



ATIVIDADE 3 DE PLN

ATIVIDADE 3

- Em grupos (podem ser os mesmos da atividade 1)
- Tarefa: classificação de inferência textual
 - Dado um par de sentenças, classificar como “*entailment*” ou “*none*”
 - Exemplo de *entailment*
 - Uma pessoa está cortando carne em pedaços
 - A pessoa não está fatiando um dente de alho em pedaços
 - Exemplo de *none*
 - Um homem está fatiando manteiga
 - Uma mulher está fatiando tofu

PASSO A PASSO

1. Ler artigo sobre a avaliação conjunta (realizada no STIL 2019)
2. Explorar conjunto de dados: ASSIN 2
 - *Os dados de treino e validação são compostos, respectivamente, por 6.500 e 500 pares de sentenças em português do Brasil, anotados para inferência e similaridade semântica. Os valores de similaridade semântica vão de 1 a 5, e as classes de inferência são entailment e none. Os dados de teste são compostos por aproximadamente 3.000 pares de sentenças, com os mesmos fenômenos linguísticos e anotações dos dados de treinamento. Todos os dados foram anotados manualmente.*

PASSO A PASSO

3. Propor uma solução de classificação baseada em *bag of words*
4. Propor uma solução de classificação baseada nas *word embeddings* pré-treinadas do NILC
5. Realizar avaliação das duas soluções, explorando configurações variadas para cada solução
 - Medidas de avaliação usuais
 - Análise de erros e avaliação qualitativa
 - Lições aprendidas sobre a tarefa
 - Lições aprendidas sobre as duas soluções propostas, destacando suas vantagens e limitações

PASSO A PASSO (UM PARÊNTESE)

- Configurações variadas
 - Números de atributos
 - Valores dos atributos
 - Possível seleção de atributos
 - Possível agregação de atributos
 - Técnicas diferentes de aprendizado de máquina

PASSO A PASSO

6. Preparar relatório (~10 páginas) e submeter no e-Disciplinas até 10/maio
 - Atenção: relatório em nível de pós-graduação!
 7. Apresentar o trabalho na aula do dia 17/maio
 - 5min por grupo
-
- Para os fortes: propor e avaliar uma 3ª solução que seja simbólica