

**Tema: Processamento de Linguagem Natural com Análise de Sentimentos e Modelagem de Tópicos para dados de Redes Sociais**

## Contexto e Motivação

As redes sociais já fazem parte do cotidiano de boa parte da população. O *Twitter* já conta com aproximadamente 126 milhões de usuários ativos diariamente e 6000 tweets por segundo sendo enviados[1]. Neste cenário, a classificação destas mensagens é de grande relevância, já que ao identificar o sentimento predominante, é possível entender a aceitação do público por parte de certo tópico e seu impacto

## Proposta

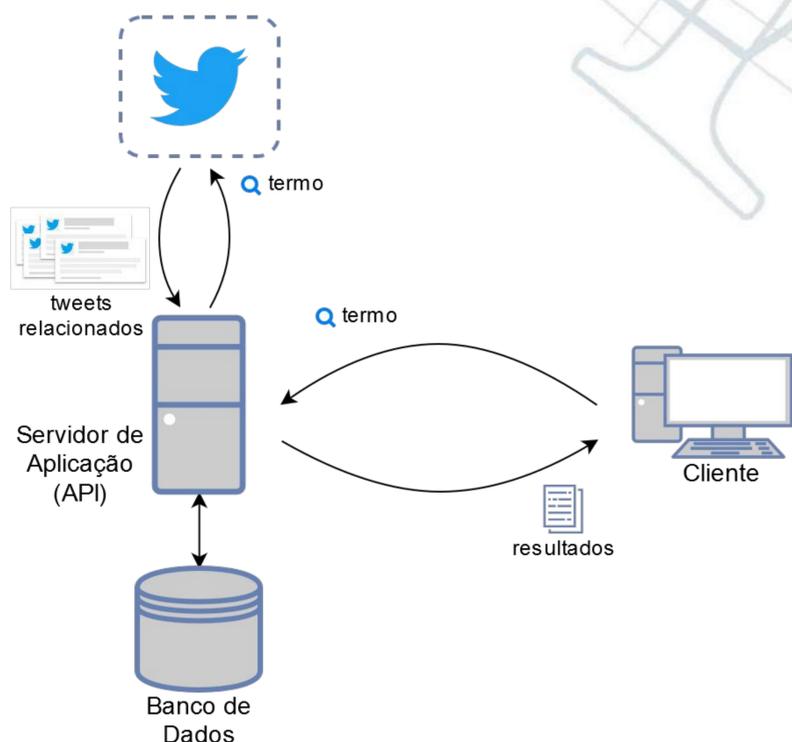
O projeto consiste em um sistema capaz de inferir a polaridade e a emoção atrelada, de forma automatizada, utilizando dados provenientes do *Twitter*. Para tal, é obtida a quantidade especificada de *tweets*, relacionados a um termo qualquer para serem processados.

De modo a oferecer uma forma de diferenciar diversos assuntos de mesmo termo de busca, foi desenvolvido um mecanismo para a criação de tópicos e separação dos textos contidos nestes.

## Projeto

O sistema desenvolvido foi dividido em um módulo para a coleta de *tweets*, modelos de *deep learning* para análise de sentimentos e emoções e um modelo estatístico para a modelagem de tópicos

A integração das partes se deu na aplicação desenvolvida para a apresentação visual. Assim o processamento todo foi feito no servidor de aplicação, e os resultados poderiam ser visualizados no lado do cliente.



## Resultados

Para os modelos de *Deep Learning*, foram feitos testes com datasets anotados por pessoas[2][3]. Os resultados seguintes são para os melhores modelos treinados de cada arquitetura.

Modelo	Acurácia	Precisão	Recall	F1-Score
<i>CNN 1D</i>	0,8524	0,8527	0,8521	0,8523
<i>LSTM Bidirecional</i>	0,8301	0,8314	0,8296	0,8297
<i>LSTM Bidirecional + CNN</i>	0,8245	0,8269	0,8239	0,8240

Resultados para a Análise de Sentimentos

Model	Acurácia	Precisão	Recall	F1-Score
<i>CNN 1D</i>	0,7044	0,6394	0,6840	0,6536
<i>LSTM Bidirecional</i>	0,7393	0,6788	0,7322	0,6962
<i>LSTM Bidirecional + CNN</i>	0,7084	0,6500	0,7062	0,6675

Resultados para o Reconhecimento de Emoções

Em relação aos modelos estatísticos, utilizados para modelagem em tópicos. A figura abaixo apresenta um exemplo de tópico obtido com o LDA, para o termo "nobel".

Tópicos resultantes do LDA	
Tópico 1	get (0.030), give (0.024), go (0.017), trump (0.015), obama (0.015), year (0.014), think (0.013), people (0.012), would (0.012), president (0.011)
Tópico 2	greta_thunberg (0.020), deserve (0.016), leader (0.015), world (0.013), say (0.013), laureate (0.012), get (0.012), great (0.011), u (0.009), well (0.009)
Tópico 3	work (0.030), winner (0.025), one (0.024), physic (0.017), year (0.014), two (0.012), reject (0.011), rejection_letter (0.011), economics (0.010), no (0.010)
Tópico 4	literature (0.044), chemistry (0.022), peter_handke (0.018), get (0.016), battery (0.015), genocide (0.015), via (0.013), winner (0.012), handke (0.012), time (0.009)
Tópico 5	ahmed (0.102), ethiopia (0.073), ethiopian (0.065), prime_minister (0.065), pm (0.041), congratulation (0.029), eritrea (0.027), ali (0.018), winner (0.016), via (0.015)

## Aplicação

Aplicação desenvolvida com React e Flask para apresentação das informações obtidas a partir da classificação dos textos e modelagem em tópicos.

A figura abaixo representa a visualização dos resultados de um tópico para o termo "nobel". Esta conta com:

- Nuvem de palavras: seleção de palavras que compõem o tópico;
- Barra de polaridade: apresenta resultados da análise de sentimentos
- Grau de emoção: apresenta resultados do reconhecimento de emoções



## Referências

- [1] SHABAN, H. Twitter reveals its daily active user numbers for the first time. *The Washington Post*, 02 2019.
- [2] GO, A.; BHAYANI, R.; HUANG, L. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 2009.
- [3] MOHAMMAD, S. et al. Semeval-2018 task 1: Affect in tweets. In: *Proceedings of the 12th international workshop on semantic evaluation*. [S.l.: s.n.], 2018. p. 1-17.

**Integrantes:** Matheus Felix Lima Dias da Silva  
Rodrigo Masaru Ohashi

**Orientador:** Prof. Dr. Ricardo Luis de Azevedo da Rocha