



PROFISSÃO  
POLICIAL

# Informática

Professor Fabio Rosar

# INFORMÁTICA

## Professor Fabio Rosar

### Sumário

<b>1</b>	<b>MINERAÇÃO DE DADOS .....</b>	<b>2</b>
1.1	TAREFAS DA MINERAÇÃO DE DADOS.....	5
<b>2</b>	<b>APRENDIZADO DE MÁQUINA .....</b>	<b>7</b>
2.1	MÉTODOS PARA O APRENDIZADO DE MÁQUINA .....	8
2.2	TÉCNICAS DE APRENDIZADO DE MÁQUINA.....	10
2.3	MÉTRICAS.....	11
2.4	ÉTAPAS.....	12
<b>3</b>	<b>QUESTÕES DE RENDIMENTO .....</b>	<b>14</b>

# MINERAÇÃO DE DADOS E APRENDIZADO DE MÁQUINA

## 1 MINERAÇÃO DE DADOS

Também conhecido por data mining. Trata de refinar a análise de uma estrutura de dados complexa no intuito de maximizar e antecipar ações estratégicas, minimizar riscos, aumentar a fidelidade e a proximidade ao cliente.

Um exemplo de Data Mining é quando alguma compra “fora do perfil” do usuário de cartão de crédito é efetuada e a operadora entra em contato com o cliente para questionar se realmente aquela compra foi feita por ele pois detectaram que foge do seu perfil de compra. A operadora sabe do perfil pois tem muitos dados que podem ser cruzados. Isso é mineração de dados.

A partir deste conceito, para que uma mineração de dados seja executada com sucesso, alguns requisitos devem ser seguidos, entre eles a inserção correta dos dados no banco, evitar redundâncias e dados desnecessários.

Nesse âmbito surgem os repositórios de dados que são extremamente organizados, os Data Marts (banco de dados setorizados, geralmente por departamentos da empresa – BD do marketing, BD de vendas, BD de clientes) e os conhecidos Data Warehouse (os maiores bancos de dados recebem esses nomes e são genéricos e completos).

A mineração de dados não foi criada com objetivo de substituir as técnicas atuais de análise de dados. Ela utiliza como base para a maioria de seus trabalhos os experimentos da Estatística, Inteligência Artificial, Máquina de Estado e Banco de dados para construir seu modelo.

Diversas pesquisas têm sido direcionadas para o desenvolvimento de técnicas com objetivo de extrair informações a partir de um grande volume de dados e

transformar estas informações em conhecimento útil. Esta área é conhecida na literatura como KDD (Fayyad et al., 1996b).

O processo de KDD surgiu no final da década de 80 com o objetivo principal de procurar conhecimento em bases de dados. Há várias definições do seu conceito, sendo a seguinte definição a mais utilizada:

“KDD é um processo, de várias etapas, não trivial, interativo e iterativo, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados” (FAYYAD, 1996).

O KDD (knowledge discovery in databases ou descoberta de conhecimento nas bases de dados) é uma tentativa de solucionar o problema causado pela chamada sobrecarga de dados proveniente da geração de grande volume de informações em nossa atualidade, o que o autor reconhece como a “era da informação”.

Para Fayyad, Piatetsky-Shapiro e Smyth (1996), o KDD refere-se ao processo completo de descoberta de conhecimento, e a mineração de dados é uma das atividades do processo. O KDD pode ainda ser definido como o processo de extração de informação a partir de algum tipo de banco de dados (estruturado ou não), possibilitando um conhecimento, previamente desconhecido, potencialmente útil e compreensível (Cardoso; Machado, 2008).

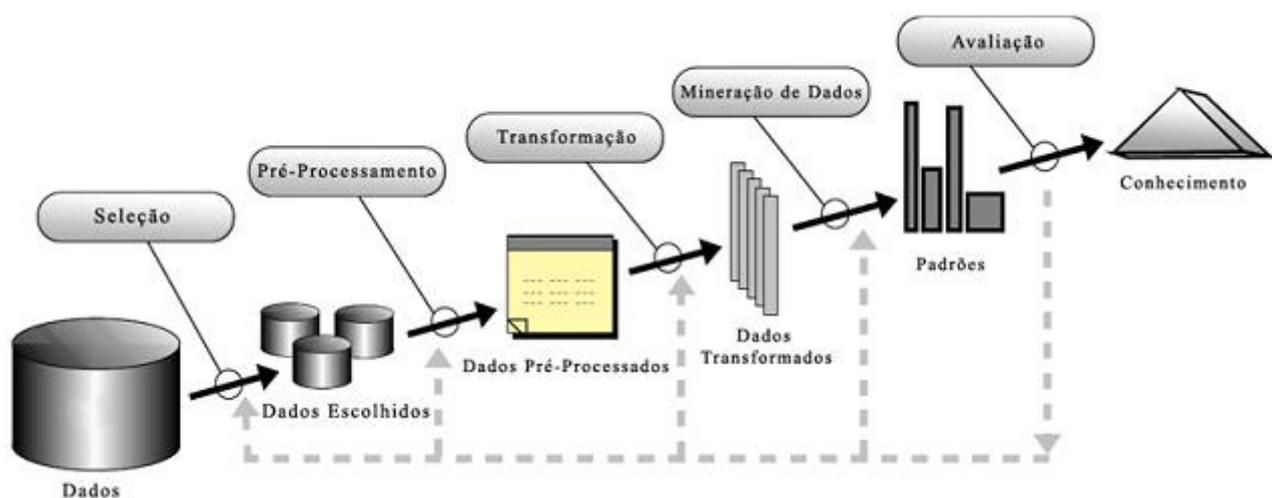


Figura 1 - Processo KDD

Dentro da etapa de mineração de dados, existem diversos processos para padronizar fases e atividades. Segundo Larose (2005) e Hand, Mannila e Smyth (2001), devido à grande quantidade de literatura disponível, o CRISP – DM (Cross-Industry Standard Process of Data Mining) pode ser considerado o padrão de maior aceitação.

Conforme Chapman et al. (2000), as fases do processo CRISP-DM são:

- **Compreensão dos negócios:** nesse ponto discute-se até definir a intenção ou objetivo que se deseja alcançar com a mineração de dados. Essa fase é fundamental para as etapas seguintes.
- **Compreensão dos dados:** a intenção nessa etapa é observar com cuidado os dados. Na observação dos dados, que pode abranger técnicas de agrupamento e de exploração visual, é fundamental identificar aqueles que são relevantes para o problema em questão, certificando-se de que as variáveis relevantes não são interdependentes (Olson; Delen, 2008).
- **Preparação dos dados:** nessa etapa, os dados, que podem ser provenientes de várias fontes e possuir formatos diversos, são preparados para que os métodos de mineração de dados possam ser aplicados. São realizados procedimentos para filtrar, combinar e especificar valores vazios em variáveis nulas (missing) que podem ter tratamentos diferenciados.
- **Modelagem:** o cerne do processo de mineração é nessa etapa. Nesse ponto, as técnicas ou algoritmos de mineração são escolhidos e configurados de acordo com os dados selecionados, dependendo dos objetivos desejados, e aplicados (McCue, 2007).
- **Avaliação:** nessa etapa a participação dos gestores da administração pública – conhecedores do domínio ou conhecedores de negócio – é fundamental para o êxito no processo de mineração. Os resultados, que podem ser colocados de forma gráfica ou agrupada, podem não fazer muito sentido para um especialista em mineração de dados, mas para o conhecedor do negócio talvez faça bastante sentido. O especialista em mineração precisa fazer testes e validações e adaptar o modelo, se for o caso.

- **Aplicação:** mesmo que o propósito do modelo seja o aumento do conhecimento sobre os dados, tal conhecimento obtido precisa ser organizado e apresentado de uma forma útil e que possibilite alteração de procedimentos ou processos. Dependendo da situação, a fase de implantação pode ser tão simples quanto gerar um relatório ou tão complexa quanto implementar um processo de mineração de dados repetível em outras áreas, processos, contratos etc. Na mineração de dados, são definidas as tarefas e as técnicas, que serão utilizadas de acordo com os objetivos do estudo, a fim de obter uma resposta para o problema (Matos; Chalmeta; Coltell, 2006). As tarefas podem ser preditivas, ou seja, buscam apontar ou predizer o valor de um atributo baseado nos valores de outros atributos, ou ainda descritivas, que buscam derivar padrões.

## 1.1 Tarefas da mineração de dados

Segundo Larose (2005), as tarefas são classificadas pelo seu potencial de realização, que consiste na especificação do que se deseja buscar nos dados, ou pelo interesse de encontrar uma categoria de padrões.

A seguir as tarefas mais comuns:

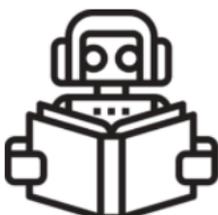
- **Classificação (preditiva):** visa demonstrar a qual classe um determinado registro ou instância pertence. É o processo de encontrar um conjunto de modelos (funções) que descrevem e distinguem classes ou conceitos. Segundo Witten, Frank e Hall (2011), é separar e conquistar, porque a tarefa de classificação identifica uma regra que abrange instâncias em uma classe (excluindo as instâncias que não estão na classe). Boa parte dos métodos de classificação utilizam técnicas estatísticas e de aprendizado de máquina. Um exemplo seria a intenção da gestão de um mercado em descobrir quais clientes teriam características de “bom comprador” ou “mau comprador”. Um modelo de classificação poderia incluir a seguinte regra: “clientes da faixa econômica B, com idade entre 50 e 60 são maus compradores”.

- **Agrupamento (descritiva):** um agrupamento (ou cluster) é uma coleção de registros similares entre si, porém diferentes dos outros registros nos demais agrupamentos. Berkhin (2002) definiu cluster como sendo uma divisão de dados em grupos de objetos semelhantes. Representar dados por meio de clusters pode significar perda de certos detalhes finos (semelhante à compactação de dados com perdas), mas alcançar a simplificação. Essa tarefa se difere da classificação porque não necessita que os registros sejam previamente categorizados (aprendizado não-supervisionado). O agrupamento não tem a pretensão de classificar, estimar ou prever o valor de uma variável, mas apenas identificar os grupos de dados similares. Um exemplo seria a aplicação para um estudo de mercado buscando identificar segmentos de mercado para um produto específico.
- **Associação (descritiva):** consiste em identificar quais instâncias estão relacionadas. É uma tarefa muito comum e utilizada em Market Basket (cesta de compras). Consideremos um mercado novamente e a intenção de entender o padrão de comportamento dos clientes, com a finalidade de descobrir o que é levado numa mesma compra por clientes para que o estabelecimento possa melhorar a organização das prateleiras e facilitar tal conjunto de compras para potencializar lucros. Um dos exemplos mais clássicos e populares de mineração de dados envolve justamente a tarefa de associação, realizada em uma das maiores redes de varejo dos Estados Unidos (o Walmart). Foi descoberto, no gigantesco armazém de dados dessas redes de varejo, 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. Os produtos foram postos próximos um do outro e o resultado foi o aumento significativo da venda de fraldas e de cervejas (Gurovitz, 2011).
- **Regressão (preditiva):** consiste numa tarefa que tem o intuito de prever o valor de uma variável contínua baseada em outras variáveis, assumindo um modelo de dependência linear ou não linear. Geralmente se utilizam abordagens estatísticas ou de redes neurais. Um exemplo simples de predição seria prever a velocidade do vento em função da temperatura, umidade, pressão atmosférica etc.

- **Sumarização (descritiva):** consiste numa tarefa que possibilita a identificação de uma descrição coerente e inteligível para os dados (ou para um subconjunto deles). Em diversas vezes é possível sumarizar os dados mesmo com alguma imprecisão, e o valor das técnicas de sumarização consiste na capacidade de descrever os dados e não necessariamente em sua precisão. É possível sumarizar os dados de uma base de dados através de tarefas de classificação, porém nem toda tarefa de classificação cria modelos que descrevem os dados de maneira que sejam facilmente interpretados.
- **Detecção de anomalias ou outliers (descritiva):** essa tarefa visa detectar desvios de comportamento, facilmente encontrados quando tais desvios são significativos nas análises. Essa tarefa é utilizada por instituições financeiras para identificar fraudes e ainda por administradores de redes da área de tecnologia da informação (TI) para localizar possíveis intrusões.

Fonte de pesquisa: MINERAÇÃO DE DADOS: UM OLHAR INSTIGANTE DE POSSIBILIDADES E APLICAÇÕES PARA ÓRGÃOS DA ADMINISTRAÇÃO PÚBLICA FEDERAL parte da Revista do Serviço Público (RSP), Brasília 73(3) 451–478 jul/set 2022. <https://repositorio.enap.gov.br/bitstream/1/7443/1/5446-Texto%20do%20Artigo-25502-1-10-20220928.pdf> acessado em janeiro de 2024.

## 2 APRENDIZADO DE MÁQUINA



Nos outros módulos temos dois exemplos que podem ser utilizados para elucidar uma parte da teoria em relação a aprendizado de máquina:

Filtro antispam: o servidor de email começa a “aprender sozinho” que determinadas empresas estão sendo apontadas como SPAM por vários dos seus usuários de e-mail. O filtro então aprende que deve reportar as próximas mensagens desta empresa como SPAM para todos os clientes.

Heurística: a regra do cenário de antivírus é que, antes da vacina, surjam os vírus. O antivírus, através o método de heurística, aprende o comportamento normal dos programas daquela máquina. Quando algum programa começa a agir de forma anômala, o antivírus tenta fazer uma varredura mas não tem a vacina. Por via das dúvidas o antivírus envia esse programa para a quarentena e avisa o usuário que aquele programa está “doidão” mas não tem certeza se é um vírus pois não tem a vacina.

O aprendizado de máquina (ou Machine Learning) é uma parte da inteligência artificial onde o computador, através de algoritmos e técnicas complexas, aprende a executar determinadas tarefas.

A constante atualização dos hardwares em relação à velocidade, capacidade e eficiência aliado ao barateamento das tecnologias influenciam muito no aprendizado de máquina. Basicamente as máquinas utilizam algoritmos para capturar os dados e aprender com estes, levando em consideração o histórico destes dados para uma tomada de decisão ou execução de uma tarefa.

## 2.1 Métodos para o Aprendizado de Máquina

Existem, basicamente, quatro métodos para que as máquinas aprendam a partir dos dados coletados.

**Aprendizado supervisionado**: nesse método o “treinamento” da máquina é feito com dados de entrada e saída conhecidos. Comparado ao ensinamento do ser humano por um tutor que baliza o que é certo e errado. Os dados possuem rótulos e nestes estão descritos o que é possível fazer. Os dados que dão entrada no sistema já sabem o que deve ser feito para a execução final (saída). Exemplo: quando o sistema detecta uma possível fraude no uso do cartão de crédito de uma pessoa. Quando foge do “padrão” automaticamente o sistema antifraude é acionado.

Quando tentamos prever uma variável dependente a partir de variáveis independentes.

Variáveis Independentes	Variáveis Dependentes
Formação e Experiência	Salário
Idade e qualidade de vida	Propensão a doença
Histórico escolar	Nota de prova

As técnicas mais conhecidas para resolver problemas de aprendizado supervisionado são:

- Regressão linear
- Regressão Logística
- Redes neurais
- Árvores de decisão

**Aprendizado não supervisionado:** ao contrário do anterior, aqui os dados não possuem rótulos. Seria um aprendizado sem tutor, autodidata. Nesse caso o sistema deve procurar nos dados coletados uma possível saída para os dados de entrada. Um exemplo do nosso mundo é a classificação de pessoas que tendem a adquirir alguma doença a partir de determinadas atitudes (alimentação, nível de stress, e tc.).

No caso do aprendizado não supervisionado não existe um padrão associado. Ex: consumidores que compram queijos e vinhos ou carne e carvão ou leite e fralda. Não dá para saber quantos perfis existem. Nesse caso, o computador terá que descobrir os perfis sem dados anotados e precisaremos de métodos de aprendizado não supervisionados. Uma opção seria observar nos registros de compras se existem padrões repetidos e que permitiriam a inferência de um grupo ou perfil de consumidor. Outra opção seria ver diretamente quais produtos são frequentemente comprados juntos e, então, aprender uma regra associativa entre eles.

Nesse caso precisamos associar os dados presentes com uma representação.

Dados	Representação
Transações de cartão	Normalidade da transação
Compras	Associar produtos
Registro de compras	Perfil de consumidor

Outros exemplos de aplicações de aprendizado não supervisionados são sistemas de recomendação de filmes ou músicas, detecção de anomalias e visualização de dados. Dentre as técnicas mais conhecidas para resolver problemas de aprendizado não supervisionado estão redes neurais artificiais, Expectativa-Maximização, clusterização k-médias, máquina de suporte vetorial (ou máquinas kernel).

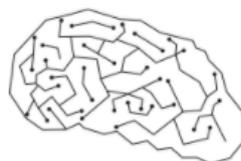
Existem outros dois métodos que não são muito utilizados:

- Aprendizado semi supervisionado
- Aprendizado por reforço

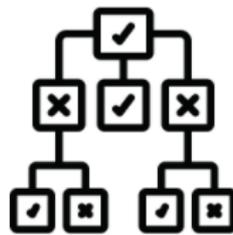
## 2.2 Técnicas de Aprendizado de Máquina

Existem várias técnicas de aprendizado de máquina que já foram citados aqui. Vou descrever algumas:

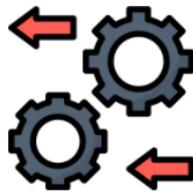
**Redes Neurais:** são baseadas em como o cérebro humano funciona. Existem diferentes tipos de Redes Neurais, mas basicamente elas consistem em um conjunto de nós (ou neurônios) dispostos em várias camadas com interconexões ponderadas entre eles. Cada neurônio combina um conjunto de valores de entrada para produzir um valor de saída, que por sua vez é passado para outros neurônios nas camadas seguintes.



**Árvore de Decisão:** essa técnica utiliza algoritmos do formato de árvore. Cada árvore possui nó, ramos, folhas e ligações entre eles. Um ramo é um conjunto de nó que testa cada atributo da classificação e cada ramo corresponde ao valor do atributo. E as folhas atribuem uma classificação para os ramos. A classificação é o resultado que se quer chegar ou prever com o algoritmo.



**Regressão:** é a técnica utilizada quando se quer prever valores de algo futuro, baseado em comportamento de variáveis passadas. Um exemplo seria prever o valor de mercado de um imóvel utilizando um algoritmo de regressão linear.



## 2.3 Métricas

Algumas métricas interessantes, para identificar erros e acertos, serão descritas logo abaixo:

**Acurácia:** essa métrica é a mais básica para medir o aprendizado, leva em consideração a proporção de predições corretas, sem levar em consideração a quantidade de erros.

Sensibilidade: é a proporção de verdadeiros positivos, ou seja, a capacidade do algoritmo em prever casos corretamente para os casos que são realmente verdadeiros.

Especificidade: A proporção de verdadeiros negativos: a capacidade do algoritmo em prever situações erradas que realmente são falsas.

Eficiência: é a média aritmética da sensibilidade e especificidade. Indica se o algoritmo está mais suscetível a ter verdadeiros positivos ou verdadeiros negativos.

## 2.4 Etapas

O processo de aprendizado de máquina pode ser dividido em 7 etapas:

1. Coleta de dados: depois da escolha do hardware a etapa de coleta de dados é crucial para o resultado. Deve levar em consideração a quantidade e a qualidade dos dados coletados.
2. Preparação dos dados: verificar se as informações são bem distribuídas ou se são tendenciosas.
3. Escolha do modelo: há uma gama muito grande de modelos de aprendizado de máquina. Os mais utilizados são:
  - a. Classificação binária: faz a máquina perguntar se sim ou se não. O algoritmo utilizado é o de regressão logística. Ex: O e-mail recebido é ou não é SPAM?
  - b. Classificação multiclasse: a máquina pode prever mais de dois resultados. Ex: consegue determinar se o produto é eletrônico, medicamento etc.
  - c. Classificação por regressão: usam o algoritmo de regressão linear e conseguem prever um valor numérico. Ex: quantos produtos serão vendidos.

4. Treinamento: etapa fundamental para preparar a máquina e aprimorar suas habilidades de previsão. A máquina aprende com seus erros e se aperfeiçoa. Considerado o pilar do aprendizado de máquina.
5. Avaliação: efetua testes para verificar se a máquina aprendeu e se está capacitada em um contexto real.
6. Aprimoramento dos parâmetros: visa melhorar a qualidade e a eficiência do modelo que foi escolhido para ser utilizado. É nessa etapa que são verificados os valores que afetam a acurácia e o tempo de treinamento.
7. Predição: é a fase em que a máquina pode ser usada para responder ao que foi treinada.

### 3 QUESTÕES DE RENDIMENTO

#### 01 (CEBRASPE | 2021 | PF | ESCRIVÃO)

Acerca dos conceitos de mineração de dados, aprendizado de máquina e bigdata, julgue o próximo item.

A análise de clustering é uma tarefa que consiste em agrupar um conjunto de objetos de tal forma que estes, juntos no mesmo grupo, sejam mais semelhantes entre si que em outros grupos.

#### Resolução

A análise de clustering, ou agrupamento, é uma tarefa fundamental em mineração de dados. Essa técnica tem como objetivo agrupar um conjunto de objetos de modo que os membros de um mesmo grupo sejam mais semelhantes entre si do que com os membros de outros grupos.

O algoritmo de clustering busca identificar padrões intrínsecos nos dados, agrupando instâncias que compartilham características comuns. Este processo é não supervisionado, o que significa que o algoritmo não requer rótulos prévios para os grupos, diferenciando-se assim do aprendizado supervisionado.

Um banco pode usar a análise de clustering para agrupar seus clientes em grupos com base em seu risco de crédito. Isso pode ajudar o banco a tomar decisões mais informadas sobre a concessão de crédito. **CERTA**

## 02 (CEBRASPE | 2018 | PF | AGENTE)

Julgue o item que segue, relativo a noções de mineração de dados, big data e aprendizado de máquina.

Pode-se definir mineração de dados como o processo de identificar, em dados, padrões válidos, novos, potencialmente úteis e, ao final, compreensíveis.

### **Resolução**

Mineração de dados pode ser definida como o processo de explorar grandes conjuntos de dados para identificar padrões válidos, novos e potencialmente úteis, que, ao final, sejam compreensíveis. O objetivo principal da mineração de dados é extrair informações valiosas a partir dos dados, revelando relações, tendências ou padrões que podem não ser aparentes em uma análise superficial.

Para atingir esse objetivo, diversas técnicas são aplicadas, incluindo algoritmos de aprendizado de máquina, análise estatística e métodos de visualização de dados. A compreensibilidade dos padrões identificados é crucial para que os resultados sejam úteis e interpretáveis pelos usuários finais, permitindo que tomem decisões informadas com base nas descobertas da mineração de dados.

Os padrões descobertos pela mineração de dados devem ser válidos, ou seja, devem ser consistentes com os dados. Também devem ser novos, ou seja, devem ser diferentes dos padrões que já são conhecidos. Além disso, os padrões devem ser potencialmente úteis, ou seja, devem ter algum valor para o usuário. Por fim, os padrões devem ser compreensíveis, ou seja, devem ser fáceis de entender e interpretar.

Para ilustrar: vamos considerar um exemplo hipotético envolvendo dados de compras online. Suponha que tenhamos um conjunto de dados que contenha informações sobre transações, como produtos comprados, valor gasto, horário da compra e localização do cliente. Aqui estão alguns padrões válidos, novos e potencialmente úteis que poderiam ser identificados por meio da mineração de dados:

1. Padrão Válido: Clientes que compram frequentemente certos produtos durante as promoções de final de semana.

2. Padrão Novo: Uma tendência emergente de clientes que começaram a comprar online em determinada região geográfica após o lançamento de uma campanha publicitária específica.

3. Padrão Potencialmente Útil: Identificação de associações entre a compra de certos produtos e a faixa etária dos clientes, permitindo personalizar ofertas com base nessas informações.

A mineração de dados nesse contexto poderia revelar esses padrões, fornecendo insights valiosos para estratégias de marketing, personalização de ofertas e otimização da experiência do cliente. Esses padrões podem não ser imediatamente óbvios, mas a análise dos dados pode revelar relações valiosas que beneficiam o negócio. **CERTA**

### **03 (CEBRASPE | 2018 | PF | AGENTE)**

Julgue o item que segue, relativo a noções de mineração de dados, big data e aprendizado de máquina.

Situação hipotética: Na ação de obtenção de informações por meio de aprendizado de máquina, verificou-se que o processo que estava sendo realizado consistia em examinar as características de determinado objeto e atribuir-lhe uma ou mais classes; verificou-se também que os algoritmos utilizados eram embasados em algoritmos de aprendizagem supervisionados.

Assertiva: Nessa situação, a ação em realização está relacionada ao processo de classificação.

#### **Resolução**

A classificação é uma tarefa comum em aprendizado de máquina supervisionado, onde o modelo é treinado com um conjunto de dados rotulado (ou seja, um conjunto de dados em que as classes já são conhecidas) para aprender a associar corretamente as

características dos objetos às suas classes correspondentes. Uma vez treinado, o modelo pode ser utilizado para classificar novos objetos com base nas características aprendidas durante o treinamento.

Alguns exemplos de como a classificação pode ser usada:

- Classificação de imagens: Um algoritmo de classificação pode ser usado para identificar o conteúdo de uma imagem, como um rosto, um objeto ou uma cena.
- Classificação de texto: Um algoritmo de classificação pode ser usado para identificar o tópico de um texto, como notícias, finanças ou entretenimento.
- Classificação de e-mail: Um algoritmo de classificação pode ser usado para identificar se um e-mail é spam ou não. **CERTA**

#### 04 (CEBRASPE | 2018 | PF | ESCRIVÃO)

CPF

NOME

DATA DE NASCIMENTO

NOME DO PAI

NOME DA MAE

TELEFONE

CEP

NUMERO

As informações anteriormente apresentadas correspondem aos campos de uma tabela de um banco de dados, a qual é acessada por mais de um sistema de informação e também por outras tabelas. Esses dados são utilizados para simples cadastros, desde a consulta até sua alteração, e também para prevenção à fraude, por meio de verificação dos dados da tabela e de outros dados em diferentes bases de dados ou outros meios de informação.

Considerando essas informações, julgue o item que segue.

Se um sistema de informação correlaciona os dados da tabela em questão com outros dados não estruturados, então, nesse caso, ocorre um processo de mineração de dados.

 **Resolução**

A correlação entre dados estruturados e não estruturados é uma estratégia na mineração de dados. No cenário descrito, os dados estruturados, organizados na tabela com campos como CPF, nome, telefone, etc., são apenas uma parte do panorama. Para obter insights mais profundos e descobrir padrões complexos, é necessário correlacionar esses dados estruturados com informações não estruturadas.

Os dados não estruturados, como comentários em redes sociais ou outras formas de feedback do cliente, podem conter nuances e contextos valiosos que os dados estruturados por si só não conseguem capturar. Ao realizar essa correlação, o sistema de informação pode identificar padrões de comportamento que são indicativos de possíveis fraudes.

Por exemplo, considerando o contexto da correlação entre os dados da tabela e os dados de mídia social, o sistema pode descobrir que uma determinada combinação de informações, como um padrão consistente de mudanças no número de telefone em conjunto com reclamações específicas em redes sociais, pode sugerir atividades fraudulentas. Essa análise integrada de dados estruturados e não estruturados pode ser fundamental para identificar padrões complexos e melhorar os mecanismos de prevenção à fraude.

Suponha que o sistema de um banco observe um aumento repentino nas transações de um determinado cliente que não segue o padrão usual de seu histórico de compras. O cliente, cujos dados estruturados indicam que ele reside em uma cidade específica, começa a realizar compras em locais e horários atípicos.

A correlação com dados não estruturados pode revelar que esse cliente está publicando em redes sociais comentários como "perdi meu cartão" ou "preciso de um cartão de crédito urgente". Esses dados não estruturados, quando correlacionados com as transações atípicas, sugerem um comportamento fora do comum e podem ser indicativos de uma possível fraude.

**Análise:**

- **Padrão de Compra:** O sistema detecta um padrão incomum de compras em locais não visitados anteriormente pelo cliente e em horários fora do padrão.

- **Correlação com Dados Não Estruturados:** A correlação com dados de mídia social revela que o cliente postou recentemente sobre a perda do cartão e a necessidade urgente de um novo.

**CERTA.**

### **05 (CEBRASPE | 2018 | PF | ESCRIVÃO)**

Em um big data, alimentado com os dados de um sítio de comércio eletrônico, são armazenadas informações diversificadas, que consideram a navegação dos usuários, os produtos comprados e outras preferências que o usuário demonstre nos seus acessos. Tendo como referência as informações apresentadas, julgue o item seguinte.

Uma aplicação que reconheça o acesso de um usuário e forneça sugestões diferentes para cada tipo de usuário pode ser considerada uma aplicação que usa machine learning.

#### **Resolução**

Nesse contexto, o machine learning pode ser aplicado para analisar padrões de comportamento do usuário, suas interações anteriores, produtos comprados e outras preferências para prever e sugerir itens relevantes. Existem várias técnicas de machine learning, como algoritmos de recomendação, clustering ou até mesmo modelos mais avançados, que podem ser empregadas para melhorar a personalização das sugestões ao longo do tempo.

Essa abordagem é comum em sites de comércio eletrônico, onde a capacidade de oferecer recomendações personalizadas com base no comportamento do usuário contribui para uma experiência mais relevante e satisfatória. **CERTA.**

### 06 (CEBRASPE | 2018 | PF | Escrivão)

Em um big data, alimentado com os dados de um sítio de comércio eletrônico, são armazenadas informações diversificadas, que consideram a navegação dos usuários, os produtos comprados e outras preferências que o usuário demonstre nos seus acessos. Tendo como referência as informações apresentadas, julgue o item seguinte.

Dados coletados de redes sociais podem ser armazenados, correlacionados e expostos com o uso de análises preditivas.

#### **Resolução**

As análises preditivas são um tipo de análise de dados que usa modelos matemáticos para **prever o comportamento futuro**. Esses modelos podem ser usados para prever uma variedade de coisas, como tendências de mercado, comportamentos dos clientes ou eventos futuros.

Os dados coletados de redes sociais podem ser usados para alimentar esses modelos. Esses dados podem incluir informações como postagens, curtidas, comentários, compartilhamentos e hashtags.

Ao analisar esses dados, as análises preditivas podem identificar padrões e tendências que podem ser usados para prever o comportamento futuro.

Por exemplo, uma empresa pode usar análises preditivas para prever quais produtos são mais propensos a serem comprados por seus clientes. Isso pode ajudar a empresa a direcionar seus anúncios de forma mais eficaz e a aumentar suas vendas. **CERTA.**

### 07 (CEBRASPE | 2018 | PF | Perito)

Acerca de banco de dados, julgue o seguinte item.

A mineração de dados se caracteriza especialmente pela busca de informações em grandes volumes de dados, tanto estruturados quanto não estruturados, alicerçados no

conceito dos 4V's: volume de mineração, variedade de algoritmos, velocidade de aprendizado e veracidade dos padrões.

### **Resolução**

O conceito dos 4Vs da mineração de dados se refere especificamente a Big Data, e não a mineração de dados em geral.

Os "4V's" no contexto do Big Data:

**1. Volume:**

- Refere-se à enorme quantidade de dados gerados e armazenados. No Big Data, lidamos com conjuntos de dados muito grandes, que podem chegar a petabytes ou além. Essa grande escala de dados cria desafios em termos de armazenamento, processamento e análise eficiente.

**2. Variedade:**

- Indica a diversidade de tipos de dados. Além dos dados estruturados encontrados em bancos de dados tradicionais, o Big Data lida com dados não estruturados, semiestruturados e complexos, como texto, imagens, vídeos, dados de sensores, entre outros.

**3. Velocidade:**

- Refere-se à taxa na qual os dados são gerados, processados e disponibilizados para análise. No contexto do Big Data, a velocidade muitas vezes significa a capacidade de lidar com dados em tempo real ou quase em tempo real, garantindo que as análises sejam feitas rapidamente para extrair insights valiosos.

**4. Veracidade:**

- Diz respeito à confiabilidade e qualidade dos dados. Garantir a precisão, integridade e confiabilidade dos dados é crucial no Big Data, pois a tomada de decisões baseada em dados incorretos pode levar a resultados inadequados.

Além desses "4V's", alguns contextos adicionam outros "V's", como Valor (a capacidade de obter valor significativo a partir dos dados) e Vulnerabilidade (a necessidade de proteger os dados contra acessos não autorizados).

Em resumo, os "4V's" são princípios fundamentais no contexto do Big Data, destacando os desafios associados à gestão e análise de grandes volumes de dados diversificados, em tempo real e com garantia de qualidade.

**ERRADA.**

### **08 (CEBRASPE | 2023 | TBG | Analista de Comercialização)**

Acerca de mineração de dados (data mining), julgue o item a seguir.

A mineração de dados não se limita às técnicas de obtenção de dados, envolvendo também a observação de padrões nos dados obtidos.

#### **Resolução**

A mineração de dados é um processo de descoberta de padrões e informações em grandes conjuntos de dados. Esse processo envolve uma série de etapas, incluindo:

**Obtenção** de dados: Os dados podem ser obtidos de uma variedade de fontes, como bancos de dados, arquivos, sensores e mídias sociais.

**Preparação** de dados: Os dados precisam ser preparados para análise, o que pode incluir limpeza, transformação e redução de dimensionalidade.

**Mineração** de dados: As técnicas de mineração de dados são usadas para identificar padrões e informações nos dados preparados.

**Interpretação** dos resultados: Os resultados da mineração de dados precisam ser interpretados e validados para garantir que sejam significativos e úteis.

Portanto, a mineração de dados não se limita às técnicas de obtenção de dados. Ela também envolve a observação de padrões nos dados obtidos.

**CERTA.**

## 09 (CEBRASPE | 2021 | TCE-RJ | Analista de Controle Externo)

Com relação a noções de mineração de dados e Big Data, julgue o item que se segue.

As regras de associação adotadas em mineração de dados buscam padrões frequentes entre conjuntos de dados e podem ser úteis para caracterizar, por exemplo, hábitos de consumo de clientes: suas preferências são identificadas e em seguida associadas a outros potenciais produtos de seu interesse.

### **Resolução**

As regras de associação na mineração de dados são utilizadas para identificar padrões frequentes em conjuntos de dados, e são especialmente úteis para caracterizar hábitos de consumo, como as preferências dos clientes. Veja:

#### 1. **Regras de Associação:**

- As regras de associação são utilizadas para descobrir relações significativas entre variáveis em grandes conjuntos de dados. Elas identificam padrões frequentes, como associações entre itens em transações de compras.

#### 2. **Hábitos de Consumo:**

- Ao aplicar regras de associação a dados de compras, é possível identificar quais produtos são frequentemente adquiridos juntos. Por exemplo, se muitos clientes que compram café também compram leite, uma regra de associação pode ser gerada indicando essa relação.

#### 3. **Caracterização de Preferências:**

- Com base nas regras de associação, é possível caracterizar as preferências dos clientes. Se uma pessoa compra regularmente produtos A e B, pode-se inferir que ela tem preferência por esses itens. Essa informação pode ser valiosa para personalizar ofertas ou recomendações.

#### 4. **Identificação de Potenciais Produtos:**

- Além de caracterizar as preferências, as regras de associação também podem ser utilizadas para identificar potenciais produtos de interesse para os clientes. Se, por exemplo, a regra mostra que clientes que compram

livros frequentemente também compram e-readers, isso sugere que e-readers podem ser de interesse para outros clientes que compram livros. Dessa forma, as regras de associação são uma ferramenta valiosa na mineração de dados para entender e caracterizar padrões de comportamento, especialmente em cenários de consumo, onde a identificação de hábitos e preferências é fundamental para estratégias de negócios, marketing e recomendações personalizadas.

**CERTA.**

### 10 (CEBRASPE | 2022 | PC-PB | Escrivão)

A coleta de dados que serão garimpados, na mineração de dados, é feita na etapa de :

- a) aplicação.
- b) análise.
- c) priorização.
- d) mineração.
- e) preparação.

#### **Resolução**

A mineração de dados é um processo de descoberta de padrões e informações em grandes conjuntos de dados. Esse processo envolve uma série de etapas, incluindo:

**Obtenção** de dados: Os dados podem ser obtidos de uma variedade de fontes, como bancos de dados, arquivos, sensores e mídias sociais.

**Preparação** de dados: Os dados precisam ser preparados para análise, o que pode incluir limpeza, transformação e redução de dimensionalidade.

**Mineração** de dados: As técnicas de mineração de dados são usadas para identificar padrões e informações nos dados preparados.

**Interpretação** dos resultados: Os resultados da mineração de dados precisam ser interpretados e validados para garantir que sejam significativos e úteis.

Portanto, a coleta de dados é feita na etapa de preparação, que é a primeira etapa do processo de mineração de dados. A coleta de dados pode ser feita a partir de uma variedade de fontes, como bancos de dados, arquivos, sensores e mídias sociais. Os dados podem ser coletados de forma manual ou automática.

Após a coleta, os dados precisam ser preparados para análise. Isso pode incluir limpeza, transformação e redução de dimensionalidade.

A limpeza dos dados consiste em remover erros e inconsistências dos dados. A transformação dos dados consiste em converter os dados para um formato que seja mais adequado para análise. A redução de dimensionalidade consiste em reduzir o tamanho dos dados sem perder informações importantes.

Após a preparação, os dados estão prontos para serem analisados.

**GABARITO E.**

### **11 (CEBRASPE | 2021 | PCDF | Agente)**

Com relação a mineração de dados, aprendizado de máquina e aplicações Python, julgue o item a seguir.

Uma das aplicações de Python é o aprendizado de máquina, que pode ser exemplificado por um programa de computador que aprende com a experiência de detectar imagens de armas e de explosivos em vídeos, tendo seu desempenho medido e melhorado por meio dos erros e de acertos decorrentes da experiência de detecção.

#### **Resolução**

O aprendizado de máquina é uma área da inteligência artificial que permite que os sistemas aprendam e melhorem de forma autônoma, sem serem explicitamente programados.

Python é uma linguagem de programação de alto nível que é amplamente usada para aprendizado de máquina. Python é uma linguagem de fácil aprendizado e possui uma ampla biblioteca e ferramentas de aprendizado de máquina.

O exemplo dado na afirmativa é um exemplo de como o aprendizado de máquina pode ser usado para detectar armas e explosivos em vídeos. O programa de computador seria treinado em um conjunto de dados de imagens de armas e explosivos. Essas imagens seriam marcadas como "arma" ou "explosivo". O programa de computador então usaria essas informações para aprender a detectar armas e explosivos em novos vídeos.

O desempenho do programa de computador seria medido e melhorado por meio dos erros e acertos decorrentes da experiência de detecção. Por exemplo, se o programa de computador detectar erroneamente uma imagem de um gato como uma arma, isso seria considerado um erro. O programa de computador então usaria essa informação para melhorar sua capacidade de diferenciar gatos de armas.

**CERTA.**

---

---

## 12 (CEBRASPE | 2021 | PCDF | Agente)

Com relação a mineração de dados, aprendizado de máquina e aplicações Python, julgue o item a seguir.

A detecção de novos tipos de fraudes é uma das aplicações comuns da técnica de modelagem descritiva da mineração de dados, a qual viabiliza o mapeamento rápido e preciso de novos tipos de golpes por meio de modelos de classificação de padrões predefinidos de fraudes.

### **Resolução**

A afirmativa está errada porque a modelagem descritiva é uma técnica de mineração de dados que busca identificar padrões e tendências nos dados existentes. Ela não é adequada para identificar novos tipos de fraudes, pois esses tipos de fraudes ainda não existem e, portanto, não podem ser detectados nos dados existentes.

Para identificar novos tipos de fraudes, é necessário usar uma técnica de mineração de dados chamada de modelagem preditiva. A modelagem preditiva usa modelos

matemáticos para prever o comportamento futuro. Esses modelos podem ser usados para prever a probabilidade de novos tipos de fraudes ocorrerem.

**ERRADA.**





## **CONCURSEIRO QUE PRETENDE SER POLICIAL NÃO FAZ RATEIO**

Todo o material desta apostila (textos e imagens) está protegido por direitos autorais do Profissão Policial Concursos de acordo com a Lei 9.610/1998. Será proibida toda forma de cópia, plágio, reprodução ou qualquer outra forma de uso, não autorizada expressamente, seja ela onerosa ou não, sujeitando-se o transgressor às penalidades previstas civil e criminalmente.