

**Universidade de São Paulo**  
**Faculdade de Filosofia, Letras e Ciências Humanas**  
**Departamento de Ciência Política**

**FLP0478 - Processamento de Língua Natural Aplicada para Ciência Política e Análise de Políticas Públicas (2024)**

**2º semestre / 2024**

**Laboratório Aula 08**

**Professora Lorena G. Barberia**

**Prazo: 21 de outubro, 2024**

A tarefa deste laboratório será a de montar o seu script em Python do zero, desde a descrição dos seus dados até a execução de um modelo logístico multinomial, classificando o posicionamento dos tweets (todas as classes) com relação às vacinas de COVID-19. O trabalho deve estar dividido nas seguintes seções

1. Introdução
2. Descrição dos Dados
  - Tweets Relevantes/Irrelevantes
  - Distribuição das Classes em Posicionamento
  - Possíveis consequências da distribuição
3. Modelo Multinomial
  - a. Divisão Treino/Teste
  - b. Quais os hiperparâmetros? (e.g. n-gram, solver, penalty, and regularization strength)
  - c. Matriz de Confusão e Relatório de Classificação
  - d. Conclusão sobre os resultados
4. K-fold e *Tuning* Dos Hiperparâmetros
  - a. Faça um K-fold Cross-validation (defina o K - mais um hiperparâmetro)
  - b. Quais as conclusões?
  - c. Altere algum(ns) dos hiperparâmetros e refaça a classificação
  - d. O resultado mudou? O que você conclui?

Vocês devem entregar um relatório de até 05 páginas em .doc ou .pdf, junto do seu script em Python (.ipynb), no moodle na caixa do “Laboratório Aula 08” de seu respectivo turno.

Dica: [Sklearn doc](#)