CENTRO UNIVERSITÁRIO SAGRADO CORAÇÃO

KETERLY GEOVANA GOUVEIA SILVA

DETECÇÃO AUTOMÁTICA DE CONTEÚDOS PRECONCEITUOSOS UTILIZANDO TÉCNICAS DE CLASSIFICAÇÃO DE TEXTOS

BAURU

2021

KETERLY GEOVANA GOUVEIA SILVA

DETECÇÃO AUTOMÁTICA DE CONTEÚDOS PRECONCEITUOSOS UTILIZANDO TÉCNICAS DE CLASSIFICAÇÃO DE TEXTOS

Monografia de iniciação científica apresentada a Pró-reitoria de Pesquisa e Pós-Graduação como parte dos pré requisitos para aprovação do conselho, sob orientação do professor Me. Patrick Pedreira Silva

BAURU

2021

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD

Silva, Keterly Geovana Gouveia

S586d

Detecção automática de conteúdos preconceituosos utilizando técnicas de classificação de textos / Keterly Geovana Gouveia Silva. -- 2021.

32f. : il.

Orientador: Prof. M.e Patrick Pedreira Silva

Monografia (Iniciação Científica em Ciência da Computação) - Centro Universitário Sagrado Coração - UNISAGRADO - Bauru - SP

1. Classificação Automática. 2. Algoritmos. 3. PLN. 4. Inteligência Artificial. I. Silva, Patrick Pedreira. II. Título.

Elaborado por Lidyane Silva Lima - CRB-8/9602

RESUMO

Com o aumento exponencial da quantidade de informações textuais torna-se relevante investigar métodos e técnicas que permitam lidar com este conteúdo de forma eficiente e automática, dessa forma, o processamento automático de textos é um grande desafio. A classificação automática de texto envolve atribuir uma ou mais categorias de documentos predefinidas. Esta investigação se propõe a examinar as técnicas associadas à categorização automática, propondo uma ferramenta computacional que permita identificar conteúdos preconceituosos em redes sociais, a partir da análise estatística e linguística de textos coletados. O intuito deste trabalho foi investigar e desenvolver técnicas de classificação automática para realizar a tarefa de detectar discursos de ódio contra a comunidade LGBTQIA+. Foi desenvolvido um website responsivo, utilizando linguagem Python, banco de dados MySQL e Flask Framework, permitindo que usuários possam classificar seus comentários ou visualizar comentários classificados pelo modelo como preconceituoso ou não.

Palavras-chave: Classificação Automática. Algoritmos. PLN. Inteligência Artificial.

Sumário

1 INTRODUÇÃO DA PESQUISA

A internet vem aumentando exponencialmente sua quantidade de informações textuais geradas nessa última década; somente nos últimos 13 anos, o número de domínios cresceu de 15.000 para 350.000.000 (O'DELL, 2011). Esse aumento de informações também se reflete no crescimento da participação dos usuários nos conteúdos da Web. Tanta informação agrupada em um ciberespaço comum, onde são criadas centenas de milhares de novos conteúdos por dia torna interessante, porém difícil e extremamente demorada, a tarefa de processar informação. Entretanto, além de informações úteis, muito conteúdo impróprio, sobretudo, com carga preconceituosa é gerado e postado nas redes sociais, causando transtornos e incentivando essa prática criminosa.

Neste contexto, surgiu uma nova área de pesquisa relacionada ao Processamento de Linguagem Natural (PLN), chamada de Classificação Automática de Textos. As pesquisas nesta área visam extrair, de certo modo, dessa crescente quantidade de dados, dados úteis que permitam compreender o conteúdo divulgado pelos usuários de internet. Deste modo, é possível classificar textos considerando o conteúdo contido em determinado documento (PANG; LEE, 2008).

A classificação automática de textos tem sido bastante explorada em diversas tarefas: indexação automática para sistemas de recuperação de informação, organização de documentos e filtragem de textos, organização e filtragem de mensagens de e-mail, filtro de notícias, recomendação de documentos, etc. A utilização de especialistas humanos para realizar categorizações de forma manual é um processo custoso e lento e, diante desse enorme conteúdo gerado, motiva a investigação de técnicas de categorização automática de documentos. Os resultados obtidos nesta área de pesquisa são promissores e satisfatórios, o que incentiva e justifica o seu estudo e aplicação na tarefa de detecção de conteúdos preconceituosos (WEISS *et al*., 2005). As redes sociais têm sido palco de manifestações de caráter ofensivo, preconceituosos, discriminatório, de grave intolerância. Escondidas no anonimato que as redes sociais permitem com o distanciamento que

promovem, algumas pessoas se sentem à vontade para expressar todo tipo de agressão e difusão de mentiras, ferindo honra e dignidade das pessoas.

Detectar esses conteúdos impróprios e agir com rapidez representa o diferencial entre o sucesso ou o fracasso no combate ao preconceito, justificando os esforços envolvidos no desenvolvimento de ferramentas que possam processar esse tipo de informação. Deste modo, cada vez mais pessoas e principalmente empresas, estão interessadas em observar as postagens de um grupo de pessoas sobre temas que podem representar conteúdos impróprios e que devem ser eliminados de suas redes sociais.

O próprio governo federal lançou em 2015, o chamado Pacto pelo Enfrentamento às Violações de Direitos Humanos na Internet. A iniciativa, com o nome de “Humaniza Redes” prevê uma ouvidoria online, onde denúncias podem ser feitas e, em seguida, encaminhadas a provedores de internet (FOLHAPRESS, 2015).

Segundo um levantamento realizado pelo projeto Comunica que Muda,iniciativa da agência Nova/sb, são alarmantes os números referentes à intolerância do internauta brasileiro. A pesquisa aponta que plataformas como Facebook, Twitter e Instagram têm um número elevado de textos sobre temas sensíveis, como racismo, posicionamento político e homofobia. Foram identificadas 393.284 menções, sendo 84% delas com abordagem negativa, de exposição do preconceito e da discriminação (COMUNICA QUE MUDA, 2016). O que se tem visto nas redes sociais é o acirramento do discurso de ódio, de intolerância às diferenças.

Como resultado do panorama político gerado a partir das eleições de 2014, houve um intenso debate nas redes, na maioria das vezes com xingamentos e discursos rasos, que incentivam o ódio e a divisão. Do total de mensagens analisadas,com cunho político cerca de 97,4% delas abordavam aspectos negativos. A segregação virtual foi materializada no muro erguido no gramado do Congresso Nacional para separar manifestantes contra e a favor do impeachment (MATSUURA, 2016).

O segundo tema com maior número de mensagens está relacionado ao ódio às mulheres, com isso, a misoginia se alastra pelas redes sociais. Assédio, pornografia de vingança, incitação ao estupro e outras violências são, por vezes, curtidas e compartilhadas, reforçando no ambiente virtual o

machismo presente na sociedade de desigualdades de gênero (MATSUURA, 2016).

Pessoas com algum tipo de deficiência, também sofrem com o preconceito nas redes sociais, havendo em muitas situações uma abordagem negativa sobre o tema. Termos como “leproso” e “retardado mental” e o uso da deficiência para “justificar” direitos são usados nessas citações preconceituosas (MATSUURA, 2016).

O racismo também tem forte presença nas redes sociais brasileiras, fato este evidenciado recentemente com casos de ataques a celebridades negras. O levantamento também mostra que existe intolerância pela aparência, homofobia, classes sociais, idade/geração, religião e xenofobia. Este cenário mostra a necessidade de estudos que permitam debater a tênue linha que separa o discurso de ódio do direito à liberdade de expressão. O direito à liberdade de expressão não é absoluto, legislações tratam o discurso de ódio explicitamente como um limitador da liberdade de expressão (MATSUURA, 2016).

Basicamente existem dois tipos de intolerância. O visível, em que o agressor vai direto ao ponto; e o invisível, mais sutil, que se esconde em comentários que podem passar despercebido, pois abordam discursos que já foram incorporados pela sociedade, mas não pelas vítimas (MATSUURA, 2016).

Nesta visão, ao contrário do que pode parecer, o Brasil tem se mostrado como um país intolerante. As redes sociais são um reflexo dessa realidade, amplificando o ódio e reafirmando os preconceitos que as pessoas já têm (MATSUURA, 2016).

Diante desta problemática, este trabalho visa proporcionar subsídios para se estudar os métodos necessários para fazer com que postagens em redes sociais possam ser processadas e classificadas automaticamente, gerando um “entendimento” do seu conteúdo semântico, conforme a justificativa e objetivos apresentados adiante.

2 OBJETIVOS

A seguir são apresentados os objetivos que nortearam o desenvolvimento desta pesquisa

2.1 OBJETIVO GERAL

Fazer um estudo experimental da categorização de textos no formato digital, utilizando técnicas estatísticas e/ou linguísticas na tarefa de detecção de conteúdos preconceituosos.

2.2 OBJETIVOS ESPECÍFICOS

a) Realizar um levantamento bibliográfico dos métodos de classificação de textos a serem utilizados;
b) Montar um corpus linguístico[1] (coleção de documentos) para a realização do processamento desejado, utilizando posts reais publicados em redes sociais;
c) Definir e estudar de uma ferramenta de mineração de dados adequada à simulação dos métodos propostos nesta investigação;
d) Definir um modelo para classificação/identificação de comentários preconceituosos;
e) Implementar o modelo proposto;
f) Testar e interpretar os resultados obtidos a partir da aplicação modelo de classificação ao corpus montado.

[1] Corpus linguístico é um conjunto de textos escritos ou falados numa língua que serve como base de análise

3 REFERENCIAL TEÓRICO

A seguir são apresentadas teorias que fundamentaram o desenvolvimento deste projeto.

3.1 INTELIGÊNCIA ARTIFICIAL

Inteligência Artificial (IA) é o nome dado a uma das áreas mais complexas e fascinantes da Ciência da Computação. A IA pode ser definida como a capacidade de uma máquina de racionar, agir, decidir, armazenar conhecimento e comunicar-se como um ser humano (GONGORA, 2007).

Rich (1988) e Sato (2009), em definições mais práticas, definem a IA como sendo o estudo de como fazer com que as máquinas pensem e realizem tarefas em que seres humanos são melhores.

Essas definições chegam a ser simplistas e óbvias diante a complexidade e até mesmo obscuridade da área. Como é possível uma máquina tornar-se inteligente? A partir de qual momento podemos considerá-la inteligente? Qual o alcance desse termo?

Alan Turing, pai da computação e IA, fez o mesmo questionamento: “Podem as máquinas pensar?”. Conseguiu respondê-lo criando o teste de Turing, onde uma máquina, para ser considerada de fato inteligente, deveria ser submetida a um teste prático chamado “O Jogo da Imitação”, com quatro participantes: o interrogador fazendo as perguntas, um homem e uma máquina ocultos respondendo-as e um júri avaliando-as. Ao final do teste, o júri escolhe qual das respostas é a do ser humano. Caso a máquina seja escolhida, é considerada inteligente (TURING, 1950).

No ano de 2014, que marcou 60 anos da morte de Turing, um chatbot imitando um garoto de 13 anos, chamado Eugene Goostman, conseguiu passar no teste convencendo 33% dos juízes de que era humano. A prova aconteceu durante o Turing Test 2014, um evento anual organizado pela Universidade de Reading, na Inglaterra, em que máquinas e softwares tentam passar no teste de Turing. Por ser o primeiro software a ser aprovado, este passa a ser um marco na história da Inteligência Artificial (ROHR, 2014).

Um dos objetivos da IA é transformar simples computadores em máquinas cognitivas, que na sua forma mais simples, interajam com seres humanos. Se um computador cognitivo pode interagir com o ambiente, logo, poderá atuar sobre ele para melhorá-lo. Atualmente, os seres humanos fornecem respostas incompletas por não terem todas as informações. Já um sistema cognitivo, permitirá a coleta de todas as informações necessárias, incluindo o que é visto, ouvido, sentido e falado, e utilizará esses dados para fornecer respostas mais precisas aos problemas enfrentados. Uma das linhas da computação cognitiva que será abordada posteriormente é o Processamento de Linguagem Natural (PLN), responsável por tornar possível a comunicação com a máquina através da língua natural do ser humano (LEE *et al*., 2015)

3.2 PROCESSAMENTO DE LINGUAGEM NATURAL (PLN)

O Processamento de Linguagem Natural (PLN) é a subárea da IA que estuda a capacidade e as limitações de uma máquina em entender a linguagem falada pelos seres humanos no dia a dia (ROSA, 2011).

O objetivo do Processamento de Linguagem Natural é fornecer aos computadores a capacidade de entender e compor textos. E "entender" um texto significa reconhecer o contexto, fazer análise sintática, semântica, léxica e morfológica, criar resumos, extrair informação, interpretar os sentidos e até aprender conceitos com os textos processados (De JESUS *et al*., 2020).

No teste de Turing citado anteriormente, um pré-requisito para a máquina participar, é a capacidade de processar línguas naturais a fim de habilitá-la a se comunicar com sucesso na língua humana, no caso o inglês (RUSSEL; NORVIG, 2004).

O PLN está voltado a três aspectos da comunicação em língua natural:

a) som: prosódia e fonologia;
b) estrutura: morfologia e sintaxe;
c) significado: semântica e pragmática.

A prosódia está relacionada aos padrões de ritmo e entonação da língua. A fonologia está relacionada com o reconhecimento dos sons que compõem as palavras de uma língua. A morfologia estuda a construção das

palavras, com seus radicais e afixos, que correspondem a partes estáticas e variantes das palavras, como as inflexões verbais. A sintaxe define a estrutura de uma frase, com base na forma como as palavras se relacionam nessa frase. A semântica associa significado a uma estrutura sintática, em termos dos significados das palavras que a compõem. A pragmática verifica se o significado associado à uma estrutura sintática é realmente o significado mais apropriado no contexto considerado (ROSA, 2011).

3.2.1 HISTÓRICO

Graças ao surgimento dos computadores no início dos anos 40, novas frentes de pesquisa nos diversos campos do conhecimento científico tiveram avanços significativos (SILVA *et al*., 2007).

Com a necessidade de fazê-los “entenderem” instruções para realizarem tarefas, surgiram as linguagens de programação, que deram vida ao início da comunicação homem-máquina (SILVA *et al*., 2007).

No início, as linguagens eram mais próximas da linguagem da máquina e mais distantes da do ser humano. Com o passar do tempo, surgiram linguagens como a Lisp e Prolog, que se destacam por distanciarem-se da representação imposta pela arquitetura do computador. Porém, embora sejam instruções mais inteligíveis que as sequências da linguagem de máquina, ainda não são instruções em linguagem natural (SILVA *et al*., 2007).

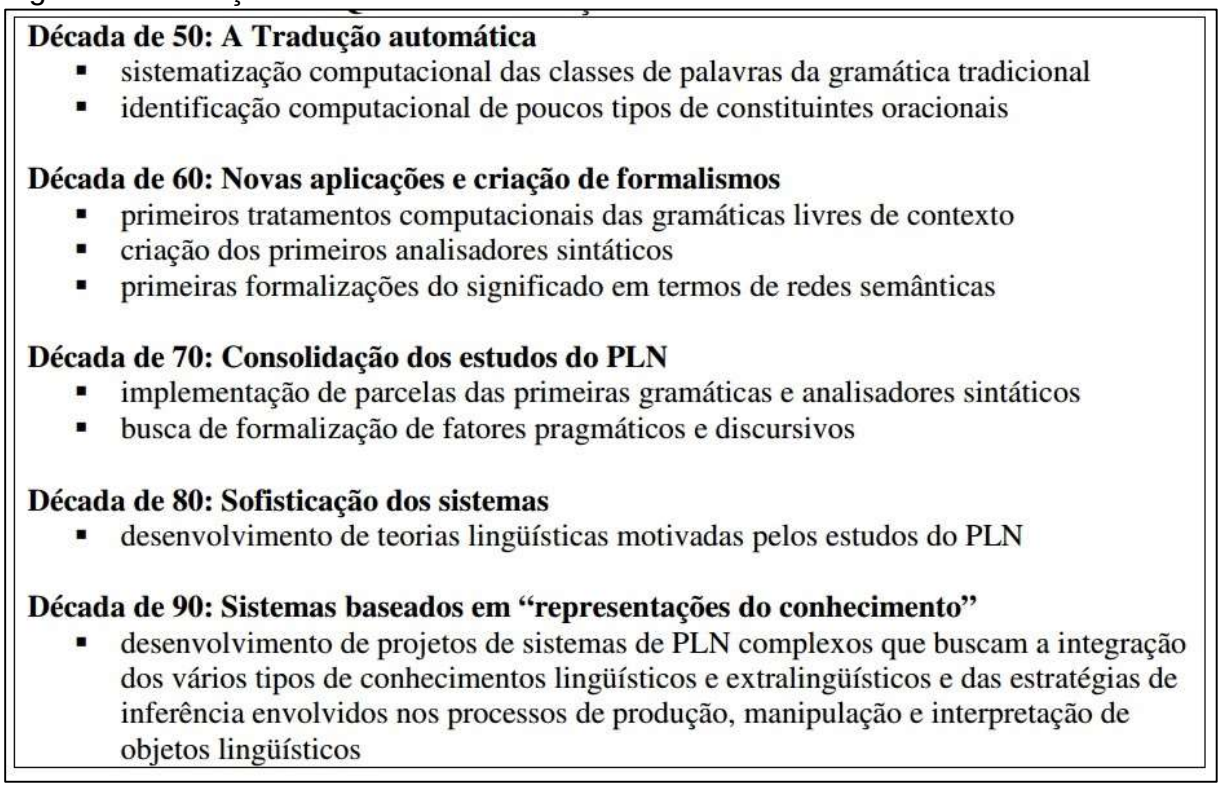
Com o objetivo de transformar computadores em instrumentos mais acessíveis, a saída foi utilizar interfaces gráficas representacionais. Um objeto gráfico significaria várias linhas de codificação na linguagem da máquina. A prova de que esta alternativa deu certo é que os computadores hoje dispõem de sofisticadas interfaces gráficas, com seus menus, ícones, janelas e cores. Esta estratégia não só resolveu o problema do contato direto com a linguagem da máquina, como também transformou os computadores em máquinas atraentes, fáceis de operar e populares, uma vez que os usuários não precisam mais utilizar comandos avançados e de difícil memorização (SILVA *et al*., 2007).

Mesmo com esse avanço no relacionamento homem-máquina, a comunicação via linguagem natural continua sendo um desafio: como criar

programas capazes de interpretar mensagens codificadas em linguagem natural e decifrá-las para a linguagem de máquina?

Com o passar dos anos, houve muitas pesquisas e desenvolvimentos nos mais diversos ramos do processamento de linguagem natural, destacando-se a tradução automática, considerada pela maioria como o marco inicial na utilização dos computadores para o estudo das línguas naturais (SILVA *et al*., 2007). A evolução do PLN é ilustrada da Figura 1.

Figura 1 - Evolução do PLN.



Década de 50: A Tradução automática
- sistematização computacional das classes de palavras da gramática tradicional
- identificação computacional de poucos tipos de constituintes oracionais

Década de 60: Novas aplicações e criação de formalismos
- primeiros tratamentos computacionais das gramáticas livres de contexto
- criação dos primeiros analisadores sintáticos
- primeiras formalizações do significado em termos de redes semânticas

Década de 70: Consolidação dos estudos do PLN
- implementação de parcelas das primeiras gramáticas e analisadores sintáticos
- busca de formalização de fatores pragmáticos e discursivos

Década de 80: Sofisticação dos sistemas
- desenvolvimento de teorias lingüísticas motivadas pelos estudos do PLN

Década de 90: Sistemas baseados em "representações do conhecimento"
- desenvolvimento de projetos de sistemas de PLN complexos que buscam a integração dos vários tipos de conhecimentos lingüísticos e extralingüísticos e das estratégias de inferência envolvidos nos processos de produção, manipulação e interpretação de objetos lingüísticos

Fonte: Silva *et al*. (2007, p. 8).
Nota: Adaptada pelo autor.

De modo geral, no PLN, buscam-se soluções para questões computacionais que requerem o tratamento computacional de uma ou mais línguas naturais, quer sejam escritas ou faladas. Mais precisamente, o PLN dedica-se a propor e desenvolver sistemas computacionais que têm a língua natural escrita como objeto primário. Para tanto, linguistas e cientistas da computação, buscam fundamentos em várias disciplinas: Filosofia da Linguagem, Psicologia, Lógica, Inteligência Artificial, Matemática, Ciência da Computação, Linguística Computacional e Linguística (SILVA *et al*., 2007).

Em PLN, os linguistas trabalham em duas frentes: utilizam o computador para desenvolver e validar teorias e dados linguísticos, e fornecem o

conhecimento necessário para o desenvolvimento de sistemas especializados. Os cientistas da computação, por sua vez, implementam ferramentas para desenvolvimento e validação dessas teorias, originando os Sistemas de Processamento de Linguagem Natural, mais detalhadamente abordados no tópico 4.3 (SILVA *et al*., 2007).

3.2.2 APLICAÇÕES

O atual estágio de desenvolvimento do PLN já permite sua aplicação em áreas como: Extração de Informação, Recuperação de Informação, Tradução Automática, Geração Automática de Texto, Geração de Linguagem Natural, Simplificação de Texto, Correção ortográfica e a que diz respeito a esta pesquisa, Interpretação de Linguagem Natural. A seguir apresenta-se uma breve explanação sobre cada uma:

a) Extração de Informação (EI): Processo no qual informações são apresentadas pela relevância da busca. A EI parte do princípio de que algumas páginas da Web que tratam de assuntos mais específicos tendem a apresentar regularidade quanto à formatação, estrutura e conteúdo podendo ser agrupadas formando classes de páginas, por exemplo, páginas de cinema, classificados ou eventos científicos. A EI extrai informações relevantes podendo tanto classificar uma página segundo um contexto de domínio como também extrair informações relevantes a este contexto estruturando as informações contidas na página e armazenando-as em bases de dados (SILVA, 2003);
b) Recuperação de Informação (RI): RI é a área da Ciência da Computação que permite o acesso fácil e rápido a informações. “A RI trata da representação, armazenamento, organização e acesso a tens de informação, de forma organizada e eficiente” (SILVA, 2003);
c) Tradução Automática (TA): Consiste no processo de transposição de palavras entre idiomas naturais, através da utilização de programas de leitura e interpretação de textos. A Tradução Automática (TA) é um dos domínios da Linguística computacional

(LC) que mais envolve conhecimento linguístico, por codificar informações de uma língua para outra(SILVA *et al*., 2007);

d) Geração Automática de Texto: Consiste no processamento de informações já existentes, mas dessa vez de modo mais estruturado e com possibilidade de adequação/manipulação para melhor armazenamento/utilização (PARDO, 2008);
e) Geração de Linguagem Natural: Área ainda iniciante na Ciência da Computação, procura construir sistemas programados de linguagem que se aproxime da Linguagem Natural humana. Tem como foco a interpretação de perguntas feitas em linguagem natural dentro de sistemas de apoio à decisão, reconhecendo estruturas semântica se transformando-as em consultas que retornam resultados relativos à questão elaborada pelo analista(SILVA *et al*., 2007);
f) Interpretação de Linguagem Natural: Consiste no processo através do qual o ser humano interage linguisticamente com a máquina, a qual é provida de programação que lhe permite interpretar enunciados dúbios sem ruídos que possam prejudicar a comunicação. O processamento de Linguagem Natural vem a facilitar a interação do software (através de sua interface) com o usuário, para que se torne mais fácil a comunicação e a passagem de conhecimento, assim quem fizer o uso de um software, possa compreender o que ele tem a oferecer e consiga saber o que o usuário está necessitando. Utilizando a linguagem natural torna-se mais simples o questionamento de uma determinada área, já que não há necessidade de se saber corretamente a implementação do sistema, o que ele irá buscar, como por exemplo, em uma consulta a um banco de dados, o usuário não precisa saber o que são tabelas e nem como elas buscam as informações, e nem o funcionamento de um banco de dados, ele apenas deseja que o resultado da pesquisa seja mostrado de forma simples e objetiva(SILVA *et al*., 2007);
g) Simplificação de Texto: Ou sumarização é a área de processamento em PLN que elabora/gera resumos (sumários) a

partir de textos completos. São sistematizações feitas a partir da extração de palavras chaves identificadas no texto original (PARDO, 2008);

h) Correção Ortográfica: Sistemas de correção ortográfica (do inglês, *spellingchecker systems*) processam um texto em uma dada língua natural com os objetivos de identificar os erros cometidos quanto à ortografia (palavras que não constam do léxico dessa língua ou usadas em contexto impróprio) e sugerir alternativas prováveis e ortograficamente corretas a cada erro identificado (FELIPPO; SILVA, 2008).

3.3 MINERAÇÃO DE DADOS

De forma geral, a Mineração de Dados ou Data Mining pode ser conceituada como a descoberta e análise inteligente de informações úteis da Web (COOLEY, 1997).

Todos os tipos de textos que compõem o dia a dia de organizações e pessoas são produzidos e armazenados em meios digitais. Além de todos os conteúdos produzidos profissionalmente por empresas, os usuários passaram a compartilhar na web seus conhecimentos, críticas, opiniões e vincular esses conteúdos a sites, blogs, redes sociais, fóruns, bate papos, dentre outros. Desta forma, usuários tendem a postar seus comentários sobre pessoas, organizações, serviços, produtos e marcas, alimentando ainda mais esse vasto banco de informações da World Wide Web (GUEDES; AFONSO; MAGALHÃES, 2010).

Sendo assim, torna-se evidente a dificuldade em filtrar e tirar informações relevantes dessa imensa massa de dados. Para usuários finais, é de grande interesse saber quais as demais opiniões sobre um produto que deseja adquirir. Para organizações, saber o que os consumidores pensam sobre sua marca e produtos é uma grande vantagem competitiva. Em virtude desse grande volume de dados eletrônicos disponíveis na internet e a necessidade de obter informações relevantes a partir deles, torna-se necessário o uso de técnicas de extração de conhecimento automáticas e eficientes, com o objetivo de recuperar e minerar conhecimentos úteis da web e

apresentá-los ao usuário em uma leitura objetiva e conclusiva, facilitando a interpretação e tomada de decisão (GUEDES; AFONSO; MAGALHÃES, 2010).

Existem três frentes que categorizam a Mineração de Dados na Web, a Mineração de Uso, a Mineração de Estrutura e a Mineração de Conteúdo. A mineração de uso aborda a mineração das informações de uso da Web, são as informações sobre como o usuário interage com a Web. Nessa categoria são tratadas questões como personalização, interfaces adaptativas e aprendizado de perfis de usuários. A mineração de estrutura aborda a mineração das informações contidas entre os documentos da Web. Os documentos da Web se relacionam basicamente através de vínculos de hipertexto, e esses vínculos escondem informações valiosas não só sobre a topologia da Web, mas também sobre como os documentos se relacionam. A mineração de conteúdo aborda a mineração dos dados contidos dentro dos documentos da Web. A grande quantidade de formatos que os dados podem assumir (textos comuns, páginas HTML, imagens, áudio, vídeo, etc.) acaba dirigindo as técnicas de mineração a serem utilizadas. (MARINHO; GIRARDI, 2005).

Esta última se estende para a Mineração de Opinião ou Opinion Mining, responsável por minerar e classificar opiniões, assunto do próximo tópico e de grande importância para esta pesquisa.

3.4 MINERAÇÃO DE OPINIÃO

A mineração de opinião, ou Opinion Mining, também conhecida como análise de sentimentos, pode ser definida como a técnica que avalia um conteúdo subjetivo emitido em linguagem natural e descobre o sentimento que é transmitido. Geralmente associado à classificação binária entre sentimentos positivos e negativos, o termo é usado de uma forma mais abrangente para significar o tratamento computacional de opinião, sentimento e subjetividade em textos (PANG; LEE, 2002). Com o advento da web como fonte de informações, grande parte dos usuários tem buscado nela textos que forneçam esse tipo de informação desejada, opiniões sobre alguma entidade de interesse como um produto específico, uma empresa, um lugar, uma pessoa, dentre outros. O objetivo principal é permitir que um usuário obtenha uma visão geral sobre o que outros pensam sobre algo em particular, sem precisar localizar e

ler cada opinião feita na web. Para atingir esse objetivo, a mineração de opinião é dividida em três grandes etapas: coleta de conteúdo, classificação e sumarização dos resultados.

3.4.1 COLETA DE CONTEÚDO

Etapa na qual é feita uma busca em fontes diversas, tais como artigos em sites, comentários em mídias sociais, anúncios, documentos dentre outras. É importante a utilização de técnicas avançadas de busca, visando identificar se o conteúdo encontrado trata-se de uma opinião ou um fato. Fatos por si só devem ser descartados, porém opiniões expressas em fatos devem ser mantidas (BECKER; TUMITAN, 2005).

3.4.2 CLASSIFICAÇÃO

A classificação é a etapa mais importante do processo e é nela que a polaridade ou orientação da opinião é definida. Esta etapa determina se uma opinião é positiva, negativa ou neutra. Na classificação ou análise de sentimentos, são as palavras opinativas que têm a maior importância, pois, através delas, é possível determinar o sentimento expresso pelo autor. Exemplos de palavras opinativas: bom, legal, ótimo, ruim, péssimo etc. (BECKER; TUMITAN, 2005).

3.4.3 SUMARIZAÇÃO DOS RESULTADOS

Etapa focada na apresentação dos resultados, que podem ser de forma textual, ou gráfica. A melhor forma de representar os resultados é a gráfica, pois facilita a visualização e entendimento dos resultados sumarizados em totais e dados estatísticos (BECKER; TUMITAN, 2005).

4 TRABALHOS CORRELATOS

Embora as áreas de Processamento de Linguagem Natural e Mineração de Opinião sejam recentes, existem muitas pesquisas e informações disponíveis. A cada dia surgem novos artigos científicos, novas metodologias para alcançar os resultados, novas ferramentas de classificação de sentimentos, novos bancos de dados para utilização, novos desafios e necessidades. Isso prova que as áreas estão aquecidas e em crescente expansão.

Os trabalhos seguintes, apesar de não lidarem diretamente com a temática proposta nesta pesquisa, trazem métodos, insights e materiais que poderão contribuir para esta investigação de iniciação científica.

Uma das pesquisas analisadas foi a intitulada: “Protótipo para Mineração de Opinião em Redes Sociais: Estudo de Casos selecionados usando o Twitter”, de autoria de Leandro Matioli Santos, que aborda a mineração da opinião nas redes sociais, estudando se é possível aplicar a técnica nesse tipo de mídia, quais os desafios e dificuldades, se os resultados são satisfatórios e relevantes (SANTOS, 2010)

Outra pesquisa analisada foi a intitulada: “Mineração de Opiniões aplicada a mídias sociais”, de Marlo Vieira dos Santos e Souza, que analisa o cenário das empresas nas mídias sociais, a inteligência competitiva, o mercado capitalista e suas tendências nessas mídias, opiniões sobre produtos, marcas, entidades, etc. (SOUZA, 2012).

Dessa forma, a intenção deste trabalho é contribuir com informações, conclusões e análises relevantes para essas áreas, direcionando para a exploração de opiniões nas mídias sociais envolvendo a questão da identificação do preconceito.

5 METODOLOGIA

Este projeto é uma pesquisa na área de Processamento de Linguagem Natural, objetivando classificar comentários com potencial conteúdo preconceituoso, criando uma interface web para que textos digitais sejam classificados automaticamente, aumentando a base de dados inicial do projeto e melhorando a precisão do algoritmo.

O projeto foi desenvolvido em duas etapas: uma fase de embasamento teórico e uma fase prática de implementação.

Na primeira etapa que corresponde à fase teórica da pesquisa, foi realizado um levantamento bibliográfico sobre temas relacionados ao projeto, além do estudo de outros conteúdos fundamentais para o andamento da pesquisa. Levantou-se material sobre:

1. Inteligência Artificial
2. Processamento de Linguagem Natural
3. Machine Learning
4. Linguagem de Programação Python
5. Documentação Twitter API
6. Bibliotecas Python: Pandas, Pymysql, Tweepy, Sqlalchemy e Sklearn
7. Flask Framework

Este estudo inicial possibilitou o andamento do projeto e o embasamento necessário para a etapa prática da pesquisa.

Na segunda etapa, que resultou em uma fase mais prática,foi feita uma seleção dos textos digitais para compor a base de dados de treinamento do algoritmo. Ao todo 500 comentários foram escolhidos e classificados manualmente, permitindo a programação das funções de classificação.

Os dados foram obtidos de algumas formas: Na plataforma do Youtube foi feita uma coleta manual, lendo comentários em vídeos e assistindo a Youtubers em um tipo de vídeo com o objetivo de reagir aos comentários do próprio canal. Neste caso, o aplicativo Google Leans permitiu que, ao apontar a câmera para a imagem ou vídeo, o texto fosse extraído. Além dessa extração

de palavras, foi possível copiar o texto para um dispositivo em específico, agilizando ainda mais o processo.

Essa mesma ferramenta foi utilizada quando a busca por textos foi feita no Google Imagens, assim como em sites que noticiavam cenas de preconceito e no próprio Twitter, onde muitas pessoas publicam prints em um *exposed* (atualmente, muitos usuários em redes sociais expõem outros usuários com prints ou informações a respeito de algum ato que gera indignação e repúdio).

A rede social Twitter teve a coleta por meio da API e de forma manual quando os resultados trazidos pela aplicação não eram satisfatórios. Quando usada a linguagem Python para a coleta desses dados, os scripts foram desenvolvidos utilizando Jupyter Notebook em arquivos de Python 3.

Além de selecionar, foi necessário armazenar os dados. Para isso, foi criado um banco de dados simples, contendo 3 campos principais: id, comentário e emoção. Este banco de dados serviu como treinamento para o modelo de classificação. Por este ser o objetivo, além de comentários preconceituosos, foram armazenados comentários que não possuíam discurso de ódio. Por se tratar de uma base de treinamento, os dados foram classificados manualmente no campo “emoção” com os valores “preconceituoso” e “não preconceituoso”. Sendo assim, os dados coletados manualmente e pelo Google Leans foram inseridos somente no campo comentário, e os dados obtidos pela API foram inseridos através da biblioteca pymysql também no campo comentários.

Ao montar a base de dados, foi possível utilizá-la para identificar a melhor biblioteca para classificar automaticamente os textos digitais inseridos como teste. Inicialmente, com base nos estudos de cursos, o código em Python foi desenvolvido com a biblioteca NLTK e o algoritmo foi treinado. Posteriormente, acompanhando um live no Youtube, o mesmo teste foi feito utilizando a biblioteca NLTK junto a Sklearn. O objetivo desta etapa era a avaliação das bibliotecas para a escolha, e não o desenvolvimento do programa efetivamente. Após as comparações, a biblioteca NLTK foi escolhida, tendo seu algoritmo re-programado, focado em testes de classificação e na análise dos resultados.

Depois da programação do algoritmo pelo Jupyter Notebook, foi iniciada a fase de criação do produto web, dividida em duas etapas de desenvolvimento: front end e back end.

O desenvolvimento front end englobou a criação do protótipo da interface através da ferramenta Figma, sendo uma base para a programação, mas possibilitando alterações futuras. Neste protótipo decidiu-se pela criação da página principal, página de classificações corretas e incorretas, pagina de classificar e a página de estatísticas do algoritmo.

Para a utilização da linguagem Python na web, o framework Flask foi escolhido e com isso deu-se início a programação com as seguintes tecnologias:

1) HTML
2) CSS
3) Bootstrap
4) Chart.js
5) MySQL

Utilizando essas tecnologias algumas rotas foram criadas e os templates desenvolvidos, inserindo informações sobre o projeto, tabela de resultados, formulário para a interação com o site e gráficos com estatísticas de erros e acertos do algoritmo, além de uma visão da quantidade de comentários obtidos com este site, e que permitiu o enriquecimento da base de dados inicial. Foi desenvolvida também uma página não visível no menu, onde o usuário valida se a classificação está ou não correta.

No desenvolvimento back-end com a linguagem Python, foi necessário analisar a adaptação do algoritmo de classificação. Por ter sido programado anteriormente utilizando Jupyter Notebook, houve uma mudança com adição de novas funções, principalmente para a interação com o banco de dados. Foi necessário também criar três novas tabelas, contendo os campos id, comentário e emoção. Uma tabela foi responsável pelo armazenamento dos comentários corretos, e as outras, dos comentários incorretos. Programaram-se funções para selecionar todos os comentários das tabelas no banco de dados além da inserção de novos comentários.

A programação backend controlou toda a lógica do site, interferindo nas tabelas de comentário, na classificação do comentário do usuário e na

alimentação dos gráficos. Os resultados (telas) do protótipo desenvolvido serão descritos na seção seguinte.

6 RESULTADOS

A seguir serão apresentados os resultados obtidos na pesquisa

6.1 ALGORITMO DE CLASSIFICAÇÃO COM NLTK

O algoritmo implementado (Figura 2) utilizou a biblioteca NLTK para a classificação dos comentários. Esta classificação trabalhou com diversas funções, sendo utilizadas para a limpeza dos dados, obtenção da frequência das palavras, extração das palavras únicas e a classificação propriamente dita. Por trabalhar com uma tabela de probabilidade, o algoritmo indicou a porcentagem em relação as classes “preconceituoso” e “não preconceituoso”. Na programação web essas porcentagens foram utilizadas em uma lógica de comparação, onde, ao invés de exibir em porcentagem para o usuário, o site faz uma exibição utilizando palavras, melhorando o entendimento do leitor.

Figura 2 – Script de classificação com NLTK

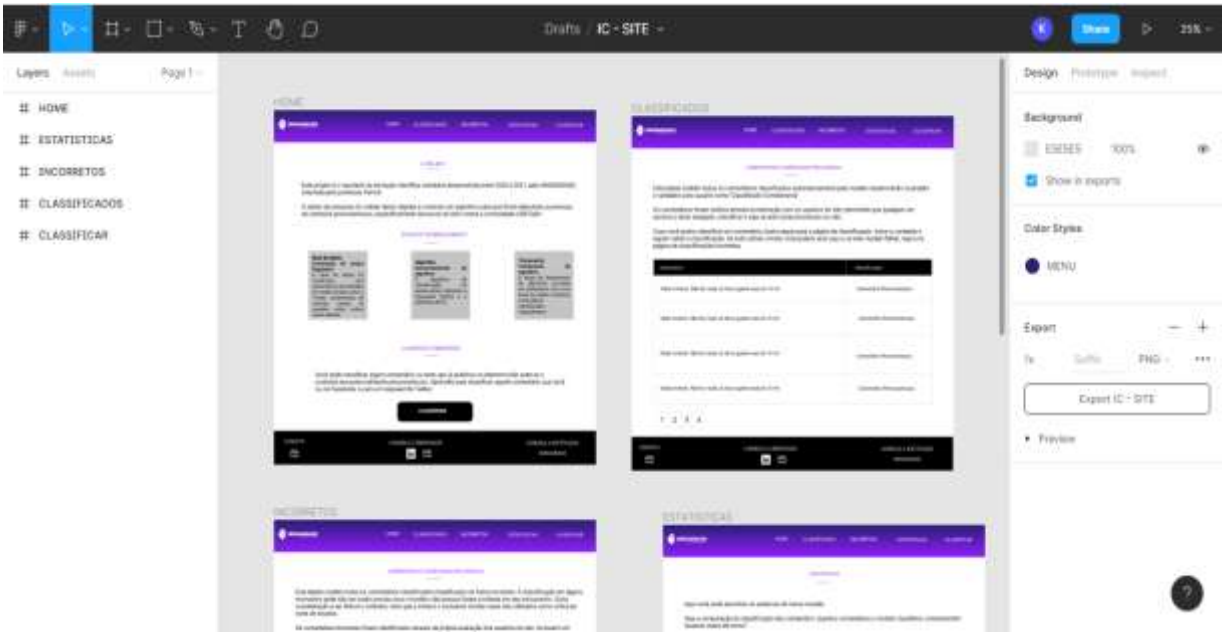
```
classificador = NaiveBayesClassifier.train(base_completa)

teste = comentario

palavraTeste = []
for palavrasTeste in teste.split():
        testeStm = [p for p in palavrasTeste.split()]
        palavraTeste.append(str(stemmer.stem(testeStm[0])))



novaTeste = extra_palavras(palavraTeste)

#classificando com probabilidade
classificador.classify(novaTeste)
dist = classificador.prob_classify(novaTeste)


def resultTeste():
        resultComentario = []
        for classe in dist.samples():
                resultComentario.append("%.2f" % (dist.prob(classe)*100))
        #novoComentario(teste, resultComentario[0], resultComentario[1])
        print(resultComentario)
        return resultComentario


return resultTeste()
```

Fonte: Elaborada pelo autor

6.2 DESIGN DE INTERFACE

A ferramenta Figma (Figura 3) foi utilizada na prototipagem da interface web. Com ela foi possível testar cores, espaçamento, margem e a disposição dos elementos. O protótipo foi a base para o desenvolvimento front end, mas não impossibilitou mudanças visuais conforme a necessidade.

Figura 3 – Design da interface utilizando Figma



Fonte: Elaborada pelo autor

6.3 TELAS DO SITE DE CLASSIFICAÇÃO

As imagens seguintes apresentam as páginas do site desenvolvido utilizando Flask com a linguagem Python, HTML, CSS, Bootstrap, Chart.js e o banco de dados MySQL.

A tela “home” (Figura 4 e Figura 5) exibe todas as informações gerais do projeto, descrevendo brevemente o objetivo, o orientador e as etapas da pesquisa, além de fazer uma chamada para que o usuário classifique um conteúdo.

Figura 4 – Página HOME com descrição do projeto

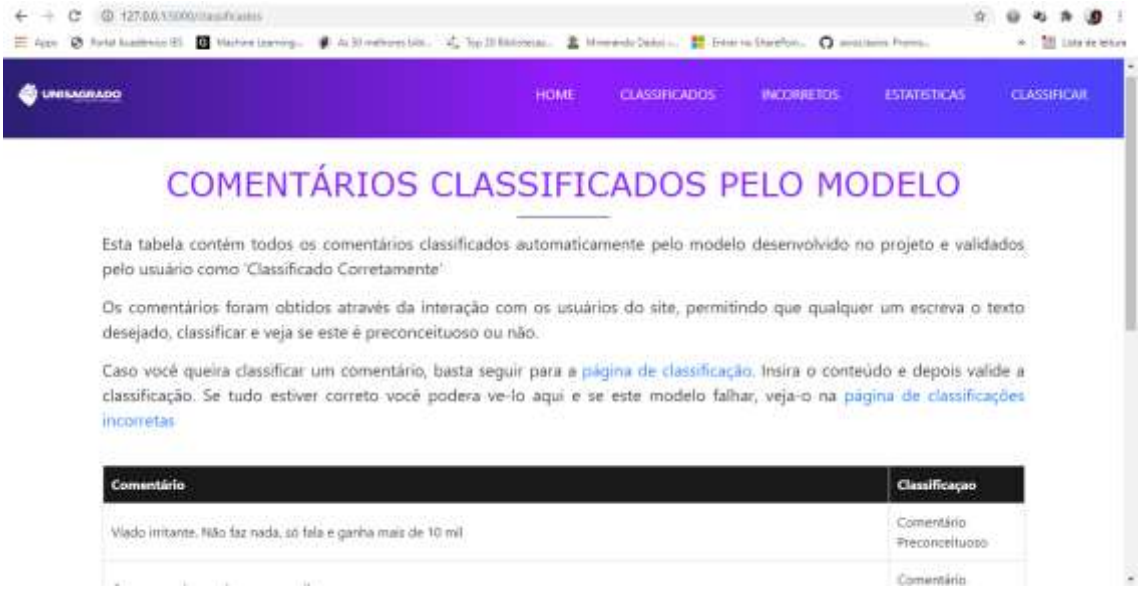
Fonte: Elaborada pelo autor

Figura 5– Página HOME com navegação para a página classificados

Fonte: Elaborada pelo autor

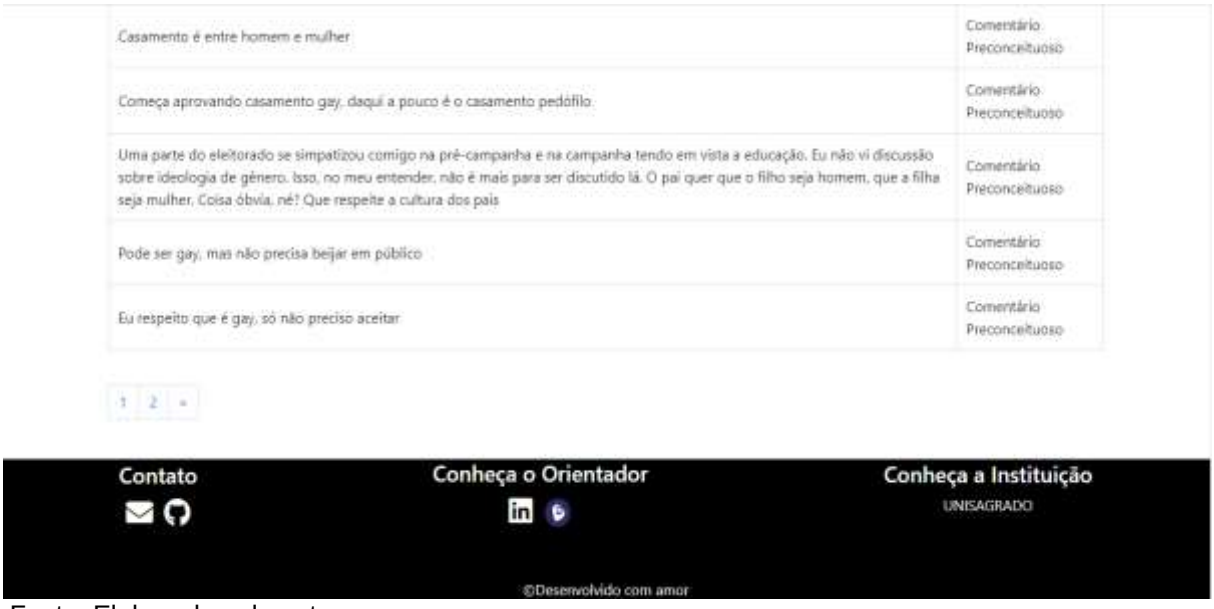
A tela “classificados” (Figura 6 e Figura 7) exibe em uma tabela todos os comentários classificados corretamente pelo modelo de acordo com a validação do próprio usuário. Além disso, ela traz uma parte informativa.

Figura 6 – Página CLASSIFICADOS



Fonte: Elaborada pelo autor

Figura 7 – Página CLASSIFICADOS com paginação de resultados



Fonte: Elaborada pelo autor

A tela “incorretos” (Figura 8) exibe em uma tabela todos os comentários classificados incorretamente pelo modelo, de acordo com a validação do próprio usuário. Além disso, ela traz uma parte informativa.

Figura 8 – Página INCORRETOS

Fonte: Elaborada pelo autor

A tela “classifique” (Figura 9) exibe um campo para que o usuário insira um conteúdo e veja sua classificação. Ao clicar para classificar, ele é redirecionado para a página de validação.

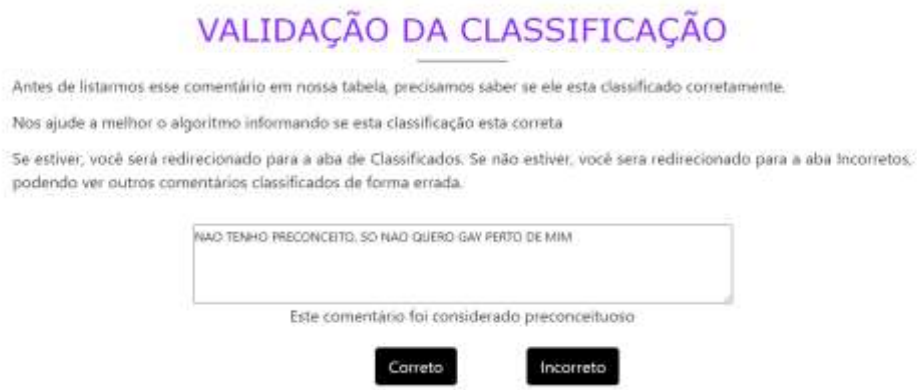
Figura 9 – Página CLASSIFICAR

Fonte: Elaborada pelo autor

A tela “validação” (Figura 10) é a página onde o usuário informa se a classificação dada pelo modelo está correta. Com essa validação ele é redirecionado para a página “classificados” caso o modelo tenha acertado, ou para a página “incorretos”. Se incorreto o comentário é salvo na base de dados de comentários incorretos e na base de dados teste, onde sua classificação é invertida para que ele seja utilizado nos treinamentos. Se correto, ele é salvo

somente na base de dados de comentários corretos, mas também é usado no treinamento.

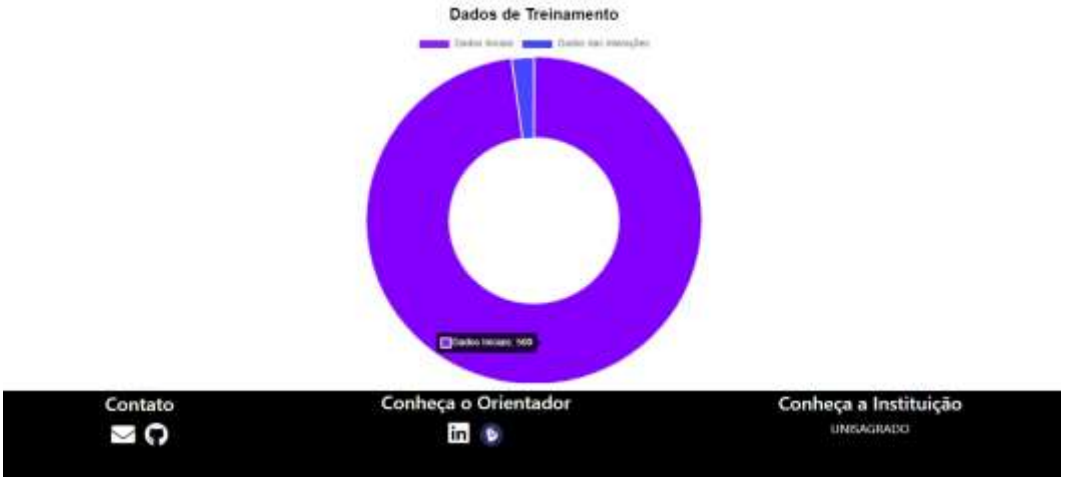
Figura 10 – Página de VALIDAÇÃO



Fonte: Elaborada pelo autor

A tela “estatística” (Figura 11) exibe dois gráficos alimentados dinamicamente usando o banco de dados. Um gráfico mostra a estatística de erros e acertos do modelo e o outro faz uma comparação entre a quantidade de dados iniciais e os novos dados obtidos com a interação dos próprios usuários.

Figura 11 – Página de ESTATÍSTICAS com gráfico de treinamento



Fonte: Elaborada pelo autor

7 CONSIDERAÇÕES FINAIS

O site necessita ser disponibilizado para mais usuários. Uma alternativa é a utilização de servidores gratuitos com o objetivo de dar continuidade a melhorias em sua interface e alterações no algoritmo se necessário.

O desenvolvimento deste projeto de Iniciação Científica trouxe, além do conhecimento nas tecnologias e ferramentas utilizadas, uma reflexão sobre o comportamento humano no mundo digital e como comentários, principalmente em redes sociais, podem ser carregados de ódio e preconceito.

A pesquisa contribuiu para o enriquecimento da lógica de programação, aprendizado sobre a linguagem Python e suas bibliotecas, além da prática de tecnologias já conhecidas. A Iniciação Cientifica também foi uma oportunidade para ter mais contato com uma área crescente da Computação, além de ser um ponto inicial para uma trajetória no mundo da pesquisa.

REFERÊNCIAS

BECKER, K.; TUMITAN, D. Introdução à Mineração de Opiniões: Conceitos, Aplicações e Desafios. UFRGS, 2005. Disponível em: <http://www.inf.ufrgs.br/~kbecker/lib/exe/fetch.php?media=minicursosbbd_versaosubmetida.pdf>. Acesso em: 10 janeiro 2018.

COMUNICA QUE MUDA, 2016. Nova/sb identifica quadro de intolerância no Brasil. Disponível em: <http://www.clubedecriacao.com.br/ultimas/comunica-que-muda-2/>. Acesso em: 10dez. 2017.

COOLEY, R. Web mining: information and pattern Discovery on the World Wide Web, Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence. IEEEXplore, 1997. Disponível em: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=632228>. Acesso em: 10 dez 2017.

FOLHAPRESS. Governo Federal Lança Ação Para Defesa de Direitos Humanos na Internet. Jornal do Comércio, 2015. Disponível em: <https://www.jornaldocomercio.com/site/noticia.php?codn=192894>. Acesso em: 25 jun 2021.

GONGORA, A. D. O que é Inteligência Artificial? UFSC, 2007. Disponível em: <http://www.egov.ufsc.br/portal/sites/default/files/anexos/6515-6514-1-PB.pdf>. Acesso em: 25 fev. 2018.

GUEDES, R.; AFONSO, D.; MAGALHÃES, L. H. de. Mineração de opiniões de usuários na busca de conhecimento. Vianna Sapiens, 2010. Disponível em: <http://www.viannajunior.edu.br/files/uploads/20131001_141137.pdf>. Acesso em: 4 dez 2017.

JESUS, A. P. F. de. *et al*. Robô de Conversação Baseado em Inteligência Artificial para Treinamento na Lei Geral de Proteção de Dados Pessoais. UNISANTA, 2020. Disponível em: <https://periodicos.unisanta.br/index.php/sat/article/view/2657>. Acesso em: 30 jun 2021.

LEE, Y. M. T. C. *et al*. O Desenvolvimento da Computação Cognitiva. ENEGEP, 2015. Disponível em: <http://www.abepro.org.br/biblioteca/TN_STO_213_261_27007.pdf>. Acesso em 18 jun 2021.

MARINHO, L. B.; GIRARDI, R. Mineração na Web. Research Gate, 2005. Disponível em: <http://www.researchgate.net/profile/Leandro_Marinho/publication/228369738_Minerao_na_Web/links/0deec539852e70896c000000.pdf>. Acesso em: 10 dez 2017.

MATSUURA, S. Brasil Cultiva Discurso de Ódio nas Redes Sociais, Mostra Pesquisa. CEERT, 2016. Disponível em:

<https://ceert.org.br/noticias/comunicacao-midia-internet/12814/brasil-cultiva-discurso-de-odio-nas-redes-sociais-mostra-pesquisa>. Acesso em: 30 jun 2021.

O'DELL, J. How Big Is the Web & How Fast Is It Growing? Disponível em: <https://mashable.com/2011/06/19/how-many-websites/>. Acesso em: 5 mar. 2013.

PANG, B.; LEE, L.; VAITHYANATHAN, S. Thumbs up? Sentiment Classification using Machine Learning Techniques. ACM DIGITAL LIBRARY, 2002. Disponível em: < http://dl.acm.org/citation.cfm?id=1118704>. Acesso em: 04 dez. 2017.

PANG, Bo;LEE, Lilian. Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval 2(1-2), pp. 1–135, June 2008.

PUBLICO. Comentários racistas publicados no Facebook divulgados à porta de quem os escreveu. 2015. Disponível em: <https://www.publico.pt/2015/12/01/tecnologia/noticia/comentarios-racistas-publicados-no-facebook-divulgados-a-porta-de-quem-os-escreveu-1716118>. Acesso em: 05 mar. 2018.

RICH, E. Inteligência Artificial. São Paulo: McGRALL-HILL, 1988.

ROHR, A. Computador convence juízes de que é garoto de 13 anos em 'teste de Turing'. Globo.com, 2014. Disponível em: <http://g1.globo.com/tecnologia/noticia/2014/06/computador-convence-juizes-que-e-garoto-de-13-anos-em-teste-de-turing.html>. Acesso em: 26 jan. 2018.

ROSA, J. L. G. Fundamentos da Inteligência Artificial. Rio de Janeiro: Gen-LTC, 2011.

RUSSEL, S. J.; NORVIG, P. Inteligência Artificial. Rio de Janeiro: Elsevier, 2004.

SANTOS, L. M. Protótipo para Mineração de Opinião em Redes Sociais: Estudo de Casos selecionados usando o Twitter. UFLA, 2010. Disponível em: <http://www.bcc.ufla.br/wp-content/uploads/2013/2010/LeandroMatioli.pdf>. Acesso em: 26 Jan. 2018.

SATO, P. O que é Inteligência Artificial? Revista Escola, 2009. Disponível em: <http://revistaescola.abril.com.br/ciencias/fundamentos/inteligencia-artificial-onde-ela-aplicada-476528.shtml>. Acesso em: 26 abr. 2015.

SILVA, B. C. D. da. *et al*. Introdução ao Processamento das Línguas Naturais e Algumas Aplicações. 2007. 119 f. Série de Relatórios do Núcleo Interinstitucional de Lingüística Computacional, São Carlos, 2007. Disponível em: <http://www.letras.etc.br/ebralc/NILCTR0710-DiasDaSilvaEtAl.pdf>. Acesso em: 14 janeiro 2018.

SOUZA, M. V. dos. S. Mineração de Opiniões aplicada a mídias sociais. PUCRS, 2012. Disponível em: <http://meriva.pucrs.br:8080/dspace/bitstream/10923/1457/1/000448645-Texto%2bCompleto-0.pdf>. Acesso em: 20 dez. 2017.

TURING, A. M. Computing machinery and intelligence. UMBC, 2002. Disponível em: <http://www.csee.umbc.edu/courses/471/papers/turing.pdf>. Acesso em: 26 abr. 2015.

WEISS, S.M.; ZHANG, T.; DAMERAU, F. Text Mining: Predictive Methods for Analyzing Unstructured Information. Springer Editora, Edição 1, 2005.