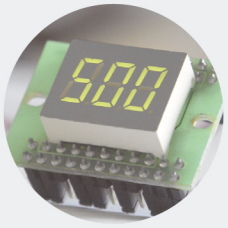
INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO

MESTRADO EM ENGENHARIA ELECTROTÉCNICA - SISTEMAS ELÉCTRICOS DE ENERGIA






Planeamento da Manutenção de equipamentos médicos com recurso a técnicas de inteligência artificial

FERNANDO SÉRGIO NICOLA DA COSTA SALGADO

julho de 2023

POLITÉCNICO
DO PORTO

INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO

MESTRADO EM ENGENHARIA ELECTROTÉCNICA - SISTEMAS ELÉCTRICOS DE ENERGIA

Planeamento da Manutenção de equipamentos médicos com recurso a técnicas de inteligência artificial

FERNANDO SÉRGIO NICOLA DA COSTA SALGADO

Junho de 2023

POLITÉCNICO
DO PORTO

PLANEAMENTO DA MANUTENÇÃO DE EQUIPAMENTOS MÉDICOS COM RECURSO A TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL

Fernando Sérgio Nicola da Costa Salgado

Departamento de Engenharia Eletrotécnica

Mestrado em Engenharia Eletrotécnica – Sistemas Elétricos de Energia

2023

Relatório elaborado para satisfação parcial dos requisitos da Unidade Curricular de DSEE – Dissertação do Mestrado em Engenharia Eletrotécnica – Sistemas Elétricos de Energia

Candidato: Fernando Sérgio Nicola da Costa Salgado, Nº 1980263, 1980263@isep.ipp.pt

Orientação científica: José Marílio Oliveira Cardoso, joc@isep.ipp.pt

Departamento de Engenharia Eletrotécnica

Mestrado em Engenharia Eletrotécnica – Sistemas Elétricos de Energia

2023

Agradecimentos

O presente trabalho foi realizado sob a orientação do Professor Doutor José Marílio Oliveira Cardoso, a quem me cabe, exprimir sincero reconhecimento pelo apoio reiteradamente prestado e disponibilidade manifestada. Dele recebi orientações e conselhos que muito contribuíram para o resultado deste trabalho e o tornaram ainda mais enriquecedor. Desde já agradeço todo o contributo prestado pelo Professor Tiago Pereira no esclarecimento de dúvidas e apoio para a realização do meu trabalho.

Agradeço ainda à minha esposa e filha, por fazerem parte da minha vida e pela forma como souberam compreender as minhas prolongadas ausências. Em particular à minha querida esposa, expressar a enorme admiração que tenho por ela, pela força, coragem e determinação demonstrada nesta fase tão desafiante da nossa vida e que me serviu de verdadeira inspiração todos os dias.

Aos meus dois amigos Paulo Fontes e Miguel Costa, pela sua amizade, pelos bons momentos que partilhamos, pelo apoio e confiança que me transmitiram.

Resumo

Esta dissertação apresenta um estudo sobre a função manutenção, considerada fundamental para as organizações, pois é essencial para garantir a continuidade dos processos de produção, a prestação de serviços com qualidade e para assegurar que os equipamentos e instalações funcionem de forma eficiente e segura.

Este trabalho tem como objetivo principal, desenvolver uma solução com recurso a técnicas de inteligência artificial para otimizar o planeamento da manutenção de equipamentos médicos, assim como suportar a gestão da manutenção nas suas tomadas de decisão.

Foram recolhidos dados de funcionamento de equipamentos médicos, aos quais foram aplicados diferentes algoritmos de aprendizagem de *machine learning* (aprendizagem automática) de modo a identificar o melhor modelo em termos de precisão e exatidão que se aplica à resolução do objetivo proposto.

Espera-se demonstrar que a solução desenvolvida contribuirá não só para a otimização do planeamento, bem como para prolongar o ciclo vida útil dos equipamentos, evitando falhas, diminuindo os tempos de inatividade e reduzindo os custos de manutenção.

Palavras-Chave

Planeamento, Manutenção Preditiva, Inteligência Artificial, *Machine Learning*, Algoritmos.

Abstract

This thesis offers a study about maintenance, considered essential for organizations, as it's essential to ensure the stability of production processes, the delivery of quality services and to ensure that equipment and installations operate efficiently and safely.

The main objective of this work is to develop a solution using artificial intelligence techniques to optimize the maintenance planning of medical equipment, as well as to support maintenance management in their decisions.

Medical equipment operating data was collected, different machine learning algorithms was applied to identify the best model in terms of precision and accuracy that be appropriate to the resolution of the proposed objective.

It's expected to demonstrate that the developed solution will contribute not only to the optimization of planning, but also to extend the useful life cycle of the equipment, avoiding failures, reducing downtime, and reducing maintenance costs.

Keywords

Planning, Predictive Maintenance, Artificial Intelligence, Machine Learning, Algorithms.

Índice

Índice de Figuras

Índice de Tabelas

Siglas e Acrónimos

ANNS - *Artificial Neural Networks* (Redes Neuronais Artificiais)

AR - *Augmented Reality* (Realidade Aumentada)

BSC - *Balanced Score Card* (Mapa de Indicadores de Desempenho)

CBM - *Condition-Based Maintenance* (Manutenção Baseada em Condições)

CMMS - *Computorized Maintenance Management Systems* (Software de Gestão de Manutenção)

CNNs - *Convolutional Neural Network* (Redes Neuronais Convolucionais)

CPS - *Cyber Physical Systems* (Sistemas Ciber Físicos)

DL - *Deep Learning* (Aprendizagem Profunda)

DT - *Decision tree* (Árvore de decisão)

ERP - *Enterprise Resource Planning* (Sistema Integrado de Gestão Empresarial)

FMEA - *Failure Mode and Effect Analysis* (Análise de Modo e Efeito de Falha)

GBDT - Gradiente Boosting Árvores Decisão *(Gradient Boosting Decision Tree Classifier)*

GMAC - Gestão da Manutenção Assistida por Computador

IA - *Artificial Intelligence* (Inteligência Artificial)

ICD - Indicadores Chave de Desempenho

IoT - *Internet of Things* (Internet das Coisas)

JIT - *Just in Time*

KNN - *K-Nearest Neighbors* (K Vizinho Mais Próximo)

LCC - *Life Cost Cycle* (Custo Ciclo de Vida)

LM - *Lean Maintenance*

Logs - *Logs* (Mensagens de Registo)

ML - *Machine Learning* (Aprendizagem de máquina)

MLPs - *Multi-Layer Perceptrons*

MTBF - *Mean Time Between Failures* (Tempo Médio entre Avarias)

MPD - Manutenção Preditiva

NLP - *Natural Language Processing* (Processamento de Linguagem Natural)

OEE - Eficácia Global Equipamentos

PHM - *Prognostics and Health Management* (Prognósticos e Gestão de Saúde)

RF - *Random Forest* (Árvores de Decisão)

RCM - *Reliability-Centered Maintenance* (Manutenção Centrada na Fiabilidade)

RL - *Reinforced Learning* (Aprendizagem Reforçada)

RNNS - *Recurrent Neural Networks* (Redes Neuronais Recorrentes)

SVM - *Support Vector Machines* (Máquina de Vetores de Suporte)

TEI - *Total Imployment Involvement* (Envolvimento Total dos Funcionários)

TPM - *Total Productive Maintenance* (Manutenção Produtiva Total)

TQC - *Total Quality Control* (Controlo Total da Qualidade)

VR - *Virtual Reality* (Realidade Virtual)

1. INTRODUÇÃO

1.1. CONTEXTUALIZAÇÃO

O rápido desenvolvimento tecnológico a que temos assistido, associado ao atual contexto económico e globalização dos mercados, está a pressionar as organizações a acelerar a inovação e procurar adotar soluções baseadas em tecnologias digitais que permitam transformar as suas atividades e os seus modelos de negócio.

No setor da saúde e nos sistemas médicos, esta transformação está a ser alavancada pela digitalização dos seus processos com recurso às tecnologias IoT *(internet of things)*, desempenhando estas um papel cada vez mais fundamental, principalmente na manutenção remota e preditiva de equipamentos médicos. As IoT permitem a gestão e monitorização remota do desempenho operacional e técnico de dispositivos médicos e a capacidade de realizar análises avançadas dos dados recolhidos dos equipamentos. Esta realidade conduz a uma gama impressionante de aplicações práticas e acréscimo de valor às diferentes partes interessadas *(stakeholders)*, entre as quais, os fabricantes de equipamentos médicos, as organizações responsáveis por assegurar a manutenção dos equipamentos e os profissionais do setor da saúde enquanto utilizadores destes equipamentos.

As tecnologias IoT suportam a implementação da estratégia de manutenção preditiva e remota, ajudando as organizações a monitorizar, manter e otimizar o desempenho em tempo real dos equipamentos pelos quais são responsáveis.

Os dados de um dispositivo médico podem ser recolhidos e analisados remotamente por profissionais de saúde e fabricantes de equipamentos com o objetivo de antecipar potenciais falhas dos mesmos. Os dispositivos médicos contêm diferentes componentes elétricos, eletrónicos, hidráulicos, pneumáticos, etc., que, como em qualquer outro equipamento, têm um determinado ciclo de vida e necessitam de ser substituídos periodicamente.

Normalmente as organizações do setor da saúde, dispõem de equipas técnicas locais dotadas com as competências necessárias para assegurar a adequada manutenção dos equipamentos e garantir o seu correto funcionamento. Contudo, uma eventual falha do equipamento pode originar um considerável tempo de inatividade e interrupção de tratamentos dos pacientes.

As tecnologias IoT podem assim contribuir para o aumento da eficiência, redução da necessidade de intervenções manuais e diminuição da probabilidade de falhas, dado que a monitorização dos dados dos equipamentos pode identificar padrões de funcionamento anómalos de componentes e consequente necessidade da sua substituição. Desta forma, as equipas de manutenção podem ser alertadas com antecedência para que as peças de reposição possam ser encomendadas e planeada a sua substituição.

1.2. OBJETIVOS

O objetivo central deste trabalho é o desenvolvimento de uma solução com recurso a técnicas de inteligência artificial para otimizar o planeamento da manutenção, bem como para suportar a gestão da manutenção nas suas tomadas de decisão. O objetivo específico, é alavancar a implementação da estratégia de manutenção remota e/ou preditiva.

Estas técnicas de inteligência artificial serão aplicadas simultaneamente a uma quantidade considerável de parâmetros que permitam monitorizar o desempenho de equipamentos médicos, aos *logs* (mensagens de registo) de funcionamento dos equipamentos e ao histórico do registo das intervenções realizadas pelas equipas técnicas de manutenção.

Com esta solução, as equipas de manutenção podem ser alertadas com antecedência para a necessidade de manutenção nos equipamentos, contribuindo para a conversão de processos aleatórios, como a manutenção corretiva, em processos planeáveis. Com a previsão de

erros e eventuais falhas de equipamentos, a manutenção pode ser planeada com antecedência e garantir a disponibilidade dos materiais necessários antes da falha do equipamento. Desta forma, diferentes desperdícios podem ser mitigados, a eficiência dos processos de manutenção será melhorada, bem como aumentará a agregação de valor para a organização.

As áreas como a gestão de manutenção e gestão de stocks serão beneficiadas por esta solução uma vez que a manutenção dos equipamentos passará a ser predominantemente planeável.

1.3. ORGANIZAÇÃO DO DOCUMENTO

Este documento está organizado em seis capítulos, onde se apresentam os principais resultados, contributos e conclusões do trabalho desenvolvido.

Neste primeiro capítulo, faz-se a contextualização geral onde se destaca a necessidade de as organizações do setor da saúde acelerarem a inovação e de procurarem adotar soluções baseadas em tecnologias digitais que permitam transformar as suas atividades e modelos de negócio.

No capítulo seguinte, é apresentado um estudo sobre o estado da arte da área da manutenção, onde é analisada a importância dos departamentos de manutenção em contexto organizacional, definição dos seus objetivos, tipos e estratégias de manutenção disponíveis para aplicação, as normas da manutenção e a importância dos sistemas de informação.

No terceiro capítulo, é apresentado um estudo sobre a evolução da manutenção industrial, o impacto e as contribuições das revoluções industriais na área da manutenção, conceitos, tecnologias e soluções atuais disponíveis no âmbito da manutenção preditiva.

No quarto capítulo, descreve-se o trabalho realizado e a solução a ser desenvolvida para a previsão antecipada de falhas em equipamentos médicos.

No quinto capítulo, apresenta-se uma análise dos modelos utilizados e consequente discussão dos resultados obtidos, avaliados pelas métricas e sua eficiência.

No sexto capítulo, para finalizar, apresenta-se o conjunto de conclusões obtidas a partir do trabalho realizado, assim como as perspetivas para eventuais trabalhos futuros. Esta abordagem, tem como objetivo identificar lacunas no trabalho atual e propor orientações para futuras investigações.

2. MANUTENÇÃO DE EQUIPAMENTOS

2.1. ENQUADRAMENTO

Qualquer estudo sobre a manutenção, deve iniciar pela sua definição e compreensão dos aspetos envolventes, de modo a contribuir para uma base sólida de conhecimento sobre o assunto. Por conseguinte, é essencial utilizar uma linguagem comum, de forma a garantir a consistência e precisão na comunicação de informações técnicas. Por outro lado, garantirá que todos os intervenientes, desde os técnicos especializados até à direção da empresa, compreendam os objetivos, metas e resultados da manutenção.

Os conceitos aqui abordados serão os mais usuais no processo de gestão da manutenção, recorrendo a definições contidas em normas portuguesas e/ou internacionais e definições de autores de referência sempre que possível.

2.2. DEFINIÇÃO DE MANUTENÇÃO

Segundo a NP EN 13306 – Terminologia da Manutenção, a manutenção é uma combinação de todas as ações técnicas, administrativas e de gestão, durante o ciclo de vida de um bem, destinadas a mantê-lo ou repô-lo num estado em que ele pode desempenhar a função requerida. Ainda segundo a mesma norma, gestão da manutenção, são todas as atividades de gestão que determinam os objetivos, a estratégia e as responsabilidades

respeitantes à manutenção e que são implementadas por diversos meios tais como o planeamento, o controlo e a supervisão da manutenção e a melhoria de métodos na organização, incluindo os aspetos económicos.

"A função manutenção envolve todas as atividades que se relacionam com a conservação das boas condições de funcionamento de equipamentos, sistemas e instalações, e a realização de intervenções corretivas sempre que ocorram falhas ou avarias, de modo que o sistema de operações (produção e/ou serviço) alcance o desempenho desejado" [1].

A manutenção pode então ser definida como o conjunto de todas as ações, tendo em vista garantir a disponibilidade dos equipamentos e instalações, de modo a contribuir para um processo produtivo, com confiabilidade, segurança, custos adequados e com preservação do meio ambiente.

2.3. IMPORTÂNCIA DA MANUTENÇÃO

A globalização dos mercados, o atual contexto económico assim como o rápido desenvolvimento tecnológico que se tem verificado, está a pressionar as empresas a acelerar a inovação e a procurar adotar soluções que permitam transformar as suas atividades e os seus modelos de negócio.

A função manutenção, é habitualmente assegurada por um departamento próprio dentro das organizações, dependendo do setor de atividade e modelo de gestão. Numa outra abordagem, poderá inclusive ser considerada como uma unidade de negócio, o que leva atualmente as organizações a adotar uma perspetiva de gestão global em termos de desempenho, qualidade, custos e serviço.

Algumas das atividades de manutenção são consideradas como despesas necessárias, contudo tendem a não ser valorizadas dado não estarem diretamente ligadas à produção ou prestação de serviços. No entanto, a manutenção é essencial para garantir a continuidade dos processos de produção e prestação de serviços, dado que uma falha na manutenção pode resultar em interrupções no fluxo de trabalho e perda de receita. Portanto, é importante que as organizações valorizem e invistam na manutenção, como uma parte crucial da operação e sucesso geral da empresa.

Os departamentos de manutenção desempenham um papel fundamental nas organizações, tendo como principal objetivo prolongar o ciclo vida útil dos equipamentos, evitando

falhas, diminuindo os tempos de inatividade e reduzindo os custos de manutenção. No desempenho natural das suas tarefas e atividades, têm de se relacionar com outros departamentos, tais como compras, operações, desenvolvimento, recursos humanos, etc.

Embora podendo não ser dos departamentos mais importantes dentro da estrutura de uma organização, têm um contributo importante na entrega do produto/serviço final ao cliente.

2.4. OBJETIVOS DA MANUTENÇÃO

Os objetivos da manutenção e da empresa estão estreitamente relacionados, tendo a estratégia de negócio e os objetivos globais da empresa um papel determinante na definição dos objetivos do departamento de manutenção. A gestão da manutenção deve trabalhar em estreita colaboração com as outras áreas da empresa, para priorizar e garantir que os objetivos de manutenção estejam alinhados com os objetivos globais da empresa e contribuam para o sucesso geral da empresa.

A figura 1 apresenta de uma forma resumida os principais objetivos da manutenção.

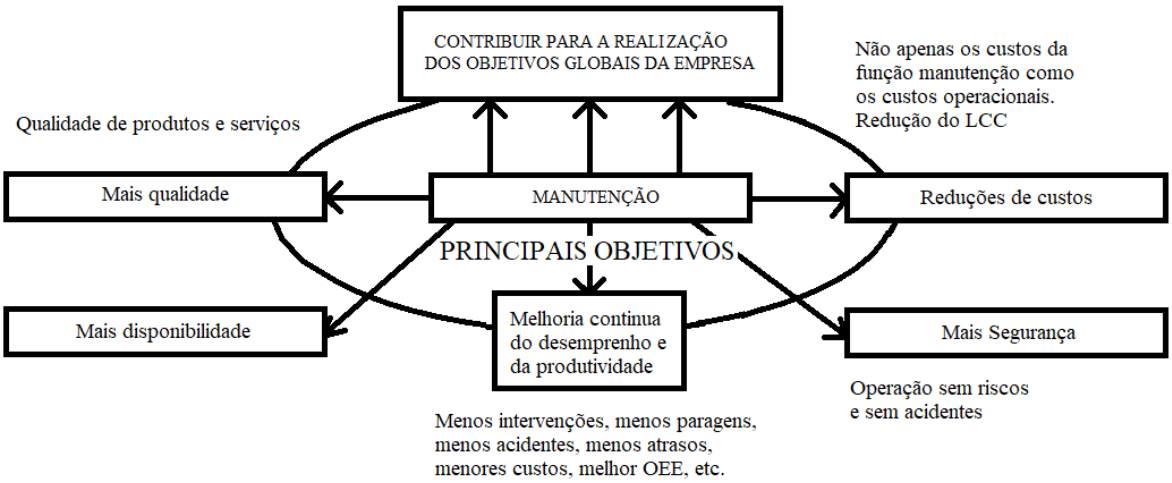


Figura 1 Principais objetivos da função manutenção [1]

Entre os objetivos mais importantes, destaca-se a disponibilidade, pois mede a capacidade de um equipamento estar disponível para cumprir com as funções para as quais foi projetado. De acordo com a norma EN 15341, a disponibilidade é calculada como a relação

entre o tempo que o equipamento está disponível para utilização e o tempo total, descontando o tempo de paragem para manutenção e o tempo de paragem não planeado.

Os objetivos da manutenção devem ser definidos e alinhados com os objetivos da empresa, tendo em conta que deve ser encontrado um equilíbrio entre os custos da indisponibilidade (associada às perdas de produção ou serviço) e os custos da manutenção (necessários para assegurar disponibilidade).

Relativamente à qualidade, é uma parte importante da manutenção, pois contribui para que os equipamentos e instalações da organização sejam mantidos em boas condições de funcionamento e em segurança para utilização. A manutenção deve-se concentrar em entregar serviços de alta qualidade, alinhados com as atuais normas e as necessidades dos clientes [1].

2.5. TIPOS E ESTRATÉGIAS DE MANUTENÇÃO

A adoção de tipos e estratégias adequadas de manutenção é essencial para otimizar o desempenho operacional e reduzir custos associados à manutenção. Neste capítulo, serão apresentados os principais tipos de manutenção programada e não programada, destacando as suas características e importância para as organizações.

2.5.1. MANUTENÇÃO PROGRAMADA

Manutenção programada (EN13306) – manutenção preventiva efetuada de acordo com um calendário preestabelecido ou de acordo com um número definido de unidades de utilização.

A manutenção programada, é uma manutenção preventiva planeada para ser realizada numa data definida e, que pode ser resultante de um diagnóstico prévio ou da implementação sistemática de um ciclo de manutenção. De uma forma simples, é uma tarefa de manutenção que pode ser calendarizada. Exemplo: “inspeção e verificação todas as 10.000 horas de operação “.

Manutenção preventiva (EN13306) – manutenção efetuada a intervalos de tempo predeterminados ou de acordo com critérios prescritos com a finalidade de reduzir a probabilidade de avaria ou de degradação do funcionamento de um bem.

De facto, este é o principal objetivo da gestão da manutenção. Contudo é importante referir que a manutenção preventiva é cara, requer normalmente a substituição sistemática de componentes, implica tempos de disponibilidade de equipamentos e por mais abrangente e cara que seja, não consegue cobrir todas as hipóteses de falha.

Num sistema produtivo "real", embora havendo equipamentos que imperativamente deverão beneficiar de um sistema de manutenção preventiva, há muitos outros que podem ser reparados unicamente em caso de avaria efetiva, isto é, ter manutenção puramente curativa.

Manutenção sistemática (EN13306) – manutenção preventiva efetuada a intervalos de tempo preestabelecidos ou segundo um número definido de unidades de utilização, mas sem controlo prévio do estado do bem.

A manutenção sistemática, é uma manutenção programada para acontecer com uma periodicidade preestabelecida. Neste caso, não é realizado um diagnóstico ou qualquer avaliação prévia do estado do equipamento. Por norma, as periodicidades destas tarefas são definidas pelos fabricantes dos equipamentos, embora se considere sempre o contributo crítico da engenharia da manutenção e das suas equipas técnicas para implementar as adequadas tarefas de manutenção face ao regime real e contexto de operação dos equipamentos.

Manutenção preditiva (EN13306) – manutenção condicionada efetuada de acordo com as previsões extrapoladas da análise e avaliação de parâmetros significativos da degradação de um bem.

A manutenção preditiva é um tipo de manutenção em que os períodos de manutenção não são fixos nem pré-estabelecidos, sendo dependente da monitorização, avaliação e/ou inspeção de parâmetros que permitem controlar o estado do equipamento.

Os parâmetros controlados têm limites e/ou padrões de funcionamento que, uma vez próximo dos seus limites ou ultrapassados, desencadeiam uma previsão extrapolada da necessidade de manutenção. Com esta estratégia as necessidades de manutenção dependem da condição real do estado do equipamento.

Naturalmente esta estratégia de manutenção implica a criação de novos tipos de tarefas, nomeadamente de medição, avaliação e/ou inspeção. Estas tarefas necessitam de recursos

que acarretam um custo que, por vezes, pode ser elevado. Por exemplo, se o custo da monitorização de determinado componente de um equipamento for superior ao custo do próprio componente então é vantajoso utilizar uma manutenção do tipo preventivo, ou mesmo reativo. Tudo depende do custo, da importância, informação disponível e características do componente em causa.

2.5.2. MANUTENÇÃO NÃO PROGRAMADA

Contrariamente à manutenção programada, a manutenção não programada é uma estratégia de manutenção que não pode ser calendarizada.

Manutenção corretiva (EN13306) – manutenção efetuada depois da deteção de uma avaria, e destinada a repor o bem num estado em que possa realizar uma função requerida.

A manutenção corretiva é uma manutenção do tipo reativo, desencadeada pela deteção de uma ou mais falhas no equipamento, onde vão ser executadas tarefas de manutenção com o objetivo de repor o equipamento no seu estado normal de funcionamento.

Manutenção corretiva de urgência (EN13306) – manutenção corretiva que é realizada imediatamente após a deteção de uma falha a fim de evitar consequências inadmissíveis.

Na prática, é uma manutenção prioritária sobre todas as outras tarefas e que tem de ser realizada com carácter de urgência para repor o equipamento no seu estado normal de funcionamento de forma a mitigar as consequências da indisponibilidade do mesmo.

2.6. NORMAS DE MANUTENÇÃO

As normas relacionadas com a manutenção assumem um importante papel, particularmente em contexto organizacional. São fundamentais para a padronização e no estabelecimento de orientações para as atividades de manutenção, promovendo a eficiência, a segurança e a qualidade dos serviços prestados. Neste capítulo, serão apresentadas algumas das normas mais importantes e relacionadas com a manutenção: a Norma da Terminologia da Manutenção EN 13306, a Norma dos Indicadores de Manutenção EN 1534, a Norma dos Contratos de Manutenção EN 13269 e a Norma da Documentação de Manutenção EN 13460.

2.6.1. NORMA DA TERMINOLOGIA DA MANUTENÇÃO EN 13306

A Norma Europeia EN 13306, também conhecida como "Manutenção - Terminologia", é uma norma internacionalmente reconhecida que tem como objetivo fornecer uma base comum para a definição e uso dos termos relacionados à manutenção. Ela abrange todos os tipos de manutenção, incluindo preventiva, corretiva e preditiva, e tem aplicabilidade a todos os tipos de bens, exceto as aplicações informáticas. A norma fornece definições claras e precisas para os termos usados na manutenção, ajudando a garantir a compreensão e a comunicação precisa entre os profissionais envolvidos na manutenção, facilitando assim a cooperação e a colaboração entre diferentes partes interessadas.

É da responsabilidade de toda a organização de manutenção estabelecer a sua estratégia de manutenção de acordo com três critérios fundamentais:

- Garantir a disponibilidade dos equipamentos, em boas condições para cumprir com a função requerida a custos ótimos;
- Considerar os requisitos, regulamentos e normas de segurança relativos aos equipamentos e profissionais da manutenção e da operação e, quando necessário, ter em conta o impacto ambiental;
- Prolongar a vida útil do equipamento e/ou que o produto ou serviço fornecido seja de qualidade, tendo em conta os custos, caso necessário.

No quadro da missão atribuída à CEN/TC 319 [15] foi necessário produzir uma norma estruturada de vocabulário da manutenção, contendo os termos principais e suas definições. A manutenção fornece uma contribuição essencial à segurança de funcionamento de um equipamento. O utilizador das normas de manutenção tem necessidade de definições formalmente corretas, para melhor compreender os requisitos da manutenção. Estes requisitos poderão ser importantes na redação dos contratos de manutenção, pois garantem que as partes envolvidas tenham compreensão comum dos termos e requisitos. Os termos contidos nesta norma indicam que a manutenção não está confinada às atividades técnicas, mas inclui todas as atividades tais como o planeamento e a gestão da documentação.

2.6.2. NORMA DOS INDICADORES DE MANUTENÇÃO EN 15341

Esta norma estabelece um sistema de gestão de indicadores para medir o desempenho da manutenção, sob a influência de diversos fatores, tais como: económicos, técnicos e organizacionais. Estes indicadores servem para avaliação e melhoria da eficiência e eficácia de forma a se atingir a excelência da manutenção dos bens imobilizados. A maioria destes indicadores aplicam-se a todas as instalações industriais e serviços (edifícios, infraestruturas, transporte, distribuição, redes, etc.).

Os indicadores escolhidos para medir e controlar o desempenho da organização, devem, segundo as melhores práticas, ser agrupados e alimentar o que se denomina por um mapa de indicadores, também conhecido como *Balanced Scorecard* (BSC), onde podem ser analisados e correlacionados entre si.

O BSC é uma ferramenta de planeamento estratégico e de gestão amplamente utilizada em empresas, indústria e governos em todo o mundo para alinhar:

a) Atividades de negócios com a visão e estratégia da organização;

b) Benchmarking interno e externo, consiste em comparar os resultados dos indicadores de uma organização com outras organizações;

c) Melhorar a comunicação interna e externa;

d) Monitorizar o desempenho das organizações em relação aos objetivos estratégicos.

Esta ferramenta deve suportar os departamentos de manutenção para monitorizar os seus processos centrais e, caso necessário, implementar ações de melhoria. O BSC dá uma visão geral sobre a estabilidade dos processos e a possibilidade de comparar uma organização com outras organizações por meio de um *ranking*.

2.6.3. NORMA DOS CONTRATOS DE MANUTENÇÃO EN 13269

Esta norma foi elaborada com o objetivo de definir a estrutura tipo para a elaboração de um contrato de prestação de serviços de manutenção, disponibilizado um conjunto de orientações e aspetos que devem ser equacionados aquando da celebração de um contrato de manutenção.

Trata-se assim de um documento que deve ser consultado como referência na elaboração de um contrato de manutenção, de forma a garantir uma abordagem estruturada, cuidada e facilitar na definição dos resultados pretendidos com as atividades de manutenção.

2.6.4. NORMA DA DOCUMENTAÇÃO DE MANUTENÇÃO EN 13460

A norma EN 13460, "Sistemas de gestão da manutenção - Requisitos para a documentação" é uma norma europeia que estabelece os requisitos para a documentação de um sistema de gestão da manutenção. Ela especifica quais os documentos que devem estar presentes no sistema de gestão da manutenção e os requisitos para cada um desses documentos. Destina-se a garantir que o sistema de gestão da manutenção seja abrangente, consistente e claramente definido, fornecendo um enquadramento para a gestão eficaz das atividades de manutenção ao longo da vida útil do equipamento.

Esta documentação é estabelecida para a fase operacional do equipamento e tem em consideração os acordos estabelecidos entre o utilizador e o fornecedor, e registados no contrato que foi realizado de acordo com a norma EN 13269.

2.6.5. NORMA QUALIFICAÇÃO PESSOAL TÉCNICO DE MANUTENÇÃO EN 15628

A norma EN15628 especifica os requisitos para a qualificação do pessoal envolvido na manutenção de instalações, infraestruturas, equipamentos e sistemas de produção. Esta norma serve como orientação para definir os conhecimentos, aptidões e competências necessárias para que os profissionais de manutenção possam desempenhar as suas tarefas de maneira eficaz e segura.

Os seguintes profissionais da organização de manutenção são abrangidos por esta Norma Europeia:

- Técnico Especialista em Manutenção;
- Supervisor de Manutenção e Engenheiro de Manutenção;
- Gestor de Manutenção (Responsável pela Função ou Serviço de Manutenção).

A norma não especifica os critérios de verificação nem a formação específica dos profissionais que estão relacionados com o setor onde desempenham funções. Geralmente,

são as empresas que garantem aos seus profissionais a necessária formação nos respetivos setores para que estes obtenham a especialização e profissionalização desejáveis.

2.7. FILOSOFIA LEAN NA MANUTENÇÃO

Depois de na década de 1970 a manutenção produtiva total (TPM) ter alavancado a manutenção para um novo patamar de desenvolvimento, envolvendo e comprometendo os operadores de produção com a manutenção das boas condições de funcionamento dos equipamentos, a manutenção foi gradualmente invadida na década de 1980 e 1990 por novas técnicas orientadas à fiabilidade, como a análise de modos de falha e efeitos (FMEA), a manutenção baseada em condições (CBM), a manutenção preditiva e a gestão de ativos.

Essas técnicas foram projetadas para melhorar a confiabilidade dos equipamentos e reduzir a necessidade de intervenções corretivas. A TPM e as técnicas orientadas à fiabilidade, são complementares e combinadas, podem dar uma visão mais completa e eficiente da manutenção.

Estas técnicas potenciaram o papel da manutenção, mas não foram capazes de a levar a assumir o papel de parceiro ativo na redução de custos e na melhoria contínua da empresa. Houve assim a necessidade de repensar a manutenção e encontrar ferramentas orientadas para estes objetivos.

Os pensamentos *Lean* que surgiram e foram aplicados no início do século passado à indústria automóvel com resultados comprovados, foram considerados e assim aplicados à área da manutenção.

A *Lean Maintenace* (LM), pode ser definida como um conjunto de operações proactivas que emprega atividades planeadas através das práticas TPM e estratégias de manutenção centrada na fiabilidade recorrendo a equipas autónomas.

A LM é suportada por um sistema descentralizado de gestão de materiais e peças de reserva que garantem o fornecimento JIT (*just-in-time*) do que é necessário e apoiada num grupo de engenharia de fiabilidade que realiza análise de causas e efeitos e análises de manutenção preditiva.

Assenta numa cultura de melhoria contínua, envolvendo todas as pessoas na preservação da função de cada equipamento.

2.8. SISTEMAS DE INFORMAÇÃO NA MANUTENÇÃO

Os *softwares* de gestão da manutenção, também conhecidos como GMAC, gestão da manutenção assistida por computador, CMMS (*Computorized Maintenance Management System*) ou ERP, (*Enterprise Resource Planning*) [2], são uma ferramenta amplamente utilizada na indústria para gerir as atividades de manutenção de uma organização. São sistemas que ajudam as empresas a gerir as operações dos seus negócios, incluindo finanças, recursos humanos, compras, produção e logística.

Podem incluir módulos específicos para gerir a manutenção, permitindo às empresas planear, programar e registar as suas atividades de manutenção. Alguns exemplos de funcionalidades que devem incluir:

- Equipamentos/objetos de manutenção: codificação e registo, com ficha estruturada de características técnicas [2]. Uma ficha estruturada de características técnicas é um documento ou registo que contém informações detalhadas sobre as características técnicas de um equipamento ou objeto de manutenção. Ela pode incluir informações como o fabricante, modelo, número de série, data de instalação, capacidade, especificações, opções, desenhos, diagramas, instruções de operação e outras informações relevantes. É uma importante ferramenta para garantir que as equipas de manutenção tenham informações precisas e permanentemente atualizadas sobre cada equipamento, o que pode ajudar a reduzir o tempo de inatividade e aumentar a eficiência.

- Planeamento de manutenção preventiva: permite definir programas de manutenção preventiva para equipamentos e instalações, incluindo entre outros processos, inspeções, testes, modificações, alterações, atualizações e substituição de peças.

- Registo de manutenção corretiva: permite registar e rastrear as atividades de manutenção corretiva, incluindo reparações e substituição de peças, bem como o tempo e os custos associados.

- Gestão de *stocks*: permite gerir o *stock* de peças e materiais para manutenção, incluindo a compra, armazenamento e alocação de recursos.

- Análise de desempenho: permite monitorizar e analisar o desempenho dos equipamentos, através do cálculo de indicadores expressivos das atividades de manutenção, os chamados ICD (Indicadores Chave de Desempenho). Estes indicadores podem permitir identificar tendências e potenciais problemas.

Exemplos de alguns indicadores chave:

a) Taxa de falha dos equipamentos;

b) MTBF – *Mean Time Between Failures*, tempo médio entre avarias;

c) Disponibilidade;

d) Tempo médio de reparação;

e) Tempo médio até reparar;

f) Tempo médio de espera;

g) Eficácia global dos equipamentos (OEE).

Alguns *softwares*, em particular os sistemas ERP, podem integrar-se com outros sistemas, como sistemas de automação de fábricas, sistemas de gestão de ativos e sistemas de gestão de projetos, permitindo uma melhor visibilidade e controle das operações de manutenção.

O software de manutenção oferece várias vantagens para as organizações, tais como:

- Melhoria da eficiência operacional: o software permite uma gestão eficiente das tarefas de manutenção, contribuindo para o aumento da produtividade e diminuição dos tempos de inatividade.

- Redução de custos: o software permite o planeamento e otimização das tarefas de manutenção, o que resulta em menores custos de manutenção.

- Maior confiabilidade: o software permite a monitorização e a prevenção de falhas, o que aumenta a confiabilidade dos equipamentos e reduz os riscos de interrupções na produção.

- Melhoria da segurança: o software permite a gestão de tarefas de manutenção de forma segura, o que reduz os riscos de acidentes e incidentes.

- Gestão de ativos: o software permite uma gestão eficiente dos ativos, incluindo a monitorização do desempenho, a previsão da vida útil e a programação da manutenção preventiva.

Concluindo, o software de manutenção é uma ferramenta fundamental pelas razões enumeradas, suportando assim as organizações na tomada de decisões sustentadas, fornecendo importantes informações para melhorar a eficiência operacional e a rentabilidade da organização.

2.9. CONCLUSÃO

Este capítulo teve como objetivo apresentar uma reflexão sobre o estado da arte da função manutenção. Abordou-se a importância da manutenção para as organizações, sendo ela essencial para garantir a continuidade dos processos de produção e serviços. Seguidamente abordaram-se os objetivos da manutenção, de modo a considerar o principal objetivo que tem como prioridade prolongar o ciclo de vida útil dos equipamentos, evitando falhas, diminuindo os tempos de inatividade, bem como a redução dos custos de manutenção.

Posteriormente, foram referidos os tipos e as estratégias de manutenção tendo em conta as normas em vigor. As normas de manutenção incluem regulamentos e padrões que devem ser seguidos para garantir a segurança e a eficiência dos equipamentos.

Deu-se uma relevante importância à filosofia *Lean* na manutenção, dado que pode contribuir para a eliminação de desperdícios e para o aumento da eficiência na manutenção de equipamentos.

Por último, abordaram-se os sistemas de informação na manutenção e sua respetiva importância.

3. EVOLUÇÃO DA MANUTENÇÃO INDUSTRIAL

As estratégias de manutenção passaram por uma evolução natural e gradual ao longo das diferentes Revoluções Industriais, conforme se pode constatar na figura 2.

Figura 2 Evolução das atividades e métodos de manutenção

A filosofia de manutenção reativa ou corretiva, é a mais comum em organizações com poucos recursos e incapacidade de reter o pessoal de manutenção. Durante anos, a indústria lutou para abandonar esta filosofia de manutenção, fundamentalmente porque é cara, originando tempos de inatividade significativos e aumento de custos.

Por sua vez, a manutenção preventiva é uma filosofia de manutenção baseada num conjunto de tarefas pré-definidas e destinadas a prevenir falhas, mitigar o risco de falha e o número e duração de paragens não programadas. A Indústria 4.0 introduziu uma nova abordagem, fábricas interligadas e orientadas por dados [25]. Com base neste conceito, a tendência dominante na manutenção é passar da manutenção reativa para a manutenção preditiva, também designada por manutenção inteligente.

3.1. INTRODUÇÃO À INDÚSTRIA 4.0

Ao longo da sua história, a humanidade passou por muitas evoluções tecnológicas que podem ser vistas como transições de um estado que tendem a ocorrer gradualmente. Quando ocorre uma mudança significativa ou disruptiva, essas transições são designadas como revoluções. Estas revoluções podem durar anos ou décadas e nem sempre implicam a extinção do estado anterior. A maioria destas revoluções foram desencadeadas pela emergência de novas tecnologias, novas ideias, ou políticas num determinado domínio, mas o seu impacto acaba por ser transversal.

O setor da indústria que teve origem no século XVIII, com a necessidade de produção em massa, tem evoluindo rapidamente, impulsionada pela automatização e globalização dos mercados. Tem vindo a ser marcado pelo desenvolvimento de tecnologias disruptivas que originam revoluções com um significativo impacto social e económico.

Antes da revolução industrial, a produção era totalmente artesanal, caracterizada pelo baixo volume de produção, produtos não uniformizados, elevados custos de produção e de baixa qualidade. Atualmente, muita da mão de obra foi substituída por automatismos, tendo a indústria progredido ao longo dos anos através de vários estágios e alterações, encontrando-se a sociedade atual a vivenciar a quarta revolução industrial.

A indústria 4.0 impacta em muitos setores de atividade, entre os quais, o da saúde. No atual contexto económico, marcado pela globalização e exigência dos mercados, também as organizações do setor da saúde estão a ser pressionadas a melhorar o desempenho e a

eficiência dos seus processos, tendo como objetivo o reforço da competitividade e a satisfação das necessidades dos seus clientes.

Conectividade, disponibilidade de dados, digitalização, novos equipamentos, aumento da qualidade de *stocks*, novas ferramentas, produção controlada, etc., são algumas das estratégias e tecnologias que contribuem para alcançar este objetivo e que de certa forma deu origem ao que se chama de Indústria 4.0. Este nome surgiu de um projeto da indústria alemã, denominado *Plattform Industrie 4.0* (Plataforma da Indústria 4.0), lançado em 2011, na Feira de Hannover [35].

A integração destas novas tecnologias digitais numa organização, tais como as tecnologias da Internet das Coisas (IoT), *Big Data*, integração de métodos de inteligência artificial (IA), realidade aumentada (AR) e virtual (VR), sistemas ciber físicos (CPS), sensores ligados em rede, entre outras, combinadas com técnicas de automação industrial, proporcionam a conexão entre o mundo real e virtual. São estas tecnologias que contribuem para a melhoria da produtividade das organizações, através da otimização dos processos e estimulam o desenvolvimento de produtos e serviços inteligentes.

Esta nova abordagem, vai permitir às organizações estar mais bem preparadas para as exigências dos atuais mercados, contribuído para a maximização da disponibilidade dos equipamentos, redução de paragens não planeadas e redução de custos através da implementação da estratégia de manutenção preditiva.

A manutenção preditiva têm-se revelado uma abordagem promissora, fornecendo importantes informações sobre a restante vida útil dos equipamentos, através de previsões baseadas na recolha de dados provenientes de vários sensores e componentes de equipamentos. Na figura 3 pode ser observada a tendência mundial para a aplicação deste tipo de manutenção inteligente com uma previsão do ano 2022 a 2030.

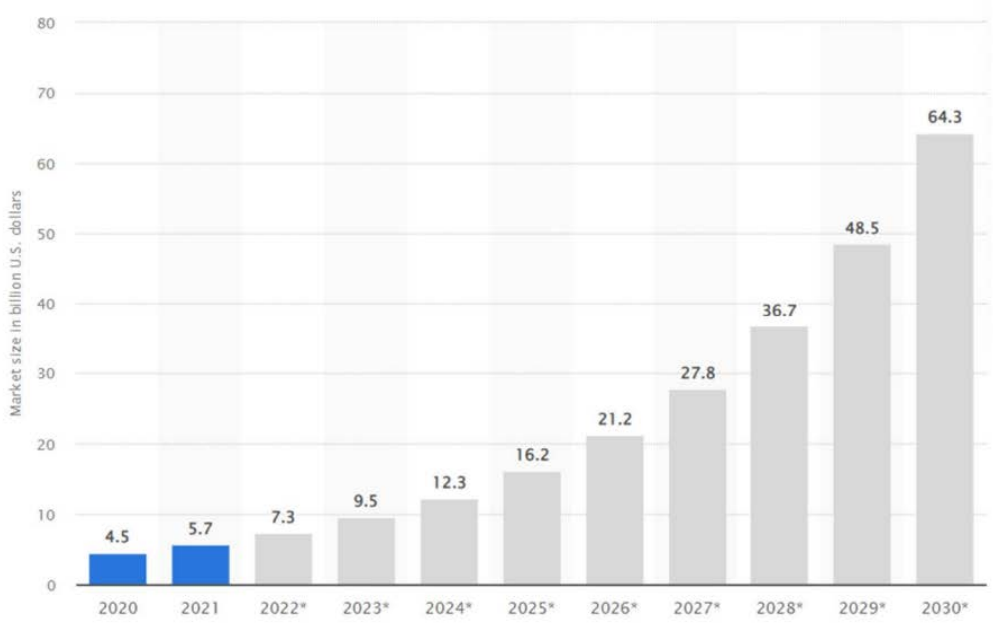


Figura 3 Dimensão do mercado da indústria 4.0 nos anos 2020 e 2021 com projeção até ao ano 2030 [14]

Esta projeção é de extrema importância para compreender o crescimento e o potencial futuro da indústria 4.0. Observa-se uma tendência de crescimento significativo na adoção e implementação da indústria 4.0, prevendo-se que o mercado continue a expandir-se rapidamente, abrangendo um número cada vez maior de setores e empresas.

3.1.1. INDÚSTRIA 4.0

A introdução da tecnologia dos computadores e automação na Indústria 3.0 teve um impacto significativo na indústria, permitindo aumentos significativos na eficiência, precisão e rapidez de produção. Atualmente, a Indústria 4.0, também conhecida como Indústria inteligente, transporta esta evolução ainda mais longe, permitindo que os computadores e equipamentos industriais comuniquem e colaborem entre si para automatizar e otimizar processos. A combinação de CPS e a IoT, permite a recolha e análise em tempo real de dados dos equipamentos e a tomada de decisões sem intervenção humana. Esta aproximação, possibilita que o conceito de fábrica inteligente se torne uma realidade, permitindo aumentos significativos na eficiência e na flexibilidade da produção.

Ao disponibilizar suporte a máquinas inteligentes que se tornam ainda mais inteligentes à medida que mais dados forem processados, as fábricas, irão tornar-se mais eficientes, produtivas e flexíveis, reduzindo os resíduos e desperdícios, o que é benéfico tanto para a

empresa como para o meio ambiente. Em última análise, é a rede dessas máquinas que se encontram conectadas digitalmente, que geram e partilham informações, que dá o verdadeiro poder à Indústria 4.0 e que permite às organizações obterem uma visão completa do seu negócio e tomar decisões sustentadas.

A Indústria 4.0, veio permitir às organizações inovar em vários aspetos dos seus negócios, desde o desenvolvimento de novos produtos até a produção, melhoria e distribuição dos mesmos. Transformou radicalmente o mundo industrial e produtivo, introduzindo novas tecnologias e abordagens que permitem uma maior intercomunicação entre os equipamentos e a automação dos processos. A IoT permite a ligação entre equipamentos e a recolha de dados em tempo real, o *Big Data* permite o armazenamento e análise de grandes volumes de dados, a inteligência computacional permite a tomada de decisões baseadas em dados e a automação permite a otimização e melhoria dos processos.

Efetivamente, esta tecnologia digital permite às empresas reagir com maior agilidade às mudanças do mercado, adaptando-se rapidamente a novos desafios e oportunidades. A automação e a inteligência artificial também permitem aumentar a eficiência operacional, através da melhoria da qualidade e rapidez dos processos, reduzindo custos e desperdícios. Como resultado, serão produzidos bens de forma mais eficiente ao longo da cadeia de valor, aumentando a competitividade da empresa.

A implementação de uma fábrica eficiente com base no modelo de Indústria 4.0 requer o uso de uma variedade de tecnologias multidisciplinares, incluindo IoT, automação avançada, inteligência artificial, robótica, *Big Data*, análise de dados, *cloud computing* e sistemas de tomada de decisão. Embora algumas dessas tecnologias já estejam maduras o suficiente para serem implementadas em massa, outras ainda precisam ser desenvolvidas e melhoradas antes de serem utilizadas em larga escala.

A comunicação entre esses dispositivos e sistemas, através da Internet, é fundamental para o funcionamento da Indústria 4.0, pois permite a coordenação e automação dos processos. No entanto, é importante notar que a implementação da Indústria 4.0 é um processo contínuo e evolutivo, e as empresas devem estar preparadas para adaptar-se às mudanças tecnológicas e às necessidades do mercado. A Figura 4 apresenta uma visão geral dos componentes da Indústria 4.0.

Figura 4 Componentes da indústria 4.0, no contexto de uma fábrica inteligente [4]

A aplicação dos componentes desta quarta revolução industrial veio atualizar a estratégia da manutenção, permitindo às empresas prolongar a vida útil dos equipamentos e reduzir os custos através da comunicação contínua e instantânea entre as diferentes máquinas/equipamentos, seja na produção ou na cadeia de abastecimento. Um dos pilares fundamentais da indústria 4.0 é a manutenção preditiva, pois permite prever e evitar avarias ou anomalias, através da recolha e análise de dados em tempo real, o que permite reduzir custos e paragens não planeadas.

Estratégias inteligentes de manutenção preditiva e remota são agora implementadas pelas empresas, auxiliando a monitorizar, manter e otimizar os seus equipamentos em tempo real de modo a antecipar falhas.

3.1.2. INDÚSTRIA 5.0

A indústria 5.0 é a quinta geração da indústria e consiste em combinar a tecnologia desenvolvida na indústria 4.0 com a colaboração humana. Ela tem como objetivo a criação de sistemas altamente automatizados e conectados que possam adaptar-se e aprender com a interação humana. A Sociedade 5.0 é uma abordagem que procura aproveitar essa colaboração para criar soluções inovadoras e melhorar a qualidade de vida. Concluindo, a indústria 5.0 não substitui a indústria 4.0, mas complementa-a, trazendo uma nova forma de colaboração humano-tecnologia para o desenvolvimento de soluções avançadas.

A indústria 4.0 foi caracterizada por avanços tecnológicos significativos, como a IoT, automação avançada, IA, computação em nuvem, entre outros, contudo a indústria 5.0

demonstra que a tecnologia não é tudo. A indústria 5.0 enfatiza a importância da colaboração entre o ser humano e a máquina e o realce nas interações humanas, onde a segurança e conforto passam a ser mais visadas, enquanto os serviços tecnológicos permitem o aumento de produtividade e do serviço na área industrial.

O avanço da indústria 4.0, exige cada vez mais por parte dos processos industriais uma maior assertividade nos resultados de inspeção, manutenção, planeamento e gestão. Este objetivo é alcançado graças ao uso de sistemas integrados, que permitem a monitorização e avaliação dos resultados através de indicadores e informações provenientes de diversas fontes.

Tecnologias multidisciplinares estão cada vez mais presentes nas indústrias, tais como, sensores inteligentes, a IoT e armazenamento em nuvem, entre outras, permitindo a conectividade de todos os objetos físicos e virtuais dentro do mesmo ambiente. Isto permite a recolha de dados em tempo real, o que é fundamental para a manutenção preditiva e otimização de processos.

A maioria dos setores industriais estão ainda na revolução 4.0 enquanto a sociedade migra para o 5.0, onde tudo já se apresenta de forma conectada e com interação do ser humano. A sociedade 5.0 atenua o medo causado pela indústria 4.0 e promove a união e sinergia do homem com a máquina, contribuindo assim para um desenvolvimento equilibrado e sustentável.

Na figura 5 apresenta-se uma visão sobre a evolução das revoluções industriais.

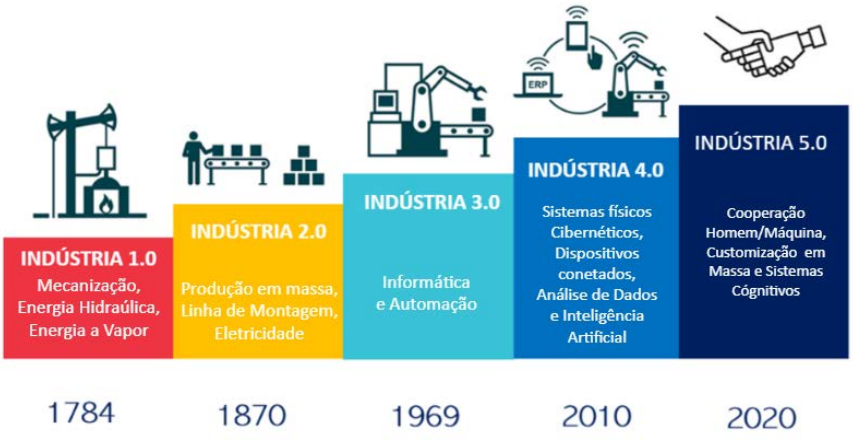


Figura 5 Evolução das revoluções industriais

No atual contexto, a função manutenção tem como principais objetivos maximizar a disponibilidade de equipamentos, reduzir os tempos de inatividade e os custos de manutenção. Suportada em diferentes tecnologias, tais como, robôs autónomos, simulações, IoT, modelos de aprendizagem automática, computação em nuvem, *Big Data* e realidade aumentada, a manutenção tende a reduzir de intervenções corretivas e a melhorar a gestão de ativos.

À medida que aumenta o controlo da manutenção, deve-se aumentar o conhecimento técnico para suporte, de modo a permitir aumentar o conhecimento dos equipamentos e as suas necessidades de manutenção, contribuindo assim para uma melhor gestão de ativos.

Neste contexto, a formação e o desenvolvimento profissional assumem um papel de extrema importância de modo a garantir que as equipas técnicas de manutenção adquiram as competências necessários para realizar as tarefas de manutenção de forma eficiente e segura. A manutenção passa a ser mais integrada, de forma a trazer a manutenção, engenharia de manutenção e gestão da manutenção para o mesmo nível e mesmo foco.

Os programas de formação habitualmente incluem formação técnica e específica nos equipamentos, formação em gestão de ativos, engenharia de manutenção, e outras aptidões relacionadas com as tarefas a desenvolver.

Paralelamente, combinar a manutenção, engenharia de manutenção e a gestão da manutenção num único nível e foco permite uma abordagem mais holística e integrada para a manutenção. Esta abordagem, garante que sejam consideradas todas as perspetivas e necessidades relacionadas à manutenção, incluindo aspetos técnicos, gestão e financeiros, contribuído assim para melhorar a eficiência, qualidade, segurança.

A manutenção baseada nas condições, também conhecida como manutenção preditiva, é apoiada por toda a tecnologia já mencionada, totalmente apoiada pela indústria 4.0 e inserida na 5.0, assumindo assim um importante papel na transformação digital das operações de manutenção.

Estas tecnologias permitem a supervisão contínua dos equipamentos e a recolha de dados em tempo real, que podem ser utilizados para prever falhas e planear de forma proativa a manutenção dos equipamentos antes da falha. Esta aproximação, permite a tomada de

decisões baseadas em dados, ao contrário de avaliações humanas, contribuindo para a maximizar a eficiência, a qualidade e a segurança, bem como a redução de custo. Portanto, muito mais que uma revolução em si, é uma oportunidade de melhoria contínua e evolução da gestão da manutenção, dado que permite que as decisões sejam baseadas em dados e análises, tornando-as mais rápidas, precisas e fundamentadas.

Conclui-se assim, que a função manutenção apoiada na indústria 4.0/5.0 e a sua tecnologia, tem melhorado significativamente o desempenho e a eficiência dos seus processos de manutenção. Contudo, é fundamental que o conhecimento básico de ferramentas e conceitos seja difundido, que as pessoas se beneficiem do desenvolvimento técnico e evoluam enquanto profissionais.

3.2. MANUTENÇÃO PREDITIVA

A manutenção preditiva (MPD) é uma recente abordagem de manutenção preventiva que visa melhorar o desempenho e a eficiência dos processos de fabricação, aumentando a vida útil dos equipamentos, contribuindo assim para uma gestão operacional sustentável.

A MPD, baseia-se na recolha de dados em tempo real a partir de sensores e outras fontes, que são analisados com recurso a técnicas de inteligência artificial e modelos de aprendizagem automática para prever o comportamento futuro do equipamento. Esta aproximação, permite que as equipas de manutenção possam planear e realizar intervenções com carácter preventivo antes que as falhas ocorram, o que pode contribuir para reduzir custos, paragens não planeadas e aumentar a disponibilidade e eficiência dos equipamentos.

A MPD têm-se revelado uma abordagem promissora, encontrando-se em franca expansão e a ser implementada por várias indústrias de diferentes setores de atividade. A implementação habitualmente é assegurada, avaliando a vida útil restante dos elementos responsáveis pela falha e permitindo a monitorização remota e em tempo real de falhas dos equipamentos.

A manutenção preditiva é a abordagem mais recente e avançada da manutenção, proporcionado a maior vida útil dos equipamentos, maior confiabilidade e mais ambientalmente e economicamente correta, pois pode contribuir para a redução de utilização de recursos e evitar a produção de resíduos desnecessários.

A abordagem de manutenção utilizada para identificar e corrigir potenciais problemas antes da falha dos equipamentos, é designada por manutenção proativa. Ela baseia-se na realização de inspeções regulares, testes, bem como através da recolha e análise de dados sobre o desempenho dos equipamentos. Esta abordagem de manutenção que é muito eficaz quando combinada com a manutenção preditiva, está a tornar-se cada vez mais popular entre as empresas.

Na Figura 6 apresenta-se uma visão sobre o processo de manutenção preditiva.

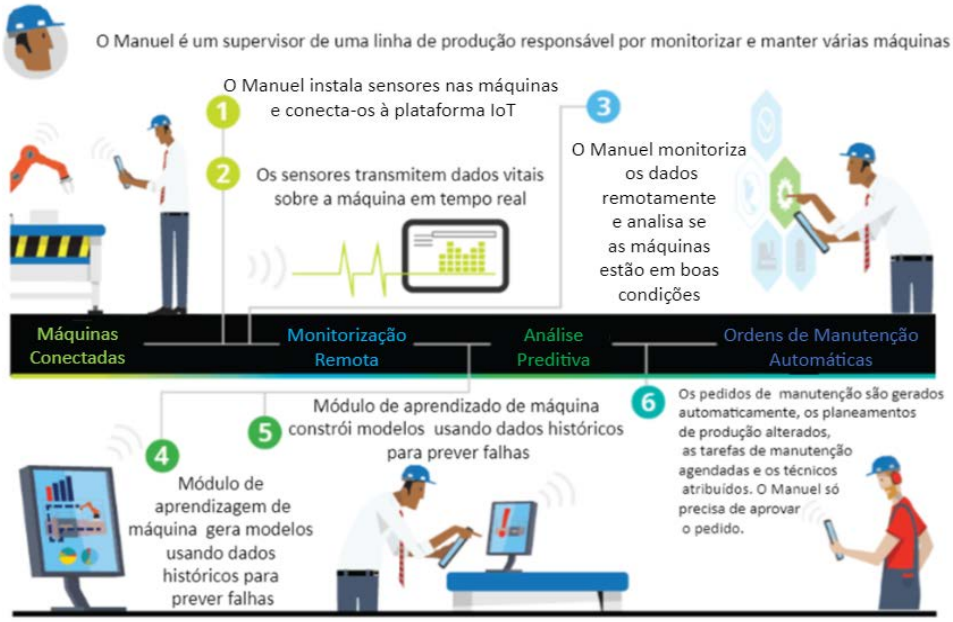


Figura 6 O processo da manutenção preditiva [21]

3.3. FERRAMENTAS DA MANUTENÇÃO PREDITIVA

No início do século XXI, e na era da quarta revolução industrial, as ferramentas computacionais e de visualização avançada baseadas nas mais recentes tecnologias são fundamentais para a transformação digital na Indústria 4.0. Estas ferramentas foram adotadas por várias empresas de diferentes setores de atividade, em várias aplicações industriais e, mais especificamente, na manutenção preditiva. Nos capítulos seguintes, são apresentadas as principais ferramentas tecnológicas aplicadas na manutenção 4.0 cada vez mais popular.

3.3.1. SISTEMAS CIBER FÍSICOS

Sistemas ciber físicos, são sistemas de informação e automação que viabilizam a troca informações, execução de comandos e a monitorização de processos produtivos à distância e em tempo real. São sistemas que permitem realizar simulações sobre o processo produtivo, num ambiente virtual, sem que o ambiente físico seja comprometido ou interrompido.

Os CPS são compostos por sensores e atuadores, para recolher e enviar dados para um sistema de controlo central, que usa algoritmos para analisar os dados e tomar decisões sobre como controlar os componentes físicos do sistema. Os dados são comunicados em tempo real ao ambiente virtual, que os representa em interfaces gráficos amigáveis e intuitivos, como se formasse um "gêmeo virtual" do mundo físico. Os CPS transmitem informações e dados em tempo real, interligados por meio do espaço cibernético, mundo virtual para o mundo real, permitindo que o mundo real possa atuar no sistema produtivo, seja no controlo, reprogramação, acompanhamento ou atuando da maneira mais conveniente no sistema produtivo.

Os CPS são usados em diversas aplicações e em diferentes setores de atividade, como automação industrial, transporte inteligente, sistemas de saúde entre outros. Os CPS podem contribuir para melhorar a eficiência, precisão e segurança das operações, permitindo que dispositivos físicos sejam monitorizados e controlados remotamente. Eles são desenhados para serem altamente confiáveis, seguros e adaptáveis a ambientes em constante mudança.

3.3.2. IOT (INTERNET OF THINGS)

A digitalização está a transformar diversos setores de atividade, entre os quais, os sistemas médicos e de saúde em todo o mundo, com as tecnologias IoT *(Internet of Things)* a desempenhar um papel cada vez mais fundamental, principalmente na manutenção remota e preditiva de equipamentos médicos.

As tecnologias IoT oferecem um suporte à manutenção preditiva e remota para auxiliar as organizações da área da saúde a monitorizar, manter e otimizar os seus equipamentos em tempo real. Os dados de desempenho de um dispositivo médico podem ser consolidados e analisados remotamente de modo a antecipar falhas.

Os equipamentos médicos contêm diferentes componentes, têm um determinado ciclo de vida e precisam ser substituídos periodicamente. Normalmente, as organizações estão dotadas de equipas técnicas locais para assegurar a manutenção dos equipamentos de modo a garantir o seu correto funcionamento, contudo, uma eventual falha no equipamento, pode conduzir a um considerável tempo de inatividade e interrupção de tratamentos.

Tendo como objetivo o aumento da eficiência dos processos de manutenção, a redução da necessidade de verificações manuais e da probabilidade de falhas, as tecnologias IoT podem ser usadas para ler os dados dos componentes dos equipamentos. Estes dados serão posteriormente utilizados para monitorizar o seu tempo de operação, identificação do tempo em falta até chegar ao fim do seu ciclo de vida e identificar um eventual padrão e/ ou regime de funcionamento anómalo ou próximo dos seus limites estabelecidos. Desta forma, será possível às equipas técnicas de manutenção planear de forma antecipada as tarefas de manutenção nos equipamentos e caso necessário solicitar as peças de reposição, contribuindo assim para um aumento da disponibilidade dos equipamentos para os utilizadores finais bem como uma maior estabilidade e segurança dos processos clínicos.

A implementação da manutenção remota e preditiva de equipamentos médicos requer que os dispositivos estejam dotados de recursos de IoT integrados. Uma hipótese será a utilização de uma Gateway IoT conectada ao equipamento médico, para que as equipas de manutenção possam interagir com os componentes dos equipamentos, ler os seus dados, permitir a realização de análises e a configuração de regras e/ou alertas necessários.

Existem diferentes níveis de certificação exigidos quando os dados e máquinas são mantidos em ambiente hospitalar, em comparação com quando os dados são transferidos através de infraestruturas 3G, 4G e agora 5G.

À medida que mais fabricantes procuram ativamente integrar a capacidade de IoT em dispositivos para permitir a manutenção remota e preditiva de equipamentos médicos, é provável que o design nativo de IoT se torne um normativo.

Como resultado da pandemia de Covid-19, também é provável que vejamos mais dispositivos médicos entrando na casa do paciente para permitir o tratamento remoto. Muitos fabricantes de dispositivos médicos estão a desenvolver novos dispositivos que podem ser transportados para casa dos pacientes, formando parcerias com especialistas em

IoT para permitir a integração efetiva de tecnologias IoT com todas as medidas necessárias de segurança de dados.

3.3.3. BIG DATA

Big Data, é o termo utilizado em inglês para descrever uma grande quantidade de dados, estruturados ou não, que podem ser difíceis de processar através de métodos tradicionais de armazenamento e análise. Estes dados podem ser fornecidos por várias fontes, como redes sociais, dispositivos móveis, sensores de IoT, entre outros. A capacidade de processar e analisar esses dados em grande escala é chamada de "*Big Data Analytics*", sendo esta uma área em franco crescimento, pois permite às organizações tomar decisões sustentadas em informações valiosas e conhecimentos adquiridos a partir da análise de dados.

A recolha, processamento e análise de *Big Data* em tempo real de sistemas ciber físicos são fases estratégicas para alavancar uma transformação inteligente na manutenção, em particular na implementação da estratégia da manutenção preditiva, planeamento e gestão de risco.

Nesta abordagem, as intervenções passam a ser planeadas e otimizadas com recurso a ferramentas de inteligência artificial, modelos de aprendizagem automática ou por meio de modelos estatísticos e abordagens baseadas nos dados e informações recolhidos pelos diferentes sensores.

Figura 7 Modelo 5V 's do *Big Data* [38]

A caraterização dos dados massivos ou *Big Data,* habitualmente é feita de acordo com três "V's", os V's de Volume, Variedade e Velocidade, aos quais se acrescentam outros "V's" complementares, como Valor e Veracidade/Validade, como se pode observar na figura 7.

Em suma, são inúmeras as possibilidades de aplicação de *Big Data Analytics* na indústria. Todavia, o ponto chave para a compreensão do papel desta tecnologia no caso da chamada Quarta Revolução Industrial, e da Indústria 4.0 em particular, está na visão de que dados obtidos e armazenados em grande quantidade representam um novo ativo organizacional, comparáveis a dinheiro em caixa, aplicações financeiras, máquinas e equipamentos, imóveis ou *stocks*. Embora estes dados sejam intangíveis, fazem parte do capital intelectual das organizações, têm valor e podem ser usados para aumentar a eficiência, receita, identificar novos mercados e oportunidades de negócios.

3.3.4. REALIDADE VIRTUAL

A realidade virtual é um ambiente digital que permite aos utilizadores ter uma experiência imersiva num mundo simulado. São utilizadas tecnologias, tais como óculos de VR, sensores de movimento e computação gráfica, que em conjunto reproduzem a presença física do utilizador num ambiente virtual ou outro tipo de dispositivo de exibição. A realidade virtual permite aos utilizadores interagir com objetos e cenários virtuais como se estivessem presentes no mundo real, criando uma sensação de presença e imersão. É muito utilizada em jogos, formações e simulações.

Figura 8 Realidade virtual

3.3.5. REALIDADE AUMENTADA

A realidade aumentada (AR) é uma tecnologia emergente desenvolvida com base na realidade virtual (VR), que combina elementos virtuais com elementos do mundo real, para criar uma experiência imersiva que permite aos utilizadores interagir simultaneamente com o mundo virtual e o mundo real. Gera informações virtuais tridimensionais através de um computador, *smartphone*, *tablets*, óculos de realidade virtual, entre outros.

A AR tem múltiplas aplicações em diferentes setores, como jogos, educação, saúde, entretenimento, formação, manutenção, entre outros. Atendendo que vários processos de produção exigem elevados padrões de qualidade e taxas de erro próximas de zero para garantir os requisitos e a segurança do utilizador final, a AR também pode fazer parte do equipamento dos operadores com interfaces para aumentar a produtividade, precisão e qualidade.

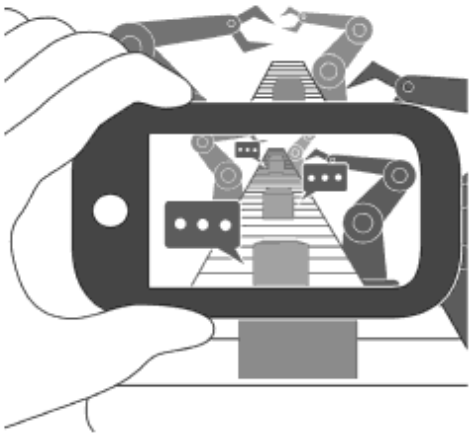


Figura 9 Realidade aumentada

3.3.6. INTELIGÊNCIA ARTIFICIAL

Inteligência Artificial (IA), é o termo chave na transição para a Indústria 4.0. É uma área da ciência que tem como objetivo estudar dispositivos e sistemas computacionais criados para interagir com seres humanos de uma forma que possa ser considerada inteligente.

A associação entre computação e inteligência foi proposta pela primeira vez por Alan Turing no seu artigo publicado em 1950 "*Computing Machinery and Intelligence*"[34]. Em 1956, *John McCarthy* [35], criou o termo "Inteligência Artificial" para descrever um mundo em que as máquinas poderiam resolver problemas que antes só podiam ser resolvidos por humanos. A tecnologia de IA é utilizada em diversas aplicações, incluindo o

processamento de linguagem, visão computacional, sistemas de tomada de decisão e tradução de idiomas.

A IA pode servir como base para a integração de vários objetos e sistemas inteligentes, de forma a aperfeiçoar a experiência dos utilizadores. Compreende criar interações mais perfeitas entre dispositivos, automatizar tarefas e processos e disponibilizar experiências personalizadas com base em necessidades e preferências individuais.

A área da IA encontra-se em constante evolução, tendo potencial de revolucionar muitos setores e transformar o paradigma de como vivemos e trabalhamos.

3.3.7. MACHINE LEARNING E DEEP LEARNING

Machine Learning (ML) e *Deep Learning* (DL) são ambos subcampos da IA, que se concentram no desenvolvimento de algoritmos e modelos que permitem que os computadores aprendam com os dados e aperfeiçoem seu desempenho ao longo do tempo. A figura 10 identifica todas as categorias, métodos e modelos de ML aplicáveis para projetos relacionados com a área da manutenção.

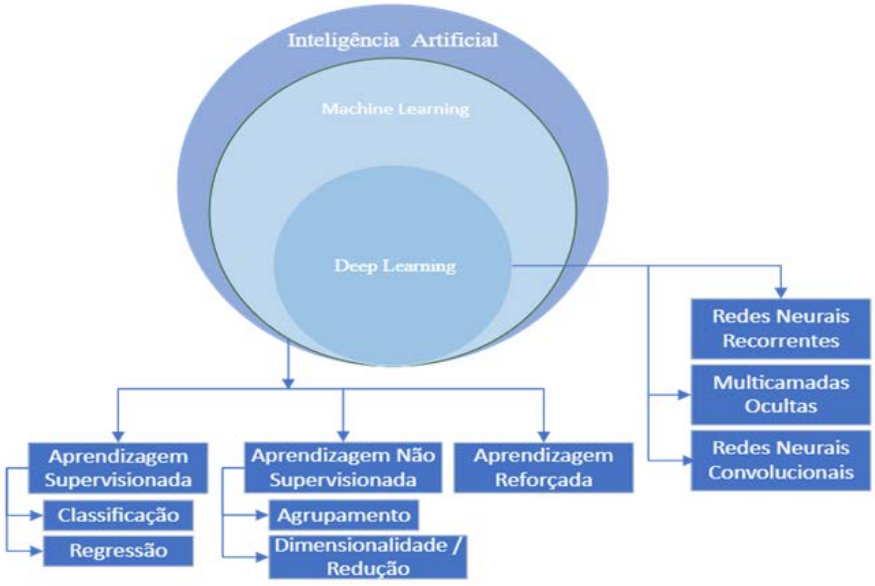


Figura 10 Classificação de algoritmos de Inteligência Artificial e categorias de linguagens de aprendizagem automática [20]

ML, compreende a utilização de algoritmos e modelos estatísticos que podem identificar padrões e tendências nos dados e gerar previsões ou decisões com base nessas informações. Existem diferentes tipos de aprendizagem automática, tais como a

aprendizagem supervisionada, aprendizagem não supervisionado e aprendizagem reforçada.

Aprendizagem supervisionada: A aprendizagem supervisionada, é um método de aprendizagem automática onde o algoritmo é treinado com dados que possuem as respostas conhecidas (para cada entrada é conhecida a saída desejada). O algoritmo compara as saídas produzidas com as saídas desejadas e procura por erros, ajustando seus parâmetros para melhorar a sua precisão. Após várias iterações 'n', o modelo deve ser capaz de classificar novos dados sem marcação. Este método é apropriado para aplicações de classificação.

Aprendizagem não supervisionada: Aprendizagem não supervisionada, é um método de aprendizagem automática onde o algoritmo não é alimentado com dados de aprendizagem rotulados. No entanto, o modelo recebe um conjunto de dados e é esperado que encontre de forma autónoma padrões ou estruturas nos dados fornecidos. É habitualmente utilizado para tarefas de agrupamento e deteção de valores atípicos (*outliers*). Difere da aprendizagem supervisionada, onde não existe diferença entre dados de aprendizagem e os dados de teste.

Aprendizagem reforçada: A aprendizagem reforçada (RL), é um método de aprendizagem automática em que um agente aprende a tomar decisões interagindo com um ambiente. Na RL, o agente recebe observações do ambiente e, em resposta, seleciona ações a serem executadas. O ambiente então faz a transição para um novo estado e o agente recebe uma recompensa ou penalidade com base na ação que realizou. O objetivo do agente é aprender uma política que maximize a recompensa cumulativa esperada ao longo do tempo.

A RL tem sido utilizada para treinar agentes para executar uma vasta gama de tarefas, como jogar, controlar robôs e otimizar a alocação de recursos.

DL é um subconjunto de ML inspirado na estrutura e função do cérebro humano, especificamente nas redes neuronais. Compreende a utilização de uma série de algoritmos chamados de redes neuronais artificiais (ANNS) para processar e analisar grandes volumes de dados complexos e não estruturados, como imagens, áudio e texto. Técnicas de DL podem ser aplicadas a equipamentos industriais em diferentes situações, como deteção de falhas, previsão de falhas, etc.

Os algoritmos de DL podem melhorar o seu desempenho ao longo do tempo, reforçando os parâmetros da rede neuronais com base nos dados processados. Os algoritmos de DL mais utilizados, são as redes neuronais convolucionais (CNNs) para processamento de imagens e redes neuronais recorrentes (RNNS) para processamento de dados sequenciais.

Concluindo, ML é um campo mais amplo que inclui várias técnicas e algoritmos para treinar modelos e fazer previsões, enquanto DL é uma abordagem específica que utiliza redes neuronais para processar e analisar dados. Ambos são amplamente utilizados em diferentes aplicações, como o processamento de linguagem natural, visão computacional ou reconhecimento de voz.

3.3.8. ALGORITMOS

Os algoritmos em ML são a lógica por trás dos modelos de aprendizagem analisados no subcapítulo anterior. Eles são responsáveis por processar os dados e fazer previsões ou decisões com base nas entradas. Diferentes algoritmos são usados para diferentes tarefas e tipos de aprendizagem. De seguida apresentam-se alguns dos algoritmos utilizados com mais frequência em ML.

Regressão linear: a regressão linear é um algoritmo de aprendizagem automática supervisionada utilizado para prever uma variável de resultado contínua (também conhecida como variável dependente) com base em uma ou mais variáveis preditivas (também conhecidas como variáveis independentes). Este algoritmo assume que existe uma relação linear entre as variáveis preditivas e a variável de resultado. A regressão linear pode ser usada para regressão linear simples ou múltipla.

A regressão linear simples é usada quando há apenas uma variável preditiva e é usada para ajustar uma linha reta aos dados. A equação da reta é representada pela seguinte equação:

$$y = b0 + b1 * x \quad (1)$$

Onde y é a variável de resultado, x é a variável preditiva, b0 é a intercetação de y e b1 representa o declive da reta.

A regressão linear múltipla é usada quando há duas ou mais variáveis preditivas e é usada para ajustar um hiperplano aos dados. A equação do hiperplano é representada pela seguinte equação:

$$y = b0 + b1x1 + b2x2 + \cdots + bn * xn \quad (2)$$

Onde y é a variável de resultado, x1, x2, ..., xn são as variáveis preditivas, b0 é a intercetação y e b1, b2, ..., bn são os coeficientes das variáveis preditivas.

Na figura11 apresenta-se uma exemplificação de uma regressão linear.

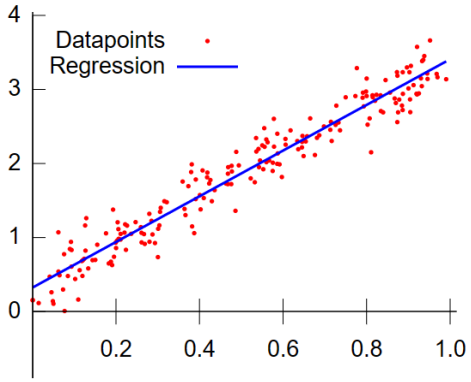


Figura 11 Exemplificação gráfica de regressão linear

A regressão linear pode ser usada para fazer previsões, identificar padrões e tendências e compreender a relação entre as variáveis. A regressão linear é um algoritmo simples, mas poderoso, amplamente utilizado em muitos campos, incluindo finanças, economia ou engenharia.

Árvores de decisão: Uma árvore de decisão (DT) é um tipo de algoritmo de aprendizagem automática supervisionado utilizado para classificação de problemas. O objetivo de um DT é prever o nome da classe de uma determinada entrada, com base nos valores dos seus recursos.

As DT são amplamente utilizadas em muitas aplicações, como análise de risco de crédito, diagnóstico médico ou processamento de linguagem natural. São fáceis de interpretar, entender e visualizar, no entanto, são propensas a *overfitting*, que é um problema comum em aprendizagem automática, quando o modelo se torna muito complexo, conduzindo a um fraco desempenho.

Na figura12 apresenta-se uma exemplificação de uma árvore de decisão.

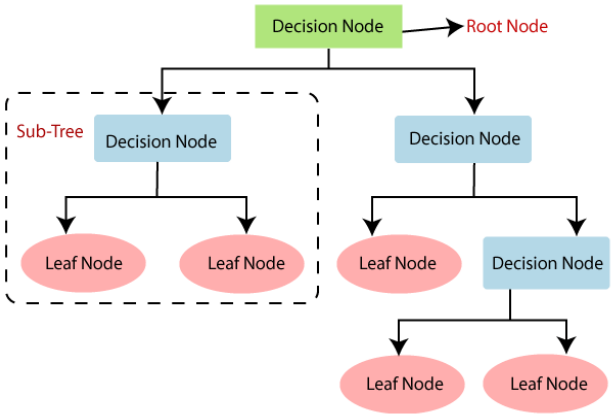


Figura 12 Exemplo de uma DT

Regressão logística: é um algoritmo de aprendizagem automática supervisionada utilizado para problemas de classificação. É idêntico à regressão linear, contudo, em vez de prever uma variável de resultado contínua, prevê a probabilidade de um resultado binário (ou seja, um valor de 0 ou 1). O objetivo da regressão logística é encontrar o melhor modelo de ajuste para descrever a relação entre as variáveis preditivas e a variável de resultado binária.

O modelo de regressão logística é representado pela seguinte equação:

$$p(x) = \frac{1}{(1 + e^{-b0-b1x1-b2x2-\cdots.bn+xn)})} \quad (3)$$

Onde *p(x)* representa a probabilidade de o resultado ser 1, x1, x2, ..., xn são as variáveis preditivas e *b0, b1, b2, ..., bn* são os coeficientes das variáveis preditivas. Os coeficientes são estimados a partir dos dados durante o processo de aprendizagem.

Uma vez que o modelo tenha sido treinado, ele pode ser usado para fazer previsões sobre novos dados. A probabilidade prevista é então transformada em um resultado binário (0 ou 1).

A regressão logística é um algoritmo simples, no entanto, poderoso, amplamente utilizado em diferentes áreas, incluindo finanças, economia e medicina. É particularmente útil para

problemas em que o resultado é binário, como na classificação de e-mails de spam e não-spam ou em diagnósticos médicos em que o resultado é doente ou não.

Support Vector Machines; (SVMs) são um tipo de algoritmo de aprendizagem automática supervisionada utilizado para problemas de classificação e regressão. O objetivo de um SVM é encontrar o melhor limite (ou hiperplano) que separa as diferentes classes nos dados. A fronteira é escolhida de forma a maximizar a margem, que é a distância entre a fronteira e os pontos de dados mais próximos de cada classe, também conhecidos como vetores de suporte.

Os SVMs são baseados no conceito de maximização da margem, que é a distância entre o limite e os pontos de dados mais próximos de cada classe. O algoritmo encontra o limite que maximiza essa distância, que é chamado de hiperplano de margem máxima.

Estes algoritmos são utilizados em muitas aplicações, como classificação de texto, classificação de imagem e bioinformática, sendo eficazes em espaços de alta dimensão, com eficiência de memória e podem lidar com limites de decisão não lineares. No entanto, eles podem ser sensíveis à escolha do *kernel* (parte central de um sistema operativo que controla todos os recursos do sistema e a comunicação entre componentes de hardware e software), às configurações de parâmetros e podem ser afetados pela presença de *outliers* (valores atípicos) e dados com ruído.

Na figura13 apresenta-se uma exemplificação de um SVM.

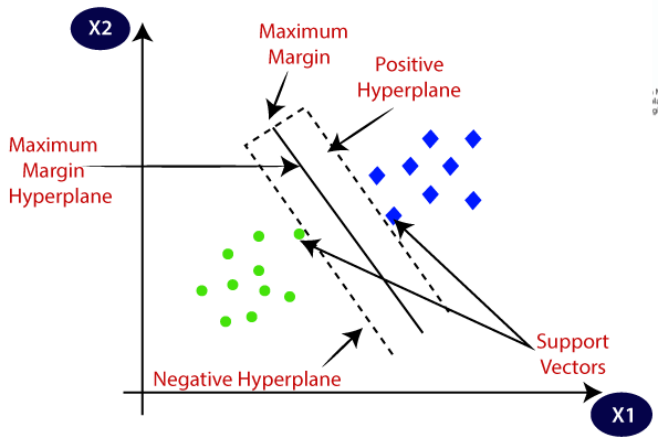


Figura 13 Exemplo de um SVM

Naive Bayes: é uma família de algoritmos probabilísticos baseados na aplicação do teorema de Bayes com suposições de independência forte entre os recursos. Eles são probabilísticos, o que significa que calculam a probabilidade de cada *tag* (marcador classificador de dados) para um determinado texto e, em seguida, imprimem a *tag* com o maior valor.

O teorema de Bayes também conhecido como Regra de Bayes ou Lei de Bayes, é utilizado para determinar a probabilidade de uma hipótese com conhecimento prévio. Depende da probabilidade condicional.

A fórmula do teorema de Bayes é dada por:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (4)$$

Onde,

P(A|B) é a probabilidade posterior: probabilidade da hipótese A no evento observado B.

P(B|A) é a probabilidade de verossimilhança: probabilidade da evidência dada que a probabilidade de uma hipótese é verdadeira.

O algoritmo *Naive Bayes* é utilizado principalmente em problemas de classificação de texto e processamento de linguagem natural (NLP). É simples e fácil de implementar, sendo relativamente rápido e eficiente para grandes conjuntos de dados. No entanto, a suposição de independência entre recursos geralmente não é verdadeira em problemas do mundo real, o que pode levar a um fraco desempenho.

K-Nearest Neighbors (KNN): é um algoritmo de aprendizagem supervisionada não paramétrico, baseado em instâncias podendo ser utilizado para problemas de classificação e regressão.

A ideia base do algoritmo KNN é encontrar o número K de pontos mais próximos no espaço de recursos para um determinado ponto de dados e, em seguida, atribuir o rótulo de classe com base na maioria dos pontos mais próximos. O número de vizinhos mais próximos (K) considerados é um parâmetro especificado pelo utilizador.

O algoritmo KNN funciona armazenando todos os casos disponíveis e classifica novos casos com base numa medida de similaridade (por exemplo, funções de distância). O caso atribuído à classe é mais comum entre seus K vizinhos mais próximos medidos por uma função de distância.

Com a ajuda do KNN, pode ser identificado facilmente a categoria ou classe de um determinado conjunto de dados. Ver figura 14.

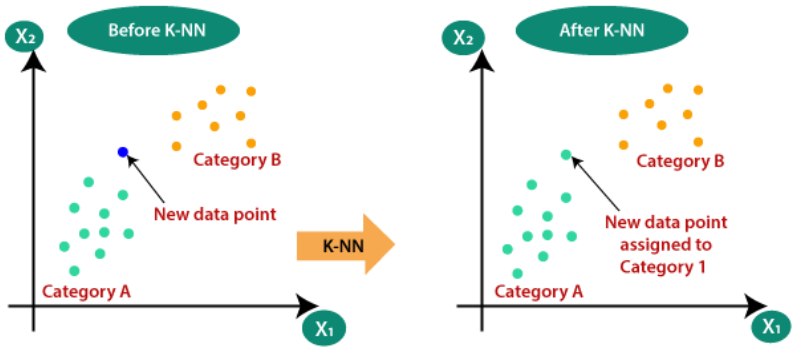


Figura 14 Exemplo de aplicação de um algoritmo KNN

O algoritmo KNN é simples e fácil de implementar, não requer que nenhuma distribuição de probabilidade seja assumida, funciona bem com um grande número de dimensões e não faz suposições sobre a linearidade dos dados. Por outro lado, requer grandes recursos de memória para armazenar todo o conjunto de dados de aprendizagem, as suas previsões podem ser lentas em grandes conjuntos de dados e podem ser sensíveis a recursos irrelevantes.

Random Forest: (RF) é um algoritmo de aprendizagem supervisionado utilizado em problemas de classificação e regressão. Consiste numa combinação de várias árvores de decisão, e é utilizado para melhorar o desempenho de uma única árvore de decisão.

A ideia básica por trás de um RF, é construir um considerável número de árvores de decisão durante o processo de aprendizagem e, em seguida, combinar as suas previsões durante o processo de teste. Cada árvore de decisão é construída utilizando um subconjunto aleatório dos dados de aprendizagem e um subconjunto aleatório dos recursos.

Ao método de construir várias árvores de decisão e combinar as suas previsões, designa-se por "*bagging*". A previsão final do RF é dada pela média das previsões de todas as árvores de decisão. Na figura15 apresenta-se uma exemplificação de um algoritmo RF.

O algoritmo RF é fácil de usar, pode ser alimentado por grandes conjuntos de dados com um elevado número de variáveis e pode ser usado para problemas de classificação e regressão.

Em resumo, o RF é uma técnica de aprendizagem que utiliza várias árvores de decisão para melhorar o desempenho e precisão de uma única árvore de decisão e prevenir o problema do *overfitting* referido na DT.

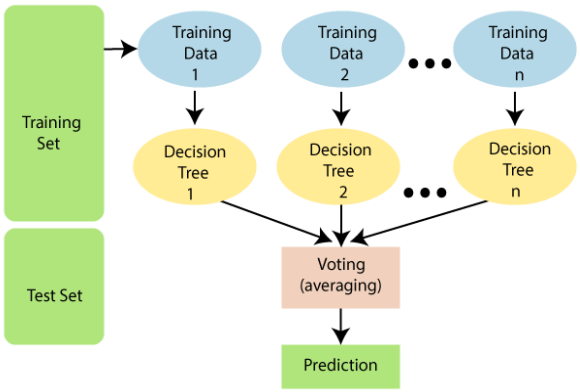


Figura 15 Exemplo de aplicação de um algoritmo *Random Forest*

Redes neuronais: é um algoritmo de DL inspirado na estrutura e função do cérebro humano. É composto por camadas de "neurónios" interligados, que processam e transmitem informações. Na figura16 apresenta-se uma exemplificação de uma rede neuronal.

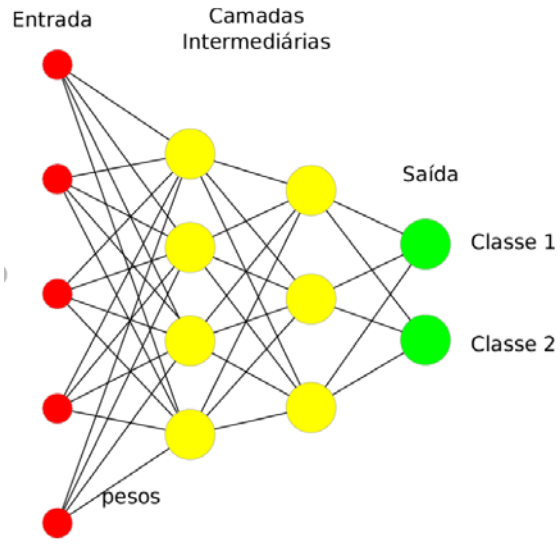


Figura 16 Rede neuronal artificial simplificada com cinco entradas, duas camadas ocultas e dois neurónios de saída

As redes neuronais são utilizadas para processar e analisar grandes volumes de dados, complexos e não estruturados, como imagens, áudio e texto, tendo demonstrado ser particularmente eficazes em tarefas em que há muitos dados e os relacionamentos entre eles não são bem compreendidos.

Existem vários tipos de redes neuronais, incluindo redes neuronais *feedforward*, redes neuronais recorrentes e redes neuronais convolucionais. As redes neuronais f*eedforward*, também conhecidas como *multi-layer perceptrons* (MLPs), são o tipo mais simples de rede neuronal, na qual a informação flui numa só direção, da entrada para a saída. As redes neuronais recorrentes, por outro lado, são projetadas para trabalhar com dados sequenciais, onde as informações podem fluir em qualquer direção. As redes neuronais convolucionais, são usadas para reconhecimento de imagem e de vídeo, onde a rede aprende a reconhecer padrões em imagens analisando as relações entre os pixels.

As redes neuronais são modelos de aprendizagem automática, mas podem ser difíceis de treinar e interpretar. Elas exigem uma grande quantidade de dados e recursos computacionais e podem ser sensíveis à escolha de hiper parâmetros.

Em resumo, as redes neuronais são um tipo de modelo de aprendizagem automática inspirado na estrutura e função do cérebro humano. Podem ser utilizadas para uma vasta gama de tarefas, sendo particularmente eficazes para problemas de elevada complexidade.

Estes são apenas alguns exemplos dos muitos algoritmos utilizados na aprendizagem automática. Cada algoritmo tem pontos fortes e pontos fracos, dependendo a escolha do problema específico e do tipo de dados a processar.

Gradient Boosting Decision Tree Classifier: (GBDT) é um tipo de algoritmo de aprendizagem automática supervisionado que combina a técnica de *boosting* com o algoritmo de árvores de decisão. Esta é uma das técnicas utilizadas para resolver problemas de classificação e regressão.

Neste algoritmo, as árvores de decisão são construídas de forma iterativa, tendo como objetivo mitigar o erro do modelo em cada iteração. Uma das vantagens do GBDT, é a sua capacidade para trabalhar com dados de alta dimensionalidade e características complexas.

3.4. CONCLUSÃO

Neste capítulo, foi apresentada uma reflexão sobre a evolução da manutenção industrial, marcada pelo avanço da tecnologia e a introdução de novos conceitos e tendências, como a Indústria 4.0 e a Indústria 5.0.

A Indústria 4.0, é caracterizada por uma forte conexão entre sistemas físicos e virtuais, automação avançada, uso de dados e análises para melhorar a eficiência e a flexibilidade da produção. Por sua vez, a Indústria 5.0 é uma evolução da Indústria 4.0, onde se inclui e valoriza a colaboração humano-máquina e a inteligência artificial.

Abordou-se também a manutenção preditiva, que é uma abordagem avançada de manutenção e que se baseia em dados e análises para prever falhas nos equipamentos antes que ocorram. Isso é possível recorrendo a tecnologias como o *Big Data*, IoT, CPS, *Machine Learning* e algoritmos avançados. Estas tecnologias permitem recolher e analisar grandes volumes de dados em tempo real, e utilizar esses dados para prever falhas e planear a manutenção de forma mais eficiente e precisa.

4. DESENVOLVIMENTO DA SOLUÇÃO

Com este capítulo, pretende-se descrever cada uma das etapas do desenvolvimento da solução implementada. A solução foi desenvolvida com recurso ao software *Microsoft Power BI*, tendo como objetivo principal a otimização do planeamento da manutenção de equipamentos médicos, com recurso a métodos de análise preditivos de modo a evitar falhas em equipamentos médicos.

A primeira fase, consistiu na recolha de dados provenientes de equipamentos médicos, assim como na recolha do histórico de intervenções realizadas nos referidos equipamentos.

Na segunda fase, procedeu-se à análise e respetivo tratamento dos dados recolhidos, tendo estes sido armazenados em tabelas no formato Excel e posteriormente importados e armazenamento num modelo de dados do *Power BI*.

A terceira fase, teve como objetivo projetar os painéis interativos no *Power BI*, tendo sido equacionada uma página inicial, páginas de navegação e páginas de conteúdo específico. Foram selecionados conjuntos de dados relevantes com o objetivo de treinar diferentes algoritmos de aprendizagem automática e obter previsões.

Esta seleção permitiu identificar o modelo mais adequado em termos de precisão e exatidão dos resultados obtidos, tendo as previsões deste modelo sido posteriormente integradas no desenvolvimento de um painel interativo específico no *Power BI*.

Os algoritmos de aprendizagem automática utilizados foram os seguintes:

- Árvores de decisão;
- Regressão linear;
- Regressão logística;
- SVM;
- *Nave Bayes;*
- KNN;
- *Random Forest;*
- Redes neuronais;
- *Gradient Boosting Decision Tree Classifier.*

Na figura 17, pode-se observar as etapas para a construção do modelo de aprendizagem:

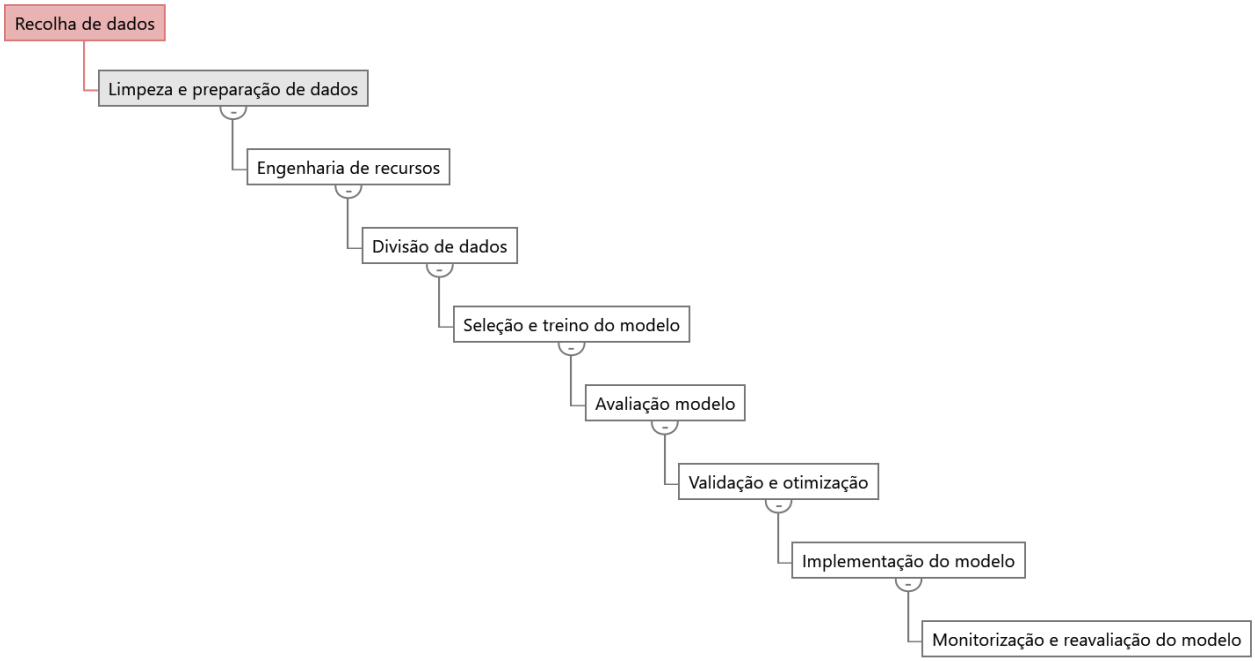


Figura 17 Etapas do desenvolvimento da solução

4.1. BASE DE DADOS

Os dados utilizados neste trabalho são provenientes de equipamentos médicos. É importante salientar, que sendo dados reais foram anonimizados antes de sua utilização neste trabalho. O processo de anonimização envolveu a remoção de informações pessoais identificáveis, tais como: nomes, identificadores únicos e outros dados sensíveis, a fim de garantir a privacidade e a confidencialidade dos dados.

A anonimização dos dados foi realizada em conformidade com as diretrizes éticas e de privacidade. Ela teve como objetivo preservar a confidencialidade dos participantes da pesquisa e cumprir com as regulamentações de proteção de dados vigentes. Ao remover informações pessoais identificáveis, garante-se a segurança e a privacidade das organizações, além de evitar qualquer divulgação não autorizada de informações sensíveis.

Os dados foram recolhidos e armazenados no formato Excel, sendo representativos do funcionamento dos equipamentos, com registos de mensagens de informação, alarmes e erros. Paralelamente, foram igualmente recolhidos e armazenados no mesmo formato dados referentes ao histórico de intervenções realizadas nos equipamentos em análise.

Os dados provenientes dos equipamentos, foram recolhidos durante o período de 01-11-2022 e 30-04-2023, tendo sido estruturados e organizados em tabelas com diferentes dados, dimensões e chaves de relacionamento entre trabelas. Posteriormente, foram transferidos para o *Power BI* para o desenvolvimento de painéis de análise interativos.

As diversas tabelas que constituem a base de dados, têm múltiplas dimensões e diferentes características, destacando-se as tabelas "*Errors_with_features*" e "*Service_reports*", a primeira relativa ao registo de eventos e parâmetros de funcionamento dos equipamentos e a segunda relativa ao registo de intervenções técnicas nos equipamentos. Estas, serviram como base para a construção dos conjuntos de dados utilizados para treinar os algoritmos propostas.

No anexo Q, poderá ser consultado o diagrama relacional entre as tabelas, incluindo as chaves primárias e estrangeiras entre as mesmas.

Para este trabalho foram utilizadas ferramentas de *Business Intelligence* e processamento estatístico, tais como o *Microsoft Power BI, Microsoft Excel, Minitab*, e a linguagem de

programação *Python* com recurso a diferentes bibliotecas, entre as quais: *Pandas, Numpy e Seaborn*.

O *Power BI*, é uma poderosa ferramenta utilizada para a visualização e análise de dados, permitindo conectar várias fontes de dados, entre as quais, arquivos CSV ou Excel, com o objetivo de desenvolver painéis interativos e relatórios para as diversas partes interessadas.

O *Minitab* é um software estatístico e de análise de dados, disponibilizando vários recursos estatísticos e ferramentas de visualização de dados, contribuindo para a análise, resolução de problemas e tomada de decisões baseada em dados.

Por último, o *Python*, é uma linguagem de programação de alto nível, interpretada e de propósito geral, destacando-se pela sua simplicidade e clareza da sintaxe, o que a torna fácil de aprender e utilizar.

4.1.1. RESTRIÇÕES DOS DADOS DISPONÍVEIS

É importante salientar que embora os dados disponíveis sejam reais, apresentam algumas limitações na sua informação, nomeadamente, no que diz respeito a falta de informação completa sobre algumas falhas, representatividade dos dados e à complexidade e/ou multidimensionalidade dos dados.

Estas limitações podem ter impacto nos algoritmos de IA e na precisão das previsões na estratégia da manutenção preditiva. Por exemplo, a falta de informação completa sobre algumas falhas, pode limitar a capacidade dos algoritmos de reconhecerem determinados padrões, enquanto a complexidade dos dados pode dificultar a identificação de relações relevantes.

Por outro lado, o facto dos dados utilizados não serem totalmente representativos dos parâmetros relevantes de funcionamento dos equipamentos, pode limitar a capacidade dos algoritmos de IA em identificar corretamente os padrões associados a diferentes tipos de falhas ou componentes.

De forma a mitigar estas limitações, foram adotadas técnicas de pré-processamento e modelação avançada. Porém, poderão permanecer ainda assim algumas limitações nos resultados, nomeadamente na confiabilidade dos resultados obtidos com os algoritmos de IA, com especial ênfase na previsão da tipologia da falha e acessórios envolvidos.

Considerando estas restrições de dados, como proposta de melhoria contínua para a implementação da estratégia manutenção preditiva, sugere-se que os fabricantes dos equipamentos médicos adaptem os protocolos de comunicação dos seus equipamentos de modo a possibilitar a recolha de dados mais relevantes sobre o seu desempenho.

4.2. ARQUITETURA DA SOLUÇÃO

A solução desenvolvida, foi projetada para atender às necessidades de análise e visualização de dados referentes ao desempenho técnico e operação de equipamentos médicos, no âmbito da aplicação da estratégia de manutenção proativa e/ou preditiva.

A primeira etapa do desenvolvimento da aplicação, consistiu na elaboração de um planeamento da arquitetura da solução, tendo sido identificados os requisitos dos dados, os painéis de controlo a serem desenvolvidos e os respetivos dados necessários.

Na figura 18 apresenta-se o mapa de ideias resumido da arquitetura da solução na fase inicial do planeamento, podendo ser consultado na secção em anexo a versão completa do mesmo.

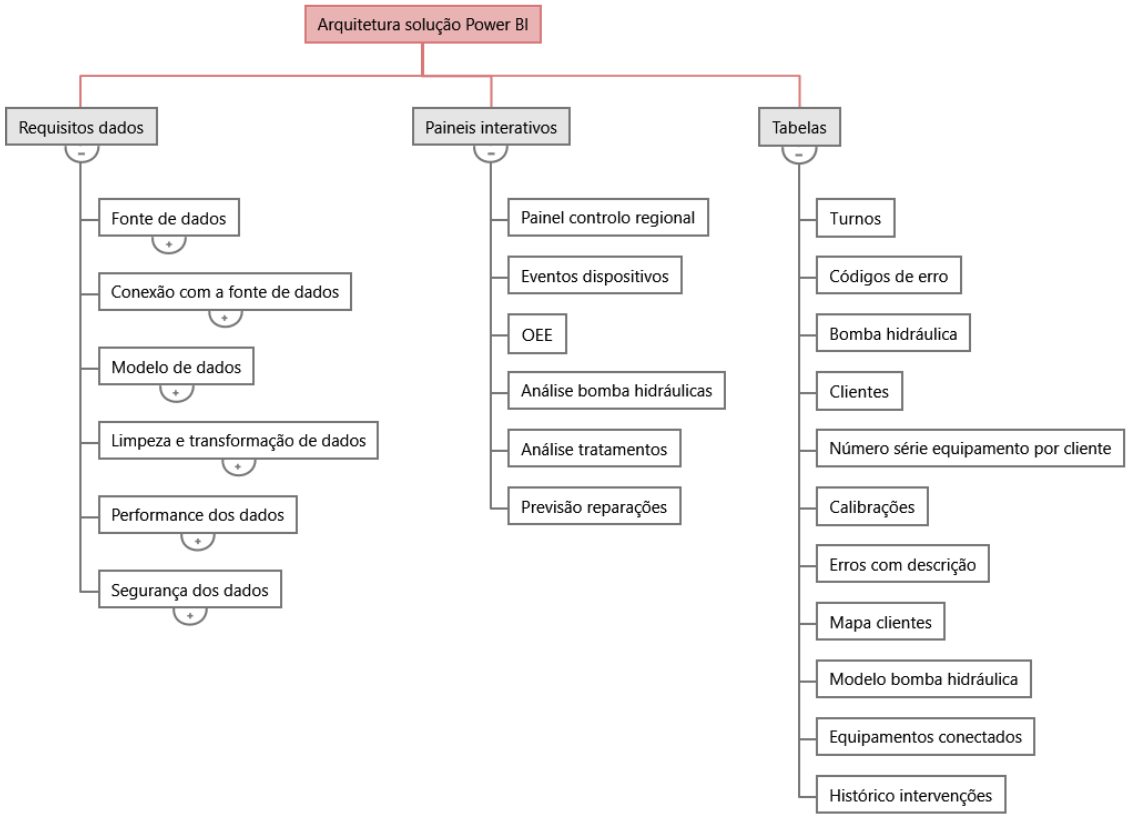


Figura 18 Mapa ideias arquitetura solução Power BI

Esta etapa, foi importante na medida que permitiu uma melhor interpretação dos dados, identificação das informações recolhidas e a forma como elas são apresentadas.

4.2.1. CONEXÃO E MODELAÇÃO DE DADOS

Esta etapa teve como objetivo identificar e selecionar os dados relevantes para a análise, o que implicou a criação de uma conexão com as fontes de dados disponíveis e a seleção dos dados necessários para responder aos objetivos da análise.

O *Power BI* enquanto ferramenta, permite a conectividade com diversas fontes de dados, como base de dados relacionais (SQL *Server, Oracle,* MySQL), arquivos CSV, Excel, *SharePoint, API's* e serviços online, entre outros.

No caso da solução desenvolvida, foi estabelecida uma conexão a uma fonte de dados no formato Excel, tendo-se posteriormente procedido à importação e armazenamento destes dados num modelo de dados do *Power BI*.

Seguidamente, procedeu-se à exploração dos dados disponíveis de forma a entender a sua estrutura, os campos, os formatos, tipos de dados, etc., tendo contribuído para a identificação dos dados relevantes a serem utilizados nas análises e no desenvolvimento dos painéis interativos.

Por fim, antes de avançar para o desenho dos painéis interativos, foi realizada uma verificação da integridade dos dados, de modo a garantir a sua qualidade. Foram realizados testes de consistência, como a validação de tipos de dados, o que incluiu a deteção e consequente limpeza dos valores inválidos, concatenação de algumas variáveis e a criação de campos calculados. Quanto à relevância dos dados, foi avaliada comparando as necessidades das análises com as informações disponíveis, de modo a garantir que estas contribuam diretamente para os resultados esperados.

4.2.2. PROCESSO DE DESIGN E LAYOUT

Para esta fase de projeção dos painéis interativos no *Power BI*, foi adotada uma abordagem que promovesse a clareza, interatividade e usabilidade dos painéis. Desta forma, antes de começar a construção dos painéis, foram identificados os requisitos da análise bem como identificado o público-alvo.

Esta abordagem, permitiu determinar a relevância da informação e quais as visualizações e interatividade mais adequadas para satisfazer as necessidades dos utilizadores.

Em relação à disposição e arquitetura da aplicação, foi desenvolvida uma página inicial, páginas de navegação e páginas de conteúdo especializado, tendo em consideração atributos como a facilidade de uso e a coerência, a fim de simplificar a navegação e a compreensão dos dados. Na figura 19 pode-se observar o ecrã inicial da aplicação.

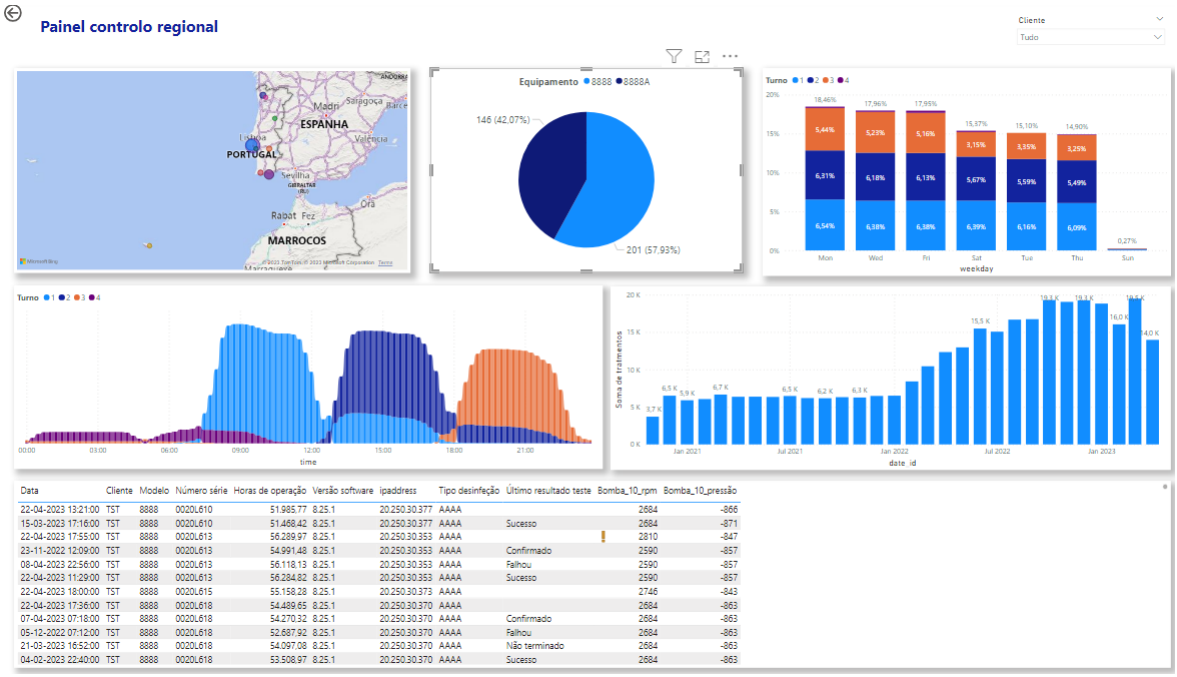


Figura 19 Página inicial aplicação desenvolvida

No que diz respeito à seleção das visualizações adequadas, foram utilizados gráficos de barras, gráficos de linhas, mapas, tabelas, entre outros, tendo como objetivo a satisfação das necessidades dos utilizadores.

Relativamente ao *layout* e à organização dos elementos visuais, a disposição foi planeada de forma tornar a informação de fácil compreensão. Os gráficos e tabelas foram organizados de maneira lógica e clara, de forma a permitir que os utilizadores identifiquem rapidamente os dados relevantes.

Quanto à utilização de elementos interativos, foram utilizados filtros, em diferentes versões, tais como: *slicers*, *drill-through*, entre outros, permitindo aos utilizadores explorar os dados de maneira interativa, filtrar informações relevantes, aprofundar em detalhe ou alternar entre diferentes perspetivas.

Por último, no que diz respeito aos elementos visuais utilizados no design de painéis, foi tido em consideração a adequação das cores, a formatação e estética de modo a tornar os painéis atraentes e de fácil compreensão.

4.3. PAINÉIS INTERATIVOS

Este capítulo tem como objetivo identificar e descrever as informações apresentadas nos painéis interativos desenvolvidos, bem como a eficácia e a sua utilidade para a solução desenvolvida.

É importante salientar que no decorrer da descrição dos painéis interativos desenvolvidos, surgiu uma dificuldade relacionada com a integração das fotos dos painéis neste capítulo. Dada a quantidade de informações apresentadas em cada painel e o tamanho da imagem, tornou-se difícil visualizar adequadamente todos os detalhes relevantes.

Desta forma, optou-se por disponibilizar uma imagem completa de cada painel na secção de anexo do presente documento, onde será possível analisar com maior detalhe todos os pormenores e informações relevantes de cada painel.

Na ferramenta levada a cabo, foram criados os seguintes painéis interativos:

1. Painel controlo regional;
2. Eventos dispositivos;
3. OEE-Eficácia Global Equipamentos;
4. Análise bombas hidráulicas;
5. Análise tratamentos;
6. Previsão reparações.

O Painel de controlo regional, cujo pormenor da estrutura se apresenta no anexo B, foi desenvolvido com objetivo de disponibilizar aos utilizadores a seguinte informação:

a) Localização geográfica dos clientes;

b) Quantificação de equipamentos por cliente e por modelo;

c) Distribuição de tratamentos por dia da semana e por turno;

d) Quantificação de tratamentos por turno;

e) Quantificação de tratamentos por mês;

f) Tabela com parâmetros relevantes sobre o desempenho dos equipamentos.

Na disposição habitual da informação nos painéis interativos, foi adotada a estratégia de posicionar a componente gráfica na parte superior da página, enquanto a informação apresentada sob a forma de tabela foi disposta na parte inferior da página. Adicionou-se um filtro que permite filtrar a informação por cliente, permitindo assim criar interatividade com os utilizadores.

Para a informação da "localização geográfica dos clientes", foi utilizado o componente Visual Mapa, permitindo aos utilizadores consultar a localização geográfica e quantidade de equipamentos por cliente. Esta informação combinada, pode ser extremamente útil para a gestão da manutenção, por exemplo, para o planeamento e alocação de recursos, priorização de solicitações por parte dos clientes ou análise de desempenho regional.

Relativamente à "quantificação de equipamentos por cliente e por modelo", foi utilizado um gráfico circular para disponibilizar esta informação, podendo esta ser aproveitada, por exemplo, para o planeamento estratégico de peças de reposição.

No que diz respeito à análise dos tratamentos, foram considerados três elementos gráficos que permitem analisar a distribuição dos tratamentos em diferentes perspetivas, sejam ela de carácter diário, semanal ou mensal.

Esta informação pode ser de extrema relevância, na medida que permite à gestão da manutenção, por exemplo, identificar padrões e tendências de utilização dos equipamentos, sejam eles diários, semanais ou mensais. Durante os períodos de maior utilização, a gestão poderá adaptar a quantidade de recursos humanos necessários de modo a garantir uma adequada cobertura.

Por fim, a tabela com a informação, data, cliente, modelo equipamento, número de série, horas de operação, software, endereço IP, tipo de desinfeção, resultado do teste, pressão

das bombas e rotação das bombas, possui várias informações úteis para a gestão e equipas de manutenção.

Por exemplo, a tabela fornece dados relevantes para a análise de falhas e a resolução de problemas, na medida que o resultado do teste pode indiciar falhas ou problemas de funcionamento nos equipamentos. Os parâmetros pressão das bombas e r.p.m., através dos valores registados e alertas, dão a informação ao utilizador da necessidade de explorar e aprofundar o desempenho deste componente no painel interativo de análise de bombas hidráulicas.

Concluindo, a informação disponibilizada neste painel de controlo regional, é de extrema relevância para a gestão e equipas de manutenção, na medida que permite uma melhor caracterização do perfil do cliente. Estas informações são essenciais para uma gestão eficiente, tomada de decisões informadas e otimização dos recursos de manutenção.

O painel “Eventos Dispositivos”, cuja visão geral da estrutura se apresenta no anexo C, foi desenvolvido com objetivo de disponibilizar aos utilizadores a seguinte informação:

a) Quantidade de eventos por equipamento;

b) Quantidade de eventos por estado;

c) Quantidade de eventos por código;

d) Quantidade de eventos por CPU;

e) Quantidade e distribuição de mensagens;

f) Tabela com informação detalhada da memória de eventos.

Para o desenvolvimento deste painel, foram considerados vários elementos filtro, de modo a proporcionar aos utilizadores a capacidade de personalizar a visualização e a análise dos dados, permitindo uma exploração interativa, análise detalhada e comparação de dados.

Os filtros utilizados foram:

a) Cliente, facilita a análise do desempenho dos equipamentos na perspetiva do cliente;

b) Turno, permite a análise de dados com base nos diferentes turnos de trabalho. Pode ajudar a identificar tendências, padrões ou problemas específicos que possam ocorrer em determinados turnos, bem como no auxílio do planeamento de recursos e alocação das equipas;

c) Número de série, permite realizar a análise de dados específicos de um dado equipamento. Esta análise permite rastrear o desempenho, histórico de eventos bem como qualquer problema recorrente num dado equipamento;

d) Modelo, permite realizar a análise de dados com base nos diferentes modelos de equipamentos. Esta informação pode ser útil para comparar o desempenho e a confiabilidade dos diferentes modelos, identificar problemas recorrentes em determinados modelos e tomar decisões com base nessa informação;

e) Criticidade, permite priorizar e analisar os equipamentos com base em diferentes graus de severidade dos eventos;

f) Tipologia mensagem, permite analisar dados com base na origem dos eventos, estando disponíveis eventos que são resultantes da operação do equipamento e eventos relativos ao equipamento;

g) Período de análise, ajusta o intervalo de tempo dos dados a serem exibidos no painel. Esta informação pode ser útil para visualizar tendências ao longo do tempo, identificar sazonalidades ou analisar o desempenho em períodos específicos.

Para visualização dos dados “Soma de eventos por equipamento”, optou-se por utilizar um gráfico de colunas horizontais. Este gráfico tem uma utilidade significativa, na medida que permite identificar os equipamentos com maior incidência de eventos, o que pode indiciar equipamentos com maior probabilidade de falhas ou problemas recorrentes.

No que diz respeito à “Soma de eventos por estado”, recorreu-se igualmente a um gráfico de colunas horizontais. Este gráfico é de especial relevância, dado que permite correlacionar de forma clara o estado de operação dos equipamentos com a incidência de eventos.

Em relação à visualização da “Soma de eventos por código”, foi adotado o mesmo modelo gráfico. Este permite visualizar de forma clara e comparativa a incidência de cada código

de erro do equipamento, contribuindo para a identificação dos erros mais frequentes e/ou problemas recorrentes que afetam o desempenho do equipamento.

O gráfico dos dados “Soma de eventos por CPU”, permite visualizar de forma clara e comparativa a incidência de eventos por CPU. Esta informação, pode contribuir para a identificação dos CPU’s que apresentam maior incidência de erros e que possam afetar o desempenho dos equipamentos.

O gráfico dos dados “Contagem de mensagens”, disponibiliza a incidência de cada mensagem emitida pelo equipamento, contribuindo para a identificação das mensagens mais frequentes, possíveis problemas recorrentes ou situações que possam carecer de especial atenção.

Por último, a tabela de dados apresentada neste painel, com a informação por equipamento, data, cliente, modelo equipamento, número de série, horas de operação, criticidade, CPU, estado de operação, código de erro, descrição, causa, correção do erro e probabilidade de falha, apresenta dados relevantes para a identificação de falhas recorrentes, análise de criticidade, priorização de intervenções, identificação de padrões e tendências e probabilidade de falha.

O painel seguinte, OEE-Eficácia Global Equipamentos, cuja visão geral da estrutura se apresenta no anexo D, foi desenvolvido com objetivo de disponibilizar aos utilizadores a seguinte informação:

a) Média do OEE por número de série;

b) Disponibilidade por ano e mês;

c) Velocidade por ano e mês;

d) Qualidade por ano e mês;

e) Tabela de dados OEE por número de série, ano e mês;

f) Distribuição geográfica do valor do indicador OEE por cliente.

O desempenho global de um equipamento, é resultante da influência de vários fatores com origem interna e externa. O indicador OEE, é um dos principais indicadores utilizado na manutenção que permite avaliar esse desempenho. Este painel tem como objetivo disponibilizar informações essenciais para realizar, avaliar e identificar oportunidades de melhoria.

Foram considerados dois elementos filtro, de modo a proporcionar aos utilizadores a capacidade de personalizar a visualização e a análise dos dados, permitindo uma exploração interativa, análise detalhada e comparação de cenários.

Filtros utilizados:

a) Cliente, análise do desempenho dos equipamentos na perspetiva do cliente;

b) Período de análise, ajuste do intervalo de tempo dos dados a serem exibidos no painel. Esta informação é útil para visualizar tendências ao longo do tempo, identificar sazonalidades ou analisar o desempenho em períodos específicos.

Para a visualização do “Valor médio do OEE por número de série”, optou-se por utilizar um gráfico de colunas horizontais. Este tipo de visualização é especialmente adequado para exibir a comparação entre diferentes categorias ou grupos, permitindo uma análise visual rápida e clara.

Este gráfico é de grande relevância para a análise, dado que disponibiliza uma visão clara e comparativa do desempenho dos equipamentos, contribuindo para identificar equipamentos com desempenho abaixo da média.

No que diz respeito à visualização da componente da “disponibilidade”, “velocidade” e “qualidade”, decidiu-se utilizar gráficos de linhas. Este tipo de visualização é especialmente adequado, quando se deseja analisar tendências, padrões ou variações de uma variável ao longo de um período.

Estes gráficos têm uma utilidade significativa, na medida que permitem, por exemplo, relativamente à disponibilidade identificar períodos de baixa disponibilidade, o que pode indiciar problemas de manutenção ou falhas recorrentes.

O gráfico da velocidade, permite à gestão e equipas de manutenção identificar padrões de utilização dos equipamentos, variações sazonais ou variações na taxa de ocupação dos equipamentos. Estas informações podem ser úteis para otimizar a alocação de recursos, planear intervenções de carácter preventivo, identificar constrangimentos na capacidade ou identificar a necessidade de investimentos em equipamentos adicionais.

O gráfico da qualidade, tem como objetivo avaliar o desempenho e a eficiência do equipamento durante o seu funcionamento. Com esta informação, é possível identificar períodos de desempenho abaixo do esperado, avaliar o impacto de eventuais ajustes ou melhorias na operação dos equipamentos, bem como identificar necessidades de formação.

Por último, a tabela de dados presente neste painel, com a informação do valor do OEE, disponibilidade, velocidade e qualidade, por cliente, número de série equipamento, ano mês, permite aos utilizadores analisar a mesma informação contida nos gráficos de forma consolidada e sob a forma de tabela.

O próximo painel, “análise de bombas hidráulicas”, cuja visão geral da estrutura se apresenta no anexo E, foi desenvolvido com o objetivo de disponibilizar aos utilizadores a seguinte informação:

a) R.p.m. Bomba P10;

b) Pressão bomba P10;

c) Impulsos/min Bomba P11;

d) Impulsos/min Bomba P12;

e) Impulsos/min Bomba P13;

f) Tabela com informação por cliente e equipamento da variação de r.p.m. e pressão;

g) Tabela com informação por cliente e equipamento, valores variáveis das bombas com informação de gestão visual de avisos.

Foram disponibilizados vários filtros, de modo a proporcionar aos utilizadores a capacidade de personalizar a visualização e a análise dos dados, permitindo uma exploração interativa, análise detalhada e comparação de cenários.

Filtros utilizados:

a) Cliente, permite realizar a análise do desempenho dos equipamentos na perspetiva do cliente;

b) Número de série, permite realizar a análise de dados específicos de um dado equipamento. Esta análise permite rastrear o desempenho individual de um dado equipamento;

c) Modelo, permite realizar a análise de dados com base nos diferentes modelos de equipamentos. Esta informação é importe para comparar o desempenho e a confiabilidade de diferentes modelos, identificar problemas comuns em determinados modelos e tomar decisões de manutenção com base nessa informação;

d) Tipo de bomba, permite realizar a análise de dados com base nos diferentes tipos de bombas hidráulicas utilizadas nos equipamentos. Permitem analisar o desempenho de cada tipo de bomba, identificar quais tipos que são mais eficientes ou confiáveis, e tomar decisões em relação à seleção e manutenção das bombas hidráulicas;

e) R.p.m. bomba P10, permite realizar a análise de dados com base na rotação específica desta bomba. Esta informação é de grande utilidade para a gestão e equipas de manutenção, dado contribuir para analisar o impacto da rotação da bomba no desempenho global do equipamento, identificar os limites ideais de rotação para obter os melhores resultados e eventualmente ajustar a rotação;

f) Período de análise, ajusta o intervalo de tempo dos dados a serem exibidos no painel. Esta informação permite visualizar tendências ao longo do tempo, identificar sazonalidades ou analisar o desempenho em períodos específicos.

Para a visualização dos dados sob a forma gráfica, optou-se pela utilização de gráficos de linhas horizontais, incluindo os limites de operação e linha de tendência. Os gráficos

construídos, permitem acompanhar ao longo do tempo o desempenho das bombas nas diferentes perspetivas equacionados, nomeadamente, r.p.m., pressão e impulsos/min.

Os limites de operação incluídos nos gráficos, permitem identificar de forma rápida se os valores se encontram dentro dos parâmetros estabelecidos. Por exemplo, se os valores da r.p.m. estiverem consistentemente dentro dos limites esperados, indica que a bomba se encontra a funcionar conforme esperado. Caso ocorram desvios frequentes, poderá ser um indício de que podem existir problemas neste componente.

Por outro lado, a linha de tendência mostra a direção geral dos dados de variação ao longo do tempo, o que permite identificar se há um aumento, diminuição ou estabilidade na variável a controlar, contribuindo para a compreensão do desempenho do componente.

A combinação dos limites de operação com a linha de tendência, dependendo do componente, poderá permitir identificar desvios significativos. Se houver pontos que ultrapassem os limites de operação ou se a linha de tendência indicar uma variação abrupta, isso pode indiciar problemas de operação ou falha no componente.

Esta informação é de especial relevância, dado que permite que as equipas de manutenção atuem de forma proativa, planeando intervenções preventivas antes da falha do componente.

A tabela com a informação do cliente, modelo, número de série, variação r.pm. e variação pressão, permite que as equipas de manutenção identifiquem de uma forma rápida padrões de variação das r.p.m. e pressão das bombas hidráulicas ao longo do tempo. Esta informação, permite agilizar a análise dos dados e as consequentes etapas do planeamento das intervenções.

Por último, a tabela com a informação do cliente, data, número de série, modelo, r.p.m. P10, Pressão P10, Pressão MAX P10, P11 impulsos/min, P12 impulsos/min e P13 impulsos/min, permite analisar a mesma informação contida nos gráficos de forma consolidada e sob a forma de tabela.

O painel análise de tratamentos, cuja visão geral da estrutura se apresenta no anexo F, foi desenvolvido com objetivo de disponibilizar aos utilizadores a seguinte informação:

a) Contagem de tratamentos por dia;

b) Contagem de tratamentos por ano e mês;

c) Contagem de tratamentos por turno;

d) Contagem de tratamentos por dia da semana.

Este painel, apenas disponibiliza informação sob a forma gráfica. Foram considerados vários elementos filtro, de modo a proporcionar aos utilizadores a capacidade de personalizar a visualização e a análise dos dados, permitindo uma exploração interativa, análise detalhada e comparação de cenários.

Filtros utilizados:

a) Cliente, permite realizar a análise do desempenho dos equipamentos na perspetiva do cliente;

b) Turno, permite realizar a análise de dados com base nos diferentes turnos de trabalho. Poderá ajudar na identificação de tendências e padrões, assim como no planeamento de recursos e alocação de equipas;

c) Modelo, viabiliza a análise de dados com base nos diferentes modelos de equipamentos. Esta informação é útil para comparar a taxa de utilização dos diferentes modelos;

d) Número de série, permite realizar a análise dos dados específicos de um dado equipamento. Esta análise permite rastrear o desempenho individual de um dado equipamento;

e) Período de análise, ajusta o intervalo de tempo dos dados a serem exibidos no painel. Esta informação é útil para visualizar tendências ao longo do tempo, identificar sazonalidades ou analisar o desempenho em períodos específicos.

Para a visualização dos dados da quantificação de tratamentos por ano e mês, assim como para a visualização dos dados da quantificação de tratamentos por dia e de dia da semana, optou-se por utilizar gráficos de colunas verticais.

Esta informação é de especial relevância, dado permitir que as equipas de manutenção, identifiquem os períodos de maior e menor utilização dos equipamentos, contribuindo para um adequado planeamento de recursos.

Relativamente à visualização da informação da contagem de tratamentos por turno, optou-se por um gráfico circular. Esta informação pode ser útil para planear e antecipar as necessidades de recursos, como mão de obra, equipamentos ou materiais, de modo a garantir a adequada disponibilidade da equipa de manutenção para satisfazer as necessidades específicas de cada turno.

Por último, o painel previsão de reparações, cuja visão geral da estrutura se apresenta no anexo D, foi desenvolvido com objetivo de disponibilizar aos utilizadores a seguinte informação:

a) Probabilidade de falha por equipamento;

b) Probabilidade de falha do componente.

Para o desenvolvimento deste painel, considerou-se duas tabelas e um elemento visual Mapa, com a informação por equipamento da probabilidade de falha, do componente e localização geográfica.

Esta informação combinada, pode ser extremamente útil para a manutenção, por exemplo, para o planeamento e alocação de recursos, priorização de atendimento ou análise de desempenho por regional.

Foram ponderados os seguintes elementos filtro, de modo a proporcionar aos utilizadores a capacidade de personalizar a visualização e a análise dos dados.

Filtros utilizados:

a) Cliente, permite realizar a análise do desempenho dos equipamentos na perspetiva do cliente;

b) Modelo, viabiliza a análise de dados com base nos diferentes modelos de equipamentos. Esta informação é útil para a comparação do desempenho dos diferentes modelos;

c) Número de série, permite realizar a análise dos dados específicos de um dado equipamento. Esta análise permite rastrear o desempenho individual de um equipamento.

A informação contida neste painel, é de especial relevância, dado que permite às equipas de manutenção ter acesso a informações preditivas, contribuindo para planear antecipadamente e de forma mais eficiente as atividades de manutenção, evitando interrupções inesperadas na operação dos equipamentos e consequentemente mitigar o tempo de inatividade não planeado.

Outra possível otimização, é contribuir para uma gestão mais eficiente das peças de reposição, evitando a rutura e/ou o excesso de peças.

4.4. CONCLUSÃO

Este capítulo teve como objetivo descrever cada uma das etapas do desenvolvimento da solução implementada. Descreveu-se em detalhe a base de dados utilizada no estudo, incluindo a sua estrutura, fonte de dados e características relevantes. Foram identificados os conjuntos de dados necessários para a análise preditiva e visualização dos dados.

Seguidamente, foi feita uma reflexão sobre as restrições associadas aos dados disponíveis, como a falta de informação completa sobre algumas falhas, a representatividade dos dados, a complexidade e multidimensionalidade dos dados.

Posteriormente, descreveu-se a conexão dos dados à aplicação *Power BI* e a modelagem dos mesmos, incluindo transformações e relacionamentos entre tabelas.

Abordou-se o processo criativo dos painéis interativos desenvolvidos, onde foram exploradas as funcionalidades e recursos disponíveis no *Power BI* para criar visualizações interativas e painéis personalizados.

Por fim, descreveu-se as informações apresentadas nos painéis interativos desenvolvidos, bem como a eficácia e a sua utilidade para as diferentes partes interessadas.

5. ANÁLISE DE RESULTADOS

Com este capítulo, pretende-se descrever o desenvolvimento, testes e análise de resultados obtidos para os modelos preditivos utilizados. Serão apresentados e discutidos os resultados obtidos a partir da aplicação de algoritmos de aprendizagem automática, considerando métricas de avaliação como a exatidão, precisão, *recall, f1-score*, entre outras.

Foram utilizados conjunto de dados para treino, teste e validação dos modelos, aos quais foram aplicados os diferentes algoritmos de aprendizagem de máquina descritos no capítulo 4, dando origem a resultados referentes aos modelos preditivos estudados.

Os resultados obtidos foram analisados em relação ao seu desempenho preditivo, tendo -se efetuado a comparação das previsões e dos modelos com os dados reais das falhas nos equipamentos e respetivos componentes. Foram utilizados gráficos e tabelas para demonstrar a precisão e a consistência dos modelos.

Por último, foram discutidas as limitações dos modelos e possíveis fontes de erro, assim como a estratégia para mitigar essas limitações. Na secção de anexos, pode ser consultado o código dos algoritmos de previsão desenvolvidos para cada um dos modelos propostos.

5.1. AVALIAÇÃO E ESCOLHA DO MODELO

Este capítulo tem como objetivo, apresentar uma análise detalhada sobre a avaliação e escolha dos modelos preditivos utilizados para o desenvolvimento da solução proposta, com foco na sua relevância para a gestão e equipas de manutenção. O desafio que se coloca, passa por prever a falha dos equipamentos e componentes, tendo como objetivo a otimização do planeamento e a eficiência das atividades de manutenção.

Os algoritmos previamente identificados na secção 4, foram aplicados a dois conjuntos de dados de teste, os quais incluem as variáveis necessárias para o desenvolvimento dos modelos. O objetivo principal desta etapa, consistiu em avaliar o desempenho dos modelos e realizar uma comparação entre eles, tendo como finalidade escolher o modelo mais adequado.

Para a realização desta avaliação, os conjuntos de dados foram divididos numa proporção de 80 % para o treino dos modelos e 20 % para o teste dos dados.

Uma das formas mais utilizadas para representar os resultados da avaliação de uma classificação os dados, é recorrer à utilização de matrizes de confusão. Estas correlacionam os valores reais com os valores previstos, sendo a saída da matriz resultante, uma matriz dois por dois, onde as linhas correspondem às classes verdadeiras e as colunas correspondem às classes previstas.

Na figura 20, pode ser observado um exemplo de uma matriz de confusão, sendo esta composta por quatro elementos: verdadeiro positivo (VP), falso positivo (FP), verdadeiro negativo (VN) e falso negativo (FN).

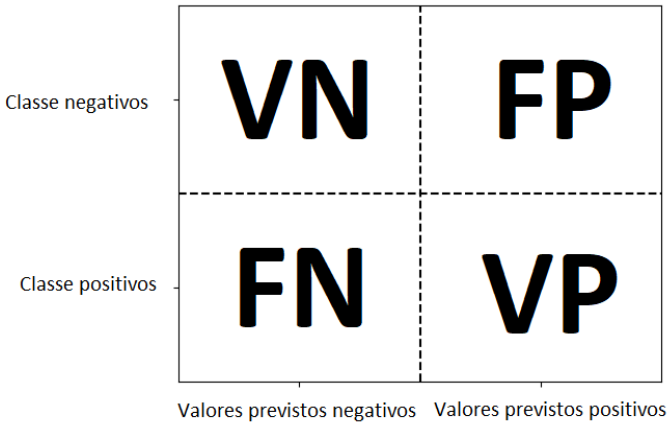


Figura 20 Matriz de confusão para classificação binária

O elemento verdadeiro positivo (VP), representa a quantidade de observações corretamente classificadas como positivas pelo modelo. O falso positivo (FP), indica as observações incorretamente classificadas como positivas pelo modelo. O verdadeiro negativo (VN), representa as observações corretamente classificadas como negativas. Por fim, o falso negativo (FN) indica as observações incorretamente classificadas como negativas

A partir destes quatro elementos, VN, FN, VP e FP, diversas métricas podem ser calculadas para avaliar o desempenho do modelo. Algumas das métricas habitualmente utilizadas são:

- Exatidão;
- Precisão;
- *Recall;*
- Curvas *precison-recall;*
- Curvas ROC;
- AUC.

Seguidamente, na figura 21, são apresentadas as matrizes de confusão resultantes da aplicação dos diferentes algoritmos ao conjunto de dados de teste para estimar o estado de falha dos equipamentos:

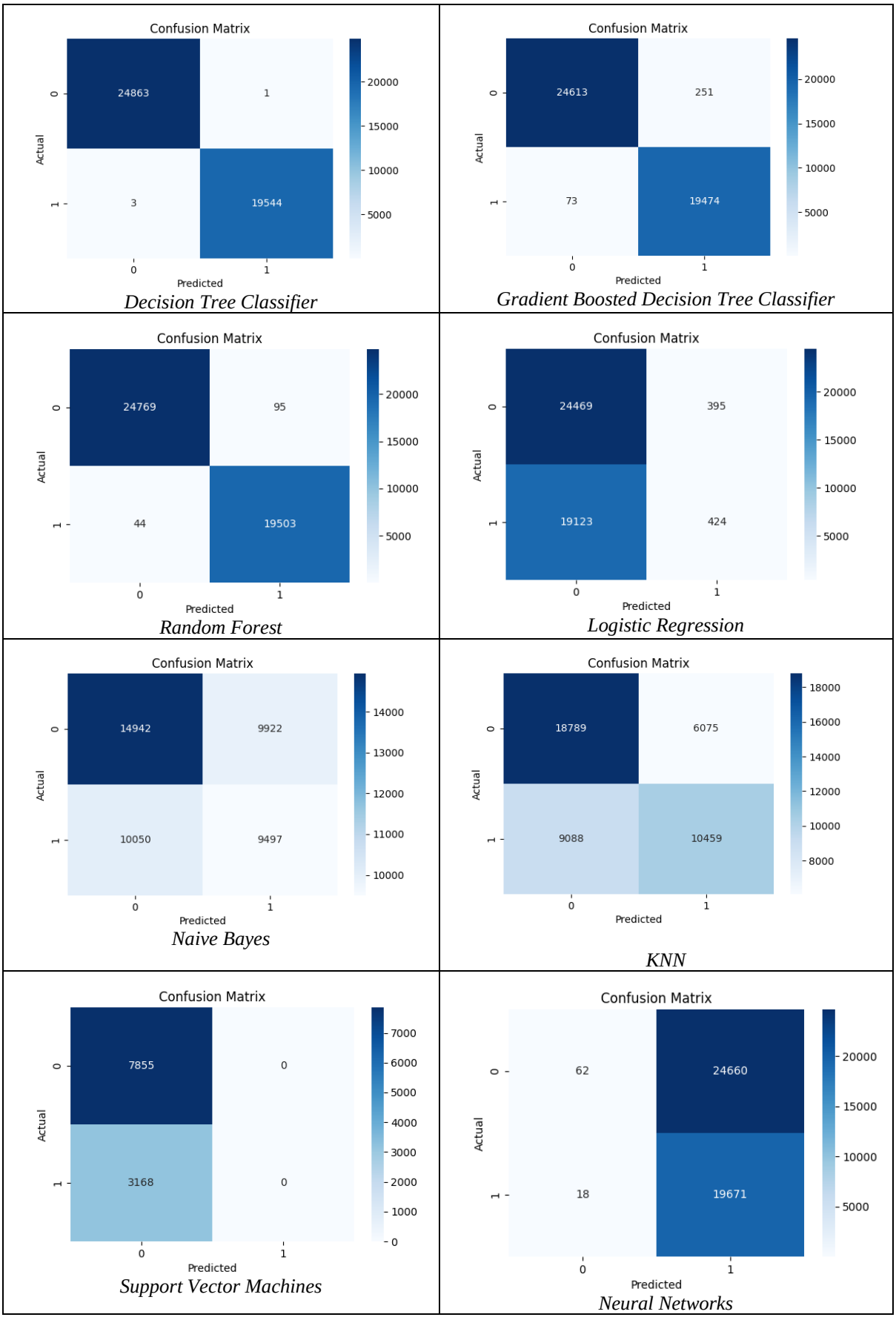


Figura 21 Matrizes de confusão para os algoritmos testados

Relativamente à a exatidão, que representa a proporção de classificações corretas em relação ao total de observações, calculada utilizando a seguinte equação.

$$Exatidão = \frac{VP + VN}{VP + FP + FN + VN} \tag{5}$$

Após a aplicação dos algoritmos ao conjunto de dados de teste, obtiveram-se os seguintes resultados de exatidão, conforme pode ser verificado na tabela 1.

Tabela 1 – Resultados da exatidão para o conjunto de dados proposto

Algoritmo	Exatidão (%)
	Conjuntos de dados do estado de falha dos equipamentos
Random Forest Classifier	100
Decision Tree Classifier	100
Gradient Boosting Decision Tree Classifier	99
Support Vector Machines	71
KNN	66
Logistic Regression	56
Naive Bayes	55
Neural Networks	44

Destacam-se os algoritmos RF, *Decision Tree Classifier e Gradient Boosting Decision Tree Classifier*, dado terem obtido a melhor exatidão, com 100 %, 100 % e 99 % respetivamente, demonstrando um desempenho superior em relação aos demais algoritmos, o que significa que estes foram capazes de realizar previsões mais precisas.

Por outro lado, os algoritmos *Support Vector Machine e KNN*, apresentaram exatidões de 71 % e 66 %, respetivamente. Embora estes valores possam ser considerados satisfatórios, indicam um desempenho relativamente inferior quando comparados com os algoritmos que obtiveram melhor desempenho.

A precisão, é outra métrica utilizada e que mede a proporção de verdadeiros positivos em relação ao total de observações classificadas como positivas, calculada através da seguinte equação:

$$Precisão = \frac{VP}{VP + FP} \tag{6}$$

Foram obtidos os seguintes resultados de precisão, após aplicação dos algoritmos ao conjunto de dado de teste:

Tabela 2 – Resultados da precisão para o conjunto de dados proposto

Algoritmo	Precisão (%)
	Conjuntos de dados de estado de falha dos equipamentos
Random Forest Classifier	100
Decision Tree Classifier	100
Gradient Boosting Decision Tree Classifier	100
Support Vector Machines	69
KNN	62
Logistic Regression	45
Naive Bayes	51
Neural Networks	44

Em termos de precisão, os algoritmos RF, o DT e o GBDT obtiveram igualmente os melhores resultados, com 100 % cada um, evidenciando um desempenho superior em relação aos demais algoritmos, o que significa que eles foram capazes de fazer previsões mais precisas.

Em contrapartida, os algoritmos KNN e SVM, obtiveram uma precisão de 69 % e 62 %, respetivamente. Quando comparados com os algoritmos que obtiveram melhor desempenho em termos de precisão, os valores embora satisfatórios, são relativamente inferiores.

O *recall*, é outra métrica utilizada e que mede quantas das amostras positivas são recolhidas pelas previsões positivas, obtendo-se através da seguinte equação:

$$Recall = \frac{VP}{VP + FN} \tag{7}$$

Para o *recall*, foram obtidos os seguintes resultados após aplicar os algoritmos aos conjuntos de dados de teste:

Tabela 3 – Resultados do *recall* para o conjunto de dados proposto

Algoritmo	Recall (%)
	Conjuntos de dados do estado de falha dos equipamentos
Random Forest Classifier	100
Decision Tree Classifier	100
Gradient Boosting Decision Tree Classifier	100
Support Vector Machines	45
KNN	54
Logistic Regression	2
Naive Bayes	49
Neural Networks	100

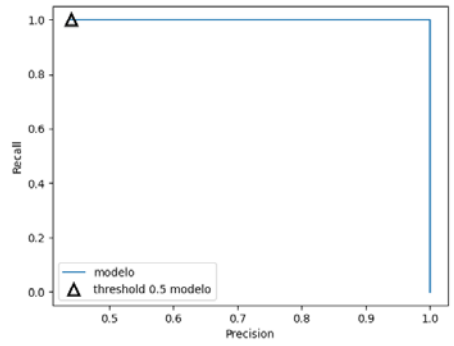
Em comparação com os demais algoritmos, o RF, DT, GBDT e *Neural Networks* alcançaram a melhor *recall*, com 100% em todos eles, evidenciando um desempenho superior.

Os Algoritmos SVM e KNN, apresentaram precisões de 45 % e 54 %, respetivamente, demonstrando um desempenho relativamente inferior quando comparados com os algoritmos que obtiveram melhor desempenho.

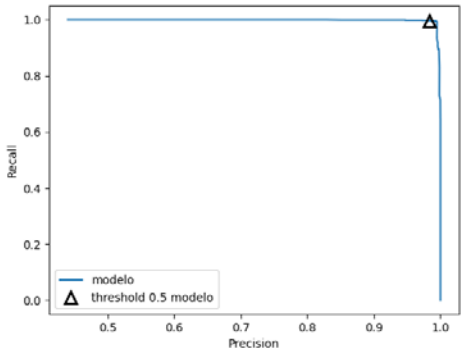
Por último, procedeu-se à avaliação dos resultados obtidos para as curvas de *precision-recall*, as curvas ROC (*Receiver Operating Characteristic*) e AUC (*Area under the curve*).

As curvas de *precision-recall* disponibilizam uma representação visual da relação entre a precisão e o *recall* para diferentes pontos de classificação. São úteis para avaliar o equilíbrio entre a capacidade do modelo em classificar corretamente os casos positivos (precisão) e a capacidade de identificar corretamente todos os casos positivos (*recall*).

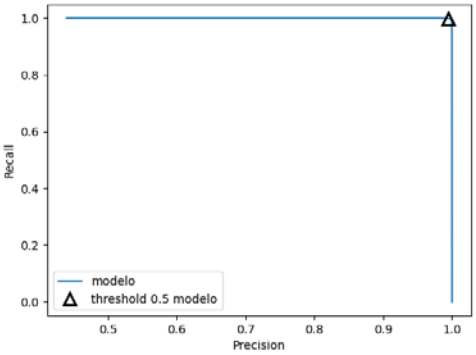
Na figura 22, pode-se observar o resultado das curvas *precision-recall* que foram obtidas para cada um dos algoritmos:



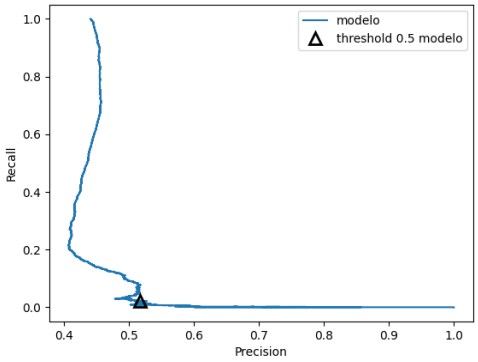
Decision Tree Classifier



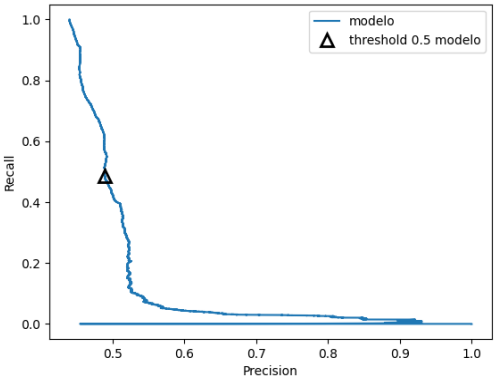
Gradient Boosted Decision Tree Classifier



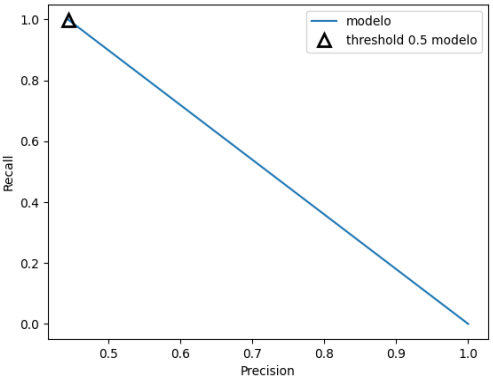
Random Forest



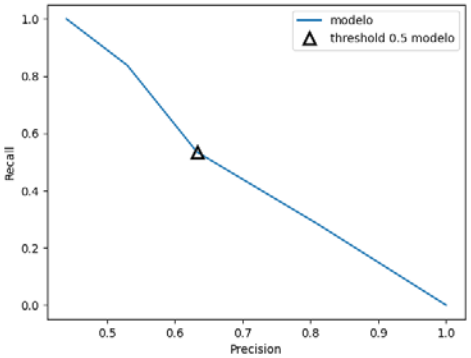
Logistic Regression



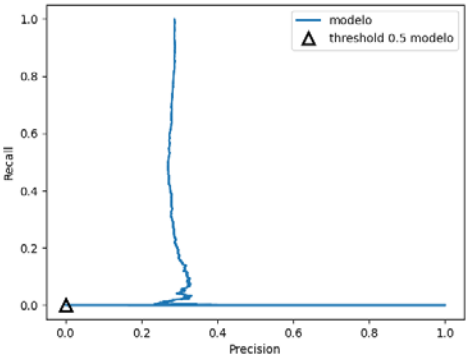
Naive Bayes



Neural networks



KNN



SVM

Figura 22 Curvas *precision-recall*

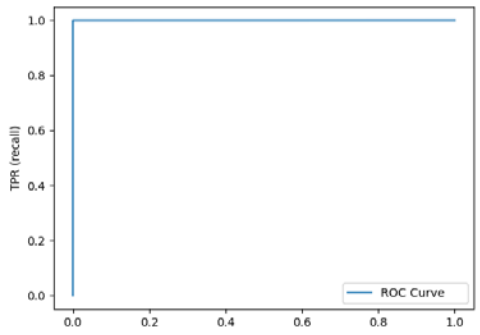
Da análise das curvas obtida, podemos observar que os algoritmos apresentaram diferentes desempenhos. Com os algoritmos RF, DT, GBDT obtiveram-se uma curva que se aproxima mais do canto superior direito, o que revela um equilíbrio superior entre a precisão e o *recall*. Os restantes algoritmos, obtiveram curvas que pela sua configuração evidenciam um desempenho inferior.

As curvas ROC, por outro lado, revelam a taxa de verdadeiros positivos em relação à taxa de falsos positivos para diferentes pontos de classificação. São importantes para avaliar o desempenho do modelo na distinção entre as classes positiva e negativa, considerando a taxa de classificações corretas e incorretas.

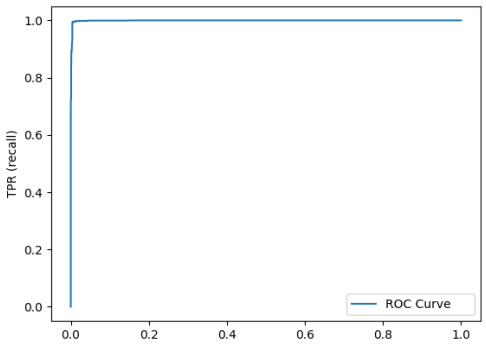
Na figura 23, pode ser observado as curvas ROC que foram obtidas para cada algoritmo.

Observa-se que os algoritmos RF, DT *e* GBDT obtiveram uma curva que se aproxima mais do canto superior esquerdo, o que indica uma maior taxa de verdadeiros positivos em relação aos falsos positivos. Os restantes algoritmos, obtiveram curvas que pela sua configuração evidenciam um desempenho inferior.

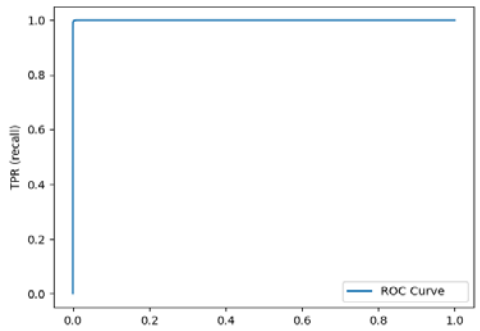
Paralelamente, foi calculada a AUC para cada modelo. A AUC é uma medida que quantifica a capacidade de um modelo distinguir entre diferentes classes, sendo uma representação numérica do desempenho das curvas ROC. Quanto maior for a AUC, melhor será o desempenho do modelo na distinção entre as classes.



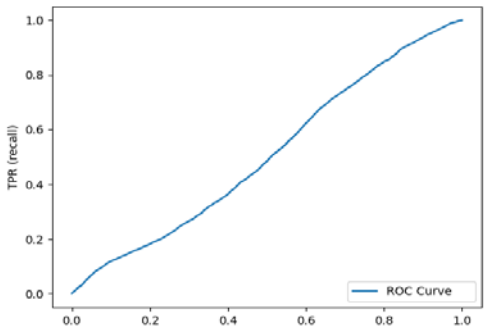
Decision Tree Classifier



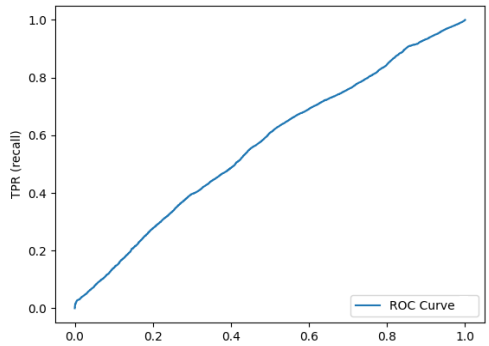
Gradient Boosted Decision Tree Classifier



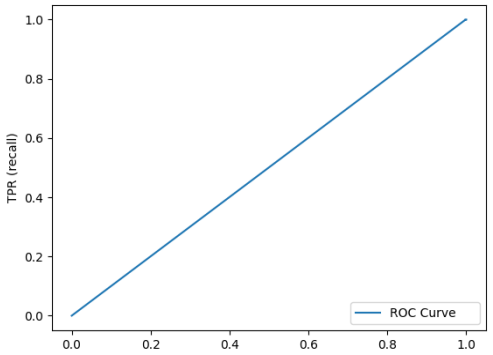
Random Forest Classifier



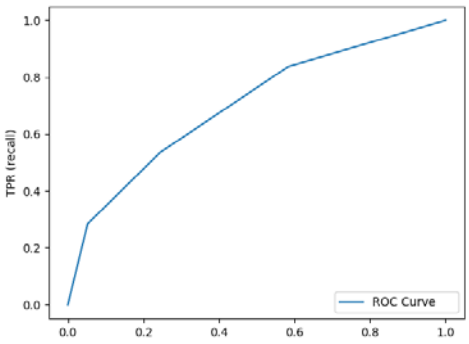
Logistic Regression



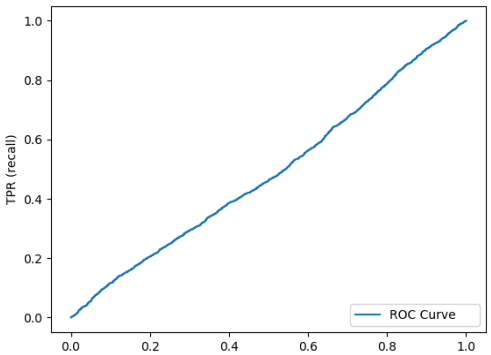
Naive Bayes



Neural Networks



KNN



SVM

Figura 23 Curvas ROC

Relativamente à AUC, apresenta-se na tabela 4 os resultados obtidos desta métrica para cada um dos algoritmos:

Tabela 4 – Resultados métrica AUC

Algoritmo	AUC (%)
	Conjuntos de dados do estado de falha dos equipamentos
Random Forest Classifier	**100**
Decision Tree Classifier	**100**
Gradient Boosting Decision Tree Classifier	**99**
Support Vector Machines	**70**
KNN	**56**
Logistic Regression	**51**
Naive Bayes	**50**
Neural Networks	**43**

Pode-se concluir que os resultados obtidos foram consistentes para as curvas ROC. Note-se que os algoritmos RF, DT e GBDT registaram um valor AUC mais elevado, seguido pelos algoritmos KNN e *Naive Bayes*. Os algoritmos *Logistic regression, Neural Networks e* SVM obtiveram valores AUC inferiores aos demais algoritmos.

Estes resultados, demonstram que os algoritmos RF, DT e GBDT alcançaram um desempenho mais satisfatório em relação às métricas analisadas, nomeadamente curvas de *precision-recall*, curvas ROC e valor AUC.

Considerando que os resultados obtidos nas métricas de exatidão e precisão apresentaram valores muito elevados, para os algoritmos RF, DT e GBDT, verificou-se a necessidade de analisar se os modelos não apresentam *overfitting*.

O *overfitting* é um fenómeno indesejado que ocorre quando um modelo apresenta um excelente desempenho com dados de treino, contudo tem um desempenho inferior relativamente a dados não utilizados, nomeadamente, os dados de validação ou teste.

Desta forma, foi realizada uma avaliação dos respetivos algoritmos, tendo-se comparado a evolução da exatidão com os dados de treino, teste e validação conforme pode ser observado nas figuras 24, 25 e 26.

Da análise gráfica dos algoritmos DT e GBDT, comprova-se que as três linhas (treino, validação e teste) estão muito próximas e sobrepostas em todo o gráfico, o que indica que o

modelo tem um comportamento consistente nos diferentes conjuntos de dados à medida que o número de estimadores aumenta, não ocorrendo *overfitting*.

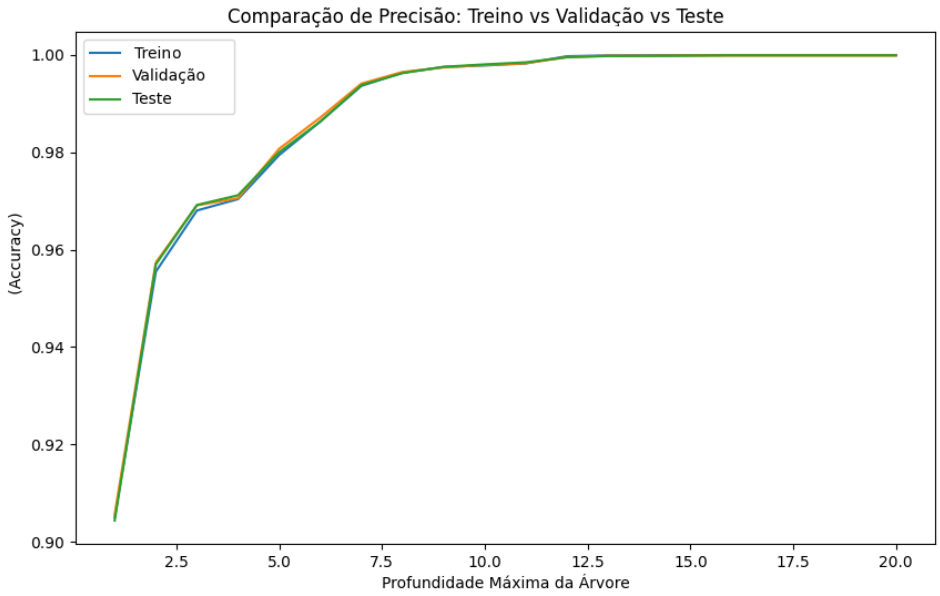


Figura 24 Comparação de 20 amostras desempenho treino, validação e teste *Decision Tree Classifier*

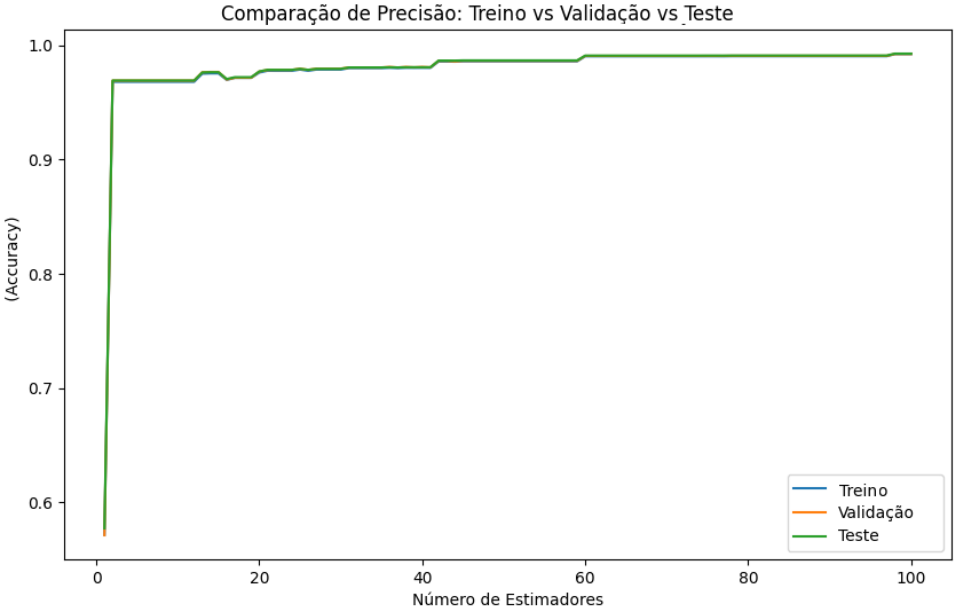


Figura 25 Comparação de 100 amostras desempenho treino, validação e teste *Gradient Boosting Decision Tree Classifier*

Por outro lado, no caso do gráfico do algoritmo RF, a linha do conjunto de treino demonstra uma precisão significativamente maior do que as linhas do conjunto de validação e teste, encontrando-se estas duas últimas relativamente próximas uma da outra.

Fica comprovado pela análise gráfica que ocorre *overfitting*, o que indica que o modelo se ajusta muito bem aos dados de treino, contudo tem um desempenho inferior relativamente a dados não utilizados, nomeadamente, os dados de validação ou teste. Desta forma, optou-se por excluir o algoritmo RF apesar do elevado desempenho em todas as métricas de avaliação.

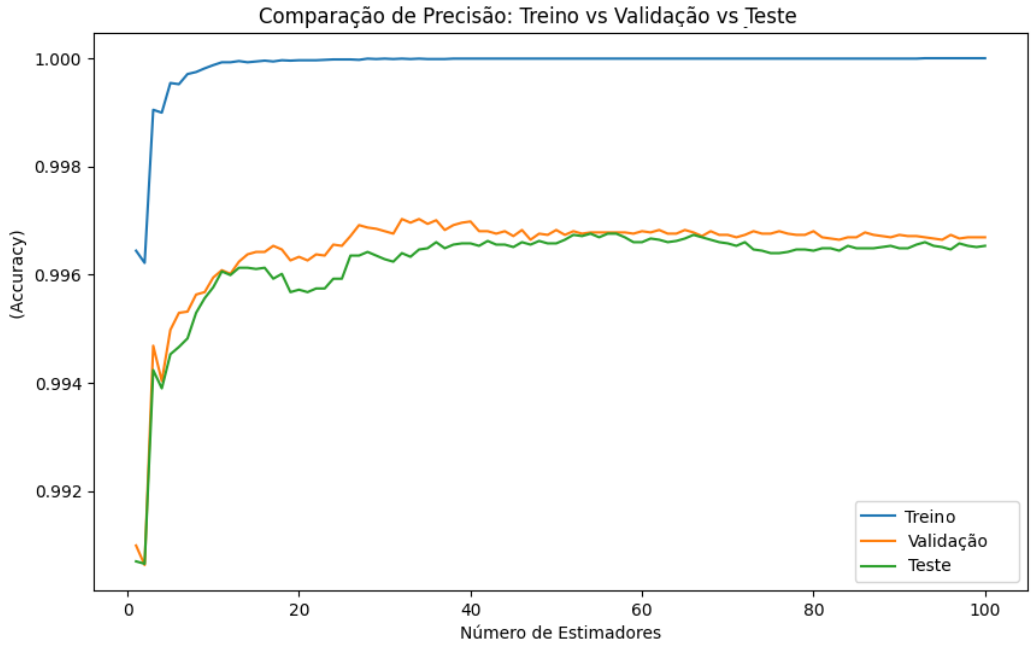


Figura 26 Comparação de 100 amostras desempenho treino, validação e teste *Random Forest*

No que diz respeito ao modelo para previsão de falha dos componentes, foi decidido utilizar exclusivamente o algoritmo GBDT para treinar o correspondente conjunto de dados de teste. Esta decisão foi tomada, considerando que a previsão da falha de componentes contempla uma classificação multi-classes e pelo facto de o treino dos diferentes algoritmos com este conjunto de dados exigir um tempo significativamente superior para o seu processamento.

Na figura 27, é a presentada a matriz de confusão resultante da aplicação deste algoritmo:

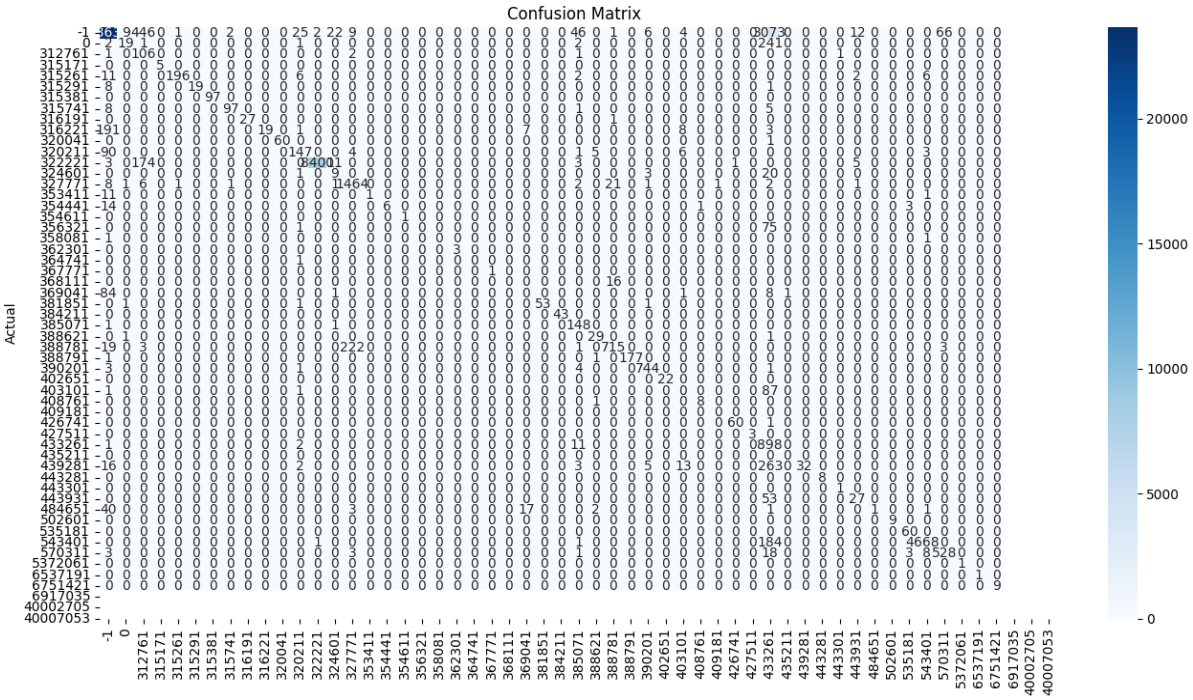


Figura 27 Matriz de confusão para o algoritmo testado

Relativamente às restantes métricas de avaliação, obtiveram-se os resultados apresentados na tabela 5:

Tabela 5 – Resultados das métricas conjunto de dados proposto

Métrica	Conjuntos de dados de falhas dos componentes (%)
Exatidão	**87**
Precisão	**95**
Recall	**87**
f1-score	**89**

O resultado da exatidão, indica que o modelo foi capaz de classificar corretamente 87 % dos exemplos, o que revela uma boa taxa de acerto do modelo. A precisão com o resultado 95 % indica a proporção de exemplos classificados como falhas dos componentes que são realmente falhas.

O *recall* com 87 %, evidencia a proporção de falhas reais que foram corretamente identificadas pelo modelo. O *f1-score* com 89 %, é uma medida que correlaciona a precisão e a *recall*, disponibilizando uma avaliação do desempenho do modelo.

Concluindo, tendo como base a análise apresentada, selecionaram-se os algoritmos DT e GBDT para realizar a previsão do estado de falha dos equipamentos, por evidenciarem um excelente desempenho nas métricas de avaliação e não apresentarem *overfitting* nos dados de entrada. Na figura 28, pode-se observar um gráfico de radar com a avaliação das métricas para cada um dos algoritmos testados.

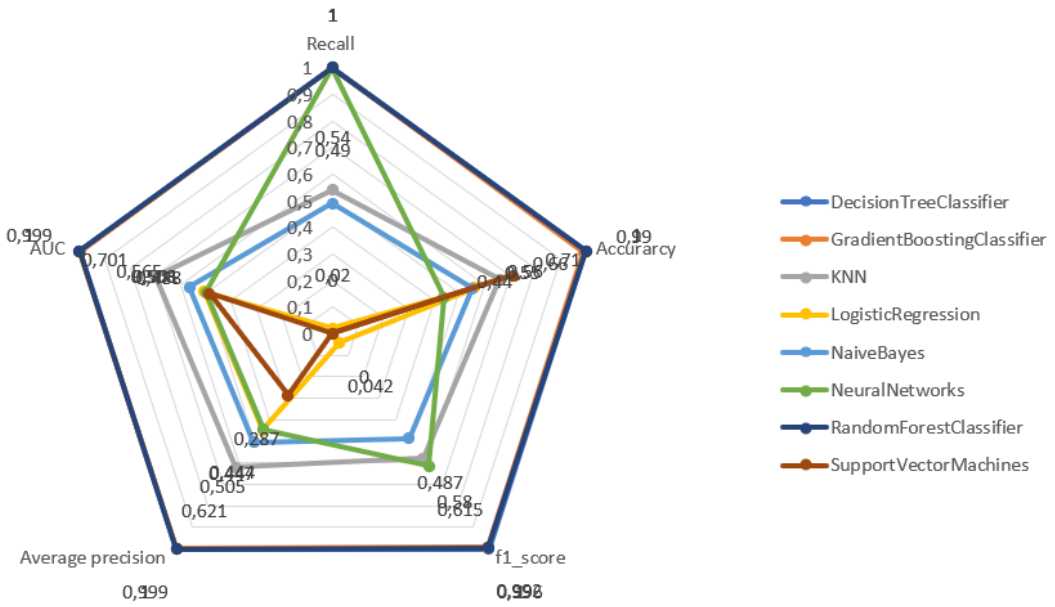


Figura 28 Gráfico de radar avaliação métrica dos algoritmos testados.

Para realizar a previsão da falha dos componentes, selecionou-se exclusivamente o algoritmo GBDT pelas razões já identificadas anteriormente.

5.2. DIVISÃO, TREINO, TESTE E VALIDAÇÃO

Este capítulo tem como objetivo descrever o processo de divisão do conjunto de dados para os algoritmos selecionados, nomeadamente o DT e o GBDT. Esta abordagem, é importante para assegurar que o modelo desenvolvido tenha a capacidade de realizar previsões ou classificações com a precisão adequada.

A divisão dos conjuntos de dados foi realizada de acordo com uma proporção de 70 % para treino e 20 % para teste e 10 % para validação. Esta divisão, é uma prática habitualmente adotada e aceite na área de aprendizagem de máquina e que permite utilizar uma parte dos dados para treinar os modelos e a outra parte para os avaliar.

De seguida, o treino dos modelos foi realizado utilizando o conjunto de dados de treino, que corresponde a 70 % dos dados disponíveis. Os algoritmos previamente selecionados, foram aplicados a este conjunto de dados, tendo como objetivo construir os modelos de forma adequada e identificar padrões nos dados.

O conjunto de teste, composto por 20 % dos dados, foi utilizado para testar a capacidade preditiva dos modelos desenvolvidos. Nesta etapa, os modelos treinados foram aplicados aos dados de teste e as previsões foram comparadas com os valores reais. Esta avaliação permitiu analisar o desempenho e a precisão dos modelos em dados não utilizados anteriormente.

O conjunto de validação, composto por 10 % dos dados, foi utilizado para avaliar o desempenho dos modelos e ajustar caso necessário.

Após a avaliação dos resultados, os conjuntos de dados que apresentaram o melhor desempenho foram separados em conjuntos de treino e teste, tendo-se obtido os resultados apresentados na Tabela 6.

Tabela 6 – Resultados da precisão para os conjuntos de dados propostos

Algoritmo	Precisão (%)	
	Falha nos Equipamentos	Falha nos Componentes
Decision Tree Classifier	100	
Gradient Boosting Decision	99	74

Os dois algoritmos estudados nesta fase, DT e GBDT, obtiveram para o modelo de previsão de falha dos equipamentos um desempenho idêntico. Atendendo que apenas foi utilizado o algoritmo GBDT para o modelo de previsão de falha de componentes, optou-se por prosseguir o estudo com implementação nas previsões apenas com este algoritmo para ambos os modelos.

5.3. ANÁLISE DOS RESULTADOS DAS PREVISÕES

Este capítulo visa, validar se a solução desenvolvida vai ao encontro das necessidades estabelecidas para este trabalho. Na figura 29, pode-se observar uma vista global do painel interativo “Prova de conceito” com todas as análises necessárias, sendo possível analisar se

os resultados obtidos estão de acordo com as expectativas e se a solução cumpre com os requisitos estabelecidos.

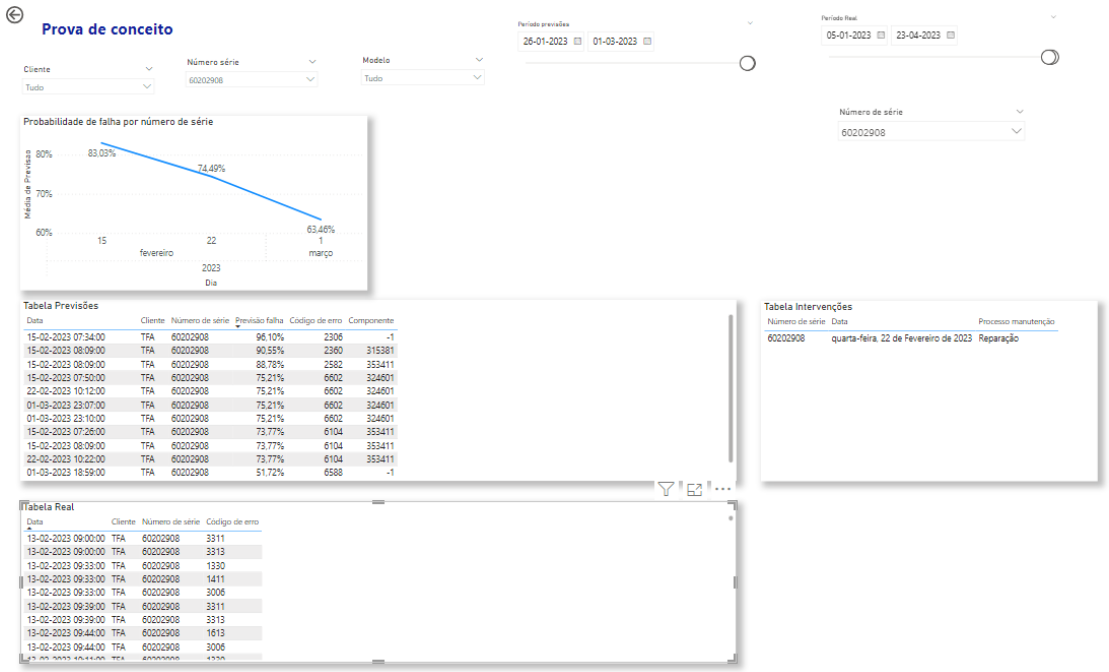


Figura 29 Vista global da prova de conceito

O painel foi desenvolvido com um gráfico de linhas, conforme ilustrado na figura seguinte, tendo como finalidade a análise da evolução da probabilidade de falha de um dado equipamento.

Figura 30 Gráfico da evolução da probabilidade de falha

Na figura 31, é possível comparar a tabela de previsões com a tabela real, onde se demonstra que a previsão está em consonância com os eventos reais. No exemplo dado, é

possível observar que havia uma probabilidade de 90,55 % de ocorrência do defeito 2360 e no componente 315381.

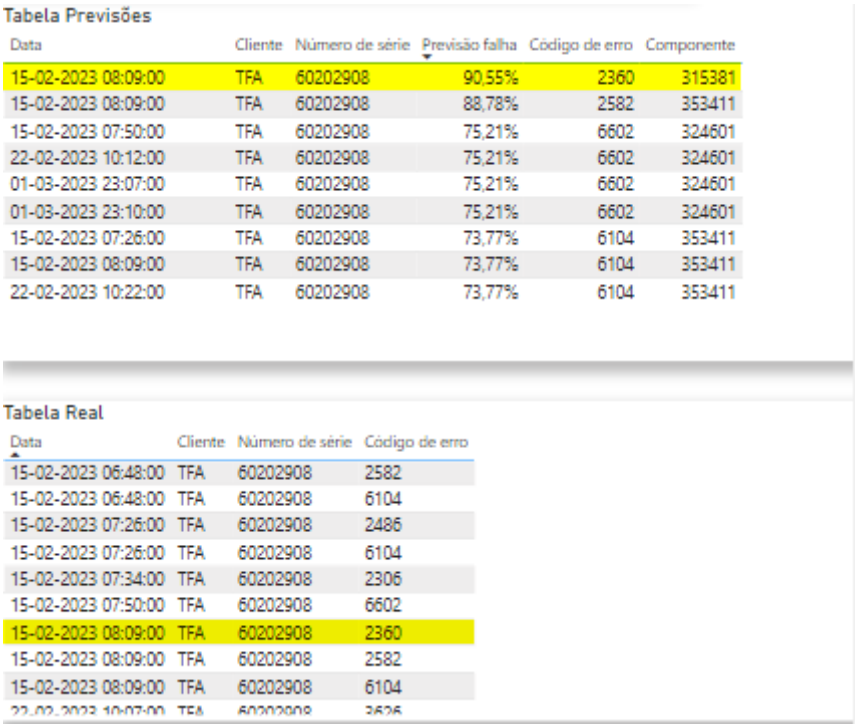


Tabela Previsões

Data	Cliente	Número de série	Previsão falha	Código de erro	Componente
15-02-2023 08:09:00	TFA	60202908	90,55%	2360	315381
15-02-2023 08:09:00	TFA	60202908	88,78%	2582	353411
15-02-2023 07:50:00	TFA	60202908	75,21%	6602	324601
22-02-2023 10:12:00	TFA	60202908	75,21%	6602	324601
01-03-2023 23:07:00	TFA	60202908	75,21%	6602	324601
01-03-2023 23:10:00	TFA	60202908	75,21%	6602	324601
15-02-2023 07:26:00	TFA	60202908	73,77%	6104	353411
15-02-2023 08:09:00	TFA	60202908	73,77%	6104	353411
22-02-2023 10:22:00	TFA	60202908	73,77%	6104	353411

Tabela Real

Data	Cliente	Número de série	Código de erro
15-02-2023 06:48:00	TFA	60202908	2582
15-02-2023 06:48:00	TFA	60202908	6104
15-02-2023 07:26:00	TFA	60202908	2486
15-02-2023 07:26:00	TFA	60202908	6104
15-02-2023 07:34:00	TFA	60202908	2306
15-02-2023 07:50:00	TFA	60202908	6602
15-02-2023 08:09:00	TFA	60202908	2360
15-02-2023 08:09:00	TFA	60202908	2582
15-02-2023 08:09:00	TFA	60202908	6104

Figura 31 Tabela comparação previsão de falhas com dados reais

Por fim, na tabela intervenções, figura 32, é possivel observar que a intervenção para resolução deste evento ocorreu passado 7 dias.

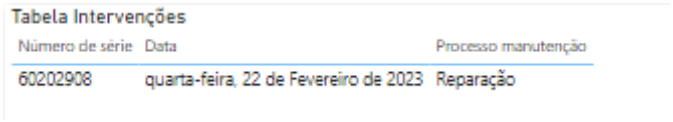


Tabela Intervenções

Número de série	Data	Processo manutenção
60202908	quarta-feira, 22 de Fevereiro de 2023	Reparação

Figura 32 Tabela registo de intervenções

Na figura 33, é possível observar outro exemplo, com uma probabilidade de 88,78 % de ocorrência do defeito 2618 e código componente 388621. O evento foi detetado no dia 1 de março 2023, tendo o equipamento sido intervencionado no dia seguinte.

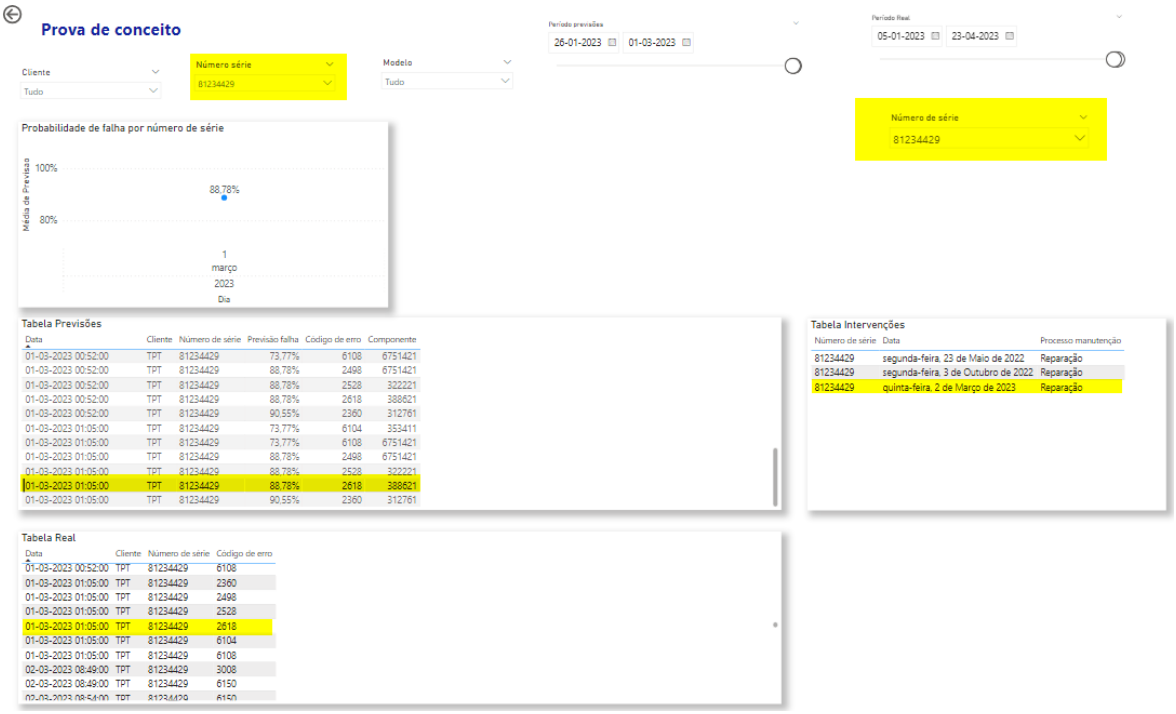


Figura 33 Exemplo 2 prova de conceito

5.4. CONCLUSÃO

Em suma, este estudo proporcionou uma avaliação detalhada dos modelos de aprendizagem automática utilizados, com base em métricas de avaliação especificas. Os algoritmos DT e GBDT, demonstraram um desempenho superior em relação aos demais algoritmos, apresentando elevada exatidão, precisão e *recall*.

Os modelos desenvolvidos, foram capazes de fazer previsões precisas e consistentes em relação às falhas nos equipamentos e componentes, contribuindo para uma deteção antecipada das falhas. Desta forma, as equipas de manutenção podem planear intervenções de carácter preventivo, evitar paragens não programadas e maximizar a disponibilidade dos equipamentos.

6. CONCLUSÕES

A função manutenção é considerada fundamental para as organizações, pois é essencial para garantir a continuidade dos processos de produção, a prestação de serviços e para assegurar que os equipamentos e instalações funcionem de forma eficiente e segura.

A importância do planeamento da manutenção tem aumentado nos últimos anos, devido ao aumento da complexidade dos equipamentos e instalações, e à necessidade de garantir a disponibilidade contínua dos mesmos. Paralelamente, o planeamento da manutenção é visto cada vez mais como uma ferramenta que contribui para a competitividade das empresas, pois permite que elas reduzam os custos de manutenção e melhorem a eficiência dos seus processos.

O presente trabalho apresenta uma abordagem inovadora para a área da manutenção, especificamente no setor da saúde. No primeiro capítulo destaca-se a necessidade de as organizações do setor da saúde acelerarem a inovação e procurarem soluções baseadas em tecnologias digitais que permitam transformar as suas atividades e modelos de negócio.

O segundo capítulo apresenta um estudo sobre o estado da arte da área da manutenção, destacando a importância dos departamentos de manutenção em contexto organizacional, os seus objetivos, tipos e estratégias de manutenção, normas e sistemas de informação.

O terceiro capítulo aborda a evolução da manutenção industrial, o impacto e as contribuições das revoluções industriais para a área da manutenção. É apresentado

igualmente a evolução dos conceitos, tecnologias e soluções atuais para implementação da abordagem da manutenção preditiva.

No capítulo 4, foi descrito o desenvolvimento da solução implementada, abordando a base de dados utilizada neste trabalho, os conjuntos de dados necessários para análise preditiva e visualização. Foi abordada a conexão dos dados à aplicação Power BI, incluindo a sua modelação. Adicionalmente, foi explorado o processo criativo dos painéis interativos desenvolvidos, utilizando as funcionalidades e recursos do Power BI para criar visualizações e painéis personalizados. Por fim, foram descritas as informações apresentadas nos painéis interativos desenvolvidos, destacando sua eficácia e utilidade para as diferentes partes interessadas.

No capítulo 5, avaliou-se detalhadamente os modelos de aprendizagem automática utilizados, com recurso a métricas específicas de avaliação. Os algoritmos DT e GBDT demonstraram um desempenho superior em relação aos demais algoritmos, apresentando elevada precisão, exatidão e *recall*. Foi demonstrado que os modelos desenvolvidos foram capazes de fazer previsões precisas e consistentes em relação às falhas nos equipamentos e componentes, permitindo uma deteção antecipada das falhas. Para este efeito, foi desenvolvido o painel interativo “Prova de conceito” que não estava inicialmente previsto, todavia veio dar um contributo substancial ao trabalho desenvolvido.

Resumindo, o presente trabalho apresenta uma visão geral da área da manutenção e da evolução da manutenção industrial, bem como disponibiliza uma solução inovadora para a previsão de falhas em equipamentos médicos. A solução desenvolvida, poderá contribuir para a otimização do planeamento, assim como o prolongamento do ciclo vida útil dos equipamentos, evitando falhas, diminuindo os tempos e os custos de manutenção.

No decorrer deste trabalho, foram encontradas diversas dificuldades, nomeadamente no que diz respeito ao processamento de dados de alguns algoritmos devido a limitações de hardware, complexidade de dados, volume de dados e na execução de *scripts* Python no Power BI.

Tendo em conta os conhecimentos adquiridos enquanto estudante do curso de Engenharia Eletrotécnica - Sistema Elétricos de Energia, o tema inteligência artificial exigiu uma grande dedicação e investigação da minha parte para que fosse possível ultrapassar as dificuldades que foram naturalmente surgindo ao longo deste trabalho.

Considerando que este trabalho é uma tese de mestrado com metas e prazos estabelecidos, não seria recomendável, nem exequível, prosseguir com outras abordagens além das aqui apresentadas. No entanto, dada a sua relevância e o valor acrescentado que poderão representar, enumera-se aqui como sugestões para investigações futuras:

- Considerar a aplicação de todos os algoritmos estudados ao conjunto de dados utilizados para previsão da falha de componentes ou modelos de aprendizagem automática com multi-classes. Embora tenha-se obtido um desempenho bastante satisfatório com o algoritmo GBDT aplicado a este conjunto de dados, reconhece-se que será necessário explorar e compreender qual o resultado para os restantes algoritmos. Sugerimos, portanto, que em trabalhos futuros dedique-se uma análise mais alargada relativamente a este ponto, considerando diferentes perspetivas e abordagens metodológicas.

- Salienta-se que outras análises merecem ser investigadas para complementar o trabalho aqui apresentado. Por exemplo, a investigação das influências externas no desempenho global dos equipamentos, pode fornecer importantes perspetivas para uma compreensão mais abrangente e contextualizada do tema aqui estudado.

- Por último, podia-se complementar o painel interativo “Reparações preditivas” com outras informações, por exemplo, tempo restante até à falha, classificação do risco da falha e a geração de alertas ou notificações automáticas por e-mail para as partes interessadas.

Referências Bibliográficas

[1] PINTO, João Paulo – Manutenção Lean, 2013. ISBN:978-972-757-877-1.

[2] CABRAL, José Paulo – Organização e gestão da manutenção, 2013. ISBN: 978-972-757-440-7.

[3] CABRAL, José Paulo – Gestão da manutenção de equipamentos, instalações e edifícios, 2021. ISBN: 978-989-752-476-9.

[4] Blucher, editora – Indústria 4.0, conceitos e fundamentos, ISBN: 8521213700 9788521213703.

[5] [18] MURPHY, Kevin P (s.d.). Machine Learning A Probabilistic Perspective. Disponivel online: http://noiselab.ucsd.edu/ECE228/Murphy_Machine_Learning.pdf, janeiro 2023.

[6] GÉRIN, E., AYDIN, O., & Mahdavi-Amiri, A. (2019). Artificial Intelligence. Manual of Digital Earth, 357–385. https://doi.org/10.1007/978-981-32-9915-3_10, janeiro 2023.

[7] MITCHELL, Tom M. - Machine Learning 1st Edition, 1997, ISBN: 978-0070428072.

[8] ALPAYDIN, Ethen - Introduction to Machine Learning, second Editions, 2009, ISBN: 978-0262012430.

[9] Amaral, Fernando – Gestão da manutenção na indústria, 2016. ISBN: 978-989-752-151-5.

[10] Norma EN 13306 de 2007 – Terminologia da manutenção.

[11] Norma EN 15341:2009 – Indicadores de manutenção.

[12] Norma EN 13269 – Norma dos contratos de manutenção.

[13] Norma EN 15628:2014 - Qualificação pessoal técnico de manutenção.

[14] Predictive Maintenance Market Size Worldwide 2020–2030|Statista. 2022, disponível online: https://www.statista.com/statistics/748080/global-predictive-maintenance-market-size, janeiro 2023.

[15] CEN/TC 319 –Normalização no domínio da manutenção no que diz respeito a normas genéricas de aplicação geral. Disponível online: https://standards.cencenelec.eu/dyn/www/f?p=205:7:0::::FSP_ORG_ID:6300&cs=1CEF1EA360BD3851719B096E5D80D730A, janeiro 2023.

[16] Decision Tree Algorithm in Machine Learning - Javatpoint. Disponível online: https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm, janeiro 2023.

[17] d. Disponível online: https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm, janeiro 2023.

[18] K-Nearest Neighbor(KNN) Algorithm for Machine Learning - Javatpoint. Disponível online: https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning, janeiro 2023.

[19] Machine Learning Random Forest Algorithm - Javatpoint. Disponível online: https://www.javatpoint.com/machine-learning-random-forest-algorithm, janeiro 2023.

[20] Machine Learning: What It is, Tutorial, Definition, Types – Javatpoint. Disponível online: https://www.javatpoint.com/machine-learning, janeiro 2023.

[21] Industry 4.0 and predictive technologies for asset maintenance | Deloitte Insights – disponível online: https://www2.deloitte.com/uk/en/insights/focus/industry-4-0/using-predictive-technologies-for-asset-maintenance.html, janeiro 2023.

[22] Predictive Maintenance and The Smart Factory | Deloitte US, disponível online: https://www2.deloitte.com/us/en/pages/operations/articles/predictive-maintenance-and-the-smart-factory.html, janeiro 2023.

[23] Deloitte_Predictive-Maintenance_PositionPaper.pdf, disponível online: https://www2.deloitte.com/content/dam/Deloitte/de/Documents/deloitte-analytics/Deloitte_Predictive-Maintenance_PositionPaper.pdf, janeiro 2023.

[24] Industry 4.0 (deloitte.com), disponível online: https://www2.deloitte.com/us/en/insights/focus/industry-4-0.html, janeiro 2023.

[25] Digital industrial transformation in the age of Industry 4.0 | Deloitte Insights, Disponível online: https://www2.deloitte.com/us/en/insights/focus/industry-4-0/digital-industrial-transformation-industrial-internet-of-things.html, janeiro 2023.

[26] Automated machine learning and predictive insights | Deloitte Insights, disponível online: https://www2.deloitte.com/xe/en/insights/topics/analytics/automated-machine-learning-predictive-insights.html, janeiro 2023.

[27] Industry 5.0: The Arising of a Concept - ScienceDirect, disponível online: https://www.sciencedirect.com/science/article/pii/S1877050922023973, janeiro 2023.

[28] Merging two revolutions: A human-artificial intelligence method to study how sustainability and Industry 4.0 are intertwined - ScienceDirect, disponível online: https://www.sciencedirect.com/science/article/pii/S0040162522007867, janeiro 2023.

[29] An integrated outlook of Cyber–Physical Systems for Industry 4.0: Topical practices, architecture, and applications - ScienceDirect, disponível online: https://www.sciencedirect.com/science/article/pii/S294973612200001X, janeiro 2023.

[30] Artificial intelligence for industry 4.0: Systematic review of applications, challenges, and opportunities - ScienceDirect, disponível online: https://www.sciencedirect.com/science/article/pii/S0957417422024757, janeiro 2023.

[31] Industry 4.0 vs. Industry 5.0: Co-existence, Transition, or a Hybrid - ScienceDirect, disponível online: https://www.sciencedirect.com/science/article/pii/S1877050922022840, janeiro 2023.

[32] A visual review of artificial intelligence and Industry 4.0 in healthcare - ScienceDirect, disponível online: https://www.sciencedirect.com/science/article/pii/S0045790622002269, janeiro 2023.

[33] Basic Artificial Intelligence Techniques: Machine Learning and Deep Learning - ScienceDirect, disponível online: https://www.sciencedirect.com/science/article/abs/pii/S0033838921000786, janeiro 2023.

[34] Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications - ScienceDirect, disponível online:

https://www.sciencedirect.com/science/article/pii/S0926580522003132, janeiro 2023.

[35] TURING, Alan – Cientista Universal, 1947, ISBN: 978-989-8974-02-0.

[36] John McCarthy | Biography & Facts | Britannica, disponível online: https://www.britannica.com/biography/John-McCarthy, janeiro 2023.

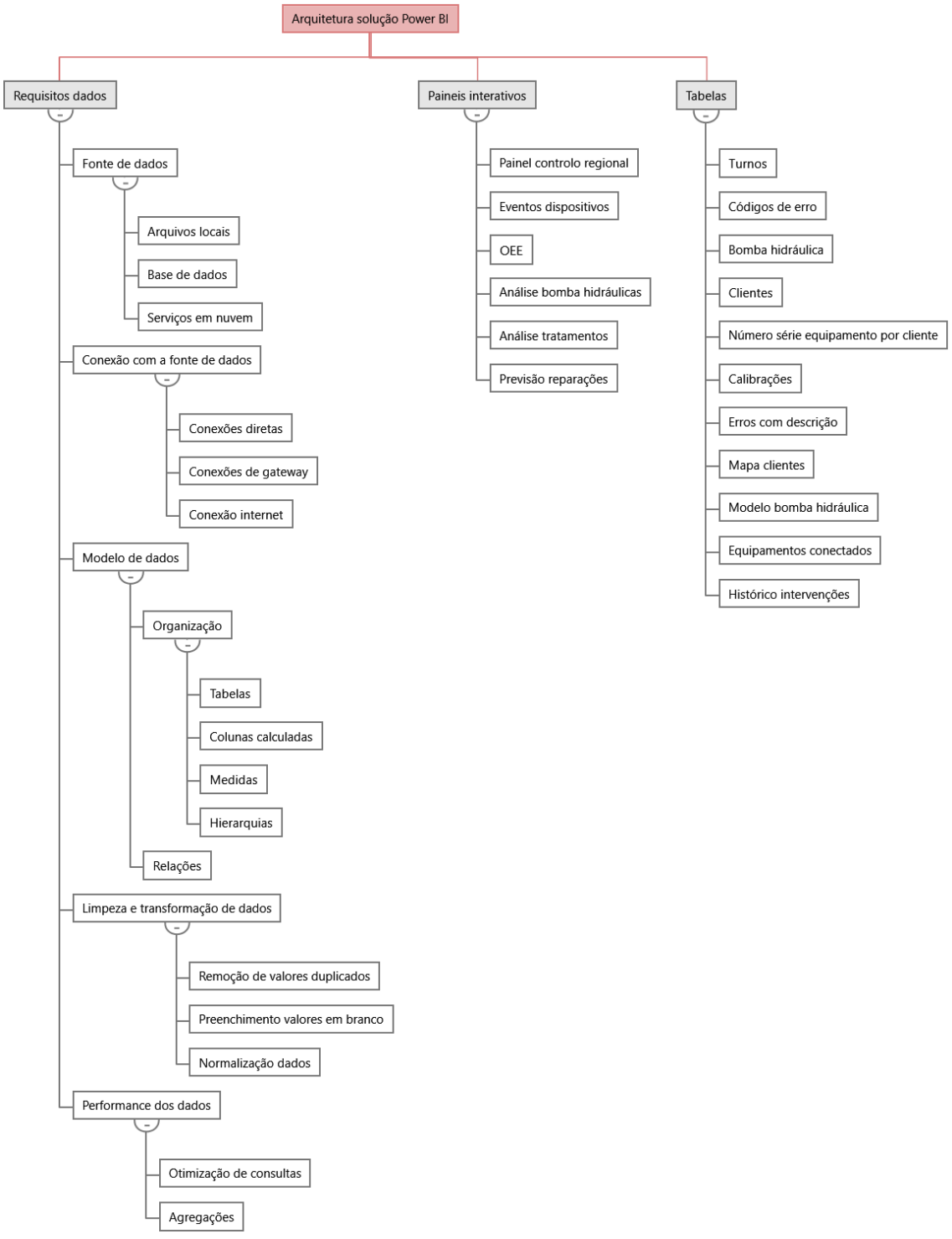
[37] Site oficial da Plattform Industrie 4.0, disponível online https://www.plattform-i40.de/en/home.html

[38] Big Data in academic libraries: literature review and future research directions, disponível online: https://www.researchgate.net/figure/The-5V-model-that-currently-defines-Big-Data_fig2_330872470

[39] MÜLLER, Andreas– Introduction to Machine Learning with Python, 2016. ISBN:978-1-449-36941-5.

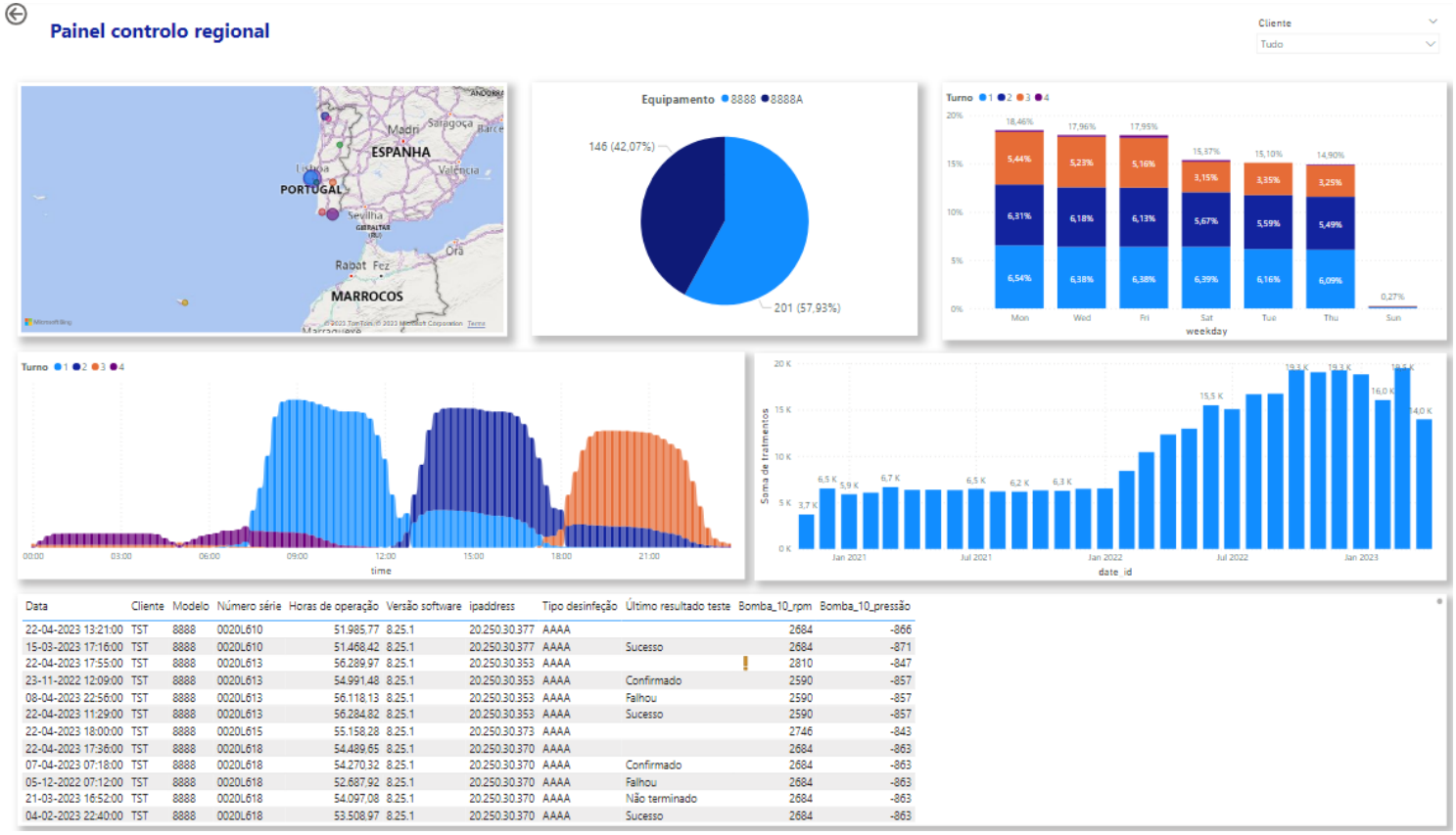
Anexo A. Mapa ideias arquitetura solução Power BI

Neste anexo é apresentado em detalhe o mapa de ideias que serviu de base para o planeamento da arquitetura da solução.



Anexo B. Painel controlo regional solução Power BI

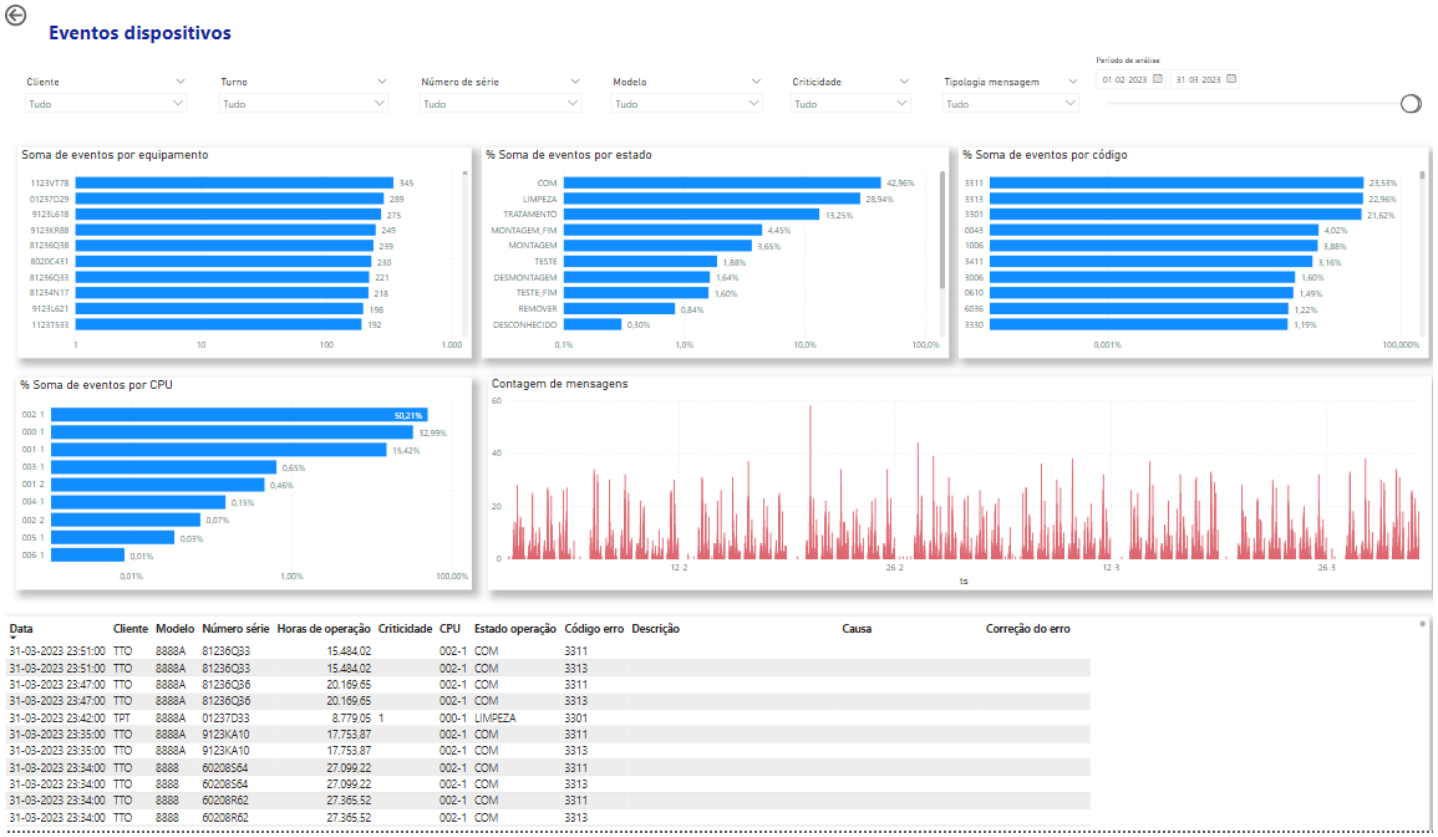
Neste anexo é apresentado o painel interativo da solução desenvolvida no Power BI, “Painel de controlo regional”.



Data	Cliente	Modelo	Número série	Horas de operação	Versão software	ipaddress	Tipo desinfeção	Último resultado teste	Bomba_10_rpm	Bomba_10_pressão
22-04-2023 13:21:00	TST	8888	0020L610	51.985,77	8.25.1	20.250.30.377	AAAA		2684	-866
15-03-2023 17:16:00	TST	8888	0020L610	51.468,42	8.25.1	20.250.30.377	AAAA	Sucesso	2684	-871
22-04-2023 17:55:00	TST	8888	0020L613	56.289,97	8.25.1	20.250.30.353	AAAA		2810	-847
23-11-2022 12:09:00	TST	8888	0020L613	54.991,48	8.25.1	20.250.30.353	AAAA	Confirmado	2590	-857
08-04-2023 22:56:00	TST	8888	0020L613	56.118,13	8.25.1	20.250.30.353	AAAA	Falhou	2590	-857
22-04-2023 11:29:00	TST	8888	0020L613	56.284,82	8.25.1	20.250.30.353	AAAA	Sucesso	2590	-857
22-04-2023 18:00:00	TST	8888	0020L615	55.158,28	8.25.1	20.250.30.373	AAAA		2746	-843
22-04-2023 17:36:00	TST	8888	0020L618	54.489,65	8.25.1	20.250.30.370	AAAA		2684	-863
07-04-2023 07:18:00	TST	8888	0020L618	54.270,32	8.25.1	20.250.30.370	AAAA	Confirmado	2684	-863
05-12-2022 07:12:00	TST	8888	0020L618	52.687,92	8.25.1	20.250.30.370	AAAA	Falhou	2684	-863
21-03-2023 16:52:00	TST	8888	0020L618	54.097,08	8.25.1	20.250.30.370	AAAA	Não terminado	2684	-863
04-02-2023 22:40:00	TST	8888	0020L618	53.508,97	8.25.1	20.250.30.370	AAAA	Sucesso	2684	-863

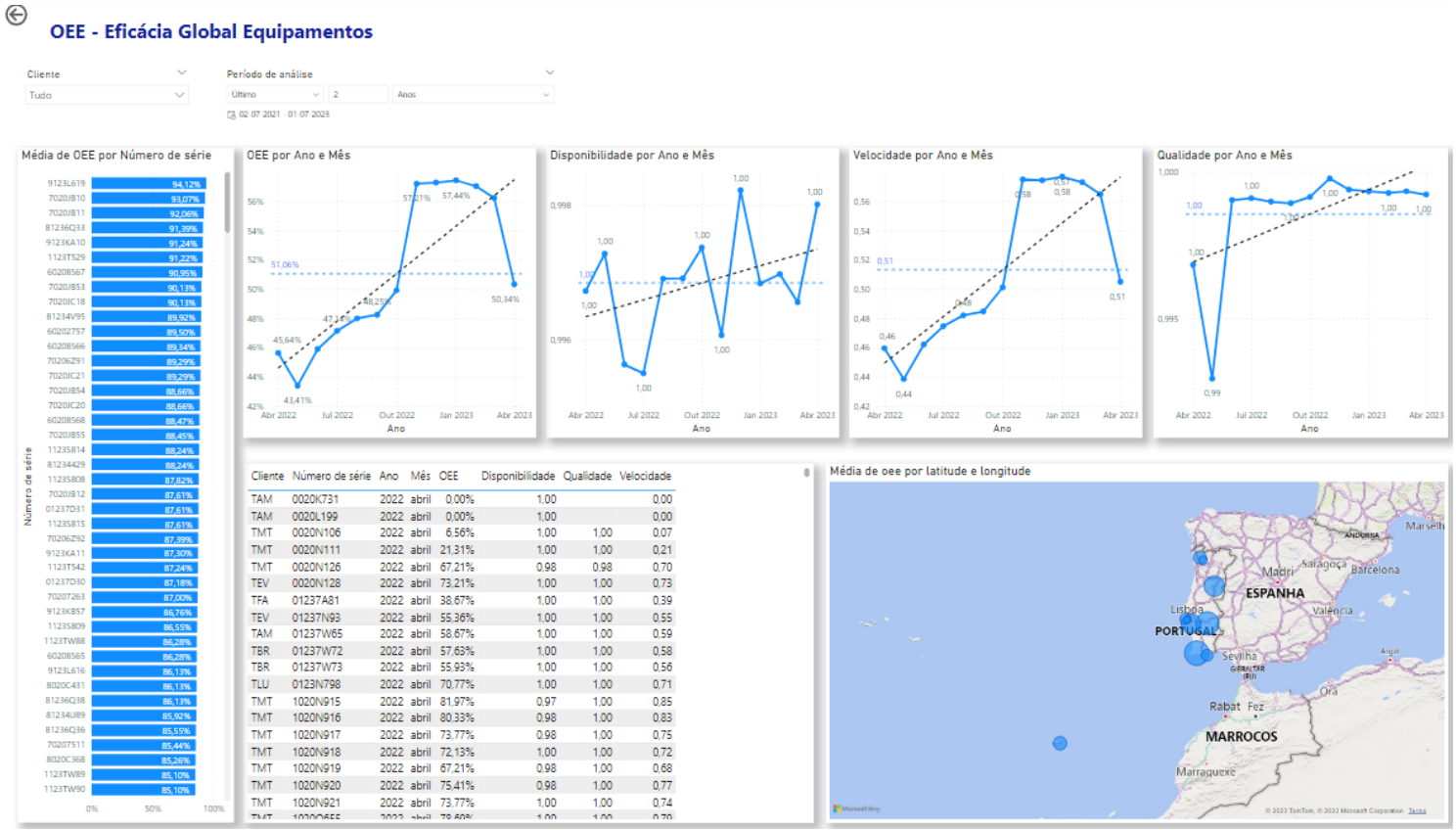
Anexo C. Painel Eventos Dispositivos

Neste anexo é apresentado o painel interativo da solução desenvolvida no Power BI, “Painel Eventos Dispositivos”.



Anexo D. Painel OEE–Eficácia Global Equipamentos

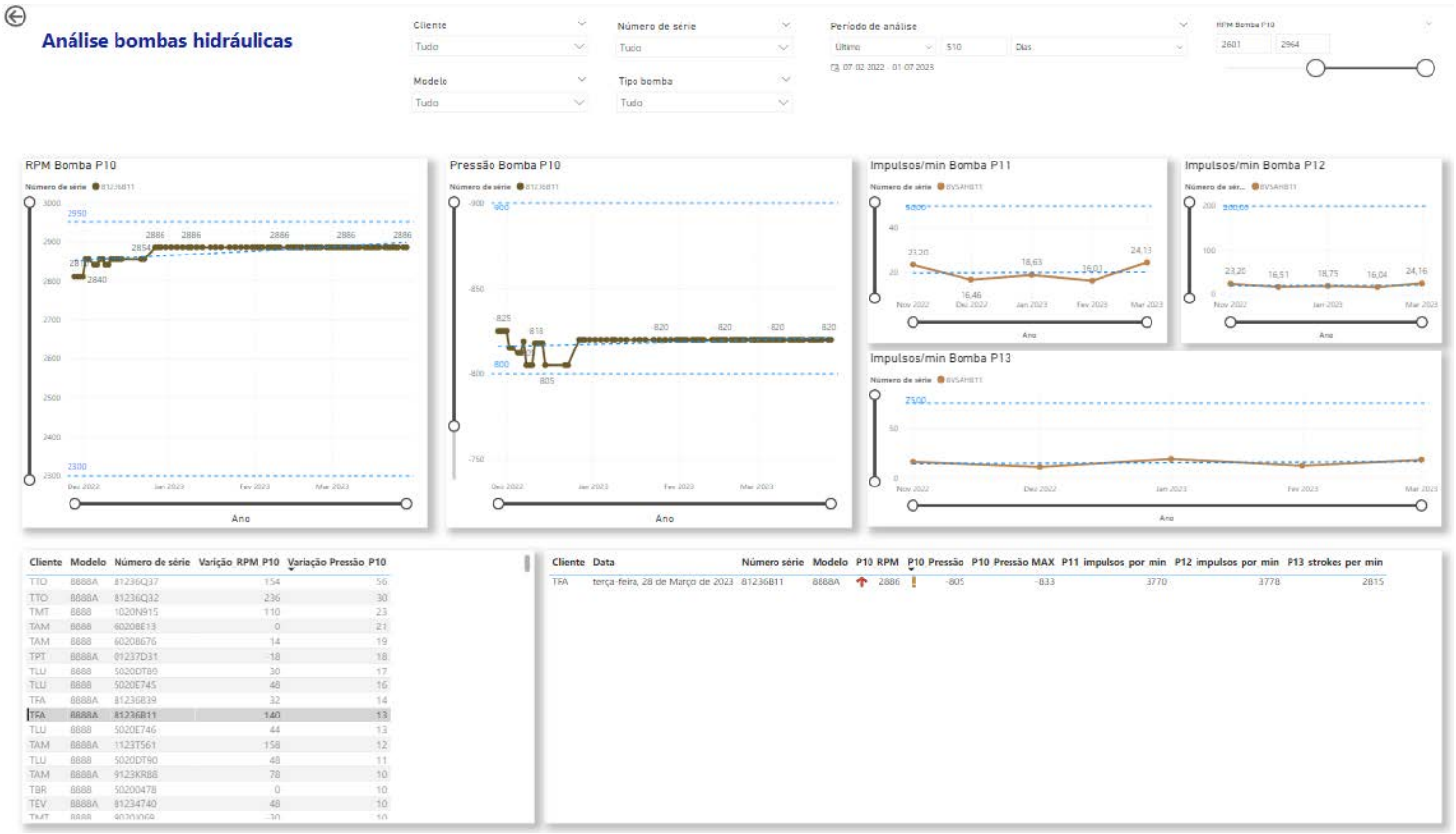
Neste anexo é apresentado o painel interativo da solução desenvolvida no Power BI, “OEE-Eficácia Global Equipamentos”.



Cliente	Número de série	Ano	Mês	OEE	Disponibilidade	Qualidade	Velocidade
TAM	0020K731	2022	abril	0,00%	1,00		0,00
TAM	0020L199	2022	abril	0,00%	1,00		0,00
TMT	0020N106	2022	abril	6,56%	1,00	1,00	0,07
TMT	0020N111	2022	abril	21,31%	1,00	1,00	0,21
TMT	0020N126	2022	abril	67,21%	0,98	0,98	0,70
TEV	0020N128	2022	abril	73,21%	1,00	1,00	0,73
TFA	01237A81	2022	abril	38,67%	1,00	1,00	0,39
TEV	01237N93	2022	abril	55,36%	1,00	1,00	0,55
TAM	01237W65	2022	abril	58,67%	1,00	1,00	0,59
TBR	01237W72	2022	abril	57,63%	1,00	1,00	0,58
TBR	01237W73	2022	abril	55,93%	1,00	1,00	0,56
TLU	0123N798	2022	abril	70,77%	1,00	1,00	0,71
TMT	1020N915	2022	abril	81,97%	0,97	1,00	0,85
TMT	1020N916	2022	abril	80,33%	0,98	1,00	0,83
TMT	1020N917	2022	abril	73,77%	0,98	1,00	0,75
TMT	1020N918	2022	abril	72,13%	1,00	1,00	0,72
TMT	1020N919	2022	abril	67,21%	0,98	1,00	0,68
TMT	1020N920	2022	abril	75,41%	0,98	1,00	0,77
TMT	1020N921	2022	abril	73,77%	1,00	1,00	0,74

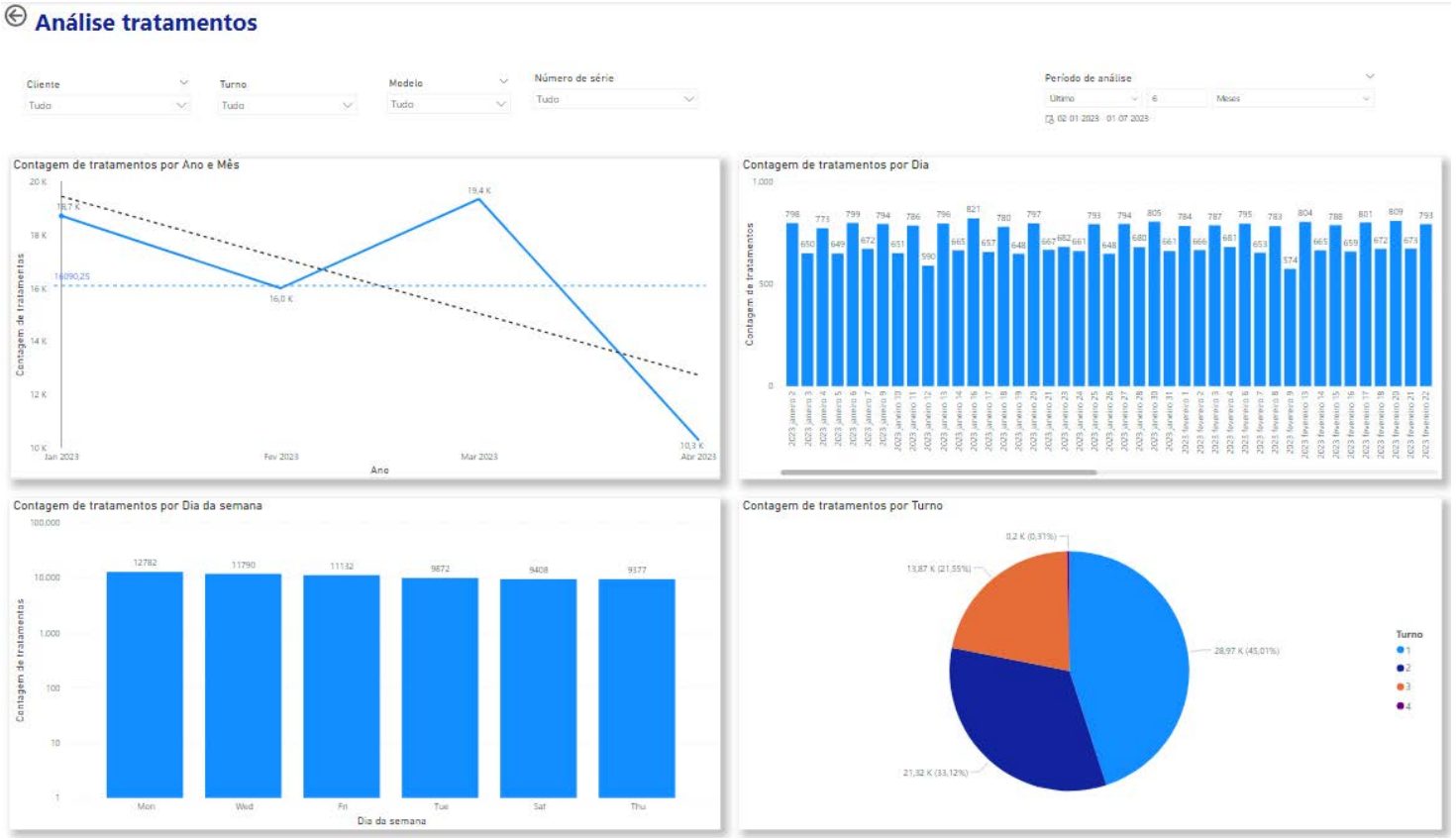
Anexo E. Painel análise bombas hidráulicas

Neste anexo é apresentado o painel interativo da solução desenvolvida no Power BI, “Painel análise bombas hidráulicas”.



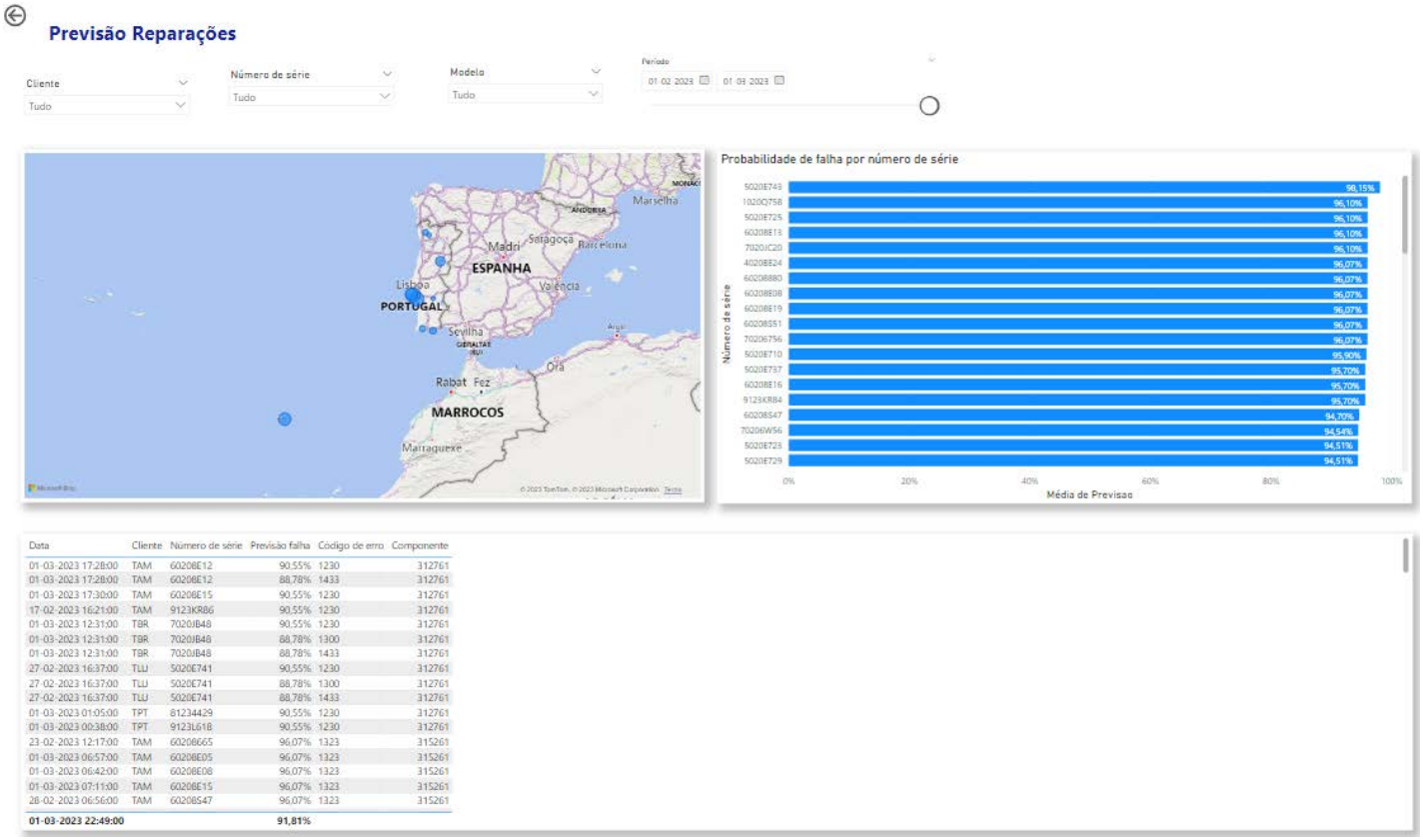
Anexo F. Painel análise de tratamentos

Neste anexo é apresentado o painel interativo da solução desenvolvida no Power BI, “Painel análise de tratamentos”.



Anexo G. Painel previsão reparações

Neste anexo é apresentado o painel interativo da solução desenvolvida no Power BI, “Painel previsão reparações”.



Anexo H. Código Python algoritmo DT

Neste anexo é apresentado o código Python desenvolvido para treinar o modelo com o algoritmo *Decision Tree Classifier*.

```
from sklearn.metrics import accuracy_score
import seaborn as sns
import mglearn
import pandas as pd
import numpy as np
from sklearn.calibration import CalibratedClassifierCV
from sklearn.metrics import precision_recall_curve, f1_score
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report,
mean_squared_error
from sklearn import tree
from IPython.display import Image

# Carregar os dados da memória de erros do equipamento em um
DataFrame
from sklearn.tree import DecisionTreeClassifier
from IPython.core.display_functions import display


dados = pd.read_excel('C:\\Users\\Fernando
Salgado\\PycharmProjects\\ ML_dataset_error_logs_total.xlsx')
dados_or = dados.copy()


colunas_factorized =
['criticality','user_device','product','testresult']
colunas_drop =
['error_name','error_code','state','type','serialnumber','ipaddres
s','machine_operating_hours',

'partition_year_month','date_id','kpi_treatment_duration','msgXsta
te','hierarchyl5','Spare Part NR','Spare Part Description',
'Status
1/0','cpu_name','cpu_shortcut','ts','sessionid','monitor_c_pess_op
erating_hours','monitor_p_pess_operating_hours',
'cnt','icc_url','typ','flag_only_dev','hierarchyl1','hierarchyl2',
'hierarchyl3','hierarchyl4',
'ebm_c_pess_operating_hours','ebm_p_pess_operating_hours','hydraul
ic_c_pess_operating_hours','hydraulic_p_pess_operating_hours',
'psu_c_pess_operating_hours','bpm_c_pess_operating_hours','btm_c_p
ess_operating_hours','bvm_c_pess_operating_hours',
'bic_strokes_in_session','so_strokes_in_session','uf_strokes_in_se
ssion','msgXvalue','softwareversion','en_us_text',
'ts_5min_trunc','error_log_ts']
```

```
# Criar map para voltar aos valores originais
mapeamentos = {}
for coluna in colunas_factorized:
    factors, unique_indices = pd.factorize(dados_or[coluna])
    dados[coluna] = factors
    mapeamentos[coluna] = {i: value for i, value in
enumerate(dados_or[coluna].unique())}


# Separar as características (X) e o rótulo (y)
X = dados.drop(colunas_drop, axis=1)  # Assume que a coluna
'falha' indica a ocorrência de falha
y = dados['Status 1/0']

# Dividir os dados em conjuntos de treinamento e teste
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)


# Criar e treinar o modelo
modelo = DecisionTreeClassifier()
modelo.fit(X_train, y_train)

# Fazer previsões
previsoes = modelo.predict(X_test)

# Avaliar o desempenho do modelo
print(classification_report(y_test, previsoes))

# Listas para armazenar a precisão dos diferentes conjuntos
train_accuracy = []
val_accuracy = []
test_accuracy = []

# Testar diferentes profundidades máximas da árvore
max_depth_range = range(1, 21)
for max_depth in max_depth_range:
    # Criar um classificador DecisionTree
    dt = DecisionTreeClassifier(max_depth=max_depth,
random_state=42)

    # Treinar o classificador
    dt.fit(X_train, y_train)

    # Fazer previsões nos conjuntos de treinamento, validação e
teste
    y_train_pred = dt.predict(X_train)
    y_val_pred = dt.predict(X_val)
    y_test_pred = dt.predict(X_test)

    # Calcular a precisão das previsões
    train_accuracy.append(accuracy_score(y_train, y_train_pred))
    val_accuracy.append(accuracy_score(y_val, y_val_pred))
    test_accuracy.append(accuracy_score(y_test, y_test_pred))
```

```
# Probabilidades
probabilidades = modelo.predict_proba(X_test)

# Exibir as probabilidades
print(probabilidades)


print("f1_score of decision tree: {:.3f}".format(f1_score(y_test,
modelo.predict(X_test))))

from sklearn.metrics import average_precision_score
ap_rf = average_precision_score(y_test,
modelo.predict_proba(X_test)[:, 1])
print("Average precision of decision tree: {:.3f}".format(ap_rf))

from sklearn.metrics import roc_auc_score
rf_auc = roc_auc_score(y_test, modelo.predict_proba(X_test)[:, 1])
print("AUC for Decision Tree: {:.3f}".format(rf_auc))

# Converter as probabilidades em DataFrame
probabilidades = pd.DataFrame(probabilidades,
columns=modelo.classes_)
prediction_result = pd.concat([X_test.reset_index(drop=True),
probabilidades], axis=1)


# Combinar as probabilidades com as demais variáveis de entrada
prediction_fatorized = pd.concat([X_test.reset_index(drop=True),
probabilidades], axis=1)


#Reverter
previsoes_restored = pd.DataFrame()

colunas_or = prediction_result.columns

for coluna in colunas_or:

    if (coluna in colunas_factorized):
        valores_previstos =
np.take(list(mapeamentos[coluna].values()),
prediction_fatorized[coluna])
        previsoes_restored[coluna] = valores_previstos
    else:
        previsoes_restored[coluna] = prediction_fatorized[coluna]
```

Anexo I. Código Python algoritmo GBDT

Neste anexo é apresentado o código Python desenvolvido para treinar o modelo com o algoritmo *Gradient Boosted Decision Tree Classifier*.

```
from sklearn.metrics import accuracy_score
import seaborn as sns
import mglearn
import pandas as pd
import numpy as np
from sklearn.calibration import CalibratedClassifierCV
from sklearn.metrics import precision_recall_curve, f1_score
from sklearn.ensemble import GradientBoostingClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report,
mean_squared_error
from sklearn import tree
from IPython.display import Image

# Carregar os dados da memória de erros do equipamento em um
DataFrame
from sklearn.tree import DecisionTreeClassifier
from IPython.core.display_functions import display


dados = pd.read_excel('C:\\Users\\Fernando
Salgado\\PycharmProjects\\ ML_dataset_error_logs_total.xlsx')
dados_or = dados.copy()


colunas_factorized =
['criticality','user_device','product','testresult']
colunas_drop =
['error_name','error_code','state','type','serialnumber','ipaddres
s','machine_operating_hours',
'partition_year_month','date_id','kpi_treatment_duration','msgXsta
te','hierarchyl5','Spare Part NR','Spare Part Description',
'Status
1/0','cpu_name','cpu_shortcut','ts','sessionid','monitor_c_pess_op
erating_hours','monitor_p_pess_operating_hours',
'cnt','icc_url','typ','flag_only_dev','hierarchyl1','hierarchyl2',
'hierarchyl3','hierarchyl4',
'ebm_c_pess_operating_hours','ebm_p_pess_operating_hours','hydraul
ic_c_pess_operating_hours','hydraulic_p_pess_operating_hours',
'psu_c_pess_operating_hours','bpm_c_pess_operating_hours','btm_c_p
ess_operating_hours','bvm_c_pess_operating_hours',
'bic_strokes_in_session','so_strokes_in_session','uf_strokes_in_se
ssion','msgXvalue','softwareversion','en_us_text',
                'ts_5min_trunc','error_log_ts'
                ]
```

```
# Criar map para voltar aos valores originais
mapeamentos = {}
for coluna in colunas_factorized:
    factors, unique_indices = pd.factorize(dados_or[coluna])
    dados[coluna] = factors
    mapeamentos[coluna] = {i: value for i, value in
enumerate(dados_or[coluna].unique())}


# Separar as características (X) e o rótulo (y)
X = dados.drop(colunas_drop, axis=1)  # Assume que a coluna
'falha'
y = dados['Status 1/0']

# Dividir os dados em conjuntos de treinamento e teste
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

# Criar e treinar o modelo
modelo = GradientBoostingClassifier()
modelo.fit(X_train, y_train)

# Fazer previsões
previsoes = modelo.predict(X_test)

# Avaliar o desempenho do modelo
print(classification_report(y_test, previsoes))

# Listas para armazenar a precisão dos diferentes conjuntos
train_accuracy = []
val_accuracy = []
test_accuracy = []

# Probabilidades
probabilidades = modelo.predict_proba(X_test)

# Exibir as probabilidades
print(probabilidades)

print("f1_score of GradientBoostingClassifier:
{:.3f}".format(f1_score(y_test, modelo.predict(X_test))))

from sklearn.metrics import average_precision_score
ap_rf = average_precision_score(y_test,
modelo.predict_proba(X_test)[:, 1])
print("Average precision of GradientBoostingClassifier:
{:.3f}".format(ap_rf))

from sklearn.metrics import roc_auc_score
rf_auc = roc_auc_score(y_test, modelo.predict_proba(X_test)[:, 1])
print("AUC for GradientBoostingClassifier: {:.3f}".format(rf_auc))


# Converter as probabilidades em DataFrame
probabilidades = pd.DataFrame(probabilidades,
columns=modelo.classes_)
```

```
# Combinar as probabilidades com as demais variáveis de entrada
prediction_result = pd.concat([X_test.reset_index(drop=True),
probabilidades], axis=1)


# Combinar as probabilidades com as demais variáveis de entrada
prediction_fatorized = pd.concat([X_test.reset_index(drop=True),
probabilidades], axis=1)


#Reverter
previsoes_restored = pd.DataFrame()
colunas_or = prediction_result.columns

for coluna in colunas_or:

    if (coluna in colunas_factorized):
        valores_previstos =
np.take(list(mapeamentos[coluna].values()),
prediction_fatorized[coluna])
        previsoes_restored[coluna] = valores_previstos
    else:
        previsoes_restored[coluna] = prediction_fatorized[coluna]
```

Anexo J. Código Python algoritmo RF

Neste anexo é apresentado o código Python desenvolvido para treinar o modelo com o algoritmo *Random Forest Classifier*.

```
import seaborn as sns
import mglearn
import pandas as pd
import numpy as np
from sklearn.calibration import CalibratedClassifierCV
from sklearn.metrics import precision_recall_curve, f1_score
from sklearn.ensemble import RandomForestClassifier,
RandomForestRegressor
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report,
mean_squared_error
from sklearn.metrics import accuracy_score
from IPython.display import Image

# Carregar os dados da memória de erros do equipamento em um
DataFrame
from IPython.core.display_functions import display

dados = pd.read_excel('C:\\Users\\Fernando
Salgado\\PycharmProjects\\ ML_dataset_error_logs_total.xlsx')
dados_or = dados.copy()


colunas_factorized =
['criticality','user_device','product','testresult']
colunas_drop =
['error_name','error_code','state','type','serialnumber','ipaddres
s','machine_operating_hours',
'partition_year_month','date_id','kpi_treatment_duration','msgXsta
te','hierarchyl5','Spare Part NR','Spare Part Description',
'Status
1/0','cpu_name','cpu_shortcut','ts','sessionid','monitor_c_pess_op
erating_hours','monitor_p_pess_operating_hours',
'cnt','icc_url','typ','flag_only_dev','hierarchyl1','hierarchyl2',
'hierarchyl3','hierarchyl4',
'ebm_c_pess_operating_hours','ebm_p_pess_operating_hours','hydraul
ic_c_pess_operating_hours','hydraulic_p_pess_operating_hours',
'psu_c_pess_operating_hours','bpm_c_pess_operating_hours','btm_c_p
ess_operating_hours','bvm_c_pess_operating_hours',
'bic_strokes_in_session','so_strokes_in_session','uf_strokes_in_se
ssion','msgXvalue','softwareversion','en_us_text',
'ts_5min_trunc','error_log_ts']


# Criar map para voltar aos valores originais
```

```
mapeamentos = {}
for coluna in colunas_factorized:
    factors, unique_indices = pd.factorize(dados_or[coluna])
    dados[coluna] = factors
    mapeamentos[coluna] = {i: value for i, value in
enumerate(dados_or[coluna].unique())}


# Separar as características (X) e o rótulo (y)
X = dados.drop(colunas_drop, axis=1)  # Assume que a coluna
'falha'
y = dados['Status 1/0']

# Dividir os dados em conjuntos de treinamento e teste
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)


# Criar e treinar o modelo
modelo = RandomForestClassifier()
modelo.fit(X_train, y_train)

# Fazer previsões
previsoes = modelo.predict(X_test)

# Avaliar o desempenho do modelo
print("Classifier report:")
print(classification_report(y_test, previsoes))

print("Métricas teste X_train com y_train:")
# Fazer previsões no conjunto de treinamento e teste
y_train_pred = modelo.predict(X_train)
y_test_pred = modelo.predict(X_test)

# Calcular a acurácia das previsões
train_accuracy = accuracy_score(y_train, y_train_pred)
test_accuracy = accuracy_score(y_test, y_test_pred)

print("Acurácia no conjunto de treinamento:", train_accuracy)
print("Acurácia no conjunto de teste:", test_accuracy)


# Probabilidades
probabilidades = modelo.predict_proba(X_test)

# Exibir as probabilidades
print(probabilidades)


print("f1_score of RandomForestClassifier:
{:.3f}".format(f1_score(y_test, modelo.predict(X_test))))

from sklearn.metrics import average_precision_score
ap_rf = average_precision_score(y_test,
modelo.predict_proba(X_test)[:, 1])
print("Average precision of RandomForestClassifier:
```

```
{:.3f}".format(ap_rf))

from sklearn.metrics import roc_auc_score
rf_auc = roc_auc_score(y_test, modelo.predict_proba(X_test)[:, 1])
print("AUC for RandomForestClassifier: {:.3f}".format(rf_auc))

# Converter as probabilidades em DataFrame
probabilidades = pd.DataFrame(probabilidades,
columns=modelo.classes_)
prediction_result = pd.concat([X_test.reset_index(drop=True),
probabilidades], axis=1)


# Combinar as probabilidades com as demais variáveis de entrada
prediction_fatorized = pd.concat([X_test.reset_index(drop=True),
probabilidades], axis=1)


#Reverter
previsoes_restored = pd.DataFrame()

colunas_or = prediction_result.columns

for coluna in colunas_or:

    if (coluna in colunas_factorized):
        valores_previstos =
np.take(list(mapeamentos[coluna].values()),
prediction_fatorized[coluna])
        previsoes_restored[coluna] = valores_previstos
    else:
        previsoes_restored[coluna] = prediction_fatorized[coluna]
```

Anexo K. Código Python algoritmo KNN

Neste anexo é apresentado o código Python desenvolvido para treinar o modelo com o algoritmo *K-Nearest Neighbors*.

```
import seaborn as sns
import mglearn
import pandas as pd
import numpy as np
from sklearn.calibration import CalibratedClassifierCV
from sklearn.metrics import precision_recall_curve, f1_score
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report,
mean_squared_error
from IPython.display import Image

# Carregar os dados da memória de erros do equipamento em um
DataFrame
from IPython.core.display_functions import display

print("Avaria 0/1 _ ML_dataset_error_logs_total.xlsx _
KNeighborsClassifier _ só matrix confusao")
dados = pd.read_excel('C:\\Users\\Fernando
Salgado\\PycharmProjects\\ ML_dataset_error_logs_total.xlsx')

dados_or = dados.copy()



colunas_factorized =
['criticality','user_device','product','testresult']
colunas_drop =
['error_name','error_code','state','type','serialnumber','ipaddres
s','machine_operating_hours',
'partition_year_month','date_id','kpi_treatment_duration','msgXsta
te','hierarchyl5','Spare Part NR','Spare Part Description',
'Status
1/0','cpu_name','cpu_shortcut','ts','sessionid','monitor_c_pess_op
erating_hours','monitor_p_pess_operating_hours',
'cnt','icc_url','typ','flag_only_dev','hierarchyl1','hierarchyl2',
'hierarchyl3','hierarchyl4',
'ebm_c_pess_operating_hours','ebm_p_pess_operating_hours','hydraul
ic_c_pess_operating_hours','hydraulic_p_pess_operating_hours',
'psu_c_pess_operating_hours','bpm_c_pess_operating_hours','btm_c_p
ess_operating_hours','bvm_c_pess_operating_hours',
'bic_strokes_in_session','so_strokes_in_session','uf_strokes_in_se
ssion','msgXvalue','softwareversion','en_us_text',
'ts_5min_trunc','error_log_ts']
```

```
# Criar map para voltar aos valores originais
mapeamentos = {}
for coluna in colunas_factorized:
    factors, unique_indices = pd.factorize(dados_or[coluna])
    dados[coluna] = factors
    mapeamentos[coluna] = {i: value for i, value in
enumerate(dados_or[coluna].unique())}


# Separar as características (X) e o rótulo (y)
X = dados.drop(colunas_drop, axis=1)  # Assume que a coluna
'falha'
y = dados['Status 1/0']


# Dividir os dados em conjuntos de treinamento e teste
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

# Criar e treinar o modelo
modelo = KNeighborsClassifier(n_neighbors=3)
modelo.fit(X_train, y_train)

# Fazer previsões
previsoes = modelo.predict(X_test)

# Avaliar o desempenho do modelo
print(classification_report(y_test, previsoes))

# Probabilidades
probabilidades = modelo.predict_proba(X_test)

# Exibir as probabilidades
print(probabilidades)

print("f1_score of KNeighborsClassifier:
{:.3f}".format(f1_score(y_test, modelo.predict(X_test))))

from sklearn.metrics import roc_curve
fpr_rf, tpr_rf, thresholds_rf = roc_curve(y_test,
modelo.predict_proba(X_test)[:, 1])


from sklearn.metrics import roc_auc_score
rf_auc = roc_auc_score(y_test, modelo.predict_proba(X_test)[:, 1])
print("AUC for KNeighborsClassifier: {:.3f}".format(rf_auc))



# Converter as probabilidades em DataFrame
probabilidades = pd.DataFrame(probabilidades,
columns=modelo.classes_)
# Combinar as probabilidades com as demais variáveis de entrada
prediction_result = pd.concat([X_test.reset_index(drop=True),
probabilidades], axis=1)
```

```
# Combinar as probabilidades com as demais variáveis de entrada
prediction_fatorized = pd.concat([X_test.reset_index(drop=True),
probabilidades], axis=1)


#Reverter
previsoes_restored = pd.DataFrame()

colunas_or = prediction_result.columns

for coluna in colunas_or:

    if (coluna in colunas_factorized):
        valores_previstos =
np.take(list(mapeamentos[coluna].values()),
prediction_fatorized[coluna])
        previsoes_restored[coluna] = valores_previstos
    else:
        previsoes_restored[coluna] = prediction_fatorized[coluna]
```

Anexo L. Código Python algoritmo *Logistic Regression*

Neste anexo é apresentado o código Python desenvolvido para treinar o modelo com o algoritmo *Logistic Regression*.

```
import seaborn as sns
import pandas as pd
import numpy as np
from sklearn.calibration import CalibratedClassifierCV
from sklearn.metrics import precision_recall_curve, f1_score
from sklearn.linear_model import LinearRegression,
LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report,
mean_squared_error
from IPython.display import Image

# Carregar os dados da memória de erros do equipamento em um
DataFrame
from IPython.core.display_functions import display

print("Avaria 0/1 _ ML_dataset_error_logs_total.xlsx _
LogisticRegression _ só matrix confusao")
dados = pd.read_excel('C:\\Users\\Fernando
Salgado\\PycharmProjects\\ \\ML_dataset_error_logs_total.xlsx')
dados_or = dados.copy()


colunas_factorized =
['criticality','user_device','product','testresult']

colunas_drop =
['error_name','error_code','state','type','serialnumber','ipaddres
s','machine_operating_hours',
'partition_year_month','date_id','kpi_treatment_duration','msgXsta
te','hierarchyl5','Spare Part NR','Spare Part Description',
'Status
1/0','cpu_name','cpu_shortcut','ts','sessionid','monitor_c_pess_op
erating_hours','monitor_p_pess_operating_hours',
'cnt','icc_url','typ','flag_only_dev','hierarchyl1','hierarchyl2',
'hierarchyl3','hierarchyl4',
'ebm_c_pess_operating_hours','ebm_p_pess_operating_hours','hydraul
ic_c_pess_operating_hours','hydraulic_p_pess_operating_hours',
'psu_c_pess_operating_hours','bpm_c_pess_operating_hours','btm_c_p
ess_operating_hours','bvm_c_pess_operating_hours',
'bic_strokes_in_session','so_strokes_in_session','uf_strokes_in_se
ssion','msgXvalue','softwareversion','en_us_text',
```

```
'ts_5min_trunc','error_log_ts']

# Criar map para voltar aos valores originais

mapeamentos = {}
for coluna in colunas_factorized:
    factors, unique_indices = pd.factorize(dados_or[coluna])
    dados[coluna] = factors
    mapeamentos[coluna] = {i: value for i, value in
enumerate(dados_or[coluna].unique())}


# Separar as características (X) e o rótulo (y)
X = dados.drop(colunas_drop, axis=1)  # Assume que a coluna
'falha'
y = dados['Status 1/0']

# Dividir os dados em conjuntos de treinamento e teste
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

# Criar e treinar o modelo
modelo = LogisticRegression()
modelo.fit(X_train, y_train)

# Fazer previsões

previsoes = modelo.predict(X_test)

# Avaliar o desempenho do modelo
print(classification_report(y_test, previsoes))

# Probabilidades
probabilidades = modelo.predict_proba(X_test)

# Exibir as probabilidades
print(probabilidades)

print("f1_score of LogisticRegression:
{:.3f}".format(f1_score(y_test, modelo.predict(X_test))))

from sklearn.metrics import average_precision_score
ap_rf = average_precision_score(y_test,
modelo.predict_proba(X_test)[:, 1])
print("Average precision of LogisticRegression:
{:.3f}".format(ap_rf))

from sklearn.metrics import roc_auc_score
rf_auc = roc_auc_score(y_test, modelo.predict_proba(X_test)[:, 1])
print("AUC for LogisticRegression: {:.3f}".format(rf_auc))

# Converter as probabilidades em DataFrame
probabilidades = pd.DataFrame(probabilidades,
columns=modelo.classes_)
# Combinar as probabilidades com as demais variáveis de entrada
prediction_result = pd.concat([X_test.reset_index(drop=True),
```

```
probabilidades], axis=1)

# Combinar as probabilidades com as demais variáveis de entrada
prediction_fatorized = pd.concat([X_test.reset_index(drop=True),
probabilidades], axis=1)


#Reverter
previsoes_restored = pd.DataFrame()

colunas_or = prediction_result.columns

for coluna in colunas_or:

    if (coluna in colunas_factorized):
        valores_previstos =
np.take(list(mapeamentos[coluna].values()),
prediction_fatorized[coluna])
        previsoes_restored[coluna] = valores_previstos
    else:
        previsoes_restored[coluna] = prediction_fatorized[coluna]
```

Anexo M. Código Python algoritmo *Naive Bayes*

Neste anexo é apresentado o código Python desenvolvido para treinar o modelo com o algoritmo *Naive Bayes*.

```
import seaborn as sns
import pandas as pd
import numpy as np
from sklearn.calibration import CalibratedClassifierCV
from sklearn.metrics import precision_recall_curve, f1_score
from sklearn.naive_bayes import GaussianNB
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report,
mean_squared_error
from IPython.display import Image

# Carregar os dados da memória de erros do equipamento em um
DataFrame
from sklearn.tree import DecisionTreeClassifier
from IPython.core.display_functions import display

print("Avaria 0/1 _ ML_dataset_error_logs_total.xlsx _ NaiveBayes
_ só matrix confusao")
dados = pd.read_excel('C:\\Users\\Fernando
Salgado\\PycharmProjects\\ ML_dataset_error_logs_total.xlsx')

dados_or = dados.copy()


colunas_factorized =
['criticality','user_device','product','testresult']
colunas_drop =
['error_name','error_code','state','type','serialnumber','ipaddres
s','machine_operating_hours',
'partition_year_month','date_id','kpi_treatment_duration','msgXsta
te','hierarchyl5','Spare Part NR','Spare Part Description',
'Status
1/0','cpu_name','cpu_shortcut','ts','sessionid','monitor_c_pess_op
erating_hours','monitor_p_pess_operating_hours',
'cnt','icc_url','typ','flag_only_dev','hierarchyl1','hierarchyl2',
'hierarchyl3','hierarchyl4',
'ebm_c_pess_operating_hours','ebm_p_pess_operating_hours','hydraul
ic_c_pess_operating_hours','hydraulic_p_pess_operating_hours',
'psu_c_pess_operating_hours','bpm_c_pess_operating_hours','btm_c_p
ess_operating_hours','bvm_c_pess_operating_hours',
'bic_strokes_in_session','so_strokes_in_session','uf_strokes_in_se
ssion','msgXvalue','softwareversion','en_us_text',
'ts_5min_trunc','error_log_ts']

# Criar map para voltar aos valores originais
```

```
mapeamentos = {}
for coluna in colunas_factorized:
    factors, unique_indices = pd.factorize(dados_or[coluna])
    dados[coluna] = factors
    mapeamentos[coluna] = {i: value for i, value in
enumerate(dados_or[coluna].unique())}


# Separar as características (X) e o rótulo (y)
X = dados.drop(colunas_drop, axis=1)  # Assume que a coluna
'falha'
y = dados['Status 1/0']

# Dividir os dados em conjuntos de treinamento e teste
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

# Criar e treinar o modelo
modelo = GaussianNB()
modelo.fit(X_train, y_train)

# Fazer previsões
previsoes = modelo.predict(X_test)

# Avaliar o desempenho do modelo
print(classification_report(y_test, previsoes))

# Probabilidades
probabilidades = modelo.predict_proba(X_test)

# Exibir as probabilidades
print(probabilidades)

print("f1_score of NaiveBayes: {:.3f}".format(f1_score(y_test,
modelo.predict(X_test))))

from sklearn.metrics import roc_auc_score
rf_auc = roc_auc_score(y_test, modelo.predict_proba(X_test)[:, 1])
print("AUC for NaiveBayes: {:.3f}".format(rf_auc))

# Converter as probabilidades em DataFrame
probabilidades = pd.DataFrame(probabilidades,
columns=modelo.classes_)
# Combinar as probabilidades com as demais variáveis de entrada
prediction_result = pd.concat([X_test.reset_index(drop=True),
probabilidades], axis=1)


# Combinar as probabilidades com as demais variáveis de entrada
prediction_fatorized = pd.concat([X_test.reset_index(drop=True),
probabilidades], axis=1)


#Reverter
previsoes_restored = pd.DataFrame()
```

```
colunas_or = prediction_result.columns

for coluna in colunas_or:

    if (coluna in colunas_factorized):
        valores_previstos =
np.take(list(mapeamentos[coluna].values()),
prediction_fatorized[coluna])
        previsoes_restored[coluna] = valores_previstos
    else:
        previsoes_restored[coluna] = prediction_fatorized[coluna]
```

Anexo N. Código Python algoritmo SVV

Neste anexo é apresentado o código Python desenvolvido para treinar o modelo com o algoritmo *Support Vector Machines*.

```
import seaborn as sns
import mglearn
import pandas as pd
import numpy as np
from sklearn.calibration import CalibratedClassifierCV
from sklearn.metrics import precision_recall_curve, f1_score
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report,
mean_squared_error
from IPython.display import Image

# Carregar os dados da memória de erros do equipamento em um
DataFrame
from IPython.core.display_functions import display

dados = pd.read_excel('C:\\Users\\Fernando
Salgado\\PycharmProjects\\ML_dataset_error_logs_total.xlsx')

dados_or = dados.copy()


colunas_factorized =
['criticality','user_device','product','testresult']
colunas_drop =
['error_name','error_code','state','type','serialnumber','ipaddres
s','machine_operating_hours',

'partition_year_month','date_id','kpi_treatment_duration','msgXsta
te','hierarchyl5','Spare Part NR','Spare Part Description',
'Status
1/0','cpu_name','cpu_shortcut','ts','sessionid','monitor_c_pess_op
erating_hours','monitor_p_pess_operating_hours',
'cnt','icc_url','typ','flag_only_dev','hierarchyl1','hierarchyl2',
'hierarchyl3','hierarchyl4',
'ebm_c_pess_operating_hours','ebm_p_pess_operating_hours','hydraul
ic_c_pess_operating_hours','hydraulic_p_pess_operating_hours',
'psu_c_pess_operating_hours','bpm_c_pess_operating_hours','btm_c_p
ess_operating_hours','bvm_c_pess_operating_hours',
'bic_strokes_in_session','so_strokes_in_session','uf_strokes_in_se
ssion','msgXvalue','softwareversion','en_us_text',
'ts_5min_trunc','error_log_ts']

# Criar map para voltar aos valores originais
mapeamentos = {}
for coluna in colunas_factorized:
```

```
    factors, unique_indices = pd.factorize(dados_or[coluna])
    dados[coluna] = factors
    mapeamentos[coluna] = {i: value for i, value in
enumerate(dados_or[coluna].unique())}


# Separar as características (X) e o rótulo (y)
X = dados.drop(colunas_drop, axis=1)  # Assume que a coluna
'falha'
y = dados['Status 1/0']


# Dividir os dados em conjuntos de treinamento e teste
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

# Criar e treinar o modelo
modelo = SVC()
modelo.fit(X_train, y_train)

# Fazer previsões
previsoes = modelo.predict(X_test)

# Avaliar o desempenho do modelo
print(classification_report(y_test, previsoes))

print("f1_score of SVC: {:.3f}".format(f1_score(y_test,
modelo.predict(X_test))))

from sklearn.metrics import average_precision_score
ap_rf = average_precision_score(y_test,
modelo.predict_proba(X_test)[:, 1])
print("Average precision of SVC: {:.3f}".format(ap_rf))

from sklearn.metrics import roc_auc_score
rf_auc = roc_auc_score(y_test, modelo.predict_proba(X_test)[:, 1])
print("AUC for SVC: {:.3f}".format(rf_auc))

# Converter as probabilidades em DataFrame
probabilidades = pd.DataFrame(probabilidades,
columns=modelo.classes_)
prediction_result = pd.concat([X_test.reset_index(drop=True),
probabilidades], axis=1)


# Combinar as probabilidades com as demais variáveis de entrada
prediction_fatorized = pd.concat([X_test.reset_index(drop=True),
probabilidades], axis=1)

#Reverter
previsoes_restored = pd.DataFrame()

colunas_or = prediction_result.columns

for coluna in colunas_or:
```

```
    if (coluna in colunas_factorized):
        valores_previstos =
np.take(list(mapeamentos[coluna].values()),
prediction_fatorized[coluna])
        previsoes_restored[coluna] = valores_previstos
    else:
        previsoes_restored[coluna] = prediction_fatorized[coluna]
```

Anexo O. Código Python algoritmo *Neural Networks*

Neste anexo é apresentado o código Python desenvolvido para treinar o modelo com o algoritmo *Neural Networks*.

```
import seaborn as sns
import pandas as pd
import numpy as np
from sklearn.calibration import CalibratedClassifierCV
from sklearn.metrics import precision_recall_curve, f1_score
from sklearn.neural_network import MLPClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report,
mean_squared_error
from IPython.display import Image

# Carregar os dados da memória de erros do equipamento em um
DataFrame
from IPython.core.display_functions import display

dados = pd.read_excel('C:\\Users\\Fernando
Salgado\\PycharmProjects\\ ML_dataset_error_logs_total.xlsx')

dados_or = dados.copy()


colunas_factorized =
['criticality','user_device','product','testresult']
colunas_drop =
['error_name','error_code','state','type','serialnumber','ipaddres
s','machine_operating_hours',
'partition_year_month','date_id','kpi_treatment_duration','msgXsta
te','hierarchyl5','Spare Part NR','Spare Part Description',
'Status
1/0','cpu_name','cpu_shortcut','ts','sessionid','monitor_c_pess_op
erating_hours','monitor_p_pess_operating_hours',
'cnt','icc_url','typ','flag_only_dev','hierarchyl1','hierarchyl2',
'hierarchyl3','hierarchyl4',
'ebm_c_pess_operating_hours','ebm_p_pess_operating_hours','hydraul
ic_c_pess_operating_hours','hydraulic_p_pess_operating_hours',
'psu_c_pess_operating_hours','bpm_c_pess_operating_hours','btm_c_p
ess_operating_hours','bvm_c_pess_operating_hours',
'bic_strokes_in_session','so_strokes_in_session','uf_strokes_in_se
ssion','msgXvalue','softwareversion','en_us_text',
'ts_5min_trunc','error_log_ts']


# Criar map para voltar aos valores originais
mapeamentos = {}
for coluna in colunas_factorized:
    factors, unique_indices = pd.factorize(dados_or[coluna])
```

```
    dados[coluna] = factors
    mapeamentos[coluna] = {i: value for i, value in
enumerate(dados_or[coluna].unique())}


# Separar as características (X) e o rótulo (y)
X = dados.drop(colunas_drop, axis=1)  # Assume que a coluna
'falha' y = dados['Status 1/0']

# Dividir os dados em conjuntos de treinamento e teste
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=0)

# Criar e treinar o modelo
modelo = MLPClassifier(random_state=42)
modelo.fit(X_train, y_train)

# Fazer previsões
previsoes = modelo.predict(X_test)

# Avaliar o desempenho do modelo
print(classification_report(y_test, previsoes))

# Probabilidades
probabilidades = modelo.predict_proba(X_test)

# Exibir as probabilidades
print(probabilidades)

print("f1_score of MLPClassifier: {:.3f}".format(f1_score(y_test,
modelo.predict(X_test))))

from sklearn.metrics import average_precision_score
ap_rf = average_precision_score(y_test,
modelo.predict_proba(X_test)[:, 1])
print("Average precision of MLPClassifier: {:.3f}".format(ap_rf))

from sklearn.metrics import roc_auc_score
rf_auc = roc_auc_score(y_test, modelo.predict_proba(X_test)[:, 1])
print("AUC for MLPClassifier: {:.3f}".format(rf_auc))

# Converter as probabilidades em DataFrame
probabilidades = pd.DataFrame(probabilidades,
columns=modelo.classes_)
# Combinar as probabilidades com as demais variáveis de entrada
prediction_result = pd.concat([X_test.reset_index(drop=True),
probabilidades], axis=1)

# Combinar as probabilidades com as demais variáveis de entrada
prediction_fatorized = pd.concat([X_test.reset_index(drop=True),
probabilidades], axis=1)


#Reverter
previsoes_restored = pd.DataFrame()
```

```
colunas_or = prediction_result.columns

for coluna in colunas_or:

    if (coluna in colunas_factorized):
        valores_previstos =
np.take(list(mapeamentos[coluna].values()),
prediction_fatorized[coluna])
        previsoes_restored[coluna] = valores_previstos
    else:
        previsoes_restored[coluna] = prediction_fatorized[coluna]
```

Anexo P. Código Python carregar modelo previsão de falha de equipamentos e componentes

Neste anexo é apresentado o código Python para carregar no Power BI o modelo de previsão de falha de equipamentos e componentes.

```
import numpy as np
import pandas as pd
import joblib


modelo_01 = joblib.load('C:\\Users\\Fernando Salgado\\GBDT.pkl')
modelo_peca = joblib.load('C:\\Users\\Fernando
Salgado\\modelo_pecas.pkl')

# Carregar os dados para fazer previsões
dados_01 = pd.read_excel('C:\\Users\\Fernando
Salgado\\vw_fact_cf_errors_with_features_delta.xlsx')
dados_peca = pd.read_excel('C:\\Users\\Fernando
Salgado\\vw_fact_cf_errors_with_features_delta.xlsx')
dados_or_01 = dados_01.copy()
dados_or_peca = dados_peca.copy()


colunas_factorized = []
colunas_drop_01 =
['error_name','error_code','state','type','serialnumber','ipaddres
s','machine_operating_hours',
'partition_year_month','date_id','kpi_treatment_duration','msgXsta
te','hierarchyl5',
'cpu_name','cpu_shortcut','ts','sessionid','monitor_c_pess_operati
ng_hours','monitor_p_pess_operating_hours',
'cnt','icc_url','typ','flag_only_dev','hierarchyl1','hierarchyl2',
'hierarchyl3','hierarchyl4',
'ebm_c_pess_operating_hours','ebm_p_pess_operating_hours','hydraul
ic_c_pess_operating_hours','hydraulic_p_pess_operating_hours',
'psu_c_pess_operating_hours','bpm_c_pess_operating_hours','btm_c_p
ess_operating_hours','bvm_c_pess_operating_hours',
'bic_strokes_in_session','so_strokes_in_session','uf_strokes_in_se
ssion','msgXvalue','softwareversion','en_us_text',
'ts_5min_trunc','error_log_ts']


colunas_drop_peca =
['error_name','error_code','state','type','serialnumber','ipaddres
s','machine_operating_hours',

'partition_year_month','date_id','kpi_treatment_duration','msgXsta
te','hierarchyl5',
```

```
'cpu_name','cpu_shortcut','ts','sessionid','monitor_c_pess_operati
ng_hours','monitor_p_pess_operating_hours',
'cnt','icc_url','typ','flag_only_dev','hierarchyl1','hierarchyl2',
'hierarchyl3','hierarchyl4',
'ebm_c_pess_operating_hours','ebm_p_pess_operating_hours','hydraul
ic_c_pess_operating_hours','hydraulic_p_pess_operating_hours',
'psu_c_pess_operating_hours','bpm_c_pess_operating_hours','btm_c_p
ess_operating_hours','bvm_c_pess_operating_hours',
'bic_strokes_in_session','so_strokes_in_session','uf_strokes_in_se
ssion','msgXvalue','softwareversion','en_us_text',
'ts_5min_trunc','error_log_ts']

dados_01 = dados_01.drop(colunas_drop_01, axis=1)  # Assume que a
coluna 'falha' indica a ocorrência
dados_peca = dados_peca.drop(colunas_drop_peca, axis=1)  # Assume
que a coluna 'falha' indica a ocorrência de falha

# Fazer previsões com o modelo
previsoes_01 = modelo_01.predict_proba(dados_01)
dados_01['Previsao'] = previsoes_01[:,1]

#previsoes = modelo.predict_proba(dados)
previsoes_peca = modelo_peca.predict(dados_peca)
dados_peca['SparePart'] = previsoes_peca


#Reverter
saida_excel = pd.DataFrame()

colunas_or = dados_or_01.columns
#print(colunas_or.toString())

for coluna in colunas_or:
    #print("col: " + coluna)
    if (coluna in colunas_factorized):
        valores_previstos =
np.take(list(mapeamentos[coluna].values()), dados_01[coluna])
        saida_excel[coluna] = valores_previstos
    else:
        saida_excel[coluna] = dados_or_01[coluna]

saida_excel['Previsao'] = previsoes_01[:,1]
saida_excel['SparePart'] = previsoes_peca#[:,1]

print("preparar excel")
df = pd.DataFrame(saida_excel)
print("Pronto para gravar em excel")
caminho_arquivo = 'C:\\Users\\Fernando
Salgado\\dataset_saida_error_codes_spares.xlsx'

df.to_excel(caminho_arquivo, index=False)
```

Anexo Q. Diagrama de relações tabelas no Power BI

Neste anexo são apresentadas o diagrama de relações tabelas no Power BI.