

CERTIFICACAO DP-100: Designing and Implementing a Data Science Solution on Azure

OBJETIVO:

Validar habilidades em modelagem, treinamento, avaliacao e operacionalizacao de solucoes de Machine Learning usando Azure.

☐ TOPICOS PRINCIPAIS PARA ESTUDO:

1. Preparacao de Dados (15–20%)

- Explorar e transformar dados usando Azure Machine Learning e notebooks.
- Trabalhar com conjuntos de dados tabulares, arquivos e imagens.
- Pre-processamento com sklearn, pandas e PySpark.

2. Modelagem (25–30%)

- Treinamento de modelos com scripts e AutoML.
- Escolha de metricas adequadas (RMSE, MAE, accuracy, etc).
- Validacao cruzada e tunagem de hiperparametros.

3. Avaliacao de Modelos (20–25%)

- Avaliar modelos de regressao, classificacao e clustering.
- Interpretabilidade com SHAP, Fairlearn e visualizacoes.
- Exportar modelos treinados (pkl, ONNX).

4. Implementacao e Operacionalizacao (30–35%)

- Registro de modelos com MLflow.
- Deployment como Web Service (ACI, AKS).
- Monitoramento, logging e versionamento.
- Pipelines automatizados com Azure ML Pipeline.

FERRAMENTAS & TECNOLOGIAS

- Azure Machine Learning Studio
- Python e Jupyter Notebooks
- Azure CLI e Azure ML SDK
- AutoML
- MLflow
- Streamlit (para interfaces)
- FAISS ou similares (para busca vetorial)

□ DICAS FINAIS

- Estude com prática: crie projetos reais (como o chatbot de PDFs).
- Explore os laboratórios práticos do Microsoft Learn.
- Use o Azure com conta trial ou conta de estudante gratuita.
- Revise os objetivos da prova no site oficial antes de agendar.