

## QUESTÕES COMENTADAS MINERAÇÃO DE DADOS (CESPE)

Apresentamos abaixo um conjunto de questões sobre o assunto que aprendemos nesta aula. Esperamos que elas ajudem na fixação da matéria. Qualquer dúvida, estamos às ordens!



### 12. Ano: 2018 Banca: CESPE Órgão: EBSEH Prova: Analista de Tecnologia da Informação

Julgue o item que se segue, a respeito de arquitetura e tecnologias de sistemas de informação.

A descoberta de novas regras e padrões em conjuntos de dados fornecidos, ou aquisição de conhecimento indutivo, é um dos objetivos de data mining.

**Comentário:** Vejam que essa é uma questão recente do CESPE. Ela apresenta uma definição consistente de mineração de dados. Veja que temos o aprendizado supervisionado ou indutivo, onde queremos prever um valor futuro para um determinada variável. Ou, por outro lado, queremos descobrir padrões nos dados fornecidos. Ambos podem ser vistos como objetivos da mineração de dados. Logo, temos uma questão correta.

**Gabarito:** C



### 13. Ano: 2018 Banca: CESPE Órgão: TCM-BA Cargo: Auditor de Contas Questão: 12

Assinale a opção correta a respeito do CRISP-DM.

A CRISP-DM é uma suíte de ferramentas proprietárias que vem se tornando um padrão da indústria para mineração de dados, uma vez que fornece um plano completo e tecnologias para a realização de um projeto de mineração de dados.

B A verificação da qualidade dos dados é uma atividade da fase de entendimento dos dados.

C Durante a fase de preparação dos dados, é realizado um inventário de requisitos, suposições e restrições de recursos.

D Na fase de avaliação dos dados, são realizadas as atividades de identificar valores especiais dos dados e catalogar seu significado.

E Na fase de preparação dos dados, são realizadas as atividades de analisar o potencial de implantação de cada resultado e estimar o potencial de melhoria do processo atual.

**Comentário:** Vamos comentar cada uma das alternativas, lembrando que o processo do CRISP-DM é organizado em um conjunto de etapas: entendimento do negócio, seleção dos dados (*data understanding*), preparação dos dados, modelagem dos dados, avaliação do processo e execução (*deployment*).



- A) O CRISP-DM propõe a ser uma referência que propõe uma visão geral do ciclo de vida de um projeto de mineração de dados, logo, a alternativa B está **incorreta**.
- B) **Correto!** Na fase de entendimento ou seleção dos dados é traçado o perfil dos dados. Neste momento é avaliada a qualidade dos dados.
- C) A preparação dos dados está mais preocupada em trazer e ajustar os dados para execução do algoritmo de mineração. A fase de **preparação de dados** consiste na preparação dos dados que visa a **coleta, limpeza, transformação, integração e formatação** dos dados definidos na etapa anterior. Logo, temos mais uma alternativa **incorreta**.
- D) Não existe uma fase de avaliação dos dados dentro do processo. Logo, temos mais uma alternativa **incorreta**.
- E) Vejam que a alternativa A está **incorreta**, as ações descritas são associadas a fase de avaliação do processo.

**Gabarito:** B



**14. Ano: 2018 Banca: CESPE Órgão: TCM-BA Cargo: Auditor de Contas Questão: 13**

A respeito das técnicas e(ou) métodos de mineração de dados,  
assinale a opção correta.

- A O agrupamento (ou clustering) realiza identificação de grupos de dados que apresentam coocorrência.
- B A classificação realiza o aprendizado de uma função que pode ser usada para mapear os valores associados aos dados em um ou mais valores reais.
- C A regressão ou predição promove o aprendizado de uma função que pode ser usada para mapear dados em uma de várias classes discretas definidas previamente, bem como encontrar tendências que possam ser usadas para entender e explorar padrões de comportamento dos dados.
- D As regras de associação identificam grupos de dados, em que os dados têm características semelhantes aos do mesmo grupo e os grupos têm características diferentes entre si.
- E Os métodos de classificação supervisionada podem ser embasados em separabilidade (entropia), utilizando árvores de decisão e variantes, e em particionamento, utilizando SVM (support vector machines).

**Comentário:** Vamos analisar cada uma das alternativas acima.

- A) A alternativa apresenta o conceito de coocorrência que está relacionado ao conceito de regra de associação, e tenta colocá-lo dentro do contexto de agrupamento. Sendo assim, temos mais uma alternativa **incorreta**.



- B) A classificação baseia-se na definição de um conjunto de rótulos que permitem classificar uma nova informação em um conjunto de classes pré-definidas. Veja que essas classes não necessariamente são valores reais. Logo, temos uma alternativa incorreta.
- C) A regressão não vai mapear valores em um conjunto de classes. Se você pensar que a regressão é uma função de múltiplas variáveis a imagem desta pode ser um conjunto fechado, porém, contínuo. Logo, temos uma alternativa errada.
- D) A alternativa D descreve uma característica da tarefa de clusterização. Logo, temos uma alternativa **incorreta**.
- E) **Essa é a nossa resposta.** Basicamente, o SVM é um algoritmo supervisionado que tenta criar uma linha (ou uma fronteira) que melhor separa os dados. Essa linha normalmente chamamos de “**Hyperplano**”. Para entender um pouco mais desse algoritmo acesse o [link](#). Veja que ele aplica o conceito de particionamento. Agora ideia da classificação é construir um conceito de separabilidade entre os novos valores de entrada e as classes ou rótulos estabelecidos.

**Gabarito:** D.



**15. Ano: 2017 Banca: CESPE Órgão: TCE-PE Cargo: Auditor de Obras Públicas Questão: 119**

Julgue o item que se refere a CRISP-DM (Cross Industry Standard Process for Data Mining).

[119] Durante a fase de entendimento do negócio, busca-se descrever claramente o problema, fazer a identificação dos dados e verificar se as variáveis relevantes para o projeto não são interdependentes.

**Comentário:** Essa para mim foi a questão mais difícil da prova. Primeiramente você teria que se lembrar das fases do CRISP, a primeira fase é o entendimento do negócio e a segunda entendimento dos dados. Vejamos o que cada uma tem como objetivo:

**Entendimento do negócio:** deve determinar os objetivos de negócio, fazer uma análise da situação atual e estabelecer os objetivos da mineração de dados. Finalizando com um plano de projeto.

**Entendimento dos dados:** Nesta etapa vamos entender os dados baseados nos requisitos. Nesta etapa podemos incluir uma coleta de dados, descrição, exploração e verificação da qualidade dos mesmos. Nesta etapa temos uma característica peculiar: identificar se as variáveis do modelo são independentes umas das outras. Quando as variáveis são independentes podemos concluir que elas não possuem informações sobrepostas. Em econometria ou análise matemática podemos pensar em variáveis que são linearmente independentes. Uma escolha cuidadosa de variáveis independentes pode fazer com que a execução dos algoritmos seja feita de forma mais eficiente.

Enfim, depois desta longa explicação teórica, podemos perceber que o examinador associou eventos de etapas diferentes dos CRISP a fase de entendimento do negócio. Logo, a alternativa está incorreta.



**Gabarito: E.**



**16. Ano: 2017 Banca: CESPE Órgão: TCE-PE Cargo: Analista De Controle Externo Área: Auditoria De Contas Públicas Questão: 119**

Em relação à análise de agrupamentos (clusterização) em mineração de dados, julgue o item seguinte.

119 O método de clustering k-means objetiva particionar 'n' observações entre 'k' grupos; cada observação pertence ao grupo mais próximo da média.

**Comentário:** O algoritmo de K-means é de fato um método de clusterização. Confesso que essa informação não foi vista no nosso curso. O algoritmo inicia com a escolha dos k elementos que formaram as sementes iniciais. Vamos tentar entender um pouco mais do funcionamento do mesmo.

Escolhidas as sementes iniciais, é calculada a distância de cada elemento em relação às sementes, agrupando o elemento ao grupo que possuir a menor distância (mais similar) e recalculando o centroide do mesmo. Quando temos mais de um elemento, imagine vários pontos em um papel, o centroide é um ponto central entre esses pontos. O processo é repetido até que todos os elementos façam parte de um dos clusters.

Depois desta etapa, fazemos ajustes nos elementos usando métodos estatísticos que tentam diminuir a dispersão dentro de cada grupo, por meio da mudança de elementos entre os grupos. O processo é interrompido quando a mudança de um elemento de um cluster para outro não gera mais ganho.

Para visualizar o algoritmo funcionando, você pode olhar na página da Wikipédia em inglês<sup>2</sup>. Para saber a fonte da questão, basta acessar a definição da Wikipédia em português: “Em mineração de dados, agrupamento k-means é um método de Clustering que objetiva particionar n observações dentre k grupos onde cada observação pertence ao grupo mais próximo da média. Isso resulta em uma divisão do espaço de dados em um Diagrama de Voronoi.”

Desta forma, a alternativa está correta.

**Gabarito: C**



**17. Ano: 2017 Banca: CESPE Órgão: SEDF Cargo: Analista de gestão educacional – Especialidade: tecnologia da informação Questão: 119**

<sup>2</sup> [https://en.wikipedia.org/wiki/K-means\\_clustering](https://en.wikipedia.org/wiki/K-means_clustering)

Com relação a data mining e data warehouse, julgue os itens que se seguem.

[119] Agrupar registros em grupos, de modo que os registros em um grupo sejam semelhantes entre si e diferentes dos registros em outros grupos é uma maneira de descrever conhecimento descoberto durante processos de mineração de dados.

**Comentário:** Se analisarmos a descrição acima temos uma definição da tarefa de agrupamento ou **clustering**.

A **clusterização** é a classificação **não-supervisionada de dados**, formando agrupamentos ou clusters. Ela representa uma das principais etapas do processo de análise de dados denominada análise de clusters. A análise de clusters envolve, portanto, a **organização de um conjunto de padrões** (usualmente representados na forma de vetores de atributos ou pontos em um espaço multidimensional – espaço de atributos) **em clusters**, de acordo com **alguma medida de similaridade**. De forma intuitiva, padrões pertencentes a um dado cluster devem ser mais “similares” entre si do que em relação a padrões pertencentes a outros clusters.

Vejam, portanto, que a alternativa está correta.

**Gabarito:** C.



**18. Ano: 2016 Banca: CESPE Órgão: TRT-08 Cargo: Analista de TI - QUESTÃO 10**

Acerca de data mining, assinale a opção correta.

A A fase de preparação para implementação de um projeto de data mining consiste, entre outras tarefas, em coletar os dados que serão garimpados, que devem estar exclusivamente em um data warehouse interno da empresa.

B As redes neurais são um recurso matemático/computacional usado na aplicação de técnicas estatísticas nos processos de data mining e consistem em utilizar uma massa de dados para criar e organizar regras de classificação e decisão em formato de diagrama de árvore, que vão classificar seu comportamento ou estimar resultados futuros.

C As aplicações de data mining utilizam diversas técnicas de natureza estatística, como a análise de conglomerados (cluster analysis), que tem como objetivo agrupar, em diferentes conjuntos de dados, os elementos identificados como semelhantes entre si, com base nas características analisadas.

D As séries temporais correspondem a técnicas estatísticas utilizadas no cálculo de previsão de um conjunto de informações, analisando-se seus valores ao longo de determinado período. Nesse caso, para se obter uma previsão mais precisa, devem ser descartadas eventuais sazonalidades no conjunto de informações.

E Os processos de data mining e OLAP têm os mesmos objetivos: trabalhar os dados existentes no data warehouse e realizar inferências, buscando reconhecer correlações não explícitas nos dados do data warehouse.



**Comentário:** Teceremos comentários sobre cada uma das alternativas acima.

A Sabemos que a mineração de dados pode acontecer sobre qualquer tipo de arquivo de dados. Lembrem-se a possibilidade de *textmining* que não tem necessidade de dados armazenados em um DW. Alternativa **errada!**

B Na alternativa B existe uma avalanche de conceitos misturados: redes neurais, que fazem parte do conjunto de assuntos relacionados a inteligência artificial; técnicas estatística e árvore de decisão. Cada técnica de mineração é usada com um propósito específico, por exemplo, a classificação vai permitir que você classifique novas entradas de acordo com um conjunto pré-determinado de saídas, que foram construídos em uma etapa anterior do processo. A questão peca por misturar vários conceitos.

C Criar clusters, ou seja, agrupar subconjuntos de dados de acordo com alguma semelhança entre eles. Essa é a nossa resposta.

D Uma serie temporal deve considerar a sazonalidade, pela lei da oferta e demanda, se você percebe que as vendas aumentam no Natal, você pode aumentar o preço ou o estoque. O fato de desconsiderar a sazonalidade torna a questão incorreta.

E Os processos de OLAP e Data mining são diferentes em relação a complexidade e resultados esperados. OLAP é uma ferramenta de consulta em bases de dados analíticas, ele visa extrair informações por meio de queries e utilizando as operações sobre os cubos de dados, mas não aplicam algoritmos específicos neste processo. Data Mining é bem mais complexo que OLAP, ele busca padrões em grandes volumes de dados por meio de técnicas estatísticas e de algoritmos de inteligência artificial, por exemplo. Sendo assim não é possível comparar de forma tão simplista quando a alternativa tentou fazer, por isso, a letra E está incorreta.

**Gabarito: C**



**19. Ano: 2016 Banca: CESPE Órgão: FUNPRESP-JUD Prova: Analista - Tecnologia da Informação**

Julgue o item subsecutivo, referente às tecnologias de bancos de dados.

Em DataMining, as árvores de decisão podem ser usadas com sistemas de classificação para atribuir informação de tipo.

**Comentário:** O algoritmo de Árvores de decisão gera uma estrutura de árvore que ajuda na classificação e predição das amostras desconhecidas. Com base nos registros do conjunto de treinamento, uma árvore é montada e, a partir desta árvore, pode-se classificar a amostra desconhecida sem necessariamente testar todos os valores dos seus atributos. O algoritmo de classificação por árvores de decisão é considerado um algoritmo supervisionado, pois é necessário saber quais são as classes de cada registro do conjunto de treinamento.

Como o algoritmo monta uma árvore, é necessário definir quais são os elementos desta árvore. Para simplificar a explicação do algoritmo, basta pensar em uma árvore como um conjunto de





nós que são conectados por ramificações. Basicamente existem três tipos de nós: o nó raiz, que inicia a árvore, os nós comuns que dividem um determinado atributo e geram ramificações e os nós-folha que contêm as informações de classificação do algoritmo. Já as ramificações possuem todos os valores possíveis do atributo indicado no nó para facilitar a compreensão e interpretação. Veja a figura abaixo:

### Árvore de Decisão para Jogar Tênis



A ideia do algoritmo é montar uma árvore onde cada nó indica o teste de um atributo. Os atributos escolhidos são chamados de atributos divisores ou atributos teste. A escolha de atributos é feita com base no maior ganho de informação, isto é, na qualidade de classificação do atributo. Deste modo, podemos dizer que o atributo que melhor classifica os dados deve ser escolhido como um nó da árvore. Para facilitar a compreensão, é comum colocar os valores das probabilidades de cada classe dentro do nó.

Aproveitamos a questão para introduzir maiores detalhes sobre árvore de decisão. Observamos ainda que a alternativa está plenamente correta.

**Gabarito:** Certo



### 20. Ano: 2016 Banca: CESPE Órgão: FUNPRES-EXE Prova: Especialista - Tecnologia da Informação

Com relação à forma como os dados são armazenados e manipulados no desenvolvimento de aplicações, julgue o item a seguir.

Na implementação de mineração de dados (data mining), a utilização da técnica de padrões sequenciais pode ser útil para a identificação de tendências.

**Comentário:** Algoritmos de padrões sequenciais identificam tipos de padrões sequenciais em restrições mínimas especificadas pelo usuário. Esta técnica procura por compras ou eventos que ocorrem em uma sequência através do tempo. Por exemplo, uma loja pode descobrir que consumidores que comprem TVs tendem também a comprar filmadoras de 8mm em 60% das vezes. Ou seja, pode identificar uma tendência de compras. Logo, temos uma alternativa correta!

**Gabarito:** Certo





**21. Ano: 2016 Banca: CESPE Órgão: TCE-SC Cargo: Auditor de TI**

Julgue os itens subsecutivos, acerca de mineração de dados.

98 Para a realização de prognósticos por meio de técnicas de mineração de dados, parte-se de uma série de valores existentes obtidos de dados históricos bem como de suposições controladas a respeito das condições futuras, para prever outros valores e situações que ocorrerão e, assim, planejar e preparar as ações organizacionais.

99 As aglomerações, tipos de informação obtidos por meio da mineração de dados, caracterizam-se por se ligarem a um único e específico evento, em torno do qual ocorrem várias ações, com produção sistêmica de informações gerenciais que apoiarão uma nova ocorrência do mesmo tipo de evento.

**Comentário:** A alternativa 98 trata dos aspectos de previsão relacionados a mineração de dados. Basicamente você deve escolher um algoritmo, parametrizar, utilizar dados históricos das bases de dados como entrada e tentar prever o futuro. O texto da questão está correto.

Sobre a questão 99, aglomerações ou clusters são grupos de indivíduos de uma amostra que possuem características semelhantes. Geralmente são definidos intervalos de valores, para cada intervalo temos um clusters, agrupamento ou aglomeração. Essa ideia de eventos em série relacionados a um evento inicial está relacionada a técnicas de associação. Questão, portanto, encontra-se incorreta.

**Gabarito: C E**



**22. ANO: 2015 BANCA: CESPE ÓRGÃO: TCU PROVA: AUDITOR FEDERAL DE CONTROLE EXTERNO – ANÁLISE DE INFORMAÇÕES.**

No que concerne a data mining (mineração de dados) e big data, julgue os seguintes itens.

[82] O uso prático de data mining envolve o emprego de processos, ferramentas, técnicas e métodos oriundos da matemática, da estatística e da computação, inclusive de inteligência artificial.

[83] Quem utiliza o data mining tem como objetivo descobrir, explorar ou minerar relacionamentos, padrões e vínculos significativos presentes em grandes massas documentais registradas em arquivos físicos (analógicos) e arquivos lógicos (digitais).

[84] A finalidade do uso do data mining em uma organização é subsidiar a produção de afirmações conclusivas acerca do padrão de comportamento exibido por agentes de interesse dessa organização.





[85] No ambiente organizacional, devido à grande quantidade de dados, não é recomendado o emprego de data mining para atividades ligadas a marketing.

**Comentários:** Vamos analisar cada uma das assertivas acima a respeito de mineração de dados.

[82] Se analisarmos qualquer definição formal de Data Mining, por exemplo: “Mineração de dados, ou data mining, é o processo de análise de conjuntos de dados que tem por objetivo a descoberta de padrões interessantes e que possam representar informações úteis”.

Segundo a teoria Data Mining é uma mistura de diferentes disciplinas: Estatística, Aprendizado de máquina e Banco de dados. Podemos então verificar que os termos listados na questão se enquadram perfeitamente dentro do contexto e das definições existentes. Portanto, alternativa correta.

[83] Vamos analisar mais uma definição de Data Mining para chegarmos a uma conclusão definitiva a respeito desta questão. “A mineração de dados é um campo interdisciplinar que reúne técnicas de aprendizado de máquina, reconhecimento de padrões, estatísticas, banco de dados e visualização para abordar a questão da extração de informações a partir de grandes bases de dados”. Vejam que em todas as definições as análises são feitas sobre base de dados digitais. Usar tecnologia para otimizar e ampliar o horizonte e descobrir padrões ou informações relevantes. Sendo assim a alternativa está incorreta pois afirma que é possível fazer análise em arquivos físicos (analógicos).

[84] A princípio você pode ser levado a acreditar que a questão está falsa, pois Data Mining não vai produzir afirmações conclusivas. Veja, porém, que o examinador usou a palavra “subsidiar” a produção. Esse é justamente a grande característica que está descrita no detalhamento da última falácia acima. Resposta da alternativa 84 é **correta**.

[85] Observem que o examinador colocou uma afirmação incorreta. Dizer que não é possível usar DM para atividades ligadas a marketing. O caso mais clássico conhecido é justamente o das “fraudas próximas das cervejas<sup>3</sup>”. Uma das maiores redes de varejo dos Estados Unidos descobriu em seu gigantesco armazém de dados que a venda de fraldas descartáveis estava associada à de cerveja. Em geral, os compradores eram homens, que saíam à noite para comprar fraldas e aproveitavam para levar algumas latinhas para casa.

**Gabarito: C E C E.**



### 23. CESPE - DEPEN - 2015 - Agente Penitenciário Federal - Tecnologia da Informação (Médio)

Acerca de datawarehouse e datamining, julgue os itens subsequentes.

<sup>3</sup> <http://exame.abril.com.br/revista-exame/edicoes/633/noticias/o-que-cerveja-tem-a-ver-com-fraldas-m0053931>



[116] Os objetivos do datamining incluem identificar os tipos de relacionamentos que se estabelecem entre informações armazenadas em um grande repositório.

[117] Datamart é a denominação atribuída a um sistema de dataware que atende a áreas específicas de negócios de organizações e que representa um subconjunto lógico do datawarehouse.

[118] O datawarehouse possibilita a análise de grandes volumes de dados, que, por sua vez, permitem a realização de uma melhor análise de eventos futuros.

**Comentário.** Vamos então comentar cada uma das afirmações acima:

116. A definição dos objetivos de **datamining** presentes na alternativa está alinhada com as definições clássicas do conceito, vejamos uma delas: "Mineração de dados é o processo de exploração de grandes quantidades de dados com o objetivo de encontrar anomalias, padrões e correlações para suportar a tomada de decisões e proporcionar vantagens estratégicas. Usando uma ampla variedade de técnicas, você pode utilizar estas informações para aumentar as receitas, reduzir custos, melhorar o relacionamento com os clientes, reduzir riscos e muito mais". Pelo exposto podemos concluir que a alternativa se encontra correta.

117. O conceito de datamart está associado a um subconjunto do data warehouse e, normalmente, é orientado para uma área de negócios da empresa ou equipe específica. Considerando que os data warehouses têm seu escopo sobre toda a empresa ou organização, as informações em data marts pertencem a um único departamento. Desta forma, podemos concluir que a afirmação está correta.

118. Do ponto de vista teórico, e talvez um pouco purista, as tarefas de data mining são divididas em **descritivas** e **preditivas**. As **descritivas** caracterizam as propriedades gerais dos dados em um banco de dados, basicamente, focam em achar padrões reconhecidos por seres humanos para descrever os dados.

Já as **preditivas** essas tarefas realizam uma inferência sobre os dados atuais para fazer previsões sobre os mesmos. Usam variáveis para **prever valores futuros ou desconhecidos de outras variáveis**. Vejam que você **não** faz análise de eventos futuros. Isso é semanticamente incoerente. Desmembrando a questão temos:

1 - "O datawarehouse **possibilita** a **análise de grandes volumes de dados**." Está frase está ok, usou a palavra **possibilita** e não disse que era **responsável**. Continuando a análise:

2 - "... que, **por sua vez**, permitem a realização de uma melhor análise de eventos futuros." Ao meu ver, esse "**que, por sua vez**" diz respeito a "**análise de grandes volumes de dados**" e não ao DW. Então substituindo ficaria assim: "**..análise de grandes volumes de dados permitem a realização de uma melhor análise de eventos futuros**"

Veja que está estranho, a análise permite uma melhor análise ... desta forma a questão encontra-se **errada**.

**Gabarito: C C E.**





**24. Ano: 2015 Banca: CESPE Órgão: TJ-DFT Prova: Técnico Judiciário - Programação de Sistemas**

Julgue o item a seguir, a respeito de datawarehouse e de datamining.

Em um processo de mineração, durante a etapa de preparação dos dados, são analisados os requisitos de negócio para consolidar os dados.

**Comentário:** Lembre-se que o processo CRISP-DM tem uma etapa específica para analisar os requisitos de negócio. Em outra etapa temos a preparação dos dados. Logo, a afirmação acima está incorreta.

**Gabarito:** E



**25. Ano: 2015 Banca: CESPE Órgão: MEC Prova: Administrador de Dados**

Acerca de data warehouse (DW), Business Intelligence (BI) e data mining, julgue o item que se segue.

Situação hipotética: Após o período de inscrição para o vestibular de determinada universidade pública, foram reunidas informações acerca do perfil dos candidatos, cursos inscritos e concorrências. Ademais, que, por meio das soluções de BI e DW que integram outros sistemas, foram realizadas análises para a detecção de relacionamentos sistemáticos entre as informações registradas. Assertiva: Nessa situação, tais análises podem ser consideradas como data mining, pois agregam valor às decisões do MEC e sugerem tendências, como, por exemplo, o aumento no número de escolas privadas e a escolha de determinado curso superior.

**Comentário:** Observem que a afirmação está correta e de acordo com o que vimos até aqui. A mineração de dados ajuda a identificar tendências sobre os dados. Essas tarefas são conhecidas como preditivas.

**Gabarito:** C



**26. Ano: 2015 Banca: CESPE Órgão: MEC Prova: Administrador de Banco de Dados**

Julgue o item seguinte, referente a data mining.

[1] Selecionar uma amostra e determinar os conjuntos de itens frequentes dessa amostra para formar a lista de previsão de subconjunto são as principais características do algoritmo de previsão.



[2] A predição em algoritmos de data mining objetiva modelar funções sobre valores para apresentar o comportamento futuro de determinados atributos.

[3] Algoritmo genético é uma das ferramentas do data mining que utiliza mecanismos de biologia evolutiva, como hereditariedade, recombinação, seleção natural e mutação, para solucionar e agrupar problemas.

Comentário: Vamos comentar cada uma das alternativas acima.

[1] Vamos comparar o conceito de previsão com o algoritmo de amostragem para entender a diferença entre eles:

**Previsão:** Esta técnica tem por objetivo a avaliação de um valor de uma variável ainda não identificada, baseando-se em dados adquiridos através do comportamento desta variável ao longo do tempo.

**Amostragem:** A ideia principal do **algoritmo de amostragem** é selecionar uma pequena amostra, que caiba na memória principal do banco de dados de transações, e determinar os conjuntos de itens frequentes daquela amostra.

**Veja que a alternativa 1 mistura os dois conceitos, logo, temos uma alternativa errada.**

[2] Agora sim! Temos uma definição consistente de previsão conforme apresentada no comentário acima. Logo, alternativa correta.

[3] **Algoritmos Genéricos** ou AGs são algoritmos de otimização e busca baseados nos mecanismos de seleção natural e genética. Enquanto os métodos de otimização e busca convencionais trabalham geralmente de forma sequencial, avaliando a cada instante uma possível solução, os AGs trabalham com um conjunto de possíveis soluções simultaneamente. Algoritmos Genéticos (AGs) são uma classe de procedimentos de pesquisa aleatórios capazes de realizar pesquisas adaptativas e robustas sobre uma ampla gama de topologias de espaço de pesquisa. Modelados após o surgimento adaptativo de espécies biológicas a partir de mecanismos evolutivos e introduzidos por Holland, AGs vêm sendo aplicados com sucesso em campos diversificados como análise de imagens, escalonamentos e projetos de engenharia.

Algoritmos genéticos são uma classe particular de algoritmos evolutivos que usam técnicas inspiradas pela biologia evolutiva como hereditariedade, mutação, seleção natural e recombinação (ou crossing over). Logo, a alternativa 3 está correta.

Gabarito: E C C

