UFRPE - Pró-Reitoria de Pesquisa e Pós-Graduação

Programa de Pós-Graduação em Informática Aplicada (PPGIA)

Disciplina: Mineração de Textos – Prof. Rinaldo Lima

**Miniprojeto em Grupo -** Classificação de Textos

Data de Entrega: 09/05/2018 - Enviar ao email: rjlima01@gmail.com



## 6a. LISTA DE EXERCÍCIOS

(1) Usando o Weka, crie os classificadores listados abaixo usando a representação Bag of Words e mostre uma tabela comparativa considerando as seguintes métricas Precision, Recall, F-measure e Acurácia.

Classificadores a se avaliar:

- Naïve Bayes
- k-NN (use k = 5)
- Decistion Tree (J48)
- Decision List (PART)
- Support Vector Machine (SMO)
- Logistic Regression

## **Considere:**

- O dataset: Polarity Detection.
- Uso de *Stoplist* e *Stemming* como pré-processamento
- Usar os métodos de ponderação das palavras (cálculo de frequência)
  - o *Multinomial*, onde se registra, na matriz de representação dos documentos, a frêquencia do token (peson = 0 até n), como está nos slides da aula.
  - o *Binário*, considera apenas se o termo aparece ou não (peso = 0 ou 1)
- Avalie o seu classificador usando 5-fold Cross Validation (N = 5)

## Material extra de apoio para os exercícios

Ver os links no arquivo Links.txt.