**FACIMP WYDEN**

**ANÁLISE DETALHADA DE DADOS: DESVENDANDO NOTÍCIAS FALSAS E VERDADEIRAS PARA IDENTIFICAR PADRÕES E MITIGAR DESINFORMAÇÃO.**

**AMANDA BRANDÃO SOUSA.**

**JOÃO LUCAS DOS SANTOS OLIVEIRA.**

**2023**

**Imperatriz-MA**

Sumário

[**1.** DIAGNÓSTICO E TEORIZAÇÃO 3](#_Toc151742826)

[**Conjunto de dados de notícias falsas e reais** 4](#_Toc151742827)

[2. PLANEJAMENTO E DESENVOLVIMENTO DO PROJETO 4](#_Toc151742828)

[2.1 Plano de trabalho (usando ferramenta acordada com o docente) 5](#_Toc151742829)

[2.2 Descrição da forma de envolvimento do público participante na formulação do projeto, seu desenvolvimento e avaliação, bem como as estratégias pelo grupo para mobilizá-los. 5](#_Toc151742830)

[2.3. Grupo de trabalho (descrição da responsabilidade de cada membro) 6](#_Toc151742831)

[2.5 Recursos previstos: 7](#_Toc151742832)

[2.6 Detalhamento técnico do projeto: 8](#_Toc151742833)

[3. ENCERRAMENTO DO PROJETO 8](#_Toc151742834)

[3.1 Relato Coletivo: 8](#_Toc151742835)

[3.1.1 Avaliação de reação da parte interessada: 9](#_Toc151742836)

[3.2 Relato de Experiência Individual (Pontuação específica para o relato individual): 9](#_Toc151742837)

[3.2.4. REFLEXÃO APROFUNDADA 11](#_Toc151742838)

[3.2.5. CONSIDERAÇÕES FINAIS 11](#_Toc151742839)

# **1.** DIAGNÓSTICO E TEORIZAÇÃO

1.1. Identificação das partes interessadas e parceiros:

Neste projeto, as partes interessadas abrangem:

* **Participantes:** Comunidade online e usuários de redes sociais que consomem notícias.
* **Parceiros:** Instituições de pesquisa, organizações de mídia, e plataformas de redes sociais, com possíveis acordos de cooperação para compartilhamento de dados e acesso às informações.

1.2. Problemática e/ou problemas identificados:

A problemática identificada refere-se à disseminação de informações falsas e sua impactante presença na sociedade digital. A escolha desta questão foi motivada pela necessidade de compreensão e mitigação do impacto das fake news, identificada a partir da análise comparativa entre notícias verdadeiras e falsas. Este estudo baseou-se em duas bases de dados do Kaggle, ambas relacionadas ao mesmo tema, porém abordando as questões de notícias de modo diferente, o que permitiu uma análise mais abrangente e multifacetada sobre o fenômeno das informações falsas na era digital. Os problemas identificados incluem:

* **Propagação de Fake News:** A presença significativa de notícias falsas circulando nas redes sociais e plataformas online.
* **Impacto na Credibilidade:** A influência negativa das fake news na confiabilidade das informações disponíveis para os usuários.
* **Necessidade de Distinção:** A demanda por métodos mais eficazes de distinguir notícias verdadeiras das falsas para promover uma compreensão precisa dos eventos.

1.3. Justificativa:

A problemática identificada sobre a disseminação de notícias falsas e sua influência na sociedade digital é crucial academicamente, alinhando-se à aprendizagem baseada em projetos. Este projeto permite a aplicação prática dos conhecimentos adquiridos, enfatizando a compreensão dos mecanismos de propagação de desinformação e a identificação de estratégias para mitigar seu impacto. A relação com o curso se manifesta na aplicação direta de conceitos teóricos em análise de dados, estatística e interpretação de conteúdo digital. A motivação para essa investigação reside na urgência de compreender e enfrentar os desafios impostos pela disseminação de informações enganosas na era digital, contribuindo para a formação crítica e ética de indivíduos aptos a discernir entre informações verdadeiras e falsas.

1.4. Objetivos/resultados/efeitos a serem alcançados:

**Análise dos Padrões de Propagação de Notícias Falsas e Verdadeiras:**

Identificar, analisar e documentar minuciosamente os padrões e mecanismos de propagação de notícias falsas e verdadeiras em ambientes digitais diversos. Isso inclui a compreensão das estratégias de disseminação, perfis de usuários envolvidos e os contextos em que essas notícias são mais prevalentes.

**Desenvolvimento de Estratégias Mitigadoras:**

Desenvolver e propor estratégias, ferramentas e abordagens inovadoras para mitigar a disseminação de informações enganosas. Essas soluções buscam fortalecer a capacidade do público em discernir entre conteúdo autêntico e desinformação, promovendo uma postura crítica e consciente ao consumir informações online.

**Conscientização e Capacitação do Público:**

Alertar e motivar ativamente o público-alvo a buscar uma maior formação nas redes sociais, incentivando a verificação de fontes e a busca por conteúdo de qualidade, diante da frequente presença de notícias falsas em proporções semelhantes às notícias verdadeiras.

1.5 **Referencial Teórico:**

**Notícias Falsas: FakeNewsNet - Fonte dos Dados Utilizados.**

Os dados analisados neste projeto foram obtidos do [FakeNewsNet](https://www.kaggle.com/datasets/algord/fake-news?rvi=1), uma plataforma hospedada no Kaggle, reconhecida por fornecer conjuntos de dados fundamentais para o estudo de notícias falsas e reais. Esses conjuntos de dados oferecem informações relevantes sobre a disseminação de informações enganosas e verdadeiras em ambientes online.

Fonte: <https://www.kaggle.com/datasets/algord/fake-news?rvi=1>

# **Conjunto de dados de notícias falsas e reais:**

Dados meticulosamente compilados e preparados, esse conjunto de dados oferece uma oportunidade única para investigar os padrões de desinformação e autenticidade nas notícias online. A plataforma Kaggle, reconhecida por sua diversidade de conjuntos de dados bem estruturados, facilitou o acesso a informações cruciais para este estudo.

Fonte: <https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset>

# 2. PLANEJAMENTO E DESENVOLVIMENTO DO PROJETO:

## 2.1 Plano de trabalho (usando ferramenta acordada com o docente)

**Ações a serem Executadas para Alcançar os Objetivos do Projeto:**

* Realização de pesquisas para conduzir uma análise detalhada.
* Utilização do Kaggle como fonte para a identificação e análise dos dados.
* Definição de prazos específicos para cada etapa do projeto, alinhados aos objetivos estabelecidos.

**Prazos para as Entregas dos Textos de Cada Item deste Roteiro de Extensão:**

**Comparação entre Fake News e Informações Verificadas:**

* Prazo de entrega do texto: 1 dia
* Prazo da análise detalhada: 1 dia
* **Velocidade de Propagação de Tweets.**

**Top 46 Domínios: Distribuição Percentual de Notícias Falsas e Verdadeiras:**

* Prazo de entrega do texto: 1 dia.
* Prazo da análise detalhada: 1 dia
* **Comparação de Engajamento Médio.**

**Porcentagem de Notícias Falsas versus Verdadeiras por Fonte (Top 46):**

* Prazo de entrega do texto: 2 dias
* Prazo da análise detalhada: 1 dia
* **Apresentação dos Dilemas e Semelhanças das Fake News.**

**4-gramas mais comuns em notícias falsas e verdadeiras:**

* Prazo de entrega do texto: 2 dias
* Prazo da análise detalhada: 1 dia
* Palavras mais frequentes.

Esse detalhamento visa destacar as pesquisas necessárias para a análise, utilizando o Kaggle como fonte principal de dados para identificação e análise. Os prazos definidos para a entrega de cada texto estão alinhados aos objetivos do projeto e à problemática identificada, variando de 1 a 2 dias para cada etapa.

## Descrição da forma de envolvimento do público participante na formulação do projeto, seu desenvolvimento e avaliação, bem como as estratégias pelo grupo para mobilizá-los.

Antes da execução deste projeto, foi conduzida uma pesquisa detalhada intitulada " UM ESTUDO DETALHADO: EXPERIÊNCIA DE JOVENS APRENDIZES". O objetivo principal era coletar dados sobre a experiência dos participantes do programa Jovem Aprendiz, abordando aspectos como faixa etária, gênero, nível de satisfação e classe social predominante entre os envolvidos.

**Processo da Pesquisa:**

* **Aquisição de Dados:** Um formulário de pesquisa (link: <https://forms.gle/gAtrecMsJK4TYV6z5>) foi desenvolvido para obter informações dos jovens aprendizes, visando garantir a confidencialidade e anonimato dos participantes. Os dados coletados incluíram informações sobre faixa etária, gênero, classe social e outros aspectos relacionados ao programa.

**Base de Dados Obtida pelo Kaggle para Análise de Notícias Falsas e Verdadeiras:**

Após a recusa inicial da empresa para análise, toda a base de dados deste projeto foi coletada por meio da plataforma Kaggle. Isso possibilitou a construção de uma base consistente e informativa para a análise proposta, alinhada às informações disponíveis nessa plataforma. O projeto foi concebido a partir de um esforço individual, considerando a necessidade de compreender e enfrentar os desafios apresentados pelas notícias falsas. A base de dados foi construída de forma individual, com foco no número expressivo de ocorrências de fake news, oferecendo um panorama detalhado e abrangente sobre essa problemática.

Durante a formulação do projeto, a coleta de informações e a definição de estratégias foram conduzidas individualmente, buscando compreender e interpretar os padrões presentes nos dados coletados. Essa abordagem direta contribuiu significativamente para a identificação de pontos-chave que nortearam a análise comparativa entre notícias falsas e verdadeiras.

## 2.3. Grupo de trabalho (descrição da responsabilidade de cada membro)

Dado que este projeto foi conduzido individualmente, desde a coleta de dados até a análise e conclusões. Sendo assim, o relato de aprendizagem e as contribuições ao projeto refletem seu esforço individual na concepção e execução de todas as etapas do trabalho.

O envolvimento único no projeto permitiu uma compreensão abrangente da temática, aprofundando-se na identificação de padrões de propagação de notícias falsas e verdadeiras, análise da velocidade de propagação de tweets, exploração da distribuição percentual por domínio, comparação do engajamento médio entre notícias falsas e verdadeiras, e a observação da porcentagem de notícias falsas versus verdadeiras por fonte.

Essa abordagem individual possibilitou a realização das análises de forma minuciosa e consistente, contribuindo para o entendimento dos dilemas apresentados pelas fake news e a compreensão das semelhanças nos dados, enfatizando a importância do alerta e da identificação dessas notícias para a conscientização do público.

**2.4. Metas, Critérios ou Indicadores de Avaliação do Projeto:**

**Análise dos Padrões de Propagação de Notícias Falsas e Verdadeiras:**

* Identificação detalhada dos mecanismos de disseminação de ambas as categorias de notícias em ambientes digitais variados.
* Avaliação dos contextos em que as notícias falsas e verdadeiras são mais prevalentes.

Critérios de Avaliação:

* Completa identificação dos padrões de propagação.
* Profundidade na análise dos contextos de disseminação.

**Desenvolvimento de Estratégias Mitigadoras:**

* Formulação de estratégias inovadoras e ferramentas para conter a disseminação de informações enganosas.
* Proposta de abordagens que fortaleçam a capacidade do público em discernir entre conteúdo autêntico e desinformação.

Critérios de Avaliação:

* Originalidade e efetividade das estratégias propostas.

**Conscientização e Capacitação do Público:**

* **Ativação de campanhas de conscientização para alertar e motivar o público-alvo sobre a importância da formação crítica nas redes sociais.**
* **Incentivo à verificação de fontes e à busca por conteúdo de qualidade.**

## Recursos previstos:

**Materiais:**

* **Computadores e Acesso à Internet:** Fundamental para a análise de dados, processamento de informações e acesso aos conjuntos de dados fornecidos.
* **Software de Análise de Dados:** Ferramentas como Python (com bibliotecas como Pandas, NLTK, Matplotlib), o Google Colab e o Visual Studio, são essenciais para análise e visualização dos dados.

**Institucionais:**

* Autorização de Uso de Dados: Garantia da utilização ética e legal dos conjuntos de dados obtidos, respeitando direitos autorais e de uso.

## Detalhamento técnico do projeto:

**Coleta de Dados do Kaggle:**

* Aquisição dos conjuntos de dados do Kaggle, obtendo informações sobre os artigos de notícias, incluindo títulos, URLs, domínios de origem, número de retuítes e rótulos (real ou falso).

**Pré-processamento dos Dados:**

* Limpeza dos dados para remoção de informações irrelevantes, tratamento de valores ausentes, padronização de formatos e remoção de duplicatas.
* Transformação dos dados para formatos adequados para análise, garantindo consistência e integridade.

**Análise Exploratória de Dados (EDA):**

* Exploração inicial dos dados para compreender sua estrutura, distribuição e características principais.
* Identificação de padrões preliminares e tendências nos conjuntos de dados de notícias falsas e verdadeiras.

**Conscientização e Capacitação do Público:**

* **Ativação do público-alvo para buscar maior formação nas redes sociais, incentivando a verificação de fontes e a busca por conteúdo de qualidade diante da frequente presença de notícias falsas.**

**Essas etapas técnicas são fundamentais para a compreensão e a construção de soluções e estratégias que ajudam a lidar com a disseminação de notícias falsas, apresentando um panorama mais crítico e informado aos usuários.**

# 3. ENCERRAMENTO DO PROJETO

## Relato Coletivo:

**Ao conduzir este estudo de maneira individual, as considerações sobre o alcance dos objetivos sociocomunitários estabelecidos foram centralizadas na compreensão aprofundada das dinâmicas de disseminação de informações falsas e verdadeiras. A análise minuciosa dos dados permitiu uma percepção mais clara das tendências e diferenças na propagação dessas informações, fornecendo insights relevantes para a compreensão do fenômeno das fake news.**

**Embora tenha sido uma abordagem individual, o objetivo foi contribuir para a conscientização sobre a importância da verificação de fontes e do pensamento crítico na sociedade. Essa análise buscou não apenas identificar padrões na propagação de informações, mas também ressaltar a necessidade de estratégias eficazes para enfrentar o desafio das fake news, promovendo um ambiente informacional mais confiável para a comunidade.**

### 3.1.1 Avaliação de reação da parte interessada:

A condução de uma avaliação da reação diante dos dados obtidos do Kaggle assume um papel crucial na validação e compreensão da análise realizada. Esta fase engloba a aplicação de metodologias como formulários estruturados, entrevistas gravadas em áudio ou vídeo, ou ainda depoimentos audiovisuais, com a finalidade de elucidar as percepções e reflexões dos participantes em relação aos dados analisados.

O propósito principal dessa etapa reside em capturar insights qualitativos e quantitativos sobre a utilização dos dados provenientes do Kaggle. Além disso, visa avaliar a significância e relevância dessas informações no contexto das notícias falsas e verdadeiras, além de compreender o impacto desses dados na percepção dos participantes.

O retorno obtido por meio dessas estratégias de avaliação é primordial para comprovar a eficácia e utilidade dos dados do Kaggle na análise desenvolvida. Esta análise qualitativa contribui para um entendimento mais amplo sobre a absorção e interpretação das informações apresentadas.

## Relato de Experiência Individual (Pontuação específica para o relato individual):

1. **Contextualização Pessoal:** Ao longo do projeto, pude conduzir todas as etapas individualmente, desde a análise e coleta de dados até a criação e visualização dos mesmos. Esse trabalho me proporcionou uma visão ampla sobre o tema das notícias falsas e verdadeiras, bem como me permitiu mergulhar profundamente na compreensão dos padrões de propagação e análise dessas informações.
2. **Responsabilidades e Contribuições:** Como responsável por todas as etapas do projeto, conduzi a coleta de dados, análise estatística, criação de gráficos e interpretação dos resultados. A autonomia nesse processo me permitiu aprofundar meu conhecimento na área, além de desenvolver habilidades técnicas valiosas no manuseio de conjuntos de dados e ferramentas analíticas.
3. **Aprendizados e Desafios:** Essa experiência foi desafiadora e enriquecedora. Aprendi a lidar com dados complexos, a identificar e interpretar padrões de propagação de notícias e a comunicar resultados de forma clara e eficaz. Enfrentei desafios na organização e interpretação dos dados, mas cada obstáculo se tornou uma oportunidade de aprendizado.

3.2.1. Contextualização

Minha participação foi fundamental na análise e compreensão de diversos aspectos relacionados à disseminação de fake news e notícias autênticas. Por meio de análise de gráficos e interpretação de dados, pude contribuir para uma compreensão mais abrangente das tendências, padrões e diferenças na propagação dessas informações.

Minha atuação focou-se na interpretação e análise dos gráficos que representavam a disseminação de notícias falsas e verdadeiras, destacando as semelhanças, diferenças e os possíveis impactos na sociedade. A análise criteriosa desses dados permitiu uma compreensão mais aprofundada sobre como as informações são disseminadas e percebidas, fornecendo insights valiosos para enfrentar o desafio das fake news.

Minha participação neste projeto não apenas enriqueceu minha compreensão sobre a problemática da desinformação, mas também contribuiu para oferecer perspectivas e insights cruciais que podem ser fundamentais na identificação e mitigação da propagação de notícias falsas.

3.2.2. Metodologia

O trabalho foi conduzido de forma individual, com o uso de plataformas como o Kaggle, Excel, Google Colab e Visual Studio para análise, processamento e visualização dos dados. Meu envolvimento nesse projeto foi desde a coleta inicial de informações até a interpretação e apresentação dos resultados.

3.2.3. Resultados e Discussão

Durante a experiência, aprofundei-me na compreensão dos dados, identificando padrões intrigantes de disseminação de informações falsas e verdadeiras. Encontrei desafios na interpretação e organização dos dados, mas cada dificuldade foi uma oportunidade valiosa de aprendizado.

Os resultados obtidos reforçaram a importância do engajamento crítico online e da conscientização sobre a verificação de fontes. Essa experiência me permitiu desenvolver habilidades analíticas e de comunicação, além de fortalecer meu interesse na análise de dados.

### REFLEXÃO APROFUNDADA

A análise detalhada dos gráficos revelou insights importantes sobre a propagação de notícias falsas e verdadeiras. Foi fascinante observar como os padrões de disseminação se assemelham e diferem entre esses tipos de informações. A experiência ressaltou a complexidade de lidar com a desinformação e a importância de promover a educação em mídia para combater esse problema. A análise dos dados visuais enfatizou a necessidade crítica de verificar fontes e desenvolver um olhar mais questionador sobre as informações consumidas diariamente. Isso reforça a necessidade contínua de abordar estratégias para mitigar os impactos das fake news, garantindo assim um ambiente informacional mais confiável para todos.

### CONSIDERAÇÕES FINAIS

Ao finalizar este estudo, algumas oportunidades e direções para trabalhos futuros se destacam:

1. **Parcerias Estratégicas:** Explorar parcerias com instituições educacionais, agências de checagem de fatos e plataformas de mídia social para desenvolver programas educacionais sobre literacia digital, capacitando as pessoas a identificar e combater a desinformação.
2. **Pesquisa Adicional:** Investir em pesquisas mais aprofundadas para aprimorar algoritmos de detecção de fake news, utilizando Inteligência Artificial e aprendizado de máquina para desenvolver ferramentas mais eficazes na identificação precoce de notícias falsas.
3. **Desenvolvimento Tecnológico:** Explorar soluções tecnológicas inovadoras, como sistemas de verificação de fontes automatizadas, análise semântica avançada e uso de blockchain para rastrear a origem e autenticidade das informações.
4. **Iniciativas de Extensão:** Realizar workshops e palestras educativas em escolas, universidades e comunidades, visando promover o pensamento crítico e fornecer ferramentas para avaliar a veracidade das notícias.
5. **Colaborações Interdisciplinares:** Fomentar colaborações interdisciplinares entre cientistas de dados, pesquisadores sociais e profissionais de comunicação para abordar a desinformação de maneira abrangente e multifacetada.
6. **Políticas de Plataformas Digitais:** Pressionar por políticas mais rígidas nas plataformas online, incentivando a transparência na distribuição de conteúdo e o combate ativo às fake news.

Essas ações têm o potencial de fortalecer a capacidade de identificação e prevenção de informações falsas, promovendo um ambiente informacional mais confiável e crítico para o público em geral.