

Question:	1	2	3	4	Total
Points:	30	25	20	25	100
Score:					

Ler com atenção:

- **Importante:** A fraude denota uma grave falta de ética e constitui um comportamento não admissível num estudante do ensino superior e futuro profissional mestre. Qualquer tentativa de fraude leva à reprovação na disciplina, tanto do facilitador como do prevaricador.
- Este teste é individual e com consulta. Não pode usar dispositivos electrónicos, incluindo o telemóvel. Só são admitidos apontamento em papel. Não pode trocar apontamentos com colegas.
- Responda às perguntas nos espaços indicados para o efeito. Se precisar de papel adicional, contacte o docente vigilante da sua sala.
- Nas **perguntas Verdadeiro/Falso com pedido de justificação**, respostas sem justificação são cotadas a zero.
- Nas **perguntas Verdadeiro/Falso sem pedido de justificação**, respostas incorretas implicam uma penalização de 30% da cotação; se não responder, tem 0 na pergunta.

1. (30%) De forma a ajudar a gerir criteriosamente a distribuição de vacinas pelas pessoas dos diversos países no mundo, suponha que a OMS quer desenvolver um sistema de apoio à decisão da tomada de vacina para a covid-19, de forma a que os que têm maior probabilidade de ter um estado severo de covid-19 e que tenham menos probabilidade de ter efeitos secundários graves de uma vacina sejam os que recebem uma vacina, independentemente do país a que pertençam. Para tal, foram recolhidos dados pessoais da população mundial que tomaram a vacina e já tiveram covid-19 (ver extrato do data set na Tabela 1 e Tabela 2; assuma que o domínio de cada variável é o conjunto de valores que constam nas suas colunas).

Id	Com	VacCov	VS	EVS	GV
34	N	N	0	2	0
22	N	S	2	0	0
65	S	S	1	1	2
7	S	N	0	2	1

Tabela 1: Legenda: Id - Idade; Com - Comorbilidades; VacCov - Vacina ou Covid-19; VS - Vida Social (frequência de discotecas, bares, etc.); EVS - Estilo de vida saudável (prática desporto, alimentação saudável, etc.); GV - Nível de gravidade de covid-19; 2-severo, 1-ligeiro, 0-nenhum; N- Não; S - Sim

Id	S	Vac	GV
38	M	Pfizer	0
45	M	Moderna	1
40	F	Janssen	2
47	M	Pfizer	1

Tabela 2: Legenda: Id - Idade; S - Sexo; Vac - Tipo de vacina; GV - Gravidade dos Efeitos Secundários da Vacina; M - Masculino; F- Feminino; 2-severo, 1-ligeiro, 0-nenhum

- (a) Diga que classe de algoritmos de Aprendizagem Computacional escolheria no desenvolvimento do sistema inteligente da OMS, se pretendesse:
- accuracy* nas predições?
 - explicabilidade nas predições?
 - equilíbrio entre explicabilidade e *accuracy*?

Solution: Exemplo de resposta:

Neural Nets

Linear regression/Logistic regression

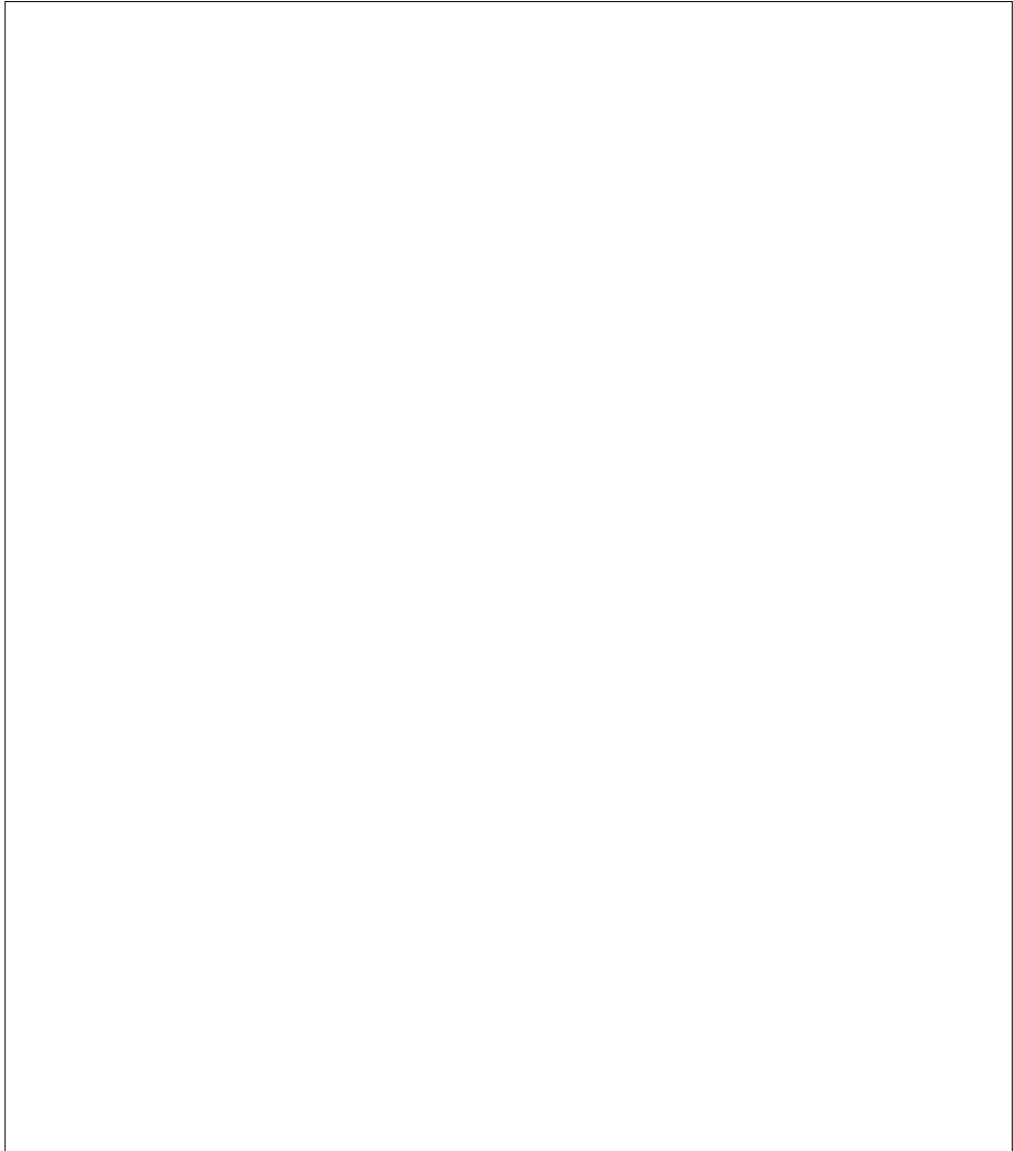
Bayes Nets/Decision Trees

- (b) Faça um esboço do grafo (sem CPTs) da(s) Rede(s) Bayesiana(s) que, para os dados pessoais de uma determinada pessoa, permita(m) obter informação probabilística sobre os possíveis cenários de covid-19 (2-severo, 1-ligeiro, 0-assintomático) e sobre os possíveis efeitos secundários de uma vacina (2-severo, 1-ligeiro, 0-nenhum).

Resposta:

Solution: Exemplo de resposta:

Um modelo Naive Bayes, tanto para os possíveis cenários de covid-19, como para os possíveis efeitos secundários de uma vacina. As variáveis (nodos) antecedentes são as representados nas respetivas tabelas.



- (c) Suponha que a OMS vai usar algoritmos de Deep Learning e pretende adicionar-lhes explicabilidade. Que soluções sugere, porquê, e como funcionariam?

Resposta:

Solution: Exemplo de resposta:

Usar surrogate models como o LIME. Produzir um data set em “torno” da instância/caso a explicar, sendo o valor da target variable dado pelo black-box model, treinar o LIME, usando um modelo explicável, como Decision Trees, com este data set, e obter os resultados e respetiva explicação usando este modelo.

- (d) Se tivesse de optar entre *Human in the Loop* e *Machine in the Loop*, diga, justificando, qual destes tipos de Inteligência Artificial colaborativa consideraria para este sistema desenvolvido pela OMS?

Solution: Exemplo de resposta:

Qualquer destes tipos de IA seriam considerados válidos desde que bem justificados. A diferença está em dar mais foco ou menos foco à participação humana. Poder-se-ia considerar o modelo em que o humano toma as decisões e a máquina ajuda a confirmar ou a alertar para outras decisões (Machine in the Loop), ou então o modelo contrário (Human in the Loop), em que a decisão se centra na máquina, aparecendo o humano como adjuvante, por vezes a retificar a decisão da máquina, por vezes a ensiná-la.

Resposta:

2. (25%) Uma equipa de cientistas de dados foi contratada para monitorizar as opiniões dos eleitores acerca de um partido político, com base na informação publicada em redes sociais. O processo começou por criar um conjunto de dados a usar como referência, constituído por 1000 publicações e algumas anotações relacionadas com a opinião transmitida. Duas dessas publicações encontram-se na tabela 3.

ID	Texto	Autor	Data
1	<i>não que não apoie várias medidas propostas, mas podia ter estado melhor no último debate!</i>	@apirata11	08-01-2022
2	<i>se continuam a defender cegamente esta medida, não vão contar com o meu voto...</i>	@janaoe1984	08-01-2022

Tabela 3: Exemplo de publicações a analisar.

- (a) Indique duas sub-tarefas da área de *Sentiment Analysis* cuja aplicação faça sentido neste cenário. Explique qual poderia ser o seu resultado para as publicações na tabela 3 e porquê.

Solution: Escolher duas das seguintes:

- Análise de subjetividade: ambas as publicações têm conteúdo subjetivo;
- Análise de polaridade: na primeira é mais difícil escolher apenas uma polaridade, enquanto que a segunda tende mais para o negativo.
- *Aspect-based Sentiment Analysis*: na primeira publicação, há um sentimento positivo relativo às medidas, e negativo relativamente ao debate; na segunda o sentimento relativamente à medida é negativo;
- Extração de quintuplos de opinião: para a primeira publicação (partido, medidas, positivo, @apirata11, 08-01-2022), (partido, debate, negativo, @apirata11, 08-01-2022) e para a segunda (partido, medida, negativo, @janaoe1984, 08-01-2022);
- Seria ainda possível realizar o reconhecimento de emoções: o resultado dependeria da emoções consideradas, por exemplo, tristeza para a primeira publicação e tristeza ou raiva para a segunda.

- (b) Indique um sub-problema de *Data Mining* adequado às tarefas referidas na pergunta anterior e justifique porque se adequa.

Solution: A análise de subjetividade, de polaridade e de emoções podem ser abordadas como um problema de classificação. Se, para algum dos anteriores, se pretender atribuir um valor numa escala contínua, o problema será de regressão.

- (c) A anotação do conjunto de referência foi realizada com recurso a *crowdsourcing*. Especificamente, uma das propriedades anotadas baseou-se na opinião independente de sete trabalhadores diferentes. Suponha que, para a publicação #1 da tabela, as respostas dos sete trabalhadores foram: $[1, 1, 0, 0, -1, 1, -1]$. Suponha ainda que, numa fase de análise de respostas, se verifica que o trabalhador #7 atribuiu sempre mesmo valor a todas as publicações (-1).

Indique um método de agregação de resultados adequado a este cenário, explique em que consiste e aplique-o ao cálculo do valor resultante para a publicação #1.

Solution: *Majority vote* seria uma boa opção, mas como o trabalhador #7 não é fiável, deve ser ignorado, e o método que o faz chama-se *honey pot*. Independentemente disso ser feito ou não, o valor resultante, por ser o mais vezes atribuído, é 1. Se se considerar que a escala contínua, também seria possível usar a média ou a mediana.

- (d) Considere que o texto das publicações é representado num modelo vetorial tradicional e que é usada a seguinte lista de *stopwords*: *que, de, o, a, os, as, em*. Quantas células do vetor usado para representar a publicação #1 da tabela serão diferentes de 0? Explique o seu raciocínio, se necessário, recorrendo a cálculos.

Solution: Após remover as *stopwords*, há 13 *tokens* diferentes na publicação, ou seja, o vetor terá 13 células diferentes de 0.

3. (20%) Considere a matriz de utilidade apresentada em baixo com *ratings* de seis *smartphones* (S1,...,S3) por quatro utilizadores (A, B, C, and D). Os ratings têm domínio [0,5]. Se utilizar um *User Based, Nearest Neighbor Collaborative Filtering Recommender System*, qual a predição para o *rating* do smartphone S3 pelo utilizador A, usando a função de predição $pred(a, p)$ descrita em baixo e considerando um conjunto de vizinhança de tamanho 2, calculado com recurso à Manhattan Distance, descrita também em baixo?

$$d(x, y) = \sum_{p \in P} \|r_{x,p} - r_{y,p}\| \quad (1)$$

$$pred(a, p) = \bar{r}_a + \frac{\sum_{b \in N} sim(a, b) * (r_{b,p} - \bar{r}_b)}{\sum_{b \in N} sim(a, b)}$$

	S1	S2	S3
A	2	3	?
B	3	2	4
C	1	2	3
D	4	5	5

4. (25%) Uma empresa que atua na área da cibersegurança desenvolveu um software de nome *Looppa* que, recorrendo a técnicas de IA, analisa todas as mensagens de texto recebidas por uma organização com o intuito de identificar intrusões maliciosas na sua rede informática.

Em particular, o sistema *Looppa* identifica mensagens contendo *links* que, se acionados, possam levar ao *phishing* de credenciais sensíveis, ou à instalação de software malicioso em qualquer computador da rede. Realiza, com esse fim, uma análise semântica de cada mensagem, interpretando tanto o seu texto como as imagens nelas contidas. O sistema consegue identificar de forma fiável, por exemplo, mensagens que pretendam parecer terem sido enviadas por entidades bancárias.

Suponha que a empresa de cibersegurança pretende que o *Looppa* siga as Orientações Éticas para uma IA de Confiança (*Ethics Guidelines for Trustworthy AI*) definidas pelo *High-Level Expert Group on Artificial Intelligence* criado por iniciativa da Comissão Europeia.

- (a) Dê exemplos de atos que o *Looppa* **não deva realizar**, ou de decisões que **não deva tomar**, para que não ponha em causa a sua conformidade com as três componentes da IA de Confiança (*Trustworthy AI*) (um exemplo para cada componente).

Solution:

De acordo com as orientações referidas, o sistema deve ser:

legal: não deverá, por conseguinte, guardar informação pessoal dos utilizadores sem a sua autorização expressa;

ético: não deverá, ainda, identificar mensagens com base em critérios de género, idade, ou raça;

robusto: não deverá, tão-pouco, filtrar informação cujo conteúdo não esteja garantidamente estabelecido como malicioso, sob pena de causar dano à organização.

Resposta:

- (b) Classifique as seguintes afirmações como Verdadeiras ou Falsas (considerando as Orientações Éticas atrás referidas):
- (i) O respeito pela autonomia humana é um imperativo ético para este sistema.
 - (ii) O respeito pelo Regulamento de Proteção de Dados em vigor é um imperativo ético para este sistema.

Verdadeira/Falsa: _____

Porquê?

Verdadeira/Falsa: _____

Porquê?

Solution:

- (i) Verdadeira. De acordo as orientações referidas, o respeito pela autonomia humana é um dos princípios éticos, assentes nos direitos fundamentais, que devem ser especificados como imperativos éticos que os profissionais no domínio da IA devem esforçar-se sempre por respeitar.
- (ii) Falsa. O respeito pelo Regulamento de Proteção de Dados é um imperativo Legal. As orientações referidas distinguem os imperativos legais dos imperativos éticos.