Compilado de decisões e resultados para a criação dos modelos de ML

1. Definição dos marcadores (labels)

**“Esta é uma das grandes dificuldades do aprendizado de máquina especialmente no que diz respeito à análise de sentimentos por dois motivos principais: a alta subjetividade envolvida na tarefa e a demanda de tempo necessária para que especialistas definam a polaridade de muitas sentenças.”**

Nessa etapa foram escolhidos 2 dicionários para avaliar qual se ajustaria melhor aos dados existentes.

* **Vader:** Foi criado com base em outros 3 dicionários diferentes (LIWC, ANEW e o GI)
* **Opinion-lexic:** Diferente do anterior por estar mais atualizado a lidar com expressões em redes sociais e reviews de produtos.

O opinion-lexic foi escolhido tanto por ter apresentado uma sensibilidade maior em sua classificação, quanto pelo fato de sua origem ter vinda de reviews de produtos, colaborando para o objetivo desse projeto que envolve classificar reviews de filmes.

A group of blue and orange bars

Description automatically generated

Nas imagens acima os valores **Standard** representam classificações feitas apenas com a remoção básica das palavras mais comuns na língua inglesa (stop words), já o **Custom** contém algumas palavras a mais que aparecem com muita frequência nos textos e não representam uma relevância significativa na classificação das polaridades.

A close-up of words

Description automatically generatedA close up of words

Description automatically generated