

Trabalho T2:

Prática com ferramenta de mineração de dados Weka.

Mesmos grupos do Trabalho T1.

No arquivo TICtotal.zip, estão os arquivos:

TICdata2000.arff – arquivo com a planilha original de dados (5822 registros), preparada para uso pelo Weka

TICeval2000.arff – arquivo com a planilha do conjunto de teste (4000 registros), preparada para uso pelo Weka

TICEvalPositivo2000.csv – arquivo de exportação do Excel, contendo apenas os registros de TICeval2000 que têm CARAVAN = 1 (238 registros)

Para exercitar, usem TICdata2000.arff como conjunto de treino e TICeval2000.arff como conjunto de teste.

Executem o algoritmo J48 (implementação do C4.5 no Weka) e capturem a árvore de decisão produzida e as correspondentes matrizes de confusão e acurácias calculadas, para cada um dos arquivos.

Com os dados preparados no T1, por cada grupo, criar um arquivo .arff (é arquivo texto) e executar o mesmo algoritmo sobre esses dados, considerando eles como conjunto de treino.

Dica: exportar pelo Excel para arquivo .csv e substituir “;” por “,”, e incluir linhas de descrição dos dados como nos outros arquivos .arff.

Capturar a árvore produzida, a matriz de confusão e a acurácia calculada.

Como conjunto de teste, usem o arquivo TICEvalPositivo2000.csv, preparado conforme as preparações que vocês fizeram para o arquivo do T1, e verifiquem quantos acertos no final vocês tiveram em relação ao arquivo original. Para esta última execução, listar o Log do Weka completo.