

Projeto Final do Bootcamp CDIA

Início: 25/08/2025 18:00:00

Prazo de entrega do código: até 01/09/2025 23:59:59

Entrega do vídeo: 01/09/2025 até 05/09/2025 23:59:59

Contextualização



Uma empresa do setor industrial contatou você para a criação de um sistema inteligente de manutenção preditiva das suas diferentes máquinas. Essa empresa forneceu um conjunto de dados contendo informações coletadas a partir de dispositivos IoT sensorizando atributos básicos de cada máquina. O objetivo é criar um sistema capaz de identificar as falhas que venham a ocorrer, e se possível, qual foi o tipo da falha. Cada amostra no conjunto de dados é composta por 8 atributos que descrevem o comportamento de desgaste da máquina e do ambiente. Além dessas características, cada amostra é rotulada com uma das 5 possíveis classes de defeitos.

O sistema deverá ser capaz de, a partir de uma nova medição do dispositivo IoT (ou conjunto de medições), prever a classe do defeito e retornar a probabilidade associada. Além disso, a

empresa espera que você extraia insights da operação e dos defeitos e gere visualizações de dados.

Descrição dos dados

Serão disponibilizados dois arquivos de dados:

- **Bootcamp_train.csv** - use o para explorar, treinar, avaliar seus modelos.
- **Bootcamp_test.csv** - esse arquivo não contém os labels, gere as predições usando seu modelo e use a seguinte API para ver o desempenho final do modelo.

Campo	Descrição
id	Identificador das amostras do banco
id_produto	Identificador único do produto. Combinação da variável Tipo e um número de identificação
tipo	Tipo de produto/máquina (L/M/H)
temperatura_ar	Temperatura do ar no ambiente (K)
temperatura_processo	Temperatura do processo (K)
umidade_relativa	Umidade relativa do ar (%)
velocidade_rotacional	Velocidade rotacional da máquina em rotações por minutos (RPM)
torque	Torque da máquina em Nm
desgaste_da_ferramenta	Duração do uso da ferramenta em minutos
falha_maquina	Indica se houve falha na máquina (1) ou não (0)
FDF (Falha Desgaste Ferramenta)	Indica se houve falha por desgaste da ferramenta (1) ou não (0)
FDC (Falha Dissipacao Calor)	Indica se houve falha por dissipação de calor (1) ou não (0)
FP (Falha Potencia)	Indica se houve falha por potência (1) ou não (0)
FTE (Falha Tensao Excessiva)	Indica se houve falha por tensão excessiva (1) ou não (0)
FA (Falha Aleatoria)	Indica se houve falha aleatória (1) ou não (0)

Orientações e reflexões sobre o problema

- A descrição do problema é curta e com poucas informações. Pense sobre o problema e como o tipo de solução que você desenvolverá pode gerar um grande impacto na indústria.
- Dados reais de empresas possuem erros (tanto de sensores, digitação, ...), como estão os dados disponíveis no desafio? Faça sua própria análise.
- Que tipo de problema/modelagem faz sentido nesse caso? Na API de visualização das métricas está considerando um problema **multirrótulo** (essencialmente uma classificação binária para cada classe de falha), isso faz sentido? Você está livre para modelar o problema da forma que preferir (binário, multiclasse, multirrótulo), mas descreva suas razões para a escolha.
- Qual a métrica que faz mais sentido para avaliar o problema que você definiu? Uma ou mais métricas?
- Como estão os rótulos do conjunto de dados? Há desbalanceamento entre as classes positivas e negativas? Investigue e apresente suas conclusões.
- Não avalie direto no conjunto de testes pela API, considere os dados da API como se fossem indisponíveis para você durante o treinamento e avaliação dos modelos. Quando tiver um bom modelo, use para uma avaliação final.

Avaliação

Realize uma apresentação documentando o processo de gestão de atividades do projeto, a metodologia utilizada, o desenvolvimento, os resultados obtidos, os aprendizados e os próximos passos.

Para a apresentação, **grave um vídeo de no máximo 10 minutos e envie o arquivo no sistema ava.**

Mais do que os resultados em si, o principal objetivo deste desafio é demonstrar de forma clara e objetiva durante a apresentação, todo o processo de pensamento e planejamento do projeto.

Itens avaliados na entrega do projeto:

- Desenvolvimento e organização do código para resolver o problema proposto;
- Documentação da solução (código, modelagem e diagramas);
- Interpretação clara e precisa do problema;

- Análise exploratória e tratamento adequado dos dados fornecidos, envolvendo análises estatísticas e tratamento dos dados;
- Metodologia de treinamento e avaliação dos modelos para o problema do projeto;

Itens avaliados na apresentação:

- Apresentação clara do projeto e defesa do ponto de vista;
- Habilidade de storytelling para engajar o público e comunicar a solução de forma eficaz;
- Justificativas para cada tomada de decisão;
- Qualidade das visualizações dos dados e dos resultados apresentados.

Qualquer código desenvolvido e utilizado para o conteúdo da apresentação deve ser disponibilizado em um **repositório do Github** de forma pública e o **link compartilhado no sistema AVA**:

Requisitos

Use de todo o conhecimento adquirido durante as aulas do bootcamp e mostre as suas competências de exploração de dados, modelagem preditiva, engenharia de software, storytelling, entre outros!

Não se esqueça dos seguintes itens:

1. Elaboração do vídeo de apresentação dos resultados para a banca;
2. Uso da linguagem de programação Python;
3. Disponibilização do código em repositório público no Github.

Obs: É permitido o uso de qualquer IDE, biblioteca de aprendizado de máquina ou de uso geral;

Extra

Os seguintes itens não são requisitos do desafio, mas seriam diferenciais se houver tempo.

- Organização do repositório em scripts python documentados em vez de jupyter notebooks;
- Disponibilização do modelo via RestAPI (qualquer biblioteca pode ser utilizada, como por exemplo, a [FastAPI](#))

- Desenvolvimento de um dashboard para visualização das análises e/ou uso do modelo (ex: usando [Streamlit](#))
- Preparar o código para ser rodado com Docker;
- Qualquer outra coisa que você queira trazer! (código limpo, uso da nuvem, CI/CD, MLOps, ...)