|  |
| --- |
| ANÁLISE DE FRAUDE EM CARTÕES DE CRÉDITO Licenciatura em Engenharia Informática | 2024/2025 | 3º ano  Laboratório de Análise de Dados | Prof. Fátima Leal  Gabriela Almeida | 47004 João Faria | 48057 Luís Meireles | 47799 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 projeto, desenvolvido no âmbito da unidade curricular Laboratório de Analise de Dados, tem como principal objetivo a aplicação prática de técnicas de análise de dados e algoritmos de aprendizagem automática, utilizando a linguagem de programação Python.

A realização do projeto divide-se em duas fases complementares: Na primeira fase, é realizada uma análise estatística exploratória de um conjunto de dados escolhido, com o intuito de compreender a sua estrutura, características e padrões relevantes. Esta análise inclui o uso de medidas estatísticas, visualizações gráficas, criação de novas variáveis e normalização dos dados, culminando num painel de controlo gráfico e numa reflexão crítica dos resultados obtidos.

Na segunda fase, são aplicados e comparados diversos modelos de Machine Learning, deforma a resolver um problema de revisão ou de classificação. Esta etapa inclui também a análise de desempenho dos modelos, técnicas de redução de dimensionalidade, validação cruzada, e eventualmente algoritmos de recomendação e ferramentas automáticas, com o objetivo de validar os métodos escolhidos.

Este relatório descreve todo o processo de desenvolvimento do trabalho, desde a seleção do conjunto de dados até à análise final dos modelos aplicados, incluindo as decisões técnicas tomadas, os desafios enfrentados e as aprendizagens adquiridas ao longo do projeto.

# Contextualização do Dataset

O dataset utilizado neste projeto foi obtido a partir da plataforma Kaggle e é amplamente conhecido por ser utilizado em estudos sobre deteção de fraude em transações financeiras. Este conjunto de dados representa transações realizadas com cartões de crédito por titulares europeus ao longo de dois dias, em setembro de 2013. O dataset contém 284.807 registos, distribuídos por 31 variáveis. Todas as variáveis presentes no dataset classificam-se como Quantitativa Contínua, exceto a variável *class*, que é uma variável Qualitativa nominal.

O principal objetivo deste dataset é servir como base para o desenvolvimento e avaliação de modelos de Machine Learning capazes de identificar transações fraudulentas de forma automática, um problema real e crítico enfrentado por instituições financeiras em todo o mundo.

# 

# Conclusão

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut sit amet felis et orci porttitor fermentum. Aenean eu vehicula elit. Pellentesque lectus tellus, congue at aliquet sit amet, placerat in est. Pellentesque non porta quam. Duis fermentum elit sit amet neque fermentum eget porta odio malesuada. Sed eget nulla vitae velit luctus laoreet in ut elit. Aliquam erat volutpat. Donec nunc augue, tristique ac tincidunt id, dignissim a nunc. Duis vitae ante ut urna malesuada tempor id in tortor. Nullam aliquam nulla et nunc viverra vel facilisis metus euismod. Donec nec dapibus orci.



upt.pt