Curso: Machine Learning Engineer

Duração: 21h
Área formativa: Cursos

Sobre o curso

Aprende a transformar modelos de Machine Learning em soluções escaláveis, monitorizadas e prontas para produção.

O papel do Machine Learning Engineer é fundamental para transformar modelos experimentais em sistemas fiáveis, escaláveis e sustentáveis em contexto empresarial.

Neste curso, trabalhas todo o ciclo de vida do Machine Learning em produção: pipelines de dados, preparação de modelos, MLOps, deployment, monitorização e gestão contínua. A abordagem é orientada à engenharia, com foco em ambientes reais e integração com equipas de desenvolvimento e operações.

No final, consegues construir, versionar, colocar em produção e monitorizar modelos de ML de forma profissional.

Metodologia

:: Sessões práticas orientadas a casos reais

:: Exercícios técnicos guiados

:: Construção progressiva de pipeline end-to-end

:: Workshop final integrador

Objectivos

Ao longo do curso vais aprender a:

:: Aplicar o ciclo de vida completo de Machine Learning em produção;

:: Desenhar e implementar pipelines de dados escaláveis;

:: Desenvolver modelos orientados à produção;

:: Validar, otimizar e preparar modelos para deployment;

:: Aplicar práticas de MLOps (versionamento, automação, reprodutibilidade);

:: Colocar modelos em produção via API ou container;

:: Monitorizar modelos e gerir drift ao longo do tempo:

:: Comunicar decisões técnicas a diferentes stakeholders.

Pré-requisitos

:: Conhecimentos básicos de Python;

:: Noções gerais de dados e estatística descritiva;

:: Familiaridade com notebooks, IDEs, linha de comandos ou APIs (nível introdutório).

Destinatários

:: Profissionais de TI ou Engenharia de Software que pretendem trabalhar com ML em produção;

:: Data Scientists que querem evoluir para engenharia e operação de modelos;

:: Data Engineers que necessitam integrar pipelines de dados com ML;

:: Programadores e analistas técnicos envolvidos em projetos de IA;

:: Profissionais que pretendem desempenhar funções de ML Engineer ou colaborar em equipas de MLOps.

Programa

Fundamentos de Machine Learning e Papel do ML Engineer

Este módulo enquadra o papel do Machine Learning Engineer no ciclo de vida de ML, distinguindo modelos experimentais de soluções prontas para produção. Analisa responsabilidades, ferramentas essenciais e desafios reais associados a dados imperfeitos, integração com equipas técnicas e monitorização contínua em ambiente empresarial.

- O que é Machine Learning Engineering?
- Diferença entre Data Scientist, Data Engineer, ML Engineer e AI Engineer
- ML Lifecycle: data → feature engineering → model → deploy → monitorização
- Tipos de modelos: supervised, unsupervised, reinforcement, generative
- ML para produção vs. ML académico
- Ferramentas fundamentais: Python, scikit-learn, pandas, NumPy, MLFlow, Evidently, Weights & Biases.
- Atividade: exploração de datasets e definição de problemas de ML

Data Ingestion e Data Engineering para ML

Focado na construção de pipelines robustos, este módulo aborda padrões de ingestão batch e streaming, modelação orientada a features e processamento distribuído. Neste módulo desenvolves pipelines escaláveis que suportam treino, re-treino e monitorização contínua em contexto de produção.

- Diferença entre Data Science datasets e production data pipelines
- Data Ingestion Patterns: batch, streaming, near-real-time
- Fontes de dados: databases, data lakes, APIs, event streams
- Data Modelling para ML (feature-oriented vs table-oriented)
- Processamento distribuído com PySpark (exemplos práticos)
- Introdução a conceitos de cloud data processing e escalabilidade
- Data validation e schema enforcement para ML pipelines
- Onde entram (e onde não entram) técnicas clássicas de Data Cleaning
- Atividade: construção de um pipeline simples de ingestão e processamento de dados com PySpark ou pandas escalável

Desenvolvimento de Modelos Orientado a Produção

Analisa a seleção de modelos sob a perspetiva operacional, avaliando trade-offs entre latência, custo, interpretabilidade e facilidade de deployment. Aprende a integrar técnicas de model interpretability com SHAP para validação pré e pós-produção.

- Visão geral dos principais tipos de modelos (sem zoo excessivo): Linear / tree-based / ensemble / neural networks
- Trade-offs de modelos em produção:
    - Latência
    - Custo computacional
    - Facilidade de deployment
    - Monitorização e retraining
- Diferença entre modelos explainable e black-box em produção
- Introdução prática a model interpretability com SHAP
- Uso de SHAP antes do deployment (validação e confiança)
- Uso de SHAP após o deployment (monitorização e auditoria)
- Preparação do modelo para produção (artefactos, metadata, inputs/outputs)
- Atividade: Treino de um modelo com análise de impacto em produção e geração de explicações com SHAP.

Otimização, Tuning e Validação

Aprende a aplicar métricas e técnicas de tuning como GridSearch e Bayesian Optimization com foco em decisões de produção. Identifica riscos como data leakage e valida modelos com rigor antes do deployment.

- Métricas de classificação e regressão: accuracy, f1, ROC-AUC, RMSE
- Métodos de tuning: GridSearchCV, RandomSearch, Bayesian Optimization
- Regularização: L1, L2, dropout (em redes neuronais)
- Deteção de data leakage e problemas de validação como riscos operacionais em ambiente produtivo
- Comparação de modelos com estatística básica
- Preparação do modelo final para deployment

- Atividade: Tuning completo de um modelo com análise comparativa

MLOps, Versionamento e Automação

Explora práticas modernas de MLOps para transformar experimentos em pipelines reprodutíveis e colaborativos. Neste módulo explora o versionamento com Git e DVC, tracking com MLflow e conceitos de CI/CD aplicados a Machine Learning.

- O que é MLOps?
- Versionamento com Git + DVC
- MLflow para tracking, modelos e experimentos
- CI/CD aplicado a ML
- Orquestração de pipelines: Kubeflow, Airflow, Prefect (visão geral)
- Boas práticas de reprodutibilidade
- Atividade: Criação de um pipeline versionado com MLflow ou DVC.

Deployment de Modelos

Aprende a colocar modelos de Machine Learning em produção via API, containers ou cloud. Aborda exportação de modelos, criação de serviços REST com FastAPI e gestão de escalabilidade e performance.

- O que é deployment e quando deve ser feito
- Exportação de modelos: pickle, onnx, MLflow models
- Criação de APIs com FastAPI
- Containers com Docker para ML
- Deployment em cloud (Azure ML, AWS Sagemaker, Vertex AI – overview)
- Latência, escalabilidade e gestão de recursos
- Atividade: deployment de um modelo como REST API com FastAPI

Monitorização, Drift Detection e Manutenção

Capacita-te para monitorizar modelos em produção, identificar data drift e concept drift e implementar retraining pipelines. Neste módulo aprende a definir alertas e métricas operacionais que garantem fiabilidade contínua.

- Monitorização de modelos em produção: métricas, dashboards e logs para deteção de degradação e anomalias ao longo do tempo
- Data Drift, Concept Drift e Performance Drift
- Técnicas de deteção de drift
- Retraining pipelines e triggers
- Alertas, métricas e limites operacionais
- Monitorização com MLflow, Prometheus ou ferramentas equivalentes
- Atividade: Simulação de drift e criação de alarme de monitorização

Workshop Final: ML Pipeline End-to-End

Consolida as competências adquiridas ao longo das sessões anteriores, construindo um pipeline completo de Machine Learning, desde ingestão de dados até deployment e monitorização inicial. No final irás desenvolver uma solução pronta para produção com versionamento, tuning e API funcional.

- Escolha do dataset ou caso fornecido
- Construção de pipeline de dados

- Treino + tuning + validação
- Versionamento e tracking
- Deployment como API
- Monitorização inicial
- Apresentação final
- Output final: Pipeline completo e apresentação com justificações técnicas