

**COLETÂNEA CONCURSOS PÚBLICOS**



# **GLOSSÁRIO DO EDITAL DO CONCURSO DA SUSEP**

**CONHECIMENTOS ESPECÍFICOS PARA O CARGO 4: ANALISTA TÉCNICO –  
ÁREA: TECNOLOGIA DA INFORMAÇÃO E CIÊNCIA DE DADOS**



**COM QUESTÕES INÉDITAS COMENTADAS  
PÓS-EDITAL 2025**

## APRESENTAÇÃO

A ideia central deste ebook preparatório para o concurso da Superintendência de Seguros Privados (SUSEP) é que o estudante possa aprender de forma rápida todo o conteúdo relacionado a tecnologia da informação (TI) do edital, portanto, para um maior aprofundamento em cada assunto é necessário utilizar materiais complementares.

Este ebook foi elaborado com auxílio de inteligência artificial e revisado e complementado pelo autor Izaias Batista dos Santos.

Siga a Kuasarnex nas redes sociais para receber notificações sobre conteúdos, cupons e oportunidades.



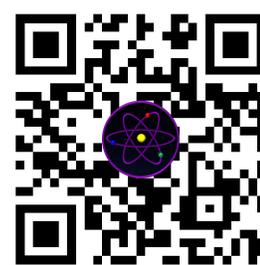
[Seguir no Instagram](#)



[Inscrição no Canal](#)



[Seguir no LinkedIn](#)



[Site Oficial](#)

**AUTOR: IZAIAS BATISTA DOS SANTOS**

Olá, seja muito bem-vindo(a)!

Sou Izaias Batista dos Santos, autor deste material preparatório especialmente desenvolvido para auxiliar você em sua jornada rumo à aprovação no concurso da SUSEP. Tenho orgulho de compartilhar um pouco da minha trajetória com você, pois acredito que conhecer quem está por trás dos conteúdos reforça a confiança no estudo e nos resultados.

Sou mestre em **Tecnologias Computacionais para o Agronegócio** pela **Universidade Tecnológica Federal do Paraná (UTFPR)**, especialista em **Engenharia de Software** pela **Pontifícia Universidade Católica de Minas Gerais (PUC-Minas)**, possuo **MBA em Gerenciamento de Projetos de TI** pelo **Instituto de Gestão em Tecnologia da Informação (IGTI)** e sou bacharel em **Sistemas de Informação** pelo **Centro Universitário Dinâmica das Cataratas (UDC)**.

Minha trajetória profissional foi construída com dedicação e propósito em diversas áreas da Tecnologia da Informação. Atuei como **coordenador de TIC no Exército Brasileiro**, **analista de sistemas sênior** na **Fundação Parque Tecnológico Itaipu**, **coordenador de projetos de software** na empresa **Eits Prognus Software Livre** e também fui **professor substituto de informática e suas tecnologias** e **técnico em tecnologia da informação e comunicações** no **Instituto Federal do Paraná**.

Atualmente, sou **analista de sistemas e processos** no **Conselho Federal de Química**, onde atuo com **gestão por processos** e exerço a função de **gestor técnico de sistemas**, sempre buscando alinhar boas práticas, inovação e eficiência para garantir resultados sustentáveis e soluções de alto impacto.

Com base em minha experiência acadêmica e profissional, preparei este e-book para oferecer a você um conteúdo claro, atualizado e voltado aos temas mais cobrados nos concursos públicos da área de Tecnologia da Informação. Meu objetivo é ajudá-lo(a) a dominar os assuntos e encurtar o caminho até a aprovação.

E, nos momentos em que bater a dúvida sobre sua capacidade de aprender e vencer esse desafio, lembre-se sempre da poderosa frase:

**“Tudo posso naquele que me fortalece!”**

Ah, e caso queira acompanhar mais conteúdos, dicas e novidades sobre concursos públicos e Tecnologia da Informação, será um prazer ter você comigo nas redes sociais!

Siga-me e vamos juntos nessa jornada:



[Seguir no Instagram](#)



[Inscrição no Canal](#)



[Seguir no LinkedIn](#)

Estarei sempre compartilhando conhecimento, informações atualizadas e motivação para que você se mantenha firme no seu propósito. Será uma honra contar com sua presença por lá!

Bons estudos e sucesso em sua caminhada!

## OUTROS MATERIAIS

Conheça outros materiais acessando o seguinte



[Acessar Conteúdos](#)

Conheça também os materiais gratuitos acessando o seguinte



[Material Gratuito](#)

## SIMBOLOGIAS UTILIZADAS

As simbologias estão estrategicamente posicionadas em nossos materiais a fim de destacar alguns assuntos

MARCADOR	EXPLICAÇÃO
 <b>Foco para discursivas!</b>	Conteúdo com maior probabilidade de ser cobrado em provas discursivas por ter alta incidência em provas anteriores
 <b>Memorize!</b>	Conteúdo que é explorado em quase todas as provas e que pode cair em discursivas
 <b>Resumo!</b>	Explicação sintetizada
 <b>Destaque!</b>	Destacar um ponto
 <b>Despenca nas provas!</b>	Sempre cai nas provas e em mais de uma questão
 <b>Cai muito!</b>	É cobrado em muitas provas
 <b>Exercício essencial!</b>	Treinamento para fixação do conteúdo apresentado
 <b>Atenção!</b>	Hora de ficar atento se estiver lendo no automático
 <b>Importante!</b>	Conteúdo relevante
 <b>Ponto chave!</b>	Ponto mais relevante do assunto
 <b>Sugestão de leitura!</b>	Recomendação de conteúdo complementar

## **CARGO 4: ANALISTA TÉCNICO – ÁREA: TECNOLOGIA DA INFORMAÇÃO E CIÊNCIA DE DADOS**

### **SUMÁRIO**

TECNOLOGIA DA INFORMAÇÃO E COMUNICAÇÃO:

#### **CAPÍTULO I - CIÊNCIA DE DADOS**

- 1.1 Aprendizado supervisionado
- 1.2 Aprendizado não-supervisionado
- 1.3 Redes neurais artificiais
- 1.4 IA generativa
- 1.5 Processamento de linguagem natural (PLN)
- 1.6 Big Data
- 1.7 Visão computacional
- 1.8 Deep learning
- 1.9 Mineração de dados
- 1.10 ETL e ELT
- 1.11 Manipulação, tratamento e visualização de dados; técnicas para lidar com valores ausentes, com outliers e com desbalanceamento dos dados; seleção, extração e criação de novas características (feature engineering)
- 1.12 Modelagem: otimização de hiperparâmetros, técnicas de regularização, métricas para avaliação e seleção de modelos, validação de modelos
- 1.13 Aplicações em Python: análise de dados, aprendizado de máquina, técnicas de classificação, técnicas de regressão, ensembling de modelos, técnicas de agrupamento, técnicas de redução de dimensionalidade, técnicas de associação e sistemas de recomendação
- 1.14 Governança e ética na IA: transparência, responsabilidade, explicabilidade, privacidade, segurança e viés
- 1.15 MLOps: Gestão de código, treinamento, implantação, monitoramento e versionamento de modelos, automação do ciclo de produção

#### **CAPÍTULO II - BANCOS DE DADOS**

- 2.1 Banco de dados: conceitos básicos, arquitetura, estrutura de dados e normalização de dados
- 2.2 Modelagem de dados: relacional, multidimensional e NoSQL
- 2.3 Linguagem de consulta estruturada (PL/SQL e SQL)
- 2.4 Linguagem de definição de dados (DDL)
- 2.5 Linguagem de manipulação de dados (DML)
- 2.6 Sistemas de gestão de banco de dados (SGDB)
- 2.7 DataWarehouse, DataMart, DataLake, DataMesh

## **CAPÍTULO III - SEGURANÇA DA INFORMAÇÃO**

- 3.1 Normas NBR ISO/IEC 27001, NBR ISO/IEC 27002 e NBR ISO/IEC 27005 em suas respectivas versões mais recentes
- 3.2 Gestão de identidades e acesso: métodos de autenticação e autorização, SSO (single sign-on), SAML (security assertion markup language), OAuth2, OpenID Connect, JWT (JSON web token)
- 3.3 Privacidade e segurança por padrão
- 3.4 Principais tipos de ataques e vulnerabilidades: DDoS, DoS, DNS spoofing, worms, ransomware, adware, spyware, keylogger, cavalo de Troia, eavesdropping, phishing, brute force, port scanning, injection (SQL, LDAP), XSS (cross-site scripting), quebra de autenticação e gerenciamento de sessão, referência insegura a objetos, cross-site request forgery, armazenamento inseguro de dados criptografados
- 3.5 Controles e testes de segurança para aplicações web e web services
- 3.6 Autenticação em múltiplos fatores (MFA)
- 3.7 Soluções para segurança da informação: firewall, IDS (intrusion detection system), IPS (intrusion prevention system), SIEM (security information and event management), proxy, IAM (identity access management), PAM (privileged access management), antivírus, antispam
- 3.8 Frameworks de segurança da informação e segurança cibernética: MITRE ATT&CK, CIS controls e NIST cybersecurity framework (NIST CSF)
- 3.9 Assinatura e certificação digital, criptografia e proteção de dados em trânsito e em repouso. 10 Segurança em nuvens e de contêineres. 11 Lei nº 13.709/2018 (Lei Geral de Proteção de Dados Pessoais) e suas alterações

## **CAPÍTULO IV - ENGENHARIA DE SOFTWARE**

- 4.1 Arquitetura de sistemas web: gRPC, WebSockets, TLS, servidores proxy, cache, DNS, balanceamento de carga, tolerância a falhas e escalabilidade em sistemas web
- 4.2 Princípios e práticas de DevOps e DevSecOps e integração e entrega contínuas (CI/CD)
- 4.3 Práticas de UX e UI design
- 4.4 Web services
- 4.5 GIT
- 4.6 Transações distribuídas
- 4.7 Arquitetura de software: arquitetura multicamadas, arquitetura orientada a serviços, microsserviços, padrões arquiteturais (SOAP e REST), DDD (domain driven design), clean architecture, serverless, arquitetura baseada em nuvem
- 4.8 Engenharia de requisitos: elicitação, especificação e gerenciamento de requisitos, prototipação, usabilidade e linguagem de modelagem UML 2.5.1
- 4.9 Qualidade de software: métricas de software, indicadores e métricas de qualidade de produtos de software, CMMI 2.0, MPS.BR — Referência Guia Geral MPS Software 2023
- 4.10 Padrões de projeto (design patterns)
- 4.11 Testes de software
- 4.12 Práticas: implementação e manutenção de software, gerência de configuração, lógica de programação e paradigmas de programação, linguagens e frameworks (Java, PHP, JavaScript, HTML, CSS, XML, JSON, Python, C#, ASP.NET, Delphi 7, Angular)
- 4.13 Contagem em projetos: IFPUG e NESMA, uso de deflatores

## **CAPÍTULO V - INFRAESTRUTURA DE TI**

- 5.1 Redes de computadores: conceitos de LAN, WAN e SDN
- 5.2 Endereçamento e protocolos da família TCP/IP
- 5.3 Gerenciamento de redes TCP/IP: arquitetura de gerenciamento, SMI, SNMP e MIB
- 5.4 Noções de infraestrutura como código e automação de infraestrutura de TI
- 5.5 Ambientes virtualizados: Docker e Kubernetes (boas práticas para infraestrutura e orquestração de contêineres) e administração e gerenciamento de ambientes virtualizados
- 5.6 Computação em nuvem: conceitos e implementação e administração de serviços de nuvem (IaaS, PaaS e SaaS)
- 5.7 Serviços de rede Microsoft Windows Server: DNS, DHCP, Radius, autenticação, certificados, Active Directory (AD)
- 5.8 Monitoramento e análise de sistemas em produção por meio do uso de ferramentas de monitoramento e logging: Nagios, Prometheus, Grafana, Elasticsearch, Kibana, application performance monitoring (APM)
- 5.9 Protocolos da camada de aplicação, como simple mail transfer protocol (SMTP), hypertext transfer protocol (HTTP), hypertext transfer protocol secured (HTTPS), SSL/TLS, lightweight directory access protocol (LDAP), network file system (NFS), server message block (SMB)
- 5.10 Administração e operação de sistemas operacionais: Microsoft Windows Server e Linux
- 5.11 Conceitos e ferramentas de orquestração e automação de infraestrutura: Puppet, Ansible

## **CAPÍTULO VI - GESTÃO DE TI**

- 6.1 Gestão estratégica
  - 6.1.1 Planejamento estratégico de negócio
  - 6.1.2 Planejamento estratégico de TI
  - 6.1.3 Alinhamento estratégico entre TI e negócio
- 6.2 BPMN (business process model and notation)
- 6.3 Gerenciamento de projetos (PMBOK 7ª edição)
  - 6.3.1 Conceitos básicos, estrutura e objetivos de um projeto conforme o guia PMBOK 7
  - 6.3.2 Ciclo de vida do projeto e ciclo de vida do produto conforme o guia PMBOK 7
    - 6.3.2.1 Domínios e Princípios conforme o guia PMBOK 7
  - 6.3.3 Processos, grupos de processos e áreas de conhecimento (PMBOK 5 e PMBOK 6)
- 6.4 Gerenciamento de serviços (ITIL 4)
  - 6.4.1 Conceitos básicos, estrutura e objetivos
  - 6.4.2 Processos e funções de estratégia, desenho, transição e operação de serviços
- 6.5 Governança de TI (COBIT 2019)
  - 6.5.1 Conceitos básicos, estrutura e objetivos
  - 6.5.2 Requisitos da informação
  - 6.5.3 Recursos de tecnologia da informação
  - 6.5.4 Domínios, processos e objetivos de controle
- 6.6 Governança de TI do SISP  
(<https://www.gov.br/governodigital/pt-br/estrategias-e-governanca-digital/sisp/documentos/guia-degovt-ic-do-sisp>)
- 6.7 Metodologias ágeis
  - 6.7.1 Scrum

6.7.1.1 Fundamentos do Scrum

6.7.1.2 Papéis e responsabilidades (Product Owner, Scrum Master, Time de Desenvolvimento)

6.7.1.3 Eventos do Scrum (Sprint, Reunião de Planejamento, Daily Scrum, Revisão da Sprint, Retrospectiva)

6.7.1.4 Artefatos do Scrum (Backlog do Produto, Backlog da Sprint, Incremento)

6.7.2 Kanban

6.7.2.1 Fundamentos do Kanban

6.7.2.2 Princípios e práticas do Kanban

6.7.2.3 Fluxo contínuo e gestão visual

6.7.2.4 Limites de trabalho em progresso (WIP)

6.7.2.5 Métricas e melhorias contínuas no Kanban

6.7.3 Técnicas de estimativa ágil (planning poker e story points)

6.8 Ferramentas e técnicas: indicadores de desempenho em projetos, MS Project e técnicas de acompanhamento de projetos

6.9 Contratações de TI: Lei nº 14.133/2021 e suas alterações e Instrução Normativa SGD/ME nº 94/2022

## **CAPÍTULO I - CIÊNCIA DE DADOS**

1.1 Aprendizado supervisionado

1.2 Aprendizado não-supervisionado

1.3 Redes neurais artificiais

1.4 IA generativa

1.5 Processamento de linguagem natural (PLN)

1.6 Big Data

1.7 Visão computacional

1.8 Deep learning

1.9 Mineração de dados

1.10 ETL e ELT

1.11 Manipulação, tratamento e visualização de dados; técnicas para lidar com valores ausentes, com outliers e com desbalanceamento dos dados; seleção, extração e criação de novas características (feature engineering)

1.12 Modelagem: otimização de hiperparâmetros, técnicas de regularização, métricas para avaliação e seleção de modelos, validação de modelos

1.13 Aplicações em Python: análise de dados, aprendizado de máquina, técnicas de classificação, técnicas de regressão, ensembling de modelos, técnicas de agrupamento, técnicas de redução de dimensionalidade, técnicas de associação e sistemas de recomendação

1.14 Governança e ética na IA: transparência, responsabilidade, explicabilidade, privacidade, segurança e viés

1.15 MLOps: Gestão de código, treinamento, implantação, monitoramento e versionamento de modelos, automação do ciclo de produção

## 1.1 Aprendizado supervisionado

O **Aprendizado Supervisionado** é um dos conceitos mais fundamentais dentro da Ciência de Dados e do campo da Inteligência Artificial (IA), sendo amplamente cobrado em concursos públicos na área de Tecnologia da Informação. Trata-se de uma abordagem que busca ensinar algoritmos a realizar previsões e classificações com base em dados históricos que já possuem respostas conhecidas, denominadas **rótulos** ou **targets**. Essa metodologia é aplicada em situações em que o objetivo é fazer previsões precisas sobre novas informações, a partir de padrões previamente aprendidos.

No contexto do setor público, o Aprendizado Supervisionado tem enorme aplicabilidade na detecção de fraudes fiscais, na previsão de demanda por serviços públicos, na classificação de documentos oficiais, no reconhecimento automatizado de padrões em processos judiciais, entre outras demandas críticas para a administração pública.

### 1) Definição e Utilidades

#### ♦ O que é Aprendizado Supervisionado?

O Aprendizado Supervisionado é uma técnica de **Machine Learning (ML)** em que o modelo é treinado com um conjunto de dados rotulados, ou seja, com exemplos nos quais tanto as **entradas (features)** quanto às saídas (**rótulos**) são conhecidas. Durante o treinamento, o algoritmo aprende a identificar padrões nos dados que permitem prever ou classificar novos exemplos desconhecidos com base no que foi aprendido.

Esse aprendizado ocorre por meio da minimização de um erro entre as previsões realizadas e os valores reais, ajustando seus parâmetros internos até alcançar uma performance satisfatória. Após treinado e validado, o modelo pode ser usado para classificar, prever ou estimar saídas para novas informações.

#### 📌 Utilidades em TI e concursos:

- Previsão de inadimplência em tributos públicos.
- Classificação automática de manifestações recebidas em ouvidorias.
- Identificação de processos com alto risco de atraso em tribunais.
- Previsão da demanda por vagas escolares em regiões específicas.
- Criação de assistentes virtuais capazes de responder perguntas baseadas em categorias.
- Apoio em auditorias eletrônicas, apontando casos suspeitos para revisão humana.

---

### 2) Exemplos de Aplicação

♦ **Na saúde pública:** Prever a probabilidade de internação hospitalar com base em dados clínicos e históricos médicos.

- ♦ **No setor tributário:** Classificar contribuintes como "alto risco" ou "baixo risco" de sonegação fiscal.
  - ♦ **Na segurança pública:** Identificar padrões em ocorrências criminais e prever áreas com maior chance de reincidência de crimes.
  - ♦ **No judiciário:** Prever o tempo de tramitação de processos e recomendar priorizações.
  - ♦ **Em licitações:** Detectar possíveis irregularidades em propostas com base em características históricas.
- 

### 3 Tecnologias Relacionadas

- **Linguagens:** Python (bibliotecas como scikit-learn, TensorFlow, Keras), R (pacotes como caret, randomForest).
  - **Ferramentas de BI e IA:** Power BI (com integração Python), Google Colab, Azure Machine Learning, AWS SageMaker.
  - **Ambientes de análise:** Jupyter Notebook, VS Code, RStudio.
  - **Plataformas públicas:** Dados Abertos, portais de transparência e sistemas internos para extração e tratamento dos dados rotulados.
- 

### 4 Tópicos mais cobrados em concursos públicos

#### ✓ Problemas típicos do Aprendizado Supervisionado

Problema	Objetivo	Exemplo prático
Classificação	Categorizar registros	Determinar se um documento é "urgente" ou "não urgente"
Regressão	Prever valores contínuos	Estimar o valor da arrecadação mensal de impostos

---

#### ✓ Funcionamento do Aprendizado Supervisionado

O processo do Aprendizado Supervisionado é composto por etapas fundamentais:

- ♦ **Preparação dos dados:** Seleção e limpeza de dados históricos com entradas (variáveis independentes) e saídas (variáveis dependentes).
- ♦ **Divisão do conjunto de dados:** Separação em conjuntos de treino e teste (geralmente 70% para treino e 30% para teste).
- ♦ **Treinamento do modelo:** Ajuste dos parâmetros do algoritmo para aprender padrões nos dados de treino.

- ♦ **Validação e avaliação:** Teste do modelo com dados nunca vistos para verificar sua eficácia com métricas adequadas.
- ♦ **Previsão:** Aplicação do modelo para prever ou classificar novos dados reais.

## ✓ Principais Algoritmos Supervisionados

### 📢 Despenca nas provas!

Algoritmo	Descrição	Quando usar
Regressão Linear	Modela relações lineares entre variáveis	Prever valores contínuos (ex.: receitas públicas)
Regressão Logística	Classificação binária	Determinar aprovações/reprovações
Árvores de Decisão	Segmentação de dados com regras	Classificações com regras claras
Random Forest	Conjunto de árvores (ensemble)	Melhorar precisão em dados complexos
Support Vector Machines (SVM)	Maximizar margens entre categorias	Dados com fronteiras bem definidas
K-Nearest Neighbors (KNN)	Classificação baseada na proximidade	Pequenas bases de dados com vizinhança próxima
Gradient Boosting	Correção iterativa de erros	Alto desempenho em dados ruidosos

## ✓ Métricas de Avaliação no Aprendizado Supervisionado

🔔 É fundamental avaliar o desempenho dos modelos para garantir precisão e confiabilidade. As principais métricas são:

Métrica	Aplicação	Significado
Acurácia	Classificação	Percentual de acertos gerais
Precisão	Classificação	Taxa de acertos entre os positivos previstos
Recall	Classificação	Capacidade de identificar todos os positivos reais
F1-Score	Classificação	Média harmônica entre Precisão e Recall
$R^2$ (R-quadrado)	Regressão	Proporção da variabilidade explicada
MAE e RMSE	Regressão	Erros médios absolutos e quadráticos

## ✓ Vantagens do Aprendizado Supervisionado

- Alta precisão quando treinado com dados de qualidade.
- Facilidade de interpretação em modelos simples (como regressão linear).
- Aplicabilidade em problemas reais do setor público e privado.
- Ampla variedade de algoritmos prontos para diferentes cenários.

## ✓ Desvantagens

- Dependência de dados rotulados (que podem ser escassos ou custosos de obter).
  - Risco de sobreajuste (**overfitting**) se o modelo aprender demais o "ruído" dos dados.
  - Necessidade de reavaliação constante para manter a performance com dados atualizados.
- 

## Resumo!

✓ O **Aprendizado Supervisionado** é a base da maioria das aplicações práticas de Machine Learning, especialmente onde há dados históricos com respostas conhecidas.

✓ Permite resolver problemas de classificação e regressão com alta assertividade.

✓ Em concursos públicos de TI, é cobrado em questões teóricas e práticas, incluindo identificação de problemas adequados, escolha de algoritmos, funcionamento e métricas de avaliação.

✓ No setor público, garante maior eficiência na previsão de demandas, combate a fraudes, automatização de classificações e muito mais.

## Lista de questões no estilo Cebraspe para aprendizado supervisionado

**1. No aprendizado supervisionado, os modelos são treinados com dados que possuem rótulos associados.**

 Comentário: O aprendizado supervisionado utiliza conjuntos de dados nos quais cada entrada tem uma resposta esperada (rótulo), permitindo que o modelo aprenda a associar padrões de entrada a saídas.

✓ Resposta: Certo

---

**2. O principal objetivo do aprendizado supervisionado é identificar padrões em dados sem a necessidade de informações de saída conhecidas.**

 Comentário: Esse conceito se refere ao aprendizado não supervisionado. No aprendizado supervisionado, as saídas (rótulos) são conhecidas e utilizadas no treinamento.

✗ Resposta: Errado

---

**3. Problemas de classificação e de regressão são exemplos típicos de aplicações do aprendizado supervisionado.**

💡 Comentário: No aprendizado supervisionado, os modelos são utilizados tanto para classificar dados em categorias quanto para prever valores contínuos.

✅ Resposta: Certo

---

**4. No aprendizado supervisionado, o modelo não precisa de dados históricos para realizar previsões futuras.**

💡 Comentário: O aprendizado supervisionado depende diretamente de dados históricos rotulados para aprender padrões e realizar previsões precisas.

❌ Resposta: Errado

---

**5. Modelos supervisionados podem ser aplicados em situações nas quais não há rótulos disponíveis nos dados de entrada.**

💡 Comentário: Sem rótulos, o aprendizado supervisionado não pode ocorrer, pois os rótulos são essenciais para guiar o treinamento do modelo.

❌ Resposta: Errado

---

**6. Em aprendizado supervisionado, o conjunto de dados geralmente é dividido em subconjuntos de treino e de teste para avaliar a capacidade de generalização do modelo.**

💡 Comentário: Dividir o conjunto de dados ajuda a verificar se o modelo consegue manter um bom desempenho em dados que não foram usados no treinamento.

✅ Resposta: Certo

---

**7. O overfitting ocorre no aprendizado supervisionado quando o modelo aprende demais sobre os dados de treino e apresenta baixo desempenho em dados novos.**

💡 Comentário: Overfitting é um problema clássico no aprendizado supervisionado e reflete a falta de capacidade de generalização do modelo.

✅ Resposta: Certo

---

**8. Em aprendizado supervisionado, quanto maior a quantidade de dados rotulados e de boa qualidade, maior a probabilidade de o modelo atingir um bom desempenho.**

💡 Comentário: Dados abundantes e precisos proporcionam ao modelo melhores condições para aprender padrões relevantes e generalizáveis.

✅ Resposta: Certo

---

**9. Modelos supervisionados são incapazes de realizar tarefas de previsão de valores contínuos, sendo limitados apenas à classificação.**

💡 Comentário: Além de classificação, o aprendizado supervisionado também abrange tarefas de regressão, que preveem valores contínuos.

✗ Resposta: Errado

---

**10. A Regressão Logística é um exemplo de algoritmo de aprendizado supervisionado utilizado em problemas de classificação.**

💡 Comentário: Apesar do nome "Regressão", esse algoritmo é amplamente usado para resolver problemas de classificação binária.

✓ Resposta: Certo

---

**11. No aprendizado supervisionado, não existe necessidade de ajustar hiperparâmetros para melhorar o desempenho do modelo.**

💡 Comentário: O ajuste de hiperparâmetros é fundamental para aprimorar a performance de muitos modelos supervisionados.

✗ Resposta: Errado

---

**12. Em classificação supervisionada, a acurácia é uma métrica que pode ser inadequada quando o conjunto de dados está desbalanceado.**

💡 Comentário: Em conjuntos desbalanceados, métricas como precisão, recall e F1-score podem ser mais apropriadas que a acurácia.

✓ Resposta: Certo

---

**13. A validação cruzada é uma técnica que busca melhorar a avaliação de modelos supervisionados ao utilizar diferentes partições dos dados para treino e teste.**

💡 Comentário: A validação cruzada distribui os dados em várias rodadas de treino e validação para obter uma avaliação mais confiável.

✓ Resposta: Certo

---

**14. Em aprendizado supervisionado, modelos simples sempre apresentam maior risco de overfitting do que modelos complexos.**

💡 Comentário: Modelos simples tendem a sofrer mais com underfitting, enquanto modelos

complexos são mais propensos ao overfitting.

✗ Resposta: Errado

---

**15. Classificação binária é um tipo de problema supervisionado no qual o modelo deve atribuir uma entre duas possíveis categorias a cada exemplo.**

💡 Comentário: Na classificação binária, os rótulos de saída são compostos por duas classes distintas, como "sim" ou "não".

✓ Resposta: Certo

---

**16. Em aprendizado supervisionado, dados ruidosos e outliers podem comprometer a capacidade do modelo de aprender padrões relevantes.**

💡 Comentário: A presença de ruídos e valores extremos pode levar o modelo a aprender padrões falsos, prejudicando o desempenho.

✓ Resposta: Certo

---

**17. Problemas supervisionados não se beneficiam de técnicas de balanceamento de classes, pois o modelo ajusta automaticamente a proporção entre categorias.**

💡 Comentário: Em situações de desbalanceamento, técnicas específicas são necessárias para evitar que o modelo favoreça a classe majoritária.

✗ Resposta: Errado

---

**18. O aprendizado supervisionado é indicado para problemas nos quais não se deseja prever saídas específicas com base nas entradas.**

💡 Comentário: O aprendizado supervisionado busca exatamente prever ou classificar saídas específicas com base nos dados de entrada.

✗ Resposta: Errado

---

**19. O K-Nearest Neighbors (KNN) é um exemplo de algoritmo de aprendizado supervisionado que pode ser usado tanto para classificação quanto para regressão.**

💡 Comentário: KNN é um algoritmo flexível, aplicável em tarefas de classificação e também em regressão.

✓ Resposta: Certo

---

**20. Em aprendizado supervisionado, quanto menor a capacidade de generalização de um modelo, melhor será seu desempenho em dados novos.**

💡 Comentário: Alta capacidade de generalização é o objetivo do aprendizado supervisionado para garantir bom desempenho em dados não vistos.

✘ Resposta: Errado

## 1.2 Aprendizado não-supervisionado

O **Aprendizado Não-Supervisionado** é uma das abordagens mais relevantes da **Ciência de Dados** e um tema cada vez mais presente em concursos públicos voltados para Tecnologia da Informação. Diferentemente do aprendizado supervisionado, em que os dados possuem respostas conhecidas (rótulos), no aprendizado não-supervisionado os algoritmos trabalham com conjuntos de dados **sem rótulos**, ou seja, sem que haja previamente categorias, classificações ou valores de saída definidos.

O objetivo central do aprendizado não-supervisionado é identificar estruturas ocultas, padrões, agrupamentos e associações nos dados sem qualquer conhecimento prévio das relações existentes. Essa abordagem é fundamental em ambientes públicos que lidam com grandes volumes de dados e precisam organizar informações para auxiliar em tomadas de decisão estratégicas, como segmentação de contribuintes, detecção de comportamentos atípicos e agrupamento de registros históricos.

### 1) Definição e Relevância

#### ♦ O que é Aprendizado Não-Supervisionado?

Aprendizado Não-Supervisionado é uma técnica de **Machine Learning (ML)** que tem como objetivo encontrar padrões, estruturas e relacionamentos em conjuntos de dados sem que haja um rótulo ou resposta conhecida associada a cada registro. Em outras palavras, o algoritmo "explora" os dados em busca de informações ocultas e organiza esses dados de maneira autônoma, baseando-se nas similaridades ou frequências dos atributos observados.

Essa abordagem é particularmente útil quando o objetivo é analisar comportamentos, dividir dados em segmentos semelhantes, reduzir dimensionalidade para simplificação de análise e encontrar associações entre variáveis, mesmo sem ter previamente definido quais resultados deveriam ser encontrados.

#### 📌 Importância para TI e concursos públicos:

- Identificação de padrões ocultos em grandes volumes de dados.
- Segmentação de usuários, processos ou documentos.
- Detecção de anomalias e comportamentos atípicos.
- Apoio na criação de políticas públicas baseadas em perfis populacionais.
- Agrupamento de serviços similares para otimização de recursos.

- Análise exploratória prévia para preparar dados antes da modelagem supervisionada.
- 

## 2 Exemplos de Aplicação

- ◆ **Saúde Pública:** Agrupamento de pacientes com perfis de sintomas semelhantes para estudo de novas doenças.
  - ◆ **Tribunais:** Detecção de padrões em processos judiciais para identificação de grupos com maior risco de atraso.
  - ◆ **Educação:** Segmentação de estudantes por perfil socioeconômico e desempenho acadêmico para políticas de apoio.
  - ◆ **Segurança Pública:** Análise de registros de ocorrências para detectar regiões com padrões de crimes semelhantes.
  - ◆ **Fiscalização Tributária:** Agrupamento de empresas com movimentações financeiras semelhantes, para análise de risco.
- 

## 3 Tecnologias Relacionadas

- **Linguagens:** Python (bibliotecas como scikit-learn, pandas, seaborn), R (pacotes como cluster, factoextra).
  - **Ambientes de análise:** Jupyter Notebook, Google Colab, RStudio.
  - **Ferramentas de BI:** Power BI (com scripts Python), Tableau.
  - **Cloud Computing:** Azure Machine Learning, Google Cloud AI Platform, AWS SageMaker.
  - **Métodos complementares:** PCA (Análise de Componentes Principais), t-SNE (t-Distributed Stochastic Neighbor Embedding), UMAP (Uniform Manifold Approximation and Projection).
- 

## 4 Tópicos mais cobrados em concursos públicos

- ✓ **Principais Problemas Resolvidos pelo Aprendizado Não-Supervisionado**



[www.kuasarnex.com](http://www.kuasarnex.com)

 [@kuasarnex](https://www.instagram.com/kuasarnex)