

Universidade de Pernambuco
Programa de Pós-Graduação em Engenharia da
Computação (PPGEC)

Proposta de Dissertação de Mestrado

Área: Computação Inteligente

Título: Mineração Textual para Classificação de Mensagens Xenófobas Utilizando Support Vector Machine e a estratégia TF-IDF

Orientador – Cleyton Mário de Oliveira Rodrigues (cleyton.rodrigues@upe.br)

Descrição – Xenofobia é o medo em relação à grupos estrangeiros. Contudo, entende-se que a Xenofobia representa algo muito mais amplo, ao destacar não só o medo, mas também (e, principalmente) o ódio, a rejeição, ou a hostilidade contra diferentes grupos étnicos, motivada geralmente pelas diferenças culturais, políticas, históricas, sociais, religiosas ou mesmo econômicas. Em geral, tenta-se justificar os ataques e as ofensas de ódio sobre a desculpa de proteger o que é seu, de forma que o país (ou aquela região, em específico) não perca sua identidade.

No contexto de algoritmos inteligentes, a Mineração de Textos é uma área interdisciplinar (envolve a Inteligência Artificial, a Aprendizagem de Máquina, a Estatística, entre outras áreas) para que se possa garimpar informações textuais em busca de informações úteis para tomadas de decisões. Geralmente, envolve buscas quantitativas e qualitativas em grandes volumes de textos (muitas vezes, não-estruturados) na tentativa de identificar padrões ou tendências no texto escrito. Para a análise textual, há diversas estratégias, como a análise estatística, que se baseia na frequência/importância dos termos presentes no texto, ou seja, os documentos são vistos através de um *bag of words*. Dessa forma, a mineração textual proposta neste trabalho prever analisar postagens em mídias sociais, para que possa classifica-la como uma mensagem xenófoba ou não. Almeja-se também otimizar acurácia das métricas de classificação, como o uso do recurso GridSearch() que permite fazer testes exaustivos com várias combinações de parâmetros, à medida que se acompanha a acurácia dos resultados alcançados.

Outrossim, além do texto em si, é preciso levar em consideração outras singularidades das postagens, com o uso de emoticons, re-postagens, abreviações entre outras características. Destaca-se como contribuições da pesquisa, além do modelo classificador, uma base de análise de postagens em português e rotulada como mensagem xenofóbica ou não que pode ser utilizada em outras pesquisas.

Referências Bibliográficas

1. EBECKEN, N; LOPES, M; COSTA, M. Mineração de Textos, chapter 13, p. 337–370. Manole, 2003.
2. LAGARZA, Cecília de. Xenofobia. Laboreal, Porto, v. 7, n. 2, p. 1-5, dez. 2011.
3. LORENA, A. C.; CARVALHO, A. C. de. Uma introdução às support vector machines. Revista de Informática Teórica e Aplicada, v. 14, n. 2, p. 43–67, 2007.
4. ONU BR. Estudo da ONU aponta aumento da população de migrantes internacionais. 2019. Disponível em: <https://nacoesunidas.org/estudo-da-onu-aponta-aumento-da-populacao-de-migrantes-internacionais/amp/>. Acesso em: 01 nov 2021.