

Ciência da Computação / Sistemas de Informação

Avaliação de Modelos de Linguagem para Mineração de Opinião na Língua Portuguesa

Thiago Salles Santos - 8º módulo de Ciência da Computação, UFLA, iniciação científica

Denilson Alves Pereira - Orientador DCC, UFLA. - Orientador(a)

Resumo

Vivemos em um cenário no qual saber compreender, analisar e prever as opiniões, atitudes e emoções expressas nas mensagens textuais das pessoas é uma atividade essencial para o mercado, para aplicações como: monitoramento de marca, pesquisa de mercado, tomada de decisões políticas, detecção de tendências, dentre outros. O objetivo deste trabalho é apresentar uma abordagem para a mineração de opinião em textos, incluindo o domínio específico da bebida cachaça, a partir de bases de dados extraídas de redes sociais e modelos de linguagem pré-treinados. Para a realização do projeto, foi primeiramente realizado uma pesquisa para o levantamento de modelos de linguagem para processamento de linguagem natural pertencentes ao estado da arte, sendo selecionados os modelos twitter-XLM-roBERTa e BERTimbau. Posteriormente, foi realizada uma busca por bases de dados, com o intuito de compor as baselines da pesquisa. As bases de dados selecionadas foram Brazilian Portuguese Sentiment Analysis Datasets e Building a Sentiment Corpus of Tweets in Brazilian Portuguese. Em seguida, foi realizada a rotulação por amostragem dos datasets extraídos de redes sociais, sobre o assunto cachaça. Com todos os dados em mãos, foi realizado um pré-processamento e a avaliação dos modelos. Para o modelo BERTimbau, foi realizado o ajuste fino para a tarefa de análise de sentimentos, já para o modelo twitter-XLM-roBERTa, foi realizada a sua avaliação a partir da ferramenta pipeline, fornecida pela plataforma Hugging Face. Com isso, foram obtidos um total de 10 modelos resultantes do ajuste fino do BERTimbau, 15 resultados de predição, utilizando os datasets das baselines e 12 resultados de predição dos datasets extraídos das redes sociais. Após a análise dos resultados, pode-se concluir que, a mineração de opinião é uma ferramenta poderosa que pode ser usada para entender as opiniões expressas em textos não estruturados, em uma variedade de aplicações, inclusive no domínio da bebida cachaça. Como trabalhos futuros, pretende-se avaliar outros modelos de linguagem e rotular mais dados do dataset sobre cachaça.

Palavras-Chave: Análise de sentimentos, Bebida cachaça, Modelos.

Instituição de Fomento: FAPEMIG; UFLA

Link do pitch: <https://youtu.be/114q2UQWGfc>