# Projeto de Pesquisa - Análise de Sentimentos em Textos de Redes Sociais Usando Técnicas de Aprendizado de Máquina

## 1. Introdução

A análise de sentimentos em textos de redes sociais é uma área relevante dentro do Processamento de Linguagem Natural (PLN) e da Inteligência Artificial. Com o crescimento exponencial das interações em plataformas como Twitter, Facebook e Instagram, surge a necessidade de compreender automaticamente as opiniões, emoções e percepções dos usuários. Essa análise permite a empresas, organizações e pesquisadores obter insights valiosos para tomada de decisão, monitoramento de reputação, análise de comportamento do consumidor e estudos sociais. Este projeto visa investigar e desenvolver modelos de aprendizado de máquina aplicados à análise de sentimentos em textos extraídos de redes sociais, com foco na língua portuguesa.

## 2. Problema

Como classificar automaticamente o sentimento (positivo, negativo ou neutro) presente em textos publicados em redes sociais, utilizando técnicas de aprendizado de máquina aplicadas à língua portuguesa?

## 3. Objetivos

### 3.1 Objetivo Geral

Desenvolver um modelo de análise de sentimentos aplicado a textos de redes sociais em português, utilizando técnicas de aprendizado de máquina.

### 3.2 Objetivos Específicos

- Levantar e selecionar bases de dados adequadas.  
- Realizar pré-processamento dos textos (limpeza, tokenização, remoção de ruídos e vetorização).  
- Implementar e comparar modelos de aprendizado de máquina e de deep learning.  
- Avaliar o desempenho dos modelos utilizando métricas como acurácia, precisão, recall e F1-Score.  
- Discutir os resultados e propor melhorias.

## 4. Metodologia

A metodologia adotada contempla as seguintes etapas:  
  
- Realização de revisão bibliográfica sobre análise de sentimentos, PLN e aprendizado de máquina.  
- Coleta de dados através de datasets públicos ou raspagem de textos de redes sociais.  
- Pré-processamento dos dados, incluindo limpeza, tokenização, remoção de stopwords, lematização e vetorização.  
- Desenvolvimento de modelos utilizando algoritmos de aprendizado de máquina (Naive Bayes, SVM, Random Forest) e deep learning (redes neurais, BERT, BERTimbau).  
- Avaliação dos modelos por meio de métricas estatísticas (acurácia, precisão, recall e F1-Score).  
- Análise crítica dos resultados obtidos e sugestões de aprimoramento.

## 5. Ferramentas e Tecnologias

As ferramentas e tecnologias que serão utilizadas incluem:  
  
- Linguagem de Programação: Python.  
- Bibliotecas: NLTK, spaCy, Scikit-Learn, Pandas, TensorFlow, Keras, HuggingFace (Transformers).  
- Ambientes de desenvolvimento: Google Colab, Jupyter Notebook e, se necessário, AWS ou Google Cloud.

## 6. Cronograma

O cronograma do projeto está organizado nas seguintes etapas:  
  
- Levantamento teórico e revisão bibliográfica: Mês 1 e 2.  
- Coleta e preparação dos dados: Mês 3.  
- Desenvolvimento e treinamento dos modelos: Mês 4 e 5.  
- Avaliação dos modelos e análise dos resultados: Mês 6.  
- Escrita do relatório final e preparação para defesa: Mês 7 e 8.

## 7. Referências

- Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. IEEE Intelligent Systems.  
- Souza, F., Nogueira, R., & Lotufo, R. (2020). BERTimbau: Pretrained BERT Models for Brazilian Portuguese. Brazilian Conference on Intelligent Systems.  
- Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. O'Reilly Media.  
- Jurafsky, D., & Martin, J. H. (2021). Speech and Language Processing. Prentice Hall.  
- Aggarwal, C. C., & Zhai, C. (2012). Mining Text Data. Springer.