

PLN para a Ciência Política e Políticas Públicas Públicas

Professora: Lorena Barberia

Departamento de Ciência Política-USP

lorenabarberia@usp.br



Tópicos da Aula

- 1 Sobre o curso
- 2 Funcionamento do curso

Sobre a disciplina

Metodologia *Project-Based Learning* ancorada no projeto *Mapping Political Elites COVID-19 Vaccine Tweets in Brazil*.

O que é uma disciplina de *Project-Based Learning* ?

- Um curso *Project-Based Learning* tem como principais características:
 - É orientado por uma questão central;
 - Apresenta de forma práticas problemas relacionados à pesquisa;
 - Tem como foco a discussão sobre um problema de pesquisa;
 - Tem maior profundidade na discussão sobre desenho de pesquisa.

Objetivos do Curso

- A disciplina tem como **principal** objetivo:
 - Capacitar o aluno a refletir sobre as implicações do desenho de pesquisa em processamento de linguagem natural.
- Outros objetivos da disciplina são:
 - Capacitar o aluno para leitura de códigos e aplicações em Python;
 - Ensinar um protocolo para a aplicação de PLN para a Ciência Política;
 - Introduzir o aluno ao Aprendizado de Máquina priorizando métodos supervisionados;
 - Capacitar o aluno na criação de bancos de dados e corpus apropriados para tarefas de aprendizado de máquina.

O que o curso não é?

- Este não é um curso:
 - Sobre Python
 - Sobre Diferentes Abordagens e/ou Métodos de Aprendizado de Máquina
 - Sobre a matemática/estatística utilizada nos modelos de aprendizado de máquina

Contexto em 2020

- Bolsonaro e seus aliados foram contra restrições durante a pandemia da COVID-19.
- As vacinas estavam sendo desenvolvidas, e sua aprovação somente ocorreu no final de 2020.
- A maioria do Congresso, do STF, dos governadores, e das lideranças políticas se posicionaram a favor da medidas sanitárias.



Figure: Bolsonaro oferece hidroxiclороquina a uma Ema

Contexto em 2021

- A vacinação se inicia em janeiro de 2021 com cobertura limitada e poucas vacinas disponíveis.
- Há diferentes imunizantes sendo aprovados e utilizados no Brasil (CoronaVac, AstraZeneca, Pfizer, J&J).
- Há incertezas sobre imunidade e novas variantes com escape vacinal.



Figure: João Dória e a primeira pessoa imunizada no Brasil

Motivação

- Os métodos de aprendizado supervisionado podem nos ajudar a classificar o debate nas redes sociais (e.g., Twitter).
- O monitoramento é ainda mais importante considerando que devido as restrições da pandemia, muitos candidatos e políticos aumentaram o uso de mídias sociais para se comunicar com sua base (e.g. parlamentos fechados para deliberação e tramitação), e o eleitorado também teve engajamento nas redes.

O Problema em Aprendizado com Supervisão

- Y = classificação (e.g. posicionamento, sentimento, etc.)
- X = vetor de características preditoras
- *Training Data Set* = $(x_1, y_1, \dots, x_N, y_N)$
- Coleta e Avaliação de outras Amostras para avaliar se o *training data set* corretamente classifica outras amostras
- Statistical Framework: $y = f(x) + \epsilon$

Nossos Objetivos

- Precisão (*Accuracy vs Precision*)
- Incerteza
- Validação e Robustez
- Replicabilidade e Transparência

Funcionamento do curso

- Modelo de aprendizado ativo.
- Etapas do curso:
 - Nivelamento de Python (aulas 1-5);
 - Introdução ao Aprendizado de Máquina (aulas 6-11);
 - Introdução ao BERT (aulas 12-15).
- Avaliação:
 - Listas semanais (Graduação e Pós-graduação);
 - Trabalho final (Pós-graduação).

Monitoria

- Como funciona:
 - Venha com dúvidas pontuais, depois de ter executado o código.
 - Duas vezes por semana em horário a ser definido com as turmas.
 - Presencial/Online.
- Monitores:
 - Pedro Schmalz (pedrosantanaschmalz@usp.br)
 - Rebeca Carvalho (rebeca.jesus.carvalho@gmail.com)
 - Guilherme Tiengo (guilherme.tiengo@usp.br)

Dúvidas?