

# Lista de Verificação de Manutenção de Dados

Bem-vindo à era da Inteligência Artificial (IA), em que a forma como você faz negócios depende de tecnologias com uso intensivo de dados, como *machine learning* e *deep learning*. Para tirar proveito dessas novas ferramentas de IA, é preciso garantir que a “casa de dados” da sua organização esteja em ordem.

Aqui está uma lista de verificação para você dar os primeiros passos rumo a uma casa de dados limpa, dividida nas duas principais fases de manutenção: treinamento e conclusão.

Siga estas etapas para se tornar um mestre em IA. Para obter mais informações de como passar da prova de conceito de IA à produção e escala completas, confira este relatório da IDC:

[Acelere e Operacionalize Implementações de IA Utilizando - Infraestrutura Otimizada de IA.](#)

## Treinamento

Na parte do treinamento que trata da preparação para IA, você desenvolverá algoritmos para compreender um conjunto de dados. O foco principal será a coleta de dados existentes e a utilização de IA para o aprendizado de uma nova habilidade.

- Determine o problema de negócios específico que você deseja solucionar usando IA (comece com projetos menores para ajudar no seu aprendizado)
- Localize os dados que podem solucionar o problema em fontes relevantes (é bastante provável que nem todos os dados estejam localizados em um só lugar)
- Prepare os seus dados com tags de metadados para reduzir significativamente o tempo necessário para encontrar dados pertinentes
- Certifique-se de que os seus dados estejam devidamente sincronizados e vinculados em todos os conjuntos de dados que serão utilizados (inclusive a sincronização de tempo)
- Sinalize todos os dados confidenciais de clientes e outros dados privados para garantir que você os esteja mantendo em absoluta segurança e esteja respeitando plenamente a devida governança e regulamentação (o processo de marcação de metadados pode ajudar nesse quesito)
- Escolha o ambiente de desenvolvimento correto para o tipo de dados que você está usando e o modo como os dados serão formatados (isto é, imagens, vídeo, texto de forma livre e áudio – cada um deles costuma ter um tipo de ambiente)
- Extraia os conjuntos de dados do seu repositório e coloque-os no ambiente de desenvolvimento
- Divida os dados em dois grupos para ajudar a melhorar o processo de desenvolvimento do modelo (mantenha um conjunto em uma pasta chamada “treinamento” e outro conjunto em uma pasta chamada “teste”)
- Mantenha a rastreabilidade dos dados acompanhando de onde e de que fontes eles vieram (considere o uso de ferramentas que possam ajudar a automatizar o processo)
- Execute as tarefas básicas de higienização de dados para preparar os dados e criar um modelo (por exemplo, inclua o preenchimento de entradas de dados ausentes e a remoção de entradas nulas)
- Use uma amostra de dados do subconjunto da qual você já sabe a resposta (isso é chamado de “conjunto de treinamento”) e identifique todas as etapas de pré-processamento necessárias para preparar os dados para fazer uma previsão
- Use os seus conhecimentos sobre esse conjunto de treinamento para calcular pontuações de precisão que podem lhe dar confiança para aplicar o mesmo modelo a novos dados para os quais o modelo nunca tenha sido explicitamente treinado

## Conclusão

Depois de desenvolver um modelo que funcione para solucionar o seu problema de negócios, você passará do treinamento para a conclusão. Nessa fase, você deve pegar esse modelo bem-sucedido e aplicá-lo em dados novos, o que requer também uma certa manutenção contínua de dados.

- Localize o seu modelo de IA próximo dos seus dados para reduzir a latência, reduzir os requisitos de largura de banda e melhorar a performance geral do modelo
- Desenvolva um processo de pipeline de dados eficiente e aplique a rotulagem de metadados aos seus dados conforme eles entram, de modo que novos dados possam ser coletados e usados para aprimorar o modelo daqui para frente
- Marque os dados de forma que fiquem vinculados e sincronizados (por exemplo, se os dados forem sequenciados no tempo, será possível sincronizar os conjuntos de dados ou vincular selecionando um campo, como o nome de um cliente, em todos os dados que entram)
- Desenvolva um plano em longo prazo para o armazenamento do ciclo de vida dos dados para saber como você gerenciará o volume e a velocidade dos dados conforme entram e são arquivados
- Considere contratar um Diretor de Dados para a manutenção do gerenciamento de dados da sua organização para IA, deeplearning e outros projetos futuros orientados a dados