

Resumos Técnicos – Fake News Detection

Resumo técnico (formato ABNT) — Detection of fake news from social media using support vector machine learning algorithms (Sudhakar & Kaliyamurthie, 2024)

SUDHAKAR, M.; KALIYAMURTHIE, K. P. Detection of fake news from social media using support vector machine learning algorithms. Measurement: Sensors, v. 32, p. 101028, 2024. O estudo propõe um sistema automatizado de detecção de notícias falsas em redes sociais utilizando técnicas de machine learning e deep learning. O objetivo é identificar postagens fraudulentas — especialmente relacionadas à pandemia de Covid-19 — a partir de textos publicados no Twitter. O conjunto de dados utilizado contém 1.375.592 tweets, que foram submetidos a um rigoroso processo de pré-processamento textual, incluindo normalização, remoção de ruído, tokenização e extração de atributos por meio da técnica TF-IDF (Term Frequency–Inverse Document Frequency). Foram implementados e comparados diferentes algoritmos de aprendizado supervisionado: Logistic Regression, Naïve Bayes, Support Vector Machine (SVM), Decision Tree e Random Forest, além de modelos de deep learning como Long Short-Term Memory (LSTM) e Convolutional Neural Network (CNN). O desempenho dos modelos foi avaliado por métricas de accuracy, precision, recall e F1-score. Os resultados evidenciam que o SVM apresentou o melhor desempenho geral, alcançando 98 % de acurácia, seguido pela Logistic Regression (95 %). O modelo Naïve Bayes obteve 74 % de acurácia, enquanto o LSTM teve desempenho inferior (65 %), indicando baixa adequação de arquiteturas profundas ao tipo de dado analisado. Conclui-se que técnicas clássicas de machine learning, especialmente SVM e Logistic Regression, são mais eficientes na detecção de fake news baseadas em texto curto, quando comparadas a modelos de deep learning aplicados ao mesmo domínio. O trabalho sugere, para pesquisas futuras, o uso de abordagens híbridas e conjuntos de dados multilíngues a fim de aprimorar a robustez dos classificadores. Palavras-chave: Fake news; Redes sociais; Machine learning; Classificação; TF-IDF; SVM.

Resumo técnico (formato ABNT) — Leveraging Machine Learning for Fake News Detection (Masciari et al., 2020)

MASCIARI, E.; MOSCATO, V.; PICARIELLO, A.; SPERLÌ, G. Leveraging Machine Learning for Fake News Detection. In: INTERNATIONAL CONFERENCE ON DATA SCIENCE, TECHNOLOGY AND APPLICATIONS (DATA 2020). Anais [...]. Lisboa: SCITEPRESS – Science and Technology Publications, 2020. p. 151–157. DOI: 10.5220/0009767401510157. O artigo apresenta um framework completo para detecção de notícias falsas que combina técnicas de machine learning e deep learning aplicadas a conteúdos textuais e multimídia provenientes de redes sociais e sites de notícias. O objetivo é classificar automaticamente notícias como verdadeiras ou falsas, a partir da análise do texto e das imagens associadas. A arquitetura proposta é composta por quatro módulos principais: (1) data ingestion, responsável pela coleta de dados por meio de APIs

e crawlers; (2) pré-processamento, que realiza limpeza, agregação e armazenamento dos dados com o uso das ferramentas Apache Kafka e Cassandra DB; (3) NLP processing, voltado à classificação textual utilizando extração de características via TF-IDF e algoritmos supervisionados; e (4) multimedia processing, destinado à identificação de imagens falsas através de técnicas de deep learning e Error Level Analysis (ELA). Os experimentos foram conduzidos sobre dois datasets amplamente utilizados: LIAR, com 12,8 mil declarações rotuladas manualmente do site Politifact.com, e FakeNewsNet (subconjunto PolitiFact), contendo cerca de 800 notícias políticas rotuladas como verdadeiras ou falsas. Após o pré-processamento e a vetorização textual, diversos algoritmos de machine learning foram comparados — Logistic Regression, Naïve Bayes, Decision Tree, Random Forest e SVM — bem como modelos de deep learning como Google BERT e Convolutional Neural Network (CNN). Os resultados demonstram que o algoritmo Logistic Regression apresentou o melhor desempenho global, com maior accuracy, precision e F1-score em ambos os conjuntos de dados. O bom desempenho é atribuído ao pré-processamento rigoroso e à seleção eficiente de features via TF-IDF. Modelos baseados em árvores mostraram menor eficiência e maior custo computacional. Conclui-se que abordagens clássicas de machine learning — especialmente Logistic Regression — são altamente eficazes na detecção de fake news textuais, enquanto o uso combinado de análise multimodal (texto + imagem) amplia o potencial de identificação de conteúdos falsos. Os autores indicam como trabalhos futuros a incorporação de características de perfis de usuários e análise dinâmica de propagação de notícias nas redes sociais. Palavras-chave: Fake news; Machine learning; Classificação; Processamento de Linguagem Natural; TF-IDF; Logistic Regression.