



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO

Rua Dom Manoel de Medeiros, s/n – Dois Irmãos 52171-900 Recife-PE

Fone: 81 3302 1000 www.bcc.ufrpe.br

DISCIPLINA: Aprendizagem de Máquinas	CÓDIGO: 14719
DEPARTAMENTO: de Computação	ÁREA: Informática
CURSO: Bacharelado em Ciência da Computação	
PROFESSOR RESPONSÁVEL: Luciano Demétrio Santos Pacífico	
DATA MÁXIMA DE ENTREGA: 11-11-2018	

Projeto da 1ª Verificação de Aprendizagem

1. Execute experimentos conforme especificado abaixo:

BASE DE DADOS

- a. Cada aluno deve fazer uso da base de dados especificada na Atividade 1 do Projeto (ML-P01).
- b. As bases de dados adotadas devem apresentar o mínimo de 500 padrões (dados da base).
- c. Não é permitido o uso do mesmo problema ou base de dados por alunos diferentes.

PRÉ-PROCESSAMENTO

- d. Ao menos um método de pré-processamento deve ser usado, gerando uma nova base de dados.
- e. O tipo de pré-processamento utilizado deve estar relacionado ao contexto da aplicação.
- f. Remoção de vírgulas, espaços em branco, identificador dos padrões, etc. **não serão considerados pré-processamento válidos.**

EXPERIMENTOS

- g. Os experimentos devem ser executados de acordo com o esquema abaixo para cada uma das bases de dados geradas (tanto a base de dados “brutos” quanto a base de dado pré-processadas):
 - i. **Deve-se executar o 10-fold cross-validation 5 vezes para cada base de dados**, com cada uma das cinco execuções partindo de uma distribuição aleatória dos dados entre cada *fold*, resultando em um total de 50 experimentos por base de dados (10 x 5).
 - ii. Em cada um dos 50 experimentos, **os conjuntos de treinamento e teste devem ser mantido o mesmo para cada algoritmo a ser testado** (mesmo ponto de partida para cada modelo), de modo a obter-se uma avaliação justa dos resultados.
 - iii. Os algoritmos a serem testados são: **Árvore de Decisão, Naïve Bayes, K-Vizinhos Mais Próximos (K-NN), K-Vizinhos Mais Próximos**

Ponderado (W-K-NN), Máquina de Vetores de Suporte (SVM) e Regressão Logística.

- iv. Deve-se testar o K-NN variando-se **3 vezes o número do parâmetro k** , assim como o W-K-NN para os mesmos 3 valores de k escolhidos.
- v. A medida de distância adotada tanto no K-NN quanto no W-K-NN é a Distância Euclidiana.
- vi. A SVM deve ser testada com as funções de kernel **Linear** e **RBF**.

ANÁLISE EXPERIMENTAL

- h. A comparação deve ser realizada em relação à **Acurácia no Conjunto de Teste**.
- i. Deve-se calcular a **média**, o **desvio padrão** e a **mediana** da **Acurácia do Conjunto de Testes para os 50 testes** realizados para cada algoritmo em cada base de dados.
- j. Uma análise comparativa entre os algoritmos utilizados e suas variações deve ser realizada, de modo a identificar qual abordagem pode ser considerada a mais confiável para a resolução dos problemas analisados.
- k. Essa análise deve ser feita de forma empírica (baseado nas medidas obtidas) e através do uso de algum teste de hipóteses estatístico (t de *Student*, Teste de Friedman, ANOVA, etc.).

MATERIAL A SER ENTREGUE AO PROFESSOR

- l. Um relatório deve ser gerado de acordo com a especificação abaixo:
 - i. O relatório deve seguir o formato de um **artigo acadêmico**, em suas partes (título, resumo, introdução, metodologia, modelos apresentados, análise experimental, conclusões e referências).
 - ii. Deve-se descrever em detalhes o tipo de pré-processamento adotado para cada base de dados, justificando a escolha desse pré-processamento.
 - iii. Os resultados experimentais devem ser descritos de forma clara e concisa, de modo a justificarem-se as conclusões obtidas quanto aos modelos e problemas estudados.
 - iv. Uma conclusão geral deve ser apresentada.
 - v. O artigo deve ser escrito fazendo-se uso do template da SBC (coluna única, artigo contendo entre 10 e 14 páginas), ou o template do IEEE (coluna dupla, artigo contendo entre 6 e 8 páginas).
- m. Os códigos fonte comentados desenvolvidos.
- n. As bases de dados originais e pré-processadas.
- o. O slide da apresentação do projeto (em .pdf).

Observações:

- 1. **O Projeto é individual.**
- 2. O tempo de apresentação será de **15 minutos por aluno**, seguidos de até cinco minutos de perguntas.
- 3. Tanto o relatório de experimentos (artigo), quanto os códigos fonte, as bases de dados (sem pré-processamento e pré-processadas) e o slide de apresentação

devem ser enviadas em um único arquivo “zipado” para o e-mail ldsp.ufrpe@gmail.com até a data máxima de entrega da atividade.

4. O artigo deve ser submetido em formato .pdf.
5. Cópias serão **ANULADAS**.