

Nome: Henrique Augusto Rodrigues
Matrícula: 675263
Turno: Tarde
Professor(s): Cristiane Neri Nobre

Questão 01 - Knowledge Discovery in Databases (KDD), "processo não trivial de identificação de padrões válidos, novos, potencialmente úteis e, finalmente, compreensíveis em dados". [Fayyad et al. (1996)].

Também considerado como um processo exploratório automático de grande banco de dados, essa segunda definição é dita por [Friedman (1997a)] secundária, pois, "uma análise secundária de dados em um grande banco de dados", o termo secundário enfatiza o fato de que, a primeira proposição de banco de dados não é a análise de dados em si. O passo central do KDD é atribuído a mineração de dados, segundo [Rokach, Lior. **Data mining with decision trees: theory and applications. 2nd edition. 2015**], o processo de KDD foi dividido em nove passos. Esse processo é iterativo em cada passo, significa que, volta em cada passo para ajustar para ajustar o que for necessário.

Começando pela determinação dos objetivos:

1 - Desenvolvendo uma compreensão do domínio do aplicativo: Passo preparatório para entender o que deve ser feito com várias decisões, processo em que, define e entende o(s) objetivo(s). Até mesmo precisando ser revisado para definir os próximos (3) passos, nesse passo é que se toma as decisões sobre algoritmos, representações, etc.

2 - Criar um conjunto de dados no qual a descoberta será realizada: Nesse passo, os dados serão utilizados para a descoberta de conhecimento, se os dados são viáveis, obter dados adicionais e integrar tudo que foi descoberto, incluindo atributos que serão considerados no processo.

Evidenciando a base para a construção do modelo, para se obter o sucesso, é preciso, nesse passo, analisar e considerar tudo o que for possível para que não seja uma falha.

3 - Processando e limpando: Essa etapa serve para a limpeza dos dados, lidar com valores que estão faltando, remover ruído e "outliers". Podendo envolver complexos métodos estatísticos ou o uso específico de algoritmos.

4 - Transformação de dados: Aqui é feita a geração de melhores dados e a mineração de dados é desenvolvida de preparada. Há alguns métodos que podem ser utilizado, um deles é a redução de dimensão, é feita a seleção e extração de características e a gravação da amostragem. O outro método é, a transformação de atributos, é feita a transferência de funções contínuas de atributos numéricos e transformação funcional.

O processo reflete leva ao entendimento da transformação necessária.

5 - Escolhendo a tarefa apropriada de mineração de dados: Entre classificação, regressão ou agrupamento. Devemos escolher qual ou quais será o melhor para o modelo e objetivo declarados nos passos anteriores.

Os dois principais são a descrição e a predição, o primeiro inclui a mineração não supervisionada e a visualização dos aspectos, Já o segundo, é a mineração de dados supervisionada.

A maioria das técnicas são baseadas no aprendizado indutivo, onde, o modelo é explicitamente ou implicitamente pela generalização por um número suficiente de exemplos treinados.

6 - Escolhendo o algoritmo para a mineração de dados: Seleciona-se a especificação do modelo a ser usado para, procurar padrões. Exemplo, para precisão, as redes neurais são melhores, já para o entendimento, a árvore de decisão é a melhor escolha. O foco do metalearning (**metalúrgico é a tradução literal, mas, qual o melhor termo nesse caso?**) É analisar se, o algoritmo foi ou não sucedido em um problema em específico.

7 - Empregando algoritmo de mineração de dados: Utilizar o algoritmo até possuir um resultado satisfatório, controlar os parâmetros com o número mínimo de instâncias em, uma única folha, ou, árvore de decisão.

8 - Avaliação: Avaliar e interpretar os padrões extraídos, respeitando o que foi decidido no primeiro passo.

Foca na compreensão e usabilidade do modelo escolhido.

9 - Utilizando o conhecimento adquirido: Aqui, determina-se o sucesso da efetividade do processo. Um dos riscos, é a perda do laboratório em que, tudo foi construído. Os dados já não

são mais estáticos e sim dinâmicos, é incorporado o conhecimento em outro sistema para ações futuras.

O conhecimento se torna ativo e sentido de que, podemos fazer mudanças no sistema e medir os efeitos.

Questão 02 - São quatro os tipos de problemas, são eles: classificação, regressão, agrupamento (clustering) e regras de associação.

Classificação: Utilizado para prever ou descrever uma classe, o seu atributo é nominal, exemplo, se quero saber se irei assistir 'Star Wars' ou não, vai depender se será as trilogias e qual delas, se será na sequência, o local em que irei assistir.

Regressão: É semelhante a **Classificação**, a diferença é que, o seu atributo é numérico.

Exemplo: O preço da assinatura do serviço streaming para assistir a franquia 'Star Wars'.

Agrupamento (clustering): Agrupa-se as instâncias de acordo com os atributos de entrada.

Exemplo: Identificar o perfil de usuário que assistem "Star Wars".

Regras de Associação: Usada para, buscar semelhança e/ou associações entre os elementos.

Exemplo: Quem assistiu a trilogia original de 'Star Wars', também assistiu 'The Mandalorian'.

Questão 03 - Os principais métodos de aprendizagem de máquina são: supervisionado, não supervisionado, semi supervisionado, reforço e deep Learning.

Aprendizado Supervisionado: É feito através de indução, tem-se uma entrada e um rótulo desejado. Exemplo, caso queira o entendimento de algo, é melhor utilizar a árvore de decisão.

Aprendizado não supervisionado: Recebe-se apenas atributos de entrada, e encontra agrupamentos. Exemplo, redes neurais, boa escolha para a procura de padrões, o perfil de quem assistiu a trilogia original de 'Star Wars' e também assistiu ao seriado 'The Mandalorian'.

Aprendizado semi supervisionado: São utilizados dados não rotulados e rotulados para, aprender melhor o modelo. Exemplo, utilizar a geometria para separar quem assistiu a primeira trilogia e a melhor de 'Star Wars' de quem assistiu as demais.

Aprendizado por reforço: Quanto maior a recompensa, teremos uma maior maximização dos resultados. Exemplo, rede neural, encontrar quem assiste 'Star Wars' e assina o serviço de streaming, agrupar e melhorar a forma de atrair mais fãs da franquia.

Deep Learning: É uma rede neural multi-camadas, utilizada para reconhecimento com ordem de complexidade muito grande. Exemplo, assistentes pessoais.

Questão 04 -

Escolha de testes de atributo

Alternativo	0	Já foi	
Bar	0	Já foi	
Sex/Sab	0,081704165945510	Já foi	
Fome	0,047660763468214	Já foi	
Cliente	0,540852082972755	Correto	
Preço	-0,0833333333333333	ERRADO	
Chuva	0,034043402477296	Já foi	
Res	0,081704165945510	Já foi	
Tipo	0	Correto	
Tempo	0,069172916546474	Já foi, mas com dúvida	
Conc	0,081704165945510	Já foi	

1) Maior ganho, coloca na raiz. O atributo cliente é a raíz.

Cálculos:

Alternativo: $1 - (6 \div 12 \times (-1 \div 2 \times \log(1 \div 2; 2) - 1 \div 2 \times \log(1 \div 2; 2))) + 6 \div 12 \times (-1 \div 2 \times \log(1 \div 2; 2) - 1 \div 2 \times \log(1 \div 2; 2))$

Bar: $1 - (6 \div 12 \times (-1 \div 2 \times \log(1 \div 2; 2) - 1 \div 2 \times \log(1 \div 2; 2))) + 6 \div 12 \times (-1 \div 2 \times \log(1 \div 2; 2) - 1 \div 2 \times \log(1 \div 2; 2))$

Sex/Sab: $1 - (7 \div 12 \times (-2 \div 6 \times \log(2 \div 6; 2) - 4 \div 6 \times \log(4 \div 6; 2))) + 5 \div 12 \times (-4 \div 6 \times \log(4 \div 6; 2) - 2 \div 6 \times \log(2 \div 6; 2))$

Fome: $1 - (7 \div 12 \times (-4 \div 6 \times \log(4 \div 6; 2) - 2 \div 6 \times \log(2 \div 6; 2))) + 5 \div 12 \times (-3 \div 6 \times \log(3 \div 6; 2) - 3 \div 6 \times \log(3 \div 6; 2))$

Cliente: $1 - (6 \div 12 \times (-2 \div 6 \times \log(2 \div 6; 2) - 4 \div 6 \times \log(4 \div 6; 2)))$

Preço: $1 - (7 \div 12 \times (-3 \div 6 \times \log(3 \div 6; 2) - 1 \div 2 \times \log(1 \div 2; 2))) + 2 \div 12 \times (-1 \div 2 \times \log(1 \div 2; 2) - 1 \div 2 \times \log(1 \div 2; 2)) + (4 \div 12 \times (-1 \div 2 \times \log(1 \div 2; 2) - 1 \div 2 \times \log(1 \div 2; 2)))$

Chuva: $1 - (5 \div 12 \times (-2 \div 6 \times \log(2 \div 6; 2) - 4 \div 6 \times \log(4 \div 6; 2))) + 7 \div 12 \times (-3 \div 6 \times \log(3 \div 6; 2) - 3 \div 6 \times \log(3 \div 6; 2))$

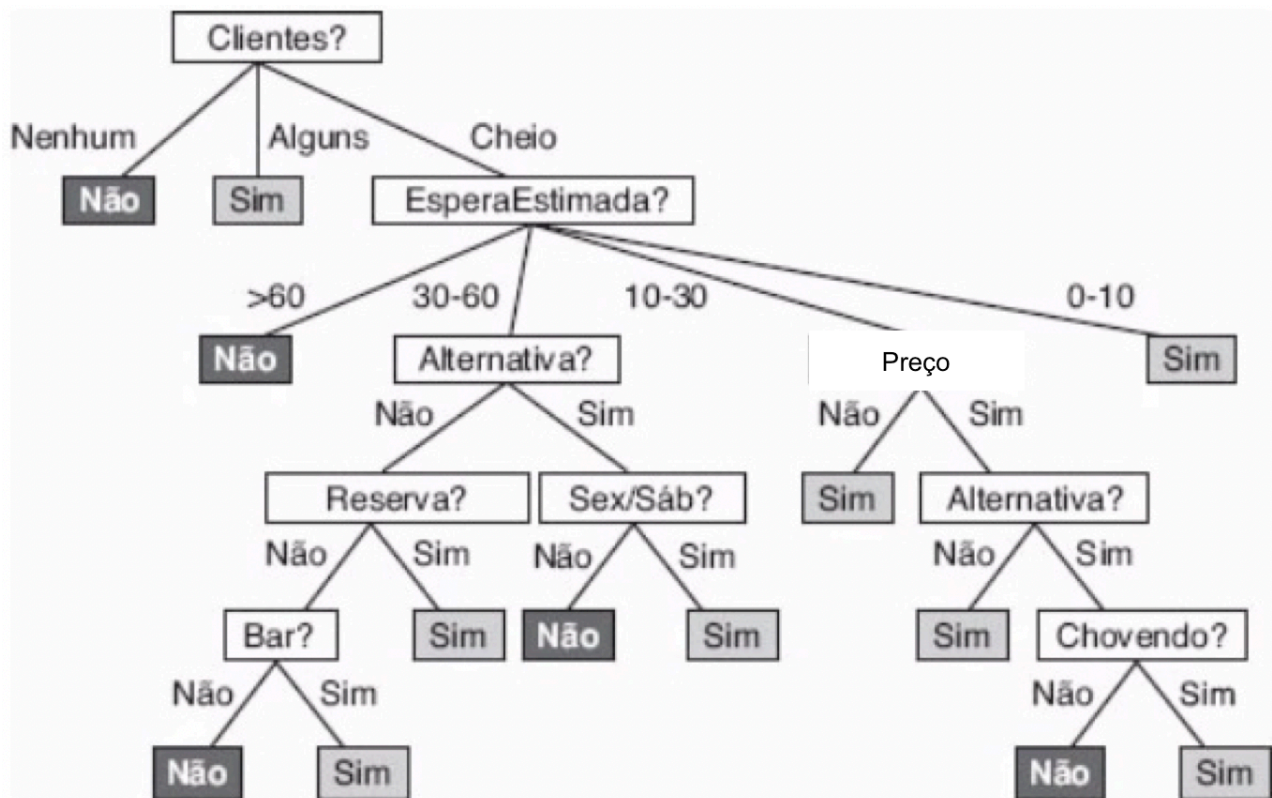
Resp: $1 - (5 \div 12 \times (-2 \div 6 \times \log(2 \div 6; 2) - 4 \div 6 \times \log(4 \div 6; 2))) + 7 \div 12 \times (-2 \div 6 \times \log(2 \div 6; 2) - 4 \div 6 \times \log(4 \div 6; 2))$

Tipo: $1 - (2 \div 12 \times (-1 \div 2 \times \log(1 \div 2; 2) - 1 \div 2 \times \log(1 \div 2; 2))) + 2 \div 12 \times (-1 \div 2 \times \log(1 \div 2; 2) - 1 \div 2 \times \log(1 \div 2; 2)) + (4 \div 12 \times (-1 \div 2 \times \log(1 \div 2; 2) - 1 \div 2 \times \log(1 \div 2; 2))) + 4 \div 12 \times (-1 \div 2 \times \log(1 \div 2; 2) - 1 \div 2 \times \log(1 \div 2; 2))$

Tempo: $1 - (6 \div 12 \times (-3 \div 6 \times \log(3 \div 6; 2) - 3 \div 6 \times \log(3 \div 6; 2))) + 2 \div 12 \times (-1 \div 6 \times \log(1 \div 6; 2) - 1 \div 6 \times \log(1 \div 6; 2)) + (2 \div 12 \times (-1 \div 6 \times \log(1 \div 6; 2) - 1 \div 6 \times \log(1 \div 6; 2))) + 2 \div 12 \times (-1 \div 6 \times \log(1 \div 6; 2) - 1 \div 6 \times \log(1 \div 6; 2))$

Conc: $1 - (6 \div 12 \times (-4 \div 6 \times \log(4 \div 6; 2) - 2 \div 6 \times \log(2 \div 6; 2))) + 6 \div 12 \times (-2 \div 6 \times \log(2 \div 6; 2) - 4 \div 6 \times \log(4 \div 6; 2))$

2)



Questão 05 - Em falta