

# AGÊNCIA NACIONAL DE MINERAÇÃO (ANM)

## CARGO 24: ESPECIALISTA EM RECURSOS MINERAIS

### ESPECIALIDADE: TECNOLOGIA DA INFORMAÇÃO

### CIÊNCIA DE DADOS

Prova Discursiva

Aplicação: 16/02/2025

## PADRÃO DE RESPOSTA DEFINITIVO

O processo de descoberta de conhecimento em bancos de dados (KDD) é composto por várias etapas sequenciais que visam transformar dados brutos em informações úteis.

i. Seleção dos dados: refere-se à escolha de dados relevantes para o objetivo da análise. Essa etapa é importante, pois os dados irrelevantes ou de baixa qualidade podem comprometer as demais etapas do KDD.

ii. Pré-processamento: refere-se ao processo de limpeza e à transformação dos dados, como a eliminação de valores ausentes, a normalização e a remoção de *outliers*. Esse processo é importante, pois visa garantir que os dados sejam adequados para a mineração e que não causem distorções nos resultados.

iii. Transformação dos dados: refere-se ao processo pelo qual os dados são transformados ou combinados em formatos adequados para análise por meio de técnicas como agregação ou redução de dimensionalidade. Essa etapa é importante, pois garante que os dados estejam adequados para as análises subsequentes.

iv. Mineração de dados: refere-se à aplicação de técnicas de aprendizado de máquina e técnicas estatísticas a fim de identificar padrões, associações e correlações nos dados. Essa etapa é importante, pois, a partir da aplicação dessas técnicas, há grande possibilidade de se encontrar padrões úteis à análise.

v. Interpretação e avaliação: refere-se à etapa de análise, interpretação e avaliação dos resultados extraídos na etapa de mineração de dados quanto à sua relevância e aplicabilidade. Essa etapa é importante, pois permite que se entenda o valor agregado do conhecimento extraído e a extensão de sua aplicabilidade.

Durante o processo de KDD, várias técnicas de mineração de dados podem ser utilizadas, dependendo do objetivo do estudo e do tipo de dados. ~~O candidato deverá citar quatro dessas técnicas.~~ **O candidato deve citar duas, dentre as técnicas a seguir:**

i. Classificação: essa técnica visa atribuir rótulos ou categorias a dados. Por exemplo, em um sistema de recomendação, um algoritmo de classificação pode prever se um cliente comprará ou não um determinado produto.

ii. *Clustering* (agrupamento): o *clustering* é uma técnica de aprendizado não supervisionado que agrupa dados semelhantes em *clusters* ou grupos. Ele é útil para explorar dados e identificar padrões ocultos.

iii. Regras de associação: essa técnica busca identificar relações entre diferentes variáveis, como no caso de descobrir que “clientes que compram pão também compram leite”.

iv. Redução de dimensionalidade: técnicas como análise de componentes principais (PCA) são usadas para reduzir a complexidade dos dados, mantendo a maior quantidade possível de informação relevante.

v. Regressão: utilizada para modelar a relação entre variáveis, a regressão é uma técnica importante para prever valores contínuos, como a previsão de vendas de um produto com base em diferentes variáveis.

A interpretação e validação dos padrões descobertos são etapas cruciais no processo de KDD, pois determinam a utilidade prática do conhecimento extraído. Após a mineração de dados, os padrões ou os modelos gerados devem ser avaliados para garantir que sejam significativos, precisos e úteis para o problema em questão. Isso envolve a análise do contexto em que os padrões serão aplicados, considerando a viabilidade e a relevância das conclusões. A validação é feita por meio de testes, como a avaliação de desempenho de modelos preditivos usando métricas de precisão, *recall* e F1-Score, ou a verificação da robustez das associações encontradas. Se os padrões forem interpretados corretamente, eles podem ser usados para a tomada de decisões estratégicas, como ajustes em campanhas de *marketing*, aprimoramento de produtos ou otimização de processos operacionais. No entanto, se os padrões não forem bem interpretados ou validados, eles podem levar a conclusões errôneas e decisões inadequadas, prejudicando a organização e comprometendo os investimentos feitos na mineração de dados.

### QUESITOS AVALIADOS

#### QUESITO 2.1 – Etapas do processo de KDD e importância de cada uma delas

Conceito 0 – Não abordou nenhuma etapa do processo de KDD nem sua importância.

Conceito 1 – Mencionou as etapas, mas não as explicou nem abordou sua importância.

Conceito 2 – Abordou, corretamente, apenas uma etapa do processo de KDD, mas não explicou sua importância.

Conceito 3 – Abordou, corretamente, apenas uma etapa do processo de KDD e sua importância.

Conceito 4 – Abordou, corretamente, duas etapas do processo de KDD, mas não explicou a importância de nenhuma delas.

Conceito 5 – Abordou, corretamente, duas etapas do processo de KDD, mas explicou adequadamente a importância de apenas uma delas.

Conceito 6 – Abordou corretamente duas ou mais etapas do processo de KDD e explicou a importância de cada uma delas.

### **QUESITO 2.2 – Técnicas de mineração de dados utilizadas durante o processo de KDD**

Conceito 0 – Não descreveu corretamente nenhuma técnica de mineração de dados utilizada durante o processo de KDD.

Conceito 1 – Descreveu corretamente apenas uma técnica de mineração de dados utilizada durante o processo de KDD.

Conceito 2 – Descreveu, de forma parcialmente correta, duas técnicas de mineração de dados utilizadas durante o processo de KDD.

Conceito 3 – Descreveu corretamente duas técnicas de mineração de dados utilizadas durante o processo de KDD.

### **QUESITO 2.3 – Explicação sobre como a interpretação e validação dos padrões descobertos impactam a utilidade do conhecimento extraído para a tomada de decisões**

Conceito 0 – Não explicou como a interpretação e validação dos padrões descobertos impactam a utilidade do conhecimento extraído para a tomada de decisões.

Conceito 1 – Explicou, de forma parcialmente inconsistente, como a interpretação e validação dos padrões descobertos impactam a utilidade do conhecimento extraído para a tomada de decisões.

Conceito 2 – Explicou corretamente como a interpretação e validação dos padrões descobertos impactam a utilidade do conhecimento extraído para a tomada de decisões.