

#### 4ª. LISTA DE EXERCÍCIOS

- *Dividir a turma em 2 equipes e implementar o que se pede.*
- *Apresentar, em até 15 minutos, os resultados obtidos.*
- *Cada membro da equipe deverá participar da apresentação*

##### **Equipe 1:**

**Implementar o algoritmo *Naive Bayes* para classificação binária, isto é, considerando apenas 2 classes, usando a representação *Bag of Words* levando em conta também:**

- O dataset: *Polarity Detection*.
- Uso de *Stoplist* e *Stemming* como pré-processamento
- Usar os métodos de ponderação das palavras (cálculo de frequência)
  - *Multinomial*, onde se registra, na matriz de representação dos documentos, a frequência do token (peso = 0 até n). Como está nos slides da aula.
  - *Binário*, considera apenas se o termo aparece ou não (peso = 0 ou 1)
- Avalie o seu classificador usando 5-fold Cross Validation (N = 5)
- Exiba os resultados usando a métrica de *Accuracy*
- Use *Add-1 smoothing* (como está nos slides da aula)
- Em vez da multiplicação de probabilidades, use a soma dos *logs*.
- Discutir os resultados encontrados.

## 1. Equipe 2:

**Implementar o algoritmo  $k$ -NN para classificação binária, isto é, considerando apenas 2 classes, usando a representação *Bag of Words* levando em conta também:**

- O dataset: *Polarity Detection*.
- Uso de *Stoplist* e *Stemming* como pré-processamento
- Usar os métodos de similaridade
  - Shared Word Count, se a palavra parece ou não (peso = 0 ou 1)
  - Word Count + Bonus (como nos slides da aula)
- Avalie o seu classificador usando 5-fold Cross Validation ( $N = 5$ ).
- Exiba os resultados usando as métricas de Accuracy e gere um gráfico da métrica *Accuracy* para  $k = \{3, 4, 5, 6, 7, 8, 9, 10\}$ 
  - Para qual valor de  $k$ , o algoritmo teve melhor desempenho? Discutir os resultados.

---

### Material extra de apoio para os exercícios

Ver os links no arquivo **Links.txt**.