Título

Relatório de Estágio

Nome aluno

(Número)

Orientadores:

Nome | ISEC

Nome | Empresa

Licenciatura em Engenharia Informática

Ramo de …

Instituto Politécnico de Coimbra

Instituto Superior de Engenharia de Coimbra

Julho de 2021

# Agradecimentos

…

# Resumo

… este documento pretende apenas servir de linha orientadora, de uso opcional, para a elaboração dos relatórios de estágio… a estrutura e formatação finais dos relatórios deverão ser definidos caso a caso pelos alunos e respetivos orientadores…

***Palavras‑chave:*** *Big Data, Serialização, Spring Boot.*

# Abstract

…

***Keywords:*** *Big Data, Serialization, Spring Boot.*

**Índice**

[Agradecimentos i](#_Toc68777782)

[Resumo iii](#_Toc68777783)

[Abstract v](#_Toc68777784)

[Índice de figuras ix](#_Toc68777785)

[Índice de tabelas xi](#_Toc68777786)

[Acrónimos xiii](#_Toc68777787)

[1 Introdução 1](#_Toc68777788)

[1.1 Entidade de acolhimento 1](#_Toc68777789)

[1.2 Instituto Superior de Engenharia de Coimbra 1](#_Toc68777790)

[1.3 Objetivos e plano de trabalhos 1](#_Toc68777791)

[1.4 Estrutura do relatório 2](#_Toc68777792)

[2 Mais um capítulo 5](#_Toc68777793)

[3 Mais um capítulo 7](#_Toc68777794)

[4 Mais um capítulo 9](#_Toc68777795)

[4.1 Requisitos 9](#_Toc68777796)

[4.2 Arquitetura 9](#_Toc68777797)

[4.3 Tecnologias e ferramentas usadas 9](#_Toc68777798)

[4.4 Implementação 9](#_Toc68777799)

[4.5 Testes e validação 9](#_Toc68777800)

[4.5.1 Método A 9](#_Toc68777801)

[4.5.2 Método B 10](#_Toc68777802)

[5 Mais um capítulo 11](#_Toc68777803)

[6 Conclusões e trabalho futuro 13](#_Toc68777804)

[6.1 Conclusões 13](#_Toc68777805)

[6.2 Trabalho Futuro 13](#_Toc68777806)

[Referências 15](#_Toc68777807)

[Anexos 17](#_Toc68777808)

[Anexo A: Proposta de Estágio A-1](#_Toc68777809)

[Anexo B: Especificação de Requisitos B-1](#_Toc68777810)

# Índice de figuras

[Figura 1.1 Explicação da figura 1](#_Toc68777811)

# Índice de tabelas

[Tabela 4.1 – Tecnologias e ferramentas usadas 9](#_Toc68777812)

# Acrónimos

…

ITS – Intelligent Transportation Systems

…

# Introdução

A tecnologia dos telemóveis revolucionou a forma como as pessoas e as empresas obtêm e partilham informação sobre mobilidade, fornecendo conjuntos massivos de trajetórias registadas, dados longitudinais e o seu potencial e precisão em tempo real. Este projeto é beneficiado pelo CGIUL e pretende analisar a gestão da mobilidade, através de dados de telemóveis, nomeadamente na questão do planeamento do quotidiano na cidade de Lisboa. Efetivamente, é fornecido conjuntos de dados de telemóveis referentes ao volume de pessoas que entram e saiam dos 11 principais pontos de entrada e saída da cidade nas horas de ponta da manhã (7:30h-10:00h) e da tarde (17:00h-19:30h), que geram congestionamentos nas principais vias de acesso. O objetivo do projeto é caraterizar esses fluxos diários durante os dois períodos referidos e a sua relação com fatores como calendário escolar, períodos de férias e ocorrência de temporal.

Contextualização… depois de ler os parágrafos que compõem este início de capítulo, deve ficar‑se a entender que existe uma determinada área/contexto, uma determinada empresa que pretende fazer ou já faz algo nessa área e que apresentou, no âmbito da unidade curricular de Projeto ou Estágio da licenciatura em Engenharia Informática (LEI) [1] do Instituto Superior de Engenharia de Coimbra (ISEC) [2], uma proposta de estágio com um determinado objetivo… que este documento pretende descrever o trabalho realizado…

…a referir no texto as figuras, quando existirem, conforme apresentado na Figura 1.1…

[](https://www.google.pt/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=0ahUKEwiW-4ij3_zSAhXB0RoKHYa2AJkQjRwIBw&url=http://mariohenriquemartins.com.br/enquadramento-o-que-voce-destaca-em-seu-argumento/&psig=AFQjCNEBh3yreV0BxghV2mg3qhOIzvt7pA&ust=1490912295373271)

Figura . Explicação da figura

… as subsecções seguintes são apenas exemplificativas…

## Entidade de acolhimento

… a usar referências quando se mencionar algo dito por outra pessoa [2]…

## Instituto Superior de Engenharia de Coimbra

…

## Objetivos e plano de trabalhos

Através da análise dos fluxos diários entre a manhã (7:30h-10:00h) e tarde (17:00h-19:30h) nos principais pontos de entrada e saída da cidade de Lisboa pretende-se atingir os seguintes objetivos:

- Caraterizar o volume total de entradas e saídas da cidade durante o período da hora de ponta;

- Caraterizar o volume de entradas e saídas da cidade durante o período da hora de ponta para cada um dos 11 pontos de entrada e saída;

- Comparar com outros períodos do dia, - Relacionar o ponto anterior com variáveis como os períodos de aulas ou férias e a existência de pluviosidade;

- Análise das zonas de destino daqueles que saem da cidade;

- Análise das zonas de origem daqueles que entram na cidade;

… referir a proposta original, que se encontra em anexo, e apresentar o que realmente aconteceu…

## Estrutura do relatório

…

…

# Estado da Arte

Neste capítulo é realizado um estudo do estado de arte, descrevendo área da previsão do fluxo de tráfego e as suas diferentes técnicas de recolha de dados. Efetivamente, também é analisado como estas técnicas podem ser relacionadas como o uso de smartphones como sensores para captar dados relativos á mobilidade.

## Metodologia

O exercício metodológico neste projeto é abarcado segundo a base de dados do Google Scholar, onde o professor orientador sugeriu a pesquisa da palavra-chave "traffic flow" e esta apresentou o maior número de artigos com 3 740 000 resultados. Depois comecei a experimentar outras palavras-chave, que são mais específicas em termos de utilização, “mobile data” e “machine learning”, e estas apresentaram menos resultados, mas ainda devolvem um número substancial de artigos, variando entre 1 780 000 para "traffic flow mobile data" e 4 230 para "machine learning traffic flow mobile data". Globalmente, estes números sugerem que existe uma quantidade significativa de investigação em curso na área do fluxo de tráfego e dos dados móveis.

Os artigos analisados foram conduzidos entre 2019 e 2023, e focam principalmente na gestão eficaz do tráfego em situações críticas, com o objetivo de ajudar a reduzir o congestionamento de tráfego, os custos de combustível e a poluição atmosférica, através do estudo do fluxo de veículos.

Tabela - Número de artigos no google scholar com base numa palavra-chave

|  |  |
| --- | --- |
| **Google Scholar** | |
| **Palavras-Chave** | **Número de artigos** |
| “traffic flow” | 3 740 000 |
| “traffic flow mobile data” | 1 780 000 |
| “traffic flow based on phone data” | 8 690 |
| “machine learning traffic flow mobile data” | 4 230 |
| “mobility data traffic flow” | 6 320 |
| “traffic prediction based on phone location” | 6 080 |

## Fluxo de Tráfego

Atualmente, o estudo do fluxo de tráfego é um campo de investigação com muitos desenvolvimentos em curso, desta forma os investigadores têm desenvolvido uma variedade de modelos e teorias para descrever e analisar o fluxo de tráfego. Estes modelos podem apresentar desde simples equações matemáticas a simulações mais complexas, como por exemplo o estudo de matrizes origem-destino que servem para determinar as rotas de desvio ideais nas redes urbanas [4], algoritmos de “Machine Learning” para classificar o congestionamento do tráfego [5] ou utilização de frameworks como a “Promotion” para prever o fluxo do tráfego [6], e assim conceber e gerir sistemas de transporte que sejam eficientes, seguros e sustentáveis.

Para além disto, a crescente utilização de telemóveis e o número de dispositivos conectados também contribuiu significativamente para o crescimento de tráfego móvel, disponibilizando uma riqueza de informações que podem ser utilizadas para prever padrões de tráfego, tais como a localização e a velocidade dos dispositivos na rede rodoviária [7][8].

### 2.2.1 Matriz Origem-Destino como sensor de tráfego

A matriz origem-destino (OD) na gestão do tráfego fornece uma imagem da distribuição espacial, com células individuais representadas através do número de viagens efetuadas entre um par de zonas de análise de tráfego (TAZs) [4]. Estes modelos estáticos e dinâmicos podem ser utilizados no planeamento de um sistema de transporte, de modo a otimizar o controlo e gestão do tráfego, dando particular atenção à necessidade e possibilidade de atualização frequente dos dados de entrada na sequência de alterações no volume do fluxo de tráfego real na rede rodoviária urbana.

### 2.2.2 Classificação do fluxo de tráfego através de algoritmos de machine learning

Efetivamente, um estudo em 2017 constatou que o custo dos engarrafamentos de trânsito para os condutores americanos foi estimado em 179 milhões de dólares [4], como tal, para ajudar a melhorar as condições de tráfego os investigadores decidiram avaliar o comportamento e eficácia de vários algoritmos de “machine learning” para classificar esse fluxo e concluíram que os Sistemas de Transporte Inteligentes (ITS) podiam ser uma alternativa associada a infraestrutura rodoviária com dispositivos IoT para captar informação.

Na verdade, estes resultados da classificação são influenciados por factores através de afinação de parâmetros, o tamanho do conjunto de dados em termos de registos, atributos selecionados, e a percentagem de dados utilizados para formação e validação. Além disso, a análise de diferentes ambientes urbanos, incluindo intersecções, autoestradas e estradas suburbanas também é fundamental para identificar os classificadores mais eficazes.

### Promotion FrameWork para prever o fluxo tráfego

A capacidade de prever o tráfego é uma das aplicações mais significativas em tecnologias de transporte inteligentes. A precisão das previsões é crucial para garantir a eficiência dos transportes públicos e para evitar congestionamentos. Tendo em conta, algoritmos para prever o tráfico, a abordagem da framework PROMOTION é uma abordagem baseada em deep learning para a previsão do fluxo de tráfego que utiliza redes neurais gráficas (GNNs) para modelar relações de espaço-tempo entre diferentes segmentos rodoviários numa rede de transportes [6].

De facto, esta framework explora técnicas inteligentes de pré-processamento baseado em gráficos, onde cada nó representa um segmento de estrada e as arestas representam a conectividade entre eles. O modelo GNN é treinado para aprender as características de cada nó e aresta, o que capta as relações espaço-temporais entre os diferentes segmentos rodoviários, melhorando a precisão da previsão do fluxo de tráfego para otimizar sistemas inteligentes de transporte (ITS) e planeamento urbano.

### Crescimento IoT e a sua utilidade para melhorar o fluxo do tráfego

Foram estimados que cerca de 77,5 Exabytes de dados móveis serão consumidos por mês até 2022, o que representa 71% de todo o tráfego IP [7]. Consequentemente, é essencial prever com precisão a carga de tráfego móvel para gerir eficientemente o grande volume de tráfego móvel. Por exemplo, foi desenvolvido um modelo baseado em "Machine Learning" que chamado Deep Traffic Predictor (DeepTP), que utiliza RNNs, LSTM, CNNs e modelos LSTM para prever a carga de tráfego móvel, como também foi descoberto que o fluxo pode ser previsto com alta precisão, otimizando a alocação de recursos e a qualidade da circulação.

### Smartphones como sensores de fluxo de tráfego

O elevado uso de smartphones levou ao desenvolvimento de novas formas de recolha de dados sobre as condições de tráfego. Um desses métodos é a utilização de smartphones como sensores de tráfego, o que oferece várias vantagens, tais como ser rentável, proporcionando uma alternativa mais barata aos sensores de tráfego tradicionais e oferecendo uma cobertura mais ampla de dados em tempo real, como informações precisas e atualizadas sobre as condições de tráfego. Desta forma, para identificar com precisão a estrada que um condutor está a circular o uso de algoritmos de map-matching [8] revelou que a trajetória dos veículos em movimento tem caraterística de memória curta e assim é provável identificar o próximo segmento de estrada que o condutor vai entrar [7] para prever com precisão o fluxo do tráfego.

Tabela - Análise comparativa de artigos baseado no fluxo do tráfego

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Artigo** | **Ano** | **Objetivo** | **Técnicas Aplicadas** | **Resultados** |
| [4] | 2023 | Gestão eficaz do tráfego em situações críticas. | Matriz Origem-Destino  LSTM | Método proposto poderia ter uma implementação prática em sistemas de atribuição dinâmica de tráfego em tempo real para aplicações ITS. |
| [5] | 2023 | Análise do fluxo de veículos para ajudar a reduzir o congestionamento de tráfego, os custos de combustível e a poluição atmosférica. | Decision Tree  Extra-tree  k-Near Neighbors  Random Forest  MLP | Resultados de classificação mais elevados foram obtidos no conjunto de dados da Primavera, o que coincide com as restrições de mobilidade estabelecidas nas cidades devido à pandemia de Covid-19 e a Decision Tree obteve a maior precisão na experiência. |
| [6] | 2022 | Utilização de gráficos convulsionais para a previsão do fluxo de tráfego urbano num ambiente com Internet de alta velocidade. | STGCN  STGNN  STFGNN  GraphWaveNet  Promotion | A promotion quando comparada com as técnicas de previsão do fluxo de tráfego de base supera as outras soluções (STGCN, STGNN, STFGNN e GraphWaveNet). |
| [7] | 2022 | Previsão da quantidade de tráfego móvel | LSTM  GRU  Random Forest  Linear Regression  Support Vector Machine | Os esquemas baseados no RNN foram mais precisos que os esquemas convencionais baseados em “Machine Learning”. Verificou-se o GRU proporcionou o desempenho mais elevado, apesar de ter a menor complexidade computacional. |
| [8] | 2019 | Demonstrar o desempenho do algoritmo Map-Matching e relacionar a precisão da estimativa e a frequência usada com os dados das torres de localização. | Map-Matching  Virtual Inductive Loop | Os resultados desta experiência mostram que o map-matching obteve uma precisão média na ordem dos 60% a 85%. A precisão da rota estimada estende uma equação linear proporcional à frequência de amostragem do local. |

# . Dataset

## Contexto

O conjunto de dados fornecidos pela Vodafone contém mais de 1 milhão de registos de “*Número de telemóveis que entram e saem da cidade de Lisboa*” e “*Observações das estações meteorológicas do IPMA*” entre setembro de 2021 e dezembro de 2022. No que diz respeito á obtenção dos dados, existem 11 eixos principais usados para manter o registo do fluxo de entrada e saída de telemóveis.

## Conteúdo

O conteúdo é dividido por 4 conjuntos, nomeadamente pela grelha dos registos móveis representado pelo conjunto1, pela localização dos pontos de entrada dos eixos, conjunto 2, pelo conjunto 3 que deveria representar o mapeamento dos troços da via representado por 11 registos, sem qualquer metadados ou outro ficheiro que os represente e por fim os dados meteorológicos do IPMA, representado pelo conjunto 4.

## 3.2.1 Conjunto 1

Eixo: nome eixo da cidade que efetuou o registo

Datetime: data do dia e hora do registo (dd/mm/aaaa hh:mm)

extract\_year\_2: ano do registo

extract\_month\_3: mês do registo

extract\_day\_4: dia do registo

C12: Número de entradas no eixo

C13: Número de saídas no eixo

## 3.2.2 Conjunto 2

id\_eixo\_viario: id do eixo

Eixo: nome do eixo

longitude: coordenada longitudinal do eixo

latitude: coordenada latitudinal do eixo

## 3.2.2 Conjunto 4

fecha: Data da última leitura

estacion: N.º da estação - 01200535 - Lisboa Geofísico, 01200579- Lisboa Gago Coutinho, 01210762 - Lisboa Tapada da Ajuda

humidade: Humidade relativa média do ar

iddireccvento: Direção média do vento - 0 a 9

intensidadeventokm: Intensidade média do vento

pressão: Pressão atmosférica

radiação: Radiação solar

temperatura: Temperatura

precacumulada: Precipitação acumulada

position: Coordenadas e tipo de entidade geográfica (ponto, linha ou polígono)

## Qualidade do conteúdo

No caso de estudo deste dataset, o objetivo é analisar dados sobre o número de telemóveis que entram e saem de um determinado eixo da cidade de Lisboa e a Vodafone detém 60% do mercado e como os dados apresentam valores decimais, pode ser devido a um fator de correção dos outros 40% do mercado.

Para além disto, existe uma incongruência nos dados do IPMA, porque os meta dados no ficheiro veem com a data de 2018 e no ficheiro excel vão desde 2020 a 2022, o que pode levantar dúvidas em relação á consistência destes dados.

## Análise do conteúdo

Totalidade dos Registos Conjunto 1: 1 048 576

Totalidade dos Registos Conjunto 4: 998206

## 3.5.1. Conjunto 1

Em relação a 2021 os registos são bastante inconsistentes, porque os registos são diferentes para cada eixo e particularmente em setembro que os registos se iniciam no dia 14 e acabam a outubro no dia 1. Para além disto, os registos deste ano indicam uma média de registos de 66.6% de 5min e 33.4% de 15min.

Tabela - Análise registos conjunto1 2021

**\*–** Registos com 5min de diferença **\*\* –** Registos com 15min de diferença

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **VODAFONE\_EIXOS** | | | | | | | | | | | | | |
| **PGIL\_VODAFONE\_EIXOS\_2021** | | | | | | | | | | | | | |
| **Mês** | **Registos** | **A1** | **A5** | **A36 Túnel do Grilo** | **Calçada de Carriche** | **IC2 (Sacavém)** | **IC16** | **IC19** | **Marginal** | **N117 (Cabos Ávila)** | **Ponte 25 Abril** | **Ponte Vasco Gama** | **Datas Registadas** |
| Setembro  **\*(5min)**  **\*\*(15min)** | 53844  35977  17867 | 4930  3288  1642 | 4947  3299  1648 | 4941  3312  1629 | 4882  3287  1595 | 4983  3319  1664 | 4841  3209  1632 | 4877  3258  1619 | 4862  3272  1590 | 4854  3237  1617 | 4868  3248  1620 | 4869  3248  1611 | 14/09/2021 até 1/10/2021 00:00 ás 23:55 |
| Outubro  **\*(5min)**  **\*\*(15min)** | 95898  63733  32165 | 8690  5722  2968 | 8626  5711  2915 | 8680  5786  2894 | 8784  5783  3001 | 8710  5841  2869 | 8729  5843  2886 | 8712  5751  2961 | 8751  5834  2917 | 8817  5902  2915 | 8687  5748  2939 | 8712  5812  2900 | 1/10/2021 até 31/10/2021 00:00 ás 23:55 |
| Novembro  **\*(5min)**  **\*\*(15min)** | 79288  52811  26477 | 7218  4821  2397 | 7152  4773  2379 | 7191  4824  2367 | 7245  4813  2432 | 7189  4786  2403 | 7204  4778  2426 | 7221  4812  2409 | 7239  4803  2436 | 7237  4795  2442 | 7166  4782  2384 | 7226  4824  2402 | 1/11/2021 até 31/11/2021 00:00 ás 23:55 |
| Dezembro  **\*(5min)**  **\*\*(15min)** | 97240  64874  32366 | 8840  5880  2960 | 8830  5926  2904 | 8800  5847  2953 | 8876  5948  2928 | 8816  5861  2955 | 8824  5907  2917 | 8958  5960  2998 | 8884  5958  2926 | 8872  5921  2951 | 8741  5791  2950 | 8799  5875  2924 | 1/12/2021 até 31/12/2021 00:00 ás 23:55 |

No que diz respeito a 2022, falta consistência entre janeiro e março com o número de registos variável e as datas registadas em fevereiro entre dia 1 e 8 e março entre dia 18 e 31. De maneira que, entre abril e dezembro os registos são consistentes, visto que existe o mesmo número de registos para cada eixo e as datas registadas são congruentes. Além disto, os registos até a março apontam uma média de registos de 66.6% de 5min e 33.4% de 15min e 100% de registos de 15min para os restantes meses do ano.

Tabela - Análise registos conjunto1 2022

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **VODAFONE\_EIXOS** | | | | | | | | | | | | | |
| **PGIL\_VODAFONE\_EIXOS\_2022** | | | | | | | | | | | | | |
| **Mês** | **Registos** | **A1** | **A5** | **A36 Túnel do Grilo** | **Calçada de Carriche** | **IC2 (Sacavém)** | **IC16** | **IC19** | **Marginal** | **N117 (Cabos Ávila)** | **Ponte 25 Abril** | **Ponte Vasco Gama** | **Datas Registadas** |
| Janeiro  **\*(5min)**  **\*\*(15min)** | 98005  65505  32500 | 8849  5920  2929 | 8941  5950  2991 | 8902  6005  2897 | 8865  5897  2968 | 8816  5953  2964 | 8917  5949  2978 | 8927  6043  2899 | 8959  5950  3009 | 8895  5959  2936 | 8921  5936  2985 | 8887  5943  2944 | 1/01/2022 até 31/01/2022 00:00 ás 23:55 |
| Fevereiro  **\*(5min)**  **\*\*(15min)** | 19525  12937  6588 | 1777  1186  591 | 1773  1176  597 | 1762  1163  599 | 1789  1190  599 | 1761  1165  596 | 1790  1169  621 | 1761  1169  592 | 1784  1177  607 | 1750  1157  593 | 1787  1195  592 | 1791  1190  601 | 01/02/2022 até 08/02/2022 00:00 ás 23:55 |
| Março  **\*(5min)**  **\*\*(15min)** | 42163  28166  13997 | 3826  2568  1258 | 3844  2567  1277 | 3802  2561  1241 | 3849  2548  1301 | 3845  2553  1292 | 3867  2567  1300 | 3888  2603  1285 | 3796  2495  1301 | 3812  2573  1239 | 3833  2572  1261 | 3801  2559  1242 | 18/03/2022 10:45 até 31/02/2022 00:00 ás 23:55 |
| Abril  **\*(5min)**  **\*\*(15min)** | 31328  -  31328 | 2848  -  2848 | 2848  -  2848 | 2848  -  2848 | 2848  -  2848 | 2848  -  2848 | 2848  -  2848 | 2848  -  2848 | 2848  -  2848 | 2848  -  2848 | 2848  -  2848 | 2848  -  2848 | 01/04/2022 até 30/04/2022 00:00 ás 23:55 |
| Maio  **\*(5min)**  **\*\*(15min)** | 31713  -  31713 | 2883  -  2883 | 2883  -  2883 | 2883  -  2883 | 2883  -  2883 | 2883  -  2883 | 2883  -  2883 | 2883  -  2883 | 2883  -  2883 | 2883  -  2883 | 2883  -  2883 | 2883  -  2883 | 01/05/2022 até 31/05/2022 00:00 ás 23:55 |
| Junho  **\*(5min)**  **\*\*(15min)** | 31680  -  31680 | 2880  -  2880 | 2880  -  2880 | 2880  -  2880 | 2880  -  2880 | 2880  -  2880 | 2880  -  2880 | 2880  -  2880 | 2880  -  2880 | 2880  -  2880 | 2880  -  2880 | 2880  -  2880 | 01/06/2022 até 30/06/2022 00:00 ás 23:55 |
| Julho  **\*(5min)**  **\*\*(15min)** | 32736  -  32736 | 2976  -  2976 | 2976  -  2976 | 2976  -  2976 | 2976  -  2976 | 2976  -  2976 | 2976  -  2976 | 2976  -  2976 | 2976  -  2976 | 2976  -  2976 | 2976  -  2976 | 2976  -  2976 | 01/07/2022 até 31/07/2022 00:00 ás 23:55 |
| Agosto  **\*(5min)**  **\*\*(15min)** | 32736  -  32736 | 2976  -  2976 | 2976  -  2976 | 2976  -  2976 | 2976  -  2976 | 2976  -  2976 | 2976  -  2976 | 2976  -  2976 | 2976  -  2976 | 2976  -  2976 | 2976  -  2976 | 2976  -  2976 | 01/08/2022 até 31/08/2022 00:00 ás 23:55 |
| Setembro  **\*(5min)**  **\*\*(15min)** | 31152  -  31152 | 2832  -  2832 | 2832  -  2832 | 2832  -  2832 | 2832  -  2832 | 2832  -  2832 | 2832  -  2832 | 2832  -  2832 | 2832  -  2832 | 2832  -  2832 | 2832  -  2832 | 2832  -  2832 | 01/09/2022 até 30/09/2022 00:00 ás 23:55 |
| Outubro  **\*(5min)**  **\*\*(15min)** | 29579  -  29579 | 2689  -  2689 | 2689  -  2689 | 2689  -  2689 | 2689  -  2689 | 2689  -  2689 | 2689  -  2689 | 2689  -  2689 | 2689  -  2689 | 2689  -  2689 | 2689  -  2689 | 2689  -  2689 | 01/10/2022 até 31/10/2022 00:00 ás 23:55 |
| Novembro  **\*(5min)**  **\*\*(15min)** | 30057  -  30057 | 2732  -  2732 | 2733  -  2733 | 2733  -  2733 | 2732  -  2732 | 2733  -  2733 | 2732  -  2732 | 2733  -  2733 | 2733  -  2733 | 2733  -  2733 | 2733  -  2733 | 2733  -  2733 | 01/11/2022 até 30/11/2022 00:00 ás 23:55 |
| Dezembro  **\*(5min)**  **\*\*(15min)** | 32725  -  32725 | 2975  -  2975 | 2975  -  2975 | 2975  -  2975 | 2975  -  2975 | 2975  -  2975 | 2975  -  2975 | 2975  -  2975 | 2975  -  2975 | 2975  -  2975 | 2975  -  2975 | 2975  -  2975 | 01/12/2022 até 31/12/2022 00:00 ás 23:55 |

## 3.5.1.1. Análise Registos

Tabela - Registos Esperados vs. Registos Validados (15min)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Registos 15min** | | | | |
| **Ano** | **Mês** | **Esperado** | **Validado** | **%Omissos** |
| 2021 | Setembro | 31680 | 17669 | 44.23 |
| 2021 | Outubro | 32736 | 31031 | 5.20 |
| 2021 | Novembro | 31680 | 25819 | 18.50 |
| 2021 | Dezembro | 32736 | 31200 | 4.69 |
| 2022 | Janeiro | 32736 | 32015 | 2.20 |
| 2022 | Fevereiro | 29568 | 6503 | 78.00 |
| 2022 | Março | 32736 | 13731 | 58.05 |
| 2022 | Abril | 31680 | 31328 | 1.11 |
| 2022 | Maio | 32736 | 31713 | 3.12 |
| 2022 | Junho | 31680 | 31680 | 0.00 |
| 2022 | Julho | 32736 | 32736 | 0.00 |
| 2022 | Agosto | 32736 | 32736 | 0.00 |
| 2022 | Setembro | 31680 | 31152 | 1.66 |
| 2022 | Outubro | 32736 | 29579 | 9.64 |
| 2022 | Novembro | 31680 | 30057 | 5.12 |
| 2022 | Dezembro | 32736 | 32725 | 0.03 |

Tabela - Registos Esperados vs. Registos Validados (1h)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Registos 1h** | | | | |
| **Ano** | **Mês** | **Esperado** | **Validado** | **%Omissos** |
| 2021 | Setembro | 7920 | 4477 | 43.48 |
| 2021 | Outubro | 8184 | 7997 | 2.28 |
| 2021 | Novembro | 7920 | 6729 | 15.04 |
| 2021 | Dezembro | 8184 | 8117 | 0.82 |
| 2022 | Janeiro | 8184 | 8184 | 0.00 |
| 2022 | Fevereiro | 7392 | 1668 | 77.43 |
| 2022 | Março | 8184 | 3531 | 56.85 |
| 2022 | Abril | 7920 | 7832 | 1.11 |
| 2022 | Maio | 8184 | 7942 | 2.96 |
| 2022 | Junho | 7920 | 7920 | 0.00 |
| 2022 | Julho | 8184 | 8184 | 0.00 |
| 2022 | Agosto | 8184 | 8184 | 0.00 |
| 2022 | Setembro | 7920 | 7788 | 1.66 |
| 2022 | Outubro | 8184 | 7436 | 9.14 |
| 2022 | Novembro | 7920 | 7535 | 4.86 |
| 2022 | Dezembro | 8184 | 8184 | 0.00 |

Tabela - Descrição geral de entradas e saídas

|  |  |  |
| --- | --- | --- |
| **Variáveis** | **Entradas** | **Saídas** |
| Mínimo | 0.0 | 0.0 |
| Média | 144.51 | 141.12 |
| Máximo | 10270.72 | 10204.88 |
| Desvio Padrão | 174.15 | 166.52 |

## 3.5.1. Conjunto 4

PGIL\_IPMA\_METEO\_OBS\_16.01.2020\_a\_08.09.2022.csv

Registos: 998206

Datas Registadas: 16/01/2020 até 08/09/2022

Datas Uteis: 14/09/2021 até 08/09/2022

Na verdade, através da aplicação da função drop\_duplicates da biblioteca pandas no ficheiro foram apresentados 931 376 registos com exatamente os mesmos valores, cerca de 93.3%, e, portanto, é suposto o dataset apresentar 66830 registos que representem dados uteis para o problema, cerca de 6.7%. No entanto, ainda assim existem dados inconsistentes com valores nulos e incoerentes com o valor -99.0 disposto por várias colunas. De modo que, é necessário dissecar os dados com valores inválidos para a mesma data na mesma estação. Com efeito, após esta dissecação foram validados 45442 registos, o que representam 4.7% do dataset.

Tabela - Registos Esperados vs. Validados (IPMA)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Registos 1H** | | | | |
| **Anos** | **Mês** | **Esperado** | **Validado** | **%Omissos** |
| 2021 e 2022 | Janeiro | 2976 | 2933 | 1.44 |
| 2021 e 2022 | Fevereiro | 2688 | 2681 | 0.26 |
| 2021 e 2022 | Março | 2976 | 2960 | 0.53 |
| 2021 e 2022 | Abril | 2880 | 2842 | 1.32 |
| 2021 e 2022 | Maio | 2976 | 2974 | 0.07 |
| 2021 e 2022 | Junho | 2880 | 2698 | 6.32 |
| 2021 e 2022 | Julho | 2976 | 2972 | 0.07 |
| 2021 e 2022 | Agosto | 2976 | 2974 | 0.07 |
| 2021 e 2022 | Setembro | 2880 | 1779 | 35.07 |
| 2021 e 2022 | Outubro | 2976 | 1422 | 52.22 |
| 2021 e 2022 | Novembro | 2880 | 1440 | 50.00 |
| 2021 e 2022 | Dezembro | 2976 | 1470 | 50.06 |

# . CRISP-DM

É importante compreender as fases do processo CRISP-DM a fim de executar corretamente e com qualidade a análise de dados. Efetivamente, a análise de dados é um processo sistemático que envolve um conjunto de atividades para descobrir padrões, que é o objetivo final da análise de dados.

O modelo CRISP-DM consiste em seis fases.

## 4.1. Compreensão do negócio

A primeira fase do modelo CRISP-DM é a compreensão do negócio, que é uma fase de alto nível para compreender as necessidades de gestão e os objetivos e requisitos do negócio. De facto, é desenvolvido um plano de projeto com base nestes objetivos e é definido um orçamento de alto nível ou uma estimativa de custos para o realizar.

**4.2. Compreensão dos dados**

A segunda fase do modelo CRISP-DM, é a compreensão dos dados. Esta fase envolve a identificação de dados relevantes de diferentes fontes, a recolha dos dados iniciais, a familiarização com os dados, a identificação de problemas de qualidade, a descoberta de novos conhecimentos sobre os dados e a deteção de subconjuntos interessantes para formar hipóteses, através de tabelas ou gráficos.

**4.3. Preparação ou Pré-processamento dos dados**

A terceira fase do modelo CRISP-DM, que é a preparação dos dados. Na fase de preparação, os dados identificados na fase anterior são carregados em um ambiente de analise e passam por um processo de tratamento, transformação, limpeza e seleção para construir um conjunto final de dados a ser usado. Com efeito, esta etapa é a que consome mais tempo e esforço no processo CRISP-DM, cerca de 80% do tempo total, pois os dados do mundo real geralmente são incompletos, ruidosos e inconsistentes e precisam ser corrigidos e melhorados para a análise posterior.

**4.4. Construção ou Modelação de modelos**

Os dados tratados e selecionados são usados na próxima etapa, que é a construção do modelo ou modelação, onde são selecionadas as técnicas e tarefas a serem aplicadas para atender aos objetivos do negócio. Na fase de construção do modelo, pode-se escolher vários modelos possíveis e fazer avaliação comparativa entre eles.

**4.5. Testes e Avaliação**

Na fase teste e avaliação, o modelo construído é testado e avaliado para ver se atende aos objetivos empresariais e para avaliar a precisão e eficiência do modelo em relação aos métodos alternativos. Essa é uma fase crítica, porque se não for descoberto nenhum padrão válido e útil, a mineração de dados pode ser considerada uma perda de tempo e esforço.

**4.6. Implementação**

Na fase de implementação, o modelo de mineração de dados selecionado e avaliado na fase anterior é posto á prova. A implementação pode envolver o uso de software específico ou ferramentas de programação, além de testes e validações contínuas para garantir que o modelo esteja a funcionar corretamente, para que no final, a documentação de todas as etapas do processo de implementação seja rigorosa e fiável para futuras referências.

Uma imagem com diagrama

Descrição gerada automaticamente

Figura - Modelo CRISP-DM. Fonte:[3]

# Mais um capítulo

… as subsecções seguintes são apenas exemplificativas…

## Requisitos

…

## Arquitetura

…

## Tecnologias e ferramentas usadas

… que se encontram descritas na Tabela 4.1.

Tabela . – Tecnologias e ferramentas usadas

|  |  |  |
| --- | --- | --- |
| **Nome** | **Descrição** | **Utilização no projeto** |
| Picasso [2] | Biblioteca *open source* que permite fazer *download* e *caching* de imagens de forma simples. | Utilizada para fazer o *download* das imagens de perfil do utilizador. |
|  |  |  |

## Implementação

…

## Testes e validação

…

### Método A

…

### Método B

…

# Mais um capítulo

…

# Conclusões e trabalho futuro

...

## Conclusões

...

## Trabalho Futuro

…

# Referências [3]

|  |  |
| --- | --- |
| [1] | “ISEC - Licenciaturas - Engenharia Informática,” [Online]. Available: https://www.isec.pt/PT/estudar/licenciaturas/EngInfor/. [Acedido em 8 4 2021]. |
| [2] | “ISEC - Instituto Superior de Engenharia de Coimbra | Instituto Politécnico de Coimbra,” [Online]. Available: https://www.isec.pt/PT/Default.aspx. [Acedido em 8 4 2021]. |
| [3] | C. Y. Wijaya, “CRISP-DM Methodology For Your First Data Science Project,” 26 Abril 2021. |
| [4] | R. Ż. Teresa Pamuła, “Estimation and prediction of the OD matrix in uncongested urban road network based on traffic flows using deep learning,” 4 Novembro 2022. |
| [5] | B. M. Eddy Sanchezdela Cruz, “Urban Traffic Flow Identification by Comparing Machine Learning Algorithms,” p. 7, 7 Fevereiro 2023. |
| [6] | A. B. G. S. J. C.-W. L. Youcef Djenouri, “Hybrid graph convolution neural network and branch-and-bound optimization for traffic flow forecasting,” 22 Setembro 2022. |
| [7] | W. L. Fazel Haq Ahmadzai, “A mobile traffic load prediction based on recurrent neural network: A case of telecommunication in Afghanistan,” 28 Maio 2022. |
| [8] | E. J. K. P. L. R. B. Juan Martín, “Traffic Monitoring via Mobile Device Location,” 17 Outubro 2019. |

# 

# Anexos

Proposta de Estágio

… conteúdo Anexo A

… conteúdo Anexo A

… conteúdo Anexo A

Especificação de Requisitos

… conteúdo Anexo B

… conteúdo Anexo B

… conteúdo Anexo B