



AVALIAÇÃO 3 – PARTE III

Realizar os passos a seguir para aplicação de uma técnica de mineração de dados usando o software Weka disponível em <http://www.baixaki.com.br/download/weka.htm>

1. Fazer download do resultado de um vestibular da UEM, escolhendo um curso do Listão Geral;
2. Preparar um arquivo texto com extensão ‘.arff’ à partir do relatório obtido no passo 1, seguindo o guia para conversão desse tipo de arquivo em formato Weka, conforme relacionado a seguir:
 - 1) Copiar arquivos de resultados dos vestibulares de 2011 e 2012 (inverno e verão) do curso escolhido para o Word;
 - 2) Excluir candidatos desistentes, cabeçalhos, número de inscrição, nome e coluna de totais;
 - 3) Substituir brancos por ‘*’;
 - 4) Substituir vários ‘*’ por um único ‘*’;
 - 5) Substituir ‘*’ por ‘;’;
 - 6) Selecionar tudo e converter para tabela;
 - 7) Salvar e abrir em planilha;
 - 8) Excluir colunas de inscrição, nome e totais (manter Escore Final);
 - 9) Alterar coluna de Escore Final para: APR (APROVADO), REP (REPROVADO) e CLAS (LISTA DE ESPERA);
 - 10) Discretizar notas das provas de conh.gerais, redação, português, língua estrangeira, artes e matemática, como no exemplo abaixo:
 - selecionar colunas de notas e eliminar decimais,
 - ordenar coluna de cada nota,
 - selecionar notas de 0 a 20 e substituir por CG_0_30,
 - selecionar notas de 21 a 30 e substituir por CG_31_50,
 - selecionar notas de 31 a 40 e substituir por CG_51_60,
 - selecionar notas de 41 a 50 e substituir por CG_61_70,
 - selecionar notas de 51 a 60 e substituir por CG_71_80,
 - selecionar notas igual ou acima de 61 e substituir por CG_81_
 - 11) Abrir arquivo no editor e incluir cabeçalho conforme abaixo:

```
@relation 'VEST_UEM'
@attribute ANO{2011, 2012}
@attribute COTISTA{SIM, NÃO}
@attribute NOTACG{CG_0_30, CG_31_50, CG_51_60, CG_61_70,
CG_71_80, CG_81_}
@attribute NOTARED{RED_0_30, RED_31_50, RED_51_60, RED_61_70,
RED_71_80, RED_81_}
...
@attribute CLASSIF{APR, REP, CLAS}
@data
```



Universidade Estadual de Maringá
Centro de Tecnologia
Departamento de Informática



Curso de Bacharelado em Ciência da Computação
Disciplina: Banco de Dados II (1752 – Turma 01)
Professora: Dra. Maria Madalena Dias

3. Abrir o arquivo ‘.arff’ na ferramenta Weka (opção do menu Aplicações -> Explorer) e executar o algoritmo de classificação J48 (opção Classify -> Choose -> trees -> J48), algumas vezes alternando com diferentes valores do grau de Confiança e com provas específicas. Execute inicialmente com os valores default e altere a Confiança para 0.6 ou 0.7 e execute novamente
4. Entregar um documento com o seguinte conteúdo:
 - Introdução do trabalho;
 - Informações sobre o vestibular curso(s) e escolhidos;
 - Resultado de cada execução realizada do algoritmo de classificação com resultados satisfatórios;
 - Análise dos resultados obtidos em cada execução do algoritmo de classificação.
 - Conclusão geral do trabalho, incluindo dificuldades encontradas.