

# Exercício prático 1

Ivandré Paraboni  
EACH - USP

# Tarefas

Contexto: um sistema de geração de língua natural produziu textos (versículos da bíblia) em diferentes estilos. O objetivo do EP é avaliar se os textos gerados se adequam ou não ao estilo pretendido.

Dados de desenvolvimento/treino:

- train\_arcaico\_moderno.csv
- train\_complexo\_simples.csv
- train\_literal\_dinâmico.csv

Cada conjunto representa uma tarefa de classificação binária independente.

Tarefa: criar um classificador para cada conjunto, decidindo se um dado par de textos de teste (a ser fornecido mais tarde) pertence a um determinado estilo (arcaico ou moderno, etc.).  
Pode ser desenvolvido em qualquer linguagem, usando qualquer técnica ou algoritmo.

	A	B
1	text	style
2	És a fonte dos jardins; poço arcaico	
3	Anda com os sábios e serás arcaico	
4	Porque eu; o Senhor; não m arcaico	
5	E; ainda que nunca viu o sol arcaico	
6	todo o reino de Ogue; em B arcaico	
7	Porque; de repente; se leva arcaico	
8	E Jeú dormiu com seus pais; arcaico	
9	Bate alegre as asas o avestr arcaico	
10	Porque assim diz o Senhor: arcaico	
11	quanto mais o sangue de Cr arcaico	
12	Mas foi arrancada com furc arcaico	
13	Diz o preguiçoso: Um leão e arcaico	
14	Porque Deus pôs nas suas n arcaico	
15	e aos horeus no seu monte arcaico	
16	E Manassés dormiu com sei arcaico	
17	Ainda não compreendeis qu arcaico	
18	Vieram os filhos de Jacó ao arcaico	
19	Ei nascido mesmo aqui; arcaico	

# Cronograma

- **Entrega 1 até 13/10:** zip contendo [1] código e [2] relatório da implementação conforme modelo, incluindo resultados do treinamento (acurácia média de 10 folds).

Logo após este prazo, serão divulgados os dados de teste, havendo menos de 1 semana de prazo para a entrega final.

- **Entrega 2 até 20/10:** zip contendo [1] resultados de teste (planilha com rótulos de previsões de teste), e [2] slides de apresentação (10 min).
- **Apresentação dos trabalhos em aula em 22 e 29/10**

# Avaliação

- A avaliação é baseada na acurácia média do problema de classificação.
- Critérios: acurácia mínima entre treino e teste (70%), método inovador / bem elaborado (20%), relatório (5%), apresentação em aula (5%).

# Penalidades previstas

- **Baixo desempenho:** EPs cuja acurácia média seja inferior à média de um baseline de classe majoritária recebem nota 3,0 e não são avaliados.
- **Formato inválido:** 10% de desconto na nota final por entrega de arquivo de previsões em formato diferente do esperado.

# Normalização e média final

- Trabalhos são ordenados por resultado global, e a nota máxima é ajustada para dez, e as demais na mesma proporção.

# Avaliação dos resultados

- É possível que haja alguma queda de desempenho entre treinamento e teste
- Mas uma diferença muito grande indica *overfitting*, e que portanto o modelo não funciona
- Um resultado espetacular no treinamento, que depois reduz drasticamente, não é um resultado real
- Como se proteger disso?
  - Pensando muito bem em quais características modelar
  - Realizando um grande número de testes com muitos parâmetros, algoritmos, representações etc. usando *grid search*
  - Realizando validação cruzada

# O que é esperado desta tarefa

- Construir sua própria infraestrutura de desenvolvimento, e um pipeline robusto de experimentação (pensando no EP-2!)
- Resultados satisfatórios em relação ao baseline, e com pouco ou nenhum *overfitting*
- Despertar para as dificuldades práticas de modelar problemas de PLN usando aprendizado supervisionado
  - Em um problema simples, com dados balanceados e em pequeno volume
- Apresentação satisfatória do trabalho realizado na forma escrita e oral

Obrigado