

Análise de sentimento: desenvolvimento de recursos para o pré-processamento de *tweets*

Leonardo Pereira dos Santos¹, Natally Lauthart Lazzery¹, Larissa Astrogildo de Freitas^{1*}
*Orientador

¹Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul (IFRS) - *Campus* Osório. Osório RS, Brasil.

A Web tem se apresentado como uma importante plataforma de comunicação. Através de redes sociais, blogs, fóruns e sites para avaliação de produtos e serviços, os usuários têm disponibilizado e encontrado cada vez mais opiniões sobre diversos assuntos. Com isso, Análise de Sentimento ou Mineração de Opinião, busca identificar o sentimento, percepção ou ação do público em relação ao alvo desta opinião. A pesquisa Análise de Sentimentos em Textos Escritos na Língua Portuguesa tem como objetivo o desenvolvimento de recursos referentes à Análise de Sentimento em textos publicados em redes sociais. O tema abordado nesta pesquisa é o processo de impeachment de Dilma Rousseff, com foco em comentários opinativos feitos no Twitter, por usuários na região de Osório/RS. Para que a Análise de Sentimento possa ser realizada, é necessário o pré-processamento destes textos, no caso, *tweets*. Para tal, são realizadas três etapas: captura de *tweets*, que consiste na coleta dos *tweets*; limpeza de *tweets*, que consiste na correção dos textos informais e na adaptação destes para uma versão mais próxima da norma culta da Língua Portuguesa; e anotação de *tweets*, que consiste na identificação do sujeito do discurso e na classificação da opinião (positiva ou negativa). Para a realização de cada uma das etapas, foram construídas ferramentas específicas. A Ferramenta de Captura de *Tweets*, responsável pela coleta dos *tweets*. Essa ferramenta é capaz de filtrar *tweets* a partir de especificações de datas (17 de abril e 12 de maio de 2016, referentes às votações na câmara de deputados e senado), idioma (português), localização (Osório/RS) e termos de busca. A Ferramenta de Limpeza de *Tweets*, responsável pela limpeza dos *tweets*. Essa ferramenta é capaz de fazer substituições de termos abreviados e do internetês por correspondências ortográficas, a partir do uso de dicionários. A Ferramenta de Anotação Manual Automatizada (FAMA), responsável pela anotação dos *tweets*. Essa ferramenta é capaz de identificar o sujeito do discurso a partir de um banco de dados contendo nomes de deputados federais, senadores, ministros, partidos políticos, vice-presidente e presidente e classificar a opinião sobre o sujeito do discurso contida nos *tweets* (positiva ou negativa). Ainda, deverão ser desenvolvidos outros recursos para automatizar as demais fases da Análise de Sentimento. Todos estes recursos serão disponibilizados em um *website*, para que possam ser aprimorados e reutilizados. Também, serão divulgados os resultados obtidos com o auxílio desses recursos a respeito do processo de impeachment.

Palavras-chave: Análise de sentimento. Twitter. Pré-processamento.

Trabalho executado com recursos do Edital nº 12/2015/Programa Institucional de Bolsas de Iniciação Científica e/ou Tecnológica (PROBICT) do IFRS, da Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação.