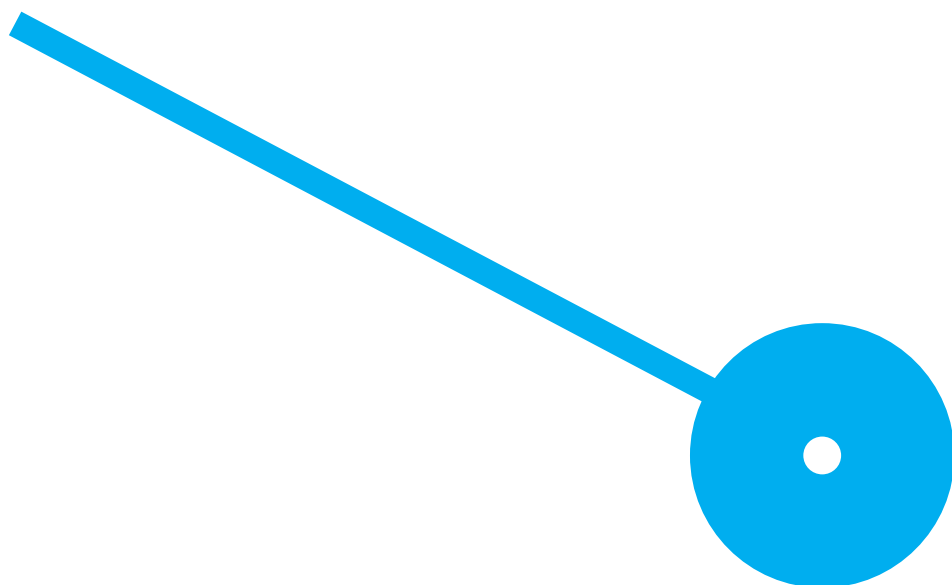
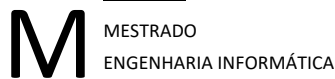


Tecnologias Escaláveis para Análise de Dados

João Oliveira (8210291), Luís Soares (8210396)

03/2025





Tecnologias Escaláveis para Análise de Dados

João Oliveira (8210291), Luís Soares (8210396)

Professor

Prof. Doutor David Rua Carneiro

Relatório de Projeto apresentado para cumprimento dos requisitos necessários à obtenção de aprovação na Unidade Curricular de Tecnologias Escaláveis para Análise de Dados na Escola Superior de Tecnologia e Gestão do Instituto Politécnico do Porto.

Declaração de integridade

Nós, **João Pedro Silva Oliveira**, estudante nº **8210291**, e **Luís Soares**, estudante nº **8210396** do Mestrado **Engenharia Informática** da Escola Superior de Tecnologia e Gestão do Instituto Politécnico do Porto, declaro que não fizemos plágio nem auto-plágio, pelo que o trabalho intitulado “**TODO: Change This**” é original e da nossa autoria, não tendo sido usado previamente para qualquer outro fim. Mais declaramos que todas as fontes usadas estão citadas, no texto e na bibliografia final, segundo as regras de referência adotadas na instituição.

Agradecimentos

Gostaríamos de expressar a nossa gratidão ao Professor Doutor Davide Rua Carneiro pela orientação, apoio e paciência durante o desenvolvimento deste trabalho. Toda a experiência e conhecimento foram essenciais para a realização deste projeto.

Agradecemos também aos nossos colegas de curso pela troca de conhecimentos, experiências e pelo espírito de camaradagem que sempre nos motivou a continuar, mesmo nos momentos mais desafiadores.

Aos nossos familiares, pelo suporte incondicional, compreensão e encorajamento ao longo desta jornada académica. Sem o vosso apoio, este trabalho não teria sido possível.

Por fim, agradecemos a todos que, direta ou indiretamente, contribuíram para a realização deste trabalho. A todos, o nosso sincero obrigado.

Abstract

Keywords:

Resumo

Palavras-chave:

Conteúdo

Lista de Figuras	vi
Lista de Tabelas	vii
Lista de Listagens	vii
Lista de Acrónimos	viii
1 Introdução	1
1.1 Contextualização	1
1.2 Apresentação do Caso de Estudo	1
1.3 Motivação e Objetivos	1
1.4 Estrutura do Relatório	1
2 Caracterização Geral do Problema	2
2.1 <i>Problem Description</i>	2
2.2 <i>Business Glossary</i>	2
2.3 <i>Data sources description</i>	2
3 ...	3
3.1 <i>Data description report</i>	3
3.2 <i>Data processing report</i>	3
4 Machine Learning	4
5 Conclusões e Trabalho Futuro	5
6 Bibliografia	6

Lista de Figuras

Lista de Tabelas

Lista de Listagens

Capítulo 1

Introdução

1.1 Contextualização

1.2 Apresentação do Caso de Estudo

1.3 Motivação e Objetivos

Motivação:

Objetivos:

1.4 Estrutura do Relatório

O presente relatório está organizado da seguinte forma:

Capítulo 2

Caracterização Geral do Problema

Esta secção deve incluir uma caracterização geral do problema bem como dos fatores mais relevantes indicados na literatura. Deve apresentar um glossário com os principais termos do domínio, bem como uma descrição das fontes de dados identificadas, que deve incluir, no mínimo, uma descrição da forma de aceder aos dados, e um dicionário de dados (por fonte)

2.1 *Problem Description*

2.2 *Business Glossary*

2.3 *Data sources description*

Capítulo 3

...

3.1 *Data description report*

Esta secção, descreve o trabalho que ocorre após a integração dos dados, e deve descrever o(s) dataset(s) resultante(s) e que serão usados como base para as tarefas de processamento e análise de dados. Entre outros aspetos, esta secção deve pelo menos analisar a quantidade e qualidade (e.g. distribuição das variáveis, dados em falta, problemas nos dados e tarefas de limpeza necessárias) dos dados disponíveis. Isto será usado como base para o trabalho a desenvolver na fase seguinte.

3.2 *Data processing report*

Esta secção deve descrever todas as tarefas de limpeza, transformação, criação (feature engineering), ou outras levadas a cabo de forma a melhorar a qualidade dos dados, e a expor da melhor forma o conhecimento existente para a fase de Machine Learning que se segue.

Capítulo 4

Machine Learning

Esta secção deve descrever as tarefas de modelação levadas a cabo, bem como uma análise crítica dos seus resultados. Ou seja, deve ser possível perceber pela leitura desta secção, que algoritmos foram utilizados, que modelos foram produzidos e qual a sua qualidade, como os modelos foram produzidos e avaliados, ou ainda qual ou quais os modelos seleccionados para colocar em produção no dashboard.

Capítulo 5

Conclusões e Trabalho Futuro

;;Elaborar uma apreciação crítica sobre o trabalho realizado, apontando os seus pontos fortes e fracos. Adicionalmente, caso se aplique, enunciar eventuais tarefas a realizar futuramente ou novas opções para estender o trabalho realizado.;;

Capítulo 6

Bibliografía