-- CONHECIMENTOS ESPECÍFICOS --

Com relação a ciência de dados e inteligência artificial, julgue os próximos itens.

51 Os modelos de aprendizagem das redes neurais incluem o aprendizado por memória e o aprendizado competitivo.

52 Na inteligência artificial, a representação completa de determinado assunto constitui uma ontologia.

53 Os processos heurísticos buscam, primeiramente, estabelecer soluções teóricas para, depois, avançar para uma única tentativa de solução.

54 O dispositivo computacional da inteligência artificial engloba, entre outros elementos, a percepção, que é a capacidade de provocar mudanças no ambiente.

55 O algoritmo de classificação KNN (K Nearest Neighbors) é utilizado na inteligência artificial para reconhecimento de padrões.

56 Na ciência de dados, as cadeias de Markov podem ser aplicadas a ferramentas de previsão do clima e de espalhamento de doenças, por exemplo.

57 Em uma rede regida pelo aprendizado Hebbiano, apenas um neurônio de saída é disparado, enquanto em uma rede regida pelo aprendizado competitivo, os neurônios podem ser excitados simultaneamente.

58 A utilização da lógica nebulosa é adequada quando há necessidade de uma variável fazer parte da solução de um problema.

59 A capacidade de classificação é uma tarefa importante no aprendizado de máquina, por ser utilizada no estabelecimento de padrões.

60 Na inteligência artificial, o agente inteligente deve ser capaz de ter autonomia, isto é, deve ter a capacidade de acumular conhecimento que seja útil em suas ações.

Considerando processos de análise e mineração de dados, julgue os itens subsecutivos.

61 A técnica da árvore de decisão utiliza, entre outras, a abordagem de predição, na qual grupos diferentes são identificados de acordo com características em comum.

62 É na fase de mineração do *data mining* que são definidos os metadados dos dados manipulados.

63 *Data mart* e *data warehouse* são termos sinônimos que se referem igualmente a um local onde é armazenada uma grande quantidade de dados.

64 Em *data mining*, a técnica de associação é uma função que determina o coeficiente de afinidade entre certos eventos.

65 Em um banco de dados NoSQL, o requisito de consistência para sistemas distribuídos é obtido quando todos os nós contêm os mesmos dados.

66 A análise de conglomerados tem como objetivo agrupar os elementos de um conjunto de dados que tenham semelhanças entre si.

67 O processo de análise por inferência busca o conhecimento dos dados para obtenção de resultados consolidados.

68 Os dados utilizados como fonte para análise e *data mining* são considerados semiestruturados quando possuem uma estrutura não homogênea, como arquivos XML e JSON.

69 Em *data mining*, um mesmo processo de análise de dados pode utilizar bancos de dados relacionais ou NoSQL, mas não simultaneamente.

70 No contexto de mineração de dados, o atributo da veracidade está associado ao grau de confiabilidade dos dados utilizados na solução.

Com base em obtenção, limpeza e pré-processamento de dados, julgue os próximos itens.

71 A *toquenização* (do inglês *tokenization*) de palavras é uma atividade inerente aos algoritmos de análise granular de palavras em documentos.

72 É na etapa de modelagem do processo de análise que ocorrem a seleção dos dados, a definição e a construção do modelo.

73 No processo de análise de dados, a identificação de correlação entre dois dados define também a causalidade entre esses dados.

74 As anomalias verificadas em um conjunto de dados devem ser descartadas de modo a se manter a harmonia do conjunto analisado.

75 A análise descritiva tem como objetivo responder à pergunta ‘o que aconteceu?’, ao passo que a análise diagnóstica visa responder à questão ‘por que isso aconteceu?’.

76 O objetivo da classificação na fase de pré-processamento é o particionamento de um grupo de documentos em subgrupos, com características em comum.

77 O pré-processamento de dados deve ocorrer em bases de informações não estruturadas, com o objetivo de diminuir a quantidade de dados a serem processados.

78 Em um processo de análise de dados, os *shadow systems* não representam dados consolidados, por isso devem ser excluídos de todas as fases do projeto.

79 A normalização de um conjunto de dados tem como objetivo representar todos os valores desse conjunto em uma escala de 0 a 1.

80 No processo de análise, as atividades de limpeza, filtragem e estruturação dos dados estão concentradas na etapa de compreensão dos dados.

Julgue os itens que se seguem, relativos a modelos de regressão e inferência estatística.

81 Nos modelos de inferência baseada em regressão linear simples, os erros são correlacionados e sua média é superior a zero.

82 Uma das ferramentas mais úteis para a avaliação da qualidade dos modelos de regressão é o gráfico de resíduos, em que os resíduos são dispostos no eixo das abscissas e os valores da variável explicativa, no eixo das ordenadas.

83 Os modelos de regressão polinomial envolvem funções polinomiais de variável explicativa.

84 A inferência estatística consiste na formulação de proposições probabilísticas a partir de evidências já observadas.

85 A visão subjetivista da inferência estatística determina que as probabilidades são aspectos reais do universo.

86 Segundo a teoria de Bayes, que deu origem às redes bayesianas, eventos passados não influenciam a probabilidade de eventos futuros correlacionados.

87 Os modelos de regressão resistente são alternativas em que os estimadores são baseados em medianas.

88 Abordagens baseadas em regras para a incerteza adicionam um fator de confusão a cada regra para acomodar a incerteza dos eventos.

89 A verdade probabilística de uma sentença complexa é calculada a partir da soma das verdades de cada componente da sentença.

90 O princípio da indiferença estatística afirma que as proposições sintaticamente simétricas em relação à evidência são consideradas proposições de igual probabilidade.

Julgue os itens a seguir, a respeito de algoritmos e técnicas supervisionadas e não supervisionadas de aprendizado de máquina e aprendizagem profunda.

91 A tarefa do agrupamento consiste em segmentar uma base de dados não rotulada em grupos que tenham algum significado ou utilidade prática.

92 No algoritmo K-*means*, a similaridade intragrupo é avaliada considerando-se o valor médio dos objetos em um grupo, que pode ser visto como o seu centro de gravidade ou o centroide.

93 Os algoritmos baseados em árvore de decisão definem modelos com uma técnica para estimar a probabilidade de um evento ocorrer sob determinada circunstância, usando-se uma estimativa *a priori* da probabilidade de sua ocorrência.

94 A distância de Minkowski entre duas cadeias de caracteres é definida como o número mínimo de operações de edição necessárias para transformar a primeira cadeia de caracteres na segunda.

95 Se os números em um conjunto estiverem igualmente espaçados, então a mediana e a média desse conjunto serão iguais.

96 Considere que o índice de calor (I) em determinado local depende da temperatura T (em graus Celsius), da umidade H e da velocidade do vento s (em m/s). Considere, ainda, que um modelo de regressão múltipla para I seja descrito por $I(H, T) = T + 5{,}9\,H - 0{,}42\,s$. Nessa situação, para um dia com temperatura de 31 °C, umidade igual a 0,84 e velocidade do vento de 12 m/s, o índice de calor médio será superior a 35,00.

97 O objetivo das regras de associação é encontrar todos os conjuntos de itens que possuem confiança mínima com máximo de dados observados.

98 O algoritmo Apriori emprega busca em profundidade e gera conjuntos de itens candidatos (padrões) de k elementos a partir de conjuntos de itens de $k - 1$ elementos, sendo os padrões não frequentes eliminados.

99 *Random forests* fornecem pontuações de distância euclidiana entre os recursos que podem ser usadas para compreender a importância relativa de cada recurso na previsão.

100 As redes neurais permitem construir modelos que sejam padronizados de acordo com o funcionamento do cérebro humano.

Julgue os itens a seguir que tratam de extração e representação de conhecimento.

101 Em mineração de dados, interpretação e explanação consiste em filtrar o conjunto de dados por meio de mecanismo que varia de acordo com a técnica de mineração utilizada.

102 As ferramentas de tratamento de dados oferecem recursos que contemplam tanto a limpeza e padronização de dados, com funcionalidades que incluem correções de erros e remoção de duplicatas, quanto a transformação de dados, com funcionalidades que incluem filtragem de dados, junções de tabelas, agregações e cálculos.

103 Se um algoritmo induzir hipóteses que apresentam baixa taxa de acerto mesmo no conjunto de treino, então as hipóteses estão superajustadas aos dados.

A respeito de ontologia, julgue os próximos itens.

104 A representação formal de ontologias é utilizada para que estas sejam consumidas por computadores, enquanto a representação gráfica é utilizada para compreensão humana.

105 Ontologias podem ser entendidas como aplicações criadas para simular a ação de especialistas humanos, com o propósito de solucionar problemas específicos em um dado domínio.

106 Na ontologia, o conjunto de axiomas do domínio é aquele que serve para modelar restrições e regras inerentes às instâncias.

Julgue os itens que se seguem, relativos à OWL e tecnologias semânticas.

107 Na *web* semântica, buscam-se mecanismos que capturem o significado das páginas, criando um ambiente no qual os computadores possam processar e relacionar conteúdos provenientes de várias fontes.

108 Na *web* semântica, as ontologias são os principais elementos que embasam as lógicas de geração de inferências e essas representações fornecem elementos contendo axiomas, que são as bases lógicas para os motores de inferências.

109 A proposta da *web* semântica é a de que seja entendida apenas pelos usuários, que serão capazes de operar eficientemente sobre as informações e compreender seus significados.

110 OWL é o padrão atual em linguagens para ontologia na *web*, recomendada como principal linguagem para construção de ontologias, segundo o consórcio W3C.

Julgue os próximos itens, relativos a linguagens, ferramentas e bibliotecas que facilitam a criação, o treinamento e a implantação de modelos de *software* com aprendizado de máquina.

111 Pandas e NumPy são bibliotecas que oferecem uma base sólida para a realização de tarefas de análise de dados, desde a importação de dados até a visualização de resultados.

112 Bibliotecas como TensorFlow e PyTorch são exclusivas para algoritmos de aprendizado profundo e não são adequadas para tarefas de aprendizado de máquina tradicionais.

113 A biblioteca Scikit-learn do Python inclui métricas de avaliação padrão, métodos para validação cruzada, busca de hiperparâmetros e outras ferramentas para avaliação do desempenho de modelos de classificação.

114 Spacy é uma biblioteca Python eficiente e rápida para processamento de linguagem natural, mas não oferece recursos como análise morfológica, entidades nomeadas em texto e integração com modelos de aprendizado de máquina pré-treinados.

115 O PyTorch permite que os desenvolvedores definam seus modelos de maneira dinâmica, o que pode ser vantajoso em cenários nos quais a estrutura do modelo tem de ser modificada.

Acerca do desenvolvimento de *pipelines* e do processamento distribuído para aprendizado de máquina, julgue os seguintes itens.

116 Kubeflow é um serviço de nuvem autônomo para gerenciamento e processamento de grandes volumes de dados.

117 Apache Spark é frequentemente usado para escalabilidade em ambientes de aprendizado de máquina.

118 MLflow é uma ferramenta exclusiva para uma única tarefa, como treinamento ou implantação de modelos, proporcionando funcionalidades para experimentação, rastreamento de parâmetros e métricas, reprodução de modelos, empacotamento e implantação.

119 Um dos propósitos da ferramenta Docker, no contexto de desenvolvimento de *pipelines* para aprendizado de máquina, é facilitar a criação e distribuição de ambientes isolados, proporcionando uma melhoria na segurança do ambiente de desenvolvimento.

120 A gestão eficiente de recursos como CPU e memória não é uma consideração crítica em *pipelines* de aprendizado de máquina, pois os recursos podem ser alocados dinamicamente, conforme necessário.

Espaço livre