



TÉCNICO
LISBOA

**Análise da viabilidade de previsão do consumo de combustível
de camiões de transporte em obras rodoviárias
através de machine learning**

Hugo Filipe da Silva Freitas Fernandes

Dissertação para a obtenção do Grau de Mestre em

Engenharia Civil

Orientadores: Professor Doutor José Manuel Coelho das Neves

Doutor Manuel Afonso Parente

Júri

Presidente: Professora Doutora Maria do Rosário Maurício Ribeiro Macário

Orientador: Professor Doutor José Manuel Coelho das Neves

Vogal: Professor Doutor Ionut Dragos Moldovan

Outubro de 2021

Para os meus filhos Tomás e Laura

DECLARAÇÃO

Declaro que o presente documento é um trabalho original da minha autoria e que cumpre todos os requisitos do Código de Conduta e Boas Práticas da Universidade de Lisboa.

AGRADECIMENTOS

Gratidão, é o sentimento que me invade a alma ao escrever estas linhas de reconhecimento.

Grato, aos meus orientadores, Professor José Neves e Doutor Manuel Parente, por todos os conhecimentos técnicos que permitiram a conclusão desta dissertação, mas mais do que a vertente académica, pela vertente humana, sempre disponível e generosa, e por terem acreditado em mim, mesmo quando o cenário não era o mais favorável.

Grato a todos aqueles que com a sua cooperação permitiram a génese e a conclusão deste trabalho, nomeadamente, à Built CoLab, na pessoa do Eng.º João Moutinho, ao Manuel Sampaio pela construção do protótipo e à JJR, nas pessoas do Eng.º André Lourenço e Eng.º Filipe Gameiro.

Grato aos meus amigos, dentro e fora do IST, que acompanharam este trajeto e estiveram presentes nos bons e maus momentos e um agradecimento especial à Ana Chambel pelo “Kit Tese” e pela ajuda e experiência na prática de documentos académicos. Ao Dr. Francisco por todo o suporte mental e orientação que estiveram na base do sucesso pessoal e académico.

Grato à minha esposa Lucinda e aos meus filhos Tomás e Laura, pelo suporte familiar e pelo sacrifício em termos de tempo e atenção disponíveis para a conclusão deste trabalho. Foi por vocês também!

Grato, duma maneira muito especial, aos meus pais Hernâni e Conceição, por sempre me terem dado todas as condições, neste longo e por vezes penoso caminho, que se conclui agora com a concretização de um sonho, também vosso, mas sobretudo por todos os valores que me transmitiram e por serem quem são, um exemplo. Muito, muito obrigado!

RESUMO

No contexto tecnológico e empresarial em que vivemos, os objetivos passam por fazer mais utilizando menos recursos através da otimização dos processos. Assim, utilizar a tecnologia mais recente em áreas como a inteligência artificial, com o intuito de termos máquinas a prever e estimar custos, configura-se como algo bastante importante. Sabendo que o custo dos combustíveis, mais concretamente dos veículos pesados de transporte de materiais para obras de pavimentação de estradas, representa uma parte relevante do custo total do processo, na presente dissertação estabeleceu-se o objetivo de analisar a viabilidade da previsão do consumo de combustível dos mesmos. Recorrendo à utilização de base de dados real decorrente da instalação de sensores num camião de obra, procedeu-se à previsão do consumo de combustível através da implementação de modelos de Inteligência Artificial, mais especificamente *machine learning*, considerando os trajetos percorridos, as velocidades médias, as cargas transportadas e o registo do consumo real associado a cada trajeto, que corresponde ao termo de comparação no qual se baseia o estudo. Esta premissa foi validada em contexto real num estaleiro de construção rodoviária em colaboração com uma empresa construtora portuguesa. Os resultados deste estudo demonstraram a viabilidade da previsão efetuada e confirmaram as vantagens associadas à combinação de sensorização para recolha de dados e inteligência artificial num contexto de construção real.

Palavras-chave: Camiões, Consumo de combustível, *Machine Learning*, Pavimentação, Sensores

ABSTRACT

In our current technological and business setting, the main goal is to do more with less, i.e., to use fewer resources, through process optimization. Therefore, the use of the latest technology in areas such as artificial intelligence, as a way to predict and estimate costs, is crucial. Knowing that the cost of fuels, more specifically heavy vehicles transporting materials for road paving works, represents a significant part of the total cost of the process, this dissertation's goal is to analyze the feasibility of predicting fuel consumption. Through the installation of sensors in a transport truck, the real numbers of fuel consumption were analyzed. This database allowed the forecast of fuel consumption through the implementation of Artificial Intelligence models, more specifically machine learning. This has enabled an evaluation, considering the following dimensions: route traveled, average speed, loads transported and recording of the actual consumption associated with each route, which corresponds to the comparison term on which the study is based. This premise was validated in a real-life context (road construction site) in collaboration with a Portuguese construction company. The results of this study support the feasibility of the concept, providing relevant insights into the advantages associated with the combination of sensorization for data collection and artificial intelligence in a real construction scenario.

Keywords: Fuel consumption, Machine Learning, Pavements, Sensors, Trucks

ÍNDICE

1. INTRODUÇÃO	1
1.1. Enquadramento.....	1
1.2. Objetivos.....	1
1.3. Metodologia	2
1.4. Organização da dissertação.....	3
2. REVISÃO DE LITERATURA.....	5
2.1. IoT e Sensorização	5
2.1.1. Aceleração e Inclinação	6
2.1.2. Sistema de Posicionamento Global	8
2.2. Machine Learning	9
2.2.1. Modelos	10
2.2.2. Tarefas de Machine Learning	14
2.2.3. Processos dos modelos	15
2.3. Casos da bibliografia.....	16
3. CASO DE ESTUDO.....	21
3.1. Implementação.....	21
3.1.1 Hardware.....	21
3.1.2 Software	24
3.1.3 Encapsulamento.....	25
3.1.4 Instalação no equipamento de obra	26
3.2. Apresentação e tratamento de dados.....	30
3.2.1 Avaliação da qualidade de modelos de machine learning	30

3.2.2 Dados não-processados	30
3.2.3 Tratamento de dados.....	32
4. RESULTADOS E DISCUSSÃO.....	35
5. CONSIDERAÇÕES FINAIS	39
5.1. Conclusões.....	39
5.2. Desenvolvimentos futuros.....	40
REFERÊNCIAS BIBLIOGRÁFICAS.....	41

LISTA DE QUADROS

Quadro 2.1 - Resumo das partes integrantes do GNSS.....	8
Quadro 2.2 - Descrição das fases referentes ao CRISP-DM	15
Quadro 3.1 – Valores extraídos da base de dados não-tratados	31
Quadro 3.2 - Intervalos e descrição da inclinação	33
Quadro 3.3 – Valores retirados da base de dados de treino e teste	33

LISTA DE FIGURAS

Figura 1.1 - Processo de recolha, consolidação e processamento de dados de campo e criação de modelos de previsão.....	2
Figura 2.1 - Representação da ação da aceleração num condensador MEMS (adaptado de Kopfinger e Ahlsén, 2019)	6
Figura 2.2 - Giroscópio tradicional	7
Figura 2.3 - Exemplo da cobertura dos satélites (Drotek@, 2021)	8
Figura 2.4 - Exemplo esquemático do funcionamento de uma <i>decision tree</i> (adaptado de Yiu@, 2021)	10
Figura 2.5 - Exemplo esquemático de uma <i>Random Forest</i> (adaptado de Yiu@, 2021)	11
Figura 2.6 - Esquema de uma <i>artificial neural network</i> (adaptado de Santos@, 2021)	12
Figura 2.7 - Representação gráfica do SVM (adaptado de Agrawal@, 2021)	13
Figura 2.8 - Tarefas do <i>machine learning</i>	14
Figura 2.9 - Diagrama referente ao CRISP-DM (Saltz et al., 2017)	15
Figura 2.10 – Parâmetros e a sua correlação utilizados nos modelos de previsão de (Almer, 2015)	18
Figura 3.1 - GNSS 7 click (Mikroe@, 2021a)	22
Figura 3.2 - IMU 15 click (Mikroe@, 2021b)	22
Figura 3.3 - NVIDIA Jetson Nano 2GB (NVidia@, 2021)	23
Figura 3.4 - Projeto desenvolvido no ambiente Tinkercad	25
Figura 3.5 - Protótipo Final.....	26
Figura 3.6 – Vista geral a) e pormenor b) da instalação da antena GNSS no camião	26
Figura 3.7 - Instalação do protótipo	27
Figura 3.8 – Fluxograma das ações solicitadas ao motorista	28
Figura 3.9 - Exemplo da folha de registos preenchida pelo motorista	29

Figura 3.10 – Exemplo de um perfil altimétrico de uma viagem aferido pelo camião sensorizado	32
Figura 4.1 - Representação gráfica das curvas REC dos vários modelos	35
Figura 4.2 - Representação gráfica do resultado dos dados estimados vs. observados	36
Figura 4.3 - Representação gráfica da relevância das varáveis para o modelo	37

LISTA DE SIGLAS

ANN – Artificial Neural Networks

CRISP-DM – Cross Industry Standard Process for Data Mining

GNSS – Global Navigation Satellite Systems

GPIO – General Purpose Input/Output

GPS – Global Positioning System

IA – Inteligência Artificial

IMU – Inertial Measurement Unit

IoT – Internet of Things

I2C – Inter-Integrated Circuit

MEMS – Micro Electro-Mechanical Systems

ML – Machine Learning

RF – Random Forests

SCP – Secure Copy Protocol

STARS – Scania Truck and Road Simulation

SVM – Support Vector Machines

1. INTRODUÇÃO

1.1. Enquadramento

Um dos principais custos em projetos lineares de infraestruturas de transporte está relacionado com o consumo de combustível de maquinaria pesada. Entre estes, os camiões de transporte de materiais de construção estão tipicamente associados a uma variabilidade muito elevada no que diz respeito ao seu consumo de combustível, uma vez que é uma função de vários fatores tipicamente difíceis de estimar. Estes últimos incluem não só aspetos relacionados com as características do equipamento, mas também com o peso da carga transportada e o trajeto através do qual o transporte de materiais é realizado. Normalmente, as equipas de que tratam do dimensionamento e gestão de projetos usam valores médios para o consumo de combustível deste tipo de equipamento, para estimar os seus requisitos de combustível e orçamento ao longo da construção. No entanto, os valores médios não são suficientemente precisos para permitir que os projetistas estimem devidamente os custos relacionados com o combustível durante todo o período de um projeto de construção. Assim, este trabalho descreve os primeiros resultados de um estudo destinado a combinar tecnologias de monitorização remota baseadas em sensores inteligentes com modelos de previsão de inteligência artificial (IA) (isto é, *machine learning* (ML)), para criar uma estrutura que visa a estimação do uso de combustível por camiões, neste tipo de projetos de construção, com maior precisão. Em última análise, o projeto é um passo rumo às tendências de digitalização atuais defendidas por conceitos como os sistemas ciber-físicos, tais como a Indústria 4.0 e 5.0.

1.2. Objetivos

Em projetos de construção e reabilitação de pavimentos rodoviários, as tarefas de transporte de materiais constituem um dos componentes de grande importância para o processo construtivo. Em especial, a capacidade de prever os custos associados às tarefas de transporte com exatidão é um fator decisivo para um bom planeamento da obra e dimensionamento do parque de máquinas associado, nomeadamente em termos de consumo de combustível. Estes custos dependem de uma série de fatores, tais como a distância de transporte, capacidade de carga dos equipamentos (e.g., camiões, *dumpers*) ou as características do percurso pelo qual se realiza a atividade de transporte de materiais de construção. Ao mesmo tempo, as tendências atuais relacionadas com conceitos tais como a Construção 4.0 promovem a utilização de sensores instalados nos equipamentos com o objetivo de registo e criação de grandes bases de dados de obras passadas.

Neste contexto, esta dissertação tem como objetivo principal a utilização de bases de dados de obras reais decorrentes da instalação de sensores em equipamentos de transporte para estudar a viabilidade da previsão do consumo de combustível dos mesmos. Esta previsão apoia-se na implementação de modelos de IA (e.g., ML), tendo em vista a análise das bases de dados acima referidas em função da caracterização dos percursos percorridos pelos camiões sensorizados (e.g., número de subidas leves e íngremes, número de descidas), das

velocidades médias atingidas pelos camiões, das cargas transportadas e do registo do consumo real associado a cada percurso, que corresponde ao "ground truth" no qual se baseia o estudo.

1.3. Metodologia

A metodologia adotada no desenvolvimento da presente dissertação começou por definir os indicadores fundamentais para prever o consumo de combustível e como efetivá-lo. Assim, foram escolhidos sensores para fazer parte de um protótipo que foi instalado num veículo de transporte (camião) a fim de monitorizá-lo e daí obter informação sob a forma de ficheiros informáticos, como descrito e esquematizado na Figura 1.1. O processo teve continuidade com a recolha de dados automatizados provenientes dos sensores e outros de entrada manual recolhidos pelo operador, seguindo-se posteriormente o processo de consolidação, o pré-processamento e análise dos dados e a aplicação dos algoritmos de *machine learning* na criação dos modelos de previsão.

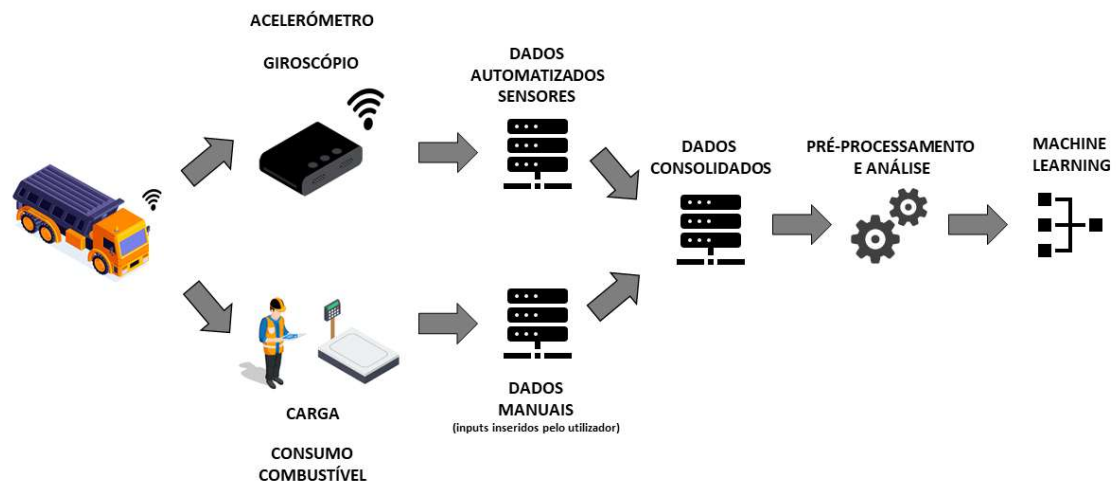


Figura 1.1 - Processo de recolha, consolidação e processamento de dados de campo e criação de modelos de previsão

Obtida essa informação, seria imprescindível tratá-la e retirar resultados, o qual foi realizado com o auxílio de um software de programação, nomeadamente, o R (R Development Core Team 2011). Este software, constituído por pacotes que lhe conferem a habilidade de simular métodos de ML, permitiu uma estimativa do consumo de combustível a qual foi comparada com o valor real observado, de modo a validar e aferir a robustez do modelo e a retirar ilações sobre relações causa-efeito de certos parâmetros com o consumo do veículo de transporte.

1.4. Organização da dissertação

A dissertação encontra-se estruturada em cinco capítulos, os quais serão sumarizados no seguimento deste subcapítulo.

O presente capítulo 1, de carácter introdutório, enquadra a importância e a atualidade do âmbito desta dissertação, bem como, apresenta os objetivos definidos e a metodologia aplicada para esse efeito. É ainda efetuada uma breve descrição da estrutura do documento da dissertação.

No capítulo 2, é feita uma descrição da temática e dos principais conceitos que lhe estão associados, bem como a revisão da literatura com o intuito de esclarecer o domínio e a atualidade do estudo desta problemática.

No seguimento, o capítulo 3 aborda o caso de estudo, nomeadamente, o fabrico do protótipo que alberga os sensores e a sua aplicação no equipamento de obra (camião). Posteriormente, o capítulo descreve o modo de obtenção e o tratamento dos dados que permitiram aferir da qualidade do modelo de estimação e a sua comparação com os dados reais.

No capítulo 4, os resultados do caso de estudo foram apresentados e analisados com recurso a quadros, gráficos e comparações permitindo o conhecimento da verdadeira robustez e verossimilhança do modelo obtido através de métodos computacionais.

Por fim, no capítulo 5 são apresentadas as considerações finais do trabalho realizado, nomeadamente, as principais conclusões obtidas, assim como as limitações e restrições ao trabalho desenvolvido e as oportunidades de incrementação e desenvolvimento para trabalhos futuros.

É importante referir também que, pelo facto desta temática abranger uma grande variedade de conceitos relativamente recentes na área das ciências tecnológicas e inexistentes no léxico em língua Portuguesa, optou-se na presente dissertação por usar diretamente a terminologia em Inglês e não fazer a tradução livre.

2. REVISÃO DE LITERATURA

Este capítulo aborda os conceitos necessários à melhor compreensão da temática abordada desta dissertação, nomeadamente, a definição das ferramentas úteis à resolução da melhor estimativa do consumo de combustíveis através da aprendizagem computadorizada de padrões associados a parâmetros previamente selecionados. Considera-se importante definir e explicar a tecnologia e ferramentas por detrás da resolução do problema levantado no âmbito desta dissertação a fim de compreender o processo e o objetivo. Analogamente, este capítulo servirá para estabelecer o ponto de partida e o que já se sabe do caminho – o que já foi analisado e executado nesta temática através da revisão bibliográfica da mesma – o caminho em si, através da elucidação dos conceitos de ML no tratamento da informação obtida e da tecnologia associada para a obtenção desses mesmos dados, e da sensorização que monitorizará os vários parâmetros e que permitirá chegar ao destino, o objetivo estabelecido.

2.1. IoT e Sensorização

Definir o conceito de internet das coisas (IoT – *Internet of Things*) e sensorização é uma condição pertencente à fase inicial da recolha de informação do trabalho associado a esta dissertação. Deste modo, a IoT e particularmente o seu conceito, na era tecnológica em que vivemos, são de uma grande atualidade. Pode-se definir IoT como a comunicação entre sensores, dispositivos ou uma combinação de ambos numa rede comum aos mesmos (Santos et al., 2016). A partir desta premissa, dispositivos distintos podem transmitir ou armazenar informação e dessa maneira tornar protótipos mais inteligentes e capazes de gerar comunicação entre si. De entre as várias áreas de desenvolvimento, a evolução da IoT permitiu a conceção e o crescimento da Indústria 4.0 como é atualmente denominada, viabilizando as interações entre vários dispositivos, melhorando substancialmente os processos de fabrico que têm génese na junção entre o mundo físico e o mundo virtual. De maneira análoga, pode-se considerar que o processo produtivo de uma fábrica passa a funcionar como um enorme sistema central nervoso, capaz de comunicar em tempo real e monitorizando vários parâmetros em simultâneo.

Com base no referido no parágrafo anterior, será abordado os tipos de sensores a utilizar existentes no mercado, as suas mais-valias bem como as diferentes aplicações associadas a estes. Adicionalmente, os parâmetros a medir tiveram como génese o facto de o consumo de combustível ser sensível a diversos fatores, uns de baixa complexidade, como a combinação peso do veículo-carga, a inclinação da estrada e a velocidade do veículo, e outros de medição mais complexa ou ambígua, tais como o tipo de condução do motorista ou as rotações do motor, e por essa razão, estes últimos fatores foram excluídos do âmbito desta dissertação.

Nos subcapítulos seguintes, proceder-se-á à descrição dos métodos de sensorização utilizados para a obtenção de informação.

2.1.1. Aceleração e Inclinação

Como referido anteriormente, a aceleração associada ao veículo pesado e a correspondente inclinação longitudinal associada ao traçado da estrada são dados de extrema importância para a estimação do consumo de combustível. A título exemplificativo, através da medição da aceleração no eixo vertical (eixo zz) com um acelerómetro, podemos perceber se o veículo circula numa via com excelentes condições de pavimentação ou, por outro lado, se a mesma é não pavimentada e com acentuada irregularidade longitudinal, pois as condições da via podem ser bastante determinantes no consumo de combustível.

Considera-se que um acelerómetro é um sensor elétrico que afere a aceleração própria de um sistema. A maneira com essa aceleração é aferida em comparação com outro sistema em queda livre, e caso este esteja em repouso, o seu valor corresponde a uma aceleração vertical de $1g = 9,81 \text{ ms}^{-2}$, caso esteja em queda livre, o valor da aceleração vertical é nulo. Estes dispositivos têm larga utilização em aplicações industriais, como por exemplo, para detetar vibrações em sistemas mecânicos.

Compostos por circuitos integrados de silício, ou MEMS (*Micro Electro-Mechanical Systems*) é a designação que pode ser dada ao funcionamento físico destes sensores. Basicamente, são considerados sistemas mecânicos, no entanto o seu comportamento é gerado por circuitos eletrónicos, sendo constituído por peças mecânicas microscópicas que se podem associar à eletrónica. É possível aferir a aceleração através da variação da capacitância do condensador MEMS, conforme demonstrado na figura 2.

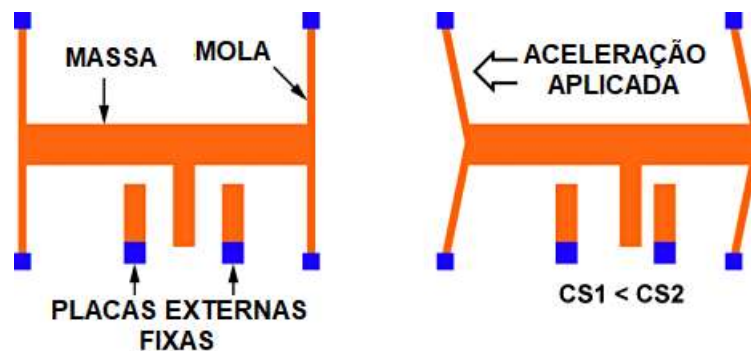


Figura 2.1 - Representação da ação da aceleração num condensador MEMS (adaptado de Kopfinger e Ahlsén, 2019)

Analisando a Figura 2.1, à medida que as molas são expostas a uma aceleração e dependendo da direção em que ela se move para frente ou para trás, a distância entre as placas fixas e as placas presas à mola altera-se e ocorre uma variação da capacitância (Kopfinger e Ahlsén, 2019). Certos fatores podem provocar uma reação na referida massa tais como vibrações, movimento ou ainda a gravidade. Existindo um movimento da massa, a distância de separação dos pinos do condensador é modificada o que origina uma variação de capacidade elétrica, compondo assim um sistema eletromecânico capaz de caracterizar movimento. Consequentemente, é gerado um circuito

capaz de detetar esta variação de capacidade e transformá-la numa grandeza quantificável, maioritariamente em Volts, com o intuito de permitir a outros dispositivos a sua leitura (Jost@, 2021).

Conforme já referido anteriormente, a inclinação é um dos fatores mais influentes no âmbito do consumo de combustível, assim sendo, uma rampa de elevada inclinação efetuada por um veículo pesado origina um aumento significativo do consumo comparativamente a descidas ou a uma condução em perfil horizontal. Tendo em conta esta influência, assume uma enorme importância a capacidade de se perceber e quantificar se o veículo pesado se encontra numa via horizontal ou inclinada e desse modo relacioná-lo com o consumo efetuado numa rota. Nesse seguimento, de modo a se obter a inclinação, são utilizados giroscópios, denominação de sensores capazes de aferir inclinações.

Por definição, o giroscópio é um dispositivo utilizado para medir a orientação e a velocidade angular. Os giroscópios tradicionais são compostos por um disco em volta de um eixo de rotação livre de assumir a orientação por ele próprio. Quando este se encontra a rodar, a orientação do eixo não é afetada devido ao princípio da conservação do momento angular. A Figura 2.2 mostra os principais componentes de um giroscópio tradicional.

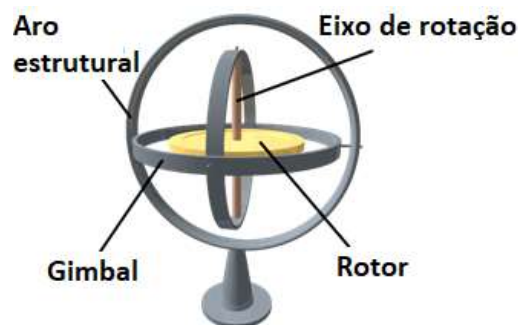


Figura 2.2 - Giroscópio tradicional

Paralelamente ao acelerómetro, o conceito de giroscópio também foi introduzido nos circuitos integrados MEMS. Podemos encontrá-los, a título de exemplo, nos *smartphones*, pois são eles que estão na base da rotação do ecrã, nomeadamente, quando há rotação numa direção, o movimento é acompanhado por uma rotação de 90° do ecrã nesse mesmo sentido.

A junção de um acelerómetro e um giroscópio no mesmo chip designa-se de IMU (*Inertial Measurement Unit*). Para além destes dois sensores, os IMU poderão ter associados também um magnetómetro. Estes dispositivos são maioritariamente utilizados em navegação de meios de transporte.

2.1.2. Sistema de Posicionamento Global

A obtenção da posição, como já foi referido anteriormente, é fundamental para o levantamento da informação e dados que permitam estimar o consumo de combustível. Sendo algo que faz parte do quotidiano, é fácil intuir que a forma mais usual de se obter a posição é com recurso ao GPS (*Global Positioning System*). Este termo foi largamente expandido e utilizado por ser pioneiro, todavia, não é o termo mais correto, pois é a denominação de apenas uma parte de um sistema global e mais complexo. A origem dos sensores de sistema de posicionamento global está associada à investigação militar nos Estados Unidos da América e somente em 1993 se tornou acessível para a utilização pública (Canales, 2018). Por ser um termo original, facilmente se criou o equívoco de invocar o seu nome quando se fala de um sistema global de posicionamento, mas este pertence apenas a uma das constelações que englobam o *Global Navigation Satellite Systems* (GNSS). A sigla GNSS designa uma constelação de satélites que fornecem cobertura acerca da posição e tempo aos dispositivos capazes de capturar este sinal e que são os recetores GNSS. Deste sistema fazem parte cinco constelações de satélites, incluindo o famoso GPS, que permite obter uma cobertura global planetária e, consequentemente, a obtenção de um vasto sinal à escala planetária e que se podem resumir no Quadro 1 e ilustrar na Figura 2.3.

Quadro 2.1 - Resumo das partes integrantes do GNSS

Designação	País
GPS	EUA
Galileo	UE
BeiDou	China
Glonass	Rússia
QZSS	Japão

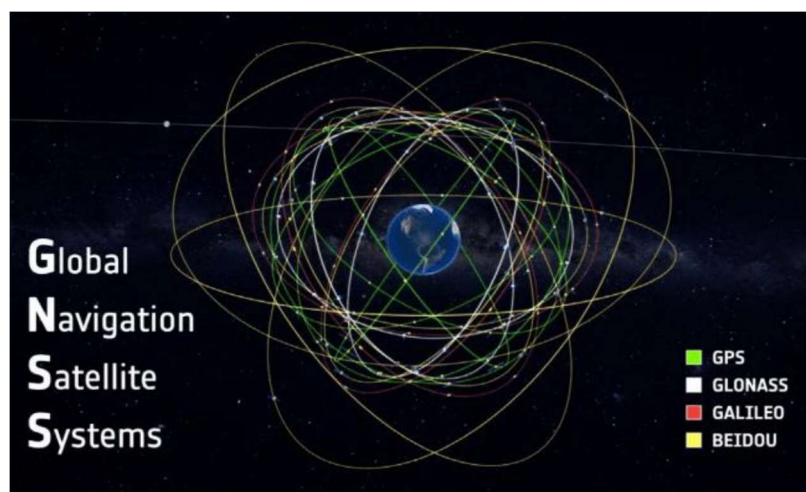


Figura 2.3 - Exemplo da cobertura dos satélites (Drotek@, 2021)

Com este sistema é possível também obter uma posição tridimensional. Contudo, é necessário que a antena do sensor consiga captar sinal de quatro constelações diferentes pois, caso contrário, o recetor consegue apenas uma posição aproximada assumindo que este se encontra à altura referente à linha do mar.

Após a receção do sinal enviado pelos diferentes satélites, o recetor calcula a distância a que estes se encontram e obtém a posição exata do utilizador, podendo alguns sensores incluir tecnologias em que a sua sensibilidade pode ir à escala dos centímetros. Dentro de uma diversidade de variáveis, este tipo de sensores, para além da posição e através do retorno das coordenadas polares tais como a latitude, longitude e altitude, permitem-nos obter a velocidade, os erros associados à leitura de posição, entre outras informações que sejam pertinentes extrair numa ótica de utilizador.

2.2. Machine Learning

O *machine learning* (ML) é um dos vários ramos da IA, tendo por base análises de bases de dados que podem automatizar a tomada de decisões (Mitchell, 1997). A IA tem vindo a ser desenvolvida desde 1956, altura em que foi abordada pela primeira vez no encontro do Dartmouth Summer Research Project sobre “Artificial Intelligence” em New Hampshire, Estados Unidos da América (Solomonoff, 1985). A IA é um ramo da ciência informática que incide na investigação, projeto e aplicação de computadores inteligentes (Poole et al, 1998). Esta tem sido desenvolvida com base numa interação multidisciplinar, de entre elas, a ciência da computação, cibernética, teoria da informação, psicologia, linguística e neurofisiologia.

A IA é uma ciência que incide na pesquisa e aplicação das leis das atividades da inteligência humana. Tem sido um assunto abordado mundialmente atingindo grandes desenvolvimentos na atualidade. Com efeito, esta tecnologia é aplicada atualmente em muitos campos, como o sistema especialista, sistema de conhecimento base, sistema de base de dados inteligente, sistema de otimização inteligente e o sistema de robô inteligente. Destes ramos, o sistema especialista é o mais antigo e com maior extensão no seu conhecimento sendo que pode ser descrito como um sistema informático que imita a capacidade de tomada de decisão de um ser humano e foi designada como a tecnologia geradora de conhecimento e de tomada de decisão do século (Lu et al., 2012). No campo da engenharia civil, especialmente na conceção e projeto, gestão da construção e tomada de decisão, muitos problemas foram influenciados por várias incertezas que podiam ser resolvidas não só na base de matemática, física e cálculos mecânicos, mas também da experiência dos utilizadores. Porém esse conhecimento e a experiência poderão, por vezes, ser incompletos e complexos, não podendo ser tratados por procedimentos tradicionais. No entanto, a IA tem as suas próprias vantagens pois pode resolver problemas complexos a um nível muito especializado, através da imitação das tomadas de decisão de especialistas. Em suma, a IA pode ter uma ampla perspetiva de aplicação na prática da engenharia civil.

A génese existencial do ML passa por aprender com padrões e tendências em extensas quantidades de informação - bases de dados - e transformar essa informação em conhecimento. Este conhecimento poderá ser

utilizado para uma melhor compreensão dos dados, para previsões de aspetos desconhecidos ou futuros, entre outros, e é este tipo de capacidade que dá origem ao apoio da tomada de decisão dos utilizadores.

2.2.1. Modelos

No presente subcapítulo, serão apresentados os modelos de ML mais relevantes para o âmbito desta dissertação e a sua escolha baseou-se noutros trabalhos feitos com sucesso nesta temática.

Decision Trees

O método das *Decision Trees* é utilizado para classificação e regressão de dados e tem como objetivo criar um modelo onde seja possível prever um valor através da aprendizagem de regras simples, tendo como referência os dados inseridos (Yiu@, 2021)

As *Decision Trees* formam o seu treino a partir dos dados, de modo a gerar uma aproximação de uma curva sinusoidal através das regras de decisão *if, then* e *else*. A Figura 2.4 apresenta um exemplo esquemático do funcionamento de uma *Decision Tree*.



Figura 2.4 - Exemplo esquemático do funcionamento de uma *decision tree* (adaptado de Yiu@, 2021)

Este modelo, tem como característica e vantagem o facto de ser de fácil visualização e interpretação, que necessita de poucos dados para ser treinado eficientemente. O *white box model* está na génese deste modelo, logo, se uma dada situação é constatada num modelo, é utilizada a lógica de modo a explicar as condições. Em contrapartida, pode existir uma elevada instabilidade e volatilidade pelo facto de necessitar de poucos dados, já que, uma ligeira variância no *dataset* pode levar a um resultado completamente díspar.

Random Forest

O método *Random Forest*, como o próprio nome indica, consiste num elevado número de árvores de decisão individuais que operam como um conjunto. Cada árvore individual neste método, exibe uma previsão de classe e, por sua vez, a classe com mais ocorrências torna-se a previsão do nosso modelo. A Figura 2.5 apresenta um exemplo esquemático de uma *Random Forest*.

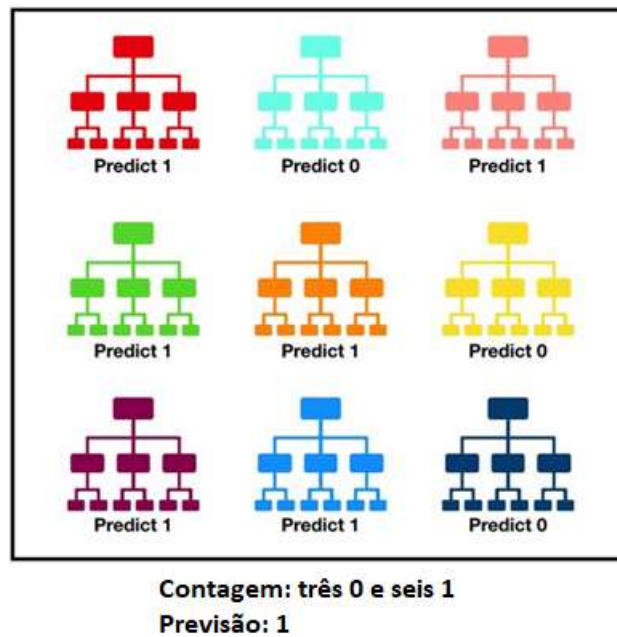


Figura 2.5 - Exemplo esquemático de uma *Random Forest* (adaptado de Yiu@, 2021)

O conceito fundamental por trás da *Random Forest* é simples, mas poderoso, pois exprime a sabedoria de uma maioria. A fundamentação científica do motivo pelo qual o modelo de *Random Forest* tem dado resultados melhores que os outros modelos está no facto de um grande número de modelos (árvores) não correlacionados que funcionam como um conjunto terem um desempenho melhor do que qualquer um desses modelos a funcionar individualmente.

A baixa correlação entre os modelos é a chave para o sucesso, pois os modelos não correlacionados podem produzir previsões em conjunto mais precisas do que qualquer uma das previsões individuais. A razão para este efeito é que as árvores se protegem umas às outras dos seus erros individuais (sob a condição de que elas não errarem constantemente na mesma direção). Embora algumas árvores possam estar erradas, a maioria das árvores estarão certas, então, como um grupo, as árvores são capazes de se mover na direção correta. Portanto, os pré-requisitos para um bom desempenho das *Random Forests* são, primeiramente, haver indícios concretos nos dados que indiquem que os modelos construídos por eles vão ter um desempenho melhor em termos de previsão do que uma previsão aleatória e depois que as previsões e, conseqüentemente, os erros feitos pelas árvores individuais tenham correlações baixas entre si.

Artificial Neural Networks

As *Artificial Neural Networks* (ANN) e correlacionam-se com as atividades do cérebro humano e podem ser caracterizadas como processos de distribuição paralela, em que simples unidades de processamento possuem uma habilidade natural de armazenar o conhecimento de experiências e com isso retirarem ilações. Consequentemente, as ANN requerem a existência de sistemas computacionais com a habilidade de aprenderem empiricamente, proporcionando a resolução de problemas mediante características não lineares. Na gênese da sua criação está a vontade humana em configurar máquinas inteligentes e qualificadas em tomadas de decisão o mais idênticas possível ao cérebro humano. Um grande número destas redes é instruído através do treino, apreendendo informação através de exemplos. O *neural computing* requer que um certo número de neurónios se conecte entre si, originando a rede neuronal. Neste processo, cada neurónio executa um processo simples, que conectados numa rede, devolverá um *output*. Por sua vez, cada *input* tem um peso associado que altera a relevância de cada neurónio. A Figura 2.6 apresenta o esquema de uma ANN.

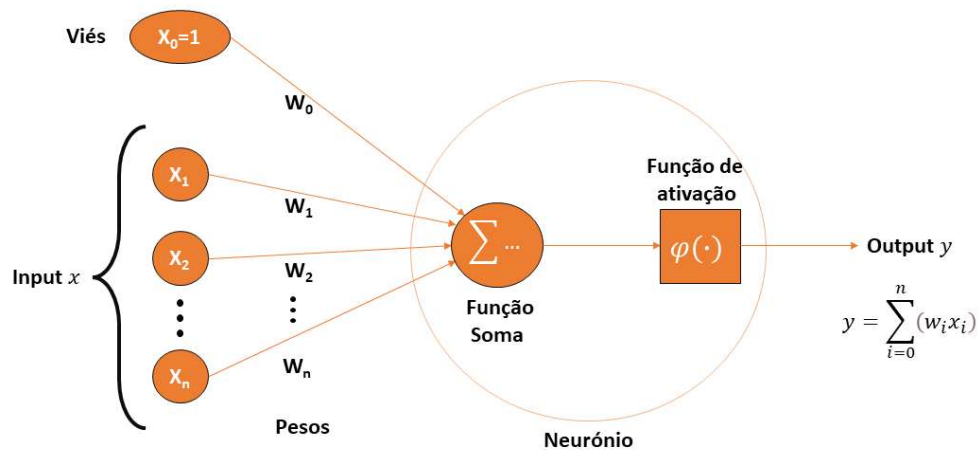


Figura 2.6 - Esquema de uma *artificial neural network* (adaptado de Santos@, 2021)

Algoritmos distintos com regras bem delineadas são aplicados no método de treino destas redes neuronais. Consequentemente, dois métodos distintos, usualmente denominados de paradigmas, são utilizados para o processo de aprendizagem – *supervised learning* e *unsupervised learning* – sendo que, no caso do primeiro, o *input* e o *output* são fornecidos. A rede processa os *inputs* e existe uma comparação entre os *outputs* desejados e os obtidos. As discrepâncias são enviadas ao sistema, que origina um ajuste no peso de cada neurónio, repetindo este processo várias vezes. Por sua vez, no segundo método, os *inputs* são introduzidos, mas não é dada a informação à máquina sobre o *output* desejado e, assim, o próprio sistema determina a melhor opção sobre o agrupamento dos dados.

Contudo, as redes neuronais artificiais também apresentam certas desvantagens. Olhando para o seu funcionamento, os dados inseridos têm de ser maioritariamente de boa qualidade, o que no mundo real é muitas vezes impossível. Pode acabar por se obter de forma lenta os resultados devido à necessidade de ter de se treinar

o modelo e fazê-lo aprender o que é necessário obter. Adicionalmente, os modelos resultantes são tipicamente uma *black box*, ou seja, não é fácil conseguir compreender o funcionamento do próprio modelo no contexto do processo usado para transformar os *inputs* em *outputs*, o que por vezes representa uma desvantagem em comparação com outros modelos de previsão.

Support Vector Machines

O modelo *Support Vector Machines* (SVM) tem como base a estatística e é característico deste uma elevada capacidade de aprendizagem, mesmo em casos que os dados são complexos, podendo ser utilizado para classificação ou para análises de regressão (Vapnik et al., 1997). Tendo como referência um certo *dataset*, o SVM encontra uma solução única e global, designando um hiper-plano que cria uma conexão entre dois tipos de dados. Numa perspetiva de espaço bidimensional este hiper-plano torna-se uma linha (Agrawal@, 2021). A Figura 2.7 ilustra uma representação gráfica do SVM.

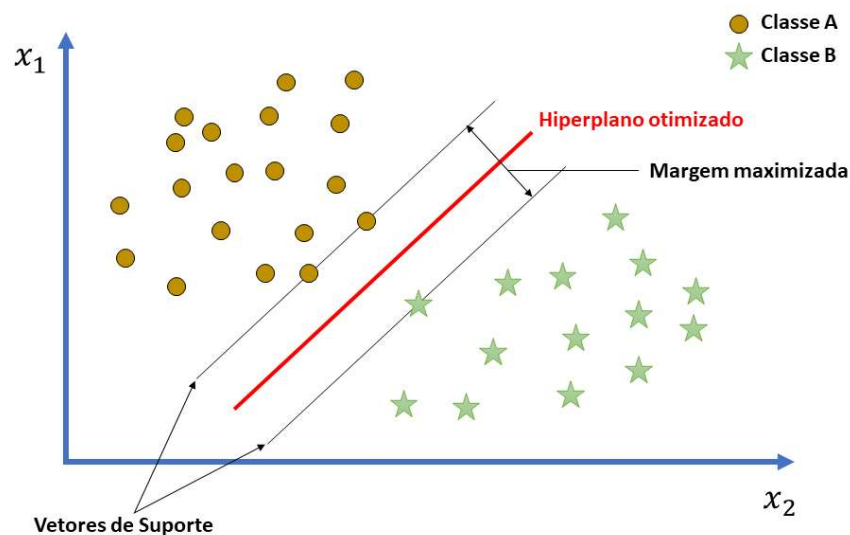


Figura 2.7 - Representação gráfica do SVM (adaptado de Agrawal@, 2021)

Este método organiza os dados por classes, utilizando planos de decisão e, conseqüentemente, encontra uma solução otimizada, maximizando a distância entre os dois planos. Isto significa que, se não houver pontos perto da linha de decisão, então não existem incertezas no que diz respeito às decisões da classificação.

Contudo, não obstante os SVMs serem máquinas de aprendizagem linear, eles acabam por ser não lineares no que concerne ao seu espaço original. A popularidade deste modelo vem do facto de ter a capacidade de combinar as vantagens de modelos lineares e não lineares, tal como a previsão dos resultados obtidos em cada domínio. Efetivamente, a sua capacidade de associar as vantagens de modelos lineares e não lineares assim como a previsão dos resultados obtidos em cada domínio, confere-lhe uma elevada popularidade.

2.2.2. Tarefas de Machine Learning

Na atualidade, existe uma larga variedade de algoritmos de ML em que a escolha da sua utilização varia conforme o propósito que o utilizador deseje. Em grande parte, eles são aplicados em previsão e descrição. A Figura 2.8 mostra os modelos preditivos e descritivos associados às tarefas do ML.

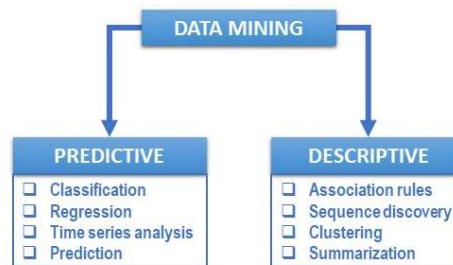


Figura 2.8 - Tarefas do *machine learning*

Analisando o modelo preditivo, este concentra-se na tomada de decisões gerando um modelo apto a prever ou estimar valores e no que concerne a previsões as mais solicitadas são as classificações e as regressões. Nestas, a classificação abarca a conversão dos dados de entradas em categorias pré-estabelecidas, para além de criar uma série de modelos ou funções que tenta caracterizar diferentes classes, seguindo as análises realizadas aos dados, de forma a poder categorizar novas entradas tendo em conta o modelo de classificações. Considerado este o objetivo mais usual, variadíssimas vezes é complementado por outros métodos tais como as redes neurais e *decision trees*. A finalidade da regressão é prever valores futuros ou incógnitos de variável dependente, originando modelos a partir de *databases* coletadas. Redes neurais artificiais, regressões múltiplas e *support vector machines* são os algoritmos geralmente aplicados neste panorama.

Por sua vez, o modelo descritivo tem no seu âmbito o incremento do conhecimento da máquina e da compreensão dos dados recebidos, concentrada na descoberta de padrões perceptíveis à ação humana. Elencando, os desafios mais comuns passam pela segmentação ou evidenciação de sequências, – onde é pretendida a formação de grupos de objetos o mais homogêneos possível, através de, por exemplo, a procura de similaridades no *dataset* passando assim a existir um potencial *clustering* durante várias séries – pela visualização – de modo a caracterizar a informação completa com o auxílio de diagramas, beneficiando assim a representação visual e interpretação de padrões e sequências – pela associação – que reconhece dependências entre variáveis evidenciando as mais relevantes – e pela sumarização – com o enfoque na elaboração de uma caracterização compacta para um maior *dataset*, de forma a permitir a geração automática de resultados.

Conclui-se que, entre a variedade de tarefas apresentadas, não existe uma técnica de uso consensual, já que o ML pode ser utilizado para responder a uma elevada quantidade de problemas e questões, e passará sempre por, iterativamente, aplicar várias técnicas de modo a determinar a que mais se adequa à qualidade dos resultados pretendidos ou esperados.

2.2.3. Processos dos modelos

No âmbito dos processos dos modelos, para facilitar a compreensão, implementação e análise de um processo é importante associarmos metodologias, sendo que a CRISP-DM (*Cross Industry Standard Process for Data Mining*) caracteriza-se por ser uma das mais utilizadas e completas destas metodologias. Esta tem a vantagem de ser transversal à indústria onde é implementada e às ferramentas utilizadas, assim os utilizadores tendem a usá-la como uma base estrutural para guiar todo o processo, independentemente da escolha das ferramentas (Saltz et al., 2017). Caracteriza-se como um processo hierárquico, sequencial e com um ciclo de vida que é desenvolvido em seis fases: compreensão do assunto, compreensão de dados, preparação de dados, modelação, avaliação e implementação. A Figura 2.9 mostra o diagrama referente ao CRISP-DM.

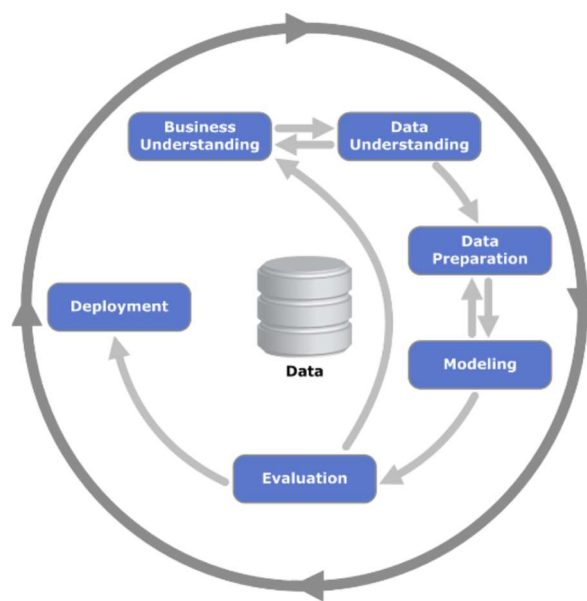


Figura 2.9 - Diagrama referente ao CRISP-DM (Saltz et al., 2017)

O Quadro 2.2 apresenta a caracterização das fases relativas ao CRISP-DM.

Quadro 2.2 - Descrição das fases referentes ao CRISP-DM

Fase	Descrição
Compreensão do assunto	Identificar e compreender os objetivos do projeto, posteriormente, converter na definição de problema de <i>data mining</i> e estabelecer um plano prévio com foco no objetivo
Compreensão de dados	Recolher e analisar os dados de modo a aferir a respetiva qualidade, descobrir pontos relevantes e detetar padrões/sequências. Posteriormente, para informação omitida pode-se formular algumas hipóteses

Fase	Descrição
Preparação de dados	Compilar o <i>dataset</i> final que irá ser utilizado durante a fase de aprendizagem (modelação) de modo a elaborar o modelo final. Selecionar registos e atributos dos dados <i>raw</i> iniciais e organizar os mesmos
Modelação	Selecionar o algoritmo de <i>machine learning</i> mais adequado e otimização dos seus parâmetros de modo a descobrir padrões no <i>dataset</i> utilizado
Avaliação	Avaliar minuciosamente todos os modelos e elaborar uma revisão referente aos passos anteriores de modo a verificar o sucesso do objetivo final
Implementação	Organizar os conhecimentos obtidos e as suas implementações, de modo a facilitar o manuseamento pelo utilizador

2.3. Casos da bibliografia

O potencial das aplicações de ML em Infraestruturas de Transporte e Geotecnia tem sido alvo de um enfoque considerável na última década (Gomes Correia et al., 2012). Com efeito, acompanhando o desenvolvimento cada vez maior de tecnologias de monitorização remota e armazenamento de dados, as bem-sucedidas aplicações de ML neste campo abrangem diferentes áreas, como por exemplo, produtividade em terraplenagem (Parente et al., 2014; Hola e Schabowicz, 2010), segurança de taludes (Tinoco et al., 2017) e resistência à compressão de *jet grouting* (Tinoco et al., 2011), gestão, manutenção e monitorização de pavimentos (Ma et al., 2019; Zhao et al., 2019). Estes podem frequentemente abordar processos específicos, tais como a estimativa da taxa de trabalho de compactação (Marques et al., 2008), ou do tempo de ciclo da escavadora (Edwards e Griffiths, 2000; Tam et al., 2002), bem como compreender uma parte fulcral de sistemas maiores e mais complexos, como os sistemas de gestão e de atribuição de frotas (Parente et al., 2016), ou sistemas de conceção, dimensionamento e gestão de pavimentos (Ma et al., 2019; Zhao et al., 2019).

Nesta última área também se tem observado um crescente número de aplicações no contexto da avaliação e manutenção das condições do pavimento (Souza et al., 2018; Nunes e Mota, 2019; Majidifard et al., 2020). Um aspeto notável destes sistemas está relacionado com o facto de alavancarem conceitos como a sensorização e *digital twins* para recolher dados, que, por sua vez, fornecem a base de dados de treino e teste. Por outras palavras, os modelos preditivos nestes sistemas de gestão de pavimentos são treinados com dados provenientes de diferentes sensores colocados em campo, quer nos veículos de inspeção (Souza et al., 2018; Nunes e Mota, 2019) quer no próprio pavimento (Ma et al., 2019; Zhao et al., 2019; Majidifard et al., 2020).

Não obstante, existe uma menor atenção na estimativa dos custos associados aos processos de construção. Em particular, a estimativa do consumo de combustível através de um modelo preditivo de aprendizagem automática é um tema que, apesar de ter tido alguns desenvolvimentos noutros domínios, como a logística e as rotas dos camiões de longo curso (Svard, 2014; Perrotta et al., 2017; Schoen et al., 2019), não foi feito um esforço equivalente no domínio da construção. Nos grandes projetos de infraestruturas de transporte, o combustível é um dos principais custos associados à execução de processos de construção e a sua estimativa baseada em dados é uma vantagem muito significativa para os gestores de projeto e projetistas. Normalmente, as decisões relativas à alocação de máquinas pesadas, planeamento de tarefas ou avaliação de desempenho são efetuadas através da utilização de estimativas de consumo de combustível baseadas na experiência profissional dos operadores ou em documentos e guias genéricos, tais como o manual de desempenho da CATERPILLAR (Caterpillar Performance Handbook, 2018). Assim, a disponibilidade de informação baseada em sensores em contexto real permite que as equipas de decisão possam gerir adequadamente os seus recursos (tanto os económicos como os físicos, como por exemplo, os equipamentos de construção) não só durante as fases de conceção e planeamento, mas também durante as fases de construção propriamente ditas, em que a atribuição de recursos pode ser ajustada em função de restrições e ocorrências imprevisíveis (por exemplo, avaria de equipamentos ou mau funcionamento do equipamento por baixa produtividade).

Grande parte dos estudos realizados acerca da previsão de consumo de combustível baseia-se em simulações através de cálculos físicos e geralmente são lentos a executar, pois simulam os componentes internos dos camiões (Sandberg, 2001; Nasser et al., 1998). Um exemplo disso, é o *Scania Truck and Road Simulation (STARS)*, que se trata de um sistema de simulação, que apenas prevê o consumo após obter certas configurações relativas ao veículo e ao condutor (Svard, 2014). Uma desvantagem acerca deste tipo de previsões é que as simulações requerem tempo e configurações manuais. À medida que se acrescentam mais parâmetros, mais demorada acaba por ser a simulação devido ao aumento de complexidade.

Existem bastantes estudos relacionados com a previsão de consumos nos diferentes tipos de transporte através de algoritmos de ML, servindo como principal fonte de inspiração o STARS. Devido à qualidade do estudo, existem muitos projetos derivados deste. Porém, muito poucos conseguiram concluir se existe a possibilidade de fazer a necessária previsão partindo dos algoritmos de ML.

Um estudo implementou um modelo inspirado no STARS usando *random forest* e *gradient boosts* (Viswanathan, 2013) e concluiu que o primeiro obtinha melhores resultados assim como a maior relevância dos parâmetros velocidade, descidas sem consumo, distância com atrelado, distância com *cruise control* e velocidade máxima. Porém, só teve em consideração os parâmetros do comportamento do condutor e não incluiu as características da estrada, do veículo nem as condições atmosféricas. Não abordou a maneira de como treinar o modelo para a escolha de rotas nem deteção de anomalias.

Um estudo realizado (Lindberg, 2014) tentou elaborar um modelo preditivo usando dados acerca da posição, quantidade de combustível e outras características inerentes ao camião, a serem enviadas com uma dada

frequência. Após várias recolhas seguindo uma estrada entre duas cidades diferentes, ao aplicar os diferentes modelos, este não terá chegado a nenhuma conclusão definitiva, existindo a dúvida de qual modelo teria melhores resultados. Porém, conseguiu concluir que o peso do veículo e a inclinação da estrada seriam as variáveis mais relevantes para as suas previsões. Devido à limitação dos dados e à sua baixa *sample rate*, os resultados publicados mostram uma acuidade de apenas 26%.

Outro estudo confirmou que era possível efetuar este tipo de previsões (Svard, 2014), delineando até, que o peso da carga do veículo e a inclinação seriam os fatores mais determinantes nos cálculos associados. Tal como o tema abordado neste relatório, o comportamento do motorista, a velocidade de ralenti e outras variáveis que influenciam o consumo não foram abordados. Foram treinados diferentes modelos, tais como *decision trees*, redes neuronais artificiais, modelos de regressão linear e SVM. Este último modelo acabou por ser o mais produtivo dentro dos seus estudos, porém os seus resultados podem não ser fiáveis, pois terá usado dados diferentes quando treinou os restantes modelos.

Foi realizado um projeto semelhante no que diz respeito a questões logísticas (Almer, 2015). Contudo, apesar dos resultados bastante positivos na utilização de modelos de ML para a previsão de consumos de combustível, com uma grande variedade de parâmetros conforme demonstrado na Figura 2.10, o seu objetivo seria um pouco diferente, quer pela natureza do processo (viagens de longa distância de transporte de bens), quer pelas

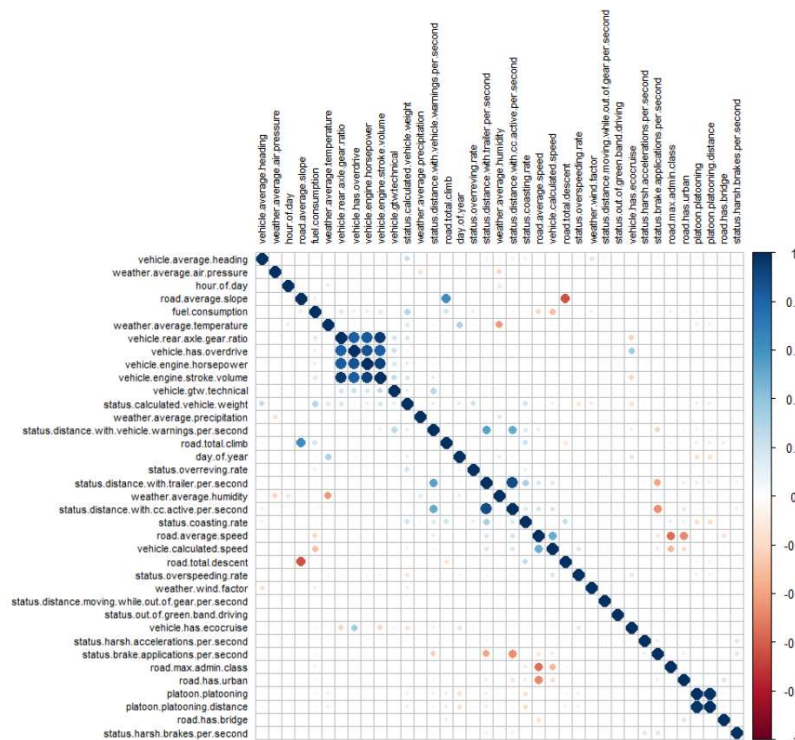


Figura 2.10 – Parâmetros e a sua correlação utilizados nos modelos de previsão de (Almer, 2015)

condicionantes do mesmo. Com efeito, as obras de vias de comunicação incluem viagens geralmente cíclicas, em terrenos variados (pavimentado, não pavimentado, estaleiro), com cargas e materiais variados, sendo no geral um ambiente bastante menos controlado. Assim, o presente projeto visa também acrescentar mais conhecimento ao estado da arte, especificamente para a área da construção e seus equipamentos.

3. CASO DE ESTUDO

No presente capítulo, será feita a apresentação do caso de estudo, todo o processo de fabrico do equipamento que monitorizou a informação sobre os parâmetros necessários para estimar e relacionar o consumo de combustível às condições (inclinação, combinação peso do veículo-carga, velocidade) em tempo real. As razões por detrás da escolha do hardware, abordando a questão da escolha dos sensores e os dados recolhidos pelos mesmos. Informação e descrição sobre a sua instalação no equipamento de obra, excertos relativos aos primeiros testes realizados, verificando o correto funcionamento dos sensores utilizados, tal como uma primeira análise a uma base de dados obtida a partir da sensorização. Posteriormente, a segunda etapa do processo, a receção e validação dos dados obtidos bem como o tratamento e ferramentas utilizadas para a angariação de resultados que irão ser comentados no capítulo seguinte. De referir que este trabalho tem a sua génese numa parceria entre a Built CoLab, um laboratório de desenvolvimento tecnológico para as empresas, o Instituto Superior Técnico e a JJR, empresa de referência no setor de obras públicas.

3.1. Implementação

3.1.1 Hardware

Tendo como premissa o conhecimento dos dados que se pretende recolher e conhecendo as ofertas existentes, acabou por se optar por três tipos de sensores diferentes:

- ⇒ GNSS receiver
- ⇒ Acelerómetro
- ⇒ Giroscópio

Para efeitos de armazenamento, leitura de sensores e futuro processamento de algoritmos de ML em tempo real, utilizou-se o microcomputador NVIDIA Jetson Nano 2GB Development Kit.

GNSS Receiver

A partir do GNSS 7 click, implementado com o módulo recetor NEO-M9N como apresentado na Figura 3.1, é possível obter uma quantidade variada de dados tais como latitude, longitude, altitude, velocidade, data e hora, que são de elevado interesse. Os erros associados à latitude, longitude e altitude serão algo interessante para se captar caso os valores a serem enviados estejam perto do valor real. A partir daqui, consegue-se obter a posição exata do camião em tempo real e ao longo do tempo, que será a principal finalidade deste sensor. Este sensor tem incluída a antena que trata da receção da informação transmitida pelos satélites.



Figura 3.1 - GNSS 7 click (Mikroe@, 2021a)

Acelerómetro e Giroscópio

A partir do IMU 6DOF 15 click, representado na Figura 3.2, que contém um ASM330LHH (acelerómetro de 3 eixos e giroscópio de 3 eixos) [22], obtém-se a aceleração nos eixos longitudinal ao veículo, transversal (i.e., eixos das rodas), e vertical (i.e., perpendicular à superfície de circulação do veículo), correspondendo aos eixos dos xx , yy e zz , assim como a inclinação nos mesmos eixos xx , yy e zz . Estes dados revelam-se de extrema importância pois permitem saber a inclinação do troço onde o camião circula perfilando-se como um dado bastante significativo para o algoritmo de *machine learning* dado que esta tem uma grande influência no consumo. Apesar de estarem instalados no *framework* de IoT e sensorização do protótipo, os dados medidos por estes sensores foram usados apenas para validar as características do percurso em ação conjunta com o GNSS não sendo usados diretamente como *inputs* para os modelos de previsão no âmbito desta dissertação, embora se considere que será pertinente fazê-lo em desenvolvimentos futuros tal como irá ser discutido posteriormente no subcapítulo destinado para o efeito.



Figura 3.2 - IMU 15 click (Mikroe@, 2021b)

Microcomputador

Após a escolha de sensores, os dados obtidos têm de ser armazenados e processados. Para este efeito, utiliza-se o microcomputador da NVIDIA o modelo Jetson Nano 2GB Development Kit, representado na Figura 3.3 [25]. Este microcomputador revela inúmeros benefícios, tal como a existência de um fórum muito ativo com imensos exemplos de projetos desenvolvidos por uma comunidade que se procura ajudar entre si. A existência de um *40-pin header* possibilita a comunicação via UART, I2C e SPI entre este *developer kit* e os sensores, tal como um slot para cartões SD, utilizado para armazenar os ficheiros recolhidos. A comunicação entre o GNSS receiver e o microcomputador é feita via UART [26], ao contrário do IMU, que comunica via I2C [27]. O microcomputador é geralmente utilizado em aplicações de *computer vision*, robótica e sistemas de IoT uma vez que tem bom desempenho a correr algoritmos de *machine learning* de forma energeticamente eficiente.



Figura 3.3 - NVIDIA Jetson Nano 2GB (NVidia@, 2021)

Em comparação com o microcomputador mais dominante do mercado, o Raspberry Pi 4, o Jetson Nano 2GB contém uma interface gráfica mais poderosa, capaz de correr as mais famosas *frameworks* presentes neste tema, como PyTorch, TensorFlow, Keras, CUDA, etc. Posto isto, pode-se concluir que o Jetson é potencialmente mais adequado a aplicações de inteligência artificial comparativamente ao RPi 4.

Após o download da imagem do sistema operativo, esta é escrita num cartão SD, e após as configurações iniciais fica disponível um ambiente de desenvolvimento Linux distribuído pela NVIDIA denominado por NVIDIA JetPack.

3.1.2 Software

Esta secção refere-se à implementação dos sensores, através do desenvolvimento de um código capaz de cumprir os requisitos. Elaborado com o apoio do BUILT CoLAB na linguagem Python, que, segundo a Northeastern University, terá sido a linguagem mais utilizada numa escola global no ano de 2020, sendo uma alternativa ao clássico C. Apesar de C ser uma linguagem de mais baixo nível, com mais controlo em si mesma e na própria máquina, sendo até mais eficiente computacionalmente, o Python, invés, é uma linguagem de alto nível muito procurada devido à inúmeras quantidades de bibliotecas existentes e de possuir uma sintaxe relativamente simples que acelera o processo de prototipagem.

Este código foi elaborado de maneira a corresponder à criação de dois ficheiros .CSV, um que terá valores a serem impressos com uma mais elevada frequência, tais como a aceleração nos três eixos (x,y,z) e o *time stamp* e outro ficheiro com valores com uma frequência menos elevada, nomeadamente, a inclinação nos três eixos (x, y, z), posição (latitude, longitude e altitude) e erro estimado, velocidade e o *time stamp*.

A criação de dois ficheiros distintos vai de encontro à pretensão de, posteriormente, serem processados os dados. Sendo que, o primeiro, será utilizado para deteção do *relenti* do motor de combustão interna para aferir os momentos em repouso, estimando a rotação do mesmo pelas vibrações do veículo em gamas de frequência superiores (entre 1000 e 6000 Hz) originárias da rotação do motor recorrendo a um filtro de frequência. Por sua vez, no segundo ficheiro são gravados os restantes dados, aos quais corresponde uma frequência de amostragem inferior.

A comunicação entre o acelerómetro e o *Jetson* foi feita através do protocolo I2C (*Inter-Integrated Circuit*), tal como o sensor de distância. Foi inserido no sistema um botão que interage com um *GPIO* (*General Purpose Input/Output*) do *Jetson* que, após acionado, acende um LED RGB, neste caso de cor verde, também controlado via *GPIO*, transmitindo ao utilizador a informação de que os valores dos sensores estão a ser recebidos e registados num ficheiro .CSV. Premindo o botão uma segunda vez, o ficheiro é marcado como pronto a enviar, para o servidor. O utilizador pode repetir o processo para obter várias *runs* e estas ficam registadas em ficheiros distintos. Assim que o protótipo tem acesso à internet, os ficheiros guardados são enviados para um servidor via SCP (*Secure Copy Protocol*) sinalizado com a cor azul intermitente do LED RGB.

Para além disso, enquanto o acesso à internet se encontra disponível, é verificada a existência de uma nova versão do código alojado no GitHub designada para utilização em campo fazendo uso de diferentes *branch* de um repositório Git.

Através da implementação de *threads*, no caso de existir uma avaria num dos sensores, e este despoletar um comportamento erróneo do *software*, é possível continuar a recolha correta de dados dos restantes sensores. Esta escolha de arquitetura, permite também que as operações que bloqueiam a execução de código não comprometam a cadência de recolha de dados de outros sensores.

Por fim, de modo a poder existir um controlo relativamente à fiabilidade do hardware, foi criado um ficheiro .CSV adicional que regista a temperatura do CPU e do GPU do microcomputador, pois o protótipo fica alojado dentro do habitáculo de um camião, podendo existir a possibilidade de enfrentar temperaturas muito elevadas durante o Verão que acabarão por diminuir o tempo de vida do equipamento, tendo como ponto crítico a temperatura do equipamento quando é iniciado, visto que o sistema liga ao mesmo tempo que é inserida a chave na ignição, sendo que nesta altura ainda não é refletido o arrefecimento da cabine pelos sistemas de climatização.

3.1.3 Encapsulamento

Seguindo o processo de fabrico, foi necessário de seguida encapsular o protótipo para proteger todo o sistema, assim como, tornar mais *user-friendly* o seu aspeto e utilização para os futuros utilizadores. Para este efeito, ainda com o apoio do BUILT CoLAB utilizou-se o Tinkercad (Autodesk Tinkercad, 2021), um software online de modelações 3D freeware de fácil utilização. Contudo, tem como desvantagens apenas estar disponível como *web app* e ter poucas funcionalidades para este tipo de modelações, sendo mais orientado para utilizadores pouco experientes. Ao conceber a estrutura da caixa será necessário executar duas aberturas, uma para o LED e outra para o botão, mais três furos para aparafusar o protótipo à caixa. Adicionalmente, terá duas aberturas laterais, uma para a saída da antena do sensor GNSS e outra para as portas do microcomputador.



Figura 3.4 - Projeto desenvolvido no ambiente Tinkercad

Após a conceção esquemática da peça, esta foi produzida com o recurso a uma impressora 3D tendo como material base o PETG, visto que, entre os disponíveis, é aquele que tem uma temperatura de transição vítrea acima das temperaturas esperadas dentro da cabine do veículo, evitando desta maneira a deformação do encapsulamento. Finalizado este processo, resta aparafusar a plataforma de madeira (onde o microcomputador e os sensores estão fixados) ao encapsulamento e obtém-se assim o protótipo final pronto a ser utilizado conforme ilustrado na Figura 3.5.



Figura 3.5 - Protótipo Final

3.1.4 Instalação no equipamento de obra

Na sequência da parceria entre o Instituto Superior Técnico, o BUILT CoLAB e a empresa de construção JJR, foi possível testar o protótipo no seu ambiente final, nos equipamentos em ambiente de obra. A aceitação dos colaboradores desta empresa face aos procedimentos que teriam de ser efetuados pelos operadores foi um ponto positivo, pois sem a aceitação e cooperação destes não seria possível testar o projeto desenvolvido.

O passo seguinte seria proceder à instalação do protótipo. Este, à priori, tem como condição ser acoplado num sistema onde não haja oscilações da caixa, sendo que a sua instalação necessita de ser executada de uma forma não amortecida, de modo a não deturpar os valores referentes ao IMU. Ao mesmo tempo, à saída do estaleiro existia uma balança que procedia à pesagem da carga relativa ao sistema camião mais carga transportada, que assegurou a medição da carga do veículo em cada viagem. Na Figura 3.6 ilustra-se a instalação da antena GNSS.

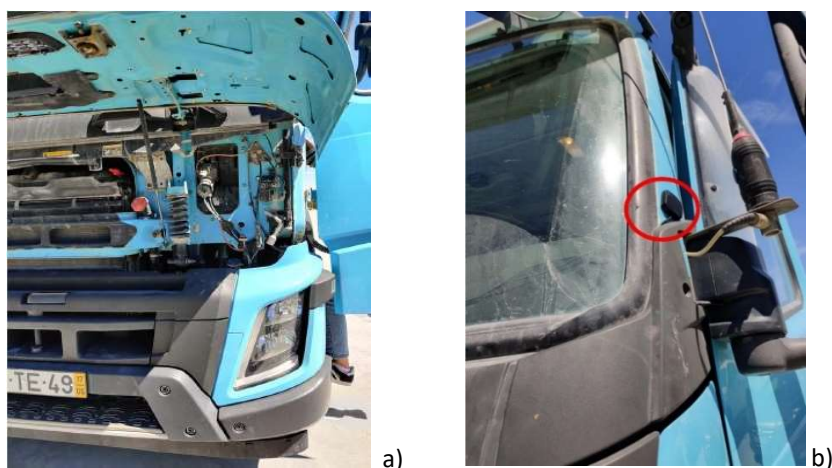


Figura 3.6 – Vista geral a) e pormenor b) da instalação da antena GNSS no camião

O primeiro passo foi proceder à instalação da antena do GNSS *receiver* no camião, optando-se por colocá-lo no exterior de modo a não reduzir a qualidade do sinal e dos dados recebidos, que aconteceria caso esta fosse instalada dentro do habitáculo. Foi possível passar os cabos por baixo do capô, desde o exterior ao interior do veículo e assim fazer a ligação ao protótipo instalado dentro do habitáculo.

Posteriormente passou-se à instalação do protótipo. Após análise aos locais mais viáveis onde este poderia ser colocado, foi decidido instalá-lo junto do motorista, já que este tem uma parte integrante do sucesso do procedimento de recolha dos dados através do manuseamento do protótipo, conforme demonstrado na Figura 3.7 .



Figura 3.7 - Instalação do protótipo

A integridade física do protótipo foi garantida com a sua aplicação sobre um material de borracha aderente, impedindo o seu deslizamento e certificando a sua fixação.

De seguida, foi necessário efetuar a conexão entre o microcomputador e a rede *Wi-Fi* existente no estaleiro, para que, quando o camião regressasse das suas viagens, poder enviar os ficheiros recolhidos para o servidor.

A alimentação do protótipo é feita através da ligação do isqueiro presente no camião que por sua vez é convertida para 5V que permite alimentar o protótipo via USB permitindo assim a sua utilização.

De modo a podermos obter o consumo real que o camião consumiu em cada viagem, foi solicitado ao motorista, antes de iniciar cada viagem, atestar completamente o depósito, utilizar as balanças que normalmente utilizam para pesar a carga, fazer a viagem, chegar ao destino e voltar a atestar o depósito, anotando a quantidade de

litros de combustível que foi atestada no caminhão para se aferir a quantidade real de combustível gasta para aplicação posterior no treino dos algoritmos de *machine learning*. Este procedimento é descrito na Figura 3.8.

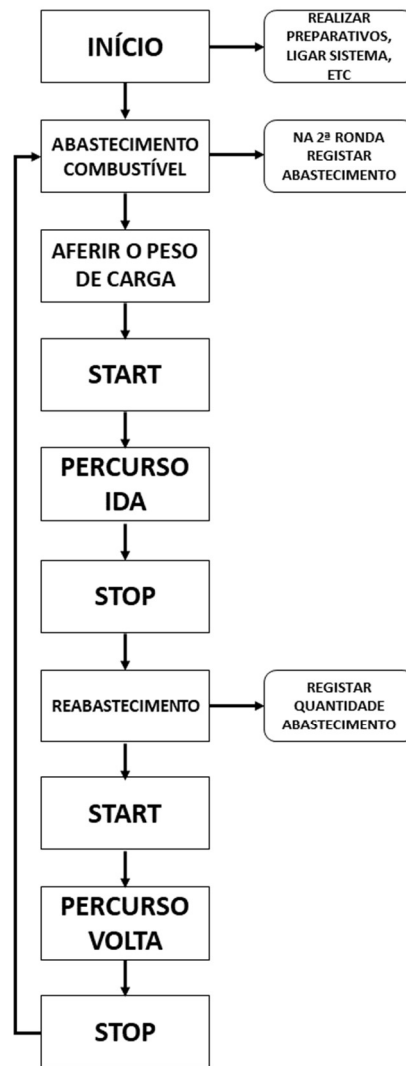


Figura 3.8 – Fluxograma das ações solicitadas ao motorista

Foi também fornecido aos motoristas uma folha de registos, onde estes terão de registar o consumo da última viagem e a peso da carga que este transporta, tal como a hora a que a viagem foi iniciada, sendo ainda disponibilizado um campo de observações com o intuito de ser reportada alguma anomalia no procedimento ou algum acontecimento relevante no percurso. Um exemplo deste tipo de registos é apresentado na Figura 3.9.

3.2. Apresentação e tratamento de dados

3.2.1 Avaliação da qualidade de modelos de machine learning

Neste trabalho, as técnicas implementadas focam-se em modelos de regressão. Como tal, a sua avaliação baseia-se principalmente em três métricas principais, nomeadamente o erro absoluto médio, MAE (Equação 3.1), que retrata o erro associado ao grau de aprendizagem de um determinado modelo, a raiz quadrada da média dos erros quadrados, a RMSE (Equação 3.2), que penaliza valores de erro mais elevados, e o coeficiente de correlação, R^2 (Equação 3.3), que compreende a correlação entre os valores observados e os valores previstos (Hastie et al., 2009).

$$MAE = \frac{\sum_{i=1}^N (y - \hat{y})}{N} \quad (3.1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y - \hat{y})^2}{N}} \quad (3.2)$$

$$R^2 = \left(\frac{\sum_{i=1}^N (y - \bar{y}) \times (\hat{y} - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^N (y - \bar{y})^2 \times \sum_{i=1}^N (\hat{y} - \bar{\hat{y}})^2}} \right)^2 \quad (3.3)$$

Onde:

y - é o vetor de output de rede computado,

\hat{y} - é o vetor de output alvo, e

N - é o número de amostras na base de dados.

Adicionalmente, a curva Característica de Erro de Regressão (REC) (Bi e Bennett, 2003) também foi adotada como uma medida da função de distribuição cumulativa do erro de diferentes algoritmos de regressão, permitindo uma análise comparativa entre estes últimos.

3.2.2 Dados não-processados

Como mencionado anteriormente, os dados são a base dos modelos ML. No entanto, o facto de estes algoritmos preditivos dependerem tão fortemente da disponibilidade de dados representa, simultaneamente, o seu maior potencial e a sua maior limitação. Com efeito, a capacidade preditiva de um modelo ML depende tanto da quantidade como da qualidade dos dados disponíveis. Por outras palavras, para se conseguir uma previsão perfeita do comportamento de uma variável-alvo, o modelo não só deve ser alimentado com dados suficientes, como estes dados devem também abranger todas as variáveis independentes que possam ter algum nível de

influência sobre esse comportamento. Além disso, a variabilidade dos dados deve ser consistente o suficiente para permitir que os algoritmos ML recolham conhecimento e discernimento sobre quanto cada variável independente afeta a variável-alvo, bem como, a maneira pela qual as variáveis independentes se ligam e se afetam entre si.

Neste contexto, pode-se facilmente encontrar vários conjuntos de dados pré-processados associados a diferentes campos que podem resultar em modelos quase perfeitos (ou seja, com precisão de previsão próxima de 100% no caso de modelos de classificação, ou margens de erro muito baixas no caso de modelos de regressão), dependendo da técnica ML adotada, como é o caso dos conhecidos VGG-16 (Simonyan e Zisserman, 2015) ou das bases de dados ResNet50 (He et al., 2016) no campo de classificação de imagem. No entanto, este nível de precisão/erro nem sempre é atingível quando se trata de dados do mundo real, uma vez que estes tendem a ser ruidosos e afetados por níveis mais elevados de falta de fiabilidade ou dados em falta. Embora exista uma variabilidade de ações que podem ser tomadas para lidar com estas questões durante a fase de depuração e processamento de dados (que serão discutidos na subsecção subsequente), os modelos típicos baseados no mundo real só muito raramente podem alcançar ajustes perfeitos aos dados, e assim, consequentemente, capacidades preditivas perfeitas.

Tendo em conta estas considerações, os dados utilizados neste trabalho resultam de um cenário real constituído por um estaleiro de construção rodoviária em Portugal. Os dados não processados são consequência das atividades sensorizadas de um veículo de transporte de materiais de pavimentação, com cerca de 21 viagens por diferentes vias e superfícies de pavimentos (tais como, estrada não pavimentada do estaleiro, estrada nacional e autoestrada), enquanto transporta diferentes materiais, nomeadamente e em grande parte, a mistura betuminosa do estaleiro onde é fabricada para um estaleiro de construção rodoviária. Além da carga e das superfícies do pavimento, as viagens apresentam também alguma variação no que diz respeito à distância total, que varia entre os 20 e os 60km. Os principais parâmetros medidos incluem tempo, dados de localização/GPS, altitude, velocidade e inclinação de 3 eixos do camião, tal como exemplificado no Quadro 3.1.

Quadro 3.1 – Valores extraídos da base de dados não-tratados

Inclinação_X	Inclinação_Y	Inclinação_Z	Latitude	Longitude	Altitude	Veloc.	Hora e data
0,778198	-0,29755	-1,31226	39,4447	-7,47812	447	0,043	2021-07-22T12:42:36.100Z
0,778198	-0,29755	-1,31226	39,4447	-7,47812	447	0,038	2021-07-22T12:42:36.400Z
0,778198	-0,29755	-1,31226	39,4447	-7,47812	447	0,038	2021-07-22T12:42:36.400Z
0,839233	-0,30518	-0,03052	39,4447	-7,47812	447	0,038	2021-07-22T12:42:36.700Z
0,839233	-0,30518	-0,03052	39,4447	-7,47812	447	0,013	2021-07-22T12:42:36.800Z

Inclinação_X	Inclinação_Y	Inclinação_Z	Latitude	Longitude	Altitude	Veloc.	Hora e data
0,839233	-0,30518	-0,03052	39,4447	-7,47812	447	0,014	2021-07-22T12:42:37.100Z
0,839233	-0,30518	-0,03052	39,4447	-7,47812	447	0,044	2021-07-22T12:42:37.500Z
0,923157	-0,28992	-1,95313	39,4447	-7,47812	447	0,004	2021-07-22T12:42:37.900Z
0,923157	-0,28992	-1,95313	39,4447	-7,47812	447	0,036	2021-07-22T12:42:38.200Z

Como pode ser facilmente deduzido, a taxa de atualização associada à estrutura IoT recolhe dados várias vezes por segundo, acabando por gerar bases de dados muito grandes sob a forma de ficheiros .csv, cada um correspondente a uma viagem do camião. Uma vez que estas viagens podem ter cerca de 70 quilómetros, os ficheiros de dados associados também aumentam proporcionalmente para cerca de 12500 entradas nesses casos. Nesta fase do projeto, a base de dados contém as mencionadas 21 rotas para este camião, transportando diferentes tipos (e pesos) de carga.

3.2.3 Tratamento de dados

A Figura 3.10 retrata um dos perfis altimétricos mais típicos através dos quais o camião realizou as suas viagens. Estes foram extrapolados integrando os dados de velocidade ao longo do tempo, e validados recorrendo aos dados da inclinação. A partir deste ponto, foi adotada uma metodologia (*sliding window*) de obter troços de 10 em 10 metros para ajustar as linhas ao longo do perfil altimétrico, permitindo a determinação da inclinação de cada secção de 10 metros. Através desta metodologia, cada viagem foi traduzida para a distância acumulada que o camião percorre em cada tipo de inclinação, de acordo com as considerações descritas no Quadro 3.2.

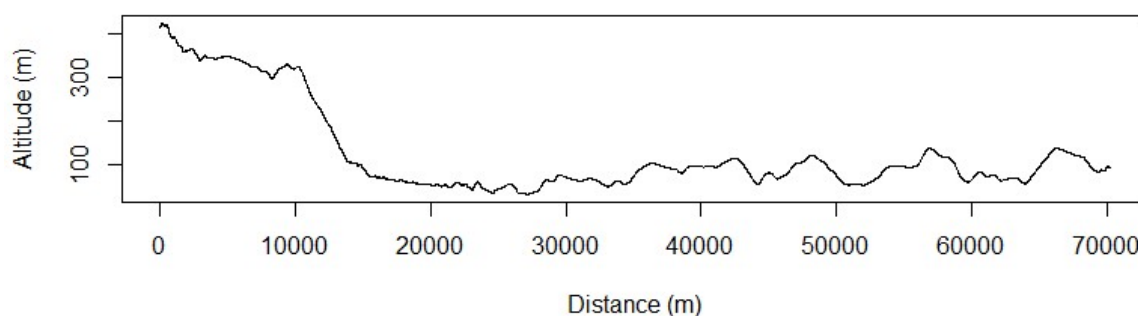


Figura 3.10 – Exemplo de um perfil altimétrico de uma viagem aferido pelo camião sensorizado

Quadro 3.2 - Intervalos e descrição da inclinação

Descrição da inclinação	Intervalo de valores	Denominação
Superfície plana	-1% < Inclinação ≤ +1%	AD_0.01n_0.01
Subida ligeira	+1% < Inclinação ≤ +5%	AD_0.01_0.05
Subida moderada	+5% < Inclinação ≤ +10%	AD_0.05_0.1
Subida íngreme	Inclinação ≥ +10%	AD_0.1
Descida ligeira	-5% < Inclinação ≤ -1%	AD_0.01_0.05n
Descida moderada	Inclinação ≤ -5%	AD_0.05n

Em última análise, esta conversão consistiu em determinar a percentagem da distância total de percurso que o camião passa em cada tipo de inclinação ao longo de cada viagem, resultando numa grande parte da base de dados de treino e teste. Estes últimos são complementados com os dados sobre a distância total da viagem (TDistance, metros), velocidade média (AvSp, metros por segundo), peso de carga (Carga, toneladas) e consumo de combustível (FConsumption, litros, variável alvo). As duas últimas características são inseridas manualmente na base de dados, uma vez que são originárias dos registos manuais realizados pelo condutor do camião entre viagens e durante cada ação de reabastecimento. O Quadro 3.3 retrata a base de dados processada que suporta os algoritmos finais de ML.

Quadro 3.3 – Valores retirados da base de dados de treino e teste

AD_0.01n_0.01	AD_0.01_0.05	AD_0.05_0.1	AD_0.1	AD_0.01_0.05n	AD_0.05n	TDistance	AvSp	Carga	FConsumption
0,32	0,25	0,01	0	0,36	0,06	66341,91	18,48	29,48	30
0,40	0,33	0,05	0	0,21	0,02	65704,09	20,52	0	23
0,20	0,33	0,13	0,01	0,23	0,11	52992,28	12,39	0	20,5
0,20	0,25	0,08	0,01	0,28	0,17	35887,31	9,15	33	25,5
0,19	0,25	0,10	0,01	0,28	0,17	36563,17	8,52	32,68	25
0,37	0,23	0,01	0	0,35	0,05	62786,77	17,40	33,76	26,5
0,43	0,30	0,05	0	0,21	0,01	62282,68	20,79	0	23,5
0,38	0,22	0,01	0	0,33	0,05	62786,77	18,91	32,46	26
0,41	0,31	0,05	0	0,22	0,01	62606,6	19,16	0	23

4. RESULTADOS E DISCUSSÃO

Os resultados foram obtidos com recurso ao pacote *rminer* (Cortez, 2010) para R. Com a utilização desta ferramenta, vários modelos na base de dados processada foram treinados e testados, nomeadamente, RF, ANN e SVM. Uma vez que a capacidade de generalização é uma preocupação fundamental, não só para implementação futura, assim como para a avaliação dos modelos, foi adotada uma abordagem de validação cruzada de 10 *runs*. A quantidade relativamente baixa de dados justifica o valor *k-fold* de 3. Isto significa que os dados são avaliados em todo o conjunto de treino, dividindo-os em 3 partes. O modelo é então treinado 3 vezes enquanto reserva uma parte diferente de cada vez, como um conjunto de dados de teste, usando assim os dados disponíveis em todo o seu potencial (Hastie et al., 2009). A Figura 4.1 mostra uma comparação entre as curvas REC dos três modelos. A análise desta figura indica que a RF parece ter um desempenho ligeiramente superior aos outros dois modelos, o que é consistente com os resultados obtidos na literatura, relativos à previsão do consumo de combustível de veículos pesados (Perrotta et al., 2017).

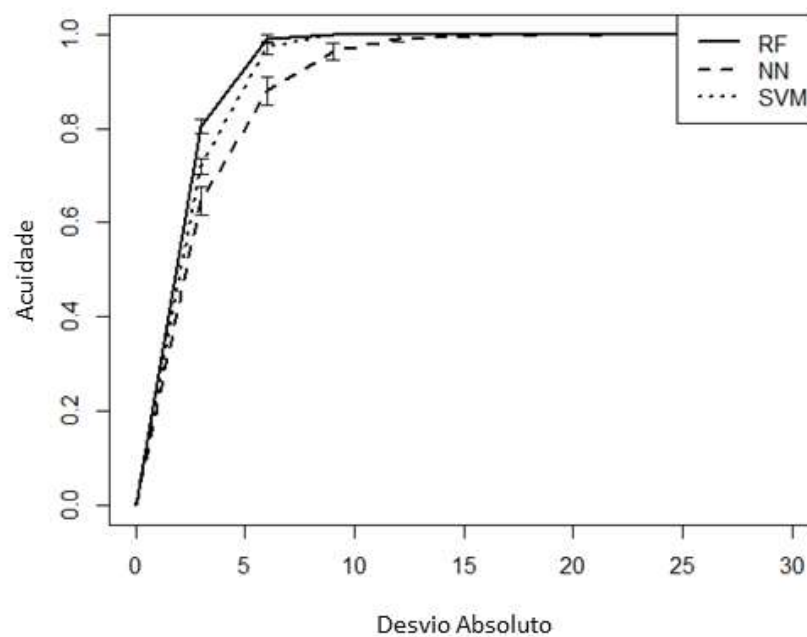


Figura 4.1 - Representação gráfica das curvas REC dos vários modelos

De facto, o modelo RF apresenta um erro ligeiramente menor, com um MAE de 1,723915 (correspondente a um erro percentual de cerca de 7,6%), um RMSE de 2,127431 e um R^2 ligeiramente melhor para os dados, com um valor de 61,4%. Embora seja expetável que este valor melhore à medida que mais dados se tornem disponíveis, ainda é considerado um valor razoável se considerarmos que o modelo se baseia inteiramente em dados reais. Esta hipótese é apoiada pelo valor do coeficiente de correlação de Pearson, R , que está em 78,4% para o modelo RF, denotando uma interdependência muito boa entre as variáveis independentes e a variável-alvo.

A Figura 4.2 mostra como os valores previstos vão de encontro aos valores observados no conjunto de dados de teste. Como se pode facilmente inferir, quanto mais perto os pontos estiverem da linha diagonal, melhor o ajuste, e consequentemente melhor o valor de R^2 . A análise do valor revela que, o modelo parece ser capaz de reproduzir o comportamento da variável-alvo especialmente bem, nos valores da gama média (ou seja, os valores de consumo de combustível de 20 a 27 litros), embora os valores mais próximos da extremidade inferior (i.e., à volta dos 17 litros), e superior (i.e., à volta dos 30 litros) estejam a ser ligeiramente sobrestimados e subestimados, respetivamente. Como expetável, isto deve-se ao facto de existir uma quantidade muito menor de registos na base de dados que se enquadram nestas extremidades, algo a aprimorar com o incremento da base de dados ao longo do tempo.

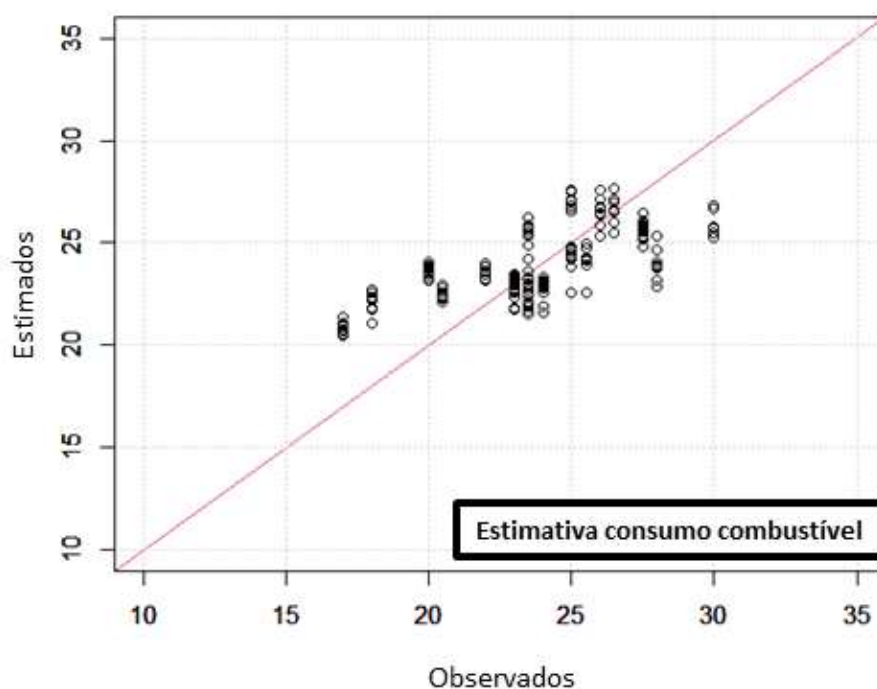


Figura 4.2 - Representação gráfica do resultado dos dados estimados vs. observados

Outro aspeto relevante para a análise é a importância relativa das variáveis para o modelo RF, representadas na Figura 4.3. A figura representa o aumento do erro quadrado médio, %IncMSE, em resultado da permuta das variáveis correspondentes. Por outras palavras, quanto maior o valor de %IncMSE, mais importante é a variável correspondente para a capacidade preditiva do modelo. Neste contexto, é interessante observar que o peso da carga transportada pelo camião é considerado pelo modelo como a sua variável mais importante, seguida da predominância de subidas ligeiras e moderadas, assim como das descidas moderadas. Intuitivamente, é fácil deduzir como estes aspetos são válidos, uma vez que quanto maior o peso e as inclinações ascendentes, maior o consumo de combustível, enquanto as inclinações descendentes implicam um valor quase nulo do consumo

de combustível. Por outro lado, a variável correspondente a subidas íngremes é considerada a menos relevante para o modelo, sendo facilmente explicável pela baixa ocorrência das secções com este tipo de inclinação ao longo do percurso do veículo, e, como tal, o modelo identifica a variável associada como tendo uma importância muito baixa. No entanto, a remoção desta variável do conjunto de dados de formação tem um impacto negativo na capacidade preditiva do modelo final, mesmo que numa escala diminuta, e como tal, foi mantida no conjunto de dados de treino.

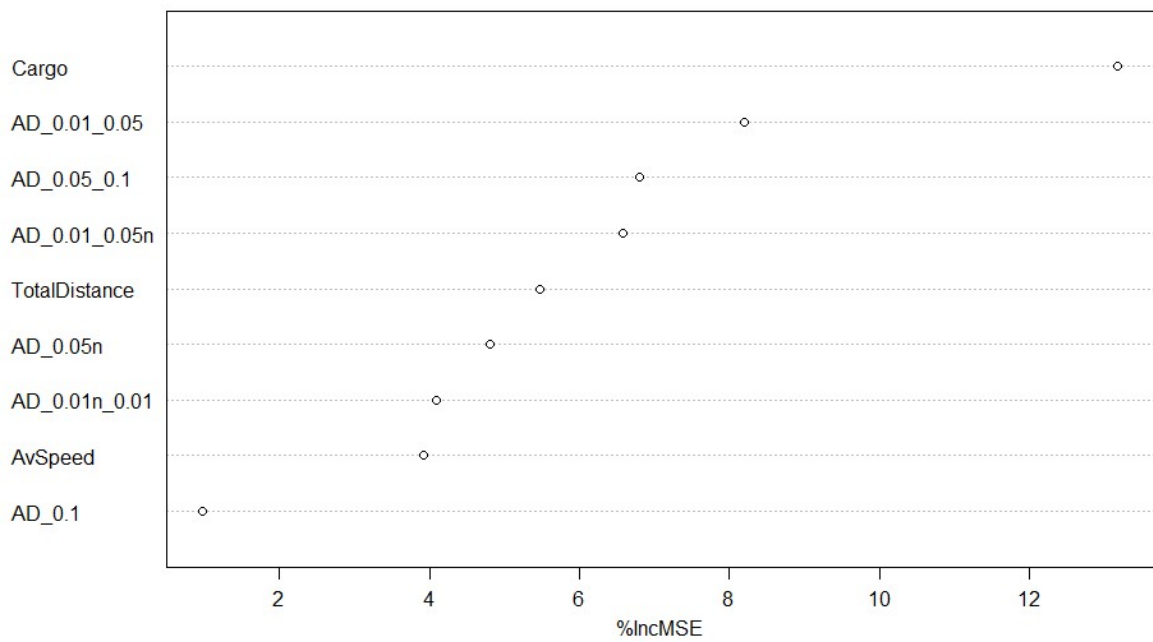


Figura 4.3 - Representação gráfica da relevância das variáveis para o modelo

5. CONSIDERAÇÕES FINAIS

5.1. Conclusões

A presente dissertação apresentou um modelo de previsão do consumo de combustível aplicado a caminhões de transporte de materiais de construção, baseado em cenário real, com o objetivo de ser utilizado como uma ferramenta para o planejamento e gestão de projetos e análise orçamental. A inovação deste trabalho foi a integração de uma estrutura conceitual de IoT para coleta e transmissão de informação na base de dados, que compreende os dados de treino e teste para os modelos preditivos. Os resultados mostraram que o consumo de combustível pode ter uma forte correlação com a carga transportada, a inclinação longitudinal das rotas e a distância total, demonstrando-se assim que constituem parâmetros de entrada fundamentais para alcançar previsões mais precisas e fiáveis do consumo de combustível. Estes resultados podem ser de especial interesse para engenheiros com funções de gestão, planejamento e tomada de decisão, uma vez que esta informação é-lhes facilmente acessível através de sistemas já existentes de SIG (por exemplo, inclinação das rotas e distância total), cadernos de encargos ou modelos BIM (por exemplo, carga).

Embora seja notório que foi escassa a quantidade de dados necessária nesta fase aos modelos de ML para serem considerados generalizáveis e assim serem implementados na prática, considera-se que o resultado da presente dissertação tem potencial e os resultados mostraram-se promissores. Acima de tudo, a metodologia utilizada foi um contributo relevante para o estado do conhecimento, no sentido que foi dado um primeiro incremento à implementação no mundo real de um *digital twin*, assim como de um sistema de ML de autoaprendizagem num paradigma de IoT, acompanhando assim as tendências atuais em matéria de automação, digitalização e Indústria/Construção 4.0.

Uma das limitações que pode ser apontada para o modelo desenvolvido é que o utilizador é obrigado a estimar a velocidade média em toda a rota, se verificar um obstáculo significativo. No entanto, esta questão pode potencialmente ser atenuada pela introdução dos dados provenientes dos acelerómetros. Com efeito, aproveitar o eixo vertical dos acelerómetros para inferir uma classificação genérica de cada tipo de superfície através da qual o camião circula (por exemplo, estrada não revestida, estrada nacional, autoestrada) pode fornecer informações sobre o comportamento do camião em diferentes ambientes (por exemplo, velocidade média, número médio de paragens completas, condições de circulação, entre outros). Posteriormente, este tipo de informação pode até ser suficientemente valiosa para que o modelo possa eventualmente substituir a necessidade de estimar a velocidade pelo utilizador, que, em vez disso, apenas poderá ter de estimar a percentagem de cada tipo de superfície em relação à distância total da viagem, à semelhança das características de inclinação da via já presentes no modelo. Por outro lado, a quantidade reduzida de dados, devido ao limite de janela temporal para recolhê-los, bem como a qualidade destes, afetada por falhas na coleta de dados dos sensores ou falhas humanas no preenchimento correto dos valores reais observados, limitaram o tamanho e a qualidade da base de dados que serviram de treino e teste ao algoritmo de ML.

5.2. Desenvolvimentos futuros

Além do que já foi referido anteriormente, os trabalhos futuros deverão incluir, naturalmente, a expansão do estudo para abranger uma maior quantidade de veículos, rotas e cargas transportadas, de modo a produzir um modelo de previsão mais robusto e generalizável. Além disso, os resultados alcançados, motivam o desenvolvimento de um sistema de aquisição de sensores em tempo real, capaz de lidar com os atuais estrangulamentos de frequência de amostragem, apoiando assim um contínuo e automático processo de treino e teste dos modelos de previsão, melhorando a sua precisão e fiabilidade, aumentando a quantidade de informação obtida a partir de sensores.

Para além disso, uma série de outros recursos poderiam ter sido explorados, tais como o *oversampling* e *undersampling*, para a introdução de um enviesamento, de maneira a seleccionar mais amostras de uma classe do que de outra ou retirá-las, de modo a compensar um desequilíbrio que possa estar presente ou suscetível de se desenvolver nos dados de uma amostra puramente aleatória, bem como a utilização de outros algoritmos de ML que obtiveram resultados interessantes na bibliografia como por exemplo as *Gradient Boost Machines*, entre outros. Será ainda interessante explorar, com uma base de dados mais extensa, a seleção manual das amostras de treino ao contrário do que foi feito, automaticamente pelo algoritmo, que neste caso se tornaria irrelevante devido ao reduzido número de amostras existentes.

REFERÊNCIAS BIBLIOGRÁFICAS

Agrawal, S. (2020), "Introduction to Support Vector Machine (SVM) | Artificial Intelligence in Plain English." <https://ai.plainenglish.io/introduction-to-support-vector-machine-svm-cd0759098471> (acesso 11 Ago., 2021).

Almér, H. (2015), "Machine learning and statistical analysis in fuel consumption prediction for heavy vehicles", Master's thesis, KTH, School of Computer Science and Communications (CSC), Stockholm.

Bi, J.; Bennett, K.P. (2003). Regression error characteristic curves. In Proceedings of the Twentieth International Conference on Machine Learning, Washington, DC, USA, 21–24 August 2003; pp. 43–50.

Canales, J. (2018). *Navigating the history of GPS*. *Nature Electronics*, 1(12), 610–611. doi:10.1038/s41928-018-0187-9.

Caterpillar (2018). Caterpillar Performance Handbook. Peoria, Illinois, U.S.A.

Cortez, P. (2010). "Data Mining with Neural Networks and Support Vector Machines using the R/miner Tool." *Advances in Data Mining - Applications and Theoretical Aspects 10th Industrial Conference on Data Mining (ICDM 2010)*, P. Perner, ed., Springer, Berlin, Germany, 571–583.

Drotek Electronics @ - "Better GNSS positioning thanks to multi-constellation and multi-frequency." <https://electronics.drotek.com/?p=6201> (acesso 11 Ago., 2021).

Edwards, D. J. e Griffiths, I. J. (2000). "Artificial intelligence approach to calculation of hydraulic excavator cycle time and output," *Min. Technol.* 109 pp. 23–29.

Gomes Correia, A., Cortez, P., Tinoco, J. e Marques, R. (2012), "Artificial Intelligence Applications in Transportation Geotechnics," *Geotech. Geol. Eng.* 31, 861–879

Hastie, T., Tibshirani, R., e Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, New York.

He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition, *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 2016-December 770–778. <https://doi.org/10.1109/CVPR.2016.90>.

Hola, B. e Schabowicz, K. (2010). "Estimation of earthworks execution time cost by means of artificial neural networks," *Autom. Constr.* 19 570–579.

Jost, D. @ (2019), "What is an accelerometer? | FierceElectronics." <https://www.fierceelectronics.com/sensors/what-accelerometer> (acesso 11 Ago., 2021).

Kopfinger, A. e Ahlsén, D. (2019). Identification of absolute orientation using inertial measurement unit. Degree Thesis, Halmstad University, 14, Halmstad.

Lindberg, J. (2014). Fuel consumption prediction for heavy vehicles using machine learning on log data. Master's thesis, KTH, School of Computer Science and Communications (CSC), Stockholm.

Lu, P., Chen, S., e Zheng, Y. (2012). Artificial intelligence in civil engineering. In *Mathematical Problems in Engineering* (Vol. 2012). <https://doi.org/10.1155/2012/145974>

Ma, X., Dong, Z., Chen, F., Xiang, H., Cao, C. e Sun, J. (2019). Airport asphalt pavement health monitoring system for mechanical model updating and distress evaluation under realistic random aircraft loads, *Constr. Build. Mater.* 226 227–237. <https://doi.org/https://doi.org/10.1016/j.conbuildmat.2019.07.174>

Majidifard, H., Adu-Gyamfi, Y., e Buttlar, W.G. (2020). Deep machine learning approach to develop a new asphalt pavement condition index, *Constr. Build. Mater.* 247 118513. <https://doi.org/https://doi.org/10.1016/j.conbuildmat.2020.118513>.

Marques, R., Gomes Correia, A., e Cortez, P. (2008). Data Mining Applied to Compaction of Geomaterials, in Eight International Conference on the Bearing Capacity of Roads, Railways and Airfields.

Mikroe @ | “GNSS 7 Click.” <https://www.mikroe.com/gnss-7-click> (acesso 11 Ago., 2021).

Mikroe @ | “6DOF IMU 15 Click.” <https://www.mikroe.com/6dof-imu-15-click> (acesso 11 Ago., 2021).

Mitchell, T. (1997). *Machine Learning*. New York: McGraw Hill. ISBN 0-07-042807-7. OCLC 36417892.

Nasser, S. H., WeiBermel, V., e Wiek, J. (1998). Computer simulation of vehicle's performance and fuel consumption under steady and dynamic driving conditions.

Nunes, D.E. e Mota, V.F.S. (2019). A participatory sensing framework to classify road surface quality, *J. Internet Serv. Appl.* 10 13. <https://doi.org/10.1186/s13174-019-0111-1>.

NVIDIA Developer @ | Jetson Nano 2GB Developer Kit. <https://developer.nvidia.com/embedded/jetson-nano-2gb-developer-kit> (acesso 11 Ago., 2021).

Parente, M., Gomes Correia, A. e Cortez, P. (2014). “Use of Data Mining Techniques in Earthwork Management: A Case Study”, *Earthwork Project Management, Slope Stability Analysis, and Wave-Based Testing Techniques*: pp. 1-8. American Society of Civil Engineers, Geotechnical Special Publication (GSP) 252 (2014). <http://dx.doi.org/10.1061/9780784478523.001>.

Parente, M., Gomes Correia, A., Cortez, P. (2016). "Metaheuristics, data mining and geographic information systems for earthworks equipment allocation", *Advances in Transportation Geotechnics III* 143 506-513. 10.1016/j.proeng.2016.06.064.

Perrotta, F., Parry, T., e Neves, L. C. (2017). "Application of machine learning for fuel consumption modelling of trucks," *2017 IEEE International Conference on Big Data (Big Data)* 3810-3815. 10.1109/BigData.2017.8258382.

Poole, D.; Mackworth, A. e Goebel, R. (1998). *Computational Intelligence: A Logical Approach*. New York: Oxford University Press. ISBN 978-0-19-510270-3.

"RStudio | Open source & professional software for data science teams - RStudio." <https://www.rstudio.com/> (acesso 11 Ago., 2021).

Saltz, J., Shamshurin, I. e Connors, C. (2017). "Predicting data science sociotechnical execution challenges by categorizing data science projects," *J. Assoc. Inf. Sci. Technol.*, vol. 68, no. 12, pp. 2720–2728, Dec. 2017, doi: 10.1002/ASI.23873.

Sandberg, T. (2001). *Heavy truck modeling for fuel consumption simulations and measurements*. Master's thesis, Linköping University.

Santos, B. P., Silva, L. A. M., Celes, C. S. F. S., Borges Neto, J. B., Peres, B. S., Augusto, M., Vieira, M., Vieira, F. M., Goussevskaia, O. N. e Loureiro, A. A. F. (2016). *Internet das Coisas: da Teoria à Prática*. Universidade Federal de Minas Gerais (UFMG), Belo Horizonte.

Santos, J. M., "REDES NEURONAIS Conceitos", [https://www.isep.ipp.pt/files/Redes%20Neuronais%20-%20Conceitos%20-%20JMS%20\(vers%C3%A3o_final\).pdf](https://www.isep.ipp.pt/files/Redes%20Neuronais%20-%20Conceitos%20-%20JMS%20(vers%C3%A3o_final).pdf) (acesso 08 Out, 2021)

Schoen, A., Byerly, A., Hendrix, B., Bagwe, R. M., Santos, E. C. d. e Miled, Z. B., "A Machine Learning Model for Average Fuel Consumption in Heavy Vehicles," in *IEEE Transactions on Vehicular Technology*, vol. 68 (2019) 6343-6351. 10.1109/TVT.2019.2916299.

Simonyan, K. e Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition, 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc. 1–14.

Solomonoff, R. J. (1985). *The Time Scale of Artificial Intelligence; Reflections on Social Effects, Human Systems Management*, Vol 5, 149-153.

Souza, V.M.A., Giusti, R. e Batista, A.J.L. (2018). Asfalt: A low-cost system to evaluate pavement conditions in real-time using smartphones and machine learning, *Pervasive Mob. Comput.* 51 121–137. <https://doi.org/https://doi.org/10.1016/j.pmcj.2018.10.008>.

Svärd, C. (2014). Predictive modelling of fuel consumption using machine learning techniques. Technical report, Scania CV AB.

Tam, C. M., Tong, T., e Tse S. (2002). "Artificial neural networks model for predicting excavator productivity," *J. Eng. Constr. Archit. Manag.* 9 446–452

TinkerCad | AutoDesk Tinkercad. <https://www.tinkercad.com/> (acesso 11 Ago., 2021).

Tinoco J., Gomes Correia, A. e Cortez, P. (2011). "Application of data mining techniques in the estimation of the uniaxial compressive strength of jet grouting columns over time". *Constr Build Mater* 25 1257–1262. 10.1016/j.conbuildmat.2010.09.027

Tinoco J., Gomes Correia, A., Cortez, P. e Toll, D. (2017). "Stability condition identification of rock and soil cutting slopes based on soft computing", *Journal of computing in civil engineering* 32 04017088.

Vapnik, V., Burges, C.J., Kauffman, L. e Smola, A. (1997). *Support Vector Regression Machines*, *Neural Information Processing Systems* 9, eds. M.C. Mozer. J.I. Jordan, T. Petsche. pp. 155-161, MIT Press. 1997

Viswanathan, A. (2013). Data driven analysis of usage and driving parameters that affect fuel consumption of heavy vehicles. Master's thesis, Linköping University, Statistics, The Institute of Technology

Yiu, T. @ (2019). "Understanding Random Forest - How the Algorithm Works and Why it Is So Effective", *Towards Data Science* <https://towardsdatascience.com/understanding-random-forest-58381e0602d2> (acesso 08 Out, 2021).

Zhao, H., Wu, D., Zeng, M., Tian, Y. e Ling, J. (2019). "Assessment of concrete pavement support conditions using distributed optical vibration sensing fiber and a neural network", *Constr. Build. Mater.* 216 214–226. <https://doi.org/https://doi.org/10.1016/j.conbuildmat.2019.04.195>