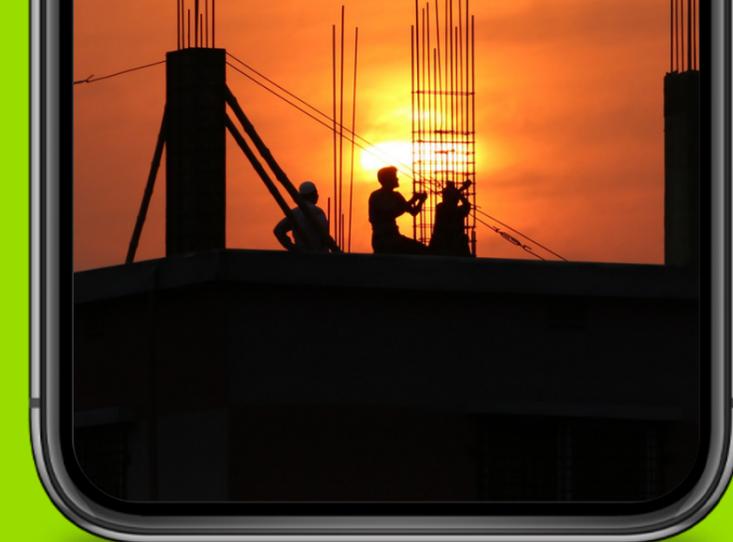
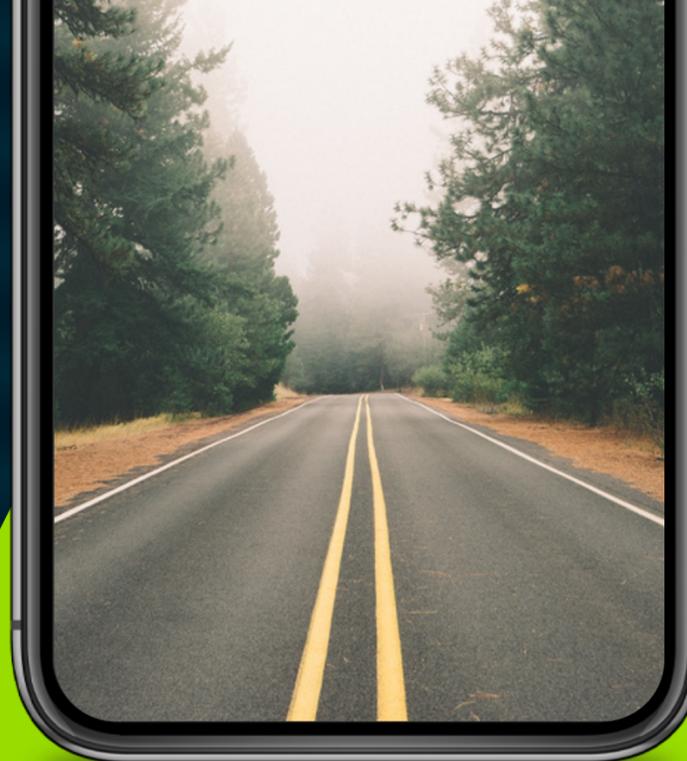


Análise de Sentimento dos usuários do Ministério de Infraestrutura no Twitter



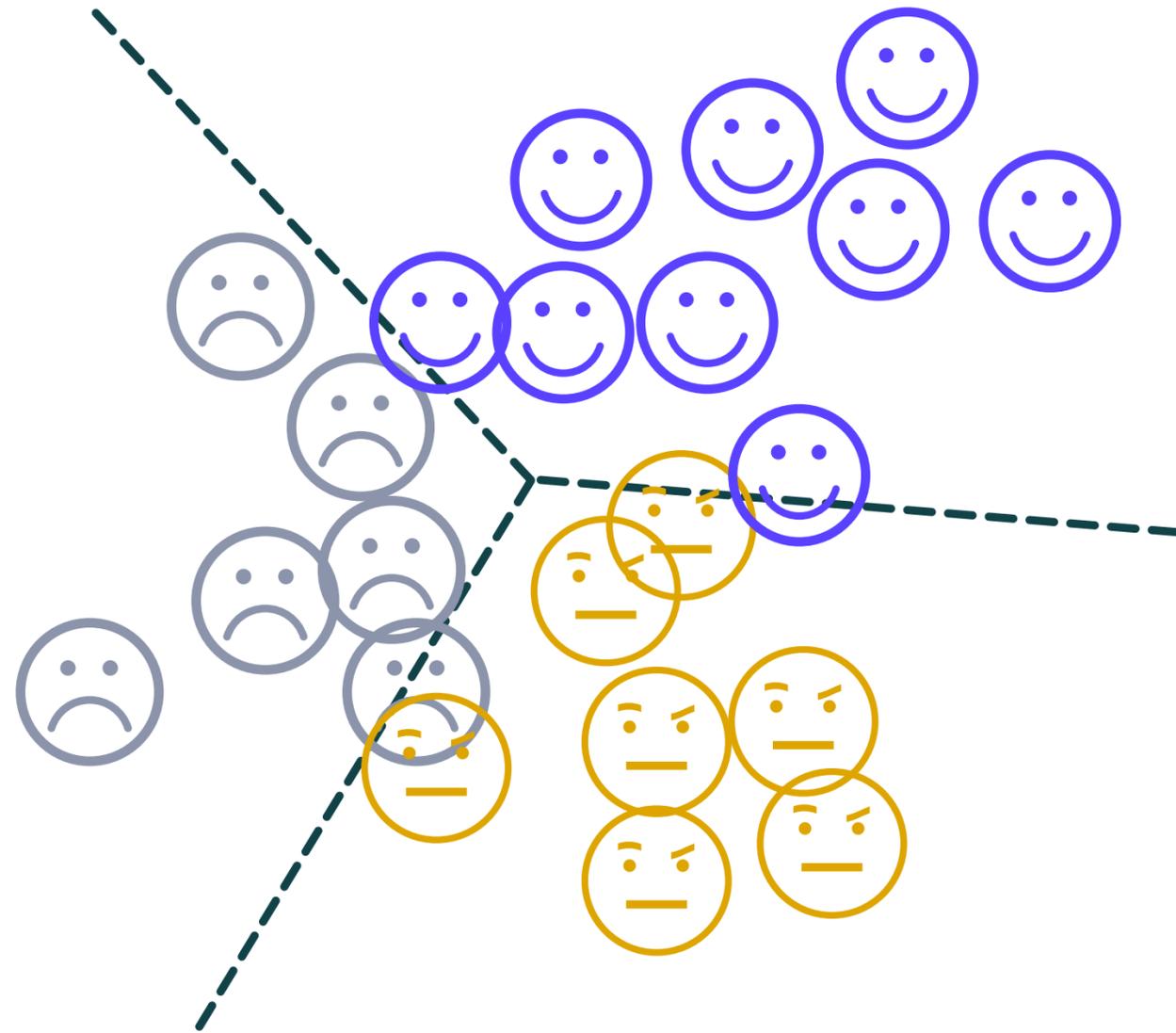
Problema

Classificar os comentários dos cidadãos, constantes no perfil institucional do Ministro de Infraestrutura no Twitter e relacionados ao tema "rodovia", a fim de facilitar a tomada de decisão e promover ações de melhoria na qualidade dos serviços oferecidos aos usuários.



 Positivo  Negativo  Questionamento

Solução



Treinamento supervisionado de modelo de Machine Learning de classificação e Análise de Sentimento com base no texto dos comentários dos twitters coletados do perfil institucional do Ministro de Infraestrutura no Twitter.

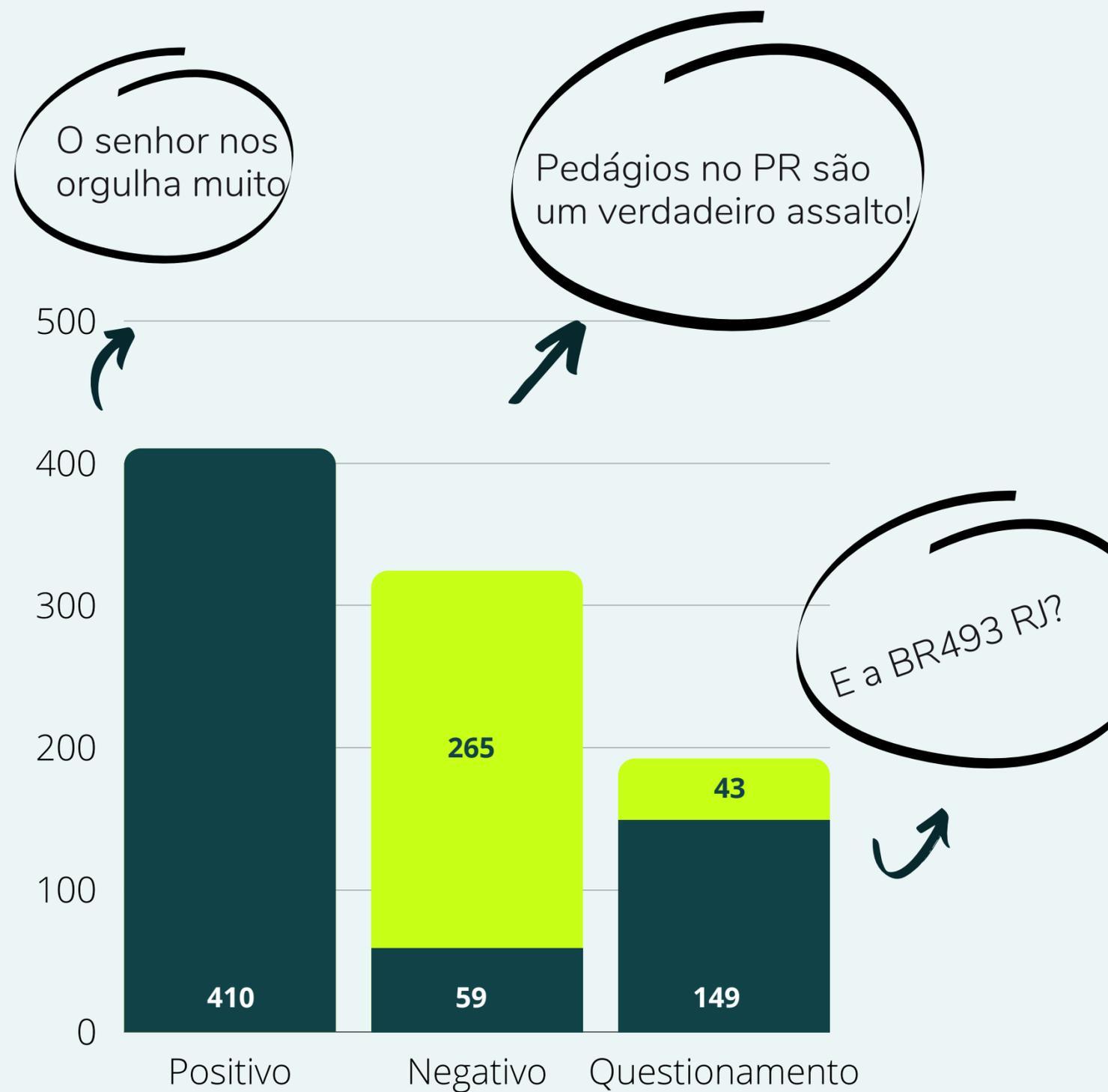
Dados



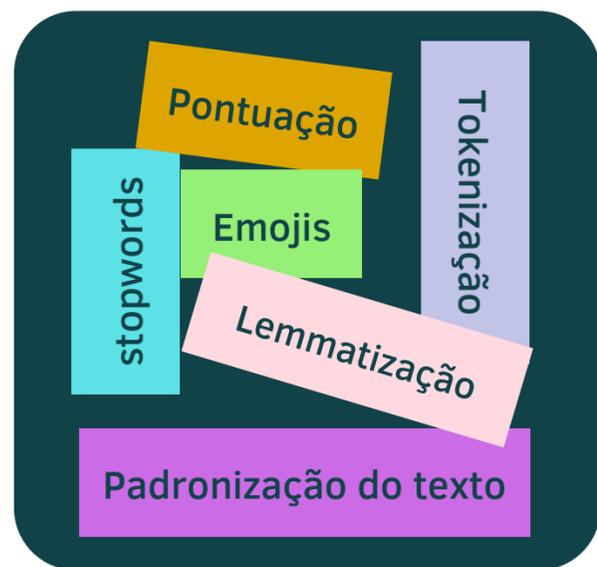
Dados originais

Enriquecimento da classe negativa com dados da base do Kaggle

Enriquecimento da classe questionamento com dados anotados manualmente



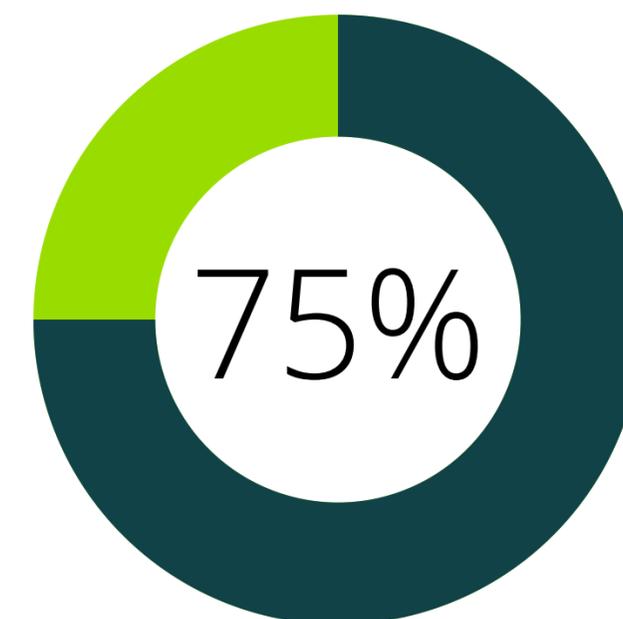
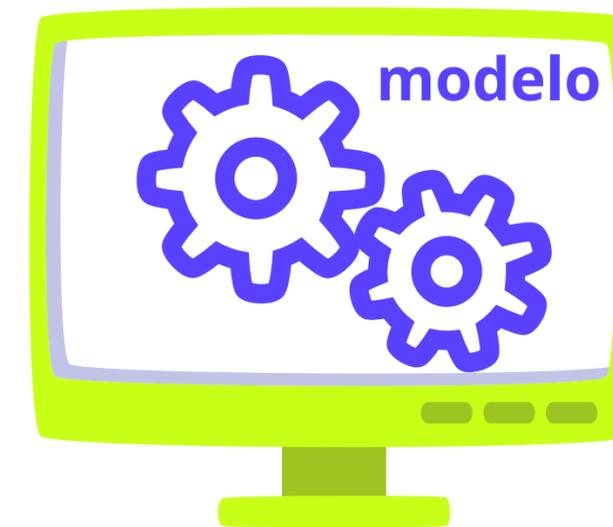
Atributos



TF-IDF

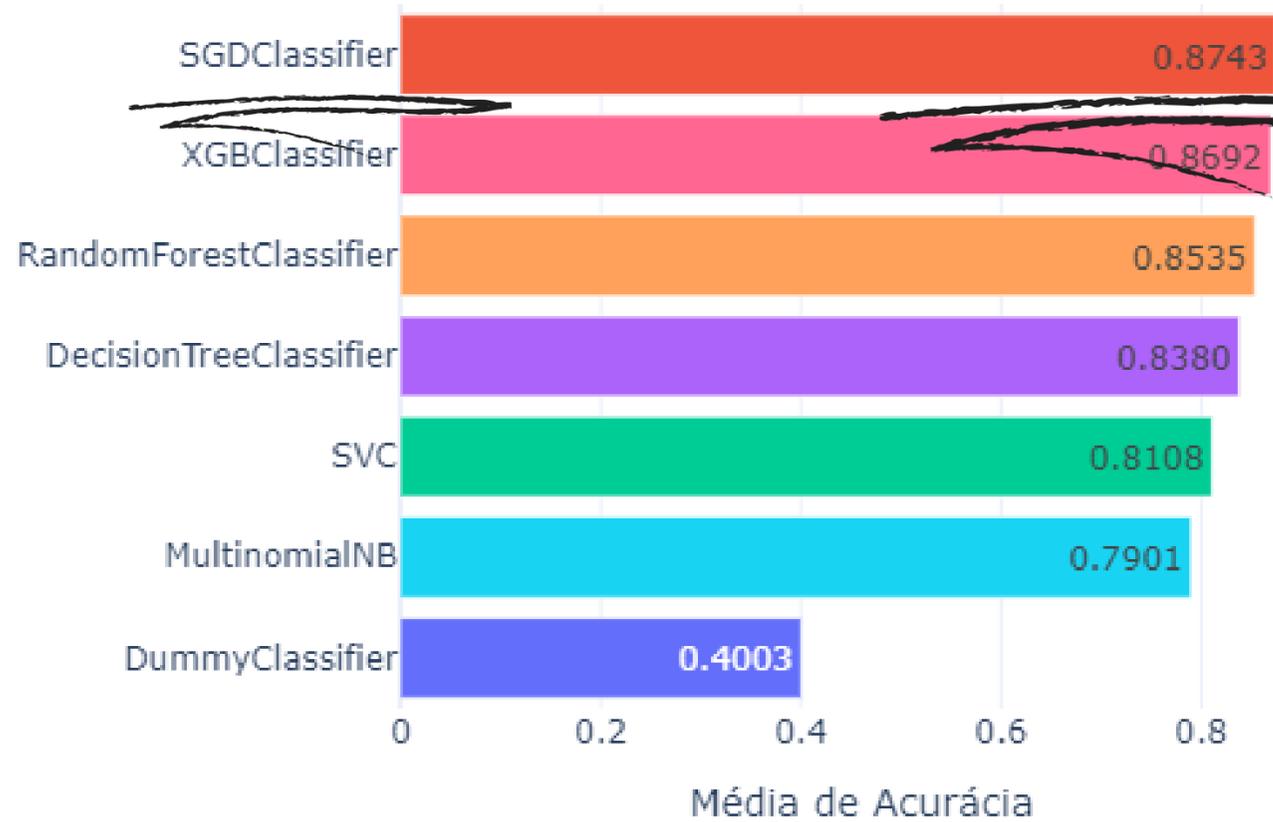


| vocabulário | | | | | |
|-------------|----------|----------|--------|-----|---------|
| 0 | ministro | parabens | :(| ... | pedagio |
| 1 | 0.2569 | 0.3551 | 0.0073 | ... | 0.0487 |
| ... | ... | ... | ... | ... | ... |
| n | 0.0941 | 0.0043 | 0.6310 | ... | 0.4427 |



Modelos e Resultado

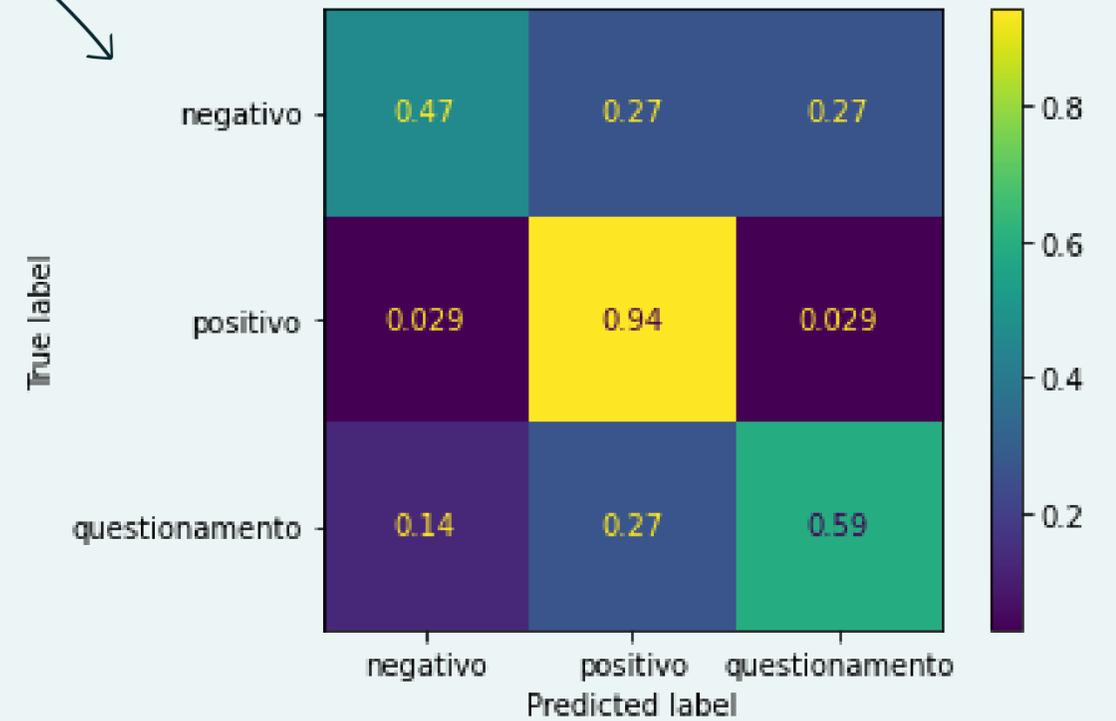
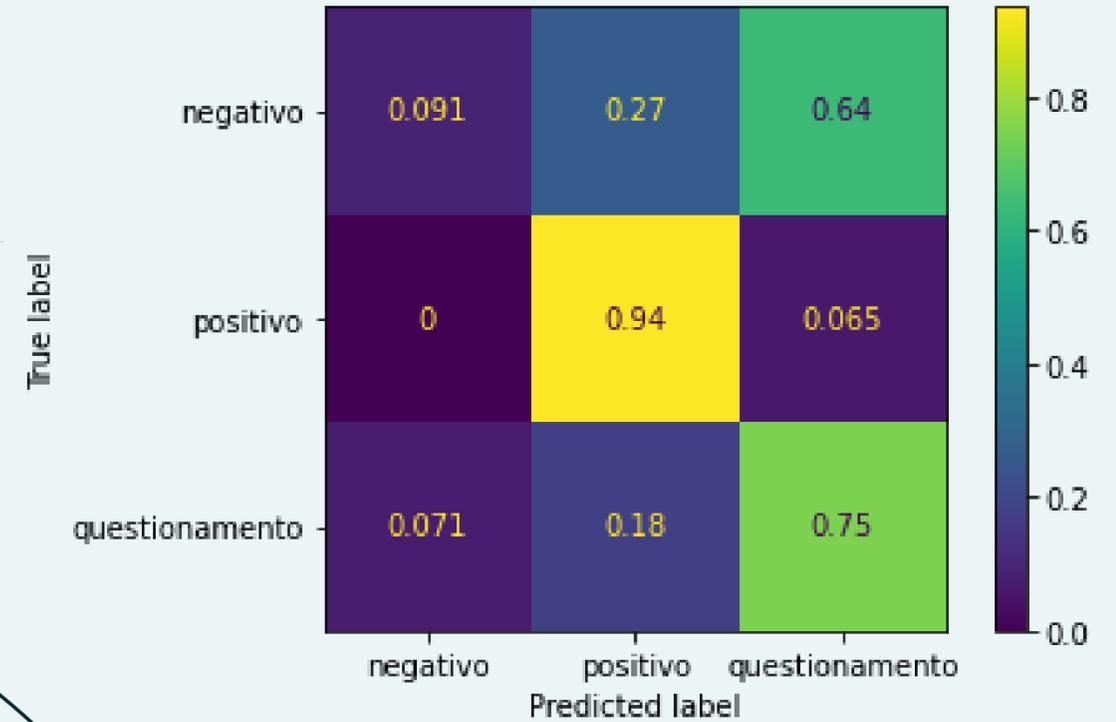
Ranking da acurácia dos modelos testados



Balaceamento

antes

depois



Interpretação

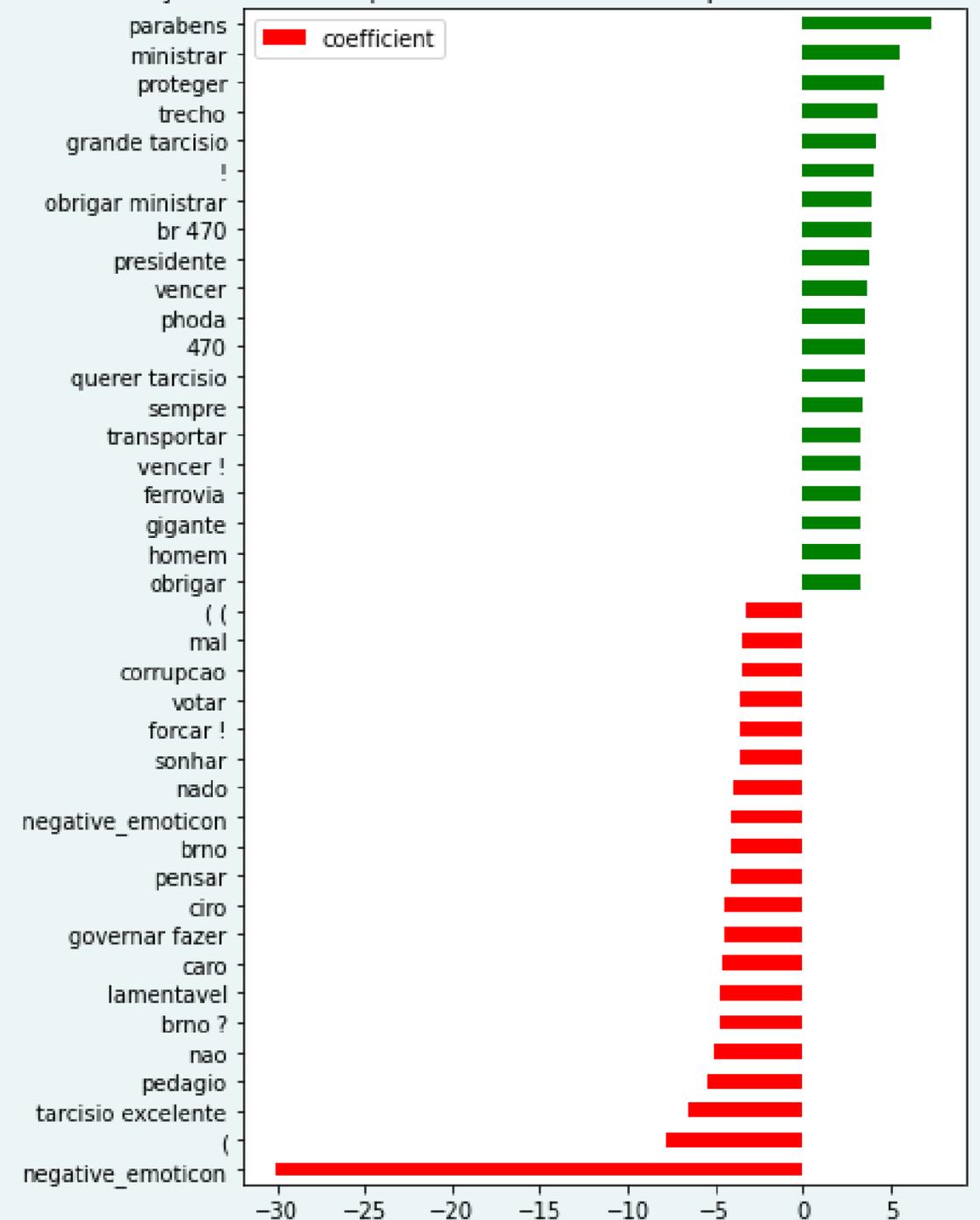
"parabens ministro thor ! senhor achado brasil !"

| y=negativo (score -6.731) top features | | y=positivo (score 6.250) top features | | y=questionamento (score -2.714) top features | |
|--|-----------|---------------------------------------|-----------|--|-----------|
| Contribution? | Feature | Contribution? | Feature | Contribution? | Feature |
| -0.353 | parabens | +2.247 | ! | +0.014 | ministrar |
| -0.413 | brasil | +2.066 | parabens | -0.374 | brasil |
| -0.708 | <BIAS> | +1.341 | senhor | -0.425 | parabens |
| -1.036 | ! | +0.919 | brasil | -0.484 | senhor |
| -1.325 | ministrar | +0.883 | ministrar | -0.563 | <BIAS> |
| -1.420 | senhor | -0.513 | achar | -0.882 | ! |
| -1.476 | parabens | -0.694 | <BIAS> | | |

"mp ferrovias ? vai deixar morrer 100bi investimento ?"

| y=negativo (score -2.495) top features | | y=positivo (score -3.751) top features | | y=questionamento (score 4.196) top features | |
|--|----------|--|----------|---|------------|
| Contribution? | Feature | Contribution? | Feature | Contribution? | Feature |
| -0.006 | ? ir | +0.734 | deixar | +2.421 | ? |
| -0.440 | deixar | -0.572 | ferrovia | +1.591 | ferrovia |
| -0.520 | ferrovia | -0.694 | <BIAS> | +0.876 | mp |
| -0.708 | <BIAS> | -3.219 | ? | +0.157 | ferrovia ? |
| -0.821 | ? | | | -0.031 | ir |
| | | | | -0.256 | deixar |
| | | | | -0.563 | <BIAS> |

TOP 20 palavras mais relevantes para as classes Positivo e Negativo"



Destques

Cidadão como termômetro!!



Lições

1 Balanceamento das classes

3 Separação da base de teste logo no início

5 Entendimento do processo de coleta e rotulação dos dados pela área de negócio

6 Biblioteca otimizada para português-br

2 Importância da pontuação e emojis

4 Vazamento de dados

7 Uso de modelos "estado da arte"

Obrigado!



Priscilla A. S. Rodrigues

Câmara dos Deputados



Marcela Gomes Corrêa

Ministério da Infraestrutura



Liliane Vieira Lopes

Instituto Brasileiro de Museus



José Renato Borelli

Instituto Federal de São Paulo