

Lógica para Predição com Inteligência Artificial

Módulo 3: Comum 1º Ano

- RESUMO EXECUTIVO

O módulo Lógica para Predição com Inteligência Artificial capacita equipes a construir modelos preditivos a partir de bases de dados reais de empresas parceiras. Em dez semanas, os alunos aprendem a explorar, tratar e modelar dados usando técnicas básicas de aprendizado de máquina supervisionado e não supervisionado, comunicando os resultados em relatórios analíticos voltados para a tomada de decisão. O produto final é um processo estruturado de treinamento de modelo funcional, com modelo pronto, análise técnica de performance e recomendações práticas para o parceiro.

- O DESAFIO

Organizações acumulam grandes volumes de dados operacionais, de clientes e de mercado, mas poucos conseguem transformar esses dados em previsões confiáveis que orientem decisões estratégicas. A distância entre ter dados e extrair valor deles é real: faltam profissionais que dominem ao mesmo tempo o ferramental técnico de análise e a capacidade de traduzir resultados em linguagem de negócio. Modelos construídos sem essa ponte tendem a ficar em relatórios internos sem impacto prático.

- OBJETIVO DO MÓDULO

Formar equipes capazes de conduzir um projeto completo de ciência de dados, desde a análise exploratória até a entrega de um modelo preditivo explicável para um parceiro real. O módulo cobre estatística aplicada, álgebra linear, aprendizado supervisionado e não supervisionado, séries temporais, visualização de dados e aspectos de governança e ética no uso de dados. As equipes também desenvolvem habilidades de comunicação para apresentar achados a audiências sem perfil técnico.

- PROBLEMAS QUE PODEMOS RESOLVER

- Previsão de demanda imprecisa que gera excesso de estoque ou rupturas em cadeias de suprimento.
- Identificação de clientes com maior probabilidade de cancelamento (churn) antes que a saída aconteça.
- Detecção de anomalias em processos operacionais ou transações financeiras que indicam fraude ou falha.
- Segmentação de base de clientes para personalização de ofertas e comunicações de marketing.
- Apoio à tomada de decisão clínica ou de saúde pública usando dados estruturados de prontuários ou pesquisas.

- POSSÍVEIS CASOS DE APLICAÇÃO

1. Casos de Sucesso Reais (O que já entregamos)

- Retenção de Clientes, Atendimento e Recursos Humanos (Churn & Turnover):

- Desenvolvemos com o BCG (Boston Consulting Group) um modelo preditivo avançado de machine learning para a identificação precoce de propensão ao cancelamento (*churn*) de clientes, auxiliando na definição de campanhas de retenção ativa.
- **Finanças, Gestão de Crédito e Avaliação de Riscos:**
- Construímos para a CVM (Comissão de Valores Mobiliários) um algoritmo preditivo para a avaliação e mitigação de riscos operacionais em Fundos de Investimento em Direitos Creditórios (FIDCs), antecipando desvios e inconformidades regulatórias.
- **Operações, Logística e Eficiência de Frotas:**
- Estruturamos para a ATIVOS um modelo matemático-preditivo focado na precificação inbound de fretes de cana-de-açúcar, mapeando correlações históricas de custos operacionais entre os períodos de safra e entressafra.
- **Vendas, Receitas e Posicionamento de Mercado:**
- Criamos para a Chilli Beans um modelo preditivo focado no reposicionamento estratégico e impulso de conversão de vendas para a categoria de lentes de grau no varejo óptico.
- **Saúde e Gestão Educacional:**
- Criamos para a Faculdade de Medicina da USP (Oncologia) modelos preditivos para o estudo epidemiológico e previsão de risco de câncer feminino, além de sistemas estatísticos voltados à recomendação dos melhores tratamentos clínicos com base em históricos de diagnósticos.

2. Outras Aplicações Possíveis (Onde mais podemos gerar valor)

- Podemos construir modelos de regressão linear para a projeção e previsão de vendas sazonais e receita bruta mensal, auxiliando indústrias e varejistas no planejamento estratégico de médio prazo.
- Podemos estruturar sistemas inteligentes de recomendação de produtos com base em algoritmos de agrupamento (*clustering*) por similaridade de comportamento de compra, elevando o *ticket* médio de e-commerce e portfólios B2B.
- Podemos desenvolver classificadores estatísticos de score de crédito para fintechs e bancos, avaliando instantaneamente o risco financeiro de proponentes com base em variáveis cadastrais e de comportamento transacional.
- Podemos projetar detectores de anomalias em dados de sensores físicos de maquinários pesados (indústrias, utilidades e portos), prevendo de forma diagnóstica a necessidade de manutenções preventivas antes da falha efetiva.

- POSSÍVEIS ENTREGÁVEIS

- Notebook (.ipynb) documentado com análise exploratória de dados, tratamento, feature engineering e modelagem.
- Modelo preditivo treinado e avaliado, com métricas de desempenho explicadas em linguagem acessível.
- Aplicação interativa com visualização dos principais insights e resultados do modelo.
- Relatório de explicabilidade do modelo, indicando as variáveis de maior influência nas previsões.
- Análise de conformidade com LGPD aplicada ao uso dos dados do parceiro.

- Apresentação executiva com recomendações práticas de uso do modelo no contexto de negócio do parceiro.

- PREMISSAS E REQUISITOS

1. Da Empresa Parceira:

- Disponibilizar base de dados real ou anonimizada com volume suficiente para treinamento e validação do modelo.
- Em caso de modelos de aprendizado supervisionado, a base de dados deve possuir uma ou mais colunas-alvo¹.
- Garantir que os dados fornecidos estejam em conformidade com a LGPD ou orientar sobre as restrições aplicáveis.
- Designar um ponto focal de negócio para alinhar o problema a ser resolvido e validar as hipóteses ao longo do módulo.
- Disponibilizar contexto sobre como a predição seria usada na prática, para guiar as escolhas de modelagem.

2. Do Projeto:

- As equipes trabalham com Python e bibliotecas amplamente adotadas na indústria (Numpy, Pandas, Scikit-Learn, Matplotlib).
- O modelo entregue é um protótipo validado, não um sistema em produção; a integração com sistemas do parceiro é de responsabilidade dele.
- A qualidade do modelo depende diretamente da qualidade e volume dos dados fornecidos pelo parceiro.
- As equipes seguem ciclo ágil de sprints quinzenais com entregas intermediárias e revisão com o parceiro.

- ESCOPO E RESTRIÇÕES

1. Escopo Incluído:

- Análise exploratória e tratamento da base de dados fornecida pelo parceiro.
- Seleção, treinamento e avaliação de modelos de aprendizado de máquina supervisionado ou não supervisionado.
- Análise de explicabilidade do modelo e identificação das variáveis mais relevantes.
- Revisão de conformidade com LGPD aplicada ao conjunto de dados utilizado.
- Criação opcional de aplicação com visualizações para comunicação dos resultados.

2. Fora do Escopo:

- Implantação do modelo em ambiente de produção ou integração com sistemas do parceiro.
- Coleta ou curadoria de novos dados além do conjunto fornecido pelo parceiro.

¹ Dentro do Aprendizado de Máquina Supervisionado, a coluna-alvo é simplesmente a "resposta certa" que você quer que o modelo aprenda a prever.

Imagine uma planilha com dados sobre carros usados:

- Colunas normais (características): Ano, marca, quilometragem e número de portas.
- Coluna-alvo (resultado): O Preço do carro.

Durante o treino, o modelo analisa o preço junto com as outras informações para entender a regra do jogo (por exemplo, "quanto mais novo e menos rodado, mais caro"). Depois de treinado, ao receber os dados de um carro novo sem preço, ele usa o que aprendeu para adivinhar esse valor sozinho.

- Desenvolvimento de APIs ou interfaces para consumo do modelo por outros sistemas.
- Retreinamento ou monitoramento contínuo do modelo após o encerramento do módulo.
- Projetos que envolvam dados sensíveis sem anonimização prévia, processada e aprovada pelo parceiro.

- **POSSÍVEIS TECNOLOGIAS / ARQUITETURA / FERRAMENTAS DE NEGÓCIOS**

- Python com Pandas, NumPy e Scikit-Learn para análise e modelagem.
- Matplotlib e Seaborn para visualização de dados; Streamlit para aplicações de visualização de dados e resultados.
- Jupyter Notebooks para documentação interativa do processo de análise.
- Git e GitLab para versionamento do código e dos notebooks.
- Ferramentas de governança: checklist LGPD e documentação de linhagem de dados.
- Frameworks de negócio: 5 forças de Porter, Análise SWOT, Canvas de Proposta de Valor, Matriz de Risco.