

IPV - ESTGV |



## Instituto Politécnico de Viseu

Escola Superior de Tecnologia e Gestão de Viseu

# Instituto Politécnico de Viseu

Escola Superior de Tecnologia e Gestão de Viseu



Aos familiares e amigos



## RESUMO

Com o contínuo envelhecimento das cidades que não são metrópoles, aliado a um constante crescimento da população nos centros urbanos mais desenvolvidos (metrópoles), em certa parte devido às oportunidades de emprego existentes, é crucial que as primeiras passem por um processo de modernização, de modo a captarem a atenção dos seus habitantes.

Para, de alguma forma, (re)capturar o interesse dos seus cidadãos, o paradigma que permite a uma cidade não metrópole atingir este objetivo será transformar a cidade numa *Smart City*. A cidade passará para um novo patamar tecnológico, mudando a maneira como as operações e serviços da cidade atuam e interagem com o cidadão, diferenciando-se assim das outras cidades. Como tal, esta cidade irá tornar-se numa cidade preparada para as tecnologias 5G experimentais, usando estas como base para a transição para uma economia baseada em conhecimento e plataformas digitais, valorizando o grande volume de dados (*Big Data*) recolhidos por vários sensores no ramo da *Internet of Things* (IoT) e o desenvolvimento de novos produtos e serviços para o crescimento e criação de novas empresas e empregos.

Dentro do domínio das *Smart Cities*, mais concretamente na temática dos resíduos, pretende-se com este projeto criar Modelos de Informação Comum, recorrendo a tecnologias *Big Data*, IoT e *Data Science*, de modo a estes serem integrados numa plataforma de gestão de dados, permitindo a um município da cidade, que será um dos clientes desta plataforma, tomar decisões consoante os dados providenciados, tudo com o intuito de otimizar os processos de gestão dos resíduos adjacentes à cidade.

Para se gerarem os Modelos de Informação Comum, aliado às tecnologias mencionadas no parágrafo anterior, será utilizado *Machine Learning*, de forma a aplicar previsão nos conjuntos de dados dentro da temática do lixo. Os conjuntos de dados estão inerentemente ligados aos resíduos da cidade de Austin, sendo que estes dados são de carácter “real”, ou seja, registados pela empresa de recolha que opera na cidade. A escolha do conjunto de dados associados à cidade de Austin prende-se com a disponibilidade em modo aberto dos dados, e por, dentro do contexto do país pertencente, Austin corresponder a uma cidade não metrópole. De modo a se retirar valor dos dados, utilizar-se-á uma tecnologia de *Business Intelligence* que permite a criação e customização de ferramentas que providenciam informação.



## ABSTRACT

With the continuous aging of cities that are not metropolises, coupled with a constant growth of the population in the most developed urban centers (metropolises), in part due to the existing job opportunities, it is crucial that the former undergoes a process of modernization, to capture the attention of its inhabitants.

To somehow (re)capture the interest of its citizens, the paradigm that allows a non-metropolis city to achieve this objective consists in transforming the city into a Smart City. The city will move to a new technological level, changing the way the city's operations and services operate and interact with the citizen, thus differentiating itself from other cities. To this end, the municipality will become a city prepared for experimental 5G technologies, using these as a basis for the transition to a knowledge-based economy and digital platforms, valuing the large volume of data (Big Data) collected by various sensors in the field the Internet of Things (IoT) and the development of new products and services for the growth and creation of new companies and jobs.

Within the domain of Smart Cities, more specifically in the theme of waste, this project aims to create Common Information Models, using Big Data, IoT and Data Science technologies, so that they are integrated into a data management platform, allowing a city council, which will be one of the customers of this platform, to make decisions according to the data provided, all with the aim of optimizing the waste management processes adjacent to the city.

In order to generate the Common Information Models, combined with the technologies mentioned in the previous paragraph, Machine Learning will be used, in order to apply prediction in datasets within the theme of garbage. The datasets are inherently linked to the waste from the city of Austin, and these data are “real” in nature, that is, recorded by the collection company that operates in the city. The motive for choosing a dataset associated with the city of Austin is related to the availability in open mode of the data, and also because, within the context of the country it belongs to, Austin corresponds to a non-metropolis city.

In order to acquire value from the data, a Business Intelligence technology will be used to allow the creation and customization of tools that provide information.



## **PALAVRAS CHAVE**

*Smart City*

*Internet of Things*

*Big Data*

*Data Science*

Modelos de Informação Comum

*Machine Learning*

*Business Intelligence*



## KEY WORDS

*Smart City*  
*Internet of Things*  
*Big Data*  
*Data Science*  
*Common Information Models*  
*Machine Learning*  
*Business Intelligence*



## **AGRADECIMENTOS**

Primeiramente, deixo um grande obrigado aos meus pais, por me terem proporcionado um percurso académico muito enriquecedor, incutindo igualmente as bases cruciais para uma vida independente como um cidadão exemplar.

Agradeço também aos meus restantes familiares, nomeadamente aos tios e avós, por partilharem todas as suas lições de vida e experiências, servindo de pilares para o homem que sou hoje.

Aos amigos de longa data e colegas de trabalho, muito obrigado por representarem o que considero como o derradeiro significado de espírito de equipa e entreajuda.

Deixo um especial agradecimento aos orientadores deste projeto, Ana Cristina Wanzeller Guedes de Lacerda e Filipe Cabral Pinto, pelo acompanhamento, disponibilidade e sugestões durante o decorrer dos trabalhos realizados.

Por fim, um obrigado à empresa Altice Labs, por disponibilizarem uma plataforma essencial para a realização deste projeto.

Do fundo do coração, a todos vós, obrigado!



# ÍNDICE GERAL

ÍNDICE GERAL .....	xiii
ÍNDICE DE FIGURAS .....	xvi
ÍNDICE DE QUADROS .....	xx
ABREVIATURAS E SIGLAS .....	xxii
1. Introdução .....	1
1.1 Contextualização .....	1
1.2 Fundamentação .....	2
1.3 Objetivos .....	3
1.4 Estrutura .....	4
2. Estado da Arte .....	5
2.1 <i>Smart City</i> .....	5
2.2 <i>Big Data</i> .....	8
2.3 <i>Internet of Things</i> .....	10
2.4 <i>Business Intelligence</i> .....	11
2.5 <i>Data Science</i> .....	13
2.6 <i>Machine Learning</i> .....	15
2.7 <i>Machine Learning</i> aplicado a resíduos municipais .....	18
3. Metodologia .....	25
4. Cenário do Projeto .....	29
4.1 Síntese Geral .....	29
4.2 Tipos de Resíduos .....	31
5. Arquitetura de API para processamento do conjunto de dados .....	37
6. Comparação de algoritmos <i>Machine Learning</i> de predição .....	42
6.1 Resultados obtidos da comparação .....	42
6.2 Conclusões da análise efetuada .....	49
7. Arquitetura de API para ingestão de dados de previsão .....	50
7.1 Resultados obtidos da ingestão .....	51

7.2	Submissão dos conjuntos de dados na plataforma de gestão .....	53
8.	Plataforma de Gestão de Dados .....	54
8.1	Breve descrição da plataforma .....	54
8.2	Consulta dos dados correspondentes aos tipos de resíduos.....	55
8.3	Caracterização dos <i>widgets</i> constituintes dos <i>Dashboards</i> BI dos tipos de resíduos existentes.....	57
8.4	Análise da informação consultável nos <i>dashboards</i> BI dos tipos de resíduos existentes.....	62
9.	Conclusões .....	75
9.1	Desafios encontrados.....	77
9.2	Trabalho futuro.....	78
	Referências.....	79
	Apêndice 1 – Figuras dos tempos de processamento da api .....	83
	Apêndice 2 – Contratempos encontrados durante a fase de comparação dos algoritmos.....	89
	Apêndice 3 – Figuras dos tempos de processamento da api de previsão.....	91
	Apêndice 4 – Cronograma do projeto na fase de Pré-Proposta .....	94
	Anexo 1 – Taxa de desemprego: total e por sexo (%) .....	95



## ÍNDICE DE FIGURAS

Figura 2-1: Áreas de atuação numa <i>Smart City</i> .....	7
Figura 2-2: Representação gráfica dos “três Vs” do Big Data.....	9
Figura 2-3: Arquitetura de um dispositivo IoT .....	10
Figura 2-4: Exemplo de um <i>dashboard</i> com informações de vendas.....	12
Figura 2-5: Domínios de conhecimento constituintes de <i>Data Science</i> .....	14
Figura 2-6: Domínios de conhecimento constituintes de <i>Machine Learning</i> .....	16
Figura 2-7: Representação do algoritmo ARIMA.....	20
Figura 2-8: Representação do algoritmo Decision Trees.....	21
Figura 2-9: Representação do algoritmo Artificial Neural Networks.....	22
Figura 2-10: Representação do algoritmo SVM .....	23
Figura 2-11: Representação do algoritmo Prophet .....	24
Figura 3-1: Representação gráfica da metodologia EDP .....	25
Figura 3-2: Representação gráfica da metodologia CRISP-DM.....	27
Figura 4-1: Conjunto de dados do lixo recolhido em Austin, Texas .....	30
Figura 4-2: Página Principal da Plataforma de Gestão de Dados da Altice Labs.....	31
Figura 4-3: Resíduos urbanos em Portugal: total e por tipo de recolha.....	33
Figura 4-4: Dados da gestão de resíduos em Portugal no ano de 2013 .....	34
Figura 5-1: Arquitetura da aplicação Flask desenvolvida.....	38
Figura 5-2: Exemplo de código utilizado na construção da aplicação Flask .....	39

Figura 7-1: Arquitetura da aplicação Flask de predição desenvolvida.....	51
Figura 7-2: Acesso aos dados dos tipos de resíduos na plataforma de gestão de dados .....	53
Figura 8-1: Consulta de dados do tipo Garbage Collections na plataforma de gestão .....	55
Figura 8-2: Consulta de dados do tipo Garbage Collections na plataforma de gestão com editor de <i>query</i> .....	56
Figura 8-3: Resultado da consulta elaborada no editor de <i>query</i> .....	56
Figura 8-4: Gráfico ‘Garbage Collections' recorded data’ .....	58
Figura 8-5: <i>Widgets</i> ‘Recorded data's average of tons processed’ e ‘Recorded data's lowest tons processed’ .....	59
Figura 8-6: Gráfico ‘Recorded data's total of tons processed per month’ .....	60
Figura 8-7: Gráfico ‘Garbage Collections' predicted data’ .....	60
Figura 8-8: <i>Widgets</i> ‘Predicted data's average of tons processed’ e ‘Predicted data's lowest tons processed’ .....	61
Figura 8-9: <i>Widget</i> ‘Predicted data of processed tons for 2021’ .....	62
Figura 8-10: <i>Dashboard</i> do resíduo Garbage Collections.....	63
Figura 8-11: <i>Dashboard</i> do resíduo Recycling – Single Stream.....	64
Figura 8-12: <i>Dashboard</i> do resíduo Sweeping.....	65
Figura 8-13: <i>Dashboard</i> do resíduo Yard Trimming .....	66
Figura 8-14: <i>Dashboard</i> do resíduo Bulk.....	67
Figura 8-15: <i>Dashboard</i> do resíduo Brush.....	68
Figura 8-16: <i>Dashboard</i> do resíduo Recycling – Paper .....	69
Figura 8-17: <i>Dashboard</i> do resíduo Recycling – Comingle .....	70

Figura 8-18: *Dashboard* do resíduo Organics..... 71

Figura 8-19: *Dashboard* do resíduo Mixed Litter..... 72

Figura 8-20: *Dashboard* do resíduo Recycled Metal..... 73

Figura 8-21: *Dashboard* do resíduo Recycling – Plastic Bags ..... 74



## ÍNDICE DE QUADROS

Quadro 4-1: Descrição dos tipos de resíduos constituintes do <i>dataset</i> .....	31
Quadro 4-2: Descrição dos tipos de resíduos do <i>dataset</i> a serem analisados .....	35
Quadro 5-1: Registo dos tempos de processamento obtidos pela API desenvolvida consoante os diferentes tipos de resíduos .....	40
Quadro 6-1: Resultados da previsão dos algoritmos para o resíduo Garbage Collections .....	43
Quadro 6-2: Resultados da previsão dos algoritmos para o resíduo Recycling – Single Stream .....	43
Quadro 6-3: Resultados da previsão dos algoritmos para o resíduo Sweeping .....	44
Quadro 6-4: Resultados da previsão dos algoritmos para o resíduo Yard Trimming.....	44
Quadro 6-5: Resultados da previsão dos algoritmos para o resíduo Bulk .....	45
Quadro 6-6: Resultados da previsão dos algoritmos para o resíduo Brush .....	45
Quadro 6-7: Resultados da previsão dos algoritmos para o resíduo Recycling – Paper.....	46
Quadro 6-8: Resultados da previsão dos algoritmos para o resíduo Recycling – Comingle ...	46
Quadro 6-9: Resultados da previsão dos algoritmos para o resíduo Organics.....	47
Quadro 6-10: Resultados da previsão dos algoritmos para o resíduo Mixed Litter.....	47
Quadro 6-11: Resultados da previsão dos algoritmos para o resíduo Recycled Metal.....	48
Quadro 6-12: Resultados da previsão dos algoritmos para o resíduo Recycling – Plastic Bags .....	48
Quadro 7-1: Registo dos tempos de processamento obtidos pela API de previsão desenvolvida consoante os diferentes tipos de resíduos .....	52



## ABREVIATURAS E SIGLAS

ML	<i>Machine Learning</i>
BI	<i>Business Intelligence</i>
EDP	<i>Engineering Design Process</i>
API	<i>Application Programming Interface</i>
ARIMA	<i>AutoRegressive Integrated Moving Average</i>
ANN	<i>Artificial Neural Networks</i>
SVM	<i>Support Vectoring Model</i>
MAPE	<i>Mean Absolute Percentage Error</i>



# 1. Introdução

Este capítulo é constituído pelas secções: Contextualização/Enquadramento do problema, Fundamentação/Motivação para resolver o problema, Objetivos que advêm do projeto proposto e Estrutura do documento.

## 1.1 Contextualização

A criação de empregos é uma das principais prioridades das economias mundiais. Em Portugal, a percentagem de pessoas desempregadas atingiu um pico de 16,2% em 2013. Este número traduz-se em milhares de trabalhadores qualificados a serem forçados a deixar o país em busca de novas oportunidades. Todavia, o desemprego em Portugal, registado em 2018, como é possível consultar no Anexo 1, rondava os 7%, o que correspondia a uma realidade menos avassaladora.

Contrariando os dados relativos ao desemprego, existe a falta de preenchimento de vagas no setor de tecnologias. Utilizando o exemplo do distrito de Aveiro, nos últimos 12 meses, centenas de vagas no ramo tecnológico não foram preenchidas (Fonte: <https://uia-initiative.eu/en/uia-cities/aveiro>). Este cenário pode não levar apenas a grandes perdas para a competitividade das empresas no mercado, assim como, de igual forma, também à perda de empregos existentes, levando as pessoas a deslocarem-se para as grandes metrópoles, devido ao investimento existente nestas cidades feito pelas grandes multinacionais.

A cidade de Aveiro enquadra-se no espectro de cidades não metrópoles, portanto, para que esta se possa potencializar, conseguindo igualmente incentivar as pessoas a residir na sua província, Aveiro encontra-se a investir em projetos na área das Tecnologias de Informação e Comunicação - TIC.

De acordo com Hollands (2008), as cidades poderão conseguir arrecadar benefícios em termos socioeconómicos, caso adotem tecnologias computacionais, como as TIC, a fim de

incentivar a inovação, a aprendizagem, o conhecimento e a solução de problemas que só as tecnologias oferecem.

Com visão assente nas tecnologias presentes nas habitações da cidade, Balakrishna (2012) considera que todos os objetos e processos integrantes podem estar conectados, recorrendo à internet e/ou a sensores em sintonia com um sistema tecnológico pessoal, com o intuito de tornar a vida humana mais eficiente. Para o autor, utilizar as TI para transformar a vida e o trabalho numa região é visto como sendo o próximo passo evolutivo e lógico a tomar, de modo a que a localidade consiga implementar, para benefício próprio e dos seus habitantes, um crescimento sustentável a médio e longo prazo.

O projeto de grandes dimensões (especificado na próxima secção) que está atualmente a decorrer pretende fomentar as tecnologias de IoT e *Big Data* na cidade, recorrendo a infraestruturas já existentes e futuras, de forma a agregar dados em tempo real dos sensores. Os dados posteriormente poderão ser visualizados numa plataforma de gestão, que poderá ser utilizada por uma câmara municipal, representados em gráficos que servirão de apoio à tomada de decisão por parte do município.

## 1.2 Fundamentação

Um dos projetos de maior relevo no ramo da computação, com investimento a decorrer em Portugal, chama-se “Aveiro STEAM City”. Em termos de valores, isto traduz-se num investimento total **de 6,1 milhões de euros, sendo que parte deste montante corresponde ao financiamento dos Fundos Comunitários de 4,9 milhões**, provenientes do Fundo Europeu de Desenvolvimento Regional (FEDER). Este investimento é resultante da aprovação da candidatura da Câmara Municipal de Aveiro ao programa comunitário *Urban Innovative Actions* (UIA), com envolvimento de instituições tais como a Altice Labs, o Instituto de Telecomunicações e a Universidade de Aveiro. Este projeto serviu de inspiração para a direção dos trabalhos a executar.

O objetivo passa por utilizar tecnologias no ramo da *Internet of Things* (IoT) que possibilitam a visualização, conhecimento e atuação nas diferentes infraestruturas da cidade. Estas tecnologias, aliadas às infraestruturas 5G, servirão de “fonte” à plataforma digital urbana que irá permitir à cidade efetuar a sua gestão, de forma eficiente.

A área de atuação onde o trabalho proposto está inserido, corresponde à área dos resíduos. Serão utilizados conjuntos de dados ligados à vertente dos resíduos que terão de ser analisados e tratados devidamente para que possam ser reencaminhados para a plataforma de gestão. Esta plataforma servirá de ponto centralizado onde o cliente, sendo neste cenário a câmara, poderá consultar os dados diretamente em tabelas, ou visualizar os dados numa componente que incorpora gráficos informativos.

Porém, de modo a enriquecer estes dados para que o cliente consiga ter um suporte válido para a tomada de decisão, seja na questão de definir uma rota de recolha do lixo de contentores mediante a consulta dos *dashboards* da plataforma, ou outras interrogações que se prendam com os tipos de dados que os parceiros providenciaram, serão aplicadas técnicas de

*Machine Learning* para complementar os conjuntos de dados tratados, que irão dar origem a dados de previsão, permitindo criar gráficos informativos que o cliente possa consultar de modo a tomar uma decisão preventiva. Estas técnicas serão aplicadas antes da inclusão dos dados na plataforma de gestão.

Falando de problemas e limitações a nível dos resíduos que existem nas cidades, em determinadas alturas, a recolha é efetuada sem que os contentores estejam cheios, ou, falando do contrário, os contentores ultrapassem a sua carga máxima, sem que a recolha seja feita. Outra inconveniência passa pela gestão da quantidade de resíduos processada, nomeadamente em épocas temporais que albergam festividades tais como o Natal e a Páscoa, onde, por vezes, é difícil para os municípios atuarem de forma acertada, não sabendo se é necessário ou não alocar mais recursos para processar resíduos municipais gerados pelos cidadãos.

É nesta perspetiva de providenciar suporte à tomada de decisão que o projeto se pretende inserir, fornecendo informações de interesse ao município, de forma a reduzir custos com a gestão dos resíduos inerentes à cidade, bem como aumentar o grau de satisfação dos munícipes relativamente tratamento dos resíduos, visando também a sua consciencialização e disponibilidade para futura colaboração na minimização deste problema.

Ao tratar os dados recolhidos dos sensores de parceiros e aliar esta componente à plataforma de gestão de dados da empresa, o município acede à plataforma e consegue retirar valor dos dados conseguindo, ultimamente, otimizar os seus processos.

O trabalho realizado permitirá a uma cidade dar um passo muito importante de forma a tornar-se numa *Smart City*, podendo, deste modo, pertencer ao grupo de cidades na vanguarda da tecnologia que, por sua vez, permite oferecer melhores condições aos seus habitantes, assim como atrair mais empresas com produtos inovadores, que possibilitam à cidade sustentar o seu crescimento.

### 1.3 Objetivos

O trabalho a ser efetuado exige que sejam utilizadas metodologias de *Data Science* para a ingestão, normalização, enriquecimento, limpeza e extração de conhecimento, a partir de dados relacionados com a gestão de cidades (*Smart Cities*). O cenário a servir de enquadramento para as metodologias a implementar prende-se com o tratamento de resíduos da cidade de Austin.

Os dados serão utilizados na criação de modelos de classificação e previsão de *Machine Learning*, de forma a apoiar a decisão para o enriquecimento e otimização da gestão da recolha de resíduos das cidades, recorrendo posteriormente a uma plataforma de gestão de dados e analítica da Altice Labs para os armazenar e visualizar. No que toca aos tipos de dados que serão tratados, estes passarão por aspetos tais como o volume de carga coletada e tipo de carga processada.

Em suma, pretende-se com os dados providenciados, neste caso, acerca de resíduos e questões que se relacionem com o seu tratamento, fornecer informação que apoie a câmara e/ou os seus parceiros na otimização da gestão operacional dos processos a nível dos resíduos. Esta informação será incorporada na plataforma integradora, detida pela empresa, a qual permitirá

interligar vários tipos de dados no âmbito de *Smart Cities*. O trabalho a ser efetuado tem a ambição de, por um lado, apoiar o melhoramento da atividade de tratamento de resíduos, e, por outro lado, contribuir para a plataforma da empresa, de forma a comprovar a utilidade da plataforma integradora detida por esta, a qual vai conciliar dados de diferentes cenários.

Para o estipulante do trabalho, em conjugação com o objetivo principal, a meta a ter em atenção será a contextualização aprofundada no termo *Data Science* e o conhecimento adquirido nesta área, sobretudo nas funções que um cientista de dados desempenha, o valor que este acrescenta à entidade empregadora, as responsabilidades que esta função exige, entre outras potenciais descobertas.

### 1.4 Estrutura

Este documento encontra-se dividido em nove capítulos principais. O capítulo de Introdução visa fazer uma exposição fundamentada do problema atual das cidades, que enfrentam um conjunto de questões relacionadas com a gestão dos recursos públicos.

Segue-se o capítulo do Estado de Arte, o qual pretende atribuir reconhecimento aos estudos que foram feitos nas áreas relacionadas com o projeto a realizar.

Com o capítulo da Metodologia, explicam-se os métodos de investigação e operações que serão aplicadas no projeto.

Transitando para o capítulo do Cenário do Projeto, neste será esclarecido o *dataset* alvo de análise, assim como o valor que se pretende retirar do mesmo.

Seguindo para a Arquitetura de API para processamento do conjunto de dados, para este será descrita a API desenvolvida que permite aplicar pré-processamento nos dados do conjunto utilizado, assim como serão apresentados os resultados obtidos no processo.

No capítulo da Comparação de algoritmos *Machine Learning* de predição, serão analisados, comparados e classificados os potenciais algoritmos responsáveis pelo enriquecimento do conjunto de dados.

Passando para o capítulo da Arquitetura de API para ingestão de dados de previsão, neste encontra-se abordado o método de criação dos conjuntos de dados enriquecidos com dados de previsão, originados pela nova API desenvolvida. Presentes igualmente estão os resultados obtidos do processamento da API, assim como certos requisitos para o registo dos dados armazenados na plataforma da empresa.

De seguida, no caso do capítulo da Plataforma de Gestão de Dados, descreve-se em que consiste a plataforma utilizada, o(s) método(s) de acesso aos dados armazenados, os *widgets* constituintes dos *dashboards* BI elaborados, e o valor retirável dos dados presentes nos *dashboards* dos tipos de resíduos.

Por fim, na Conclusão é feita uma prospeção do trabalho implementado, assim como uma previsão do trabalho futuro em vista derivado deste projeto.

## 2. Estado da Arte

O capítulo “Estado da Arte” é constituído por secções que abordam os seguintes temas do projeto: *Smart City*, *Big Data*, *Internet of Things*, *Machine Learning*, *Data Science* e *Business Intelligence*

### 2.1 *Smart City*

O termo *Smart City*, de acordo com Batty et al. (2012), surgiu durante a última década como uma fusão de ideias sobre como as Tecnologias da informação e Comunicação (TIC) podem melhorar o funcionamento das cidades, otimizando a eficiência (na utilização de recursos), aumentando a competitividade (da cidade) e fornecendo novas maneiras pelas quais os problemas da pobreza, privação social e ambiente poluído podem ser abordados.

Este conceito tem surgido cada vez mais em parceria com o tópico IoT, apesar de este primeiro ainda não estar devidamente definido. De acordo com Zanella et al. (2014), a aplicação do paradigma IoT a um contexto urbano serve de resposta ao forte impulso demonstrado por muitos governos nacionais na subscrição de soluções de TIC na gestão de assuntos públicos, adotando desta forma o chamado conceito de *Smart City*. Continuam a afirmar que, apesar de ainda não existir uma definição formal de “*Smart City*”, o objetivo final passa por fazer um melhor uso dos recursos públicos, aumentando a qualidade dos serviços oferecidos aos cidadãos e reduzindo os custos operacionais de administrações públicas.

No que toca ao reconhecimento de uma definição geral para *Smart City*, a mesma ideia é, de certa forma, partilhada por Chourabi et al. (2012), onde admitem que embora exista um aumento na frequência do uso da frase “*Smart City*”, ainda não existe consenso sobre o conceito entre vários profissionais envolvidos na área.

Porém, Caragliu, Del Bo, & Nijkamp (2011) acreditam firmemente que uma cidade (*city*) é inteligente (*smart*) quando esta investe em capital humano e social e em Tecnologias de

Informação e Comunicação (TIC), permitindo desta forma um crescimento económico sustentável e alta qualidade de vida, com uma gestão inteligente dos recursos naturais, por vias de governação participativa.

Ainda que um dos principais focos das *Smart Cities* seja os seus habitantes, não podemos descartar o potencial de negócio (e forte parceria) com a Câmara Municipal que uma destas cidades disponibiliza para as empresas na área das Tecnologias da Informação. Atualmente, segundo Höjer & Wangel (2015), as *Smart Cities* são um conceito dinamizado pelo sector dos negócios. Descrevem como sendo um termo que atrai enorme interesse por parte de empresas envolvidas nas TIC e infraestruturas. No aspeto comercial, defendem que fornecer soluções TIC a uma *Smart City* possui bastante potencial para definir um conceito que pode ser orientado ao sector público de administradores da cidade.

Relativamente às peças essenciais constituintes de uma *Smart City*, Kumar & Dahiya (2017) defendem que uma *Smart City* é composta por seis elementos fundamentais: pessoas inteligentes, economia inteligente, mobilidade inteligente, ambiente inteligente, habitação inteligente e governação inteligente, destacando as pessoas como sendo o “motor” do sistema. O principal destaque é atribuído às "pessoas inteligentes", pois sem a sua participação e envolvimento ativo num sistema *Smart City*, este nunca iria funcionar.

A mesma linha de pensamento é partilhada por Anthopoulos (2015), onde descreve uma *Smart City* como sendo um espaço urbano com funcionalidades inovadoras que se encontram agrupadas em seis dimensões.

A primeira dimensão corresponde à das Pessoas, referindo-se à descoberta e atendimento aos requisitos atuais e futuros. De seguida tem-se a da Habitação, onde o objetivo passa por melhorar a qualidade de vida e a coerência social, bem como a eficiência em relação a energia, comida, água e afins. A próxima dimensão é a do Meio Ambiente, de onde constam aspetos como a proteção, controlo de resíduos e emissões, e resiliência contra as alterações climáticas. Posteriormente, a dimensão da Governação, que se refere à garantia da utilidade urbana e disponibilidade de serviços. Seguidamente tem-se a Economia, no que toca de crescimento sustentável e competitividade da cidade. Por fim, a dimensão da Mobilidade, que aborda a transportação e o controlo de problemas no tráfego.

Estas dimensões podem, de certa forma, ser encontradas no artigo escrito por Gaur et al. (2015), onde explicam que o conceito de *Smart City* é implementado num ambiente urbano complexo, incorporando vários sistemas complexos de gestão da infraestrutura, comportamento humano, tecnologia, estruturas sociais e políticas, e economia. Para estes autores, uma cidade inteligente providencia uma maneira inteligente de gerir componentes, tais como o transporte, saúde, energia, casas e edifícios e o meio ambiente.

Consultando a seguinte figura, Figura 2-1, é possível visualizar de forma sintetizada quais as áreas de atuação de uma *Smart City* que estão intrinsecamente ligadas às dimensões explicitadas previamente:

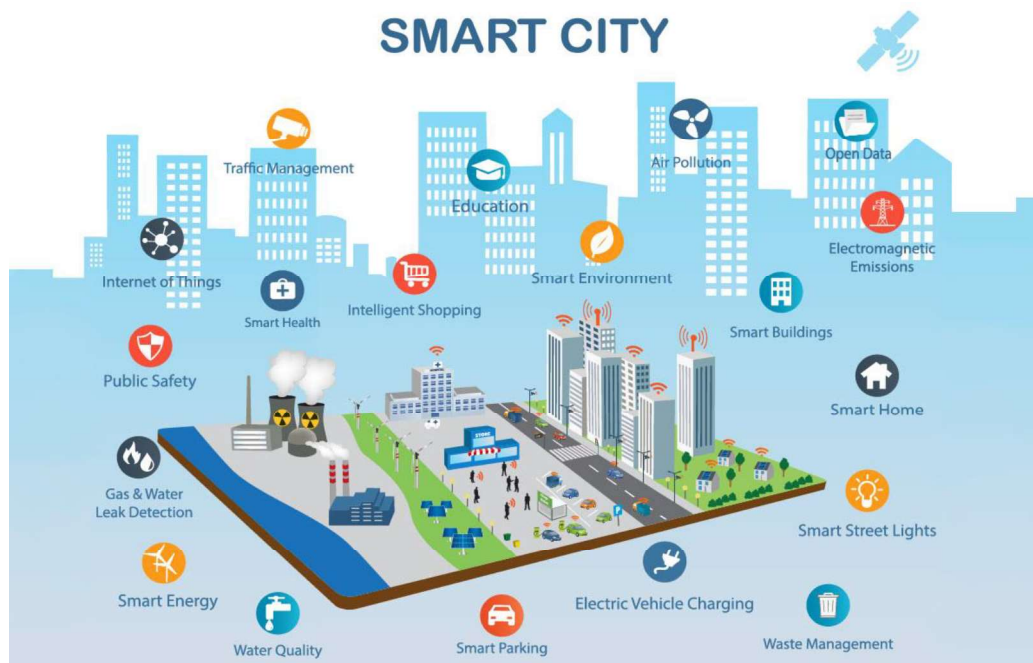


Figura 2-1: Áreas de atuação numa Smart City (Retirado de: <https://www.digitalsign.pt/smartsities/img/smarts-cities-infrastructure-iot.jpg> a 25 de Janeiro de 2020)

Focando na perspectiva de gestão do meio ambiente de uma *Smart City*, e enquadrando de certa forma com o cenário-alvo deste projeto, Mohanty, Choppali, & Kougiianos (2016) consideram como crucial reduzir as emissões de carbono e os resíduos urbanos para aumentar a sustentabilidade e a eficiência da cidade.

Continuando no mesmo tópico, um estudo feito por Anagnostopoulos, Zaslavsky, & Medvedev (2015) consistia em propor e aplicar um algoritmo robusto que otimizava dinamicamente a rota dos camiões de recolha de resíduos da cidade, com base em cenários homogéneos e heterogéneos.

Notando casos de estudos a nível internacional, na temática dos resíduos, temos primeiramente um sistema desenvolvido por Sharma, Singha, & Dutta (2015) na cidade de Jalandar, na Índia. Este sistema consistia em monitorizar os contentores da cidade, atribuindo um indicador a cada caixote, no qual era descrito o quão cheio ou vazio estava cada contentor. O sistema de monitorização permitia visualizar em tempo real o grau de enchimento dos caixotes, pelo que era possível atuar rapidamente e eficientemente nos contentores que precisavam de ser esvaziados.

Similarmente, Folianto, Low, & Yeow (2015) implementaram um sistema inteligente de classificação de enchimento dos contentores do lixo, indicando se os caixotes estavam cheios ou vazios. Para além desta componente, os sensores encarregados de enviar os dados dos caixotes para o sistema tinham um ciclo de trabalho configurável, de modo a reduzir o consumo de energia adjacente a estes.

Na área metropolitana de Copenhaga, Dinamarca, Gutierrez et al. (2015) criaram um sistema uniformizado que verifica, com base em estimativas calculadas pelos autores, o grau de enchimento para cada um dos três mil e quarenta e seis contentores do lixo monitorizados,

fornecendo uma rota de recolha de lixo a cada uma das dezoito equipas de recolha simuladas. Todos estes dados recolhidos seriam armazenados num único servidor, onde este estaria encarregue de processar os dados e calcular as rotas.

## 2.2 *Big Data*

Uma *Smart City* abrange vários domínios do conhecimento. Um destes é, inegavelmente, *Big Data*, tal como abordado por Batty (2013), onde afirma existir uma coincidência entre o que agora está a ser denominado de *Smart Cities* e *Big Data*, sendo que a inteligência destas cidades corresponde principalmente às formas pelas quais os sensores podem gerar novos fluxos de dados em tempo real e como as bases de dados geradas posteriormente podem ser integradas de forma a criar valor.

Al Nuaimi et al. (2015) defendem que a principal força do conceito de *Big Data* é a alta influência que este terá em vários aspetos de uma *Smart City* e, conseqüentemente, na vida das pessoas. Descrevem este conceito como um que possui altos níveis de crescimento, particularmente em relação à taxa projetada de crescimento anual na quantidade de dados globais gerados por ano. Devido a tal, defendem, muitos governos começaram a utilizar *Big Data* para servir de suporte ao desenvolvimento e sustentabilidade das *Smart Cities* existentes mundialmente.

Esta linha de pensamento é expandida por Kitchin (2014), esclarecendo que o “hype” e a esperança do *Big Data* residem na transformação no conhecimento e na governança das cidades, através da criação de um dilúvio de dados que procura fornecer um entendimento e controlo em tempo real muito mais sofisticado, numa escala mais ampla e de granulação mais fina.

No que toca à definição do termo, Sagioglu & Sinanc (2013) decretam *Big Data* como sendo conjuntos de dados massivos de estrutura grande, variada e complexa. Afirmam que estes conjuntos trazem de forma integrada dificuldades a nível de armazenamento de processos, e análise e visualização de resultados.

Segundo os autores, os processos de pesquisa em grandes quantidades de dados permitem revelar padrões e correlações úteis para empresas ou organizações, tudo com o propósito de obter *insights* mais aprofundados e obter uma vantagem sobre a concorrência.

Expandindo nesta temática, Davenport, Barth, & Bean (2012) acreditam que é inegável o fato de cada vez mais as empresas trabalharem com dados volumosos e não-estruturados. De acordo com os autores, as empresas estão a tirar proveito de *Big Data* para entender o seu negócio a um nível mais granular, nomeadamente através da recolha de dados de dispositivos, em tempo-real.

Madden (2012) complementa esclarecendo que cada vez mais as organizações lidam com coleções de dados na ordem dos *Petabytes*, provenientes de fluxos de dados, históricos de transações e sensores.

Em termos teóricos, como consta Russom (2011), a maioria das definições de *Big Data* concentram-se apenas no tamanho dos dados no momento em que são armazenados. O autor, não descartando a importância desta componente, complementa afirmando que atributos como

a variedade e a velocidade dos dados são de igual forma importantes quando se quantifica *Big Data*.

É nesta perspetiva que surgiu a definição dos “três Vs” do *Big Data*: volume, variedade e velocidade. A figura seguinte, Figura 2-2, consiste em representar graficamente os tópicos associados a cada um destes “Vs”:

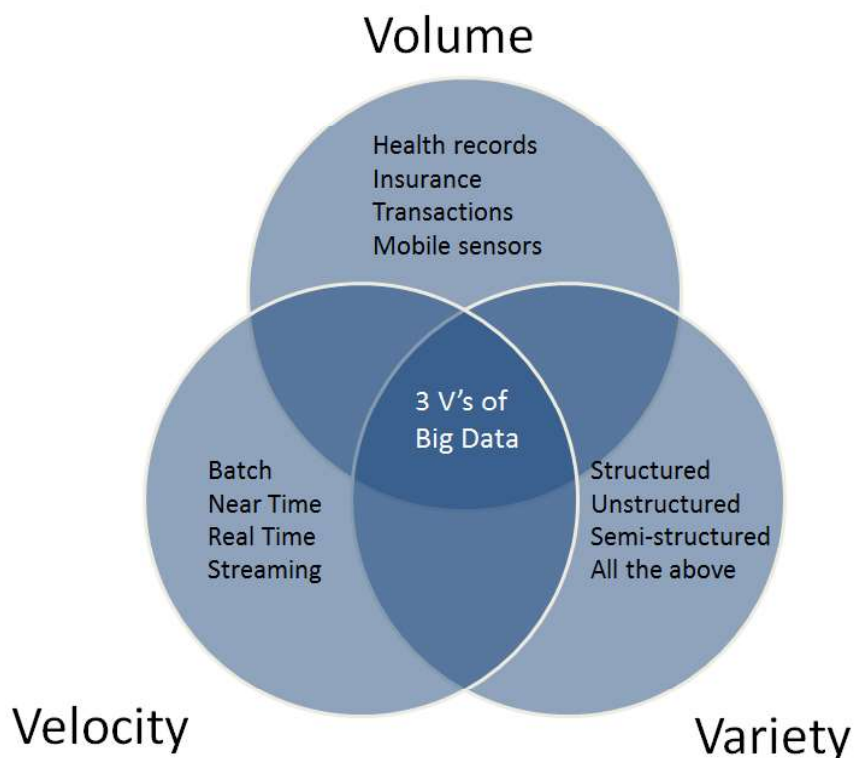


Figura 2-2: Representação gráfica dos “três Vs” do Big Data (Retirado de: [https://www.researchgate.net/figure/The-3-Vs-of-Big-Data-Volume-variety-and-velocity-have-been-widely-accepted-as-the-key\\_fig1\\_292142503](https://www.researchgate.net/figure/The-3-Vs-of-Big-Data-Volume-variety-and-velocity-have-been-widely-accepted-as-the-key_fig1_292142503) a 25 de Janeiro de 2020)

Cada uma destas dimensões assume determinadas temáticas:

- **Volume:** Refere-se ao tamanho dos dados presentes nas tabelas e ficheiros do sistema, gerados por registos e transações, podendo atingir valores na ordem dos *Terabytes* e *Petabytes*.
- **Variedade:** Relaciona-se com a estruturação dos dados recolhidos, ou seja, se estes estão estruturados, semiestruturados ou não-estruturados.
- **Velocidade:** Esta dimensão está interligada à forma como os dados chegam ao sistema, podendo estes chegar em tempo-real ou por *batch*.

Para Russom, os “três Vs” do *Big Data* constituem uma definição mais correta do termo, e rejeitam a ideia de que *Big Data* se refere apenas a volume de dados.

## 2.3 Internet of Things

As áreas de *Internet of Things* (IoT) e *Big Data*, quando conciliadas, permitem a criação de sistemas muito poderosos, devido aos dispositivos IoT captarem milhares de dados sobre várias métricas, para posteriormente serem armazenados recorrendo a *Big Data*. Como tal, uma *Smart City* deverá usufruir dos sensores ligados à área da IoT.

Numa definição proposta por Xia et al. (2012), os autores sugerem que IoT refere-se à rede de objetos do quotidiano que estão conectados entre si. Continuam afirmando que a IoT consegue potencializar a presença destes artefactos na Internet, integrando todos os objetos interativos em um ou vários sistemas, o que permite desenvolver uma rede distribuída capaz de comunicar com seres humanos e outros dispositivos.

Entretanto, Chui, Löffler, & Roberts (2010) consideram que o conceito por detrás de IoT é constituído por sensores e atuadores embutidos em objetos físicos conectados por redes com e/ou sem fios, podendo partilhar o mesmo endereço IP (*Internet Protocol*).

Estas redes, afirmam, produzem enormes volumes de dados que são transmitidos para os computadores, onde estes irão analisar o recolhido. Os autores acreditam que estes sensores e atuadores são ferramentas capazes de entender a complexidade do ambiente onde se situam, atuando rapidamente consoante o requerido pela situação. A Figura 2-3, corresponde a uma possível arquitetura constituinte de um sensor na área da IoT:

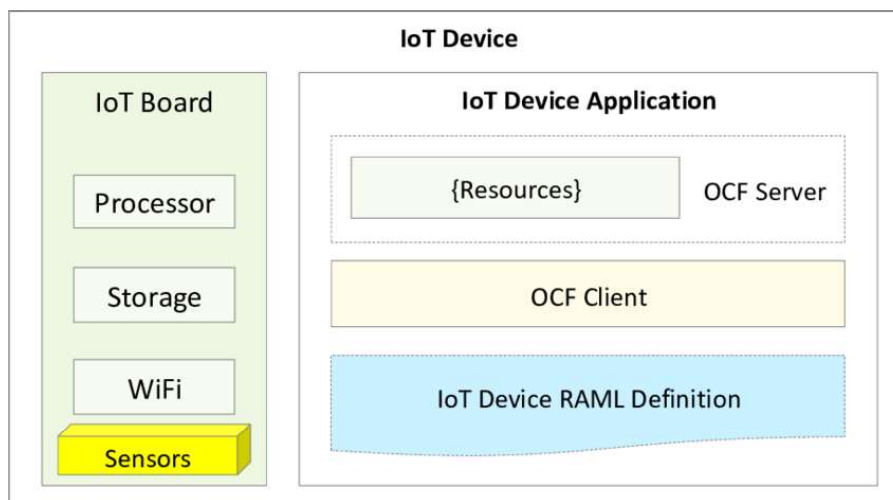


Figura 2-3: Arquitetura de um dispositivo IoT (Retirado de:

<https://www.researchgate.net/publication/329545895/figure/fig4/AS:702465641033730@1544492268810/Functional-architecture-of-Internet-of-Things-IoT-device.png> a 26 de Janeiro de 2020)

Expandindo no tópico de IoT, Rathore et al. (2016) sugerem que poderão haver muitas pessoas hesitantes no que toca à forma como a IoT pode ser usada para estabelecer e construir uma *Smart City*. De modo a atingir este objetivo, os autores sugerem que é preciso incorporar muitos sensores em locais diferentes (da cidade) para coletar os dados e analisá-los para dar um melhor uso. O objetivo final é alcançar casas inteligentes, estacionamento inteligente,

sistemas de climatização, gestão do tráfego de veículos, população ambiental e sistemas de vigilância.

Para demonstrar o valor criado pelas tecnologias IoT, Wortmann, & Flüchter (2015) recorreram a uma simples lâmpada para exemplificar como tal pode surgir. A principal função de uma lâmpada é fornecer luz num local específico. Porém, de acordo com os autores, se forem aplicadas tecnologias IoT à lâmpada, esta, adicionalmente, poderá detetar a presença humana (em horas onde, por exemplo, um parque público está encerrado) e funcionar como um sistema de segurança de baixo custo, onde, no caso de intrusão, ativa um modo de luz intermitente e envia um alerta para que alguém possa averiguar a situação.

Outro exemplo dado dentro desta temática, Pang et al (2015) explicam que, quando um lote de banana é transportado do Brasil para a Suécia, a jornada tem uma duração de, tipicamente, 6 semanas, de onde constam meios de transportação tais como navios, camiões e comboios. Como descrevem os autores, colher as bananas antes de estarem maduras e amadurecê-las durante o transporte (em ambientes controlados) é um método comum para lidar com a longa jornada que a fruta tem de percorrer.

O que pode acontecer é, por vezes, a colheita ser feita já de forma tardia. Caso isto aconteça, este incidente reduz significativamente a vida útil das bananas, levando a que muitas destas acabem por se estragar antes de chegarem ao destinatário. Se este incidente puder ser detetado em tempo útil, recorrendo a sensores integrados num sistema IoT, o destino poderá ser alterado para um local mais curto, onde os autores dão o exemplo de França, para se fazer a colheita.

Portanto, encontrando-se de acordo com a previsão de vida útil das bananas transportadas, é possível prevenir um risco desagradável mitigando, conseqüentemente, o prejuízo que poderia advir desta situação.

No que toca a relações com outras temáticas, o ramo da IoT está intrinsecamente ligado a *Big Data*, devido à quantidade de dados gerados pelos dispositivos. De acordo com Al-Ali et al. (2017), a grande quantidade de dados armazenados nas diferentes cidades de um país apresenta vários desafios no armazenamento, organização e análise de dados.

Os autores continuam afirmando que a IoT e *Big Data* são candidatos naturais para enfrentar estes desafios, pois as tecnologias de IoT podem monitorar e controlar o consumo de energia de eletrodomésticos em larga escala. De igual forma, os autores defendem que a tecnologia *Big Data* pode ser utilizada para recolher e analisar grandes quantidades de dados.

## **2.4 Business Intelligence**

Mesmo após o tratamento dos dados recolhidos, poderá haver a necessidade de se implementarem mecanismos destinados aos utilizadores não envolvidos nas áreas de TI, para que estes possam adquirir um conhecimento mais aprofundado nas medidas a tomar para a gestão da cidade. Para o efeito, agrega-se mais uma componente de possível relevância: *Business Intelligence* (BI).

Este tópico é definido por Santos & Ramos (2006), explicitando que os sistemas de BI utilizam os dados disponíveis nas organizações para disponibilizarem informação relevante para a tomada de decisão. As autoras defendem que estes sistemas agregam um conjunto de ferramentas de interrogação e exploração dos dados com ferramentas que permitem gerar relatórios, de forma a produzir informação que será posteriormente utilizada pela gestão de topo das organizações, no suporte à tomada de decisão.

A apoiar esta definição temos Hočevar & Jaklič (2010), que consideram BI como sendo um processo organizado e sistemático pelo qual uma organização reúne, analisa e propaga informações de fontes internas e externas relevantes para as suas atividades de negócio, de modo a dar suporte à tomada de decisão.

A seguinte figura, Figura 2-4, corresponde a uma amostra de um *dashboard* que pode ser criado recorrendo a uma plataforma BI:



Figura 2-4: Exemplo de um dashboard com informações de vendas (Retirado de: <https://i.pinimg.com/originals/77/21/fa/7721facfef29bd80a17a2a8007168add.png> a 27 de Janeiro de 2020)

No que toca às empresas que utilizam estes sistemas, Gangadharan & Swami (2004) referem que BI não se restringe apenas às empresas multinacionais com capacidade para grandes investimentos. Defendem que as próprias Pequenas e Médias Empresas (PMEs) podem investir nestes sistemas inteligentes, de forma a recolherem e analisarem informações relevantes.

Porém, Moss & Atre (2003) alertam para orientações que indicam se o investimento e adoção de sistemas de BI é justificável. Primeiramente, os Fatores de Negócio, onde é essencial

identificar os motivos de negócios de modo a implementar BI, ou seja, os objetivos estratégicos da empresa. Os objetivos da solução de BI devem estar alinhados de acordo os objetivos empresariais. De seguida, os Requisitos de Análise dos Negócios, sendo que as informações necessárias para alcançar os objetivos estratégicos definidos previamente devem estar clarificados.

O próximo passo prende-se com a Análise de Custos e Benefícios, onde deve ser efetuada uma avaliação dos custos de implementação e manutenção do sistema de BI e uma avaliação dos benefícios esperados. Os benefícios tangíveis da ferramenta devem ser avaliados financeiramente, enquanto os benefícios inatingíveis devem ser definidos de forma qualitativa. Por fim, a Avaliação do Risco, onde é efetuada uma definição dos riscos relacionados à complexidade do sistema, integração do sistema no negócio e sistemas de informação já existentes.

Expandindo nos benefícios de BI, Babu (2012) afirma que as ferramentas de BI permitem eliminar suposições dentro de uma organização, melhorar a comunicação entre os departamentos desta, e permite à empresa atuar rapidamente em casos que esta considere ser digna de atuação (e. g. redução da produção de um certo produto, esvaziar contentores monitorizados no local de trabalho), tudo de acordo com a informação providenciada pela ferramenta.

Num estudo onde o autor desta tese e colegas, Santos et al. (2019), investigaram o tópico, concluiu-se que as possibilidades que os relatórios (*reports*) de BI oferecem, seja por informações claramente compreensíveis ou por interação facilitada, juntamente com painéis intuitivamente projetados para permitir uma avaliação mais acertada por parte do utilizador, fazem com que BI seja crucial para que as organizações tenham um futuro enriquecedor.

### **2.5 Data Science**

*Data Science* corresponde à ciência que recorre à utilização de métodos científicos computacionais para explorar dados, usando técnicas de forma a encontrar padrões que possam ser relevantes.

A figura seguinte, Figura 2-5, efetua uma exposição das áreas de conhecimento que são englobadas por *Data Science*. De notar que alguns destes domínios podem ser, de igual forma, inequivocamente associados a *Machine Learning*. Isto sucede devido à aparente dependência gerada entre os dois termos, nomeadamente na implementação de projetos na área da investigação:

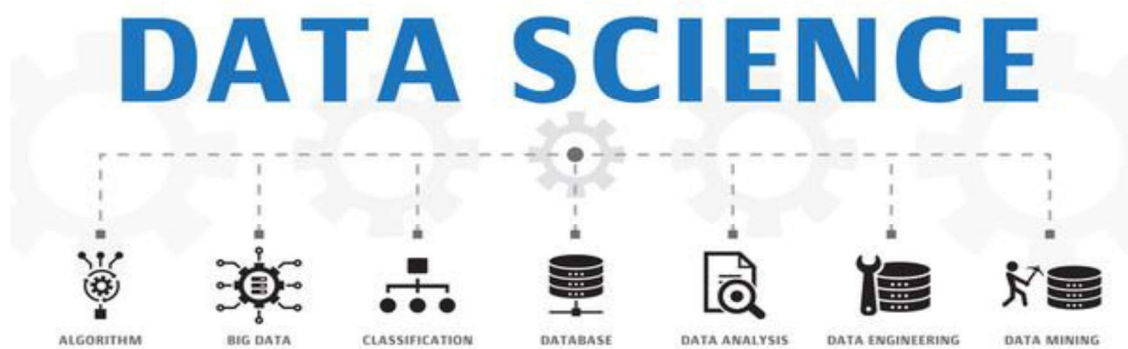


Figura 2-5: Domínios de conhecimento constituintes de Data Science (Retirado de: [https://t3.ftcdn.net/jpg/02/27/72/50/240\\_F\\_227725036\\_wCuI6jWNn2NY6y019FFJBL1QFitVmIUZ.jpg](https://t3.ftcdn.net/jpg/02/27/72/50/240_F_227725036_wCuI6jWNn2NY6y019FFJBL1QFitVmIUZ.jpg) a 26 de Janeiro de 2020)

De acordo com Tansley & Tolle (2009), as funções de um *Data Scientist* podem ser fragmentadas em três papéis distintos:

- **Data Engineer:** Pessoas que desempenham esta função operam, sobretudo, com código a baixo-nível, podendo ter conhecimento na área ML;
- **Data Analyst:** Corresponde ao grupo de pessoas encarregues de explorar dados recorrendo a métodos de estatística e análise;
- **Data Steward:** Os trabalhadores que exercem esta função são responsáveis por gerir e preservar os dados com valor para a empresa.

De referir que, num contexto mais generalista, não existe uma definição absoluta de *Data Science*. Como podemos ver sintetizado por Van Der Aalst (2016), um cientista de dados deve recorrer a técnicas de análise focadas em dados, como ML, de modo a obter valor para a organização. Porém, Waller & Fawcett (2013) introduzem *Data Science* como um conceito onde se aplicam métodos quantitativos e qualitativos para resolver problemas relevantes, de forma a prever resultados.

Aprofundando um pouco no tópico *Data Science* e *Data Mining*, e sobretudo como estes se relacionam, Provost & Fawcett (2013) referem que o primeiro consiste sobretudo num conjunto de princípios fundamentais que apoiam e orientam a extração de informações e conhecimentos a partir dos dados. É da opinião destes que o conceito intrinsecamente mais relacionado a *Data Science* é *Data Mining*, onde este segundo corresponde à extração do conhecimento dos dados recorrendo a tecnologias.

A apoiar as declarações destes dois cientistas temos Han, Kamber, & Pei (2011), constatando que *Data Mining* é o processo de descobrir padrões e conhecimentos de interesse, através da análise e extração de conhecimento a partir de conjuntos de dados volumosos.

De igual forma, um estudo do Departamento de Engenharia Informática da Universidade de Illinois (UI, 2006) defende que *Data Mining* corresponde ao processo de extração de padrões ou informações interessantes, quer estas se tratem de assuntos não triviais, implícitos, anteriormente desconhecidos e/ou potencialmente úteis, a partir de grandes quantidade de dados.

Falando de casos de estudo na temática dos resíduos, temos um projeto elaborado por Verma, Wei, & Kusiak (2013), onde, a partir de um conjunto de dados recolhidos na cidade de Des Moines, Iowa, foi utilizado *Data Mining* para prever Partículas Não Dissolvidas (TSS) na água dos esgotos.

## 2.6 *Machine Learning*

Uma das formas mais populares de obter informação valiosa através do processamento de grandes volumes de dados é denominada de *Machine Learning* (ML). Como refere Alpaydin (2020), ML consiste em programar computadores de forma a otimizar critérios de desempenho, recorrendo a dados baseados em acontecimentos prévios. Continua, dizendo que a utilidade desta componente é particularmente elevada quando simplesmente não é possível desenvolver um programa para resolver um determinado problema, sem se efetuar previamente um estudo dos dados.

Expandindo na definição de ML, Ayodele (2010) categoriza o termo como constituído pelos seguintes tipos:

- ***Supervised Learning***: Neste tipo de ML, o algoritmo gera uma função que mapeia as entradas para as saídas desejadas, tendo já conhecimento das entradas existentes no sistema. Dentro do tipo *Supervised Learning*, existem duas categorias de funcionamento distintas, a **classificação** e a **regressão**. Na classificação, o algoritmo deve recorrer a uma função que mapeia um determinado vetor numa das várias classes existentes (ou seja, a amostra pertence a duas ou mais classes), tendo por objetivo o facto de que, com os dados já classificados, o algoritmo consiga aplicar essa classificação a amostras que ainda não estão rotuladas, observando vários exemplos de entrada e saída da função. Por sua vez, na regressão, os algoritmos recorrem a saídas que possuem variáveis numéricas que podem abranger um número infinito de valores, entre dois intervalos especificados nas saídas, de modo a controlar o valor da saída calculado.
- ***Unsupervised Learning***: No caso deste tipo de ML, o funcionamento dos algoritmos têm apenas por base as observações de saídas, não existindo um padrão para os dados de entrada que se relacionem diretamente com o resultado final. Neste caso, o objetivo é encontrar um padrão nos atributos de entrada de forma a se obter um dado de saída.
- ***Semi-supervised Learning***: Este tipo de aprendizagem combina exemplos de entradas padronizadas e não-padronizadas de modo a gerar uma função ou classificação apropriada ao tipo de dados em análise.
- ***Reinforcement Learning***: Os algoritmos associados a este tipo aprendem de acordo com as suas configurações iniciais. O algoritmo é “recompensado” por executar corretamente e “multado” por executar incorretamente. Para que a aprendizagem seja feita da maneira mais eficaz possível, maximizando as recompensas e minimizando as penalizações, não é feita em qualquer circunstância intervenção após as configurações iniciais.
- ***Transduction***: Este tipo de ML é semelhante no seu funcionamento ao *Supervised Learning*, tendo como diferença chave a construção de duas ou mais funções, tentando

prever novas saídas com base em entradas e saídas de dados alocados para treino do algoritmo.

- **Learning to Learn:** No caso deste tipo de ML, o algoritmo aprende a aplicar uma possível solução com base no histórico de tentativas prévias.

Alpaydın (2020) forneceu um exemplo que consiste na transformação de voz em texto para justificar a aplicação de algoritmos de ML. Com base nos grandes volumes de dados recolhidos de dicções e sotaques de vários utilizadores, o autor sugere que estes algoritmos conseguem aprender o que o utilizador está a dizer e mapear o que foi pronunciado em palavras. Isto assenta sobretudo na capacidade do algoritmo em se adaptar às circunstâncias em que as palavras ou as frases foram ditas.

Continuando a mesma linha de pensamento, Dietterich (1997) expande nas potencialidades da aplicação de ML em diversos cenários. O autor afirma que novas técnicas de ML estão a ser aplicadas a problemas, de forma a solucioná-los, nomeadamente, no reconhecimento da fala dos seres humanos.

Como consta a figura seguinte, Figura 2-6, ML é um tópico que integra várias dimensões, sendo que parte destas são, de igual forma, partilhadas por *Data Science*:



Figura 2-6: Domínios de conhecimento constituintes de Machine Learning (Retirado de: <https://neilpatel.com/wp-content/uploads/2019/05/ilustracao-com-diagrama-sobre-machine-learning-1.jpeg> a 26 de Janeiro de 2020)

No que toca à relação com o tópico Big Data, segundo o que Al-Jarrah et al. (2015) constam, existem várias *frameworks* de computação capazes de aplicar ML, onde a combinação de técnicas, como por exemplo *Deep Learning* e *Parallel Training*, fornecem várias maneiras possíveis de processar *Big Data*.

*Deep Learning* é definido por LeCun, Bengio, & Hinton (2015) como sendo uma tecnologia que analisa grandes volumes de dados com o objetivo de descobrir uma estrutura complexa recorrendo ao algoritmo de retropropagação do erro (*backpropagation*), de forma a indicar como um computador deve alterar os seus parâmetros internos.

Transitando para *Parallel Training*, Dong, Krzyżak, & Suen (2003) referem que esta é uma tecnologia à qual se recorre para reduzir o tempo de treino despendido pelo computador. A técnica é, sobretudo, aplicada na fase de otimização do sistema.

De acordo com George, Haas, & Pentland (2014), conciliar *Big Data* com ML permite desenvolver uma ferramenta ou *framework* que não só analisa padrões nos dados, como também pode fornecer a probabilidade preditiva de um evento acontecer.

Num estudo aplicado à área da saúde, Beam & Kohane (2018) defendem que exemplos recentes demonstraram que *Big Data* e ML podem criar algoritmos com desempenho semelhante aos médicos humanos. Segundo os autores, embora ML e *Big Data* possam parecer incompatíveis no início, os investigadores acreditam que estes se relacionam muito com os modelos estatísticos tradicionais que são reconhecidos pela maioria dos médicos.

Transitando para as métricas assentes no *Supervised Learning*; este é constituído pelas categorias de classificação e regressão. Como tal, estas categorias possuem as suas próprias métricas.

Aprofundando nas métricas de classificação, um estudo da autoria de Tharwat (2018) revela que estas contêm métricas que avaliam a **accuracy**, que representa o número de amostras classificadas corretamente para as classes positivas e negativas, **precision**, associada à proximidade das medições entre os valores reais e previstos, **recall**, que representa o rácio de amostras positivas classificadas corretamente em função do número total de amostras positivas, e a **confusion matrix**, que corresponde à tabela que mostra as frequências de classificação para cada cenário do modelo em teste, atribuindo valores num intervalo entre 0 e 1.

A *confusion matrix* abrange quatro tipos de frequências distintas, sendo estas:

- **True Positive (TP)**: Esta frequência ocorre quando o modelo prevê dados corretamente de acordo com a amostra de dados reais;
- **True Negative (TN)**: Similar à frequência TP, com a diferença assente na previsão de dados na negativa;
- **False Positive (FP)**: A atribuição da frequência FP é feita quando o modelo previu incorretamente um dado positivo;
- **False Negative (FN)**: Possui funcionamento similar à frequência FN, onde a distinção ocorre na previsão incorreta de um dado negativo.

O autor consta igualmente que as métricas *accuracy*, *precision* e *recall* são calculadas consoante os valores das colunas da matriz de confusão, de forma a classificar o modelo em análise.

No caso da componente das métricas de regressão, existe um estudo elaborado por Hyndman (2006), de onde se destaca, mais concretamente, a **Mean Absolute Percentage Error**, correspondente ao acrónimo **MAPE**. A seguinte expressão (2-1) demonstra a constituição da métrica referida:

$$MAPE = \frac{\sum \frac{|A-F|}{A} \times 100}{N} \quad (2-1)$$

A métrica consiste no cálculo do erro percentual dos valores previstos comparativamente aos valores reais. Usando a expressão (2-1) como ponto de partida, o valor da métrica é obtido através do somatório ( $\Sigma$ ) das divisões entre a subtração de A (Valores Atuais) com F (Valores

Previstos) e A, multiplicando o resultado obtido por 100, finalizando com a divisão do resultado por N (Número Total de Amostras).

O autor enunciou a vantagem da utilização desta métrica, sendo esta relacionada com o facto de os erros de percentagem serem independentes da escala, significando isto, portanto, que as métricas que têm por base o cálculo destes erros são frequentemente utilizadas para comparar o desempenho da previsão de algoritmos que lidam com dados de séries temporais. Portanto, o funcionamento desta métrica é transversal a algoritmos compatíveis com métricas de classificação, desde que estes algoritmos suportem o processamento de dados *time series*.

Outro estudo feito por Goodwin & Lawton (1999) apoia a escolha da MAPE em detrimento das outras métricas baseadas na média, mais concretamente a *Mean Squared Error (MSE)* e a *Mean Absolute Error (MAE)*, por os autores defenderem que, na necessidade de se efetuar uma comparação entre os desempenhos de algoritmos de previsão em diferentes séries temporais, as medidas de precisão como a MAE e a MSE não devem ser utilizadas, devido à possibilidade de existirem grandes variações na escala das observações das séries. Os autores sugerem que, para estes cenários, deve-se recorrer a medidas livres de unidade, onde a métrica MAPE é a métrica mais utilizada para o efeito.

Tendo em conta que estas afirmações vão de encontro com os vários aspetos deste projeto, nomeadamente o facto de os conjuntos de dados em análise (e, posteriormente, a serem enriquecidos com dados de previsões) possuírem dados de recolha ao longo do tempo, e os algoritmos de previsão que estão a ser comparados focarem-se essencialmente na previsão de eventos futuros, foi decidido que a métrica MAPE seria a métrica a implementar.

## 2.7 *Machine Learning* aplicado a resíduos municipais

Em termos dos principais casos de estudo na área onde este projeto se enquadra, neste caso, os resíduos, Kannangara et al. (2018) aplicaram algoritmos de ML, mais propriamente, Redes Neurais (*Neural Networks*) e Árvores de Decisão (*Decision Tree*), de forma a prever os resíduos sólidos gerados nos municípios de Ontário, no Canadá, com base em conjuntos de dados gerados de anos anteriores pelos municípios da província.

Num projeto de alertas de enchimento dos contentores do lixo, Baby et al. (2017) conciliaram a componente de ML, de forma a obter informação sobre os hábitos de geração de lixo numa determinada região, com o intuito de prever a quantidade de lixo que seria gerada futuramente.

Outro caso de estudo onde foi aplicado ML, Torres-García et al. (2015) recorreram a algoritmos de reconhecimento, de forma a desenvolver um Separador de Resíduos Inteligente (IWS). Este sistema analisava latas de plástico e latas de alumínio, conseguindo detetar automaticamente o tipo de lata em análise.

Destacando os Resíduos Sólidos Municipais – MSW, Rostami & Baghban (2018) decidiram efetuar um estudo aprofundado sobre como os MSW podem ser utilizados como fonte de energia. Ao consultarem artigos de outros autores, descobriram que o Valor de Temperatura Elevada – HHV de cada MSW é crucial para determinar como este pode ser aproveitado.

Porém, não existia um modelo cientificamente testado e comprovado que permitia calcular o parâmetro HHV. O modelo proposto pelos autores poderá ser consultado no tópico SVM.

Tendo em conta que grande parte dos estudos elaborados em cenários similares ao deste projeto recorreram a algoritmos do tipo *Supervised Learning*, foi decidido que o foco na aplicação de algoritmos de ML ao conjunto de dados em estudo seria vocacionado para os algoritmos do tipo especificado.

A apoiar esta decisão existe um estudo feito por Caruana & Niculescu-Mizil (2006), onde os autores esclarecem que os algoritmos de previsão (do tipo *Supervised Learning*) são aplicáveis em muitos cenários, onde, na aplicação de um modelo com estas características, as diferentes métricas de desempenho são apropriadas para diferentes domínios.

Dada esta afirmação dos autores, não é de todo surpreendente que diferentes algoritmos tenham melhores resultados que outros num cenário específico, e o contrário aconteça num cenário diferente. Como tal, é necessário aprofundar conceitos tais como as métricas de classificação dos algoritmos, que permitem analisar quais os algoritmos que obtiveram melhor desempenho, assim como os próprios algoritmos associados a *Supervised Learning*.

Dos algoritmos existentes, os escolhidos recaíram sobre os que colegas de profissão, a nível internacional, utilizaram para aplicar um modelo a um determinado problema, tendo disponibilizado o estudo elaborado para a comunidade científica poder consultar em plataformas apropriadas (p. e.: *Google Scholar*, *Scitepress*).

A seleção final é constituída pelos seguintes:

- **ARIMA (*AutoRegressive Integrated Moving Average*)** - A lógica associada a este algoritmo é de, como justifica Shen et al. (2011), um grupo de dados ordenados de séries temporais que é formado ao longo do tempo ser descrito por um modelo matemático correspondente onde, seguidamente, os dados futuros são previstos de acordo com o modelo e valores anteriores e presentes dos dados de séries temporais.

Segundo o autor, o acrónimo do algoritmo corresponde aos principais aspetos do modelo em si, sendo estes a Auto Regressão – *AutoRegressive*, a Integração – *Integrated*, e a Média Variável – *Moving Average*. A Auto Regressão corresponde à componente que usa a relação dependente entre uma observação com as restantes. A Integração trata de subtrair uma observação obtida com a observação da etapa anterior, procedendo desta forma para todos os registos existentes, com o propósito de criar uma série temporal. Por fim, a Média Móvel recorre à dependência entre uma observação e um erro residual obtido através da média calculada entre as observações.

Uma notação padrão denominada  $p, d, q$  é utilizada pelo ARIMA. Esta notação corresponde a parâmetros que são substituídos por valores inteiros, de modo a calcular os valores das previsões. Os parâmetros da notação são estruturados da seguinte forma:  $p$ , onde este corresponde ao número de observações incluídas no modelo,  $d$ , onde se contabiliza o número de vezes que as observações brutas são diferenciadas, e  $q$ , responsável por calcular o valor da Média Variável.

Num estudo feito por Li et al. (2013), os autores utilizaram o algoritmo ARIMA para prever a meteorologia. Justificam a escolha deste algoritmo com a premissa de que, como quase todos os dados no campo da meteorologia são de séries temporais, as condições

meteorológicas futuras podem ser melhor previstas através de um algoritmo que considere o espaço temporal.

A seguinte figura, Figura 2-7, corresponde a um esquema representativo do funcionamento do algoritmo ARIMA:

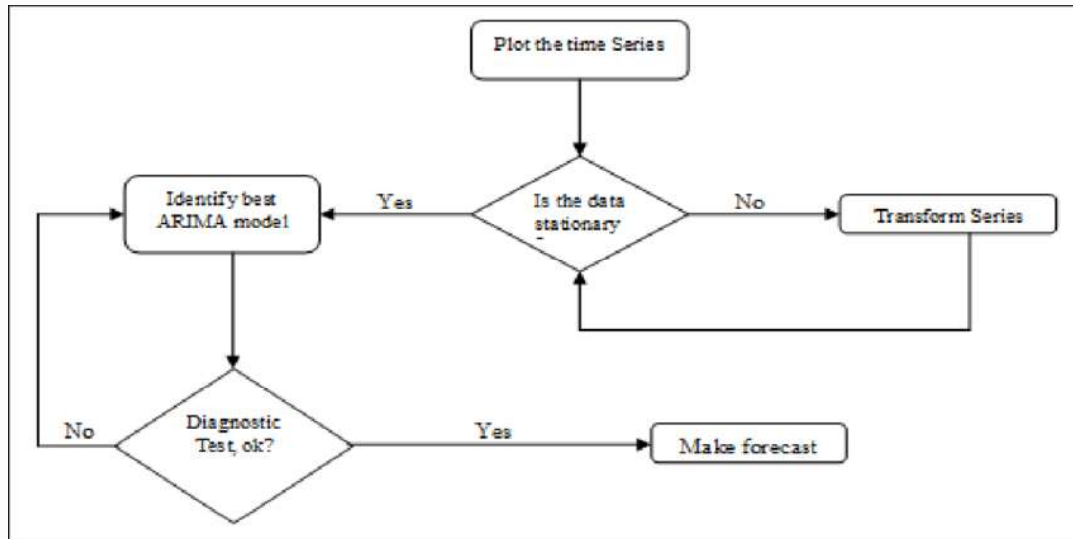


Figura 2-7: Representação do algoritmo ARIMA (Retirado de:

[https://www.researchgate.net/figure/fig-1-The-steps-of-ARIMA-model-building-methodology-is-presented-in-a-flow-chart-in\\_fig1\\_334742152.png](https://www.researchgate.net/figure/fig-1-The-steps-of-ARIMA-model-building-methodology-is-presented-in-a-flow-chart-in_fig1_334742152.png) a 15 de Março de 2020)

- **Decision Trees** – No caso deste algoritmo, sintetizado por Harrag, El-Qawasmeh, & Pichappan (2009), uma *Decision Tree* (Árvore de Decisão) é uma ‘árvore’ na qual os nós internos são categorizados por atributos, onde os ‘ramos’ que surgem destes são classificados pelo peso que um atributo tem no *dataset*, e as ‘folhas’ são rotuladas por categorias.

Para definir o ganho de informação com precisão, é utilizada uma técnica chamada entropia, que caracteriza o grau de impureza de um dado conjunto de dados. A entropia é calculada pela seguinte expressão (2-2).

$$H(S) = \sum -p(c)\log_2 p(c) \quad (2-2)$$

Onde  $S$  é o conjunto de dados onde a entropia será calculada,  $c$  o conjunto de classes em  $S$  (e. g.  $\{yes, no\}$ ) e  $p(c)$  a percentagem de número de elementos da classe  $c$  para o número de elementos no conjunto  $S$

Quando  $H(S) = 0$ , significa que os elementos do conjunto  $S$  são da mesma classe. Quando  $H(s) = 1$ , significa que 50% das classes são de um tipo (e. g. *yes*) e os restantes 50% das classes são do outro tipo (e. g. *no*).

No final, é calculado o ganho, que é a diferença entre a entropia calculada antes da separação do conjunto  $S$  e a entropia calculada depois da separação para o atributo  $A$ . A seguinte expressão (2-3) apresenta o cálculo do ganho.

$$IG(A,S) = H(S) - \sum p(t)H(t) \quad (2-3)$$

Onde  $H(S)$  é a entropia do conjunto  $S$ ,  $p(t)$  a percentagem do número de elementos em  $t$  para o número de elementos no conjunto  $S$  e  $H(t)$  a entropia do subconjunto  $t$ .

O algoritmo analisa um documento de teste testando recursivamente os pesos que os atributos dos nós internos têm no vetor de documento, até chegar a um resultado final. Os autores utilizaram este algoritmo para classificar diferentes manuscritos em linguagem árabe.

Na Figura 2-8 é possível consultar um esquemático referente ao funcionamento do algoritmo *Decision Trees*:

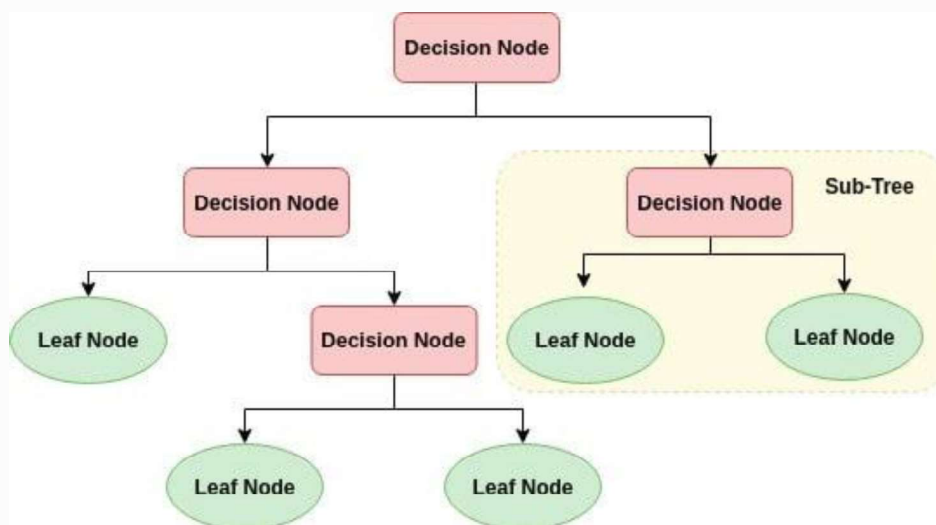


Figura 2-8: Representação do algoritmo Decision Trees (Retirado de: [https://res.cloudinary.com/dyd911kmh/image/upload/f\\_auto,q\\_auto:best/v1545934190/1\\_r5ikdb.png](https://res.cloudinary.com/dyd911kmh/image/upload/f_auto,q_auto:best/v1545934190/1_r5ikdb.png) a 15 de Março de 2020)

- **Artificial Neural Networks** – De acordo com Yegnanarayana (2009), uma *Artificial Neural Network* (Rede Neuronal Artificial) é constituída por unidades de processamento computacionais interligadas. Na generalidade, este modelo possui uma componente para *inputs*, e outra para *outputs*. A cada valor recebido no *input* é atribuído um peso, sendo posteriormente calculado o somatório do peso. Dependendo do valor total do peso, será emitido um sinal, onde este é encarregue de gerar um *output* final. A partir do momento que a rede neuronal tiver uma camada escondida, é utilizado o *Deep Learning*, que corresponde a uma técnica de utilização de uma ou mais camadas escondidas numa rede neuronal, com o objetivo de encontrar padrões mais detalhados, efetuando, assim, previsões mais acertadas.

Um dos principais desafios da utilização desta técnica centra-se no grande poder computacional que estas redes necessitam para o treino do modelo. Porém, com a natural evolução da tecnologia, a possibilidade de utilização destas técnicas aumenta, visto que a tendência no que toca aos custos do aluguer dos servidores necessários tem tendência a diminuir.

Um estudo foi feito por Xie, Ma, & Bai (2009) recorrendo a *Artificial Neural Networks*, com o intuito de prever a qualidade do ar ao nível dos edifícios de uma cidade, utilizando dados da *Occupant Symptom Metric - PIAQ* que foram obtidos em estudos anteriores.

O esquema que sintetiza o funcionamento do algoritmo encontra-se representado pela Figura 2-9:

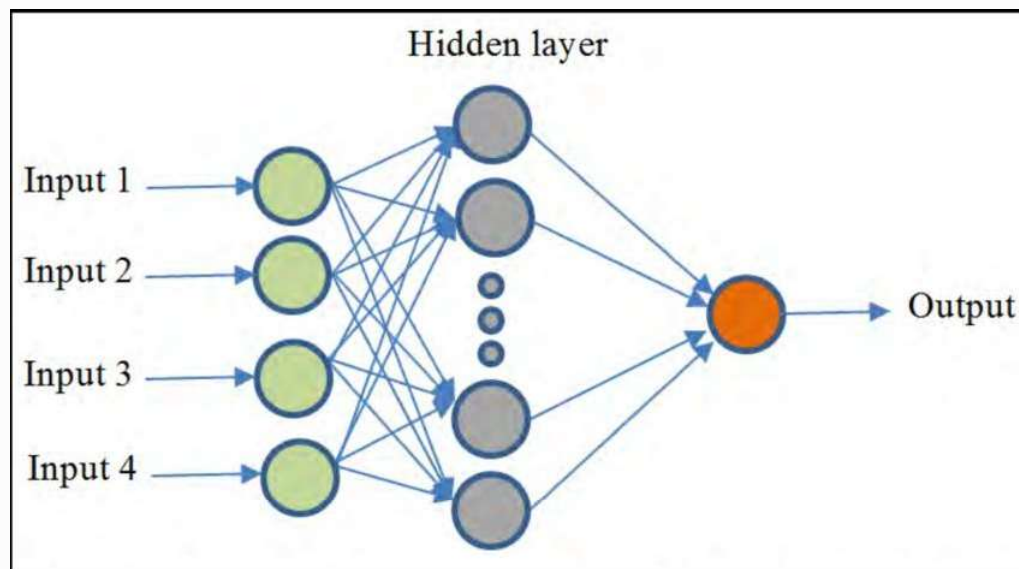


Figura 2-9: Representação do algoritmo Artificial Neural Networks (Retirado de: <https://www.researchgate.net/figure/fig9/AS:395936795774976@1471410096154/Schematic-Diagram-of-the-ANN-Structure-to-Model-Engines-with-a-Single-Output.png> a 16 de Março de 2020)

- **SVM (*Support Vectoring Model*)** – Este algoritmo, como constam Olatunji, Selamat, & Raheem (2011), recorre a funções não lineares para mapear *inputs* num espaço multidimensional, classificando os *inputs* em várias categorias. A relação que este modelo constrói entre os *inputs* e os *outputs* a serem gerados é feita posteriormente recorrendo às categorias previamente definidas, determinando o *output* de acordo com a categoria do *input*.

Os seguintes conceitos do SVM constituem as principais componentes do seu funcionamento. Estes correspondem a *Hyperplane*, Vetores de Suporte e Margem. O *Hyperplane* é um espaço de decisão que divide conjuntos de objetos pertencentes a classes diferentes. Por sua vez, os Vetores de Suporte correspondem aos pontos de dados mais próximos do *Hyperplane*, onde a linha de separação é definida pelo último de acordo com os pontos de dados dos Vetores. Por fim, a Margem é definida como a lacuna existente entre duas linhas definidas pelos pontos de dados dos Vetores, onde a preferência recai sempre pela maior Margem.

O principal objetivo do algoritmo passa por dividir o conjunto de dados em classes de modo a encontrar um *Maximum Marginal Hyperplane* – MMH. Para o efeito, o SVM irá gerar *Hyperplanes* de forma iterativa, agrupando as classes o mais eficientemente possível.

Só após a conclusão das segregações é que o algoritmo irá proceder à escolha do *Hyperplane* que separa as classes da forma matematicamente mais acertada.

Num estudo elaborado por Rostami & Baghban (2018), os autores decidiram propor um modelo baseado em Máquinas de Vetores de Suporte - SVM que permitia estimar o Valor de Temperatura Elevada - HHV dos resíduos produzidos em municípios. Para maximizarem a precisão dos seus resultados, e minimizarem o grau de incerteza, os autores recorreram a conjuntos de dados alojados num servidor dedicado ao armazenamento de informações de vários resíduos municipais, dos quais constavam valores HHV, para o efeito desejado.

A Figura 2-10 demonstra um gráfico que representa o funcionamento do SVM:

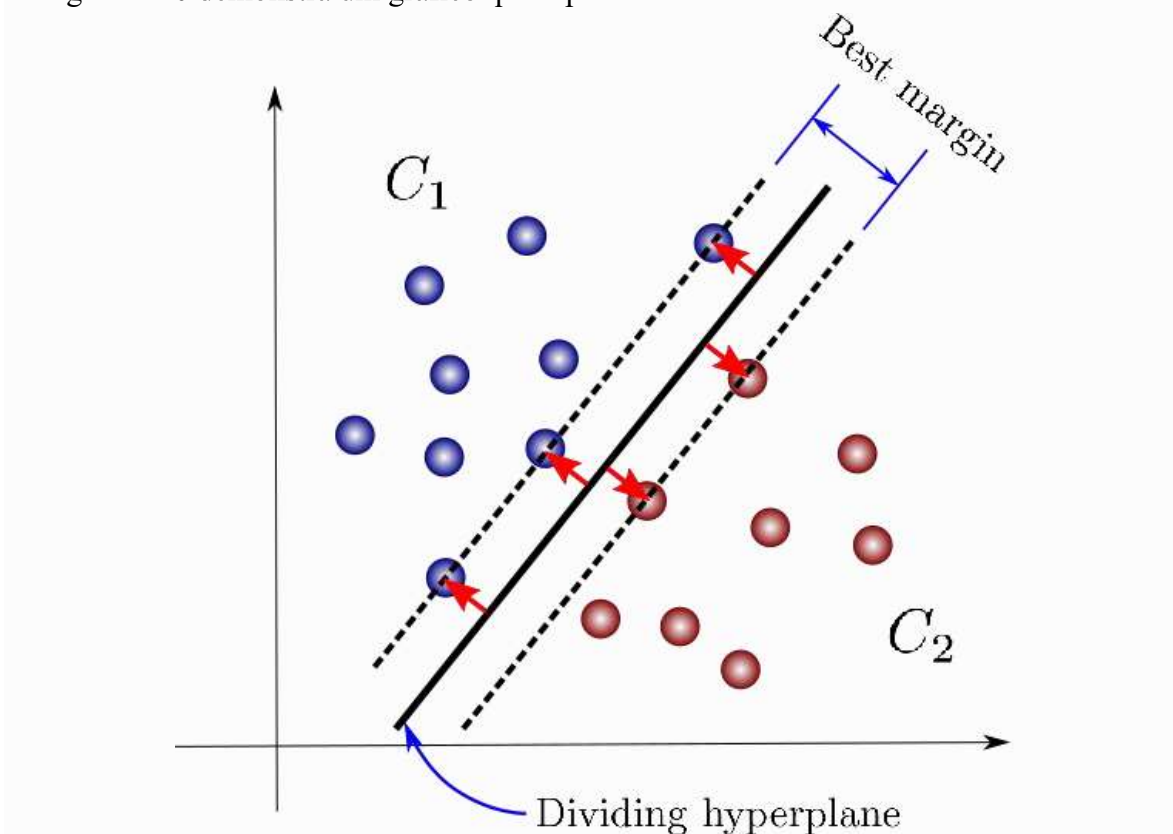


Figura 2-10: Representação do algoritmo SVM (Retirado de: [https://miro.medium.com/max/1088/1\\*6U9NrruycDBsPOyivpn8UQ.png](https://miro.medium.com/max/1088/1*6U9NrruycDBsPOyivpn8UQ.png) a 16 de Março de 2020)

- **Prophet** – Por último, refere-se um modelo que muito recentemente tem crescido em popularidade, particularmente entre programadores com conhecimentos nas linguagens Python e R.

O *Prophet* é um algoritmo desenvolvido pela equipa de *Data Scientists* por detrás do Facebook. De acordo com o disponibilizado pela fonte: “O *Prophet* é um procedimento para prever dados de séries temporais tendo por base um modelo aditivo onde tendências não lineares são ajustadas de acordo com a sazonalidade anual, semanal e diária...” (Fonte: <https://facebook.github.io/prophet/>).

O algoritmo é baseado numa plataforma de modelação e computação estatística denominada *Stan*. O *Stan* recorre a funções de cálculo de probabilidade e a matrizes de álgebra linear para determinar os registos futuros. Neste, os utilizadores especificam funções de densidade dos *logs* suportados pela linguagem de programação R.

Destaca-se o estudo feito por Papacharalampous & Tyrallis (2018), onde concluíram que o modelo constituinte do *Prophet* deve ser utilizado para calcular previsões a longo termo (de acordo com o estudo dos autores, o espaço temporal corresponde a mais de 30 dias).

Por ser uma tecnologia recente, com um número de estudos científicos a nível internacional relativamente baixo, decidiu-se que, devido ao conjunto de dados do projeto possuir dados desde Agosto de 2004, o *Prophet* seria incluído no comparativo de algoritmos *Machine Learning*.

De acordo com a figura em baixo disponibilizada, Figura 2-11, é possível consultar o método de funcionamento, num nível sintetizado, do algoritmo de *Machine Learning Prophet*:

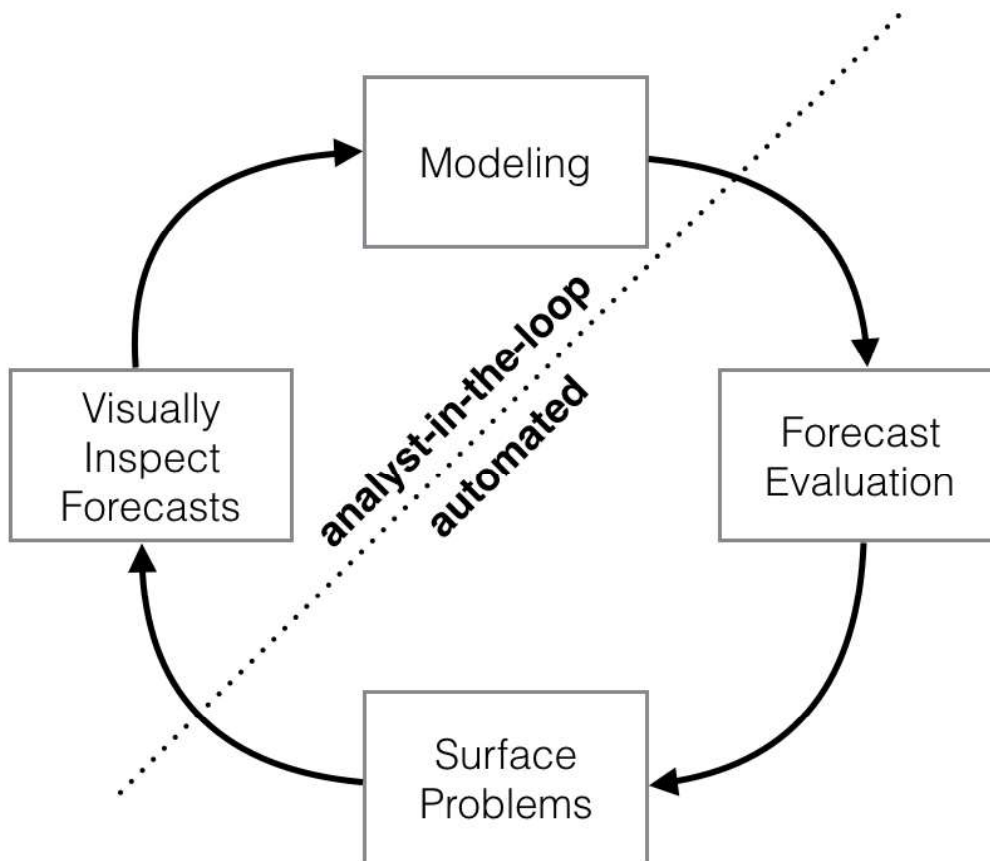


Figura 2-11: Representação do algoritmo Prophet (Retirado de: <https://research.fb.com/wp-content/uploads/2017/02/image-2.png?w=884> a 17 de Março de 2020)

### 3. Metodologia

Na elaboração do projeto, a metodologia de Investigação e Desenvolvimento - I&D - que servirá de base para este será a *Engineering Design Process* (EDP). Esta metodologia é composta por uma série de etapas que os engenheiros na área de investigação seguem para encontrar uma solução para um determinado problema. Na Figura 3-1 podem ser encontradas em termos gráficos as etapas constituintes da EDP:

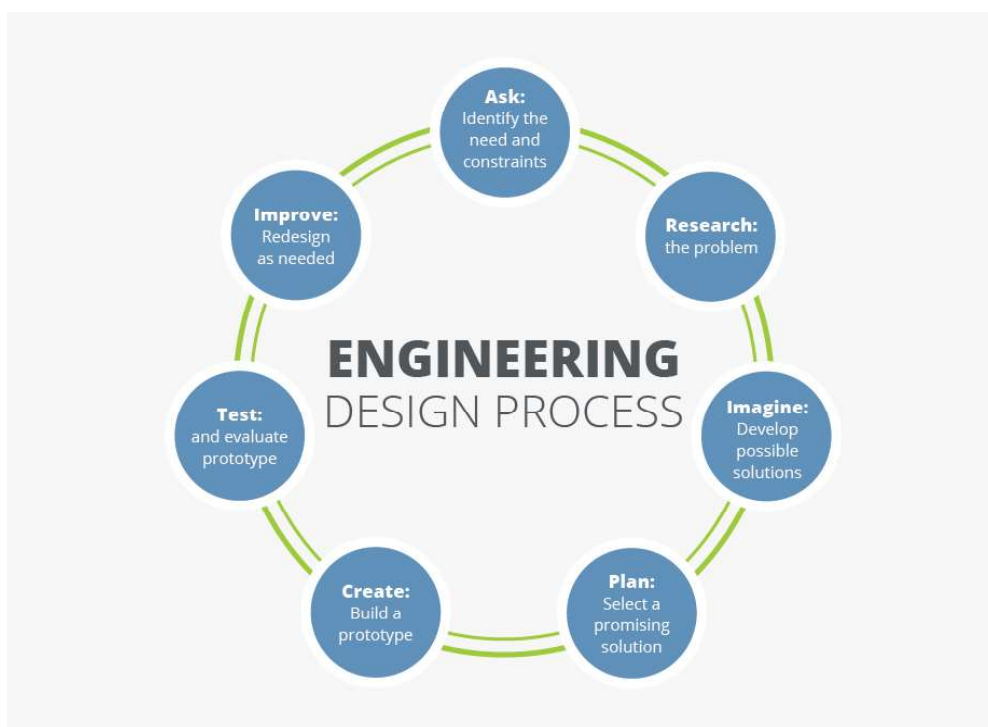


Figura 3-1: Representação gráfica da metodologia EDP (Retirado de: [https://www.netclipart.com/pp/m/82-821491\\_the-engineering-design-process-engineering-design-process.png](https://www.netclipart.com/pp/m/82-821491_the-engineering-design-process-engineering-design-process.png) a 19 de Dezembro de 2019)

A primeira etapa consiste em identificar o problema, onde é estabelecido o desafio presente, assim como as limitações existentes. De seguida, seguem-se as etapas de investigação sobre o problema, onde se efetua um trabalho de pesquisa sobre situações semelhantes, e de desenvolvimento de algumas soluções, que podem surgir mediante a pesquisa previamente efetuada. Segue-se depois a fase do planeamento, onde se traça um plano de acordo com as soluções arrecadadas. A quinta etapa corresponde à criação, onde se segue o plano elaborado e se constrói um protótipo de acordo com tal. Após a implementação do protótipo, surge a sexta etapa, que consiste no teste e avaliação do protótipo, verificando se esta solução resolveu de forma satisfatória o problema. Em caso negativo, surge a sétima etapa do EDP, que corresponde à fase de melhoramento. Esta etapa passa por verificar que aspetos estão menos bem tratados, ou até subestimados, e melhorar o protótipo consoante as novas necessidades identificadas.

De acordo com a metodologia de I&D especificada acima, a metodologia de trabalho adotada segue os seguintes passos: primeiramente, procedeu-se à revisão da literatura, de modo a melhorar o conhecimento na área em estudo, rever o progresso feito até ao momento e verificar novos estudos que se possam relacionar com o projeto; de seguida, passou-se para a recolha dos conjuntos de dados que se enquadram na vertente dos resíduos, junto da Altice Labs e parceiros associados; o próximo passo prendeu-se com a exploração e estudo dos conjuntos recolhidos a nível dos tipos de dados presentes e valores registados; decorrido o anterior, procedeu-se à preparação dos dados recorrendo a algoritmos de *Data Science*; por fim, chega a implementação de *Data Mining*, onde os dados foram modelados e processados criando, consequentemente, os Modelos de Informação Comum.

A nível das operações de *Data Mining* previamente descritas, a metodologia adotada foi a *Cross Industry Standard Process for Data Mining*, também conhecida pela sigla CRISP-DM. Esta metodologia é principalmente utilizada em cenários onde se pretende extrair valor dos dados arrecadados, seja este adquirido através da descoberta de padrões desconhecidos, ou pelo tratamento de dados não-estruturados.

Segundo Wirth & Hipp (2000) esta metodologia revela-se útil não só para a aplicação de técnicas de *Data Science*, como e. g. *Data Mining* e *Machine Learning*, mas, de igual forma, para o planeamento, documentação e comunicação dos processos utilizados.

As tarefas integrantes da CRISP-DM podem ser executadas numa ordem não necessariamente linear, porém, não adotar a sequência padrão poderá, consequentemente, fazer com que seja necessário visitar com mais frequência tarefas realizadas anteriormente, podendo mesmo, no pior cenário, repetir determinadas ações. A figura seguinte, Figura 3-2, representa graficamente as tarefas desencadeadas pela norma CRISP-DM:

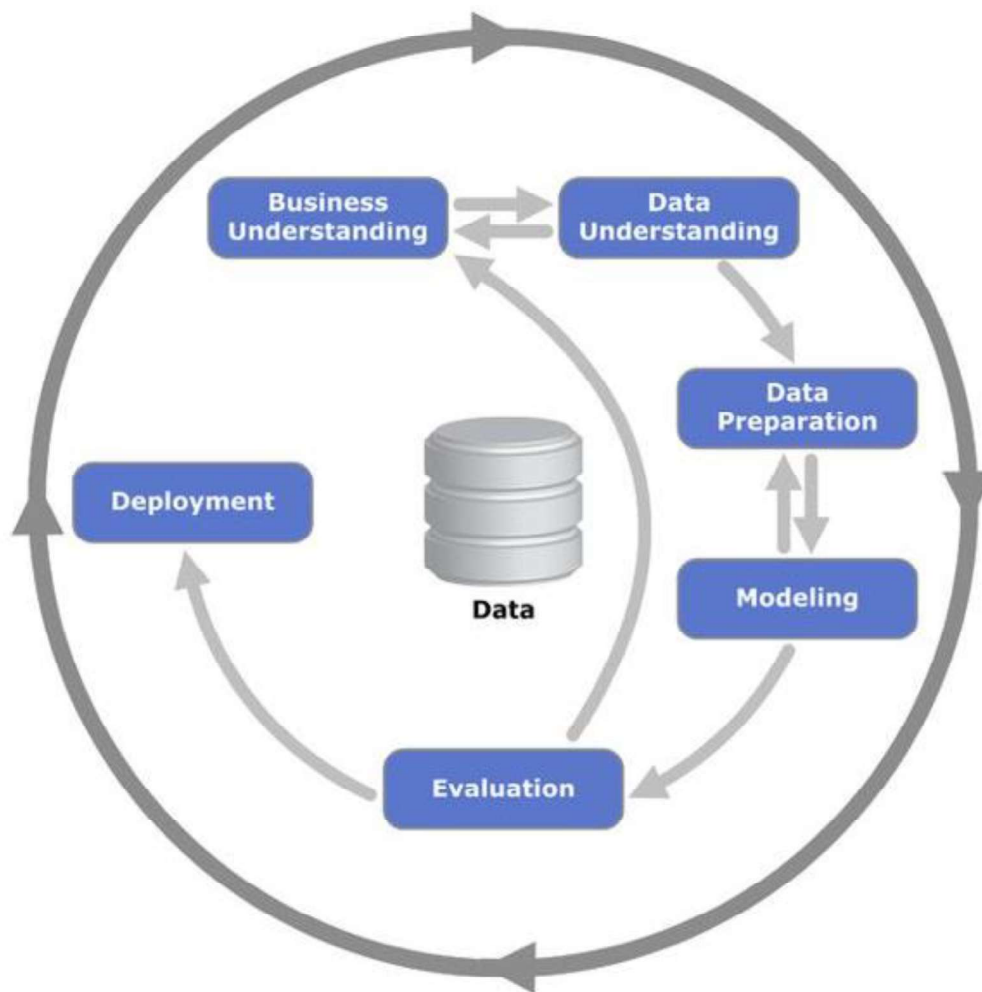


Figura 3-2: Representação gráfica da metodologia CRISP-DM (Retirado de: <http://www.datascience-pm.com/wp-content/uploads/2018/09/crisp-dm-wikicommons.jpg> a 1 de Dezembro de 2019)

A CRISP-DM é constituída por seis principais elementos, que, no seu cerne, podem ser consideradas como etapas de um projeto, desde a sua constituição, até à sua finalização:

1. **Business Understanding** (Entender o negócio) – Em primeiro lugar, é necessário fazer uma análise aprofundada dos objetivos e necessidades do negócio. A organização pode ter objetivos e restrições concorrentes que devem ser adequadamente manuseados. O foco desta etapa passa por descobrir fatores importantes que podem influenciar o resultado do projeto.
2. **Data Understanding** (Entender os dados) - A segunda fase do CRISP-DM requer que os dados sejam devidamente documentados nos recursos do projeto. Isto significa que todas as fontes de dados a utilizar devem ser registadas e analisadas, de modo a saber-se como e quando estas serão integradas no projeto.
3. **Data Preparation** (Preparar os dados) – Nesta etapa do projeto é crucial elaborar uma tomada de decisão acerca dos dados/fontes que efetivamente serão usados para análise. Critérios como a relevância e qualidade dos dados, assim como certas restrições técnicas a nível de limites nos volumes ou tipos de dados irão servir de base à escolha. Após o

estabelecimento concreto das fontes, serão aplicados aos dados técnicas como a seleção, limpeza, construção e formatação adequadas.

4. **Modelling** (Modelação dos dados) - A seleção de uma técnica de modelação, como por exemplo a *decision-tree*, a *random forest*, ou a *neural network*, corresponde ao passo seguinte, juntamente com a criação do cenário de teste para validar a qualidade do modelo.
5. **Evaluation** (Avaliação) – Chegando a esta fase, os resultados dos modelos criados serão alvos de uma avaliação no contexto das intenções do negócio. Dependendo da avaliação, novos objetivos a serem concretizados podem surgir devido aos novos padrões descobertos e/ou devido ao valor retirado dos dados. Este elemento é um processo iterativo, onde a decisão de considerar os resultados como válidos ou não deve ser tomada nesta etapa antes de se transitar para a fase final.
6. **Deployment** (Produção) – Por fim, nesta etapa, as informações finais aglomeradas devem ser expostas de forma perceptível aos *stakeholders*. Terá de se gerar um consenso acerca da estratégia a ser implementada, de acordo com os Modelos de Informação Comum apresentados. Se os modelos forem efetivamente relevantes, estes serão documentados para posteriormente entrarem em fase de produção.

Em síntese, de acordo com o levantamento do Estado da Arte no que toca a *Data Science*, fazendo um paralelismo com a estrutura da metodologia apresentada, é possível perceber que a adoção da metodologia CRISP-DM foi uma escolha viável para a implementação deste projeto de mestrado.

Analisando detalhadamente o que consiste *Data Science*, consegue-se perceber que os projetos na área começam com a análise dos negócios. Após esta etapa, estes projetos focam-se de seguida nos dados que devem ser reunidos, explorados e preparados. Sabendo que, após a extração de conhecimento dos dados, estes projetos precisam de ser avaliados pelo valor que possuem para a empresa no mundo real, é inegável que a metodologia CRISP-DM fornece uma forte orientação para as atividades a desempenhar por um cientista de dados.

## 4. Cenário do Projeto

Neste capítulo apresenta-se uma síntese do cenário em estudo e descreve-se os tipos de resíduos em análise.

### 4.1 Síntese Geral

Enquadrado na temática dos resíduos, o cenário que servirá de análise para o projeto será a recolha dos vários tipos de lixo de uma cidade. Neste caso, a cidade à qual os dados do *dataset* se referem prende-se com a cidade de Austin, Texas.

O conjunto de dados utilizado, Figura 4-1, contém registos diários do lixo recolhido pela entidade encarregada do processo, de nome *Austin Resource Recovery*, sendo estes datados desde Agosto de 2004 até à atualidade (o *dataset* é alvo de uma constante atualização diária). Para o mesmo dia, existem tipicamente vários registos, devido ao tipo de resíduos que é processado. Estes tipos de resíduos poderão ser colheitas de contentores (*Garbage Collections*), lixo orgânico (*Organics*), lixo na estrada (*Sweeping*), entre outros.

Na totalidade, o conjunto de dados é composto por mais de 670 mil registos, onde cada registo é constituído por oito colunas:

- **Report Date:** O dia em que a informação da recolha do lixo foi registada;
- **Load Type:** O tipo de carga específica que foi recolhida nesse dia;
- **Load Time:** A data e hora em que os resíduos foram recolhidos;
- **Load Weight:** O peso que foi coletado pela entidade apropriada;
- **Drop off Site:** A localização onde os resíduos foram entregues para disposição ou reciclagem;
- **Route Type:** A categoria de coleção à qual os resíduos processados pertencem;
- **Route Number:** O identificador do *Route Type* registado;
- **Load Id:** O identificador do registo da recolha efetuada.

## 4 – Cenário do Projeto

REPORT_DA ↓	LOAD_TYPE	LOAD_TIME	LOAD_WEIGHT	DROPOFF_SITE	ROUTE_TYPE	ROUTE_NUM...	LOAD_ID
03/25/2020	ORGANICS	2020 Mar 25 04:1...	11,760	ORGANICS BY GO...	YARD TRIMMING...	OW11	866618
03/25/2020	ORGANICS	2020 Mar 25 11:2...	12,280	ORGANICS BY GO...	YARD TRIMMING...	OW13	866623
03/25/2020	MIXED LITTER	2020 Mar 25 05:2...	3,500	TDS LANDFILL	KAB	KAB04	866597
03/25/2020	GARBAGE COLLE...	2020 Mar 25 02:0...	10,740	TDS LANDFILL	GARBAGE COLLE...	PAW79	866627
03/25/2020	ORGANICS	2020 Mar 25 11:4...	14,700	ORGANICS BY GO...	YARD TRIMMING...	OW11	866617
03/25/2020	YARD TRIMMING	2020 Mar 25 02:2...	14,200	HORNSBY BEND	YARD TRIMMINGS	YW03	866642
03/25/2020	GARBAGE COLLE...	2020 Mar 25 07:3...	24,940	TDS LANDFILL	GARBAGE COLLE...	PAW74	866626
03/25/2020	GARBAGE COLLE...	2020 Mar 25 01:0...	24,920	TDS LANDFILL	GARBAGE COLLE...	PAW73	866621
03/25/2020	YARD TRIMMING	2020 Mar 25 09:5...	14,440	HORNSBY BEND	YARD TRIMMINGS	YW04	866630
03/25/2020	SWEEPING	2020 Mar 25 12:0...		ONION CREEK	STREET CLEANIN...	SB63	866640
03/25/2020	YARD TRIMMING	2020 Mar 25 04:1...	15,580	HORNSBY BEND	YARD TRIMMINGS	YW02	866639
03/25/2020	GARBAGE COLLE...	2020 Mar 25 11:0...	15,460	TDS LANDFILL	GARBAGE COLLE...	PAW74	866625
03/25/2020	SWEEPING	2020 Mar 25 12:0...		ONION CREEK	STREET CLEANIN...	SB51	866637
03/25/2020	YARD TRIMMING	2020 Mar 25 04:1...	19,320	HORNSBY BEND	YARD TRIMMINGS	YW04	866631

< Previous   Next >      Showing Rows 1-14 out of 679,225

Figura 4-1: Conjunto de dados do lixo recolhido em Austin, Texas (Retirado de: <https://data.austintexas.gov/Utilities-and-City-Services/Waste-Collection-Diversion-Report-daily-/mbnu-4wq9> a 25 de Março de 2020)

Com este *dataset* enriquecedor, será possível aplicar algoritmos de *Machine Learning* (após o devido tratamento prévio dos dados), de modo a obter previsão sobre os registos de recolha dos próximos meses.

O objetivo com esta ação passa por demonstrar ao município e à entidade de recolha os potenciais resíduos que terão de ser recolhidos, onde, com base na informação providenciada, as entidades possam tomar decisões tais como perceber se compensa investir em mais pontos de recolha para o resíduo X.

Aliando estes dados à componente de *Business Intelligence* (BI), assente na plataforma de gestão de dados da empresa Altice Labs, representada pela Figura 4-2, será possível implementar *dashboards* informativos de modo a facilitar a perceção do significado dos dados, significando isto que será possível retirar valor dos mesmos, com o intuito de auxiliar o município, a entidade de recolha e outros potenciais clientes nas futuras decisões que estes tenham de tomar.

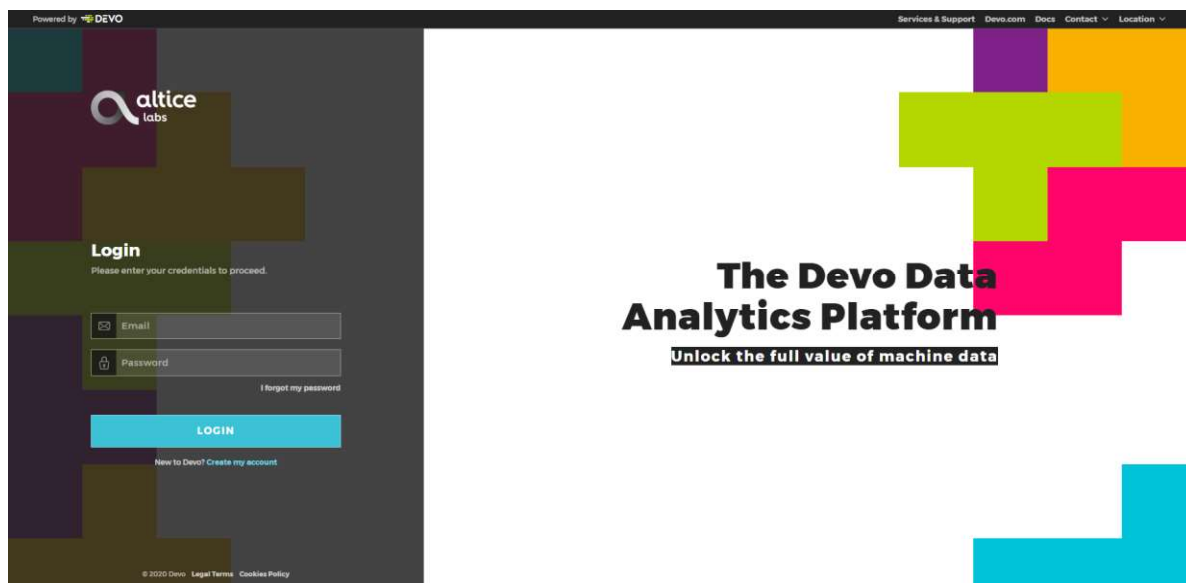


Figura 4-2: Página Principal da Plataforma de Gestão de Dados da AltiCE Labs (Retirado de: <https://alticelabs.devo.com/login> a 5 de Março de 2020)

## 4.2 Tipos de Resíduos

No conjunto de dados escolhido é possível encontrar tipos distintos de resíduos processados. Os seguintes podem ser consultados no Quadro 4-1:

Quadro 4-1: Descrição dos tipos de resíduos constituintes do dataset

Nome do Tipo de Resíduo	Descrição
<b>GARBAGE COLLECTIONS</b>	Corresponde à recolha do lixo presente nos contentores localizados em zonas de habitação.
<b>RECYCLING - SINGLE STREAM</b>	Este tipo de resíduos refere-se a um sistema no qual um tipo de camião específico recolhe todas as fibras de papel, plásticos e metais de outros contentores, ao invés de serem recolhidos de forma separada (a separação é depois realizada no centro de recolha).
<b>SWEEPING</b>	O seguinte tipo de lixo corresponde ao varrimento das ruas do território abrangido pela cidade.
<b>YARD TRIMMING</b>	Este tipo de lixo abrange o processo de jardinagem feita em jardins considerados de carácter público.

<b>BULK</b>	Representa o aglomerado de produtos agrícolas e químicos que foram coletados pela entidade.
<b>BRUSH</b>	Corresponde à escovagem e recolha de, essencialmente, folhas caídas das árvores da cidade.
<b>RECYCLING - PAPER</b>	Este tipo de resíduos refere-se a um sistema no qual um tipo de camião específico recolhe única e especificamente todas as fibras de papel presentes nos contentores apropriados.
<b>RECYCLING - COMINGLE</b>	Este tipo de resíduos refere-se a um sistema no qual um tipo de camião específico recolhe única e especificamente todos os resíduos derivados de produtos alimentares presentes nos contentores apropriados.
<b>DEAD ANIMAL</b>	O seguinte tipo de lixo representa o processamento de animais defuntos nas estradas.
<b>ORGANICS</b>	Corresponde ao processamento de lixo orgânico.
<b>TIRES</b>	Esta categoria pertence aos pneus de qualquer género de veículo que são considerados inutilizáveis.
<b>MIXED LITTER</b>	Significa lixo misturado, relaciona-se com o tipo <b>RECYCLING - SINGLE STREAM</b> .
<b>LITTER</b>	Corresponde a lixo (não foi possível encontrar uma especificação mais concreta para este tipo).
<b>MULCH</b>	Esta categoria representa o processamento de adubo.
<b>RECYCLED METAL</b>	Este tipo de resíduos refere-se a um sistema no qual um tipo de camião específico recolhe única e especificamente todos os materiais metálicos presentes nos contentores apropriados.
<b>BAGGED LITTER</b>	Corresponde a lixo ensacado (não foi possível encontrar uma especificação mais concreta para este tipo).
<b>RECYCLING - PLASTIC BAGS</b>	Este tipo de resíduos refere-se a um sistema no qual um tipo de camião específico recolhe única e especificamente todos os sacos de plástico presentes nos contentores apropriados.

<b>YARD TRIMMING - CHRISTMAS TREES</b>	Este tipo de lixo abrange o processo de jardinagem feita na época natalícia em jardins considerados de carácter público, relaciona-se com o tipo <b>YARD TRIMMING</b> .
<b>MATTRESS</b>	Esta categoria representa o processamento de colchões abandonados que se encontram junto a contentores de recolha.
<b>CONTAMINATED RECYCLING</b>	Corresponde ao processamento de lixo contaminado.
<b>CONTAMINATED YARD TRIMMING</b>	Corresponde ao processamento de lixo de jardim contaminado.

De modo a que o cenário em estudo se enquadre o mais possível com a realidade portuguesa, deverão ser analisados apenas os tipos de resíduos cujo processamento seja registado em Portugal.

Através da consulta de fontes de informação fidedignas, as quais estão representadas pelas Figuras 4-3 e 4-4, é possível obter-se uma perceção de quais os resíduos que são registados, assim como as respetivas quantidades:

t (toneiada) [ver tabela completa](#)

Territórios	Tipo de recolha					
	Total		Recolha indiferenciada		Recolha selectiva	
Anos <span>+</span>	2002	2018	2002	2018	2002	2018
Portugal	4.595.146	5.221.646	4.368.770	4.157.752	212.665	1.063.894
— Continente	4.358.399	4.944.673	4.158.853	3.950.821	199.546	993.852
+ Norte	1.431.512	1.644.625	1.360.146	1.384.433	71.365	260.192
+ Centro	891.040	996.139	862.955	853.495	28.085	142.644
+ Área Metropolitana de Lisboa	1.370.787	1.516.864	1.288.022	1.112.881	82.765	403.984
+ Alentejo	382.690	381.770	375.067	320.138	7.622	61.632
+ Algarve	282.371	405.275	272.663	279.875	9.708	125.400
— Região Autónoma dos Açores	95.208	150.525	81.497	114.894	0	35.631
+ Região Autónoma dos Açores	95.208	150.525	81.497	114.894	0	35.631
— Região Autónoma da Madeira	141.539	126.448	128.420	92.037	13.119	34.411
+ Região Autónoma da Madeira	141.539	126.448	128.420	92.037	13.119	34.411

Figura 4-3: Resíduos urbanos em Portugal: total e por tipo de recolha (Retirado de: <https://www.pordata.pt/Municipios/Res%3%adduos+urbanos+total+e+por+tipo+de+recolha-65-241> a 3 de Abril de 2020)

Consultando o portal oficial da PORDATA (Base de dados de Portugal Contemporânea), consegue-se chegar à conclusão de que, para além da quantidade astronómica de lixo que é produzido em Portugal, tanto para continente como para os arquipélagos, a recolha indiferenciada é claramente superior à recolha seletiva (diferenciada), apesar de as toneladas de resíduos processados por métodos seletivos ter, no mínimo, quintuplicado em algumas zonas (exemplo: região Centro e Alentejo).

Fazendo comparações diretas com os tipos de resíduos existentes no *dataset*, a **recolha indiferenciada** pode ser incorporada pelos tipos: “GARBAGE COLLECTIONS”, “RECYCLING - SINGLE STREAM”, “SWEEPING”, “BULK” e “MIXED LITTER”.

No caso da **recolha seletiva**, os seguintes tipos serão os mais adequados: “YARD TRIMMING”, “RECYCLING - PAPER”, “RECYCLING - COMINGLE”, “ORGANICS”, “RECYCLED METAL” e “RECYCLING - PLASTIC BAGS”.

Parameter	Mainland	Azores	Madeira	Total
<b>Population</b>				
Total (inhabitants)	9,947,599	247,495	262,202	10,457,295
<b>Waste generation</b>				
Total (thousand tonnes)	4,363	128	108	4,598
Total (kg/capita/year)	438.6	514.3	410.6	439.7
<b>Waste composition (%)</b>				
Organics	39	36	NA	NA
Paper and cardboard	13	14	NA	NA
Plastic	10	12	NA	NA
Glass	6	10	NA	NA
Metals	2	3	NA	NA
Wood	1	1	NA	NA
Other	29	24	NA	NA
<b>Type of MSW collection (% by weight)</b>				
Separate collection	13	12	14	13
Undifferentiated collection	87	88	86	87
<b>Separate collection (% by weight)</b>				
Biodegradable municipal waste (BMW)	16.6	0	0	15.9
Paper and cardboard	31.5	37	37	31.7
Packaging (made of plastic, metals, and cardboard)	17.9	29	29	18.4
Glass	34.1	33.9	33.9	34
Batteries	0.01	0.03	0.06	0.01
<b>Waste management (% of MSW)</b>				
Recycling (multi-material)	13	12.7	8.3	12.9
Recycling (organic)	13.4	5.4	0	12.9

Figura 4-4: Dados da gestão de resíduos em Portugal no ano de 2013 (Retirado de: <https://ec.europa.eu/environment/waste/plans/index.htm> a 3 de Abril de 2020)

Na elaboração de uma pesquisa sobre o setor da gestão de resíduos em Portugal na *webpage* da comissão europeia, é possível obter-se um ficheiro em formato PDF com informações sobre os resíduos processados, de forma detalhada. Focando na secção dos parâmetros em análise (referência à secção **Parameter** da tabela), consegue-se reconhecer tipos de resíduos que vão de encontro aos que estão presentes no conjunto de dados em análise, sendo estes:

- “ORGANICS” (Organics);
- “RECYCLING - PAPER” (Paper and cardboard);
- “RECYCLING - PLASTIC BAGS” (Plastic);
- “RECYCLED METAL” (Metal);
- “GARBAGE COLLECTIONS” (Undifferentiated collection);
- “RECYCLING - COMINGLE” (Separate collection);
- “RECYCLING - SINGLE STREAM” (Undifferentiated collection);
- “MIXED LITTER” (Undifferentiated collection);
- “BRUSH” (Biodegradable municipal waste);
- “SWEEPING” (Biodegradable municipal waste).

Em conclusão, tendo por base os resíduos presentes nas *factsheets* das entidades consultadas, os tipos de resíduos que serão analisados, pré-processados, aplicados em *Machine Learning* e utilizados como objeto de consulta da plataforma de gestão de dados (mencionada na secção anterior) serão os seguintes (Quadro 4-2):

Quadro 4-2: Descrição dos tipos de resíduos do dataset a serem analisados

Nome do Tipo de Resíduo	Descrição
<b>GARBAGE COLLECTIONS</b>	Corresponde à recolha do lixo presente nos contentores localizados em zonas de habitação.
<b>RECYCLING - SINGLE STREAM</b>	Este tipo de resíduos refere-se a um sistema no qual um tipo de camião específico recolhe todas as fibras de papel, plásticos e metais de outros contentores, ao invés de serem recolhidos de forma separada (a separação é depois realizada no centro de recolha).
<b>SWEEPING</b>	O seguinte tipo de lixo corresponde ao varrimento das ruas do território abrangido pela cidade.
<b>YARD TRIMMING</b>	Este tipo de lixo abrange o processo de jardinagem feita em jardins considerados de carácter público.

<b>BULK</b>	Representa o aglomerado de produtos agrícolas e químicos que foram coletados pela entidade.
<b>BRUSH</b>	Corresponde à escovagem e recolha de, essencialmente, folhas caídas das árvores da cidade.
<b>RECYCLING - PAPER</b>	Este tipo de resíduos refere-se a um sistema no qual um tipo de camião específico recolhe única e especificamente todas as fibras de papel presentes nos contentores apropriados.
<b>RECYCLING - COMINGLE</b>	Este tipo de resíduos refere-se a um sistema no qual um tipo de camião específico recolhe única e especificamente todos os resíduos derivados de produtos alimentares presentes nos contentores apropriados.
<b>ORGANICS</b>	Corresponde ao processamento de lixo orgânico.
<b>MIXED LITTER</b>	Significa lixo misturado, relaciona-se com o tipo <b>RECYCLING - SINGLE STREAM</b> .
<b>RECYCLED METAL</b>	Este tipo de resíduos refere-se a um sistema no qual um tipo de camião específico recolhe única e especificamente todos os materiais metálicos presentes nos contentores apropriados.
<b>RECYCLING - PLASTIC BAGS</b>	Este tipo de resíduos refere-se a um sistema no qual um tipo de camião específico recolhe única e especificamente todos os sacos de plástico presentes nos contentores apropriados.

## 5. Arquitetura de API para processamento do conjunto de dados

No capítulo anterior foram analisados os 21 diferentes tipos de resíduos que constituem o conjunto de dados. Da totalidade existente, foram escolhidos os 12 tipos que, com devida fundamentação, se enquadram com a realidade em Portugal.

Tendo em conta que este *dataset* incorpora tipos de resíduos que não são relevantes para o projeto, viu-se a necessidade de implementar uma aplicação que permitisse separar os diferentes dados de resíduos, juntando estes num respectivo ficheiro para cada tipo de resíduo existente. Esta solução acaba por juntar o útil ao agradável, no sentido em que, para além de fazer a limpeza dos dados existentes, acaba por diminuir o *strain* de processamento a ser feito pelos algoritmos de *Machine Learning*, que está entregue aos recursos da máquina.

Como, ao separar-se os dados de diferentes tipos em ficheiros, se estão a originar vários ficheiros de menor tamanho comparativamente a um único ficheiro de tamanho elevado (por possuir centenas de milhares de registos), acaba-se por otimizar o processamento que a máquina terá de fazer para os 12 tipos de resíduos em análise, seja pelos tempos de processamento no total, ou pelos recursos que a máquina terá de alocar para efetuar a operação.

A *framework* escolhida para criar a aplicação do sistema foi o *Flask*. Esta ferramenta é baseada na linguagem de programação *Python*. A escolha é justificada com a familiaridade que o autor tem com a tecnologia, devido ao mesmo ter a perceção de que o *Flask* possui bibliotecas que lhe permitiam implementar o pretendido. A tecnologia tem também um carácter minimalista e de muito poucas restrições, o que significa que qualquer programador pode implementar exatamente o que precisa recorrendo a bibliotecas externas. Estes fatores fazem do *Flask* uma ferramenta flexível e extensível.

De modo a comprovar as afirmações feitas no parágrafo anterior, temos um estudo feito por Aslam et al. (2015), onde os autores concluíram que, devido à *web* ser o auxílio mais utilizado no que toca à procura de suporte para resolução de problemas, a *framework Flask* é uma

solução viável para os requisitos de qualquer tipo de utilizador. Os autores consideram o *Flask* como sendo uma *framework* maleável o suficiente para fornecer uma solução para qualquer tipo de problema.

A figura seguinte, Figura 5-1, representa um esquema da arquitetura do sistema desenvolvido com o auxílio da aplicação *Flask*:

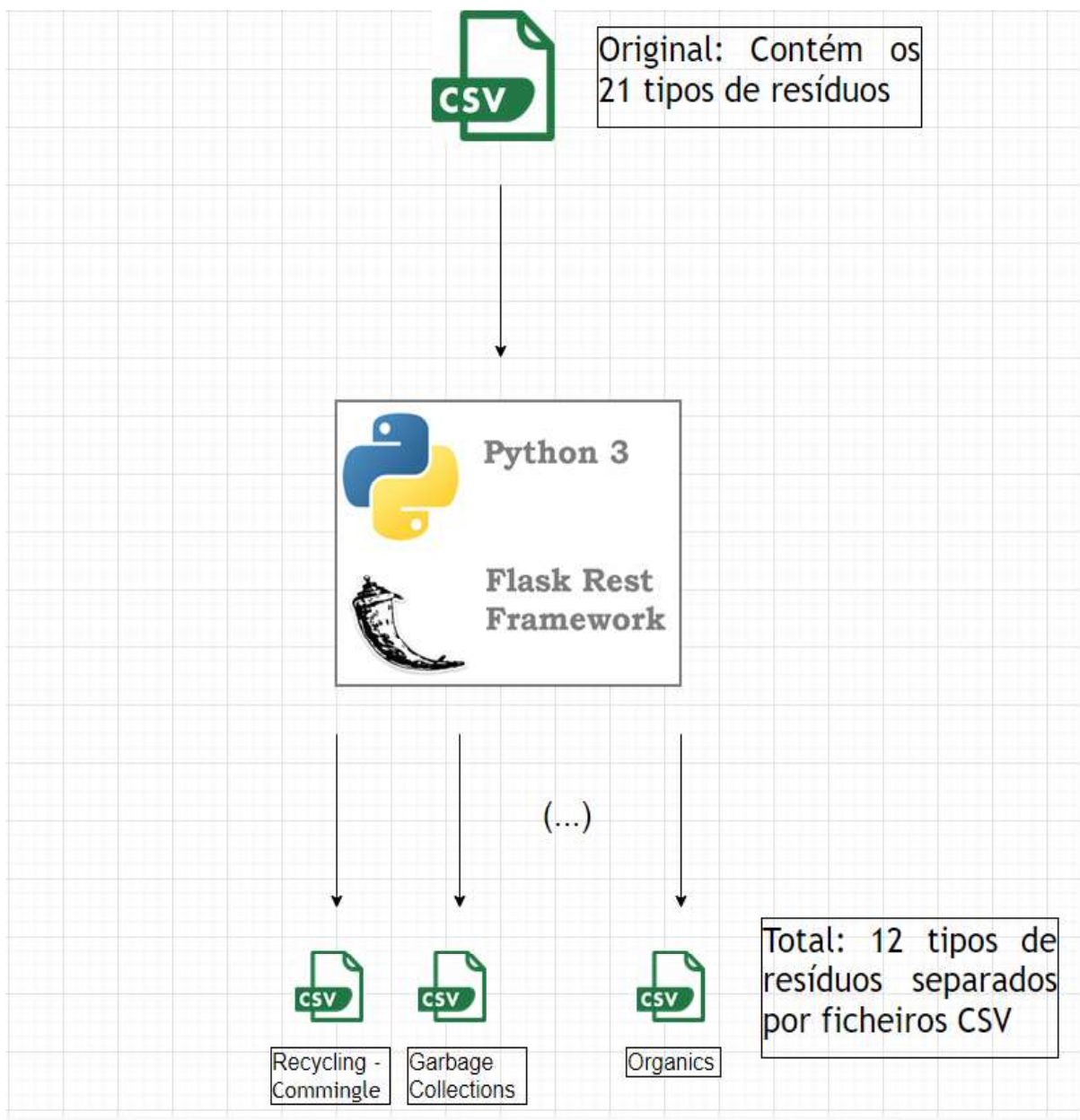


Figura 5-1: Arquitetura da aplicação Flask desenvolvida (Esquema elaborado em: <https://app.diagrams.net/>)

O ficheiro em formato CSV inicial, após ser processado pela aplicação em Flask, dá origem a 12 ficheiros no mesmo formato, onde cada um destes contém os dados do tipo de resíduo correspondente ao nome do ficheiro. O processamento dos dados é efetuado com o auxílio de

bibliotecas utilizadas em *Data Science*, sendo uma destas o *'Pandas'*, uma biblioteca particularmente popular na comunidade *Python*.

A figura abaixo disponibilizada, Figura 5-2, representa um exemplo de uma parte do código desenvolvido, necessário para o funcionamento da aplicação *Flask* projetada:

```

from flask import Flask
import pandas as pd
#import io, csv
import itertools

from datetime import *

app = Flask(__name__)

#####
#-----GARBAGE COLLECTIONS-----#
#####
@app.route('/garbage_collections')
def process_garbage_collections():
    start = datetime.now()
    data = pd.read_csv('Waste_Collection__Diversion_Report__daily__csv', header=0, error_bad_lines=False)
    stocklist = list(data.values)
    #print(stocklist[999])
    f = open('GARBAGE_COLLECTIONS_DATA.csv', 'a')
    #writer = csv.DictWriter(f, fieldnames=["y", "ds"])
    #writer.writeheader()
    f.write('ds,y\n')
    for i in stocklist: #itertools.islice(stocklist, 25000)
        #print(i[2])
        #print(i[3])
        if pd.isna(i[3]) == True or i[1] != 'GARBAGE COLLECTIONS':
            print('A row ' + str(i) + ' nao tem dados de toneladas recolhidas ou não é do Load Type pretendido')
        else:
            #'02/17/2017 10:53:00 AM'
            timestamp = datetime.strptime(i[2], '%m/%d/%Y %I:%M:%S %p').strftime('%Y-%m-%d %H:%M:%S')
            #print(timestamp)
            if ',' not in str(i[3]):
                i[3] = str(i[3]) + ',000'

            if len(str(i[3])) < 8 and str(i[3]).count(",") < 2 and str(i[3]).count(".") < 2:
                #print(i[3])
                #print(str(i[3]).count("."))
                #print(str(i[3]).count(",") < 2)
                #if ',' in str(i[3]) and '.' in str(i[3]):
                #    #return 'Caught a ridiculous value...'
                val = str(i[3]).replace(",",".")
                if float(val) < 50:
                    valore = str(val) + '\n'
                    #print(valore)
                    #print("-----")
                    #f.write(",".join([timestamp, str(i[3]), '\n']))
                    #if float(valore) < 50:
                    f.write(str(timestamp) + ', '+ valore)

    f.close()
    end = datetime.now()
    return 'Success. Total Processing Time of GARBAGE COLLECTIONS data was ' + str(end - start)

```

Figura 5-2: Exemplo de código utilizado na construção da aplicação *Flask* (Figura retirada de código-fonte do projeto no dia 5 de Maio de 2020)

Neste exemplo encontram-se presentes elementos necessários para o pré-processamento do conjunto de dados original, onde se deve destacar, especificamente, as linhas de código que lidam com os registos do espaço temporal dos dados. Estas linhas de código contêm uma das bibliotecas potencializadas pelo próprio *Flask*.

De acordo com os tipos de resíduos pretendidos, foram medidos os tempos que a API demorou para processar cada tipo de resíduo (as figuras correspondentes aos tempos de processamento da API por cada resíduo, durante o seu funcionamento, podem ser consultadas no Apêndice 1), sendo que estes tempos foram registados e fornecidos pela própria aplicação *Flask*.

Os resultados foram documentados e sintetizados no quadro disponibilizado em baixo, Quadro 5-1:

Quadro 5-1: Registo dos tempos de processamento obtidos pela API desenvolvida consoante os diferentes tipos de resíduos

Nome do Tipo de Resíduo	Tempo de Processamento Registado
<b>GARBAGE COLLECTIONS</b>	3 minutos e 20 segundos
<b>RECYCLING - SINGLE STREAM</b>	4 minutos e 13 segundos
<b>SWEEPING</b>	4 minutos e 19 segundos
<b>YARD TRIMMING</b>	4 minutos
<b>BULK</b>	4 minutos e 38 segundos
<b>BRUSH</b>	4 minutos e 19 segundos
<b>RECYCLING - PAPER</b>	4 minutos e 18 segundos
<b>RECYCLING - COMINGLE</b>	4 minutos e 51 segundos
<b>ORGANICS</b>	4 minutos e 34 segundos
<b>MIXED LITTER</b>	4 minutos e 52 segundos

<b>RECYCLED METAL</b>	4 minutos e 23 segundos
<b>RECYCLING - PLASTIC BAGS</b>	4 minutos e 59 segundos

Tal como se pode verificar, os tempos de processamento do conjunto de dados são relativamente curtos, com uma variância bastante baixa.

Os tempos de cada resíduo mantiveram-se sempre entre os 4-5 minutos, com a exceção do resíduo **GARBAGE COLLECTIONS**, que arrecadou um tempo de processamento mais baixo na ordem dos 3 minutos e 20 segundos.

Estes resultados comprovam a utilidade e flexibilidade tanto da API desenvolvida, como da *framework* em si.

## 6. Comparação de algoritmos *Machine Learning* de predição

Neste capítulo do relatório efetua-se a comparação entre todos os algoritmos alvos de análise, utilizando os *datasets* baseados no cenário.

### 6.1 Resultados obtidos da comparação

Para a análise proposta, foi analisado o comportamento dos algoritmos escolhidos na fase de pesquisa e elaboração do Estado da Arte, para cada conjunto de dados dos resíduos municipais existentes. Após os algoritmos calcularem, respetivamente, as suas previsões, o resultado final do comportamento é depois quantificado pela métrica **MAPE** (métrica escolhida derivada do estudo do Estado da Arte).

De modo a obter-se uma classificação geral com o propósito de se descobrir quais foram os algoritmos que se comportaram melhor que outros, foram definidos critérios pontuais para o efeito. Foi estabelecido que, para o algoritmo com melhor performance, correspondendo neste caso ao valor de MAPE menor, seria atribuído a este o maior número de pontos, enquanto que, para o algoritmo com o maior valor de MAPE, sendo neste cenário o menos eficaz, a penalização corresponde ao menor número de pontos atribuído.

Simplificando o esquema de classificação dos algoritmos de *Machine Learning*, obtemos o seguinte critério:

- **1º Classificado – 5 pontos** (Valor de MAPE mais baixo);
- **2º Classificado – 4 pontos;**
- **3º Classificado – 3 pontos;**
- **4º Classificado – 2 pontos;**
- **5º Classificado – 1 ponto** (Valor de MAPE mais elevado).

Para os resultados obtidos para cada tipo de resíduo, encontram-se disponibilizados, ao longo do capítulo, os quadros que contêm os valores da MAPE (*Mean Absolute Percentage Error*) de cada algoritmo, assim como os pontos atribuídos aos algoritmos mediante os valores de MAPE que foram registados.

No resíduo **GARBAGE COLLECTIONS**, todos os algoritmos tiveram um desempenho relativamente semelhante, à exceção do *Artificial Neural Networks* (ANN), com valor de MAPE de 93.437%, cujas previsões não tiveram o mesmo grau de precisão que a dos seus oponentes. Continuando neste resíduo, falta salientar os algoritmos *Prophet* (60.720%) e *Support Vectoring Model* (SVM) (60.403%), estes que tiveram um resultado bastante similar, tendo o último arrecadado uma ligeira vantagem. Os resultados obtidos e classificação atribuída aos algoritmos podem ser consultados no Quadro 6-1:

Quadro 6-1: Resultados da previsão dos algoritmos para o resíduo Garbage Collections

Algoritmos <i>Machine Learning</i>	ANN	ARIMA	<i>Decision Trees</i>	<i>Prophet</i>	SVM
<b>Valor MAPE (%)</b>	93.437	62.373	67.745	60.720	60.403
<b>Pontos Atribuídos</b>	1	3	2	4	5

De acordo com os resultados do resíduo **RECYCLING – SINGLE STREAM**, o algoritmo que claramente não conseguiu calcular previsões ao nível dos seus rivais foi o *Prophet*, com MAPE correspondente a 210.811%. Destaque para o *AutoRegressive Integrated Moving Average* (ARIMA), com 59.361%, *Decision Trees*, com 59.524%, e SVM, com 58.693%, por conseguirem resultados bastante competitivos, com, uma vez mais, o último dos três a obter uma performance um pouco melhor. O Quadro 6-2 expõe a sumarização dos resultados:

Quadro 6-2: Resultados da previsão dos algoritmos para o resíduo Recycling – Single Stream

Algoritmos <i>Machine Learning</i>	ANN	ARIMA	<i>Decision Trees</i>	<i>Prophet</i>	SVM
<b>Valor MAPE (%)</b>	85.225	59.361	59.524	210.811	58.693
<b>Pontos Atribuídos</b>	2	4	3	1	5
<b>Classificação Pontual</b>	3	7	5	5	10

Comparando os resultados obtidos para o resíduo **SWEEPING**, é possível verificar que, na generalidade, não existem algoritmos com níveis de performance relativamente semelhantes. Algoritmos a destacar: ANN (88.289%), por ter a pior performance dos cinco, e *Prophet* (29.639%), pelo motivo contrário, com um resultado claramente melhor que os restantes. Consultando o seguinte quadro, Quadro 6-3, é possível analisar os valores obtidos, tanto de MAPE como de classificação, para os algoritmos analisados:

Quadro 6-3: Resultados da previsão dos algoritmos para o resíduo Sweeping

Algoritmos <i>Machine Learning</i>	ANN	ARIMA	<i>Decision Trees</i>	<i>Prophet</i>	SVM
<b>Valor MAPE (%)</b>	88.289	59.869	39.127	29.639	48.961
<b>Pontos Atribuídos</b>	1	2	4	5	3
<b>Classificação Pontual</b>	<b>4</b>	<b>9</b>	<b>9</b>	<b>10</b>	<b>13</b>

Tendo em atenção a MAPE dos algoritmos para **YARD TRIMMING**, destaque para os algoritmos ARIMA, com 42.574%, *Decision Trees*, com 40.462%, e SVM, com 40.287%, por terem resultados bastante semelhantes. Por uma terceira vez, o SVM leva ligeira vantagem. Em contrapartida, os algoritmos *Prophet* (148.110%) e ANN (92.061%) não obtiveram resultados ao nível dos três melhores. O Quadro 6-4 descreve os registos obtidos pelos algoritmos em teste:

Quadro 6-4: Resultados da previsão dos algoritmos para o resíduo Yard Trimming

Algoritmos <i>Machine Learning</i>	ANN	ARIMA	<i>Decision Trees</i>	<i>Prophet</i>	SVM
<b>Valor MAPE (%)</b>	92.061	42.574	40.462	148.110	40.287
<b>Pontos Atribuídos</b>	2	3	4	1	5
<b>Classificação Pontual</b>	<b>6</b>	<b>12</b>	<b>13</b>	<b>11</b>	<b>18</b>

Com níveis de performance distintos para os cinco algoritmos, excetuando *Decision Trees* (61.242%) e SVM (60.403%), é possível ver que o *Prophet*, com valor de MAPE de 38.907%, foi o algoritmo que obteve o melhor nível de previsão para o resíduo **BULK**. A disparidade nos níveis de precisão do algoritmo ANN (87.252%), comparativamente aos restantes, começa a ser visível (para os *datasets* dos resíduos municipais). O quadro seguinte, Quadro 6-5, sintetiza os resultados obtidos para este tipo de resíduo:

Quadro 6-5: Resultados da previsão dos algoritmos para o resíduo Bulk

Algoritmos <i>Machine Learning</i>	ANN	ARIMA	<i>Decision Trees</i>	<i>Prophet</i>	SVM
<b>Valor MAPE (%)</b>	87.252	55.920	61.242	38.907	60.403
<b>Pontos Atribuídos</b>	1	4	2	5	3
<b>Classificação Pontual</b>	7	16	15	16	21

Para o resíduo **BRUSH**, existe uma clara distinção entre os algoritmos com bons resultados: ARIMA, com 41.899%, *Decision Trees*, com 43.433%, e SVM, com 49.862%, e os algoritmos com resultados aquém dos melhores: ANN, com 82.722%, e *Prophet*, com 84.579%. Dos três melhores algoritmos, o que demonstrou níveis de precisão mais robustos foi efetivamente o ARIMA. Os pontos de classificação e valor MAPE de cada algoritmo referentes ao resíduo BRUSH podem ser consultados no Quadro 6-6:

Quadro 6-6: Resultados da previsão dos algoritmos para o resíduo Brush

Algoritmos <i>Machine Learning</i>	ANN	ARIMA	<i>Decision Trees</i>	<i>Prophet</i>	SVM
<b>Valor MAPE (%)</b>	82.722	41.899	43.433	84.579	49.862
<b>Pontos Atribuídos</b>	2	5	4	1	3
<b>Classificação Pontual</b>	9	21	19	17	24

No caso do resíduo **RECYCLING – PAPER**, o algoritmo *Prophet* (34.614%) destaca-se por revelar uma melhor precisão nas suas previsões que os restantes algoritmos, notando que estes quatro obtiveram um valor de MAPE relativamente semelhante. Entre estes, o algoritmo com registo mais negativo é o ARIMA, com 71.687%. O quadro seguinte, Quadro 6-7, revela os resultados obtidos para este tipo de resíduo:

Quadro 6-7: Resultados da previsão dos algoritmos para o resíduo Recycling - Paper

Algoritmos <i>Machine Learning</i>	ANN	ARIMA	<i>Decision Trees</i>	<i>Prophet</i>	SVM
<b>Valor MAPE (%)</b>	69.859	71.687	62.291	34.614	63.506
<b>Pontos Atribuídos</b>	2	1	4	5	3
<b>Classificação Pontual</b>	<b>11</b>	<b>22</b>	<b>23</b>	<b>22</b>	<b>27</b>

Continuando para o resíduo **RECYCLING – COMINGLE**, é possível notar uma pior performance por parte do algoritmo SVM, cujo valor de MAPE foi 102.292%, enquanto que, por sua vez, o algoritmo *Prophet*, com 57.866%, consegue obter um valor de MAPE melhor quando comparado, especificamente, com os algoritmos ANN (65.049%) e ARIMA (66.233%). De uma forma sumariada, é possível consultar os registos obtidos no seguinte quadro, Quadro 6-8:

Quadro 6-8: Resultados da previsão dos algoritmos para o resíduo Recycling - Comingle

Algoritmos <i>Machine Learning</i>	ANN	ARIMA	<i>Decision Trees</i>	<i>Prophet</i>	SVM
<b>Valor MAPE (%)</b>	65.049	66.233	81.732	57.866	102.292
<b>Pontos Atribuídos</b>	4	3	2	5	1
<b>Classificação Pontual</b>	<b>15</b>	<b>25</b>	<b>25</b>	<b>27</b>	<b>28</b>

Contrastando com o resíduo anterior, os algoritmos *Decision Trees*, com 54.556%, e SVM, com 53.866%, quando efetuando previsão para dados do resíduo **ORGANICS**, conseguem obter melhores níveis de previsão, levando este último uma ligeira vantagem. Por sua vez, os algoritmos *Prophet* (83.419%) e ANN (86.376%) não conseguiram obter os mesmos níveis de precisão quando comparados com os do tipo de resíduo processado previamente. O seguinte quadro, Quadro 6-9, sumariza os resultados dos algoritmos:

Quadro 6-9: Resultados da previsão dos algoritmos para o resíduo Organics

Algoritmos <i>Machine Learning</i>	ANN	ARIMA	<i>Decision Trees</i>	<i>Prophet</i>	SVM
<b>Valor MAPE (%)</b>	86.376	54.844	54.556	83.419	53.866
<b>Pontos Atribuídos</b>	1	3	4	2	5
<b>Classificação Pontual</b>	<b>16</b>	<b>28</b>	<b>29</b>	<b>29</b>	<b>33</b>

Para o resíduo **MIXED LITTER**, é possível determinar um grupo de algoritmos que não obteve resultados favoráveis, sendo este constituído por ARIMA (96.959%), *Decision Trees* (97.096%) e SVM (108.739%). Contrariamente a estes três, o algoritmo *Prophet*, com valor de MAPE de 35.383%, obteve um valor de MAPE consideravelmente satisfatório. Os registos obtidos no comparativo entre algoritmos para este tipo de resíduo podem ser vistos no Quadro 6-10:

Quadro 6-10: Resultados da previsão dos algoritmos para o resíduo Mixed Litter

Algoritmos <i>Machine Learning</i>	ANN	ARIMA	<i>Decision Trees</i>	<i>Prophet</i>	SVM
<b>Valor MAPE (%)</b>	64.881	96.959	97.096	35.383	108.739
<b>Pontos Atribuídos</b>	4	3	2	5	1
<b>Classificação Pontual</b>	<b>20</b>	<b>31</b>	<b>31</b>	<b>34</b>	<b>34</b>

No caso do **RECYCLED METAL**, destaque para os maus resultados obtidos pelos algoritmos ARIMA, com valor obtido de 106.761%, e SVM, com valor de 146.483%. Continuando os bons resultados obtidos com o resíduo anterior, o *Prophet* (25.035%) volta a obter níveis de precisão nas suas previsões muito agradáveis. O quadro seguinte, Quadro 6-11, contém os resultados do comparativo para os dados do resíduo RECYCLED METAL:

Quadro 6-11: Resultados da previsão dos algoritmos para o resíduo Recycled Metal

Algoritmos <i>Machine Learning</i>	ANN	ARIMA	<i>Decision Trees</i>	<i>Prophet</i>	SVM
<b>Valor MAPE (%)</b>	58.777	106.761	40.483	25.035	146.483
<b>Pontos Atribuídos</b>	3	2	4	5	1
<b>Classificação Pontual</b>	<b>23</b>	<b>33</b>	<b>35</b>	<b>39</b>	<b>35</b>

Para este último resíduo, o algoritmo *Decision Trees*, com valor de MAPE de 42.073%, obteve o melhor resultado, quando comparado aos restantes. Por sua vez, o algoritmo ANN (99.282%) demonstrou não estar ao mesmo nível que os restantes no que toca à previsão de dados do resíduo **RECYCLING – PLASTIC BAGS**. Para o resíduo em questão, disponibiliza-se o Quadro 6-12 para permitir uma análise dos registos obtidos:

Quadro 6-12: Resultados da previsão dos algoritmos para o resíduo Recycling – Plastic Bags

Algoritmos <i>Machine Learning</i>	ANN	ARIMA	<i>Decision Trees</i>	<i>Prophet</i>	SVM
<b>Valor MAPE (%)</b>	99.282	70.044	42.073	72.064	46.870
<b>Pontos Atribuídos</b>	1	3	5	2	4
<b>Classificação Pontual</b>	<b>24</b>	<b>36</b>	<b>40</b>	<b>41</b>	<b>39</b>

## 6.2 Conclusões da análise efetuada

Após uma comparação extensiva dos algoritmos de *Machine Learning* deste estudo, englobando todos os tipos de resíduos no conjunto de dados, que serão alvos de análise na plataforma de ingestão de dados da Altice Labs, a classificação dos algoritmos, tendo em conta o esquema pontual estabelecido, foi a seguinte:

- **1º Classificado:** Prophet (41 pontos)
- **2º Classificado:** Decision Trees (40 pontos)
- **3º Classificado:** Support Vectoring Model (39 pontos)
- **4º Classificado:** AutoRegressive Integrated Moving Average (36 pontos)
- **5º Classificado:** Artificial Neural Networks (24 pontos)

Mediante esta classificação, é possível afirmar que o algoritmo que obteve, na sua generalidade, melhores níveis de precisão nas suas previsões foi o **Prophet**, conseguindo os níveis mais baixo de MAPE em seis dos doze tipos de resíduos em análise. Como tal, de todos os algoritmos comparados neste estudo, este foi o algoritmo escolhido para desencadear previsões de dados para os conjuntos de dados de resíduos municipais, nos quais os ficheiros de teste deste comparativo foram baseados.

De notar que, apesar de não demonstrarem os mesmos níveis de precisão do algoritmo *Prophet*, os algoritmos *Decision Trees* e *Support Vectoring Model* (SVM) podem ser considerados como excelentes alternativas para aplicação de *Machine Learning* em dados de resíduos municipais.

Porém, de acordo com os resultados deste estudo, o algoritmo que, efetivamente, não esteve ao mesmo nível dos seus “adversários” foi o *Artificial Neural Networks* (ANN).

Dos doze diferentes tipos de resíduos municipais existentes, este algoritmo em nenhuma ocasião conseguiu obter os melhores níveis de precisão, pelo que, devido aos algoritmos de *Machine Learning* que existem atualmente como alternativa, a sua aplicação em outros *datasets* deste género não é recomendada.

## 7. Arquitetura de API para ingestão de dados de previsão

De acordo com os resultados obtidos no comparativo dos algoritmos de *Supervised Learning*, foi decidido que o *Prophet* seria o algoritmo encarregue de adicionar valor aos conjuntos de dados dos resíduos municipais.

Como tal, de modo a se atingir o pretendido, recorreu-se mais uma vez à ferramenta *Flask* para, efetivamente, conciliar uma API com o *Prophet*, cujas capacidades estão assentes na previsão de valores futuros, de modo a serem gerados novos *datasets* com os dados dos resíduos do passado e presente, agregados aos dados do futuro.

A figura seguinte, Figura 7-1, corresponde a um esquema da arquitetura do sistema desenvolvido com o auxílio da aplicação *Flask* e do algoritmo *Prophet*:

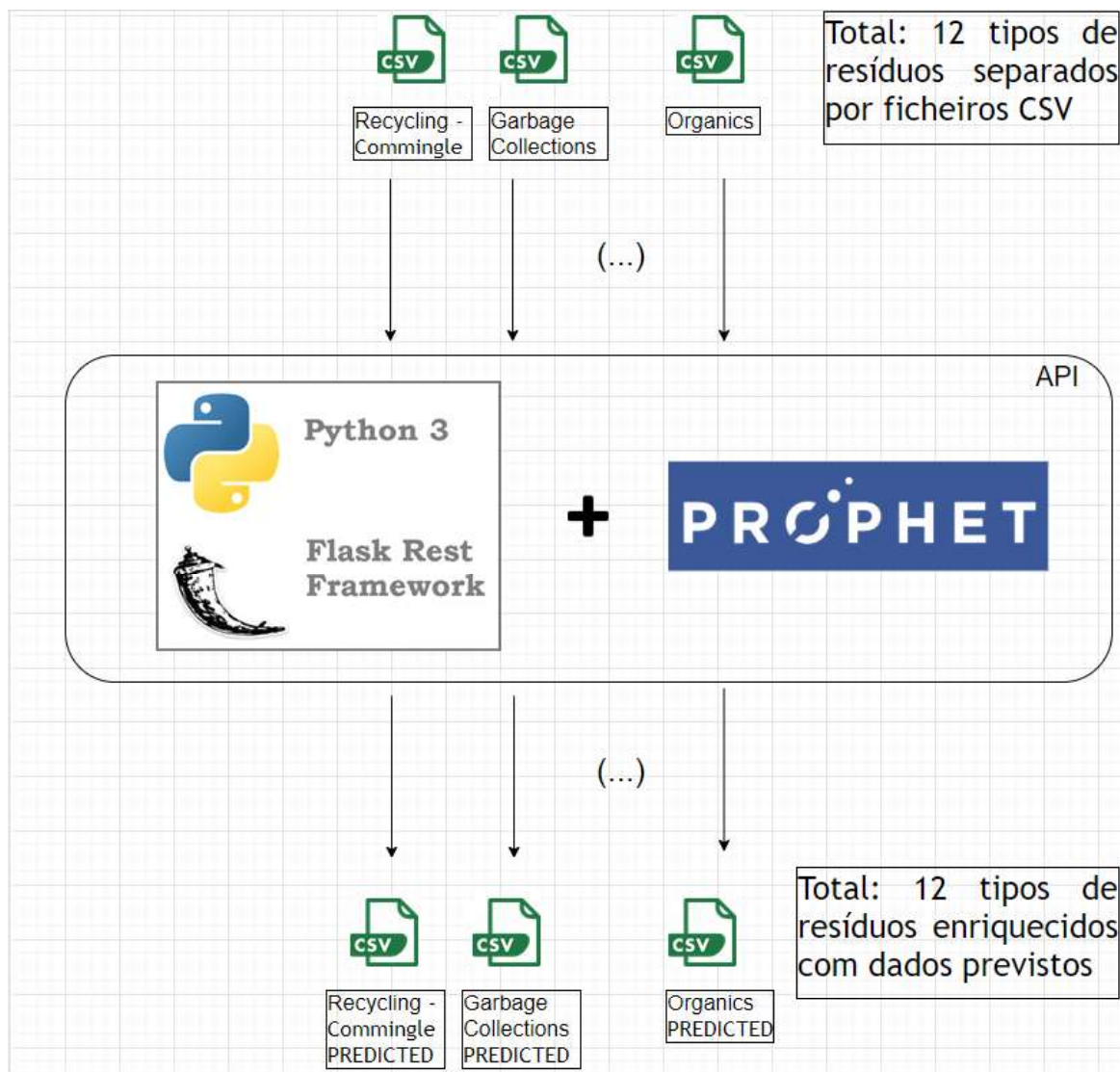


Figura 7-1: Arquitetura da aplicação Flask de previsão desenvolvida (Esquema elaborado em: <https://app.diagrams.net/>)

Para cada um ficheiro de um tipo de resíduo em formato CSV, será gerado um novo ficheiro CSV com o acréscimo do sufixo PREDICTED que irá conter os dados que foram sendo registados pela empresa de recolha dos resíduos de Austin, e, de igual forma, os dados de previsão calculados pelo algoritmo *Prophet*. É de notar que este último calculou valores de previsão cujo espaço temporal se estende até ao final do ano de 2021. Estes registos estarão presentes no ficheiro CSV num formato uniforme.

## 7.1 Resultados obtidos da ingestão

Consoante os tipos de resíduos existentes, foram medidos os tempos que a API demorou para processar cada tipo de resíduo, de modo a conciliar os dados obtidos durante o tempo e os dados previstos pelo algoritmo (as figuras que detalham os tempos de processamento da API

de previsão por cada resíduo podem ser consultadas no Apêndice 3). Os resultados obtidos podem ser consultados no Quadro 7-1:

Quadro 7-1: Registo dos tempos de processamento obtidos pela API de previsão desenvolvida consoante os diferentes tipos de resíduos

Nome do Tipo de Resíduo	Tempo de Processamento Registado
<b>GARBAGE COLLECTIONS</b>	17 minutos e 31 segundos
<b>RECYCLING - SINGLE STREAM</b>	10 minutos e 17 segundos
<b>SWEEPING</b>	1 minuto e 25 segundos
<b>YARD TRIMMING</b>	10 minutos e 17 segundos
<b>BULK</b>	2 minutos e 44 segundos
<b>BRUSH</b>	3 minutos e 56 segundos
<b>RECYCLING - PAPER</b>	2 minutos e 3 segundos
<b>RECYCLING - COMINGLE</b>	1 minuto e 38 segundos
<b>ORGANICS</b>	15 segundos
<b>MIXED LITTER</b>	5 segundos
<b>RECYCLED METAL</b>	4 segundos
<b>RECYCLING - PLASTIC BAGS</b>	35 segundos

Tendo em conta os resultados obtidos, é possível verificar que os tempos de processamento da API desenvolvida têm uma variância considerável para os vários resíduos existentes, contemplando valores na ordem dos segundos (Organics, Mixed Litter, Recycled Metal e Recycling – Plastic Bags) e valores na ordem da dezena de minutos (Garbage Collections, Recycling – Single Stream e Yard Trimming). Estes resultados são justificados com o facto de existirem resíduos que possuem uma quantia de registos consideravelmente baixa (na ordem das centenas), quando comparados com os resíduos que possuem registos na ordem dos dezenas de milhares.

Como, no caso dos resíduos com mais registos, existe mais informação acessível ao algoritmo *Prophet* para este poder prever eventos futuros com um maior nível de certeza, é expectável que este modelo demore mais tempo a conciliar todos os eventos existentes no conjunto de dados a enriquecer. Porém, tendo em conta que estão a ser tratados e previstos grandes quantidades de dados, deve-se salientar a flexibilidade disponibilizada pelo *Flask* aquando da sua integração com o algoritmo de *Machine Learning Prophet*.

## 7.2 Submissão dos conjuntos de dados na plataforma de gestão

Após a geração dos ficheiros de dados, o próximo passo consistiu em armazenar os dados na plataforma de gestão da empresa Altice Labs mencionada no capítulo 4. Para o efeito, foi necessário atribuir identificadores (na plataforma correspondem a *tags*) consoante o tipo de resíduo aos quais os dados são referentes. Para o efeito, seguiu-se a nomenclatura sugerida pela empresa de a primeira *tag* corresponder ao domínio dos dados, e a segunda *tag* corresponder ao parceiro que fornece os dados.

O resultado final da criação de tabelas na plataforma de gestão pode ser consultado na Figura 7-2:

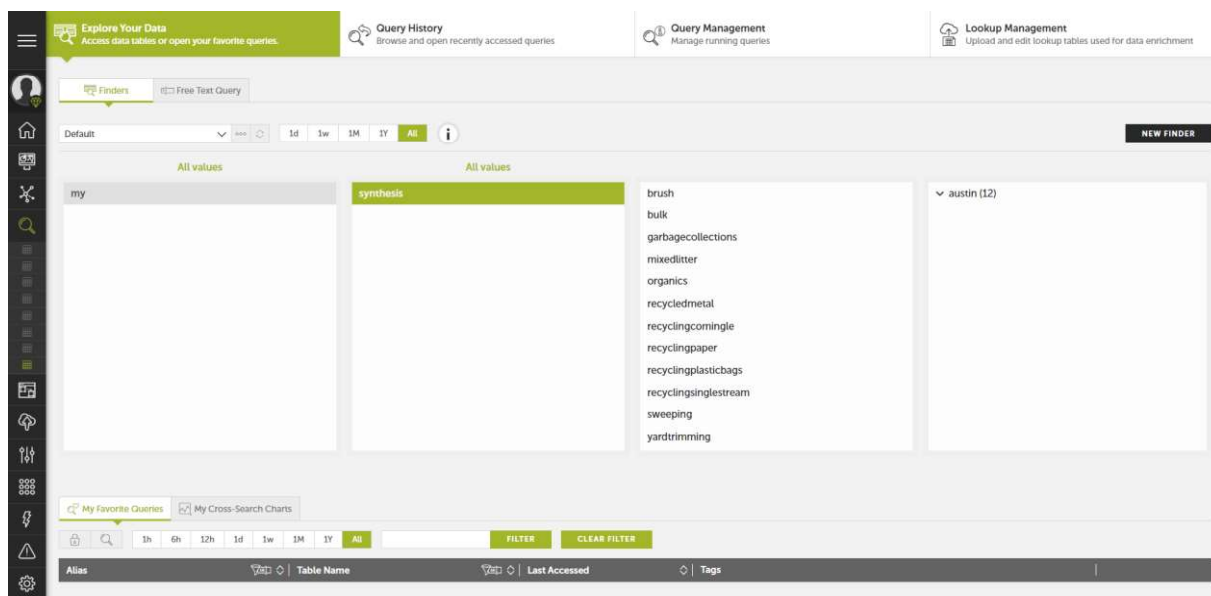


Figura 7-2: Acesso aos dados dos tipos de resíduos na plataforma de gestão de dados (Retirado de: <https://alticelabs.devo.com/welcome> a 7 de Agosto de 2020)

Fornecendo um exemplo, caso o utilizador pretenda consultar os dados correspondentes ao **Recycled Metal**, este deve aceder à *tag* final *my.synthesis.recycledmetal.austin*, onde *my.synthesis* corresponde a uma *tag* predefinida que o sistema atribui, *recycledmetal* corresponde ao domínio onde os dados se inserem, e *austin* corresponde ao parceiro que forneceu os dados.

## 8. Plataforma de Gestão de Dados

Neste capítulo serão exploradas as características da plataforma utilizada, assim como a consulta dos dados dos resíduos na plataforma, aliada à criação de “valor” para os clientes da plataforma, recorrendo a *dashboards* na temática de *Business Intelligence*.

### 8.1 Breve descrição da plataforma

A arquitetura da plataforma utilizada foi projetada para armazenar, analisar e fornecer informações sobre os dados com rapidez, sendo escalável consoante o inevitável crescimento da quantidade de dados que são inseridos na plataforma. A plataforma de gestão de dados combate os elevados tempos de indexação de dados na ingestão destes fornecendo uma solução *cloud-native*, que permite a obtenção de maior flexibilidade e escalabilidade quando comparada com soluções *cloud* do género *lift-and-shift*.

Tendo também por base uma arquitetura *multitenant* (múltiplos domínios), a plataforma oferece suporte a milhares de organizações, fornecendo uma instância dedicada a cada cliente na qual o mesmo pode efetuar a gestão dos dados presentes, assim como a gestão dos utilizadores do domínio. A interação com o domínio da organização pode ser realizada através da interface do utilizador, ou pode ser efetuada por uma API proprietária do cliente capaz de comunicar com uma REST API disponibilizada pela plataforma.

O cliente pode consultar os dados presentes na plataforma recorrendo a um sistema de *tags*, que tem o objetivo de classificar os tipos de dados a serem analisados por tabelas, ou pode aceder aos dados através de um *query editor* que suporta a linguagem de programação *Language-Integrated Query* - LINQ.

## 8.2 Consulta dos dados correspondentes aos tipos de resíduos

Como especificado previamente, para se poderem consultar os dados submetidos na plataforma, existem duas formas distintas de aceder a estes, ou através do sistema de *tags*, ou através da elaboração de uma *query* no editor dedicado.

O acesso feito diretamente pelo sistema de *tags* corresponde à escolha de uma das *tags* disponibilizadas pela interface, esta que se encontra representada pela Figura 7-2 presente no capítulo 7. Acedendo, por exemplo, aos dados do tipo **GARBAGE COLLECTIONS** (associados à *tag my.synthesis.garbagecollections.austin*), obtém-se o seguinte resultado, como demonstra a Figura 8-1:

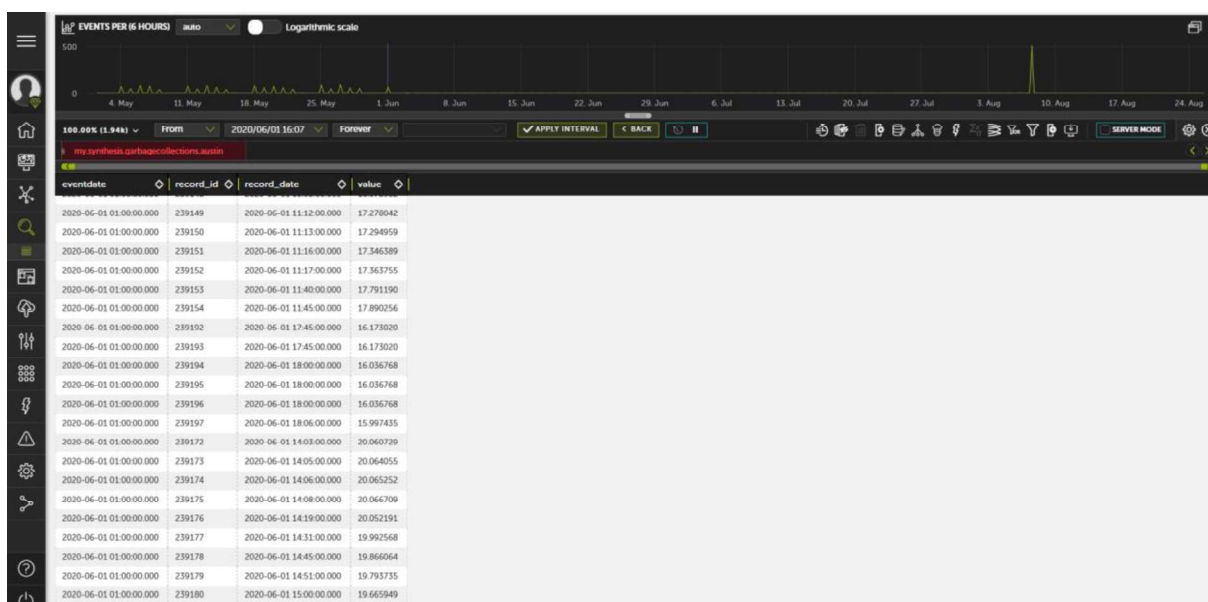


Figura 8-1: Consulta de dados do tipo Garbage Collections na plataforma de gestão (Retirado de: <https://alticelabs.devo.com/welcome> a 24 de Agosto de 2020)

De notar que, à semelhança dos restantes tipos de resíduos, a tabela é constituída por quatro colunas, sendo estas:

- **Eventdate:** Coluna criada por omissão pela plataforma; pode corresponder à data de inserção dos registos na plataforma, ou pode ser configurada de modo a assimilar as datas presentes nos registos inseridos (esta última opção foi efetivamente a utilizada para todos os tipos de resíduos);
- **Record\_id:** Corresponde ao identificador do registo presente na tabela;
- **Record\_date:** Este campo contém as datas de registo dos valores da tabela, sendo aplicado tanto para os valores registados pela *Austin Resource Recovery*, como para os valores previstos pela API de previsão;
- **Value:** O seguinte campo contém os valores associados ao registo, igualmente aplicável tanto para os valores registados pela entidade de Austin, como para os valores previstos pelo algoritmo ML.

A pesquisa pelo editor da *query* pressupõe a utilização da linguagem LINQ, cujo formato pode ser representado como demonstrado pela Figura 8-2:

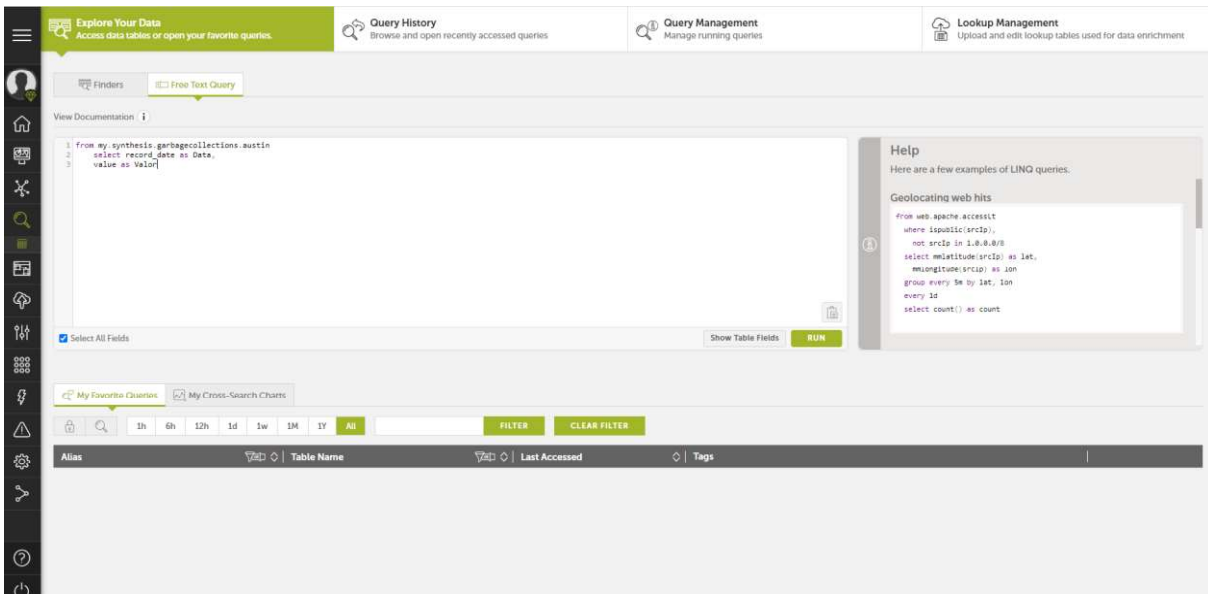


Figura 8-2: Consulta de dados do tipo Garbage Collections na plataforma de gestão com editor de query (Retirado de: <https://alticelabs.devo.com/welcome> a 24 de Agosto de 2020)

Para o caso apresentado na Figura 8-2, a *query* tem como alvo a tabela que contém os dados do tipo de resíduo **GARBAGE COLLECTIONS**, acessível através da *tag* apropriada, com a adição dos campos *record\_date* e *value* serem representados pelos nomes Data e Valor, respetivamente. O resultado da pesquisa elaborada encontra-se representado na Figura 8-3:

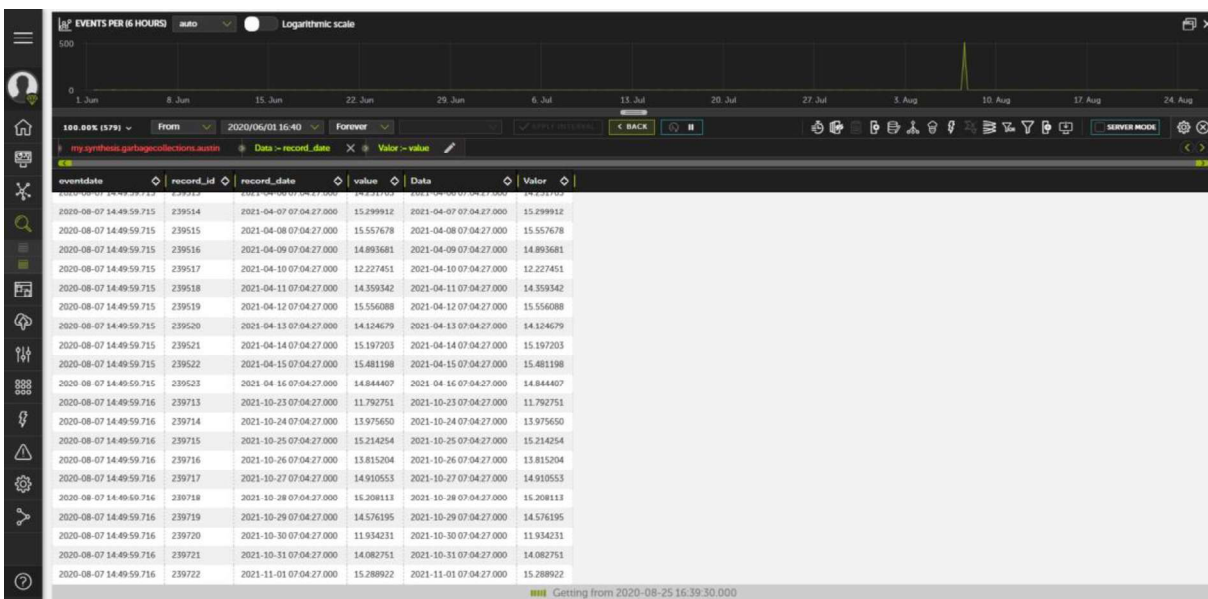


Figura 8-3: Resultado da consulta elaborada no editor de query (Retirado de: <https://alticelabs.devo.com/welcome> a 24 de Agosto de 2020)

### 8.3 Caracterização dos *widjets* constituintes dos *Dashboards* BI dos tipos de resíduos existentes

De modo a se aplicar a componente de *Business Intelligence* (BI) nos dados tratados dos resíduos, é necessário transitar-se para a componente de ‘*Activeboards*’ da plataforma. Nesta, é possível construir-se *dashboards* com recurso a um conjunto de *widjets* que englobam tipos de gráficos tais como os gráficos de linhas, gráficos de barras, *SimpleValues*, entre outros.

A construção da informação que terá que ser representada pelos gráficos está dependente da utilização do editor de *query* em linguagem LINQ mencionado previamente, sendo que para cada *widjet* existente, tem-se de igual forma um editor dedicado. Foram criados *dashboards* para todos os tipos de resíduos presentes na plataforma, totalizando doze.

No que toca aos tipos de gráficos que caracterizam o(s) *dashboard(s)*, estes focam-se essencialmente na análise dos dados captados pela empresa de recolha do lixo, correspondendo aos gráficos que se encontram na parte superior do *dashboard*, e focam-se igualmente na análise dos dados previstos para o final de 2020 e totalidade de 2021, representados pelos gráficos presentes na parte inferior do *dashboard*.

Começando pelos gráficos encarregues da informação dos dados não previstos, utilizando o *dashboard* associado ao resíduo **GARBAGE COLLECTIONS** como exemplo, no gráfico de nome ‘*Garbage Collections' recorded data*’ encontra-se representada a informação de todos os registos captados pela *Austin Resource Recovery* desde Setembro de 2019 até Junho de 2020 (esta última data corresponde ao momento em que se obteve o conjunto de dados para se aplicar o devido pré-processamento dos dados e eventual enriquecimento com novos dados de previsão).

O facto de não se mostrar mais informação prévia a Setembro de 2019 (dado existirem registos datados desde Agosto de 2004) prende-se com a limitação do próprio *widjet*, sendo que este apenas permite mostrar informação com data no máximo de até 365 dias antes do dia atual. A figura seguinte, Figura 8-4, permite visualizar com melhor detalhe a informação presente no gráfico:

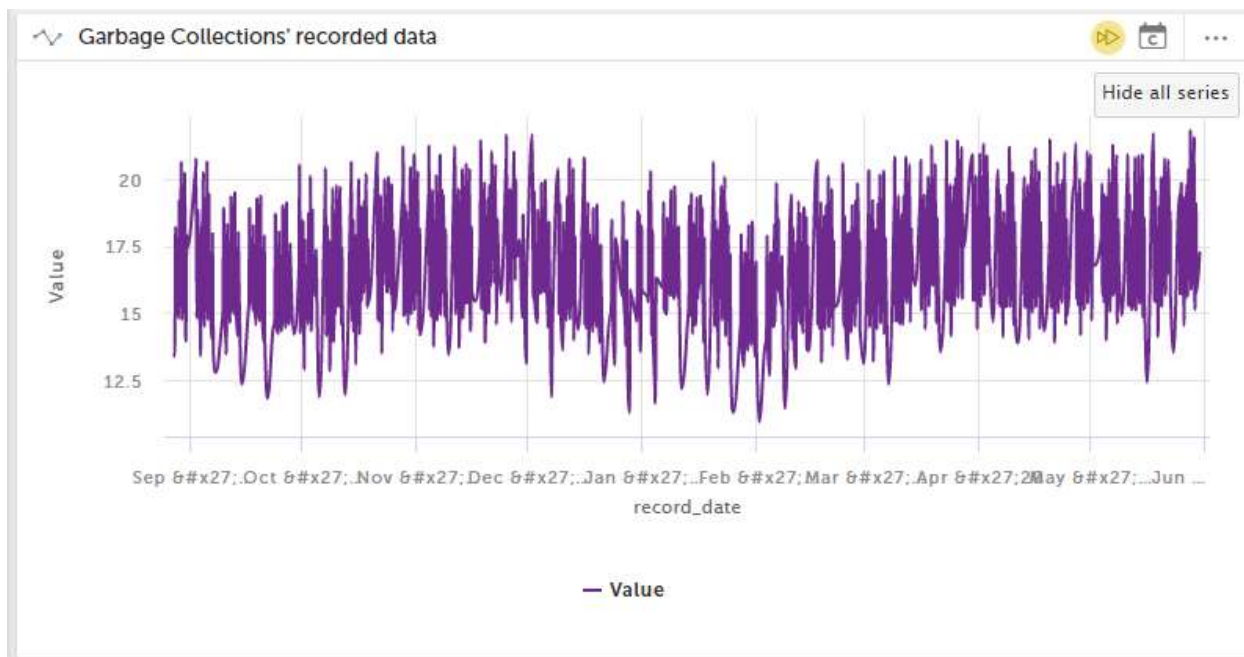


Figura 8-4: Gráfico ‘Garbage Collections' recorded data’ (Retirado de: <https://alticelabs.devo.com/welcome#/dashboard/1617985> a 28 de Agosto de 2020)

Seguidamente, nos *widets* de nome ‘*Recorded data's average of tons processed*’ e ‘*Recorded data's lowest tons processed*’ é possível consultar, respetivamente, a média de toneladas processadas por dia de acordo com os dados processados pela companhia e o valor mínimo registado pela companhia de processamento do resíduo **GARBAGE COLLECTIONS**.

A Figura 8-5 mostra em detalhe os valores correspondentes às métricas dos *widets* configuradas:

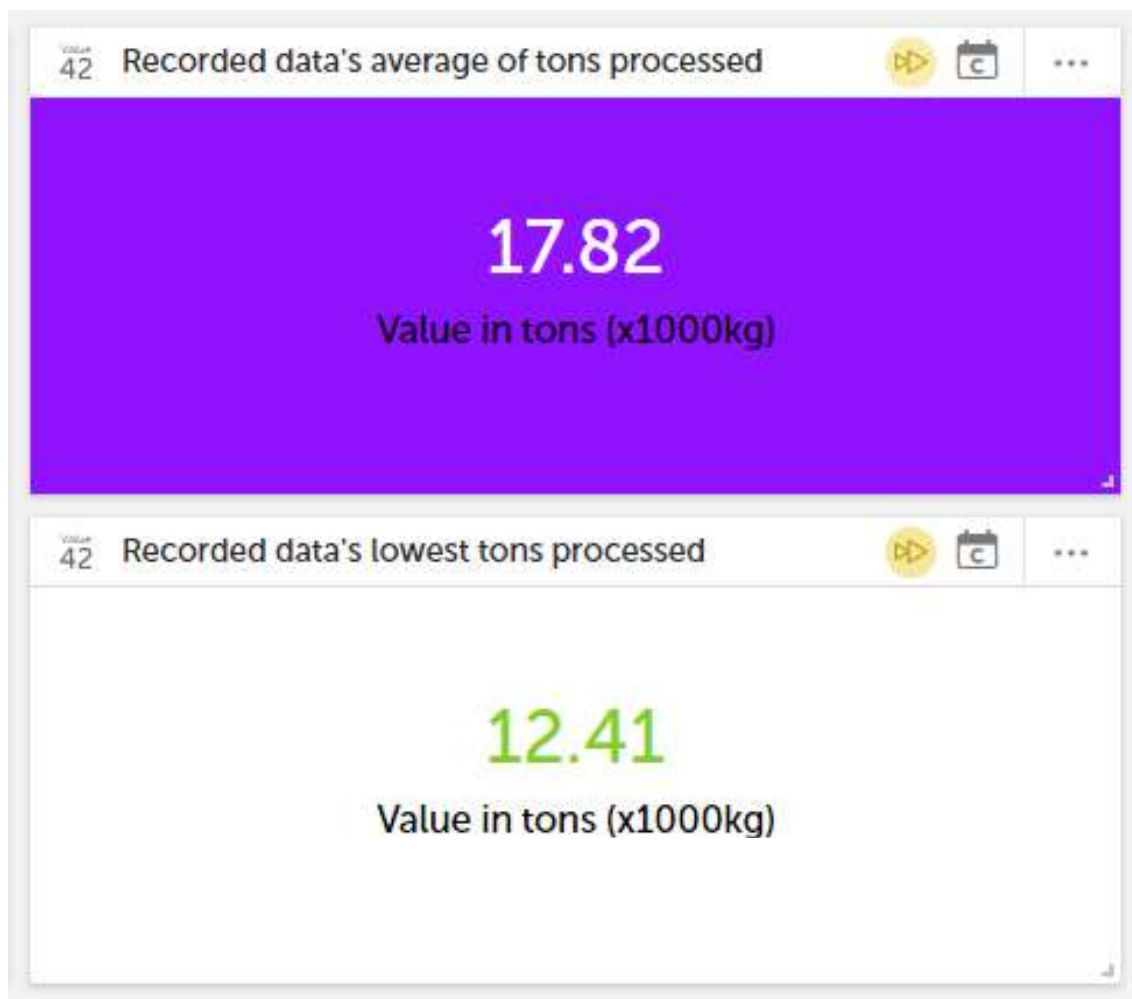


Figura 8-5: Widgets ‘Recorded data's average of tons processed’ e ‘Recorded data's lowest tons processed’ (Retirado de: <https://alticelabs.devo.com/welcome#/dashboard/1617985> a 28 de Agosto de 2020)

De modo a terminar a descrição dos *widgets* alusivos aos dados não previstos, no gráfico ‘Recorded data's total of tons processed per month’ encontra-se presente a informação agregada por mês (datado a partir de Setembro de 2019, pelas mesmas razões que o gráfico ‘Garbage Collections' recorded data’) do total de toneladas processadas até Junho de 2020, exclusive.

A Figura 8-6 demonstra o resultado obtido após a configuração deste *widget* para o resíduo GARBAGE COLLECTIONS:

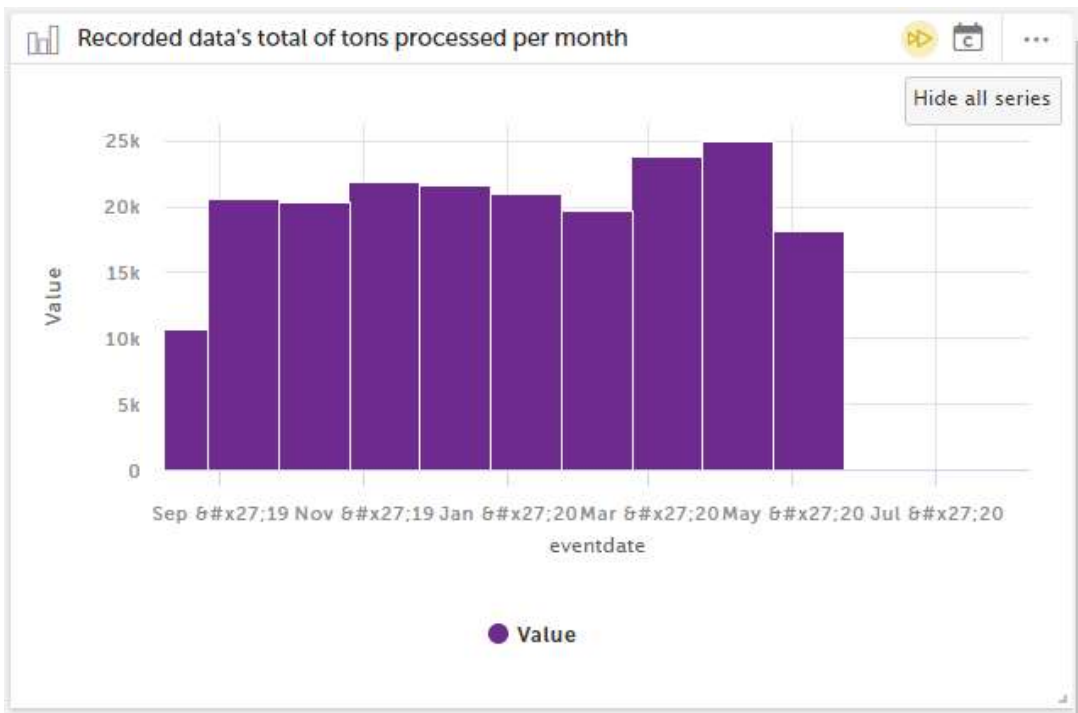


Figura 8-6: Gráfico ‘Recorded data's total of tons processed per month’ (Retirado de: <https://alticelabs.devo.com/welcome#/dashboard/1617985> a 28 de Agosto de 2020)

Transitando então para os gráficos associados aos dados de previsão; no gráfico ‘Garbage Collections' predicted data’ estão presentes os dados previstos pelo algoritmo de ML Prophet, datados desde Junho de 2020 até Dezembro de 2021.

A seguinte figura, Figura 8-7, representa o *widget* associado, assim como a informação disponibilizada pelo mesmo:

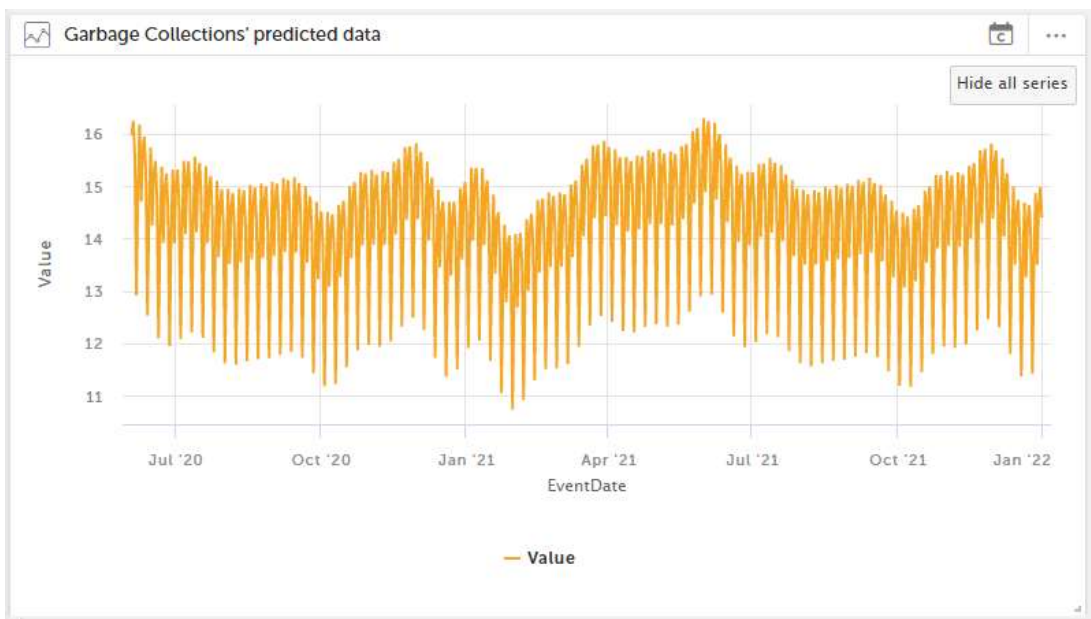


Figura 8-7: Gráfico ‘Garbage Collections' predicted data’ (Retirado de: <https://alticelabs.devo.com/welcome#/dashboard/1617985> a 28 de Agosto de 2020)

De seguida, nos *widgets* 'Predicted data's average of tons processed' e 'Predicted data's lowest tons processed' é possível consultar, respetivamente, os valores referentes à média de toneladas que poderão ser processadas por dia e ao mínimo de toneladas que poderão ser processadas, registados pelo algoritmo de Machine Learning no que toca ao resíduo GARBAGE COLLECTIONS.

A Figura 8-8 revela os valores correspondentes às métricas dos *widgets* anteriormente especificadas:

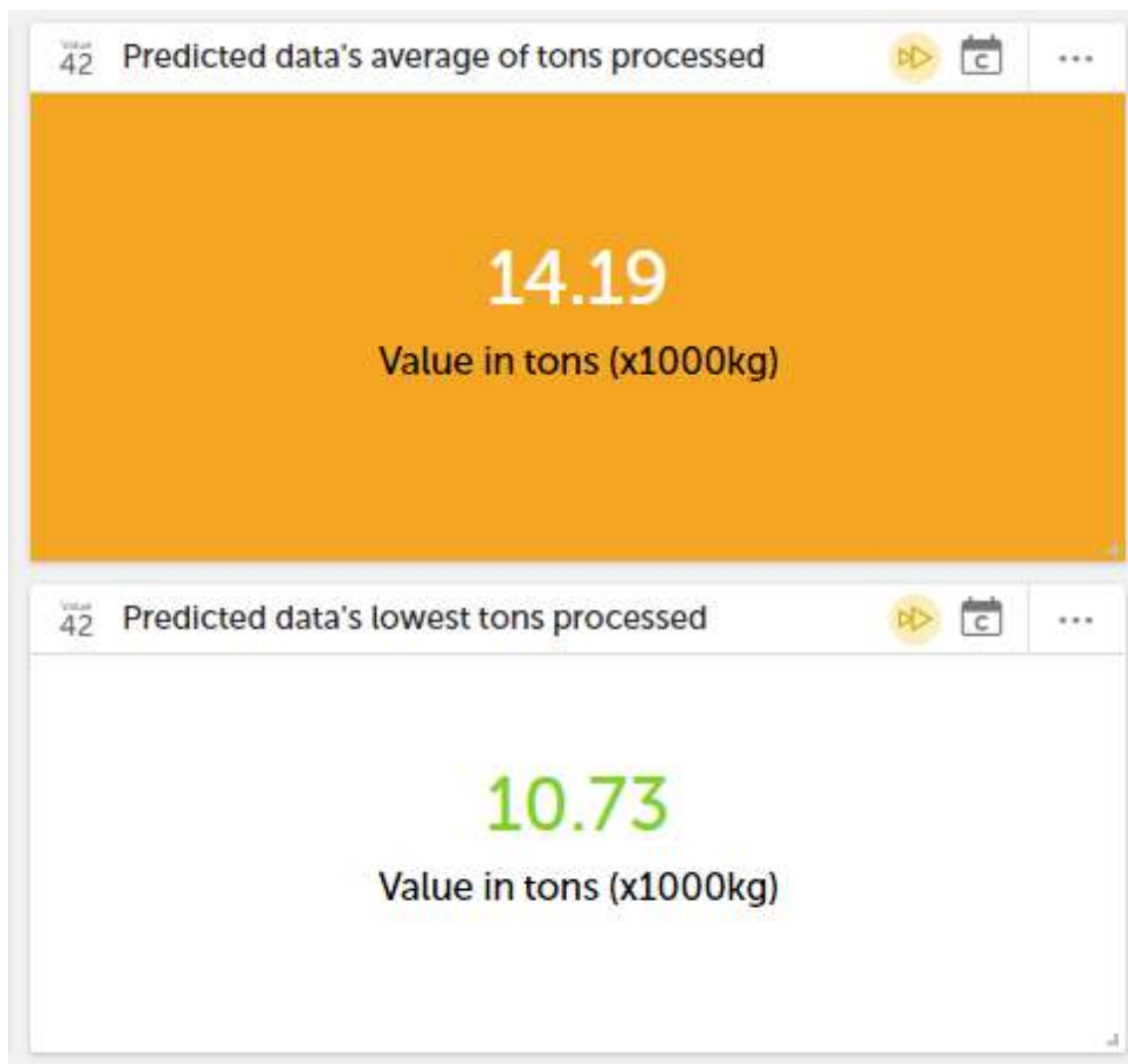


Figura 8-8: Widgets 'Predicted data's average of tons processed' e 'Predicted data's lowest tons processed' (Retirado de: <https://alticelabs.devo.com/welcome#/dashboard/1617985> a 28 de Agosto de 2020)

Por fim, no *widget* 'Predicted data of processed tons for 2021' é possível encontrar uma tabela que contém todos os registos previstos para o ano 2021, onde, consoante o valor presente no registo, é a este atribuída uma cor mediante a categoria de valores onde se insere. Existem no total quatro categorias, sendo que, para os registos cujos valores se enquadrem na categoria de valores mais baixa, os valores destes registos são destacados com a cor verde,

enquanto que, no outro extremo, para os registos cujos valores se enquadrem na categoria de valores mais alta, os valores destes registos são destacados com a cor vermelha. O objetivo deste *widget* passa por salientar os dias em que poderá haver uma maior atividade ou uma maior necessidade de recursos alocados.

A informação presente no *widget* descrito acima pode ser consultada na seguinte figura, Figura 8-9:

EventDate ^	Value
2021-01-01 07:04:27.000	14.498076000000001
2021-01-02 07:04:27.000	11.90537
2021-01-03 07:04:27.000	14.101115
2021-01-04 07:04:27.000	15.350231
2021-01-05 07:04:27.000	13.958215
2021-01-06 07:04:27.000	15.056013
2021-01-07 07:04:27.000	15.350403
2021-01-08 07:04:27.000	14.708755
2021-01-09 07:04:27.000	12.049688
2021-01-10 07:04:27.000	14.17306

Figura 8-9: Widget ‘Predicted data of processed tons for 2021’ (Retirado de: <https://alticelabs.devo.com/welcome#/dashboard/1617985> a 28 de Agosto de 2020)

## 8.4 Análise da informação consultável nos *dashboards* BI dos tipos de resíduos existentes

Após a construção de *dashboards* para todos os tipos de resíduos, é então possível obter os denominados *insights* sobre os dados armazenados. Estes encontram-se detalhados mediante cada categoria de resíduo existente.

No que respeita a **GARBAGE COLLECTIONS**, de acordo com o gráfico que contém os dados registados pela *Austin Resource Recovery* nos últimos 365 dias, é possível ver que as oscilações do fluxo de quantidade de lixo processada por dia mantêm-se constante. Porém, consegue-se estabelecer que o período onde a quantidade de lixo processada diariamente

atinge os valores mais baixos situou-se entre meados de Dezembro de 2019 e início de Fevereiro de 2020.

No que toca à média de toneladas processada pela entidade por dia, tendo em conta que os registos têm início em Agosto de 2004, este valor corresponde a 17.82 toneladas. No caso do valor mínimo registado, este corresponde a 12.41 toneladas.

Transitando para o gráfico que contém o total por mês de toneladas processadas nos últimos 365 dias, verifica-se que o mês com o maior valor corresponde a Abril, com um total de 24994 toneladas, enquanto que o mês com o menor valor registado é Agosto, com 10730 toneladas.

Em termos de *widgets* que contêm valores de previsão, é possível verificar que o algoritmo prevê um pico de toneladas de lixo a serem processadas para finais de Maio e inícios de Junho, destacando de igual forma o intervalo previsto de menor quantidade a ser processada entre meados de Janeiro de 2021 e meados de Fevereiro de 2021.

Tendo em conta o valor da potencial média de toneladas processadas, este, quando comparado com a média descrita previamente, desce para as 14.19 toneladas. Por sua vez, o potencial valor mínimo previsto é de 10.73 toneladas.

Por fim, analisando o *widget* com os dados previstos para o ano 2021, retira-se a conclusão de que o algoritmo prevê os maiores valores de toneladas processadas para os dias 31 de Maio, 3 de Junho e 7 de Junho, sendo estes 16.29, 16.24 e 16.21 toneladas respetivamente. No polo contrário, o algoritmo prevê os menores valores para os dias 23 de Janeiro, 30 de Janeiro, e 6 de Fevereiro, com os valores 11.05, 10.73 e 10.91 toneladas respetivamente.

A figura seguinte, Figura 8-10, corresponde ao *dashboard* que engloba os *widgets* do tipo de resíduo analisado:

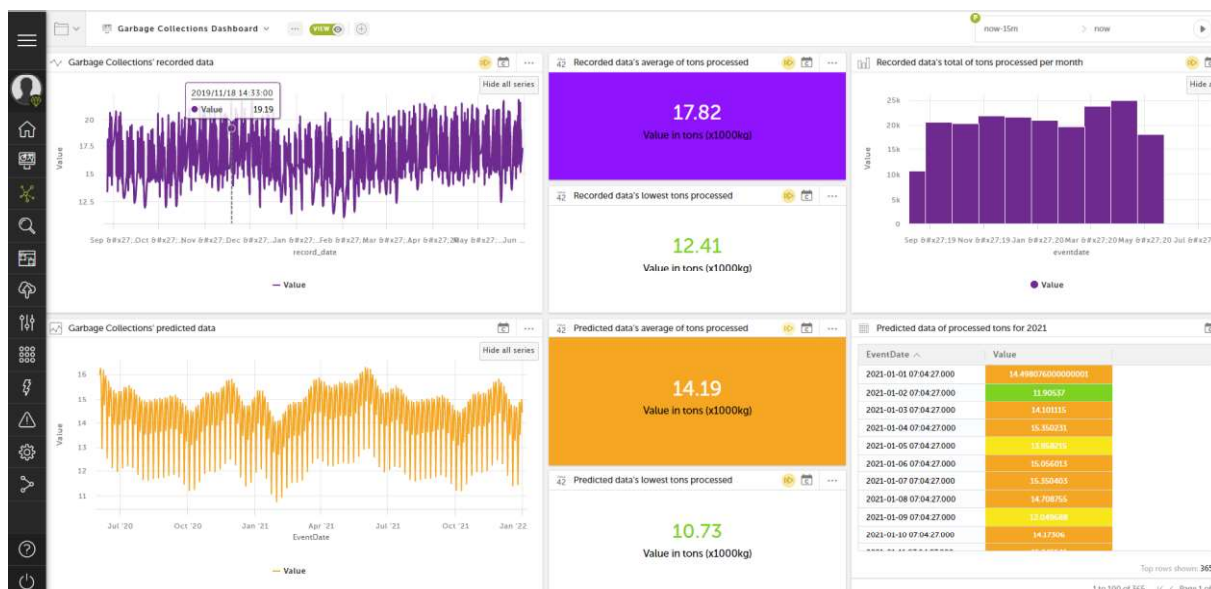


Figura 8-10: Dashboard do resíduo Garbage Collections (Retirado de: <https://alticelabs.devo.com/welcome#/dashboard/1617985> a 28 de Agosto de 2020)

Abordando os insights do resíduo **RECYCLING – SINGLE STREAM**, consultando o gráfico correspondente ao *display* dos últimos 365 dias de dados, comparativamente ao resíduo anterior, as quantidades de lixo processadas são consideravelmente mais baixas. Contudo, a nível de discrepância nas oscilações, estes partilham a mesma consistência. Em termos de período onde se registam os valores mais elevados, este situa-se entre inícios de Dezembro de 2019 e meados de Janeiro de 2020.

Referindo a média de toneladas processada pela entidade por dia, o valor é de 10.30 toneladas, sendo que, no caso do valor mínimo registado, este corresponde a 6.87 toneladas. No caso do gráfico que contém o total por mês de toneladas processadas nos últimos 365 dias, registou-se, à semelhança do resíduo anterior, Abril como sendo o mês com o maior valor, totalizando 11330 toneladas, onde o mês com o menor valor é Agosto, com 4521 toneladas.

Falando dos *widjets* que lidam com os valores de previsão, consegue-se constatar que o algoritmo prevê três picos para o potencial máximo de toneladas de lixo a serem processadas, sendo estes inícios de Dezembro de 2020 e inícios de Janeiro e Junho de 2021. O destaque para o intervalo previsto de menor quantidade a processar é em meados de Outubro de 2021.

Falando do valor da potencial média de toneladas processadas, após comparação com a média dos dados processados pela empresa, prevê-se um decréscimo para as 6.48 toneladas, sendo que o potencial valor mínimo previsto sofre igualmente uma queda, para as 4.94 toneladas.

Terminando a análise para este resíduo com o *widjet* com os dados previstos para o ano 2021, conclui-se que o algoritmo prevê os maiores valores de toneladas processadas para os dias 7 e 11 de Janeiro, sendo estes 7.50 e 7.51 toneladas respetivamente. Em contraste, o algoritmo prevê os registos mais baixos para os dias 10 e 17 de Outubro, com os valores 4.94 e 5.03 toneladas respetivamente.

Na seguinte figura, Figura 8-11, encontra-se o respetivo *dashboard* do resíduo analisado:

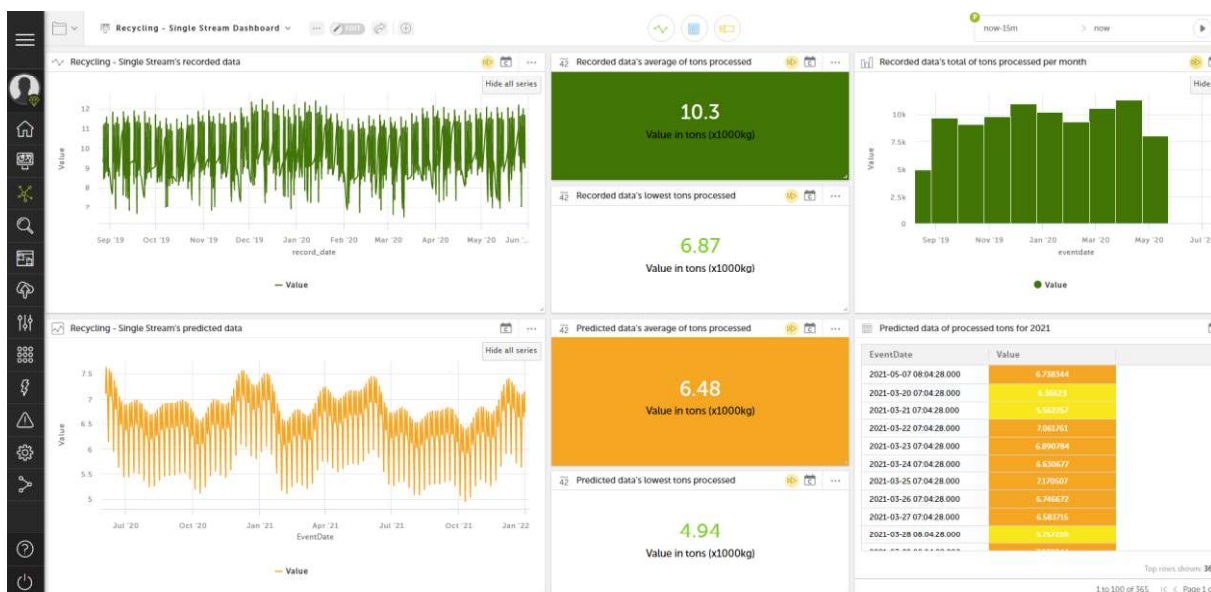


Figura 8-11: Dashboard do resíduo Recycling – Single Stream (Retirado de: <https://alticelabs.devo.com/welcome#/dashboard/1617985> a 28 de Agosto de 2020)

Focando no resíduo **SWEEPING**, no gráfico correspondente aos últimos 365 dias, comparativamente aos dois resíduos anteriormente analisados, é possível demarcar notáveis oscilações no fluxo de quantidade de lixo processada por dia, destacando um período de baixa quantidade de lixo processada diariamente entre meados de Novembro de 2019 e finais de Janeiro de 2020.

Em termos de média de toneladas processada pela entidade por dia, o valor registado corresponde a 13.14 toneladas. Falando do valor mínimo registado, este corresponde a 10.81 toneladas.

No que toca ao gráfico que contém o total por mês de toneladas processadas, conclui-se que o mês com o maior valor registado é Março, com um valor correspondente de 1071 toneladas. Por sua vez, o mês com o menor valor registado é Agosto, totalizando 401 toneladas.

Transitando para os *widjets* que demonstram informações de previsão, verifica-se que o algoritmo prevê um intervalo de maiores toneladas de lixo a serem processadas entre finais de Maio de 2021 e finais de Julho de 2021, onde, em contraste, o intervalo previsto de menor quantidade a ser processada está estabelecido entre finais de Novembro de 2020 e meados de Janeiro de 2021.

De acordo com o valor da potencial média de toneladas processadas, este, quando comparado com a média previamente constatada, desce para as 10.68 toneladas. Continuando, o algoritmo de ML definiu um potencial valor mínimo previsto de 6 toneladas.

Por forma a terminar a análise para este resíduo, analisando o *widjet* com os dados previstos para o ano 2021, conclui-se que o algoritmo prevê os maiores valores de toneladas processadas para os dias 23 e 30 de Maio, com os valores, respetivamente, de 16.32 e 16.43 toneladas. No outro extremo, o algoritmo prevê os valores mínimos para 10 e 17 de Dezembro, sendo os valores respetivos de 6.02 e 6.01 toneladas.

O *dashboard* com os *widjets* do tipo de resíduo analisado pode ser visto na Figura 8-12:

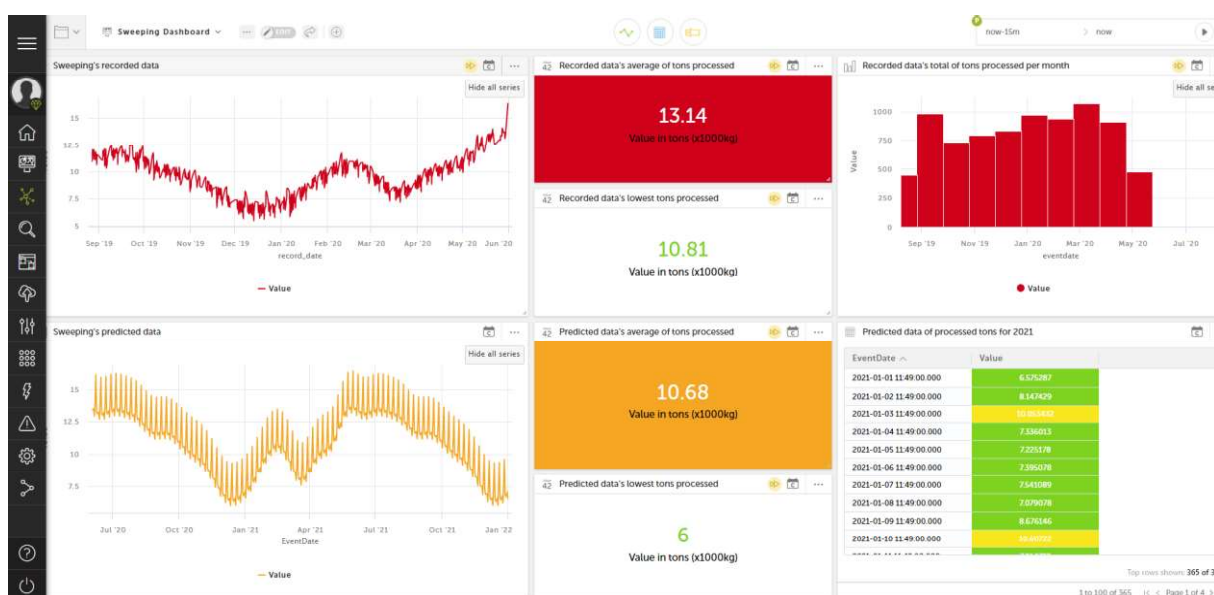


Figura 8-12: Dashboard do resíduo Sweeping (Retirado de: <https://alticelabs.devo.com/welcome#/dashboard/1617985> a 28 de Agosto de 2020)

Saltando para o **YARD TRIMMING**, tendo em atenção o gráfico que contém os dados registados pela empresa de recolha no último ano, conseguem-se destacar dois períodos no fluxo de quantidade de lixo processada por dia em que a quantidade de lixo processada diariamente atinge os seus valores mais baixos, o primeiro sendo entre finais de Setembro de 2019 e início de Outubro de 2019, e o segundo compreendido entre finais de Outubro de 2019 e meados de Novembro de 2019.

Virando para a média de toneladas processada pela entidade por dia, o valor registado é de 11.01 toneladas, onde o valor mínimo é de 8.61 toneladas.

No caso do gráfico que contém o total de toneladas processadas por mês, verifica-se que o mês com o menor valor corresponde a Fevereiro, com o valor de 1046 toneladas. De referir que o mês com o maior valor registado é Março, com 3485 toneladas.

Falando dos *widjets* vocacionados para os valores previstos, consegue-se verificar um pico de toneladas de lixo a serem processadas em meados de Março de 2021. No outro lado do espetro, o intervalo previsto de menor quantidade a ser processada está situado entre finais de Julho de 2021 e finais de Agosto de 2021.

Comparativamente à média de toneladas processada pela entidade, a média calculada pelo algoritmo desce para as 8.28 toneladas, sendo que o potencial valor mínimo previsto é de 4.04 toneladas.

Para concluir, de acordo com o *widjet* dos dados previstos para 2021, é possível verificar que o algoritmo prevê os maiores valores de toneladas processadas para os dias 22 e 23 de Março, sendo estes 12.32 e 12.30 toneladas respetivamente, onde, no caso dos menores valores, os dias correspondem a 8, 15 e 22 de Agosto, traduzindo-se em 4.15, 4.04 e 4.18 toneladas respetivamente.

Na Figura 8-13 pode-se consultar o *dashboard* que contém os dados referentes ao resíduo YARD TRIMMING:

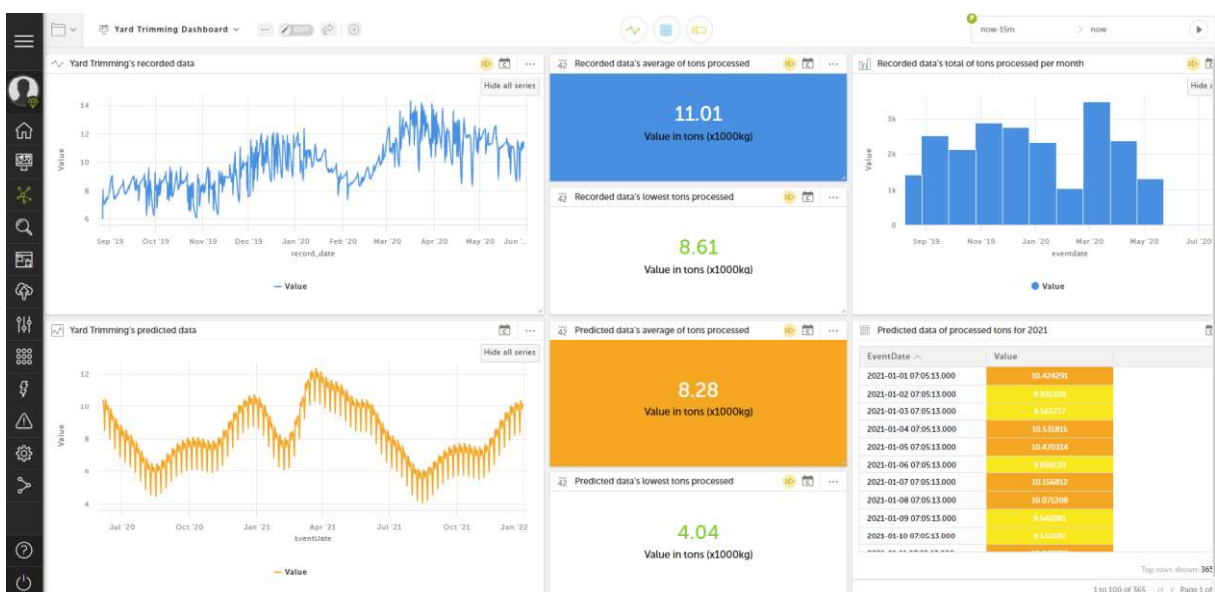


Figura 8-13: Dashboard do resíduo Yard Trimming (Retirado de: <https://alticelabs.devo.com/welcome#/dashboard/1617985> a 28 de Agosto de 2020)

No que toca ao gráfico **BULK**, consultando o gráfico que analisa dados registados pela *Austin Resource Recovery* no último ano, verifica-se que as oscilações do fluxo de quantidade de lixo processada por dia mantém-se minimamente constantes até ser atingido o mês de Março, onde o maior valor registado durante esse período sucedeu-se no dia 21 de Setembro de 2019.

Em termos da média de toneladas processada pela empresa por dia, este valor corresponde a 8.93 toneladas. No caso do valor mínimo registado, este corresponde a 8.4 toneladas.

Focando no gráfico referente ao total por mês de toneladas processadas nos últimos 365 dias, é possível validar que o mês com o maior valor corresponde a Setembro, com um total de 2143 toneladas, onde, contrariamente ao prévio, o mês com o menor valor registado corresponde a Maio, com 116 toneladas.

Analisando os *widgets* que contêm valores previstos, conseguem-se identificar três picos previstos de toneladas de lixo a serem processadas, sendo estes 1 de Novembro de 2020, 7 de Fevereiro de 2021 e 31 de Outubro de 2021.

Segundo o valor da potencial média de toneladas processadas, ao se comparar este com a média mencionada anteriormente, observa-se uma ligeira descida para as 8.7 toneladas, sendo que a descida entre valor mínimo derivado das recolhas e valor mínimo previsto é mais acentuada, correspondendo o último a 7.62 toneladas.

Por fim, no que toca ao *widget* com os dados previstos para o ano 2021, verifica-se que para 31 de Outubro de 2021 (dia correspondente ao maior valor), o registo de toneladas previsto é de 10.23 toneladas. No extremo oposto, o algoritmo prevê os menores valores para os dias 26 de Março e 2 de Abril, com os valores respetivos de 7.62 e 7.61 toneladas.

A figura seguinte, Figura 8-14, contém os *widgets* constituintes do *dashboard* do resíduo **BULK**:

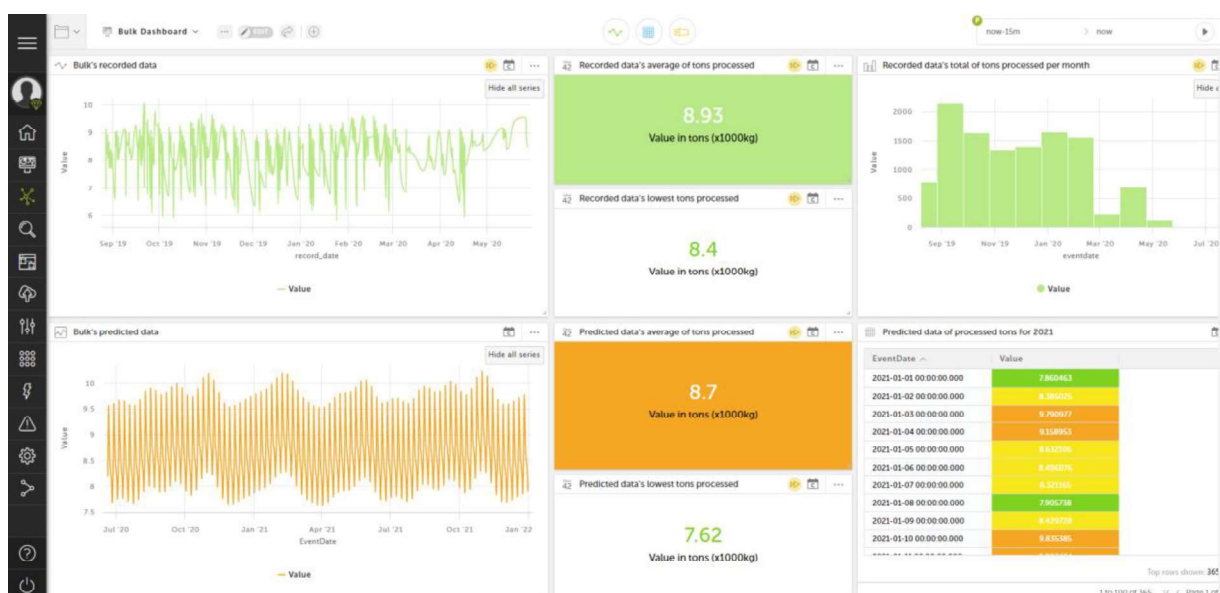


Figura 8-14: Dashboard do resíduo Bulk (Retirado de: <https://alticelabs.devo.com/welcome#/dashboard/1617985> a 28 de Agosto de 2020)

No caso do resíduo **BRUSH**, de acordo com o gráfico correspondente ao *display* dos dados do último ano, conseguem-se identificar seis picos negativos distintos do fluxo apresentado pelo gráfico de linhas em questão. O período de maior atividade que é possível estabelecer situa-se entre meados de Fevereiro de 2020 e finais de Março de 2020.

Contabilizando a média de toneladas processada pela entidade por dia, o valor obtido é de 6.31 toneladas, destacando que, no caso do valor mínimo registado, este assume o valor de 5.1 toneladas.

Tendo em atenção o gráfico responsável por caracterizar o total por mês de toneladas processadas nos últimos 365 dias, obtém-se Setembro como o mês com o maior valor registado, totalizando 1263 toneladas. Revelando uma grande discrepância para o mês anteriormente mencionado, Maio, o mês com o menor registo, totaliza quase 82 toneladas.

Falando dos *widgets* que lidam com os valores de previsão do algoritmo *Prophet*, consegue-se constatar que o algoritmo prevê um intervalo de maiores toneladas de lixo a serem processadas entre inícios de Fevereiro de 2021 e finais de Março de 2021, onde, em contraste, o intervalo previsto de menor quantidade a ser processada está estabelecido entre meados de Julho de 2021 e inícios de Agosto de 2021.

Comparativamente à média de toneladas processada pela entidade, a média calculada pelo algoritmo revela uma descida para as 5.65 toneladas, sendo que o potencial valor mínimo previsto situa-se nas 4.74 toneladas.

De modo a terminar a análise para este resíduo com o *widget* com os dados previstos para o ano 2021, conclui-se que o algoritmo prevê o maior registo de toneladas processadas no dia 14 de Fevereiro, situando-se nas 6.31 toneladas. Contrastando com este registo, o algoritmo prevê os registos mais baixos para os dias 23 e 30 de Julho, com os valores 4.79 e 4.74 toneladas respetivamente.

O *dashboard* e *widgets* associados do resíduo BRUSH pode ser consultado na Figura 8-15:

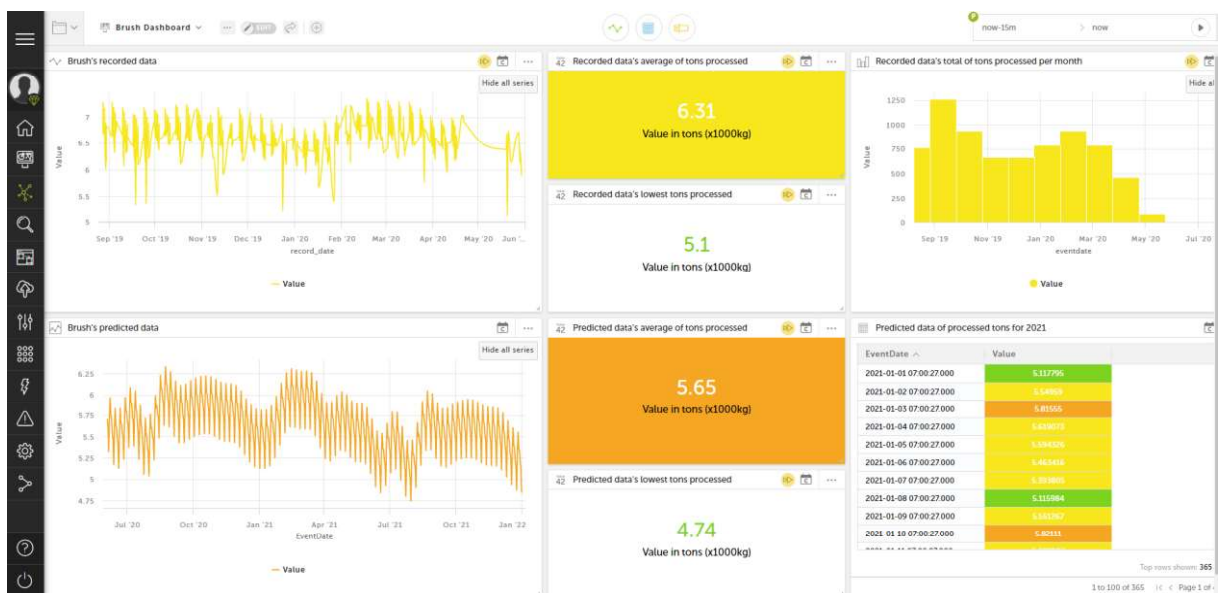


Figura 8-15: Dashboard do resíduo Brush (Retirado de: <https://alticelabs.devo.com/welcome#/dashboard/1617985> a 28 de Agosto de 2020)

De destacar que, nos insights referentes ao resíduo **RECYCLING – PAPER**, devido à existente limitação de mostrar apenas dados do último ano (este resíduo possui dados registados pela companhia até 2008), não foi possível obter todos os infográficos desejados. Esta restrição é imposta nos *widgets* que permitem a obtenção do gráfico que contém os dados registados pela *Austin Resource Recovery* nos últimos 365 dias, assim como o gráfico que contém o total por mês de toneladas processadas no último ano.

No entanto, esta limitação não ocorre para os *widgets* responsáveis pelo cálculo da média e valor mínimo. Referindo a média de toneladas processada pela entidade por dia, este valor corresponde a 2.24 toneladas. No caso do valor mínimo registado, este corresponde a 2.2 toneladas.

Referindo os *widgets* que contêm valores de previsão, é possível verificar que o algoritmo prevê dois picos de toneladas de lixo a serem processadas, sendo que o primeiro encontra-se entre finais de Novembro de 2020 e finais de Dezembro de 2020, e o segundo situa-se entre finais de Novembro de 2021 e finais de Dezembro de 2021. Destaca-se o intervalo previsto de menor quantidade a ser processada entre meados de Junho de 2021 e finais de Julho de 2021. Comparando com o valor da média calculada assente nos valores recolhidos, o valor da potencial média de toneladas processadas desce ligeiramente para as 2.06 toneladas. Todavia, o potencial valor mínimo previsto tem uma notável queda para as 0.53 toneladas.

Para terminar, efetuando uma análise sobre o *widget* com os dados previstos para o ano 2021, conclui-se que o *Prophet* prevê o maior valor de toneladas processadas para o dia 9 de Dezembro, sendo este 3.32 toneladas. Porém, foram previstos os menores valores para os dias 4 e 11 de Julho, com os respetivos registos de 0.53 e 0.57 toneladas.

Na seguinte figura, Figura 8-16, pode-se averiguar o *dashboard* do resíduo RECYCLING - PAPER:

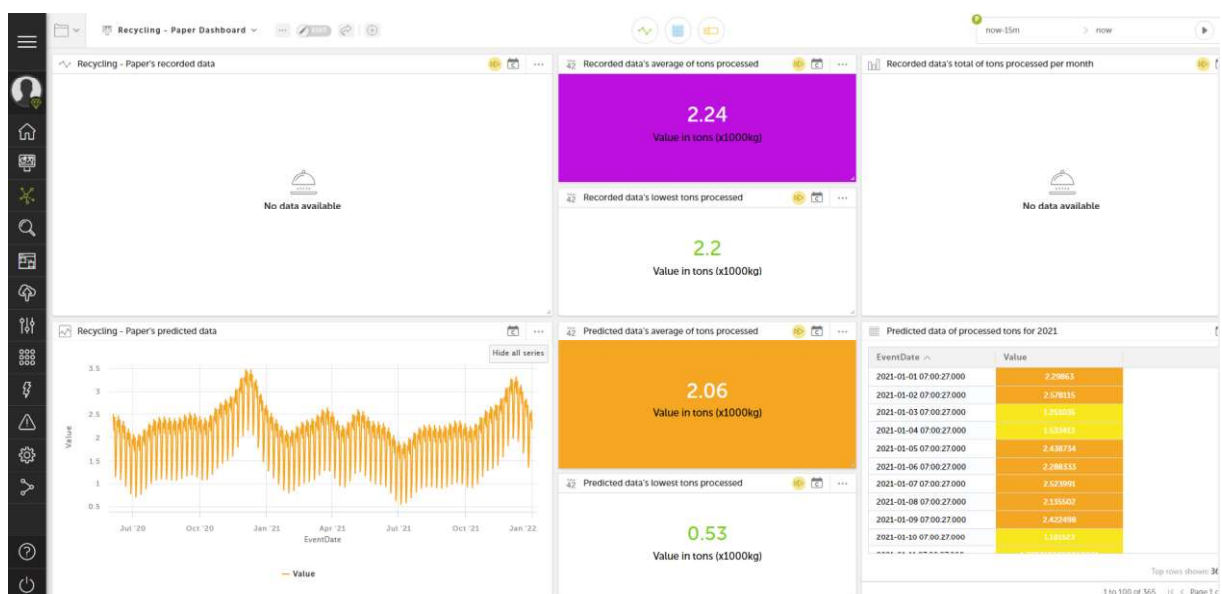


Figura 8-16: Dashboard do resíduo Recycling – Paper (Retirado de: <https://alticelabs.devo.com/welcome#/dashboard/1617985> a 28 de Agosto de 2020)

Repetindo-se o mesmo cenário para o resíduo **RECYCLING – COMINGLE**, não foi possível retirar informação do *widget* associado ao gráfico que contém os dados registados nos últimos 365 dias, e do *widget* responsável pelo gráfico que contém o total por mês de toneladas processadas. Isto prende-se com o facto de os dados estarem inseridos num espaço temporal similar ao resíduo RECYCLING – PAPER.

Focando no *widget* capaz de revelar a média de toneladas processada pela entidade por dia, o valor calculado é de 4.43 toneladas. Falando do valor mínimo registado pela empresa de recolha, este foi de 4.03 toneladas.

Em termos de *widgets* com foco assente nos valores previstos, consegue-se verificar um pico de toneladas de lixo a serem processadas entre finais de Dezembro de 2020 e meados de Janeiro de 2021. De notar que, excetuando o período mencionado, as oscilações do fluxo de quantidade de lixo processada por dia revelam-se constantes.

Comparativamente à média de toneladas processada pela entidade, a média calculada pelo algoritmo desce de forma ligeira para as 4.32 toneladas, onde o potencial valor mínimo previsto é de 3.69 toneladas.

Por fim, tendo em atenção o *widget* dos dados previstos para 2021, é possível verificar que o algoritmo prevê o maior valor de toneladas processadas para o dia 2 de Janeiro, sendo este 5.06 toneladas, onde, no caso do menor valor, o dia correspondente é 3 de Março, com um valor previsto de 3.72 toneladas.

De acordo com a figura seguinte, Figura 8-17, é possível consultar o *dashboard* e os seus *widgets* incorporantes, encarregues de analisar os dados referentes ao tipo RECYCLING - COMINGLE:

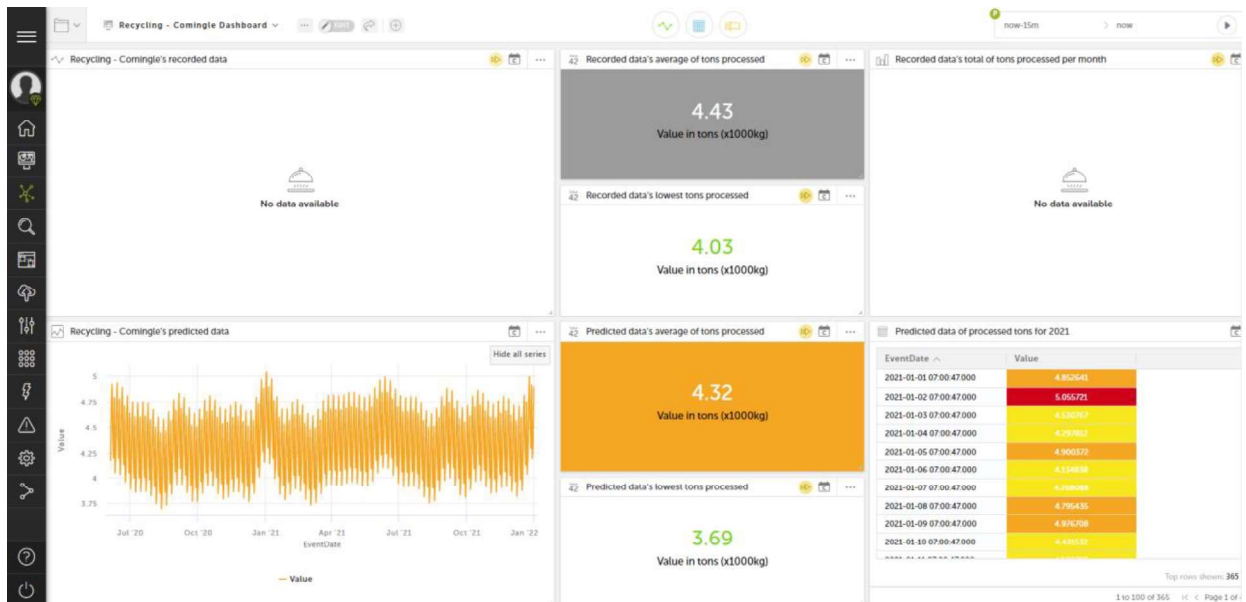


Figura 8-17: Dashboard do resíduo Recycling – Comingle (Retirado de: <https://alticelabs.devo.com/welcome#/dashboard/1617985> a 28 de Agosto de 2020)

No que diz respeito ao resíduo **ORGANICS**, tendo em conta o gráfico interligado aos dados registados pela *Austin Resource Recovery* no último ano, destaca-se de forma imediata do fluxo de dados representado o registo do dia 30 de Março de 2019, com um valor associado de 20.11 toneladas.

Contabilizando a média de toneladas processada por dia, o valor registado é 11.52 toneladas. No caso do valor mínimo registado, este situa-se nas 9.75 toneladas de lixo orgânico.

Para o gráfico que contém o total de toneladas processadas por mês, verifica-se que o mês com o menor valor corresponde a Agosto, com 518 toneladas. No entanto, e revelando uma enorme discrepância comparativamente ao mês mencionado, o mês com o maior valor registado é Março, com 11 mil e 919 toneladas, correspondendo a 11 mil e 401 toneladas de diferença entre os dois meses.

Falando dos *widgets* vocacionados para os valores previstos, consegue-se verificar um pico de toneladas de lixo a serem processadas entre finais de Março de 2021 e meados de Abril de 2021. No outro lado do espectro, destacam-se dois intervalos previstos de menor quantidade a ser processada, sendo que o primeiro encontra-se situado entre inícios de Agosto de 2020 e inícios de Setembro de 2020, e o segundo situa-se entre inícios de Agosto de 2021 e inícios de Setembro de 2021.

Comparativamente à média de toneladas processada pela entidade, a média calculada pelo algoritmo desce para as 10.26 toneladas, sendo que o potencial valor mínimo previsto, quando comparado com o mínimo registado pela entidade, desce abruptamente para as 4.79 toneladas. Para concluir, de acordo com o *widget* dos dados previstos para o ano de 2021, é possível verificar que o algoritmo prevê os maiores valores de toneladas processadas para os dias 29 de Março e 5 de Abril, sendo estes, respetivamente, 15.54 e 15.46 toneladas, onde, no caso dos menores valores, os dias correspondem a 22 e 29 de Agosto, traduzindo-se em 4.91 e 4.90 toneladas respetivamente.

O *dashboard* correspondente ao tipo de resíduo ORGANICS é consultável na Figura 8-18:

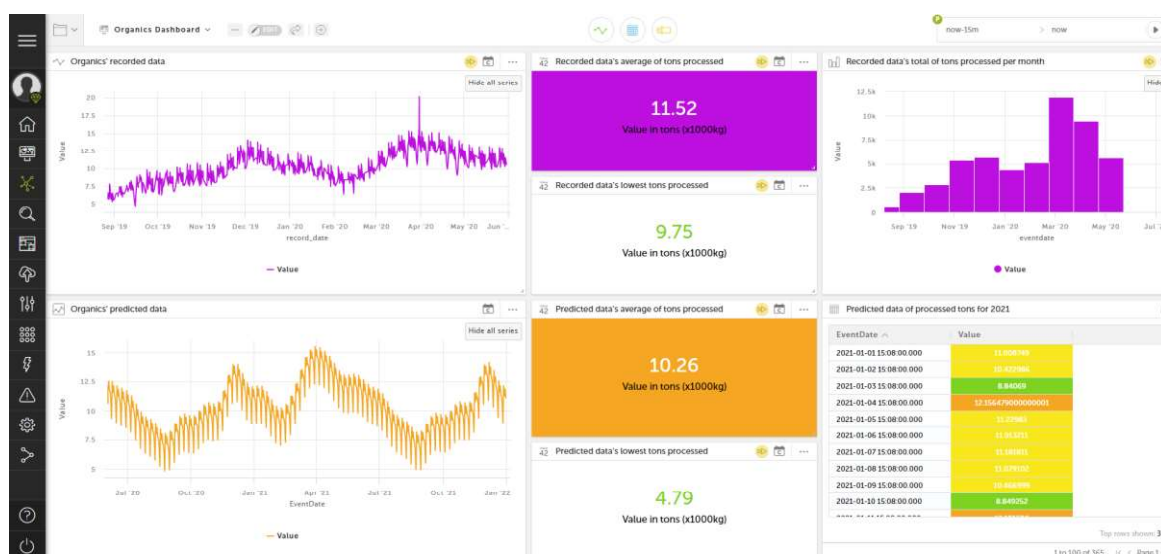


Figura 8-18: Dashboard do resíduo Organics (Retirado de: <https://alticelabs.devo.com/welcome#/dashboard/1617985> a 28 de Agosto de 2020)

Virando foco para o resíduo **MIXED LITTER**, consultando o gráfico correspondente ao *display* dos últimos 365 dias de dados, é possível verificar a existência de um nível elevado de discrepância nas oscilações do gráfico. Isto é justificado com a reduzida existência de registos, comparativamente a outros tipos de resíduos, no período de tempo a ser analisado. Em termos de período onde se registam os valores mais elevados, este situa-se entre os dias 2 e 3 de Setembro de 2019. Por sua vez, os registos com valores mais baixos correspondem aos dias 21 de Novembro de 2019 e 29 de Maio de 2020.

Transitando para a média de toneladas processada por dia, o valor é de 3.22 toneladas, sendo que, para o caso do valor mínimo registado, este corresponde a 2.07 toneladas.

No caso do gráfico do total por mês de toneladas processadas no último ano, registou-se o mês de Janeiro como o mês com o maior valor, totalizando 54.01 toneladas. Inversamente, o mês com o menor valor registado corresponde a Agosto, com 35.62 toneladas.

Falando dos *widgets* que lidam com os valores de previsão, consegue-se constatar a existência de vários picos de potencial recolha de lixo, válido tanto para alturas com maior movimentação, como para menor movimentação. Porém, a similaridade partilhada pelos picos de alta e baixa movimentação é, de acordo com o algoritmo de *Machine Learning*, a ocorrência destes picos durante a estação do ano do Outono e Inverno.

Falando do valor da potencial média de toneladas processadas, após comparação com a média dos dados processados pela empresa, prevê-se um acréscimo ligeiro para as 3.38 toneladas, todavia, o potencial valor mínimo previsto desce, transitando para as 1.57 toneladas.

Encerrando a análise com o *widget* com os dados previstos para o ano 2021, conclui-se que o algoritmo prevê o maior valor de toneladas processadas para o dia 18 de Janeiro, sendo este 4.91 toneladas. Em sentido inverso, o algoritmo prevê os registos mais baixos para os dias 11 de Novembro e 23 de Dezembro, com os valores de 1.66 e 1.70 toneladas respetivamente.

Na figura seguinte, Figura 8-19, pode-se analisar o *dashboard* do resíduo MIXED LITTER:

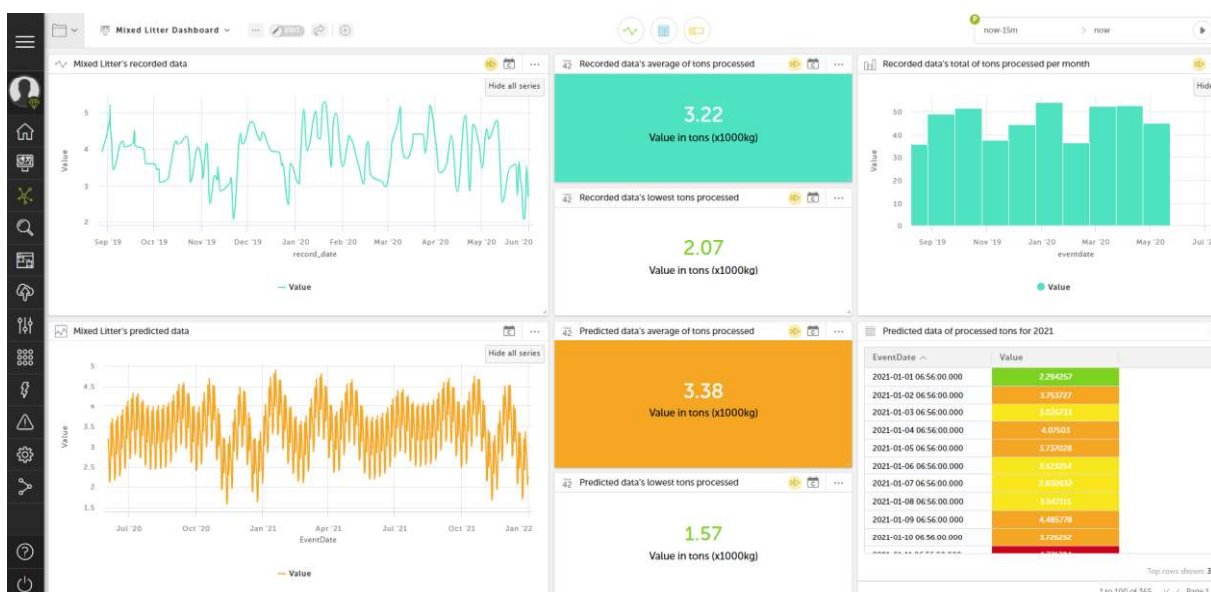


Figura 8-19: Dashboard do resíduo Mixed Litter (Retirado de: <https://alticelabs.devo.com/welcome#/dashboard/1617985> a 28 de Agosto de 2020)

Concentrando agora no resíduo **RECYCLED METAL**, de acordo com o gráfico correspondente ao *display* dos dados do último ano, verifica-se que a quantidade de registos existente para este tipo de resíduo é, comparativamente aos restantes, bastante limitada. Contudo, é possível identificar dois picos de processamento deste resíduo, correspondendo estes aos dias 13 de Janeiro de 2020 e 30 de Maio de 2020.

Contabilizando a média de toneladas processada pela *Austin Resource Recovery* por dia, o valor obtido é de 7.26 toneladas, destacando que, no caso do valor mínimo registado, este assume o valor de 5.66 toneladas.

Tendo em atenção o gráfico responsável por caracterizar o total por mês de toneladas processadas nos últimos 365 dias, obtém-se Maio como o mês com o maior valor registado, totalizando 29.05 toneladas. Mais ainda, o *widget* permite verificar que, para os meses de Dezembro e Abril, não foi recolhida qualquer quantia do tipo metal.

Falando dos *widgets* que lidam com os valores de previsão do algoritmo *Prophet*, consegue-se constatar que o algoritmo prevê um intervalo de maiores toneladas de lixo a serem processadas referente à primeira quinzena de Janeiro de 2021, contrastando com a potencial estabilidade nas oscilações do gráfico apresentado.

Comparativamente à média de toneladas processada pela companhia, a média calculada pelo algoritmo revela uma descida não acentuada para as 6.82 toneladas, sendo que o potencial valor mínimo previsto situa-se nas 4.03 toneladas.

Para concluir, recorrendo ao *widget* com os dados previstos para o ano 2021, conclui-se que o algoritmo prevê o maior registo de toneladas processadas para o dia 9 de Janeiro, correspondendo este às 14.32 toneladas. Contrariamente a este registo, o algoritmo prevê o registo mais baixo para o dia 22 de Julho, com o valor associado de 4.14 toneladas.

Provisionando a Figura 8-20, consegue-se consultar o *dashboard* e respetivos *widgets*, encarregues de analisar os dados referentes ao tipo RECYCLED METAL:

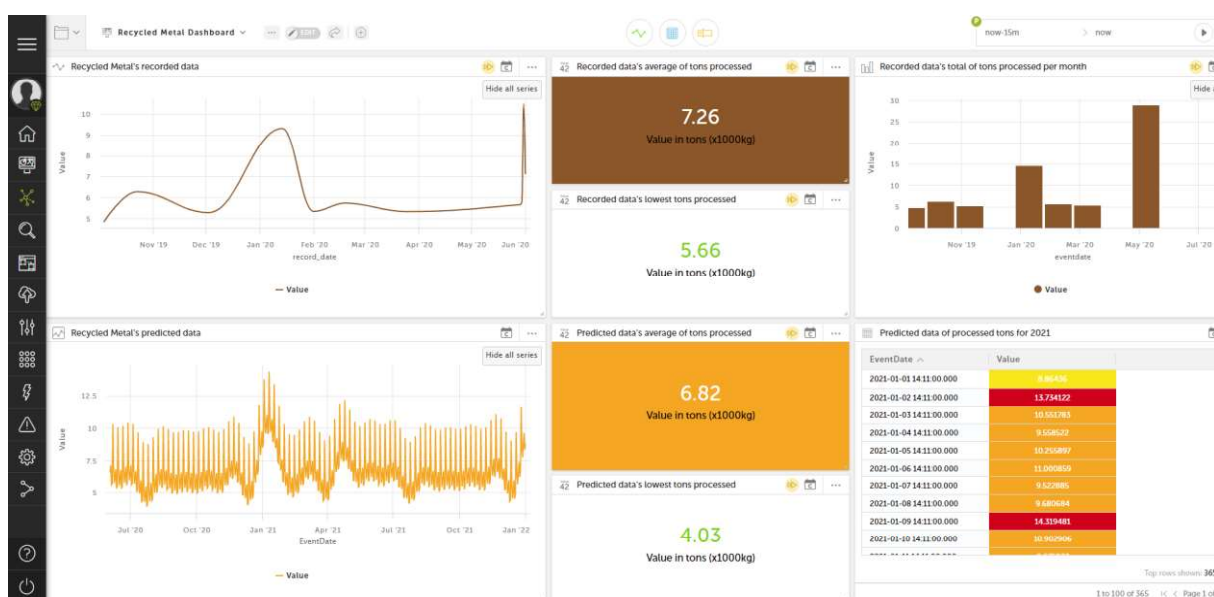


Figura 8-20: Dashboard do resíduo Recycled Metal (Retirado de: <https://alticelabs.devo.com/welcome#/dashboard/1617985> a 28 de Agosto de 2020)

Termina-se a análise aos *insights* dos dados com o resíduo **RECYCLING – PLASTIC BAGS**. Tendo em atenção o gráfico que contém os dados registados pela entidade dedicada à recolha dos resíduos no último ano, consegue-se identificar de forma intuitiva o padrão de crescimento constante no que toca ao processamento de sacos de plástico.

Virando para a média de sacos de plástico processados pela empresa por dia, o valor registado é de, aproximadamente, 5591 sacos de plástico. A nível de mínimo registado, este traduz-se em, aproximadamente, 5316 sacos de plástico.

Para o caso do gráfico que contém o total de sacos processados por mês, verifica-se que o mês com o menor valor corresponde a Agosto, totalizando cerca de 73228 sacos de plástico. Contrariamente, o mês com o maior valor registado é Abril, com cerca de 165601 sacos de plástico.

Mencionando os *widgets* vocacionados para os valores previstos, consegue-se verificar, de acordo com o algoritmo, a continuação do crescimento constante do número de sacos a serem processados por dia. Isto traduz-se numa média diária de sacos potencialmente processados pela entidade de cerca de 6003 sacos de plástico, cifrando o potencial valor mínimo na ordem dos 5435 sacos de plástico.

Para terminar, analisando o *widget* dos dados previstos para o ano de 2021, é possível verificar que o *Prophet* prevê os maiores valores de sacos de plástico processados para os dias 25 e 26 de Dezembro, sendo estes 6614 e 6615 sacos de plástico respetivamente. Contudo, no caso dos menores valores previstos, os dias potencialmente mais calmos correspondem a 6 e 13 de Janeiro, traduzindo-se em 5626 e 5635 sacos de plástico respetivamente.

Na Figura 8-21 é possível consultar o *dashboard* que contém *widgets* que analisam os dados do resíduo RECYCLING – PLASTIC BAGS:

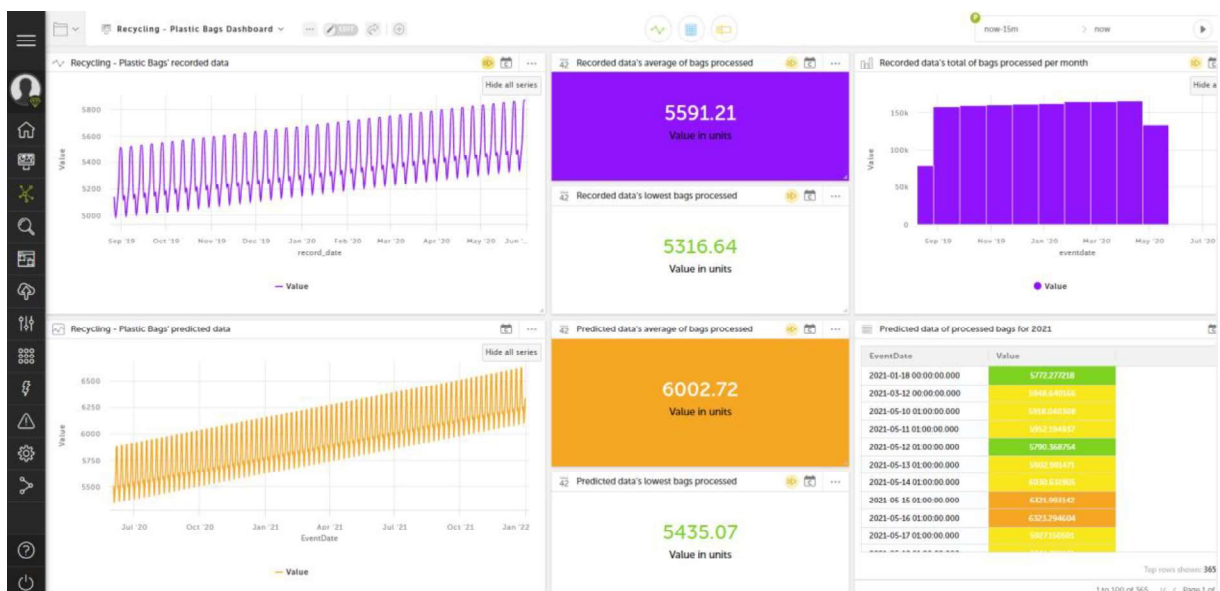


Figura 8-21: Dashboard do resíduo Recycling – Plastic Bags (Retirado de: <https://alticelabs.devo.com/welcome#/dashboard/1617985> a 28 de Agosto de 2020)

## 9. Conclusões

A transformação de uma cidade numa *Smart City* é possível através da dinamização do sector dos negócios, nomeadamente, por parte de empresas cuja área de negócio se prenda com as tecnologias e/ou infraestruturas. Como tal, considera-se como necessário o fornecimento de soluções, por parte destas empresas, que potenciem uma gestão mais eficaz dos recursos públicos de uma cidade, reduzindo os custos operacionais de administrações públicas.

Atualmente, um dos projetos internacionais a potencializar o investimento em empresas tecnológicas tem o nome de “Aveiro STEAM City”. Este tem como principal objetivo a utilização de tecnologias no ramo da *Internet of Things* - IoT que possibilitam a visualização, conhecimento e atuação nas diferentes infraestruturas associadas aos recursos públicos de uma cidade.

Utilizando o projeto descrito como inspiração, o trabalho desenvolvido consistiu na criação de um sistema capaz de interpretar um conjunto de dados providenciado por uma companhia responsável pelo processamento dos resíduos de uma cidade, aplicando-se as implementações necessárias para se organizar os dados na maneira mais eficiente possível, com o propósito de serem aplicados algoritmos de *Machine Learning* (ML) nos dados. Os dados enriquecidos com ML foram posteriormente incluídos numa plataforma de gestão de dados, com o propósito de se retirar valor dos dados, recorrendo a ferramentas na área de *Business Intelligence* (BI).

Para se averiguar qual o algoritmo de ML com melhor capacidade de enriquecer o(s) conjunto(s) de dados com valores de previsão, foi efetuada uma comparação entre algoritmos dentro do tipo de ML denominado *Supervised Learning*, sendo estes compostos por *AutoRegressive Integrated Moving Average – ARIMA*, *Decision Trees*, *Artificial Neural Networks*, *Support Vectoring Model – SVM* e *Prophet*.

Após a conclusão desta etapa, foi criada a devida componente que permite o enriquecimento dos conjuntos de dados com valores de previsão baseados nos atuais. Esta componente recorreu a bibliotecas inerentes ao *Flask*, para o efeito de serem utilizadas para a componente de

pré-processamento do conjunto de dados original, com a integração do algoritmo *Prophet*, capaz de calcular os dados num espaço temporal superior ao atual.

A etapa final consistiu na inclusão dos novos conjuntos de dados na plataforma de gestão de dados, detida pela Altice Labs, com a posterior aplicação da componente de BI sobre os dados, potencializada pela mesma. A componente de BI é caracterizada por permitir a criação de *dashboards* que contêm informação sintetizada em determinados tipos de *widgets*, com o propósito de facilitar a identificação de padrões e informações relevantes para qualquer tipo de utilizador.

No que toca ao alinhamento com os requisitos inicialmente, pode-se considerar que o trabalho realizado atingiu os requisitos impostos. O trabalho executado no decorrer da tese permitiu a obtenção de informação que possibilita aos vários parceiros da empresa otimizar determinados aspetos da gestão operacional dos processos a nível dos resíduos. Em suma, o trabalho realizado permite apoiar, com sucesso, o incremento positivo no que toca ao processamento/tratamento dos resíduos municipais.

Com o auxílio da tecnologia de ML implementada, o cliente final, mais concretamente o município, ao consultar os gráficos referentes aos valores de previsão, consegue antecipar em que intervalos de tempo poderão existir maiores e/ou menores níveis de atividade de resíduos processados, efetivamente otimizando os custos inerentes à recolha do lixo dos contentores da cidade.

Este trabalho, ao utilizar este género de plataformas e tecnologias, pretende igualmente apelar para a enorme potencialidade que existe em retirar informação extremamente relevante dos dados sobre as diferentes métricas existentes numa cidade, não apenas de dados associados a resíduos. Trabalhos realizados nesta vertente permitem um crescimento económico e de qualidade de vida dos cidadãos, graças a uma gestão inteligente dos recursos da cidade, possibilitada por este género de plataformas.

Como estabelecido no Estado da Arte deste trabalho, o objetivo final de uma cidade que pretenda integrar o lote de *Smart Cities* passa por utilizar os recursos públicos de uma forma eficiente, efetivamente permitindo aos cidadãos uma maior qualidade nos serviços em catálogo. Considera-se, com isto, que o trabalho desenvolvido permite a uma cidade atingir o paradigma de *Smart City*.

A nível de objetivos pessoais, é possível reportar que as expectativas foram mais que satisfeitas. Devido à consciencialização obtida, com o decorrer deste trabalho, sobre a importância das funções que um *Data Scientist* possui, considera-se que não se teria entendido de uma melhor forma o significado de desempenhar as funções de um cientista de dados, caso não se tivesse levado a cabo um trabalho nesta vertente e dimensão.

Um dos aspetos a retirar após a conclusão do trabalho prende-se com a experiência adquirida com as áreas de ML e BI. Com uma aposta cada vez maior de entidades públicas e privadas neste género de sistemas, existe a convicção de que a realização deste trabalho irá representar-se na forma de uma mais-valia futura, a nível profissional.

## 9.1 Desafios encontrados

No que toca a desafios a ultrapassar originados pelo projeto, destaca-se o paralelismo que foi necessário fazer entre os resíduos processados pela entidade residente nos Estados Unidos da América, e os resíduos cujo processamento fosse registado em Portugal. Este desafio é justificado pela necessidade de existir um cenário de estudo enquadrado o mais possível com a realidade portuguesa, devido aos trabalhos que estão a ser realizados em diferentes cidades. Como existem dados que não são do interesse do projeto mencionado, considerou-se que não compensaria dedicar mais tempo e recursos ao estudo destes dados. Para o efeito, foi realizado um estudo intensivo dentro dos devidos portais certificados a nível nacional e internacional, de modo a se comprovar o interesse existente na utilização do conjunto de dados alvo do cenário.

Outro desafio revelou-se na fase de comparação dos algoritmos de *Machine Learning* do estudo. Inicialmente estava previsto a utilização de um *setup* para a comparação. Porém, certas limitações não contabilizadas foram encontradas, o que levou a uma pesquisa mais intensiva das alternativas existentes, com a devida implementação a ser efetuada posteriormente (mais informações sobre estes contrastes podem ser consultadas no Apêndice 2).

O último desafio a nível técnico surgiu com a aplicação de *Business Intelligence* nos dados dos resíduos, onde foi necessário realizar experimentações com os vários *widgets* providenciados pela plataforma de gestão, de forma a se atingir um consenso relativamente ao equilíbrio entre informações pertinentes a serem reveladas pelos gráficos, e facilidade de interpretação por parte do utilizador final dos *widgets* constituintes dos *dashboards* BI criados. Este desafio requereu um estudo mais aprofundado da documentação providenciada pela plataforma da empresa, assim como um estudo das limitações existentes nas funcionalidades de interesse.

Por fim, certamente o fator mais inesperado, o desafio imposto pela pandemia causada pelo COVID-19. Este justifica-se devido ao período de incerteza existente durante os momentos iniciais do vírus propagado mundialmente, quer em termos da continuidade do projeto da tese proposto, como da conferência CAPSI2020 onde um artigo baseado na tese foi aceite, entre outros fatores distintos.

A pandemia que se propagou em paralelismo com os desenvolvimentos do trabalho a implementar constituiu um desafio de compreensão em múltiplas dimensões como, por exemplo, a saúde física, saúde psicológica e os esforços económicos. Considera-se que foi necessário realizar um esforço adicional, tanto individual, como por parte dos orientadores, para se prosseguirem as etapas do trabalho a realizar dentro dos prazos estabelecidos. Para o efeito, recorreu-se a novas dinâmicas organizacionais, com vista à entrega do mesmo dentro dos prazos estabelecidos.

## 9.2 Trabalho futuro

Como trabalho futuro, propõe-se a aplicação do sistema criado a outros conjuntos de dados dentro do âmbito das *Smart Cities*, como, por exemplo, o tratamento de águas residuais, os consumos de eletricidade, e os gases poluentes. Esta sugestão visa a melhoria da qualidade de vida dos cidadãos, assim como a redução de custos associados às diferentes métricas presentes numa cidade.

Mais focada em *Machine Learning* (ML), esta proposta sugere a continuação do estudo e aplicações de algoritmos dentro do tipo *Supervised Learning* como, por exemplo, o *Naive Bayes* e o *K-Nearest Neighbor* – KNN, comparando os resultados obtidos por estes no que toca à previsão de novos valores baseados no conjunto de dados já utilizado.

Sugere-se, de igual forma, a criação de novos cenários recorrendo a dados recolhidos em tempo real, com sensores de parceiros que já se encontrem instalados pelas diversas zonas das cidades. Estes sensores, contrariamente aos conjuntos de dados disponibilizados por parceiros, têm a possibilidade de servir de fonte direta ao sistema desenvolvido e/ou à plataforma de gestão de dados utilizada.

Como sugestão para a área de *Business Intelligence* (BI), aconselha-se a elaboração de uma pesquisa intensiva sobre alternativas BI existentes no mercado, com o propósito de se investigar que ferramentas estas possuem, de modo a se retornar o máximo de valor sobre os dados analisados.

Em termos académicos, sugere-se a criação de um segundo artigo científico, baseado na comparação entre os algoritmos de ML utilizados neste projeto, dado que o primeiro artigo foi elaborado com foco na aplicação da ferramenta *Flask* ao conjunto de dados da *Austin Resource Recovery*.

## REFERÊNCIAS

- Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications*, 6(1), 25.
- Al-Jarrah, O. Y., Yoo, P. D., Muhaidat, S., Karagiannidis, G. K., & Taha, K. (2015). Efficient machine learning for big data: A review. *Big Data Research*, 2(3), 87-93.
- Al-Ali, A. R., Zualkernan, I. A., Rashid, M., Gupta, R., & Alikarar, M. (2017). A smart home energy management system using IoT and big data analytics approach. *IEEE Transactions on Consumer Electronics*, 63(4), 426-434.
- Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.
- Anagnostopoulos, T., Zaslavsky, A., & Medvedev, A. (2015, April). Robust waste collection exploiting cost efficiency of IoT potentiality in Smart Cities. In *2015 International conference on recent advances in internet of things (RIoT)* (pp. 1-6). IEEE.
- Anthopoulos, L. (2015, August). Defining smart city architecture for sustainability. In *proceedings of 14th electronic government and 7th electronic participation conference (IFIP2015)* (pp. 140-147).
- Aslam, F. A., Mohammed, H. N., Mohd, J. M., Gulamgaus, M. A., & Lok, P. S. (2015). Efficient Way Of Web Development Using Python And Flask. *International Journal of Advanced Research in Computer Science*, 6(2).
- Ayodele, T. O. (2010). Types of machine learning algorithms. *New advances in machine learning*, 3, 19-48.
- Babu, K. V. S. N. (2012). Business intelligence: Concepts, components, techniques and benefits. *Components, Techniques and Benefits (September 22, 2012)*.
- Baby, C. J., Singh, H., Srivastava, A., Dhawan, R., & Mahalakshmi, P. (2017, March). Smart bin: An intelligent waste alert and prediction system using machine learning approach. In *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)* (pp. 771-774). IEEE.
- Balakrishna, C. (2012, September). Enabling technologies for smart city services and applications. In *2012 sixth international conference on next generation mobile applications, services and technologies* (pp. 223-227). IEEE.
- Batty, M., Axhausen, K. W., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., ... & Portugali, Y. (2012). Smart cities of the future. *The European Physical Journal Special Topics*, 214(1), 481-518.
- Batty, M. (2013). Big data, smart cities and city planning. *Dialogues in Human Geography*, 3(3), 274-279.
- Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. *Jama*, 319(13), 1317-1318.
- Caragliu, A., Del Bo, C., & Nijkamp, P. (2011). Smart cities in Europe. *Journal of urban technology*, 18(2), 65-82.

- Caruana, R., & Niculescu-Mizil, A. (2006, June). An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning* (pp. 161-168).
- Chourabi, H., Nam, T., Walker, S., Gil-Garcia, J. R., Mellouli, S., Nahon, K., ... & Scholl, H. J. (2012, January). Understanding smart cities: An integrative framework. In *2012 45th Hawaii international conference on system sciences* (pp. 2289-2297). IEEE.
- Chui, M., Löffler, M., & Roberts, R. (2010). The internet of things. *McKinsey Quarterly*, 2(2010), 1-9.
- Davenport, T. H., Barth, P., & Bean, R. (2012). *How'big data'is different*. MIT Sloan Management Review.
- Dietterich, T. G. (1997). Machine-learning research. *AI magazine*, 18(4), 97-97.
- Dong, J. X., Krzyżak, A., & Suen, C. (2003, July). A fast parallel optimization for training support vector machine. In *International Workshop on Machine Learning and Data Mining in Pattern Recognition* (pp. 96-105). Springer, Berlin, Heidelberg.
- Folianto, F., Low, Y. S., & Yeow, W. L. (2015, April). Smartbin: Smart waste management system. In *2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)* (pp. 1-2). IEEE.
- Gangadharan, G. R., & Swami, S. N. (2004, June). Business intelligence systems: design and implementation strategies. In *26th International Conference on Information Technology Interfaces, 2004*. (pp. 139-144). IEEE.
- Gaur, A., Scotney, B., Parr, G., & McClean, S. (2015). Smart city architecture and its applications based on IoT. *Procedia computer science*, 52, 1089-1094.
- George, G., Haas, M. R., & Pentland, A. (2014). Big data and management.
- Goodwin, P., & Lawton, R. (1999). On the asymmetry of the symmetric MAPE. *International journal of forecasting*, 15(4), 405-408.
- Gutierrez, J. M., Jensen, M., Henius, M., & Riaz, T. (2015). Smart waste collection system based on location intelligence. *Procedia Computer Science*, 61, 120-127.
- Han, J., Kamber, M., & Pei, J. (2011). Data mining concepts and techniques third edition. *The Morgan Kaufmann Series in Data Management Systems*, 83-124.
- Hollands, R. G. (2008). Will the real smart city please stand up? Intelligent, progressive or entrepreneurial?. *City*, 12(3), 303-320.
- Hočevar, B., & Jaklič, J. (2010). Assessing benefits of business intelligence systems—a case study. *Management: journal of contemporary management issues*, 15(1), 87-119.
- Höjer, M., & Wangel, J. (2015). Smart sustainable cities: definition and challenges. In *ICT innovations for sustainability* (pp. 333-349). Springer, Cham.
- Hyndman, R. J. (2006). Another look at forecast-accuracy metrics for intermittent demand. *Foresight: The International Journal of Applied Forecasting*, 4(4), 43-46.
- Kannangara, M., Dua, R., Ahmadi, L., & Bensebaa, F. (2018). Modeling and prediction of regional municipal solid waste generation and diversion in Canada using machine learning approaches. *Waste management*, 74, 3-15.
- Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal*, 79(1), 1-14.
- Kumar, T. V., & Dahiya, B. (2017). Smart economy in smart cities. In *Smart Economy in Smart Cities* (pp. 3-76). Springer, Singapore.

## REFERÊNCIAS

- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444.
- Li, L., Ma, Z., Liu, L., & Fan, Y. (2013). Hadoop-based ARIMA algorithm and its application in weather forecast. *International Journal of Database Theory and Application*, 6(5), 119-132.
- Madden, S. (2012). From databases to big data. *IEEE Internet Computing*, 16(3), 4-6.
- Mining, W. I. D. (2006). Data mining: Concepts and techniques. *Morgan Kaufmann*.
- Mohanty, S. P., Choppali, U., & Kougianos, E. (2016). Everything you wanted to know about smart cities: The internet of things is the backbone. *IEEE Consumer Electronics Magazine*, 5(3), 60-70.
- Moss, L. T., & Atre, S. (2003). *Business intelligence roadmap: the complete project lifecycle for decision-support applications*. Addison-Wesley Professional.
- Olatunji, S. O., Selamat, A., & Raheem, A. A. A. (2011). Predicting correlations properties of crude oil systems using type-2 fuzzy logic systems. *Expert Systems with Applications*, 38(9), 10911-10922.
- Pang, Z., Chen, Q., Han, W., & Zheng, L. (2015). Value-centric design of the internet-of-things solution for food supply chain: Value creation, sensor portfolio and information fusion. *Information Systems Frontiers*, 17(2), 289-319.
- Papacharalampous, G. A., & Tyrallis, H. (2018). Evaluation of random forests and Prophet for daily streamflow forecasting. *Advances in Geosciences*, 45.
- Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big data*, 1(1), 51-59.
- Rathore, M. M., Ahmad, A., Paul, A., & Rho, S. (2016). Urban planning and building smart cities based on the internet of things using big data analytics. *Computer Networks*, 101, 63-80.
- Rostami, A., & Baghban, A. (2018). Application of a supervised learning machine for accurate prognostication of higher heating values of solid wastes. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 40(5), 558-564.
- Russom, P. (2011). Big data analytics. *TDWI best practices report, fourth quarter*, 19(4), 1-34.
- Sagiroglu, S., & Sinanc, D. (2013, May). Big data: A review. In *2013 International Conference on Collaboration Technologies and Systems (CTS)* (pp. 42-47). IEEE.
- Santos, B., Sérgio, F., Abrantes, S., Sá, F., Loureiro, J., Wanzeler, C. and Martins, P. (2019). Open Source Business Intelligence Tools: Metabase and Redash. In *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management - Volume 1: KDIR*, ISBN 978-989-758-382-7, pages 467-474.
- Santos, M. Y., & Ramos, I. (2006). *Business Intelligence: Tecnologias da informação na gestão de conhecimento*. FCA-Editora de Informática, Lda.
- Sharma, N., Singha, N., & Dutta, T. (2015). Smart bin implementation for smart cities. *International Journal of Scientific & Engineering Research*, 6(9), 787-791.
- Shen, C. W., Lee, H. C., Chou, C. C., & Cheng, C. C. (2011). Data mining the data processing technologies for inventory management. *Journal of computers*, 6, 784-791.

- Tansley, S., & Tolle, K. M. (2009). *The fourth paradigm: data-intensive scientific discovery* (Vol. 1). A. J. Hey (Ed.). Redmond, WA: Microsoft research.
- Tharwat, A. (2018). Classification assessment methods. *Applied Computing and Informatics*.
- Torres-García, A., Rodea-Aragón, O., Longoria-Gandara, O., Sánchez-García, F., & González-Jiménez, L. E. (2015). Intelligent waste separator. *Computación y Sistemas*, 19(3), 487-500.
- Van Der Aalst, W. (2016). Data science in action. In *Process Mining* (pp. 3-23). Springer, Berlin, Heidelberg.
- Verma, A., Wei, X., & Kusiak, A. (2013). Predicting the total suspended solids in wastewater: a data-mining approach. *Engineering Applications of Artificial Intelligence*, 26(4), 1366-1372.
- Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *Journal of Business Logistics*, 34(2), 77-84.
- Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (pp. 29-39). Citeseer.
- Wortmann, F., & Flüchter, K. (2015). Internet of things. *Business & Information Systems Engineering*, 57(3), 221-224.
- Xia, F., Yang, L. T., Wang, L., & Vinel, A. (2012). Internet of things. *International Journal of Communication Systems*, 25(9), 1101.
- Xie, H., Ma, F., & Bai, Q. (2009, August). Prediction of indoor air quality using artificial neural networks. In *2009 Fifth International Conference on Natural Computation* (Vol. 2, pp. 414-418). IEEE.
- Yegnanarayana, B. (2009). Artificial Neural Networks. PHI Learning Unip. Ltd.
- Zanella, A., Bui, N., Castellani, A., Vangelista, L., & Zorzi, M. (2014). Internet of things for smart cities. *IEEE Internet of Things journal*, 1(1), 22-32.

# APÊNDICE 1 – FIGURAS DOS TEMPOS DE PROCESSAMENTO DA API

localhost:5000/garbage\_collectio x +

localhost:5000/garbage\_collections

Aplicações myaltice Bem-vindo a Portal... Altice Labs | Experi... Dashboard - Wiki IoT Data Atendimento ao Co... Portal -- pvprod / 1... bruno-g-santos@al... IoT Ser

Success. Total Processing Time of GARBAGE COLLECTIONS data was 0:03:20.471815

```
Anaconda Prompt (anaconda3) - python flask_split.py
A row ['02/20/2020', 'RECYCLING - SINGLE STREAM', '02/21/2020 01:00:00 PM', '9,880'
'BALCONES RECYCLING', 'RECYCLING - SINGLE STREAM', 'RHAU53', 862141] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020', 'RECYCLING - SINGLE STREAM', '02/21/2020 09:00:00 AM'
'11,280', 'BALCONES RECYCLING', 'RECYCLING - SINGLE STREAM', 'RFAU53', 862270] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020', 'RECYCLING - SINGLE STREAM', '02/21/2020 01:00:00 PM', '7,040'
'TDS - MRF', 'RECYCLING - SINGLE STREAM', 'RHAU50', 862122] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/22/2020', 'ORGANICS', '02/22/2020 10:00:00 AM', '500', 'ORGANICS BY GOSH'
'YARD TRIMMINGS-ORGANICS', 'OBS00', 862272] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020', 'RECYCLING - SINGLE STREAM', '02/21/2020 10:00:00 AM'
'13,840', 'BALCONES RECYCLING', 'RECYCLING - SINGLE STREAM', 'RHAU51', 862124] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020', 'SWEEPING', '02/21/2020 12:00:00 AM', nan, 'ONION CREEK'
'STREET CLEANING-DOWNTOWN', 'SDT01', 862254] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020', 'RECYCLING - SINGLE STREAM', '02/21/2020 10:00:00 AM'
'15,460', 'BALCONES RECYCLING', 'RECYCLING - SINGLE STREAM', 'RHAU51', 862127] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020', 'RECYCLING - SINGLE STREAM', '02/21/2020 10:00:00 AM', '9,580'
'BALCONES RECYCLING', 'RECYCLING - SINGLE STREAM', 'RFAU52', 862269] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020', 'RECYCLING - SINGLE STREAM', '02/21/2020 10:00:00 AM'
'11,780', 'BALCONES RECYCLING', 'RECYCLING - SINGLE STREAM', 'RFAU52', 862266] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020', 'DEAD ANIMAL', '02/20/2020 02:12:00 PM', '200', 'TDS LANDFILL'
'DEAD ANIMAL', 'DA01', 862275] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020', 'RECYCLING - SINGLE STREAM', '02/21/2020 01:00:00 PM', '4,400'
'BALCONES RECYCLING', 'RECYCLING - SINGLE STREAM', 'RHAU51', 862128] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020', 'SWEEPING', '02/21/2020 12:00:00 AM', nan, 'ONION CREEK'
'STREET CLEANING-DOWNTOWN', 'SDT01', 862254] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020', 'RECYCLING - SINGLE STREAM', '02/21/2020 12:00:00 AM', '520'
'TDS - MRF', 'KAB', 'KAB06', 862252] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020', 'RECYCLING - SINGLE STREAM', '02/21/2020 10:00:00 AM'
'12,380', 'BALCONES RECYCLING', 'RECYCLING - SINGLE STREAM', 'RFAU51', 862263] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
127.0.0.1 - - [04/Jun/2020 17:38:27] "[37mGET /garbage_collections HTTP/1.1] 200 -
```

Resultado do processamento dos dados do resíduo Garbage Collections

localhost:5000/recycling\_single\_ x +

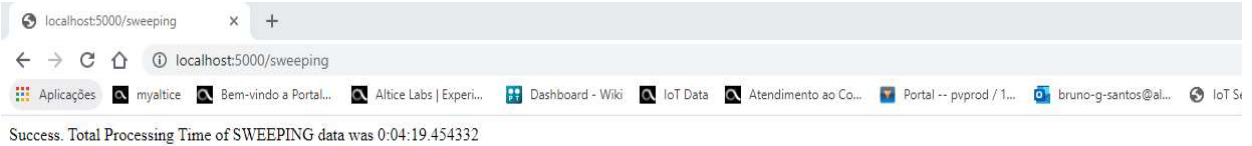
localhost:5000/recycling\_single\_stream

Aplicações myaltice Bem-vindo a Portal... Altice Labs | Experi... Dashboard - Wiki IoT Data Atendimento ao Co... Portal -- pvprod / 1... bruno-g-santos@al... IoT Ser

Success. Total Processing Time of RECYCLING - SINGLE STREAM data was 0:04:13.513558

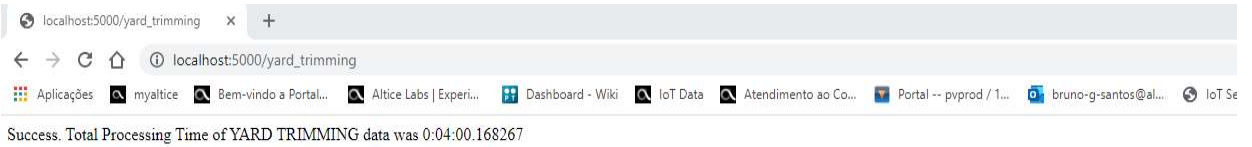
```
Anaconda Prompt (anaconda3) - python flask_split.py
A row ['02/21/2020', 'GARBAGE COLLECTIONS', '02/21/2020 02:45:00 PM', '18,480'
'TDS LANDFILL', 'GARBAGE COLLECTION', 'PAF70', 862260] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020', 'YARD TRIMMING', '02/21/2020 12:55:00 PM', '11,220'
'HORNBY BEND', 'YARD TRIMMINGS', 'YF01', 862273] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/22/2020', 'DEAD ANIMAL', '02/22/2020 01:08:00 PM', '20', 'TDS LANDFILL'
'DEAD ANIMAL', 'DA01', 862276] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020', 'ORGANICS', '02/21/2020 01:20:00 PM', '7,940'
'ORGANICS BY GOSH', 'YARD TRIMMINGS-ORGANICS', 'OF1', 862274] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020', 'SWEEPING', '02/21/2020 12:00:00 AM', nan, 'ONION CREEK'
'STREET CLEANING-BLVD', 'SB46', 862257] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020', 'SWEEPING', '02/21/2020 12:00:00 AM', nan, 'ONION CREEK'
'STREET CLEANING-BLVD', 'SB03', 862256] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020', 'SWEEPING', '02/21/2020 12:00:00 AM', nan, 'ONION CREEK'
'STREET CLEANING-BLVD', 'SB133', 862255] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/22/2020', 'ORGANICS', '02/22/2020 10:00:00 AM', '500', 'ORGANICS BY GOSH'
'YARD TRIMMINGS-ORGANICS', 'OBS00', 862272] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020', 'SWEEPING', '02/21/2020 12:00:00 AM', nan, 'ONION CREEK'
'STREET CLEANING-DOWNTOWN', 'SDT01', 862254] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020', 'DEAD ANIMAL', '02/20/2020 02:12:00 PM', '200', 'TDS LANDFILL'
'DEAD ANIMAL', 'DA01', 862275] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020', 'SWEEPING', '02/21/2020 12:00:00 AM', nan, 'ONION CREEK'
'STREET CLEANING-DOWNTOWN', 'SDT01', 862253] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020', 'GARBAGE COLLECTIONS', '02/21/2020 01:45:00 PM', '23,240'
'TDS LANDFILL', 'GARBAGE COLLECTION', 'PAF72', 862262] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020', 'GARBAGE COLLECTIONS', '02/21/2020 02:30:00 PM', '22,120'
'TDS LANDFILL', 'GARBAGE COLLECTION', 'PAF73', 862264] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020', 'GARBAGE COLLECTIONS', '02/21/2020 02:30:00 PM', '19,420'
'TDS LANDFILL', 'GARBAGE COLLECTION', 'PAF71', 862261] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
127.0.0.1 - - [04/Jun/2020 16:14:30] "[37mGET /recycling_single_stream HTTP/1.1] 200 -
```

Resultado do processamento dos dados do resíduo Recycling – Single Stream



```
Anaconda Prompt (anaconda3) - python flask_split.py
A row ['02/22/2020' 'ORGANICS' '02/22/2020 10:00:00 AM' '500' 'ORGANICS BY GOSH'
'YARD TRIMMINGS-ORGANICS' 'OBS00' 862272] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'13,840' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RHAU51' 862124] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'SWEEPING' '02/21/2020 12:00:00 AM' nan 'ONION CREEK'
'STREET CLEANING-DOWNTOWN' 'SDT01' 862254] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'15,460' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RHAU51' 862127] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM' '9,580'
'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RFAU52' 862269] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'11,780' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RFAU52' 862266] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'DEAD ANIMAL' '02/20/2020 02:12:00 PM' '200' 'TDS LANDFILL'
'DEAD ANIMAL' 'DA01' 862275] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 01:00:00 PM' '4,400'
'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RHAU51' 862128] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'SWEEPING' '02/21/2020 12:00:00 AM' nan 'ONION CREEK'
'STREET CLEANING-DOWNTOWN' 'SDT01' 862253] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'GARBAGE COLLECTIONS' '02/21/2020 01:45:00 PM' '23,240'
'TDS LANDFILL' 'GARBAGE COLLECTION' 'PAF72' 862262] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'GARBAGE COLLECTIONS' '02/21/2020 02:30:00 PM' '22,120'
'TDS LANDFILL' 'GARBAGE COLLECTION' 'PAF73' 862264] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'GARBAGE COLLECTIONS' '02/21/2020 02:30:00 PM' '19,420'
'TDS LANDFILL' 'GARBAGE COLLECTION' 'PAF71' 862261] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 12:00:00 AM' '520'
'TDS - MRF' 'KAB' 'KAB06' 862252] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'12,380' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RFAU51' 862263] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
127.0.0.1 - - [04/Jun/2020 16:20:11] "[37mGET /sweeping HTTP/1.1] 200 -
```

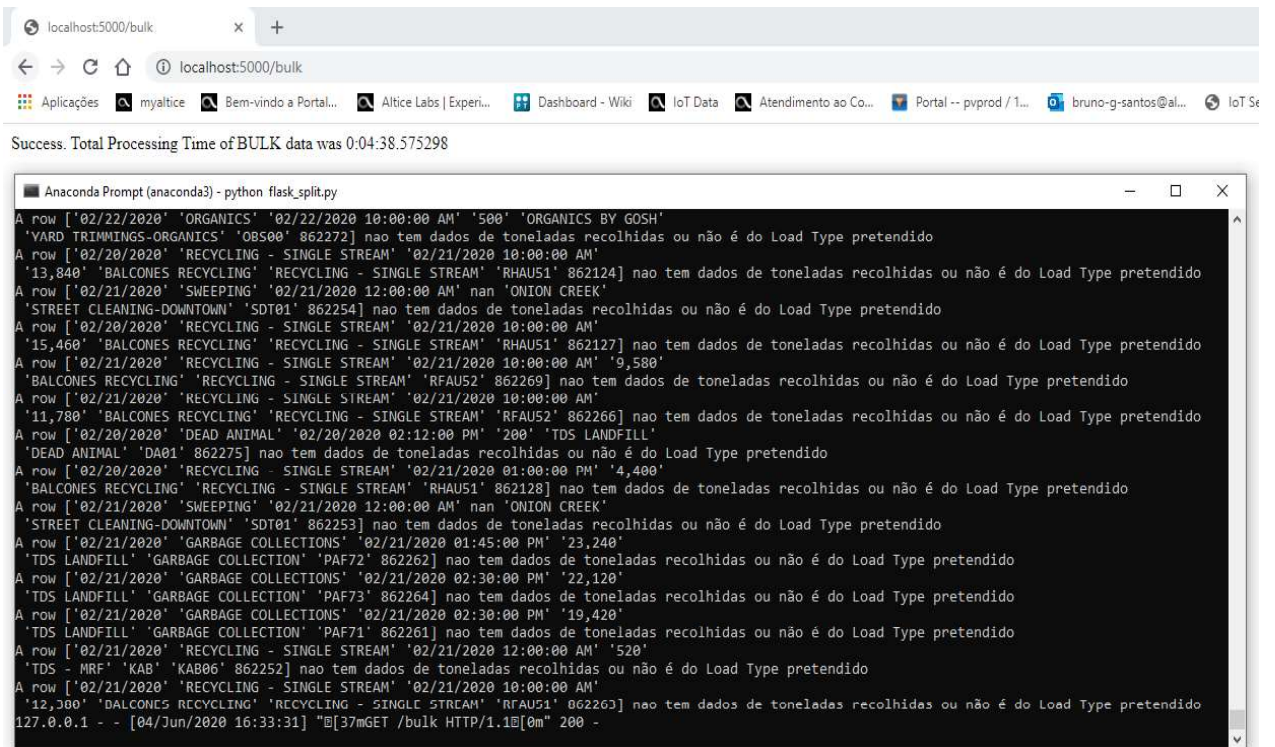
## Resultado do processamento dos dados do resíduo Sweeping



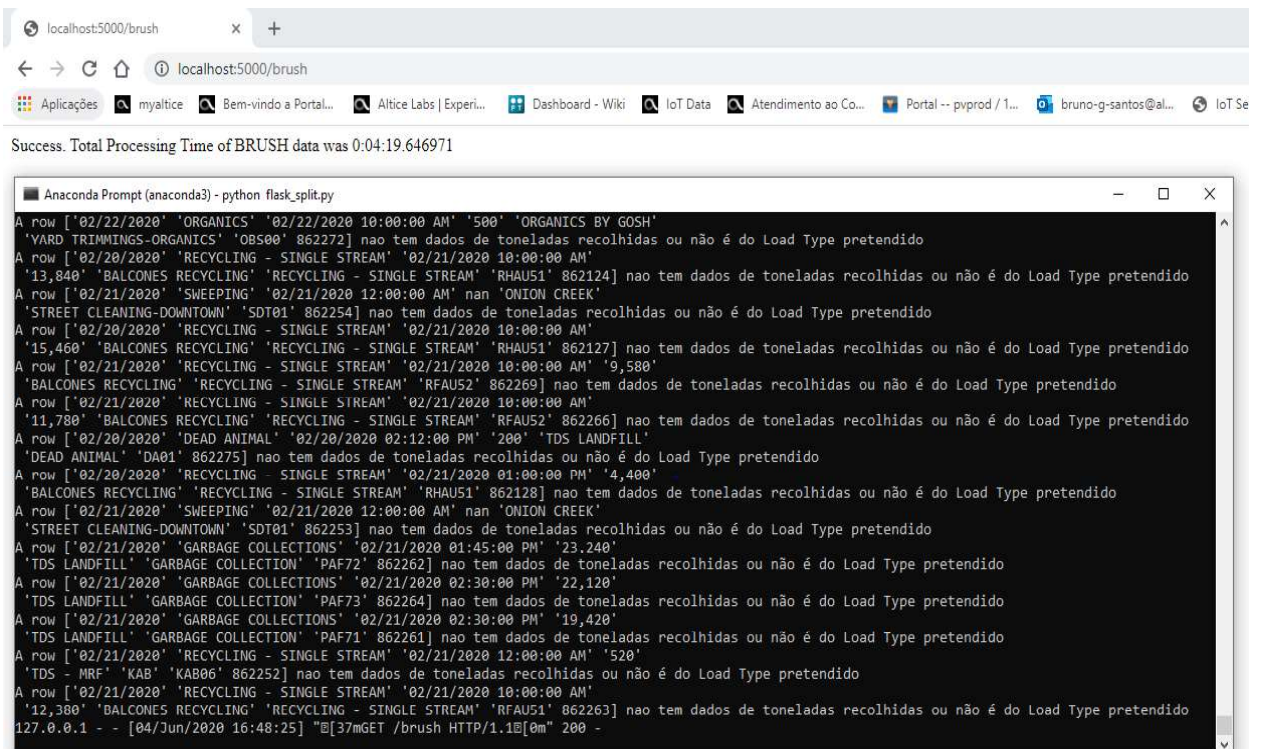
```
Anaconda Prompt (anaconda3) - python flask_split.py
A row ['02/22/2020' 'ORGANICS' '02/22/2020 10:00:00 AM' '500' 'ORGANICS BY GOSH'
'YARD TRIMMINGS-ORGANICS' 'OBS00' 862272] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'13,840' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RHAU51' 862124] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'SWEEPING' '02/21/2020 12:00:00 AM' nan 'ONION CREEK'
'STREET CLEANING-DOWNTOWN' 'SDT01' 862254] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'15,460' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RHAU51' 862127] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM' '9,580'
'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RFAU52' 862269] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'11,780' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RFAU52' 862266] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'DEAD ANIMAL' '02/20/2020 02:12:00 PM' '200' 'TDS LANDFILL'
'DEAD ANIMAL' 'DA01' 862275] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 01:00:00 PM' '4,400'
'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RHAU51' 862128] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'SWEEPING' '02/21/2020 12:00:00 AM' nan 'ONION CREEK'
'STREET CLEANING-DOWNTOWN' 'SDT01' 862253] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'GARBAGE COLLECTIONS' '02/21/2020 01:45:00 PM' '23,240'
'TDS LANDFILL' 'GARBAGE COLLECTION' 'PAF72' 862262] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'GARBAGE COLLECTIONS' '02/21/2020 02:30:00 PM' '22,120'
'TDS LANDFILL' 'GARBAGE COLLECTION' 'PAF73' 862264] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'GARBAGE COLLECTIONS' '02/21/2020 02:30:00 PM' '19,420'
'TDS LANDFILL' 'GARBAGE COLLECTION' 'PAF71' 862261] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 12:00:00 AM' '520'
'TDS - MRF' 'KAB' 'KAB06' 862252] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'12,380' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RFAU51' 862263] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
127.0.0.1 - - [04/Jun/2020 16:27:51] "[37mGET /yard_trimming HTTP/1.1] 200 -
```

## Resultado do processamento dos dados do resíduo Yard Trimming

# APÊNDICE 1 – FIGURAS DOS TEMPOS DE PROCESSAMENTO DA API



Resultado do processamento dos dados do resíduo Bulk



Resultado do processamento dos dados do resíduo Brush



```
Anaconda Prompt (anaconda3) - python flask_split.py
A row ['02/22/2020' 'ORGANICS' '02/22/2020 10:00:00 AM' '500' 'ORGANICS BY GOSH'
'YARD TRIMMINGS-ORGANICS' 'OBS00' 862272] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'13,840' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RHAU51' 862124] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'SWEEPING' '02/21/2020 12:00:00 AM' nan 'ONION CREEK'
'STREET CLEANING-DOWNTOWN' 'SDT01' 862254] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'15,460' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RHAU51' 862127] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM' '9,580'
'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RFAU52' 862269] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'11,780' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RFAU52' 862266] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'DEAD ANIMAL' '02/20/2020 02:12:00 PM' '200' 'TDS LANDFILL'
'DEAD ANIMAL' 'DA01' 862275] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 01:00:00 PM' '4,400'
'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RHAU51' 862128] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'SWEEPING' '02/21/2020 12:00:00 AM' nan 'ONION CREEK'
'STREET CLEANING-DOWNTOWN' 'SDT01' 862253] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'GARBAGE COLLECTIONS' '02/21/2020 01:45:00 PM' '23,240'
'TDS LANDFILL' 'GARBAGE COLLECTION' 'PAF72' 862262] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'GARBAGE COLLECTIONS' '02/21/2020 02:30:00 PM' '22,120'
'TDS LANDFILL' 'GARBAGE COLLECTION' 'PAF73' 862264] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'GARBAGE COLLECTIONS' '02/21/2020 02:30:00 PM' '19,420'
'TDS LANDFILL' 'GARBAGE COLLECTION' 'PAF71' 862261] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 12:00:00 AM' '520'
'TDS - MRF' 'KAB' 'KAB06' 862252] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'12,380' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RFAU51' 862263] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
127.0.0.1 - - [04/Jun/2020 16:54:27] "[37mGET /recycling_paper HTTP/1.1@0m] 200 -
```

## Resultado do processamento dos dados do resíduo Recycling – Paper



```
Anaconda Prompt (anaconda3) - python flask_split.py
A row ['02/22/2020' 'ORGANICS' '02/22/2020 10:00:00 AM' '500' 'ORGANICS BY GOSH'
'YARD TRIMMINGS-ORGANICS' 'OBS00' 862272] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'13,840' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RHAU51' 862124] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'SWEEPING' '02/21/2020 12:00:00 AM' nan 'ONION CREEK'
'STREET CLEANING-DOWNTOWN' 'SDT01' 862254] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'15,460' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RHAU51' 862127] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM' '9,580'
'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RFAU52' 862269] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'11,780' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RFAU52' 862266] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'DEAD ANIMAL' '02/20/2020 02:12:00 PM' '200' 'TDS LANDFILL'
'DEAD ANIMAL' 'DA01' 862275] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 01:00:00 PM' '4,400'
'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RHAU51' 862128] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'SWEEPING' '02/21/2020 12:00:00 AM' nan 'ONION CREEK'
'STREET CLEANING-DOWNTOWN' 'SDT01' 862253] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'GARBAGE COLLECTIONS' '02/21/2020 01:45:00 PM' '23,240'
'TDS LANDFILL' 'GARBAGE COLLECTION' 'PAF72' 862262] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'GARBAGE COLLECTIONS' '02/21/2020 02:30:00 PM' '22,120'
'TDS LANDFILL' 'GARBAGE COLLECTION' 'PAF73' 862264] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'GARBAGE COLLECTIONS' '02/21/2020 02:30:00 PM' '19,420'
'TDS LANDFILL' 'GARBAGE COLLECTION' 'PAF71' 862261] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 12:00:00 AM' '520'
'TDS - MRF' 'KAB' 'KAB06' 862252] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'12,380' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RFAU51' 862263] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
127.0.0.1 - - [04/Jun/2020 17:00:34] "[37mGET /recycling_comingle HTTP/1.1@0m] 200 -
```

## Resultado do processamento dos dados do resíduo Recycling – Comingle

## APÊNDICE 1 – FIGURAS DOS TEMPOS DE PROCESSAMENTO DA API

```
localhost:5000/organics
Success. Total Processing Time of ORGANICS data was 0:04:34.103742

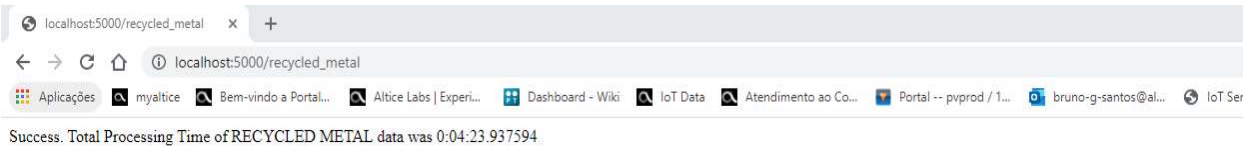
Anaconda Prompt (anaconda3) - python flask_split.py
A row ['02/20/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 01:00:00 PM' '7,040'
'TDS - MRF' 'RECYCLING - SINGLE STREAM' 'RHAU50' 862122] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'13,840' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RHAU51' 862124] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'SWEEPING' '02/21/2020 12:00:00 AM' nan 'ONION CREEK'
'STREET CLEANING-DOWNTOWN' 'SDT01' 862254] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'15,460' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RHAU51' 862127] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM' '9,580'
'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RFAU52' 862269] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'11,780' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RFAU52' 862266] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'DEAD ANIMAL' '02/20/2020 02:12:00 PM' '200' 'TDS LANDFILL'
'DEAD ANIMAL' 'DA01' 862275] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 01:00:00 PM' '4,400'
'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RHAU51' 862128] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'SWEEPING' '02/21/2020 12:00:00 AM' nan 'ONION CREEK'
'STREET CLEANING-DOWNTOWN' 'SDT01' 862253] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'GARBAGE COLLECTIONS' '02/21/2020 01:45:00 PM' '23,240'
'TDS LANDFILL' 'GARBAGE COLLECTION' 'PAF72' 862262] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'GARBAGE COLLECTIONS' '02/21/2020 02:30:00 PM' '22,120'
'TDS LANDFILL' 'GARBAGE COLLECTION' 'PAF73' 862264] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'GARBAGE COLLECTIONS' '02/21/2020 02:30:00 PM' '19,420'
'TDS LANDFILL' 'GARBAGE COLLECTION' 'PAF71' 862261] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 12:00:00 AM' '520'
'TDS - MRF' 'KAB' 'KAB06' 862252] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'12,380' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RFAU51' 862263] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
127.0.0.1 - - [04/Jun/2020 17:06:08] "[37mGET /organics HTTP/1.1]0m" 200 -
```

Resultado do processamento dos dados do resíduo Organics

```
localhost:5000/mixed_litter
Success. Total Processing Time of MIXED LITTER data was 0:04:52.897559

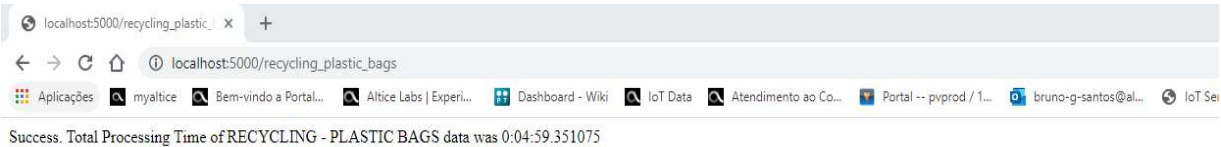
Anaconda Prompt (anaconda3) - python flask_split.py
A row ['02/22/2020' 'ORGANICS' '02/22/2020 10:00:00 AM' '500' 'ORGANICS BY GOSH'
'YARD TRIMMINGS-ORGANICS' 'OBS00' 862272] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'13,840' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RHAU51' 862124] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'SWEEPING' '02/21/2020 12:00:00 AM' nan 'ONION CREEK'
'STREET CLEANING-DOWNTOWN' 'SDT01' 862254] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'15,460' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RHAU51' 862127] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM' '9,580'
'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RFAU52' 862269] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'11,780' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RFAU52' 862266] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'DEAD ANIMAL' '02/20/2020 02:12:00 PM' '200' 'TDS LANDFILL'
'DEAD ANIMAL' 'DA01' 862275] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 01:00:00 PM' '4,400'
'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RHAU51' 862128] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'SWEEPING' '02/21/2020 12:00:00 AM' nan 'ONION CREEK'
'STREET CLEANING-DOWNTOWN' 'SDT01' 862253] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'GARBAGE COLLECTIONS' '02/21/2020 01:45:00 PM' '23,240'
'TDS LANDFILL' 'GARBAGE COLLECTION' 'PAF72' 862262] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'GARBAGE COLLECTIONS' '02/21/2020 02:30:00 PM' '22,120'
'TDS LANDFILL' 'GARBAGE COLLECTION' 'PAF73' 862264] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'GARBAGE COLLECTIONS' '02/21/2020 02:30:00 PM' '19,420'
'TDS LANDFILL' 'GARBAGE COLLECTION' 'PAF71' 862261] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 12:00:00 AM' '520'
'TDS - MRF' 'KAB' 'KAB06' 862252] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'12,380' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RFAU51' 862263] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
127.0.0.1 - - [04/Jun/2020 17:12:24] "[37mGET /mixed_litter HTTP/1.1]0m" 200 -
```

Resultado do processamento dos dados do resíduo Mixed Litter



```
Anaconda Prompt (anaconda3) - python flask_split.py
A row ['02/22/2020' 'ORGANICS' '02/22/2020 10:00:00 AM' '500' 'ORGANICS BY GOSH'
'YARD TRIMMINGS-ORGANICS' 'OBS00' 862272] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'13,840' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RHAU51' 862124] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'SWEEPING' '02/21/2020 12:00:00 AM' nan 'ONION CREEK'
'STREET CLEANING-DOWNTOWN' 'SDT01' 862254] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'15,460' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RHAU51' 862127] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM' '9,580'
'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RFAU52' 862269] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'11,780' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RFAU52' 862266] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'DEAD ANIMAL' '02/20/2020 02:12:00 PM' '200' 'TDS LANDFILL'
'DEAD ANIMAL' 'DA01' 862275] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 01:00:00 PM' '4,400'
'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RHAU51' 862128] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'SWEEPING' '02/21/2020 12:00:00 AM' nan 'ONION CREEK'
'STREET CLEANING-DOWNTOWN' 'SDT01' 862253] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'GARBAGE COLLECTIONS' '02/21/2020 01:45:00 PM' '23,240'
'TDS LANDFILL' 'GARBAGE COLLECTION' 'PAF72' 862262] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'GARBAGE COLLECTIONS' '02/21/2020 02:30:00 PM' '22,120'
'TDS LANDFILL' 'GARBAGE COLLECTION' 'PAF73' 862264] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'GARBAGE COLLECTIONS' '02/21/2020 02:30:00 PM' '19,420'
'TDS LANDFILL' 'GARBAGE COLLECTION' 'PAF71' 862261] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 12:00:00 AM' '520'
'TDS - MRF' 'KAB' 'KAB06' 862252] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'12,380' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RFAU51' 862263] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
127.0.0.1 - - [04/Jun/2020 17:17:28] "[37mGET /recycled_metal HTTP/1.1[0m] 200 -
```

## Resultado do processamento dos dados do resíduo Recycled Metal



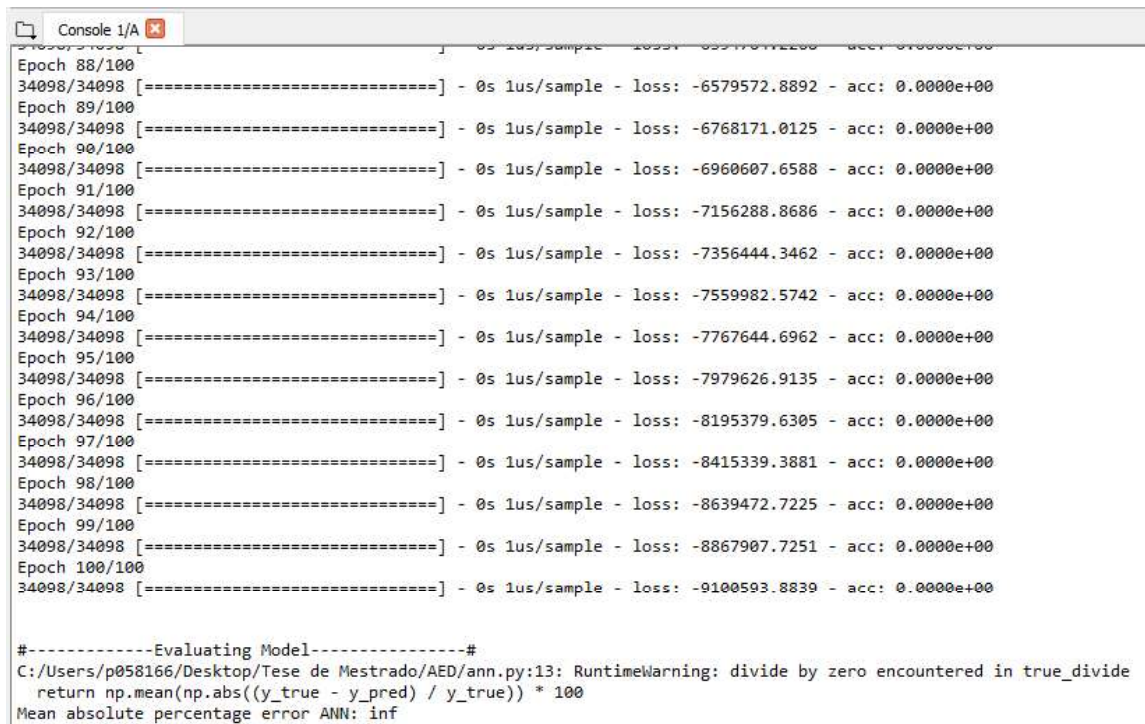
```
Anaconda Prompt (anaconda3) - python flask_split.py
A row ['02/22/2020' 'ORGANICS' '02/22/2020 10:00:00 AM' '500' 'ORGANICS BY GOSH'
'YARD TRIMMINGS-ORGANICS' 'OBS00' 862272] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'13,840' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RHAU51' 862124] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'SWEEPING' '02/21/2020 12:00:00 AM' nan 'ONION CREEK'
'STREET CLEANING-DOWNTOWN' 'SDT01' 862254] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'15,460' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RHAU51' 862127] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM' '9,580'
'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RFAU52' 862269] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'11,780' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RFAU52' 862266] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/20/2020' 'DEAD ANIMAL' '02/20/2020 02:12:00 PM' '200' 'TDS LANDFILL'
'DEAD ANIMAL' 'DA01' 862275] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 01:00:00 PM' '4,400'
'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RHAU51' 862128] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'SWEEPING' '02/21/2020 12:00:00 AM' nan 'ONION CREEK'
'STREET CLEANING-DOWNTOWN' 'SDT01' 862253] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'GARBAGE COLLECTIONS' '02/21/2020 01:45:00 PM' '23,240'
'TDS LANDFILL' 'GARBAGE COLLECTION' 'PAF72' 862262] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'GARBAGE COLLECTIONS' '02/21/2020 02:30:00 PM' '22,120'
'TDS LANDFILL' 'GARBAGE COLLECTION' 'PAF73' 862264] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'GARBAGE COLLECTIONS' '02/21/2020 02:30:00 PM' '19,420'
'TDS LANDFILL' 'GARBAGE COLLECTION' 'PAF71' 862261] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 12:00:00 AM' '520'
'TDS - MRF' 'KAB' 'KAB06' 862252] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
A row ['02/21/2020' 'RECYCLING - SINGLE STREAM' '02/21/2020 10:00:00 AM'
'12,380' 'BALCONES RECYCLING' 'RECYCLING - SINGLE STREAM' 'RFAU51' 862263] nao tem dados de toneladas recolhidas ou não é do Load Type pretendido
127.0.0.1 - - [04/Jun/2020 17:33:21] "[37mGET /recycling_plastic_bags HTTP/1.1[0m] 200 -
```

## Resultado do processamento dos dados do resíduo Recycling – Plastic Bags

## APÊNDICE 2 – CONTRATEMPOS ENCONTRADOS DURANTE A FASE DE COMPARAÇÃO DOS ALGORITMOS

Durante a fase de configuração e teste da métrica MAPE, surgiu um pequeno percalço. Devido a alguns conjuntos de dados possuírem milhares de registos de recolhas, a matriz originada com os valores de previsão pode ultrapassar uma constituição na ordem dos 100000x100000.

Tendo em conta que, para além de existir uma matriz de previsão, existe de igual forma uma matriz com os dados reais e de tamanho igual à matriz de previsão, e que, para se calcular a percentagem do erro, é necessário comparar cada registo das duas matrizes, isto traduz-se numa exigência na capacidade de processamento da máquina tanto em tempo, como em “poder”, que, por esta ser limitada, a metodologia acaba por devolver um valor infinito, como se pode consultar na figura em baixo:



```
Console 1/A
Epoch 88/100
34098/34098 [=====] - 0s 1us/sample - loss: -6579572.8892 - acc: 0.0000e+00
Epoch 89/100
34098/34098 [=====] - 0s 1us/sample - loss: -6768171.0125 - acc: 0.0000e+00
Epoch 90/100
34098/34098 [=====] - 0s 1us/sample - loss: -6960607.6588 - acc: 0.0000e+00
Epoch 91/100
34098/34098 [=====] - 0s 1us/sample - loss: -7156288.8686 - acc: 0.0000e+00
Epoch 92/100
34098/34098 [=====] - 0s 1us/sample - loss: -7356444.3462 - acc: 0.0000e+00
Epoch 93/100
34098/34098 [=====] - 0s 1us/sample - loss: -7559982.5742 - acc: 0.0000e+00
Epoch 94/100
34098/34098 [=====] - 0s 1us/sample - loss: -7767644.6962 - acc: 0.0000e+00
Epoch 95/100
34098/34098 [=====] - 0s 1us/sample - loss: -7979626.9135 - acc: 0.0000e+00
Epoch 96/100
34098/34098 [=====] - 0s 1us/sample - loss: -8195379.6305 - acc: 0.0000e+00
Epoch 97/100
34098/34098 [=====] - 0s 1us/sample - loss: -8415339.3881 - acc: 0.0000e+00
Epoch 98/100
34098/34098 [=====] - 0s 1us/sample - loss: -8639472.7225 - acc: 0.0000e+00
Epoch 99/100
34098/34098 [=====] - 0s 1us/sample - loss: -8867907.7251 - acc: 0.0000e+00
Epoch 100/100
34098/34098 [=====] - 0s 1us/sample - loss: -9100593.8839 - acc: 0.0000e+00

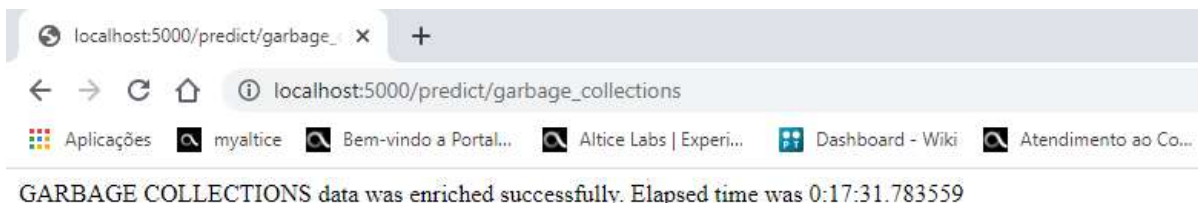
#-----Evaluating Model-----#
C:/Users/p058166/Desktop/Tese de Mestrado/AED/ann.py:13: RuntimeWarning: divide by zero encountered in true_divide
  return np.mean(np.abs((y_true - y_pred) / y_true)) * 100
Mean absolute percentage error ANN: inf
```

### Teste da métrica MAPE com o algoritmo *Artificial Neural Networks*

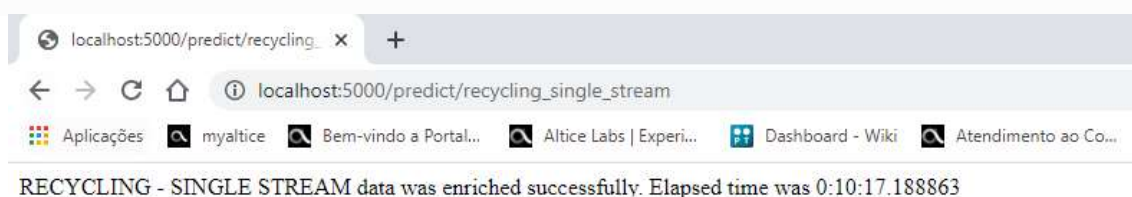
Como solução para esta situação, resolveu-se, a partir dos *datasets* dos tipos de resíduos originais, retirar amostras para se proceder ao teste dos algoritmos de previsão, constituindo um total de 100 registos por resíduo. As amostras foram incorporadas num novo ficheiro em formato CSV, continuando, de igual forma, a separação dos dados por tipo de resíduo correspondente. A figura seguinte representa um esquema dos ficheiros gerados para este teste em específico:



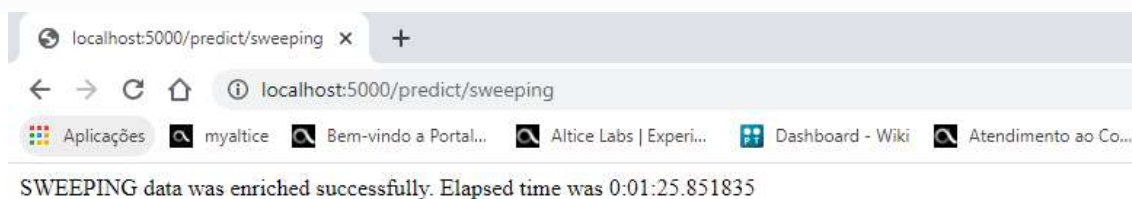
## APÊNDICE 3 – FIGURAS DOS TEMPOS DE PROCESSAMENTO DA API DE PREVISÃO



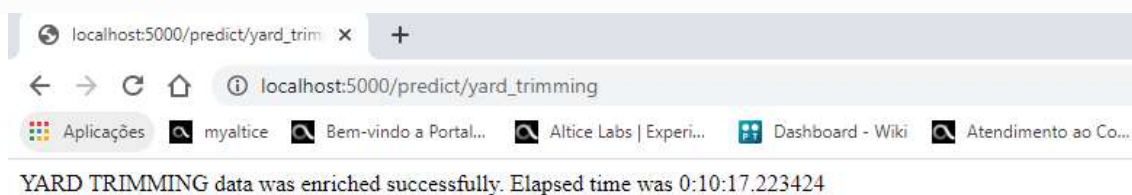
### Resultado da ingestão de dados do resíduo Garbage Collections



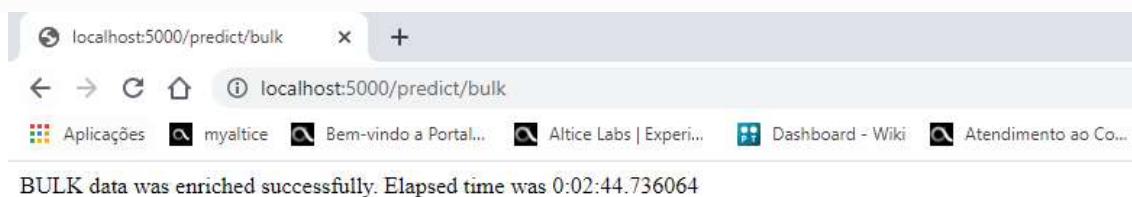
### Resultado da ingestão de dados do resíduo Recycling – Single Stream



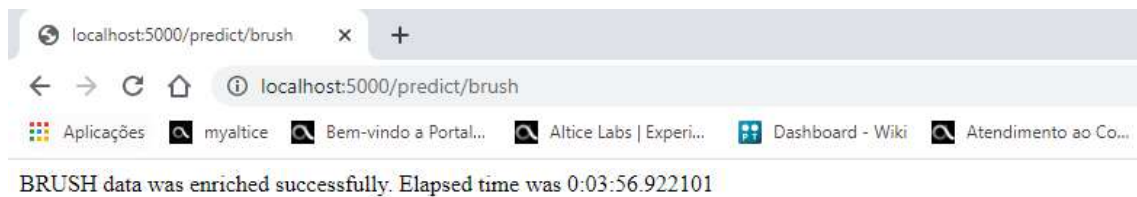
### Resultado da ingestão de dados do resíduo Sweeping



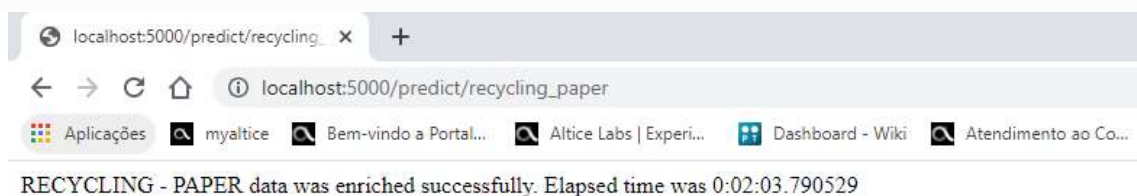
### Resultado da ingestão de dados do resíduo Yard Trimming



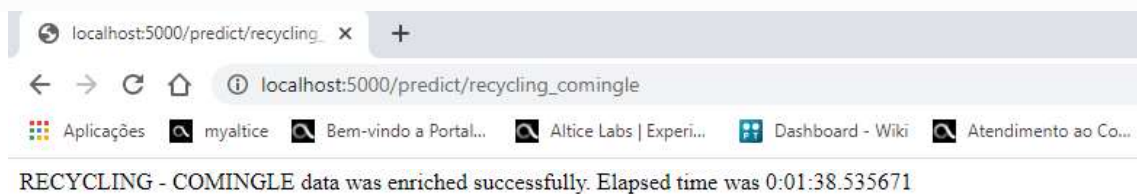
### Resultado da ingestão de dados do resíduo Bulk



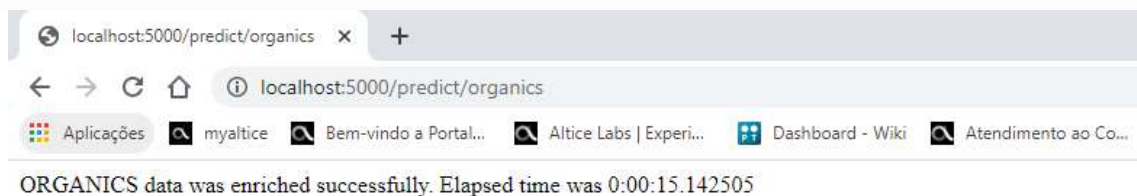
### Resultado da ingestão de dados do resíduo Brush



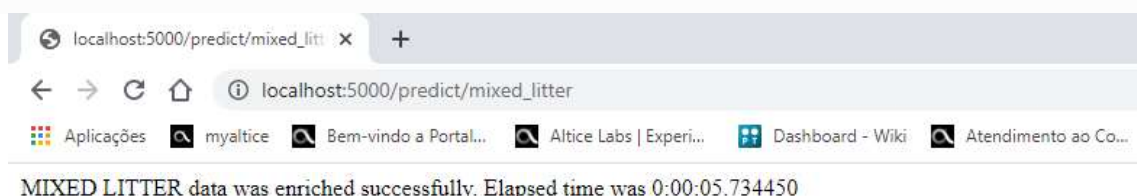
### Resultado da ingestão de dados do resíduo Recycling – Paper



### Resultado da ingestão de dados do resíduo Recycling – Comingle

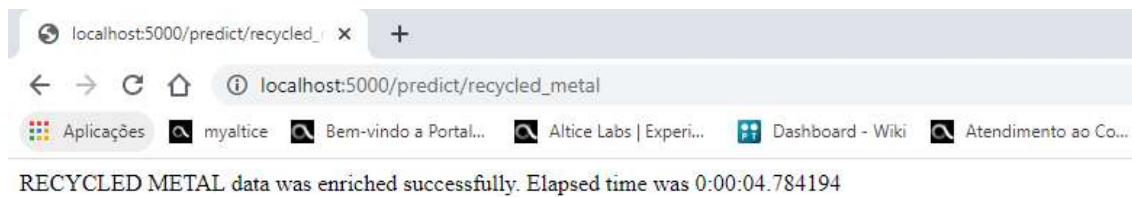


### Resultado da ingestão de dados do resíduo Organics

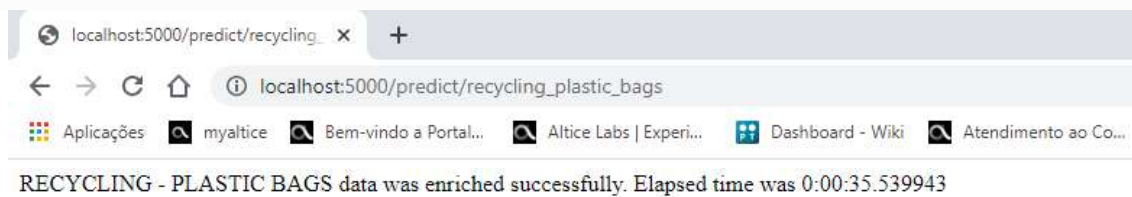


### Resultado da ingestão de dados do resíduo Mixed Litter

## APÊNDICE 3 – FIGURAS DOS TEMPOS DE PROCESSAMENTO DA API DE PREVISÃO



### Resultado da ingestão de dados do resíduo Recycled Metal



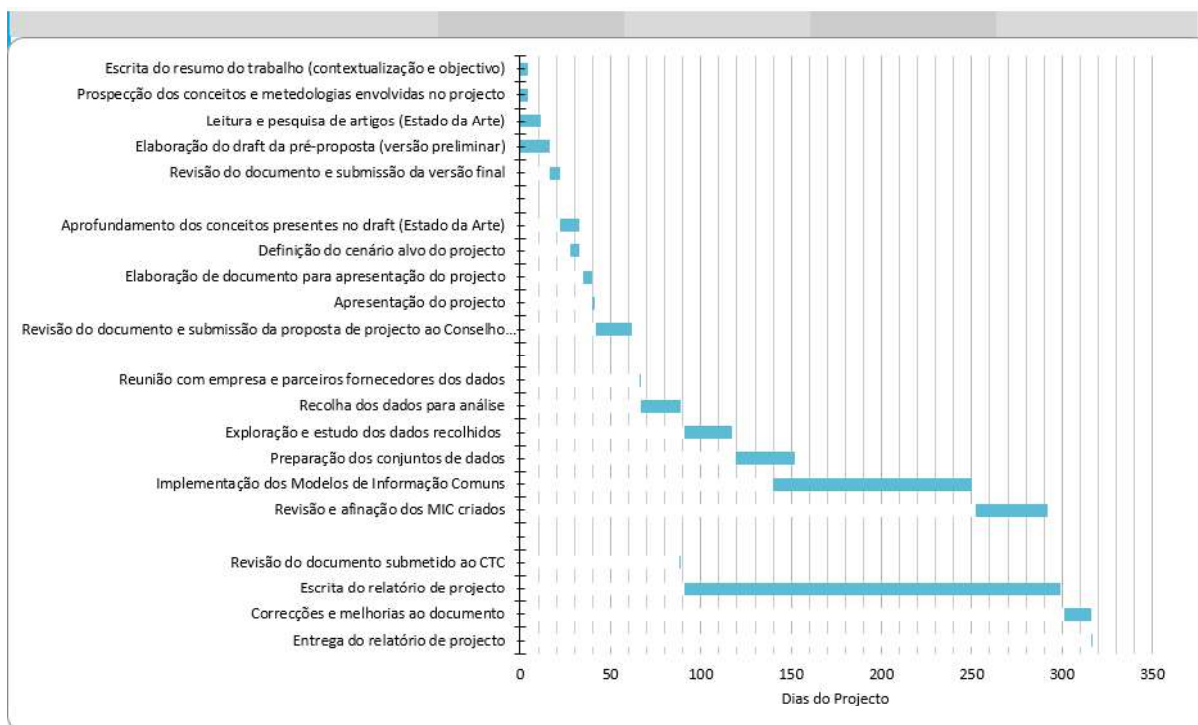
### Resultado da ingestão de dados do resíduo Recycling – Plastic Bags

# APÊNDICE 4 – CRONOGRAMA DO PROJETO NA FASE DE PRÉ-PROPOSTA

## CRONOGRAMA DO PROJECTO

\* célula calculada automaticamente

NOME DA TAREFA	DATA INICIO	DATA FIM	COMEÇA NO DIA*	DURAÇÃO* (WORK DAYS)
<b>Início do Projecto</b>				
Escrita do resumo do trabalho (contextualização e objectivo)	11/4	11/7	0	4
Prospecção dos conceitos e metodologias envolvidas no projecto	11/4	11/7	0	4
Leitura e pesquisa de artigos (Estado da Arte)	11/4	11/14	0	11
Elaboração do draft da pré-proposta (versão preliminar)	11/4	11/19	0	16
Revisão do documento e submissão da versão final	11/20	11/25	16	6
<b>Proposta Final do Projecto</b>				
Aprofundamento dos conceitos presentes no draft (Estado da Arte)	11/26	12/6	22	11
Definição do cenário alvo do projecto	12/2	12/6	28	5
Elaboração de documento para apresentação do projecto	12/9	12/13	35	5
Apresentação do projecto	12/14	12/14	40	1
Revisão do documento e submissão da proposta de projecto ao Conselho Técnico Científico	12/16	1/4	42	20
<b>Implementação do Projecto</b>				
Reunião com empresa e parceiros fornecedores dos dados	1/9	1/9	66	1
Recolha dos dados para análise	1/10	1/31	67	22
Exploração e estudo dos dados recolhidos	2/3	2/28	91	26
Preparação dos conjuntos de dados	3/2	4/3	119	33
Implementação dos Modelos de Informação Comuns	3/23	7/10	140	110
Revisão e afinação dos MIC criados	7/13	8/21	252	40
<b>Elaboração do Relatório de Projecto</b>				
Revisão do documento submetido ao CTC	1/31	1/31	88	1
Escrita do relatório de projecto	2/3	8/28	91	208
Correcções e melhorias ao documento	8/31	9/14	301	15
Entrega do relatório de projecto	9/15	9/15	316	1
<b>Defesa do Projecto</b>				
Preparação da apresentação do projecto	9/16	10/5	317	20
Defesa e apresentação do projecto	10/9	10/9	340	1



## ANEXO 1 – TAXA DE DESEMPREGO: TOTAL E POR SEXO (%)

Taxa - %

Anos	Sexo		
	Total	Masculino	Feminino
2012	15,5	15,6	15,5
2013	16,2	16,0	16,4
2014	13,9	13,5	14,3
2015	12,4	12,2	12,7
2016	11,1	11,0	11,2
2017	8,9	8,4	9,4
2018	7,0	6,6	7,4

↳  
Taxa de desemprego: total e por sexo (%)  
Fontes de Dados: INE - Inquérito ao Emprego  
Fonte: PORDATA  
Última actualização: 2019-02-07