

PLN para a Ciência Política e Políticas Públicas Públicas

Professora: Lorena Barberia

Semana 7



Tópicos da Aula

- 1 Visão Geral de Classificação Supervisionada
- 2 Classificação dos Documentos
- 3 Aprendizado Estatístico com um Modelo de Classificação

Desenho de Pesquisa: Seleção

- Amostra;
- *Operacionalização do Conceito + Classificação (Training Data Set);*
- Modelo de Aprendizado.

Classificação Supervisionada

- Nosso objetivo é selecionar uma amostra de documentos classificados para criar um modelo estatístico, e utilizar este modelo para classificar outros documentos fora da amostra.
- Definimos um modelo tal que nossa previsão y para o documento i é gerado por uma função $f_{\mu}(W_i,)$, onde W_i é a representação do documento e μ são os parâmetros do modelo.
- Aprendemos então os parâmetros do modelo observando amostras aleatórias da entrada W e da saída y (chamadas de **amostras de treinamento**):

$$\hat{y} = f_{\mu}(W_i)$$

Classificação Supervisionada

- São 03 etapas necessárias:
 - 1 Classificação dos documentos (Definição e Categorização);
 - 2 Treinamento do modelo (Aprendizado do Modelo);
 - 3 Validação (Inferência).

Avaliação da Classificação

		Predicted Class		
		Positive	Negative	
Actual Class	Positive	True Positive (TP)	False Negative (FN) Type II Error	Sensitivity $\frac{TP}{(TP + FN)}$
	Negative	False Positive (FP) Type I Error	True Negative (TN)	Specificity $\frac{TN}{(TN + FP)}$
		Precision $\frac{TP}{(TP + FP)}$	Negative Predictive Value $\frac{TN}{(TN + FN)}$	Accuracy $\frac{TP + TN}{(TP + TN + FP + FN)}$

Table 1. Confusion matrix with advanced classification metrics



Características de uma Estratégia de Classificação com Maior Confiança/Menor Viés

- 1 Objetivo, concordância entre conceito e mensuração
- 2 Regras definidas ex-ante
- 3 Confiabilidade
- 4 Validade
- 5 Replicável

Os 3 Conceitos e Classificação dos Documentos

- 1 **Conceito:** Conteúdo Relevante sobre Vacinas e Vacinação contra SARS-CoV-2;
- 2 **Conceito:** Posicionamento sobre Vacinas e Vacinação contra SARS-CoV-2;
- 3 **Conceito:** Conteúdo Relevante sobre Vacinas e Vacinação contra SARS-CoV-2 em Crianças e Adolescentes.

Relevância

- Postagens em que os conteúdos se referiam a vacinas e vacinação contra a COVID-19 receberam pontuação 1 (**Sim**), enquanto postagens que continham apenas palavras-chave, mas não abordavam vacinas e vacinação contra a COVID-19 receberam pontuação 0 (**Não**).

Irrelevância

Postagens consideradas **não relevantes** incluem mensagens referentes a:

- Vacinação em animais;
- Vacinas como metáfora para se referir a outro tema (por exemplo, transparência como vacina contra a corrupção, etc); ou
- A mensagem aborda as vacinas e vacinação contra a COVID-19 de forma indireta.

04 categorias de Posicionamento (*Stance*)

- O conteúdo escrito das postagens foi classificado como pertencente a uma de quatro categorias:
 - 1 Favorável
 - 2 Neutro
 - 3 Desfavorável
 - 4 Não Aplicável

Posicionamento Favorável (*Stance*)

- Os tweets foram classificados como **“Favoráveis”** se a mensagem:
 - Apela ao desenvolvimento e aprovação de vacinas, enfatizando a urgência e a necessidade das vacinas;
 - Defende a vacinação da população com uma ou todas as vacinas disponíveis;
 - Manifesta confiança no desenvolvimento de vacinas e da ciência;
 - Apoia acordos para o desenvolvimento ou aquisição de vacinas;
 - Apoia aprovações de vacinas e o início de campanhas de imunização; ou
 - Apoia medidas em favor da exigência da obrigatoriedade da vacinação e/ou restrições para não vacinados.

Posicionamento Neutro (*Stance*)

- Os tweets foram classificados como “**Neutro**” se a mensagem:
 - Não expressa uma posição clara sobre vacinas ou vacinação.

Posicionamento Desfavorável (*Stance*)

- Os tweets foram classificados como **“Desfavoráveis”** se a mensagem critica ou questiona:
 - Aprovação, aquisição e adoção de vacinas ou campanhas de vacinação;
 - Uma marca/tipo de vacina;
 - A eficácia ou manifesta uma falta de confiança em vacinas e as evidências baseadas em ensaios clínicos;
 - As vacinas por seus efeitos secundários ou pela falta de segurança;

Posicionamento Desfavorável (*Stance*)

- Os tweets foram classificados como “**Desfavoráveis**” se a mensagem critica ou questiona:
 - O posicionamento de organizações internacionais de saúde, empresas e laboratórios farmacêuticos, instituições de saúde, e agências reguladoras públicas de saúde com respeito a suas aprovação e uso das vacinas da COVID-19;
 - A restrição de atividades na ausência de cobertura vacinal em massa e a eficácia do isolamento social para combater a propagação do vírus; ou
 - Vacinação obrigatória, passaportes de vacinas, etc.

Não Aplicável (*Stance*)

- Os tweets foram classificados como “**Não Aplicáveis**” caso:
 - Tenham sido classificados anteriormente como **irrelevantes**, pois não abordavam vacinas e vacinação contra a COVID-19.

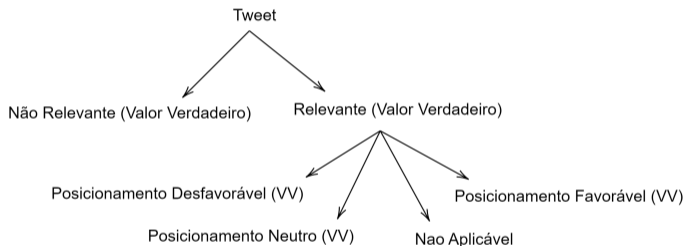
03 categorias de Posicionamento (*Stance*)

- 1 Favorável
- 2 Neutro
- 3 Desfavorável

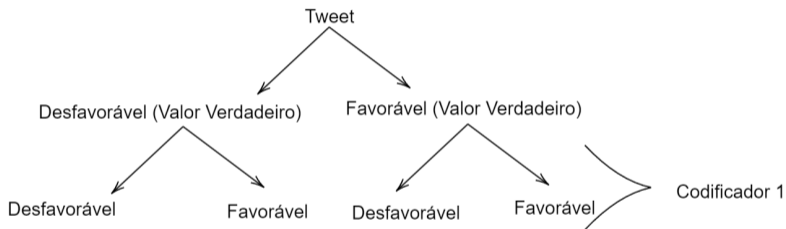
05 categorias de Posicionamento (*Stance*)

- 1 **Muito Favorável**
- 2 **Pouco Favorável**
- 3 **Neutro**
- 4 **Pouco Desfavorável**
- 5 **Muito Desfavorável**

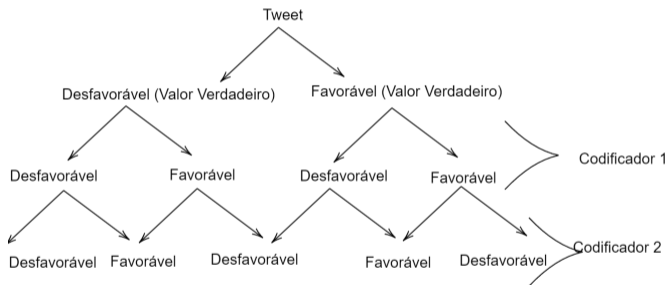
A Estratégia de Classificação de Relevância e Posicionamento sobre Vacinas da COVID-19 no Projeto



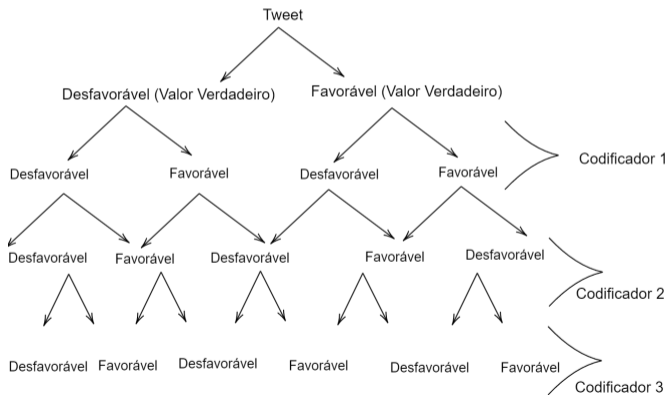
Exemplo: Implementação de Classificação com 1 classificador de Posicionamento Binário



Exemplo: Implementação de Classificação com 2 classificadores de Posicionamento Binário



Exemplo: Implementação de Classificação com 3 classificadores de Posicionamento Binário



Aprendizado Estatístico com um Modelo de Classificação

$$Y = f(X) + \varepsilon$$

O Problema em Aprendizado com Supervisão

- Y = classificação (e.g. posicionamento);
- 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: $y = f(x) + \epsilon$.



Resultados de Classificação: 06 "experimentos"

Tweets classificados	Sem. 2	Sem. 3	Sem. 4	Sem. 5	Sem. 6	Total
Um codificador	113	84	241	302	182	922
Dois codificadores	527	549	352	251	8	1687

Table: Número de tweets classificados por semana sobre relevância e posicionamento.

Resultados de Classificação: Concordância sobre Relevância dos Tweets

Sobre vacinação	Sem. 2	Sem. 3	Sem. 4	Sem. 5	Sem. 6	Total
Grau de concordância	87%	85%	83%	89%	75%	85%
Tweets avaliados	527	549	352	251	8	1687

Table: Concordância de tweets avaliados por dois codificadores.

Resultados de Classificação: Concordância sobre Posicionamento dos Tweets

Posicionamento	Sem. 2	Sem. 3	Sem. 4	Sem. 5	Sem. 6	Total
Sem filtro - relevância	68%	66%	60%	65%	50%	65 %
Com filtro - relevância	75%	73%	69%	72%	67%	73%
Tweets avaliados	527	549	352	251	8	1687

Table: Concordância de tweets avaliados por dois codificadores - Posicionamento.

Resultados de Classificação: Detalhamento da Concordância sobre Posicionamento dos Tweets

Posicionamento	Sem. 2	Sem. 3	Sem. 4	Sem. 5	Sem. 6	Total
Favorável	83%	87%	78%	87%	50%	81%
Neutro	14%	10%	18%	10%	50%	13%
Desfavorável	2%	3%	3%	1%	-	2%
Não aplicável	1%	1%	1%	1%	-	4%
Tweets avaliados	307	302	181	143	4	937

Table: Concordância de tweets avaliados por dois codificadores - Posicionamento.

Resultados de Classificação: Detalhamento da Discordância sobre Posicionamento dos Tweets

Posicionamento	Sem. 2	Sem. 3	Sem. 4	Sem. 5	Sem. 6	Total
Favorável	39%	39%	38%	40%	50%	63%
Neutro	40%	38%	41%	38%	50%	25%
Desfavorável	4%	3%	2%	2%	-	3%
Não aplicável	17%	20%	19%	19%	-	9%
Avaliações consideradas	220	246	180	126	4	776

Table: Discordância de tweets avaliados por dois codificadores - Posicionamento.

Laboratório 7