



Aula 1

Apresentação de IAA-P / Introdução aos Dados

Filipe Silva (DETI/fmsilva@ua.pt)

Resumo

- **Objetivos da unidade curricular**
- **Inteligência Artificial (IA) vs Aprendizagem Automática (AA/ML)**
- **Funcionamento e ferramentas software**
- **Avaliação da componente prática**
- **Ambiente Python:**
 - ▶ Anaconda (ambientes e bibliotecas)
 - ▶ Jupyter Notebooks (desenvolvimento interativo)

Objetivos

■ Objetivos globais da UC:

- ▶ Compreender os fundamentos da aprendizagem supervisionada, não-supervisionada e profunda (*deep learning*, DL)
- ▶ Implementar e avaliar modelos de ML para diferentes tipos de problemas (e.g., classificação, regressão, clustering, redução de dimensionalidade)
- ▶ Avaliar & analisar o desempenho (métricas adequadas e técnicas de validação) com desafios reais (*overfitting*, dados desbalanceados, qualidade dos dados)
- ▶ Planear e executar soluções estruturadas, seguindo o pipeline de IA orientado para dados: **dados** → **modelo** → **avaliação** → **impacto/valor**

■ Pipeline da Aprendizagem Automática:

- ▶ **Dados:** aquisição, limpeza, análise e preparação dos dados
- ▶ **Modelo:** seleção do algoritmo, treino, validação e optimização
- ▶ **Valor:** avaliação, interpretação e impacto da solução no problema real

Objetivos

■ Objetivos globais da UC:

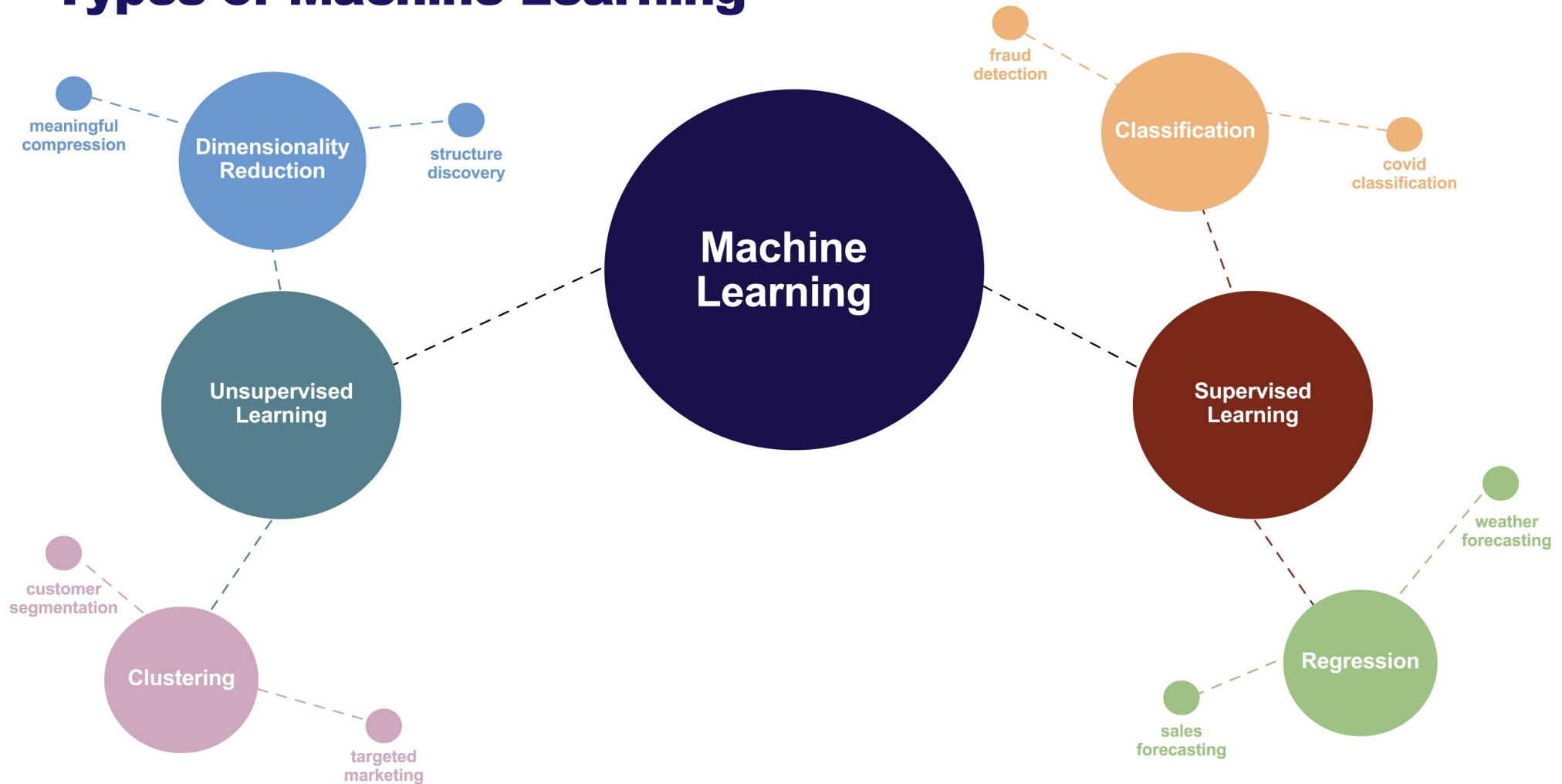
- ▶ Compreender os fundamentos da aprendizagem supervisionada, não-supervisionada e profunda (*deep learning*, DL)
- ▶ Implementar e avaliar modelos de ML para diferentes tipos de problemas (e.g., classificação, regressão, clustering, redução de dimensionalidade)
- ▶ Avaliar & analisar o desempenho (métricas adequadas e técnicas de validação) com desafios reais (*overfitting*, dados desbalanceados, qualidade dos dados)
- ▶ Planear e executar soluções estruturadas, seguindo o pipeline de IA orientado para dados: **dados → modelo → avaliação → impacto/valor**

■ Pipeline da Aprendizagem Automática:

- ▶ **Dados:** aquisição, limpeza, análise e preparação dos dados
- ▶ **Modelo:** seleção do algoritmo, treino, validação e optimização
- ▶ **Valor:** avaliação, interpretação e impacto da solução no problema real

Paradigmas de Aprendizagem em IAA

Types of Machine Learning



Source: imagem gerada por IA (Google Gemini)

IA Clássica vs Aprendizagem Automática

■ IA Clássica

- ▶ Conhecimento explicitamente programado
- ▶ Sistema segue regras, modelos ou estados definidos pelo humano
- ▶ **Exemplos:**
 - Pesquisa em espaço de estados
 - Planeamento (STRIPS)
 - Redes de Bayes e MDPs

O comportamento do sistema é “desenhado” antes da execução

■ Aprendizagem Automática

- ▶ Comportamento aprendido a partir de dados
- ▶ Modelo é estimado, não programado explicitamente
- ▶ **Exemplos:**
 - Classificação, regressão, agrupamento (*clustering*)
 - Redes neuronais profundas

Sistema melhora com a experiência (dados)

Enquanto a IA clássica define como resolver o problema, o ML aprende essa solução a partir dos dados

Avaliação da Componente Prática

$$\text{Nota Final} = 0.3 \cdot T1 + 0.3 \cdot T2 + 0.4 \cdot E$$

■ T1 – Trabalho Individual

- ▶ **Objetivo:** trabalho escrito (EN), de natureza teórico-técnica, que visa explorar um tema ao nível dos **dados** (e.g., tratamento, processamento, etc).
- ▶ **Secções obrigatórias:** (I) *Critical Analysis*; (ii) *Acknowledgement of used AI tools*
- ▶ **Entregável:** documento no formato artigo IEEE (máximo 4 páginas)
- ▶ **Critérios de avaliação:** qualidade do documento, rigor técnico/científico e a capacidade de análise crítica do tema.

■ **Datas Importantes:**

- ▶ **Escolha do tema:** 14 a 20/02/2026 → seleção do tema no elearning
- ▶ **Submissão:** até 13/03/2026 → entrega final do documento

Avaliação da Componente Prática

$$\text{Nota Final} = 0.3 \cdot T1 + 0.3 \cdot T2 + 0.4 \cdot E$$

■ T2 – Trabalho de Grupo (3 elementos)

- ▶ **Objetivo:** resolver um problema de **aprendizagem automática**, proposto pelo grupo, para promover o uso de ferramentas, o desenvolvimento de pipelines, a avaliação de desempenho, e a discussão técnica dos resultados.
- ▶ **Entregáveis:** (1) código, (2) documento no formato artigo IEEE (max 6 paginas), (3) ficheiro da apresentação oral, (4) avaliação pelos pares dentro do grupo, e (5) avaliação da apresentação de cada grupo pelos colegas de turma
- ▶ **Critérios de avaliação:** capacidade de tomar decisões informadas, justificar as decisões tomadas, e refletir sobre limitações.

■ **Datas Importantes:**

Avaliação qualitativa

- ▶ **Submissão 1:** 07/03/2026 – definição do grupo, tópico, objetivos
- ▶ **Submissão 2:** 13/03/2026 – resultados preliminares (dados, modelos)
- ▶ **Submissão 3:** 29/05/2026 – documento final, código e apresentação
- ▶ **Apresentação:** última semana de aulas

Ambiente Python

■ Anaconda 3 (Python)

- ▶ Distribuição completa que inclui **Python**, **conda**, **bibliotecas** pré-instaladas e ferramentas como o **Jupyter Notebook** e o **Anaconda Navigator**.
- ▶ Orientada para a ciência de dados e aprendizagem automática, que inclui as bibliotecas **NumPy**, **Pandas**, **scikit-learn**, **seaborn**, **matplotlib**.
- ▶ Facilita a gestão de ambientes e dependências, evitando conflitos de versões no desenvolvimento de projetos de AA.

■ Jupyter Notebooks (incluído no Anaconda 3)

- ▶ Ambiente interativo para programação em Python, combinando código, resultados, visualizações e texto explicativo num único documento
- ▶ Ideal para desenvolvimento de modelos e comunicação de resultados; pode ser iniciado a partir do Anaconda Navigator ou da linha de comandos

Estas ferramentas suportam todo o processo de Aprendizagem Automática, da análise de dados à avaliação dos modelos!

{ Anaconda / conda → gerir ambiente Python (Python + libs)
Jupyter Notebook → fornecer interface de desenvolvimento
Kernel → ligar o notebook a um Python específico

Aula 1 – Introdução ao ML

Compreender e Preparar Dados

- Descarregue para a pasta de trabalho o ficheiro de dados “**adult.csv**” e o notebook **IML_Session1_Student.ipynb**
- **Objetivos de Aprendizagem**
 - ▶ Carregar e inspecionar um conjunto de dados (dataset)
 - ▶ Identificar diferentes tipos de dados (números, categorias, texto, etc)
 - ▶ Realizar uma análise descritiva (estatística, distribuição, tendências)
 - ▶ Detetar valores em falta (*missing values*) e inconsistências
 - ▶ Tomar decisões justificadas de seleção de variáveis (*features*)
 - ▶ Guardar um conjunto de dados “limpos” para a próxima aula

Garbage in, garbage out

Valores em falta: apagar ou preencher com média?

Inconsistências: forma → “Portugal”, “PT” | conteúdo → idade negativa!

Seleção de variáveis: o que é irrelevante vs essencial?