UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE INFORMÁTICA

RODOLFO COSTA CEZAR DA SILVA

Classificação Automática de Textos em Língua Portuguesa Com Traços de Racismo no Twitter

Goiânia
2018

UNIVERSIDADE FEDERAL DE GOIÁS

INSTITUTO DE INFORMÁTICA

AUTORIZAÇÃO PARA PUBLICAÇÃO DE TRABALHO DE CONCLUSÃO DE CURSO EM FORMATO ELETRÔNICO

Na qualidade de titular dos direitos de autor, **AUTORIZO** o Instituto de Informática da Universidade Federal de Goiás – UFG a reproduzir, inclusive em outro formato ou mídia e através de armazenamento permanente ou temporário, bem como a publicar na rede mundial de computadores (*Internet*) e na biblioteca virtual da UFG, entendendo-se os termos "reproduzir" e "publicar" conforme definições dos incisos VI e I, respectivamente, do artigo 5º da Lei nº 9610/98 de 10/02/1998, a obra abaixo especificada, sem que me seja devido pagamento a título de direitos autorais, desde que a reprodução e/ou publicação tenham a finalidade exclusiva de uso por quem a consulta, e a título de divulgação da produção acadêmica gerada pela Universidade, a partir desta data.

Título: Classificação Automática de Textos em Língua Portuguesa Com Traços de Racismo no Twitter

Autor(a): Rodolfo Costa Cezar da Silva

Goiânia, 06 de Dezembro de 2018.

Rodolfo Costa Cezar da Silva – Autor

Dra. Deborah Silva Alves Fernandes – Orientadora

RODOLFO COSTA CEZAR DA SILVA

Classificação Automática de Textos em Língua Portuguesa Com Traços de Racismo no Twitter

Trabalho de Conclusão apresentado à Coordenação do Curso de Ciência da Computação do Instituto de Informática da Universidade Federal de Goiás, como requisito parcial para obtenção do título de Bacharel em Ciência da Computação.

Área de concentração: Ciência da Computação.

Orientadora: Profa. Dra. Deborah Silva Alves Fernandes

Goiânia
2018

RODOLFO COSTA CEZAR DA SILVA

Classificação Automática de Textos em Língua Portuguesa Com Traços de Racismo no Twitter

Trabalho de Conclusão apresentado à Coordenação do Curso de Ciência da Computação do Instituto de Informática da Universidade Federal de Goiás como requisito parcial para obtenção do título de Bacharel em Ciência da Computação, aprovada em 06 de Dezembro de 2018, pela Banca Examinadora constituída pelos professores:

Prof. Dra. Deborah Silva Alves Fernandes
Instituto de Informática – UFG
Presidente da Banca

Profa. Dra. Nádia Félix Felipe da Silva
Universidade Federal de Goiás – UFG

Rodolfo Costa Cezar da Silva

Graduando em Ciência da Computação na UFG - Universidade Federal de Goiás. Durante a graduação foi monitor das disciplinas de Cálculo I e II. Participou do Programa Ciências Sem Fronteiras em 2016, onde cursou por 1 ano Ciência da Computação em Indiana University of Pennsylvania, localizada na cidade de Indiana, no estado da Pennsylvania, nos Estados Unidos.

Dedico este trabalho a todos que prezam por um mundo sem preconceito. Que este seja um pequeno passo em direção a esse mundo.

Agradecimentos

Aos meus pais, Maisa e Vilson, por serem pessoas em quem me espelhar. Pelo apoio e motivação durante toda a trajetória que me trouxe até aqui, e pelos valores ensinados, que eu carrego com orgulho até hoje e sem dúvida, carregarei para o resto da minha vida.

A minha irmã Amanda, minha inspiração pessoal, que apesar da distância se faz presente em meus dias, tornando-os mais alegres.

A minha querida companheira Gabriella, que me sustentou emocionalmente todos os dias com palavras de motivação, e gestos de carinho e amor. Desejo tudo de melhor pra nós em nossa futura trajetória.

A Profa. Dra. Deborah Fernandes pelo auxílio durante a escrita dessa trabalho, por ser uma ótima professora, orientadora, profissional e por ser, sem ao menos saber, uma grande motivação pra continuar na busca do conhecimento e por ser um ótimo exemplo a ser seguido.

Ao Governo Federal e a Universidade Federal de Goiás por oferecerem um curso de qualidade e gratuito.

A todos que direta ou indiretamente contribuíram para que essa conquista fosse possível.

“Ninguém nasce odiando outra pessoa por causa da cor da sua pele, ou sua origem, ou sua religião. As pessoas têm que aprender a odiar, e se elas podem aprender a odiar, elas podem ser ensinadas a amar, pois o amor chega mais naturalmente ao coração humano do que o seu oposto.”

Nelson Mandela,

.

Resumo

Da Silva, Rodolfo. **Classificação Automática de Textos em Língua Portuguesa Com Traços de Racismo no Twitter**. Goiânia, 2018. 41p. Relatório de Graduação. Instituto de Informática, Universidade Federal de Goiás.

O uso das redes sociais está inserido no cotidiano das pessoas atualmente. Usuários fazem dela uma plataforma para expressar suas opiniões e sentimentos. O Brasil é o segundo país com o maior número de usuários do Twitter, ficando atrás apenas dos EUA, e muitas vezes essa plataforma é utilizada para incitar ódio e denegrir pessoas por causa de sua raça ou etnia, portanto faz-se necessário o uso de técnicas de Análise de Sentimentos para detectar e classificar comentários desta natureza.
O objetivo deste trabalho é detectar traços de racismo em mensagens em língua portuguesa no Twitter. Para atingir este objetivo geral, dividiu-se-o em um conjunto de objetivos específicos, que incluem caracterização dos dados, experimentos com classificadores distintos sobre dois conjuntos de dados com características diferentes, a fim de identificar através de várias métricas, qual combinação apresenta uma melhor performance.

Palavras–chave

Aprendizado de Máquina, Racismo, Twitter, Análise de Sentimento.

Abstract

Da Silva, Rodolfo. **Automatic Classification of Portuguese Messages With Traces of Racism on Twitter**. Goiânia, 2018. 41p. Relatório de Graduação. Instituto de Informática, Universidade Federal de Goiás.

The use of Social Networks is inserted in the daily lives of people today. Users turn it into a platform to express their opinions and feelings. Brazil is the second country with the largest number of Twitter users, behind only the US, and often this platform is used to incite hatred and denigrate people because of their race or ethnicity, so it is necessary to use techniques to detect and classify comments of this nature.
The objective of this work is to detect traces of racism in Portuguese language messages on Twitter. To achieve this general objective, it was divided into a set of specific objectives, which include characterization of the data, experiments with distinct classifiers on two sets of data with different characteristics, in order to identify through several metrics, which combination presents a better performance.

Keywords

Machine Learning, Racism, Twitter, Sentiment Analysis.

Sumário

Lista de Figuras

Lista de Tabelas

CAPÍTULO 1

Introdução

As Redes Sociais Online (RSO) conduziram o Processamento de Linguagem Natural (PLN) para a Análise de Sentimentos (AS), cujo objetivo principal é automatizar a busca de opinião (o boca-boca na Internet) sobre algum tema no grande volume de dados não estruturados disponíveis eletronicamente na Internet[15].

Opinião não é um fato objetivo, mas um conceito formado por diversas experiências vivenciadas por um indivíduo ou por um grupo. A natureza dinâmica do conceito é percebida em cada relação estabelecida entre os entes sociais ou a partir de um fato novo observado [15]. Tal comportamento dinâmico é potencializado quando o universo de dados e relações aumenta vertiginosamente, o que é observado com a absorção das RSO no cotidiano mundial.

Não obstante, tal potencialidade ressalta aspectos positivos e negativos do cognitivo humano. Esse artigo perpassa pelo interesse nos meios de automatização do reconhecimento daqueles negativos relacionados aos discursos de ódio e preconceito. A análise de sentimentos é um desses meios [15].

Análise de Sentimentos, também conhecida por Mineração de Opinião, é a área de estudo que explora opiniões, sentimentos, avaliações, apreços, atitudes e emoções das pessoas sobre entidades tais como produtos, serviços, organizações, indivíduos, problemas, eventos, tópicos e seus atributos [15].

Essa técnica tem sido amplamente utilizada em uma variedade de aplicações, em diversos nichos de pesquisa. Cada aplicação utiliza as técnicas para fazer predições, auxiliar na tomada de decisões, classificar sentimento público, entre outras finalidades.

Em [21] há a descrição de técnicas e abordagens que visam auxiliar sistemas de informação orientados a opinião. O foco da pesquisa apresentada nessa referência é introduzir métodos que tratam os desafios apontados pelas aplicações que envolvem análise de sentimento e mineração de opinião tais como subjetividade e ambiguidade da linguagem natural, e trazer aplicações práticas para esses métodos.

O trabalho descrito em [24] propõe um sistema de auxílio a tomada de decisão

que realiza compra e venda de ações da Bovespa[1].Esse sistema utilizou de técnicas de análise de sentimentos sobre *tweets* que continham palavras relacionadas a ações de nove empresas brasileiras expressivas no mercado de ações. Os resultados obtidos mostraram que apesar das dificuldades que a área apresenta, o investimento nesta é promissor.

No trabalho de [18], é aplicado a análise de sentimentos e aprendizado de máquina para definir a correlação entre o sentimento do público e o "sentimento do mercado de ações". São usados dados coletados no Twitter para tentar prever as variações dos valores de ações da bolsa de valores baseadas no sentimento dos usuários e dados sobre os valores anteriores das ações. O modelo proposto obteve uma acurácia de 75.56% usando *Self Organizing Fuzzy Neural Networks* (SOFNN) aplicados a dados do Twitter e valores do índice DJIA, que é um dos principais indicadores dos movimentos do mercado norte-americano.

Em [28], os autores descrevem a realização de uma análise do sentimento público do Twitter sobre as eleições presidenciais dos Estados Unidos de 2012. Foi investigada a polaridade dos sentimentos sobre os candidatos e o efeito desta na opinião pública. Através de análises em tempo-real, os pesquisadores concluíram que a opinião pública muda a medida em que eventos políticos e notícias sobre os candidatos surgem.

O objetivo deste trabalho é detectar traços de racismo em textos em língua portuguesa no Twitter. A detecção de racismo, assim como outros problemas do Processamento de Linguagem Natural, sofre do problema de ambiguidade e subjetividade, portanto foi definido nos capítulos a seguir, algumas noções sobre racismo, e como foi delimitado o escopo desse problema. No capítulo 2 são apresentados trabalhos relacionados. A fundamentação teórica é descrita no capítulo 3, em que são apresentadas definições sobre racismo, racismo na Internet e são apresentados alguns casos de racismo na Internet. Análise de Sentimentos é abordada no capítulo 4 . Metodologia de desenvolvimento, que abrange a arquitetura de trabalho, coleta, informações sobre o conjunto de dados, classificação são abordados no capítulo 5, e os resultados e discussões são apresentados no capítulo 6. Conclusões e possibilidades para trabalhos futuros são descritos nos capítulos 7 e 8

[1]Bovespa é a bolsa de valores brasileira.

CAPÍTULO 2

Trabalhos relacionados a Racismo nas Redes Sociais

Um sistema de classificação automática de textos com traços racistas é proposto em [7]. Os autores treinaram um classificador *Support Vector Machine*(SVM) com os padrões encontrados a partir de *Bag-of-Words*(BOW) e bigramas.
O corpus de 3 milhões de palavras foi divido em *datasets* de diversos tamanhos que continham o mesmo número de documentos racistas e não-racistas. Utilizando um conjunto de treinamento de 2 mil documentos e outro de teste com 410 documentos, concluíram que a técnica de SVM combinada com BOW obteve melhor resultado comparada a técnica de SVM combinada com bigramas cujas taxas de precisão foram de 87.33% e 84.77%, respectivamente.

Modelos para classificação de discurso de ódio da Internet são apresentados por [29], especificamente no Twitter, para uma gama de características como etnia, deficiência e orientação sexual . Para tal, utilizaram técnicas para extração de características dos textos, relação sintática e gramatical entre palavras. O artigo explora a influência de características diferentes na tarefa de classificação, e para isso testa todos as combinações de características para verificar qual leva a um melhor resultado. A combinação de bigramas até 4-gramas combinada com o gênero obteve os melhores resultados, com 73.66% de acurácia.

A abordagem descrita no trabalho realizado por [9] teve como finalidade a detecção automática de comentários racistas em redes sociais holandesas. Foram extraídos 5759 comentários de páginas de redes sociais que continham orientação racista. Esses dados foram classificados em três categorias (*"racist", "non-racist", "invalid"*) por dois anotadores, e um terceiro anotador que era acionado quando havia discordância entre os dois primeiros. Técnicas de dicionário LIWC(*Linguistic Inquiry and Word Count*) e SVM (*Support Vector Machine*) foram adotadas para classificação automática de comentários. Além do dicionário LIWC, foi usado um dicionário expandido que contém palavras relacionadas a discurso racista. Esse método obteve uma taxa *F-score* de 0.46 quando comparado com as anotações feitas manualmente.

Pesquisadores propuseram em [14] mapear e mensurar a ocorrência de *cyber-bullying* contra professores no Twitter através de técnicas de aprendizado de máquina. Os dados foram coletados durante uma semana através da API do Twitter, foram também pré-processados e classificados em três categorias: positivo, negativo, e neutro. Para classificação utilizaram um classificador bayesiano (*Naive-Bayes*) que obteve uma acurácia de 87.1%.

O artigo de [22] visa detectar automaticamente discurso de ódio no Twitter através de classificadores *Long-Short-Term Memory*(LSTM) que são unidades de uma Rede Neural Recorrente (RNN). O modelo proposto também leva em consideração características comportamentais do usuário, i.e, se o usuário tem tendência de publicar mensagens contendo discurso de ódio. As mensagens são classificadas em três categorias: N, R, e S (*"Neutral", "Racism", e "Sexism"*, respectivamente). A tendência do usuário (T) é calculada baseada na proporção entre as mensagens neutras (Na), racistas (Ra) e sexistas (Sa) sobre todos as mensagens publicadas pelo usuário (M) (e.g: $T(Ra) = \frac{Ra}{M}$). O modelo descrito obteve acurácia de 92.95% quando incorpora-se informações sobre a tendência (T) do usuário, enquanto o modelo sem esses dados obteve uma acurácia menor de 90.89%.

O artigo descrito em [20] propõe a utilização do aprendizado de máquina para a detecção automática da ocorrência de *bullying* no Twitter. Neste trabalho, 2000 *tweets* coletados são rotulados quando à presença ou não de *bullying*. O autor faz uma comparação entre vários classificadores como Regressão Logística, *Support Vector Machines*, *Naive-Bayes*, e a melhor acurácia encontrada é de 72.8%.

CAPÍTULO 3

Fundamentação Teórica

3.1 Racismo

3.1.1 Definições Gerais

Apesar da população brasileira ser formada por aproximadamente 50.47% de pessoas pretas e pardas, segundo o Censo Demográfico do IBGE de 2010[1], o racismo ainda é uma questão recorrente no Brasil.

Embora seja um problema frequente, sua definição não é tão trivial, tendo em vista que ideias racistas podem ser expressas e percebidas de várias maneiras. Segundo [17], "racismo é o conjunto de teorias e crenças que estabelecem uma hierarquia entre as raças e etnias. É uma doutrina ou sistema político fundado sobre o direito de uma raça (considerada pura ou superior) de dominar as outras. Por fim, é um preconceito extremado contra indivíduos pertencentes a uma raça ou etnia diferente, considerada inferior".

Tratando-se de leis brasileiras, o crime de injúria racial está associado ao uso de palavras depreciativas referentes à raça ou cor com a intenção de ofender a honra da vítima, já o crime de racismo, previsto na Lei nº 7.716/1989, implica em conduta discriminatória dirigida a um determinado grupo ou coletividade e, geralmente, refere-se a crimes mais amplos.

3.2 Racismo na Internet

O racismo sempre existiu, porém com o desenvolvimento da Internet e das redes sociais, as pessoas passaram a ter uma plataforma para expressar e propagar suas opiniões, crenças e sentimentos com maior facilidade e visibilidade para outros usuários. Alguns casos podem ser citados:

[1] Disponível em : https://www.ibge.gov.br/

- Em 2005, o estudante de Letras na Universidade de Brasília (UnB), Marcelo Mello discutia o sistema de cotas para negros na sua universidade pelo Orkut (rede social descontinuada em 2014). Durante essa discussão, Marcelo se referiu aos negros e afrodescendentes como “burros”, “urubus”, “macacos subdesenvolvidos”, entre outras ofensas [12].
- A jornalista Maria Júlia Coutinho, a Maju do "Jornal Nacional", foi vítima de comentários preconceituosos na página oficial do programa no Facebook (figura 3.1), em julho de 2016, logo quando se destacou pela sua cobertura da previsão do tempo.
- No dia 30 de junho de 2018, um *youtuber* conhecido (figura 3.2) teceu comentários durante a Copa do Mundo FIFA de 2018, sobre o jogador da seleção francesa, Kylian Mbappé, associando o jogador com suas possíveis habilidades de realizar arrastões na praia.

Figura 3.1: *Comentário racista na página do Facebook de Maju*

Figura 3.2: *Youtuber faz comentário racista sobre jogador francês no Twitter.*

Nos exemplos citados acima, é possível perceber que o racismo pode se apresentar de forma explícita, mas também de maneira subjetiva e ambígua, o que torna a tarefa de identificar traços de racismo automaticamente bastante desafiadora. A fim de elucidar melhor o problema da subjetividade e ambiguidade do racismo, tome a figura 3.2. A frase foi publicada pelo *youtuber* durante um jogo da Copa do Mundo FIFA de 2018, e pode ter várias interpretações, dentre elas : (a) que o jogador tem muita força física, e poderia dominar várias pessoas e/ou é muito rápido, ou (b) que a aparência do jogador se encaixa nos padrões do *youtuber* de assaltantes que praticam arrastão[2].

[2] Arrastão é uma tática de roubo coletivo urbano.

CAPÍTULO 4

Análise de Sentimentos

Análise de sentimentos, também chamada de mineração de opinião, é uma área que estuda as opiniões, sentimentos, avaliações, atitudes e emoções sobre entidades, tais como produtos, serviços, organizações, indivíduos, eventos, tópicos e seus atributos [10]. Uma opinião regular expressa o sentimento sobre apenas uma entidade ou aspecto [13], por exemplo : *"Essa geladeira é muito boa!!"*, enquanto uma opinião comparativa relaciona diversas entidades baseados nos aspectos que elas compartilham entre si, por exemplo : *"A imagem da TV Sony é melhor do que a Samsung"*.

4.1 Níveis de Análise

Em geral, a Análise de Sentimentos trabalha em diferentes níveis de granularidade, e que pode ser feita em três níveis :

Nível de documento em que o trabalho consiste em classificar se um documento como um todo expressa um sentimento positivo ou negativo. Este nível de análise assume que cada documento expressa opiniões sobre uma única entidade [15].

Nível de sentença consiste em analisar a polaridade do sentimento de apenas uma sentença. Assume-se que em um documento pode conter várias sentenças que podem possuir um sentimento individual [15]. Cabe ressaltar que, em geral, postagens e comentários em mídias sociais seguem um padrão de sentenças curtas [3].

Nível de entidade e aspecto em que a granularidade é menor, tenta definir um sentimento sobre uma entidade, analisa diretamente a opinião em si, e é baseado na ideia em que uma opinião consiste em um *sentimento* e um *alvo*. O objetivo desse tipo de análise é descobrir as entidades e as respectivas opiniões sobre elas separadamente [15].

O alvo da opinião é importante e nos ajuda a entender a problemática da análise de sentimentos. A frase *"Embora o atendimento não seja bom, eu amo esse restaurante"* é um exemplo de uma sentença com sentimento positivo, porém não se pode afirmar que ela é totalmente positiva pois, se o alvo foi o "atendimento" então a opinião é negativa [15].

4.2 Abordagens

4.2.1 Aprendizado de Máquina

O aprendizado de máquina pode ser supervisionado, semi-supervisionado, ou não supervisionado. O primeiro consiste em fornecer um conjunto de dados de treino previamente coletados e anotados (com alguma polaridade de sentimentos) à um classificador que pode então, classificar novos dados após esse treinamento. Um dos métodos que podem ser utilizados é o *Support Vector Machines* (SVM) que classifica textos em alguma polaridade dado certo treinamento. Seu princípio é determinar separadores lineares no espaço de busca, chamados de *hiperplanos*. Através dos dados de treinamento, a SVM cria um hiperplano que divide as classes [15].

Outro método classificador é a Regressão Logística, um modelo probabilístico que atribui pesos (B) para atributos (X), dado um conjunto de treino conhecido (y). Esse encontra o conjunto de pesos que maximiza a probabilidade de $P(X|B,y)$. Para o aprendizado de máquina supervisionado também pode ser adotado um classificador *Naive-Bayes*, que é uma especificação da Regressão Logística que usa o Teorema de Bayes para determinar o conjunto de pesos. Um dos componentes necessários para os classificadores supervisionados são conjuntos de dados rotulados para o treinamento. Esses dados geralmente são manualmente anotados por seres humanos, ou por algum tipo de serviço de colaboração coletiva (*crowdsourcing*, do inglês) como o *Figure-Eight*[1].

Na abordagem não-supervisionada, também conhecida como aprendizado por observação e descoberta, não há o uso de dados de treinamento, nesse tipo de abordagem o classificador deve encontrar padrões automaticamente sem conhecimento prévio dos dados. Esse tipo de método geralmente usa abordagens léxicas para classificação que utilizam dicionários léxicos de sentimentos. Tais dicionários associam uma palavra com um significado quantitativo, que varia de $[-1,1]$ de acordo com sua polaridade, ou qualitativo, que associam uma palavra a uma certa polaridade (positivo/negativo, feliz/triste) [15].

É possível utilizar algoritmos de aprendizado para descobrir padrões nos dados a partir de alguma caracterização de regularidade, sendo esses padrões denominados *clusters* [6]. Exemplos que estão contidos em um mesmo *cluster* são mais similares do que exemplos contidos em *clusters* diferentes. O processo de formação de *clusters* é conhecido por *clustering*, e pode ser feitas usando várias técnicas que incluem : *Cluster* K-Médio, *Cluster* Hierárquico e Mapas Auto-Organizadores.

Existe também a abordagem semi-supervisionada, que parte do pressuposto que os conjuntos dos dados disponíveis para treinamento são formados por uma parte rotulada,

[1] https://www.figure-eight.com/

em menor número, e outra não rotulada, em maior número [27]. A ideia é utilizar os exemplos rotulados para obter informações sobre o problema e utilizá-las para guiar o processo de aprendizado a partir dos exemplos não rotulados [4].

4.2.2 Abordagem Baseada em *Léxicon*

Há a adoção de um *léxicon* para a análise de sentimentos que conta e atribui pesos a palavras relacionadas a sentimentos (*sentiment words*) que foram avaliadas e categorizadas. A coleta da lista de palavras pode ser feita através de duas abordagens automáticas. Apesar da técnica automatizada ser menos trabalhosa, ela pode cometer alguns erros, portanto deve-se combinar as duas técnicas.

Umas das abordagens é a baseada em corpus, em que um conjunto base de *sentiment words*(palavras de sentimento) com polaridade conhecida explora padrões de co-ocorrência para identificar novas palavras de sentimento e sua respectiva polaridade em um grande corpus, que pode ser previamente criado, ou utilizar serviços que disponibilizam corpora, como a WordNet[2], HowNet[3], entre outros.

O método baseado em dicionário explora recursos lexicográficos através de grandes base de dados léxicas disponibilizadas, como o WordNet e outras. A estratégia é coletar, manualmente, um conjunto de palavras de sentimento e suas polaridades, e expandir esse conjunto através de sinônimos e antônimos. O novo conjunto é usado para gerar novas palavras de sentimento, iterativamente [8].

[2]https://wordnet.princeton.edu
[3]http://www.keenage.com/

CAPÍTULO 5

Metodologia de desenvolvimento

5.1 Arquitetura de trabalho

A figura 5.1 ilustra o modelo da arquitetura proposta para o experimento. O processo se inicia através de um *crawler*, que é um software desenvolvido para extrair *tweets* (e informações associadas) do servidor do Twitter em tempo real. A medida em que essas informações são capturadas, elas são armazenadas em um banco de dados junto com outras informações sobre o *tweet*, como por exemplo texto, usuário, data, latitude, longitude, entre outras. Após o período de coleta, um pré-processamento é realizado sobre o conjunto de dados. O conjunto de dados é caracterizado quanto a frequência de palavras e localização geográfica dos *tweets*. Após isso, dividiu-se o conjunto de dados em subconjuntos que foram submetidos a classificadores.

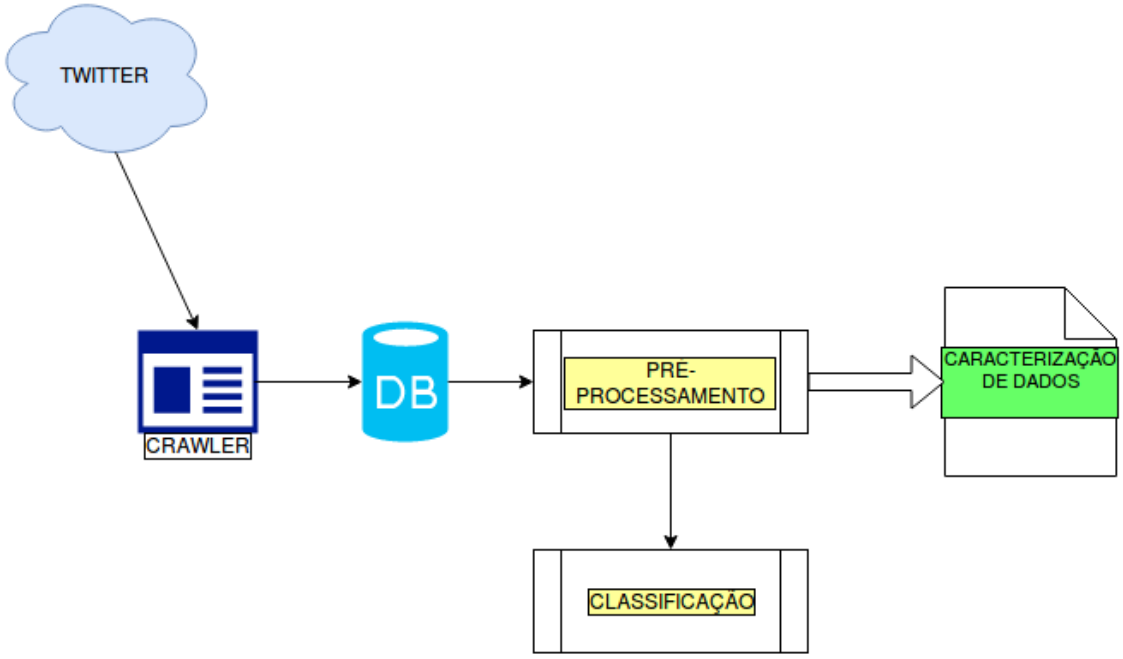


Figura 5.1: *Arquitetura proposta para análise e classificação de mensagens com traços de racismo.*

5.2 Dataset

5.2.1 Coleta

A coleta de dados foi realizada entre os dias 04/06/2018 e 04/10/2018 no Twitter. Durante esse período, foram coletados 106739 *tweets*. Optou-se pela rede social Twitter pelo fato de esta ser popular no Brasil com cerca de 27.7 milhões de usuários [2], por se tratar de *microblogging* com mensagens limitadas a 280 caracteres, e o fato da plataforma oferecer uma *Application Programming Interface*(API) que permite, de forma fácil, coletar dados em tempo real.

Os *tweets* coletados continham termos[1] que foram definidos previamente através de leituras, investigação de textos e de casos sobre racismo na Internet, redes sociais, jornais, palavras usadas em sistemas de buscas online, e também por meio de um questionário próprio (desenvolvido e disponibilizado no Google Forms) que ficou disponível durante uma semana no Twitter, a rede social alvo da pesquisa, trazendo à tona outros termos antes desconhecidos. O questionário era composto por três perguntas simples : "Você utiliza alguma rede social online?", "Você já sofreu por e/ou presenciou algum tipo de comentário que continha traços racistas em seu conteúdo?"e "O que caracterizava aquele texto como racista? Algum(ns) termo(s) em específico?". O questionário foi respondido por 14 pessoas que eram usuários da rede social em questão, e foi extramente útil para um melhor conhecimento da problemática, pois cada respondente que já sofreu/presenciou poderia dar uma contribuição única e pessoal sobre quais termos caracterizavam tal episódio vivenciado como sendo de cunho racista.

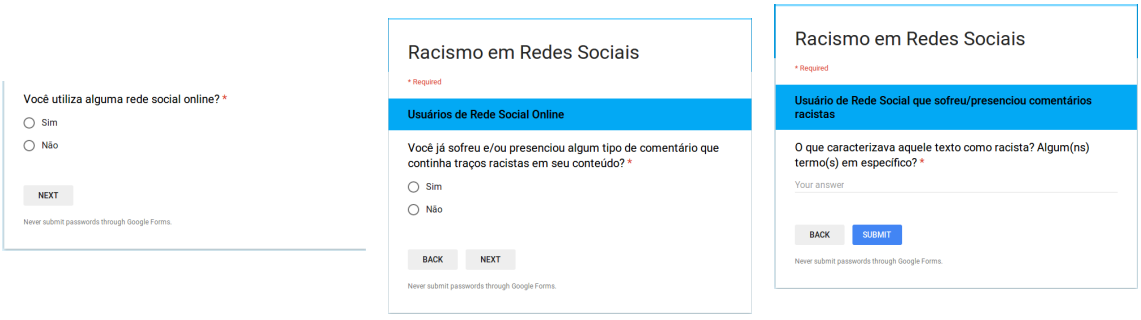


Figura 5.2: *Perguntas referentes ao questionário proposto.*

Em um projeto contra preconceito, o Professor Luiz Henrique Rosa, com a ajuda de cerca de 440 alunos de 11 turmas da Escola Municipal Herbert Moses, no Rio de Janeiro, realizaram um levantamento de 360 palavras de cunho racista[25]. O intuito era

[1]Termos de busca: "senzala", "gorila", "cabelo de bombril", "cabelo de esfregão", "nariz de nego", "nariz de nega", "tinha que ser preto", "tinha que ser preta", "preto da senzala", "preta da senzala", "preto da macumba", "nega macaca", "preto macaco", "preta macaca", "preta nojenta", "preto nojento", "criola", "crioula", "crioulo".

incluir os termos levantados pelo professor e alunos, mas ao tentar contactar o professor e a escola, não obteve-se sucesso.

Um dos obstáculos foi o fato da API coletar *tweets* cujo o nome de usuário, e não o texto do *tweets* em si, continham os termos de busca, *i.e*, um *tweet* enviado por um usuário chamado "gorilainvest"era coletado pelo *crawler*, mesmo que o corpo da mensagem não apresentasse nenhum comentário com traços racistas. Outra dificuldade encontrada foi a lapidação do conjunto de termos que seriam buscados. Nas coletas de teste, os termos “nega” e “neguinha” estavam incluídos nos termos de busca, mas após análise visual dos dados, percebeu-se que tais termos apenas tratavam de palavras de cunho racista em contextos específicos. O primeiro termo, na maioria da vezes, se referia a uma conjugação do verbo “negar”, e o segundo retornava várias mensagens de caráter apreciativo e afetuoso, e não discriminativo, portanto eles foram retirados do conjunto final de termos de busca.

5.2.2 Pré-processamento e rotulação

Do montante total de *tweets* coletados, decidiu-se realizar um pré-processamento a fim de remover as mensagens que continham *links*, pois observou-se que essas tratavam de mensagens que continham notícias que eram redirecionadas para outros sítios. Após removê-las restaram 83900 *tweets* de interesse. Também foram removidos os nomes de usuários do corpo das mensagens, ou seja, palavras que iniciavam com “@”, a fim de remover ruídos antes de utilizar os dados em um classificador.
Baseado no método utilizado em [20], 2022 *tweets* foram rotulados por dois anotadores[2] quanto à presença de traços de racismo nas mensagens. O conjunto de dados que foi apresentado aos anotadores continham apenas *tweets* com textos distintos que foram apresentados de forma aleatória, em que cada anotador rotulava o *tweet* como “sim” caso a mensagem contivesse traços de racismo, ou “não” em caso contrário. Do montante rotulado, 349 *tweets* foram rotulados como “sim” e 1676 como “não”. A Tabela 5.1 ilustra uma amostra dos dados anotados.

5.3 Caracterização de dados

Nuvem de palavras são utilizadas como uma maneira visual de apresentar a frequência de ocorrência de um certo termo ou palavra. Quanto maior o número de ocorrências de uma palavra, maior será a fonte utilizada para representá-la na nuvem de palavras. A figura 5.3 ilustra a nuvem de palavras do conjunto de dados em estudo.

[2]Os anotadores eram estudantes de cursos de graduação da Universidade Federal de Goiás, do sexo masculino e feminino, com idades de 22 e 25 anos.

Tabela 5.1: *Tabela de rotulação de dados*

texto	contém traços de racismo?
Antonio Rüdiger é um macaco, merece todo desrespeito do mundo, o pisão no pescoço do Pavard foi proposital, vai toma no cu verme preto.	sim
(...) Não estamos mais, qual o problema? Seu preto macaco, Gorila	sim
Ah pronto, charuto preto da Macumba agora quer falar merda	sim
alguem corta a Internet dessa escrava que fugiu da senzala	sim
ainda nao superei a morte da gorila koko	não
Alguém já leu "casa grande e senzala"?	não
Agora sem ser está descursao de hoje este cara do boné preto não sei como vc aceita e nojento de mais	não

Figura 5.3: *Nuvem de frequência de palavras.*

Através de uma análise visual, constatou-se que dentre as palavras mais frequentes estão : "gorila", "crioulo"e "senzala". Exemplos de tweets contendo essas palavras podem ser vistos na tabela 5.1. Geralmente, essas palavras são utilizadas como forma de denegrir pessoas negras através de comparação com animais, ou são palavras que fazem alusão ao período escravocrata.

5.3.1 Geolocalização dos *Tweets*

Dentre o conjunto de dados total, cerca de 64%(53589 *tweets*) continham informações geográficas(latitude e longitude) disponibilizadas pelos usuários, e que foram coletadas pelo *crawler*. O mapa apresentado na figura 5.4 representa os locais com maior concentração de publicação durante o período de coleta no Brasil.

É possível notar que a maior concentração de publicações analisadas é advinda das Regiões Sul e Sudeste, e o menor número em regiões como Norte e Nordeste do país. Isso pode ocorrer por inúmeras razões, como a taxa de acesso à Internet nessas regiões, nível de analfabetismo, proporção de população negra, entre outras.

Segundo pesquisa realizada em [1], regiões como Norte e Nordeste enfrentam

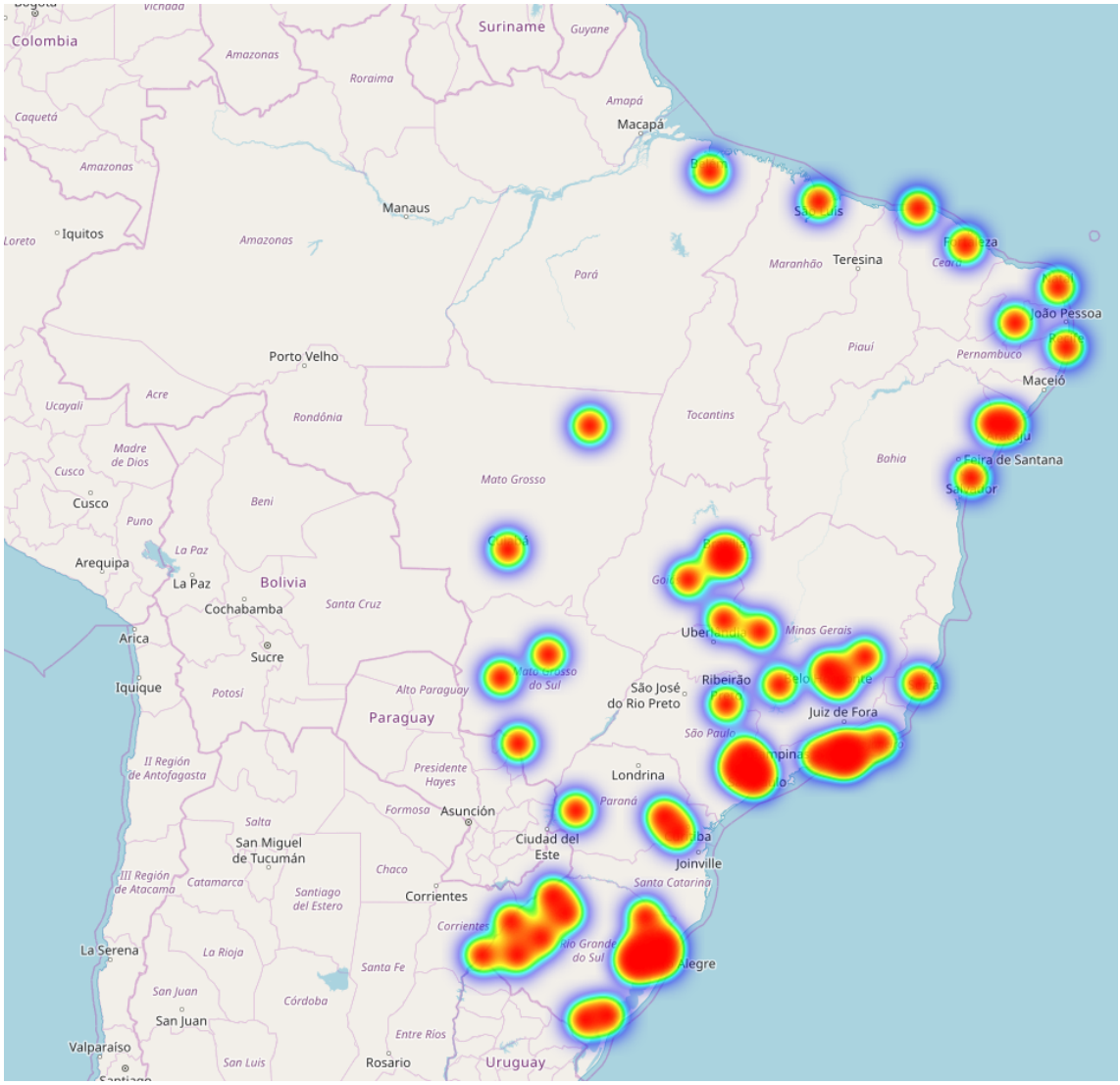
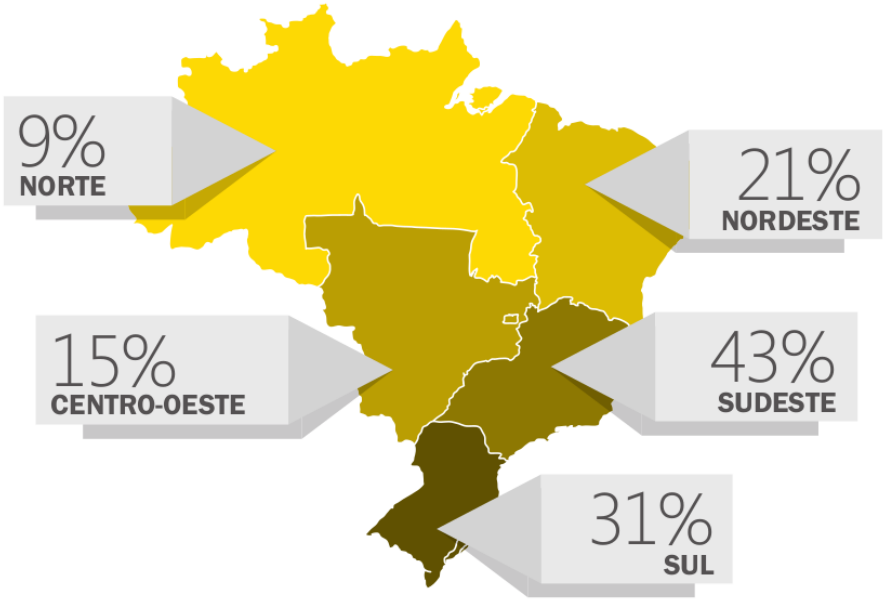


Figura 5.4: *Geolocalização dos tweets.*

dificuldades para o acesso à Internet devido a obstáculos como preço e disponibilidade, mas também destaca como obstáculo a falta de inclusão de indivíduos que não têm interesse ou não veem necessidade em acessar a rede. A figura 5.5 ilustra a proporção de empresas provedoras de serviços de Internet nas grandes regiões do país.

O Censo Demográfico de 2010 do IBGE aponta que as regiões Sul e Sudeste tem população minoritariamente pretas e pardas, correspondendo a 20.6% [3] e 46.6% [4] de suas respectivas populações. De acordo com o mesmo Censo Demográfico, regiões Norte e Nordeste tem população majoritariamente preta e parda, correspondo a 73.52% e 69%, respectivamente.

[3] Disponível em : ftp://ftp.ibge.gov.br/Censos/Censo$_Demografico_2010/Resultados_do_Universo/ods/Grandes_regioes/sul$
[4] Disponível em : ftp://ftp.ibge.gov.br/Censos/Censo$_Demografico_2010/Resultados_do_Universo/ods/Grandes_regioes/su$



Fonte: Pesquisa TIC Provedores 2014

Figura 5.5: *Proporção de empresas de serviços de Internet por regiões do Brasil.*

5.4 Classificação

5.4.1 Problema de classes desbalanceadas

No conjunto de dados de trabalho original existe o problema conhecido na literatura como "o problema de classes desbalanceadas", que consiste em uma desproporção no número de dados de uma das classes do treinamento. Na problemática abordada temos essa desproporção nas classes de *tweets* que contém traços de racismo (349 *tweets*) e nas classes de *tweets* que não contém traços de racismo (1676 *tweets*). Segundo [19] , modelos que são criados sob tais condições têm tendências de serem modelos com alta acurácia global, porém triviais, que predizem quase sempre a classe majoritária e não caracteriza um modelo bem representativo.

Uma das soluções para este problema é a utilização da técnica de balanceamento artificial como descrito em [26] , que consiste na remoção de instâncias da classe majoritária, a fim de construir um modelo mais robusto e mais preciso.

Modelos de classificação devem ser avaliados utilizando diversas métricas sensíveis à distribuição, e não apenas sua acurácia, pois um modelo de classificação pode ser bem preciso, porém não ser bem representativo, tendo em vista que as classes podem estar desbalanceadas. Considere, por exemplo, um conjunto de dados hipotético de tamanho 100, com 90 instâncias rotuladas como uma classe x, e 10 instâncias rotuladas como uma classe y. Se um classificador categorizar todas as 100 instâncias como sendo pertencentes a classe x, ele terá uma acurácia igual a 90%, porém não será um modelo bem

representativo. Para uma análise completa e fiel de um modelo, deve-se avaliar métricas como *F-score, Recall e ROC*.

Para o cálculo dessas métricas, faz-se necessário definir algumas terminologias, tais como :

- **True positive (TP)** : é a classificação correta da classe positiva. Por exemplo, a classe real é positiva e o modelo classificou como positiva.
- **True negative (TN)** : é uma classificação correta da classe negativa. Por exemplo, a classe real é negativa e o modelo classificou como negativa.
- **False positive (FP)** : é uma classificação errada da classe positiva. Por exemplo, a classe real é negativa e o modelo classificou como positiva.
- **False negative (FN)** : é uma classificação errada da classe negativa. Por exemplo, a classe real é positiva e o modelo classificou como negativa.
- **Precision (Precisão)** : A precisão é fornecida pelo o número de vezes que uma classe foi predita corretamente dividida pelo número de vezes que a classe foi predita:

$$Precision = \frac{TP}{TP+FP} \tag{5-1}$$

- **ROC :** acrônimo para *Receiver Operating Characteristic Curve*, essa métrica traça retas de **TP** e **FP** e mede a área sobre a curva (*Area Under Curve* - AUC). AUC sobre ROC serve para medir quão bom o classificador é, quanto mais próximo de 1 for o valor, mais representativo é o modelo.

Tendo essas terminologias bem definidas, pode-se calculas as métricas *F-score, Recall* e *ROC*.

$$Recall = \frac{TP}{TP+FN} \tag{5-2}$$

$$F-score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{5-3}$$

5.4.2 Dados de Teste

Para fins de análise e comparação, o montante total de *tweets* será particionado em dois *datasets* menores, cada um com suas características inerentes. O primeiro *dataset*, que será referenciado daqui em diante como **dataset1**, é um conjunto de dados em que foi aplicada a técnica de balanceamento artificial descrita na seção 5.4.1, além dos pré-processamentos descritos na seção 5.2.2. O segundo *dataset* de estudo, que será referenciado daqui em diante como **dataset2**, é um conjunto de dados que tem as

mesmas características do **dataset1**, com uma particularidade a mais que é a remoção de *stopwords*[5]. A tabela 5.2 mostra a característica de cada um dos *datasets* :

Tabela 5.2: *Características dos datasets*

dataset	número de instâncias	Características
dataset1	748	-nomes de usuários removidos -sem links
dataset2	748	-nomes de usuários removidos -sem links -stopwords removidas

Neste trabalho foram analisados dois classificadores : Naive-Bayes e Regressão Logística (*Logistic Regression*) através da ferramenta WEKA[6] (uma coleção de algoritmos de aprendizado de máquina desenvolvida em Java na Universidade de Waikato, Nova Zelândia) . Os atributos selecionados para serem submetidos aos classificadores foram : o texto do *tweet*, as *hashtags*, e o campo contendo a rotulação. Os dois classificadores nessa ferramenta foram configurados com a opção da ferramenta chamada *N-gram-tokenizer* que divide as palavras em unigramas(mínimo) até trigramas(máximo). Os dois classificadores foram aplicados sobre os dois datasets(*dataset1* e *dataset2*) descritos anteriormente, e testados sob três modos de teste, que são :

- ***k-fold-cross-validation***: Também denominada de validação cruzada de k partes, consiste em dividir o conjunto todo em k subconjuntos mutuamente exclusivos do mesmo tamanho, onde $k-1$ subconjuntos são utilizados para treino e 1 subconjunto para teste. Esse processo se repete k vezes alternando de forma circular o subconjunto de teste até que o modelo seja treinado e testado com todas as partes. Para este trabalho, foi escolhido $k = 10$.
- **Divisão Percentual**: Consiste em usar uma porção de dados para o treino, e utilizar o restante para teste. Para este trabalho, foi utilizado 75% de dados para treinamento e 25% para teste.
- ***Training set***: Consiste em testar o modelo sob as instâncias que foram utilizadas para treiná-lo.

[5] stopwords são palavras que carregam baixo valor semântico relevante para um classificador. *Stopwords* podem ser preposições, artigos, pronomes, etc.

[6] https://www.cs.waikato.ac.nz/ml/weka/

CAPÍTULO 6

Resultados e Discussões

6.1 Remoção de *stopwords*

Segundo [5] a remoção de *stopwords*, um método de pré-processamento amplamente utilizado na literatura, ajuda a produzir resultados mais precisos, alegando que essas palavras não carregam nenhum valor semântico relevante ao classificador.

Outros autores como [23], [16] e [11] afirmam que as *stopwords* carregam, de fato, informações relevantes aos classificadores e removê-las degrada a performance dos mesmos, tendo em vista que por se tratar de uma rede social com serviço de *microblogging*, usuários do Twitter tendem a fazer o uso de abreviações, gírias, expressões irregulares e etc.

Nos experimentos realizados, foram utilizados dois datasets que se diferenciavam por terem *stopwords* removidas ou não (*dataset2* e *dataset1*, respectivamente), e notou-se que a remoção de *stopwords* degradou a performance dos classificadores, seja Regressão Logística ou *Naive Bayes*, em todas as opções de teste.

Realizados os experimentos, foram analisadas as métricas acurácia, *F-score, Recall* e ROC, a fim de comparar como diferentes classificadores se comportam sob diferentes *datasets* e diferentes tipos de teste para o domínio proposto.

6.2 Resultados

As tabelas 6.1, 6.2 e 6.3 tratam dos resultados referentes aos dois classificadores aplicados ao dois *datasets*, sob os testes 10-*fold cross-validation*, Divisão Percentual(75-25), e *Trainning set*, respectivamente.

Tabela 6.1: *Resultados sob o teste 10-fold cross-validation.*

	10-fold cross-validation							
	dataset1				dataset2			
	Acurácia	F-score	Recall	ROC	Acurácia	F-score	Recall	ROC
Regressão Logística	76.336%	0.763	0.763	0.805	73.798%	0.737	0.738	0,798
Naive Bayes	66.176%	0.625	0.662	0.789	66.577%	0.630	0.666	0.786

Tabela 6.2: *Resultados sob o teste Divisão Percentual(75-25).*

	Divisão Percentual(75-25)							
	dataset1				dataset2			
	Acurácia	F-score	Recall	ROC	Acurácia	F-score	Recall	ROC
Regressão Logística	75.4011%	0.755	0.754	0.810	71.657%	0.716	0.717	0.771
Naive Bayes	66.176%	0.625	0.662	0.789	66.310%	0.632	0.663	0.768

Tabela 6.3: *Resultados sob o teste Trainning Set.*

	Trainning Set							
	dataset1				dataset2			
	Acurácia	F-score	Recall	ROC	Acurácia	F-score	Recall	ROC
Regressão Logística	91.310%	0.913	0.913	0.977	91.042%	0.909	0.910	0.976
Naive Bayes	97.326%	0.973	0.973	0.99	97.362%	0.973	0.973	1

Uma outra maneira de analisar o desempenho de um classificador, além das métricas descritas acima, é através da chamada "Matriz de Confusão", também conhecida como "Matriz de Erro", em que cada coluna representa as instâncias de uma classe prevista pelo classificador, e as linhas representam as instâncias reais de uma classe. Nas figuras 6.1 e 6.2 são apresentadas, respectivamente, as matrizes de confusão da melhor acurácia encontrada (97.326%)- obtida através do classificador *Naive Bayes* com modo de teste *Trainning set* para o *dataset1*- e pior acurácia encontrada (66.176%) -obtida através do classificador *Naive Bayes* com modo de teste *10-fold-cross-validation* para o *dataset1*-.

Figura 6.1: *Matriz de confusão para a melhor acurácia.*

Figura 6.2: *Matriz de confusão para a pior acurácia.*

Para o *dataset1*, obteve-se acurácias de 73.36% e 66.17% para os classificadores Regressão Logística e Naive Bayes, respectivamente, sob o teste *10-fold-cross validation*. Para o teste de Divisão Percentual(75-25), obteve acurácia de 75.40% para o classificador Regressão Logística, e 66.17% para Naive Bayes. No último modo de teste, *Trainning Set*, obteve-se acurácia para Regressão Logística e Naive Bayes, de 91.3% e 97.32%, respectivamente.

Para o *dataset2*, que teve as *stopwords* removidas, obteve-se acurácias de 73.79% e 66.57% para os classificadores Regressão Logística e Naive Bayes, respectivamente.

Para o teste utilizando Divisão Percentual(75-25) utilizando Regressão Logística e Naive Bayes, foram alcançadas as respectivas acurácias de 71.65% e 66.31%. No último modo de teste, *Trainning Set*, foram encontradas acurácias de 91.04% e 97.36% para os classificadores Regressão Logística e Naive Bayes, respectivamente.

CAPÍTULO 7

Conclusão

Povos indígenas habitavam o território do Brasil antes dos portugueses chegarem por volta do ano de 1500. Povos africanos involuntariamente migraram para o Brasil, pois foram capturados e trazidos para o Brasil para serem escravizados por senhores de engenho e barões do café. Todos esses movimentos migratórios, voluntários ou não, deram origem à população brasileira, bastante miscigenada, que possui uma grande riqueza cultural dentro de apenas um território. Apesar dessa miscigenação, da maioria numérica da população brasileira ser considerada preta ou parda, de 130 anos passados desde a sanção da Lei Áurea que aboliu a escravidão no Brasil, e da evolução tecnológica e de pensamento, problemas relacionados ao racismo ainda persistem.

Durante o desenvolvimento desta pesquisa, foi possível perceber que o racismo é bem mais presente do que se imagina no Brasil. Por se tratar de um país fundado na miscigenação e diversidade, acredita-se que não existiriam práticas discriminatórias contra uma pessoa baseada em cor de pele, mas esse trabalho traz à tona um pouco da realidade de uma parte da sociedade brasileira que trata um problema social tão sério como vitimismo e/ou brincadeira.

Este trabalho proporcionou um melhor compreendimento do processo de tradução de um problema social, como o racismo, em uma modelagem computacional através do uso de aprendizado de máquina e Análise de Sentimentos, evidenciando as dificuldades encontradas e que devem ser exploradas ao tratar de problemas relacionados a linguagem natural, como : uso de gírias, abreviações, subjetividade e ambiguidade em textos, processo de rotulação, remoção ou não de *stopwords*, quais classificadores são mais apropriados para cada tipo de problema, entre outros.

CAPÍTULO 8

Trabalhos futuros

Tendo em vista a dificuldade em classificar comentários/mensagens como "contendo traços de racismo" ou não, em trabalhos futuros pretende-se melhorar o processo de identificação de elementos textuais que podem configurar racismo, aplicando conhecimentos na área de ontologia para melhor delinear o domínio, assim como testar a aplicação de Babelnet[1], que é uma ferramanta que pode ser utilizada para criar uma rede semântica, expandir dicionários de sinônimos, etc.

Em futuras oportunidades pretende-se aprimorar o processo de anotação de dados, a fim de tentar reduzir o problema de subjetividade e ambiguidade, utilizando um modelo de rotulação com três anotadores, e cada mensagem podendo ter até três votos. Para uma mensagem ser rotulada com um certo rótulo x, esta teria que ter recebido dois votos x por dois anotadores distintos. Caso uma mensagem receba um voto x e um voto y, um terceiro anotador seria acionado para ter o *Voto de Minerva*[2] e decidir o desempate.

Almeja-se também trabalhar com outros tipos de abordagens de Análise de Sentimento, como por exemplo RNN (Recurrent Neural Networks), descrito em [22], a fim de comparar com os métodos propostos neste trabalho.

[1] https://babelnet.org/

[2] Voto de Minerva é o que decide uma votação que de outra forma estaria empatada.

Referências Bibliográficas

[1] **Panorama setorial da Internet Acesso à Internet no Brasil: Desafios para conectar toda a população**. Technical report, 2016.

[2] ASLAM, S. **Twitter by the Numbers: Stats, Demographics & Fun Facts**, 2018.

[3] BENEVENUTO, F. **Métodos para Análise de Sentimentos em mídias sociais**. 2015.

[4] BRUCE, R. **A Bayesian Approach to Semi-Supervised Learning**. 2001.

[5] C. SILVA.; B. RIBEIRO. **The importance of stop word removal on recall values in text categorization**. 2003.

[6] DECKER, K. M.; FOCARDI, S. **Technology Overview: A Report on Data Mining**. 1995.

[7] GREEVY, E.; SMEATON, A. F. **Classifying racist texts using a support vector machine**. In: *Proceedings of the 27th annual international conference on Research and development in information retrieval - SIGIR '04*, 2004.

[8] HAILONG, Z.; WENYAN, G.; BO, J. **Machine learning and lexicon based methods for sentiment classification: A survey**. In: *Proceedings - 11th Web Information System and Application Conference, WISA 2014*, 2014.

[9] HILTE, L.; LODEWYCKX, E.; VERHOEVEN, B.; DAELEMANS, W. **A Dictionary-based Approach to Racism Detection in Dutch Social Media**. 2005.

[10] HIRST, G.; LIU, B. **SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES Sentiment Analysis and Opinion Mining Sentiment Analysis and Opinion Mining**. 2012.

[11] HU, X.; TANG, J.; GAO, H.; LIU, H. **Unsupervised sentiment analysis with emotional signals**. In: *Proceedings of the 22nd international conference on World Wide Web - WWW '13*, 2013.

[12] ISTOÉ INDEPENDENTE. **O criminoso da internet**, 2015.

[13] JINDAL, N.; LIU, B. **Mining Comparative Sentences and Relations**. 2006.

[14] JOSÉ DE ALENCAR, R. **ESTUDO DA OCORRÊNCIA DE CYBERBULLYING CONTRA PROFESSORES NA REDE SOCIAL TWITTER POR MEIO DE UM ALGORITMO DE CLASSIFICAÇÃO BAYESIANO**. p. 5–1, 2012.

[15] LIU, B. **Sentiment Analysis and Opinion Mining**. 2012.

[16] MARTÍNEZ-CÁMARA, E.; MONTEJO-RÁEZ, A.; MARTÍN-VALDIVIA, M. T.; UREÑA, L. A.; UREÑA-LÓPEZ, U. **SINAI: Machine Learning and Emotion of the Crowd for Sentiment Analysis in Microblogs**. Technical report, 2013.

[17] MARTINS, I. C. **O racismo nas redes sociais : O mundo virtual é feito por pessoas de carne e osso!**, 2014.

[18] MITTAL, A.; GOEL, A. **Stock Prediction Using Twitter Sentiment Analysis**. 2012.

[19] MONARD, M. C.; BATISTA, G. E. A. P. A. **Learning with Skewed Class Distributions**. 2003.

[20] MONTEIRO E SILVA, G. **Aplicando Técnicas de Aprendizado de Máquina para Detecção Automática de Bullying no Twitter**. 2017.

[21] PANG, B.; LEE, L. **Opinion Mining and Sentiment Analysis**. *Foundations and Trends® in InformatioPang, B., & LeFoundations and Trends® in Information Retrieval, 1(2), 91–231.*, 1(2):91–231, 2006.

[22] PITSILIS, G. K.; RAMAMPIARO, H.; LANGSETH, H. **Detecting Offensive Language in Tweets Using Deep Learning**. 2018.

[23] SAIF, H.; HE, Y.; ALANI, H. **Semantic sentiment analysis of twitter**. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2012.

[24] SILVA ALVES, D. **Uso de técnicas de Computação Social para tomada de decisão de compra e venda de ações no mercado brasileiro de bolsa de valores**. 2015.

[25] SOUZA, M. **Para projeto contra preconceito, professor lista 360 termos racistas**.

[26] SPINELLI SCHIAVONI, A. **UM ESTUDO COMPARATIVO DE MÉTODOS PARA BALANCEAMENTO DO CONJUNTO DE TREINAMENTO EM APRENDIZADO DE REDES NEURAIS ARTIFICIAIS**. 2010.

[27] VICENTE, B.; DE LIMA, A.; MACHADO, V. P.; DE MELO, R.; VERAS, S. **Abordagem Semi-supervisionada para Rotula ao de Dados**.

[28] WANG, H.; CAN, D.; KAZEMZADEH, A.; BAR, F.; NARAYANAN, S. **A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle**. *Jeju, Republic of Korea*, p. 115–120, 2012.

[29] WASEEM, Z.; HOVY, D. **Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter**. p. 88–93, 2016.

Referências Bibliográficas

[1] **Panorama setorial da Internet Acesso à Internet no Brasil: Desafios para conectar toda a população**. Technical report, 2016.

[2] ASLAM, S. **Twitter by the Numbers: Stats, Demographics & Fun Facts**, 2018.

[3] BENEVENUTO, F. **Métodos para Análise de Sentimentos em mídias sociais**. 2015.

[4] BRUCE, R. **A Bayesian Approach to Semi-Supervised Learning**. 2001.

[5] C. SILVA.; B. RIBEIRO. **The importance of stop word removal on recall values in text categorization**. 2003.

[6] DECKER, K. M.; FOCARDI, S. **Technology Overview: A Report on Data Mining**. 1995.

[7] GREEVY, E.; SMEATON, A. F. **Classifying racist texts using a support vector machine**. In: *Proceedings of the 27th annual international conference on Research and development in information retrieval - SIGIR '04*, 2004.

[8] HAILONG, Z.; WENYAN, G.; BO, J. **Machine learning and lexicon based methods for sentiment classification: A survey**. In: *Proceedings - 11th Web Information System and Application Conference, WISA 2014*, 2014.

[9] HILTE, L.; LODEWYCKX, E.; VERHOEVEN, B.; DAELEMANS, W. **A Dictionary-based Approach to Racism Detection in Dutch Social Media**. 2005.

[10] HIRST, G.; LIU, B. **SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES Sentiment Analysis and Opinion Mining Sentiment Analysis and Opinion Mining**. 2012.

[11] HU, X.; TANG, J.; GAO, H.; LIU, H. **Unsupervised sentiment analysis with emotional signals**. In: *Proceedings of the 22nd international conference on World Wide Web - WWW '13*, 2013.

[12] ISTOÉ INDEPENDENTE. **O criminoso da internet**, 2015.

[13] JINDAL, N.; LIU, B. **Mining Comparative Sentences and Relations**. 2006.

[14] JOSÉ DE ALENCAR, R. **ESTUDO DA OCORRÊNCIA DE CYBERBULLYING CONTRA PROFESSORES NA REDE SOCIAL TWITTER POR MEIO DE UM ALGORITMO DE CLASSIFICAÇÃO BAYESIANO**. p. 5–1, 2012.

[15] LIU, B. **Sentiment Analysis and Opinion Mining**. 2012.

[16] MARTÍNEZ-CÁMARA, E.; MONTEJO-RÁEZ, A.; MARTÍN-VALDIVIA, M. T.; UREÑA, L. A.; UREÑA-LÓPEZ, U. **SINAI: Machine Learning and Emotion of the Crowd for Sentiment Analysis in Microblogs**. Technical report, 2013.

[17] MARTINS, I. C. **O racismo nas redes sociais : O mundo virtual é feito por pessoas de carne e osso!**, 2014.

[18] MITTAL, A.; GOEL, A. **Stock Prediction Using Twitter Sentiment Analysis**. 2012.

[19] MONARD, M. C.; BATISTA, G. E. A. P. A. **Learning with Skewed Class Distributions**. 2003.

[20] MONTEIRO E SILVA, G. **Aplicando Técnicas de Aprendizado de Máquina para Detecção Automática de Bullying no Twitter**. 2017.

[21] PANG, B.; LEE, L. **Opinion Mining and Sentiment Analysis**. *Foundations and Trends® in InformatioPang, B., & LeFoundations and Trends® in Information Retrieval, 1(2), 91–231.*, 1(2):91–231, 2006.

[22] PITSILIS, G. K.; RAMAMPIARO, H.; LANGSETH, H. **Detecting Offensive Language in Tweets Using Deep Learning**. 2018.

[23] SAIF, H.; HE, Y.; ALANI, H. **Semantic sentiment analysis of twitter**. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2012.

[24] SILVA ALVES, D. **Uso de técnicas de Computação Social para tomada de decisão de compra e venda de ações no mercado brasileiro de bolsa de valores**. 2015.

[25] SOUZA, M. **Para projeto contra preconceito, professor lista 360 termos racistas**.

[26] SPINELLI SCHIAVONI, A. **UM ESTUDO COMPARATIVO DE MÉTODOS PARA BALANCEAMENTO DO CONJUNTO DE TREINAMENTO EM APRENDIZADO DE REDES NEURAIS ARTIFICIAIS**. 2010.

[27] VICENTE, B.; DE LIMA, A.; MACHADO, V. P.; DE MELO, R.; VERAS, S. **Abordagem Semi-supervisionada para Rotula ao de Dados**.

[28] WANG, H.; CAN, D.; KAZEMZADEH, A.; BAR, F.; NARAYANAN, S. **A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle**. *Jeju, Republic of Korea*, p. 115–120, 2012.

[29] WASEEM, Z.; HOVY, D. **Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter**. p. 88–93, 2016.