



Universidade do Minho

Licenciatura em Engenharia Informática

NÚMERO _____

NOME _____

Aprendizagem e Decisão Inteligentes

3º ano, 2º semestre

Ano letivo 2023/2024

Prova escrita, época de recurso

4 de junho de 2024



GRUPO 1

(6 valores)

RESPONDA ÀS QUESTÕES DESTE GRUPO EM FOLHA DE TESTE SEPARADA.

QUESTÃO 1

A resolução de problemas por aplicação de técnicas de *Machine Learning* envolve a identificação do tipo de problema em questão.

Para problemas de classificação, dê 3 exemplos de métricas de qualidade, explicando o seu significado.

QUESTÃO 2

O trabalho em grupo envolveu a realização de duas tarefas. Uma delas envolvia a seleção de um problema cujos dados eram escolhidos pelo grupo de trabalho. A outra tarefa desenvolvia-se a partir de um *dataset* atribuído.

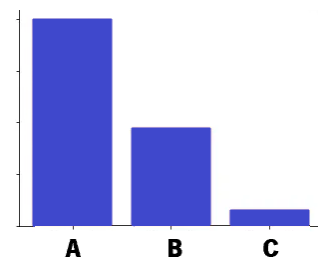
Considere a tarefa cujo *dataset* foi atribuído em função do número do grupo de trabalho:

- Sendo que o atributo *target* era discreto, explique que transformações se poderiam fazer aos dados para aplicar técnicas de previsão de resultados contínuos (regressão).
- Identifique duas técnicas de *Machine Learning* que permitam lidar com este *dataset* como sendo um problema de regressão e explique como o resultado dos modelos deveriam ser interpretados.

QUESTÃO 3

Considere o gráfico à direita que representa a frequência dos 3 valores (A/B/C), representados no eixo xx', de um atributo discreto que se apresenta claramente desequilibrado no que respeita à ocorrência de cada valor, identificada no eixo yy'.

Indique e explique 3 formas de preparação de dados que permitam minimizar o impacto que este desequilíbrio pode provocar na criação dos modelos.



GRUPO 2

(8 valores)

PARA CADA AFIRMAÇÃO APRESENTADA, RESPONDA ASSINALANDO A SUA VERACIDADE (V) OU FALSIDADE (F).

JUSTIFIQUE A RESPOSTA EXCLUSIVAMENTE NO ESPAÇO DISPONIBILIZADO.

NÃO SÃO CONSIDERADAS RESPOSTAS PARA AS QUAIS NÃO EXISTA JUSTIFICAÇÃO.

QUESTÃO 1

☐

Em Inteligência Artificial, a abordagem designada por *Soft Computing* difere da denominada *Hard Computing* por não ser tolerante à imprecisão, à incerteza e à verdade parcial.

QUESTÃO 2

- ☐ Numa rede neuronal artificial, o axónio é o ponto de ligação entre sinapse e neurónio, cujo valor representa a importância do sinal recebido, que pode ser estimulador, inibidor ou nulo.

QUESTÃO 3

- ☐ As metodologias para análise de dados CRISP-DM e SEMMA diferem, apenas, na quantidade de etapas que definem para o processo, respetivamente, 6 e 5 etapas.

QUESTÃO 4

- ☐ Na fase de preparação de dados, a discretização de atributos contínuos reduz o conhecimento que esses atributos representam.

QUESTÃO 5

- ☐ O nodo «Partitioning» do KNIME implementa uma estratégia de validação de modelos por cruzamento dos dados, dividindo o *dataset* em dois subconjuntos, um para treino e outro para teste do modelo.

QUESTÃO 6

- ☐ Algoritmos de segmentação/ *clustering* são técnicas de *Machine Learning* que se baseiam no cálculo de distâncias pelo que não são aplicáveis a atributos binários.

QUESTÃO 7

- ☐ Técnicas de *Machine Learning* suportadas por paradigmas de aprendizagem sem supervisão, como a segmentação/ *clustering*, não são adequadas para a resolução de problemas de classificação.

QUESTÃO 8

- ☐ Na fase de preparação de dados, um valor de correlação próximo de 0 (zero) significa que ambas as variáveis em estudo não apresentam correlação entre si.

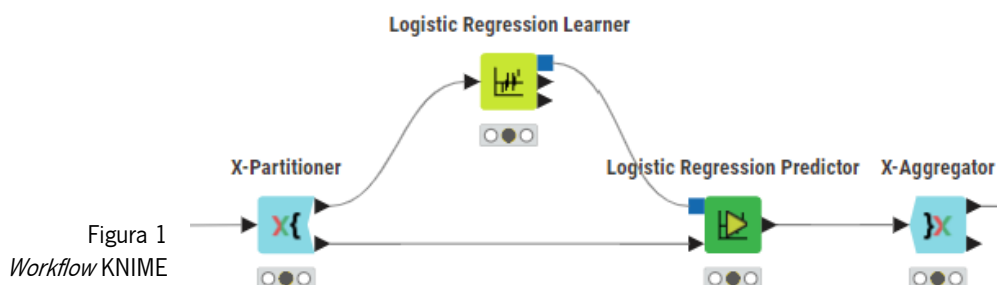
GRUPO 3

(6 valores)

AS QUESTÕES DESTE GRUPO LISTAM AFIRMAÇÕES VERDADEIRAS (V) OU FALSAS (F).
 PARA CADA QUESTÃO DESTE GRUPO, UMA AFIRMAÇÃO INCORRETAMENTE ASSINALADA ANULA OUTRA
 ASSINALADA CORRETAMENTE. UMA AFIRMAÇÃO NÃO ASSINALADA NÃO É CONSIDERADA.

QUESTÃO 1

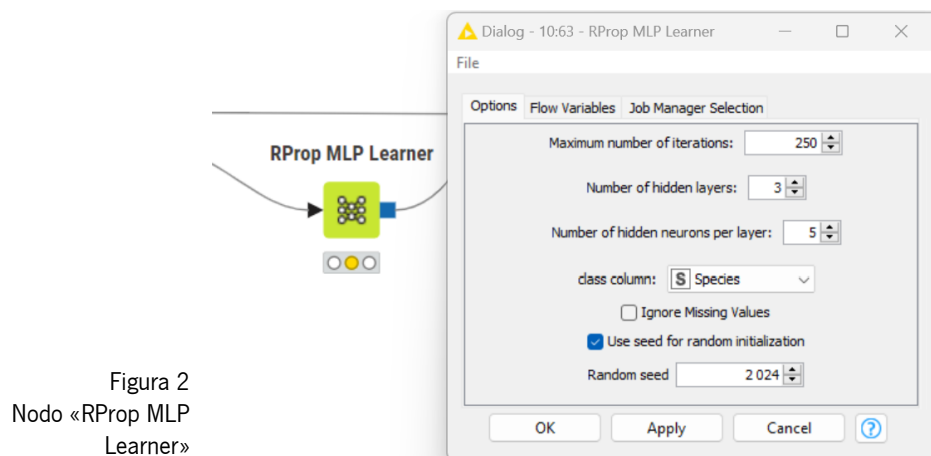
Considere o excerto de *workflow* KNIME apresentado na Figura 1.



- ☐ Os nós «X-Partitioner» e «X-Aggregator» estabelecem o ciclo de iterações de validação do modelo.
- ☐ No fluxo da Figura 1 é aplicada uma estratégia de validação *cross validation*.
- ☐ É usado o nó «X-Partitioner» para garantir a repetibilidade das experiências.
- ☐ O fluxo da Figura 1 está mal configurado porque falta ligar 2 dos 3 outputs.
- ☐ O nó «Logistic Regression Learner» constrói um modelo de previsão de valores contínuos.
- ☐ O nó «Logistic Regression Predictor» constrói um modelo de previsão de valores discretos.
- ☐ No fluxo da Figura 1, o sinal significa que os nós estão mal configurados.
- ☐ O nó «Logistic Regression Learner» constrói um modelo de classificação.

QUESTÃO 2

Considere o excerto de *workflow* KNIME apresentado na Figura 2.

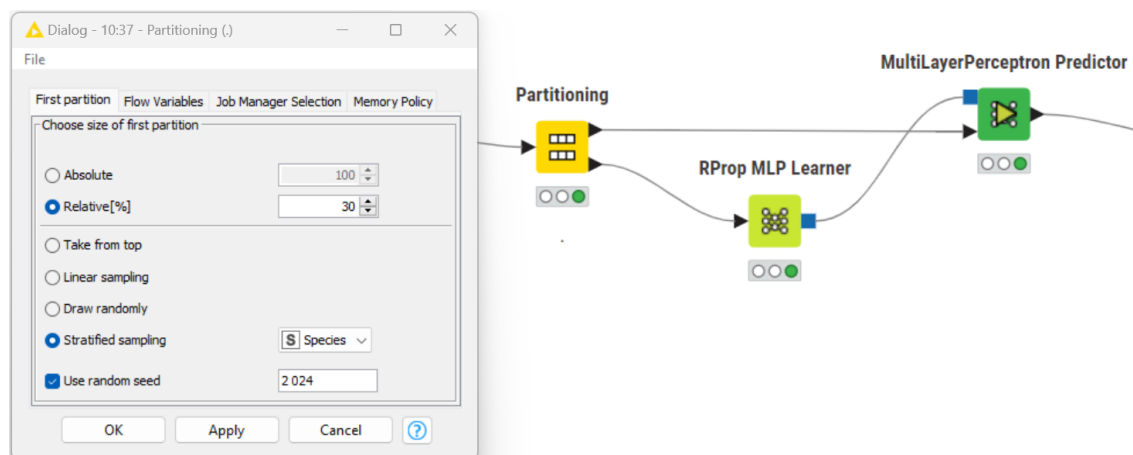


- ☐ O nódo «RProp MLP Learner» constrói um modelo suportado por Redes Neurais Artificiais.
- ☐ A rede neuronal criada é composta por 250 neurónios na camada de entrada.
- ☐ A configuração do nódo cria uma rede neuronal com 3 camadas intermédias.
- ☐ A rede neuronal criada é composta por 3 neurónios em cada camada intermédia.
- ☐ O nódo «RProp MLP Learner» constrói um modelo baseado em Redes Neurais Recorrentes.
- ☐ A configuração do nódo cria uma rede neuronal com 5 camadas.
- ☐ A configuração do nódo cria uma rede neuronal com 5 camadas intermédias.
- ☐ A rede neuronal criada é composta por 5 neurónios em cada camada intermédia.

QUESTÃO 3

Considere o excerto de *workflow* KNIME apresentado na Figura 3.

Figura 3
Workflow KNIME



- ☐ O nodo «Partitioning» particiona o *dataset* em 2 subconjuntos de dados.
- ☐ O fluxo da Figura 3 está errado porque envia os dados de teste para o nodo de treino.
- ☐ O nodo «Partitioning» divide o *dataset* em 2 partes, uma com 70% e outra com 30% dos dados.
- ☐ A configuração do nodo «Partitioning» garante a repetibilidade dos modelos criados pelo nodo «RProp MLP Learner».
- ☐ O fluxo da Figura 3 é inadequado porque está a enviar 70% dos dados para teste do modelo.
- ☐ A opção «Stratified sampling» assinalada na configuração do nodo «Partitioning» garante a repetibilidade das experiências.
- ☐ O nodo «Partitioning» divide o *dataset* em 2 partes, uma com 30% dos dados e outra com 2 024 registos retirados aleatoriamente do *dataset*.
- ☐ O fluxo da Figura 3 cria um modelo baseado em Redes Neurais Artificiais para o qual usa 70% dos dados para treino e 30% dos dados para teste.