

Matéria: Tecnologia da Informação

Assunto: Aprendizado de Máquina, Processamento de Linguagem Natural, Big Data, Programação

Resumo Teórico do Assunto

Para ter sucesso nas questões sobre Aprendizado de Máquina, Processamento de Linguagem Natural, Big Data e Programação, é fundamental compreender os conceitos básicos e a inter-relação entre essas áreas.

1. Fundamentos de Dados e Conhecimento

No contexto da Tecnologia da Informação e do Aprendizado de Máquina, a informação é frequentemente organizada em uma hierarquia:

- **Dado (Data):** É a matéria-prima, o elemento bruto, sem contexto ou significado intrínseco. São fatos isolados, números, textos, imagens capturados e armazenados. Ex: "25", "João", "São Paulo".
- **Informação (Information):** Surge quando os dados são processados, organizados e contextualizados, ganhando significado. Responde a perguntas como "quem?", "o quê?", "onde?", "quando?". Ex: "João tem 25 anos e mora em São Paulo."
- **Conhecimento (Knowledge):** É o nível mais alto, resultante da análise e interpretação da informação, permitindo identificar padrões, relações, tendências e insights. Responde a "como?" e "por que?", possibilitando a tomada de decisões e a formulação de regras. Ex: "Pessoas com 25 anos em São Paulo tendem a comprar produtos X."

2. Aprendizado de Máquina (Machine Learning - ML)

Aprendizado de Máquina é um subcampo da Inteligência Artificial que permite aos sistemas aprenderem a partir de dados, identificar padrões e tomar decisões com mínima intervenção humana.

Tipos de Aprendizado de Máquina:

- **Aprendizado Supervisionado:** O modelo é treinado com um conjunto de dados que já possui "rótulos" ou "respostas" corretas. O objetivo é aprender a mapear entradas para saídas, fazendo previsões ou classificações em novos dados. Exemplos: Regressão (prever um valor contínuo) e Classificação (prever uma categoria).
- **Aprendizado Não Supervisionado:** O modelo é treinado com dados que não possuem

rótulos. O objetivo é encontrar padrões ocultos, estruturas ou agrupamentos nos dados. Exemplos: **Agrupamento (Clustering)**, como dividir clientes em grupos com base em seus comportamentos, e Redução de Dimensionalidade.

- **Aprendizado por Reforço:** Um agente aprende a tomar decisões em um ambiente para maximizar uma recompensa. Não há um conjunto de dados pré-definido, mas sim um processo de tentativa e erro.

Técnicas de Preparação de Dados:

- **Redução de Dimensionalidade:** Processo de reduzir o número de atributos (colunas) em um conjunto de dados, mantendo a maior parte da informação relevante. Isso ajuda a lidar com o "paradoxo da dimensionalidade" e melhora o desempenho de modelos.

- * **Análise de Componentes Principais (PCA - Principal Component Analysis):** Uma técnica popular de redução de dimensionalidade que transforma um conjunto de variáveis correlacionadas em um conjunto menor de variáveis não correlacionadas, chamadas componentes principais, que capturam a maior parte da variância dos dados originais.

3. Processamento de Linguagem Natural (NLP)

Processamento de Linguagem Natural (NLP) é um campo da IA que se concentra na interação entre computadores e a linguagem humana. O objetivo é permitir que os computadores compreendam, interpretem e gerem linguagem humana de forma útil.

- **Análise de Sentimentos:** Uma técnica de NLP que visa determinar a polaridade emocional (positiva, negativa, neutra) de um texto, como opiniões em redes sociais ou avaliações de produtos.

- **Tokenização:** Um passo fundamental no pré-processamento de texto em NLP. Consiste em dividir um texto em unidades menores, chamadas **tokens**. Geralmente, os tokens são palavras, mas podem ser também frases, caracteres ou subpalavras, dependendo do contexto.

4. Big Data

Big Data refere-se a conjuntos de dados tão grandes e complexos que os métodos tradicionais de processamento de dados são inadequados. Sua definição original é frequentemente caracterizada pelos "3 V's":

- **Volume:** A enorme quantidade de dados gerados e armazenados.
- **Variedade:** A diversidade de tipos e formatos de dados (estruturados, semi-estruturados, não estruturados) provenientes de múltiplas fontes.
- **Velocidade:** A rapidez com que os dados são gerados, coletados e precisam ser processados e analisados.

Outros "V's" adicionais incluem Veracidade (qualidade e confiabilidade dos dados) e Valor (capacidade de extrair insights úteis).

5. Programação (TypeScript)

TypeScript é um superset de JavaScript, o que significa que todo código JavaScript válido é também um código TypeScript válido. A principal diferença é que TypeScript adiciona tipagem estática opcional, o que ajuda a identificar erros durante o desenvolvimento e melhora a manutenibilidade do código.

- **Funções:** Blocos de código reutilizáveis que executam uma tarefa específica. Em TypeScript/JavaScript, podem receber parâmetros e retornar um valor.

* Ex: ``function nomeDaFuncao(parametro: tipo): tipoRetorno { ... }``

- **Método `map()`:** Um método de array em JavaScript/TypeScript que cria um **novo array** populado com os resultados da chamada de uma função fornecida em cada elemento do array original. Ele não modifica o array original.

* Sintaxe: ``arrayOriginal.map(elemento => expressaoComElemento)``

- **`console.log()`:** Uma função amplamente utilizada para exibir mensagens, variáveis ou o resultado de expressões no console do navegador ou ambiente de execução (como Node.js), útil para depuração e visualização de saída.

Compreender esses conceitos e suas aplicações permitirá analisar e responder às questões de forma eficaz.

Questões de Provas Anteriores

Fonte: [escrituario_agente_de_tecnologia.pdf](#), Página: 13

pcimarkpci MjgwNDowMTRkOjE0YTU6OTI1ODozOGQ2OjNhMGM6NTM0MzplZmI1:U3V
uLCAYnyBKdWwgMjAyNSAyMzo0NzozMSAtMDMwMA==
www.pciconcursos.com.br

AGENTE DE TECNOLOGIA - Microrregião 158 -TI

13

GABARITO 1

BANCO DO BRASIL

39

No processo de aprendizagem de máquina, busca-se identificar padrões compreensíveis em bases de dados. Desse modo, supõe-se a existência de uma hierarquia de três níveis: a base, com os itens elementares, captados e armazenados por recursos de Tecnologia da Informação; o nível intermediário, no qual esses itens são processados, com significados e contextos bem definidos; e o nível mais alto, contendo os padrões ou os conjuntos cuja for -

mulação pode envolver ou relacionar os níveis inferiores. Os três níveis dessa hierarquia, listados, respectivamente, da base para o nível mais alto, são:

- (A) conhecimento, informação e dado
- (B) dado, conhecimento e informação
- (C) dado, informação e conhecimento
- (D) informação, conhecimento e dado
- (E) informação, dado e conhecimento

40

Analise com atenção o código a seguir, escrito em TypeScript 4.0.

```
function segredo(a: number[]) {  
  return a.map(x=>x*2);  
}
```

```
console.log(segredo([1,2,3]));
```

O que será exibido no console quando o código acima for executado?

- (A) "Executed JavaScript Failed:"
- (B) [2, 4, 6]
- (C) [6]
- (D) 6
- (E) 12

41

Uma organização decidiu monitorar a opinião do público sobre ela nas redes sociais. Para isso, processou as mensagens com referências ao seu nome, a fim de possibilitar o uso de uma técnica de processamento de linguagem natural conhecida como análise de sentimentos.

Após transformar cada mensagem em uma string, um dos passos importantes nessa técnica é a tokenização, que consiste em

- (A) colocar todos os caracteres da mensagem em minúsculas.
- (B) colocar todos os verbos da mensagem no infinitivo.
- (C) dividir o texto da mensagem em palavras isoladas.
- (D) eliminar todos os marcadores HTML ou XML da mensagem.
- (E) substituir todos os caracteres acentuados da mensagem por suas versões sem acento.

42

Sabendo que existe, na organização em que trabalha, uma base de dados formada por uma grande tabela que

contém apenas o id do cliente e colunas do tipo booleano indicando se um cliente possuía ou já tinha possuído cada produto da organização, um funcionário de TI resolveu dividir os clientes em grupos apenas com base nessa informação, utilizando aprendizado de máquina.

Para essa tarefa, o funcionário de TI deve utilizar o aprendizado de máquina

- (A) independente
- (B) não supervisionado
- (C) por recompensa
- (D) por reforço
- (E) supervisionado

43

Organizações modernas estão sujeitas a uma grande quantidade de dados, principalmente se tratam com grandes quantidades de clientes. Muitas das demandas de tecnologia de informação a que essas organizações estão submetidas passaram a ser organizadas e tratadas dentro do conceito de Big Data.

Além do grande volume de dados, o Big Data, em sua definição original, considera também a(s) seguinte(s) propriedade(s):

- (A) falta de qualidade, apenas
- (B) variedade, apenas
- (C) velocidade, apenas
- (D) variedade e velocidade
- (E) velocidade e falta de qualidade

44

Ao fazer a preparação para analisar dados em um banco de dados típico de Big Data, um profissional de TI percebeu que o número de atributos, ou colunas, era muito maior do que poderia processar com a ferramenta de análise disponível, sendo necessário, portanto, utilizar uma técnica de redução de dados para prepará-los para análise.

Uma técnica indicada, nesse caso, é a

- (A) Amostragem Aleatória
- (B) Amostragem Estratificada
- (C) Análise de Componentes Principais
- (D) Deduplicação
- (E) Imputação

RASCUNHO