

## ACH2118 Introdução ao Processamento de Língua Natural

Semestre 2020-2 - Exercício prático 1 – Detecção de posicionamentos morais

Estagiário PAE: Samuel Caetano da Silva ([samuel.caetano.silva@usp.br](mailto:samuel.caetano.silva@usp.br))

1. O objetivo do trabalho é implementar um classificador de posicionamentos de natureza moral (sobre legalização da posse de armas e aborto) a partir de texto com base nos dois conjuntos de dados disponível na atividade.
2. O trabalho é desenvolvido em duas partes: a primeira entrega consiste em desenvolver o classificador e reportar resultados médios de medida F1 macro usando validação cruzada de 10 partições sobre o conjunto de dados de treinamento completo, a ser avaliado de acordo com os critérios discutidos a seguir. A segunda entrega consiste em atualizar o relatório reportando os resultados do classificador aplicado ao conjunto de teste a ser fornecido.
3. Entrega 1 (treinamento): uma pasta ZIP contendo o código completo da implementação; e o relatório descritivo do classificador proposto e seus resultados de validação cruzada sobre os dados de treinamento, conforme modelo disponível no Tidia
4. Entrega 2 (teste): (a) relatório atualizado com resultados completos (precisão, cobertura e F1) com base nos dados de teste, e (b) uma breve apresentação (10 slides em formato PDF) do trabalho realizado.

**Importante:** a entrega deve ser feita por apenas um integrante de cada grupo, certificando-se de que o arquivo realmente foi carregado com sucesso, que é a versão correta, e que não está corrompido.

Critérios de avaliação. Para obter um bom resultado, observe atentamente como seu trabalho será avaliado: resultado (50%), inovação (30%), relatório (10%) e apresentação (10%). Detalhes serão discutidos em aula e apresentados nos slides da disciplina.

Para a avaliação 1, programadores Python devem usar

```
result = cross_val_score(clf, x_train, y_train, scoring='f1_macro', cv=10).mean()
```

E para a avaliação final, use `classification_report`:

[https://scikit-learn.org/stable/modules/generated/sklearn.metrics.classification\\_report.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.classification_report.html)