

Question:	1	2	3	4	Total
Points:	35	15	25	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.

Pergunta 1 (35 %)

Considere um modelo de regressão linear traduzido pela seguinte equação:

$$g(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (1)$$

- (a) Mencione uma característica deste modelo regressivo linear que faz com que seja considerado dos mais interpretáveis. Justifique a sua resposta.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Nome: _____ Número: _____

(b) Indique, justificando, uma desvantagem dos modelos regressivos lineares.

[illegible]

-
- This image shows a full page of white paper with horizontal dotted lines. The lines are evenly spaced and run across the width of the page, providing a guide for handwriting practice. There are no margins, text, or other markings on the page.

Nome: _____ Número: _____

- (d) Que métodos de visualização são usados para a interpretabilidade de um modelo regressivo linear e de que forma contribuem para a sua interpretabilidade?

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Pergunta 2 (15 %)

Classifique as seguintes frases como Verdadeiro ou Falso (nota importante, para cada frase: se der a resposta correcta, recebe 100%; se não der uma resposta, recebe 0%; se der uma resposta incorrecta, recebe uma penalização de 30% do valor da pergunta):

(a) Active Learning (Aprendizagem Ativa) é uma técnica de Inteligência Artificial que pertence à categoria de Supervised Learning (Aprendizagem Supervisionada).

(a) _____

(b) Em Active Learning, na seleção de uma instância para que o oráculo a anote, as estratégias de *Smallest Margin*, *Least Confident* e *Maximum Entropy* podem conduzir à seleção de instâncias diferentes.

(b) _____

(c) Collaborative-based Recommender Systems nunca sofrem do *cold start problem*.

(c) _____

(d) O LIME é um método global de Inteligência Artificial Explicável.

(d) _____

(e) Saber qual a função de utilidade usada por um agente artificial não contribui para a interpretabilidade do seu comportamento.

(e) _____

Pergunta 3 (25 %)

De forma a melhorar o atendimento num *call-center*, uma equipa especializada foi contratada para analisar transcrições de conversas realizadas. Entre outras tarefas, foi-lhes pedido para fazer um estudo ao nível de diferentes tipos de conversa e de interações dos clientes.

- (a) Indique como o processo de Data Mining poderia ser instanciado para este cenário, indicando explicitamente em que consistiria cada um dos passos.

Resposta:

- Data Collection: recolha de um conjunto de (transcrições de) conversas anteriores (e anotação de um sub-conjunto que poderia servir para dados de treino, validação e teste);
- Data Preprocessing: limpeza das conversas e representação num formato textual simples; identificação das features a extrair (e.g., palavras / n-gramas usados, entidades, dependências, sentimento, embedding) e sua representação (e.g., num vetor numérico);
- Analytical Processing: classificação automática das conversas ou das interações, supervisionada pelos dados anotados; ou clustering de conversas ou interações com base nas suas representações.

- (b) Suponha que a análise também inclui o desenvolvimento de um *chatbot* baseado em *Information Retrieval*, com uma função de similaridade dada pelo coseno de vetores (ver equação 2), e interações representadas através de um modelo vetorial tradicional, onde o único pré-processamento realizado é a remoção de sinais de pontuação.

$$\cos(\vec{x}, \vec{y}) = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\| \times \|\vec{y}\|} = \frac{\sum_{i=0}^n x_i \times y_i}{\sqrt{\sum_{i=0}^n x_i^2} \times \sqrt{\sum_{i=0}^n y_i^2}} \quad (2)$$

Numa primeira experiência, foi criado apenas o índice apresentado na tabela 1.

ID	Pergunta	Resposta
1	o que é o cartão de pessoa coletiva?	é o novo documento que contém o número de identificação de pessoa coletiva (NIPC).
2	como recebo o cartão de pessoa coletiva?	após a inscrição, recebe um e-mail com o código de acesso.
3	qual a validade do cartão de pessoa coletiva?	a validade depende da validade dos elementos nele constantes.
4	como se escolhe a firma?	a partir da bolsa de nomes ou outro nome aprovado através de um certificado de admissibilidade válido.
5	desejo falar com um assistente	a sua chamada será transferida para um assistente.

Tabela 1: Índice para a versão inicial do *chatbot*.

Com base nesse índice, indique a resposta que o *chatbot* selecionaria para a interação: “falar para uma pessoa”, justificando com os cálculos necessários.

Resposta: Será dada a resposta definida para a pergunta com ID 5, porque é aquela que tem maior similaridade com a interação.

$$\text{sim}(q, d1) = \frac{1}{\sqrt{2}\sqrt{10}} = \frac{\sqrt{20}}{20}$$

$$\text{sim}(q, d2) = \frac{1}{\sqrt{2}\sqrt{7}} = \frac{\sqrt{14}}{14}$$

$$\text{sim}(q, d3) = \frac{1}{\sqrt{2}\sqrt{8}} = \frac{1}{4}$$

$$\text{sim}(q, d4) = 0$$

$$\text{sim}(q, d5) = \frac{1}{\sqrt{2}\sqrt{5}} = \frac{\sqrt{10}}{10}$$

- (c) Se mais tarde se pretender pesar os termos de acordo com a sua relevância, mantendo o pré-processamento, explique como poderia calcular o TF-IDF do termo ‘o’ na primeira pergunta do índice, e qual o seu valor.

Resposta: Com base em frequências absolutas e ignorando logaritmos,

$$TF('o', d1) = 2$$

$$IDF('o') = \frac{1}{2}$$

$$TF - IDF('o', d1) = 1$$

Nome: _____ Número: _____

- (d) Para resolver este problema, poderia ainda optar-se por uma representação baseada em *sentence embeddings*, por exemplo, com recurso a um modelo de linguagem neuronal como o BERT. Indique duas vantagens desta opção relativamente às representações vetoriais tradicionais, e duas desvantagens ou questões éticas associadas.

Resposta: Vantagens: estes modelos são mais escaláveis porque representam qualquer sequência de texto com um vetor de tamanho fixo, denso, sem que seja necessário uma célula por cada palavra do vocabulário; estes vetores não consideram apenas as palavras utilizadas, mas têm por base uma representação do seu significado, aprendido com base nas co-ocorrências destas palavras em grandes coleções de texto.

Desvantagens: os vetores utilizados não são interpretáveis, ou seja, não possível explicar a relação de cada feature com o texto representado; isto para além de várias questões éticas associadas ao treino destes modelos, que vão desde a sua tendência para vieses (e.g., género, raça, religião), à sua pegada ecológica.

Pergunta 4 (25 %)

Uma empresa que atua na área médica está a desenvolver um equipamento de monitorização cardíaca destinado a ser facilmente transportável por um humano no seu dia-a-dia. O equipamento consiste num conjunto de sensores incorporados num colete que o utilizador deverá trazer vestido. O colete com os sensores está ligado por fios finos e flexíveis a um aparelho que recolhe os sinais e comunica sem fios com o telemóvel pessoal do utilizador. O aparelho é pequeno e pode transportar-se, por exemplo, num cinto apropriado. O conjunto colete+aparelho são pouco intrusivos, permitindo um uso razoavelmente confortável.

O aparelho referido tem duas funções: (i) recolhe e armazena parâmetros relevantes dos sinais cardíacos recolhidos pelos sensores, (ii) envia regularmente esses parâmetros, através do telemóvel, para um sistema central. O sistema central, por sua vez, alerta o utilizador, através de uma App do telemóvel, para qualquer situação anómala que detete com base nos parâmetros enviados.

O sistema central dispõe de modelos treinados por sistemas de *machine learning* a partir de bases de dados massivas de sinais cardíacos. Estes modelos são capazes de prever episódios cardíacos perigosos com uma antecedência razoável, pelo que os alertas enviados para o aparelho portátil podem salvar a vida do utilizador.

Suponha que a empresa pretende que este sistema 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.

Classifique as seguintes frases como Verdadeiras ou Falsas (nota importante, para cada frase: se der a resposta correcta, recebe 100%; se não der uma resposta, recebe 0%; se der uma resposta incorrecta, recebe uma penalização de 30% do valor da pergunta):

- (a) Se o sistema, por algum motivo, enviar alertas falsos para o utilizador, avisando-o de uma situação anómala sem ter evidências suficientes de que tal esteja a acontecer, ou se não enviar alertas em situações garantidamente anómalas, estará a desrespeitar a componente ética das orientações atrás referidas.

(a) **Falsa**

- (b) Se o sistema guardar informação pessoal dos utilizadores sem o seu consentimento expresso, estará a desrespeitar a componente de solidez das orientações atrás referidas.

(b) **Falsa**

- (c) Se o sistema guardar informação pessoal dos utilizadores sem o seu consentimento expresso, estará a desrespeitar a componente legal das orientações da IA de confiança.

(c) **Verdadeira**

- (d) Suponha que o sistema envia uma notificação ao utilizador informando-o de que deve dirigir-se a um serviço médico para realizar um exame num determinado prazo e o utilizador não cumpre; suponha que o sistema, nessa situação, estimando que o doente está em risco de vida, começa a enviar alertas sucessivos ao utilizador, bloqueando-lhe serviços do telemóvel, para o forçar a dirigir-se a um serviço médico. Neste caso, estaremos perante um exemplo de uma situação em que existe um conflito ético entre o princípio do *respeito pela autonomia humana* e o princípio da *explicabilidade*.

(d) **Falsa**

Nome: _____ Número: _____

- (e) Considere a situação descrita na alínea anterior. Suponha que o sistema consegue detectar a localização de telemóveis de outras pessoas na proximidade física do utilizador e envia mensagens para esses telemóveis solicitando que essas pessoas insistam junto do utilizador para a necessidade de este se dirigir a um serviço médico. Nesta situação, o sistema estará a pôr em causa um dos requisito essenciais para uma IA de confiança.

(e) **Verdadeira**

- (f) O respeito pelo princípio da equidade é um imperativo legal para este sistema.

(f) **Falsa**