Certificado

O artigo intitulado

**"COMBATE À DESINFORMAÇÃO COM FACTUAL:
UMA SOLUÇÃO BASEADA EM MODELOS DE LINGUAGEM
DE GRANDE ESCALA"**

de

***Verônica Souza dos Santos[1], Paulo Henrique Batista Rodrigues[1],
Geraldo Pereira Rocha Filho[2], Edna Dias Canedo[1],
Frederico de Almeida Meirelles Palma[3]
e Fábio Lúcio Lopes de Mendonça[1]***

[1]Universidade de Brasília – UnB, Brasil
[2]Universidade Estadual do Sudoeste da Bahia (UESB), Brasil
[3]Coordenação Geral da Rede SUAS – MDS, Brasil

foi selecionado como

Melhor Artigo

na

21ª Conferência Ibero Americana WWW/Internet (CIAWI 2024)

O Comité da Conferência, com base numa apreciação imparcial, considera este artigo como um dos melhores trabalhos.

Paula Miranda
Program Chair

iadis
international association for development of the information society

COMBATE À DESINFORMAÇÃO COM FACTUAL: UMA SOLUÇÃO BASEADA EM MODELOS DE LINGUAGEM DE GRANDE ESCALA

Verônica Souza dos Santos[1], Paulo Henrique Batista Rodrigues[1], Geraldo Pereira Rocha Filho[2], Edna Dias Canedo[1], Frederico de Almeida Meirelles Palma[3] e Fábio Lúcio Lopes de Mendonça[1]
[1]Universidade de Brasília – UnB, Brasil
[2]Universidade Estadual do Sudoeste da Bahia (UESB), Brasil
[3]Coordenação Geral da Rede SUAS – MDS, Brasil

RESUMO

Com o crescimento das tecnologias digitais e das redes sociais, a disseminação de fake news tornou-se um problema significativo, afetando a sociedade em áreas como política, saúde e processos democráticos. Esta pesquisa aborda duas lacunas fundamentais: (i) a ausência de uma categorização sistemática das diferentes formas de desinformação; e (ii) a falta de modelos de grandes linguagens de modelos (LLMs) capazes de generalizar eficientemente em múltiplos domínios e conjuntos de dados heterogêneos. Diante disso, este trabalho apresenta o FACTUAL, uma solução baseada em LLMs para detectar e mitigar fake news. O FACTUAL utiliza o modelo LLaMA e frameworks assíncronos como Asyncio e AioSQLite para processar grandes volumes de dados textuais de forma eficiente. A solução inclui uma camada de banco de dados para armazenar os resultados e um middleware que integra a análise de IA oferecendo classificações confiáveis e justificativas claras. Os resultados demonstram que o FACTUAL é eficaz na identificação de padrões de desinformação, apresentando uma boa precisão e confiabilidade nos resultados.

PALAVRAS-CHAVE

Fake News, Desinformação, Redes Sociais, Verificação de Autenticidade, Manipulação de Informações

1. INTRODUÇÃO

A disseminação de fake news, embora não seja um fenômeno recente, aumentou com o crescimento das tecnologias digitais e a popularização das redes sociais. Aliado com o avanço da internet, as fakes news se tornaram mais frequentes, já que qualquer indivíduo agora pode disseminar informações sem controle, utilizando redes sociais, tais como Facebook e Twitter (Alnabhan and Branco, 2024). Um exemplo relevante no Brasil ocorreu durante as eleições presidenciais de 2018, quando uma série de informações falsas sobre urnas eletrônicas foi amplamente divulgada, gerando desconfiança pública e afetando o debate político nacional (Bernardi, 2021; iG São Paulo, 2018). As fakes news, caracterizadas pela propagação rápida de informações fabricadas ou distorcidas, alcançam e influenciam um público maior do que as notícias verdadeiras. Esse conteúdo, frequentemente apresentado como factual, tem sido utilizado para manipular opiniões e obter vantagens políticas, sociais e econômicas, com impactos diretos sobre a percepção pública e os processos decisórios em diversas áreas.

No Brasil, o Marco Civil da Internet, regulamentado pela Lei nº 12.965/2014, estabelece diretrizes para a liberdade de expressão e proteção de dados na internet. No entanto, essa legislação ainda apresenta limitações no combate às fake news, especialmente no que diz respeito à remoção ágil de informações falsas e à responsabilização de provedores de conteúdo. Além das questões locais, eventos globais como as eleições presidenciais dos EUA em 2016 demonstraram a gravidade global do problema, evidenciando como a desinformação afeta os processos democráticos e reforçando a necessidade de regulamentações e tecnologias mais eficazes (Alnabhan and Branco, 2024; Sharma et al., 2019). Esses cenários, tanto local quanto global, destacam a urgência de novas pesquisas e abordagens para mitigar a propagação de fake news, especialmente em situações em que a rapidez da disseminação pode gerar consequências graves.

Detectar fake news não é uma tarefa trivial devido a diversos fatores, tais como a ambiguidade e subjetividade do conteúdo, a rapidez com que essas informações se propagam, a criação de conteúdo sofisticado que dificulta a diferenciação entre verdadeiro e falso, e a falta de rótulos e dados confiáveis (Alnabhan and Branco, 2024; Sharma et al., 2019). Em razão disso, a Inteligência Artificial (IA) surge como uma solução promissora para o combate à desinformação. Utilizando Modelos de Linguagem de Grande Escala (LLMs), a IA consegue analisar grandes volumes de dados em tempo real, identificando padrões e sinais de fake news. A complexidade dos LLMs permite lidar com a ambiguidade e a subjetividade do conteúdo, facilitando a detecção de informações falsas mesmo em meio a textos sofisticados e rapidamente disseminados (Sun et al., 2024; Wu et al., 2024). LLMs, como o LLAMA, Mistral e Phi, possuem uma capacidade avançada de compreender nuances no conteúdo textual, ajudando a identificar informações potencialmente falsas e fornecendo uma análise aprofundada do contexto, mesmo em casos de ambiguidade e conteúdo sofisticado. Além disso, tecnologias como o processamento de linguagem natural permitem que modelos de IA compreendam o contexto e o conteúdo de um texto, auxiliando na verificação da veracidade das informações.

Diversos trabalhos têm explorado o problema da detecção de fake news em diferentes contextos e abordagens. Em (Hu et al., 2024) e (Wang et al., 2023) exploram o uso de LLMs e abordagens multiespecializadas para ampliar a detecção de fake news em múltiplos domínios. No entanto, essas abordagens enfrentam limitações relacionadas à complexidade computacional e à introdução de vieses. Por outro lado, os trabalhos de (Kaliyar et al., 2021) e (Segura-Bedmar and Alonso-Bartolome, 2022) focam em arquiteturas baseadas em redes neurais convolucionais (CNNs) e abordagens multimodais, combinando texto e imagem para aumentar a acurácia na classificação de notícias falsas, embora o desequilíbrio de classes nos datasets utilizados limite sua eficácia em alguns cenários. Em um contexto diferente, raja2023fake aborda a detecção de fake news em línguas de poucos recursos, utilizando aprendizado por transferência com modelos transformadores, enquanto dou2021user propõe um framework que combina preferências de usuários e redes sociais para melhorar a detecção de fake news. No entanto, a complexidade computacional e a dependência de dados históricos ainda são desafios comuns a essas abordagens, limitando sua aplicabilidade em cenários reais e escaláveis.

Vale destacar, no entanto, que esta pesquisa foca em duas lacunas principais: (i) a falta de uma categorização das diferentes formas de desinformação, como desinformação deliberada, informações falsas não intencionais e rumores; e (ii) a escassez de modelos de LLMs capazes de generalizar de maneira eficaz em múltiplos domínios e conjuntos de dados heterogêneos. Essas lacunas tornam-se essenciais para aprimorar as capacidades de detecção de fake news em cenários complexos e variados, e esta pesquisa explora tais lacunas.

Com isso em mente, este trabalho apresenta o FACTUAL - Fake News Analysis with Confidence and Trust Using AI and LLM - uma solução baseada em LLMs para detectar e mitigar a disseminação de fake news. O FACTUAL foi modelado com base no modelo LLaMA, que oferece uma capacidade de análise de dados textuais por meio do framework Ollama, permitindo identificar padrões e sinais característicos de desinformação de forma eficiente. Além disso, o FACTUAL foi desenvolvido sobre frameworks assíncronos como Asyncio e AioSQLite, otimizando a integração com bancos de dados e facilitando a execução paralela de tarefas, o que resulta em uma maior escalabilidade e eficiência no processamento de grandes volumes de dados.

O restante deste trabalho está organizado da seguinte forma. Na Seção 2, discutimos os trabalhos relacionados. A Seção 3 descreve o desenvolvimento do FACTUAL, enquanto a Seção 4 aborda a avaliação de desempenho. Por fim, a Seção 5 apresenta as conclusões e direções futuras.

2. TRABALHOS RELACIONADOS

Esta seção explora diferentes trabalhos que abordam o problema da detecção de fake news. O trabalho de (Ma et al., 2024) propõe o framework Event-Radar, focado na detecção de fake news em mídias multimodais. O Event-Radar combina aprendizado multi-view e inconsistências a nível de eventos entre texto e imagem, modelando grafos de eventos para capturar relações sujeito-predicado nas notícias. Utilizando distribuições Beta, o modelo ajusta a credibilidade das modalidades. Suas limitações incluem a falta de exploração de relações causais entre eventos e a dependência de ferramentas externas para reconhecimento de entidades nomeadas.

Os trabalhos de (Hu et al., 2024) e (Wang et al., 2023) compartilham abordagens centradas no uso de LLMs e técnicas avançadas de detecção de fake news em múltiplos domínios. Em (Hu et al., 2024), os autores exploram a utilização do GPT-3.5 para atuar como conselheiros no processo de detecção de fake news. Nesse contexto, os LLMs fornecem justificativas e insights que complementam o desempenho de modelos menores, como o BERT. A proposta inclui a criação de uma Rede de Orientação Adaptativa de Justificativas (ARG), que integra as percepções geradas pelos LLMs para melhorar a capacidade de julgamento dos modelos mais leves. Embora essa abordagem adicione interpretabilidade aos modelos menores, sua dependência de infraestrutura computacional robusta representa uma limitação significativa. Além disso, a ausência de um mecanismo explícito de verificação de fatos pode deixar lacunas na detecção em casos mais complexos.

De forma semelhante, o estudo de (Wang et al., 2023) explora um sistema que busca ampliar a capacidade de detecção de fake news em múltiplos domínios, como política, entretenimento e finanças. Os autores utilizam soft-labels, uma abordagem que captura melhor as nuances de diferentes áreas, e um mecanismo de LeapGRU, projetado para ignorar palavras irrelevantes. O modelo aplica uma análise mais especializada, dividindo as notícias por domínio e utilizando grupos de especialistas para extrair características específicas. Embora esse método forneça uma análise direcionada, as limitações incluem a complexidade no tratamento de textos longos e a possível introdução de vieses, devido a interpretações subjetivas dos especialistas.

Os trabalhos de (Kaliyar et al., 2021) e (Segura-Bedmar and Alonso-Bartolome, 2022) focam em modelos multimodais e no uso de redes neurais convolucionais (CNNs) para detecção de fake news, utilizando diferentes modalidades de entrada. Em (Kaliyar et al., 2021), os autores propõem uma combinação do BERT com uma CNN, que utiliza blocos paralelos com diferentes tamanhos de kernel para capturar dependências semânticas de longa distância. Essa combinação é eficaz para lidar com ambiguidades em textos de mídias sociais, mas sua complexidade limita a aplicabilidade em larga escala, especialmente em ambientes com recursos restritos. Além disso, o treinamento desses modelos requer muitos dados rotulados, o que pode ser um desafio em domínios com baixa disponibilidade.

De forma semelhante, o artigo de (Segura-Bedmar and Alonso-Bartolome, 2022) apresenta uma abordagem multimodal que combina texto e imagem para a detecção de fake news. Utilizando o dataset Fakeddit, que contém seis categorias de notícias, e uma arquitetura baseada em CNN, os autores combinam modalidades de entrada para enriquecer a representação da notícia e melhorar a acurácia. No entanto, uma limitação significativa está no desequilíbrio das classes no dataset utilizado, o que impacta diretamente o desempenho em detectar classes minoritárias, resultando em uma redução de precisão e recall.

O estudo de (Raja et al., 2023) propõe uma abordagem específica para a detecção de fake news em línguas de poucos recursos, utilizando aprendizado por transferência com modelos transformadores pré-treinados, como o mBERT e o XLM-RoBERTa, ajustados para línguas dravidianas. A metodologia envolve fine-tuning adaptativo, permitindo que o modelo ajuste suas representações de acordo com os dados dessas línguas, usando transfer learning a partir de corpora em inglês. Apesar de alcançar bons resultados, o desempenho do modelo é fortemente dependente de dados em inglês, o que representa uma barreira quando se trata de generalização para línguas menos exploradas.

Em (Dou et al., 2021), os autores propõem o framework UPFD (User Preference-aware Fake News Detection), que considera as preferências dos usuários, baseadas em históricos de postagens, e o contexto da propagação de notícias nas redes sociais. O framework utiliza uma Rede Neural Gráfica (GNN) para combinar essas informações e prever a credibilidade das notícias. No entanto, o desempenho do UPFD está diretamente relacionado à qualidade dos dados históricos dos usuários, que podem não estar sempre presentes. Além disso, a fusão de múltiplas fontes de informação torna o framework computacionalmente complexo, enfrentando desafios de escalabilidade.

3. FACTUAL - ANÁLISE DE FAKE NEWS COM CONFIANÇA UTILIZANDO IA E LLMS

Esta seção apresenta o FACTUAL, uma solução baseada em LLM para detectar e mitigar a disseminação de fake news. Desenvolvido com base em modelos de linguagem de grande escala, o FACTUAL visa fornecer uma análise confiável de conteúdos potencialmente falsos, permitindo que os usuários tomem decisões informadas e em tempo real. O objetivo geral do FACTUAL é oferecer uma estrutura eficiente e escalável para o treinamento de grandes modelos de linguagem, proporcionando uma solução eficaz para a detecção de fake

news. Especificamente, o FACTUAL busca maximizar o desempenho em tarefas de processamento de linguagem natural, como a geração, tradução e compreensão de texto, ao mesmo tempo que permite a detecção em tempo real e a mitigação da desinformação de forma eficiente.

3.1 Visão Geral do Funcionamento do FACTUAL

A Figura 1 apresenta o funcionalmento da arquitetura do FACTUAL. O FACTUAL é composto por quatro componentes principais: (i) a camada de banco de dados; (ii) os modelos de IA; (iii) o middleware; e (iv) os frameworks de execução. A Camada de Banco de Dados (Rótulo 1, Figura 1) é responsável pelo armazenamento e gerenciamento dos dados utilizados no processo de detecção. Essa camada no FACTUAL utiliza um banco de dados SQLite para armazenar os registros relacionados a fake news que mantém informações comparativas, em que são gravadas as saídas da análise realizada pelo FACTUAL. Já a Camada de Modelos e Frameworks de IA (Rótulo 2, Figura 1) concentra o núcleo da análise do FACTUAL, utilizando o modelo LLaMA 3.2 para processar grandes volumes de dados textuais. O framework Ollama é utilizado para integrar o modelo ao FACTUAL, permitindo a análise e a identificação de padrões característicos de desinformação. O uso de um modelo de linguagem avançado como o LLaMA oferece ao FACTUAL a capacidade de compreender nuances e ambiguidade no conteúdo textual, aumentando a precisão da detecção de fake news em cenários complexos.

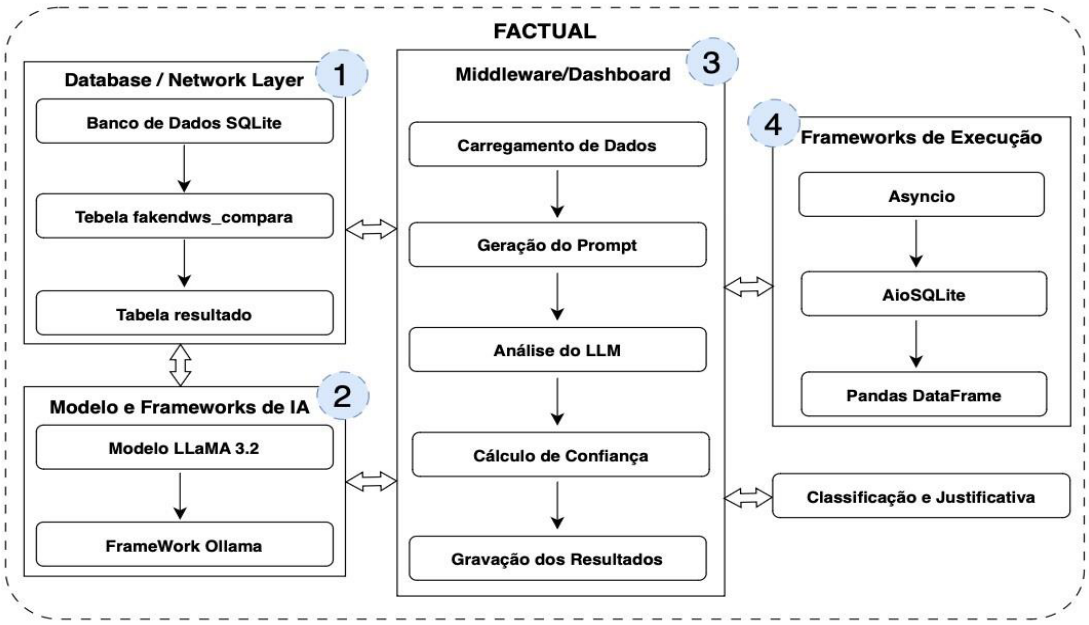


Figura 1. Visão geral do funcionamento do FACTUAL

O Middleware (Rótulo 3, Figura 1) desempenha um papel central na integração entre a camada de dados e os modelos de IA. Esse componente é responsável por tarefas como o carregamento de dados do banco de dados, a geração de prompts apropriados para o modelo de IA, a análise dos resultados obtidos e o cálculo de confiança, o que permite ao sistema avaliar a veracidade das informações analisadas. Além disso, o middleware grava os resultados no banco de dados para posterior consulta, fechando o ciclo de análise do FACTUAL. por fim, a camada de Frameworks de Execução (Rótulo 4, Figura 1) é desenvolvida para otimizar o desempenho do FACTUAL, utilizando frameworks assíncronos como Asyncio e AioSQLite para garantir a execução paralela das tarefas. Esses frameworks permitem que o FACTUAL processe múltiplos pedidos de análise simultaneamente, sem sobrecarregar o sistema, garantindo eficiência e escalabilidade. Os resultados são organizados em Pandas DataFrames, facilitando a visualização e a classificação dos dados, além de fornecer justificativas claras para as conclusões sobre a veracidade das informações analisadas.

3.2 Camada de Banco de Dados no FACTUAL

A camada de banco de dados do FACTUAL é responsável pela estruturação e armazenamento das informações necessárias para o processo de detecção de fake news. O sistema utiliza um banco de dados SQLite, escolhido pela sua leveza e integração com outras camadas do sistema. Dois componentes principais compõem essa camada: (i) a tabela "fakenews_compara"; e (ii) a tabela de resultados. A tabela "fakenews_compara" armazena os dados de entrada, que são informações textuais e outras variáveis coletadas de fontes que potencialmente disseminam fake news. Essa tabela mantém um registro das informações que serão analisadas pelo modelo de linguagem, proporcionando a base para o início do processo de detecção.

Após a análise realizada pelos modelos, os resultados são armazenados na tabela de resultados, que contém as saídas geradas pelo sistema após a detecção de fake news. Nessa tabela, são gravados dados como a classificação das notícias (verdadeiras ou falsas), as justificativas fornecidas pelos modelos e o nível de confiança atribuído a cada classificação. Esses registros são utilizados para a posterior visualização e análise, garantindo que o sistema mantenha um histórico das avaliações e suas justificativas. A escolha por uma base de dados SQLite está atrelada à necessidade de um sistema leve que permita o rápido acesso aos dados e que lide com operações de escrita e leitura simultâneas, sem comprometer o desempenho. Essa solução também facilita a integração com frameworks assíncronos, como o AioSQLite, que permitem o processamento de múltiplas requisições simultâneas, tornando o FACTUAL escalável e apto a processar grandes volumes de dados de maneira eficiente.

3.3 Processo de Classificação de Notícias na Camada de Modelos e Frameworks de IA

A camada de modelos e frameworks de IA é responsável pelo processamento e análise das informações textuais no FACTUAL. O principal componente dessa camada é o modelo de linguagem LLaMA 3.2, que realiza a classificação das notícias enviadas pelo usuário, determinando se são verdadeiras ou falsas. O primeiro passo no processo é a inicialização do FACTUAL, em que o modelo LLaMA 3.2 e o framework Ollama são carregados e conectados. Essa etapa é essencial para preparar o ambiente de execução, garantindo que o sistema esteja pronto para processar as notícias. Uma vez configurado, o FACTUAL pode começar a receber as notícias a serem analisadas. O processo de análise de notícias ocorre de maneira iterativa, passando por cada notícia individualmente. O texto da notícia é enviado ao modelo LLaMA 3.2, por meio do framework Ollama, que processa o conteúdo e gera uma resposta contendo a probabilidade de que a notícia seja falsa. A partir dessa resposta, o sistema calcula o nível de confiança da análise, utilizando as informações fornecidas pelo modelo para gerar uma métrica que reflete a certeza do sistema em sua classificação.

Com base na probabilidade retornada pelo modelo LLaMA 3.2, a notícia é classificada como fake, verídica ou indeterminada. O sistema compara a probabilidade de ser fake com um limiar previamente definido. Se a probabilidade ultrapassar esse valor, a notícia é marcada como fake; caso contrário, ela é classificada como verídica ou indeterminada. Além da classificação, o sistema armazena o resultado da análise, o nível de confiança associado à decisão e gera uma justificativa clara ao usuário. O framework Ollama facilita essa interação, fornecendo a interface para o envio dos prompts ao modelo e a recepção dos resultados. Esses resultados, que incluem a classificação e justificativa, são enfileirados em uma tarefa assíncrona, permitindo que o sistema processe múltiplas requisições de forma eficiente.

3.4 Fluxo de Processamento e Integração do Middleware/Dashboard no FACTUAL

A camada de Middleware/Dashboard do FACTUAL gerencia o fluxo de dados entre as diferentes partes do sistema, facilitando a interação entre o usuário, o modelo de IA e o banco de dados. A principal função dessa camada é carregar os dados, gerar prompts para o modelo LLaMA 3.2, processar os resultados e retornar as respostas ao usuário. O diagrama de sequência apresentado na Figura 2 apresenta o fluxo de processamento que ocorre dentro do FACTUAL. O processo começa com o usuário submetendo uma notícia, que é então carregada da tabela "fakenews_compara" no banco de dados SQLite. As notícias incluem informações como título, texto, URL e editor. Essas informações são organizadas e preparadas para o envio ao modelo LLaMA

3.2 por meio da geração de prompts. O middleware organiza e estrutura essas informações, gerando o prompt que será enviado ao modelo de IA para análise.

Após a geração do prompt, o modelo LLaMA 3.2 realiza a classificação da notícia como verdadeira (fato) ou falsa (fake) ou indeterminado e gera uma justificativa para a classificação realizada. Essa resposta é então enfileirada em uma tarefa assíncrona, utilizando o framework Asyncio. A capacidade de processar múltiplas tarefas de forma assíncrona é fundamental para que o FACTUAL consiga lidar com um fluxo constante de requisições, conforme apresentado na Figura 2.

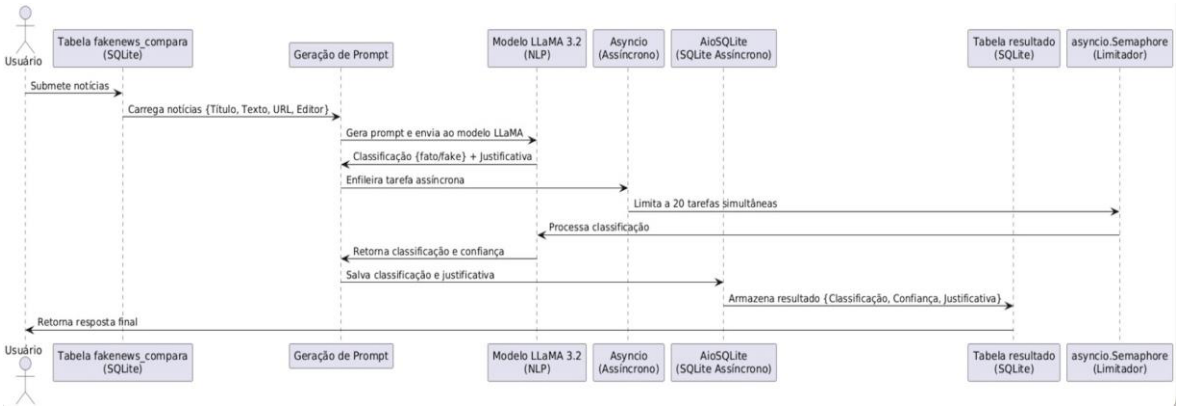


Figura 2. Diagrama de sequência para classificação de notícias

A seguir, o middleware recupera a classificação gerada pelo modelo, juntamente com a confiança, e armazena esses dados na tabela de resultados, retornando a resposta ao usuário. O sistema utiliza AioSQLite, uma extensão assíncrona do SQLite, que habilita operações de I/O não bloqueantes, otimizando latência e throughput em ambientes com alta concorrência. Isso permite que múltiplas requisições de leitura e escrita sejam processadas simultaneamente, maximizando a eficiência ao lidar com grandes volumes de dados. A tabela de resultados armazena a classe prevista, a confiança e uma justificativa gerada pelo modelo. O uso de índices no banco otimiza as consultas, garantindo rápido acesso ao histórico de análises, mesmo em cenários de escalabilidade. O sistema também implementa asyncio.Semaphore, que limita corrotinas concorrentes, prevenindo sobrecargas e controlando o uso de recursos de forma eficiente.

3.5 Execução Assíncrona e Escalabilidade na Camada de Frameworks de Execução do FACTUAL

A Camada de Frameworks de Execução do FACTUAL é responsável por otimizar o processamento e garantir a escalabilidade do sistema, gerenciando as operações de forma assíncrona. Esta camada utiliza frameworks como Asyncio e AioSQLite para processar múltiplas tarefas simultaneamente, sem sobrecarregar o sistema, e lidar com grandes volumes de dados que é uma das características desta pesquia. O Asyncio é um framework de execução assíncrona que permite o processamento de várias requisições em paralelo. No FACTUAL, o Asyncio é utilizado para gerenciar a execução de tarefas simultâneas, conforme apresentado no diagrama da Figura 2. Isso é necessário para garantir que o modelo de IA possa analisar diversas notícias de maneira eficiente, mesmo em momentos de alta demanda.

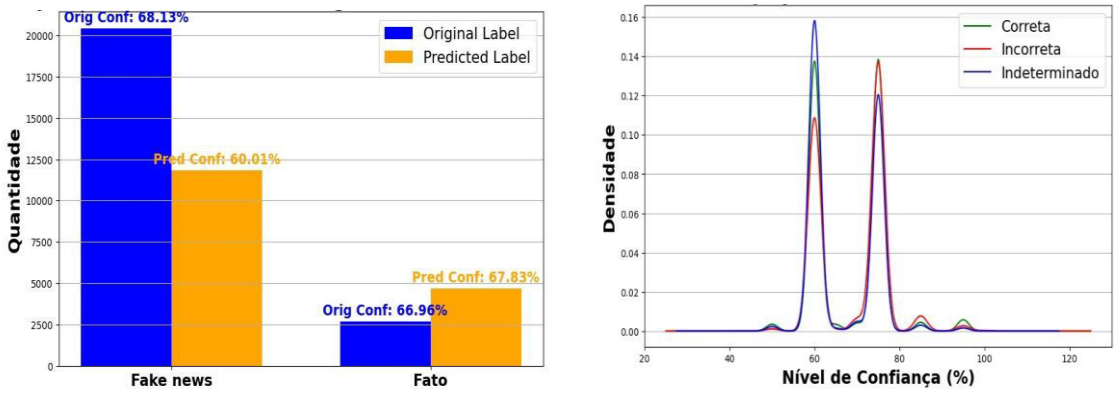
Além de processar múltiplas tarefas em paralelo, a camada de execução utiliza AioSQLite, uma extensão assíncrona do SQLite. O AioSQLite permite ao FACTUAL interagir com o banco de dados de forma eficiente, armazenando os resultados da classificação de fake news em tempo real, sem interrupções no fluxo de trabalho. O semáforo assíncrono controla o número de tarefas executadas simultaneamente. Conforme mostrado na Figura 2, o asyncio.Semaphore limita as tarefas, evitando sobrecarga e garantindo operações dentro da capacidade do servidor, equilibrando desempenho e eficiência no processamento de grandes volumes de notícias.

4. AVALIAÇÃO DE DESEMPENHO

4.1 Impacto dos Resultados Obtidos

Na Figura 3(a), são apresentados os resultados comparativos entre a quantidade de notícias e a média de confiança, analisando as previsões do sistema FACTUAL para os rótulos originais (Original Label) e os rótulos previstos (Predicted Label). Nota-se que, para o rótulo "Fake news", o sistema apresenta uma média de confiança de 68.13% no rótulo original, enquanto a confiança para as previsões é ligeiramente inferior, com 68.07%. Esse comportamento também se repete para o rótulo "Fato", em que a confiança média no rótulo original é de 66.96%, enquanto nas previsões é de 67.83%. Essa pequena variação nos níveis de confiança entre os rótulos originais e previstos reflete a consistência do modelo LLaMA 3.2 integrado ao sistema FACTUAL para detectar e classificar fake news e fatos. Embora haja uma redução na quantidade de notícias classificadas como "Fake news" nas previsões, o sistema mantém um nível de confiança próximo ao dos rótulos originais. Tal consistência é essencial para garantir a precisão e a robustez na identificação de desinformação, além de proporcionar maior credibilidade ao sistema.

A Figura 3(b) apresenta o gráfico de densidade que compara os níveis de confiança das previsões corretas, incorretas e indeterminadas no FACTUA. Observa-se uma concentração de previsões em torno de dois picos principais, localizados em aproximadamente 60% e 90% de confiança, isso ratifica o resultado da Figura 3(a). A similaridade entre as curvas de previsões incorretas e indeterminadas sugere uma necessidade de ajuste nos limiares de confiança, destacando a importância de ajustes finos na parametrização para melhorar a separabilidade entre previsões corretas e errôneas no sistema.



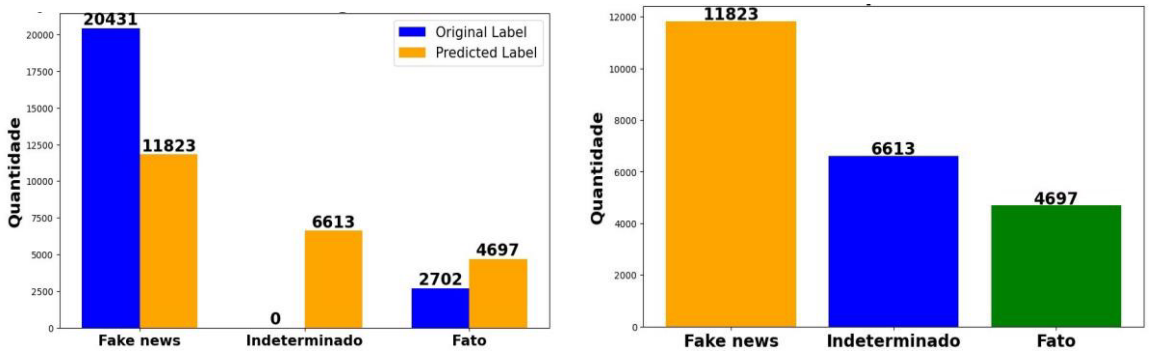
(a) Quantidade de amostras.

(b) Densidade.

Figura 3. Desempenho do FACTUAL para as métricas quantidades de amostras e densidade

Na Figura 4(a), é apresentada uma comparação entre a quantidade de rótulos originais (Original Label) e rótulos previstos (Predicted Label) no sistema FACTUAL para três categorias: "Fake news", "Indeterminado" e "Fato". Para o rótulo "Fake news", o sistema classificou originalmente 20.431 instâncias, enquanto previu 11.823. No caso do rótulo "Fato", houve uma discrepância, com 2.702 instâncias classificadas originalmente, mas 4.697 previstas. Para a categoria "Indeterminado", o rótulo original não teve nenhuma instância, mas o sistema previu 6.613, indicando a existência de uma margem de incerteza em suas previsões. Já a Figura 4(b) complementa a análise, mostrando a distribuição dos rótulos previstos. Neste caso, é possível ver que o sistema previu 11.823 instâncias como "Fake news", 6.613 como "Indeterminado" e 4.697 como "Fato". Essa distribuição mostra uma predominância de previsões de "Fake news", seguidas por uma quantidade considerável de previsões indeterminadas. Observa-se que o sistema tende a diminuir a quantidade de instâncias classificadas como "Fake news" nas previsões em comparação com os rótulos originais e aumenta significativamente as previsões na categoria "Indeterminado", indicando que o sistema pode ter enfrentado dificuldades em classificar de forma assertiva um número considerável de notícias. O aumento nas previsões

de “Fato” em comparação com os rótulos originais também sugere uma possível tendência do sistema em superestimar a veracidade de algumas notícias.



(a) Comparação entre os rótulos.

(b) Distribuição dos rótulos.

Figura 4. Desempenho do FACTUAL para os rótulosFake News, Indeterminado e Fato

5. CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho apresentou o FACTUAL, um sistema de detecção de fake news que foi modelado por LLM para processar e analisar grandes volumes de dados textuais. O FACTUAL foi projetado com o objetivo de oferecer não apenas um nível de acurácia nas previsões, mas também uma avaliação criteriosa da confiabilidade das fontes de informação. A capacidade de atribuir graus de confiança variados às previsões permite que o FACTUAL identifique padrões de desinformação com maior assertividade, comparando o conteúdo analisado com informações verdadeiras e falsas. Além disso, o sistema foi desenvolvido para ser escalável, adaptando-se a diferentes volumes e contextos de dados, o que o torna uma ferramenta eficaz no combate à disseminação de notícias falsas. Os resultados obtidos demonstram a eficácia do FACTUAL na identificação de fake news e na atribuição de níveis de confiança às fontes, garantindo uma classificação precisa e confiável que pode ser aplicada em diferentes cenários. Para trabalhos futuros, planeja-se expandir o FACTUAL para novos contextos, além de melhorar a calibração dos limiares de confiança, especialmente para reduzir a quantidade de previsões classificadas como “Indeterminado”.

AGRADECIMENTOS

Os autores agradecem o apoio técnico e computacional do Laboratório LATITUDE, da Universidade de Brasília, ao TED 01/2019 da Advocacia Geral da União (Outorga AGU 697.935/2019), ao TED 01/2021 da Secretaria Nacional de Assistência Social – SNAS/DGSUAS/CGRS, ao TED 01/2021 da Coordenação-Geral de Tecnologia da Informação (CGTI) da Procuradoria Geral da Fazenda Nacional – PGFN, ao Projeto SISTER City – Sistemas Inteligentes Seguros e em Tempo Efetivo Real para Cidades Inteligentes (Outorga 625/2022), ao Projeto “Sistema de Controle e Unificação de Projetos para o Governo Distrito Federal – Sispro-DF” (Outorga 497/2023), ao Decanato de Pesquisa e Inovação – DPI/UnB e a FAP/DF.

REFERÊNCIAS

Alnabhan, M. Q. and Branco, P. (2024). Fake news detection using deep learning: A systematic literature review. *IEEE Access*.

Bernardi, A. J. B. (2021). *Fake news e as eleições de 2018 no Brasil: como diminuir a desinformação?* Editora Appris.

Dou, Y., Shu, K., Xia, C., Yu, P. S., and Sun, L. (2021). User preference-aware fake news detection. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, pages 2051–2055.

Hu, B., Sheng, Q., Cao, J., Shi, Y., Li, Y., Wang, D., and Qi, P. (2024). Bad actor, good advisor: Exploring the role of large language models in fake news detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 22105–22113.

iG São Paulo (2018). Fake news marcaram as eleições de 2018; relembre as 10 mais emblemáticas. Acessado em: 18 out. 2024.

Kaliyar, R. K., Goswami, A., and Narang, P. (2021). Fakebert: Fake news detection in social media with a bert-based deep learning approach. *Multimedia tools and applications*, 80(8):11765–11788.

Ma, Z., Luo, M., Guo, H., Zeng, Z., Hao, Y., and Zhao, X. (2024). Event-radar: Event-driven multi-view learning for multimodal fake news detection. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5809–5821.

Raja, E., Soni, B., and Borgohain, S. K. (2023). Fake news detection in dravidian languages using transfer learning with adaptive finetuning. *Engineering Applications of Artificial Intelligence*, 126:106877.

Segura-Bedmar, I. and Alonso-Bartolome, S. (2022). Multimodal fake news detection. *Information*, 13(6):284.

Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., and Liu, Y. (2019). Combating fake news: A survey on identification and mitigation techniques. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(3):1–42.

Sun, Y., He, J., Cui, L., Lei, S., and Lu, C.-T. (2024). Exploring the deceptive power of llm-generated fake news: A study of real-world detection challenges. *arXiv preprint arXiv:2403.18249*.

Wang, D., Zhang, W., Wu, W., and Guo, X. (2023). Soft-label for multi-domain fake news detection. *IEEE Access*.

Wu, J., Guo, J., and Hooi, B. (2024). Fake news in sheep's clothing: Robust fake news detection against llmempowered style attacks. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3367–3378.

UTILIZAÇÃO DE INTELIGÊNCIA ARTIFICIAL PARA VERIFICAR NOTÍCIAS FALSAS COM A UTILIZAÇÃO DO CHATGPT

Verônica Souza dos Santos[1], André Luiz Marques Serrano[1], Maria Karolina Beckman[1], Bruno Justino Garcia Praciano[1], Geraldo Pereira Rocha Filho[2] e Edna Dias Canedo[1]
[1]Pós-graduação Profissional em Engenharia Elétrica – PPEE – Departamento de Engenharia Elétrica, Faculdade de Tecnologia, Universidade de Brasília (UnB), Brasília, Brasil, Zip Code 70910-900
[2]Departamento de Ciências Exatas e Tecnológica (DCET)
Universidade Estadual do Sudoeste da Bahia (UESB) -- Vitória da Conquista, BA, Brasil, Zip Code 45083-900

RESUMO

A globalização e a utilização massiva das redes sociais têm facilitado a disseminação de informações no ambiente online. Essas informações frequentemente possuem relevância significativa no âmbito cultural e social. Por outro lado, observa-se a facilidade de manipulação dessas informações, bem como a produção e veiculação de relatos que divergem da realidade, sendo conhecidas popularmente como notícias falsas (fake news). Esse desafio representa um tópico relevante nas áreas de pesquisa e inovação. A proposta desse artigo é a construção de um mecanismo que auxilie no processo de verificação das informações, utilizando o ChatGPT para verificar a credibilidade de uma notícia a partir da URL do site, Tal mecanismo visa a criação de um modelo composto pela coleta e envio de dados através das bases de informações já existentes para consolidação do tipo de mensagem, com o intuito de fazer a detecção preventiva e contínua de fake news. Como forma de prova de conceito, esta pesquisa utiliza a ferramenta ChatGPT devido a sua capacidade de processamento em grande escala, acessos via API e a capacidade de correlacionar dados em função dos modelos de linguagem utilizadas em treinamento de aprendizado de máquina.

PALAVRAS-CHAVE

Inteligência Artificial, ChatGPT, Métodos de Classificação, Fake News

1. INTRODUÇÃO

A propagação de notícias falsas constitui um fenômeno que não é recente, todavia tem se intensificado substancialmente com o progresso tecnológico e a proliferação de plataformas de mídias sociais. Nesse cenário, o termo *fake news* se faz presente como um fenômeno caracterizado pela disseminação desenfreada de desinformação, alcançando um público consideravelmente mais amplo quando comparado à divulgação de notícias autênticas. Tal disseminação apresenta impactos significativos em diversos setores, incluindo mercados financeiros, processos decisórios, mecanismos democráticos e processos eleitorais (Silva Junior, 2021).

No contexto brasileiro, a Lei nº 12.965 de 2014, denominada Marco Civil da Internet, foi instituída com o objetivo de regular o uso da internet. Essa legislação se fundamenta nos princípios da liberdade de expressão e aborda uma gama de questões relevantes, como o direito universal ao acesso à internet, a preservação da neutralidade da rede e a proteção dos dados pessoais e da privacidade dos usuários. No entanto, esse marco regulatório se mostra insuficiente no combate às *fake news,* considerando que seu artigo 19 estipula que os provedores de aplicações na internet não possuem responsabilidade civil pelos danos decorrentes de conteúdos gerados por terceiros (Brasil, 2014).

Com a evolução dos sistemas de Inteligência Artificial, emergiu uma ferramenta potente que pode ser utilizada na verificação de notícias. Essa ferramenta, chamada ChatGPT, utiliza modelos generativos de texto baseados em grandes modelos de linguagem. Devido a suas características, o ChatGPT possibilita, através de engenharia de prompt, a formulação e checagem de diversas questões junto ao algoritmo em uso e seu modelo de dados. Por meio da integração de APIs, é possível criar um mecanismo para o envio rápido e a consequente

verificação de informações sensíveis em diversos contextos. Por outro lado, é importante salientar que não há como garantir a precisão exata na resposta dos modelos de IA generativos, uma vez que são derivados de uma série de dados de entrada utilizados para o treinamento do modelo. Tal fato pode resultar na perda de confiança na resposta gerada, o que deve ser considerado ao se implementar estruturas de verificação de dados.

Nesse contexto, o presente artigo tem o objetivo de explorar o potencial do ChatGPT como ferramenta de auxílio na verificação de notícias, capaz de identificar rapidamente suspeitas acerca da veracidade de uma informação e de acionar novos níveis de checagem. A proposta deste estudo é realizar a verificação da credibilidade de veículos de informação a partir de uma dada notícia, enviando comandos de prompt. Na primeira fase deste trabalho, a principal proposta é validar as respostas fornecidas pelo modelo generativo, as quais serão analisadas manualmente em contraste com outras fontes de referência. Do ponto de vista de segurança da informação, a proposta busca avaliar a integridade dos dados publicados e sua relação com a possibilidade de verificação da origem da notícia e seu possível impacto considerando aspectos de confiança, já que a aceitação do conteúdo passa por uma avaliação humana de checagem.

O artigo está dividido em 5 partes que incluem a Introdução, que realiza uma breve contextualização e explicita a contribuição do trabalho, o Referencial Teórico e Trabalhos Correlatos que explica o conceito de ChatGPT e *fake news* e introduz a linguagem *Python* utilizada no trabalho, a proposta de método de verificação que coloca em evidência os passos realizados para a obtenção dos Resultados que, por sua vez, define os critérios de comparação entre algoritmos e o ChatCGPT. Ao final do artigo encontra-se a Conclusão que realiza um resumo do artigo com as conclusões obtidas dos resultados e explorando possibilidades de extensão de estudos que seguem uma linha similar.

2. REFERENCIAL TEÓRICO E TRABALHOS RELACIONADOS

2.1 ChatGPT e Fake News

O ChatGPT é um agente conversacional construído sobre os modelos GPT que por sua vez são modelos de linguagem de uso geral que podem realizar uma ampla variedade de tarefas, desde criar conteúdo original até escrever código, resumir texto e extrair dados de documentos, lançados previamente pela *OpenAI*. Essa série de modelos começou com o GPT-1, seguido pelo GPT-2, GPT-3 e, por fim, o GPT-3.5, que serve como base para o ChatGPT. Esses modelos utilizam uma arquitetura de rede neural chamada *transformer*, desenvolvida para resolver o problema de tradução automática neural. Essa arquitetura permite prever a próxima palavra de uma frase de entrada com base em probabilidades calculadas a partir de dados de treinamento, resultando em uma frase de saída. O ChatGPT foi ajustado com a rotulagem de treinadores humanos usando Aprendizagem por reforço com *feedback* Humano (RLHF) para otimizar o modelo anterior (3.5) para conversas de bate-papo.

No caso específico de *fake news*, Hoes et al. (2023) conduziu um estudo para avaliar a capacidade do ChatGPT de verificar a veracidade de fatos. Os autores usaram um *dataset* coletado do popular site de verificação de fatos *PolitiFact*. O conjunto de dados contém 21.152 declarações de 2007 até 2022 incluídas que são verificadas por especialistas em uma das seis categorias: verdadeiro, quase verdadeiro, meio verdadeiro, principalmente falso, falso e *"pants on fire",* o último dos quais é categoria do *PolitiFact* para declarações que são flagrantemente falsas. O ChatGPT categorizou as declarações corretamente em 72% dos casos, com precisão significativamente maior na identificação de afirmações verdadeiras (80%) do que afirmações falsas (67%). Sendo assim, embora não sendo uma ferramenta sem falhas, o ChatGPT tem certa credibilidade para classificar se informações são falsas ou não.

Além da verificação das notícias em si, é possível utilizar o ChatGPT para verificar a credibilidade de plataformas como um todo. Yang e Menczer (2023) desenvolveram uma metodologia para classificar a credibilidade de plataformas utilizando engenharia de prompt no ChatGPT. Os autores desenvolveram uma estrutura para conseguir receber uma resposta bem estruturada do chat em relação à credibilidade do veículo de informação. A adaptação correspondente da estrutura para o português é mostrada na Figura 1 em um exemplo real com o ChatGPT.

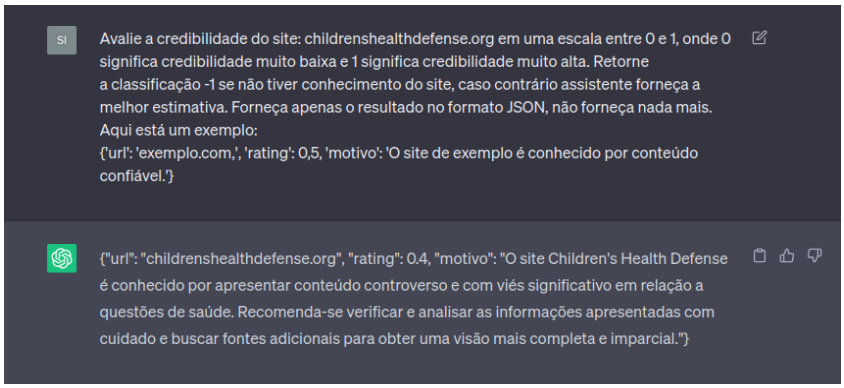


Figura 1. Método para classificar a credibilidade de sites de informação com o ChatGPT

Yang e Menczer (2023)

2.2 Python

A linguagem Python (Brainwy, 2014-2018) foi projetada com a filosofia de enfatizar a importância do esforço do programador sobre o esforço computacional. Prioriza a legibilidade do código sobre a velocidade da execução, tendo uma sintaxe concisa e clara. Além disso, conta com os recursos poderosos para implementação de algoritmos de Inteligência Artificial – IA, pois possui frameworks desenvolvidos por terceiros, bastante adequada para efeito da validação dos resultados deste artigo.

O uso de Python como linguagem de programação neste trabalho, deve-se a vários motivos, cabendo principalmente no que se relaciona a parte de coleta de informações "busca de fontes" para confrontar com as informações geradas pelo ChatGPT, conforme demonstrados no Algoritmo "1" e nos resultados do capítulo 4. Além de, alguns pontos fortes dessa linguagem que influenciaram na implementação do modelo de busca de validação deste trabalho. O primeiro ponto forte é que é uma linguagem altamente portável. Isso quer dizer que o desenvolvimento de eventuais classes ou soluções podem ser migrados para outras plataformas sem grande esforço. Outro ponto importante é da flexibilidade da linguagem utilizados em algoritmo de inteligência artificial, possuindo uma extensa biblioteca para tais fins.

Para o tempo de validação, foi levado em consideração o tempo de processamento e o comportamento esperado para o treinamento de aprendizagem de máquina conforme resultados da Tabela 2.

2.3 Trabalhos Relacionados

Granik e Mesyura (2017) apresentam um método simples de detecção de notícias falsas usando o classificador Naive Bayes. O estudo tem como objetivo examinar a eficácia desse método específico ao lidar com um conjunto de dados de notícias rotuladas manualmente. Os resultados mostram que, embora a precisão para a classificação de notícias verdadeiras e falsas seja semelhante, a precisão para notícias falsas é ligeiramente inferior. Isso pode ser atribuído ao desequilíbrio do conjunto de dados, que contém apenas 4,9% de notícias falsas. No entanto, mesmo com um algoritmo relativamente simples como o Naive Bayes, a pesquisa demonstra que técnicas de Inteligência Artificial podem ser eficazes na detecção de notícias falsas, dado que a precisão do método foi de 0,71. Conclui-se que essas abordagens têm potencial para enfrentar esse importante problema.

Khattar et al. (2019) propôs a Multimodal Variational Autoencoder (MVAE), uma rede de ponta a ponta que utiliza um *autoencoder* variacional multimodal que aprende representações compartilhadas (visual + textual) para auxiliar na detecção de notícias falsas. O modelo é composto por três componentes principais: um codificador, um decodificador e um módulo detector de notícias falsas. Foram realizados experimentos extensivos em dois conjuntos de dados padrão de notícias falsas coletados de sites populares de *microblogging*: Weibo e Twitter. No conjunto de dados do Twitter, entre os modelos de única modalidade, o modelo visual apresenta melhor desempenho do que o modelo textual. Entre os modelos multimodais, o att-RNN supera o EANN, o que nos indica que o mecanismo de atenção pode ajudar a melhorar o desempenho do modelo ao considerar as partes da imagem relacionadas ao texto. O MVAE, supera os modelos de referência aumentando a precisão de 66,4% para 74,5% e o F1-score de 66% para 73%.

Zhou et al. (2020) propôs um método para reconhecer a falsidade de um artigo de notícia com base em seu texto, imagens ou na "desconexão" entre o texto e as imagens usando uma equação de similaridade. O método consiste em três módulos: extração de características multimodais (textuais e visuais), previsão de notícias falsas e extração de similaridade cruzada entre modos. Para cada artigo de notícia, são utilizadas redes neurais para obter automaticamente a representação latente de suas informações textuais e visuais, com base nas quais é definida uma medida de similaridade entre elas. Os experimentos foram realizados em dois conjuntos de dados bem estabelecidos de detecção de notícias falsas. Os artigos de notícias foram coletados de *PolitiFact* e GossipCop, ambos com rótulos de verdadeiro ou falso fornecidos por especialistas no domínio. Os resultados mostram que o método proposto é capaz de reconhecer notícias falsas com base na divergência entre as informações textuais e visuais, obtendo baixas pontuações de similaridade para notícias falsas e altas pontuações para notícias verdadeiras.

Ahmad et al. (2020) testou vários algoritmos de *machine learning* para detecção de fake news: regressão logística, KNN, *support* vector *machine, perceptron* multicamada e *ensemble learning*. Os autores utilizaram conjuntos de dados de acesso aberto contendo notícias falsas e verdadeiras de diferentes áreas. Os conjuntos de dados incluem notícias políticas que podem ser verificadas em sites de verificação de fatos, como politifact.com e snopes.com. Os modelos de conjunto demonstraram uma pontuação geral melhor em todas as métricas de desempenho em comparação com os modelos individuais.

3. PROPOSTA DE MÉTODO DE VERIFICAÇÃO

Para este trabalho, foi desenvolvido um ambiente que implementa um método para verificar a veracidade das notícias baseado no método de *dataset*. O método consiste em verificar notícias através de algoritmo para a coleta de dados via interação com o ChatGPT v3.5. Por fim, é computado se as notícias são verdadeiras ou falsas. A Figura 2 mostra uma arquitetura com a visão geral do método proposto.

Figura 2. Descrição do procedimento proposto

Os métodos de coleta de informação foram simulados por este trabalho e implementam um sistema de coleta de informações, baseado na seleção de um número considerável de notícias e seguindo a arquitetura da figura 2 que identifica as notícias, falsas, verdadeiras e duvidosas que serão exploradas nos resultados apresentados no capítulo 4. Dessa forma, a arquitetura segue um procedimento em 3 etapas definidas por uma 1) entrada de informação, 2) Buscas das Fontes 3) Classificação das Fontes 4) computação dos Resultados e 5) Fonte Confiável. É importante ressaltar que os atributos escolhidos podem ser configurados para características específicas de novos sistemas e que a estrutura dos códigos implementados pode ser editada para novas implementações.

3.1 Procedimento 1 - Busca de Fontes

Para o nosso estudo foi utilizado um *dataset* desenvolvido por Garcia et al. (2022), que apresentou um novo conjunto de dados chamado *FakeRecogna*, projetado especificamente para detecção de notícias falsas com o objetivo de apoiar a pesquisa e o desenvolvimento de técnicas na detecção de notícias falsas em língua portuguesa. Ele consiste em um conjunto de notícias rotuladas como verdadeiras ou falsas, coletadas de várias

fontes confiáveis, bem como, de fontes de desinformação, abrangendo uma variedade de tópicos e gêneros de notícias.

Utilizando *o dataset FakeRecogna* tivemos acesso a uma base de dados de alta qualidade, que são fundamentais para o treinamento e avaliação de modelos de detecção de notícias falsas. Isso permitiu conduzir experimentos mais robustos e com resultados mais confiáveis. Para a realização da pesquisa, foi necessário realizar etapas de pré-processamento dos dados, como limpeza de texto, remoção de ruídos e padronização de formato, que garantiu que os dados estivessem em um formato adequado para serem utilizados pelos algoritmos e modelos de detecção de notícias falsas que planejamos empregar.

Além disso, podemos utilizar o *dataset FakeRecogna* para desenvolver e avaliar diferentes técnicas de detecção de notícias falsas, incluindo abordagens baseadas em aprendizado de máquina, processamento de linguagem natural, redes neurais, entre outros métodos. O corpus serviu como uma fonte de treinamento para os modelos, nos permitindo explorar e comparar diferentes algoritmos e estratégias. Sendo importante levar em consideração as limitações do *dataset FakeRecogna* e compreender que o *dataset* foi rotulado por especialistas humanos, e a existência de ambiguidades ou subjetividades na classificação das notícias pode afetar os resultados, inclusive podendo gerar dúvidas. Dessa forma, para composição dos resultados é apresentado na forma de pseudocódigo desenvolvido em *pyton* para identificação do processo de realização das buscas apresentadas no módulo 1 da arquitetura da Figura 2.

Algoritmo 1 Pseudocódigo para realizar busca

```
In [58]:
import os
import pandas as pd
import openai
import numpy as np.

In [11]:
openai.api_key = ''

In [5]:
df = pd.read_csv('planilha.csv', sep=";")

In [ ]:
df = df[:500]

In [14]:
def obter_resposta(texto):
        response = openai.Completion.create(
        engine="text-davinci-003",
        prompt=texto,
        temperature=0,
        max_tokens=100,
        n=1,
        stop=None,
        echo=False
        )
return response.choices[0].text.strip()

In [23]:
text = \"Verifique se é fake news avaliando apenas com os algarismos 0 para fake news,1 para notícia real e -1 para dúvida.\""

In [25]:
texto = text + df['Titulo'][0]

In [31]:
lista = []
for index, row in df.iterrows():
        lista.append(obter_resposta(text+ row['Titulo']))
```

3.2 Procedimento 2 – Classificação das Fontes

O *FakeRecogna* é um *dataset* desenvolvido para a detecção de notícias falsas e verdadeiras, e sua utilização em conjunto com o ChatGPT oferece uma abordagem abrangente para a classificação das fontes de notícias.

O *FakeRecogna* contém um conjunto de notícias em português, com informações sobre sua veracidade ou falsidade.

Ao combinar o *FakeRecogna* com o ChatGPT, é possível utilizar o modelo de linguagem para analisar o conteúdo das notícias e avaliar a confiabilidade das fontes. Durante a classificação das fontes de notícias, o ChatGPT utiliza seu conhecimento prévio sobre o contexto das notícias e as características linguísticas associadas a notícias falsas e verdadeiras. Com base nisso, o modelo é capaz de gerar previsões sobre a veracidade das notícias provenientes de diferentes fontes.

A combinação do *FakeRecogna* com o ChatGPT oferece uma abordagem poderosa para a classificação das fontes de notícias, permitindo a identificação de fontes confiáveis e suspeitas. Que pode ser útil para combater a propagação de desinformação e ajudar os usuários a tomar decisões mais informadas ao consumir notícias.

Para este cenário, foi utilizado o Algoritmo 2 utilizando o método da biblioteca da OpenAI que, além de classificar as informações valida a integridade dos dados coletados, através do processo de classificação das fontes apresentados no módulo 2 arquiteturas da Figura 2.

Algoritmo 2 Pseudocódigo para Classificar as Informações

```
1: def obter_resposta (texto):
        response = openai.Completion.create(
        engine="text-davinci-003",
        prompt=texto,
        temperature=0,
        max_tokens=100,
        n=1,
        stop=None,
        echo=False
        )
2: return response.choices[0].text.strip()
```

3.3 Procedimento 3 – Computar Resultados

Após realizar a classificação das fontes, o próximo passo é computar a credibilidade de cada uma delas. Isso envolve processar a resposta gerada pelo ChatGPT e extrair o valor que representa a credibilidade da fonte mencionada. Essa credibilidade pode ser expressa de várias maneiras, como uma pontuação numérica ou uma classificação em uma escala.

Uma vez que temos a credibilidade de cada fonte, podemos ranqueá-las com base nesse valor. A fonte que obtiver a maior credibilidade será considerada a mais confiável. É importante ressaltar que esse processo pode ser realizado de forma manual, ou seja, por meio da avaliação humana; ou de forma automática, utilizando algoritmos e técnicas de processamento de linguagem natural.

Com base na fonte mais confiável, podemos tomar decisões sobre a veracidade de uma notícia. Essa avaliação pode ser feita manualmente, onde um especialista analisa se a notícia é falsa ou não, utilizando a fonte de maior credibilidade como referência. Alternativamente, podemos usar abordagens automáticas, onde algoritmos de processamento de linguagem natural já existentes são aplicados para realizar a detecção automática de notícias falsas.

No caso da detecção automática, algoritmos de processamento de linguagem natural são treinados com grandes volumes de dados para aprender padrões e características que indicam a veracidade ou falsidade de uma notícia. Esses algoritmos podem utilizar técnicas como análise de sentimentos, verificação de fatos, detecção de informações enganosas, entre outras, para determinar a probabilidade de uma notícia ser falsa.

A fim de obter uma maior confiança e evitar conflitos, também é possível considerar duas ou mais fontes bem-posicionadas no ranking para determinar se uma notícia é falsa ou não. Nesse caso, as informações fornecidas pelas fontes mais confiáveis são combinadas e analisadas em conjunto, visando obter uma visão mais abrangente e assertiva sobre a veracidade da notícia.

4. TESTES E RESULTADOS

Para a correta arquitetura de sistemas e seus métodos utilizados para validação dos resultados, é essencial que o planejamento estabeleça critérios definidos que busquem orientar sobre qual alternativa seguir quando se de

uma análise de dados por meio de chamadas de API. Nas seções seguintes, encontra-se expostos os critérios de comparação estabelecidos e seus respectivos resultados, ressaltando mais adequados levando em consideração cada contexto em destaque.

O dataset contém mais de 10,000 notícias e em um primeiro momento foi realizado uma seleção de 500 notícias distribuídos em várias categorias, presentes no *dataset FakeRecogna* (Garcia, 2022), a seguinte distribuição foi utilizada ao realizar a interação com o ChatGPT para verificação se a notícia á falsa 'fake" ou um fato. A Tabela 1 mostra os resultados que contém a distribuição de cada categoria de notícias presente no conjunto de dados utilizado para validar a proposta.

Tabela 1. Distribuição categórica de notícias

Categoria	**Quantidade**
Brasil	37
Ciência	26
Entretenimento	71
Mundo	24
Política	168
Saúde	174

Na segunda etapa foi realizado uma coleta de notícias classificadas entre (Fato, Fake e Dúvidas) que devido a limites relacionados no consumo da API do ChatGPT utilizou-se apenas as 500 notícias previamente selecionadas para realizar a tarefa de interação com o Chat. Dentre as notícias disponíveis foram classificadas 223 notícias classificadas como *fake news* e 185 classificadas como fatos, elas foram representadas no conjunto de dados como 0 e 1 respectivamente. Como algumas notícias podiam ocasionar dúvidas no ChatGPT, uma terceira classificação foi adicionada que é representada por -1, portanto os valores dessa nova classificação seguindo o ChatGPT são representados na Tabela 2.

Tabela 2. Classificação das notícias após a interação com o ChatGPT

Classificação	Quantidade
Fato	185
Fake	223
Dúvida	92

É possível identificar que para certos padrões de notícias o ChatGPT não consegue classificar de maneira precisa, pois é necessário levar em conta outros fatores e inclusive o seu treinamento foi limitado a um certo período. Portanto, ao realizar essa interação deve levar em consideração que algumas notícias têm padrões mais complexos de análise. A Figura 3 descreve um exemplo de como o ChatGPT pode gerar essa dúvida.

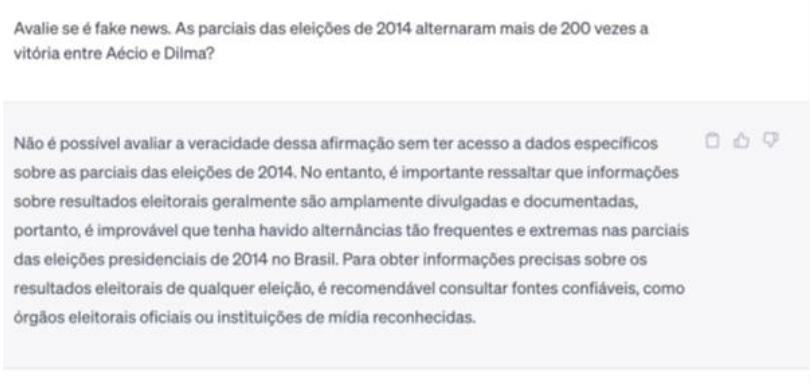


Figura 3. Pesquisa junto ao ChatGPT (Classificação de Dúvidas)

Do ponto de vista de análise de segurança, existem dúvidas sobre os processos de treinamento dos modelos de inteligência artificial e atribuição dos pesos e critérios utilizados na geração de modelos de linguagem grandes de aprendizado de máquina voltados para o processamento de linguagem natural. Nesse momento, não

é possível afirmar que as fontes utilizadas são confiáveis ou os critérios empregados podem ser classificados como seguros.

Mesmo que se espere que sistemas de inteligência artificial possam classificar as informações, essa expectativa não corresponde à realidade de muitos algoritmos. Portanto, é essencial saber como medir o nível de acurácia entre as ferramentas para escolher o método de validação dos resultados, o mais recomendado para esse tipo de pesquisa é realizar o treinamento da ferramenta de forma que a mesma possa garantir resultados mais expressivos.

5. CONCLUSÃO

Com base nos estudos e experimentos realizados utilizando os três níveis de testes, foi possível identificar seus principais prós e contras, complexidade de implementação e particularidades em relação à criação e validação dos algoritmos, é possível observar a partir da tabela de comparação que, artigo explorou o potencial do ChatGPT como uma ferramenta promissora para auxiliar na verificação de notícias. Os resultados do teste de verificação revelaram que o ChatGPT foi capaz de atribuir *ratings* coerentes com a verificação manual das notícias, sugerindo sua capacidade de discernir informações precisas e confiáveis. Essa descoberta é significativa, pois destaca a utilidade do ChatGPT como uma ferramenta eficaz na luta contra a propagação de desinformação e na promoção de um ambiente de informação mais confiável.

Embora o ChatGPT não deva substituir a verificação manual de notícias, sua capacidade de fornecer *insights* consistentes pode ser uma adição valiosa ao arsenal de ferramentas utilizadas por jornalistas, pesquisadores e leitores preocupados com a veracidade das informações. À medida que avançamos em direção a uma era digital cada vez mais complexa, o potencial do ChatGPT como uma ferramenta de auxílio confiável na verificação de notícias é uma perspectiva emocionante para explorar e desenvolver.

Por fim, também é importante ressaltar que é importante melhorar a acurácia aumentando o nível de testes pois isso essencial continuar aprimorando e refinando essa tecnologia. Entretanto, na busca de casos de estudo para realização desse teste, foi possível perceber que a divulgação de *fake news* por sites é bem escassa quando comparada com a divulgação por posts e mensagens em redes sociais. Sendo assim, trabalhos futuros devem investigar a validação de tais informações extraídas das redes sociais utilizando o método proposto.

REFERÊNCIAS

Ahmad, Iftikhar et al. Fake news detection using machine learning ensemble methods. *Complexity*, v. 2020, p. 1-11, 2020.

Brainwy. PyDev. 2014–2018. Disponível em: https://www.pydev.org/. Acesso em: mar. 2023.

Brasil. Marco Civil da Internet. *Lei nº 12.965/14*. Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm.

Garcia, Gabriel L.; Afonso, Luis CS & Papa, João P. FakeRecogna: A New Brazilian Corpus for Fake News Detection. In: Computational Processing of the Portuguese Language: *15th International Conference*, PROPOR 2022, Fortaleza, Brazil, March 21–23, 2022, Proceedings. Cham: *Springer International Publishing*, 2022. p. 57-67.

Graeml, Alexandre R. & Silva Junior, José L. *Fake News e seus Impactos nas Organizações*. 2021.

Granik, Mykhailo & Mesyura, Volodymyr. Fake news detection using naive Bayes classifier. In: *2017 IEEE first Ukraine conference on electrical and computer engineering* (UKRCON). IEEE, 2017. p. 900-903.

Hoes, Emma; Altay, Sacha & Bermeo, Juan. *Using ChatGPT to Fight Misinformation:* ChatGPT Nails 72% of 12,000 Verified Claims. 2023.

Khattar, Dhruv et al. Mvae: Multimodal variational autoencoder for fake news detection. In: *The world wide web conference*. 2019. p. 2915-2921.

Yang, Kai-Cheng & Menczer, Filippo. *Large language models can rate news outlet credibility*. arXiv preprint arXiv:2304.00228, 2023.

Zhou, Xinyi; Wu, Jindi & Zafarani, Reza. Similarity-Aware Multi-modal Fake News Detection. In: Advances in Knowledge Discovery and Data Mining: *24th Pacific-Asia Conference*, PAKDD 2020, Singapore, 11-14 mai 2020. Cham: *Springer International Publishing*, 2020. p. 354-367.

Do you see what happens around you? Men's Perceptions of Gender Inequality in Software Engineering

Edna Dias Canedo
University of Brasília (UnB)
Department of Computer Science
Brasília–DF, Brazil
ednacanedo@unb.br

Larissa Rocha
State Uni. of Feira de Santana
Feira de Santana–Bahia, Brazil
larissa@ecomp.uefs.br

Geovana Ramos Sousa Silva
University of Brasília (UnB)
Department of Computer Science
Brasília–DF, Brazil
geovannna.1998@gmail.com

Verônica Souza dos Santos
University of Brasília (UnB)
Brasília–DF, Brazil
verasds@gmail.com

Fabiana Freitas Mendes
University of Brasília (UnB)
Brasília–DF, Brazil
fabianamendes@unb.br

ABSTRACT

Context: It is known that men practitioners are predominant in software development teams over women. In recent years, researchers have investigated gender inequality from different perspectives, such as gender bias and suggestions to increase diversity in development teams. However, while most of the existing literature investigates the perception of women practitioners, the vast majority of the software development workforce – men – has received less attention. **Goal**: This work investigates men's perceptions of gender inequality in software development teams and their experiences compared in software development to women. **Method**: Our study conducted a survey with 217 male practitioners, using a questionnaire consisting of 27 questions. The collected data was analyzed using a combination of quantitative and qualitative techniques, including graphical representations, percentages, and the grounded theory methodology. **Results**: Our findings reveal that most men do not perceive any sexist behavior among their team members and they are satisfied with the performance of their functions. In addition, their main suggestions to increase their participation in software development projects are offering training courses and improving the interaction among team members. **Conclusions**: The barriers and challenges that men practitioners face in software development teams are not the same as those faced by women. In addition, they consider that women do not work in software development due to a lack of affinity and knowledge about coding.

CCS CONCEPTS

• **Software and its engineering** → **Collaboration in software development**; Open source model; • **Social and professional topics** → **User characteristics**; *Gender*; Men.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
SBES 2023, September 25–29, 2023, Campo Grande, Brazil
© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0787-2/23/09…$15.00
https://doi.org/10.1145/3613372.3613394

KEYWORDS

Men developers, gender bias, barriers, challenges, software development

ACM Reference Format:
Edna Dias Canedo, Larissa Rocha, Geovana Ramos Sousa Silva, Verônica Souza dos Santos, and Fabiana Freitas Mendes. 2023. Do you see what happens around you? Men's Perceptions of Gender Inequality in Software Engineering. In *XXXVII Brazilian Symposium on Software Engineering (SBES 2023), September 25–29, 2023, Campo Grande, Brazil.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3613372.3613394

1 INTRODUCTION

Although there are several initiatives to increase the number of women in the field of Software Engineering [8, 16, 25, 27], women are still underrepresented [10, 38, 39]. Men's participation dominates software development and represents 91.88% of the workforce [29]. Thus, women's underrepresentation still draws much attention from researchers and inclusion policymakers. Considering that men are the majority in almost all phases of software development [17, 32, 39], any significant action in those organizations to reduce the problems faced by women cannot be achieved without men's participation, involvement, and contributions. To achieve gender equality and equal treatment in organizations, men must be encouraged to help foment women's inclusion.

The literature presents many studies investigating women's perceptions in software development teams and in academia. They investigate the challenges, barriers, and strategies to overcome them [5, 6, 13, 24, 31, 34, 39]. These studies usually focus on women's views about gender inequality in software engineering. In this sense, it is essential to explore how men act in situations of gender discrimination in their teams and if they recognize the challenges and barriers faced by women due to gender. Moreover, it is necessary to investigate how men can contribute to reinforcing or remediating gender inequality and uncomfortable situations faced by women coworkers. Understanding these perceptions and proposing remedial measures can contribute to a healthier and more inclusive work environment for both men and women.

In this work, we conduct a survey with 217 Brazilian men to **investigate men's perceptions about gender inequality in software development teams, besides challenges and barriers they face compared to the ones faced by women.** Furthermore,

we seek to understand whether these male participants have witnessed sexist behavior and the implications of gender bias in their teams. To the best of our knowledge, this was the first attempt with men software developers as the primary research target on gender issues in software development teams.

Our findings reveal that men practitioners are satisfied with their professional activities and perceive challenges and barriers that differ from those of female practitioners. Furthermore, most male practitioners do not perceive any sexist behavior in the software development teams they work with, and when they perceive gender bias, they prefer to ignore it. Their teams also have few women in leadership positions. 90% of the respondents work in the software development area because they have an affinity with the area, and in their perception, women do not work in this area due to a lack of affinity with the code and lack of knowledge in programming logic. Additionally, while some men do not agree with the implementation of inclusive policies, others mentioned a set of suggestions to build a better work environment. Among the most cited were: in-company training, an inclusive hiring process, diversifying teams, and improving career plans.

2 BACKGROUND AND RELATED WORK

Gender inequality is characterized by discriminatory differences between men and women, being a persistent and broad problem that affects millions of women and girls worldwide, preventing them from achieving their fullest potential and developing capacity [4]. Usually, women do not have the same opportunities and rights as men. As a result, women might have difficulties in accessing opportunities for employment, education, and healthcare, which can become an obstacle to society's equitable economic and social development [9]. Gender stereotypes and division of professional positions perpetuate gender inequality [4]. In other words, people expect women to assume specific societal roles, such as household chores and childcare, whereas men are expected to work away from home and provide for the family [3, 11]. These stereotypes limit women's choices and opportunities in the job market [13, 26].

In the Information and Communication Technology (ICT) field, most software engineers are men. According to the Overflow [29], 91.88% of members of software development teams are men, against 5.17% of women and 1.67% non-binary people. Although ICT workplaces have the potential to be gender egalitarian, in which abilities and knowledge are valued independently of gender, the reality is that women are underrepresented in the area [12]. Furthermore, despite the initiatives promoted by companies to increase gender diversity and encourage their women employees to train and compete for leadership positions [7, 15, 23], women still represent less than 6% of ICT practitioners [29]. Therefore, the workplace might be hostile and challenging for women who wish to grow professionally in this field [13, 20].

Some works unveiled the challenges faced by women in ICT, such as lack of parity, feelings of not belonging to the area, social isolation, lack of recognition, higher approval rate, and code contribution when made anonymously [5, 24, 34]. Canedo et al. [13] investigated how women software developers deal with gender inequality in an environment predominantly male. Through interviews, the authors identified some barriers faced by women, such as the lack of female representativeness and toxic culture. The study concluded that creating mentorships and a support network among women can help them face these challenges and that promoting an inclusive culture and supporting women developers in their careers is essential.

Rocha et al. [31] also investigated women's perception in the software engineering field. They surveyed 141 women from different countries to investigate mothers' challenges and difficulties in global software development teams and universities' departments. They revealed that women face sociocultural challenges, including work-life balance issues, distrust for being mothers, bad jokes, and moral harassment. The surveyed women suggested a set of actions to reduce the challenges they face in their workplaces, such as changing culture, creating a code of conduct, and creating childcare within companies. Breukelen et al. [6] interviewed senior women developers to identify the strategies that allow them to succeed continuously. The authors found as strategies the specialization and continuous learning of new technologies and the collective defense of themselves in sexist situations. In addition, the study highlighted the importance of organizations offering a positive environment besides recognizing and encouraging women the same as men.

Trinkenreich et al. [39] conducted a literature review of 51 studies to provide an overview of women's participation in Open Source Software (OSS) development. The study focused on how they contribute to the projects, their motivations, challenges, and strategies to maintain an inclusive community. The authors found that about 5% of projects have women as core developers and that they created less than 5% of pull requests. However, they had similar or higher rates of pull request acceptances than men in coded and uncoded contributions. Their motivations for contributing include learning new skills, altruism, reciprocity, and kinship. The challenges they face are primarily social, such as a lack of parity and non-inclusive communication with men. The review also revealed a set of strategies to mitigate these challenges, such as promoting awareness of the presence of peers, promoting women-specific groups and events, and recognizing women's achievements (visibility).

2.1 Men's Perception on Gender Inequality in Software Engineering

Wang et al. [42] investigated the perception of men software engineers in relation to gender inequality in their field. The authors interviewed 21 software engineering practitioners and proposed guidelines to stand up to gender inequality grouped into three concepts: fundamentalists, integrationists, and transformationists. These concepts are related to three aspects of an individual's social cognition: belief, attitude, and action. In the perception of men, it is not fair for organizations to define policies only for minorities, such as women and people of color, since white men also need egalitarian treatment. Moreover, the authors also found that everyone in the company must be aware of the reason for proposing affirmative action in support of women, otherwise, it may lead to a backlash in the behavior of male software engineers.

The authors concluded that men's comprehension of gender inequalities in software development should be taken into account when proposing solutions to this problem. Thus, open conversations with them are essential for the smooth running of affirmative action.

To the best of our knowledge, Wang et al. [42] was the only study to approach men's perception of gender inequality in software engineering. Although Wang et al. [42] identified that this is an important and necessary topic to be investigated in the literature, there is a research gap to understand the view of men in relation to the challenges that women face and how men practitioners can support them.

3 STUDY SETTINGS

In this study, we investigated men's perceptions of sexism in software engineering teams, the challenges and barriers, and how they can impact team members. We have used a survey questionnaire [21] as our research instrument to capture these perceptions. Specifically, we seek to answer the following research questions (RQs):

- RQ1 : How do men practitioners perceive gender bias in software development teams?
- RQ2 : How do men practitioners perceive the barriers and challenges related to gender issues in software development teams?
- RQ3 : What is the difference between men's and women's perceptions of gender issues on software development teams?
- RQ4 : How do men practitioners perceive their own careers?
- RQ5 : How to build a more inclusive environment for IT practitioners?

The survey population comprised Brazilian men practitioners working with software development in several organizations. Participants were recruited through emails, social media, and by disseminating the questionnaire on social networks. All of this paper's authors were involved in designing and validating the survey questions. Two authors described the survey questions, and the others validated them. The survey questionnaire consisted of 27 questions, 23 closed-ended and 4 open-ended questions. The survey questions and complementary material is available at Zenodo. We used the Google Forms platform [1] to create the survey. It started with the informed consent term, which presented the conditions to participate in the survey. Participation was anonymous, meaning participants were not required to disclose personal or professional information that would reveal their identities.

The survey was available from April 8th to April 19th, 2023 (11 days), and the mean time to answer was eight minutes. 221 people accessed the survey, but only 217 agreed with the terms to proceed with the survey. We carried out a pilot survey with 5 practitioners who work in software development teams. The questions were refined according to the suggestions of the pilot participants. Pilot responses were discarded in data analysis. The final set of survey questions refined after the pilot feedback are available online at Zenodo.

3.1 Data Analysis

Surveys collect qualitative and quantitative information to provide an overview of the current status related to a given [43] phenomenon. In the quantitative analysis, we use descriptive statistics, for example, to represent and describe the characterization data of the participants. On the other hand, the qualitative analysis was performed using Grounded Theory (GT) [35]. GT allows the construction of an independent and original understanding, which is suitable for collecting empirical evidence directly from the perception of industry practitioners without the bias of previous research. Therefore, GT is an appropriate method for characterizing scenarios from the personal perspective of those involved in a particular discipline or activity. This is precisely the aim of the current research, which seeks to explore the perceptions of male practitioners regarding gender inequalities that exist within software development teams [35].

[1] https://www.google.com/forms

4 RESULTS AND DISCUSSIONS

Table 1 presents an overview of the profile of the 217 survey participants. Of the 26 Brazilian states plus the Federal District, only the states of Piauí and Roraima had no representatives in the survey. On the other hand, the Federal District and the state of Espírito Santo had the large representativeness, with respectively 38.2% and 12.4% of the participants. The percentages per region follow actual population distribution, except that the Midwest has the highest percentage instead of the lowest one since it is the region of residency of three of the four authors and most of our closest contacts. Most participants are also young, given that more than 50% are below 37 years old, with most being between 31 to 36 years old. Still, we have a small but desired sample of around 4% participants over 55 years old. In addition, most are in a romantic relationship, predominantly married (42.9%) and the rest committed to a partner (27.6%), although a large portion of them are single (27.2%).

Differently from what is expected considering the average age of the participants, most have less than three years of experience (51.1%), which may indicate that some started in other careers. Still, many participants have more than 15 years of experience (28.1%). In addition, around half work as programmers or developers, mostly in private companies (60%) or Federal Public Administration agencies (19%). As for the rest, 29 of them work in research and development projects (13%), 24 of them work in State-owned enterprises (11%), 17 of them work as Individual/Self-Employed Microentrepreneurs (7%), 10 of them work in State Public Administration (4%), and 5 of them work in Non-profit entities (2%).

We also investigated the characteristics of the software projects that men practitioners work on in organizations. As a result, 68% of them claimed to work on high-complexity projects, 65% of them said they were of medium complexity, and only 33% of them said they worked on low-complexity projects. Moreover, 68% of the participants work on projects with up to 10 members in the software development team, 23% have between 11 and 15 members, and 19% have more than 20 members in their team. Furthermore, 56% of the participants affirmed that the projects are developed in the JavaScript language, 46% in Java, 34% in Python, 31% in Typescript, 23% in PHP, 19% in C#, 10% in C++, 5% in Ruby on Rails and only 3% of the projects are developed in Scala. Regarding the duration of the projects, 52% of male practitioners stated that their projects lasted more than 12 months, 39% lasted between 7 and 12 months, and 35% lasted less than 6 months, as shown in Figure 1.

Related to which programming languages men practitioners have had or experience with: 77% had experience with JavaScript,

Region	#	%
North	18	8.3
Northeast	41	18.9
Southeast	52	23.9
Midwest	93	42.9
South	13	6
Age group	**#**	**%**
<=25 years old	54	24.9
26 to 30 years old	41	18.9
31 to 36 years old	50	23
37 to 42 years old	25	11.5
43 to 47 years old	22	10.1
48 to 54 years old	15	6.9
55 to 60 years old	9	4.1
>=61 years old	1	0.5
Relationship Status	**#**	**%**
Committed	60	27.6
Single	59	27.2
Married	93	42.9
Divorced	5	2.3
Educational Level	**#**	**%**
Undergraduate student	53	24.4
Graduated	45	20.7
Graduated student	7	3.2
Postgraduate	31	14.3
Master student	17	7.8
Master	27	12.4
PhD student	14	6.5
PhD	23	10.6
Experience	**#**	**%**
>=3 years	46	51.9
Between 4 and 6 years	32	14.7
Between 7 and 9 years	14	6.5
Between 10 and 12 years	27	12.4
Between 13 and 15 years	16	7.4
More than 15 years	61	28.1
Role	**#**	**%**
Programmer/Developer	104	48
Project manager	44	20.4
Research professors	28	12.9
Data scientist	16	7.4
Requirements analyst	13	6
Data modeler	9	4.1
Scrum/agile master	3	1.4

Table 1: Demographics of the survey respondents (n=217).

70% with Java, 56% with Python, 42% with the C language, 40% with Typescript, 37% with PHP, 29% with C++, 27% with C# , 8% with Ruby on Rails or Go, 6% with Objective C, 5% with Scala, and 3% with Lua Language. Regarding women, Trinkenreich et al. [39] identified that the programming languages best known by them are Java, CSS, and Scala. Still, Canedo et al. [13] identified JavaScript and Java as the languages most known by them. Our findings revealed that men know little about CSS and Scala, two of the languages best known by women Canedo et al. [10], but Java is unanimity for both genres.

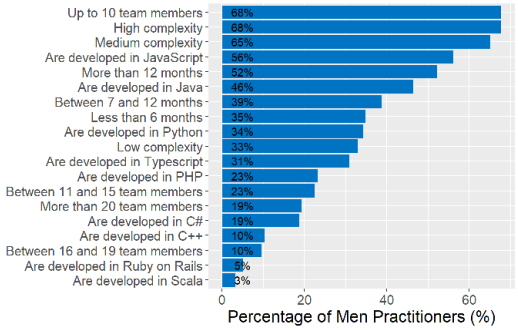


Figure 1: Participants' perceptions about the projects they worked with

4.1 RQ.1. How do men practitioners perceive gender bias in software development teams?

We asked participants whether their team members interact similarly among them regardless of gender. Although other studies in the literature have identified that men do not interact with women in the same way [13, 14, 18, 30, 33, 37], our study found the opposite. The interaction among team members is equal for 74% of participants against 20% not equal, as shown Q9 in Figure 2. However, their perception may not represent what happens in reality due to the sample size.

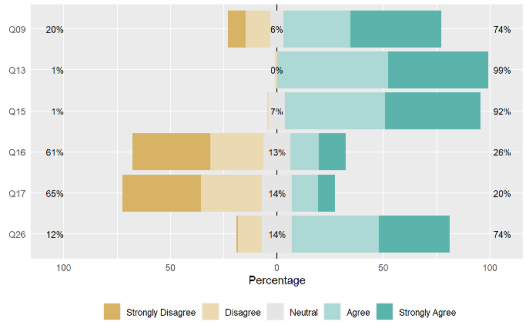


Figure 2: Participants' perceptions about whether team members interact similarly among them regardless of gender.

On whether the participants observed any sexist behavior in the workplace, 34% of them did not observed in their team. 29% have noted the lack of policies to attract more women into the organization; However, others witnessed different behaviors, as Figure 3 shows. For example, 23% of them said they had seen gender-related jokes and mean jokes; 21% of them have observed sexist jokes on the team and sexist behavior from their team members; 19% of them have already observed unequal treatment among team members; 17% observed moral harassment and feelings of superiority

on the part of men; 15% have already observed sexist attitudes and competitiveness between men and women.

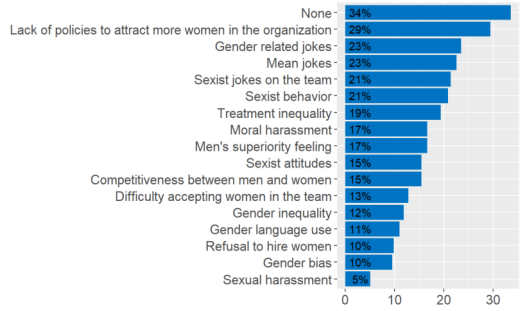


Figure 3: Participants' perceptions about whether there is sexist behavior in their team.

Canedo et al. [13] identified that gender bias and moral harassment are the sexist behavior most perceived by women in the Brazilian software industry. Nonetheless, the perception of men is quite different from that, as Figure 3 shows. Our findings on men's perception of the existence of sexist behavior in software development teams also differ from other studies [19, 20, 38, 40, 41].

RQ.3 Summary: Male practitioners in software development teams fail to recognize or acknowledge sexist behavior, and assume that the way men interact with each other is the same as how they interact with women.

4.2 RQ.2. How do men practitioners perceive the barriers and challenges related to gender issues in software development teams?

Figure 4 shows the barriers that participants face or have already faced in their software development teams. In summary, 47% mentioned the lack of experience in the field; 40% reported communication problems with other team members; 28% said not feeling useful or not creating something important; 24% mentioned being treated differently by some team members; 20% said cultural differences, such as cultural aspects of each Brazilian state, e.g., accent, taste in music, among others; 20% reported difficulty in leading teams; and lastly only 17% mentioned not facing any barrier from other teammates.

On the other side, the most cited barriers identified by the women were gender bias in task allocation, difficulty in leading the team, and the lack of a leadership profile, respectively [13]. But, *task allocation* was not even mentioned by men and the *lack of experience* was mentioned by only one of the participants in that study. Although our finding differs from other studies that investigated women's perception of the barriers they face in Software Engineering [11–13, 18, 23, 34, 37–39], we believe the studies are complementary.

Regarding the challenges (Figure 5), 47% of men practitioners reported experiencing challenges due to lack of training opportunities; 45% due to lack of knowledge; 38% due to lack of women in

Figure 4: Participants' perceptions about the barriers they faced

the organization; 27% faced a lack of trust; 25% faced lack of opportunities; 21% faced lack of attractiveness; and 20% of participants faced lack of concern from their software development team. In the women's view [13], lack of trust was the most significant challenge faced by women. Aksekili and Stettina [1] also identified the lack of trust as one of the challenges that women face in agile software development teams.

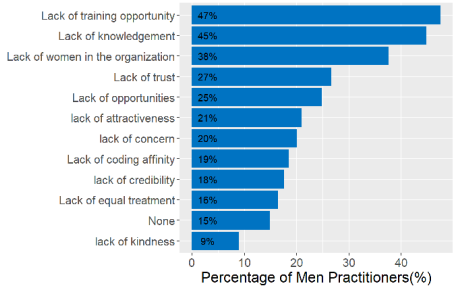


Figure 5: Participants' perceptions about the challenges they faced

Our study also showed that 65% of men practitioners did not suffer any barrier or prejudice to be accepted because they were newbies or inexperienced when they arrived at the software development team, as shown Q17 in Figure 2. Conversely, the same does not happen to women. Recent studies [13, 37] reveled that women in software development teams have reported facing challenges in gaining acceptance from male colleagues.

Still related to gender issues, this study also investigated whether there are women in leadership positions in the companies the respondents work for. As a result, 30% of the men stated that there are no women leading software development teams where they work for. 15% of the respondents affirmed that there is only one woman in this position. For 27%, there are between 2 and 4 women in leadership positions. For 12%, there are between 5 and 15 women. For 10%, there are more than 15; and 6% could not answer. Thus, 45% of respondents work in companies with no or at most one woman in a leadership position. Women are already a minority on software development teams, and women in leadership positions are even rarer.

When there are women on the teams, they also said that they work on different projects, with high (58%), medium (56%), and low

complexity (29%), and in the most diverse languages. According to the men's perceptions, the most cited languages in which women work are JavaScript (42%), Java (36%), Python (23%), Typescript (20%), C# (12%), and PHP (11%). Besides, 49% of projects have up to 10 development team members, 19% have between 11 and 15 team members, and only 12% of projects have more than 20 team members. 37% of respondents said that projects where women work last more than 12 months, 26% of them said that projects last between 7 and 12 months, and 19% said that projects where women work last less than 6 months, as shown in Figure 6.

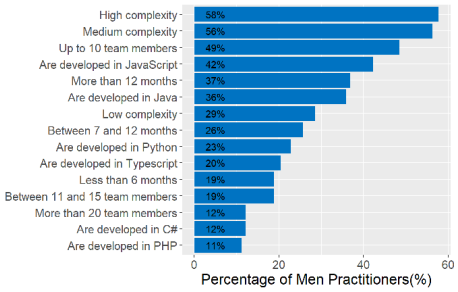


Figure 6: Participants' perceptions about the projects their women coworkers worked with

This result confirms other findings in the literature Canedo et al. [10, 12], Catolino et al. [14], Overflow [29], Trinkenreich et al. [39] in which the authors also identified that women work on projects that are developed using different programming languages, such as JavaScript, Java, Python, Typescript and PHP.

RQ.3 Summary: Men practitioners and women practitioners perceive barriers and challenges differently. In addition, men did not suffer any barriers at the beginning of their careers.

4.3 RQ.3. What is the difference between men's and women's perceptions of gender issues on software development teams?

We asked male practitioners if they had ever received or perceived any harassment from a colleague and how they resolved the situation. Most participants reported that they have never noticed any harassment. Less than 5% of the participants mentioned having already faced or perceived harassment. For example, respondent #R07 and #R160 said, respectively:

"The person responsible for training me in a new position refused to train me and was always dismissive of me in every way, leaving a project in my charge and trying to sabotage the execution of the project. I solved it by finding a way to perform my function, ignoring this colleague, and studying and working overtime to handle the demand."

Table 2: Categories and subcategories from question coding

Category	Subcategory	# Cited
Technical Skills	Affinity with the area	65
	Lack of code affinity	58
	Lack of knowledge	39
	Pressure from the area	22
	Professional development	13
	Lack of experience	10
	Professional experience	8
	Complexity of the area	7
	Ease of other areas	6
	Lack of interest	5
	Difficulty in learning	5
	Lack of capacity	5
Organizational Environment	Lack of opportunities	18
	More attractive salary	8
	Hostile environment	6
	Toxic culture	6
Interpersonal and Personal Aspects	Fear	8
	Leadership profile	5
	Prejudice	4
	Gender bias in task allocation	3
	Capacity of solving problems	3
	Humane personality	4
	Impostor syndrome	3
	Empathy	3

" I had to deny the demands requested by my project manager. This had negative consequences for the organization. I thought it was important to set a boundary at the beginning of the relationship. Over time, the manager realized that harassment didn't work for me, although this behavior only works with job security."

Although it is not common for men to experience harassment from bosses and co-workers, the same does not happen to women. In a recent study Canedo et al. [13], most women reported experiencing harassment from a male colleague within their software development team. In addition, our results also confirm the findings of Trinkenreich et al. [39], which identified that gender bias and sexist behavior are very present in the Open Source Software community.

Additionally, we investigated why some professionals go to documentation, testing, modeling, or leadership instead of software development and if, in the men's perception, this is more frequent with men or women. We identified three categories of possible reasons: 1) Technical skills, 2) Organizational environment, and 3) Personal and interpersonal aspects. Table 2 presents the categories and subcategories identified in the data analysis. The majority of respondents cited a lack of interest or affinity for the field as the primary reason for changing areas. Of the total participants, 61% mentioned this change occurs more frequently among women, 18% mentioned it occurs equally between men and women, and only 9% reported it occurs more often among men. For example, respondent #R107 and #R199 said, respectively:

"[...] Due to the difficulty women have in writing code or the difficulty in thinking about the logic for implementing an algorithm."

"[...] Some people don't know how to program and prefer to go to these areas. Women usually don't like and don't know how to code and prefer to document. In my team, the project manager already knows this and assigns these tasks to them."

As can be seen in the male statements, some men think that women do not have the aptitude for programming tasks, which may represent discriminatory behavior on their part. Conversely, Canedo et al. [13] investigated which activities women who participate in software development teams feel more motivated to perform, they reported that coding is the activity they like the most. However, more studies are needed to better understand this issue.

RQ.3 Summary: Only few men practitioners noticed or suffered any kind of harassment by a member of the team. Men believe that women do not have an affinity with coding and that is why they prefer to work in other areas.

4.4 RQ.4. How do men practitioners perceive their own careers?

We asked male practitioners which activities they felt most motivated to carry out in a software development project. As a result, 84% reported programming or development, 38% reported project management, 37% requirements elicitation, 35% software maintenance and evolution software, 34% project modeling, 27% testing, and only 14% of them reported project documentation. Along these lines, Canedo et al. [13] found that both women and men reported that coding is the activity they feel most motivated to perform.

Regarding the reasons that led men practitioners to participate in a software development team, 90% of male practitioners reported that it was due to affinity with the software development area, 76% due to personal satisfaction, 58% due to career growth opportunities, 51% due to better salaries, 28% due to guaranteed job opportunities, and only 8% of male practitioners said it was because of the encouragement of their families. The summary of these reasons is shown in Figure 7.

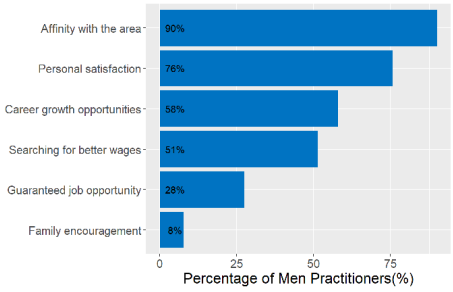


Figure 7: Reasons that led participants to join a software development team

Almost all participants (99%) confirmed that their contributions are usually well received by the other software development team members (Q13 in Figure 2). On the other hand, other studies revealed that women do **not** feel that their contributions are well received by development teams in which they work [5, 10]. Also, 92% of men practitioners stated that pull requests, comments, suggestions for improvements, bug fixes, and updates are well received by the software development team (Q15 from Figure 2), contrary to what women perceive as seen in other works [36, 39].

For 61% of respondents, both men and women do not make mean jokes with other team members. Only 26% perceive that this behavior occurs among men and women on their teams (Q16 in Figure 2). This result confirms the findings of Canedo et al. [13] in which most women claimed to have problems related to mean and sexist jokes in the team in which they work in the Brazilian software industry.

We asked male practitioners whether their career progression met their expectations, and 74% strongly agreed or agreed, whereas 12% disagreed (Q26 in Figure 2). Nevertheless, in another study [13], most women mentioned they do not believe they have the career development they expect. They believe that once they are women, they do not have the technical profile that software organizations value. They also reported that career progression for men is easier and faster, as stakeholders prefer men in the most critical positions and do not believe that women can repay the organization the benefit they receive from assuming a leadership position.

Regarding the level of satisfaction of male practitioners concerning their performance in their activities in software development teams, 87% of the respondents were satisfied or very satisfied, 9% were somewhat satisfied, and only 3% were dissatisfied or very dissatisfied. This result allows us to infer that men are quite satisfied with their performance in the software development teams in which they work in organizations. Similarly, women follow men's perception Canedo et al. [13], since they demonstrated that were satisfied with their performance in the activities carried out in the team they work.

RQ.3 Summary: Men practitioners are motivated to exercise programming activities and participate in software development teams because they have an affinity with the area. In addition, men think that their career progression meets their expectations and that team members well receive their contributions.

4.5 RQ.5. How to build a more inclusive environment for IT practitioners?

The participants who were surveyed offered suggestions for creating a more inclusive work environment for IT professionals. This was an open question, so they were free to give their suggestions or not. 12.5% (25) of respondents did not answer this question, and from those who responded, ~8% said the environment is already inclusive and did not propose improvements. In total, 177 men provided a suggestion, and from these, 18.6% (33) cited the need for **in-company training** in two ways: (i) to make employees aware of the problems faced by minority groups; and (ii) to provide professional qualification training. They suggested training focused on

leaders, newly hired, and team members. Training could be delivered in workshops or informal meetings. For example, respondent #R171 and #R04 said, respectively:

> *"[...] I worked at a company that frequently had conversations about inclusion. This company also led to closer relations between employees through informal meetings."*

> *"Training for area leaders and company team managers to promote equality and respect within their teams so that they can identify and correctly deal with situations of disrespect."*

Almost ~16% (28) of respondents suggested a more **inclusive hiring process**. However, while some of them suggested the implementation of quotas for minority groups (they cited black people, women, and LGBTQIA+), others explicitly mentioned that do not agree with quotas policies. The ones who suggested quotas affirmed that it is extremely important that there are more inclusion policies, which help speed up the breakdown of barriers to inclusion. For example, respondent #R87 stated:

> *"I believe that it is necessary to have more job openings aimed at women, LGBTQIA+, and black people. A work environment with a lot of diversity is very good."*

On the other hand, less than 3% (5) of respondents believe that **people should be hired for their technical skills** and thus there is no need for quotas. Two of them categorically stated that the companies should not implement programs only for women. For example, #R189 said:

> *"[...] they cannot commit the stupidity of hiring a professional solely and exclusively for meeting a quota or for a specific gender characteristic."*

Similarly, less than 2% (3) of respondents suggested **policies to attract both men and women**. They said that companies should have an equal policy for all company members, whether men or women. A complementary result was found in Wang et al. [42], where they found that there are men classified as fundamentalists, who advocates male dominance in software development. They do not support a hiring plan favoring underrepresented groups.

Conversely, 14 respondents (~8%) recommended that companies should adopt an **inclusive culture** and **inclusion and diversity policies**. They indicated that this culture and policies could come from a top-down perspective so that diversity can be treated responsibly. Hence, managers and human resources personnel should lead inclusive actions. Some of the actions mentioned were: having ambassadors of the cause; applying practical actions in the team's day-to-day; and creating clear strategies and goals. For instance, #R126 declared:

> *"(the company should) invest time and resources to develop or reformulate a culture that aligns management, development, and operation processes with more human values focused on excellence, performance, sustainability, and personal development"*

Another suggestion to build a more inclusive environment was **diversifying teams**. 11 (6.2%) respondents believe that the number of women and men in the teams could be better distributed. One respondent believes that projects that have more women are one step ahead in the issue of empathy with users and the team itself. He also believes that women are able to better focus on solving problems and not on pointing the finger at the culprits; they can better understand when someone on the team is not well; and they are more receptive to criticism than men. Along those lines, the Cloverpop platform in a study named *Hacking Diversity With Inclusive Decision Making*[2] found that the more diverse a team, the better decisions it makes 87% of the time. Besides, companies that invest in gender equality and diversity policies show better results, including their profitability[28].

Around 5.6% of the respondents mentioned the importance of promoting a healthier work environment. They believe that in order to foster inclusivity, the workplace should be relaxed, less formal, pleasant, and safe. They also mentioned the need for **respect and equal treatment** for all (2.3%). A place where everyone feels respected and valued. Another suggestion was to offer **flexible work hours and remote work** as a way to include more people (~3.4% of respondents). For example, R#189 commented:

> *"If possible, and the nature of the company's activities allow, offering flexible working conditions, such as adaptable hours and remote work so that everyone can feel more comfortable and adapt better [...]"*

Ten respondents (5.6%) recommended **improving career plans and policies for progression and salary equalization**. According to them, the organization must create a career path for both men and women that allows them to progress. It is worth mentioning that among them, only one mentioned the need for a plan that includes different realities and that can analyze the organization's situation to understand its challenges and propose solutions.

Eight respondents mentioned **improving communication** as a suggestion to create a more inclusive environment and another 10 suggested actions on how this improvement could happen, totaling ~10.2% of respondents. They proposed eight actions, which are: creating a space for exchanging experience, improving the interaction between IT and business, increasing interaction through pair programming and code review, promoting activities for collaborative knowledge, providing interdisciplinary activities, promoting collaboration between team members, and providing a common area.

They also mentioned that organizations could provide greater **support to newly hired employees** (less than 2%), with basic things like help setting up the machine environment and initial

[2]The Cloverpop study is available at: https://www.cloverpop.com/hacking-diversity-with-inclusive-decision-making-white-paper

introduction of team members and their skills; they could **invest in qualification** by offering access to growth opportunities (less than 3); and **invest in trainee programs** (1.1%). Other suggestions were cited only once or twice: providing constant feedback and monitoring individual performance, offering health plans, improving business and project management, fostering the use of inclusive language, more enforcement against harassment, investing in marketing to reach more women, stopping with mean jokes, not tolerating prejudice, punishment for bad behavior, and providing complaint support tools. Table 3 presents all suggestions that received two or more citations. Compared to a study on women's perceptions [13], the suggestions were similar between genres. However, women were more emphatic to suggest that organizations must have programs to promote gender diversity among their employees, besides conducting campaigns and lectures on the importance of having gender diversity.

Table 3: Suggestions to build a more inclusive environment

ID	Suggestions	# Cited
1	In-company training	33
2	Inclusive hiring process	28
3	Inclusive culture and inclusion and diversity policies	14
4	Diversifying teams	11
5	Healthier environment	10
6	Improving career plans /policies for progression /salary equalization	10
7	Improving communication	8
8	Flexible work hours/remote work	6
9	Hiring process by technical skills	5
10	respect and equal treatment	5
11	Investment in qualification	5
12	Supporting newly hired employees	3
13	Policies to attract both men and women	3
14	Trainee programs	2
15	Providing feedback	2
16	Improving business and project management	2
17	Investing in marketing to reach more women	2

RQ.3 Summary: While some men do not perceive the need for policies for building a more inclusive environment, many of them provided some suggestions, such as in-company training and an inclusive hiring process. Even with some men saying that do not agree with quotas policies, many said that this is an important action towards a more inclusive work environment.

4.6 Threats to Validity

As with any empirical work, this work has many limitations and threats to validity. *Content validity* is a subjective evaluation of the appropriateness of the instrument according to the target population [22]. Although the survey questionnaire has been partially used in a similar study that investigated Brazilian women's perceptions, two authors of this work created the survey, and the other two reviewed the questions to mitigate underlying threats. Moreover, before sharing the questionnaire, we conducted a pilot with five men practitioners, which resulted in some changes in the questionnaire.

The *external validity* aims to attest if the results can be generalized to other people, places, or times. Although this research has a large sample (217 men spread across Brazilian states), we cannot guarantee that it represents the whole country. Furthermore, the results may not apply to other countries due to cultural differences. Lastly, results can get outdated as time passes since cultural changes highly impact individuals' perceptions, especially when dealing with gender issues. In the *internal validity*, even though we have applied scientific rigor in the study's execution, it is inevitable that the subjectivity of the researchers would influence the result's interpretation. Given the delicate nature of some questions, even if we could attest that the analysis of the results was free from subjectivity, we cannot guarantee that our participants expressed fully honest opinions.

5 CONCLUSIONS

This survey's results suggest significant differences in the perceptions of men and women practitioners in software development. The responses of the 217 male participants showed a much more positive perception of their work and careers than the perceptions reported in previous studies of women in the field.

Works focusing on women's experiences in software development found that they faced several barriers, especially at the beginning of their careers. In our work, the men respondents reported fewer difficulties in their careers and felt welcomed by their teams, unlike women. In addition, they were less aware of sexism than women surveyed in other works. Some respondents even had sexist perceptions themselves, such as believing that women do not like or are not capable of programming activities. This disparity in perception raises important questions about the nature of gender bias and its impact on women's experiences in software development.

We observed that some men still do not perceive the need for an inclusive hiring process and others do not agree to a hiring process explicitly targeted at women and minority groups. On the other hand, the most suggested recommendation to build an inclusive environment was in-company training. We believe that training aimed at this end could be useful and serve to raise men's awareness about aspects of gender and inclusion. It is clear from these results that much work needs to be done to address the issue of gender inequality in the software development industry. While progress has been made in recent years, particularly in increasing the number of women in leadership positions, there is still a long way to go to create a truly equitable workplace. As organizations seek to address this issue, they must consider the findings of this study, in order to create an environment that is welcoming and supportive to all people. By doing so, we can create a more diverse and inclusive industry that can better meet the challenges of the 21st century.

ACKNOWLEDGMENTS

We thank all the man practitioners who responded to the survey. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

ARTIFACT AVAILABILITY

The data that support the findings of this study are openly available in Zenodo at https://doi.org/10.5281/zenodo.7888625 [2].

REFERENCES

[1] Asli Yüksel Aksekili and Christoph Johann Stettina. 2021. Women in agile: The impact of organizational support for women's advancement on teamwork quality and performance in agile software development teams. In *Lean and Agile Software Development: 5th International Conference, LASD 2021, Proceedings 5*. Springer, https://doi.org/10.1007/978-3-030-67084-9_1, 3–23.

[2] Anonymous. 2023. *Supplementary Material for Men's Perceptions of Gender Inequality in Software Engineering*. Zenodo. https://doi.org/10.5281/zenodo.7888625

[3] Aleteia Araujo, Maristela Holanda, Carla Castanho, Carla Koike, Roberta B Oliveira, Edna Canedo, and Mirella M Moro. 2022. Pandemia de COVID-19 tem Gênero. In *Anais do XVI Women in Information Technology*. SBC, Anais do XVI Women in Information Technology (WIT 2022 –SBC), 10.5753/wit.2022.223044, 110–121.

[4] Elisa Bellotti, Dominika Czerniawska, Martin G. Everett, and Luigi Guadalupi. 2022. Gender inequalities in research funding: Unequal network configurations, or unequal network returns? *Soc. Networks* 70 (2022), 138–151. https://doi.org/10.1016/j.socnet.2021.12.007

[5] Amiangshu Bosu and Kazi Zakia Sultana. 2019. Diversity and Inclusion in Open Source Software (OSS) Projects: Where Do We Stand?. In *2019 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, ESEM 2019, Porto de Galinhas, Recife, Brazil, September 19-20, 2019*. IEEE, https://doi.org/10.1109/ESEM.2019.8870179, 1–11.

[6] Sterre Van Breukelen, Ann Barcomb, Sebastian Baltes, and Alexander Serebrenik. 2023. "STILL AROUND": Experiences and Survival Strategies of Veteran Women Software Developers. *CoRR* abs/2302.03723 (2023), 1–13. https://doi.org/10.48550/arXiv.2302.03723 arXiv:2302.03723

[7] Danielle Brown and Melonie Parker. 2019. Google diversity annual report 2019. https://diversity.google/ (Date last accessed 16-April-2019).

[8] Danielle Brown and Melonie Parker. 2023. 2022 Diversity Annual Report. https://diversity.google/ (Date last accessed 7-April-2023).

[9] John A. Bullinaria. 2018. Agent-Based Models of Gender Inequalities in Career Progression. *J. Artif. Soc. Soc. Simul.* 21, 3 (2018), 1–28. https://doi.org/10.18564/jasss.3738

[10] Edna Dias Canedo, Rodrigo Bonifácio, Márcio Vinicius Okimoto, Alexander Serebrenik, Gustavo Pinto, and Eduardo Monteiro. 2020. Work Practices and Perceptions from Women Core Developers in OSS Communities. In *ESEM '20: ACM / IEEE International Symposium on Empirical Software Engineering and Measurement, Bari, Italy, October 5-7, 2020*, Maria Teresa Baldassarre, Filippo Lanubile, Marcos Kalinowski, and Federica Sarro (Eds.). ACM, https://doi.org/10.1145/3382494.3410682, 26:1–26:11.

[11] Edna Dias Canedo, Angélica Toffano Seidel Calazans, Geovana R. S. Silva, and Eloisa Toffano Seidel Masson. 2022. ICT Practitioners' Perception of Working from Home During the Covid-19 Pandemic: Exploring Gender Differences. In *SBES 2022: XXXVI Brazilian Symposium on Software Engineering, Brazil, October 5 - 7, 2022*, Marcelo de Almeida Maia, Fabiano A. Dorça, Rafael Dias Araújo, and Edna Dias Canedo (Eds.). ACM, https://doi.org/10.1145/3555228.3555248, 47–57.

[12] Edna Dias Canedo, Heloise Acco Tives Leão, Madianita Bogo Marioti, Fabiano Fagundes, and José Antonio Siqueira de Cerqueira. 2019. Barriers Faced by Women in Software Development Projects. *Inf.* 10, 10 (2019), 309. https://doi.org/10.3390/info10100309

[13] Edna Dias Canedo, Fabiana Freitas Mendes, Anderson Jefferson Cerqueira, Márcio Vinicius Okimoto, Gustavo Pinto, and Rodrigo Bonifácio. 2021. Breaking one barrier at a time: how women developers cope in a men-dominated industry. In *35th Brazilian Symposium on Software Engineering, SBES 2021, Joinville, Santa Catarina, Brazil, 27 September 2021 - 1 October 2021*, Cristiano D. Vasconcellos, Karina Girardi Roggia, Vanessa Collere, and Paulo Bousfield (Eds.). ACM, https://doi.org/10.1145/3474624.3474638, 378–387.

[14] Gemma Catolino, Fabio Palomba, Damian A. Tamburri, Alexander Serebrenik, and Filomena Ferrucci. 2019. Gender diversity and women in software teams: how do they affect community smells?. In *Proceedings of the 41st International Conference on Software Engineering: Software Engineering in Society, ICSE 2019, Montreal, QC, Canada, May 25-31, 2019*, Rick Kazman and Liliana Pasquale (Eds.). ACM, https://doi.org/10.1109/ICSE-SEIS.2019.00010, 11–20. https://doi.org/10.1109/ICSE-SEIS.2019.00010

[15] Rosanna Durruthy. 2018. 2018 LinkedIn Workforce Diversity Report. https://careers.linkedin.com/diversity-and-inclusion/workforce-diversity-report (Date last accessed 16-April-2019).

[16] Rosanna Durruthy. 2023. LinkedIn's 2022 Workforce Diversity Report. https://www.linkedigcom/pulse/linkedins-2022-workforce-diversity-report-rosanna-durruthy/ (Date last accessed 7-April-2023).

[17] Zixuan Feng, Mariam Guizani, Marco A Gerosa, and Anita Sarma. 2023. The State of Diversity and Inclusion in Apache: A Pulse Check. *arXiv preprint arXiv:2303.16344* 1 (2023), 1–11.

[18] Karen S Garner and Chantelle Y Van Staden. 2022. Women in Engineering–Barriers to Participation in the South African Context. In *2022 31st Annual Conference of the European Association for Education in Electrical and Information Engineering (EAEEIE)*. IEEE, https://ieeexplore.ieee.org/document/9820174, 1–4.

[19] Mariam Guizani, Bianca Trinkenreich, Aileen Abril Castro-Guzman, Igor Steinmacher, Marco Gerosa, and Anita Sarma. 2022. Perceptions of the State of D&I and D&I Initiative in the ASF. In *Proceedings of the 2022 ACM/IEEE 44th International Conference on Software Engineering: Software Engineering in Society* (Pittsburgh, Pennsylvania) *(ICSE-SEIS '22)*. ACM, https://doi.org/10.1145/3510458.3513008, 130–142.

[20] Nasif Imtiaz, Justin Middleton, Joymallya Chakraborty, Neill Robson, Gina R. Bai, and Emerson R. Murphy-Hill. 2019. Investigating the effects of gender bias on GitHub. In *Proceedings of the 41st International Conference on Software Engineering, ICSE 2019, Montreal, QC, Canada, May 25-31, 2019*, Joanne M. Atlee, Tevfik Bultan, and Jon Whittle (Eds.). IEEE / ACM, https://doi.org/10.1109/ICSE.2019.00079, 700–711.

[21] Barbara A. Kitchenham and Shari Lawrence Pfleeger. 2008. Personal Opinion Surveys. In *Guide to Advanced Empirical Software Engineering*, Forrest Shull, Janice Singer, and Dag I. K. Sjøberg (Eds.). Springer, 10.1007/978-1-84800-044-5_3, 63–92. https://doi.org/10.1007/978-1-84800-044-5_3

[22] Barbara A Kitchenham and Shari L Pfleeger. 2008. Personal opinion surveys. *Guide to advanced empirical software engineering* 1 (2008), 63–92.

[23] Karina Kohl and Rafael Prikladnicki. 2021. Challenges Women in Software Engineering Leadership Roles Face: A Qualitative Study. In *Proceedings of the 23rd International Conference on Enterprise Information Systems, ICEIS 2021, Volume 2*, Joaquim Filipe, Michal Smialek, Alexander Brodsky, and Slimane Hammoudi (Eds.). SCITEPRESS, https://doi.org/10.5220/0010413602050212, 205–212.

[24] Karina Kohl and Rafael Prikladnicki. 2022. Benefits and Difficulties of Gender Diversity on Software Development Teams: A Qualitative Study. In *SBES 2022: XXXVI Brazilian Symposium on Software Engineering, 2022*, Marcelo de Almeida Maia, Fabiano A. Dorça, Rafael Dias Araújo, and Edna Dias Canedo (Eds.). ACM, https://doi.org/10.1145/3555228.3555253, 21–30.

[25] Arvind Krishna. 2023. IBM 2021 ESG Report. https://www.ibm.com/impact/files/reports-policies/2021/IBM_2021_ESG_Report.pdf (Date last accessed 7-April-2023).

[26] Amanda Lee and Jeffrey C. Carver. 2019. FLOSS participants' perceptions about gender and inclusiveness: a survey. In *Proceedings of the 41st International Conference on Software Engineering, ICSE 2019, Montreal, QC, Canada, May 25-31, 2019*, Joanne M. Atlee, Tevfik Bultan, and Jon Whittle (Eds.). IEEE / ACM, https://doi.org/10.1109/ICSE.2019.00077, 677–687. https://doi.org/10.1109/ICSE.2019.00077

[27] Lindsay-Rae McIntyre. 2023. Global Diversity & Inclusion Report. https://www.microsoft.com/en-us/diversity/default.aspx (Date last accessed 7-April-2023).

[28] International Labour Office. 2019. *Women in business and management: The business case for change*. International Labour Organisation (ILO), https://www.ilo.org/global/publications/books/WCMS_700953/lang--en/index.htm.

[29] Stack Overflow. 2022. Developer Survey Results 2022. https://survey.stackoverflow.co/2022/ (Date last accessed 22-April-2023).

[30] Rocco Palumbo and Rosalba Manna. 2020. A portrait of workplace discrimination in Italy: empirical evidence from a nationwide survey. In *Evidence-based HRM: A Global Forum for Empirical Scholarship*, Vol. 8. Emerald Publishing Limited, https://doi.org/10.1108/EBHRM-03-2019-0029, 92–112.

[31] Larissa Rocha, Edna Dias Canedo, Claudia Pinto Pereira, Carla Bezerra, and Fabiana Freitas Mendes. 2023. Investigating the Perceived Impact of Maternity on Software Engineering: a Women's Perspective. In *Proceedings of the 16th International Conference on Cooperative and Human Aspects of Software Engineering*. IEEE, 10.1109/CHASE58964.2023.00023, 1–13.

[32] Davide Rossi and Stefano Zacchiroli. 2022. Worldwide Gender Differences in Public Code Contributions and how they have been affected by the COVID-19 pandemic. In *44th IEEE/ACM International Conference on Software Engineering: Software Engineering in Society ICSE (SEIS) 2022, Pittsburgh, PA, USA, May 22-24, 2022*. IEEE/ACM, https://doi.org/10.1109/ICSE-SEIS55304.2022.9794118, 172–183. https://doi.org/10.1109/ICSE-SEIS55304.2022.9794118

[33] Camila Sarmento, Tiago Massoni, Alexander Serebrenik, Gemma Catolino, Damian A. Tamburri, and Fabio Palomba. 2022. Gender Diversity and Community Smells: A Double-Replication Study on Brazilian Software Teams. In *IEEE International Conference on Software Analysis, Evolution and Reengineering, SANER 2022, Honolulu, HI, USA, March 15-18, 2022*. IEEE, https://doi.org/10.1109/SANER53432.2022.00043, 273–283.

[34] Lizandra Lígia Soares Bezerra Silva, Suzana Sampaio, and Marcelo Luiz Monteiro Marinho. 2022. Gender Diversity in Technology Sector: Barriers and Perceptions Success of Women in Porto Digital. In *Proceedings of the XIV Latin American Women in Computing 2022 co-located with XLVIII Latin American Computer Conference 2022 (CLEI 2022), Antioquia, Colombia, October 19-20, 2022 (CEUR Workshop Proceedings, Vol. 3321)*, María Elena García-Díaz and Maristela Holanda (Eds.). CEUR-WS.org, http://ceur-ws.org/Vol-3321/paper7.pdf, 65–76.

[35] Klaas-Jan Stol, Paul Ralph, and Brian Fitzgerald. 2016. Grounded theory in software engineering research: a critical review and guidelines. In *Proceedings of the 38th International Conference on Software Engineering, ICSE 2016, Austin, TX, USA, May 14-22, 2016*, Laura K. Dillon, Willem Visser, and Laurie A. Williams (Eds.). ACM, https://doi.org/10.1145/2884781.2884833, 120–131. https://doi.org/10.1145/2884781.2884833

[36] Josh Terrell, Andrew Kofink, Justin Middleton, Clarissa Rainear, Emerson Murphy-Hill, Chris Parnin, and Jon Stallings. 2017. Gender differences and bias in open source: Pull request acceptance of women versus men. *PeerJ Computer Science* 3 (2017), e111.

[37] Bianca Trinkenreich, Ricardo Britto, Marco Aurélio Gerosa, and Igor Steinmacher. 2022. An Empirical Investigation on the Challenges Faced by Women in the Software Industry: A Case Study. In *44th IEEE/ACM International Conference on Software Engineering: Software Engineering in Society ICSE (SEIS) 2022, Pittsburgh, PA, USA, May 22-24, 2022*. IEEE/ACM, https://doi.org/10.1109/ICSE-SEIS55304.2022.9793931, 24–35.

[38] Bianca Trinkenreich, Marco Aurélio Gerosa, and Igor Steinmacher. 2022. Women in Open Source: We Need to Talk About It. *Computer* 55, 12 (2022), 145–149. https://doi.org/10.1109/MC.2022.3200860

[39] Bianca Trinkenreich, Igor Wiese, Anita Sarma, Marco Aurélio Gerosa, and Igor Steinmacher. 2022. Women's Participation in Open Source Software: A Survey of the Literature. *ACM Trans. Softw. Eng. Methodol.* 31, 4 (2022), 81:1–81:37. https://doi.org/10.1145/3510460

[40] Yi Wang and David Redmiles. 2019. Implicit Gender Biases in Professional Software Development: An Empirical Study. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*. IEEE, 10.1109/ICSE-SEIS.2019.00009, 1–10.

[41] Yi Wang and Min Zhang. 2020. Reducing Implicit Gender Biases in Software Development: Does Intergroup Contact Theory Work?. In *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (Virtual Event, USA) *(ESEC/FSE 2020)*. ACM, 10.1145/3368089.3409762, 580–592.

[42] Yi Wang, Xinyue Zhang, and Xinyue Zhang. 2023. Fundamentalists, Integrationists, & Transformationists: An Empirical Theory of Men Software Engineers' Orientations in Gender Inequalities. In *45th IEEE/ACM International Conference on Software Engineering: Software Engineering in Society ICSE (SEIS) 2023, Melbourne, Australia, May 14-20, 2023*. IEEE/ACM, 10.1109/ICSE-SEIS58686.2023.00009, 1–12.

[43] Claes Wohlin, Per Runeson, Martin Höst, Magnus C. Ohlsson, and Björn Regnell. 2012. *Experimentation in Software Engineering*. Springer, https://doi.org/10.1007/978-3-642-29044-2.

PLATAFORMA IoT PARA PREDIÇÃO DE FALHAS EM CONGELADORES ATRAVÉS DE MODELO AUTO-REGRESSIVO INTEGRADO DE MÉDIA MOVEL (ARIMA)

Bruno Soares Rabelo
*Universiade de Brasília - UnB **
*bruno.rabelo@redes.unb.br **

Luciano Moura
*Caixa Seguradora **
*Luciano.Moura@gmail.com **

Verônica Souza dos Santos
*Universiade de Brasília - UnB **
*Veronica.santos@redes.unb.br **

Francisco L. de Caldas Filho
*Universiade de Brasília - UnB **
*francisco.lopes@redes.unb.br**

Rafael T. de Sousa Jr.
*Universiade de Brasília - UnB **
*desousa@unb.br**

Fábio Lúcio Lopes de Mendonça
*Universiade de Brasília - UnB **
*fabio.mendonca@redes.unb.br **

* Only for Final Camera-Ready Submission

RESUMO

A estrutura geral das redes da Internet das Coisas (IoT) mantém-se como um assunto interessante para pesquisa e inovação, particularmente para responder aos desafios do monitoramento geral de dispositivos em redes. Este artigo trata da estrutura e funcionamento de uma plataforma de IoT proposta para o monitoramento preventivo e contínuo da temperatura interna em congeladores a temperatura ultrabaixa (*ultralow temperature freezers)*, utilizados em armazenamento de vacinas. Na operação desses dispositivos, é relevante investigar a questão da predição de falhas e paradas inesperadas através de tecnologias de inteligência artificial. Assim, a proposta deste artigo é a criação, por concepção conjunta de hardware e software, de um modelo de sistema IoT composto por dispositivos para coleta local de dados com sensores de temperatura nos congeladores, com envio dos dados coletados por intermédio de um middleware IoT para um servidor em nuvem que consolida a coleta de dados dos dispositivos e realiza o monitoramento contínuo (24/7) de temperatura em congeladores. Essa aplicação IoT procura identificar indícios de falha através da utilização de métodos estatísticos com algoritmos específicos, utilizando modelos preditivos para series temporais. A validação da proposta utiliza dados de congeladores que se encontram distribuídos em diversas instituições de saúde no Brasil, sendo este o cenário para a utilização do sistema IoT proposto.

PALAVRAS-CHAVE

IoT, inteligência artificial, séries temporais, análise preditiva, congeladores de ultrabaixa temperatura, conservação de vacinas.

1. INTRODUÇÃO

Nas últimas décadas, a vacinação demonstrou-se o método com maior eficácia e custo-benefício no combate a doenças trasmissíveis, pois resulta em redução comprovada do número de hospitalizações, o que é associado a custos elevados de medicamentos e da infraestrutura da rede de saúde para tratamento dessas doenças (CROSEWSKI, Fernanda et all; 2018).

Nesse sentido, em 1973 foi criado no Brasil o Programa Nacional de Imunizações (PNI) com o objetivo principal de coordenação nacional das ações de imunização. O PNI vem guiando as diversas etapas que compõem o processo de vacinação no país (DOMINGUES, Carla Magda et al; 2020), o que inclui a especificação do Manual da Rede de Frio, referenciado a partir das normas da Agência Nacional de Vigilância Sanitária - ANVISA, das orientações técnicas da Organização Mundial de Saúde (OMS) e da Organização Pan-Americana de Saúde (Opas). Este documento visa oferecer embasamento técnico para garantir a qualidade e excelência nos processos de fabricação, armazenamento, transporte, manipulação e distribuição de imunobiológicos envolvendo toda a cadeia participante, desde o laboratório fabricante até a ponta da cadeia no momento da distribuição à população (Ministério da Saúde; 2017).

Um estudo realizado em 2013 (PEREIRA, D. D. S. et al; 2013) analisou a taxa de utilização e perda de vacinas no Programa Nacional de Imunizações no período entre 2007 e 2010 de um município da região metropolitana de Curitiba (PR). Este estudo revelou que 68,91% das 2.101.103 doses de vacinas distribuídas ao município sofreram perda técnica.

Outro estudo publicado em 2021 (PATINE, Flávia dos Santos et al; 2021), analisou as perdas de vacinas em uma Região de Saúde do Nordeste Paulista. Foram analisadas as fichas de notificação de alteração de temperatura enviadas ao Grupo de Vigilância Epidemiológica do estado das 192 salas de vacinação analisadas de 67 municípios, compreendendo o período de 2010 a 2017. Neste estudo observou-se uma perda de 41,4% das doses distribuídas, sendo 70,1% das perdas por motivos estruturais.

A Organização Mundial de Saúde, relata em (WORLD HEALTH ORGANIZATION et al; 2006) 16 tipos de imunizantes que devem obrigatoriamente serem armazenados entre 2°C e 8°C. Dentre as vacinas citadas, encontram-se algumas bem difundidas como BCG (Tuberculose), Hepatite B, Meningocócica, etc. Entretanto, algumas vacinas precisam de temperaturas entre -15°C e -25°C em alguma etapa da cadeia do frio para que suas propriedades sejam preservadas.

Já em (GUIMARÃES, Eliete Albano de Azevedo et al; 2018), foram analisados 24 artigos que abordavam evidências relacionadas a eventos críticos na manutenção da cadeia do frio de conservação de vacinas. Este estudo relatou que as ocorrências mais críticas se vinculam à falta de monitoramento e supervisão adequada. Estes dados reforçam a necessidade da aplicação de tecnologia de modo a tornar o processo de monitoramento da cadeia do frio mais seguro, preciso, rastreável e confiável.

Após a criação do programa de imunização, em 1974, pela Organização Mundial de Saúde, impôs-se a necessidade do monitoramento contínuo da cadeia do frio nos processos ligados à produção, estoque e distribuição de imunizantes (LLOYD, John; CHEYNE, James; 2018). Os primeiros métodos de monitoramento, que surgiram em 1976, eram baseados em registros manuais e periódicos em papel, apenas grandes instituições possuíam registradores automáticos em discos giratórios de papel (LLOYD, John; CHEYNE, James; 2018). Esse método de registro, de forma manual e periódica, é utilizado até hoje em diversas unidades de imunização (PATINE, Flávia dos Santos et al; 2021).

Mas, por volta de 1980, algumas empresas nos EUA desenvolveram métodos de indicativos automáticos da exposição de vacinas à temperatura inadequada, métodos baseados em reações químicas e mudança na coloração de rótulos e embalagens (CHOI, Seung Yeoun et al; 2019). Estes sistemas se mostraram mais confiáveis que o método pioneiro, mas ainda não garantiam a rastreabilidade necessária, por não identificar o momento exato da exposição nem preservar os dados históricos de temperatura. Após 20 anos, os mesmos criadores desenvolveram um dos primeiros sistemas de monitoramento de temperatura remoto e contínuo baseado na Internet (LLOYD, John; CHEYNE, James; 2018).

Essa evolução indica a necessidade de avançar no desenvolvimento de soluções tecnológicas para área de monitoramento contínuo da rede de frio que tragam maior confiabilidade e precisão, dando oportunidade para o emprego de inteligência artificial e internet das coisas nesse domínio.

Por tais razões, este trabalho propõe e valida uma plataforma IoT capaz de monitorar os refrigeradores da cadeia do frio, 24 horas por dia durante os 7 dias da semana. Nessa proposta, o monitoramento usa dispositivos compostos de uma sonda de temperatura e unidades de comunicação, unidades que são respectivamente responsáveis por coletar os dados da temperatura interna dos equipamentos e enviá-los via middleware IoT aos servidores onde serão processados e armazenados. Por sua vez, no servidor a aplicação IoT realiza a

consolidação e análise dos dados com o objetivo de detectar, usando algoritmos de estatística, de forma preditiva, a tendência à falha destes refrigeradores.

Especificamente, este trabalho tem um aspecto de pesquisa aplicada, pois visa obter uma solução prática funcional para aprimorar os métodos de monitoramento da cadeia de frios usada na conservação de vacinas, buscando reduzir a ocorrência de falhas técnicas e humanas, aumentando a precisão e rastreabilidade de todo o processo. A proposta tem como requisito a aplicação homogênea em todas as etapas da cadeia de frios, desde o estoque, transporte, até distribuição de vacina para doenças imunopreveníveis.

Além da preocupação com a cadeia de frios no que diz respeito à temperatura, a inclusão de um sistema computadorizado operando na infraestrutura de rede dos hospitais e coletando dados, muitas vezes críticos, implica atender a requisitos do ponto de vista da segurança cibernética. Um dos principais e mais recentes ataques cibernéticos de grande impacto, o ataque de wannacry, deflagrado em 12 de maio de 2017, atingiu mais de 600 organizações de saúde pertencentes ao Serviço Nacional de Saúde (NHS) da Inglaterra (GHAFUR, Saira et al; 2019). Os impactos foram diversos nas unidades de saúde, desde a interrupção no uso de equipamentos, como os de ressonância magnética, cancelamento de consultas médicas e cirurgias e até mesmo o retorno do preenchimento manual e em papel dos prontuários (GHAFUR, Saira et al; 2019). Sendo assim, a proposta apresentada neste artigo utiliza técnicas de segurança cibernética como criptografia de dados, firewall e autenticação de requisitantes para garantir os requisitos de segurança cibernética dos sistemas de saúde.

Em síntese, este trabalho tem como principal objetivo a concepção e validação de sistema de monitoramento contínuo da cadeia de frios de vacinas, baseado em princípios de IoT e computação em nuvem, capaz de garantir os níveis necessários de segurança cibernética, assegurar a rastreabilidade e, através de probabilidade e estatística, prever a quebra dos equipamentos de refrigeração responsáveis pelo armazenamento de vacinas. Assim, com alavancamento na teoria estatística sua implementação em inteligência artificial, busca-se realizar uma plataforma IoT em nuvem e com um conjunto de hardware para coleta, processamento, armazenamento e exibição de dados de temperatura, coletados em tempo real nos equipamentos de cadeia de frios dos sistemas de imunoprevenção.

2. TRABALHOS RELACIONADOS

Existem diversas propostas de sistemas de predição de falhas em unidades refrigeradoras na literatura. Em (CHOI, Seung Yeoun et al; 2019) os autores utilizam a potência acumulada e o fator de potência para identificar baixo volume do líquido refrigerante. Utiliza também a temperatura da parede do freezer para detectar eventos de porta aberta. Neste estudo, obteve respectivamente,33% e 37.6% de chance de erro.

Em (WANG, Dong; ZHANG, Lun; ZHANG; 2017). os autores avaliam parâmetros como temperatura de refrigeração, de congelamento, tubo de escape, tubo de retorno, do compressor e a vibração do compressor para determinar a eficiência de refrigeração do refrigerador.

Em (HOSSEINI, Sayed Saeed et al; 2020) os autores utilizam o consumo elétrico para monitorar anomalias em eletrodomésticos, especificamente em refrigeradores. O enfoque principal é redução do consumo elétrico através da detecção em tempo real destas anomalias utilizando plugs inteligentes.

Estudos (WANGANOO, Leena; SHUKLA, Vinod Kumar; 2020) estimam um crescimento de 80% da demanda da cadeia do frio em rede alimentícia. Neste artigo, os autores propõem um sistema NB-IoT (Internet das Coisas de banda estreita) para o monitoramento da cadeia do frio em todas as etapas que a compreende, processamento, transporte, armazenamento e distribuição. O sistema proposto é composto por um módulo de hardware e um servidor em nuvem para processamento e armazenamento dos dados. Neste artigo, os autores apenas propõem um conceito e deixam como trabalhos futuros à simulação com dados mais fiéis ou aplicação em campo para validar as caraterísticas que afirmam atingir utilizando esta tecnologia.

Além do monitoramento para tornar o processo mais rastreável e seguro, alguns autores já defendem a utilização de IoT para mudar a manutenção de um cenário reativo para um cenário preditivo (BERGONZI, Luca et al; 2017). Acreditam que utilizando os benefícios da IoT na manutenção a nível industrial, a tomada de ação e a predição à falhas trará grandes economias em toda a cadeia produtiva. Neste artigo (BERGONZI, Luca et al; 2017) os autores aplicaram a IoT na predição de falhas de vitrines refrigeradas analisando os principais tipos de falhas através de sensores de temperatura e pressão. A solução apresentada foi baseada em simulações dos dados dos sensores, tratamento e sinalização das falhas, deixando como oportunidade a confecção de hardware eficiente para coleta e envio dos dados.

Recentemente, um estudo (ROMEIRO, Lucas AW et al; 2021) foi publicado defendendo a aplicação de tecnologia para monitoramento de equipamentos na indústria farmacêutica. Nesta referida pesquisa, os autores

se preocuparam em prover uma solução tecnológica segura, eficiente e de baixo custo. Foram abordados conceitos de Edge Computing para tornar a rede de sensores mais eficiente energeticamente. Este artigo reforça a necessidade de mais emprego tecnológico na cadeia do frio com maior ênfase num correto funcionamento dos dispositivos instalados em campo.

Percebe-se que os trabalhos encontrados na literatura propõem análise de variáveis que dependem da intervenção física no equipamento refrigerador ou modificações das suas características iniciais, o que pode se tornar um grande desafio na tentativa de colocar em prática o sistema de monitoramento. Diferente do que foi encontrado na literatura, este artigo propõe um sistema preditivo que utilize apenas a temperatura interna do refrigerador, variável de fácil acesso e coleta.

3. PROPOSTA E MODELO DE ARQUITETURA

Este trabalho tem como um dos seus objetivos a redução de custo e simplificação da aplicação de IoT à ultra-low-temperature freezers para armazenamento de vacinas. Sendo assim, foi utilizado nesta pesquisa apenas um sensor de temperatura, no interior do equipamento. Consequentemente, a análise preditiva de falhas foi baseada em apenas uma grandeza e métodos estatísticos para análise do padrão de dados (ROMEIRO, Lucas AW et al; 2021).

Para a coleta de dados, foi projetado e prototipado um dispositivo IoT com um conjunto de hardware composto por um módulo microcontrolador ESP-32 do fabricante Expressif (Figura **1**). Esse microcontrolador possui duplo core de processamento, 520KB de memória RAM, capacidade de comunicação Wi-Fi e Bluetooth, um sensor calibrado de temperatura DS18B20 (Figura **2**), este com range de medição de -55 à 125°C (+- 0.5 °C de precisão) e uma fonte de alimentação 5V corrente contínua.

Figura 1. Microcontrolador ESP32

Figura 2. Sensor de temperatura DS18B20

Este hardware é controlado por um módulo desenvolvido em linguagem C++, visando a coleta dos dados de temperatura minuto a minuto com *timestamp* e envio a cada 5 minutos ao servidor em nuvem. No Hardware é feita paginação dos dados antes do envio à nuvem, o que visa salvaguardar os dados em caso de indisponibilidade do servidor ou das comunicações.

O servidor, por sua vez, é responsável por receber os dados, realizar validação, processamento e armazenamento. A aplicação, que opera *on-line*, foi desenvolvida em PHP 7.4 sob o *framework* Codeigniter, utiliza base de dados SQL e está disponível na estrutura cloud da Amazon AWS. Os dados são trafegados na

base de dados através do protocolo MQTT, escolha feita visando a escalabilidade da plataforma para conexões em massa de dispositivos.

Figura 3. Arquitetura Proposta

O arranjo demonstrado na Figura 3 foi instalado em 122 equipamentos divididos em 10 grandes instituições de saúde e pesquisa do Brasil na data de 01 de setembro de 2021, conforme Tabela 1. Os critérios utilizados para a seleção dos hospitais e equipamentos foram:

- Operação nacional
- Equipamentos com faixa de operação entre -40 e 15°C
- Equipamentos em boas condições de operação
- Equipamentos do tipo Refrigeradores, Freezers e Ultra-freezers

Até a data de submissão do presente artigo, foram coletados cerca de 260 mil registros minuto a minuto de cada um dos equipamentos. Estes dados foram processados e armazenados em banco de dados com indexação, particularmente com o identificador do sensor na base de dados (*id*) e o *timestamp* do momento de medição.

Tabela 1. Segmentação de Equipamentos

Tipo	Range	Quantidade
Refrigeradores	-15 à 10°C	91
Freezers	-40 à -15°C	18
Ultra-freezers	-150 à -40°C	13

Os dados coletados ficam disponíveis para consulta em banco de dados e na aplicação *web* desenvolvida. Desta forma, é facilitada a consulta e exportação dos dados para estudo através de gráficos e relatórios (Figura 4). Detalhes técnicos e do desenvolvimento da plataforma desenvolvida em PHP não serão abordados neste artigo por não fazer parte dos resultados efetivos da predição de quebras de equipamentos.

Figura 4. Exemplo de Gráfico de Temperatura composto a partir dos dados coletados

4. CARACTERÍSTICAS DOS DADOS

Após coleta de dados, foi dado início à análise destes dados a partir de modelos estatísticos. Para simplificar a análise de dados, utilizou-se a plataforma Google Colab e algumas bibliotecas em Python:

- Pandas - Manipulação e análise de dados
- Http - Realizar requisições HTTP
- Statsmodels - Estimativa de dados através de modelos
- Matplot - Plotar gráficos

Através da plataforma web, foi escolhido um equipamento dentre os 122 monitorados para iniciar a parametrização do código em Python para análise dos dados. O sensor escolhido possui dados em ciclo estável e uma anomalia pode ser percebida verificando alterações bruscas nas séries de dados.

Após importação dos dados, foram analisados alguns parâmetros visando verificar a autocorrelação. Os dados apresentaram um perfil cíclico conforme demonstrado na Figura 5.

De acordo com o teorema central do limite, temos um n suficientemente grande, ao qual tem-se a suposição de normalidade atendida.

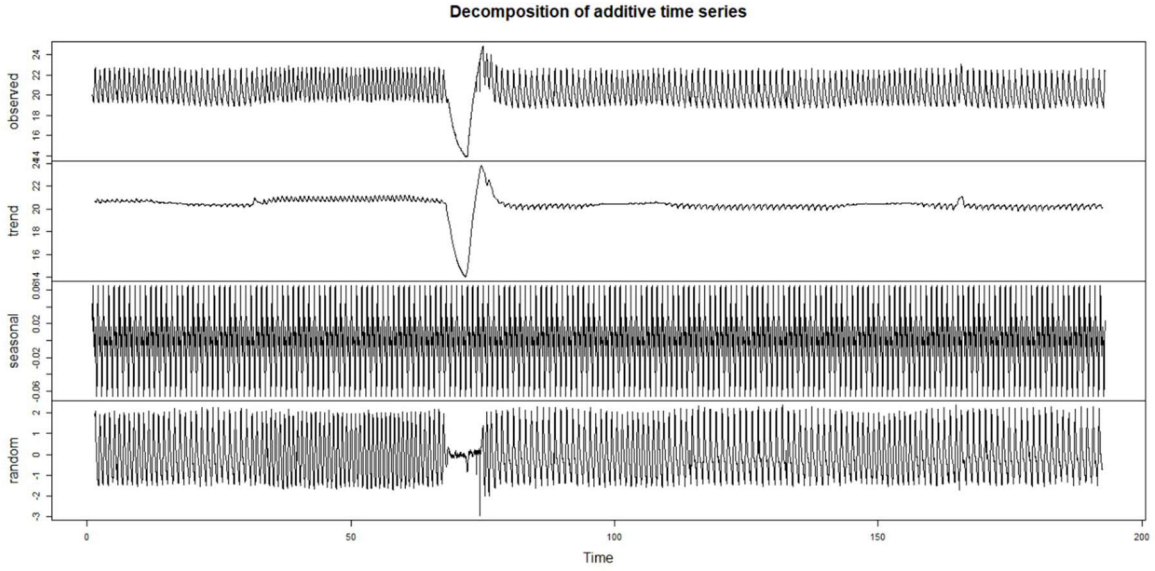


Figura 5. Gráfico de decomposição

Observa-se que os dados seguem um padrão apesar de possuir um pico ao longo da série, com isso, temos uma sazonalidade constante, resíduos que foram de fato afetados pelo pico, mas também possuem uma característica de constante.

A escolha deste cenário possibilitou a validação da proposta, a predição da temperatura mesmo em casos de instabilidade. Para a coleta de dados e utilização no Google Colab foi utilizada a biblioteca HTTP do Python para importar os dados através da API (Application programming interface) desenvolvida na plataforma PHP.

5. TESTES E RESULTADOS

Para iniciar a modelagem de dados, utilizou-se o ACF e o PACF (Morettin, P.; 2018) para determinar os fatores de correção a serem utilizados nos modelos. Conforme demonstrado na Figura **6**, é utilizado índice autorregressivo igual a 1, o que significa que a função de autocorrelação será utilizada para determinar a ordem do termo de média móvel.

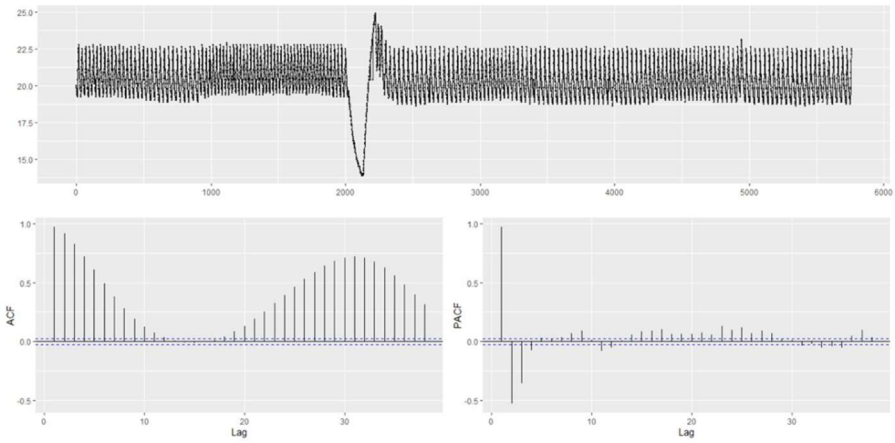


Figura 6. Autocorrelação da massa de dados

Foram testados 3 modelos diferentes para predição de dados:

- AR – *Autoregressive*
- ARMA - *Autoregressive moving average*
- ARIMA - *Autoregressive integrated moving average*

Na primeira simulação com o modelo AR, foi configurada uma janela de aprendizado de 120 minutos para projetar a temperatura dos ciclos seguintes. Após execução da predição a partir do modelo AR, obteve-se o seguinte resultado (Figura 7):

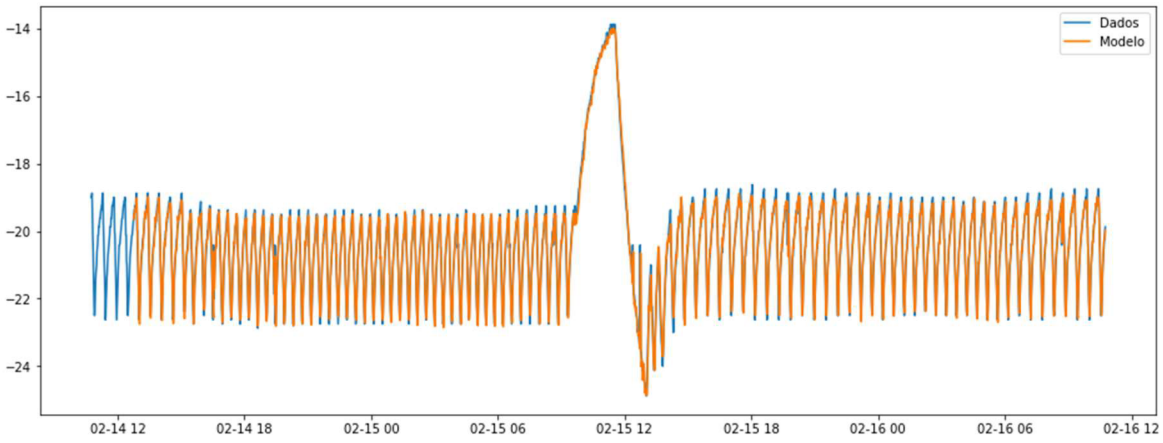


Figura 7. Modelagem de dados utilizando algoritmo AR

Com este modelo, obteve-se o erro quadrático médio de 0.05590729565537138 que é satisfatório para a aplicação que esta pesquisa deseja realizar, tendo em vista que a precisão do sensor utilizado é de +- 0.5°C.

Para fins comparativos, o modelo ARMA (Figura 8) foi parametrizado e a mesma fonte de dados do modelo AR foi utilizada para treinamento e teste do modelo. Para o treinamento, foram utilizados 2/3 dos registros e para teste o restante dos dados.

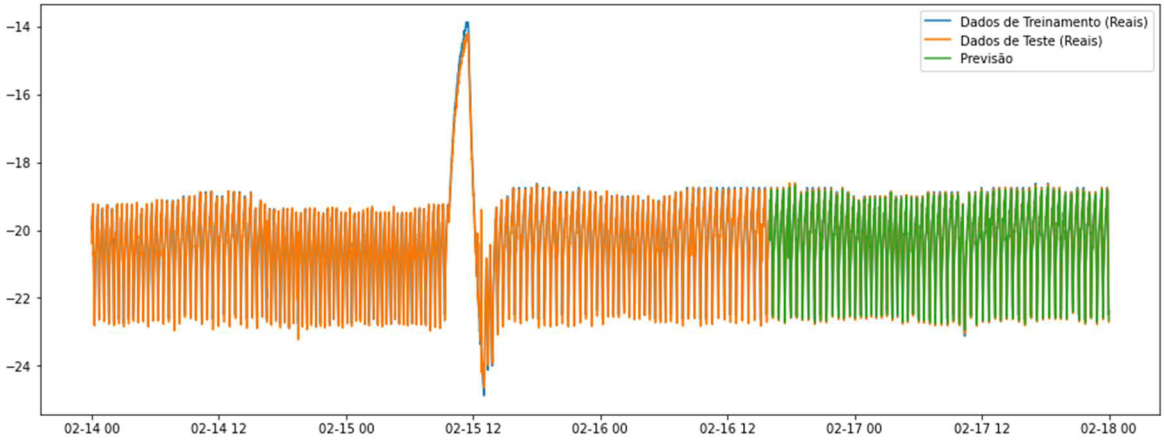


Figura 8. Modelagem de dados utilizando algoritmo ARMA

Já neste modelo, obteve-se o erro quadrático médio de 0.05264543882656802 ainda menor que o erro apresentado pelo modelo AR.

Com base nos gráficos de ACF e PACF (Figura 6), podemos dizer que o modelo que melhor se molda aos dados é ARIMA(3,1,12), com isso a predição dos dados será feita com base neste modelo (Figura 9).

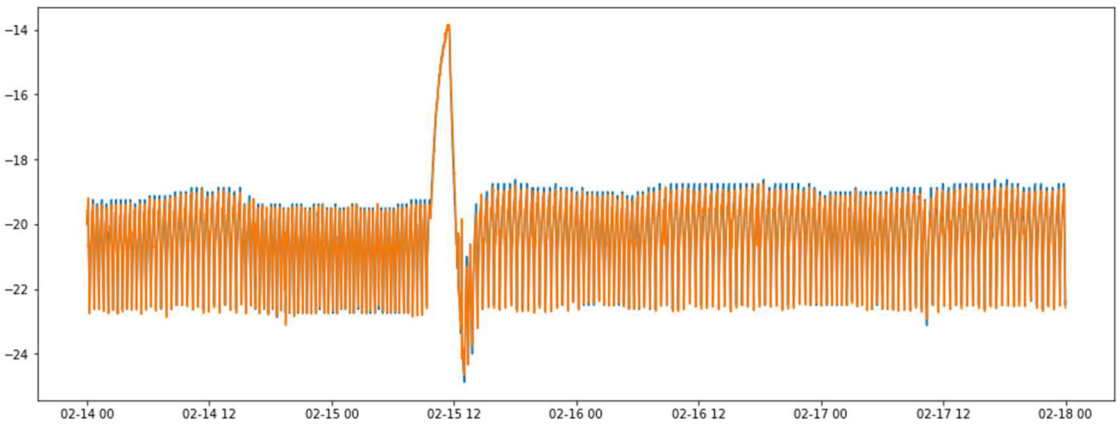


Figura 9. Modelagem de dados utilizando algoritmo ARIMA

Utilizando o modelo, foi realizada a projeção dos dados para 10 minutos à frente do fim da massa de dados. Com isso, o modelo prevê 10 registros e foi obtido o resultado apresentado na imagem (Figura 10).

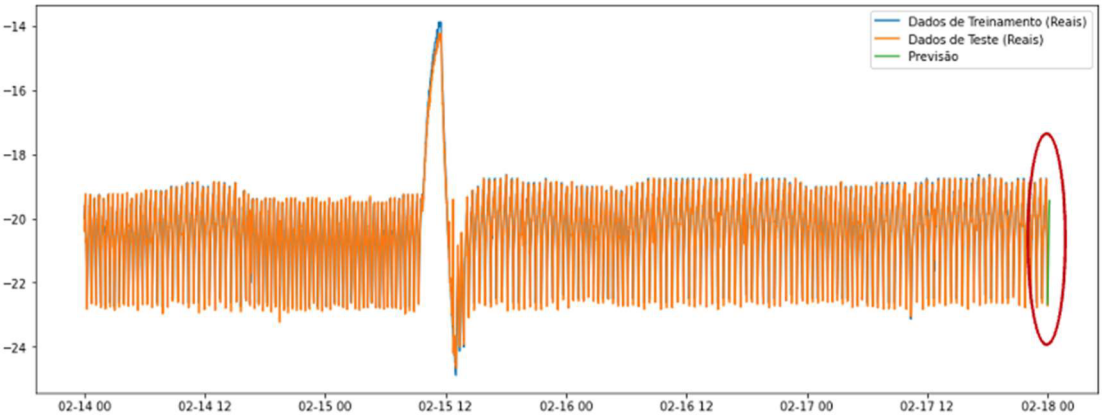


Figura 10. Predição de dados a partir do modelo ARIMA

Na Figura 10, pode-se perceber, em verde, a curva de temperatura projetada a partir do modelo matemático. Após predição destes dados foi realizada comparação dos dados projetados com dados medidos reais coletados após amostragem inicial de dados. Os resultados apresentados na Tabela 2 foram encontrados.

Tabela 2. Comparativo de dados projetados e medidos

Data e Hora	Val. Previstos	Val. Medidos (Reais)	Erro
18/02/2022 00:00	-22.2551	-21.8750	0.6199
18/02/2022 00:01	-21.9963	-21.6250	0.6287
18/02/2022 00:02	-21.6878	-21.2500	0.5622
18/02/2022 00:03	-21.3038	-21.0000	0.6962
18/02/2022 00:04	-20.8640	-20.8750	0.0110
18/02/2022 00:05	-20.4115	-20.6250	0.2135
18/02/2022 00:06	-20.0231	-20.3750	0.3519
18/02/2022 00:07	-19.7114	-20.2500	0.5386
18/02/2022 00:08	-19.4673	-20.2500	0.7827
18/02/2022 00:09	-19.2764	-20.0000	0.7236
18/02/2022 00:10	-19.1526	-19.8750	0.7224

Para fins práticos, a predição de 10 minutos é insuficiente pois dificultará a tomada de decisão caso alguma anomalia na temperatura seja prevista. Porém, com a frequência dos dados em minutos, ao projetar 6 horas de dados a frente o modelo tende à se aproximar da média da série, conforme demonstra a Figura 11. Este fenômeno será reduzido com a chegada de mais dados que alimentará o modelo melhorando suas projeções.

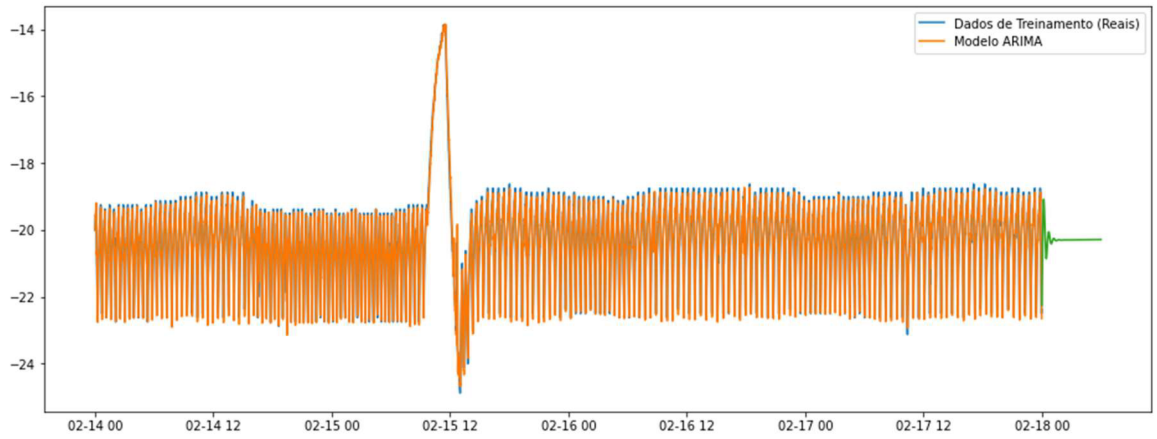


Figura 11. Predição de 6 horas de dados a partir do modelo ARIMA

Uma alternativa é o agrupamento dos dados de entrada ao modelo com uma periodicidade maior para que a janela de predição possua um intervalo que possibilite a tomada de decisão após notificação.

Desta forma, os dados de entrada foram agrupados (mediana) com frequência de 30 minutos para possibilitar, caso seja identificado algum indício de anomalia, enviar a notificação em tempo hábil facilitando a tomada de decisão pelo responsável do setor.

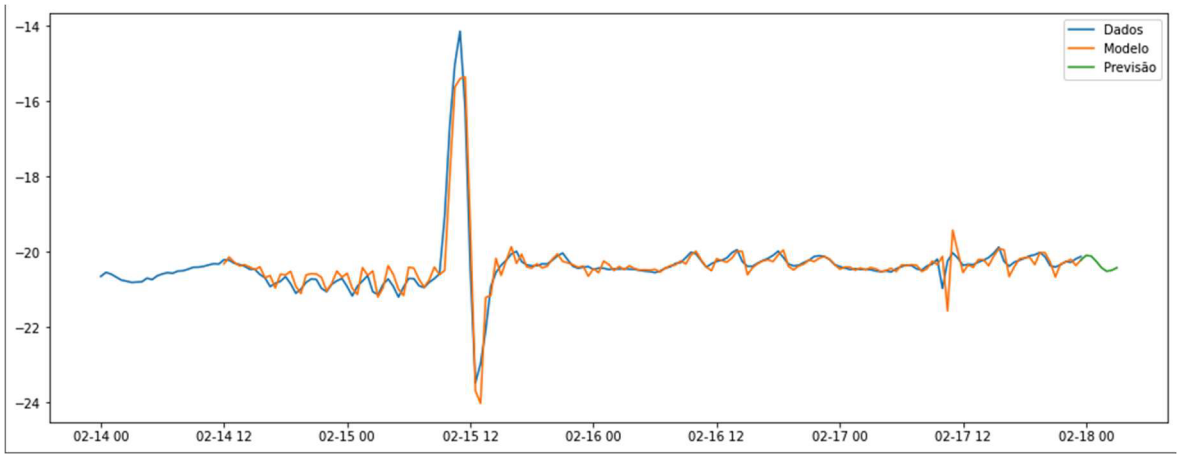


Figura 12. Predição de 3 horas de dados com frequência de 30 minutos a partir do modelo ARIMA

O modelo foi aplicado na massa de dados agrupados a cada 30 minutos e obteve-se o demonstrado na Figura 12 com erro quadrático de 0.02448108857967491.

Para melhoria da modelagem, foi realizada análise da sazonalidade da série. Realizando uma diferença sazonal (Figura 13) não obtivemos resultado melhor dos modelos tendo em vista o erro quadrático médio que resultou em 0.0566 (ARIMA(4,1,2)), valor maior que o obtido no modelo ARIMA(3,1,12).

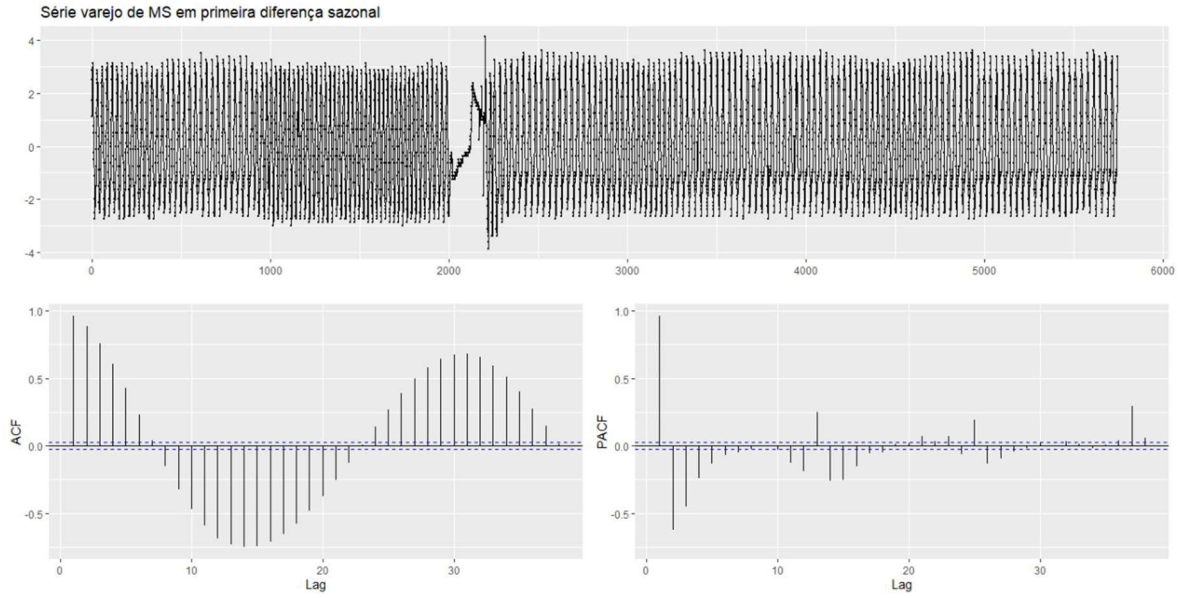


Figura 13. Primeira Diferença Sazonal

6. CONCLUSÕES E TRABALHOS FUTUROS

É notória a carência de tecnologia empregada ao sistema de saúde no que diz respeito ao monitoramento da cadeia do frio. A implementação de um sistema como o proposto neste artigo evitará diversas ocorrências de desperdícios de vacinas. Além dos imunizantes, este mesmo sistema pode ser aplicado no acondicionamento de medicamentos, alimentos, materiais termoláveis como embalagens etc.

Esse artigo apresenta um modelo preventivo, com duas contribuições, sendo a primeira um estudo de predição de falhas e paradas inesperadas através de tecnologias de inteligência artificial, permitindo a criação de um sistema IoT com suporte em nuvem, que realiza a coleta de dados dos dispositivos realizando um

monitoramento contínuo de temperatura em freezers que se encontram distribuídos em diversas instituições de saúde no Brasil.

A segunda contribuição é a apresentação da possibilidade de abstrair o middleware que possui uma camada de aplicação a fim de identificar indícios de falha através da utilização de métodos estatísticos com algoritmos específicos, utilizando modelos preditivos para series temporais.

Como trabalhos futuros, abre-se a janela de oportunidade da conversão e aplicação do algoritmo desenvolvido em ambiente computacional em nuvem para notificações em tempo real dos usuários da plataforma. Somada à esta conversão, será necessário criar um ecossistema de gestão de acessos e máquina de notificações para viabilizar o aviso prévio de anomalias através de SMS, Push Notification, e-mail etc.

Existe também a necessidade da análise de outras grandezas e outros tipos de equipamentos críticos no âmbito da saúde. Por exemplo, os resultados deste trabalho podem ser aplicados e testados no monitoramento da linha de gases em ambiente hospitalar, através da medição da pressão das linhas de oxigênio, óxido nitroso, vácuo etc.

7. AGRADECIMENTOS

Os autores agradecem ao CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico (Outorgas PQ-2 312180/2019-5 INCT em Cibersegurança 465741/2014-2), ao Conselho Administrativo de Defesa Econômica (Outorga CADE 08700.000047/2019-14), à Procuradoria Geral da União (Outorga AGU 697.935/2019), à Procuradoria Geral da Fazenda Nacional (Outorga PGFN 23106.148934/2019-67) e à UnB - Universidade de Brasília (Outorga COPEI 7129). A validação deste trabalho foi garantida pela Vitae Soluções em Engenharia que abriu sua infraestrutura e base de dados para coleta e análise dos dados apresentados neste artigo.

REFERÊNCIAS

CROSEWSKI, Fernanda; LAROCCA, Liliana Müller; CHAVES, Maria Marta Nolasco. Avoidable losses of immunobiologicals in the local instance: reflections on the nursing work process. Saúde em Debate, v. 42, n. 116, p. 203-213, 2018.

DOMINGUES, Carla Magda Allan Santos et al. 46 anos do Programa Nacional de Imunizações: uma história repleta de conquistas e desafios a serem superados., Cadernos de Saúde Pública, v. 36, 2020. 1999.

Brasil. Ministério da Saúde Secretaria de Vigilância em Saúde. Departamento de Vigilância de Doenças Transmissíveis., Coordenação Geral do Programa Nacional de Imunizações. Manual de Rede de Frio. Brasília, DF: Ministério da Saúde; 2017.

PEREIRA, D. D. S. et al. Analysis of the utilization rate and loss of vaccines in the National Immunization Program., Cad Saud Colet, v. 21, n. 4, p. 420-4, 2013.

PATINE, Flávia dos Santos et al. Analysis of vaccine loss due to temperature change., Revista Brasileira de Enfermagem, v. 74, 2021.

WORLD HEALTH ORGANIZATION et al. Temperature sensitivity of vaccines., World Health Organization, 2006.

GUIMARÃES, Eliete Albano de Azevedo et al. Eventos críticos na manutenção da conservação de vacinas., Rev. enferm. UFPE on line, p. 1781-1789, 2018.

LLOYD, John; CHEYNE, James. The origins of the vaccine cold chain and a glimpse of the future., Vaccine, v. 35, n. 17, p. 2115-2120, 2017.

CHOI, Seung Yeoun; KIM, Sean Hay. Finding the Simple Rule for Diagnosing the Operation Fault of Freezer., International Journal of Mechanical Engineering and Robotics Research, v. 8, n. 1, 2019.

WANG, Dong; ZHANG, Lun; ZHANG, Shengjian. Evaluation model of refrigerator cooling performance., In: 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). IEEE, 2017. p. 1565-1569.

HOSSEINI, Sayed Saeed et al. A Practical Approach to Residential Appliances on-Line Anomaly Detection: A Case Study of Standard and Smart Refrigerators., IEEE Access, v. 8, p. 57905-57922, 2020.

GHAFUR, Saira et al. A retrospective impact analysis of the WannaCry cyberattack on the NHS., NPJ digital medicine, v. 2, n. 1, p. 1-7, 2019.

WANGANOO, Leena; SHUKLA, Vinod Kumar. Real-Time Data Monitoring in Cold Supply Chain Through NB-IoT., In: 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE, 2020. p. 1-6.

BERGONZI, Luca et al. Data and knowledge in iiot-based maintenance application. In: ASME International Mechanical Engineering Congress and Exposition., American Society of Mechanical Engineers, 2017. p. V011T15A029.

ROMEIRO, Lucas AW et al. Low-Cost Monitoring Device for Cold-Chain using Edge Computing. In: 2021 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN). IEEE, 2021. p. 274- 279. https://www.statsmodels.org/dev/tsa.html

Morettin, P. A. (2018). Análise de séries temporais: modelos lineares univariados. Brasil: BLUCHER..