

Detecção de Fake News através de Machine Learning

Letícia A. de Azevedo

Graduando em Ciência da Computação

Centro Universitário IESB

Brasília, DF

lelealvesaz@gmail.com

Lucas R. Pereira

Graduando em Ciência da Computação

Centro Universitário IESB

Brasília, DF

lucas_rabelo.pereira@hotmail.com

Tercilio dos S. A. Fonseca

Graduando em Ciência da Computação

Centro Universitário IESB

Brasília, DF

terciliosantos@gmail.com

Tiago do N. Batista

Graduando em Engenharia da Computação

Centro Universitário IESB

Brasília, DF

tiagonbt@gmail.com

Welisson H. Lima

Graduando em Ciência da Computação

Centro Universitário IESB

Brasília, DF

welissonsacr2000@gmail.com

Abstract—Fake news is false news published by media outlets posing as true information. They spread through mass sharing on social media and have the power to change people's minds based on their content. With the use of Artificial Intelligence we can train a model to detect if a news is true or false.

Resumo—Fake news são notícias falsas publicadas por veículos de comunicação se passando por uma informação verdadeira. Elas se espalham por meio de compartilhamento em massa nas redes sociais e têm o poder de mudar a opinião das pessoas conforme seu conteúdo. Com o uso da Inteligência Artificial podemos treinar um modelo para detectar se uma notícia é verdadeira, ou falsa.

Index Terms—Palavras-chave: Fake news, Notícias falsas, Inteligência artificial, Machine learning.

I. INTRODUÇÃO

Notícia é uma informação jornalística que trata sobre um fato importante para a população. Ela abrange várias áreas importantes como, política, saúde, economia, educação e outros. Com o avanço da tecnologia, as notícias se tornaram mais

acessíveis por meio dos sites e redes sociais, no entanto não se sabe até que ponto essa notícia é verdadeira. Em 2018 foi notável a circulação em massa de Fake News através do aplicativo WhatsApp por meio de robôs para disseminar informações duvidosas nas eleições presidenciais do presidente do Brasil, Jair Bolsonaro. Este trabalho objetiva aplicar técnicas de machine learning e processamento de linguagem natural para identificar se uma notícia é falsa ou verdadeira.

II. CONTEXTO

Esta modalidade de informação possui um alto poder viral¹, se espalhando 70% mais rápido que uma notícia verdadeira, atingindo principalmente as classes mais vulneráveis e com menor grau de alfabetização, que em grande parte, se informam através de redes sociais. Por possuírem uma linguagem simples e apelarem para o emocional, o espectador ou leitor

¹**viral:**Diz-se de algo que se espalha como se fosse um vírus. Porto Editora. Dicionário Infopédia da Língua Portuguesa.

consome este material sem confirmar, em muitos casos, a veracidade dos fatos.

No Brasil e no mundo, um dos principais alvos dos criadores de Fake News é a esfera política, onde são utilizadas informações muitas vezes verdadeiras misturadas com dados falsos com o uso de palavras, imagens, datas etc. Isso credibiliza a informação pelo leitor/espectador, que em muitos casos não faz uma verificação em outras fontes confiáveis e compartilha com colegas de trabalho, amigos e/ou familiares.

Os criadores deste tipo de informação normalmente são ex-jornalistas, profissionais da área de tecnologia, profissionais de marketing etc. Sendo contratados por adversários políticos, entidades político-partidárias, empresários etc., para produzirem conteúdos adulterados gerando descrédito a pessoa pública ou partido político alvo.

Apesar de existir desde os primórdios, nos últimos anos vem se tornando um grande desafio a ser enfrentado. Neste ano (2022), estima-se que o uso e criação de Fake News no processo eleitoral brasileiro será de muitas, uma das grandes preocupações, como ocorreu nas eleições de 2018(Eleições Estaduais e Presidenciais) e 2020(Eleições Municipais), o que poderia comprometer uma disputa responsável aos cargos públicos concorridos.

O combate a desinformação vem sendo feito por vários órgãos brasileiros, um destes é o TSE (Tribunal Superior Eleitoral), que trabalha no cumprimento de leis, regimentos e buscando o comprometimento de plataformas digitais com acordos para tentar diminuir a disseminação de informações falsas no período eleitoral. Estas plataformas criam algoritmos que avaliam palavras, número de compartilhamentos, público-alvo, grupos de usuários, região do país. Estreitando assim as possibilidades de disparo em massa de mensagens com conteúdo duvidoso.

Existem também diversos veículos de checagem de notícias, que realizam buscas nas mais variadas mídias sempre identificando e classificando informações na medida do que é verdadeiro, ou que possa conter conteúdo verdadeiro misturado a conteúdo adulterado. Estas agências de checagem permitem ao leitor/espectador esclarecer-se e evitar conteúdos danosos a sociedade.

Segundo a pesquisa do comitê gestor da internet no país, cerca de 81% dos brasileiros tem acesso à internet, e 66% usam para ler notícias nas mídias sociais, dessa forma, a população fica vulnerável ao acesso conteúdos falsos

III. PROBLEMA

As fake news são um grande problema em vários países, inclusive no Brasil. Na era tecnológica, onde as redes sociais são usadas pela maioria da população, a disseminação acontece em grande escala, inclusive na época das eleições, onde a quantidade de notícias falsas se intensificam com o objetivo de influenciar o voto do eleitor.

Os principais veículos de imprensa tentam combater a disseminação dessas notícias e informar sobre a importância de verificar se uma notícia realmente é verdadeira antes de compartilhá-la na internet. Entretanto, esses esforços não estão

sendo suficientes, visto que a quantidade de notícias falsas se espalham de maneira alarmante, segundo o jornal do comércio..

IV. OBJETIVO GERAL

Com o auxílio da inteligência artificial, alguns modelos serão treinados para detectar, com uma acurácia satisfatória, tendo como requisito a veracidade da notícia, classificando-a em verdadeira ou falsa.

V. OBJETIVO ESPECÍFICO

- * Alimentar um base de dados com notícias falsas e verdadeiras.
- * Separar um percentual de 70% para consulta e 30% para treinamento.
- * Treinar algoritmo SVM-Support Vector Machine com Dataset de Notícias.
- * Treinar algoritmo Naive Bayes com Dataset de Notícias.
- * Treinar algoritmo de Vetorização com Dataset de Notícias.
- * Treinar algoritmo Chi-Square com Dataset de Notícias.
- * Treinar algoritmo Floresta Aleatória com Dataset de Notícias.
- * Verificar a acurácia de cada um dos algoritmos.
- * Determinar qual algoritmo foi mais eficiente no programa de treinamento e consulta.

VI. REFERENCIAL TEÓRICO

A. SVM - Support Vector Machine

- A máquina de vetores de suporte têm como entrada um conjunto de dados e consegue prever, para cada dado de entrada, qual de duas possíveis classes a entrada faz parte. O SVM é classificado como linear binário não probabilístico.

B. Naive Bayes

- Naive Bayes é uma técnica simples para construir modelos e classificadores, representado por vetores, onde possui os rótulos que são gerados a partir de um conjunto finito. Uma das suas vantagens é que ele requer apenas um conjunto pequeno de dados a serem treinados para estimarem os parâmetros necessários para uma classificação.

C. Floresta Aleatória

- Florestas aleatórias ou florestas de decisão aleatória que opera construindo uma infinidade de árvores de decisão na hora do treinamento do modelo. Nas tarefas de classificação a saída da floresta aleatória é a classe que foi selecionada pela maioria das árvores. Já a regressão, a previsão média das árvores individuais é retornada. Geralmente as árvores aleatórias corrigem o hábito das árvores de decisão de se ajustarem ao seu conjunto de treinamento.

D. Vetorização

- É a distribuição representativa de dados em vetores a fim de auxiliar algoritmos de aprendizado a obter uma melhor performance agrupando itens similares [Mikolov et al. 2013].

E. Chi-Square

- A Chi-Squared (χ^2) é um teste de hipótese estatística que é válido para realizar quando a estatística de teste é qui-quadrado distribuída (χ^2) sob a hipótese nula. É utilizado para determinar se há diferença estatisticamente significativa entre as frequências esperadas e as observadas.

VII. TRABALHOS CORRELATOS

Esta seção apresenta os trabalhos correlatos que associam a aplicação da inteligência artificial para detectar fake news.

- FakeChatBot - É uma aplicação em português que faz o uso de chatbots para auxiliar usuários a detectar notícias falsas na Internet. A utilização da aplicação é simples, os usuários inserem um trecho textual uma api recebe este texto e realiza uma busca em um banco de dados. Este banco de dados contém notícias falsas. Por fim retorna ao usuário se a notícia é falsa ou verdadeira [De Lima Araujo et al.2018].
- Hierarchical attention networks (HAN)- Este modelo cria uma representação efetiva do texto e, diferentemente de outras arquiteturas de redes, permite a visualização dos dados de saída. Através de um mapa de calor obtêm-se uma indicação do porquê da classe escolhida, dando destaque as palavras e sentenças consideradas mais importantes para a classificação. Neste projeto também usa-se de modelos pré-treinados para a criação das palavras vetorizadas, usando o algoritmo Word2vec. Após a construção, o modelo foi treinado e avaliado com base em um banco de dados de notícias pré classificadas.[GUARISE, L et al.2019]

VIII. RESULTADOS ESPERADOS

Criação de um software capaz de detectar com pelo menos 80% de acurácia na veracidade da notícia pesquisada.

IX. REFERÊNCIAS

1. Laura B. M. R., Michele C. B., Adriano F. Arquitetura de machine learning para análise de reportagens textuais em redes sociais para a detecção de fake news - Hortolândia – São Paulo – SP – Brasil
2. GUARISE, L. Detecção de notícias falsas usando técnicas de deep learning. 2019. 49 f.Monografia (Graduação) – Instituto de Ciências Matemáticas e de Computação (ICMC/USP), São Carlos – SP.
3. O que são Fake News? Reportagem por Lorraine Vilela Campos, Equipe Brasil Escola.<https://brasilecola.uol.com.br/curiosidades/o-que-sao-fake-news.htm>
4. O que esperar da produção de fake News nas eleições de 2022? Reportagem de Gabriela Amorim. Jornal Brasil

de Fato. <https://www.brasildefatoba.com.br/2022/03/04/o-que-esperar-da-producao-de-fake-news-nas-eleicoes-de-2022>

5. Datafolha: 60% dizem que Fake News podem influenciar muito o resultado das eleições. Reportagem por Joelmir Tavares. <https://www1.folha.uol.com.br/poder/2022/03/datafolha-60-dizem-que-fake-news-podem-influenciar-muito-o-resultado-das-eleicoes.shtml>