Verificação de mensagens ofensivas por meio de deep learning e bag-of-words

Nome dos Alunos:

- **Gabriel Alves Barbosa**
- **Gabriel Luis Silva Pereira**
- **Rodrigo Dias Moreira**
- **Yan Victor S. Azevedo**
- **Bruno Terra Faria Trindade**

É notório que atualmente as redes sociais se tornaram ambientes nocivos, onde muitas vezes mensagens de ódio são propagadas e usuários mal intencionados agridem outros usuários. Neste sentido, propõe-se uma IA que fosse capaz de reconhecer estas mensagens a fim de bloqueá-las o mais rápido possível.

Para começar, teríamos que implementar um período de aprendizado para tal IA. Primeiramente, procuraremos por datasets já rotulados disponibilizados na internet contemplando discursos de ódio em diferentes contextos e plataformas. Com base neles, os datasets seriam organizados e padronizados em um modelo Bag-of-Words. Aqui, seria feito um estudo sobre as possíveis técnicas a serem utilizadas no modelo, buscando filtrar palavras ou expressões que possam ser utilizadas como atributos relevantes para um melhor resultado de predição.

Após a obtenção dos dados, será iniciado o treino do método escolhido. Neste ponto, métricas de avaliação como Precisão, Recall (dentre outras) deverão ser implementadas e avaliadas. Assim, será feito o treino e avaliação de um modelo por meio de um Rede Neural (Perceptrons multi-camada) e, a partir dos resultados obtidos nos testes do modelo construído, será contraposto contra outros modelos construídos com base em diferentes técnicas de otimização do modelo Bag-of-Words, em busca de uma melhor precisão no resultado.