Classificação de manchetes de cunho violento por meio de mineração de texto

Tiago Battiston **Hugo Lara Urdaneta**
Depto de Engenharia de Controle Automação e Computação, UFSC, Blumenau,
E-mail: tiago.battiston@gmail.com, hugo.lara.urdaneta@ufsc.br,

RESUMO

02/12/2025

O processo de **mineração de texto** pode ser definido como a extração de conhecimentos e padrões não triviais a partir de documentos textuais não estruturados [2]. Trata-se da identificação automática de padrões relevantes em dados textuais, permitindo a descoberta de informações implícitas nos textos.

Este trabalho tem como objetivo aplicar técnicas de **Web Scraping** [3] para extrair informações de manchetes de portais de notícias, processá-las por meio de ferramentas de Processamento de Linguagem Natural (PLN) e, posteriormente, treinar um classificador capaz de identificar se tais manchetes estão associadas a temas relacionados à violência. O **Web Scraping** caracteriza-se como o processo de extração de informações relevantes presentes em páginas da web e sua conversão de formato não estruturado para estruturado, de modo a possibilitar análises subsequentes [3].

Nossa aplicação busca obter informações pertinentes das manchetes de portais de notícias brasileiros e classificá-las quanto ao seu conteúdo temático. Para isso, o algoritmo foi avaliado a partir de um corpus contendo manchetes previamente rotuladas.

O algoritmo implementado baseia-se no workflow apresentado na Figura 1. Inicialmente, aplicou-se a técnica de **web scraping** para coleta de dados em portais de notícias. Em seguida, esses dados foram tratados, estruturados e armazenados em um **dataframe**.

Posteriormente, utilizou-se um vetor contendo mais de 400 palavras associadas a violência e crimes, incluindo variações de gênero e tempo verbal. A partir da combinação entre a base de dados e o vetor de palavras-chave, realizou-se uma pré-classificação para facilitar a etapa de classificação manual.

Com a base categorizada segundo a temática das manchetes, procedeu-se ao treinamento do modelo por meio de técnicas de aprendizado de máquina, resultando na construção de um classificador. Por fim, o modelo treinado foi aplicado a novos dados extraídos de portais de notícias, sem submetê-los à etapa de pré-classificação.

Para o treinamento e teste do classificador, foi necessária a criação de um conjunto de dados validado manualmente, contendo manchetes rotuladas como violentas ou não violentas.

Na etapa de web scraping, criou-se uma lista para armazenar todos os dados extraídos, definindo-se previamente o número de páginas a serem acessadas pelo algoritmo. O resultado dessa etapa foi um dataframe contendo as colunas: **título, data, categoria, resumo, link, URL da imagem e página**.

A etapa seguinte consistiu na rotulação dos dados de acordo com o valor esperado, identificando se os títulos eram violentos ou não violentos. Para isso, empregou-se a técnica de

detecção de palavras-chave, adicionando ao dataframe indicadores referentes à presença e à frequência desses termos nos títulos.

Após a classificação inicial da base, realizou-se o pré-processamento textual, com o objetivo de remover elementos que não contribuem de forma significativa para as análises. Em seguida, gerou-se uma matriz TF-IDF composta por um vocabulário de 1.000 termos.

Considerando que o conjunto apresentava uma quantidade significativamente maior de notícias não violentas do que violentas, e visando evitar viés no aprendizado de máquina, aplicou-se a técnica SMOTE para balanceamento dos dados.

Para o treinamento do classificador, utilizou-se o modelo **MultinomialNB**, baseado no algoritmo **Multinomial Naive Bayes**. A partir dos resultados da predição no conjunto de teste, procedeu-se à validação da acurácia e ao cálculo das métricas por classe. Por fim, elaborou-se a matriz de confusão, permitindo identificar os valores de Verdadeiros Negativos, Falsos Positivos, Falsos Negativos e Verdadeiros Positivos.

As métricas de desempenho obtidas foram: **Acurácia:** 86% **Precisão:** 91% **Recall:** 86% **F1-score:** 88%

Observa-se que, apesar do desbalanceamento das classes, o classificador apresentou desempenho satisfatório quanto às predições.

Ao aplicar o classificador em um cenário real, com novas notícias sem rótulos de comparação, observaram-se bons resultados na identificação de notícias violentas. Contudo, devido à ambiguidade linguística e à subjetividade envolvida na definição do que constitui violência, é importante analisar cuidadosamente o limiar de confiança adotado. Foram observados casos de classificação correta com índice de confiança superior a 80%, bem como situações com valores em torno de $54,9\%$, nas quais ocorreu um falso positivo para notícia violenta.

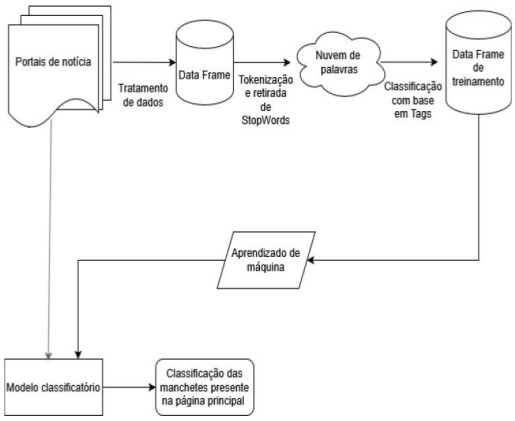

Figura 1: Workflow de construção e validação do modelo classificatório

Palavras-chave: *Mineração de texto, Violência, Classificação*

Referências

[1] TAVARES, C. N. "Uma análise discursiva de criminalidades nas manchetes do meia hora e do crónica: produção de sentidos em enunciados e a sua aplicação no ensino´´, Universidade do Estado do Rio de Janeiro, 13, 2018.

[2] Tan, Pang-Ning and Steinbach, Michael and Karpatne, Anuj and Kumar, Vipin, "Introduction to Data Mining (2nd Edition)´´, 2018, Pearson.

[3] Zhao, B. (2017). Web Scraping. In: Schintler, L., McNeely, C. (eds) Encyclopedia of Big Data. Springer, Cham.