

CARLOS EDUARDO GALVÃO

PEDRO HENRIQUE ARAÚJO MARQUES DA SILVA

THOMAS HENRIQUE DE SOUZA SANTOS

**Sankofa.AI: Analisando o Racismo Digital com Inteligência
Artificial**

CAMPINAS
2025

IFSP - INSTITUTO FEDERAL DE SÃO PAULO - CÂMPUS CAMPINAS

CURSO TÉCNICO INTEGRADO AO MÉDIO EM INFORMÁTICA

Sankofa.AI: Analisando o Racismo Digital com Inteligência Artificial

Orientador: Prof. Márcio André Miranda
Co orientador: Tatiane Salles

CAMPINAS
2025

RESUMO

Este projeto inovador propõe uma solução tecnológica robusta para combater a injúria racial nas redes sociais, um desafio crescente que afeta a segurança e bem-estar de grupos minorizados no ambiente digital. Nosso objetivo é desenvolver um software avançado, utilizando a linguagem Python e técnicas de Inteligência Artificial, para identificar, analisar e organizar discursos de cunho racista. A metodologia envolverá o processamento e a classificação de dados extraídos de plataformas digitais, com o armazenamento das informações em um banco de dados dedicado. Essa base de dados permitirá a geração de relatórios detalhados e visualizações sobre padrões e recorrências dos discursos de ódio, fornecendo um recurso valioso. Esperamos criar uma base de dados acessível que não só subsidiará pesquisas acadêmicas e iniciativas antirracistas, mas também servirá como um recurso educacional aberto para promover maior conscientização. A ferramenta auxiliará na formulação de políticas públicas eficazes no combate ao racismo virtual, alinhando-se diretamente com o Objetivo de Desenvolvimento Sustentável 10 da ONU, que visa a redução das desigualdades. Assim, contribuímos para a construção de um ambiente digital mais equitativo e seguro para todos.

Palavras Chaves: Racismo digital; Inteligência artificial; Estudos sociais.

ABSTRACT

This project addresses the issue of racial slurs on social media, a growing phenomenon that directly impacts the safety of marginalized groups in digital environments. The primary objective is to develop software capable of identifying, analyzing, and organizing racially motivated discourse, with a focus on supporting research and initiatives to combat online racism. The methodology involves the use of the Python programming language and Artificial Intelligence techniques for the processing and classification of data extracted from social media platforms. A database will be implemented to store the analyzed discourse, enabling the generation of reports and visualizations on patterns and recurrences of such content. The expected outcome is the creation of a data repository that can support researchers, educators, and anti-racist organizations, fostering greater awareness and contributing to the development of public policies. The results will be made available through the tool as an open educational resource. The project is aligned with Sustainable Development Goal 10, which aims to reduce inequalities within and among countries, particularly regarding racial equity in digital spaces.

Keywords: Digital racism; Artificial intelligence; Social studies.

SUMÁRIO

1 INTRODUÇÃO.....	1
1.1 Raça e Racismo.....	2
1.2 Racismo Digital x Racismo Algorítmico.....	4
1.3 Justificativa.....	5
2 OBJETIVOS.....	5
2.1 Objetivos Específicos.....	6
2 MATERIAL E MÉTODOS.....	6
2.2 Desktop (cliente).....	6
2.3 Servidor (Back-end).....	7
2.3.1 API Principal.....	7
2.3.2 Coletores.....	7
2.4 Modelo de Inteligência Artificial.....	8
2.5 Metodologia Scrum.....	8
2.5.1 Duração dos sprints.....	8
2.5.2 Produto.....	8
2.5.3 Frentes de trabalho.....	9
2.5.4 Product backlog.....	9
Figura 1: Scrum Planning Board.....	9
2.6 Tabela SWOT (FOFA).....	10
Figura 2: Tabela SWOT (Strengths, Weaknesses, Opportunities, Threats) desenvolvida para este projeto.....	10
3 RESULTADOS.....	10
3.1 Treinamento do modelo.....	11
Figura 3: Nuvem de palavras gerada pelo modelo.....	11
Figura 4: Matriz confusão.....	11
Figura 5: Curva ROC.....	12
3.2 Aplicação Desktop.....	13
3.3 Raspadores de dados.....	13
3.4 APIs próprias.....	13
4 CONCLUSÕES.....	14
REFERÊNCIAS.....	14

1 INTRODUÇÃO

A era digital revolucionou a forma como as pessoas se comunicam, trocam ideias e interagem entre si. No entanto, por tratar-se de um ambiente novo e pouco regulamentado, as redes sociais tornaram-se palco da disseminação de diversos tipos de preconceito, entre eles o racismo. É devido à baixa fiscalização desses meios de comunicação que personalidades como David Duke (liderança da Ku Klux Klan) se sentem livres para disseminar seus preconceitos trajados de opinião. Esse fenômeno pode ser comprovado analisando o seguinte trecho de Silva(2022)

Manifestações online racistas anti-negritude são questões brasileiras e globais, ligadas a diferentes níveis de formação estratégica de grupos de ódio, em espaços de fóruns como o *Stormfront* (lançado em 1996), o *4chan* (lançado em 2003) ou plataformas de mídias sociais hegemônicas como o Facebook e Twitter. Em 2007, o supremacista branco David Duke, liderança na Ku Klux Klan, tinha a visão de que “A internet dá, a milhões, acesso à verdade que muitos nem sabiam que existia. Nunca na história do homem pôde informação poderosa viajar tão rápido e tão longe. Eu acredito que a Internet vai gerar uma reação em cadeia de esclarecimento racial que vai chacoalhar o mundo pela velocidade de sua conquista intelectual.” A história da apropriação intensa da internet pelo supremacismo branco, que viu o novo meio de comunicação para a descentralização da abordagem e recrutamento de grupos suscetíveis ao extremismo foi consideravelmente subnotificada até recentemente. (SILVA, 2022, [sem página]).

A discriminação racial está presente na internet nas mais variadas formas, desde as microagressões até as formas mais diretas. Portanto, para se fazer uma análise completa acerca de como as questões raciais se apresentam nas redes, é preciso conceituar e entender alguns dos termos muito utilizados.

1.1 Raça e Racismo

Para compreender o que se caracteriza como injustiça racial, é essencial entender o que é raça. Segundo o Dicionário Michaelis, *raça* refere-se a “[...] cada um dos grandes grupos em que se dividem os seres humanos segundo características físicas hereditárias comuns, como a cor da pele, forma do crânio, tipo de cabelo etc” (Michaelis, 2025). Além disso, é preciso entender que tal termo já possuiu diferentes significações ao longo de contextos históricos distintos. Por exemplo, no seu surgimento o vocábulo foi usado para diferenciar grupos de humanos com base em supostas evidências de características hereditárias.

Somente nos séculos posteriores a palavra passou a ser utilizada de forma pejorativa para justificar a subjugação de alguns povos perante outros.

Esse uso adquirido pela expressão “raça” gerou nos séculos seguintes diversas formas de preconceito racial, que ao passar do tempo, foram até mesmo sendo institucionalizadas. Tem-se como exemplo desse fenômeno o colonialismo europeu, que utilizou a ideia de superioridade racial (branca) para criar uma justificativa ideológica para a opressão de pessoas que não se enquadram na raça branca. Como resultado, surgia o racismo institucionalizado, no qual o preconceito deixou de ser individual e tornou-se um sistema econômico que refletiu em desigualdades sociais, econômicas e políticas que persistem no presente.

Portanto, racismo compreende qualquer tipo de exclusão, discriminação ou desvalorização de uma raça ou etnia em específico. Após muitas lutas de diversos povos, o racismo tornou-se crime. Entretanto, caracterizar o racismo não é uma tarefa fácil. Em muitos casos tal preconceito aparece de forma sutil nas relações interpessoais, o que é conhecido como racismo velado. Segundo Petry e Dias (2018), o racismo velado tem sua reprodução enraizada no dia a dia da população e isso cria um sistema de agressões que dificulta a denúncia por parte das vítimas. Ademais, a segregação racial também está presente nas esferas públicas, materializando-se em casos de violência policial, discriminação de servidores públicos e até mesmo na baixa representatividade em cargos políticos. Essa marginalização gera efeitos para além dos pessoais, como vemos em Ribeiro e Meireles (2023):

A partir do lugar de mulher negra, servidora pública da Universidade, pretendo, a partir da discriminação racial sofrida nesse ambiente de trabalho, apontar aspectos do racismo velado que me levaram ao sofrimento psíquico e, posterior, adoecimento mental, para problematizar como a discriminação racial rebate diretamente na saúde mental da população negra, não se restringindo ao processo de adoecimento individual; mas coletivo, enquanto fenômeno estrutural que é reproduzido nas relações sociais. (Ribeiro; Meireles, 2023, Revista ABPN, Edição: 16)

Conforme Rodrigues (2023), as mídias sociais têm uma enorme responsabilidade na recente ascensão da extrema direita ao redor do mundo, pois as mesmas ajudam a disseminar preconceitos e engajar discursos de ódio. Isso se mostra real quando faz-se uma análise de casos como o movimento *White Lives Matter* (WLM), que, também segundo Rodrigues (2023), é um movimento que, assim como o Ku Klux Klan, age para criar uma ideia de guerra racial no psicológico das pessoas. Essa estratégia de criar um inimigo em comum é amplamente utilizada por

movimentos radicais de direita, que visam manipular determinada massa em prol da realização de alguma ação em favor do regime.

Em virtude da fácil disseminação de informações não confirmadas ou errôneas que a internet possibilita, esses movimentos criam fóruns e páginas nas redes sociais para que se possa agir de forma discriminatória livremente sem que haja consequências. Parte dessa impunidade se dá devido ao fato de ser extremamente complexa a fiscalização da web, com isso muitos países acabam por não ter legislações que tratem sobre essa problemática e, por consequência, os casos de discriminações, não apenas as raciais, aumentaram conforme o espaço virtual se difundiu pelo globo. E, segundo (Silva, 2022), criar uma regulamentação estatal não seria possível por se tratar de uma utopia já que a internet não tem dono.

Em síntese, é considerada racista qualquer injúria sofrida ou feita por um indivíduo que seja baseada majoritariamente em questões de raça e etnia. O racismo está fortemente presente no cotidiano dos seres humanos, manifestando-se através de atitudes como piadas, apelidos, ofensas e pensamentos, tanto nas relações entre os indivíduos como entre os órgãos públicos ou privados. Adicionalmente, com o avanço da rede mundial, surgiram novas formas de racismo e dificuldades na sua fiscalização, bem como novos movimentos neonazistas que criam uma guerra entre as raças, e isso acaba tirando diversas vidas direta e indiretamente.

1.2 Racismo Digital x Racismo Algorítmico

Racismo digital ou ciberracismo refere-se às discriminações realizadas por pessoas nas plataformas digitais, como, por exemplo, frases racistas, xingamentos, assédio, piadas ofensivas, entre muitos outros comentários públicos que gerem hostilidade contra um grupo de pessoas específico na internet. No entanto, o racismo algorítmico ocorre quando uma inteligência artificial acaba discriminando uma certa raça de pessoas, intencionalmente ou não. Isso se dá uma vez que o algoritmo pode aprender com dados que já têm preconceitos históricos ou sociais, como, por exemplo:

O incidente da câmera fotográfica: Um dos primeiros casos de racismo algorítmico foi apontado em janeiro de 2010, quando uma **marca de câmeras fotográficas não reconhecia rostos asiáticos** por um recurso

que evita selfies com olhos fechados. A marca desenvolveu um recurso que avisa se as pessoas estavam com os olhos fechados e rostos pessoas asiáticas acabavam recebendo o alerta também (Gomes, 2024)

Assim, observa-se a distinção entre o racismo digital e o racismo algorítmico, que refletem as desigualdades enraizadas na sociedade e demandam uma atenção cuidadosa para garantir que as tecnologias sejam mais inclusivas, justas e livres de discriminação.

1.3 Justificativa

Essa pesquisa justifica-se a partir do momento que o crescimento da internet se dá junto a um aumento de discriminações, nos seus mais diversos aspectos, e medidas de fiscalização não avançam na mesma velocidade. Portanto, para mitigar diversos problemas que isso tem gerado, como por exemplo, assassinatos, atentados, desigualdade social, suicídios, entre outros. É importante que novas formas de supervisão sejam criadas para facilitar a observação do Estado e até mesmo de instituições privadas.

Devido a essa demanda por uma ferramenta que ajude a obter dados sobre a discriminação racial nos ambientes digitais nas suas mais variadas formas, sendo elas microagressões ou macroagressões, este projeto propõe desenvolver um instrumento com o objetivo de facilitar pesquisas e vigilância de casos de racismo.

2 OBJETIVOS

Com essa pesquisa, a principal meta a ser alcançada será rastrear e evidenciar os casos de racismo nas redes em suas mais variadas formas. Espera-se que esta ferramenta contribua para possibilitar, de uma certa forma, a fiscalização por órgãos responsáveis e, ao mesmo tempo, gerar dados para pesquisar na área.

2.1 Objetivos Específicos

a. **Analisar comentários das redes sociais**

Coletar, filtrar e classificar comentários públicos extraídos de redes sociais com o objetivo de identificar padrões linguísticos e comportamentais associados à injúria racial e ao discurso de ódio.

b. **Gerar relatórios com os dados analisados**

Desenvolver mecanismos automatizados de geração de relatórios que apresentem estatísticas, recorrência de termos racistas, redes sociais de maior incidência, e perfis de conteúdo mais críticos, permitindo interpretações quantitativas e qualitativas.

c. **Desenvolver uma aplicação desktop amigável**

Projetar e implementar uma aplicação de fácil uso, com interface gráfica acessível, que permita a visualização dos dados, filtragem de informações, geração de relatórios e exportação de resultados por parte dos usuários.

2 MATERIAL E MÉTODOS

O software será desenvolvido em múltiplas frentes, sendo elas: desktop, servidor e modelo de inteligência artificial. Além disso, será feito o uso da metodologia ágil Scrum para melhor gestão de tempo, equipe e entregas.

2.2 Desktop (cliente)

A plataforma desktop será desenvolvida com o Tauri, um *framework* que permite criar aplicações desktop e *mobile* de maneira leve, rápida e segura, utilizando a linguagem Rust no *back-end*, e o *front-end* agnóstico.

O Tauri possui uma performance superior em relação a outras soluções como o Electron, além da segurança proporcionada pelo Rust e a diminuição significativa do tamanho das aplicações geradas. A aplicação será compatível com múltiplos sistemas operacionais, como Windows e Linux, aumentando seu alcance para os usuários. Como suporte, serão utilizadas React, Vue, Svelte e outras tecnologias. O

front-end agnóstico será executado sobre o WebView nativo do sistema, facilitando a integração com aplicações web. Além disso, a segurança proporcionada pela sua base em Rust é um ponto forte do Tauri, permitindo controle do gerenciamento de memória e falhas comuns.

Em suma, o Tauri mostra-se uma ótima opção para atender os requisitos técnicos e as perspectivas de negócios projetadas.

2.3 Servidor (Back-end)

O *back-end* do projeto é separado em dois componentes principais, a Interface de Programação de Aplicações (API) principal e os coletores.

2.3.1 API Principal

A API principal atuará como a *API Gateway* do sistema, sendo responsável por intermediar a comunicação entre o cliente e os serviços internos, como os coletores de dados. Para sua implementação, utilizaremos a linguagem Python para o desenvolvimento do servidor HTTP quanto na integração com o banco de dados e nas requisições ao nosso modelo de classificação. Como base de dados, utilizaremos o Firestore, o banco NoSQL da Google, que armazenará todas as informações do sistema — desde dados de usuários até os resultados das análises realizadas.

2.3.2 Coletores

Na arquitetura proposta, os “coletores” são responsáveis pela captura de dados em redes sociais, utilizando APIs oficiais ou Web Scraping, conforme a disponibilidade das plataformas. Para desenvolvê-los, será utilizado tecnologias que facilitem a integração com cada uma das redes, e submeter o conteúdo a uma fila de processamento, para que assim possam ser classificadas pelo modelo de inteligência artificial.

2.4 Modelo de Inteligência Artificial

O modelo de inteligência artificial será um *fine-tuning* do BERTimbau, um modelo pré-treinado projetado para compreender a língua portuguesa à base do BERT, ajustado para classificar comentários que indicam ou não presença de injúria racial.

O ajuste fino adapta um modelo pré-treinado a uma tarefa específica com um conjunto de dados especializado mais pequeno. Esta abordagem requer muito menos dados e computação em comparação com o treino de um modelo de raiz, o que a torna uma opção mais acessível para muitos utilizadores. (Hugging Face, 2025)[tradução nossa]

Para realizar esse ajuste, será utilizada uma base de dados disponível na internet, contendo informações sobre diversos casos de discurso de ódio, incluindo episódios relacionados ao racismo.

2.5 Metodologia Scrum

Para que seja possível realizar todas as propostas do trabalho dentro do prazo a equipe utilizou da metodologia SCRUM.

O Scrum é um framework de gerenciamento que as equipes usam para se auto-organizar e trabalhar em direção a um objetivo em comum. A estrutura descreve um conjunto de reuniões, ferramentas e funções para uma entrega eficiente de projetos. (AWS, 2024)

Segundo os princípios de funcionamento desse framework de gerenciamento, foi definido alguns pontos:

2.5.1 Duração dos sprints

Após uma análise de quanto tempo a equipe teria para desenvolver o projeto ficou decidido que cada sprint possuiria uma duração de 2 semanas.

2.5.2 Produto

Também foi estudado qual seria o produto entregue ao final do projeto, para isso criou-se o MVP (minimum viable product) da aplicação para saber o que deveria ser entregue.

2.5.3 Frentes de trabalho

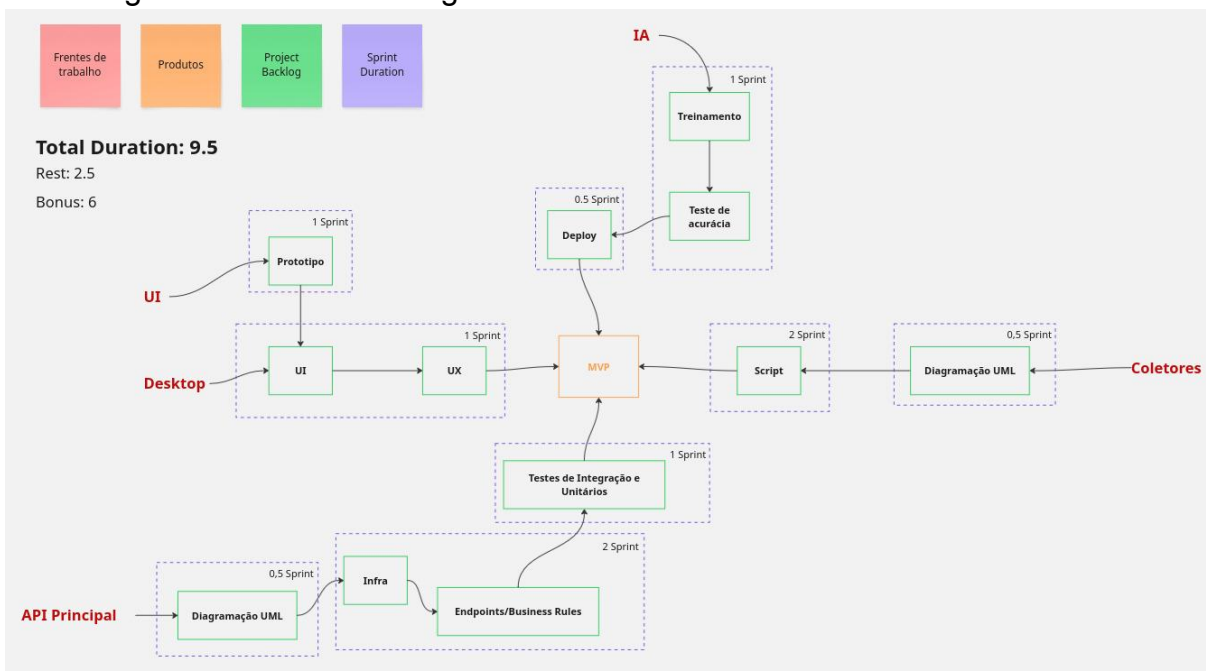
Após definir o que seria entregue, foi decidido quais seriam as frentes de trabalho que a equipe iria desenvolver para tornar o processo de desenvolvimento mais fluido.

2.5.4 Product backlog

Após o levantamento de requisitos para definir o que seria necessário para ser desenvolvido em cada frente de trabalho e, com isso propô-se os product 's backlog da aplicação. Ao mesmo tempo, definiu-se quantos sprints serão necessários para cada backlog.

Após definir esses pontos, criou-se um projeto na plataforma Miro para juntar todas as informações de uma forma visualmente didática como apresentado na figura 1.

Figura 1: Scrum Planning Board.



Fonte: Os autores.

2.6 Tabela SWOT (FOFA)

Na figura 2, é apresentado a matriz SWOT (Strengths, Weaknesses, Opportunities, Threats) ou FOFA (Forças, Oportunidades, Fraquezas, Ameaças) desenvolvido pela equipe.

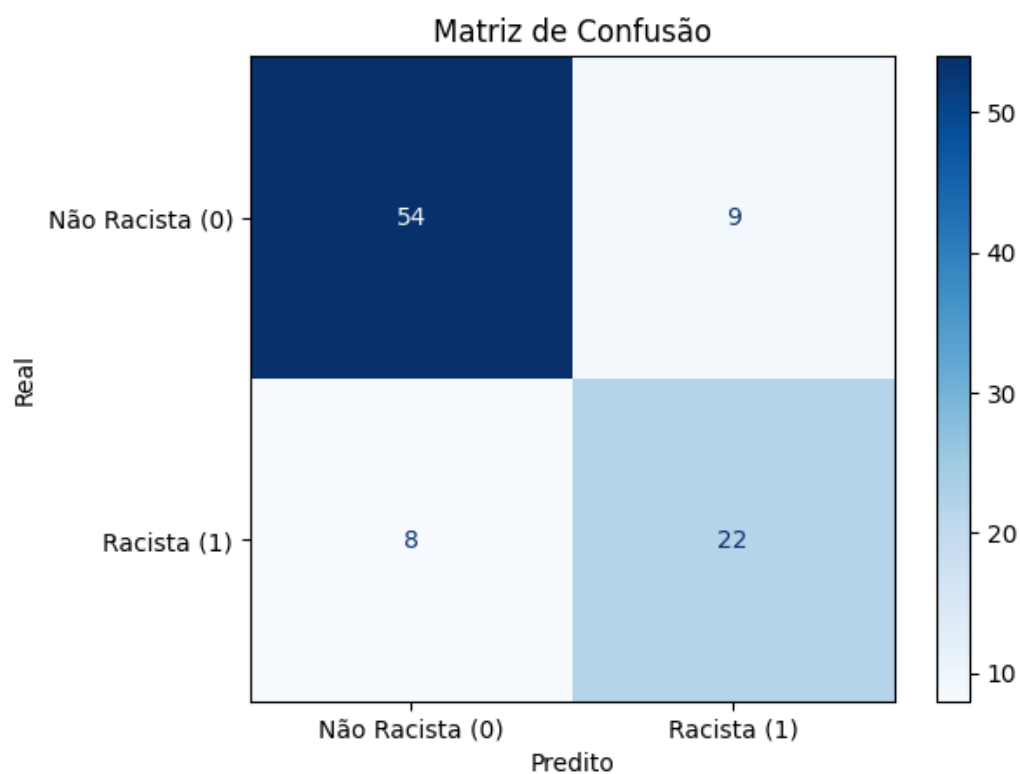
Figura 2: Tabela SWOT (Strengths, Weaknesses, Opportunities, Threats) desenvolvida para este projeto.



Fonte: Os autores.

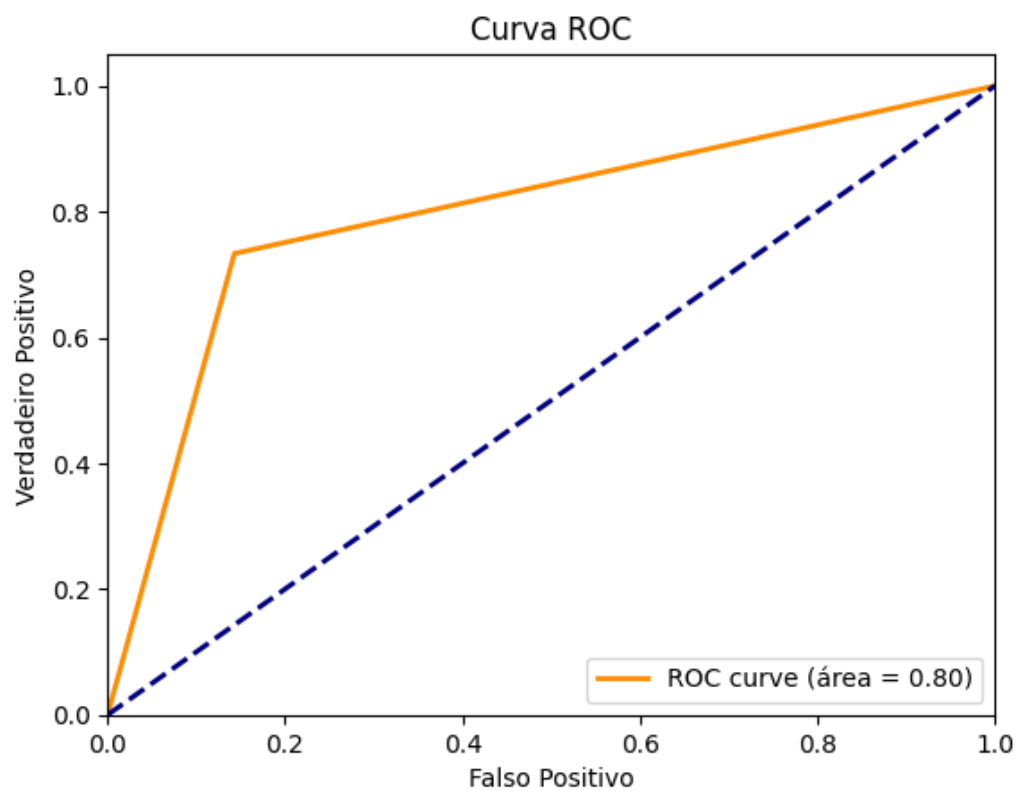
3 RESULTADOS

Até o momento, a plataforma SankofaAI ainda está em desenvolvimento contínuo e contando com o aprimoramento constante das ferramentas como por exemplo: o modelo de inteligência artificial e os *scrappers* (raspadores de dados), *para que assim possam ser concluídos os objetivos propostos anteriormente.*



Fonte: Os autores

Figura 5: Curva ROC



Fonte: Os autores

E com o auxílio de ferramentas nativas do *Trainer* (classe utilizada para treinar o modelo), foi constatado que a acurácia média da I.A estava com o valor 0,820 em uma escala de 0 a 1, um nível de confiança bem alto o que somado aos outros índices corrobora com o fato de que o modelo está confiável.

3.2 Aplicação Desktop

Atualmente, o aplicativo está com todo o seu frontend concluído. As telas foram desenvolvidas com base nos requisitos da ideia inicial desse projeto e até o presente não estão funcionais, pois ainda não foram integradas com as APIs do projeto.

3.3 Raspadores de dados

Os raspadores de dados (scrappers) estão concluídos. Porém, no processo de testes percebemos que a raspagem dos dados da rede social X será pouco efetiva para o projeto, já que a plataforma mencionada tem muitos mecanismos que proíbem a raspagem de dados, até mesmo proibindo o seu internet protocol (IP). Logo, o uso dos dados dessa rede foi e será pouco eficiente na ferramenta SankofaAI. Todavia, não houveram problemas com a raspagem dos dados das outras redes Instagram e BlueSky. Sendo assim os coletores estão concluídos, mas com uma leve redução na quantidade de dados disponibilizados.

Para fins de esclarecimento o aplicativo X disponibiliza uma API própria, contudo a mesma é inacessível com os planos mais básicos custando cerca de 100 dólares por mês e tendo uma quantidade muito limitada de publicações. Algo que não satisfaria os requisitos do projeto.

3.4 APIs próprias

As APIs próprias que irão cuidar da comunicação I.A e backend além de gerenciar os coletores e enviar para a fila de processamento já estão prontas.

4 CONCLUSÕES

Conclui-se que o projeto SankofaAI é uma importante resposta para o racismo digital que não acabará com o mesmo, mas que irá possibilitar a identificação para fins de pesquisa e criação de políticas públicas, dos casos que contenham discursos de ódio com teor racial na internet. Entretanto, para o funcionamento pleno da aplicação assim como descrito nos objetivos do projeto será necessário o prazo completo que foi estipulado no início para que seja possível integrar todas as partes da aplicação. Sendo assim, será necessário trabalhar no desenvolvimento até meados de dezembro.

REFERÊNCIAS

AWS. O que é Scrum? Amazon Web Services, 2024. Disponível em: <https://aws.amazon.com/pt/what-is/scrum/#:~:text=O%20Scrum%20%C3%A9%20um%20framework,uma%20entrega%20eficiente%20de%20projetos>. Acesso em: 30 abr. 2025.

Fine-tuning. Hugging Face. Disponível em: <https://huggingface.co/docs/transformers/training>. Acesso em: 30 abr. 2025.

GOMES, Graziely Ketlen de Souza. O que é racismo algorítmico? Politize!, 14 mar. 2024. Disponível em: <https://www.politize.com.br/o-que-e-racismo-algoritmico/>. Acesso em: 23 abr. 2025.

MICHAELIS. Dicionário Brasileiro da Língua Portuguesa. Disponível em: <https://michaelis.uol.com.br/moderno-portugues/>. Acesso em: 29 abr. 2025.

PETRY, Caroline; DIAS, Felipe da Veiga. O racismo velado sob a ótica da teoria do etiquetamento e cifra oculta. [S.l.]: [s.n.], 2018.

RIBEIRO, M.; MEIRELES, F. P. Racismo velado nos atos discricionários de gestão e a saúde mental de servidoras negras de uma universidade federal do sudeste brasileiro: Veiled racism in discretionary acts of management and mental health of black servants at a federal university in south southeast Brazil. *Revista da Associação Brasileira de Pesquisadores/as Negros/as (ABPN)*, [S. l.], v. 16, n. Edição Especial, 2023. Disponível em: <https://abpn.emnuvens.com.br/site/article/view/1602>. Acesso em: 29 abr. 2025.

RODRIGUES, Marcos Alexandre Fernandes. Racismo, segregação e morte: análise dialógica do discurso das organizações Ku Klux Klan e White Lives Matter em mídias digitais. 2023. Dissertação (Mestrado em Letras) – Universidade Federal do Rio Grande, Instituto de Letras e Artes, Rio Grande, 2023.

Tauri 2. Tauri. Disponível em: <https://v2.tauri.app>. Acesso em: 23 abr. 2025.

SILVA, Tarcizio. *Racismo algorítmico: inteligência artificial e discriminação nas redes digitais*. PubPub, 2022. Disponível em: <https://racismo-algoritmico.pubpub.org>. Acesso em: 02 jul. 2025.