

# Reconhecimento de padrões – interpretação de comentários

A pasta DADOS contém os seguintes arquivos:

- 1) PALAVRASpc.txt- Lista de palavras vetorizadas, uma por linha (n= 9538).
- 2) WWRDpc.dat- Vetores com 100 coordenadas, um por linha, correspondentes às palavras do item anterior (n=9538) ;
- 3) WTEXpc.dat- Vetores com 100 coordenadas de textos classificados (n=10400). Cada vetor de texto é calculado como a média dos vetores das palavras que ocorrem nele, sendo que as palavras que não estão na lista apresentada devem ser ignoradas.
- 4) CLtx.dat- Classificação de cada texto de WTEXpc. Textos considerados como **positivos** são classificados com 1 (um) na linha correspondente de CLtx e 0 (zero) caso a supervisão indique como sendo um texto negativo.

## O Trabalho

A proposta do trabalho é desenvolver um modelo de reconhecimento de padrões para classificar comentários, com base nos dados fornecidos. O roteiro sugerido consiste em:

- a) Treinar e **testar** um modelo de classificação baseado nos vetores do item 3 com as classes do item 4.
- b) Determinar um novo conjunto de textos (positivos e negativos) e verificar a habilidade de classificação do modelo em casos reais. Discutir resultados.
- c) Propor uma ferramenta de classificação executável com entrada livre para o usuário (uma-a-uma).
- d) O Trabalho pode ser realizado em equipes de até **três pessoas**.

## Avaliação

- e) Apresentar **relatório** descrevendo todas as etapas desenvolvidas.
- f) A avaliação levará em conta a qualidade e a originalidade da solução, bem como a programação envolvidas.
- g) Entregar relatório impresso contendo os tópicos:
  - . Resumo;
  - . Apresentação e introdução referenciando recursos utilizados;
  - . Obtenção e classificação dos padrões;
  - . Extração de características;
  - . Escolha do classificador;
  - . Testes de performance;
  - . Aplicação do modelo em comentários não utilizados para o treinamento;
  - . Conclusão.
- h) Apresentação da solução para a turma no último dia da disciplina para discussão.