

Universidade Federal de Pernambuco – UFPE

Programa de Pós Graduação em Ciência da Computação

Soluções de Mineração de Dados

IN1166

Prof: Leandro Maciel

Aluno: Irvin Soares Bezerra

Entrega da execução das tarefas de preparação dos dados identificadas na análise exploratória. Entregar relatório e scripts.

Etapas de pré-processamento para a base Dry Bean

- 1 – A base de dados apresenta 17 variáveis, sendo 14 variáveis do tipo float, 2 variáveis do tipo integer e uma variável object, contendo a classe resposta
- 2 – Não foi necessário completar dados ausentes, todas as variáveis da base estão integras sem dados ausentes, nulos, NaN ou caracteres especiais.
- 3 – As variáveis Area, Perimeter, MajorAxisLength, MinorAxisLength, Eccentricity, ConvexArea e EquivDiameter apresetam valores em uma escala maior que as demais variáveis.
- 4 – De acordo com o domínio do problema, será necessário classificar um dado grão de acordo com suas características físicas, para tal será necessário adotar modelos de classificação.
- 5 – Para realizar tal classificação, antes de aplicar um modelo os dados foram normalizados. Visando evitar problemas de overfitting e enviesamento no momento de treinamento do modelo.
- 6 – A análise estatística das variáveis não demonstrou a presença de valores discrepantes dos demais valores
- 7 – A variável categórica Class foi transformada em 7 variáveis binárias.
- 7 – A base de dados foi dividida em treino e teste da seguinte forma: 75% para Treino e 25 % para Teste.