

1 Grupo1

Questão1: Em sistemas de aprendizagem automática (machine learning), o treino de um modelo representa a capacidade de usar os dados para ensinar a máquina a resolver problemas. Neste contexto, explique o que entender por overfitting e como a preparação de dados de treino permite abordar esta característica.

Overfitting é um problema comum em Machine Learning, ocorrendo quando o modelo é treinado para se ajustar excessivamente aos dados de treino, em vez de capturar a estrutura geral dos dados. Isto pode levar a alta accuracy nos dados de treino, mas uma performance má nos dados de teste. Noutras palavras, o modelo memoriza os dados de treino em vez de aprender a generalizar padrões. A preparação dos dados de treino pode ajudar a abordar o overfitting, limitando a complexidade do modelo e fornecendo um conjunto de dados mais representativo. Isto inclui técnicas de divisão do conjunto de dados de treino, validação e teste para avaliar o desempenho do modelo em diferentes conjuntos de dados. Além disso, a normalização dos dados e a remoção de dados duplicados ou irrelevantes podem ajudar a reduzir a variância do modelo. Outra abordagem é a regularização que impõe penalizações na complexidade do modelo para evitar o overfitting, através de técnicas L1 e L2 que adicionam termos de penalização à função de custo de modelo para reduzir os pesos excessivos.

Questão2: Em sistemas de Machine Learning, a criação de modelos pode envolver a necessidade de transformar dados contínuos em discretos, implicando sempre redução de conhecimento. Explique por que razão se admite reduzir o conhecimento em análise, complementando a explicação com 3 exemplos.

A transformação de dados contínuos em discretos é uma técnica comum que envolve a conversão de variáveis numéricas contínuas em categorias discretas. Isto pode ser necessário para que os algoritmos de Machine Learning possam lidar com os dados, que podem ser difíceis ou impossíveis de serem processados na sua forma original. Embora essa transformação possa envolver a redução de informações, ela pode ser útil em casos como:

- Simplificação de modelos: ao agrupar as idades dos pacientes em categorias como "jovem", "adulto" e "idoso", pode-se reduzir a complexidade do modelo, tornando-o mais fácil de entender e interpretar.
- Melhoria da interpretabilidade: ao categorizar o nível de satisfação do cliente em "insatisfeito", "satisfeito" e "muito satisfeito", pode-se tornar mais fácil para uma empresa entender o que precisa fazer para melhorar a satisfação do cliente.
- Limitações de recursos: ao trabalhar com grandes conjuntos de dados, a transformação pode ajudar a reduzir o tamanho dos dados, permitindo

que o modelo seja executado num menor tempo.

Questão3: Em sistemas de machine learning, o treino de um modelo pode envolver a necessidade de discretizar os dados, transformando-os em intervalos(binning). Explique a diferença entre binning de igual altura e igual largura e descreva em que condições se deve optar por uma ou outra estratégia.

O binning é um processo no qual os valores contínuos de uma variável são divididos num conjunto finito de intervalos ou categorias. Existem duas estratégias comuns, estando a diferença entre elas na forma como os intervalos são definidos:

- Binning de igual altura: divide os valores em intervalos de forma a que cada intervalo contenha aproximadamente o mesmo número de observações. Esta técnica pode ser útil quando os dados têm uma distribuição assimétrica ou quando é importante garantir que cada intervalo tenha a mesma representatividade amostral. No entanto, esta estratégia pode resultar em intervalos com larguras diferentes, o que pode levar a variações estatísticas de cada intervalo.
- Binning de igual largura: divide os valores em intervalos de forma a que cada intervalo tenha a mesma largura. Isto pode ser útil quando se quer garantir que todos os intervalos tenham a mesma amplitude ou quando se deseja criar um conjunto de regras de decisão com base nos intervalos. No entanto, esta estratégia pode ser sensível a outliers ou variações nas densidades dos dados, resultando em intervalos com diferentes número de observações.

A escolha entre os dois tipos de binning depende do conjunto de dados e dos objetivos do modelo. Em geral, o binning de igual altura pode ser preferível quando a distribuição dos dados é assimétrica ou quando se quer garantir que cada intervalo tenha aproximadamente o mesmo número de observações. Já o binning de igual largura pode ser preferível quando se deseja garantir que todos os intervalos tenham a mesma amplitude ou quando se quer criar um conjunto de regras de decisão com base nos intervalos. O binning de altura pode ser preferível uma vez que evita o amontoar de valores, impede a dispersão de valores frequentes por diferentes intervalos e cria intervalos separados para outliers.

Questão4: Em qualquer metodologia de análise de dados, a preparação de dados é considerada com bastante ênfase e detalhe. O tratamento de dados nominais é uma das estratégias de preparação de dados. Explique em que consiste e qual a diferença entre as estratégias de label encoding e one-hot encoding na preparação de dados.

- Label Encoding: esta estratégia envolve a atribuição de um número inteiro para cada categoria de uma variável nominal. No entanto, pode

levar a problemas de interpretação já que a atribuição de números pode sugerir uma ordem natural entre as categorias, o que não é necessariamente verdade

- One-hot Encoding: envolve a criação de uma variável binária (com valores 0 e 1) para cada categoria de uma variável nominal. Cada observação é representada por uma combinação de valores binários. Esta estratégia evita problemas de interpretação pois não há ordenação implícita de categorias.

A escolha entre as duas técnicas depende do conjunto de dados e dos objetivos do modelo. Em geral, o label encoding pode ser útil quando as categorias de uma variável nominal têm uma ordem natural. Já o one-hot encoding é geralmente preferido quando não há ordem natural entre as categorias, ou quando as categorias são igualmente importantes para o modelo. Importa ainda referir que o one-hot encoding pode levar a um aumento no número de variáveis, o que pode tornar o modelo mais complexo. Além disso, alguns modelos podem ter dificuldades com a presença de variáveis altamente correlacionadas, o que pode ser um problema com o one-hot encoding.

Questão5: Em qualquer metodologia de análise de dados, a preparação dos dados é preponderante porque os dados recolhidos do mundo real podem apresentar-se incompletos, inconsistentes ou com lixo/ruído. Indique, justificando, que tarefas de preparação de dados se realizam na ocorrência de missing values.

- Ignorar registos: não é aconselhável se a quantidade de missing values for alta
- Ignorar atributos: não é aconselhável se os atributos forem importantes
- Preencher manualmente os dados
- Preencher dados com o mesmo valor: pode criar tendências
- Preencher com valor médio, medianas ou moda: pode ser eficaz em variáveis numéricas e categóricas, mas pode não ser adequada para dados com distribuição não normal ou valores extremos
- Preencher com valor mais frequente
- Preencher com resultados de modelos estatísticos: pode ser mais eficaz em dados complexos com muitas variáveis, mas pode demorar mais e exige habilidades estatísticas avançadas

Questão6: Em qualquer metodologia de análise de dados, a preparação de dados é considerada com bastante ênfase e detalhe. O tratamento de dados numéricos, sejam contínuos ou discretos, é uma estratégia de preparação de dados. Explique em que consiste e qual a diferença entre as estratégias de normalization e standardization na preparação dos dados.

Ambas as técnicas têm como objetivo tornar os dados mais comparáveis e mais fáceis de serem interpretados por algoritmos de machine learning. A normalização é um processo de escala em que os valores dos dados são transformados para uma escala de 0 a 1. Essa transformação é feita através da divisão de cada valor pelo valor máximo encontrado no conjunto de dados. A normalização é especialmente útil quando os valores das variáveis estão em escalas diferentes, evitando que algumas variáveis tenham mais peso na análise do que outras. Por outro lado, a standardization é um processo de escala em que os valores dos dados são transformados para uma distribuição normal com média zero e desvio padrão um. Esta transformação é feita através da subtração da média do conjunto de dados e divisão pelo desvio padrão. A standardização é especialmente útil quando os valores das variáveis têm escalas diferentes e distribuições diferentes, pois transforma todos os valores em unidades equivalentes e torna mais fácil comparar a relação entre elas. A principal diferença é a escala em que os dados são transformados. A normalização transforma os dados numa escala de 0 a 1, enquanto que a standardization transforma os dados numa distribuição normal. A normalização é útil quando os valores estão em escalas diferentes, mas a padronização é útil quando os valores das variáveis têm distribuições diferentes.

Questão7: Em sistemas de aprendizagem automática, a criação de um modelo representa a faculdade da máquina em usar os dados que caracterizam o problem para desenvolver mecanismos para resolver tais problemas. Neste contexto e tendo em consideração a fase de preparação dos dados:

- a) Explique em que consiste a seleção de atributos;
- b) Apresente sinteticamente 3 métodos de seleção de atributos;
- c) Discuta a implicação da aplicação da seleção de atributos no conjunto de dados original.

a) A seleção de atributos é uma etapa em que se identificam os atributos mais relevantes para a construção de um modelo de aprendizagem. O objetivo é reduzir a dimensão do conjunto de dados, eliminando atributos irrelevantes ou redundantes, e melhorar a precisão e desempenho do modelo.

b) Existem diversos métodos de seleção de atributos, de entre eles, filter methods - utilizam técnicas estatísticas para avaliar a correlação entre cada atributo e o target do modelo, sendo a seleção feita com base no valor da medida de relevância, como a correlação de Pearson, em que os atributos com maiores valores de relevância são selecionados -, wrapper methods - usam o modelo de aprendizagem para avaliar a importância dos atributos, através da construção

de um modelo para cada subconjunto de atributos e avaliação do desempenho em cada caso; a seleção é feita com base no desempenho do modelo em cada subconjunto, considerando o menor conjunto de atributos que fornece o melhor desempenho - e *embedded methods* - incorporam a seleção de atributos no próprio processo de treino do modelo; o objetivo é selecionar os atributos que contribuem com mais para a precisão do modelo.

c) A seleção de atributos pode implicar a redução do tamanho do conjunto de dados, tornando-o mais fácil e rápido de ser processado pelo modelo. Isto pode levar a uma melhoria no desempenho do modelo, diminuindo a complexidade e melhorando a capacidade de generalização. No entanto, a seleção de atributos pode também levar à perda de informação útil e diminuir a *accuracy* do modelo. Desta forma, é importante avaliar cuidadosamente os métodos de seleção de atributos para garantir que o conjunto de dados resultante contenha a quantidade adequada de informação para o modelo resolver o problema em questão.

Questão8: Tome em consideração o dataset escolhido pelo grupo para a realização do trabalho prático.

a) Indique e descreva sucintamente a metodologia de análise de dados adotada pelo grupo para enquadrar o desenvolvimento do trabalho.

b) Descreva como o grupo adequou o desenrolar do trabalho prático à metodologia de análise de dados selecionada.

c) Identifique, justificando, 3 técnicas de aprendizagem que tenham sido utilizadas pelo grupo para a criação de modelos.

a) O CRISP-DM é uma metodologia de análise de dados iterativa, o que significa que as suas etapas podem ser revisitadas e ajustadas conforme necessário. As seis etapas principais são:

- Business Understanding: envolve entender os objetivos do projeto, definir as perguntas a serem respondidas e determinar os critérios de sucesso.
- Data Understanding: são reunidos os dados disponíveis e identificadas as falhas que o dataset possa ter.
- Data Preparation: os dados são preparados para serem usados nos modelos de aprendizagem da máquina, incluindo a limpeza, integração, seleção e transformação de dados.
- Modeling: as técnicas de aprendizagem são utilizadas com os dados preparados e testadas para se encontrar qual o modelo que melhor se ajusta aos dados e objetivos do projeto.
- Evaluation: o modelo é avaliado usando os dados de teste para verificar a sua capacidade de generalização. O modelo pode ser ajustado e otimizado nesta etapa para melhorar o seu desempenho.
- Development: o modelo é implementado em ambiente de produção e usado para resolver o problema do negócio.

b) Quando à etapa "Business Understanding" o grupo procurou compreender o problema de negócio, os objetivos e requisitos do projeto. Neste caso, o problema é que os consumidores têm dificuldade em escolher qual smartwatch comprar devido à grande variedade disponível no mercado, e o objetivo é obter insights sobre os smartwatches populares na Índia para ajudar as empresas a melhorar o design, recursos e preços desses dispositivos. Na próxima etapa "Data Understanding", o grupo procurou obter e explorar os dados disponíveis sobre smartwatches, através dos respectivos nodos do KNIME (Statistics, Linear Correlation, Exploration). Quando à "Data Preparation", o grupo preparou os dados para análise, incluindo a limpeza e a seleção de dados relevantes. Na etapa de "Modeling", o grupo selecionou e aplicou técnicas de aprendizagem para explorar os dados e obter insights, em problemas de classificação e regressão, e utilizando, por exemplo, modelos de clustering, linear regression, ANN, etc. Na "Evaluation" o grupo avaliou os modelos desenvolvidos com base métricas de desempenho relevantes, como accuracy e o erro. A fase de "Deployment" por não fazer sentido neste contexto não foi realizada.

c) Foi utilizada a técnica "Decision Tree" para resolver um problema de classificação, onde se produziram grupos que classificavam como cheap ou expensive o preço dos smartphones, sendo depois aplicado ao dataset de teste. Além disso, utilizou-se a técnica "Linear Regression" como um problema de regressão para determinar o Rating dos smartwatches. Já o Clustering foi utilizado para classificar os smartwatches como Low end, Middle end ou High end, no que concerne à sua qualidade.

Questão9: Tome em consideração o dataset escolhido pelo grupo para a realização do trabalho prático.

a) Indique, justificando, 3 razões que sustentaram a escolha do dataset.

b) Descreva sucintamente o problem que os dados descreviam e o objetivo a alcançar.

a) O dataset foi escolhido, principalmente, porque apresenta um range variado de valores, sendo uma oportunidade de aplicar técnicas de manipulação de dados. Além disso, este dataset é sobre um mercado em rápido crescimento e de uma indústria em constante evolução, sendo pertinente para os dias de hoje. Finalmente, o dataset apresentava um número considerável de variáveis.

b) O conjunto de dados de smartwatches fornece uma oportunidade para as empresas da indústria obterem insights sobre os smartwatches populares na Índia. Ao analisar este conjunto de dados, as empresas podem entender os fatores que afetam as vendas de smartwatches e as preferências dos clientes na Índia. Esses insights podem ser usados para melhorar o design, recursos e preços desses dispositivos para atender às necessidades dos clientes.

2 Grupo2

Questão1

Quando se estudam sistemas de aprendizagem, existem três tipos de paradigmas de aprendizagem a saber, aprendizagem supervisionada, aprendizagem não supervisionada e aprendizagem por reforço.

Questão2

Na fase de análise de dados, um valor de correlação próximo de 0 significa que ambas as variáveis em estudo não têm uma correlação forte entre si.

Questão3

Na preparação de dados, utiliza-se discretização para reduzir o número de valores de um atributo contínuo, dividindo-o em intervalos.

Questão4

Hold-Out é um método de validação de modelos de machine learning que se baseia no particionamento dos dados, dividindo-os em dados de treino e de teste.

Questão5

Clustering é um processo de análise de dados através do qual se divide um conjunto de dados em segmentos/clusters de menor dimensão, identificando grupos de dados similares entre si, mas diferentes de outros segmentos/clusters.

Questão6

Uma Rede Neuronal Artificial é um sistema computacional de base conexionista para a resolução de problemas, concebido com base num modelo simplificado do SNC, definido por uma estrutura hierárquica de unidades computacionais designadas neurónios artificiais com capacidade de aprendizagem.

3 Grupo3

Questão1: Quando se estudam sistemas de aprendizagem, distinguem-se 3 técnicas de validação - supervisionada, não supervisionada e por reforço.

Falsa. Quando se estudam sistemas de aprendizagem, distinguem-se 3 paradigmas de aprendizagem - supervisionada, não supervisionada e por reforço.

Questão2: Precision e Recall são técnicas de treino de modelos que visam medir o desempenho dos modelos criados.

Falsa. Precision e Recall são métricas de avaliação do desempenho dos modelos criados, não técnicas de treino.

Questão3: Um objetivo da preparação dos dados é o de transformar os datasets de modo que os atributos que os compõem estejam formatados para os procedimentos de análise que os vão trabalhar.

Verdadeira.

Questão4: Neurónio é a unidade computacional simples de composição de redes neurais artificiais, identificado pela sua posição na rede e caracterizado pelo valor do estado.

Verdadeira.

Questão5: Numa rede neuronal artificial, a sinapse é o ponto de ligação entre um axónio e um neurónio, cujo valor representa a importância do sinal que este recebe, podendo ser estimulador, inibidor ou nulo.

Verdadeira.

Questão6: Uma rede neuronal artificial é um sistema computacional de base conexionista, definido por uma estrutura interligada de neurónios com capacidade de generalização de conceitos.

Verdadeira.

Questão7: Uma Matriz de Confusão é uma métrica de avaliação de desempenho de modelos de aprendizagem por reforço.

Falsa. Uma Matriz de Confusão é uma métrica de avaliação de desempenho de modelos de aprendizagem supervisionada.

Questão8: Uma Matriz de Confusão é uma estratégia de avaliação do desempenho de modelos baseados em Árvores de Decisão.

Verdadeira.

Questão9: Num dataset, os atributos representados por dados categóricos ou ordinais não apresentam qualquer característica diferenciadora entre si. Falsa. Os atributos categóricos e ordinais têm características diferenciadoras entre si e precisam de ser tratados de forma diferente dos atributos numéricos.

Questão10: Uma Árvore de Decisão é uma técnica de aprendizagem automática em que cada ramo representa a seleção de uma alternativa e cada folha representa uma decisão.

Verdadeira.

Questão11: Num sistema de aprendizagem automática, a fase de modelação tem como um dos seus objetivos retirar os atributos que apresentam valores inconsistentes.

Falsa. Na fase de preparação dos dados é que são retirados os atributos que apresentam valores inconsistentes.

Questão12: Nenhuma técnica de segmentação necessita que seja específica a quantidade de segmentos (clusters) a procurar nos dados.

Falsa. Por exemplo, a técnica de particionamento necessita que o valor de k seja definido.

Questão13: Machine Learning é um paradigma de computação em que a característica essencial do sistema se revela pela capacidade de aprender de modo autónomo, a partir de dados.

Verdadeira.

Questão14: Um dos principais objetivos da preparação de dados consiste em transformar os datasets por forma a que os dados neles contidos estejam adequadamente expostos ao procedimento de análise de dados.

Verdadeira.

Questão15: Técnicas de aprendizagem de regressão linear têm como objetivo

prever o valor de um resultado com base nos valores de uma única variável de previsão.

Falsa?. Técnicas de aprendizagem de regressão linear têm como objetivo prever o valor de um resultado com base nos valores de uma única variável de previsão, contudo, existe um tipo de regressão linear - múltipla - em que o resultado pode ser obtido através de múltiplas variáveis de previsão.

Questão16: A matriz de confusão ilustrada apresenta um valor de precisão de 350/410 e de recall de 350/390.

Falsa. O valor do recall é 350/410 e da precisão 350/390.

n=650

		PREVISÃO		
		SIM	NÃO	
ATUAL	SIM	350	60	410
	NÃO	40	200	240
		390	260	

Figure 1: Matriz de Confusão (questão 16)

4 Grupo4

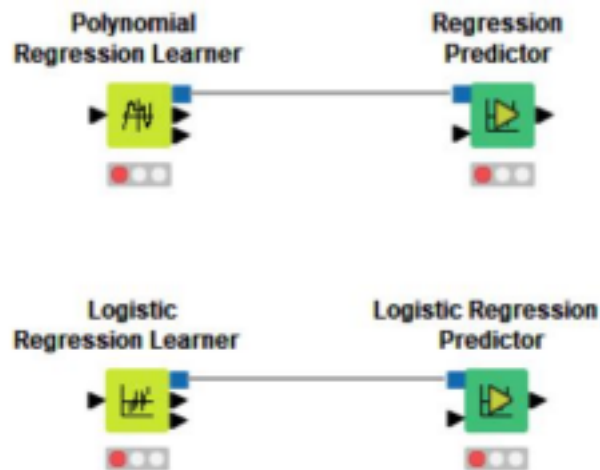


Figure 2: Nodos KNIME (grupo 4)

Questão1: Indique a funcionalidade que implementam os nodos KNIME da figura.

Os nodos "Logistic Regression Learner" e "Logistic Regression Predictor" são utilizados para construir e aplicar modelos de regressão logística, em fluxos de trabalho de análise de dados no KNIME. O "Learner" é responsável por treinar o modelo de regressão logística, ou seja, aprender a relação entre as variáveis de entrada e a variável de saída, com base nos dados de treino. Usa a otimização do gradiente descendente para ajustar os parâmetros do modelo. Já o "Predictor" é responsável por aplicar o modelo de regressão logística treinado a novos dados e fazer previsões de classificação da variável de saída. Usa os coeficientes do modelo treinado para calcular as probabilidades de pertencer a cada classe e retorna a classe com maior probabilidade como a previsão final.

Os nodos "Polynomial Regression Learner" e "Regression Predictor" são utilizados para realizar a análise de regressão polinomial. O "Learner" é usado para treinar um modelo a partir de um conjunto de dados de treino e o "Predictor" aplica o modelo treinado para prever os valores de resposta para um conjunto de dados de teste.

Questão2: Identifique a técnica de aprendizagem que implementam os nodos.

Os nodos "Logistic Regression Learner" e "Logistic Regression Predictor" implementam a técnica de aprendizagem supervisionada de Regressão Logística. Os nodos KNIME Polynomial Regression Learner e Regression Predictor implementam a técnica de aprendizagem de regressão polinomial. Essa técnica é utilizada para modelar a relação entre uma variável dependente e uma ou mais variáveis independentes, por meio de uma equação polinomial de grau n .

Questão3: Descreva as principais diferenças dos nodos.

Os nodos da Logistic Regression e da Polynomial Regression implementam técnicas de aprendizagem diferentes e, portanto, apresentam diferenças significativas na sua funcionalidade e aplicação. A regressão logística é uma técnica de aprendizagem supervisionada que se aplica a problemas de classificação binária, ou seja, aqueles em que os dados de entrada pertencem a uma das duas categorias possíveis. O objetivo do modelo de regressão logística é encontrar a relação entre os atributos de entrada e a probabilidade de pertencer a uma determinada categoria. Os coeficientes encontrados pelo modelo são usados para calcular a probabilidade de cada entrada pertencer à categoria positiva ou negativa. Este tipo de modelo é útil em situações em que é necessário classificar os dados com base em atributos conhecidos. Já a regressão polinomial é uma técnica de aprendizagem supervisionada que se aplica a problemas de regressão, ou seja, aqueles em que o objetivo é prever um valor contínuo de saída a partir de um conjunto de atributos de entrada. A regressão polinomial encontra a relação entre os atributos de entrada e a saída por meio de uma equação polinomial de grau n . O modelo polinomial tenta ajustar uma curva polinomial aos dados de entrada, de forma a fazer previsões mais precisas. Esta técnica é útil quando há uma

relação não-linear entre os dados de entrada e a saída desejada.

5 Grupo4

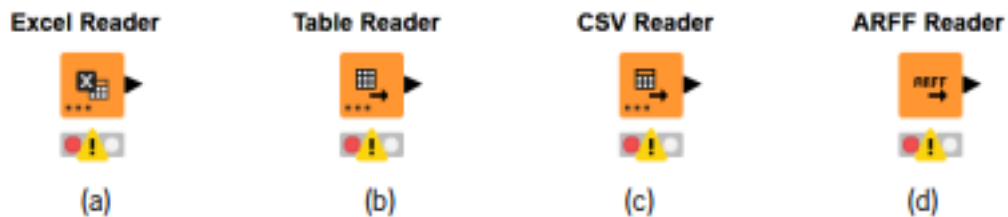


Figure 3: Nodos KNIME (grupo 5.1)

Questão1

O nodo da Figura 1 (a) permite ler dados de 1 ou mais ficheiros de EXCEL. Verdadeira.

O nodo da Figura 1 (a) lê apenas 1 ficheiro de EXCEL. Falsa.

O nodo da Figura 1 (b) lê os dados em formato nativo do Knime. Verdadeira.

O nodo da Figura 1 (c) permite ler dados de 1 ou mais ficheiros EXCEL. Falsa.

O nodo da Figura 1 (c) lê os dados de ficheiros EXCEL. Falsa.

O nodo da Figura 1 (d) lê os dados em formato nativo do Knime. Falsa.

Os nodos das Figura 1 (a), (b) e (c) são nodos de leitura de dados. Verdadeira.

O nodo da Figura 1 (d) não é um nodo de leitura de dados. Falsa.

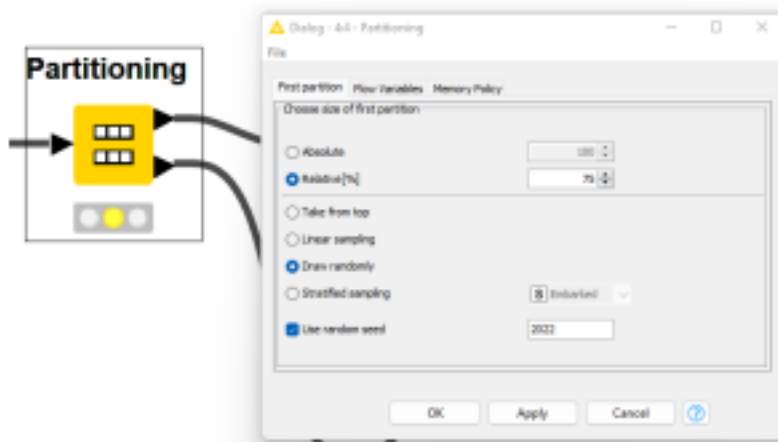


Figure 4: nodos knime (grupo 5.2)

Questão2

O nodo da Figura 2 (Partitioning) divide os dados de entrada em 2 partições de teste. Falsa.

O nodo da Figura 2 (Partitioning) divide os dados de entrada em 2 partições de treino. Falsa.

Na Figura 2 aplica-se uma estratégia de validação designada hold-out validation. Verdadeira.

O nodo da Figura 2 implementa uma estratégia de validação conhecida por leave-one-out validation. Falsa.

Na Figura 2 aplica-se uma estratégia de validação designada cross validation com $k=75$. Falsa.

Na Figura 2 aplica-se uma estratégia de validação designada leave-one-out validation com $k=2022$. Falsa.

Se os dados de entrada não forem balanceados a opção “Linear sampling” deveria estar selecionada. Falsa (Use Stratified Sampling).

As opções selecionadas na configuração do nodo garantem a replicabilidade do modelo criado. Verdadeira.

A configuração do nodo divide os dados de entrada em 2 partições, uma com 75% e outra com 25% dos dados. Verdadeira.

Na configuração do nodo, a opção “Use random seed” não tem significado quando selecionado “Draw randomly”. Falsa.

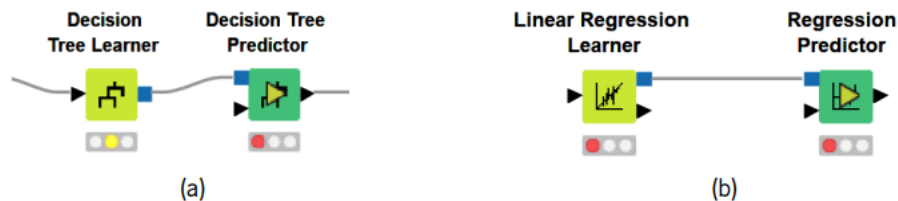


Figure 5: nodos knime (grupo 5.3)

Questão3

Na Figura 3 (a), o nodo da esquerda cria uma Árvore de Decisão. Verdadeira.

Na Figura 3 (a), o nodo da direita cria uma Árvore de Decisão. Falsa.

O nodo à esquerda da Figura 3 (b) cria um modelo de Regressão Linear. Verdadeira.

O nodo à direita da Figura 3 (b) cria um modelo de Regressão Linear. Falsa.

A sequência (a) da Figura 3 permite criar um modelo de previsão contínua. Falsa.

A sequência (b) da Figura 3 permite criar um modelo de previsão contínua. Verdadeira.

A sequência (a) da Figura 3 permite criar um modelo de previsão discreta. Verdadeira.

A sequência (b) da Figura 3 permite criar um modelo de previsão discreta. Falsa.



Figure 6: nodos knime (grupo 5.4)

Questão4

A sequência (a) da Figura 4 permite criar um modelo de previsão contínua. Falsa.

A sequência (b) da Figura 4 permite criar um modelo de previsão contínua. Verdadeira.

A sequência (a) da Figura 4 permite criar um modelo de previsão discreta. Verdadeira.

A sequência (b) da Figura 4 permite criar um modelo de previsão discreta. Falsa.

Ambas as sequências da Figura 4 usam técnicas de regressão pelo que os modelos resultantes são de previsão contínua. Falsa.