashtag* restringe o universo de coleta, além de que, os *tweets* que possuem a *hashtag* com a emoção, são mais propensos a expressarem no texto. Com isso, por mais que o algoritmo foi treinado com uma quan-

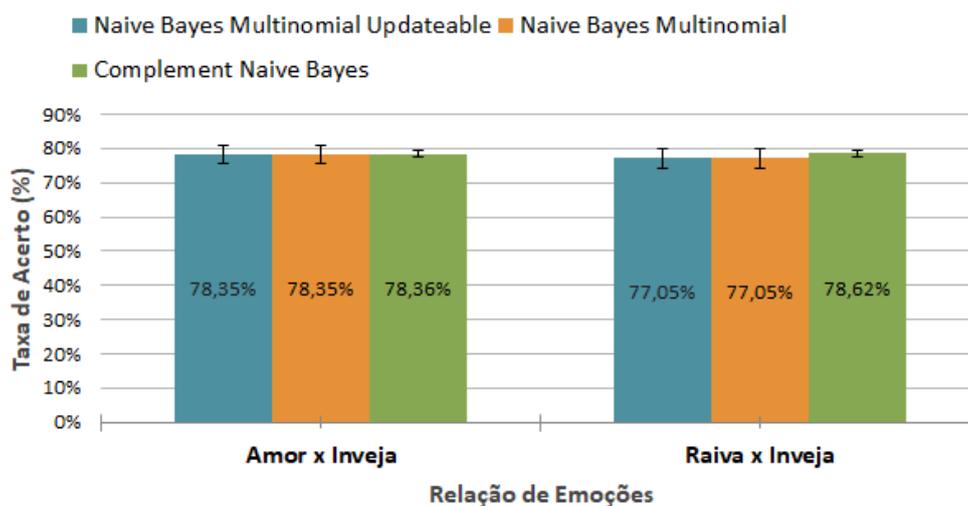


Figura 7. Base de dados sem hashtag: Acurácia dos melhores algoritmos para detectar emoções que foram consideradas como neutras, podendo ser negativas ou positivas. A partir dos resultados em cada relação, os algoritmos apresentam uma média de taxa de acerto: Naive Bayes Multinomial Updateable 77,77% (+-2), Naive Bayes Multinomial 77,77% (+-2) e Complement Naive Bayes 78,49% (+-2), com 95% de confiança.

tidade de dados bem superior, a complexidade das sentenças são diferentes em cada base. Assim, a variação da taxa de acerto pode ser entendida pela variação da complexidade, pois sentenças buscadas por *hashtags*, como: “meu deus minha empresa fez a primeira venda #feliz” tendem a apresentarem as emoções mais explicitamente do que os *tweets* obtidos apenas pela palavra da emoção.

Observando as relações neutras, nota-se um comportamento semelhante ao das relações próximas (positivas x positivas ou negativas x negativas). A Figura 7 exibe a taxa de acerto das relações com emoções neutras. O algoritmo Complement Naive Bayes teve uma precisão de 78,62% na relação Raiva x Inveja, em sequência, 78,36% na relação Amor x Inveja. Nota-se que existe uma pequena diferença na taxa de acerto dos algoritmos em ambas relações, se observarmos a Figura 8. Mas ao analisar as médias de execução contida na descrição das Figuras, fica evidente que os algoritmos não apresentam nenhuma superioridade na precisão.

Já quando observa-se as relações consideradas próximas (positivas x positivas ou negativas x negativas), na Figura 8, nota-se que houve uma leve variação na taxa de acerto se comparado com os valores obtidos na base anterior, Figura 5. Mais especificamente, o algoritmo Complement Naive Bayes apresentou 72,95% de acerto na relação Amor x Feliz, já na nova base, Figura 8, obteve 76,09%, uma pequena variação positiva de 3%, ou seja, a quantidade de dados não contribuiu para um aumento significativo na precisão.

Diferente das relações distintas (positiva x negativa), que apresentam uma piora na taxa de acerto com a mudança da base, as relações semelhantes possuem um comportamento diferente, revelam taxas de acerto semelhantes em ambas as bases. Isso sugere que com o aumento de dados para classificação, o algoritmo está mais propenso a encontrar outros padrões que diferencie as classes de emoções, mas apenas o aumento

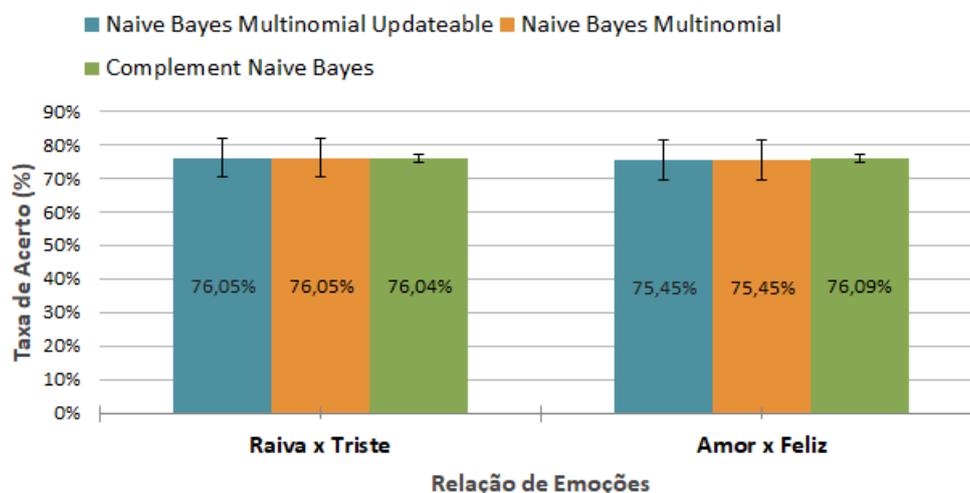


Figura 8. Base de dados sem hashtag: Acurácia dos melhores algoritmos para detectar emoções (positivas x positivas e negativas x negativas). A partir dos resultados em cada relação, os algoritmos apresentam uma média de taxa de acerto: Naive Bayes Multinomial Updateable 75,75% (+-5), Naive Bayes Multinomial 75,45% (+-5) e Complement Naive Bayes 76,06% (+-1), com 95% de confiança.

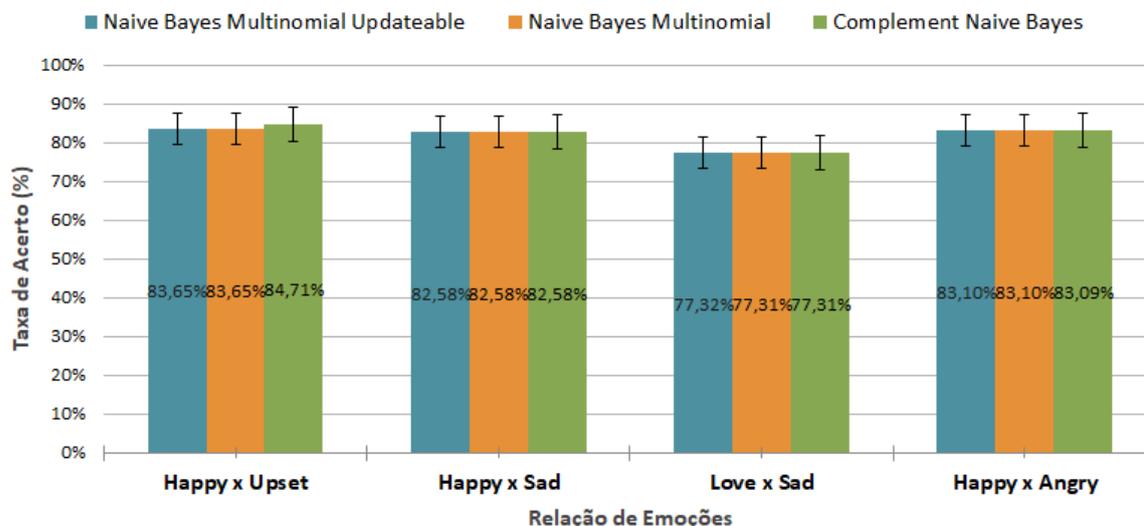


Figura 9. Base de dados em inglês: Acurácia dos melhores algoritmos para detectar emoções (positivas x negativas). Para os resultados em cada relação, os algoritmos apresentam uma média de taxa de acerto: Naive Bayes Multinomial Updateable 81,66% (+-4), Naive Bayes Multinomial 81,41% (+-4) e Complement Naive Bayes 81,92% (+-4), com 95% de confiança.

expressivo da quantidade de dados não foi suficiente para um crescimento significativo da precisão dos algoritmos. Isso porque não somente a quantidade de dados aumentou, mas também a complexidade das sentenças e a quantidade de ruído, dificultando o processo de classificação. Sabe-se também que ao analisar emoções semelhantes, a classificação tende a apresentar valores não muito satisfatórios, quando comparados com resultados

Tabela 5. Base de dados em inglês: Palavras consideradas relevantes para cada emoção. As cores representam palavras semelhantes encontradas em classes diferentes.

| # | LOVE | HAPPY | SAD | UPSET |
|----|-----------------------|-------------------------|---------------------------|---------------------------|
| 1 | E Coracao | Birthday | EChorando | Sad |
| 2 | E Sorridente | E Coracao | E RostoDesapontado | E RostoDesapontado |
| 3 | Very | E DoisCoracoes | Rip | :(|
| 4 | EChorando | :D | E Coracao | E Coracao |
| 5 | E DoisCoracoes | E Sorridente | :(| Sick |
| 6 | God | E CoracaoAmarelo | Bad | E RostoCansado |
| 7 | ESoprandoBeijo | Dear | Down | Liberals |
| 8 | Feeling | Happiness | Shit | EChorando |
| 9 | Happy | E CoracaoBatendo | Hate | Down |
| 10 | :(| Chorando | Very | Shit |

| # | ANGRY | ENVY | SARCASM |
|----|---------------------|----------------------------|----------------------------|
| 1 | Sad | EChorando | Understand |
| 2 | ERaiva | E Coracao | E Sorridente |
| 3 | White | E Sorridente | Trump |
| 4 | Trump | Jealousy | Saucy |
| 5 | E Coracao | Trump | Sense |
| 6 | EChorando | E OlhosArredondados | Sarcastic |
| 7 | Bad | Boast | EChorando |
| 8 | E Sorridente | Sins | E OlhosArredondados |
| 9 | Democrats | Lucky | E Coracao |
| 10 | Rage | Greed | Intended |

de relações com emoções distintas. Assim, semelhantemente como ocorrido nos dados coletados por *hashtag*, os algoritmos também não encontraram grandes padrões para diferenciar classes semelhantes.

4.3. Resultados da Base de Dados em Inglês

Conforme apresentado na Seção anterior, também serão aplicados apenas os três principais algoritmos que apresentaram os melhores resultados, sendo eles: Naive Bayes Multinomial Updateable, Naive Bayes Multinomial e Complement Naive Bayes.

Quando analisamos os resultados obtidos nas relações com emoções distintas, presente na Figura 9, percebe-se um comportamento similar nas bases com *tweets* em português. Já observando as relações individualmente, primeiramente temos o algoritmo Complement Naive Bayes com uma precisão de 84,71% na relação Happy x Upset, em seguida, o mesmo algoritmo apresentou 83,09% na relação Happy x Angry. Agora, se compararmos a relação Happy x Sad (82,58%) com a taxa de acerto da mesma relação em português, Feliz x Triste (72,49%), temos um ganho de 10% na precisão. Para entender o resultado, é importante observar as palavras envolvidas em cada classe. Na Tabela 5, percebe-se que as emoções possuem poucas palavras em comum, isso é interessante, pois os classificadores conseguem diferenciar cada classe de forma mais eficiente. Com exceção da palavra “E**Coracao**”, presente em todas emoções, os *emojis* [“E**DoisCoracoes**”, “E**Sorridente**”, “E**CoracaoAmarelo**”, “E**CoracaoBatendo**”] aparecem somente na classe Happy. Outro fato que justifica a boa classificação da relação Happy x Sad, é a

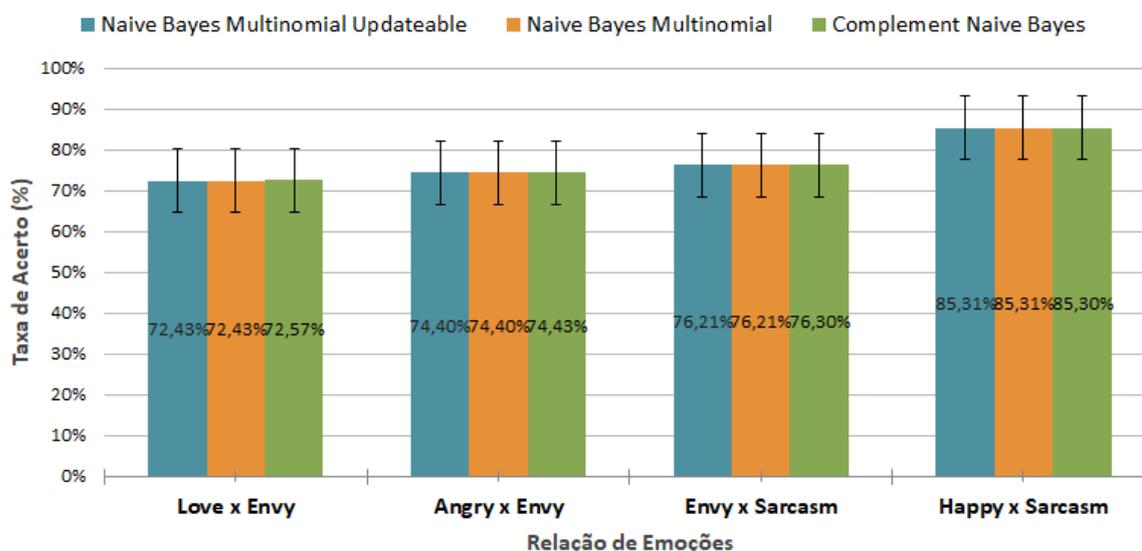


Figura 10. Base de dados em inglês: Acurácia dos melhores algoritmos para detectar emoções que foram consideradas como neutras, podendo ser negativas ou positivas. A partir dos resultados em cada relação, os algoritmos apresentam uma média de taxa de acerto: Naive Bayes Multinomial Updateable 77,08% (+-7), Naive Bayes Multinomial 77,08% (+-7) e Complement Naive Bayes 77,15% (+-7), com 95% de confiança.

presença de *emoticons*. A classe Happy contém o *emoticon* “:D”, já na classe Sad o *emoticon* “:(”. Esses *emoticons* ajudam a confirmar que nesse contexto, as emoções representam sentimentos opostos, assim, facilitando a classificação (“:D” está mais relacionado com emoções positivas, já “:(” com negativas).

Considerando as relações neutras, a relação Happy x Sarcasm teve um resultado acima da média se comparados com as outras relações da Figura 10. Com o algoritmo Naive Bayes Multinomial Updateable é possível ter uma taxa de acerto de 85,31% com a relação Happy x Sarcasm. Em geral, os classificadores detectaram padrões que diferenciam Sarcasm de outras emoções, como Happy e Envy. Ao observar a Tabela 5, é possível notar que diferentemente da emoção Love, Sarcasm não possui *emojis* como primeiras palavras de relevância, sendo [“EChorando” (7^o), “EOlhosArredondados” (8^o), “Ecoracao” (9^o)]. Pode parecer contraditório os algoritmos apresentarem uma boa taxa de acerto, visto que em quase todas as emoções, os *emojis* foram imprescindíveis, mas, quando observa-se a outra emoção da relação (Happy), a situação é oposta: [“Birthday”, (1^o) “ECoracao” (2^o), “EDoisCoracoes” (3^o), “:D” (4^o)], nota-se que os *emojis* estão entre as primeiras palavras relevantes. Dessa forma, por mais que existam *emojis* que se repetem nas duas relações, ao se comparar a relação Happy x Sarcasm, existem outras palavras presentes na classe Sarcasm, que são “mais relevantes” que os *emojis*. Assim, explicando a diferenciação da emoção Sarcasm de outras.

Já na relação Envy x Sarcasm, obtém-se uma taxa de acerto de 76,30% com o algoritmo Complement Naive Bayes. Observando as palavras relevantes, é possível identificar que o contexto Envy é bem diferente de Happy, isso reflete na diminuição da taxa de acerto, se comparado com a relação Happy x Sarcasm. A palavra “Trump” aparece nas

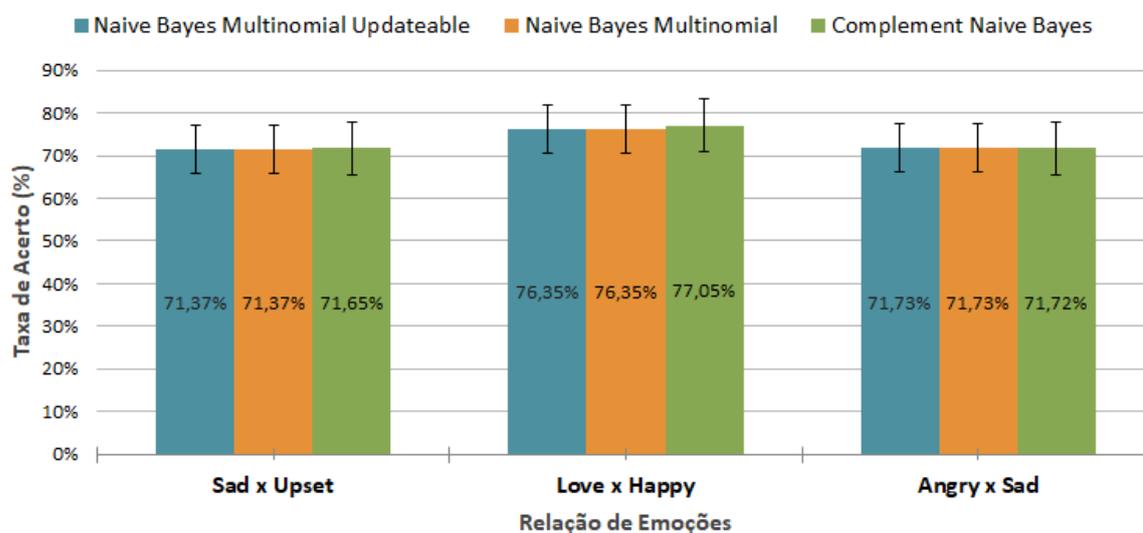


Figura 11. Base de dados em inglês: Acurácia dos melhores algoritmos para detectar emoções (positivas x positivas e negativas x negativas). A partir dos resultados em cada relação, os algoritmos apresentam uma média de taxa de acerto: Naive Bayes Multinomial Updateable 77,08% (+-5), Naive Bayes Multinomial 73,15% (+-5) e Complement Naive Bayes 73,47% (+-6), com 95% de confiança.

emoções Angry, Envy e Sarcasm. Além disso, nota-se que todos os *emojis* que aparecem em Envy também aparecem em Sarcasm. Mesmo com uma redução da taxa de acerto, os classificadores possuem uma precisão maior se comparados com outras relações mais próximas, como Sad x Upset.

Por fim, as relações com emoções próximas apresentaram resultados semelhantes ao da base de dados sem *hashtag*. O algoritmo Complement Naive Bayes apresenta um acerto de 77,05% na relação Love x Happy, conforme mostra a Figura 11. Já observando a relação Amor x Feliz, identifica-se uma taxa de acerto praticamente similar de 76,06% com o mesmo algoritmo. Já a relação Angry x Sad exibe uma taxa de acerto de 71,73% com o Naive Bayes Multinomial Updateable, mas ao observar a relação Raiva x Triste, o resultado é 76,05%. Nesse caso, o resultado obtido em inglês é um pouco inferior ao obtido em português, mas em linhas gerais seguem a mesma tendência.

Similarmente como ocorrido na base sem *hashtags*, ao observar os resultados individuais em cada relação de sentimento, é possível notar uma pequena redução na taxa de acerto. Um dos motivos que justifica a variação, é a maneira como a coleta dos dados é realizada. A principal diferença está no critério selecionado para a busca dos *tweets*. A base de dados em inglês é construída através da coleta de *tweets* com uma palavra específica, onde aumenta a probabilidade de se obter *tweets* em diversos contextos, que não necessariamente estão relacionados com a emoção que se deseja classificar. Outro fator, é que tanto no inglês como no português, as emoções podem ser expressas de formas distintas, assim, as diferenças dos resultados podem estar relacionadas com a subjetividade de quem escreveu.

Por meio dos resultados obtidos, é possível verificar que, na maioria dos casos, não

há uma diferença significativa ao se comparar as classificações em *tweets* em português brasileiro e inglês. Mesmo com variações nos resultados, constatou-se que o método Naive Bayes pode ser aplicado em *tweets* independente do idioma.

4.4. Análise das Variações dos Resultados

Mesmo com uma metodologia equivalente para a coleta dos dados, é possível notar a ocorrência de variações na acurácia dos algoritmos em diferentes base de dados. De forma mais detalhada, independentemente da quantidade de dados, da base utilizada ou até mesmo o idioma do texto, sempre é encontrado uma maior precisão na classificação de relações distintas do que relações com emoções semelhantes.

Nota-se que a precisão dos algoritmos está muito mais relacionada com o contexto do que com o idioma em si. Fatores como grande volume de dados para treino, por si só não garante um aumento da eficiência dos algoritmos. A quantidade de palavras distintivas como *emojis*, *emoticons* ou padrões no geral, muitas vezes são mais valiosos para a classificação das emoções. Por isso, classificar automaticamente sentimentos/emoções semelhantes ainda é desafiador, pois, tais sentenças tendem a conter palavras idênticas ou sinônimos, o que dificulta a tarefa. Outro fator que influencia os resultados é a qualidade do texto utilizado. Por isso, é imprescindível processar os dados eliminando ruídos ou textos irrelevantes que não contribuem para a classificação. Além disso, ao ampliar o contexto dos *tweets*, ou seja, coletar não só pela *hashtag*, mas sim pelas palavras das emoções, verificou-se uma diminuição da acurácia dos algoritmos. Como é o caso da relação Feliz x Triste (81,35%) na base de dados com *hashtag*, que ao considerar a base sem *hashtag* a precisão diminuiu para 72,49%. Assim, mesmo com uma quantidade superior de dados para treino, o algoritmo não aumentou a precisão. Isso se deu pelo aumento da complexidade das sentenças a serem classificadas, pois os *tweets* coletados pelas *hashtags* tendem a expressarem de forma mais explícita as emoções utilizadas na busca. Contudo, os resultados apresentados pelos algoritmos utilizados não estão relacionados diretamente com o idioma e sim com características presentes no texto.

Ao observar a taxa média de acerto de cada algoritmo (Naive Bayes Multinomial Updateable, Naive Bayes Multinomial e Complement Naive Bayes), nota-se que nenhum se mostra superior, i.e. todos apresentam resultados similares na classificação de emoções. Conforme exemplo na Figura 11, a média de cada algoritmo é diferente, mas ao observar o intervalo de confiança de 95%, verifica-se que os valores não apresentam variações estatisticamente significativas. Assim como no exemplo, os algoritmos apresentam comportamento similar em todos os outros cenários aqui estudados. Ademais, vale ressaltar que os intervalos de confiança foram calculados através da média da precisão dos algoritmos. Assim, em alguns cenários existem poucas relações de emoções, como é o exemplo da Figura 5, que possui apenas (Amor x Feliz) e (Triste x Chateado). Logo, o grau de liberdade para o intervalo de confiança é muito pequeno (igual a 1), ou seja, qualquer variação gera em um alto desvio padrão resultando em um amplo intervalo. Nesse exemplo, o algoritmo Naive Bayes Multinomial Updateable possui uma média de 70,03% (+-31). Já observando outros cenários em que existem mais relações, como a Figura 9, o intervalo de confiança tende a ser menor, pela mesma razão apresentada. Assim, é possível concluir que o intervalo de confiança obtido em amostras pequenas não é muito descritivo, devido ao amplo intervalo e ao curto grau de liberdade. Isso não invalida as análises ou a precisão dos algoritmos, pois observando os métodos em outros cenários, nota-se uma tendência e

um comportamento similar em várias bases de dados que alcançaram uma precisão maior ao classificar emoções distintas em relação às similares.

Por fim, o fato de que os três principais algoritmos são baseados no modelo de Bayes, fez com que eles apresentassem resultados similares. Mesmo que individualmente cada um possui evoluções distintas, por exemplo, os modelos Multinomiais utilizam uma distribuição multinomial em vez da abordagem tradicional que considera condições independentes para a rerepresentação das *features*. Já o Complement Naive Bayes, apresenta uma evolução com objetivo de evitar com que o viés presente nos dados afete a estimativa de pesos do modelo, fazendo com que sejam superiores ao modelo tradicional [Rennie et al. 2003]. Essas evoluções e variações individuais, quando aplicadas nos dados não foram suficientes para afirmar qual é o melhor modelo dentre os três. Mas em contrapartida, dentre os dez algoritmos escolhidos inicialmente, foram os que apresentaram os melhores resultados e se mostraram consistentes e eficientes em diferentes contextos.

5. Conclusão

Na literatura existem vários trabalhos que abordam problemas voltados para a análise de sentimentos, que focam em classificar textos em sentimentos positivos ou negativos. Já este trabalho, busca detectar não apenas os sentimentos, mas as emoções presentes em um texto. Para isso foram utilizados dez algoritmos de classificação, como árvores de decisão, classificadores modelo Naive Bayes, em outros modelos, para classificação de emoções em *tweets*. Apesar de ser um tema bastante estudado, ainda existem vários desafios para a descoberta de sentimentos e/ou emoções, especialmente, em textos na língua portuguesa. Ao todo, o estudo considerou sete emoções: triste, chateado, amor, feliz, raiva, inveja, ironia. Além do mais, construiu-se três bases de dados para a validação dos algoritmos em diferentes cenários. A primeira base contém 12.814 *tweets* no idioma português brasileiro, nessa base os dados foram coletados através do uso de *hashtags*. Já a segunda base possui 175.116 *tweets* que também estão no português brasileiro, além da quantidade de dados, outro ponto que difere é a maneira com que a coleta foi realizada. A busca foi feita através da palavra em si, por exemplo: “Triste” em vez de “#Triste”. Por fim, foi construída uma base de dados com 193.826 *tweets* em inglês, utilizando o mesmo critério de busca, por exemplo: “Sad” ao em vez de “#Sad”. Após a coleta, os dados foram separados em relações de emoções (positiva x negativa, negativa x negativa) e (positiva x positiva), e então aplicados os algoritmos selecionados, a fim de detectar não só a precisão ao classificar, mas também se o algoritmos conseguem distinguir emoções semelhantes. De forma geral, as relações com emoções distintas foram as que obtiveram melhores resultados, chegando a 85% em alguns casos. Constatou-se também que os algoritmos baseados no modelo de Bayes (Naive Bayes Multinomial Updateable, Naive Bayes Multinomial, Complement Naive Bayes) apresentaram os melhores resultados dentre os dez, além de se mostrarem consistentes ao serem aplicados em idiomas distintos.

Para trabalhos futuros, pretende-se expandir os métodos para detectar múltiplas emoções. Sabe-se que um texto pode conter várias emoções ao mesmo tempo, como por exemplo: Triste e Raiva. Assim, nesse contexto, os algoritmos detectam quantas e quais são as emoções presentes em cada sentença. Além do mais, planeja-se aplicar outros métodos comparando com os resultados já obtidos e quantificar a importância dos *emojis* para a detecção das emoções.

Agradecimentos. Este trabalho foi parcialmente financiado pela CAPES.

Referências

- Araujo, M., Gonçalves, P., and Benevenuto, F. (2015). Métodos para análise de sentimentos em mídias sociais. In *Proceedings of of Simpósio Brasileiro de Multimídia e Web (WEBMEDIA)*, pages 27–30.
- Araujo, M., Reis, J., Pereira, A., and Benevenuto, F. (2016). An evaluation of machine translation for multilingual sentence-level sentiment analysis. In *Proceedings of the 31st Annual ACM Symposium on Applied Computing*, pages 1140–1145. ACM.
- Batista, R. d. A., Bagatini, D. D., and Frozza, R. (2018). Classificação automática de códigos ncm utilizando o algoritmo naïve bayes. *iSys-Revista Brasileira de Sistemas de Informação*, 11(2).
- Bouazizi, M. and Ohtsuki, T. (2016). Sentiment analysis in twitter: From classification to quantification of sentiments within tweets. In *2016 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6.
- Caetano, J. A., Lima, H. S., dos Santos, M. F., and Marques-Neto, H. T. (2017). Utilizando análise de sentimentos para definição da homofilia política dos usuários do twitter durante a eleição presidencial americana de 2016. In *Procs. of the VI Brazilian Workshop on Social Network Analysis and Mining*, pages 480–491, São Paulo, Brasil.
- Dwivedi, S., Kasliwal, P., and Soni, S. (2016). Comprehensive study of data analytics tools (rapidminer, weka, r tool, knime). In *Procs. of Symposium on Colossal Data Analysis and Networking (CDAN)*, pages 1–8. IEEE.
- França, T. C. d. and Oliveira, J. (2014). Análise de sentimento de tweets relacionados aos protestos que ocorreram no brasil entre junho e agosto de 2013. In *of Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)*, pages 128–139. IEEE.
- Garg, T. and Khurana, S. S. (2014). Comparison of classification techniques for intrusion detection dataset using weka. In *Recent Advances and Innovations in Engineering (ICRAIE), 2014*, pages 1–5. IEEE.
- Garimella, K., Morales, G. D. F., Gionis, A., and Mathioudakis, M. (2018). Quantifying controversy on social media. *ACM Transactions on Social Computing*, 1(1):3.
- Kansaon, D. P., Brandão, M. A., and de Paula Pinto, S. A. (2018). Análise de sentimentos em tweets em português brasileiro. In *Procs. of the VII Brazilian Workshop on Social Network Analysis and Mining*, pages 52–63, Natal, Brasil.
- LaFree, G. (1999). Declining violent crime rates in the 1990s: Predicting crime booms and busts. *Annual Review of Sociology*, 25(1):145–168.
- Matsuda, Y. (2017). Development of emotion teaching interfaces using emoticons and emojis. In *2017 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, volume 2, pages 253–258.
- Moraes, R. L. and Sampaio, J. O. (2017). Um método de indexação e organização de informações baseado em um sociograma—aplicação em domínios específicos. *iSys-Revista Brasileira de Sistemas de Informação*, 10(2):48–84.

- Pandey, A. K. and Rajpoot, D. S. (2016). A comparative study of classification techniques by utilizing weka. In *Procs. of International Conference on Signal Processing and Communication (ICSC)*, pages 219–224. IEEE.
- Rennie, J. D. M., Shih, L., Teevan, J., and Karger, D. R. (2003). Tackling the poor assumptions of naive bayes text classifiers. In *Proceedings of the Twentieth International Conference on International Conference on Machine Learning, ICML'03*, pages 616–623. AAAI Press.
- Ribeiro, F. N., Araújo, M., Gonçalves, P., André Gonçalves, M., and Benevenuto, F. (2016). Sentibench - a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science*, 5(1):1–29.
- Rout, J. K., Choo, K.-K. R., Dash, A. K., Bakshi, S., Jena, S. K., and Williams, K. L. (2018). A model for sentiment and emotion analysis of unstructured social media text. *Electronic Commerce Research*, 18(1):181–199.
- Sharma, N., Pabreja, R., Yaqub, U., Atluri, V., Chun, S. A., and Vaidya, J. (2018). Web-based application for sentiment analysis of live tweets. In *Procs. of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, pages 120:1–120:2, Delft, The Netherlands.
- Vilares, D., Alonso, M. A., and Gómez-Rodríguez, C. (2017). Supervised sentiment analysis in multilingual environments. *Information Processing & Management*, 53(3):595–607.
- Witten, I. H., Frank, E., and Pal, C. J. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann.
- Yadollahi, A., Shahraki, A. G., and Zaiane, O. R. (2017). Current state of text sentiment analysis from opinion to emotion mining. *ACM Computing Surveys (CSUR)*, 50(2):25.