|  |
| --- |
| Análise de Fraude em Cartões de Crédito Engenharia Informática | 2024/2025 | 3ºano  Laboratório de Análise de Dados | Prof. Fátima Leal  Luís Meireles | 47799  João Faria | 48057  Mónica Fino | 47366  Paulo Brochado |  Tiago Mesquita | 47558 |

### ÍNDICE

[**Introdução 4**](#_heading=h.3rdcrjn)

[**Título 5**](#_heading=h.ktxp7cbj5jqh)

[**Conclusão 6**](#_heading=h.927it0515gle)

# Introdução

O presente dataset contém registos de transações realizadas com cartões de crédito por clientes de uma instituição financeira. Ele foi disponibilizado para estudos de **deteção de fraudes** e é amplamente utilizado na área de **aprendizado de máquina** e **ciência de dados**.

O objetivo principal deste conjunto de dados é fornecer informações suficientes para a criação de modelos preditivos que possam diferenciar transações legítimas de fraudulentas com alta precisão.

# Características do Dataset

O dataset pertence ao **domínio financeiro**, mais especificamente ao setor de **pagamentos eletrônicos e segurança bancária**. Ele é utilizado para desenvolver e aprimorar sistemas de monitoramento de transações, que têm como função detetar fraudes em tempo real.

As instituições financeiras e empresas de tecnologia utilizam técnicas de **Machine Learning** para mitigar os riscos de fraude e evitar perdas financeiras tanto para os bancos quanto para os clientes.

O conjunto de dados possui:

* **Número total de registos (linhas):** 284.807 transações
* **Número total de variáveis (colunas):** 31 atributos

O dataset contém **apenas variáveis numéricas**, que podem ser classificadas conforme a sua natureza estatística:

| **Variável** | **Tipo** | **Classificação Estatística** | **Descrição** |
| --- | --- | --- | --- |
| Time | Quantitativa | Contínua | Tempo decorrido desde a primeira transação (em segundos) |
| Amount | Quantitativa | Contínua | Valor da transação em dólares |
| V1 a V28 | Quantitativa | Contínua | Variáveis transformadas por PCA (não possuem interpretação direta) |
| Class | Qualitativa | Nominal | Indica se a transação é fraudulenta (1) ou legítima (0) |

As variáveis do dataset podem ser classificadas em **quantitativas** e **qualitativas**. As quantitativas representam valores numéricos e podem ser **contínuas** (como Time, Amount e V1 a V28, que podem assumir qualquer valor dentro de um intervalo). Não há variáveis **discretas** no dataset. Já as variáveis qualitativas representam categorias, podendo ser **nominais** (como Class, que identifica se a transação é fraudulenta ou não, sem hierarquia). O dataset não contém variáveis **ordinais**, ou seja, não há categorias com uma ordem natural definida.



upt.pt