

# Identificação da Auto Eficácia dos Usuários do Twitter em Relação ao Tema “Geografia” com o Uso da Análise de Sentimentos

Sidney Felipe Moraes , Professor Seiji Isotani , Bruno E. Penteado  
 Pós-Graduação EAD em Computação Aplicada à Educação (ICMC/USP)  
 Universidade de São Paulo (USP)

## INTRODUÇÃO

- **TEMA:** Auto Eficácia através da Mineração de dados e Análise de Sentimentos.
- **PROBLEMA:** Possível identificação e inferência do senso de auto eficácia através da mineração de dados dos posts de usuários da rede social Twitter com o auxílio da análise de sentimentos .
- **JUSTIFICATIVA:** Maior conhecimento e exploração dos dados gerados pelas redes sociais, assertividade nas ações educacionais, maiores evidências para a melhora pedagógica em sala de aula.

## MATERIAS E MÉTODOS

### MATERIAIS:

- Uso da API Pública da rede social *Twitter* para coleta de 9.052 mensagens, no período de 15 a 22 de junho de 2020. Do total, 2.572 tiveram seus sentimentos classificados manualmente, restando 6.480 que serão classificadas pelos algoritmos propostos.
  - Uso do software de mineração *Rapidminer*.
  - Uso da ferramenta de visualização de dados *Microsoft Power BI*.
- MÉTODOS:** Método *KDD*, processo de descoberta de conhecimento através de dados. Utilizado todas as etapas do método.

## OBJETIVOS

- Utilizar da mineração de dados da rede social Twitter com a busca pelo termo “geografia”, categorização manual do sentimento para o aprendizado de máquina.
- Aplicação de técnicas de mineração de dados ( Balanceamento, *Stemming*, *Stop Words*, etc.) para pré-processamento e modelagem dos dados.
- Aplicação da técnica de Análise de sentimentos através do algoritmo de classificação Polinomial e com uso de recursos avançados como a técnica de topic modeling LDA ( Latent Dirichlet Allocation) para clusterização de termos principais em relação aos tweets. Com isso, inferir quais os possíveis tópicos em relação aos sentimentos processados.
- Avaliação das métricas de Acurácia, Average, Error, e análise gráfica das previsões e classificações para estudo dos resultados.

## DISCUSSÃO / LICÕES APRENDIDAS

- Apenas técnicas de mineração, análise de sentimentos e LDA não garantem uma eficácia na inferência da identificação da auto eficácia.
- Maior volume de dados para melhor treino do algoritmo através de uma API full.
- Refinamento na etapa de pré-processamento para tratar termos de gírias e sacarmos.
- Estudo mais aprofundado de métodos e técnicas para identificação da auto eficácia.

## RESULTADOS

- Algoritmo *NAIVE BAYES* resultou em maior índice de acurácia nas previsões dos sentimentos ( P/N/NE) ( 84,86%)
- Em 50,73% dos Tweets foram positivos para o termo geografia
- Previsões de sentimento NE ( Neutras), tiveram o menor índice de acurácia (74,67%)
- Diversos termos iguais tiveram a predição diferente, alertando sobre a qualidade dos posts, gírias, sacarmos e correntes.

accuracy: 84.86% +/- 1.77% (micro average: 84.86%)

	true NE	true N	true P	class precision
pred. NE	669	138	89	74.67%
pred. N	91	1063	58	87.71%
pred. P	61	73	1127	89.37%
class recall	81.49%	83.44%	88.46%	

Figura 1-Avaliação do modelo predito.

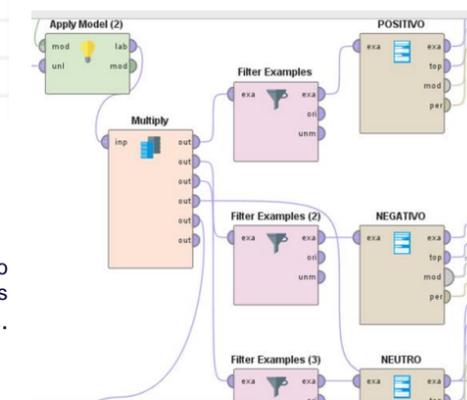


Figura 2-Mineração de dados com o uso do software *Rapidminer* e aplicação dos tópicos LDA nas categorias que foram preditas.



Figura 3 – Visualização dos sentimentos e termos LDA para possível identificação da auto-eficácia