

Parte 1

Bem vindos!

Aprendizado Supervisionado

Não esqueça de gravar a aula!



Sergio Polimante



- Estoquista -> Machine Learning Engineer e Professor
- Bacharel em Ciências e Tecnologia
- Engenheiro Robótico
- Mestre em engenharia da Informação

- Ciências da Computação: +15 anos
- Dados: +8 anos
- Eletrônica, Robótica, Logística, Precificação, Supply & Demand, Marketing, e mais.
- Professor de música, cursinho, SENAI, Coderhouse, FIAP, FGV.

- Pai do Leonardo e marido da Jéssica
- Músico Amador

Sergio Polimante



- Estoquista-> Machine Learning Engineer e Professor
- Bacharel em Ciências e Tecnologia
- Engenheiro Robótico
- Mestre em engenharia da Informação

- Ciências da Computação: +15 anos
- Dados: +8 anos
- Eletrônica, Robótica, Logística, Precificação, Supply & Demand, Marketing, e mais.
- Professor de música, cursinho, SENAI, Coderhouse, FIAP, FGV.

- Pai do Leonardo e marido da Jéssica
- Músico Amador

Avaliação

30% - Projeto feito em Aula em Grupo

70% - Questões conceituais sobre o projeto, individual.

Quem são vocês?

- ☐ Você é um/a profissional técnico/a ou de negócios?
- ☐ Se for técnico, se identifica como um profissional de dados?
- ☐ Já possui conhecimentos técnicos em IA e ML?
- ☐ Vocês possuem times ou profissionais específicos para dados?
- ☐ Sua empresa já possui um produto de dados em deploy? (dashboards, modelos de ML, etc)
- ☐ Qual sua expectativa ao finalizar essa fazer essa disciplina?

Compromissos

Participação ativa: Comprometem-se a participar ativamente das aulas, fazer perguntas e contribuir para discussões em sala de aula.

Dúvidas são bem-vindas: Não hesitem em fazer perguntas. Não existe pergunta ruim.

Confirmação de entendimento: Quando o professor perguntar se entenderam um tópico, assegurem-se de que compreenderam plenamente antes de prosseguir.

Compartilhe suas Experiências: Enriqueça a aula contribuindo com casos reais de sua experiência profissional!

Feedback sobre o curso: Forneçam feedback honesto sobre o curso, destacando o que funcionou bem e áreas que podem ser aprimoradas.

Planejamento das Aulas



Como Machine Learning
pode gerar valor para o seu negócio?

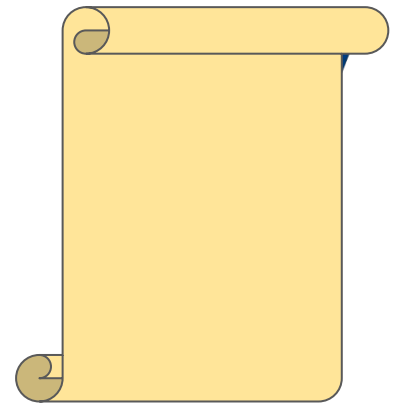
O que é Machine Learning

Definição: Sistemas que melhoram através da experiência

A mudança da programação explícita para o aprendizado com dados

Componentes-chave: dados, algoritmos, métricas de avaliação

Contexto histórico: De métodos estatísticos ao ML moderno



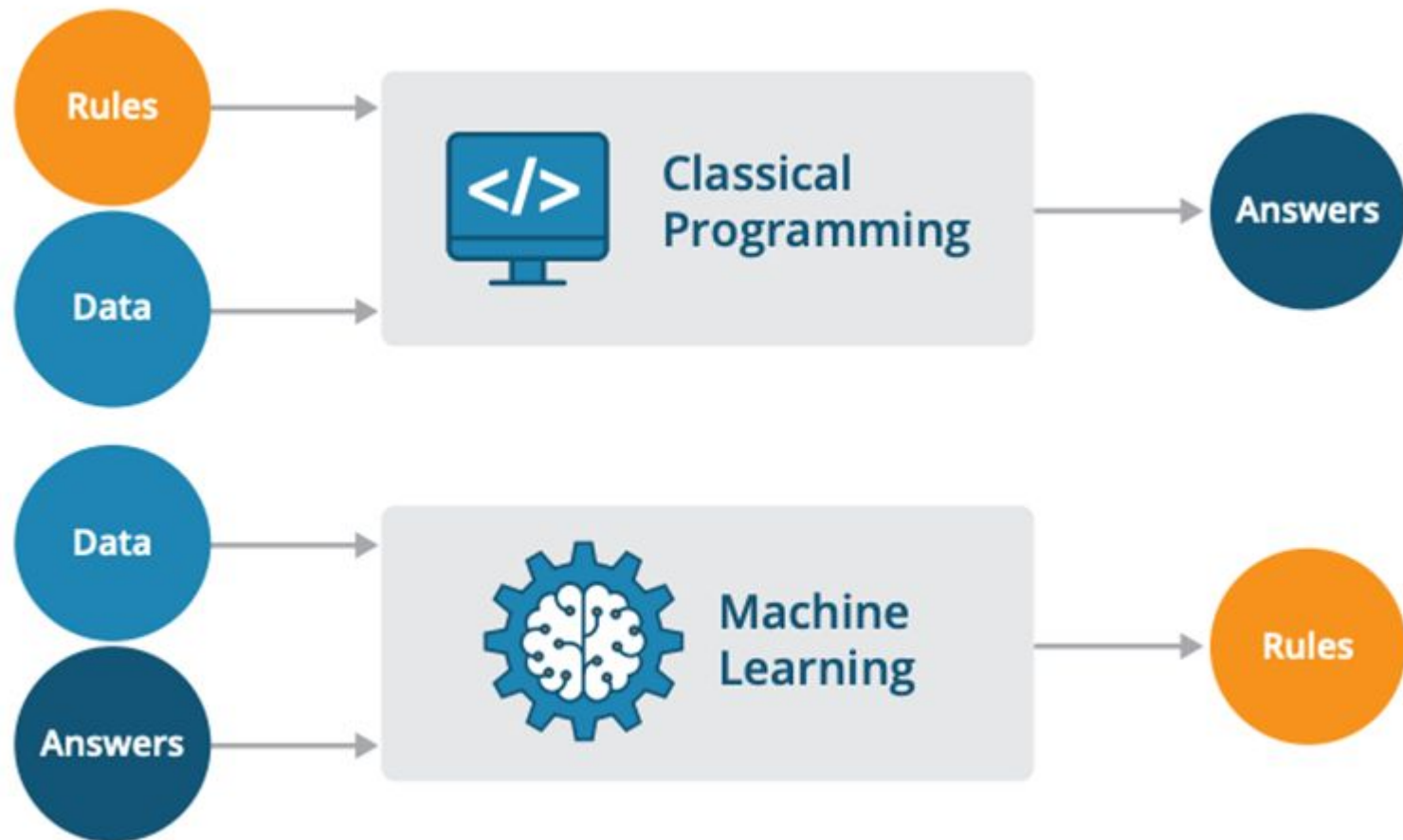
O que é Machine Learning

O que é Machine Learning?

“Machine Learning é um subconjunto da inteligência artificial (IA) que permite que sistemas aprendam e melhorem com a experiência, sem serem explicitamente programados.”

[Google](#)

O que é Machine Learning?



Problema: Detecção de Spam

Programação Clássica (Regras + Dados → Respostas)

Regras: Especialistas de negócio definem regras com base em experiências passadas.

Dados: Dados do cliente (tempo de uso, reclamações, fatura, etc.)



Saída = Resposta: Sistema classifica se o cliente vai ou não cancelar.

Exemplo de regras:

- Se o cliente tem mais de 2 reclamações no último mês → risco de churn
- Se a fatura caiu mais de 50% em 2 meses → risco de churn
- Se não usa o app há mais de 30 dias → risco de churn

Problema: Detecção de Spam

Machine Learning (Dados + Respostas → Regras aprendidas)

Dados: Dados do cliente + resposta (churn ou não churn)

Resposta (target): Dados do cliente (tempo de uso, reclamações, fatura, etc.)



Saída: O modelo aprende padrões reais de comportamento que levam ao churn. Usamos o modelo para prever o churn em outros clientes

O modelo **pode** aprender as seguintes regras (não necessariamente essas):

- Clientes que reclamam e têm redução de uso têm alto risco.
- Certos perfis demográficos cancelam após atualizações do plano.
- Clientes com queda sutil no uso, mas sem contato com suporte, podem estar "quietamente insatisfeitos".

O que é Machine Learning

IA

vs

ML

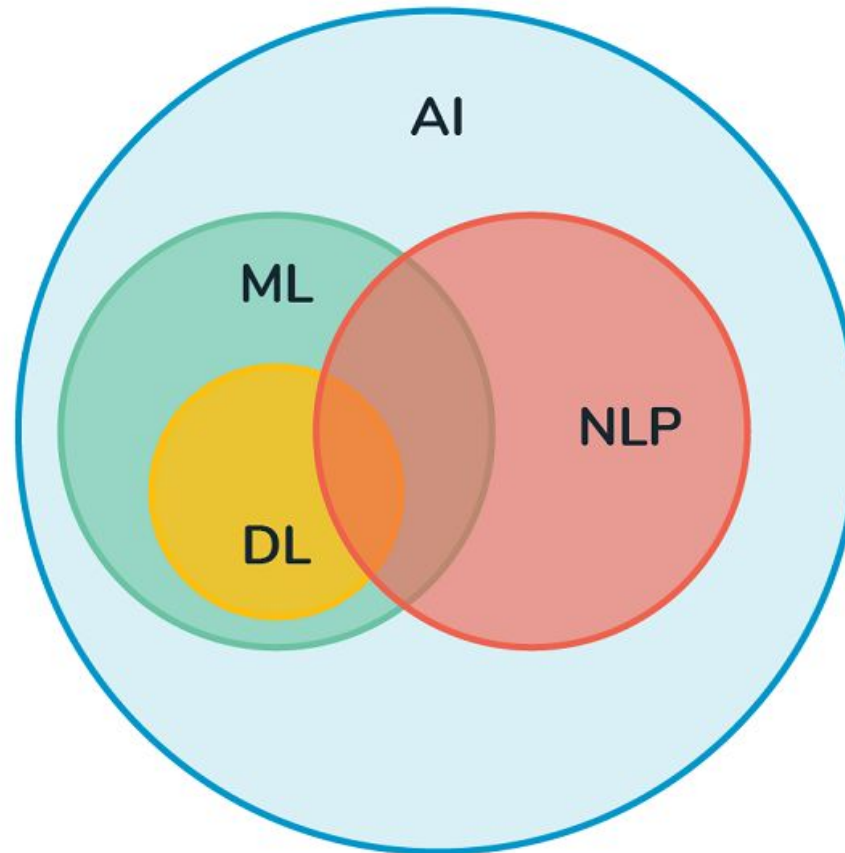
"Inteligência Artificial é o campo da ciência da computação que se concentra na criação de sistemas e máquinas capazes de realizar tarefas que normalmente exigiriam inteligência humana, como reconhecimento de padrões, aprendizado, raciocínio, resolução de problemas e tomada de decisões."





"Machine Learning é um subcampo da inteligência artificial que se concentra no desenvolvimento de algoritmos e modelos computacionais que permitem que sistemas e máquinas aprendam e melhorem seu desempenho em uma tarefa específica sem serem explicitamente programados."

"Um programa de computador é dito aprender a partir da experiência E em relação a uma classe de tarefas T e medida de desempenho P , se o seu desempenho em tarefas em T , medido por P , melhora com a experiência E ."

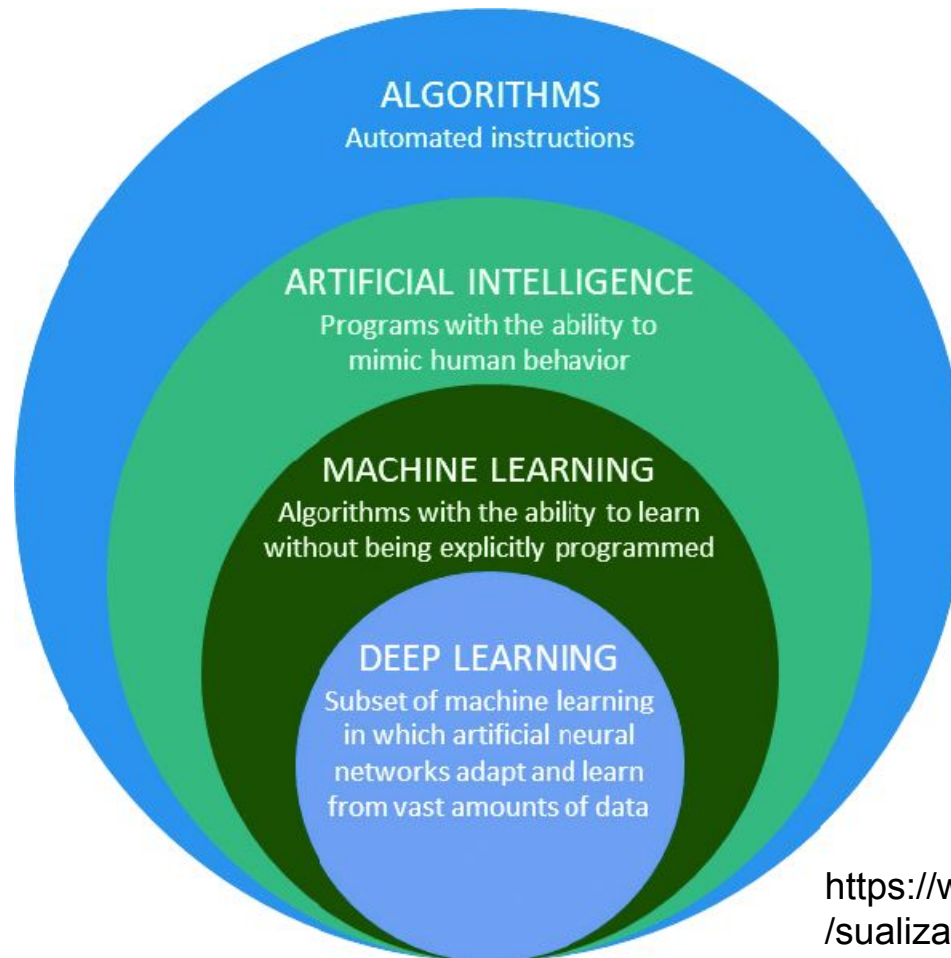
- Tom M. Mitchell

O que é Machine Learning



-  Artificial intelligence
-  Machine learning
-  Language Processing
-  Deep learning

O que é Machine Learning



https://www.researchgate.net/figure/sualization-of-algorithms-vs-artificial-intelligence-vs-machine-learning-vs-deep_fig1_339997962

O Deep Blue é inteligente?

O Deep Blue foi o algoritmo para jogar Xadrez desenvolvido pela IBM.

Usa combinação de hardware de alta potência e software de xadrez personalizado para avaliar milhões de posições por segundo, usando algoritmos para selecionar a melhor jogada com base na análise de possibilidades.

1996: Deep Blue venceu o campeão mundial de xadrez, Garry Kasparov.

1997: Deep Blue derrotou Kasparov em um confronto de seis partidas, tornando-se o primeiro computador a vencer um campeão mundial de xadrez sob regras padrão.

Professor: “O Deep blue é inteligente?”

Aluno: “Sim”

Professor: * ensina ao aluno como o algoritmo funciona *

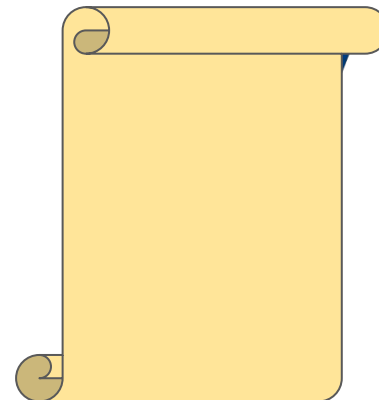
Aluno: “Não”

Componentes do Machine Learning

- **Dados**
 - Informações coletadas que servem como base para o aprendizado.
 - Podem ser de todo tipo: tabelas, imagens, sons, vídeos, etc.
- **Algoritmos**
 - Conjuntos de instruções que processam os dados para identificar padrões.
 - São os algoritmos que produzem os modelos de machine learning.
 - Em outras palavras, o modelo de machine learning é um produto da execução do algoritmo sobre uma base de dados.
- **Métricas de Avaliação**
 - Critérios utilizados para medir a performance e precisão dos modelos.
 - São importantes para validar a qualidade do modelo, como sua performance e seu valor de negócio.

Vamos ver mais a frente cada um desses assuntos com mais detalhes.

Business Value of Machine Learning



Valor Empresarial do Aprendizado de Máquina (20 minutos)

- **Estudos de Caso:** 3-4 exemplos breves de diferentes indústrias
 - Varejo: Otimização de estoque e previsão de demanda
 - Bancário: Avaliação de crédito e detecção de fraudes
 - Saúde: Auxílio diagnóstico e estratificação de risco de pacientes
 - Marketing: Segmentação de clientes e campanhas direcionadas
- **Geradores de Valor:**
 - Redução de custos através da automação
 - Crescimento de receita através da otimização
 - Gestão de riscos através da previsão
 - Inovação através de insights
- **Realidades de Implementação:**
 - A lacuna entre capacidades de ML e integração empresarial
 - Papéis em projetos de ML (analista de negócios, cientista de dados, engenheiro)

O Valor de Negócio do Machine Learning

Estudo de Casos

O Valor de Negócio do Machine Learning

Estudo de Casos

Varejo: SPAR – Previsão de Demanda e Redução de Desperdício

- Desafio: Melhorar a previsão de demanda e reduzir o desperdício de alimentos perecíveis.
- Solução: Implementação de modelos de machine learning para prever demanda e otimizar o reabastecimento.
- Resultados:
 - Precisão de previsão de estoque superior a 90%.
 - Redução de produtos não vendidos para apenas 1%.
 - Entrega de frutas e vegetais às lojas três dias mais cedo.
- Fonte: [VKTR.com](https://vktr.com) (outros exemplos de varejo no link)

O Valor de Negócio do Machine Learning

Estudo de Casos

Bancário: Banco Americano – Classificação Automatizada de Documentos

- Desafio: Gerenciar e classificar um grande volume de documentos de forma eficiente.
- Solução: Utilização de inteligência artificial para classificar e consolidar documentos automaticamente.
- Resultados:
 - 35 milhões de documentos classificados e consolidados em duas semanas.
 - Melhoria significativa na eficiência operacional e redução de custos.
- Fonte: [datamatics.com](https://www.datamatics.com)

O Valor de Negócio do Machine Learning

Estudo de Casos

Bancário – Avaliação de Crédito e Detecção de Fraudes

- Estudo de Caso: Banco do Reino Unido
- Desafio: Melhorar a precisão na avaliação de crédito e reduzir inadimplências.
- Solução: Implementação de modelos de Machine Learning para análise de crédito.
- Resultados:
 - Aumento de 83% na detecção de inadimplência em comparação com métodos tradicionais.
 - Manutenção do mesmo nível de recusas de crédito.
 - Redução de perdas financeiras e melhoria na concessão de crédito.

https://kortical.com/case-studies/ai-finance-credit-score-machine-learning?utm_source=chatgpt.com

O Valor de Negócio do Machine Learning

Estudo de Casos

Marketing: Coca-Cola – Campanha Personalizada com IA

- Desafio: Aumentar o engajamento dos consumidores através de campanhas personalizadas.
- Solução: Uso de machine learning e processamento de linguagem natural para análise de dados e personalização de campanhas.
- Resultados:
 - Aumento de mais de 2% nas vendas.
 - Engajamento nas redes sociais cresceu 870%, com milhões de consumidores compartilhando suas garrafas personalizadas online.
- Fonte: mosaikx.com (Mais estudos de caso no link)

O Valor de Negócio do Machine Learning

Estudo de Casos

Marketing – Segmentação de Clientes e Campanhas Direcionadas

- Estudo de Caso: Instacart
- Desafio: Identificar segmentos de clientes para campanhas de marketing mais eficazes.
- Solução: Utilização de algoritmos de Machine Learning para segmentação de clientes com base em dados de compras.
- Resultados:
 - Criação de campanhas de marketing mais direcionadas.
 - Aumento na taxa de conversão e fidelização de clientes.
 - Melhoria na experiência do cliente e aumento nas vendas.

<https://cadenaser.com/cmadrid/2025/04/15/el-hospital-de-fuenlabrada-participa-en-un-estudio-que-utiliza-la-ia-para-detectar-cancer-oculto-en>

O Valor de Negócio do Machine Learning

Estudo de Casos

Saúde: Rede Hospitalar dos EUA – Previsão de Alta e Redução de Internações

- Desafio: Melhorar a gestão de leitos e reduzir a duração das internações hospitalares.
- Solução: Desenvolvimento de modelos de machine learning para prever altas hospitalares e transferências para UTI.
- Resultados:
 - Redução média de 0,67 dia na duração da internação por paciente.
 - Economia estimada entre US\$ 55 e US\$ 72 milhões anuais.
 - Mais de 200 profissionais de saúde utilizam a ferramenta diariamente.
- Fonte: arxiv.org

O Valor de Negócio do Machine Learning

Estudo de Casos

Saúde – Auxílio Diagnóstico e Estratificação de Risco de Pacientes

- Estudo de Caso: Hospital Infanta Leonor
- Desafio: Detectar câncer oculto em pacientes com eventos trombóticos venosos.
- Solução: Desenvolvimento de um modelo preditivo com IA para avaliação de risco de câncer.
- Resultados:
 - 94% de precisão na identificação correta de pacientes sem câncer.
 - Detecção precoce de câncer em 6,9% dos pacientes estudados.
 - Melhoria na sobrevivência dos pacientes e redução de exames desnecessários.

<https://cadenaser.com/cmadrid/2025/04/15/el-hospital-de-fuenlabrada-participa-en-un-estudio-que-utiliza-la-ia-para-detectar-cancer-oculto-en>

Como o Machine Learning Agrega Valor ao Negócio?

Como o Machine Learning Agrega Valor ao Negócio

- Otimização de Processos e Redução de Custos
- Melhoria na Tomada de Decisão
- Personalização da Experiência do Cliente
- Previsão de Tendências e Comportamentos
- Aumento da Segurança e Detecção de Fraudes
- Inovação e Desenvolvimento de Novos Produtos

Como o Machine Learning Agrega Valor ao Negócio

Melhoria na Tomada de Decisão

- Analisar grandes volumes de dados rapidamente, fornecendo insights valiosos.
- Apoia decisões estratégicas com base em dados concretos.
- Tomada de decisões automatizadas, aumentando precisão e tempo de resposta.
- Exemplos:
 - Um banco utiliza ML para analisar dados de crédito e prever inadimplência, permitindo decisões de concessão de crédito mais precisas e reduzindo perdas financeiras.
 - Uma empresa de varejo emprega ML para analisar tendências de mercado e ajustar seu mix de produtos, aumentando as vendas em 10%.

Como o Machine Learning Agrega Valor ao Negócio

Otimização de Processos e Redução de Custos

- Automatiza tarefas repetitivas, aumentando a eficiência operacional.
- Identifica processos ineficazes e desperdícios, promovendo economia.
- Exemplos
 - Uma empresa de logística implementa ML para otimizar rotas de entrega, considerando tráfego e condições climáticas, resultando em uma redução de 15% nos custos operacionais e entregas mais rápidas.
 - Uma fábrica de manufatura utiliza ML para prever falhas em máquinas, permitindo manutenção preditiva e reduzindo o tempo de inatividade em 20%.

Como o Machine Learning Agrega Valor ao Negócio

Personalização da Experiência do Cliente

- Oferece recomendações de produtos e serviços alinhadas às preferências individuais.
- Aumenta a satisfação e fidelização dos clientes.
- Exemplos:
 - Uma plataforma de streaming emprega ML para recomendar conteúdos com base no histórico de visualização do usuário, aumentando o tempo médio de uso.
 - Um e-commerce utiliza ML para personalizar ofertas e promoções, resultando em um aumento na taxa de conversão.

Como o Machine Learning Agrega Valor ao Negócio

Previsão de Tendências e Comportamentos

- Antecipação de demandas e comportamentos do mercado.
- Adaptação proativa às mudanças, mantendo a competitividade.
- Exemplo:
 - Uma rede de supermercados usa ML para prever a demanda por produtos sazonais, ajustando estoques e evitando perdas por vencimento, o que resulta em uma economia significativa.
 - Uma empresa de moda analisa dados de redes sociais com ML para antecipar tendências, lançando coleções alinhadas às preferências dos consumidores.

Como o Machine Learning Agrega Valor ao Negócio

Aumento da Segurança e Detecção de Fraudes

- Monitoramento contínuo para identificar atividades suspeitas.
- Redução de riscos e proteção dos ativos da empresa.
- Exemplos:
 - Uma fintech implementa ML para monitorar transações em tempo real, identificando padrões suspeitos e prevenindo fraudes, reduzindo perdas.
 - Um banco utiliza ML para detectar atividades incomuns em contas correntes, alertando clientes e evitando transações fraudulentas.

Como o Machine Learning Agrega Valor ao Negócio

Inovação e Desenvolvimento de Novos Produtos

- Identifica oportunidades de mercado através da análise de dados.
- Acelera o desenvolvimento de soluções inovadoras.
- Exemplos:
 - Uma empresa de saúde utiliza ML para analisar dados genéticos e desenvolver tratamentos personalizados, acelerando o processo de pesquisa e desenvolvimento de novos medicamentos.
 - Uma fabricante de alimentos emprega ML para criar novas receitas baseadas em preferências dos consumidores, lançando produtos que atendem melhor às demandas do mercado.

Fonte: [Forbes](#)

Está mais claro como ML pode agregar valor no seu negócio?

Quais os Principais Desafios na Implantação de Machine Learning nas Empresas?

Desafios: Técnicos e Integração

- **Sistemas Legados Incompatíveis:**
 - Infra estruturas antigas que não suportam tecnologias modernas de ML.
 - Dificuldade na integração devido à falta de compatibilidade com novas ferramentas.
- **Acesso Limitado via APIs:**
 - Sistemas legados que não oferecem interfaces de programação adequadas.
 - Complicações na extração e inserção de dados essenciais para o ML.
- **Capacidade Computacional Restrita:**
 - Recursos insuficientes para processar grandes volumes de dados.
 - Necessidade de investimentos em hardware e infraestrutura.

Desafios: Qualidade e Governança de Dados

- Dados Despadronizados e Desorganizados:
 - Inconsistências que comprometem a eficácia dos modelos de ML.
 - Dificuldade na limpeza e preparação dos dados.
- Falta de Integração entre Sistemas:
 - Presença de silos de dados que impedem uma visão holística.
 - Desafios na consolidação de informações de diferentes fontes.
- Ausência de Políticas Claras de Acesso e Manipulação:
 - Riscos relacionados à privacidade e segurança dos dados.
 - Falta de diretrizes para o uso ético e responsável dos dados.

Desafios: Resistência Cultural e Organizacional

- Medo da Perda de Empregos:
 - Funcionários receosos de serem substituídos por soluções de ML.
 - Resistência à adoção de novas tecnologias.
- Desconfiança na Qualidade dos Algoritmos:
 - Stakeholders céticos quanto à precisão e confiabilidade dos modelos.
 - Necessidade de transparência e explicabilidade nos resultados.
- Desconforto com Abordagens Data-Driven:
 - Dificuldade em substituir decisões baseadas na experiência por análises de dados.
 - Conflitos entre intuição e insights gerados por ML.
- Expectativas Irrealistas sobre a IA:
 - Crença de que a IA resolverá todos os problemas sem considerar os custos e riscos.
 - Falta de compreensão sobre as limitações e desafios do desenvolvimento de modelos.

Desafios: Custo de Implementação

- Investimento Inicial Elevado:
 - Aquisição de softwares e contratação de especialistas.
 - Despesas com migração e adaptação de sistemas existentes.
- Retorno sobre o Investimento (ROI) Não Imediato:
 - Resultados tangíveis podem demorar a aparecer.
 - Desafios na mensuração do impacto financeiro direto.
- Atualização ou Substituição de Infraestruturas Legadas:
 - Processos complexos e dispendiosos.
 - Necessidade de planejamento cuidadoso para minimizar interrupções.
- Custos Contínuos de Manutenção:
 - Gastos recorrentes com atualização de modelos e infraestrutura.
 - Necessidade de monitoramento constante para garantir desempenho.

Desafios: Time Inadequado e Escassez de Talentos

- Falta de Conhecimento no Time Interno:
 - Equipes sem experiência prática em ML.
 - Necessidade de treinamentos e capacitação contínua.
- Dificuldade na Contratação de Especialistas:
 - Alta demanda e competição por profissionais qualificados.
 - Desafios na atração de talentos para áreas específicas.
- Curva de Aprendizado Íngreme para Novatos:
 - Complexidade dos conceitos e ferramentas de ML.
 - Tempo necessário para alcançar proficiência.
- Dificuldade na Retenção de Talentos:
 - Ambientes de trabalho que não promovem crescimento e inovação.
 - Necessidade de uma cultura organizacional saudável e motivadora.

Times de Dados

Quem são os profissionais que atuam no
ecossistema de Dados?

Cientista de Dados (Data Scientist):

Os cientistas de dados aplicam técnicas estatísticas e de aprendizado de máquina para extrair insights dos dados e desenvolver modelos preditivos. Eles traduzem os resultados em informações acionáveis.

Tecnologias/Ferramentas: Python, R, TensorFlow, scikit-learn, Jupyter, SQL.



Analista de Dados (Data Analyst):

Os analistas de dados interpretam os dados para fornecer informações que ajudam a tomar decisões de negócios. Eles criam relatórios, painéis e visualizações de dados.

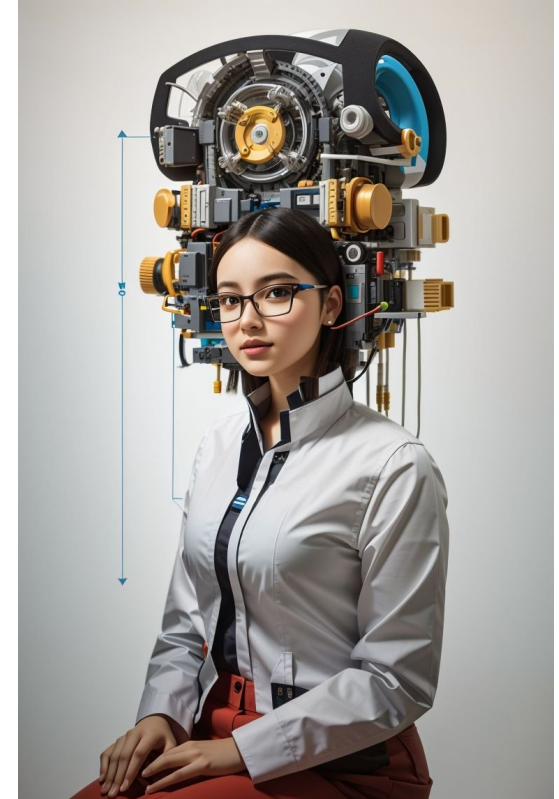
Tecnologias/Ferramentas: Excel, SQL, Tableau, Power BI, Google Analytics.



Engenheiro de Dados (Data Engineer):

Os engenheiros de dados projetam e mantêm pipelines de dados, coletam e armazenam dados de fontes diversas, garantem a qualidade dos dados e preparam-nos para análise.

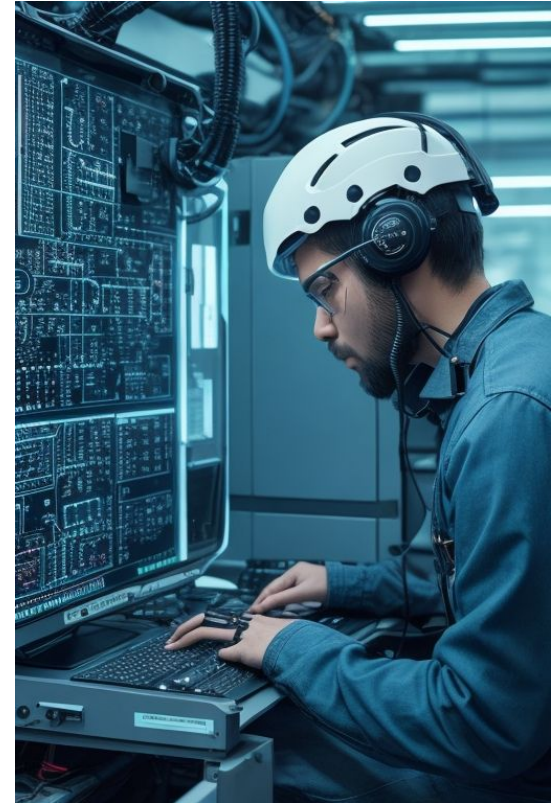
Tecnologias/Ferramentas: Apache Hadoop, Apache Spark, SQL, ETL (Extract, Transform, Load) tools, como Apache NiFi ou Talend.



Engenheiro de Machine Learning (Machine Learning Engineer)

Os engenheiros de machine learning projetam, desenvolvem e implantam modelos de aprendizado de máquina em aplicações ou sistemas em produção.

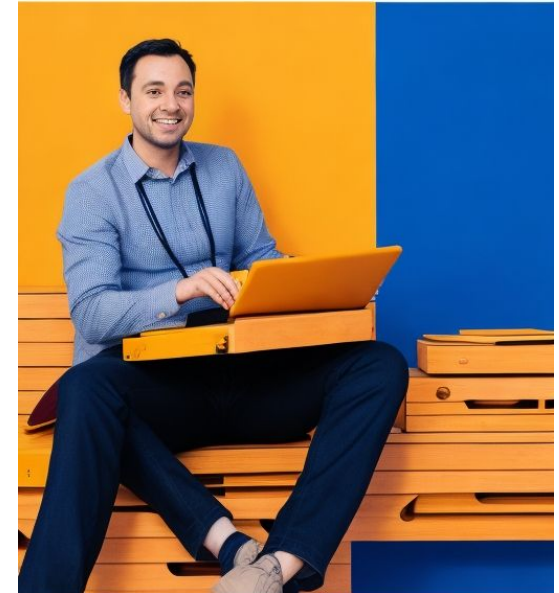
Tecnologias/Ferramentas: Python, TensorFlow, PyTorch, Docker, Kubernetes, mlflow.



Gerente de Dados e Análise (Data and Analytics Manager):

O Gerente de Dados e Análise lidera a estratégia de dados e análise da organização, supervisionando as equipes de dados, definindo objetivos e prioridades, e garantindo que as iniciativas de dados estejam alinhadas com os objetivos de negócios.

Tecnologias/Ferramentas: Além de habilidades de liderança e gestão, os Gerentes de Dados e Análise podem ter conhecimento em ferramentas de BI (Business Intelligence), ferramentas de visualização de dados, planejamento estratégico de dados e habilidades de comunicação para traduzir insights técnicos em ações de negócios.



Annka Tionaagieniks
Tonta WageWaigen~Ancits

MATFIIA NONHR MAGEN? Boiis. 35

O que é Aprendizado Supervisionado?

O que é Aprendizado Supervisionado?

- Modelo treinado para **determinar saídas com base nos dados de entrada.**
- **Objetivo:** aprender função que mapeia entrada para saída, permitindo previsões precisas.
- Utiliza rótulos (dados anotados) durante o treinamento.
- Foco na capacidade de generalização para novos dados.

Por que o Aprendizado é Chamado de "Supervisionado"?

Presença de Dados Rotulados: O modelo é treinado com conjuntos de dados que incluem entradas e suas respectivas saídas corretas, permitindo que aprenda a mapear entradas para saídas desejadas.

Processo de Correção: Durante o treinamento, o modelo faz previsões e as compara com as saídas reais, ajustando-se com base nos erros para melhorar sua precisão.

Analogia com Ensino Tradicional: Assim como um aluno aprende com a orientação de um professor, o modelo aprende com os dados rotulados que atuam como supervisores.

Objetivo de Generalização: Após o treinamento, o modelo deve ser capaz de aplicar o conhecimento adquirido para prever corretamente saídas de novos dados não vistos anteriormente.

O que é Aprendizado Supervisionado?

- **Objetivo:** atribuir categoria/classe (classificação) ou valor (regressão) a um conjunto de dados com base em suas características.
- **Classificação:** Prever categorias ou classes.
- **Regressão:** Prever valores numéricos.

Dados de Treino

A	B	C	D	E	F	G
classe	nome	sexo	idade	tarifa	porto	sobreviveu
3	Braund, Mr. Owen Harris	male	22	725	S	0
1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	7128	C	1
3	Heikkinen, Miss. Laina	female	26	792	S	1
1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	5310	S	1
3	Allen, Mr. William Henry	male	35	805	S	0
3	Moran, Mr. James	male		845	Q	0
1	McCarthy, Mr. Timothy J	male	54	5186	S	0
3	Palsson, Master. Gosta Leonard	male	2	2107	S	?
3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27	1113	S	?
2	Nasser, Mrs. Nicholas (Adele Achem)	female	14	3007	C	?
3	Sandstrom, Miss. Marguerite Rut	female	4	1670	S	?

target,
y

Dados de entrada para predição

Previsão,
prediction,
y_hat

Features vs Target

Features

- Também chamadas de variáveis preditoras ou variáveis independentes.
- Dados de entrada do modelo.
- Categóricas e Numéricas.

Target:

- Também chamadas de variável de saída, variável alvo, variável resposta, e outros.
- Dados de saída do modelo.
- **y**: variável de saída alvo (target).
- **y_hat**: variável de saída predita (prediction)

Features vs Target

Features

- Categóricas e Numéricas.

Target

- **y**: variável de saída alvo (target).
- **y_hat**: variável de saída predita (prediction)

Dados de Treino

A	B	C	D	E	F	G
classe	nome	sexo	idade	tarifa	porto	sobreviveu
3	Braund, Mr. Owen Harris	male	22	725 S		0
1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	7128 C		1
3	Heikkinen, Miss. Laina	female	26	792 S		1
1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	5310 S		1
3	Allen, Mr. William Henry	male	35	805 S		0
3	Moran, Mr. James	male		845 Q		0
1	McCarthy, Mr. Timothy J	male	54	5186 S		0
3	Palsson, Master. Gosta Leonard	male	2	2107 S		?
3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27	1113 S		?
2	Nasser, Mrs. Nicholas (Adele Achem)	female	14	3007 C		?
3	Sandstrom, Miss. Marguerite Rut	female	4	1670 S		?

target,
y

Dados de entrada para predição

Previsão, prediction, y_hat

Processo de Treinamento e Previsão

Dados de Treinamento:

- Usados no treinamento do modelo de aprendizado de máquina.
- Incluem entradas e rótulos para permitir que o modelo aprenda.

Dados de Validação:

- Não obrigatórios no treinamento do modelo.
- Podem ser utilizados para ajustar hiperparâmetros e evitar o overfitting.
- Contêm entradas e rótulos para avaliação, mas não para ajuste do modelo.

Dados de Teste:

- Não utilizados no treinamento do modelo.
- Reservados para avaliar o desempenho do modelo após o treinamento.
- Consistem em entradas e rótulos, usados exclusivamente para avaliação.



Validação vs Teste

VALIDATION DATASET

Propósito: O conjunto de validação é usado durante o processo de treinamento do modelo para tomar decisões sobre hiperparâmetros, arquitetura do modelo e outros aspectos do desenvolvimento do modelo.

Uso: Ele ajuda a avaliar o desempenho do modelo durante o treinamento e orienta ajustes para melhorar o desempenho em dados não vistos.

Modificações: Com base nos resultados da validação, você pode fazer alterações no modelo, como ajustar hiperparâmetros, adicionar regularização ou alterar a arquitetura.

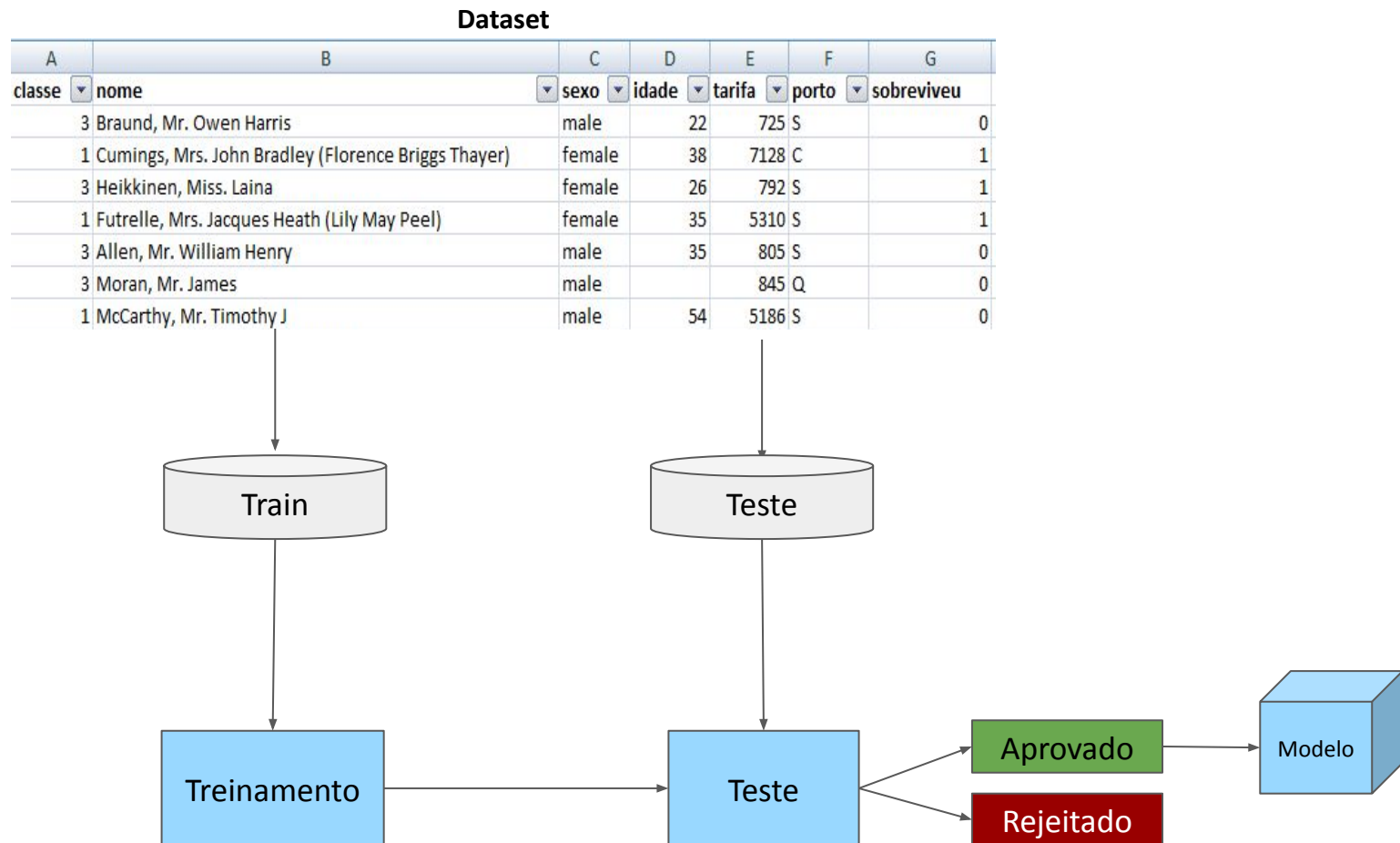
TEST DATASET

Propósito: O conjunto de teste é completamente separado e não é usado durante o processo de desenvolvimento do modelo. Seu único propósito é fornecer uma avaliação imparcial do desempenho de generalização do modelo.

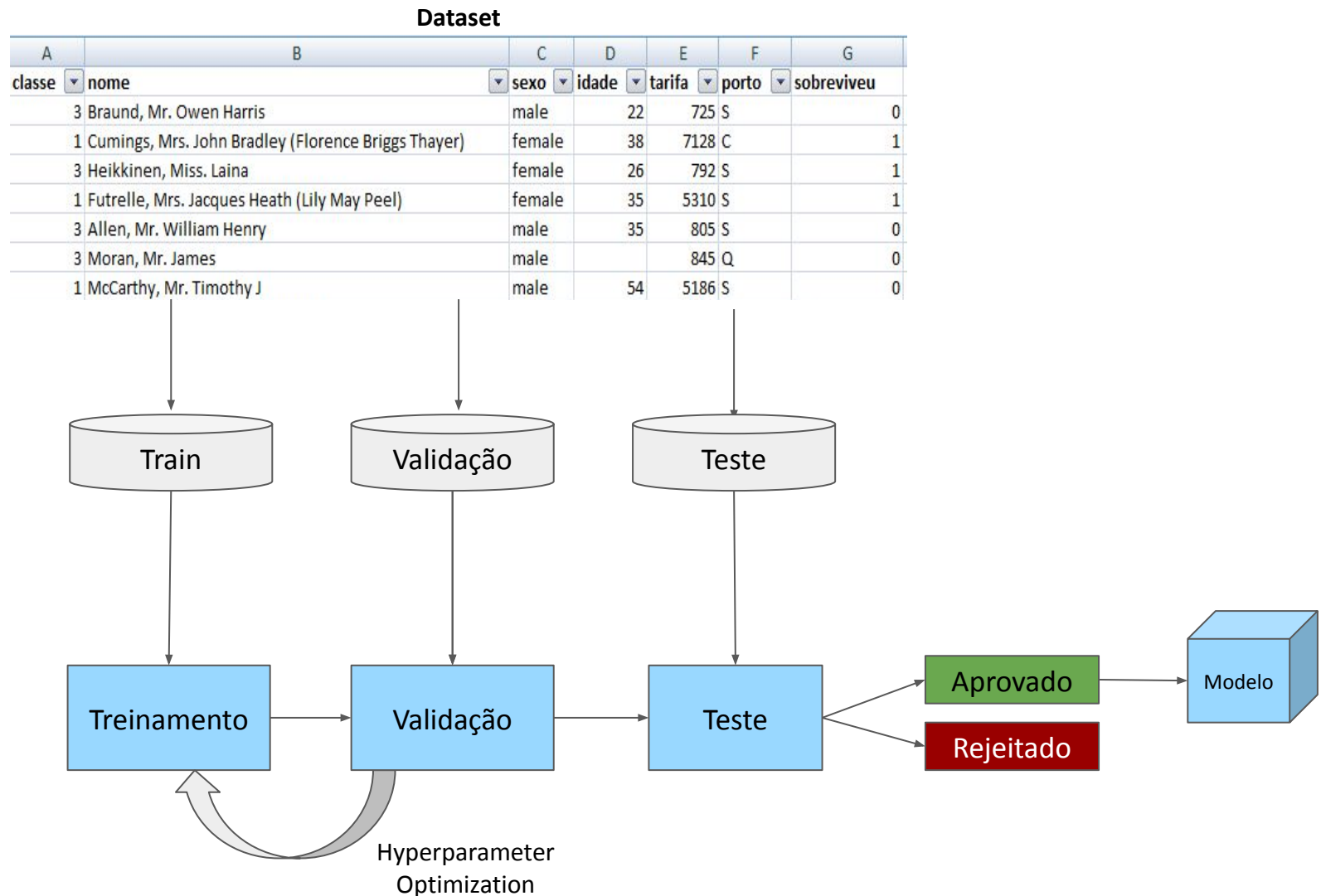
Uso: Ele é usado apenas após você ter concluído o desenvolvimento do modelo, incluindo a otimização de hiperparâmetros, com a ajuda do conjunto de validação.

Avaliação: O conjunto de teste simula cenários do mundo real ao avaliar o quão bem o modelo treinado se sai em dados que ele nunca viu antes. Ele fornece uma estimativa precisa de como o modelo se sairá em aplicações práticas.

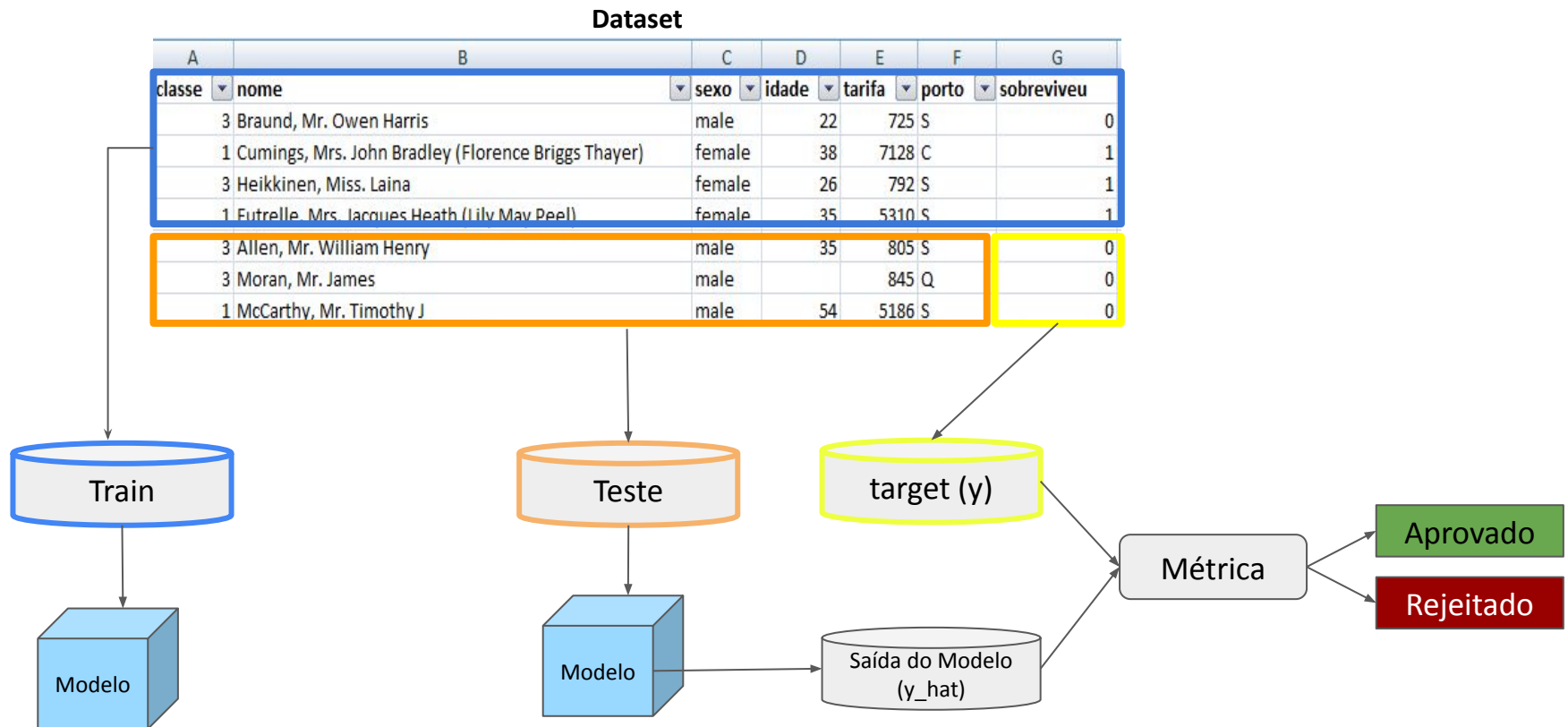
Processo de produção e utilização do modelo



Processo de produção e utilização do modelo



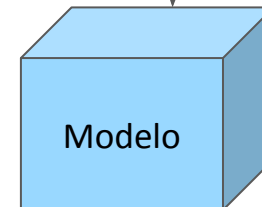
Processo de produção e utilização do modelo



Processo de produção e utilização do modelo

A	B	C	D	E	F
classe	nome	sexo	idade	tarifa	porto
3	Braund, Mr. Owen Harris	male	22	725	S

0. Preparação das Features
1. Treinamento do Modelo
2. Validação
3. Utilização (predição)



G
sobreviveu
0

Training vs Hyperparameter Optimization

TREINAMENTO

Processo que otimiza os parâmetros ‘treináveis’ do modelo.

Atualização de Parâmetros: Os parâmetros do modelo são atualizados usando técnicas de otimização, como descida de gradiente, para minimizar a perda.

OTIMIZAÇÃO DE HIPERPARÂMETROS

Processo que otimiza os parâmetros ‘estruturais’ do modelo.

Seleção de Hiperparâmetros: Diferentes configurações de hiperparâmetros, como taxa de aprendizado, número de camadas ocultas ou função de ativação, são escolhidas para experimentação.

Treinamento com Hiperparâmetros: Vários modelos são treinados com diferentes combinações de hiperparâmetros usando o conjunto de treinamento e validação.

Comparação de Modelos: Os modelos são comparados para determinar qual combinação de hiperparâmetros produz o melhor desempenho no conjunto de validação.

Seleção do Modelo Final: O modelo com os hiperparâmetros ideais, com base na avaliação no conjunto de validação, é selecionado como o modelo final.

Principais Ferramentas

Scikit-Learn: Biblioteca amplamente usada para aprendizado de máquina com uma ampla gama de algoritmos de classificação e regressão.

TensorFlow: Uma das principais bibliotecas de aprendizado de máquina da Google, usada para redes neurais e aprendizado profundo.

PyTorch: Biblioteca popular para pesquisa em aprendizado profundo e construção de modelos de redes neurais.

Keras: Interface de alto nível para redes neurais que roda em cima de frameworks como TensorFlow e Theano.

XGBoost: Implementação eficiente de Gradient Boosting para classificação e regressão.

LightGBM: Framework de Gradient Boosting com alta eficiência e escalabilidade.

CatBoost: Biblioteca de Gradient Boosting otimizada para lidar com variáveis categóricas.

H2O.ai: Plataforma de IA que inclui ferramentas para aprendizado de máquina supervisionado.

Caret: Pacote em R e python para treinamento e avaliação de modelos de aprendizado de máquina.

Quais são as outras técnicas de aprendizado?

Aprendizado Não Supervisionado

O que é Aprendizado Não Supervisionado?

É uma abordagem de Machine Learning que se concentra em descobrir padrões ou estruturas ocultas em **dados não rotulados**.

O principal objetivo é permitir que o algoritmo descubra padrões, estruturas ou relações intrínsecas nos dados sem orientação externa.

As principais aplicações de aprendizado não supervisionado são:

- Clusterização
- Redução de Dimensionalidade
- Redes Generativas Adversárias (GANs) - ★ **Generative AI** ★

Clusterização

O objetivo é agrupar dados semelhantes em clusters ou grupos, onde os elementos dentro de um cluster são mais parecidos entre si do que com elementos de outros clusters.

Funcionamento

- Os algoritmos de clusterização dividem o conjunto de dados em clusters, com base na similaridade ou proximidade dos pontos de dados.
- Não há supervisão ou rótulos nas amostras; o algoritmo deve identificar padrões automaticamente.

Clusterização

Segmentação de Clientes: O clustering ajuda as empresas a agrupar os clientes com base em seu comportamento de compra, dados demográficos ou preferências, possibilitando estratégias de marketing direcionadas.

Detecção de Anomalias: O clustering pode ser usado para identificar padrões ou anomalias incomuns nos dados, como detecção de fraude em transações financeiras ou segurança de rede.

Segmentação de Imagem: O clustering é empregado em visão computacional para segmentar imagens em regiões ou objetos, sendo útil em imagens médicas, reconhecimento de objetos e outros.

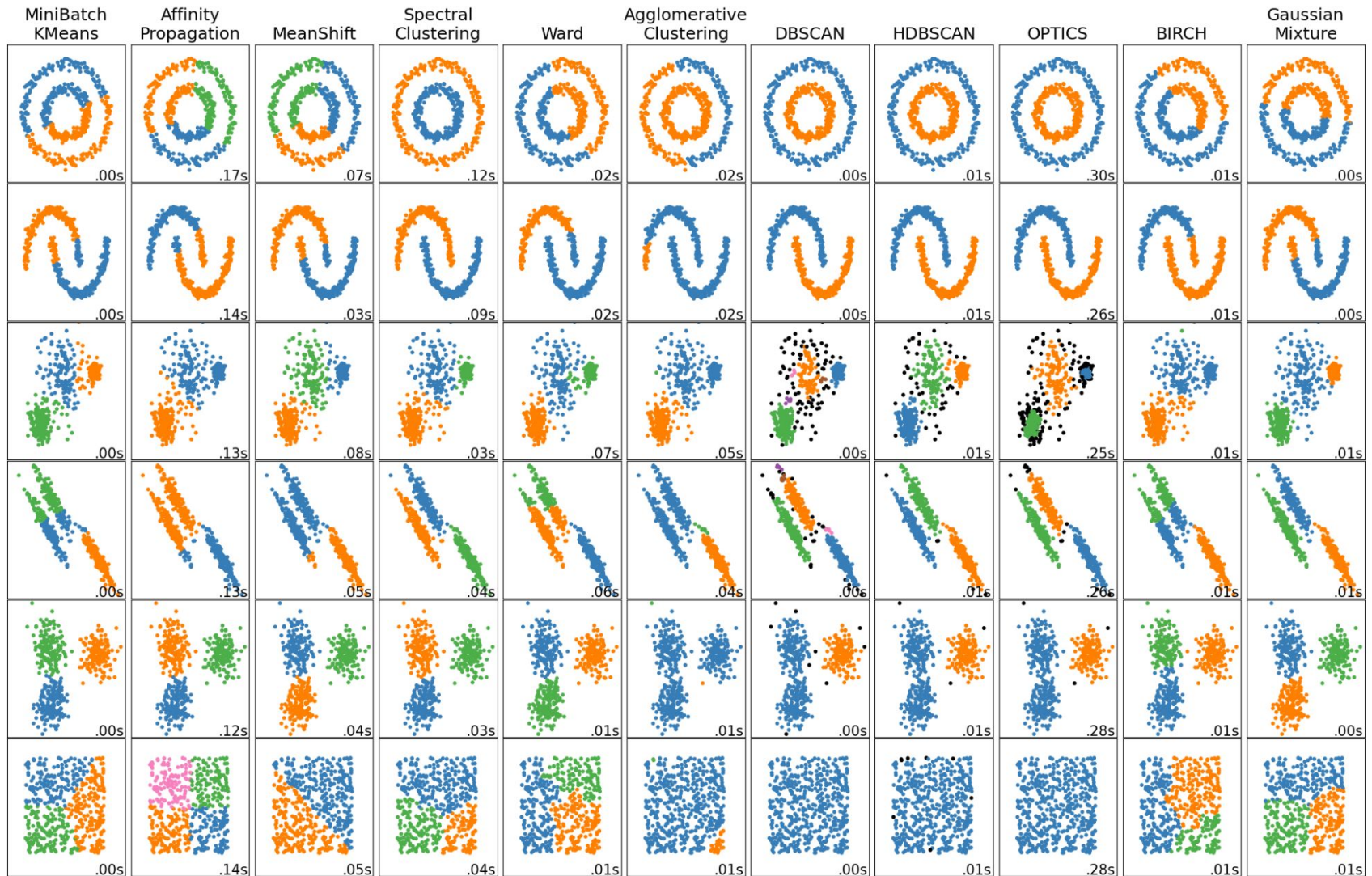
Sistemas de Recomendação: O clustering de usuários com preferências ou comportamentos semelhantes ajuda a construir sistemas de recomendação personalizados, como em comércio eletrônico e plataformas de conteúdo.

Clusterização de Documentos: Agrupar documentos por tópicos ou temas simplifica a organização de conteúdo, como na categorização de artigos de notícias ou classificação de e-mails.

Processamento de Linguagem Natural (NLP): O clustering é usado na análise de texto para modelagem de tópicos, análise de sentimentos e sumarização.

E muitos outros: Genômica e Bioinformática, Análise de Dados Espaciais, Reconhecimento de Fala, Análise de Redes, Controle de Qualidade e Manufatura, Segmentação de Mercado, Saúde, Astronomia

Clusterização - Performance de Diferentes Algoritmos



DEMONSTRAÇÃO

Clusterização de Clientes

 colab

Aprendizado Por Reforço

O que é Aprendizado por Reforço

É um paradigma de aprendizado de máquina no qual um **agente** autônomo interage com um **ambiente** e aprende a tomar **ações** que **maximizem** uma **recompensa** acumulada ao longo do **tempo**.

É comumente usado para treinar sistemas a tomar decisões sequenciais em situações dinâmicas e incertas.

Aplicações de robôs aprendendo sozinho a andar, executar tarefas, etc.

Algoritmos que aprendem a jogar algum jogo sozinho.

Como o Aprendizado por Reforço funciona

Agente e Ambiente: Reinforcement Learning envolve um "agente" (por exemplo, um robô, um programa de computador ou um sistema automatizado) que interage com um "ambiente" (o mundo real, um jogo de computador, uma simulação).

Observação: informações que o agente recebe do ambiente para tomar decisões. Usa esse estado para tomar a decisão.

Objetivo: O objetivo do agente é aprender a tomar ações de maneira que maximize uma recompensa acumulada ao longo do tempo.

Ações: escolhas que um agente faz para interagir com um ambiente.

Recompensas: Após cada ação, o agente recebe uma "recompensa," que pode ser positiva, negativa ou neutra, dependendo da qualidade da ação. Por exemplo, no contexto de negócios, uma recompensa poderia ser lucro, satisfação do cliente, eficiência operacional, etc.

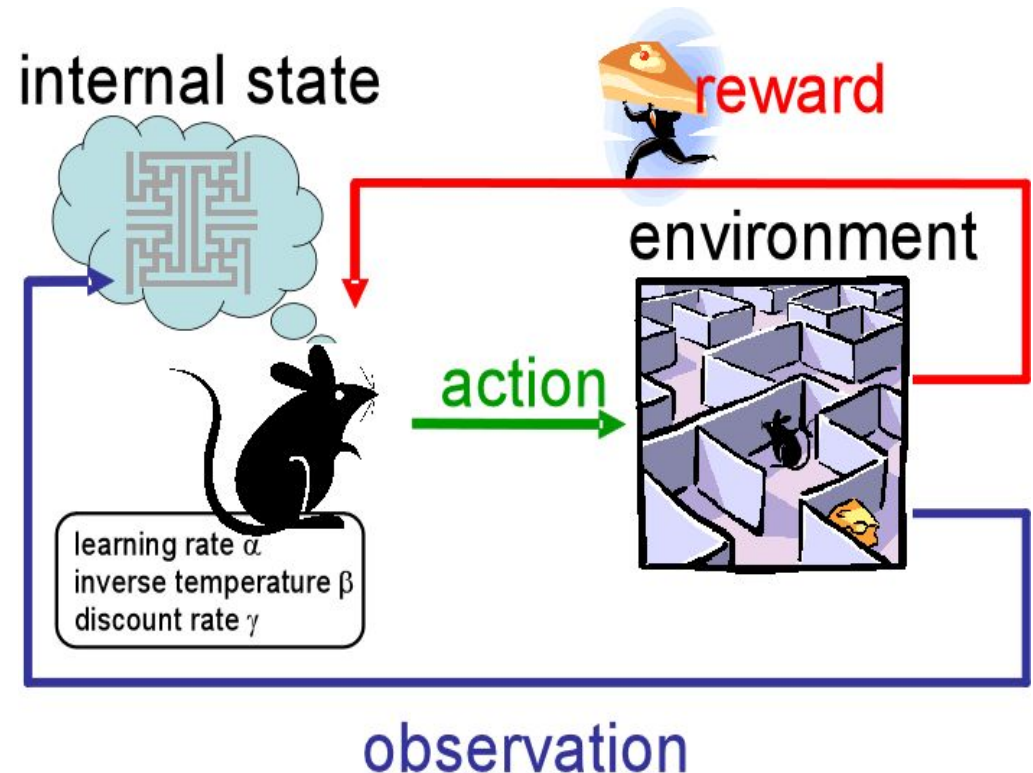
Exploração e Aproveitamento: "Exploração" (tentar ações desconhecidas) e "Aproveitamento" (escolher ações conhecidas e eficazes). Importante porque o agente precisa equilibrar a busca por novas estratégias com a escolha das melhores estratégias já aprendidas.

Treinamento: O agente aprende por tentativa e erro, realizando ações e recebendo recompensas. Com o tempo, ele ajusta sua política para melhorar o desempenho.

Como o Aprendizado por Reforço funciona

Os algoritmos que controlam o agente podem ser variados, inclusive redes neurais.

Os algoritmos que modelam as épocas (gerações) são algoritmos de otimização, como algoritmos genéticos.



[fonte](#)

Aplicações



Aplicações

Robótica: Empresas como a Boston Dynamics usam o aprendizado por reforço para ensinar robôs a executar tarefas complexas. Os robôs podem aprender a andar, correr e realizar tarefas físicas por meio de algoritmos de AR.

Veículos Autônomos: Empresas como Tesla, Waymo (subsidiária da Alphabet) e Uber utilizam aprendizado por reforço para treinar carros autônomos. Esses veículos aprendem a navegar, tomar decisões de direção e responder a várias condições de estrada autonomamente.

Finanças: O aprendizado por reforço é usado para negociação algorítmica e otimização de portfólio. Empresas como QuantConnect e Numerai aplicam o AR para desenvolver estratégias de negociação e aprimorar a tomada de decisões financeiras.

Automação Industrial: Empresas como a Siemens usam o aprendizado por reforço na automação fabril para otimizar os processos de produção e reduzir o tempo de inatividade, treinando robôs e máquinas para melhorar a eficiência.

Publicidade Online: Plataformas de publicidade usam RL para otimizar estratégias de lances e recomendação de anúncios. O Google Ads é um exemplo, onde o RL ajuda a definir lances ideais para palavras-chave em leilões de anúncios em tempo real.

Sistemas de Recomendação: Netflix e YouTube empregam o aprendizado por reforço para recomendações de conteúdo. Algoritmos otimizam quais programas ou vídeos sugerir com base no histórico de visualização e preferências dos usuários.

Produção de ML

"In machine learning, feature engineering is where most of the real work happens."
- Pedro Domingos,
author of "The Master Algorithm."

Produção de Machine Learning

Preparação e Planejamento

- Definir o Problema
- Coleta de Dados
- Limpeza e Pré-processamento de Dados
- Exploração de Dados (EDA)
- Divisão de Dados
- Feature Engineering
- Data augmentation
- Seleção de Modelo (AutoML)
- Problem Framing

Treinamento e Avaliação do Modelo

- Avaliação do Modelo
- Ajuste do Modelo (Tuning)
- Validação Final
- Implantação do Modelo

Monitoramento e Manutenção

- Monitoramento de Produção
- Manutenção e Atualização
- Documentação
- Comunicação de Resultados
- Governança de Dados e Ética

Preparação e Planejamento

1. Definir o Problema:

- Identifique claramente o problema de negócios que deseja resolver com machine learning.
- **Product Owner, Product Manager, Business Owner, Business Analyst, Data Scientist**

2. Coleta de Dados:

- Reúna os dados relevantes para o problema. Isso pode envolver fontes internas ou externas.
- **Engenheiro de Dados**

3. Limpeza e Pré-processamento de Dados:

- Elimine valores ausentes, outliers e ruído nos dados.
- Padronize, normalize ou crie recursos a partir dos dados brutos.
- **Engenheiro de Dados, Cientista de Dados**

4. Exploração de Dados (EDA):

- Realize análise exploratória de dados para entender as características dos dados.
- Visualize os dados e identifique tendências e insights preliminares.
- **Cientista de Dados, Analista de Dados**

Pré-processamento e Limpeza de Dados

Exploração de Dados (EDA):

- Realize uma análise exploratória de dados para entender a estrutura e as características dos dados.
- Visualize os dados por meio de gráficos e estatísticas descritivas.
- Identificação de features correlacionadas, desbalanceadas, incompletas, outliers, tipos de dados, formato dos dados.

Lidando com Dados Ausentes:

- Identifique e trate valores ausentes de forma apropriada, seja por remoção, imputação de valores ou técnicas avançadas, como preenchimento com média ou mediana.

Lidando com Outliers:

- Detecte e lide com valores atípicos que possam distorcer os resultados do modelo. Isso pode envolver remoção, transformação ou imputação.
- Os outliers podem significar algo, como podem ser apenas sujeira (ex: idade: 150 anos).

Balanceamento de Classes (se necessário):

- Se estiver trabalhando com um conjunto de dados desequilibrado, considere técnicas de balanceamento, como oversampling ou undersampling, para evitar viés no modelo.

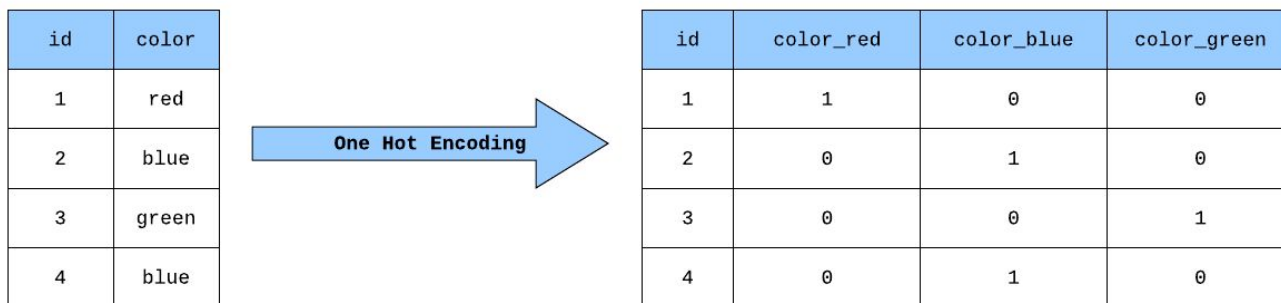
Preparação e Planejamento

5. Divisão de Dados:
 - a. Separe os dados em conjuntos de treinamento, validação e teste.
 - b. **Cientista de Dados**
6. Feature Engineering:
 - a. Crie ou selecione recursos relevantes para o modelo.
 - b. **Cientista de Dados**

Feature Engineering

Codificação de Variáveis Categóricas:

- Converta variáveis categóricas em representações numéricas que possam ser usadas em modelos. Isso inclui técnicas como **one-hot encoding**, label encoding e frequency encoding.



Feature Engineering

Codificação de Variáveis Categóricas:

- Converta variáveis categóricas em representações numéricas que possam ser usadas em modelos. Isso inclui técnicas como one-hot encoding, **label encoding** e frequency encoding.

SAFETY-LEVEL (TEXT)	SAFETY-LEVEL (NUMERICAL)
None	0
Low	1
Medium	2
High	3
Very-High	4

Height	Height
Tall	0
Medium	1
Short	2

Melhor para variáveis categóricas ordinárias.

Feature Engineering

Codificação de Variáveis Categóricas:

- Converta variáveis categóricas em representações numéricas que possam ser usadas em modelos. Isso inclui técnicas como one-hot encoding, label encoding e **frequency encoding**.

Numerical value	Animal		Numerical value	Animal_freq
1.5	cat	Frequency encoding →	1.5	0.5
3.6	cat		3.6	0.5
42	dog		42	0.25
7.1	crocodile		7.1	0.25

Melhor para variáveis categóricas de alta cardinalidade.

Data Augmentation

Modificação dos dados atuais para criar mais dados.

Vantagem é criar novas representações de um dado para um **mesmo label**.

Data augmentation pode acontecer em datasets do tipo:

- Imagem
- Texto
- Áudio
- Séries temporais

Data Augmentation

Imagem

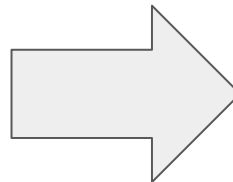
- Rotation, scaling, and translation.
- Flipping (horizontal and vertical).
- Cropping and padding.
- Gaussian noise and blur.
- Color jittering and brightness adjustments.
- Shearing and perspective transformations.

Data Augmentation

Imagem



Aquisição de Dados



Anotação dos Dados

Data Augmentation

Text Data Augmentation:

- Synonym replacement.
- Random word insertion.
- Random word deletion.
- Random word swapping.
- Back translation (translating text to another language and back).

Audio Data Augmentation:

- Time stretching and compression.
- Pitch shifting.
- Adding background noise.
- Speed perturbation.

Augmentation Techniques for Time-Series Data:

- Time warping.
- Smoothing and filtering.
- Resampling and interpolation.

Preparação e Planejamento

7. Seleção de Modelo:

- a. Escolha o algoritmo de machine learning adequado para o problema com base na análise exploratória.
- b. Problem Framing: Regressão, Classificação, Previsão?
- c. **Cientista de Dados**

8. Treinamento Inicial:

- a. Treine um modelo inicial usando o conjunto de treinamento. Geralmente com AutoML.
- b. **Cientista de Dados**

Preparação e
Planejamento

Treinamento e Avaliação
do Modelo

Avaliação do Modelo:

- Use o conjunto de validação para avaliar o desempenho do modelo.
- Utilize métricas relevantes, como precisão, recall, F1-score ou erro médio quadrático (RMSE).
- **Cientista de Dados**

Ajuste do Modelo (Tuning):

- Ajuste os hiperparâmetros do modelo para melhorar o desempenho.
- **Cientista de Dados**

Validação Final:

- Avalie o modelo em um conjunto de teste independente para obter uma avaliação final.
- **Product Owner, Product Manager, Business Owner, Business Analyst, Data Scientist**

Preparação e Planejamento

Treinamento e Avaliação do Modelo

Monitoramento e Manutenção

Implantação do Modelo:

- Implante o modelo em um ambiente de produção, como um servidor web, aplicativo ou sistema.
- **Engenheiro de ML (MLops)**

Monitoramento de Produção:

- Monitore o desempenho do modelo em produção para detecção de drift de dados, degradação de modelo ou comportamento anormal.
- **Engenheiro de ML (MLops), Cientista de Dados**

Manutenção e Atualização:

- Faça atualizações periódicas do modelo com novos dados e reentrene-o conforme necessário.
- **Engenheiro de ML (MLops), Cientista de Dados**

Documentação:

- Documente todo o processo, desde a coleta de dados até a implantação do modelo.

Comunicação de Resultados:

- Comunique continuamente os resultados e insights derivados do modelo para as partes interessadas.
- **Product Owner, Product Manager, Business Owner, Business Analyst, Data Scientist**

Governança de Dados e Ética:

- Mantenha práticas de privacidade de dados, segurança e ética ao lidar com dados sensíveis.
- **Product Owner, Product Manager, Business Owner, Business Analyst, Data Scientist**

Dúvidas?

Avaliação

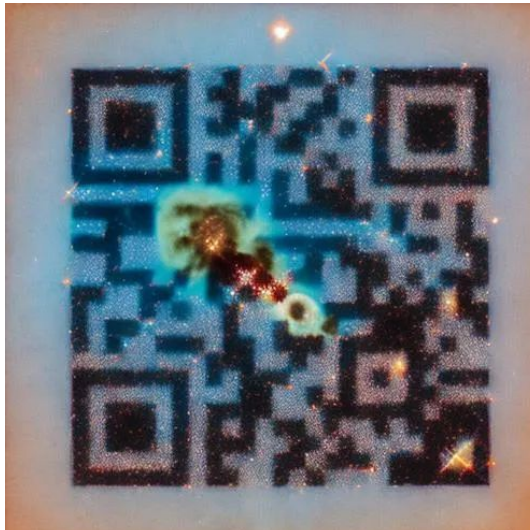


[Link para o Formulário](https://forms.gle/1o4RWqi1xYAY8pfp7)

<https://forms.gle/1o4RWqi1xYAY8pfp7>

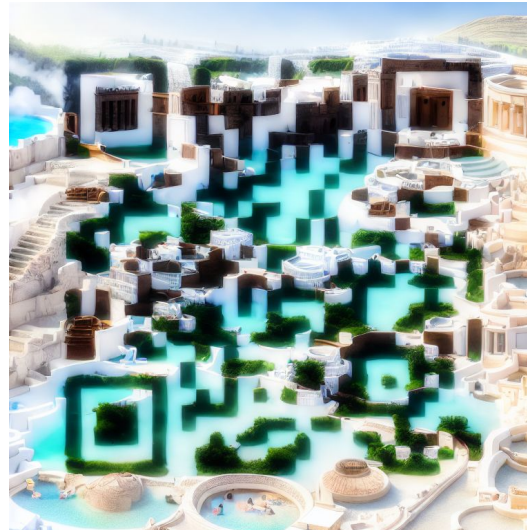
Qual é a aula que está avaliando? *

- ☐ Aula 1
- ☐ Aula 2
- ☐ Aula 3
- ☐ Aula 4
- ☐ Aula 5



[openart](#)

“deep space as seen from the Hubble telescope”



[huggingface](#)

“Sky view of highly aesthetic, ancient greek thermal baths in beautiful nature”

