Universidade Federal de Santa Maria – Centro de Tecnologia
Peso: 2
Disciplina: Mineração de Dados (ELC1098)
Professor: Dr. Joaquim Assunção
Modalidade REDE

Atenção: Leia atentamente as questões e entregue um **txt** ou **rtf** ou **doc** contendo as respostas.

1) Ao longo do semestre, estudamos 4 tipos diferentes de algoritmos para mineração de dados; associação, regressão, classificação e agrupamento. Alguns desses tipos não precisam ser treinados com a supervisão de um rótulo, em especial temos um tipo que aprende implicações entre variáveis. Cite o referido tipo e quais as três principais métricas para definir a quão boa são as saídas desses algoritmos (2):

_____
_____
_____
_____

2) Uma empresa decidiu fazer modelos de agrupamento para ajudar na tomada de decisão do gerente. Dois times foram incumbidos de gerar os agrupamentos. Ambos os times usufruíram do mesmo processo de extração, limpeza e organização dos dados; o que gerou o mesmo conjunto. Além disso, os times também conversaram para chegar na configuração ideal para o parâmetro k (nenhum outro parâmetro foi alterado). Porém, chegando no dia da entrega, havia agrupamentos diferentes. Cite um motivo para o resultado ser diferente (2.5). Essa diferença tende a ser significativa para grandes conjuntos? Justifique (1.5).

_____
_____
_____
_____
_____
_____
_____
_____
_____

3) Você entrou em um projeto de pesquisa que coleta dados de saúde pública. No final, você recebeu um conjunto de dados com N registros de M pessoas e Z características fisiológicas. As duas últimas colunas do conjunto são campos que dizem o diagnóstico da pessoa e o quão severo é o caso. Sua tarefa é criar um modelo para identificar as principais características para os casos severos da doença X. Antes de criar o modelo, quais seriam seus primeiros passos (3)? Tendo em vista que seria interessante ter um modelo autoexplicativo, qual algoritmo você ponderaria para gerar tal modelo (1)?

_____