

PARIKH, Shivam B.; ATREY, Pradeep K. Media-rich fake news detection: A survey. In: **2018 IEEE conference on multimedia information processing and retrieval (MIPR)**. IEEE, 2018. p. 436-441.

O texto introduz que as fake News são um tópico que cresceu bastante ao redor do mundo. Essa afirmação é seguida de um exemplo de uma fake News titulada: “Os palestinos reconhecem o Texas como parte do México?” (<https://www.snopes.com/fact-check/palestinians-texas-mexico/>), que ganhou muito poder ao ser espalhada em várias redes sociais que tinham o objetivo de fazer com que as pessoas acreditassem nela. Os fatores título e imagem ajudaram a convencer os leitores a acreditar que a história era verdadeira.

As redes sociais se provam como o maior veículo de dispersão de fake news. A notícia sobre os palestinos reconhecerem o Texas teve mais de 1.5 milhões de likes apenas em 4 dias após ser publicada. Estudos apresentados no artigo mostram que grande parte das notícias que circulam por redes sociais são formadas de manchete (normalmente texto em negrito), conteúdo multimídia (imagem, vídeo, áudio etc.), e conteúdo (a real notícia em texto). Os dois primeiros componentes são muito mais efetivos e destacados, um estudo apresentado pelo autor demonstra que 70% dos usuários da rede social Facebook apenas leem a manchete antes de comentar e compartilhar. Resolver esse problema de disseminação de fake news não é um trabalho fácil, na medida que a mídia tradicional vem perdendo espaço para as redes sociais e a internet, no qual qualquer pessoa pode divulgar qualquer coisa.

De acordo com o artigo o aumento das fake news também tem o fator a forma como as pessoas estão consumindo notícias. Em 2017 cerca de dois terços dos Estados Unidos veem notícias por mídias sociais e outros meios como e-mail, sites, podcasts e rádios. O autor segue falando sobre os tipos de dados presentes em notícias, sendo eles: texto, multimídia, hyperlinks e áudio.

Após isso o texto fala sobre as fake News, no qual existem algumas categorizações dessas notícias falsas, no qual podem ser baseadas em visual, usuário, postagem, rede, conhecimento, estilo e posição. Grande parte dos existentes detectores de notícias falsas de baseiam na extração de características da notícia. Em uma abordagem baseada em linguística, são extraídos n-gramas (uma sequência contígua de n itens de uma determinada amostra de texto ou fala), pontuação, legibilidade e sintaxe a fim de classificar a notícia. Em métodos baseados em fraudes é aplicada a Teoria da estrutura retórica e Modelo Vetorial em Sistemas de Recuperação da Informação para criar uma árvore hierárquica da informação. Na detecção baseada em análise de agrupamento de dados (clustering) é usado um algoritmo para comparar e dividir grandes conjuntos de dados. Nos modelos preditivos é usado notícias para treinar uma rede que pode chegar a 70% de precisão na base de dados de treino e 56% na base de dados de teste. Na abordagem baseada em conteúdo é usado análise léxica e semântica para verificar outras fontes e ver como o texto é formado fazendo, assim a classificação.

Alguns sites são grandes bases de dados de fake news e o trabalho cita BuzzFeedNEws, LIAR, PHEME e CREDBANK.